

Water Resources Research

RESEARCH ARTICLE

10.1029/2020WR027902

Key Points:

- Data-driven release-availability functions are more accurate than generic reservoir schemes when forced with observed inflows
- Data-driven schemes fail to provide robust simulation improvements under conditions of inflow bias common to large-scale hydrological models
- Water supply vulnerability varies widely across reservoir schemes, but large-scale hydrological simulations fail to expose this sensitivity

Supporting Information:

- Supporting Information S1

Correspondence to:

S. W. D. Turner,
sean.turner@pnpl.gov

Citation:

Turner, S. W. D., Doering, K., & Voisin, N. (2020). Data-driven reservoir simulation in a large-scale hydrological and water resource model. *Water Resources Research*, 56, e2020WR027902. <https://doi.org/10.1029/2020WR027902>

Received 8 MAY 2020

Accepted 8 SEP 2020

Accepted article online 14 SEP 2020

Data-Driven Reservoir Simulation in a Large-Scale Hydrological and Water Resource Model

Sean W. D. Turner¹ , Kenji Doering² , and Nathalie Voisin^{1,3} 

¹Energy and Environment Directorate, Pacific Northwest National Laboratory, Seattle, WA, USA, ²Department of Biological and Environmental Engineering, Cornell University, Ithaca, NY, USA, ³Department of Civil and Environmental Engineering, University of Washington, Seattle, WA, USA



Abstract Large-scale hydrological and water resource models (LHMs) are used increasingly to study the vulnerability of human systems to water scarcity. These models rely on generic reservoir release schemes that often fail to capture the nuances of operations at individual dams. Here we assess whether empirically derived release-availability functions tailored to individual dams could improve the simulation performance of an LHM. Seasonally varying, linear piecewise relations that specify water release as a function of prevailing storage levels and forecasted future inflow are compared to a common generic scheme for 36 key reservoirs of the Columbia River Basin. When forced with observed inflows, the empirical approach captures observed release decisions better than the generic scheme—including under conditions of drought. The inclusion of seasonally varying inflow forecasts used by reservoir operators adds further improvement. When exposed to biases and errors inherent in the LHM, data-driven policies fail to offer a robust improvement; inclusion of forecasts deteriorates LHM reservoir simulation performance in some cases. We perform sensitivity analysis to explain this result, finding that the bias inherent in LHM streamflow is amplified by a reservoir model that relies on forecasts. To harness the potential of interpretable, data-driven reservoir operating schemes, research must address LHM flow biases arising from inaccuracies in climate input, runoff generation, flow routing, and water withdrawal and consumption data.

1. Introduction

Large-scale hydrological and water resource models (LHMs) require a water storage and release scheme to simulate the flow regulating effects of dams (Dang et al., 2020; Nazemi & Wheater, 2015b). Designing such a scheme is challenging, because reservoir operations often depend on complex and undocumented decision processes. In state-of-the-art LHMs, such as PCR-GLOBWB (Wada et al., 2014), H08 (Hanasaki et al., 2008), HiGW-MAT (Pokhrel et al., 2015), WaterGAP (Döll et al., 2009), WBM (Wisser et al., 2010), and VIC/CLM-MOSART-WM (Voisin et al., 2013), a generic, inflow-and-demand-based scheme is deployed; the reservoir's primary purpose (e.g., irrigation, flood control, and water supply) and characteristics of its inflow, storage capacity, and demand, rather than its record of operation, are used to define seasonally varying water release, following the methods of Hanasaki et al. (2006), Haddeland et al. (2006), Döll et al. (2009), Biemans et al. (2011), and Voisin et al. (2013). Important constraints, such as the proportion of water to be allocated for environmental flow, are parameterized arbitrarily and uniformly across dams. This reservoir scheme (herein referred to as a “generic release scheme,” following Masaki et al., 2017) has allowed scientists to simulate key flow-regulating behaviors of a large fleet of reservoirs within a consistent framework, relying on data sets with global coverage (e.g., the Global Reservoirs and Dams Database—Lehner et al., 2011) to describe reservoir characteristics and operating purposes where operational records are scarce (Wada et al., 2017). Yet simulated flow variability downstream of dams is sensitive to the reservoir algorithm adopted (Masaki et al., 2017), and substantial errors in simulated water releases are inevitable when locally important operational nuances are neglected (Yassin et al., 2019). Flow errors can compound over time through storage memory. Each erroneous release decision provides an erroneous inflow to the next reservoir, propagating the error downstream. Although unimportant when assessing regionally aggregated water scarcity at an annual scale (Haddeland et al., 2014), such errors jeopardize an LHM's ability to simulate drought impacts, since the state of reservoirs at the onset of drought—and their operations throughout drought events—has a significant bearing on likelihood and severity of shortfalls in water supply (Turner &

©2020. The Authors.

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

Galelli, 2016). In contrast to generic release schemes, a set of empirically derived release-availability functions (herein referred to as “data-driven scheme”) can be parameterized for individual reservoirs replete with sufficiently long observational records of releases, inflows, and storage levels. Data-driven schemes have been demonstrated to outperform generic release schemes in an isolated number of reservoirs and have thus been proposed as a viable alternative that could be implemented in LHM (Coerver et al., 2018; Mateo et al., 2014; Yassin et al., 2019). To our knowledge, the improvements available from such an approach have yet to be evaluated for a large system with multiple reservoirs in cascade. We also find no published research exploring how such a scheme might affect simulated water supply reliability and vulnerability—metrics that underpin rigorous drought impact assessment (Hashimoto et al., 1982; McMahon et al., 2006).

To evaluate the potential benefits of adopting a data-driven reservoir release scheme in LHMs, we compare generic and data-driven reservoir simulations for a major river basin of the United States. Simulation performances are evaluated using standard goodness-of-fit metrics and errors in cumulative water volumes released during drought. The LHM adopted in this study tracks unmet water demands and is therefore also used here to explore sensitivity of water supply reliability and vulnerability to the choice of reservoir scheme. Our study therefore indicates the importance of the reservoir release scheme to drought analyses, in addition to revealing the improvements in flow simulation made available by advancing from a generic to a data-driven approach. Additionally, we compare two possible settings for a data-driven release model: with and without representation of perfect inflow forecasts. In the latter model, forecast lead times vary seasonally according to availability of predictive skill (e.g., longer lead times in early spring when snowpack depths indicate incoming flows weeks ahead). All experiments are conducted using both observed (off-line) and LHM-simulated (online) inflows. Crucially, this allows us to isolate the influence of hydrological model inflow bias on simulation performance. LHM are used increasingly in national water scarcity assessments as well as multisector dynamics research examining the coevolution and interactions of water, energy, and land systems (e.g., Behrens et al., 2017; Hejazi et al., 2015; Schewe et al., 2014; Voisin et al., 2018; Wada et al., 2016). This study offers a first glimpse into the potential impact and benefits of implementing site-specific operations into reservoir schemes to support these applications.

2. Methods

2.1. Domain, Models, and Data

For this study we use the Columbia River Basin (CRB) of the U.S. Pacific Northwest as the spatial domain. The CRB covers 668,000 km², encompassing territory of five U.S. states as well as British Columbia, Canada. The hydrology of basin is diverse, but overall snowmelt controlled with a large peak freshet in late spring. The CRB is appropriate for this study because the river system is highly regulated, featuring 120 reservoirs of significant storage capacity (>10 million cubic meters, MCM) and operated for various purposes, including hydropower generation, flood control, and water supply for irrigation. The CRB is also rich in lengthy records of reservoir inflow, storage levels, and release; of the 120 reservoirs of the CRB with >10 MCM capacity, we obtain sufficient observational records for 36 reservoirs to develop a data-driven release scheme (>10 years, daily resolution storage, and at least one of inflow or release).

The LHM adopted in this study is an amalgamation of two state-of-the-art models—the Variable Infiltration Capacity (VIC) model (Liang et al., 1994) and the Model for Scale Adaptive River Transport and Water Management (MOSART-WM) (Voisin et al., 2013, 2017). The VIC hydrological simulations applied in this study are available through the Livneh et al. (2013) daily CONUS near-surface gridded meteorological and derived hydrometeorological data set, which provides spatially distributed runoff simulated with gridded climate observations for the period 1915–2011. We select the period 1980–2011 and then aggregate runoff to (1/8)^o. The MOSART component of the model (Li et al., 2013) routes runoff accounting for heterogeneous land and river channel effects, while the water management component (WM) (Voisin et al., 2013) introduces reservoir storage and withdrawals of water for multiple human uses. Exogenous water demand is first met within each grid cell by allocating local surface water and ground water; surface water supply is augmented where needed through withdrawals from reservoir releases. Water demands in this study are set to a 2010 profile (as in Voisin et al., 2018)—a common approach due to uncertainty in water demand trend through time (e.g., BPA, 2011). Since actual water demands would have varied substantially over a 30-year period, a fixed demand profile clearly introduces nonnegligible error into the simulated flows. These errors

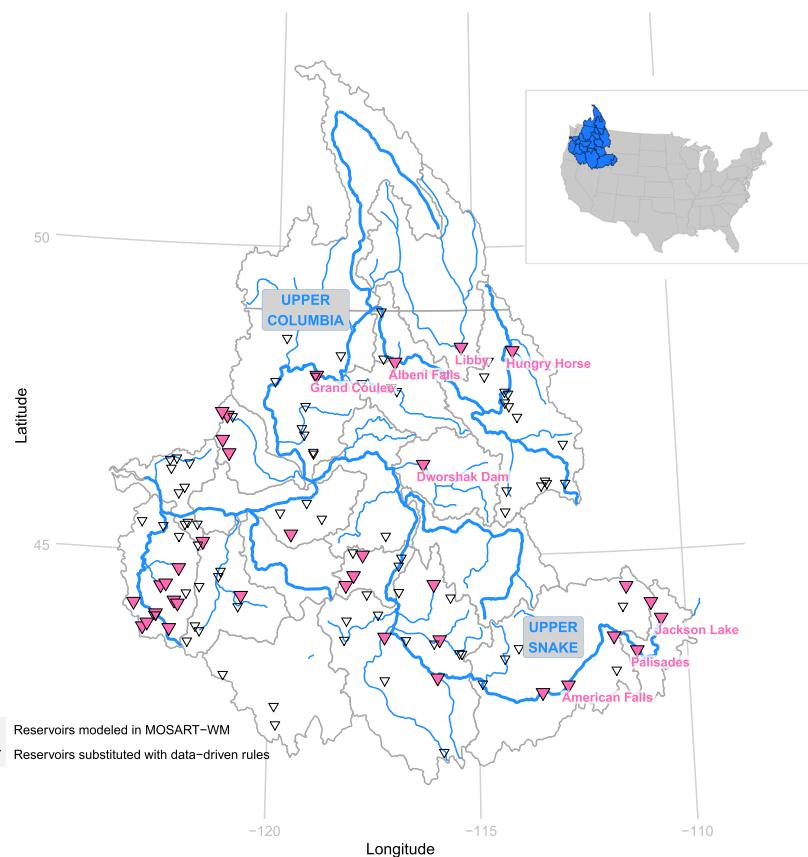


Figure 1. Study region. Columbia River basin, with 117 dams modeled in MOSART-WM. The data-driven operating rules with and without forecasts are implemented on the 36 dams highlighted. The eight labeled dams are those with storage capacity greater than 1,000 MCM.

accompany other well-documented errors that afflict an LHM simulation, including climatic forcing (Beck et al., 2017), runoff generation (reflecting the combined error from a variety of physical process representations) (Fekete et al., 2012), flow routing errors, and, of course, all upstream reservoir operations.

The generic release scheme deployed in MOSART-WM follows Hanasaki et al. (2006) and Biemans et al. (2011), with the additional enhancement of dynamic storage targets (Voisin et al., 2013). For flood control reservoirs, a monthly release pattern is initialized to match long-term annual inflow adjusted for interannual variability. Monthly releases are further adjusted based on long-term mean monthly water demands expected to be fulfilled with releases from this reservoir. For irrigation reservoirs, monthly release patterns are derived for each reservoir so as to store as much water as possible before the irrigation season and then release during the irrigation season following monthly water demand patterns. Depending on the reservoir capacity and ability to store the inflow and match the release patterns, daily releases are further adjusted for minimum environmental flow, minimum and maximum reservoir capacity. Dynamic storage targets allow for release patterns for irrigation and flood control purposes to be combined. These operations aim to reduce flood-season spill and then target high storage levels prior to the irrigation season. Releases are constrained on a daily time scale to account minimum environmental flows and minimum and maximum storage constraints. The approach deployed in MOSART-WM is common, and our study is generalizable across the majority of LHMs that include reservoirs.

To simulate reservoir releases with a data-driven scheme, we substitute generic operations with a constrained linear piecewise function parameterized for each of the 36 reservoirs in the CRB with sufficient observational records (see Turner et al., 2020). This includes eight very large reservoirs with storage capacity greater than 1,000 MCM, or 1 km^3 (Figure 1). Records of reservoir inflows, releases, and storages are

Table 1
Summary of Experiments Conducted

Experiment	Type	Forcing	Reservoir scheme
1	Off-line	Observed inflows and observed storage	A. Generic B. Data-driven, 1-week inflow C. Data-driven, dynamic inflow horizon
2	Off-line	Observed inflows	A. Generic B. Data-driven, one-week inflow C. Data-driven, dynamic inflow horizon
3	Online	Inflows simulated in LHM	A. Generic B. Data-driven, one-week inflow C. Data-driven, dynamic inflow horizon

obtained from the US Bureau of Reclamation (2019a, 2019b) and US Army Corps of Engineers (2016). The data-driven scheme consists of 52 piecewise functions, which are trained for each dam using relatively recent data (post-1995) to avoid training to outdated decision schemes, such as past rules involving less stringent environmental flow requirements. The piecewise function defines water release as a function of current-period reservoir storage plus incoming water. A different function is trained for each week of the water year (thus 52 functions). The functions are interpretable and parsimonious, while being constrained to realistic reservoir operator behavior—a strategy we use to avoid overfit in a data-scarce environment inhospitable to split-sample calibration validation (see Turner et al., 2020) (a k -fold cross-validation has been performed to ensure the efficacy of this strategy—see the supporting information). The scheme can be simulated at daily resolution by determining the week-ahead release on each day, implementing a day's worth of that release, recalculating a new week-ahead release the following day, and so on. The use of available water (storage plus inflow) rather than storage only is in part designed to avoid overfrequent depletion and spill of reservoirs with low storage relative to annual inflow (see Shin et al., 2019), since in such systems the available water will made up predominantly by inflow to drive a corresponding release.

A unique aspect of this data-driven scheme is that operations can be simulated with a “horizon curve” that is generated during the training procedure. The horizon curve specifies the inferred forecast lead time (in number of weeks) being adopted by the operator. This lead time varies by throughout the year in accordance with availability of predictive information for making a release decision. For example, early spring is typically associated with longer forecast lead times, owing to the availability of snowpack information that provides skillful, long-range inflow forecasts. The horizon curve specifies this change in forecast lead time adopted as a function of week of water year. When a horizon curve is implemented in the operating policy, release is a function of storage plus inflow out to the lead time specified by the horizon curve. The release decision is simulated in practice by assuming that the observed future inflow (perfect forecast) may act as a proxy for the forecast available to the operator at the time of making a release decision. One problem is that future inflows are not known in advance in an LHM simulation; they are generated by the model itself during simulation. An iterative approach is therefore required to determine the perfect inflow forecast for a given dam on each day (similar to the approach used in Haddeland et al., 2006, which employs a 12-month forecast). Running MOSART-WM in iterative mode involves extracting from each run (1980–2011) the simulated inflows to each reservoir, determining from these inflows the inflow forecast for the specified horizon at each dam, and then feeding those inflow forecasts back into the next simulation. As the iterations progress, the forecasted flows begin to converge, starting with upstream dams and progressing downstream. A simulation that neglects the horizon curves is run first to initialize the inflow forecasts. Convergence is achieved in 5 to 10 iterations, reflecting the degree of cascading of the reservoirs updated in the simulation.

2.2. Experimental Setup

Three experiments are conducted to determine the performances of the release models in different settings of interest (Table 1). Experiments 1 and 2 are conducted off-line: Each reservoir is simulated in isolation (i.e., not within the LHM) and is forced with observed daily inflow time series. In Experiment 1, the storage is reset to observed storage each day. This setting provides the clearest indication of the model accuracy,

because simulated releases are determined at each time step using information consistent with that used by the operator in reality. In Experiment 2, storage is fully simulated, so errors in release are allowed to accumulate and compound through time as simulated storage diverges from observed. This experiment provides a useful benchmark that will allow us to determine the contribution of storage error accumulation to model performance deterioration.

Experiment 3 is conducted online. The candidate data-driven models for 36 dams are embedded in MOSART-WM and are forced with simulated flows. The differences in model performances between Experiment 3 (online) and Experiment 2 (off-line) indicate the extent to which inflow biases inherent in the hydrological model affect the performance of the reservoirs under different operating configurations. This is important because the candidacy of data-driven release schemes has thus far been based on assessments of their performances in off-line mode (Coerver et al., 2018; Turner et al., 2020; Yassin et al., 2019) or against a benchmark of no reservoir representation (Yassin et al., 2019).

Each experiment tests three distinct reservoir release schemes: generic (Scheme A), as described in section 2.1; data-driven based on time of year, storage and 1-week (current period) inflow (Scheme B); and data-driven based on time of year, storage, and inflow forecasts of varying predefined lead times (dynamic horizon—Scheme C), using the horizon curves reported in Turner et al. (2020). Inflow forecasts have been shown to strongly influence seasonal reservoir operations, particularly for high elevation dams in Western United States fed by snowmelt (Turner et al., 2020). We adopt the forecast-based model in this study to understand whether the benefits of representing forecasts are maintained when reservoirs are simulated within the LHM. The data-driven release scheme (C) deploys forecasts based on perfect flow ahead. For the off-line simulations (Experiments 1 and 2) the regulated inflow forcing is known in advance from the observations, so the perfect forecast can be extracted directly and deployed to inform the release decision. For the online experiment, the LHM must be simulated using the iteration mode described in section 2.1. The parameters for both data-driven models (Schemes B and C) are optimized off-line using observations and held constant across all experiments. The generic scheme (Scheme A) is calibrated using long-term mean inflow with the reservoir-specific demand parameter held consistent across experiments.

2.3. Model Evaluation

2.3.1. Daily and Seasonal Metrics for Releases at Individual Reservoirs

For each experiment, models are evaluated using goodness-of-fit metrics computed at daily resolution using simulated versus observed time series of releases. We use release rather than storage because the release integrates storage errors and release decision error (since release is a function of storage in the data-driven reservoirs schemes). The computed metrics are the commonly used normalized root mean squared error (nRMSE), the transformed (normalized) RMSE (nTRMSE), and the slope of the flow duration curve (SFDCE). There are numerous ways in which the RMSE can be normalized; here we divide by the standard deviation of observations. For the nTRMSE, the release time series are first transformed so that the result is weighted by performance during periods of low release (we use Box-Cox transform with exponent of 0.3, as adopted in Van Werkhoven et al., 2009). The SFDCE is absolute error in the slope of the middle section of the flow duration curves between 30th and 70th percentile of flows, capturing error in the variability of the releases (Van Werkhoven et al., 2009). To evaluate model performance during drought, we evaluate the observed and simulated cumulative release volumes and then compare these volumes for all dams on a log-log plot (observed vs. simulated). The release volumes are computed for two separate time periods. The first is during the summer (July through September) that follows the spring (April through June) with the lowest inflow across all years. This captures the model performance for the season of largest water demand during the year with weakest flow augmentation and storage recharge from snowmelt. The second time period analyzed is the full water year (October–September) with the lowest overall inflow, capturing the model's propagated performance across all seasons with competing objectives such as wintertime storage drawdown for flood control and spring refill for water supply. These lengthy periods are deemed sensible due the prolonged nature of drought and its effect on reservoir storages.

2.3.2. Propagation to Water Supply Risk Metrics

Two analyses are conducted to understand the impacts of these reservoir policies on likelihood and severity of water supply shortfall. The first risk analysis adopts the online LHM simulation (Experiment 3). Here we

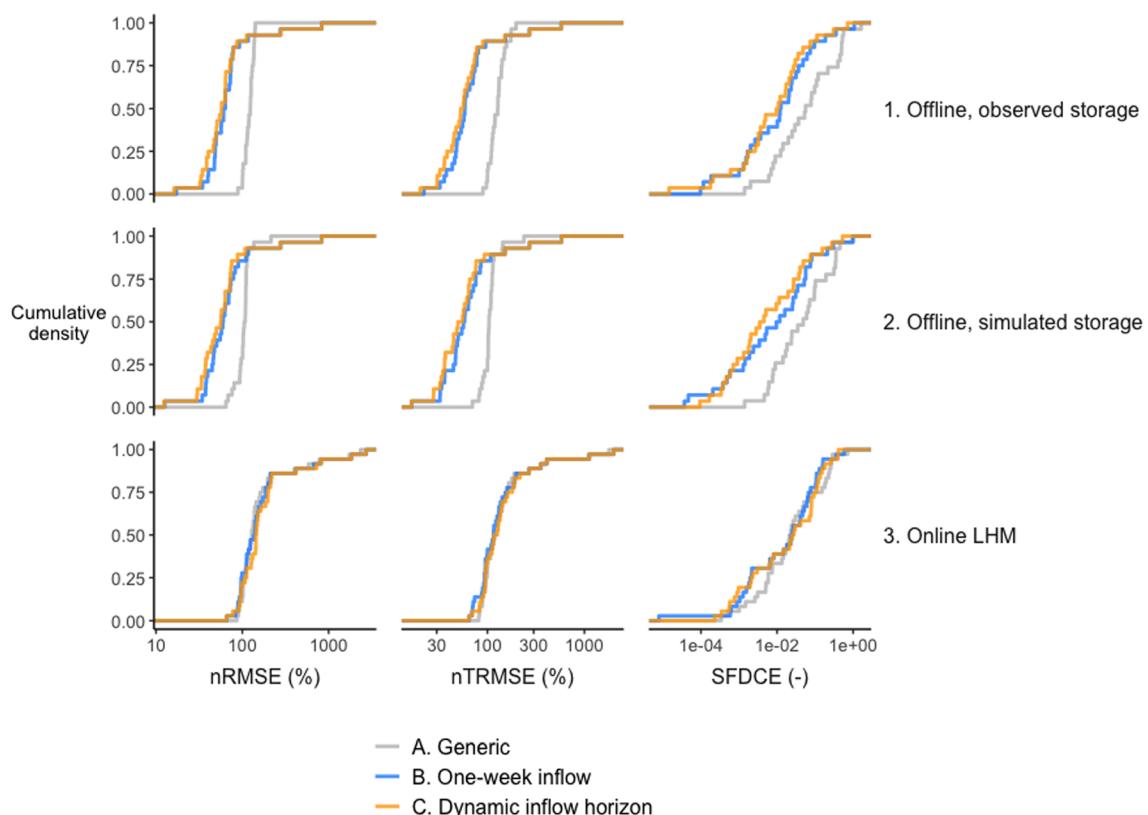


Figure 2. Empirical cumulative distribution plots across all 36 dams for each experiment and performance metric.

explore water supply risk from the record of demand shortfalls simulated in the model. Each grid cell of the LHM is associated a time-varying demand for water, which is met from local river abstractions and from local reservoir releases. If water is unavailable to meet that demand, a shortfall is registered. We compute the reliability, resilience, and vulnerability of simulated water supply at each grid cell, using the definitions provided in Hashimoto et al. (1982) and updated in McMahon et al. (2006). Reliability is number of days experiencing shortfall as a proportion of total days of simulation (reliability = 1 implies no shortfall periods). Resilience is the inverse of the average duration of continuous supply shortfall sequences (resilience = 1 indicates all shortfall periods are of 1-day duration). Vulnerability is the average severity of shortfall (as a proportion of demand) measured across all continuous shortfall sequences (vulnerability = 1 indicates total shortfall incurred at some point in all shortfall sequences). There are no observational records of water demand shortfall that can be used to determine which reservoir scheme best reproduces these metrics, so this part of our analysis reveals little about model accuracy. Instead, it indicates the impact of alternative water reservoir schemes on metrics that are important for determining the vulnerability of human systems to drought.

The second analysis is conducted off-line, focusing on the storage behaviors of the eight dams included in the study with storage capacity greater than 1,000 MCM. Each reservoir is simulated with each scheme (A, B, and C) using both the observed inflow data (as in Experiment 2) and with 100 distinct, 30-year inflow replicate sequences with weekly resolution. These inflow replicates are generated using the nonparametric nearest-neighbor bootstrap (Lall & Sharma, 1996). We measure water supply risk using the number of simulated years for which reservoir levels are drawn below active storage, which would potentially lead to severe supply curtailment. The inflow replicates sequences provide uncertainty distributions on this metric, which are used to infer robustness of differences across reservoir schemes. These results are therefore intended to examine the importance of water reservoir representation for emerging applications of LHMs relating to drought impacts on regional sectors, like water deliveries to agriculture.

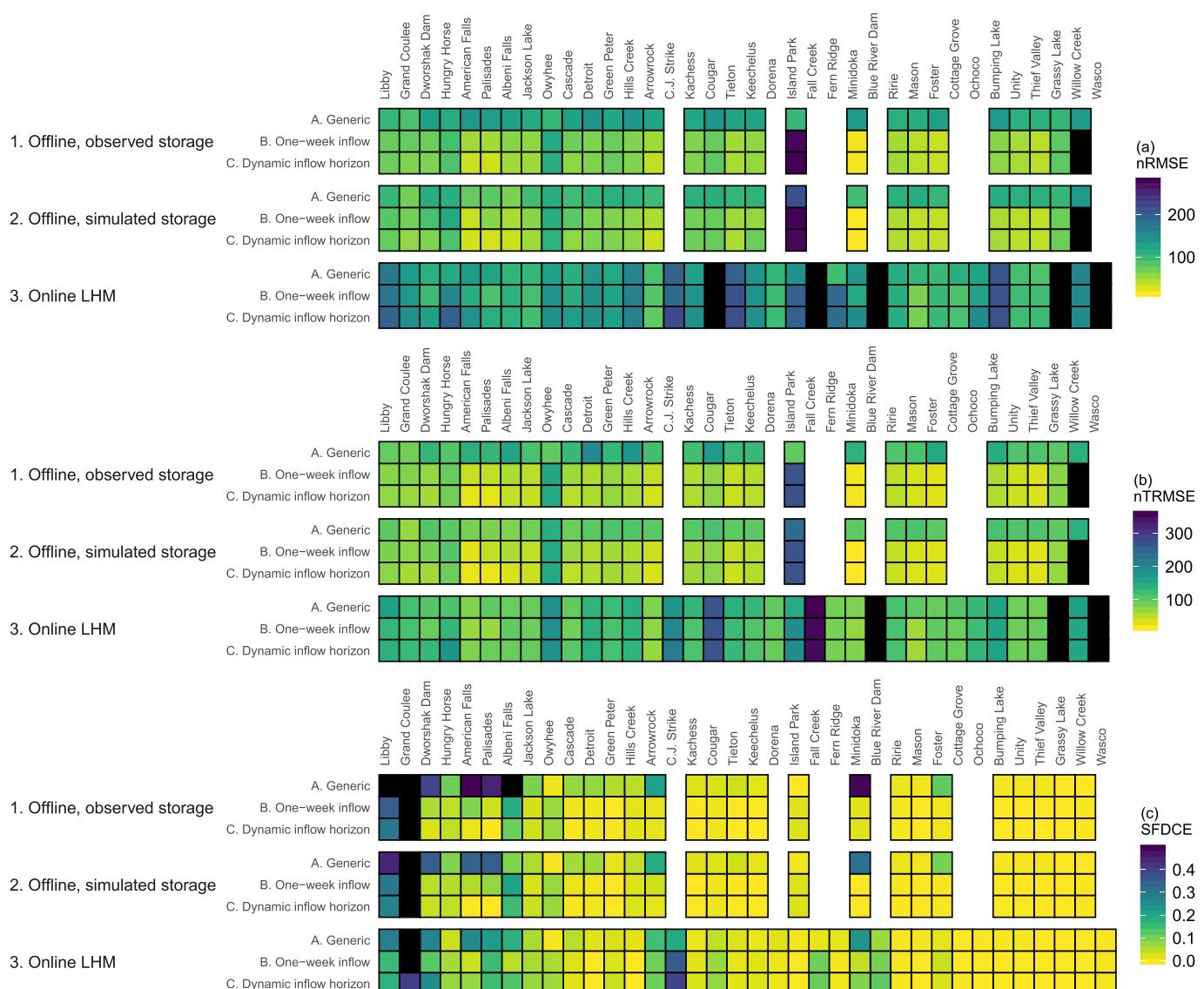


Figure 3. Model performances for all configurations, computing by (a) normalized root-mean-squared error (nRMSE, %), (b) normalized transformed RMSE (nTRMSE, %), and (c) slope of the flow duration curve error (SFDCE, —). Results for seven dams are missing for the first two experiments where daily observed inflows are unavailable. Dams are ordered (left to right) by storage capacity (largest to smallest).

3. Results

3.1. Model Performances

Results demonstrate a powerful influence of LHM inflow simulation error on the performances of reservoir schemes deployed. For off-line Experiments 1 and 2 (observed inflows), the data-driven release scheme without forecasting (Scheme B) significantly outperforms generic rules (Scheme A), while the addition of dynamic forecast horizons in the data driven model (Scheme C) provides further marginal improvement (Figure 2). The cumulative distribution of model performance metrics across all dams indicates a robust reduction (>90% of the CDF) in simulation errors for the off-line experiments. In other words, the additional efforts of acquiring operational data and training bespoke models for these reservoirs appear to pay off handsomely. However, this finding does not hold in Experiment 3 (online, LHM simulated flows). LHM simulations indicate no robust improvements, with cumulative distribution functions of goodness-of-fit scores closely aligned and overlapping.

Scores obtained for each individual dam are displayed in Figure 3 (dams are ordered according to storage capacity, with largest to the left). For off-line Experiments 1 and 2, both nRMSE (Figure 3a) and nTRMSE (Figure 3b) are reduced in 26–27 out of 29 cases depending on experiment and metric (daily inflow records are unavailable for eight of these reservoirs, so results are reported for only 28 of 36 dams in the off-line experiments). The two cases for which data-driven rules score significantly higher nRMSE and nTRMSE (Island Park and Willow Creek) indicate inconsistent inflow, release, and storage level data employed in model training (a data-driven scheme is only viable if the available operational data are accurate). The addition of the horizon curve to the data-driven model (Experiment 1, Scheme C) improves simulation performance on Experiments 1B and 2B in 75–95% of dams (depending on metric evaluated). For some dams, like American Falls and Palisades of the Upper Snake, the performance improvements are substantial, with an approximate 40% reduction in nTRMSE scores. These results further demonstrate the importance of considering the contribution of inflow forecasts to release decision making (expanding on the evidence presented in Turner et al., 2020). Very high SFDCE scores for the large dams under the generic scheme (Experiment 1, Scheme A) are reduced markedly by the data-driven schemes (Schemes B and C) (e.g., more than 50% reduction in SFDCE for Dworshak, American Falls, and Palisades).

Both nRMSE and tRMSE deteriorate significantly when using the LHM simulated flows (Figures 3a and 3b). This is an expected result, because the inflows generated within the LHM simulation are subject to significant error. The perhaps unexpected result is that both nRMSE and nTRMSE are relatively uniform for each reservoir across the three models tested (reflecting the result reported for Experiment 3 in Figure 2). Also interesting is that, contrary to marginally improving simulation results, the addition of the forecast (exposed by the difference between Schemes B and C) appears to erode performance in a number of cases in Experiment 3. At Dworshak Dam, for example, the addition of the forecast reduces the nRMSE by 10% in Experiment 1 and by 30% in Experiment 2. Forecasts have the opposite effect when the models are simulated online in the LHM (Experiment 3), increasing nRMSE by about 30%. In section 4, we shall explore and illuminate the inflow conditions under which an accurate, forecast-based model will tend to underperform relative to a less accurate model that neglects forecasts. This result carries significant ramifications for the applicability of forecast-based schemes to inform water releases in LHM.

This finding that the benefits of a data-driven model are partially lost in the LHM is supported further if we examine the ability of models to reproduce cumulative release volumes under conditions of drought (Figure 4). In off-line mode, the data-driven model improves vastly on the generic scheme, which has a tendency to underestimate reservoir outflows during drought (Experiment 1 in particular). This is likely caused by underestimation of the minimum flow constraint to meet environmental provisions in the CRB. Results here demonstrate the benefit of the data-driven model in capturing this feature correctly. Nonetheless, the drought-based release volumes recorded for Experiment 3 (LHM) exhibit similar levels of error across both generic (Scheme A) and data-driven schemes (Schemes B and C). For example, the total summer release following low-flow spring conditions (Figure 4a, top-right panel) is overestimated for Grand Coulee, Libby, Hungry Horse, and Albeni Falls but underestimated for Jackson Lake and Palisades. Although not identical in magnitude, the direction of error is similar with the data-driven models (Figure 4a, bottom-right panel), suggesting that upstream inflow biases of the LHM are more influential than the differences in reservoir model structure in driving simulated releases under conditions of drought. Similar results emerge when we compare total water releases for driest water year of record (Figure 4b)—demonstrating that the findings are not the chance feature of drought metric examined.

3.2. Impact of Release Scheme on Water Supply Reliability, Resilience, and Vulnerability

In MOSART-WM, each grid cell is assigned water demands that can be supplied through local stream abstraction or water delivery from a reservoir within a prespecified distance. Whenever this demand cannot be met by available water, a shortfall event is recorded. The record of shortfall during such events can be used to characterize the reliability, resilience, and vulnerability of water supply (see section 2.3.2). We find that for Experiment 3 (LHM simulation), resilience and vulnerability metrics are available only within the Upper Snake region, because no supply-demand shortfalls occur elsewhere during the study period of 1980–2011 (Figure 5). Shortfalls occurring in the Upper Snake arise during a single dry year, and so the calculated scores for reliability, resilience, and vulnerability should be understood in this limited context. We find a marginal difference in reliability across reservoir schemes, while resilience and vulnerability

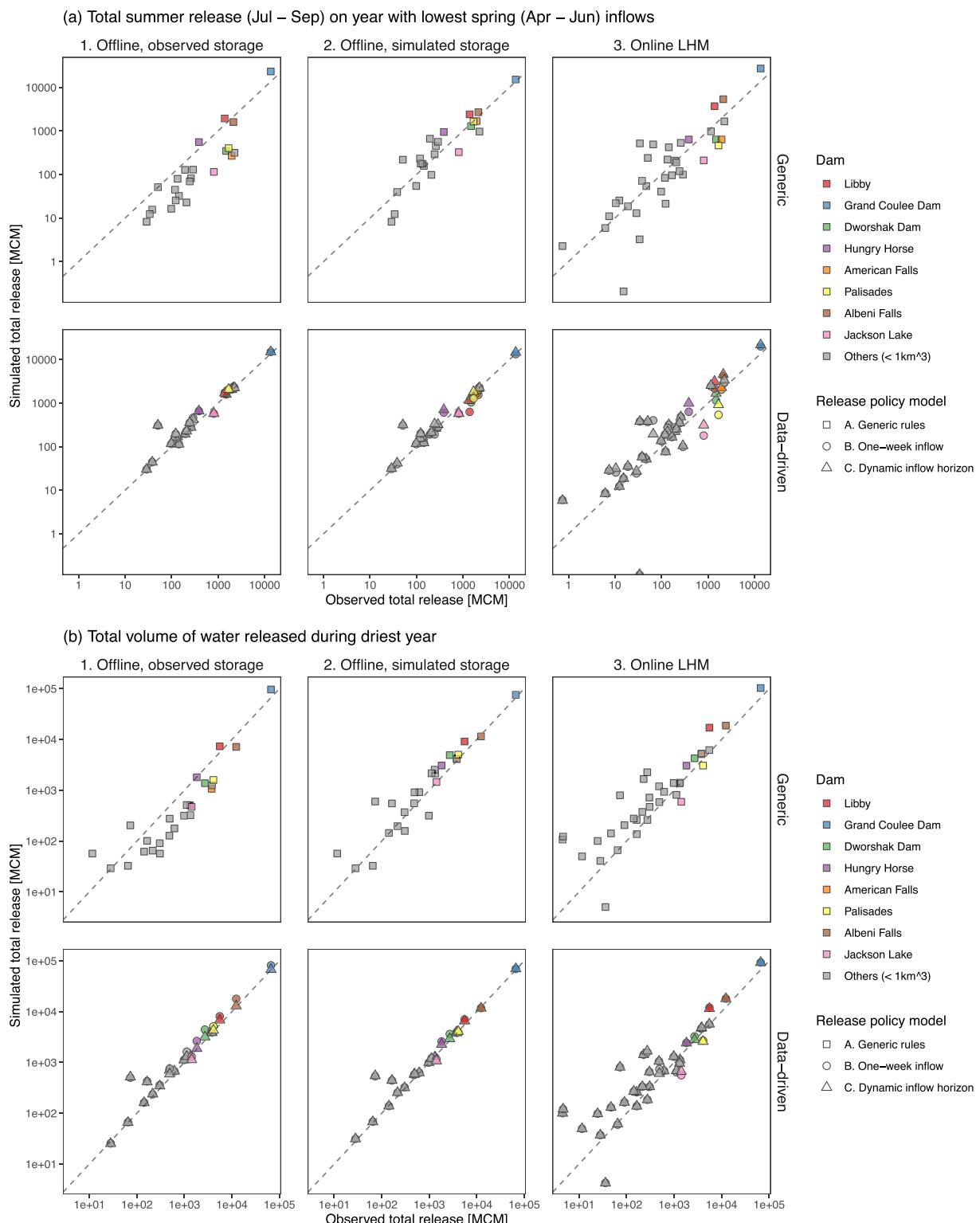


Figure 4. Simulated versus observed cumulative water releases across all experiments for two drought conditions.

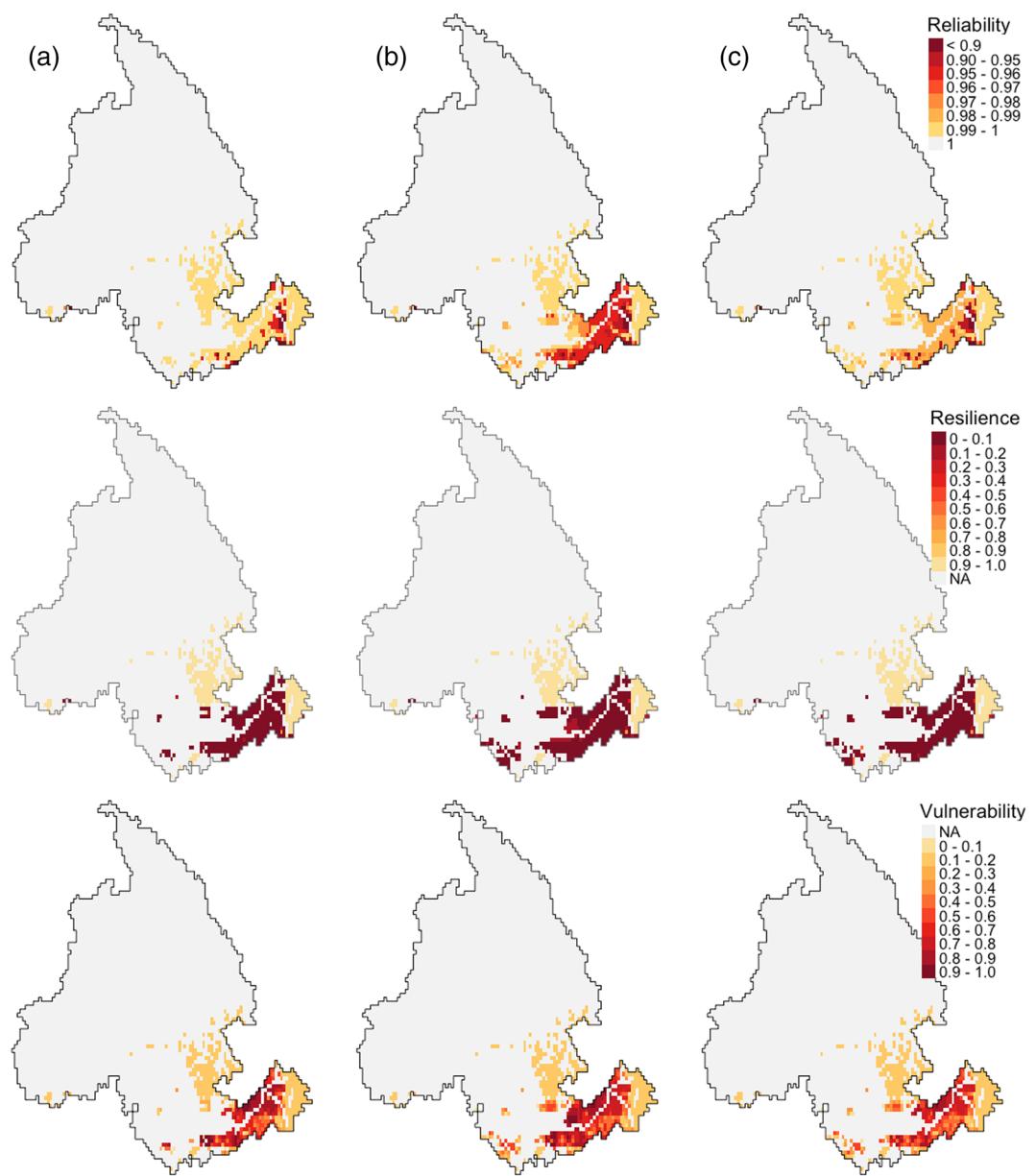


Figure 5. Implications of different reservoir release schemes on reliability, resilience, and vulnerability. Resilience and vulnerability are computed from the time series of shortfalls between supply and demand. These metrics are assigned NA values where reliability is equal to 1 (since the metrics cannot be computed without a shortfall period). Schemes are A (generic), B (data-driven, 1-week inflow), and C (data-driven, dynamic inflow forecast).

remain relatively consistent. This result is in line with the drought year analysis reported in Figures 3 and 4, wherein the nTRMSE (a metric weighted to low-flow periods) and the drought-year cumulative releases are similar across reservoir schemes for Experiment 3.

Despite the marginal differences in vulnerability reported above, stochastic storage simulations for the eight large dams (>1,000 MCM storage capacity) reveal marked differences in likelihood of reservoirs being drawn below active storage capacity—an indicator of vulnerability, since loss of active storage would likely be associated with significantly impaired ability to meet water demands (Figure 6). The impact is most striking for Dworshak Dam, which is rarely drawn below active storage with the generic release scheme (A) and always drawn below active storage (30 out of 30 years in all 100 simulations) when simulated using data-driven schemes (B and C). In each of the other dams except Libby and Hungry Horse, which remain within

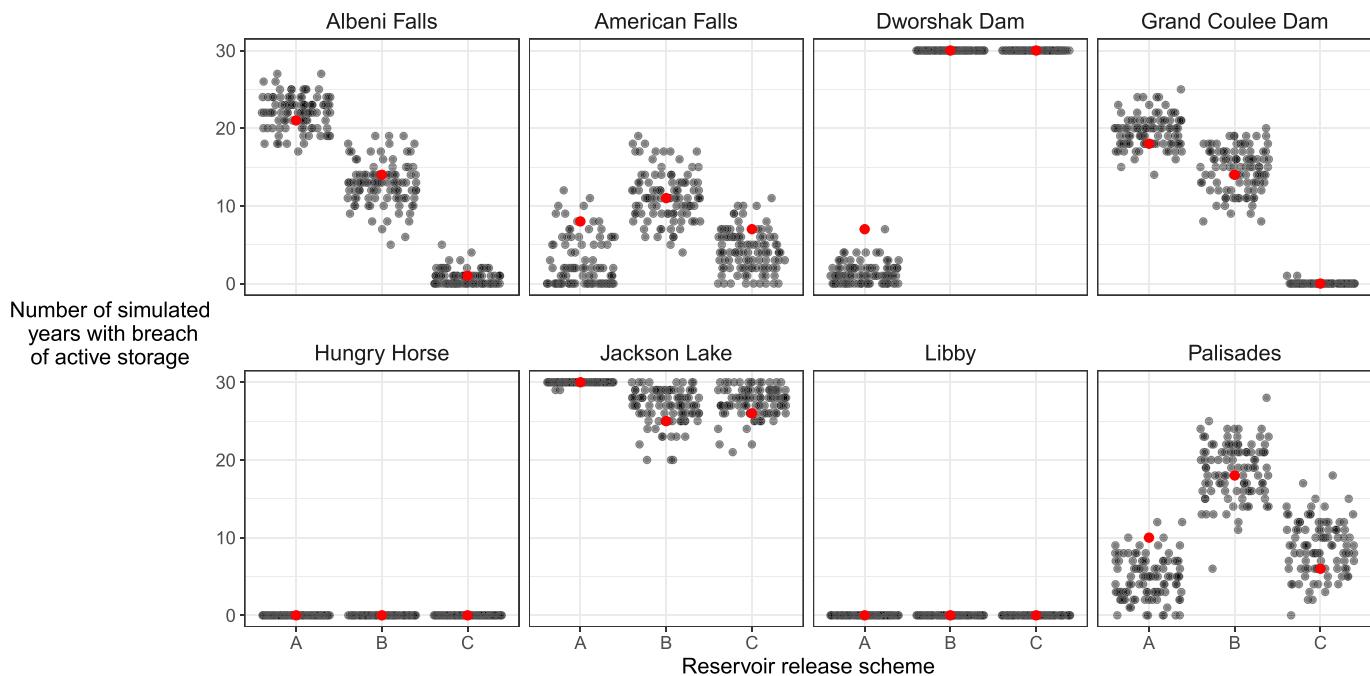


Figure 6. Comparison of release schemes across observed flows (red point) and using 100, 30-year replicate sequences (black points) for number of simulated years with a breach of active storage. For a given reservoir, each scheme is exposed to the exact same set of inflow sequences. Within each column (A, B, and C) points are positioned randomly on the horizontal axis to avoid overlap.

active storage limits across all simulations, the release scheme causes a clear difference in the distribution of the number of years resulting in a breach of active storage. Although the direction of impact varies from case to case, these results indicate strong sensitivity of reservoir vulnerability to release scheme chosen. Interestingly, these differences are often unexposed when comparing schemes using only the observed inflow sequence (red points). Simulation of observed inflows in each of the schemes as applied to American Falls, for example, lead to a similar number of years with drawdown below active storage. The robust difference between schemes emerges clearly, however, with the simulation across 100 replicate sequences.

4. Discussion and Conclusions

4.1. Need and Prospects for Data-Driven Reservoir Schemes in LHM

LHMs address a host of science problems at a range of spatial domains, from river basins to the entire globe. These models have been used to evaluate global water availability (Bierkens, 2015; Döll et al., 2012; Wada et al., 2011), explore impacts of humans and climate on terrestrial water storage and groundwater depletion (Wada et al., 2014), support water forecasting models (Koster et al., 2004), and corroborate remotely sensed water fluxes (Scanlon et al., 2018). LHMs will be relied upon increasingly to provide water availability projections for the study of climate and extreme event impacts on expansive infrastructural systems and economic sectors (Kraucunas et al., 2015; Miara et al., 2017). Integrated power grids or crop-growing regions, for example, often span multiple river basins and political regions. If one wishes to quantify the impacts of a climate trend or extreme event on these water-dependent sectors, then water availability must be simulated coherently throughout the entire spatial domain to preserve realistic spatial and temporal correlation patterns (Voisin et al., 2017). Although LHMs provide this capability, their clout is limited by coarse spatial resolution, low fidelity, and lack of observational data for calibration and verification—particularly for human decision making and when flood and drought impacts are relevant (Nazemi & Wheater, 2015a, 2015b). The question is whether these new applications envisaged would benefit from LHMs furnished with more accurate, bespoke, data-driven water management schemes, where data permit.

The results reported above offer three insights relevant to the advancement of data-driven reservoir schemes in LHMs. First, data-driven release-availability functions—despite being more accurate as measured by

standard error metrics—may not guarantee more realistic reservoir releases in an LHM simulation. They may only provide more accurate releases subject to the inflows, which are often significantly biased. This does not mean the community should avoid implementing data-driven rules in LHMs. Data-driven reservoir release rules would enrich LHMs, even if the benefits cannot yet be realized in full. However, it is essential for users of these models to be aware of the limitations; adopting data-driven rules may not yet guarantee more realistic streamflow simulation, nor reservoir variations for water quality modeling purposes, for example. Continual advancement of other LHM features that control streamflow will be necessary to harness the benefits of realistic, data-driven reservoir release schemes.

A second insight is that water supply vulnerability is highly sensitive to the choice of reservoir scheme—but that the associated implications may not readily emerge from an LHM unless forced with many drought sequences to generate a sufficiently large and representative sample of shortfalls between supply and demand. Taking full advantage of data-driven reservoir schemes to study water supply vulnerability implies a need for expansive uncertainty analysis and stress testing. In water supply reservoir operations, there is a trade-off between reliability and vulnerability (Hashimoto et al., 1982). The operator may deliberately cut back the water supply at the onset of drought (causing a minor shortfall) in order to hedge against the risk of a much larger shortfall if releases were maintained (Draper & Lund, 2004). The system is then less reliable (shortfalls are more frequent) but also less vulnerable (shortfalls are less severe). Capturing this type of behavior in an LHM reservoir scheme is essential for adequate characterization of water supply risk during drought—and, importantly, the generic scheme (some variants of which judge hydrological condition only at the beginning of the operational year) is likely at a significant disadvantage for representing such conditions. Yet results in this study suggest that the import of a different reservoir scheme may only be exposed through a process of stress-testing involving a large sample of input scenarios—a nontrivial challenge at the scale of a regional or global LHM. These models ingest distributed climate input across thousands of grid cells; such data cannot be easily synthesized without the aid of extremely computationally intensive climate models. Off-line analytics, such as the approach proposed in this work involving stochastic simulation of reservoirs, could be performed in postprocessing of an LHM to support more in-depth analysis of water supply vulnerability. For example, one could extract the LHM inflow simulations for a set of independent, key reservoirs and then expose those reservoirs to spatially correlated synthetic inflow sequences derived from a multisite stochastic flow model.

While we do advocate for further development and testing of data-driven water management schemes for LHMs, we would caution against the use of models that include forecast contribution to the release decision at this time. The third key insight from our analysis is that data-driven models that adopt a forecast horizon may actually erode the performance of an LHM, despite being marginally more accurate when simulated with observed inflows. This finding carries important implications for the prospects of adopting realistic reservoir schemes in LHMs. An explanation is offered in the following section.

4.2. Sensitivity of Reservoir Simulation Performance to Inflow Error

We shall now demonstrate through a small numerical experiment that a forecast-based model tends to be less robust to error in inflow than a similar data-driven model that neglects forecasts. This is an off-line experiment akin to Experiment 2 (observed inflow and simulated storage), with the difference that errors are added to the inflow to explore sensitivity of performance with respect to level of error. For each of the eight largest reservoirs included in our study, we perform the following sensitivity test. First, an inflow error time series is computed by subtracting the daily observed inflow from the daily inflow simulated by the LHM. We then develop a stochastic model of the resulting error time series. This is done by deseasonalizing the data and then fitting a first-order autoregressive model (AR1) to the daily time series to remove autocorrelation. Error time series with similar error can then be synthesized by bootstrapping from the model residuals and then reapplying autocorrelation and the seasonal signal. On each of 1,000 error series generated, we scale the data by a randomly sampled error factor (between 0 and 1.5, to be applied to all points of the flow time series) before adding the error time series back onto the observed inflows. A sample with error factor of 1 has error commensurate with the LHM simulated inflow, while an error factor of 0 has no error and is equal to the observed inflow (i.e., equivalent to Experiment 2). We simulate each reservoir 1,000 times with these varying inflow series. This is done for the data-driven release scheme with 1-week inflow (Scheme B) and again with the dynamic inflow horizon (Scheme C). Normalized RMSE scores are

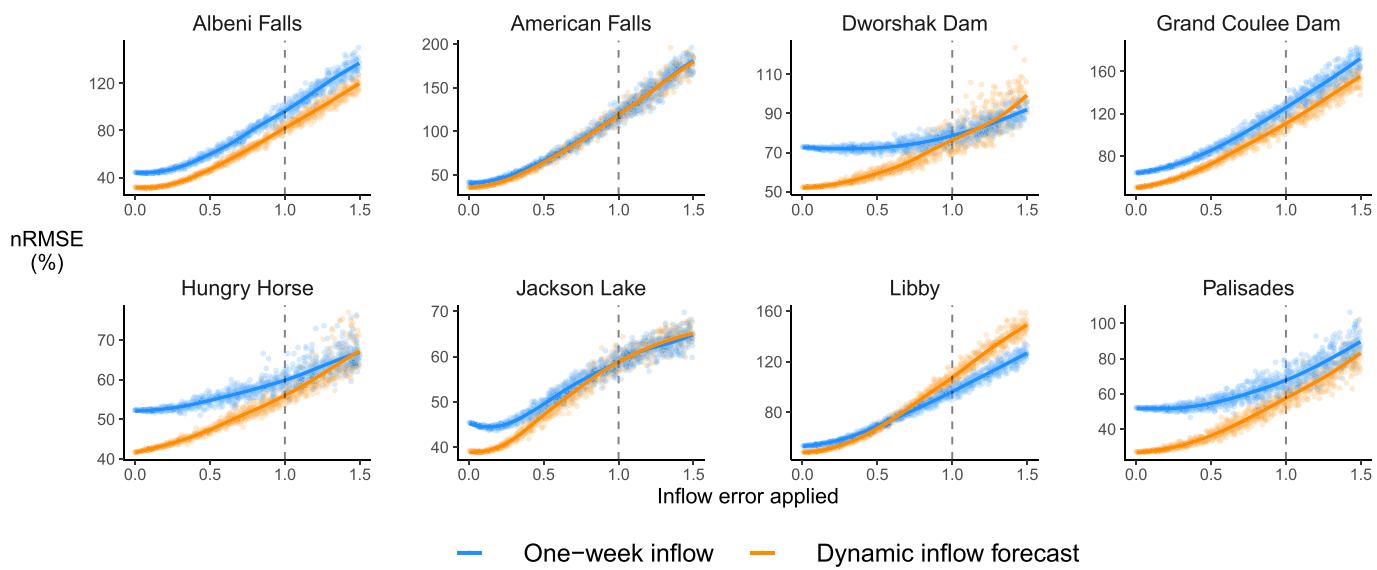


Figure 7. Stochastic error simulation results for 1,000 simulations using Scheme B (data-driven, 1-week inflow) and C (data-driven, dynamic inflow forecast). The horizontal axis represents the error factor, where 0 is commensurate with observed inflow and 1 is commensurate with the inflow error in the LHM. Values greater than 1 on the horizontal axis denote inflow errors exceeding the LHM errors. nRMSE (%) is computed on simulated release relative to observations.

computed for simulated releases relative to observed (Figure 7). When the error is 0, results are unequivocal: The forecast model (Scheme C) better represents the observed release. But as inflow error is introduced (moving from left to right along each horizontal axis), something interesting happens. For most reservoirs, the nRMSE of scheme C release rises at a sharper rate than that of Scheme B. The reason for this behavior is quite intuitive: Scheme C relies more heavily on inflow to inform its decision, so errors in that input must impact that model to a greater extent than a model less reliant on those inputs. When the error reaches the equivalent error of the LHM (error factor = 1) the forecast-based scheme (C) is outperformed by the less accurate Scheme B in half of these reservoirs—namely, at American Falls, Dworshak Dam, Jackson Lake, and Libby. The extent of this behavior varies depending on the importance of forecasting and the coincidence of inflow bias with seasons when forecasts are deployed in

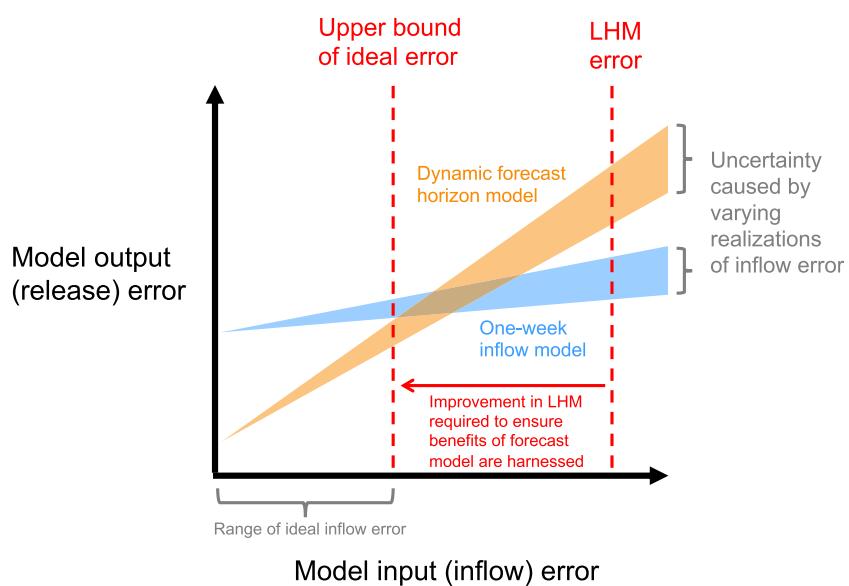


Figure 8. Conceptual model describing the relationship between reservoir inflow error and model output error, highlighting the greater sensitivity of a forecast-dependent scheme to errors in inflow. An error level that guarantees superiority of the forecast-driven scheme is denoted as “ideal error.”

the release model. Albeni Falls, for example, is simulated with significantly shorter forecast lead times than the other reservoirs, leading to a very marginal difference between the two models in terms of sensitivity to inflow error with this particular reservoir.

This sensitivity test replicates the observed results of our study, wherein the addition of the forecast actually deteriorates the performance of some reservoirs simulated in the LHM. One can generalize this result to a simple conceptual model to highlight how a level of improvement in LHM inflow could allow the modeler to harness the benefits of a forecast scheme (Figure 8). This concept is somewhat analogous to the findings of reservoir optimization research, wherein deterioration of forecast skill beyond a certain point compromises the optimality of release decisions (Turner et al., 2017; Zhao et al., 2019).

4.3. Conclusions and Future Work

Data-driven reservoir release schemes have emerged as a possible enhancement to LHMs—a natural progression given the increasing availability of observations from in situ devices and remote sensing. Today, data-driven reservoir schemes can be implemented in LHMs wherever sufficiently lengthy operational records are available to train the parameters. In future, the availability of reservoir storage levels and bathymetry from satellite remote sensing may unlock data-driven reservoir schemes at a global scale (Avisse et al., 2017; van Bemmelen et al., 2016; Zhao & Gao, 2019). But are LHMs sufficiently advanced to take full advantage of these new data? The results of this study suggest perhaps not.

Our results demonstrate that an interpretable, data-driven scheme based on release-availability functions simulates reservoir releases more accurately than generic release scheme—including during the driest years of record. The addition of seasonally varying forecast horizons in a data-driven model offers further improvement still. But these simulation improvements are available only under conditions of accurate model inputs. When exposed to the errors and bias inherent in LHM flows, the data-driven reservoir release schemes perform no better than generic. When inflow forecasting is included, data-driven schemes may even perform worse, owing to overreliance on inputs subject to significant bias. Further corroborating evidence generated using alternative LHMs, model inputs (e.g., water demands), spatial regions, and data-driven schemes is required to strengthen and confirm these conclusions. Nonetheless, the results indicate a need for further improvement of LHM flow accuracy, requiring attention on runoff generation, groundwater and soil moisture interaction, evaporation, snow cover, and other elements of human influence on the water cycle.

Our study shows that key drought metrics, such as reliability and vulnerability, are sensitive to the reservoir scheme deployed but that these differences may only be exposed through extensive sensitivity testing uncommon in LHM study. Drought impact studies that rely on improved reservoir models in LHMs may therefore require off-line sensitivity analysis using stochastic inflow sequences. The parameters of the reservoir operating rules—data-driven or otherwise—will also be subject to significant uncertainty. Uncertainty characterization and sensitivity analysis performed on the reservoir model parameters (as in Zajac et al., 2017) will be particularly important in regions lacking records of reservoir inflow and release. The appropriate tools and methods for executing such analysis in the context of large-scale water resource simulation could be defined in future work. Improvements to LHM flow simulation accompanied with data-driven human-system models and appropriate uncertainty analysis could unlock new capabilities, including rigorous drought impact assessment across large regions, and better representation of water resources management accounting for localized institutional and environmental regulations.

Data Availability Statement

Hydrologic simulations (Livneh data) were provided by the NOAA/OAR/ESRL PSL, Boulder, Colorado, USA, from their web site (at <https://psl.noaa.gov/>). Software used to generate data-driven reservoir policies is available on Github (<https://github.com/IMMM-SFA/horizon>). Simulation results supporting this study may be accessed online (at <https://zenodo.org/record/3983584>).

References

- Avisse, N., Tilmant, A., Müller, M. F., & Zhang, H. (2017). Monitoring small reservoirs' storage with satellite remote sensing in inaccessible areas. *Hydrology and Earth System Sciences*, 21, 6445–6459. <https://doi.org/10.5194/hess-21-6445-2017>
- Beck, H. E., van Dijk, A. I. J. M., de Roo, A., Dutra, E., Fink, G., Orth, R., & Schellekens, J. (2017). Global evaluation of runoff from ten state-of-the-art hydrological models. *Hydrology and Earth System Sciences*, 21(6), 2881–2903. <https://doi.org/10.5194/hess-21-2881-2017>

Acknowledgments

This research was supported by the U.S. Department of Energy Office of Science as part of the Integrated Multiscale, Multisector Modeling project (Grant No. 59534). This work was authored by the Pacific Northwest National Laboratory, managed by Battelle (Contract No. DE-AC05-76RL01830) for the U.S. Department of Energy.

- Behrens, P., van Vliet, M. T. H., Nanninga, T., Walsh, B., & Rodrigues, J. F. D. (2017). Climate change and the vulnerability of electricity generation to water stress in the European Union. *Nature Energy*, 2(8), 17114. <https://doi.org/10.1038/nenergy.2017.114>
- Biemans, H., Haddeland, I., Kabat, P., Ludwig, F., Hutjes, R. W. A., Heinke, J., et al. (2011). Impact of reservoirs on river discharge and irrigation water supply during the 20th century. *Water Resources Research*, 47, W03509. <https://doi.org/10.1029/2009WR008929>
- Bierkens, M. F. (2015). Global hydrology 2015: State, trends, and directions. *Water Resources Research*, 51, 4923–4947. <https://doi.org/10.1002/2015WR017173>
- Bonneville Power Administration, & US Army Corps of Engineers (2011). Climate and hydrology datasets for use in the RMJOC agencies' longer-term planning studies. Part III—Reservoir operations assessment: Columbia basin flood control and hydropower.
- Coerver, H. M., Rutten, M. M., & van de Giesen, N. C. (2018). Deduction of reservoir operating rules for application in global hydrological models. *Hydrology and Earth System Sciences*, 22(1), 831–851. <https://doi.org/10.5194/hess-22-831-2018>
- Dang, T. D., Chowdhury, A. F. M. K., & Galelli, S. (2020). On the representation of water reservoir storage and operations in large-scale hydrological models: Implications on model parameterization and climate change impact assessments. *Hydrology and Earth System Sciences*, 24, 397–416. <https://doi.org/10.5194/hess-24-397-2020>
- Döll, P., Fiedler, K., & Zhang, J. (2009). Global-scale analysis of river flow alterations due to water withdrawals and reservoirs. *Hydrology and Earth System Sciences*, 13(12), 2413–2432. <https://doi.org/10.5194/hess-13-2413-2009>
- Döll, P., Hoffmann-Dobrev, H., Portmann, F. T., Siebert, S., Eicker, A., Rodell, M., et al. (2012). Impact of water withdrawals from groundwater and surface water on continental water storage variations. *Journal of Geodynamics*, 59–60, 143–156. <https://doi.org/10.1016/j.jog.2011.05.001>
- Draper, A. J., & Lund, J. R. (2004). Optimal hedging and carryover storage value. *Journal of Water Resources Planning and Management*, 130(1), 83–87. [https://doi.org/10.1061/\(ASCE\)0733-9496\(2004\)130:1\(83\)](https://doi.org/10.1061/(ASCE)0733-9496(2004)130:1(83))
- Fekete, B. M., Loosar, U., Pietroniro, A., & Robarts, R. D. (2012). Rationale for monitoring discharge on the ground. *Journal of Hydrometeorology*, 13(6), 1977–1986. <https://doi.org/10.1175/JHM-D-11-0126.1>
- Haddeland, I., Heinke, J., Biemans, H., Eisner, S., Flörke, M., Hanasaki, N., et al. (2014). Global water resources affected by human interventions and climate change. *Proceedings of the National Academy of Sciences*, 111(9), 3251–3256. <https://doi.org/10.1073/pnas.1222475110>
- Haddeland, I., Skaugen, T., & Lettenmaier, D. P. (2006). Anthropogenic impacts on continental surface water fluxes. *Geophysical Research Letters*, 33, L08406. <https://doi.org/10.1029/2006GL026047>
- Hanasaki, N., Kanae, S., & Oki, T. (2006). A reservoir operation scheme for global river routing models. *Journal of Hydrology*, 327(1–2), 22–41. <https://doi.org/10.1016/j.jhydrol.2005.11.011>
- Hanasaki, N., Kanae, S., Oki, T., Masuda, K., Motoya, K., Shirakawa, N., et al. (2008). An integrated model for the assessment of global water resources—Part 1: Model description and input meteorological forcing. *Hydrology and Earth System Sciences*, 12(4), 1007–1025. <https://doi.org/10.5194/hess-12-1007-2008>
- Hashimoto, T., Stedinger, J. R., & Loucks, D. P. (1982). Reliability, resiliency, and vulnerability criteria for water resource system performance evaluation. *Water Resources Research*, 18(1), 14–20. <https://doi.org/10.1029/WR018i001p00014>
- Hejazi, M. I., Voisin, N., Liu, L., Bramer, L. M., Fortin, D. C., Hathaway, J. E., et al. (2015). 21st century United States emissions mitigation could increase water stress more than the climate change it is mitigating. *Proceedings of the National Academy of Sciences*, 112(34), 10,635–10,640. <https://doi.org/10.1073/pnas.1421675112>
- Koster, R. D., Suarez, M. J., Liu, P., Jambor, U., Berg, A., Kistler, M., et al. (2004). Realistic initialization of land surface states: Impacts on subseasonal forecast skill. *Journal of Hydrometeorology*, 5(6), 1049–1063. <https://doi.org/10.1175/JHM-387.1>
- Kraucunas, I., Clarke, L., Dirks, J., Hathaway, J., Hejazi, M., Hibbard, K., et al. (2015). Investigating the nexus of climate, energy, water, and land at decision-relevant scales: The Platform for Regional Integrated Modeling and Analysis (PRIMA). *Climatic Change*, 129(3–4), 573–588. <https://doi.org/10.1007/s10584-014-1064-9>
- Lall, U., & Sharma, A. (1996). A nearest neighbor bootstrap for resampling hydrologic time series. *Water Resources Research*, 32(3), 679–693. <https://doi.org/10.1029/95WR02966>
- Lehner, B., Liermann, C. R., Revenga, C., Vörösmarty, C., Fekete, B., Crouzet, P., et al. (2011). High-resolution mapping of the world's reservoirs and dams for sustainable river-flow management. *Frontiers in Ecology and the Environment*, 9(9), 494–502. <https://doi.org/10.1890/100125>
- Li, H., Wigmosta, M. S., Wu, H., Huang, M., Ke, Y., Coleman, A. M., & Leung, L. R. (2013). A physically based runoff routing model for land surface and Earth system models. *Journal of Hydrometeorology*, 14, 808–828. <https://doi.org/10.1175/JHM-D-12-015.1>
- Liang, X., Lettenmaier, D. P., Wood, E. F., & Burges, S. J. (1994). A simple hydrologically based model of land surface water and energy fluxes for general circulation models. *Journal of Geophysical Research*, 99(D7), 14415. <https://doi.org/10.1029/94JD00483>
- Livneh, B., Rosenberg, E. A., Lin, C., Nijssen, B., Mishra, V., Andreadis, K. M., et al. (2013). A long-term hydrologically based dataset of land surface fluxes and states for the conterminous United States: Update and extensions. *Journal of Climate*, 26(23), 9384–9392. <https://doi.org/10.1175/JCLI-D-12-00508.1>
- Masaki, Y., Hanasaki, N., Biemans, H., Schmied, H. M., Tang, Q., Wada, Y., et al. (2017). Intercomparison of global river discharge simulations focusing on dam operation—Multiple models analysis in two case-study river basins, Missouri–Mississippi and Green–Colorado. *Environmental Research Letters*, 12, 055002. <https://doi.org/10.1088/1748-9326/aa57a8>
- Mateo, C. M., Hanasaki, N., Komori, D., Tanaka, K., Kiguchi, M., Champathong, A., et al. (2014). Assessing the impacts of reservoir operation to floodplain inundation by combining hydrological, reservoir management, and hydrodynamic models. *Water Resources Research*, 50, 7245–7266. <https://doi.org/10.1002/2013WR014845>
- McMahon, T. A., Adeloye, A. J., & Zhou, S. L. (2006). Understanding performance measures of reservoirs. *Journal of Hydrology*, 324(1–4), 359–382. <https://doi.org/10.1016/j.jhydrol.2005.09.030>
- Miara, A., Macknick, J. E., Vörösmarty, C. J., Tidwell, V. C., Newmark, R., & Fekete, B. (2017). Climate and water resource change impacts and adaptation potential for US power supply. *Nature Climate Change*, 7(11), 793–798. <https://doi.org/10.1038/nclimate3417>
- Nazemi, A., & Wheater, H. S. (2015a). On inclusion of water resource management in Earth system models—Part 1: Problem definition and representation of water demand. *Hydrology and Earth System Sciences*, 19(1), 33–61. <https://doi.org/10.5194/hess-19-33-2015>
- Nazemi, A., & Wheater, H. S. (2015b). On inclusion of water resource management in Earth system models—Part 2: Representation of water supply and allocation and opportunities for improved modeling. *Hydrology and Earth System Sciences*, 19(1). <https://doi.org/10.5194/hess-19-63-2015>
- Pokhrel, Y. N., Koirala, S., Yeh, P. J. F., Hanasaki, N., Longuevergne, L., Kanae, S., & Oki, T. (2015). Incorporation of groundwater pumping in a global Land Surface Model with the representation of human impacts. *Water Resources Research*, 51, 78–96. <https://doi.org/10.1002/2014WR015602>

- Scanlon, B. R., Zhang, Z., Save, H., Sun, A. Y., Schmied, H. M., Van Beek, L. P., et al. (2018). Global models underestimate large decadal declining and rising water storage trends relative to GRACE satellite data. *Proceedings of the National Academy of Sciences*, 115(6), E1080–E1089. <https://doi.org/10.1073/pnas.1704665115>
- Schewe, J., Heinke, J., Gerten, D., Haddeland, I., Arnell, N. W., Clark, D. B., et al. (2014). Multimodel assessment of water scarcity under climate change. *Proceedings of the National Academy of Sciences*, 111(9), 3245–3250. <https://doi.org/10.1073/pnas.1222460110>
- Shin, S., Pokhrel, Y., & Miguez-Macho, G. (2019). High-resolution modeling of reservoir release and storage dynamics at the continental scale. *Water Resources Research*, 55, 787–810. <https://doi.org/10.1029/2018WR023025>
- Turner, S. W., & Galelli, S. (2016). Water supply sensitivity to climate change: An R package for implementing reservoir storage analysis in global and regional impact studies. *Environmental Modelling & Software*, 76, 13–19. <https://doi.org/10.1016/j.envsoft.2015.11.007>
- Turner, S. W. D., Bennett, J. C., Robertson, D. E., & Galelli, S. (2017). Complex relationship between seasonal streamflow forecast skill and value in reservoir operations. *Hydrology and Earth System Sciences*, 21(9), 4841–4859. <https://doi.org/10.5194/hess-21-4841-2017>
- Turner, S. W. D., Xu, W., & Voisin, N. (2020). Inferred inflow forecast horizons guiding reservoir release decisions across the United States. *Hydrology and Earth System Sciences*, 24(3), 1275–1291. <https://doi.org/10.5194/hess-24-1275-2020>
- U.S. Army Corps of Engineers (2016). National inventory of dams, Washington, D.C., available at: <http://nid.usace.army.mil/>
- US Bureau of Reclamation (2019a). Reclamation hydromet, available at: <https://www.usbr.gov/pn/hydromet/arcread.html>
- US Bureau of Reclamation (2019b). USBR Water API v1, available at: water.usbr.gov/api/web/app.php/api/
- van Beemelen, C. W. T., Mann, M., de Ridder, M. P., Rutten, M. M., & van de Giesen, N. C. (2016). Determining water reservoir characteristics with global elevation data. *Geophysical Research Letters*, 43, 11,278–11,286. <https://doi.org/10.1002/2016GL069816>
- Van Werkhoven, K., Wagener, T., Reed, P., & Tang, Y. (2009). Sensitivity-guided reduction of parametric dimensionality for multi-objective calibration of watershed models. *Advances in Water Resources*, 32(8), 1154–1169. <https://doi.org/10.1016/j.advwatres.2009.03.002>
- Voisin, N., Hejazi, M. I., Leung, L. R., Liu, L., Huang, M. Y., Li, H. Y., & Tesfa, T. (2017). Effects of spatially distributed sectoral water management on the redistribution of water resources in an integrated water model. *Water Resources Research*, 53, 4253–4270. <https://doi.org/10.1002/2016WR019767>
- Voisin, N., Kintner-Meyer, M., Wu, D., Skaggs, R., Fu, T., Zhou, T., et al. (2018). Opportunities for joint water-energy management: Sensitivity of the 2010 Western U.S. electricity grid operations to climate oscillations. *Bulletin of the American Meteorological Society*, 99(2), 299–312. <https://doi.org/10.1175/BAMS-D-16-0253.1>
- Voisin, N., Li, H., Ward, D., Huang, M., Wigmosta, M., & Leung, L. R. (2013). On an improved sub-regional water resources management representation for integration into earth system models. *Hydrology and Earth System Sciences*, 17, 3605–3622. <https://doi.org/10.5194/hess-17-3605-2013>
- Wada, Y., Bierkens, M. F., De Roo, A., Dirmeyer, P. A., Famiglietti, J. S., Hanasaki, N., et al. (2017). Human-water interface in hydrological modelling: Current status and future directions. *Hydrology and Earth System Sciences*, 21(8), 4169–4193. <https://doi.org/10.5194/hess-21-4169-2017>
- Wada, Y., de Graaf, I. E. M., & van Beek, L. P. H. (2016). High-resolution modeling of human and climate impacts on global water resources. *Journal of Advances in Modeling Earth Systems*, 8, 735–763. <https://doi.org/10.1002/2015MS000618>
- Wada, Y., Van Beek, L. P. H., Viviroli, D., Dürr, H. H., Weingartner, R., & Bierkens, M. F. (2011). Global monthly water stress: 2. Water demand and severity of water stress. *Water Resources Research*, 47, W07518. <https://doi.org/10.1029/2010wr009792>
- Wada, Y., Wisser, D., & Bierkens, M. F. (2014). Global modeling of withdrawal, allocation and consumptive use of surface water and groundwater resources. *Earth System Dynamics Discussions*, 5(1), 15–40. <https://doi.org/10.5194/esd-5-15-2014>
- Wisser, D., Fekete, B. M., Vörösmarty, C. J., & Schumann, A. H. (2010). Reconstructing 20th century global hydrography: A contribution to the Global Terrestrial Network-Hydrology (GTN-H). *Hydrology and Earth System Sciences*, 14, 1–24. <https://doi.org/10.5194/hess-14-1-2010>
- Yassin, F., Razavi, S., Elshamy, M., Davison, B., Saprizia-Azuri, G., & Wheater, H. (2019). Representation and improved parameterization of reservoir operation in hydrological and land-surface models. *Hydrology and Earth System Sciences*, 23(9), 3735–3764. <https://doi.org/10.5194/hess-23-3735-2019>
- Zajac, Z., Revilla-Romero, B., Salamon, P., Burek, P., Hirpa, F. A., & Beck, H. (2017). The impact of lake and reservoir parameterization on global streamflow simulation. *Journal of Hydrology*, 548, 552–568. <https://doi.org/10.1016/j.jhydrol.2017.03.022>
- Zhao, G., & Gao, H. (2019). Towards global hydrological drought monitoring using remotely sensed reservoir surface area. *Geophysical Research Letters*, 46, 13,027–13,035. <https://doi.org/10.1029/2019GL085345>
- Zhao, Q., Cai, X., & Li, Y. (2019). Determining inflow forecast horizon for reservoir operation. *Water Resources Research*, 55, 4066–4081. <https://doi.org/10.1029/2019WR025226>