https://doi.org/10.5194/hess-2023-215
Preprint. Discussion started: 12 September 2023

Hydrology and Earth System Sciences
Discussions

Use of satellite remote sensing to validate reservoir operations in global hydrological models: a case study from the CONUS

Kedar Otta[1], Hannes Müller Schmied[2,3], Simon N. Gosling[4], Naota Hanasaki[1,5]

[1]National Institute for Environmental Studies, Tsukuba, 305-8506, Japan
[2]Institute of Physical Geography, Goethe University Frankfurt, 60438 Frankfurt am Main, Germany
[3]Senckenberg Leibniz Biodiversity and Climate Research Centre (SBiK-F), 60325 Frankfurt am Main, Germany
[4]School of Geography, University of Nottingham, Nottingham NG7 2RD, United Kingdom
[5]The University of Tokyo, Tokyo, 113-0033, Japan

Correspondence to: Naota Hanasaki (hanasaki@nies.go.jp)

Abstract

Although river discharge simulations from global hydrological models have undergone extensive validation, there has been less validation of reservoir operations, primarily because of limited observational data. However, recent advancements in satellite remote sensing technology have facilitated the collection of valuable data regarding water surface area and elevation, thereby providing the ability to validate reservoir storage. In this study, we sought to establish a methodology for validation and intercomparison of reservoir storage within global hydrological model simulations using satellite-derived data. Accordingly, we chose two satellite-derived reservoir operation products, DAHITI and GRSAD, to create monthly time series storage data for seven reservoirs in the contiguous United States (CONUS) , with access to long-term ground truth data (the total catchment area accounts for about 9% of CONUS). We assessed two global hydrological models that participated in the Inter Sectoral Model Intercomparison Project (ISIMIP) Phase 3 project, H08 and WaterGAP2, with three distinct forcing datasets: GSWP3-W5E5 (GW), CR20v3-W5E5 (CW), and CR20v3-ERA5 (CE). The results indicated that WaterGAP2 generally outperforms H08; the CW forcing dataset demonstrated superior results compared with GW and CE; the DAHITI showed better consistency with ground observations than GRSAD if temporal

https://doi.org/10.5194/hess-2023-215
Preprint. Discussion started: 12 September 2023


Hydrology and Earth System Sciences
Discussions

coverage is sufficient. Overall, our study emphasizes the potential uses of satellite remote sensing data in reservoir operations validation and underscores the importance of normalization and decomposition techniques for improved validation efficacy. The results highlight the relative performances of different hydrological models and forcing datasets, yielding insights concerning future advancements in reservoir simulation and operational studies.

1 Introduction

Artificial reservoirs play an integral role in the hydrological cycle and water resource management (Grill et al., 2019). Significant reservoirs have been incorporated into global hydrological models (GHMs) such as H08 (Hanasaki et al., 2008b, 2018), WaterGAP (Döll et al., 2009), LPJmL (Biemans et al., 2011), PCR-GLOBWB (Wada et al., 2011), and CWatM (Burek et al., 2020). The complex, condition-dependent nature of reservoir operations (i.e., the storage and release of upstream water for downstream advantages) has led to the development and implementation of several algorithms into GHMs (Haddeland et al., 2006; Hanasaki et al., 2006). Consequently, validation and intercomparison of reservoir operations for these models are particularly important.

When new reservoir operation algorithms were introduced (Haddeland et al., 2006; Hanasaki et al., 2006) and subsequently implemented into GHMs (Hanasaki et al., 2008a, 2008b, 2018; Döll et al., 2009; Biemans et al., 2011; Wada et al., 2011; Burek et al., 2020; Pokhrel et al., 2012), they were validated using in situ observations. Various model intercomparison projects have enabled an understanding of the relative benefits of specific models and algorithms. The Inter-Sectoral Impact Model Intercomparison Project (ISIMIP; Warszawski et al., 2014) facilitates the intercomparison of numerous GHMs under uniform simulation settings. Several studies have validated and intercompared hydrological variables such as river discharge (Zaherpour et al., 2018; Kumar et al., 2022), irrigation water demand (Wada et al., 2013), and terrestrial water storage (Pokhrel et al., 2021). The first reservoir operation intercomparison was conducted by Masaki et al. (2017), who examined the effects of dams on simulated river discharge. They found substantial variations in model simulations, but their research was restricted to the Green-Colorado River and the Missouri Mississippi River basins due to the global unavailability of observed gauged data.

https://doi.org/10.5194/hess-2023-215
Preprint. Discussion started: 12 September 2023


Hydrology and Earth System Sciences
Discussions

Prior validations and intercomparisons of reservoir operations in multiple GHMs have been restricted to regions with extensive gauge observation records, such as the Green-Colorado and the Missouri-Mississippi Rivers (Masaki et al. 2017). Thus, it is challenging to assess the performance of GHMs in simulating reservoir operations and identifying superior models. Concurrently, satellite remote sensing has emerged as a valuable tool for global validation, irrespective of geographical location (Alsdorf and Lettenmaier, 2003). A few studies have developed methodologies to determine reservoir storage using satellite-derived altimetry and surface area data (Busker et al., 2019; Gao et al., 2012). However, there remains a need to establish a method for utilization of the latest satellite-based datasets for GHM validation.

This study was conducted to assess the feasibility of using satellite data to evaluate the performances of reservoir operation simulations in a multi-model and multi-forcing framework. We focused on seven strategically selected reservoirs across the Contiguous United States (CONUS) region. In accordance with the methodology developed by Gao et al. (2012) and Busker et al. (2019), we used remote sensing altimetry and surface area products to determine reservoir storage and its components in preparation for satellite-based reservoir storage observations. We used the outputs of two GHMs that participated in ISIMIP Phase 3a (Frieler et al., 2023) for stylized simulations of reservoir operations. Our research questions were as follows:

1. Can satellite-based storage estimation data serve as a surrogate for ground truth data?
2. Can we determine which GHMs or meteorological forcings perform better than others in model intercomparison projects, solely by satellite-based storage estimation?
3. Do the findings on reservoir storage validation with satellite data align with ground observations?
4. Are certain satellite products superior to others?

To address these questions, we developed a comprehensive framework that thoroughly validated and intercompared multi-model and multi-forcing simulated reservoir storage with available satellite and ground observations.

2 Materials and Methods

https://doi.org/10.5194/hess-2023-215
Preprint. Discussion started: 12 September 2023


Hydrology and Earth System Sciences
Discussions

2.1 Simulation data

2.1.1 ISIMIP Phase 3a

The third-round framework of the ISIMIP, Phase 3a, is focused on the evaluation and enhancement of impact models within the context of climate change (Frieler et al., 2023). As of June 9, 2023, nine models have participated in the global water sector, but only a few have completed simulations that include reservoir outputs. In this study, we utilized two GHMs, namely H08 and WaterGAP2 (WGP), on the basis of three meteorological forcings that are bias-adjusted combinations of two reanalyses, beginning in 1979 for W5E5 and ERA5, respectively (Lange et al., 2022):

1. GSWP3 combined with W5E5 (GSWP3+W5E5, hereafter referred to as GW)
2. 20CRv3 combined with W5E5 (20CRv3+W5E5, hereafter referred to as CW)
3. 20CRv3 combined with ERA5 (20CRv3+ERA5, hereafter referred to as CE)

These forcing data are globally available at 0.5° × 0.5° spatial resolution at daily intervals. The combination of the two models and three forcings yields six model simulations, with model and forcing names combined (e.g., H08 forced by GW results in H08_GW). Additional details regarding the simulation protocol can be found at https://protocol.isimip.org/ and in the work by Frieler et al. (2023).

2.1.2 H08 model

The H08 model is a grid-cell-based GHM designed to address the impacts of human activities on the global hydrological cycle. H08 comprises six sub-models: land surface hydrology, river routing, reservoir operation, crop growth, environmental flow, and anthropogenic water withdrawal (Hanasaki et al., 2008a, b). The model was subsequently updated to include groundwater recharge and abstraction, aqueduct water transfer, local reservoir, seawater desalination, and return flow and delivery loss schemes (Hanasaki et al., 2018). By incorporating these submodules and schemes, H08 simulates natural and anthropogenic hydrological processes at a spatial resolution of 0.5° on a daily scale by resolving water and energy balance. Specifically, H08 includes explicit flow regulation of 963 major global reservoirs. The modeling of release from the reservoir is based on the work of Hanasaki et al. (2006). Reservoirs primarily used for irrigation are classified accordingly; all other reservoirs are considered non-irrigation reservoirs. The water demand for irrigation reservoirs is presumably more affected by the seasonal cycle due to the

https://doi.org/10.5194/hess-2023-215
Preprint. Discussion started: 12 September 2023


Hydrology and Earth System Sciences
Discussions

seasonal nature of irrigation water requirements. Land surface parameters were optimized based on climatic zones using the method proposed by Yoshida et al. (2022).

2.1.3 WaterGAP model

WaterGAP (WGP) is a GHM that comprises two primary components (Müller Schmied et al., 2021). The WGP Global Water Use Models calculate water use estimates for five sectors: irrigation, domestic, manufacturing, cooling water for electricity generation, and livestock. In contrast, the WGP Global Hydrology Model uses water balance equations to calculate changes in water storage compartments and water flows between them. It considers fluxes such as groundwater recharge, evapotranspiration, and river discharge, along with net abstractions from surface water and groundwater, as calculated in a linking module from the sectoral water use models. Its calculations are performed with a daily time step. The reservoir operation has been described by Döll et al. (2009) and Müller Schmied et al. (2021). The reservoir algorithm follows the method of Hanasaki et al. (2006), differentiating between reservoirs used for irrigation and other purposes, and considering both reservoirs and regulated lakes. Contrary to the method of Hanasaki et al. (2006), the annual release from a reservoir also depends on the long-term average mean streamflow of the grid cell where the reservoir is located, considering the water balance of the reservoir. In the model version used in ISIMIP3 (WaterGAP 2.2e), 1255 "global" reservoirs with storage volumes of ≥ 0.5 km^3 and 5722 "local" reservoirs (with smaller storage volumes) are included. However, only the global reservoirs are managed with the reservoir algorithm.

The primary aim of WGP is the provision of reliable estimates of renewable water resources on a global scale. To accommodate uncertainties in GHMs, a calibration routine is applied in WGP. This calibration ensures that the long-term annual simulated river discharge closely matches observed discharge within a ± 10% tolerance at grid cells representing calibration stations. Calibration is performed using observed discharge data from a selection of 1509 discharge observation stations, which have been collated from three data sources (Müller Schmied and Schiebener, 2022).

2.2 Reservoir specification data

https://doi.org/10.5194/hess-2023-215
Preprint. Discussion started: 12 September 2023


Hydrology and Earth System Sciences
Discussions

We adopted reservoir parameters such as dam name, location (longitude and latitude), storage capacity (Sc), and maximum surface area (Ac) from data provided in the ISIMIP3a protocol. These data are primarily obtained from the Global Reservoir and Dam Database (GRanD) v1.3, which was developed by the Global Water System Project (Lehner et al., 2011). The data also include a set of dams provided by Dr. Jida Wang from Kansas State University. This collaboration has resulted in a comprehensive database of 7330 dams, either constructed or under construction, spanning the years 286 to 2020. The cumulative global storage capacity of the database is approximately 7000.5 km^3.

2.3 Ground observation data

Reservoir storage is always precisely monitored by dam operators, but the long-term time series are seldom published openly. This has been the primary obstacle in global reservoir modeling and analysis in the past. ResOpsUS (Steyaert et al., 2022) is an exhaustive dataset containing historical information about reservoir inflows, outflows, and storage time series for 679 major reservoirs across the United States. The data, with daily temporal resolution, enable detailed analysis of reservoir dynamics. However, the temporal coverage varies among reservoirs based on factors such as construction date and data availability. The dataset spans the years from 1930 to 2020, with the most robust data for the period from 1980 to 2020. Notably, reservoirs in the dataset contain more than half of the total storage capacity of large reservoirs in the U.S., with a minimum storage threshold of 0.1 km^3.

2.4 Satellite data

2.4.1 DAHITI (Surface water level time series)

The Database for Hydrological Time Series over Inland Waters (DAHITI) is a web service that offers valuable information on water levels, surface area, and volume variations in rivers, lakes, and reservoirs (Schwatke et al., 2015; 2019; Busker et al., 2019). DAHITI uses satellite altimetry technology to measure water levels in inland bodies, extending beyond its initial application in sea-level monitoring. The methodology utilizes an amalgamation of extended outlier rejection, a Kalman filter, and cross-calibrated multi-mission altimeter data. These data are collected from satellites such as Envisat, ERS-2, Jason-1, Jason-2, TOPEX/Poseidon, and SARAL/AltiKa, considering their respective uncertainties. This

https://doi.org/10.5194/hess-2023-215
Preprint. Discussion started: 12 September 2023


Hydrology and Earth System Sciences
Discussions

comprehensive approach facilitates more accurate estimation of water level time series (Schwatke et al., 2015). The data are available from 1992 to present. Temporal resolution varied within the period, but in this study, a simple monthly mean of the available data was considered.

In addition to water levels, DAHITI provides surface area time series for lakes and reservoirs, utilizing optical imagery (Schwatke et al., 2019). The temporal resolution of these measurements varies according to the imagery used; for instance, Landsat provides data every 16 days, whereas Sentinel-2 provides data every 10 days. Different satellite missions can be merged to further enhance this resolution. These surface area time series are processed using a blend of 10-m (Sentinel-2) and 30-m (Landsat) spatial resolution imagery (Schwatke et al., 2019). The methodology initially designates a broad area of interest. Subsequently, it combines five remote-sensing-based water indices to compute the water mask, based on land-water differentiation. The final step involves extraction of the water surface area time series (Schwatke et al., 2019). In this study, we used altimetry data from DAHITI to estimate reservoir storage, primarily because water surface area data were not universally available for all reservoirs under consideration.

2.4.2 GRSAD (Surface area time series)

The Global Reservoir Surface Area Dataset (GRSAD) is a creation of Zhao and Gao (2018) and Gao and Zhao (2020). This dataset provides a monthly time series of water surface area data for 6,817 reservoirs worldwide, collectively representing a storage capacity of 6,099 km^3 (Zhao and Gao, 2018). The time frame of this dataset ranges from 1984 to 2015.

GRSAD builds upon the earlier work of Pekel et al. (2016); it includes automatic corrections for disruptions caused by clouds, cloud shadows, and terrain shadows. The determination of maximum surface area extent is based on a 500-m outward extension from GRanD shapefiles (Lehner et al., 2011). These shapefiles are developed from Shuttle Radar Topography Mission (SRTM; Jarvis et al., 2008) data, identifying water regions as flat zones with uniform elevation. As a result, any surface area beyond the 500-m threshold is not considered part of the reservoir.

https://doi.org/10.5194/hess-2023-215
Preprint. Discussion started: 12 September 2023

The dataset primarily uses 30-m Landsat satellite imagery; it does not incorporate data from other satellite sources. Although it provides extensive information regarding reservoir surface areas, it does not offer altimetry or volumetric change data.

2.4.3 GRBD (Bathymetry)

The Global Reservoir Bathymetry Dataset (GRBD) constitutes another category of satellite product (Li et al., 2020). This dataset utilizes multi-source satellite imagery and altimetry data to create detailed bathymetry information for 347 reservoirs worldwide, representing approximately 50% of the global storage capacity. In addition to bathymetry data, GRBD offers valuable relationships such as Area-Elevation (A-h, refer to Section 2.6) and key reservoir parameters such as Sc.

2.5 Reservoir selection

The process of identifying common reservoirs across H08, WGP, ResOpsUS, GRSAD, and GRBD datasets is streamlined by the shared use of the GRanD ID. This sharing facilitates integration and comparison of data across the different datasets. However, DAHITI uses a unique identification system, thereby requiring individual examination of each reservoir for data availability. Accordingly, a meticulous selection procedure was conducted. First, common reservoirs among H08, WGP, ResOpsUS, GRSAD, and GRBD were identified. Then they were searched on the DAHITI website. After a comprehensive review, only seven reservoirs listed in Table 1 were found in all datasets and were thus selected as the foundation for analysis. This considerable shrink in the number of reservoirs is attributed to the availability of ground observation data (i.e. data are available in the CONUS only). The locations of these reservoirs in the H08 and WGP models within the 0.5° × 0.5° grids are indicated in Table S1. Table S2 displays the Sc utilized in our study for these reservoirs. Identifying common reservoirs to all datasets is a prerequisite for this study, which evaluates the agreement of satellite products and the performance of GHMs.

https://doi.org/10.5194/hess-2023-215
Preprint. Discussion started: 12 September 2023

Hydrology and Earth System Sciences
Discussions

Table 1: Specifications of dams and corresponding reservoirs considered in this study. Year corresponds to the initial year of reservoir operation, H_{dam} corresponds to dam height, and A_c corresponds to the maximum water surface area of the reservoir. Longitude and latitude indicate the location of the dams.

Dam name	Lake name	GRanD ID	Hydro lake ID	DAHITI ID	River	Lon	Lat	Year	H_{dam} (m)	A_c (km^2)	Main purpose
Hoover	Lake Mead	610	809	204	Colorado River	-114.74	36.02	1935	223	580.95	Water supply
Glen Canyon	Lake Powell	597	802	107	Colorado River	-111.49	36.94	1963	216	120.75	Hydro-electricity
Fort Peck	Fort Peck Lake	307	721	11112	Missouri River	-106.41	48.00	1957	78	814.09	Flood control
Toledo Bend	Toledo Bend Lake	1269	838	10247	Sabine River	-93.57	31.17	1966	34	599.62	Hydro-electricity
Structure 193	Lake Okeechobee	1957	69	57	Taylor Creek	-81.10	26.94	1972	11	1418.77	Flood control
Wesley E. Seale	Lake Corpus Christi	1317	9615	13139	Nueces River	-97.87	28.05	1958	25	59.14	Recreation
Coolidge	San Carlos Lake	656	9440	13130	Gila River	-110.52	33.18	1929	77	15.47	Irrigation

https://doi.org/10.5194/hess-2023-215
Preprint. Discussion started: 12 September 2023

Hydrology and Earth System Sciences
Discussions

2.6 Reservoir storage calculation from satellite data

For most reservoirs, as depicted in Figure 1, there is a linear relationship between the water surface area (A) and the water surface elevation (h), represented as:

$$h = a \times A + b, \quad (1)$$

where a and b represent the slope and intercept, respectively, obtained from a linear regression (Gao et al., 2012; Busker et al., 2019). These parameters are supplied by the GRBD dataset in our study (refer to Table S2).

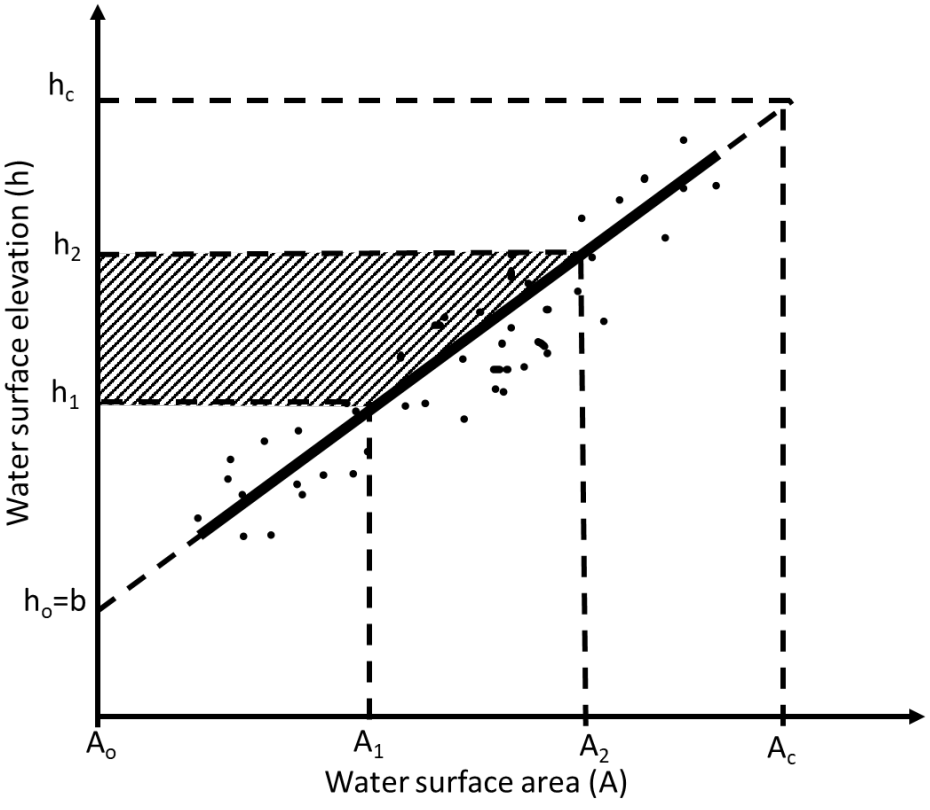


Figure 1: Relationship between water surface area (A), water surface elevation (h), and the change in storage volume (ΔS).

The volume change for a particular period is then calculated as the area of a trapezoid, as described by Gao et al. (2012):

$$\Delta S = \frac{(h2 - h1) \cdot (A1 + A2)}{2} \quad (2)$$

https://doi.org/10.5194/hess-2023-215
Preprint. Discussion started: 12 September 2023


Hydrology and Earth System Sciences
Discussions

where ΔS represents the volume change, A1 and A2 are surface areas at the start and end of the period, and h1 and h2 are their respective water surface elevations.

Gao et al. (2012) extended the linear relationship to Sc of the reservoir, resulting in the expression of the corresponding maximum surface area (Ac) and maximum water surface elevation (hc) as:

$$Si = Sc - \frac{(hc - hi) \cdot (Ac + Ai)}{2}, \quad (3)$$

where Si represents the volume of water stored in the reservoir, corresponding to water surface area Ai and water surface elevation hi. However, because hc is unknown, the storage estimation from GRSAD data is computed using the linear equation described in eq. 1, as follows:

$$hc = a \times Ac + b \quad (4)$$

Busker's Method (Busker et al., 2019) extends the linear relationship toward the minimum storage (i.e., zero storage); thus, the corresponding surface area is also zero, and the water surface elevation is the bed elevation of the reservoir.

$$Si = \frac{(hi - b) \cdot Ai}{2}, \quad (5)$$

Busker's method requires fewer parameters; Ac and hc are unnecessary. Additionally, the storage volume can be computed using only h or A by substituting the linear A-h relationship (eq. 1):

$$Si = \frac{(hi-b)^2}{2a} = \frac{a.(Ai)^2}{2} \quad (6)$$

Equation 6 is applicable for GRSAD and DAHITI datasets, which contain time series of both surface area and elevation.

2.7 Analysis

2.7.1 Outline

Initially, satellite data were compared with ground observations to determine compatibility with evaluations of model simulations. Subsequently, simulated reservoir storage from the two ISIMIP3a models, H08 and WaterGAP, was validated against two satellite datasets, GRSAD and DAHITI. Reservoir storage data were examined in three forms: raw, normalized, and decomposed. The methodologies for normalization and decomposition are described below. Refer to Table 2 for the data utilized.

https://doi.org/10.5194/hess-2023-215
Preprint. Discussion started: 12 September 2023

Hydrology and Earth System Sciences
Discussions

Table 2. Reservoir storage data used in this study.

Category	Name/ Description	Acronym	
GHMs	H08	H08	
	WaterGAP2.2e	WGP	
Input forcings	GSWP3+W5E5	GW	
	20CRv3+W5E5	CW	
	20CRv3+ERA5	CE	
Simulations	e.g., H08 forced by GW	H08_GW	
Ground observation	ResOpsUS	Grd_obs	
Reservoir volume from satellite data		**Raw Storage**	**Normalized Storage**
	GRSAD area + Sc from GRBD + Gao's Method	GRSAD_GRBD	GRSAD
	GRSAD area + Sc from ISIMIP + Gao's Method	GRSAD_ISIMIP	GRSAD
	GRSAD area + Busker's Method [Sc not needed]	GRSAD_Busker	GRSAD
	DAHITI elevation + Busker's method [Sc not needed]	DAHITI_Busker	DAHITI

2.7.2 Normalization

We normalized the monthly storage time series using the following equation:

$$S_{norm,i} = \frac{\mathrm{Si}-\mathrm{min(S)}}{\mathrm{max(S)}-\mathrm{min(S)}}, \qquad (7)$$

where min(s) and max(s) represent the minimum and maximum values of the available monthly storage time series, respectively. Different datasets and methods utilized to calculate S with GRSAD lead to the same normalized storage (Table 2) because the resulting volumes are linearly proportional to each other. Thus, by normalizing the monthly storage time series, information about the absolute value of reservoir storage is omitted; only the rate of change information is retained.

2.7.3 Decomposition

https://doi.org/10.5194/hess-2023-215
Preprint. Discussion started: 12 September 2023

Hydrology and Earth System Sciences
Discussions

The monthly storage time series ($S_{y,m}$) was decomposed (Figure 2) into annual average storage (S_y), mean annual seasonal variability (hereafter referred to as seasonal variability or $\bar{S}_m$), and residuals ($e_{y,m}$), as follows:

$$S_{y,m} = S_y + \bar{S}_m + e_{y,m}\,, \qquad (8)$$

S_y denotes the annual average storage volume for a reservoir from January to December, computed by averaging the 12 monthly storage values. $\bar{S}_m$ is determined by calculating the mean storage value for each month after subtracting the mean annual storage for that specific year; thus, it represents storage fluctuation within a year due to seasonal factors. $e_{y,m}$ constitutes the residual storage value after removing both the annual average storage and the seasonal variability, thus representing the storage component not attributable to annual or seasonal variations.

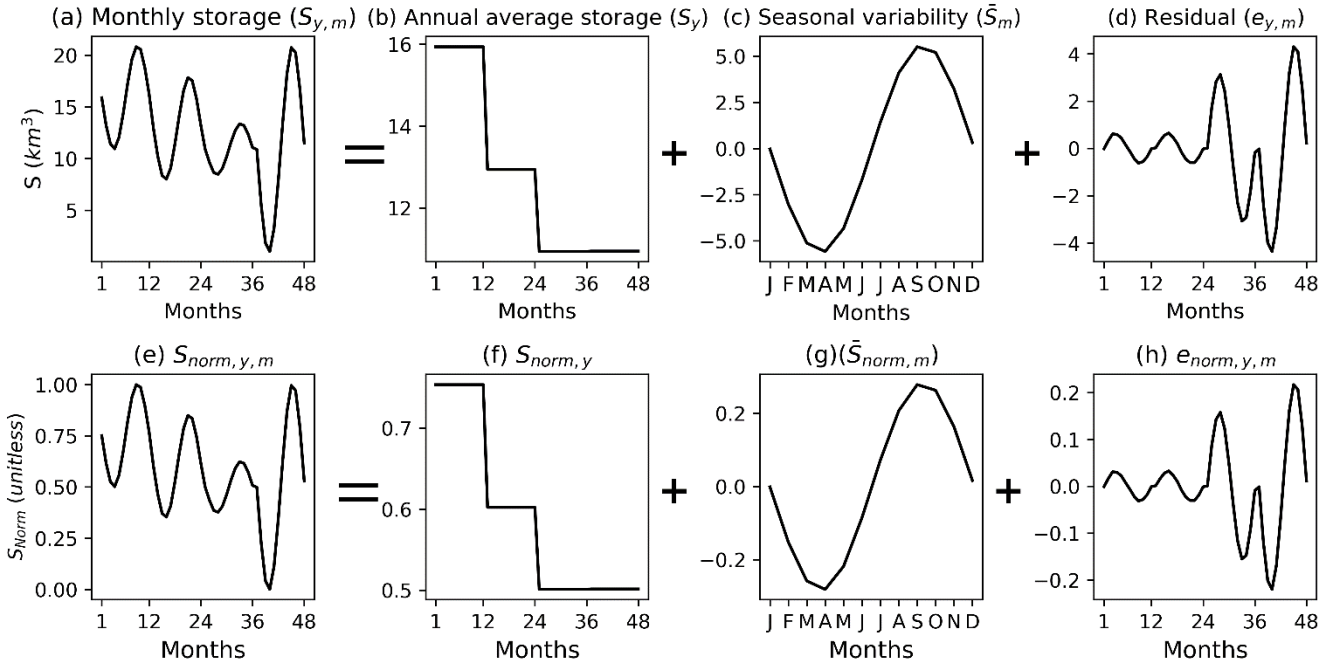


Figure 2: Components of volumetric storage investigated in this study. Raw (b-e) and normalized values (f-i).

2.8 Validation

https://doi.org/10.5194/hess-2023-215
Preprint. Discussion started: 12 September 2023


Hydrology and Earth System Sciences
Discussions

Two metrics, Pearson's correlation coefficient (r) and Nash–Sutcliffe efficiency (NSE), were used to validate time series data. Any months corresponding to missing values in either observation or simulation were excluded from the validation process (the percentage of missing values will be reported later).

r, which measures synchronicity in value fluctuations between simulations (s) and evaluations (e), is calculated as:

$$\mathbf{r} = \frac{\sum_{i=1}^{n}(s_i-\bar{s}).(e_i-\bar{e})}{\sqrt{\sum_{i=1}^{n}(s_i-\bar{s})^2}.\sqrt{\sum_{i=1}^{n}(e_i-\bar{e})^2}}, \quad (9)$$

where $\bar{s}$ and $\bar{e}$ represent the arithmetic means of the simulation (s) and evaluation (e) data, respectively. r ranges from -1 to 1, with 1 indicating a perfect positive correlation, -1 representing a perfect negative correlation, and 0 denoting no correlation. A 'two-sided' t-test (using the Wald test with a t-distribution of the test statistic) is used to determine the statistical significance of the correlation by calculating the p-value. The null hypothesis is that the slope of the regression line is zero; the alternative hypothesis is that the slope is non-zero. If $p < 0.05$, the correlation is considered statistically significant.

NSE, which measures the degree of matching between the values of evaluations (e) and simulations (s), is calculated as:

$$NSE = 1 - \frac{\sqrt{\sum_{i=1}^{n}(s_i-\bar{s})^2}}{\sqrt{\sum_{i=1}^{n}(e_i-\bar{e})^2}}, \quad (10)$$

where $\bar{s}$ and $\bar{e}$ represent the arithmetic means of the simulation (s) and evaluation (e) data, respectively. NSE values range from -∞ to 1, with 1 indicating a perfect match.

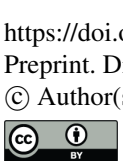
3 Results

3.1 Validation of satellite data

3.1.1 Monthly reservoir storage

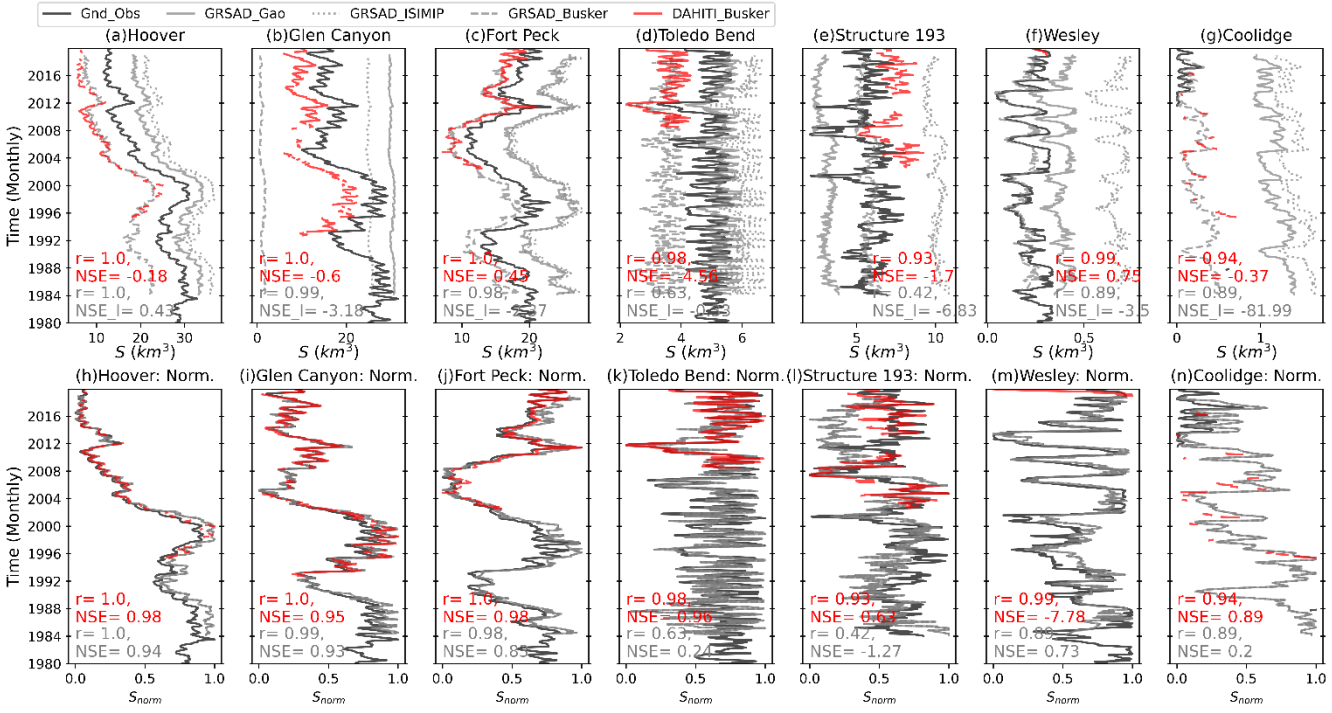
The monthly time series of storage volume for the seven selected reservoirs (reservoir storage, *S*) from two remote sensing datasets were compared with ground-based observations (Figure 3). The volumes calculated using satellite data (GRSAD and DAHITI) significantly fluctuated depending on the data source and the calculation method utilized for raw storage (Figure 3a-g). Intriguingly, even when the

https://doi.org/10.5194/hess-2023-215
Preprint. Discussion started: 12 September 2023



Hydrology and Earth System Sciences
Discussions
Open Access
EGU

same surface area data from GRSAD were used, the storage estimates varied according to the methodologies adopted (gray lines). However, after normalization, the satellite-derived reservoir volumes showed good alignment with ground observations (Figure 3h-n).



320 **Figure 3: Monthly reservoir storage from satellite data and ground observation. Raw monthly reservoir storage (a-g) and normalized storage (h-n) for the seven selected US reservoirs from ground observation (black) and two satellite data GRSAD (gray) and DAHITI (red). For GRSAD, three volumes are obtained by different combinations of data and methods (Table 2). Correlation coefficients (r) and NSE values for GRSAD_Gao and DAHITI_Busker are shown in the figure and Tables S4-5.**

325 Several factors contribute to the variances in satellite-derived *S*, utilizing different methods and data. For instance, the difference between *S* derived from GRSAD_ISIMIP and GRSAD_Gao can be attributed to the varying reservoir storage capacities used (Table S2). Intriguingly, *S* calculated using Busker's method (GRSAD_Busker), which does not consider the maximum storage parameters such as *Ac*, *hc*, and *Sc*, was closest to the observed storage.

330 The raw storage volume (*S*) calculated using DAHITI_Busker and GRSAD_Busker displayed considerable agreement for Hoover, Fort Peck, Toledo Bend, and Coolidge (Figure 3a, c, d, and g). This agreement is promising because these calculations used entirely different satellite products: surface area

https://doi.org/10.5194/hess-2023-215
Preprint. Discussion started: 12 September 2023


Hydrology and Earth System Sciences
Discussions

imagery and water level altimetry. However, discrepancies were evident for Glen Canyon and Structure 193 (Figure 3b and e).

For Glen Canyon, temporal storage variability was lost when surface area data from GRSAD were used, but not when surface elevation data from DAHITI were used. Glen Canyon, with its significant differences in surface area parameters considered in GRSAD and GRBD, showed limitations in the linear A-h relationship (Li et al., 2021). Although Li et al. (2021) previously described this issue and manually corrected a few reservoirs (including Glen Canyon), some persistent problems with GRSAD data remain.

Structure 193, also known as Lake Okeechobee, had a shallow average depth of 2.7 m but an extensive surface area (1900 km^2). Therefore, its A-h relationship was considerably different from the schematic shown in Figure 1, such that it had a very small value and b was negative (Equation 1, Table S2).

As clearly seen in the panels for the Wesley and the Coolidge Dams, the temporal coverage of DAHITI is quite limited (Figure 3f, g, m, n). For such cases, statistics should be viewed with care.

In summary, the raw satellite-based storage time series exhibited considerable uncertainty due to factors such as estimation of reservoir surface area, estimation of h0 and b (Figure 1), estimation of *hc* and *Ac*, and temporal coverage. The success in normalization is largely due to the proportionate contraction and expansion of different reservoir areas, if a significant portion of the area is considered. Consequently, normalization enables qualitative validation, including sign of change and timing of high/low peaks, of the abilities of hydrological models to simulate reservoir operations.

3.1.2 Decomposed monthly reservoir storage from satellite and ground observation

The normalized time series for satellite-derived and ground-based volumes were decomposed into annual mean storage, seasonal variability, and residuals (Figure 4). The decomposed raw storage (*S*) is depicted in Figure S1. The correlation coefficient and NSE are displayed in Table S6.

https://doi.org/10.5194/hess-2023-215
Preprint. Discussion started: 12 September 2023

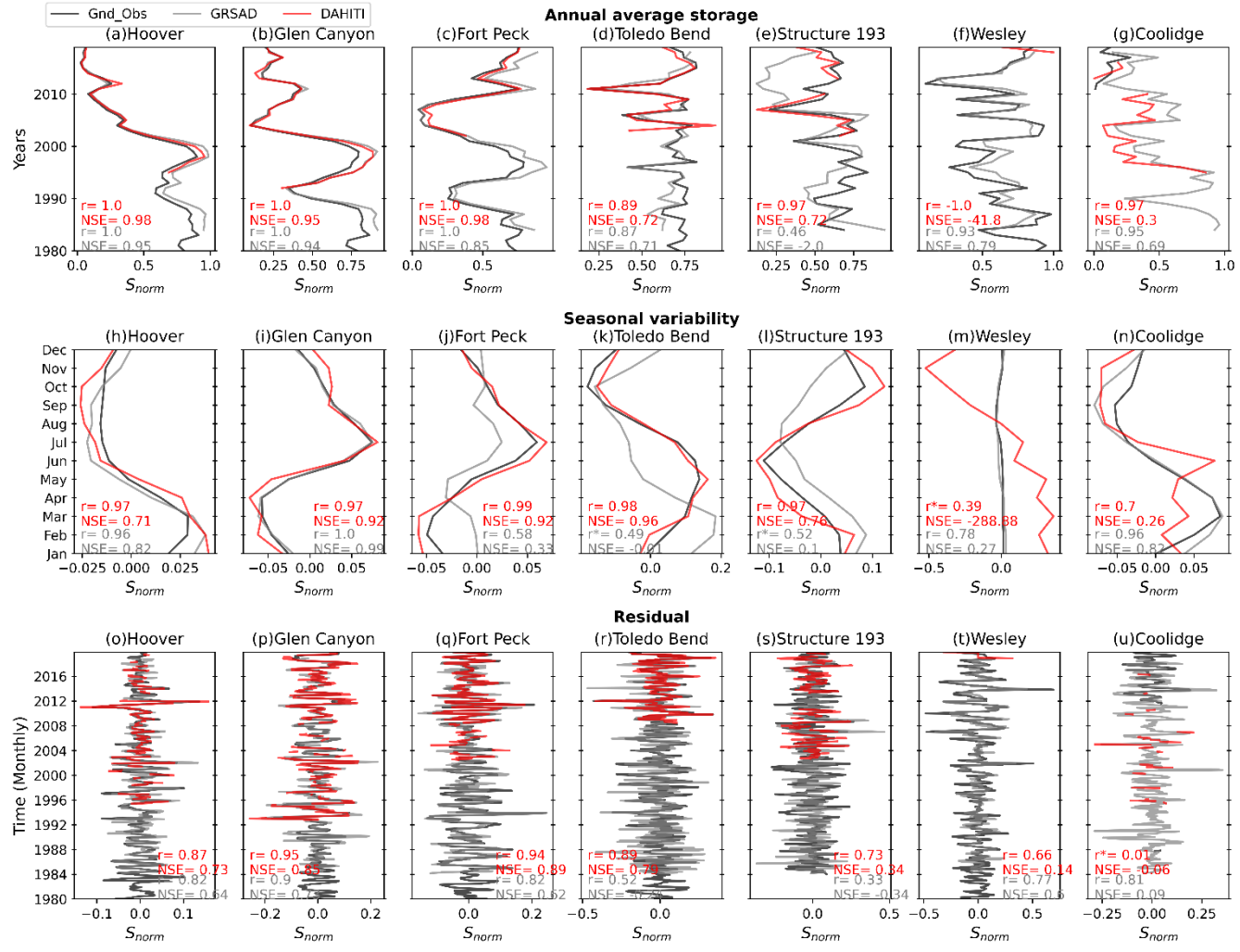


Figure 4: Decomposed normalized reservoir storage components from satellite data and ground observation. Annual mean storage (a-g), seasonal variability (h-n), and residuals (o-u) from satellite-derived data for normalized storage compared with ground observations. Corresponding correlation coefficients (r) and NSE values for DAHITI and GRSAD are annotated for each reservoir in the figure and Table S8. r* indicates insignificant correlation (p > 0.05).

The annual average of normalized reservoir storage (S_{norm}) is consistent with ground observations for both GRSAD and DAHITI (Figure 4a-g). For 10 of 14 cases, correlation coefficient and NSE values for GRSAD and DAHITI exceed 0.5. DAHITI exhibits poor $S_{norm,y}$ agreement for Structure 193; reservoirs with small surface area variability relative to water depth typically display poor agreement with ground observations (Busker et al., 2019). The agreement of DAHITI for Wesley and Coolidge dams also diverged from ground truth because temporal coverage was limited. Data for Wesley are available only

https://doi.org/10.5194/hess-2023-215
Preprint. Discussion started: 12 September 2023


Hydrology and Earth System Sciences
Discussions

from 12/2018 to 12/2019, or 13 months of the 480-month analysis period. Data for Coolidge are available for 78 months between 06/1995 and 05/2016 (252 months). Although both DAHITI and GRSAD demonstrate high correlation coefficients and NSE values, DAHITI agrees better in most reservoirs (Table S6), with the exception of the Coolidge Dam—despite a high correlation with DAHITI, the NSE is low due to limited availability of ground observation data.

The seasonal variability of normalized reservoir storage (Figure 4h-n) reveals strong alignment between ground observations and DAHITI, but not GRSAD, for Fort Peck, Toledo Bend, and Structure 193. This finding is corroborated by the results in Table S6, which indicate that correlations and NSE values are generally higher for DAHITI than for GRSAD. For Toledo Bend and Structure 193, the correlation is not statistically significant ($p > 0.05$) for GRSAD, although long-term data are available. However, GRSAD agrees better than DAHITI for Wesley and Coolidge because of the latter's limited temporal coverage.

The residual of normalized storage also exhibits strong agreement between satellite data and ground observations for most reservoir-satellite combinations, with high correlations and NSE values (Figure 4o-u). For reservoirs with sufficient temporal coverage (i.e., excluding Wesley and Coolidge), these values are generally higher for DAHITI than for GRSAD.

Overall, satellite-derived decomposed storage components (annual storage, seasonal variability, and residual) consistently compared well with ground-based observation storage components; correlation (> 0.7) and NSE (> 0.5) values were high (Moriasi et al., 2007). In most cases, the performance of annual storage was prominent among the decomposed components, particularly for GRSAD-based S_{norm} (Table S6).

These satellite-derived components of decomposed normalized monthly storage compared well against their ground observation counterparts. The annual storage, seasonal variability, and residuals, calculated after normalization of the original monthly storage, are suitable for validation of model simulations. DAHITI is highly reliable when sufficient, continuous data are available (for instance, data for > 5 years). When DAHITI data are unavailable or limited, GRSAD remains a viable (although less robust) alternative. Short-term data (< 3 years) and highly discontinuous data, such as Wesley for DAHITI and Coolidge for ground observation, should not be used for validation.

https://doi.org/10.5194/hess-2023-215
Preprint. Discussion started: 12 September 2023


Hydrology and Earth System Sciences
Discussions

3.2 Validation of simulated reservoir storage from ISIMIP3a

The simulated normalized monthly reservoir storage was initially validated against satellite-derived observations. The following sections compare the annual storage and seasonal variability, calculated using the normalized time series of monthly reservoir storage from ISIMIP3a simulations, with their respective counterparts from two satellite products. Finally, the consistency of the validation metric evaluated against satellite data is compared with the consistency of ground observations.

3.2.1 Monthly storage

The model simulations were reasonably consistent with satellite-based observations for $S_{norm,y,m}$ (Figure 5a-g). In particular, simulations for Fort Peck had high correlations with both GRSAD and DAHITI. Conversely, $S_{norm,y,m}$ for Structure 193 performed relatively poorly against DAHITI but performed better against GRSAD. On average, the model performed better when compared with GRSAD than when compared with DAHITI (Figure 5h and i). However, for Toledo Bend (Figure 3d), performance relative to DAHITI surpassed performance relative to GRSAD (Figure 5h and i). Notably, the performances of simulations decreased after 2005 (Figure 5a-g).

https://doi.org/10.5194/hess-2023-215
Preprint. Discussion started: 12 September 2023

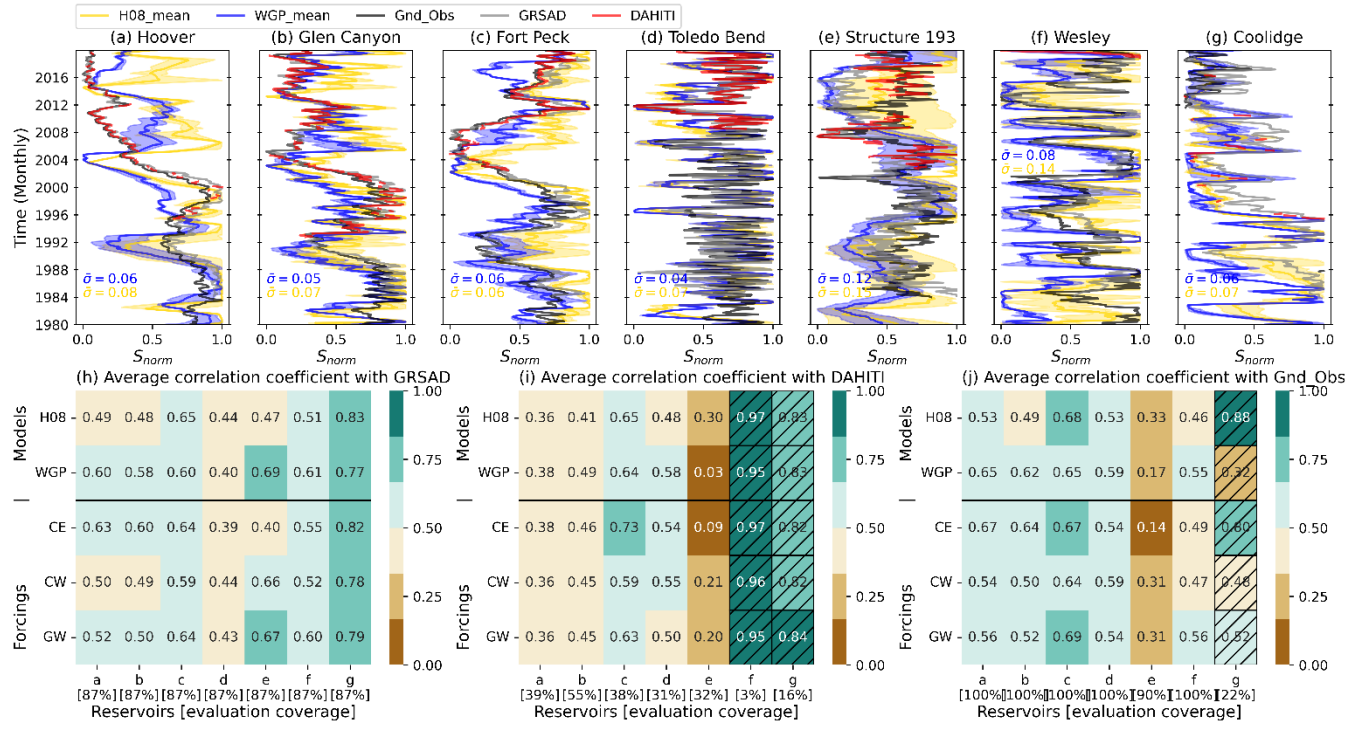
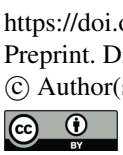


Figure 5: Validation of simulated monthly normalized reservoir storage. (a)-(g) Model simulations compared with satellite data and ground truth for monthly normalized reservoir storage. Color shading indicates mean variation among three forcing datasets, representing sensitivity to input forcings ($\bar{\sigma}$), for H08 (yellow) and WGP (blue). (h)-(j) Average correlation coefficient with three evaluation datasets: GRSAD, DAHITI, and ground observation, respectively, for each reservoir (a)-(g). Colors indicate correlation classification. Values in square brackets indicate percentage temporal coverage from 01/1980 to 12/2019 of reservoir storage for each reservoir's evaluation data. Reservoirs with hatch marks had < 30% coverage and were not included in subsequent analyses.

Among the GHMs, the performance of WGP ($\bar{r}$ > 0.5 for 8/12) was superior to the performance of H08 ($\bar{r}$ > 0.5 for 4/12) (Figure 3h and i). Compared with WGP, H08 was generally more sensitive ($\bar{\sigma}$) to input forcings (Figure 3a-g). Among the three forcings, GW ($\bar{r}$ > 0.5 for 7/12) and CE ($\bar{r}$ > 0.5 for 7/12) performed better than CW ($\bar{r}$ > 0.5 for 6/12). Direct comparison of r between CE and GW showed that CE had higher values (CE > GW for 7/12). Noteworthy is the decline in simulation performance since 2005 for undiscovered reasons. Because most DAHITI data included this period, performance relative to DAHITI is generally poor. Considering its long-term consistent coverage, GRSAD demonstrates better consistency with ground observations (i.e., Figure 5h displays better alignment with Figure 5j than with Figure 5i). Thus, in the validation of $_{Snorm,y,m}$ for ISIMIP3a, GRSAD is a more reliable evaluation data source than DAHITI.

https://doi.org/10.5194/hess-2023-215
Preprint. Discussion started: 12 September 2023



Hydrology and Earth System Sciences
Discussions
Open Access
EGU

430

3.2.2 Annual average storage

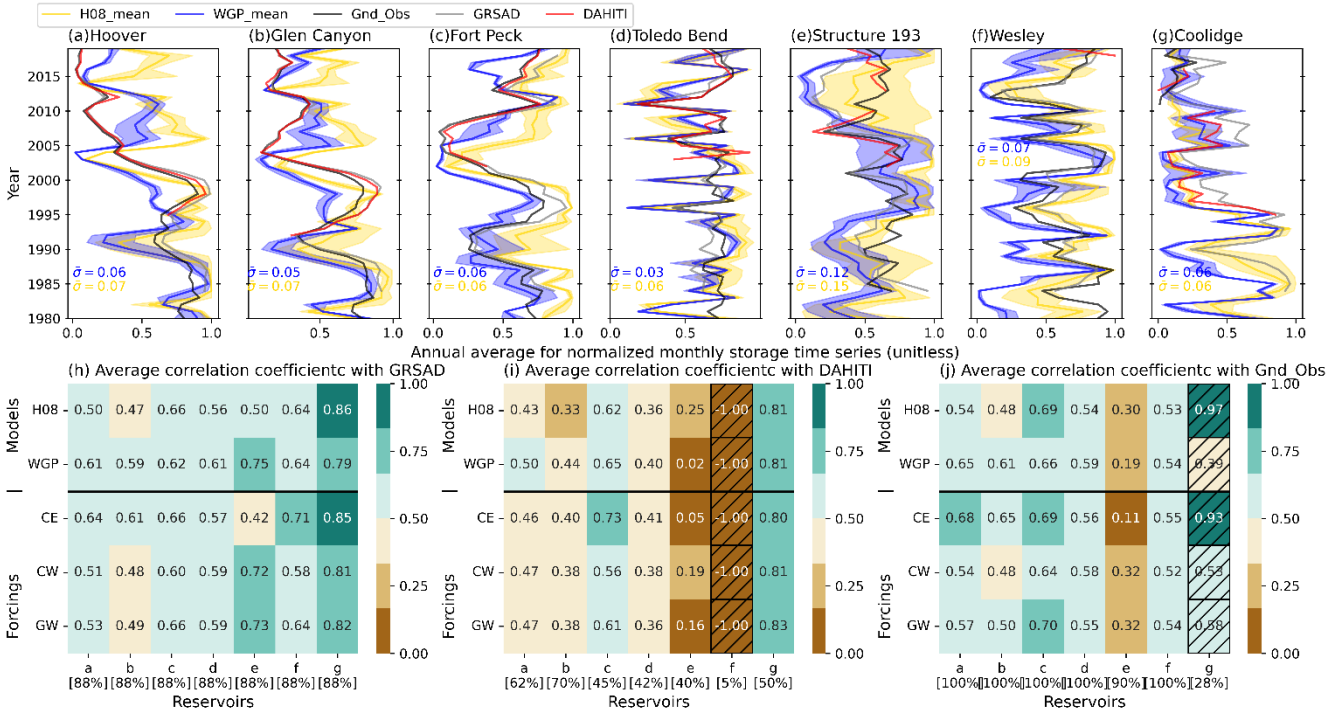


Figure 6: Validation of simulated annual average normalized reservoir storage. Same as Figure 5 but for annual average normalized reservoir storage ($S_{norm,y}$).

435

The annual storage simulations were consistent with satellite observations in most cases (Figure 6). In particular, simulations for Fort Peck and Coolidge demonstrated good agreement with both DAHITI and GRSAD (Figure 6c and g). For most reservoirs, the average correlation coefficient was > 0.5 for simulations across two models and three forcings compared with GRSAD (Figure 6h); this finding was 440 consistent with results from ground observation comparisons (Figure 6j). Because there were obvious discrepancies between DAHITI and ground observations (Figure 6i-j), we only used GRSAD for further comparisons of GHMs and forcings in this subsection.

The performances of the two GHMs concerning $S_{norm,y}$ correlations with satellite data were nearly equivalent; WGP ($\bar{r}$ > 0.5 for 7/7) was superior to H08 ($\bar{r}$ > 0.5 for 6/7) (Figure 6h). Additionally, H08 445 displayed a slightly larger standard deviation, compared with WGP (Figure 6a-g), indicating that it had

https://doi.org/10.5194/hess-2023-215
Preprint. Discussion started: 12 September 2023

Hydrology and Earth System Sciences
Discussions

substantially more interannual variability with input forcings. Among the input forcings, CE performed best in terms of the correlation coefficient, followed by GW and CW. WGP_CE correlation coefficients were considerably higher than the correlation coefficients of other GHM-Forcing combinations for most satellites (Figure S3).

The GHMs readily captured the interannual variation of reservoir storage. However, there were some limitations in satellite data that hindered validation of reservoir storage. For instance, validation relative to DAHITI was inconsistent with ground observations and GRSAD, possibly because > 50% of DAHITI data covered the period after 2005. As discussed earlier, the model's performance deteriorated during this period, confirmed by comparisons with ground observations. Thus, although DAHITI outperformed GRSAD relative to observations (as discussed in Section 3.1.2), correlations of simulated $S_{norm,y}$ were consistent with ground observations for GRSAD but not for DAHITI.

3.2.3 Seasonal variability

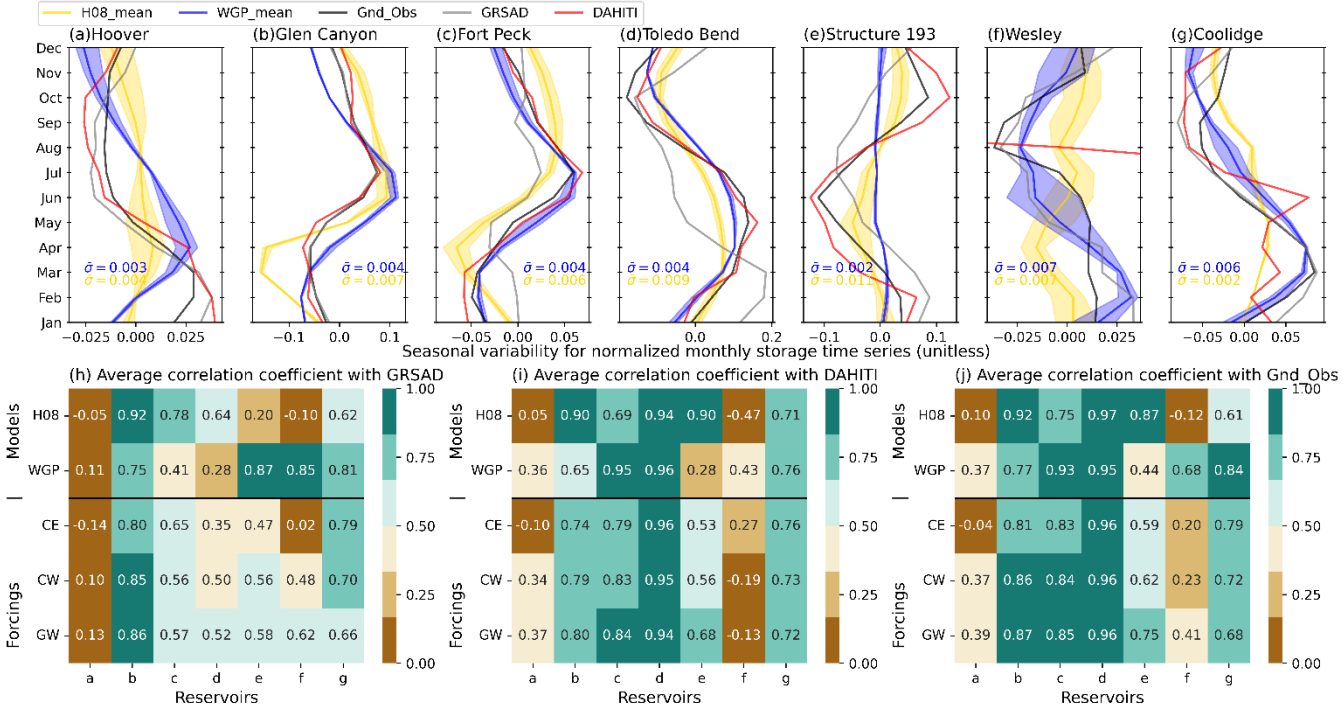


Figure 7: Validation of simulated monthly seasonal variability of normalized reservoir storage. Same as Figure 5, but for seasonal variability of normalized reservoir storage ($S_{norm,m}$). For Wesley (f), the DAHITI-derived storage is fully represented in Figure 4m.

https://doi.org/10.5194/hess-2023-215
Preprint. Discussion started: 12 September 2023


Hydrology and Earth System Sciences
Discussions

The model simulations adequately captured the seasonal cycle of reservoir storage in many instances. The correlations of simulations with both satellite datasets were generally high, such that many values exceeded 0.5 (21/35 for GRSAD and 24/35 for DAHITI), except the Hoover and Wesley dams. For instance, the simulated peak timing of the Hoover Dam (April) lagged the satellite products (from January to March) by 1-3 months, resulting in weaker correlations.

Both H08 ($\bar{r} > 0.5$ for 9/14) and WGP ($\bar{r} > 0.5$ for 8/14) performed particularly well in terms of simulating monthly variability in most reservoirs (Figure 7h and i); WGP was superior to H08 (for 9/14 cases). H08 demonstrated less robust performance for the Hoover and Wesley Dams, but WGP displayed relatively strong correlations with observations for these reservoirs. Therefore, WGP demonstrated superior overall performance compared with H08. Moreover, H08 exhibited greater variability according to input forcings, compared with WGP. Among the input forcings (Figure 7h and i), GW simulations ($\bar{r} > 0.5$ for 11/14) outperformed CW ($\bar{r} > 0.5$ for 9/14) and CE ($\bar{r} > 0.5$ for 8/14). Even in terms of r values, GW performed best for 8/14 cases among the three forcings (Figure 7h and i). This result is consistent with the evaluation relative to ground observations, where 5/7 reservoirs had the highest correlation for GW (Figure 7j).

Comparison of simulations showed that DAHITI and GRSAD aligned with ground observations; DAHITI demonstrated relative consistency. However, there were instances of contradictory outcomes, such as WGP validation for the Fort Peck Dam and the Toledo Bend Dam, where a weak correlation with GRSAD differed from a strong correlation with DAHITI. Cases with very short data availability periods, such as Wesley and Coolidge for DAHITI, should be excluded from validation. In these instances, GRSAD should be used because it can appropriately capture the seasonal variation due to its long-term data availability. However, when DAHITI has sufficient temporal coverage, it outperforms GRSAD. Opposite conclusions resulted from the H08-WGP comparison for Structure 193. In this case, the simulated seasonal cycle of WGP for Structure 193 closely correlated with GRSAD, but the amplitude was considerably smaller. This discrepancy leads to questions regarding the reliability of a single evaluation data source. Therefore, in the absence of ground-based observations, multiple satellite data products and metrics should be used to increase confidence in validation and intercomparison results.

https://doi.org/10.5194/hess-2023-215
Preprint. Discussion started: 12 September 2023


Hydrology and Earth System Sciences
Discussions

When a sufficient number of reliable satellite products were available, it would be possible to calculate the mean and ranges of satellite data ensemble.

As expected, the seasonality of reservoir storage was relatively stable compared with annual storage and residuals. This phenomenon is evident from the high correlations of seasonal variability (Figure 7) relative to annual storage (Section 3.2.2, Figure 6).

3.3 Uncertainties

The present study was conducted on the basis of numerous works that enabled satellite monitoring of artificial reservoirs, including the estimation of absolute storage volumes by Gao et al. (2012). Consequently, this study inherited some uncertainties from the previous efforts. There were four main issues. First, the methods in previous studies assumed a linear A-h relationship for reservoir storage. This relationship is not genuinely linear, particularly when the reservoir is near full or empty. The previous approaches also required knowledge of water surface elevation at storage capacity; such information is not currently available in published global reservoir inventories. Therefore, significant uncertainties may arise when calculating reservoir storage using these parameters (Gao et al., 2012). Furthermore, the records in the inventories are not necessarily error-free or consistent with information in other inventories. This systematic approach demands extensive quality checks among global reservoir inventories. Second, the limitations in area-based satellite products (i.e. GRSAD). Discrepancies in the consideration of water surface area extents (i.e., distinguishing between reservoir and river), as noted in the case of the Glen Canyon Dam (Section 3.1.2), lead to concerns about the reliability of water surface area datasets. Third, the limitations in altitude-based satellite products (i.e. DAHITI). This study found that DAHITI agrees better with the ground observation than GRSAD, but the advantage is largely deteriorated by its high frequency of missing data. Further technical advancement in data processing is expected for more extensive spatio temporal coverage of this type of data. Due to these concerns, we abandoned the use of absolute storage estimates. Although we showed that the timing and rate of rise and fall can be validated, a significant limitation of our study was the loss of storage magnitude information. Finally, since 2005, there has been a clear discrepancy between simulation results and ground observations. This issue should

https://doi.org/10.5194/hess-2023-215
Preprint. Discussion started: 12 September 2023

be further examined from various aspects, including validation with other variables and at different locations.

4 Conclusions

In this study, we examined the feasibility of using satellite reservoir storage estimations as an alternative to ground truth, applying them to validate two global hydrological models from the ISIMIP. The critical findings for the specific research questions we posed are as follows.

The first question was, "Can satellite-based storage estimation data serve as a surrogate for ground truth data?" Based on a detailed comparison of two satellite products and ground observations, we found that the satellite products can be used as a surrogate for ground truth when two key criteria are met. First, because there is significant uncertainty when converting raw satellite observations (i.e., water surface area and water level altimetry) into absolute reservoir storage volumes, both satellite and simulation data should be normalized before comparison. Second, the satellite observation period must be sufficiently long (i.e., 5 years) to correctly capture long-term trends and sample monthly storage variation. The normalized reservoir storage can be further decomposed into annual average storage, mean monthly storage, and residuals. As expected, seasonal variability exhibits the highest correlations, followed by annual storage and residuals. For seven reservoirs in CONUS, the two ISIMIP3a models, H08 and WGP, demonstrated satisfactory performance in terms of normalized annual average storage and seasonal variability.

The second question was, "Can we determine which GHMs or meteorological forcings perform better than others in model intercomparison projects, solely by satellite-based storage estimation?" We found that, overall, WGP demonstrated slightly better performance compared with H08 (Figures 5-7), although differences between the two models were minor. Considering the forcing data, CE and GW exhibited the best performances for annual storage and seasonal variability, respectively. Therefore, it is challenging to identify the superior model. Also, the model and forcing performance varies by reservoirs.

The third and fourth questions were, "Do the findings align with ground observations?" and, "Are certain satellite products superior to others?" These questions can be answered simultaneously. We found general agreement between satellite-based and ground observation-based validation results, indicating overall

https://doi.org/10.5194/hess-2023-215
Preprint. Discussion started: 12 September 2023


Hydrology and Earth System Sciences
Discussions

reliability. Comparisons of DAHITI and GRSAD revealed that DAHITI demonstrates better consistency with ground observations if temporal coverage is sufficient. However, with respect to simulations, the extended temporal coverage of GRSAD provides better agreement with ground observations for annual storage and residuals. Therefore, to increase confidence in the results, multiple satellite datasets should be utilized for model validation and intercomparison efforts.

To our knowledge, this study is the first effort to use multiple satellite-based products to validate and intercompare multiple models for reservoir operations across forcings on a continentall scale. Often, reservoir operation records are not disclosed, especially for basins that flow across multiple countries (Vu et al., 2022). Our study demonstrated the feasibility of extending the spatial coverage of validation and intercomparison on a global scale.

To facilitate further research and applications, we offer four recommendations. First, the latest satellite techniques must be incorporated to reduce the uncertainties (i.e., on the accuracy and stability of data retrieval) discussed in Section 3.3. Second, more models and forcings should be included to enhance the comprehensiveness of the study by expanding the ensemble of simulations. Third, although this study exclusively focused on the CONUS region, future studies should be performed on a global scale, including reservoirs without ground observations. Finally, an integrated platform combining multiple satellite products with a common ID is needed to synchronize reservoirs and lakes with existing inventories such as GRanD and Hydrolakes.

There is considerable potential for improvement to enhance accuracy and precision in GHMs. Although GHM simulations provide valuable insights, there remain significant uncertainties in the representation of reservoir dynamics. By refining the models, incorporating more accurate input data, and considering additional factors that influence reservoir behavior, better alignment between simulations and real-world observations can be achieved. This ongoing effort to enhance satellite-based validation will lead to more reliable reservoir storage assessments and predictions.

Code and data availability

The code and data associated with this study can be accessed at https://doi.org/10.5281/zenodo.8291850.

https://doi.org/10.5194/hess-2023-215
Preprint. Discussion started: 12 September 2023


Hydrology and Earth System Sciences
Discussions

Supplementary material

Supplementary material related to this article is available online.

Author contributions

KO and NH conceptualized and designed the experiments, which were then conducted by KO. KO, HMS, and NH developed the model code and performed the simulations. KO wrote the manuscript, with contributions from NH, HMS, and SNG.

Competing interests

The corresponding author declares that neither they nor their co-authors have any competing interests.

Financial support

This research was supported by the Japan Society for the Promotion of Science (KAKENHI; grant numbers 21H05002 and 22H01604).

References

Alsdorf, D. E. and Lettenmaier, D. P.: Tracking fresh water from space, Science, 301, 1491–1494, https://doi.org/10.1126/science.1089802, 2003.

Biemans, H., Haddeland, I., Kabat, P., Ludwig, F., Hutjes, R. W. A., Heinke, J., Von Bloh, W., and Gerten, D.: Impact of reservoirs on river discharge and irrigation water supply during the 20th century, Water Resour. Res., 47, 1–15, https://doi.org/10.1029/2009WR008929, 2011.

Burek, P., Satoh, Y., Kahil, T., Tang, T., Greve, P., Smilovic, M., Guillaumot, L., Zhao, F., and Wada, Y.: Development of the Community Water Model (CWatM v1.04) - A high-resolution hydrological model for global and regional assessment of integrated water resources management, Geosci. Model Dev., 13, 3267–3298, https://doi.org/10.5194/gmd-13-3267-2020, 2020.

Busker, T., De Roo, A., Gelati, E., Schwatke, C., Adamovic, M., Bisselink, B., Pekel, J. F., and Cottam, A.: A global lake and reservoir volume analysis using a surface water dataset and satellite altimetry, Hydrol. Earth Syst. Sci., 23, 669–690, https://doi.org/10.5194/hess-23-669-2019, 2019.

Döll, P., Fiedler, K., and Zhang, J.: Global-scale analysis of river flow alterations due to water withdrawals and reservoirs, Hydrol. Earth Syst. Sci., 13, 2413–2432, https://doi.org/10.5194/hess-13-2413-2009, 2009.

Frieler, K., Volkholz, J., Lange, S., Schewe, J., Mengel, M., Rivas López, M. R., Otto, C., Reyer, C. P. O., Karger, D. N., Malle, J. T., Treu, S., Menz, C., Blanchard, J. L., Harrison, C. S., Petrik, C. M., Eddy, T. D., Ortega-Cisneros, K., Novaglio, C., Rousseau, Y., Watson, R. A., Stock, C., Liu, X., Heneghan, R., Tittensor, D., Maury, O., Büchner, M., Vogt, T., Wang, T., Sun, F., Sauer, I. J., Koch, J., Vanderkelen, I., Jägermeyr, J., Müller, C., Klar, J., del Valle, I. D., Lasslop, G., Chadburn, S., Burke, E., Gallego-Sala, A., Smith, N., Chang, J., Hantson, S., Burton, C., Gädeke, A., Li, F., Gosling, S. N., Müller Schmied, H., Hattermann, F., Wang, J., Yao, F., Hickler, T., Marcé, R., Pierson, D., Thiery, W., Mercado-Bett, D., Forrest, M., and Bechtold, M.: Scenario set-up and forcing data for impact model evaluation and impact attribution within the third round of the Inter-Sectoral Model Intercomparison Project (ISIMIP3a), EGUsphere [preprint], 1–83, https://doi.org/10.5194/egusphere-2023-281, 2023.

Gao, H. and Zhao, G.: Global Reservoir Surface Area Dataset (GRSAD), https://doi.org/10.18738/T8/DF80WG, 2020.

Gao, H., Birkett, C., and Lettenmaier, D. P.: Global monitoring of large reservoir storage from satellite remote sensing, Water Resour. Res., 48, 1–12, https://doi.org/10.1029/2012WR012063, 2012.

Grill, G., Lehner, B., Thieme, M., Geenen, B., Tickner, D., Antonelli, F., Babu, S., Borrelli, P., Cheng, L., Crochetiere, H., Ehalt Macedo, H., Filgueiras, R., Goichot, M., Higgins, J., Hogan, Z., Lip, B., McClain, M. E., Meng, J., Mulligan, M., Nilsson, C., Olden, J. D., Opperman, J. J., Petry, P., Reidy Liermann, C., Sáenz, L., Salinas-Rodríguez, S., Schelle, P., Schmitt, R. J. P., Snider, J., Tan, F., Tockner, K., Valdujo, P. H., van Soesbergen, A., and Zarfl, C.: Mapping the world's free-flowing rivers, Nature, 569, 215–221, https://doi.org/10.1038/s41586-019-1111-9, 2019.

Haddeland, I., Skaugen, T., and Lettenmaier, D. P.: Anthropogenic impacts on continental surface water fluxes, Geophys. Res. Lett., 33, 2–5, https://doi.org/10.1029/2006GL026047, 2006.

Hanasaki, N., Kanae, S., and Oki, T.: A reservoir operation scheme for global river routing models, J. Hydrol., 327, 22–41, https://doi.org/10.1016/j.jhydrol.2005.11.011, 2006.

Hanasaki, N., Kanae, S., Oki, T., Masuda, K., Motoya, K., Shirakawa, N., Shen, Y., and Tanaka, K.: An integrated model for the assessment of global water resources - Part 1: Model description and input meteorological forcing, Hydrol. Earth Syst. Sci., 12, 1007–1025, https://doi.org/10.5194/hess-12-1007-2008, 2008a.

Hanasaki, N., Kanae, S., Oki, T., Masuda, K., Motoya, K., Shirakawa, N., Shen, Y., and Tanaka, K.: An integrated model for the assessment of global water resources - Part 2: Applications and assessments, Hydrol. Earth Syst. Sci., 12, 1027–1037, https://doi.org/10.5194/hess-12-1027-2008, 2008b.

Hanasaki, N., Yoshikawa, S., Pokhrel, Y., and Kanae, S.: A global hydrological simulation to specify the sources of water used by humans, Hydrol. Earth Syst. Sci., 22, 789–817, https://doi.org/10.5194/hess-22-789-2018, 2018.

Jarvis, A., Reuter, H. I., Nelson, A., and Guevara, E.: Hole-filled seamless SRTM data V4, International Centre for Tropical Agriculture (CIAT), available from http://srtm.csi.cgiar.org., 2008.

Kumar, A., Gosling, S. N., Johnson, M. F., Jones, M. D., Zaherpour, J., Kumar, R., Leng, G., Schmied, H. M., Kupzig, J., Breuer, L., Hanasaki, N., Tang, Q., Ostberg, S., Stacke, T., Pokhrel, Y., Wada, Y., and Masaki, Y.: Multi-model evaluation of catchment- and global-scale hydrological model simulations of drought characteristics across eight large river catchments,

Adv. Water Resour., 165, 104212, https://doi.org/10.1016/j.advwatres.2022.104212, 2022.

Lange, S., Mengel, M., Treu, S., and Büchner, M.: ISIMIP3a atmospheric climate input data (v1.0), https://doi.org/10.48364/ISIMIP.982724, 2022.

Lehner, B., Liermann, C. R., Revenga, C., Vörömsmarty, C., Fekete, B., Crouzet, P., Döll, P., Endejan, M., Frenken, K., Magome, J., Nilsson, C., Robertson, J. C., Rödel, R., Sindorf, N., and Wisser, D.: High-resolution mapping of the world's reservoirs and dams for sustainable river-flow management, Front. Ecol. Environ., 9, 494–502, https://doi.org/10.1890/100125, 2011.

Li, Y., Gao, H., Zhao, G., and Tseng, K. H.: A high-resolution bathymetry dataset for global reservoirs using multi-source satellite imagery and altimetry, Remote Sens. Environ., 244, 111831, https://doi.org/10.1016/j.rse.2020.111831, 2020.

Li, Y., Gao, H., Allen, G. H., and Zhang, Z.: Constructing reservoir area-volume-elevation curve from TanDEM-X DEM data, IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens., 14, 2249–2257, https://doi.org/10.1109/JSTARS.2021.3051103, 2021.

Messager, M. L., Lehner, B., Grill, G., Nedeva, I., and Schmitt, O.: Estimating the volume and age of water stored in global lakes using a geo-statistical approach, Nat. Commun., 7, 1–11, https://doi.org/10.1038/ncomms13603, 2016.

Moriasi, D. N., Arnold, J. G., Van Liew, M. W., Bingner, R. L., Harmel, R. D., and Veith, T. L.: Model evaluation guidelines for systematic quantification of accuracy in watershed simulations, Trans. ASABE, 50, 885–900, https://doi.org/10.13031/2013.23153, 2007.

Müller Schmied, H. and Schiebener, L.: The global water resources and use model WaterGAP v2.2e: streamflow calibration and evaluation data basis, https://doi.org/10.5281/ZENODO.7255968, 2022.

Müller Schmied, H., Caceres, D., Eisner, S., Flörke, M., Herbert, C., Niemann, C., Asali Peiris, T., Popat, E., Theodor Portmann, F., Reinecke, R., Schumacher, M., Shadkam, S., Telteu, C. E., Trautmann, T., and Döll, P.: The global water resources and use model WaterGAP v2.2d: Model description and evaluation, Geosci. Model Dev., 14, 1037–1079, https://doi.org/10.5194/gmd-14-1037-2021, 2021.

Pekel, J. F., Cottam, A., Gorelick, N., and Belward, A. S.: High-resolution mapping of global surface water and its long-term changes, Nature, 540, 418–422, https://doi.org/10.1038/nature20584, 2016.

Pokhrel, Y., Hanasaki, N., Koirala, S., Cho, J., Yeh, P. J. F., Kim, H., Kanae, S., and Oki, T.: Incorporating anthropogenic water regulation modules into a land surface model, J. Hydrometeorol., 13, 255–269, https://doi.org/10.1175/JHM-D-11-013.1, 2012.

Pokhrel, Y., Felfelani, F., Satoh, Y., Boulange, J., Burek, P., Gädeke, A., Gerten, D., Gosling, S. N., Grillakis, M., Gudmundsson, L., Hanasaki, N., Kim, H., Koutroulis, A., Liu, J., Papadimitriou, L., Schewe, J., Müller Schmied, H., Stacke, T., Telteu, C. E., Thiery, W., Veldkamp, T., Zhao, F., and Wada, Y.: Global terrestrial water storage and drought severity under climate change, Nat. Clim. Chang., 11, 226–233, https://doi.org/10.1038/s41558-020-00972-w, 2021.

Schwatke, C., Dettmering, D., Bosch, W., and Seitz, F.: DAHITI - An innovative approach for estimating water level time series over inland waters using multi-mission satellite altimetry, Hydrol. Earth Syst. Sci., 19, 4345–4364, https://doi.org/10.5194/hess-19-4345-2015, 2015.

Schwatke, C., Scherer, D., and Dettmering, D.: Automated extraction of consistent time-variable water surfaces of lakes and reservoirs based on Landsat and Sentinel-2, Remote Sens., 11, 1010, https://doi.org/10.3390/rs11091010, 2019.

Steyaert, J. C., Condon, L. E., W.D. Turner, S., and Voisin, N.: ResOpsUS, a dataset of historical reservoir operations in the contiguous United States, Sci. Data, 9, 1–8, https://doi.org/10.1038/s41597-022-01134-7, 2022.

Vu, D. T., Dang, T. D., Galelli, S., and Hossain, F.: Satellite observations reveal 13 years of reservoir filling strategies, operating rules, and hydrological alterations in the Upper Mekong River basin, Hydrol. Earth Syst. Sci., 26, 2345–2364, https://doi.org/10.5194/hess-26-2345-2022, 2022.

Wada, Y., Van Beek, L. P. H., Viviroli, D., Drr, H. H., Weingartner, R., and Bierkens, M. F. P.: Global monthly water stress: 2. Water demand and severity of water stress, Water Resour. Res., 47, 1–17, https://doi.org/10.1029/2010WR009792, 2011.

Wada, Y., Wisser, D., Eisner, S., Flörke, M., Gerten, D., Haddeland, I., Hanasaki, N., Masaki, Y., Portmann, F. T., Stacke, T., Tessler, Z., and Schewe, J.: Multimodel projections and uncertainties of irrigation water demand under climate change, Geophys. Res. Lett., 40, 4626–4632, https://doi.org/10.1002/grl.50686, 2013.

Yoshida, T., Hanasaki, N., Nishina, K., Boulange, J., Okada, M., and Troch, P. A.: Inference of parameters for a global hydrological model: identifiability and predictive uncertainties of climate-based parameters, Water Resour. Res., 58, https://doi.org/10.1029/2021WR030660, 2022.

Zaherpour, J., Gosling, S. N., Mount, N., Schmied, H. M., Veldkamp, T. I. E., Dankers, R., Eisner, S., Gerten, D., Gudmundsson, L., Haddeland, I., Hanasaki, N., Kim, H., Leng, G., Liu, J., Masaki, Y., Oki, T., Pokhrel, Y., Satoh, Y., Schewe, J., and Wada, Y.: Worldwide evaluation of mean and extreme runoff from six global-scale hydrological models that account for human impacts, Environ. Res. Lett., 13, https://doi.org/10.1088/1748-9326/aac547, 2018.

Zhao, G. and Gao, H.: Automatic correction of contaminated images for assessment of reservoir surface area dynamics, Geophys. Res. Lett., 45, 6092–6099, https://doi.org/10.1029/2018GL078343, 2018.