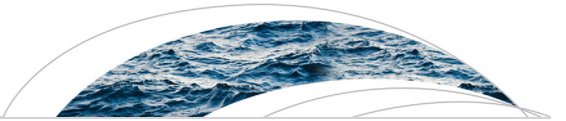
AGU PUBLICATIONS

Water Resources Research



RESEARCH ARTICLE

10.1002/2014WR016794

Basin-scale runoff prediction: An Ensemble Kalman Filter framework based on global hydrometeorological data sets

Christof Lorenz[1], Mohammad J. Tourian[2], Balaji Devaraju[3], Nico Sneeuw[2], and Harald Kunstmann[1,4]

[1]Karlsruhe Institute of Technology, Institute of Meteorology and Climate Research (IMK-IFU), Garmisch-Partenkirchen, Germany, [2]Institute of Geodesy, University of Stuttgart, Stuttgart, Germany, [3]Institute of Geodesy, University of Hannover, Hannover, Germany, [4]Institute of Geography, University of Augsburg, Augsburg, Germany

Key Points:
- Data-assimilation framework based on publicly available data sets
- Estimates of basin-scale runoff through a least-squares prediction
- Validation of runoff predictions with observations over 16 large river basins

Correspondence to:
C. Lorenz,
christof.lorenz@kit.edu

Citation:
Lorenz, C., M. J. Tourian, B. Devaraju, N. Sneeuw, and H. Kunstmann (2015), Basin-scale runoff prediction: An Ensemble Kalman Filter framework based on global hydrometeorological data sets, *Water Resour. Res.*, *51*, 8450–8475, doi:10.1002/2014WR016794.

Received 10 DEC 2014
Accepted 23 SEP 2015
Accepted article online 1 OCT 2015
Published online 28 OCT 2015

Abstract In order to cope with the steady decline of the number of in situ gauges worldwide, there is a growing need for alternative methods to estimate runoff. We present an Ensemble Kalman Filter based approach that allows us to conclude on runoff for poorly or irregularly gauged basins. The approach focuses on the application of publicly available global hydrometeorological data sets for precipitation (GPCC, GPCP, CRU, UDEL), evapotranspiration (MODIS, FLUXNET, GLEAM, ERA interim, GLDAS), and water storage changes (GRACE, WGHM, GLDAS, MERRA LAND). Furthermore, runoff data from the GRDC and satellite altimetry derived estimates are used. We follow a least squares prediction that exploits the joint temporal and spatial auto- and cross-covariance structures of precipitation, evapotranspiration, water storage changes and runoff. We further consider time-dependent uncertainty estimates derived from all data sets. Our in-depth analysis comprises of 29 large river basins of different climate regions, with which runoff is predicted for a subset of 16 basins. Six configurations are analyzed: the Ensemble Kalman Filter (Smoother) and the hard (soft) Constrained Ensemble Kalman Filter (Smoother). Comparing the predictions to observed monthly runoff shows correlations larger than 0.5, percentage biases lower than $\pm$ 20%, and NSE-values larger than 0.5. A modified NSE-metric, stressing the difference to the mean annual cycle, shows an improvement of runoff predictions for 14 of the 16 basins. The proposed method is able to provide runoff estimates for nearly 100 poorly gauged basins covering an area of more than 11,500,000 km^2 with a freshwater discharge, in volume, of more than 125,000 m^3/s.

1. Introduction

The percentage of geographically and temporally accessible runoff is of crucial importance for sufficient freshwater supply. Early estimates on the human appropriation of accessible runoff have been presented in *Postel et al*. [1996], and they raised awareness on the problem of declining water availability per capita, primarily due to a growing world population. The increase in the population of 45% stands opposed to an increase of accessible runoff of only 10% within 30 years. *Falkenmark and Rockström* [2004] estimated an increase of water requirements by about 1,300 m^2 per capita per yr for each additional person. *Vörösmarty et al*. [2010] stress that 80% of the world's population is exposed to high levels of threat to water security and that 65% of the total continental discharge is classified as moderately to highly threatened. Accordingly, recent hydrological and hydrometeorological research must aim at an improved observation, modeling, and understanding of the terrestrial water cycle, in particular runoff.

The importance of long-term monitoring of hydrological variables was recognized by the World Meteorological Organization (WMO) already in 1980, which led to the initiation of the Global Runoff Data Centre (GRDC). The GRDC collects and harmonizes global runoff observations from national hydrological services and makes them available to the public. However, over many catchments around the world, runoff is not gauged [*Blöschl et al*., 2013].

In *Dai and Trenberth* [2002], it is reported that the annual runoff rate over the unmonitored areas is comparable to that over the monitored areas. Furthermore, studies by *Sivapalan et al*. [2003], *Shiklomanov et al*. [2002], *Milzow et al*. [2011], and *Fekete et al*. [2012] stress that the number of active river gauges is steadily decreasing. Figure 1a exemplarily shows river basins without any measurements and basins with more than 5 years of missing data during the period 2000–2010. Since a similar decline can be observed for rainfall

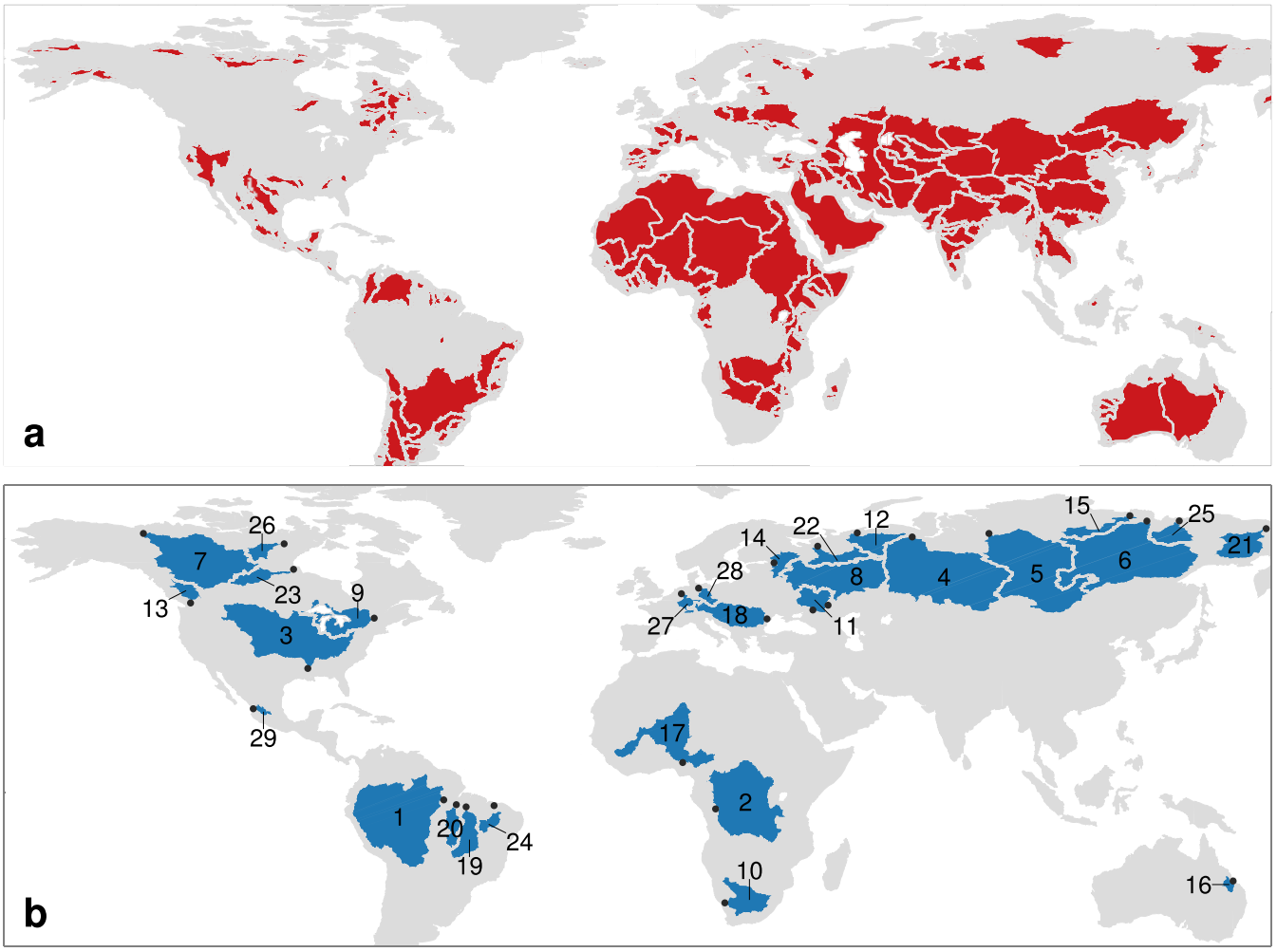


Figure 1. (a) Distribution of ungauged and poorly gauged catchments. The dark red areas indicate catchments without any measurements between 2002 and 2012 while the observed runoff time-series over the light red basins contain more than 5 years of missing values during that period. The white catchments can be considered to be dischargeless (e.g., desserts). (b) Location of the 29 basins given in Table 2 where the dot in each basin is located at the respective river gauge. The 16 dark blue shaded basins are used for validating the runoff predictions in section 4.

gauges [*Lorenz and Kunstmann*, 2012], the hydrological community has to face a significantly shrinking database for the most important water cycle variables.

The serious problem of runoff data availability and its importance for hydrological cycle studies led the International Association of Hydrological Sciences (IAHS) to launch the study framework Prediction in Ungauged Basins (PUB). The goal of 10 years of PUB was to predict the hydrological cycle in ungauged basins by improving the understanding of climatic and landscape controls on hydrological processes [*Blöschl et al.*, 2013]. It has been realized that on top of data quantity data quality also tremendously affects the performance of modeling [*McMillan et al.*, 2010].

With respect to river runoff, one of the motives of PUB was to seek for spaceborne alternatives, as traditional runoff data acquisition at in situ gauges is typically costly and time consuming. These alternatives indeed have the potential to be highly valuable for hydrology [*Alsdorf et al.*, 2007]. In terms of runoff, it has been already demonstrated by *Koblinsky et al.* [1993] that satellite-based altimetry can be used to describe river level variations. Therefore, many studies made use of satellite data in order to analyze the water cycle over ungauged basins [e.g., *Khan et al.*, 2011]. However, *Alsdorf et al.* [2007] already indicated high uncertainties in satellite derived hydrological data due to poor spatial and temporal resolutions and the inabilities to e.g., penetrate clouds or smoke. In terms of precipitation, e.g., *AghaKouchak et al.* [2012] analyzed

different systematic and random errors in satellite-derived precipitation data sets over the United States. It was further stressed by *Fekete et al*. [2012] that there is still no substitute for in situ discharge monitoring. Thus, the combination and harmonization of spaceborne and terrestrial data is a big challenge of the current hydrological research.

Another major task of the hydrological community is the closure of the terrestrial water budget. Even if the basic equation of the terrestrial water cycle

$$P-ET-R-\dot{M}=0 \tag{1}$$

with precipitation P, evapotranspiration ET, runoff R, and water storage changes $\dot{M}$ appears simple, it is known from many studies that a real closure on larger spatial scales can not be achieved with current data sources.

For a long time, a major issue of analyzing the hydrological cycle was missing observation-based data of the terrestrial water storage. In essence, direct observation of the water storage components (e.g., snow, ice, surface water, soil, (un)saturated underground storages) was not possible on large spatial scales [*Rodell and Famiglietti*, 1999]. Measurements of water storage changes by measurements of changes in groundwater levels and in soil water saturation are not reliable due to the insufficiently known storage coefficients and also the inadequate density of monitoring points on large spatial scales [*Riegger et al*., 2012].

Only after the launch of the twin satellites of the Gravity Recovery and Climate Experiment GRACE, closing the water budget on shorter time scales became possible [*Tapley et al*., 2004]. *Sheffield et al*. [2009] combined satellite-based precipitation, model-based evapotranspiration, and GRACE data to close the water budget over the Mississippi River basin over the time period of 2003–2005. *Rodell et al*. [2004a] closed the water budget using GRACE at basin scales to estimate evapotranspiration as the residual of the water budget. In *Syed et al*. [2005], [2007], and [2009], GRACE data have been used for deriving freshwater discharge using a coupled land-atmosphere water balance.

In a similar study, *Sneeuw et al*. [2014] compared the performance of estimated runoff derived from water balance closure with those from models and in situ data. In a more extensive analysis, *Lorenz et al*. [2014] compared combinations of different gridded observations, atmospheric reanalysis models, land surface hydrological models, partially model-based data sets and GRACE to derive and evaluate an ensemble of hydrological and hydrometeorological budget-based runoff estimates. However, most of these studies did not close the water budget explicitly, but estimated a single water cycle variable from the remaining ones. Thus, all the errors and uncertainties in the input data sources inevitably propagate to the estimated variable.

Other approaches, which make use of data assimilation techniques, are presented in e.g., *Pan and Wood* [2006]. The authors developed a Constrained Ensemble Kalman Filter (CEnKF), which enforced water budget closure, thereby, providing a constrained best estimate of the basin-scale water budget. Similarly, *Pan et al*. [2012] combined estimates of the terrestrial water budget from different data sources and enforced the water balance constraint using the previously developed constrained Kalman filter technique.

In general, the term data assimilation is used for combining observation data with hydrological, hydrometeorological, or land surface models. The application usually ranges from extensive global reanalysis models, which simulate the whole atmosphere and the land surface [e.g., *Dee et al*., 2011], to models which focus on the land surface only [e.g., *Reichle et al*., 2002; *Rodell et al*., 2014a]. However, it is well known that our current model systems suffer from biases or errors in the simulated water cycle variables, which finally lead to imbalances in the water budgets [e.g., *Lorenz and Kunstmann*, 2012]. We therefore propose a data assimilation framework, where the mathematical description is kept as simple as possible. In fact, the approach is based on the terrestrial water budget equation (1) only, which can be considered as the most simple hydrological model. On the other hand, we want to exploit all the advantages and strengths from widely used global available data sets for the major water cycle variables, and thus, use real data instead of complex model equations within an Ensemble Kalman Filtering framework. The algorithm is, therefore, proposed as an alternative postprocessing tool, which can be used for combining, correcting, and predicting basin-scale time series of the four major water cycle variables. In this study, however, we focus on the estimation of runoff as an performance evaluation step of the proposed approaches.

The data sets that are used are presented in section 2, while we describe the methods in section 3. In section 4, different configurations of the assimilation framework are analyzed and the best setup is used for

Table 1. Summary of the Observation and Model-Based Data Sets Containing Precipitation (P), Actual Evapotranspiration (ET), Water Storage Changes ($\dot{M}$), Runoff From River Gauges (R_{obs}), and Satellite Altimetry (R_{alt})

			Resolution		
Variable	Data Set	Version	Spatial	Temporal	Time Period
P	GPCC	6.0	0.5 °× 0.5°	1 month	1901–2010
	GPCP	2.2	2.5° × 2.5°	1 month	1979–present
	CRU	3.22	0.5° × 0.5°	1 month	1901–2013
	DEL	3.02	0.5° × 0.5°	1 month	1900–2010
ET	ERA interim		0.75° × 0.75°	1 month, 1 day, 6 h	1979–present
	GLDAS	NOAH 3.3	1.0° × 1.0°	1 month, 3 h	1948–present
	GLEAM	v1B	0.25° × 0.25°	1 day	1984–2007
	MOD16	A2	0.5° × 0.5°	1 year, 1 month, 8 days	2000–2013
	FLUXNET MTE		0.5° × 0.5°	1 month	1980–present
$\dot{M}$	GRACE GFZ	R5		1 month	2002–present
	GRACE CSR	R5		1 month	2002–present
	MERRA LAND	1.0	$\frac{1}{2}^\circ \times \frac{2}{3}^\circ$	1 month, 1 day, 1 h	1980–present
	GLDAS	NOAH 3.3	1.0° × 1.0°	1 month, 3 h	1948–present
	WGHM	NOUSE	0.5° × 0.5°	1 month	1960–2009
R_{obs}	GRDC				
R_{alt}	*Tourian et al*. [2013]				

estimating runoff over 16 large river basins. These estimates are then validated against monthly runoff observations during the period 2005-2010 using the common performance metrics correlation, relative bias (PBIAS), and the Nash-Sutcliffe Efficiency (NSE) coefficient. However, as runoff over many catchments follows a dominant annual cycle, we further analyze a NSE-type metric, which relates the estimates to the mean annual cycle from historic data. In section 5, the performance of the presented approach is compared with similar studies, while section 6 contains a conclusion of the major findings and an outlook.

2. Data

A detailed description of the applied data sets is given in *Lorenz et al*. [2014] and *Sneeuw et al*. [2014]. Here only a brief overview of the different data sets of the four water cycle variables is provided. The most important information is further summarized in Table 1.

2.1. Runoff

In order to attain maximum data coverage, both in space and time, data are collected from different sources, namely the Global Runoff Data Center (GRDC), the ArcticRIMS project Water Survey Canada, the United States Geological Survey, United States Army Corps of Engineers (USACE), the ORE HYBAM project, the Department of Water, Land and Biodiversity Conservation, Government of Australia, and the Department of Water Affairs and Forestry, Republic of South Africa.

The runoff database is further enhanced by using estimates from satellite altimetry over the Amazon, Ob, Don, and Danube basins. The methods for deriving runoff from altimetry data are described in *Tourian et al*. [2013]. They proposed a statistical approach based on quantile functions to infer a functional relation between altimetric water level and historic river gauge data. Using such a statistical function, the water level measurements from satellite altimetry is then mapped to an estimation of runoff at river gauges.

2.2. Precipitation

Four different observation based precipitation data sets, namely the data from the Global Precipitation Climatology Center (GPCC) [*Schneider et al*., 2008], the Global Precipitation Climatology Project (GPCC) [*Adler et al*., 2003], the Climatic Research Unit (CRU) [*Harris et al*., 2014], and the University of Delaware (DEL) [*Matsuura and Willmott*, 2012], are used. From these data sets, GPCC, CRU, and DEL are based on gauge observations only. The low spatial resolution GPCP product is a combination of different rainfall sensors. Currently, the GPCP includes microwave, infrared, and gauge based observations of precipitation [e.g., *Huffman et al*., 2009]. The largest number of stations are included in the GPCC product, while CRU and DEL are based on a much smaller number of rainfall gauges. Despite the pure number of stations, also the spatial coverage is significantly different between these three data sets. An analysis of the different precipitation products is presented in e.g., *Lorenz and Kunstmann* [2012].

Table 2. River Basins Which Were Used in This Study Including the Respective River Gauge, its Location, the Area of the Basin, and the Mean Annual Discharge Between the Period 1980–2010 as Provided From GRDC[a]

	River	Station	Latitude	Longitude	Area (km^2)	Discharge (m^3/s)
1	Amazon	Obidos	−1.95	−55.51	46,72,876	1,71,915
2	Congo	Kinshasa	−4.3	15.3	36,15,546	38,652
3	Mississippi	Vicksburg	32.32	−90.91	29,38,538	18,661
4	Ob	Salekhard	66.57	66.53	29,26,321	12,939
5	Yenisei	Igarka	67.48	86.5	24,54,961	19,388
6	Lena	Kyusyr	70.7	127.65	24,17,932	17,761
7	Mackenzie	Arctic Red River	67.46	−133.75	16,66,073	9194
8	Volga	Volgograd Power Plant	48.81	44.59	13,45,070	8143
9	St. Lawrence	Lasalle	45.42	−73.62	9,43,769	8575
10	Orange	Vioolsdrif	−28.76	17.72	8,28,475	169
11	Don	Razdorskaya	47.54	40.65	3,78,180	685
12	Pechora	Oksino	67.6	52.2	3,04,670	4857
13	Fraser	Hope	49.38	−121.45	2,28,874	2657
14	Neva	Novosaratovka	59.84	30.53	2,25,651	2490
15	Olenek	7.5 km d/s of mouth of Pur	72.12	123.22	1,99,723	1257
16	Fitzroy (west Australia)	The Gap	−23.09	150.11	1,26,986	138
17	Niger	Lokoja	7.8	6.76	21,00,508	5070
18	Danube	Ceatal Izmail	45.22	28.72	7,71,277	6490
19	Tocantins	Tucurui	−3.76	−49.65	7,52,993	10966
20	Rio Xingu	Altamira	−3.22	−52.21	4,45,289	8062
21	Kolyma	Kolymskaya	68.73	158.72	4,21,802	3262
22	Severnaya Dvina (northern)	Ust-Pinega	64.15	41.92	3,30,709	3392
23	Churchill	Above Red Head Rapids	58.12	−94.63	2,99,391	337
24	Rio Parnaiba	Luzilandia	−3.45	−42.37	2,97,049	692
25	Yana	Ubileynaya	70.75	136.08	2,20,949	1127
26	Thelon	below outlet of Schultz Lake	64.78	−97.05	1,71,346	924
27	Rhine	Rees	51.75	6.4	1,69,706	2392
28	Elbe	Neu Darchau	53.23	10.89	1,34,037	701
29	Rio Santiago	El Capomal	21.83	−1.05	1,26,986	161

[a]The location of the basins is shown in Figure 1b. From the 29 catchments, the first 16 are used for validating the runoff predictions.

2.3. Evapotranspiration

In contrast to the precipitation data (which are all based on similar observations), evapotranspiration data are derived from quite different modeling and data merging approaches. A comprehensive overview over different global evapotranspiration data sets can be found in *Mueller et al.* [2011] and *Mueller et al.* [2013]. Here we use three evapotranspiration-only products, namely the Global Land-surface Evaporation: the Amsterdam Methodology (GLEAM) [*Miralles et al.*, 2011], the Fluxnet Multi Tree Ensemble (MTE) [*Jung et al.*, 2009, 2010], and data from the Moderate-resolution Imaging Spectroradiometer (MOD16) [*Mu et al.*, 2007, 2011].

Besides these products, we also use modeled evapotranspiration from version 2 of the Global Land Data Assimilation System (GLDAS2) [*Rodell et al.*, 2004b] (i.e., the NOAH3.3 realization) and the ERA interim Reanalysis from the European Centre for Medium-Range Weather Forecasts (ERA I) [*Simmons et al.*, 2006; *Berrisford et al.*, 2009; *Dee et al.*, 2011].

GLEAM is based on multisatellite observations and combines a wide range of remotely sensed observations within a Priestley and Taylor-based framework [*Miralles et al.*, 2011]. In contrast to this multisatellite-approach, MOD16 is based on remote sensing data from the MODIS satellites and global meteorological data. The data sets are combined using the Penman-Monteith method [*Mu et al.*, 2007]. Finally, the MTE data set provides global evapotranspiration by empirical upscaling of eddy-covariance measurements from the FLUXNET network [*Jung et al.*, 2009].

2.4. Water Storage Changes

2.4.1. GRACE Data

In this study, GRACE observations of the temporal gravity field changes are applied for deriving basin-scale water storage changes. In particular, 10 years (2003–2012) of GRACE release 5 data from the two data processing centers GeoForschungsZentrum Potsdam (GFZ) [*Dahle et al.*, 2013] and the Center for Space Research, The University of Texas at Austin (CSR) [*Bettadpur*, 2012] are used. The spherical harmonic coefficients are filtered with a regularization type filter, as described in *Sneeuw et al.* [2014]. Then, the spectral

data are transformed to a $0.5^\circ \times 0.5^\circ$ grid using the equations from *Wahr et al*. [1998] and then aggregated over catchments to derive catchment-specific time series. As these time series represent water storage *anomalies*, the data have to be numerically differentiated in order to arrive at the required flux quantity, water storage *changes*. This is done by applying the method of three-point central differences. Further details on the preprocessing of GRACE-data can be found in *Sneeuw et al*. [2014] and *Lorenz et al*. [2014].

2.4.2. Land Surface Models

As GRACE-data are only available from 2002 on, we also need some legacy data for deriving the statistical inter and intra-catchment relationships between water storage changes and the other water cycle variables (see section 3.2). Therefore, we use data from GLDAS2. Additionally, we also apply data from the land-only replay of the Modern-Era Retrospective analysis for Research and Applications (MERRA LAND) [*Reichle et al*., 2011] and the WaterGAP Global Hydrology Model (WGHM) [*Döll et al*., 2003, 2014]. In contrast to the two land surface models MERRA LAND and GLDAS2, which do not contain a routing scheme (R. Reichle, 2013, personal communication, M. Rodell, 2012, personal communication), WGHM is tuned against observed discharge at 724 gauging stations [*Döll et al*., 2003]. It is assumed that this improves the reliability of the water availability estimates from the model. Data from WGHM have been used for various water budget studies [e.g., *Ramillien et al*., 2006; *Schmidt et al*., 2006; *Kusche et al*., 2009; *Forootan et al*., 2012; *Döll et al*., 2014].

2.5. Data Consistency

For the comparison of the runoff estimates with observations, we have remapped all the input fields to the same grid resolution ($0.5^\circ \times 0.5^\circ$) using a first-order conservative interpolation [*Jones*, 1999]. From these fields, area-weighted averages were computed over the study regions shown in Figure 1 and Table 2.

The water storage changes $\dot{M}$ computed from GRACE via the central difference scheme are an approximation of the *true* derivatives. In the spectral domain they differ by a *sinc* function. In fact, the numerical derivatives from the central difference scheme are not compatible yet to the hydrological signals (P, ET and R in equation (1)). The hydrological water storage changes should hence be filtered using a filter that resembles the *sinc* function in the frequency domain. *Swenson and Wahr* [2006] proposed to smooth the time series of the other water cycle variables according to

$$\tilde{F}_t = \frac{1}{4}F_{t-1} + \frac{1}{2}F_t + \frac{1}{4}F_{t+1}, \tag{2}$$

where F are time series of precipitation, evapotranspiration, and observed runoff. The indices $t-1$, t, and $t+1$ refer to the previous, current, and following month. We apply this filter to all the time series of precipitation, runoff and evapotranspiration.

3. Methods

3.1. Overview

For estimating basin-scale water cycle variables, different formulations of an Ensemble Kalman Filter (EnKF) [*Evensen*, 1994; *Houtekamer and Mitchell*, 1998; *Evensen*, 2003,] framework are applied, which successively predict and correct the state vector containing precipitation, evapotranspiration, runoff, and water storage changes

$$\mathbf{X}_t = \left[\mathbf{P}_t^\mathsf{T} \quad \mathbf{ET}_t^\mathsf{T} \quad \mathbf{R}_t^\mathsf{T} \quad \dot{\mathbf{M}}_t^\mathsf{T}\right]^\mathsf{T} \tag{3}$$

with t being the time-index, $\mathbf{P}_t$, $\mathbf{ET}_t$, $\mathbf{R}_t$, and $\dot{\mathbf{M}}_t$ four subvectors with dimensions $[N \times 1]$ and N the number of catchments. A summary of the different input parameter and formulas for the Ensemble Kalman Filter (EnKF), the hard (CEnKF_h) and soft (CEnKF_s) constrained Ensemble Kalman Filter, the Ensemble Kalman Smoother (EnKS), and the hard (CEnKS_h) and soft (CEnKS_s) constrained Ensemble Kalman Smoother are presented in Tables 3 and 4, respectively.

It can be construed that all the filter equations are linear. With the assumption that the errors of the variables are normally distributed, a single-state Kalman filter (SSKF) is sufficient to achieve the same results. Instead we present an ensemble Kalman filter (EnKF) approach, which will allow for its extension to assimilate on the global scale. Predicting and estimating grid-point values globally results in huge (spatial) covariance matrices,

Table 3. Overview of the Different Parameters and Variables for the Assimilation Framework

Parameter	Equation	Dimension
State vector	$\mathbf{X}_t = \left[\mathbf{P}_t^{\mathrm{T}} \quad \mathbf{ET}_t^{\mathrm{T}} \quad \mathbf{R}_t^{\mathrm{T}} \quad \dot{\mathbf{M}}_t^{\mathrm{T}}\right]^{\mathrm{T}}$	$4N\times1$
Annual cycle	$\tilde{\mathbf{X}}_m = \frac{1}{T/12}\sum_{i,m}^{T/12}\mathbf{X}_{i,m}$	$4N\times1$
Residuals	$\mathbf{r}_t = \mathbf{X}_t - \tilde{\mathbf{X}}_t$	$4N\times1$
Auto-covariance	$\mathbf{\Sigma} = \frac{1}{T-1}\sum_{t=1}^{T}\mathbf{r}_t\mathbf{r}_t^{\mathrm{T}}$	$4N\times4N$
Cross covariance	$\mathbf{\Sigma}_\Delta = \frac{1}{T-2}\sum_{t=2}^{T}\mathbf{r}_t\mathbf{r}_{t-1}^{\mathrm{T}}$	$4N\times4N$
Prediction matrix	$\hat{\mathbf{A}} = \mathbf{\Sigma}_\Delta\ \mathbf{\Sigma}^{-1}$	$4N\times4N$
Prediction noise	$\mathbf{Q}_{\mathrm{P}} = \mathbf{\Sigma} - \mathbf{\Sigma}_\Delta\ \mathbf{\Sigma}^{-1}\ \mathbf{\Sigma}_\Delta^{\mathrm{T}}$	$4N\times4N$
Control input	$\mathbf{B} = \left[-\hat{\mathbf{A}} \quad \mathbf{I}\right]$	$4N\times8N$
	$\mathbf{U}_t = \left[\tilde{\mathbf{X}}_{t-1} \quad \tilde{\mathbf{X}}_t\right]^{\mathrm{T}}$	$8N\times1$
Observation vector	$\mathbf{Y}_t = \left[\mathbf{P}_{t,\mathrm{obs}}^{\mathrm{T}} \quad \mathbf{ET}_{t,\mathrm{obs}}^{\mathrm{T}} \quad \mathbf{R}_{t,\mathrm{obs}}^{\mathrm{T}} \quad \dot{\mathbf{M}}_{t,\mathbf{obs}}^{\mathrm{T}} \quad \mathbf{R}_{t,\mathrm{alt}}^{\mathrm{T}}\right]^{\mathrm{T}}$	$5N\times1$
Observation covariance	$\mathbf{Q}_{\mathrm{O},t} = \begin{bmatrix} \mathbf{\Sigma}_{\mathrm{P},t} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{\Sigma}_{\mathrm{ET},t} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{\Sigma}_{\mathrm{R_{obs}},t} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{\Sigma}_{\dot{\mathrm{M}},t} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{\Sigma}_{\mathrm{R_{alt}},t} \end{bmatrix}$	$5N\times5N$
Observation relation matrix	$\mathbf{H}_t = \begin{bmatrix} \mathbf{I} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{I} \\ \mathbf{0} & \mathbf{0} & \mathbf{I} & \mathbf{0} \end{bmatrix}$	$5N\times4N$
Constraints	$\mathbf{G}\mathbf{X}_t = \mathbf{0}$	$N\times4N$
	$\mathbf{G} = \left[\mathbf{I} \quad -\mathbf{I} \quad -\mathbf{I} \quad -\mathbf{I}\right]$	
Augmented obs. rel. matrix	$\mathbf{H}_{t,\mathrm{aug}} = \left[\mathbf{H}_t \quad \mathbf{G}\right]^{\mathrm{T}}$	$6N\times4N$
Augmented observation vector	$\mathbf{Y}_{t,\mathrm{aug}} = \left[\mathbf{Y}_t \quad \mathbf{0}\right]^{\mathrm{T}}$	$6N\times1$
Augmented observation covariance	$\mathbf{Q}_{\mathrm{O},t,\mathrm{aug}} = \begin{bmatrix} \mathbf{Q}_{\mathrm{O},t} & \mathbf{0} \\ \mathbf{0} & \mathbf{Q}_{\mathrm{WB},t} \end{bmatrix}$	$6N\times6N$

which might cause computational issues, for example, during inversion. Therefore, we propose and proceed with the EnKF framework.

3.2. Derivation of the Prediction Equation

On monthly timescales, it can be assumed that the basin-scale water cycle variables follow a (more or less) distinct annual cycle. Therefore, the *anomalies* at time-step t are expressed as

$$\mathbf{r}_t = \mathbf{X}_t - \tilde{\mathbf{X}}_t \tag{4}$$

where $\mathbf{X}_t$ is the state vector from (3). $\tilde{\mathbf{X}}_t$ represents the long-term mean annual cycle for each of the four water cycle variables for every basin through

$$\tilde{\mathbf{X}}_m = \frac{1}{T/12}\sum_{i}^{T/12}\mathbf{X}_{i,m}, \tag{5}$$

where the indices i and m refer to different years and months, respectively. It is further assumed that the prediction from time-step t–1 to t is expressed through a stochastic process of first order

$$\mathbf{r}_t = \mathbf{A}\mathbf{r}_{t-1}\boldsymbol{\epsilon}_t, \tag{6}$$

where $\mathbf{A}$ is a (yet unknown) prediction function and $\boldsymbol{\epsilon}_t$ some zero-mean white prediction noise with covariance matrix $D\{\boldsymbol{\epsilon}\} = \mathbf{Q}_{\mathrm{P}}$ and with the auto- and cross-covariance matrices

Table 4. Overview of the Assimilation Algorithm[a]

Parameter	Equation
Initial state	$\mathbf{X}^{+}_{0,d} = \mathbf{X}_0 + \boldsymbol{\epsilon}_d$ with $\boldsymbol{\epsilon}_d \sim \mathcal{N}(\mathbf{0}, \mathbf{Q}_P)$
Prediction step	$\mathbf{X}^{-}_{t,d} = \hat{\mathbf{A}}\mathbf{X}^{+}_{t-1,d} + \mathbf{B}\mathbf{U}_{t-1} + \varepsilon_d$ with $\boldsymbol{\epsilon}_d \sim \mathcal{N}(\mathbf{0}, \mathbf{Q}_P)$
Prediction covariance	$\hat{\boldsymbol{\Sigma}}^{-}_{X,t} = \frac{1}{D-1}\sum_{d=1}^{D}\left(\mathbf{X}^{-}_{t,d} - \overline{\mathbf{X}}^{-}_{t}\right)\left(\mathbf{X}^{-}_{t,d} - \overline{\mathbf{X}}^{-}_{t}\right)^{\mathsf{T}}$
Observation Innovations	$\mathbf{V}_{t,d} = \mathbf{Y}_t - \mathbf{H}_t\mathbf{X}^{-}_{t,d} + \boldsymbol{\nu}_{t,d}$ with $\boldsymbol{\nu}_{t,n} \sim \mathcal{N}(\mathbf{0}, \mathbf{Q}_{O,t})$
Kalman Gain	$\mathbf{K}_t = \hat{\boldsymbol{\Sigma}}^{-}_{X,t}\mathbf{H}_t^{\mathsf{T}}\left(\mathbf{H}_t\hat{\boldsymbol{\Sigma}}^{-}_{X,t}\mathbf{H}_t^{\mathsf{T}} + \mathbf{Q}_{O,t}\right)^{-1}$
Correction step	$\mathbf{X}^{+}_{t,d} = \mathbf{X}^{-}_{t,d} + \mathbf{K}_t\mathbf{V}_{t,d}$
Corr. covariance	$\hat{\boldsymbol{\Sigma}}^{+}_{X,t} = \frac{1}{D-1}\sum_{d=1}^{D}\left(\mathbf{X}^{+}_{t,d} - \overline{\mathbf{X}}^{+}_{t}\right)\left(\mathbf{X}^{+}_{t,d} - \overline{\mathbf{X}}^{+}_{t}\right)^{\mathsf{T}}$
Klaman Gain (smoother)	$\mathbf{K}_t = \hat{\boldsymbol{\Sigma}}^{+}_{X,t}\hat{\mathbf{A}}\left(\hat{\boldsymbol{\Sigma}}^{-}_{X,t+1}\right)^{-1}$
Smoothing step	$\mathbf{X}_{t,d} = \mathbf{X}^{+}_{t,d} + \mathbf{K}_t\left(\mathbf{X}_{t+1,d} - \mathbf{X}^{-}_{t+1,d}\right)$
Smoothed covariance	$\hat{\boldsymbol{\Sigma}}_{X,t} = \hat{\boldsymbol{\Sigma}}^{+}_{X,t} + \mathbf{K}_t\left(\hat{\boldsymbol{\Sigma}}_{X,t+1} - \hat{\boldsymbol{\Sigma}}^{-}_{X,t+1}\right)\mathbf{K}_t^{\mathsf{T}}$

[a]For the constrained predictions, the observation vector ($\mathbf{Y}_t$), the observation relation matrix ($\mathbf{H}_t$), and the observation covariance matrix ($\mathbf{Q}_{O,t}$) are replaced with the augmented parameters ($\mathbf{Y}_{t,\mathrm{aug}}$, $\mathbf{H}_{t,\mathrm{aug}}$, $\mathbf{Q}_{O,t,\mathrm{aug}}$).

$$\boldsymbol{\Sigma} = D\{\mathbf{r}_t\} \text{ and } \boldsymbol{\Sigma}_\Delta = D\{\mathbf{r}_t, \mathbf{r}_{t-1}\}. \quad (7)$$

The prediction matrix $\mathbf{A}$ is given through the so-called *least squares prediction* or *multiple ordinary least squares regression*, which is described in e.g., *Moritz* [1980] and *Kurtenbach et al*. [2012]. As the *true* process dynamics $\mathbf{A}$ are not known, we are looking for an approximation $\hat{\mathbf{A}}$, with which the prediction from time-step t to $t+1$ reads as

$$\hat{\mathbf{r}}_t = \hat{\mathbf{A}}\mathbf{r}_{t-1}, \quad (8)$$

where $\hat{\mathbf{r}}$ is the predicted state. The approximated prediction matrix should minimize the prediction error

$$\mathbf{e} = \mathbf{r}_t - \hat{\mathbf{r}}_t \quad (9)$$

in a least squares sense, i.e.,

$$\mathbf{e}\mathbf{e}^{\mathsf{T}} \rightarrow \min \quad (10)$$

According to (9), the total error can be expressed as

$$\begin{aligned}\mathbf{e}\mathbf{e}^{\mathsf{T}} &= (\mathbf{r}_t - \hat{\mathbf{r}}_t)(\mathbf{r}_t - \hat{\mathbf{r}}_t)^{\mathsf{T}} \\ &= \left(\mathbf{r}_t - \hat{\mathbf{A}}\mathbf{r}_{t-1}\right)\left(\mathbf{r}_t - \hat{\mathbf{A}}\mathbf{r}_{t-1}\right)^{\mathsf{T}} \\ &= \mathbf{r}_t\mathbf{r}_t^{\mathsf{T}} - \mathbf{r}_t\mathbf{r}_{t-1}^{\mathsf{T}}\hat{\mathbf{A}}^{\mathsf{T}} - \hat{\mathbf{A}}\mathbf{r}_{t-1}\mathbf{r}_t^{\mathsf{T}} + \hat{\mathbf{A}}\mathbf{r}_{t-1}\mathbf{r}_{t-1}^{\mathsf{T}}\hat{\mathbf{A}}^{\mathsf{T}}\end{aligned} \quad (11)$$

The corresponding error covariance matrix is given through

$$D\{\mathbf{e}\} = E\{\mathbf{e}\mathbf{e}^{\mathsf{T}}\} = \boldsymbol{\Sigma} - \boldsymbol{\Sigma}_\Delta\hat{\mathbf{A}}^{\mathsf{T}} - \hat{\mathbf{A}}\boldsymbol{\Sigma}_\Delta^{\mathsf{T}} + \hat{\mathbf{A}}\boldsymbol{\Sigma}\hat{\mathbf{A}}^{\mathsf{T}}. \quad (12)$$

An estimate for $\mathbf{A}$ is found by minimizing the trace of the error covariance matrix, which is the case for

$$\hat{\mathbf{A}} = \boldsymbol{\Sigma}_\Delta\boldsymbol{\Sigma}^{-1}. \quad (13)$$

For a more detailed derivation, see e.g., *Kurtenbach et al*. [2012]. As it is assumed that this prediction matrix remains constant over time, the time-index has been omitted. Inserting (13) into (12) yields

$$D\{\mathbf{e}\} = \boldsymbol{\Sigma} - \boldsymbol{\Sigma}_\Delta\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_\Delta = \mathbf{Q}_P \quad (14)$$

which is the error covariance matrix of the predictor $\hat{\mathbf{A}}$ or, in the common Kalman notation, the covariance matrix of the prediction noise. As the *true* auto- and cross-covariances between anomalies of precipitation, evapotranspiration, runoff, and water storage changes are also unknown, they are approximated by the empirical sample covariance matrices

$$\hat{\boldsymbol{\Sigma}} = \frac{1}{T-1}\sum_{t=1}^{T}\mathbf{r}_t\mathbf{r}_t^{\mathsf{T}} \quad (15)$$

$$\hat{\boldsymbol{\Sigma}}_\Delta = \frac{1}{T-2}\sum_{t=2}^{T}\mathbf{r}_t\mathbf{r}_{t-1}^{\mathsf{T}} \quad (16)$$

The two matrices are shown in Figures 2a and 2b.

Putting (4) into (6), the process model reads as [*Tourian*, 2013]

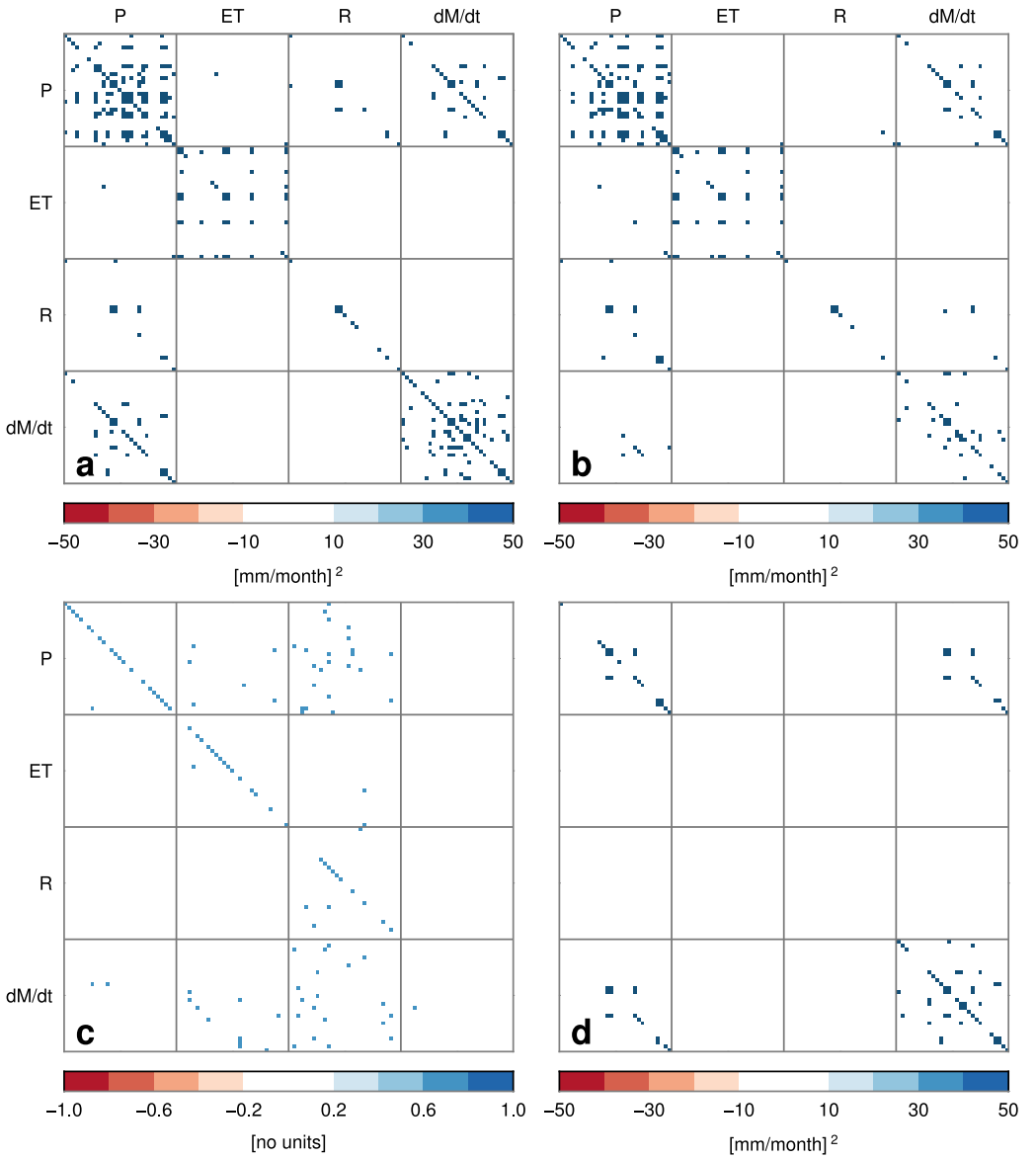


Figure 2. Temporal (a) Auto-covariance (top left) and (b) cross-covariance (top right) matrix ($\boldsymbol{\Sigma}$ and $\boldsymbol{\Sigma}_{\Delta}$, see equations (15) and (16)) between anomalies of precipitation, evapotranspiration, runoff, and water storage changes. The anomalies are computed with respect to the ensemble mean annual cycle from all available data sources. The matrices thus represent the overall mean covariance structure of the water cycle variables. Both matrices consist of 16 submatrices, of which each has the dimension [29×29]. These submatrices therefore represent the auto- and (temporal) cross-covariance between the different study regions and water cycle variables. (c) Least squares prediction matrix $\hat{\mathbf{A}}$ (bottom left) and (d) the corresponding error matrix $\mathbf{Q}_{\mathrm{P}}$ (bottom right).

$$\begin{aligned}\left(\mathbf{X}_t-\tilde{\mathbf{X}}_t\right)&=\hat{\mathbf{A}}\left(\mathbf{X}_{t-1}-\tilde{\mathbf{X}}_{t-1}\right)+\boldsymbol{\epsilon}_t\\&=\hat{\mathbf{A}}\mathbf{x}_{t-1}-\hat{\mathbf{A}}\tilde{\mathbf{X}}_{t-1}+\boldsymbol{\epsilon}_t\\\Rightarrow\ \mathbf{X}_t&=\hat{\mathbf{A}}\mathbf{X}_{t-1}-\hat{\mathbf{A}}\tilde{\mathbf{X}}_{t-1}+\tilde{\mathbf{X}}_t+\ \boldsymbol{\epsilon}_t\end{aligned}\tag{17}$$

which can be rewritten as

$$\mathbf{X}_t=\hat{\mathbf{A}}\mathbf{X}_{t-1}+\mathbf{B}\mathbf{U}_{t-1}+\ \boldsymbol{\epsilon}_t\tag{18}$$

with

$$\mathbf{B}=\begin{bmatrix}-\hat{\mathbf{A}} & \mathbf{I}\end{bmatrix} \text{ and } \mathbf{U}_t=\begin{bmatrix}\tilde{\mathbf{X}}_{t-1}\\ \tilde{\mathbf{X}}_t\end{bmatrix} \tag{19}$$

Thus, the prediction is based on the covariances between the state variables (through the least squares prediction matrix) and the mean annual cycle through the control input $\mathbf{B}_{t-1}\mathbf{U}_{t-1}$.

The control input is computed from the collection of data sets, which were described in section 2. For each water cycle variable, the annual cycle is given through the *mean* of all respective data sources. As we are mainly interested in the performance of the methodology as a predictor, only past data from 1980 to 2002 are used for computing the mean annual cycle.

3.3. Derivation of the Observation Equation

The first observation group relates the state vector to observed precipitation, evapotranspiration, runoff, and water storage changes. Thus, the framework allows the assimilation of an arbitrary number of observation data of water cycle variables. These variables can be further used as constraints, if they are assumed to be highly precise.

For assimilating runoff observations from satellite altimetry $\mathbf{R}_{t,\mathrm{alt}}$, another observation group is added to the observation vector. The runoff estimates in the state vector are directly related to the altimetry observations, which then gives the full observation equation through

$$\underbrace{\begin{bmatrix}\mathbf{P}_{t,\mathrm{obs}}\\ \mathbf{ET}_{t,\mathrm{obs}}\\ \mathbf{R}_{t,\mathrm{obs}}\\ \dot{\mathbf{M}}_{t,\mathrm{obs}}\\ \mathbf{R}_{t,\mathrm{alt}}\end{bmatrix}}_{\mathbf{Y}_t}=\underbrace{\begin{bmatrix}\mathbf{I} & \mathbf{0} & \mathbf{0} & \mathbf{0}\\ \mathbf{0} & \mathbf{I} & \mathbf{0} & \mathbf{0}\\ \mathbf{0} & \mathbf{0} & \mathbf{I} & \mathbf{0}\\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{I}\\ \mathbf{0} & \mathbf{0} & \mathbf{I} & \mathbf{0}\end{bmatrix}}_{\mathbf{H}_t}\underbrace{\begin{bmatrix}\mathbf{P}_t\\ \mathbf{ET}_t\\ \mathbf{R}_t\\ \dot{\mathbf{M}}_t\end{bmatrix}}_{\mathbf{X}_t}+\boldsymbol{v}_t \tag{20}$$

or, as in the common Kalman literature,

$$\mathbf{Y}_t=\mathbf{H}_t\mathbf{X}_t+\boldsymbol{v}_t, \tag{21}$$

where $\boldsymbol{v}_t$ is the zero-mean white observation noise with covariance matrix $D\{\boldsymbol{v}\}=\mathbf{Q}_{\mathrm{O}}$. The computation of this covariance matrix is discussed in section 3.5. If the observed time series contain missing values, the observation relation matrix has to be adjusted accordingly. Therefore, the time-index t is used here to indicate a *dynamic* observation relation matrix, which changes depending on the number of available observations.

Similar to the control input in section 3.2, the observations are computed from the collection of input data sources shown in Table 1. It is common practice in hydrology to use an ensemble mean instead of single models or data sets [e.g., *Kirtman et al.*, 2013]. Nevertheless, high-quality observations over e.g., a single catchment (if available) could be used here as well. This will be further analyzed in future studies.

3.4. Closure of Catchment-Scale Water Budgets

Water budget closure between the four water cycle variables can be achieved by adding state constraints to the assimilation scheme. These constraints adjust the estimates of precipitation, evapotranspiration, runoff, and water-storage changes in a way that the catchment-scale water budgets are closed. There are several methods for adding such constraints to the assimilation scheme [e.g., *Simon and Chia*, 2002; *Simon*, 2010]. Here we discuss and apply two straightforward approaches for adding water budget constraints.

Both methods are based on the augmentation of the observation vector by adding some *pseudo observations* of the water balance closure. The terrestrial water balance equation (1) can be written as a state constraint through

$$\begin{bmatrix}\mathbf{I} & -\mathbf{I} & -\mathbf{I} & -\mathbf{I}\end{bmatrix}\begin{bmatrix}\mathbf{P}_t\\ \mathbf{ET}_t\\ \mathbf{R}_t\\ \dot{\mathbf{M}}_t\end{bmatrix}=\mathbf{0}, \tag{22}$$

where $\mathbf{0}$ is a $[N\times 1]$ vector which contains only zeros and $\mathbf{I}$ are $[N\times N]$ identity matrices.

When we write these constraints in shorter matrix notation

$$\mathbf{G}\mathbf{X}_t = \mathbf{0} \tag{23}$$

with $\mathbf{G} = [\mathbf{I} \quad -\mathbf{I} \quad -\mathbf{I} \quad -\mathbf{I}]$, we can augment the observation equation (20) through

$$\begin{bmatrix} \mathbf{Y}_t \\ \mathbf{0} \end{bmatrix} = \begin{bmatrix} \mathbf{H}_t \\ \mathbf{G} \end{bmatrix} \mathbf{X}_t + \begin{bmatrix} v_t \\ \omega_t \end{bmatrix}, \tag{24}$$

where ω_t is some zero-mean white noise with covariance $\mathbf{Q}_{\mathrm{WB}}$ for the pseudo observations of the water budget closure.

If the water budget should be perfectly closed, the method is usually referred to as *perfect observations* or *hard constraints*. This can be easily achieved by assuming the pseudo observations to be noise free, i.e., $\mathbf{Q}_{\mathrm{WB}} = \mathbf{0}$. However, this assumption can lead to numerical issues, as the augmented observation error covariance matrix is singular. Furthermore, by considering the large differences between the data sources of precipitation, evapotranspiration, and water storage changes, such a constraint might be too strict.

Therefore, one could allow some small imbalances. Such constraints are thus called *imperfect observations* or *soft constraints*. In this case, we add some small nonzero observation errors, i.e., $\omega_t \sim \mathcal{N}(\mathbf{0}, \mathbf{Q}_{\mathrm{WB}})$. In order to be consistent with the basin-specific characteristics, this error is chosen to be 10% of the mean annual cycle of runoff [*Fekete et al*., 2012].

3.5. A Multidata Approach for Estimating Basin-Scale Errors

As there are plenty of data sources, it is difficult to identify the "best" data sets for each of the water cycle variables. It was shown in *Lorenz et al*. [2014] that there is no such data set that performs consistently well on the global scale. Moreover, a proper description of the observation errors is essential in order to successfully apply any Kalman Filtering approach. Therefore, an ensemble-based approach has been chosen, where numerous realizations of observations are estimated, based on the range of the input data sources.

First, the *range* of the collection of input data sets for the water cycle variable F (which is either precipitation, evapotranspiration, or water storage changes) is computed for every catchment n and every time-step t. This range is assumed to be a proxy for the *uncertainty* of each of the four water cycle variables. Therefore, all precipitation anomalies from K different data sets are stored in a vector:

$$\mathbf{f}_{n,t} = [f_{n,t,1} \quad f_{n,t,2} \quad \cdots \quad f_{n,t,K}]^{\mathsf{T}} \tag{25}$$

where the $f_{n,t,k}$ are the anomalies with respect to the mean annual cycle $\tilde{F}_t$. From this vector, the range is given through

$$\Delta f_{n,t} = \frac{1}{2}\left[\max\left(\mathbf{f}_{n,t}\right) - \min\left(\mathbf{f}_{n,t}\right)\right] \tag{26}$$

The sample covariance and correlation matrices are then estimated through

$$\mathbf{\Sigma}_{\Delta f} = \frac{1}{T-1}\sum_{t=1}^{T}\left(\Delta\mathbf{f}_t - \overline{\Delta\mathbf{f}}\right)\left(\Delta\mathbf{f}_t - \overline{\Delta\mathbf{f}}\right)^{\mathsf{T}} \tag{27}$$

$$\boldsymbol{\rho}_{\Delta f} = \left(\mathbf{\Sigma}_{\Delta f}^{(\mathrm{diag})}\right)^{-\frac{1}{2}} \mathbf{\Sigma}_{\Delta f} \left(\mathbf{\Sigma}_{\Delta f}^{(\mathrm{diag})}\right)^{-\frac{1}{2}} \tag{28}$$

where $\mathbf{\Sigma}_{\Delta f}(\mathrm{diag})$ is the diagonal variance matrix of $\mathbf{\Sigma}_{\Delta f}$. These correlation matrices thus describe the relationship between the *uncertainties* of the water cycle anomalies from different catchments. In order to use as much data as possible, the matrices are computed from past data only (i.e., until 2002). Then, the correlation matrices are rescaled with the ensemble standard deviation from actual data (i.e., from 2003). Although multiple data sets for each of the hydrological variables are used, there is not enough data for estimating reliable ensemble standard deviations for *every* time-step. Therefore, a *cyclostationary* approach has been chosen:

$$\sigma_{F,n,m} = \sqrt{\frac{1}{K(T/12)-1}\sum_{k=1}^{K}\sum_{i=1}^{T/12}\left(F_{n,k,i,m} - \tilde{F}_{n,k,m}\right)^2} \tag{29}$$

where the index i represents a single year and m a single calendar month (e.g., January, February, etc.). K is the number of data sets for a single variable and $\bar{F}_{n,k,m}$ the long-term mean annual cycle of the kth data set over catchment n for month m. The standard deviations for the other calendar months are computed likewise. The error covariance matrix is then given by

$$\boldsymbol{\Sigma}_{P,m} = \left(\mathbf{I} \;\; \boldsymbol{\sigma}_{F,m}\right) \;\; \boldsymbol{\rho}_{\Delta f} \left(\mathbf{I} \;\; \boldsymbol{\sigma}_{F,m}\right) \tag{30}$$

where $\mathbf{I}$ is the $N \times N$ unit matrix and

$$\boldsymbol{\sigma}_{F,m} = \left[\, \sigma_{F,1,m} \quad \sigma_{F,2,m} \quad \cdots \quad \sigma_{F,N,m} \,\right] \tag{31}$$

The matrices for each water cycle variable are then combined in the full observation error covariance matrix

$$\mathbf{Q}_{O,t} = \begin{bmatrix} \boldsymbol{\Sigma}_{P,t} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Sigma}_{ET,t} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \boldsymbol{\Sigma}_{R_{obs},t} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \boldsymbol{\Sigma}_{\dot{M},t} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \boldsymbol{\Sigma}_{R_{alt},t} \end{bmatrix}, \tag{32}$$

where $\boldsymbol{\Sigma}_{R_{obs}}$ and $\boldsymbol{\Sigma}_{R_{alt}}$ are the covariance matrices for the runoff observations from GRDC and satellite altimetry, respectively. The mean monthly index m has been replaced with the time-index t in order to account for the monthly errors in the runoff observations from GRDC and satellite altimetry. According to e.g. *Fekete et al.* [2012], traditional discharge measurements usually have an accuracy of 5–10%. We thus compute the errors of the runoff observations from GRDC to be 5% of the reported values. The errors from satellite altimetry are obtained by propagating the errors of the altimetric water level measurements through the rating curve model, as described in *Tourian et al.* [2013].

Obviously, correlations between the uncertainties of different water cycle variables are neglected. As the data come from totally different sensors, we think that this is a reasonable assumption. However, it should be further investigated how, for example, a fully populated observation error covariance matrix affects the performance of the assimilation framework.

Figure 3 shows histograms of random perturbations derived from the covariance matrix in (32). These random values are used for computing an ensemble of perturbed observations within the EnKF. Here we only show some exemplary realizations for the three catchments Amazon, Danube, and Yana for January and July. It is obvious that the chosen methodology is able to produce different perturbations for different catchments, variables, and months. For example, the precipitation perturbations in January over the Amazon basin show a wide spread as only about 73% are within the range between −50 and 50 mm/month. On the contrary, in July, almost 100% are within the same interval. Over the Yana basin, almost 100% of the precipitation perturbations are between −10 and 10 mm/month. The method is thus able to provide perturbations which depend on the climatic conditions, but also on the spread of the different data sets over the basins. However, it should be mentioned that a limited number of rainfall gauges in a basin can lead to very similar precipitation estimates from different products. This suggests a good agreement between these data sources, even if the true precipitation can be totally different.

3.6. Computation of the Least Squares Prediction Matrix

According to equations (13) to (16), the least squares prediction matrix $\hat{\mathbf{A}}$ and its error covariance matrix $\mathbf{Q}_P$ are computed from the auto- and cross-covariance matrices between anomalies of precipitation, evapotranspiration, runoff, and water storage changes. In order to capture reliable covariance structures, a collection of different data sources (which are listed in Table 1) has been used instead of single data sets. As it is avoided to use the same data twice, the covariance matrices are derived from a different period (1980 until 2002) than the period during which runoff is estimated (2003 until 2010). The two matrices $\boldsymbol{\Sigma}$ and $\boldsymbol{\Sigma}_\Delta$ are shown in Figures 2a and 2b.

It is emphasized that these matrices are derived from the anomalies, which consist of random errors, climatic variability and extreme events. They thus represent covariances mainly due to features which deviate significantly from the annual cycle. Both the auto- and cross-covariance matrix show high positive inter-

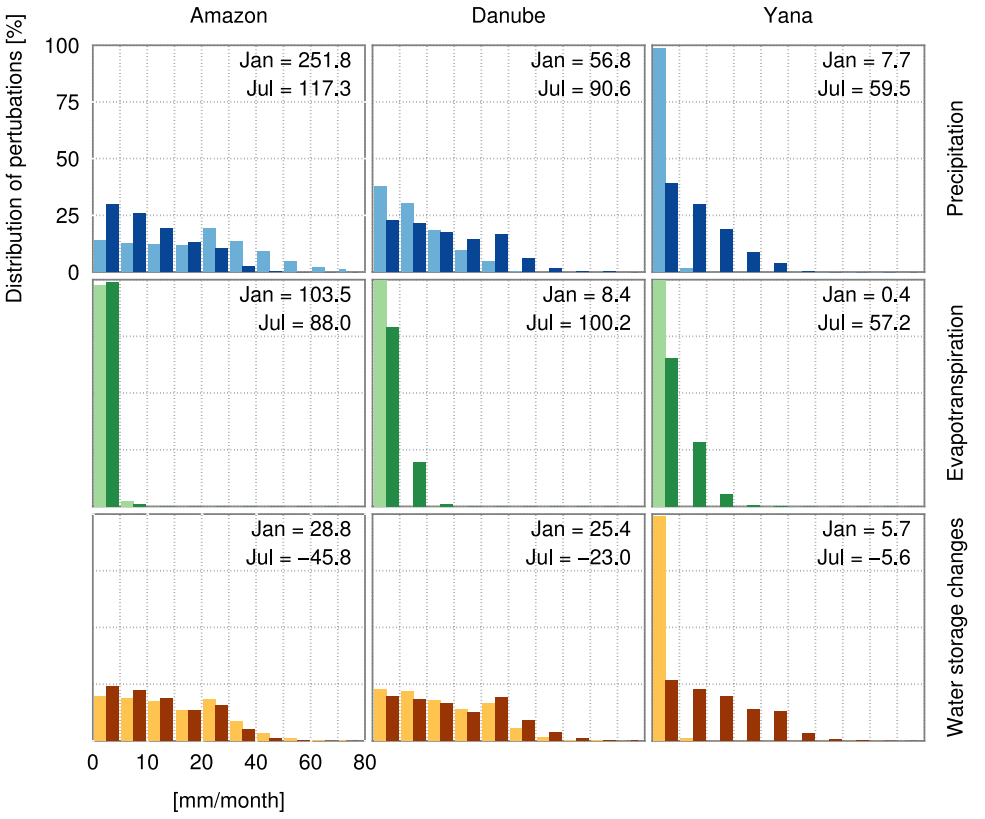


Figure 3. Histograms of 10000 random realizations drawn from a multivariate normal distribution with zero mean and the observation error covariance matrix from (32). The histograms show the distributions of the random perturbations for the three catchments (left column) Amazon, (middle column) Danube, and (right column) Yana for (top row) precipitation, (middle row) evapotranspiration, and (bottom row) water storage changes. The left bars (bright colors) correspond to the perturbations of January while the right bars (dark colors) show the perturbations in July. The figure only shows the absolute value of the perturbations as they are symmetric around zero. The numbers in the top right corner of every histogram are the ensemble mean values of precipitation, evapotranspiration, and water storage changes for January and July, respectively, in mm/month.

catchment covariances throughout most basins especially for precipitation, evapotranspiration, and water storage changes. These characteristics have been already addressed in *Riegger et al*. [2012], where it is concluded that inter-catchment correlations can be expected e.g., for neighboring catchments due to local climatic conditions.

The off-diagonal elements in $\mathbf{\Sigma}$ and $\mathbf{\Sigma}_\Delta$ for runoff are comparatively small. This indicates that there are no such strong inter-catchment correlations (or covariances) for runoff. Thus, at the anomaly level, climatic and physical variations in time series of runoff are rather local phenomena.

The inter-variable covariances are generally lower. Some strong relationship between precipitation, water storage changes, and, to a certain extent, runoff can be detected. This is indicated by relatively high covariances on the main-diagonals of the corresponding matrix blocks. At the anomaly level, this is intuitive as, for example, an exceptionally strong precipitation event might lead to high runoff and an increased water storage. The second case can be clearly identified by the high covariances on the main diagonals of the upper right and lower left submatrices, respectively. In terms of precipitation and runoff, such a strong relationship can only be observed over some catchments. As an example, the gauges of the four basins Amazon, Tocantins, Rio Xingu, and Rio Parnaiba are all located at the North-Eastern part of South America. Over these catchments, there are high inter-catchment and inter-variable covariances between precipitation, runoff, and water storage changes.

On the contrary, evapotranspiration generally does not show such high covariances with any of the other water cycle variables. This indicates that there is no strong coupling between the anomalies of evapotranspiration and other water cycle variables. The reason for this is the stable periodic characteristic of the

evapotranspiration time series. Compared to the other water cycle variables, the total signal power (i.e., the root mean square RMS) of the evapotranspiration anomalies are very low and their dynamics possess a random behavior (not shown here).

When comparing the auto- and cross-covariance matrices in Figures 2a and 2b, they both show similar structures, even if the covariances in $\mathbf{\Sigma}_\Delta$ are generally lower. The top right submatrix in Figure 2b shows, as hydrologically expected, a strong cross-covariance between precipitation and water storage changes with a time-lag of 1 month. On the other hand, the bottom left submatrix does not show such high covariances. Thus, after a strong precipitation event, the impact on the water storage can be seen even 1 month later. There is no such strong coupling between e.g., anomalies of precipitation and runoff, even if that could have been expected. This suggests that, based on the empirical covariances, significant deviations from the mean annual cycle of precipitation do not necessarily lead to exceptional runoff months.

The least squares prediction matrix and its corresponding error covariance matrix can be derived from the auto- and cross-covariance matrices according to equations (13) and (14). The two matrices $\hat{\mathbf{A}}$ and $\mathbf{Q}_\mathrm{P}$ are shown in Figures 2c and 2d.

Instead of just using fully populated covariance matrices, we can distinguish between three cases:

1. Covariances between catchments and variables (case 1)
2. Covariances between catchments (case 2)
3. Covariances between variables (case 3)

In the first case, it is assumed that there are covariances between both the catchments and the water cycle variables. In case two and three, either the covariances between the water cycle variables (case 2) or the catchments (case 3) are neglected. In Figures 2c and 2d, only the matrices for case 1 are shown. The other two cases are derived from these matrices by removing all submatrices except for those on the main diagonal (case 2) or all elements except for those on the main diagonals of each submatrix (case 3). It will be discussed in section 4.1.3 if these three cases lead to different predictions. Besides that, initial tests confirmed that using equation (8) with different configurations of the least squares prediction matrices can explain about 70% of the runoff anomaly variance (not shown). However, this must not hold true for the other water cycle variables, which might be less predictable through such a covariance-based least squares approach.

3.7. Performance Metrics

For validating the estimated time series of precipitation, evapotranspiration, runoff, and water storage changes, numerous performance metrics can be evaluated. In hydrology, it is common to use correlation, the percentage bias (PBIAS) [*Gupta et al*., 1999], and the Nash-Sutcliffe Efficiency (NSE) [*Nash and Sutcliffe*, 1970]. From these, correlation describes the level of common information content between two time series. As it is insensitive to the amplitude and the mean value, the PBIAS between an observed (Y_t) and predicted (X_t) time series,

$$\mathrm{PBIAS} = \frac{\sum_{t=1}^{T}(X_t - Y_t)}{\sum_{t=1}^{T} Y_t} = 1 - \frac{\bar{X}}{\bar{Y}}, \tag{33}$$

gives precise information about the relative difference between the long-term means. Finally, both metrics are summarized in the $\mathrm{NSE}_{\mathrm{mean}}$ coefficient

$$\mathrm{NSE}_{\mathrm{mean}} = 1 - \frac{\sum_{t=1}^{T}(X_t - Y_t)^2}{\sum_{t=1}^{T}(Y_t - \bar{Y})^2} \tag{34}$$

which is highly sensitive to the agreement in phase, amplitude, and mean between two data sets. The $\mathrm{NSE}_{\mathrm{mean}}$ further represents the normalized mean squared error between e.g., an observed and predicted time series [*Lorenz et al*., 2014].

In this study, an alternative formulation of the NSE coefficient, which takes the long-term annual cycle of a variable into account, is proposed:

$$\mathrm{NSE}_{\mathrm{cycle}} = 1 - \frac{\sum_{t=1}^{T} (X_t - Y_t)^2}{\sum_{t=1}^{T} \left(Y_t - \tilde{Y}\right)^2} \tag{35}$$

where $\tilde{Y}$ is not the long-term mean, but the long-term mean annual cycle. Similar to the common Nash-Sutcliffe Efficiency, (35) gives information if the estimated time-series X_t is a better predictor than the long-term annual cycle from historic data $\tilde{Y}$ ($\mathrm{NSE}_{\mathrm{cycle}} > 0$). If $\mathrm{NSE}_{\mathrm{cycle}} < 0$, the estimated time-series can not improve the long-term mean annual cycle with respect to the observations. This indicates that the estimated residuals (with respect to the mean annual cycle) then do not capture the short- and long-term climatic variability in the observations.

4. Results and Discussion

In this section, the performance of the presented approach will be discussed. First, different set-ups of the assimilation framework are analyzed in order to find an appropriate configuration. Therefore, several tests are run with both the unconstrained and hard Constrained Ensemble Kalman Filter (hereafter EnKF and $\mathrm{CEnKF_h}$). For these tests, the framework estimates runoff over 16 of the 29 catchments, where continuous runoff observations during the period 2005–2010 are available (see Figure 1 and Table 2). The runoff observations from the GRDC for these 16 catchments are removed from 2005, i.e., they are not assimilated. Thus, all following performance metrics are derived from the comparison between observed runoff and the estimates from the data assimilation framework during the period 2005–2010.

4.1. Sensitivity Analysis

4.1.1. Ensemble Size

In an Ensemble Kalman Filtering framework, the *true* state covariance matrices are approximated by empirical sample covariance matrices. Therefore, it has to be investigated if and how different ensemble sizes have an impact on the estimates. Studies like e.g., *Mitchell et al*. [2002] provide a comprehensive analysis of this issue. Here, estimated and observed runoff are compared with respect to the ensemble size. Figure 4a thus shows the $\mathrm{NSE}_{\mathrm{mean}}$ between observed and estimated runoff for the first five catchments out of the 16 study regions, namely the Amazon, Congo, Mississippi, Ob, and Yenisei basins.

From these catchments, there are catchments with (Amazon and Ob) and without (Congo, Mississippi, and Yenisei) runoff observations from altimetry. Furthermore, these catchments cover the main climatic regimes, i.e., the tropics (Amazon and Congo), mid- (Mississippi) and high-latitudes (Ob and Yenisei). This allows a basic analysis with respect to different runoff characteristics.

The figure clearly shows that $\mathrm{NSE}_{\mathrm{mean}}$ remains rather constant throughout all ensemble sizes, especially over the Amazon, Ob, and Yenisei basins. The largest variations can be observed for the Mississippi basin when ensemble sizes smaller than 5000 are used. However, these variations are negligible compared to the magnitude of the performance metric. The impact of the ensemble size on the unconstrained (dark colors) and constrained (bright colors) are very similar. From this analysis we have chosen an ensemble size of 10000 as Figure 4a clearly shows that using larger ensembles does not lead to improved results.

4.1.2. Perturbation Analysis

As any Ensemble Kalman Filtering approach involves random drawings of perturbations, the estimates from repeated runs of the EnKF can show a certain spread. In order to analyze this uncertainty within the proposed framework, the EnKF and $\mathrm{CEnKF_h}$ are run with a fixed ensemble size of 10000 for 500 times. Figure 4b shows the distribution of the Nash-Sutcliffe Efficiency between observed and estimated runoff for the 16 study regions.

Foremost, the NSE values show a very small spread, even if there are large differences between the catchments. For most catchments, the interquartile range (IQR) of the NSE is below 0.04. Larger values can be observed for the Orange or Don basins, where the IQR can reach values up to 0.06. Runoff over these catchments usually does not follow a typical annual cycle and has a lower overall magnitude. Therefore, even

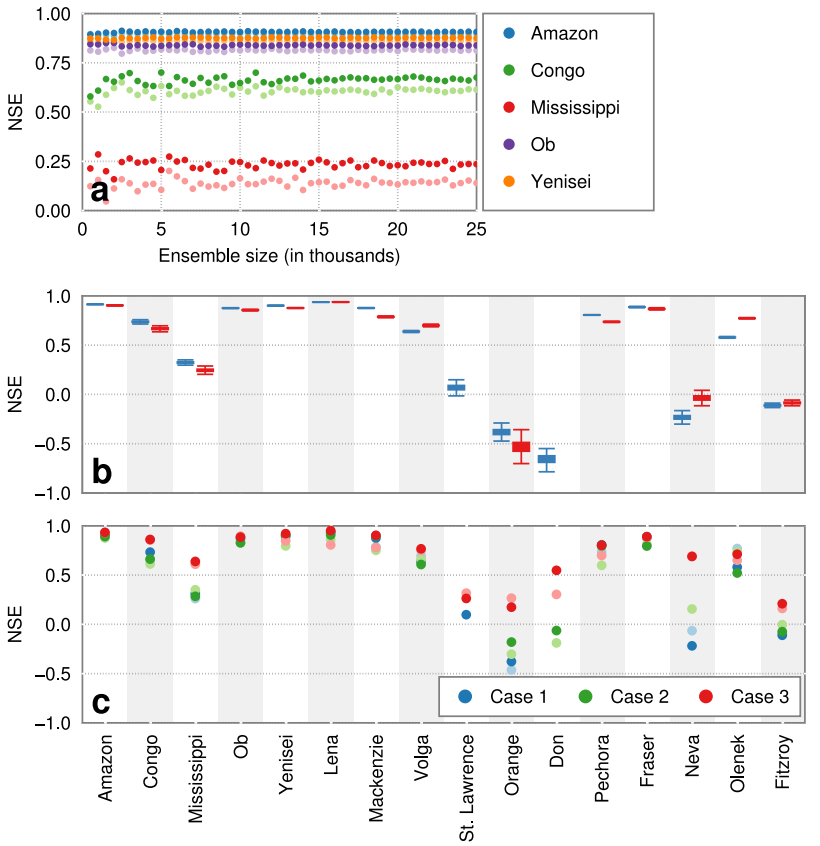


Figure 4. Sensitivity analysis showing the NSE values between observed and predicted runoff. (a) NSE derived from multiple runs of the filter using different ensemble sizes (the darker and brighter colors are derived from the EnKF and CEnKF$_h$, respectively). (b) Box-and-whisker plots from 500 runs of the EnKF (blue) and the CEnKF$_h$ (red) with the same settings and an ensemble size of 10000. The boxes represent the 25% and 75% percentile of the distribution of performance metrics. The top and bottom whiskers cover approximately 99.3% of the data, if a normal distribution is assumed. (c) NSE from the EnKF (dark colors) and the CEnKF$_h$ (bright colors) using a fully populated prediction matrix (case 1, blue), and prediction matrices which neglect correlations between water cycle variables (case 2, green) or catchments (case 3, red).

small changes in the predicted time-series can result in significantly different performance metrics. But compared to the inter-catchment variability of NSE values, this spread appears negligible.

Comparing the performance metrics from the unconstrained (EnKF) and the constrained (CEnKF$_h$) estimates clearly shows that there is no significant difference in terms of spread. Thus, even if the water budgets are forced to be closed, the estimated runoff time-series from multiple runs of the CEnKF$_h$ show a (negligible) variability.

4.1.3. Structure of the Least Squares Prediction Matrix

It is investigated how the different covariance structures in the least squares prediction matrix influence the results. This analysis thus helps to identify if the estimates benefit from fully populated covariance matrices (covariances between both catchments and water cycle variables, case 1) or if covariances between catchments only (case 2) or water cycle variables only (case 3) provide better results.

Figure 4c shows the NSE between observed runoff and the estimates from the EnKF and CEnKF$_h$ over 16 catchments for different configurations of the prediction matrix. Most catchments show positive values for all three cases, while the highest values are clearly achieved by case 3. In addition, using an unconstrained (darker dots) or constrained (brighter dots) does not lead to significantly different metrics. The worst performance can be observed for the Orange, Don, and Fitzroy basins, where case 1 and 2 lead to negative NSE values. The reasons for this are again the low overall magnitudes and irregular dynamics of the runoff time-series over these catchments.

The best performance can be achieved when correlations only between the water cycle variables are assumed (case 3). This can be understood from case 2, where the runoff estimates for a specific catchment

depend only on runoff from other basins. From a climatic point of view, this would make sense as for example, the tropical catchments often show similar runoff characteristics. However, this relationship is obviously not strong enough for providing reasonable estimates.

On the other hand, it is noted that the use of fully populated prediction matrix (case 1) does not lead to better results. For instance, the NSE values for case 1 do not always lie between those from case 2 and 3. The reason for this behavior is not yet fully understood, and therefore, needs further investigation. Henceforth, we will use covariance structures as in case 3 for the proposed methods.

4.2. Prediction Performance

After the identification of a suitable set-up, this section discusses the overall prediction performance of the proposed method. The filter is run with an ensemble size of 10000 and case 3 of the prediction matrix in six different configurations (see section 3.1). The estimated time-series for the 16 study regions are shown in Figures 5 and 6, while the performance metrics with respect to the runoff observations are summarized in Figure 7.

In terms of correlation, an overall good agreement between the dynamics of estimated and observed runoff can be identified. Over most catchments (except for Orange and Fitzroy), the different configurations led to correlations larger than 0.6. The catchments with high amplitudes and a dominant annual cycle in the runoff time-series (Amazon, Congo, Mississippi, Ob, Yenisei, Lena, Mackenzie, Volga, Pechora, Fraser, Neva, and Olenek) show correlations of at least 0.8. This good agreement can be explained by the repeating annual cycle of runoff, which is reproduced well by all six set-ups (see Figures 5 and 6). The main reason for this is the strong control input, which forces the runoff estimates to follow the mean annual cycle. However, several catchments show significant climatic variations, which are revealed by the anomaly time-series in the right columns of Figures 5 and 6. For example, over the Amazon basin, even if there is a dominant annual cycle, the anomalies still show some significant dynamics, which are reproduced quite well by the estimates. Also over Congo, Mississippi, Ob, Volga, Fraser, Neva, and Olenek, the estimated runoff anomalies show significant long-term variations, which agree well with the observations.

For Lena, Pechora, and Olenek, the runoff observations show a significant second peak during the late summer months, that can not be captured by the estimates. Over the Mackenzie basin, the estimates can only capture the annual cycle but none of the exceptional events. Additionally, e.g., the sudden drop in the constrained anomalies during 2008 can not be found in the observations.

It must be noted that the anomalies for many of the catchments with a dominant annual cycle are rather small, which complicates the estimation of exceptional events. This also explains why the correlations over these basins, despite the disagreement of the anomalies, are still very high.

Over catchments with lower overall magnitudes and irregular time-series dynamics (St. Lawrence, Orange, Don, Fitzroy), several configurations achieve correlations of 0.5 and more. Thus, the runoff estimates are in principle able to capture a large part of the climatic variability in the observations. But when looking at the time-series for these four basins (Figures 5 and 6), the estimates do not agree well with the observations. Over Don, both the estimates and the observations show a peak in runoff during the summer months during the years 2005 and 2006. In 2007 and 2009, there are no such peaks in the observations, but still in the estimated runoff. Nevertheless, on the anomaly level (right columns in Figures 5 and 6), observed runoff shows a slight decreasing trend during the period 2005 until 2010 which is also roughly reproduced by the estimates. That being said, if runoff over such basins is estimated from a water balance approach, it is mandatory that all input variables are highly consistent in both their magnitudes and dynamics. Otherwise, due to the biases in the estimates of the other water cycle variables, some artificial short-term variations are introduced, which finally result in both low correlation values and high relative biases.

Overall, the estimates are not able to reproduce extreme events in the runoff time-series (e.g., Mississippi in 2008). This can be explained by several reasons. First of all, runoff is predicted using a least squares prediction, which exploits covariance information. In fact, this information reflects only the averaged statistical relationship between the variables and basins. The statistical relationships of extreme events are thus not well represented. Moreover, the filtering of the input data, as described in section 2.5, significantly dampens the peaks in the time-series. An extreme event could disappear simply due to the smoother input time-series. On the other hand, the predictions show a significant increase during heavy flooding events over the

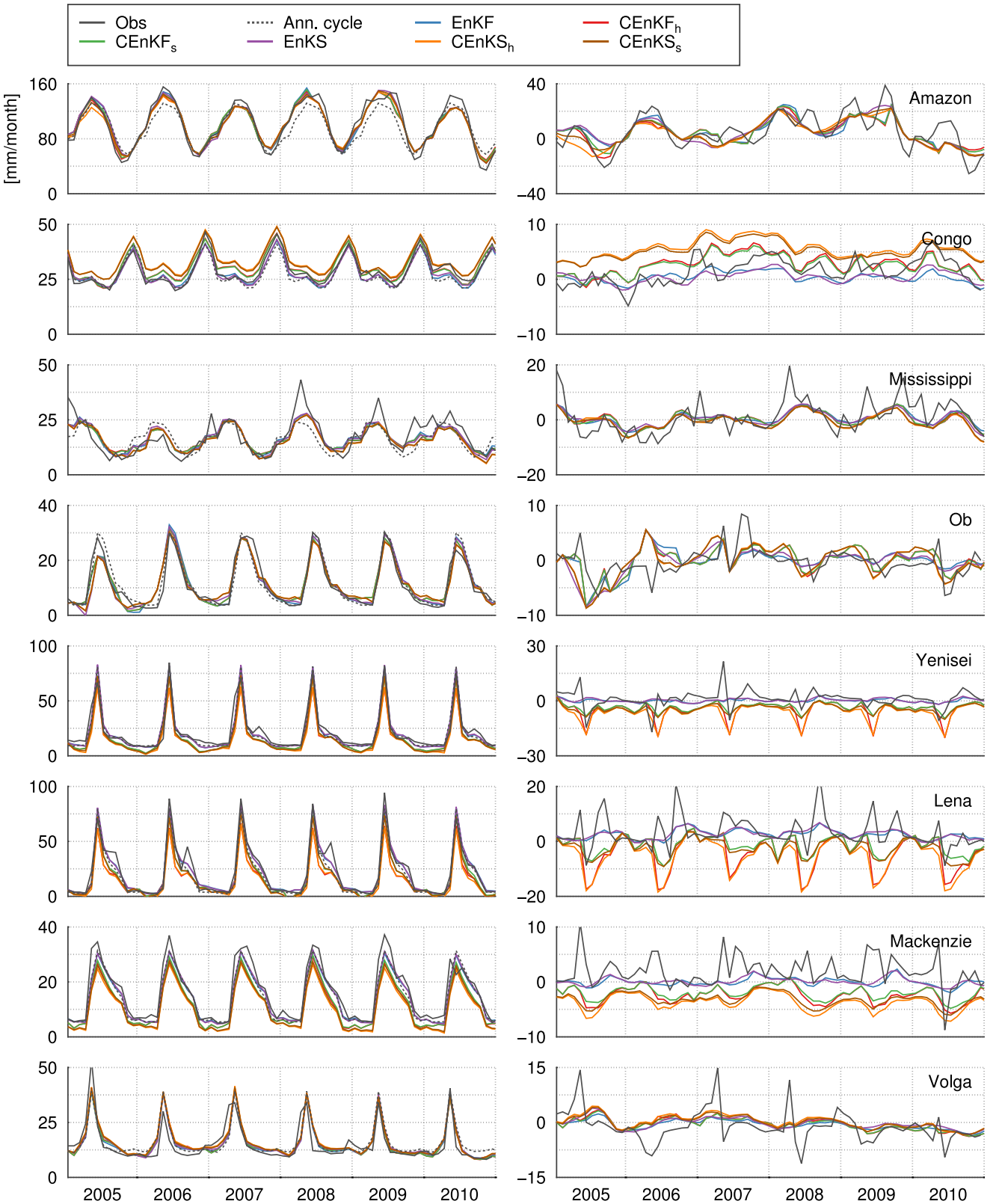


Figure 5. Time-series of observed (black) and predicted runoff from different configurations of the Kalman Filter and Smoother for the study regions 1–8, showing the (left column) full runoff time-series and (right column) runoff anomalies, where the mean annual cycle from historic runoff data has been removed.

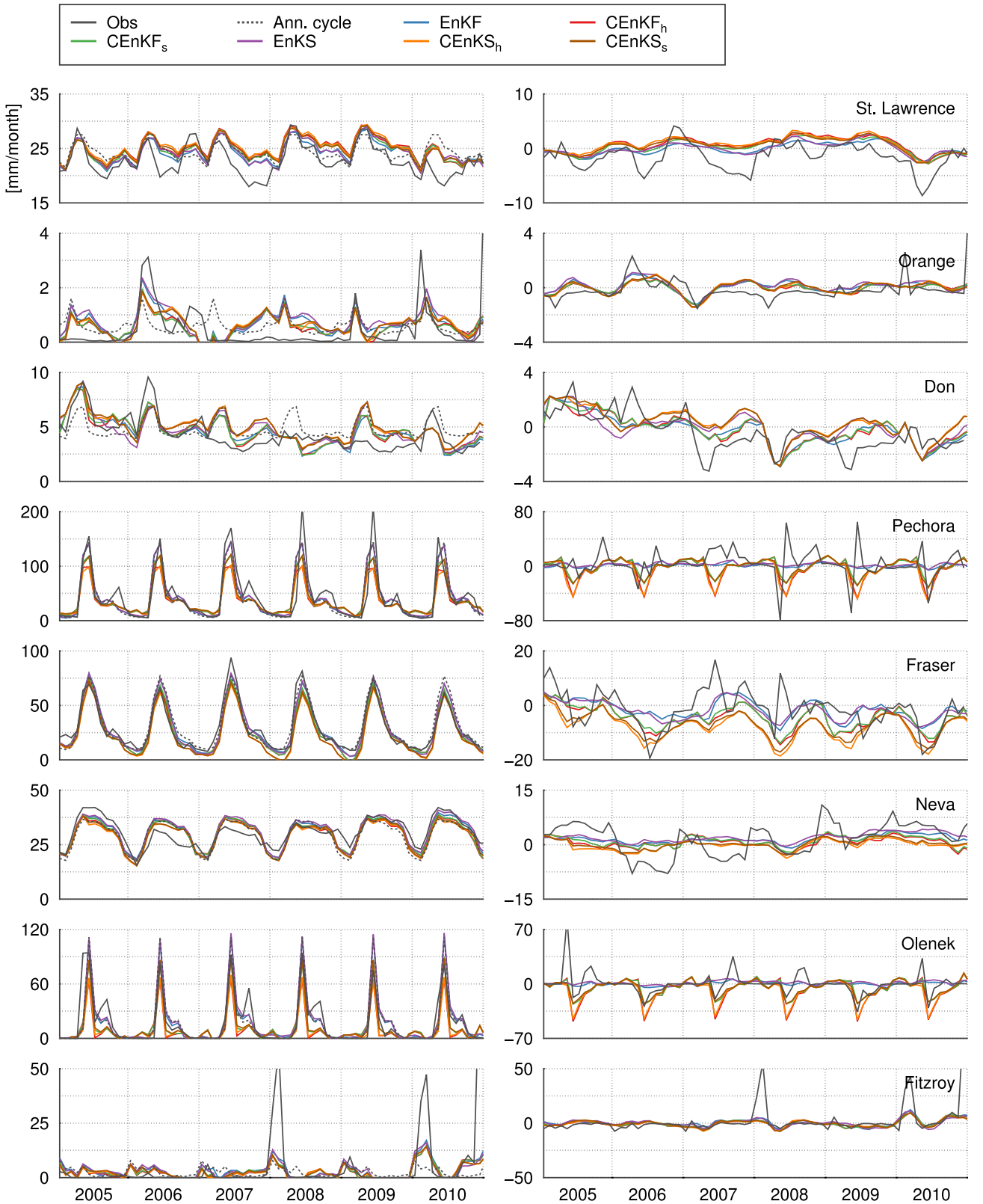


Figure 6. Time series of observed (black) and predicted runoff from different configurations of the Kalman Filter and Smoother for the study regions 9–16, showing the (left column) full runoff time-series and (right column) runoff anomalies, where the mean annual cycle from historic runoff data has been removed.

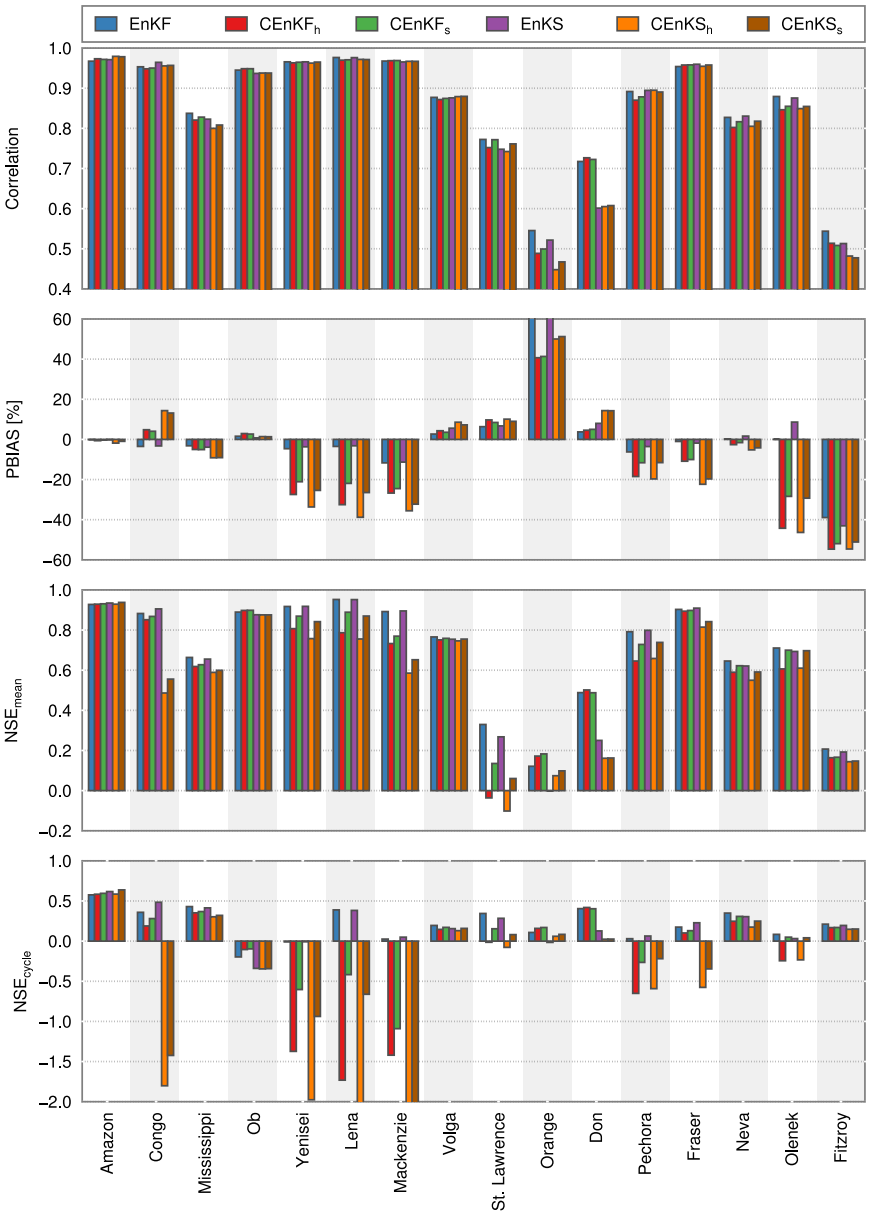


Figure 7. Performance metrics between observed and predicted runoff over all 16 study regions. The statistics are based on the period between 2005 and 2010. The colors indicate different configurations of the assimilation framework.

Fitzroy river in January 2008, February/March 2010 and December 2010/January 2011 (source: bom.gov.au). These were induced by heavy rain events, which can be also detected in the input precipitation data sets (not shown). The presented approach can therefore capture such extreme events, if these also occur in e.g., the precipitation time-series. This, however, must not always be the case: persistent rainfall during the end of the previous and the beginning of the next month can lead to a flooding event, even if (due to the monthly resolution) there are no exceptional features in the precipitation time-series.

When analyzing the performance of the different configurations, there is hardly any difference between the unconstrained, constrained, filtered, and smoothed estimates. The largest spread in terms of correlation can be observed over the Don basin (CEnKF$_h$ and CEnKF$_s$: 0.72, CEnKS$_h$ and CEnKS$_s$: 0.60), where the correlations of the filtered are significantly higher than the smoothed predictions. This holds also for the unconstrained and constrained estimates. Other significant differences can only be observed over the Fitzroy basin (EnKF: 0.54, CEnKS$_h$ and CEnKS$_s$: 0.48). For most catchments, however, the different configurations have only little impact on the dynamics of the estimated time-series.

Despite a slight underestimation of runoff over most catchments, the PBIAS values in Figure 7 show a different picture as there are significant differences between the metrics from the six set-ups. Even if there are many catchments where most configurations provide PBIAS values $< \pm$ 10% (Amazon, Congo, Mississippi, Ob, Volga, St. Lawrence, Don, Pechora, Fraser, and Neva), it is obvious that the constrained estimates show generally larger biases. This can be particularly observed over the high-latitude catchments Yenisei, Lena, Mackenzie, Don, Pechora, Fraser, Olenek, and Fitzroy, where the PBIAS values of the unconstrained and constrained estimates differ by more than 10%. An explanation for this can be a high uncertainty of the runoff estimates. When using such constraints within an assimilation framework, each variable in the state vector receives a correction, so that the water budgets are closed. This correction is based on the relative uncertainty of the respective variable. Thus, uncertain variables receive stronger corrections. The runoff estimates over the catchments showing large differences between the unconstrained and constrained predictions can therefore be assumed to be more uncertain than e.g., the predictions over the Amazon, Ob, or Volga basins. As a consequence, forcing the water budgets to be perfectly closed seems to be too strict. This assumption is supported by the better performance of the soft constrained estimates over most basins.

Over some of the catchments, however, the anomalies of the unconstrained estimates are small or almost zero throughout the whole period (Yenisei and Mackenzie), i.e., they are based mainly on the control input. When compared with observed runoff, the peaks in the unconstrained estimates are sometimes over- and sometimes underestimated. Thus, the errors cancel out, which leads to an overall small PBIAS value.

In contrast to most other catchments, the estimates over the Congo, Mackenzie, and Fraser basins show some distinct biases. For Mackenzie and Fraser, the constrained estimates significantly underestimate runoff over nearly the whole study period. The estimates from the $CEnKS_h$ and $CEnKS_s$ predict the lowest runoff over both basins. In these cases, constraining and smoothing results in a significant bias with respect to the observations. The same holds true for the Congo basin, where especially the estimates from $CEnKS_h$ and $CEnKS_s$ overestimate runoff.

In the hydrological community, NSE_{mean} values > 0.5 usually indicate an acceptable performance of runoff estimates [e.g., *Santhi et al*., 2001; *Moriasi et al*., 2007]. This is the case for 13 of the 16 estimated catchments. Several catchments even show NSE_{mean} values > 0.8 (Amazon, Congo, Ob, Yenisei, Lena, Mackenzie, Pechora, and Fraser). These are mainly basins with large amplitudes and a rather stable annual cycle in the runoff time-series. For such catchments, it is thus suggested to analyze the NSE_{cycle} metric (bottom row of Figure 7). This metric obviously shows significantly smaller values than the NSE_{mean}. Still, there are positive values over 14 of the 16 study regions. This inherently means that the estimates over these 14 catchments agree clearly better with the runoff observations than the values of the mean annual cycle.

It must be mentioned that over catchments with a less dominant annual cycle (St. Lawrence, Orange, Don, or Fitzroy) the annual cycle from historic data is obviously a bad predictor. Thus, positive NSE_{cycle} values can be expected even if the estimates show only good agreement during a limited period or for distinct peaks. In such cases, using the NSE_{mean} gives more information about the performance of an estimated time-series.

Significant improvements in terms of positive NSE_{cycle} values can be found over Amazon, Congo, Mississippi, Lena (EnKF and EnKS), Volga, St. Lawrence, Don, Neva, and Fitzroy. Even more than the classical NSE_{mean} metric, the NSE_{cycle} is very sensitive with respect to the agreement of extreme events and long-term climatic variations. From the anomaly time-series in Figures 5 and 6 it is obvious that the predictions successfully reproduce the long-term variations in the residual runoff observations over Amazon, Congo, Mississippi, Lena (EnKF and EnKS), Don, and Neva.

It becomes evident that even with high NSE_{mean} values over Ob, Yenisei, Lena, Mackenzie, Pechora, and Fraser, several configurations provide worse estimates than the mean annual cycle from historic data. Thus, the variability on top of the annual cycle, which is induced by the assimilation, does not improve the estimates. Ob seems to be a special case, as the anomalies of the estimates and the observations show similar long-term variations. In the end, the large differences, especially during 2005 and 2006, between the estimates and observations seem to result in a negative NSE_{cycle} value.

In the case of Yenisei, Lena, Pechora, Fraser, and Olenek, constraining the estimates dampens the maximum of runoff during the summer months. This can be identified by the negative peaks in the anomaly time-

series in Figures 5 and 6, where the hard constrained estimates from $CEnKF_h$ and $CEnKF_s$ assume the lowest runoff. Over the Olenek basin, using soft constraints actually leads to a better agreement, even if the secondary peaks during the late summer months are still missing. This is not the case for the Yenisei, Lena, Pechora, and Fraser basins, where the unconstrained estimates provide better results. Especially for the Lena basin, the long-term variability in the anomaly estimates from the EnKF and EnKS agrees well with the observations, even if the peaks in the anomalies are higher. A closer look reveals similar features over the Yenisei, Pechora, and Fraser basin. Over all these basins, the soft constrained estimates are able to capture some of the negative peaks in the residual time-series, but also show several unrealistic variations. Thus, using such constraints can lead to a better representation of short-term features, but the unconstrained estimates are able to capture the climatic long-term variations.

Interestingly, over all high-latitude catchments with a distinct snow accumulation phase (i.e., where runoff is close to zero), most configurations are able to reproduce the low runoff during the winter months. It can be assumed that these phases do not change drastically from year to year as the observations show only little deviations from the long-term annual cycle. Thus, during these phases, the estimates are mainly based on the control input (i.e., the mean annual cycle), which explains the good agreement with the observations.

5. Comparison to Similar Studies

In *Troy et al*. [2011], the water budgets of several basins over Northern Eurasia have been estimated using multiple data sources. They further estimate basin-scale runoff from a weighted water-balance approach, which is roughly comparable to our method. During their validation period between 1984 and 1999, they achieve NSE_{mean} values of 0.94 over the Lena and Yenisei and 0.89 over the Ob basin with respect to observed monthly runoff. Even if our performance metrics are based on a different period, we achieve similar values (Lena: 0.95 from EnKF and EnKS, Yenisei: 0.92 from EnKF and EnKS, and Ob: 0.90 from $CEnKF_h$).

Recently, *Gudmundsson and Seneviratne* [2014] published an approach for estimating gridded runoff over Europe. They also present a catchment-scale validation for several European river basins, including the Rhine and Elbe basin. These two basins are not included in our performance analysis due to missing data during the selected study period. However, when estimating runoff using our ensemble based approaches, we obtain NSE_{mean} values of 0.66 (Rhine, EnKF) and 0.62 (Elbe, $CEnKF_h$ and $CEnKF_s$), which is in good agreement with the findings from *Gudmundsson and Seneviratne* [2014].

In *Kang et al*. [2014], the Variable Infiltration Capacity (VIC) [*Liang et al*., 1994] model has been applied to the Fraser river basin. During their validation period from 1969 to 2006, the model estimated monthly runoff with respect to the observations from the gauging station at Hope with a NSE_{mean} of 0.85. In our study, the runoff estimates for the Fraser basin during the period 2005–2010 show a comparable performance ($NSE_{mean} = 0.90$ from EnKF, $CEnKF_s$, and EnKS).

Riegger and Tourian [2014] proposed a methodology, that is based on the characterization of the relationship between runoff and water storage changes. They simulated monthly runoff over five catchments, namely the Mackenzie ($NSE_{mean} = 0.93$), Ob (0.95), Lena (0.87), Yukon (0.81), and Yenisei (0.63) basins. Our approach is able to estimate runoff with a similar performance (Mackenzie: 0.90 from EnKF and EnKS, Ob: 0.90 from $CEnKF_h$, Lena: 0.95 from EnKF and EnKS, and Yenisei: 0.92 from EnKF and EnKS).

6. Conclusion

In this study, a data-assimilation framework is proposed, which can be employed for predicting and correcting catchment-scale time-series of runoff. One of the major design parameters of the framework is the use of as much real data as possible instead of using complex model equations. Thus, the approach is based on the terrestrial water budget equation only, which is included in an Ensemble Kalman Filtering framework. The prediction scheme predicts precipitation, evapotranspiration, runoff, and water storage changes using the so-called least squares prediction method. By this, we can exploit temporal and spatial covariance structures between different catchments and water cycle variables. For the observation equations of the dynamic process model, the most recent versions of widely used data sources for precipitation, evapotranspiration, runoff, and water storage changes are applied. We further use estimated runoff from satellite

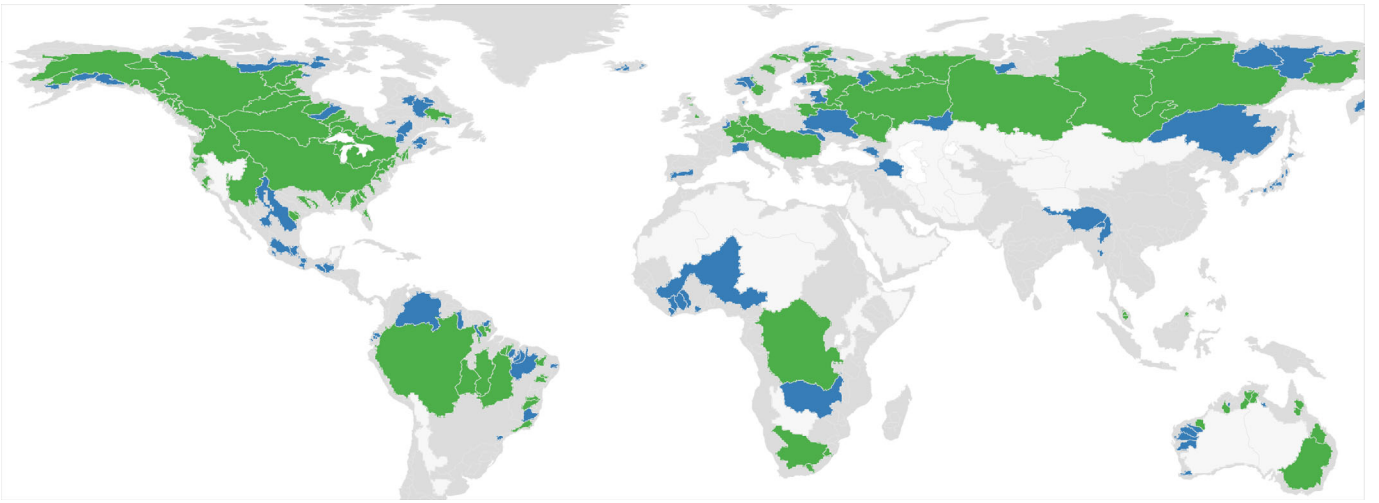


Figure 8. River basins for which runoff can be predicted using the Ensemble Kalman Filter approach based on global hydrometeorological data sets. For the green areas, more than 5 years of data after 2002 are available. The blue areas are (currently) poorly gauged basins with less than 5 years of data after 2002, but more than 5 years of continuous runoff measurements during the period 1980–2002. These catchments cover an area of more than 11,500,000 km^2 and provide a mean annual discharge volume of more than 1,25,000 m^3/s of freshwater resources.

altimetry in order to both fill the gaps in the time-series of gauge-based runoff observations but also to improve the predictions.

The filter is first run with different ensemble sizes and three different structures of the prediction matrix for finding a suitable set-up of the framework. The predictions from the filter are then compared with observed runoff from GRDC on a monthly basis. So far, the best results can be achieved by assuming correlations between water cycle variables only (case 3).

In order to enforce water budget closure between the estimated parameters, we further add appropriate water budget constraints to the framework. Therefore, it is distinguished between hard constraints, which assume perfect closure of the water budgets, and soft constraints, which allow a small well defined imbalance. In the end, the framework is run in six different configurations: the Ensemble Kalman Filter, the hard and soft Constrained Ensemble Kalman Filter, the Ensemble Kalman Smoother, and the hard and soft Constrained Ensemble Kalman Smoother.

In order to assess the performance of our data-assimilation approach, runoff is predicted over 16 large river basins and validated against in situ data. The performance analysis shows that the proposed method is able to estimate runoff with correlations larger than 0.8 for 12 of the 16 study regions. In terms of the PBIAS, values are less than $\pm$ 20% for most of the catchments (except for Orange and Fitzroy). Finally, the NSE_{mean} values are larger than 0.5 for 13 of the 16 catchments. As the runoff time-series of several catchments are dominated by a strong annual cycle, we further analyze the $NSE_{cycle,}$ which relates the performance of the predictions to the long-term mean annual cycle. For 14 of the 16 study basins, at least several configurations achieve NSE_{cycle} -values larger than 0. The time-series clearly show that the reason for this good performance is a promising agreement between the observed and predicted long-term variations in the runoff time-series.

The validation further shows that the different aforementioned configurations lead to similar time-series and performance metrics for most of the studied catchments. The different configurations thus have only a minor impact on the dynamics of the predicted time-series. As a general pattern, the constrained configurations seem to suffer from larger biases. This holds especially true for the Yenisei, Lena, Mackenzie, Pechora, Fraser, and Olenek basins. This results in rather negative NSE_{cycle} values. The soft constrained configurations provide slightly better results, which highlights the uncertainty in the data sources of water cycle variables. Enforcing a strict closure of water cycle from uncertain data might thus lead to numerical artifacts.

As a conclusion, the method is able to provide runoff estimates over the catchments shown in Figure 8, where only few or even no runoff observations are available after the year 2002. These catchments cover an

area of more than 11,500,000 km^2 and provide a discharge volume of more than 125,000 m^3/s of freshwater resources.

The performance analysis emphasizes the promising performance of the proposed method for predicting runoff. There is still room for improvement. Over Ob, Yenisei, Lena, Mackenzie, Pechora, and Fraser, several configurations provide worse predictions than the mean annual cycle from historic data, despite the high NSE_{mean} values. The reason for this has to be analyzed in more detail, as the bad performance of the constrained predictions indicates large inconsistencies in the data sources of the water cycle variables.

Furthermore, the predictions are not able to reliably reproduce extreme events in the runoff time-series as it is not appropriate to predict e.g., flooding events from monthly data. Also because the least squares prediction further exploits covariances, which describe the overall mean relationships between water cycle variables and catchments, Future studies must address an appropriate statistical characterization of exceptional events in the time-series of the water cycle variables.

Overall, the presented configurations of the data assimilation framework allow, despite their limitations and shortcomings, to both fill data gaps and extend the streamflow time-series for basins with discontinued observations. Thereby, the framework is an alternate option for predicting runoff of ungauged and poorly gauged basins.

Acknowledgments

The authors would like to thank the Helmholtz Alliance Remote Sensing and Earth System Dynamics and the priority program 1257 *Mass Transport and Mass Distribution in the System Earth* of the Deutsche Forschungsgemeinschaft DFG for funding this research. We would further like to thank the following data providers: Global Runoff Data Centre GRDC, Global Precipitation Climatology Project GPCP, Global Precipitation Climatology Centre GPCC, University of Delaware UDEL, Numerical Terradynamic Simulation Group at the University of Montana NTSG UMT, Max Planck Institute for Biogeochemistry MPI BGC, European Centre for Medium-Range Weather Forecasts ECMWF, University of Texas at Austin, Center for Space Research CSR, NASAS Goddard Earth Science Data and Information Services Center GES DISC, NOAAS Earth System Research Laboratory Physical Sciences Division ESRL PSD and NCARS Computational and Information Systems Laboratory Research Data Archive CISL rda. Further thanks are due to Diego Miralles (University of Bristol) for kindly providing the GLEAM evapotranspiration data. All graphics in the document were produced with the Generic Mapping Tools GMT [*Wessel and Smith*, 1991].

References

Adler, R. F., et al. (2003), The version-2 global precipitation climatology project GPCP monthly precipitation analysis (1979–present), *J. Hydrometeorol.*, *4*, 1147–1167, doi:10.1175/1525-7541(2003)004⟨1147:TVGPCP⟩2.0.CO;2.

AghaKouchak, A., A. Mehran, H. Norouzi, and A. Behrangi (2012), Systematic and random error components in satellite precipitation data sets, *Geophys. Res. Lett.*, *39*, L09406, doi:10.1029/2012GL051592.

Alsdorf, D. E., E. Rodríguez, and D. P. Lettenmaier (2007), Measuring surface water from space, *Rev. Geophys.*, *45*, RG2002, doi:10.1029/2006RG000197.

Berrisford, P., D. Dee, K. Fielding, M. Fuentes, P. Kallberg, S. Kobayashi, and S. Uppala (2009), *The ERA-Interim Archive Version 1.0*, *Era Rep. Ser. 1*, European Centre for Medium-Range Weather Forecasts, Reading, U. K. [Available at http://old.ecmwf.int/publications/library/do/references/list/782009.]

Bettadpur, S. (2012), UTCSR level-2 processing standards document for level-2 product release 0005, *Tech. Rep. Rev. 4.0*, Cent. for Space Res., Univ. of Tex. Austin, Tex.

Blöschl, G., M. Sivapalan, and T. Wagener (2013), *Runoff Prediction in Ungauged Basins: Synthesis Across Processes, Places and Scales*, Cambridge Univ. Press, Cambridge.

Dahle, C., F. Flechtner, C. Gruber, D. König, R. König, G. Michalak, and K.-H.Neumayer (2013), GFZ GRACE level-2 processing standards document for level-2 product release 0005, *Sci. Tech. Rep. 1.1*, GFZ Ger. Res. Cent. for Geosci., Potsdam.

Dai, A., and K. E. Trenberth (2002), Estimates of freshwater discharge from continents: Latitudinal and seasonal variations, *J. Hydrometeorol.*, *3*, 660–687, doi:10.1175/1525-7541(2002)003⟨0660:EOFDFC⟩2.0.CO;2.

Dee, D. P., et al. (2011), The ERA-Interim reanalysis: Configuration and performance of the data assimilation system, *Q. J. R. Meteorol. Soc.*, *137*, 553–597, doi:10.1002/qj.828.

Döll, P., F. Kaspar, and B. Lehner (2003), A global hydrological model for deriving water availability indicators: Model tuning and validation, *J. Hydrol.*, *270*, 105–134, doi:10.1016/S0022-1694(02)00283-4.

Döll, P., H. M. Schmied, C. Schuh, F. T. Portmann, and A. Eicker (2014), Global-scale assessment of groundwater depletion and related groundwater abstractions: Combining hydrological modeling with information from well observations and GRACE satellites, *Water Resour. Res.*, *50*, 5698–5720, doi:10.1002/2014WR015595.

Evensen, G. (1994), Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics, *J. Geophys. Res.*, *99*, 10,143–10,162, doi:10.1029/94JC00572.

Evensen, G. (2003), The Ensemble Kalman Filter: Theoretical formulation and practical implementation, *Ocean Dyn.*, *53*, 343–367, doi:10.1007/s10236-003-0036-9.

Falkenmark, M., and J. Rockström (2004), *Balancing Water for Humans and Nature: The New Approach in Ecohydrology*, Earthscan., London, U. K.

Fekete, B. M., U. Looser, A. Pietroniro, and R. D. Robarts (2012), Rationale for monitoring discharge on the ground, *J. Hydrometeorol.*, *13*, 1977–1986, doi:10.1175/JHM-D-11-0126.1.

Forootan, E., J. L. Awange, J. Kusche, B. Heck, and A. Eicker (2012), Independent patterns of water mass anomalies over Australia from satellite data and models, *Remote Sens. Environ.*, *124*, 427–443, doi:10.1016/j.rse.2012.05.023.

Gudmundsson, L., and S. I. Seneviratne (2014), Towards observation based gridded runoff estimates for Europe, *Hydrol. Earth Syst. Sci. Discuss.*, *11*, 12,883–12,932, doi:10.5194/hessd-11-12883-2014.

Gupta, H., S. Shrooshian, and P. O. Yapo (1999), Status of automatic calibration for hydrologic models: Comparison with multilevel expert calibration, *J. Hydrol. Eng.*, *4*, 135–143, doi:10.1061/(ASCE)1084-0699(1999)4:2(135).

Harris, I., P. D. Jones, T. J. Osborn, and D. H. Lister (2014), Updated high-resolution grids of monthly climatic observations—The CRU ts 3.10 Dataset, *Int. J. Climatol.*, *34*, 623–642, doi:10.1002/joc.3711.

Houtekamer, P. L., and H. L. Mitchell (1998), Data assimilation using an ensemble Kalman filter technique, *Mon. Weather Rev.*, *126*, 796–811, doi:10.1175/1520-0493(1998)126⟨0796:DAUAEK⟩2.0.CO;2.

Huffman, G. J., R. F. Adler, D. T. Bolvin, and G. Gu (2009), Improving the global precipitation record: GPCP Version 2.1, *Geophys. Res. Lett.*, *36*, L17808, doi:10.1029/2009GL040000.

Jones, P. (1999), First-and second-order conservative remapping schemes for grids in spherical coordinates, *Mon. Weather Rev.*, *127*, 2204–2210, doi:10.1175/1520-0493(1999)127⟨2204:FASOCR⟩2.0.CO;2.

Jung, M., M. Reichstein, and A. Bondeau (2009), Towards global empirical upscaling of FLUXNET eddy covariance observations: Validation of a model tree ensemble approach using a biosphere model, *Biogeosciences*, *6*, 2001–2013, doi:10.5194/bg-6-2001-2009.
Jung, M., et al. (2010), Recent decline in the global land evapotranspiration trend due to limited moisture supply, *Nature*, *467*, 951–954, doi:10.1038/nature09396.
Kang, D. H., X. Shi, H. Gaon, and S. J. Déry (2014), On the changing contribution of snow to the hydrology of the Fraser River Basin, Canada, *J. Hydrometeorol.*, *15*, 1344–1365, doi:10.1175/JHM-D-13-0120.1.
Khan, S. I., H. Yang, J. Wang, K. K. Yilmaz, J. J. Gourley, R. F. Adler, G. R. Brakenridge, F. Policell, S. Habib, and D. Irwin (2011), Satellite remote sensing and hydrologic modeling for flood inundation mapping in Lake Victoria Basin: Implications for hydrologic prediction in ungauged basins, *IEEE Trans. Geosci. Remote Sens.*, *49*, 85–95, doi:10.1109/TGRS.2010.2057513.
Kirtman, B. P., et al. (2013), The North American Multimodel Ensemble: Phase-1 seasonal-to-interannual prediction; Phase-2 toward developing intraseasonal prediction, *Bull. Am. Meteorol. Soc.*, *95*, 585–601, doi:10.1175/BAMS-D-12-00050.1.
Koblinsky, C. J., R. T. Clarke, A. C. Brenner, and H. Frey (1993), Measurement of river level variations with satellite altimetry, *Water Resour. Res.*, *29*, 1839–1848, doi:10.1029/93WR00542.
Kurtenbach, E., A. Eicker, T. Mayer-Gürr, M. Holschneider, M. Hayn, M. Fuhrmann, and J. Kusche (2012), Improved daily GRACE gravity field solutions using a Kalman smoother, *J. Geodyn.*, *59–60*, 39–48, doi:10.1016/j.jog.2012.02.006.
Kusche, J., R. Schmidt, S. Petrovic, and R. Rietbroek (2009), Decorrelated GRACE time-variable gravity solutions by GFZ, and their validation using a hydrological model, *J. Geod.*, *83*, 903–913, doi:10.1007/s00190-009-0308-3.
Liang, X., D. P. Lettenmaier, E. F. Wood, and S. J. Burges (1994), A simple hydrologically based model of land surface water and energy fluxes for general circulation models, *J. Geophys. Res.*, *99*, 14,415–14,428, doi:10.1029/94JD00483.
Lorenz, C., and H. Kunstmann (2012), The hydrological cycle in three state-of-the-art reanalyses: Intercomparison and performance analysis, *J. Hydrometeorol.*, *13*, 1397–1420, doi:10.1175/JHM-D-11-088.1.
Lorenz, C., B. Devaraju, M. J. Tourian, N. Sneeuw, J. Riegger, and H. Kunstmann (2014), Large-scale runoff from landmasses: A global assessment of the closure of the hydrological and atmospheric water balances, *J. Hydrometeorol.*, *15*, 2111–2139, doi:10.1175/JHM-D-13-0157.1.
Matsuura, K., and C. J. Willmott (2012), *Terrestrial Precipitation: 1900-2010 Gridded Monthly Time Series (Version 3.02)*, Cent. for Clim. Res., Univ. of Del., Del. [Available at climate.geog.udel.edu/∼climate/html_pages/Global2011/Precip_revised_3.02/.]
McMillan, H., J. Freer, F. Pappenberger, T. Krueger, and M. Clark (2010), Impacts of uncertain river flow data on rainfall-runoff model calibration and discharge predictions, *Hydrol. Processes*, *24*, 1270–1284, doi:10.1002/hyp.7587.
Milzow, C., P. E. Krogh, and P. Bauer-Gottwein (2011), Combining satellite radar altimetry, SAR surface soil moisture and GRACE total storage changes for hydrological model calibration in a large poorly gauged catchment, *Hydrol. Earth Syst. Sci.*, *15*, 1729–1743, doi:10.5194/hess-15-1729-2011.
Miralles, D. G., R. A. M. D. Jeu, J. H. Gash, T. R. H. Holmes, and A. J. Dolman (2011), An application of GLEAM to estimating global evaporation, *Hydrol. Earth Syst. Sci. Discuss.*, *8*, 1–27, doi:10.5194/hessd-8-1-2011.
Mitchell, H. L., P. L. Houtekamer, and G. Pellerin (2002), Ensemble size, balance, and model-error representation in an Ensemble Kalman Filter, *Mon. Weather Rev.*, *130*, 2791–2808, doi:10.1175/1520-0493(2002)130⟨2791:ESBAME⟩2.0.CO;2.
Moriasi, D. N., J. G. Arnold, M. W. V. Liew, R. L. Bingner, R. D. Harmel, and T. L. Veith (2007), Model evaluation guidelines for systematic quantification of accuracy in watershed simulations, *Trans. ASABE*, *50*, 885–900. [Available at swat.tamu.edu/media/90109/moriasimodeleval.pdf.]
Moritz, H. (1980), *Advanced Physical Geodesy*, Wichmann, Karlsruhe.
Mu, Q., 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, doi:10.1016/j.rse.2011.02.019.
Mu, Q., M. Zhao, and S. W. Running (2011), Improvements to a MODIS global terrestrial evapotranspiration algorithm, *Remote Sens. Environ.*, *115*, 1781–1800, doi:10.1016/j.rse.2011.02.019.
Mueller, B., et al. (2011), Evaluation of global observations-based evapotranspiration datasets and IPCC AR4 simulations, *Geophys. Res. Lett.*, *38*, L06492, doi:10.1029/2010GL046230.
Mueller, B., et al. (2013), Benchmark products for land evapotranspiration: LANDFLUX-EVAL multi-data set synthesis, *Hydrol. Earth Syst. Sci.*, *17*, 3707–3720, doi:10.5194/hess-17-3707-2013.
Nash, J. E., and J. V. Sutcliffe (1970), River flow forecasting through conceptual models: Part 1. A discussion of principles, *J. Hydrol.*, *10*, 282–290, doi:10.1016/0022-1694(70)90255-6.
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, doi:10.1175/JHM495.1.
Pan, M., 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. Clim.*, *25*, 3191–3206, doi:10.1175/JCLI-D-11-00300.1.
Postel, S. L., G. C. Daily, and P. R. Ehrlich (1996), Human appropriation of renewable fresh water, *Science*, *271*, 785–788, doi:10.1126/science.271.5250.785.
Ramillien, G., F. Frappart, A. Güntner, T. Ngo-Duc, A. Cazenave, and K. Laval (2006), Time variations of the regional evapotranspiration rate from Gravity Recovery and Climate Experiment (GRACE) satellite gravimetry, *Water Resour.Res.*, *42*, W10403, doi:10.1029/2005WR004331.
Reichle, R., D. B. McLaughlin, and D. Entekhabi (2002), Hydrologic data assimilation with the Ensemble Kalman Filter, *J. Hydrometeorol.*, *130*, 103–114, doi:10.1175/1520-0493(2002)130⟨0103:HDAWTE⟩2.0.CO;2.
Reichle, R. H., R. D. Kostera, G. J. M. D. Lannoy, B. A. Forman, Q. Liu, S. P. P. Mahanama, and A. Touré (2011), Assessment and enhancement of MERRA LAND surface hydrology estimates, *J. Clim.*, *24*, 6322–6338, doi:10.1175/JCLI-D-10-05033.1.
Riegger, J., and M. J. Tourian (2014), Characterization of runoff-storage relationships by satellite gravimetry and remote sensing, *Water Resour. Res.*, *50*, 3444–3466, doi:10.1002/2013WR013847.
Riegger, J., M. J. Tourian, B. Devaraju, and N. Sneeuw (2012), Analysis of GRACE uncertainties by hydrological and hydro-meteorological observations, *J. Geodyn.*, *59*, 16–27, doi:10.1016/j.jog.2012.02.001.
Rodell, M., and J. Famiglietti (1999), Detectability of variations in continental water storage from satellite observation of time dependent gravity field, *Water Resour. Res.*, *35*, 2705–2723, doi:10.1029/1999WR900141.
Rodell, M., J. Famiglietti, J. Chen, S. I. Seneviratne, P. Viterbo, S. Holl, and C. R. Wilson (2004a), Basin scale estimates of evapotranspiration using GRACE and other observations, *Geophys. Res. Lett.*, *31*, L20504, doi:10.1029/2004GL020873.
Rodell, M., et al. (2004b), The global land data assimilation system, *Bull. Am. Meteorol. Soc.*, *85*, 381–394, doi:10.1175/BAMS-85-3-381.
Santhi, C., J. G. Arnold, J. R. Williams, W. A. Dugas, R. Srinivasan, and L. M. Hauck (2001), Validation of the SWAT model on a large river basin with point and nonpoint sources, *J. Am. Water Resour. Assoc.*, *37*, 1169–1188, doi:10.1111/j.1752-1688.2001.tb03630.x.
Schmidt, R., et al. (2006), GRACE observations of changes in continental water storage, *Global Planet. Change*, *50*, 112–126, doi:10.1016/j.gloplacha. 2004.11.018.

Schneider, U., T. Fuchs, A. Meyer-Christoffer, and B. Rudolf (2008), Global precipitation analysis products of the GPCC, Deutscher Wetterdienst, Global Precipitation Climatology Centre, Deutscher Wetterdienst, Offenbach, Germany, 13 pp. [Available at http://gpcc.dwd.de.]

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, doi:10.1029/2009GL037338.

Shiklomanov, A. I., R. B. Lammers, and C. J. Vörösmarty (2002), Widespread decline in hydrological monitoring threatens pan-Arctic research, *Eos Trans. AGU*, *83*, 13–17, doi:10.1029/2002EO000007.

Simmons, A., S. Uppala, D. Dee, and S. Kobayashi (2006), ERA-Interim: New ECMWF reanalysis products from 1989 onwards, *ECMWF Newsl.*, *110*, 25–35. [Available at www.ecmwf.int.]

Simon, D. (2010), Kalman filtering with state constraints: A survey of linear and nonlinear algorithms, *IET Control Theory Appl.*, *4*, 1303–1318, doi:10.1049/iet-cta.2009. 0032.

Simon, D., and T. Chia (2002), Kalman filtering with state constraints, *IEEE Trans. Aerosp. Electron. Syst.*, *39*, 128–136, doi:10.1109/7.993234.

Sivapalan, M., et al. (2003), IAHS decade on Predictions in Ungauged Basins (PUB), 2003–2012: Shaping an exciting future for the hydrological sciences, *Hydrol. Sci. J.*, *48*, 857–880, doi:10.1623/hysj.48.6.857.51421.

Sneeuw, N., C. Lorenz, B. Devaraju, M. J. Tourian, J. Riegger, H. Kunstmann, and A. Bárdossy (2014), Estimating runoff using hydro-geodetic approaches, *Surv. Geophys.*, *4*, 1303–1318, doi:10.1007/s10712-014-9300-4.

Swenson, S., and J. Wahr (2006), Estimating large-scale precipitation minus evapotranspiration from GRACE satellite gravity measurements, *J. Hydrometeorol.*, *7*, 252–270, doi:10.1175/JHM478.1.

Syed, T. H., J. S. Famiglietti, J. Chen, M. Rodell, S. I. Seneviratne, P. Viterbo, and C. R. Wilson (2005), Total basin discharge for the Amazon and Mississippi River basins from GRACE and a land-atmosphere water balance, *Geophys. Res. Lett.*, *32*, L24404, doi:10.1029/2005GL024851.

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, doi:10.1029/2007GL031254.

Syed, T. H., J. S. Famiglietti, and D. P. Chambers (2009), GRACE-based estimates of terrestrial freshwater discharge from basin to continental scales, *J. Hydrometeorol.*, *10*, 22–40, doi:10.1175/2008JHM993.1.

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, doi:10.1126/science.1099192.

Tourian, M. J. (2013), Application of spaceborne geodetic sensors for hydrology, PhD thesis, Institute of Geodesy, Univ. Stuttgart, Stuttgart, Germany.

Tourian, M. J., N. Sneeuw, and A. Bárdossy (2013), A quantile function approach to discharge estimation from satellite altimetry (ENVISAT), *Water Resour. Res.*, *49*, 1–13, doi:10.1002/wrcr.20348.

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. Clim.*, *24*, 3272–3293, doi:10.1175/2011JCLI3936.1.

Vörösmarty, C. J., et al. (2010), Global threats to human water security and river biodiversity, *Nature*, *467*, 555–561, doi:10.1038/nature09440.

Wahr, J., M. Molenaar, and F. Bryan (1998), The time-variability of the Earth's gravity field: Hydrological and oceanic effects and their possible detection using GRACE, *J. Geophys. Res.*, *103*, 30,205–30,230, doi:10.1029/98JB02844.

Wessel, P., and W. H. F. Smith (1991), Free software helps map and display data, *Eos Trans. AGU*, *72*, 441–446, doi:10.1029/90EO00319.