Journal of Hydrology 603 (2021) 126843

Contents lists available at ScienceDirect

Journal of Hydrology

journal homepage: www.elsevier.com/locate/jhydrol

Research papers

# Water storage and release policies for all large reservoirs of conterminous United States

Sean W.D. Turner [a,*], Jennie Clarice Steyaert [b], Laura Condon [b], Nathalie Voisin [a,c]

[a] Pacific Northwest National Laboratory, Richland, WA, USA
[b] University of Arizona, Tucson, AZ, USA
[c] University of Washington, Seattle, WA, USA

ARTICLE INFO

This manuscript was handled by Marco Borga, Editor-in-Chief, with the assistance of Andrea Castelletti, Associate Editor

*Keywords:*
Large scale hydrology
Reservoir operations
Data-driven
Water management
dams

ABSTRACT

Large-scale hydrological and water resource models (LHMs) require water storage and release schemes to represent flow regulation by reservoirs. Owing to a lack of observed reservoir operations, state-of-the-art LHMs deploy a generic reservoir scheme that may fail to represent local operating behaviors. Here we introduce a new dataset of bespoke water storage and release policies for 1,930 reservoirs of conterminous United States. The Inferred Storage Targets and Release Functions (ISTARF-CONUS) dataset relies on a new inventory of observed daily reservoir operations (ResOpsUS) to generate reservoir operating rules for 595 data-rich reservoirs. These functions are developed in a standardized form that allows for extrapolation of operating schemes to 1,335 data-scarce reservoirs—leading to the first inventory of empirically derived reservoir operating policies for all large CONUS reservoirs documented in the Global Reservoir and Dams (GRanD) database. Evaluation of the new scheme in daily simulations forced with observed inflow demonstrates substantial and robust improvement for both release and storage relative to the popular Hanasaki method. Performance of the extrapolation approach for data-scarce reservoirs is evaluated with leave-one-out validation and is shown to also offer modest gains on average over Hanasaki. ISTARF-CONUS may be readily adopted in any LHM featuring large reservoirs of the conterminous United States.

## 1. Introduction

Reservoir operations must be represented in large scale hydrological and water resource models (LHMs) for reasonable estimation of the spatial and temporal distribution of water resources (Nazemi and Wheater, 2015; Wada et al., 2017). Lack of widespread data on water release decision-making has necessitated generic models of reservoir operations. Such generic schemes are broadly representative of reservoirs that fulfil a specific purpose rather than being tailored to each reservoir's actual operations (Masaki et al., 2017). These generic schemes include the seminal methods of Haddeland et al. (2006) and Hanasaki et al. (2006) (herein Han-06), which have been adapted for a variety of LHM settings (e.g., Biemans et al., 2011; Voisin et al., 2013; Wada et al., 2014). Equipped with these schemes, LHMs have been applied successfully to a range of water research topics at global and continental scales, including analyses of water scarcity (Veldkamp et al., 2017), terrestrial water storage (Pokhrel et al., 2021), power grid reliability (Voisin et al., 2020), and the global value of services provided by dams (Boulange et al., 2021).

Although generic schemes capture the general river regulating effects of reservoirs, they remain insufficient for emerging applications that require accurate storage and release simulation at sub-monthly resolution—including analysis of flood and drought impacts (Masaki et al., 2018; Veldkamp et al., 2018). Empirically derived reservoir operations informed by local data can address these shortcomings but have thus far been confined to isolated river basins for which a complete set of observed reservoir operations records can be obtained (Yassin et al., 2019; Turner et al., 2020a; Coerver et al., 2018; Macian-Sorribes and Pulido-Velazquez, 2017; Fleischmann et al., 2021). Universal coverage of bespoke reservoir simulation at a large, multi-basin spatial scale is an aspiration that has remained unfulfilled due to the lack of a national scale reservoir operations inventory.

Here we leverage a new dataset of observed daily reservoir operations in conterminous United States (CONUS) to develop the first national-scale inventory of empirically derived reservoir operations for application in LHMs. We infer operating policies using a novel

* Corresponding author.
*E-mail address:* sean.turner@pnnl.gov (S.W.D. Turner).

https://doi.org/10.1016/j.jhydrol.2021.126843
Received 13 April 2021; Received in revised form 2 July 2021; Accepted 15 August 2021
Available online 21 August 2021

algorithm—STARFIT (Storage Targets And Release Function Inference Tool)—that provides a standardized and parsimonious formulation of each reservoir's inferred operating policy. We apply STARFIT to 595 large, data-rich reservoirs located in CONUS and in Canadian river basins that drain into CONUS (herein CONUS+). The standardized formulation allows for extrapolation to data-scarce reservoirs and thus the creation of a complete dataset of operations for all 1,930 large reservoirs of CONUS+ included in the Global Reservoir and Dams (GRanD) database (Lehner et al., 2011). We evaluate the performances of both data-rich reservoirs and the extrapolation scheme relative to Han-06 for release and storage simulation accuracy with observed inflow forcing. The new weekly reservoir policy parameters dataset is named ISTARF-CONUS (Inferred Storage Targets and Release Functions for CONUS+) and is accompanied by an open-source software package that can be used to generate reservoir policies directly from daily reservoir operations time series (github.com/IMMM-SFA/starfit). These policies may be applied down to a daily temporal resolution in LHMs.

## 2. Method

### 2.1. Policy inference rationale and overview

The STARFIT algorithm generates a 19-parameter reservoir operating policy comprising weekly water storage targets and release functions (Table 1). The algorithm is designed to provide realistic simulation of water release decisions throughout year and under varying water availability conditions. STARFIT policies are relatively parsimonious compared to the data-driven methods of Yassin et al. (2019) with 72 parameters (six parameters defined independently for each month of the year) or Turner et al. (2020b) with 208 parameters (four parameters defined independently for 52 weeks of the year). Unlike models that are fitted separately for individual seasons, months or weeks of the year, STARFIT employs harmonic regression to give smooth continuity across periods with relatively few degrees of freedom. The parameters of these harmonics may be easily interpreted for seasonal timing and relative importance of key operating features, such the use of a seasonal flood or conservation pool. A robust and parsimonious model is particularly important for reservoir operations inference given data challenges, which preclude traditional split sample or leave-one-out validation. These techniques are inhibited by the relatively short records available (perhaps featuring only one or two extreme events) and by possible undocumented changes in actual operations occurring over period of record. In other words, a model may perform poorly in a validation period not because it is overfitted, but because actual objectives and constraints on operations can shift at unknown points in the historical record (discussion on these challenges is offered in Turner et al., 2020a).

**Table 1**
Summary of data for each reservoir included in the Inferred Storage Targets and Release Functions for Conterminous United States (ISTARF-CONUS) dataset.

| Data type | Information / parameters provided |
|---|---|
| Reservoir detail | Reservoir name<br>GRanD ID<br>Capacity (MCM) |
| Inference detail | Fit type (fully fitted / storage-only / extrapolated)<br>ID similarly purposed neighboring reservoir (if not fully fitted) |
| Inflow | Long term mean inflow ($m^3/s$) – GRanD<br>Long term mean inflow ($m^3/s$) – Observed (if available). |
| *Policy parameters* | |
| Upper bound of NOR | $\mu^{\uparrow}$, $\alpha^{\uparrow}$, $\beta^{\uparrow}$, $\widehat{S}^{\uparrow}_{min}$, $\widehat{S}^{\uparrow}_{max}$ |
| Lower bound of NOR | $\mu^{\downarrow}$, $\alpha^{\downarrow}$, $\beta^{\downarrow}$, $\widehat{S}^{\downarrow}_{min}$, $\widehat{S}^{\downarrow}_{max}$ |
| Seasonal release pattern | $\alpha_1$, $\alpha_2$, $\beta_1$, $\beta_2$ |
| Release adjustment | $c$, $p_1$, $p_2$ |
| Release constraints | $R_{min}$, $R_{max}$ |

STARFIT is developed to be flexible to deal with varying length, quality, and completeness of observational data. The policy inference algorithm requires neither gap-free time series nor a complete set of input variables. STARFIT is also designed to generate policies that are transferrable to data-scarce reservoirs. Weekly-varying storage targets are specified as percent of storage capacity; release functions return release decisions as percent deviation from long term mean inflow. Since both reservoir storage capacity and mean annual inflow are available for all large CONUS + reservoirs via GRanD, STARFIT's standardized policy formulation allows one to copy a reservoir's fitted policy and impose it on a similarly purposed, neighboring reservoir by updating these two specifications. This use of mean annual inflow also gives the user the ability to adjust the operations to account for inflow bias common to LHMs—a key feature of Han-06 that is retained in our data-driven approach (further detail in Section 4). Reservoirs with storage data only (i.e., no inflow or release records) can be fitted with storage targets and then have the release functions added by copying from similarly purposed, nearby reservoirs. This leads to three types of reservoir policy in ISTARF-CONUS: fully fitted (storage and release functions generated from data, 449 reservoirs); partially fitted (storage fitted from data, release functions extrapolated from similarly purposed, neighboring reservoirs, 146 reservoirs); and extrapolated (1,335 reservoirs).

The key parameters of a reservoir policy in ISTARF-CONUS are conveyed in Fig. 1 and elaborated in the following subsections. Following Turner et al. (2020b), observed reservoir operations data—namely daily records of storage, inflow, and release—are first converted to a weekly resolution, allowing for reasonable back-calculation of inflow or release if either is missing (using change in storage from start to end of each week). This expands the number of reservoirs available for policy inference. For each reservoir, the algorithm first fits two 5-parameter harmonic functions that define the upper and lower bounds of a seasonally varying Normal Operating Range (NOR) of storage (Section 2.1.1). This range provides three operating settings: normal operations, below-range operations, and above-range operations. If storage levels are within the NOR (normal operations), water release is defined using a seasonal release pattern (4 parameter harmonic function) and with a single linear adjustment to account for water availability (3 parameters) (Section 2.1.2). If storage levels drop below the NOR, the reservoir is assumed to be at risk of being depleted and so release is curtailed to a minimum outflow rate determined from observations. If storage levels rise above the NOR, the operator seeks to draw the reservoir back down within a pre-defined time period by increasing the release, subject to a below spill-crest maximum (also defined using release observations). Minimum and maximum release constitute the final two parameters of the 19-parameter scheme. All parameters associated with this model are derived solely from observations of daily storage, inflow, and release; the actual operating purposes of each reservoir emerge from this process rather than being an input to the algorithm.

#### 2.1.1. Normal operating range (NOR)

The NOR for a given reservoir is a set of weekly-varying maximum and minimum storage values. These values are given as % of reservoir capacity, and they dictate when normal release operations are applied. The NOR is defined by first aggregating storage to weekly average values and then standardizing as a function of maximum storage capacity. The general model adopted to define the NOR is:

$$\widehat{S}^{\uparrow}_t = \begin{cases} \mu^{\uparrow} + \alpha^{\uparrow}\sin 2\pi\omega t + \beta^{\uparrow}\cos 2\pi\omega t, & \widehat{S}^{\uparrow}_{min} \leq \mu^{\uparrow} + \alpha^{\uparrow}\sin 2\pi\omega t + \beta^{\uparrow}\cos 2\pi\omega t \leq \widehat{S}^{\uparrow}_{max} \\ \widehat{S}^{\uparrow}_{min}, & \mu^{\uparrow} + \alpha^{\uparrow}\sin 2\pi\omega t + \beta^{\uparrow}\cos 2\pi\omega t \leq \widehat{S}^{\uparrow}_{min} \\ \widehat{S}^{\uparrow}_{max}, & \mu^{\uparrow} + \alpha^{\uparrow}\sin 2\pi\omega t + \beta^{\uparrow}\cos 2\pi\omega t \geq \widehat{S}^{\uparrow}_{max} \end{cases} \tag{1}$$

for the upper bound, and

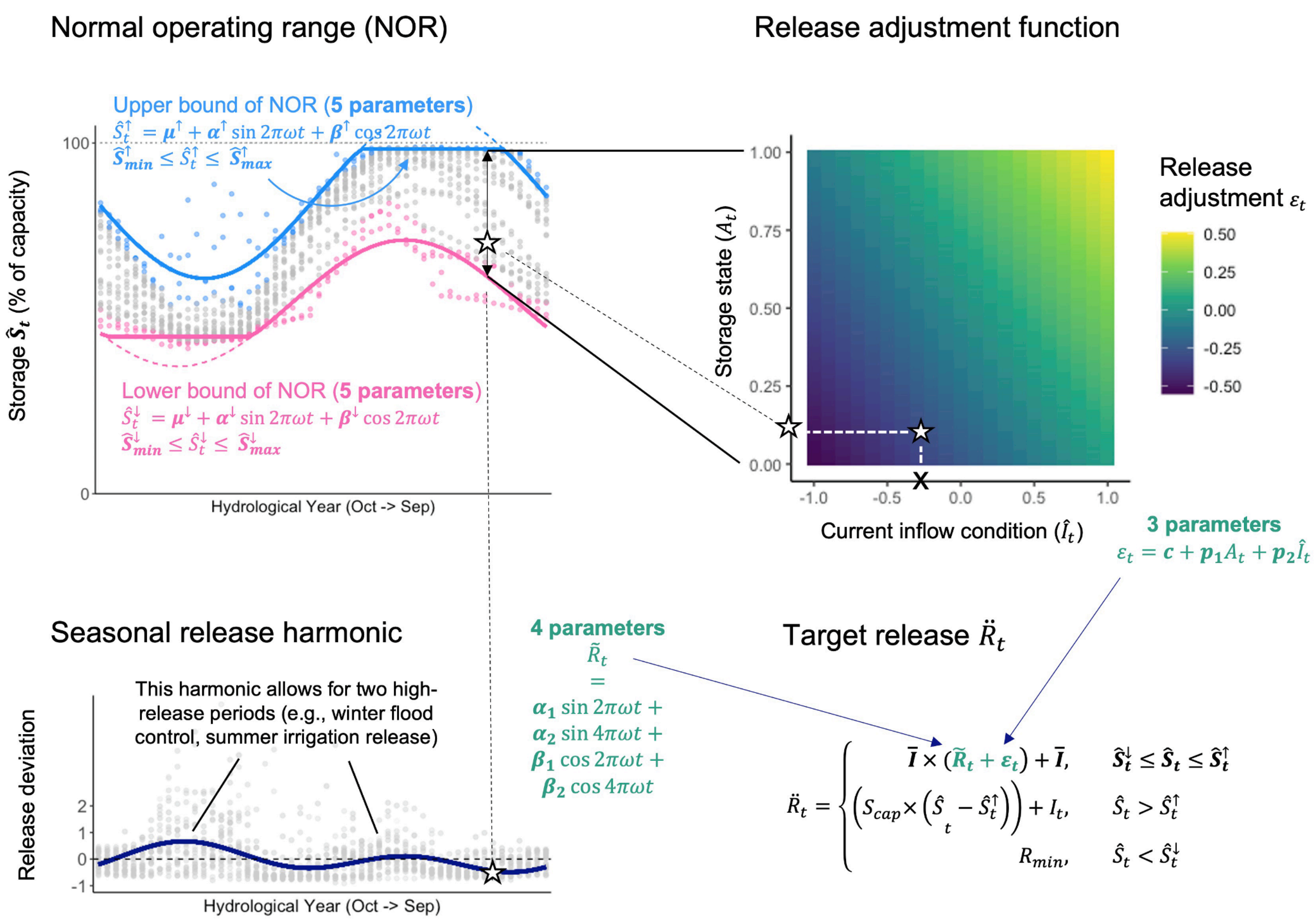


**Fig. 1.** Overall concept of a 19-parameter policy in ISTARF-CONUS. Starting from the upper left panel, upper and lower extremes of weekly storage averages are used to determine the bounds of the reservoir's Normal Operating Range (10 parameters fitted to observed storage). In simulation, if storage lies within this range, the seasonal release harmonic (lower-left panel; defined with 4 parameters, fitted to observed weekly release volumes) and release adjustment function (upper-right panel; 3 parameters, fitted to the residuals from the seasonal harmonic) are used to determine a target release. The remaining two parameters of the 19-parameter policy are the maximum and minimum release values, which act as constraints on the target release. The star symbol on each graphic illustrates how the release decision would be derived for a late summer week in which storage is low within the normal operating range.

$$\widehat{S}_t^{\downarrow} = \begin{cases} \mu^{\downarrow} + \alpha^{\downarrow}\sin 2\pi\omega t + \beta^{\downarrow}\cos 2\pi\omega t, & \widehat{S}_{min}^{\downarrow} \le \mu^{\downarrow} + \alpha^{\downarrow}\sin 2\pi\omega t + \beta^{\downarrow}\cos 2\pi\omega t \le \widehat{S}_{max}^{\downarrow} \\ \widehat{S}_{min}^{\downarrow}, & \mu^{\downarrow} + \alpha^{\downarrow}\sin 2\pi\omega t + \beta^{\downarrow}\cos 2\pi\omega t \le \widehat{S}_{min}^{\downarrow} \\ \widehat{S}_{max}^{\downarrow}, & \mu^{\downarrow} + \alpha^{\downarrow}\sin 2\pi\omega t + \beta^{\downarrow}\cos 2\pi\omega t \ge \widehat{S}_{max}^{\downarrow} \end{cases} \quad (2)$$

for the lower bound. $\widehat{S}_t^{\uparrow}$ and $\widehat{S}_t^{\downarrow}$ are fitted upper and lower bounds of the NOR in week $t$, $\omega$ is the frequency (set as 1/52 to reflect the weekly resolution of the model), and $\mu$, $\alpha$, and $\beta$ are fitted parameters controlling the baseline, amplitude, and phase of two harmonics that define the upper and lower bounds of the NOR. These upper and lower bounds of the NOR are fitted to the three highest storage values (blue points, Fig. 1 upper left panel) and three lowest storage values (pink points, same figure) for each week across all years, respectively. The use of three years protects against a NOR that envelops one-off extreme years while also providing a sufficiently wide NOR for reservoirs that have short records of perhaps just ten years' observed data. Resulting harmonics characterize observed storage behavior with a cyclic component to represent seasonal flood and conservation pools. Each harmonic is fitted with constraining parameters $\widehat{S}_{min}^{\uparrow}$ and $\widehat{S}_{max}^{\uparrow}$ ($\widehat{S}_{min}^{\downarrow}$ and $\widehat{S}_{max}^{\downarrow}$ for the lower harmonic), which provide flexibility to capture deep, sharp peaks and troughs in the NOR with possible multi-week periods during which the bounds remain fixed (see the lower bound of the NOR at the start of the hydrological year in Fig. 1, for example). These bounding constraints mean the function cannot be fitted using standard multiple linear regression. We instead use a local, gradient-free optimization algorithm (Powell, 1994) to identify the five parameters of each harmonic. The optimization searches for the constrained harmonic that minimizes root mean squared error between the harmonic and the selected storage observations (i.e., the three highest or lowest depending on whether the upper or lower harmonic is sought). To guide the algorithm to a successful solution, the harmonic parameters ($\mu^{\uparrow}, \alpha^{\uparrow}, \beta^{\uparrow}$ ; $\mu^{\downarrow}, \alpha^{\downarrow}, \beta^{\downarrow}$) are first initialized using multiple linear regression (assuming no constraints); each constraint is then initialized to lie just within the amplitude of that harmonic, ensuring that the objective function is sensitive to its adjustment. The optimization algorithm is then run to find all parameters of the harmonics. Examples of fitted operating ranges defined by the upper and lower bound harmonics are given in Fig. 2.

### 2.1.2. *Release function*

The NOR is used to identify the reservoir's release operation mode on each time step of simulation. The state of the reservoir may be normal (within NOR), above-normal, or below-normal. Operations are designed to guide storage levels back into the NOR should they deviate outside it. This can happen if there is a flood, causing storage to breach the upper bound of the normal operating range, or a prolonged dry spell, causing reservoir drawdown below the NOR. If storage exceeds the NOR, the STARFIT algorithm sets a desired release to draw storage back into range within one week. If this desired release exceeds a pre-determined maximum release, it will be constrained to the maximum

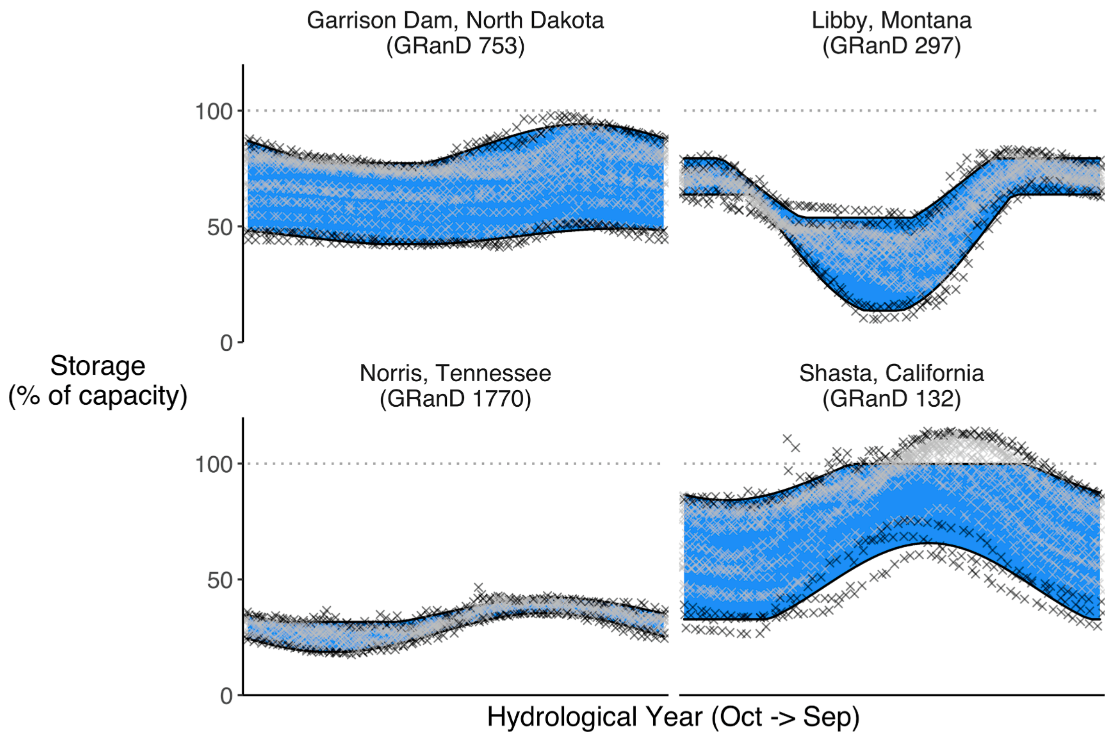


**Fig. 2.** Normal Operating Range fitted to four distinct reservoirs located in different CONUS states and river basins. Upper and lower constrained harmonic functions are given in solid black lines as fitted to three upper and lower extremes of weekly average storage values. Capacity is based on reported capacity in GRanD database, which may in some cases represent spill crest rather than full operational capacity (hence above-capacity periods shown for Shasta).

release—meaning flood conditions can still cause the reservoir to fill, to capacity, causing uncontrolled spill. If storage drops below NOR, the operator has limited ability to refill the reservoir (negative release is impossible). In this instance STARFIT applies a predefined minimum release value (also determined from observed records), which for many reservoirs represents an environmental release requirement. In simulation, all releases are also constrained by basic mass balance. For example, release cannot exceed water available in storage plus inflow.

For most time periods of a reservoir simulation, storage will lie within the normal operating range. During such periods, we assume the operator schedules the release based on three sources of information: (1) week of year, defined using the epidemiological week (EPI week, beginning Sunday), herein referred to as the operating week; (2) reservoir inflow volume during the operating week, and (3) existing state of storage at the start of the operating week, which we define here as the position of current storage within the normal operating range (i.e., storage at lower bound of NOR means storage state = 0%; storage at upper bound of NOR means storage state = 100%). Other sources of information, such as known releases from upstream reservoirs operated in coordination, or accurate inflow forecasts, are often important in reality (Rouge et al., 2021), but are not considered in the present approach, which focuses on a method that may be readily adopted within existing LHMs (discussion on challenges of forecast implementation in LHMs is offered in Turner et al., 2020a).

One can use data to determine a simple relation that defines release as a function of storage and inflow. This release relationship should vary by time of year, since releases will be adjusted depending on various seasonally varying objectives (e.g., irrigation season, snowmelt season, etc.). To avoid reconstructing this release relationship for multiple periods (leading to a heavily parameterized model), we adopt the following two-step approach. First, average weekly release is modeled using a four-parameter harmonic function (see example in lower left panel of Fig. 1). Second, the residuals from this average release pattern are fitted with single linear relationship that provides release adjustment from seasonal average as a function of current inflow and storage state.

To fit these release function parameters, weekly release volume time series are first standardized around 0 by presenting as fractional deviation from long term mean inflow:

$$\widehat{R}_t = \frac{R_t - \overline{I}}{\overline{I}} \tag{3}$$

Where $R_t$ is the time series of weekly release volumes and $\overline{I}$ is the long-term mean of inflow (same units). Since $\overline{I} \approx \overline{R}$ (i.e., long-term mean inflow approximates long term mean release), the harmonic used to model a typical seasonal release pattern may be fitted without an intercept parameter (i.e., centered on zero):

$$\widetilde{R}_t = \alpha_1 \sin 2\pi\omega t + \alpha_2 \sin 4\pi\omega t + \beta_1 \cos 2\pi\omega t + \beta_2 \cos 4\pi\omega t \tag{4}$$

Where $\alpha_1$, $\alpha_2$, $\beta_1$, and $\beta_2$ are the four parameters. Unlike for the constrained harmonics that define the NOR, these harmonics may be identified with multiple linear regression. The residuals of this harmonic ($\varepsilon_t = \widehat{R}_t - \widetilde{R}_t$) are then used to fit the following linear model, which allows for an adjustment to the seasonal release depending on available water in storage and inflow:

$$\varepsilon_t = c + p_1 A_t + p_2 \widehat{I}_t \tag{5}$$

$$\widehat{I}_t = \frac{I_t - \overline{I}}{\overline{I}} \tag{6}$$

Where $A_t$ is the availability status, equal to the fractional position of the storage within the normal operating range at the start of the operating week ($0 \le A_t \le 1$), $\widehat{I}_t$ is standardized inflow, and $c$, $p_1$, and $p_2$ are regression coefficients. By separating inflow and storage as separate inputs rather than a single water availability value, the model has flexibility to account for operations driven by either variable. The model is discarded (i.e., all three parameters set to zero) if it performs with an R-squared value less than 0.3 (suggesting poor fit) or if the $p_1$ or $p_2$ coefficients are negative (which is illogical, since it would suggest lower release with greater water available). If the linear model is discarded the release will be a function time of year only, following the harmonic function without further adjustment for water availability.

With the parameters of the release harmonic and adjustment function set, the implementation of the release decision making is:

$$\ddot{R}_t = \begin{cases} min\left(\bar{I} \times \left(\widetilde{R}_t + \varepsilon_t\right) + \bar{I},\ R_{max}\right), \widehat{S}_t^{\downarrow} \leq \widehat{S}_t \\ \end{cases}$$

$$\leq \widehat{S}_t^{\uparrow} min\left(\left(S_{cap} \times \left(\widehat{S}_t - \widehat{S}_t^{\uparrow}\right)\right) + I_t,\ R_{max}\right), \widehat{S}_t > \widehat{S}_t^{\uparrow} R_{min}, \widehat{S}_t < \widehat{S}_t^{\downarrow} \quad (7)$$

Where $\ddot{R}_t$ is the targeted release and $\widehat{S}_t$ is the storage volume ($S_t$) as a percentage of storage capacity $S_{cap}$. Parameters $R_{max}$ and $R_{min}$ are the final two parameters of the 19-parameter model. These parameters are determined using the 95th and 5th percentile values of the release observations (the software that supports this study allows this setting to be adjusted easily). Actual release implemented in simulation is then equal to the targeted release $\ddot{R}_t$ subject to mass balance constraints. If the reservoir exceeds capacity, all excess water must be spilled. If the targeted release cannot be met with existing water available, it is constrained to physical water available in storage plus inflow.

$$R_t = max\left(min\left(\ddot{R}_t, I_t + S_t\right), I_t + S_t - S_{cap}\right) \quad (8)$$

### 2.2. Application of STARFIT to ResOpsUS

The ResOpsUS inventory of observed reservoir operations includes daily storage, release, and inflow records with varying levels of length and completeness for 678 large CONUS + reservoirs. Some records include lengthy, gap-free time series of daily storage, release, and inflow, while others feature significant gaps, or may lack observations for a key variable. Although the STARFIT algorithm can glean a policy from incomplete records or cases with a missing inflow or release variable, not all records in 678 are suitable for policy inference. After removing pre-1995 data to avoid fitting policies to out-of-date operations, we discard records that have less than at least ten full years of daily storage values (for NOR identification) and five full years of either inflow or release (for release function identification). Some reservoirs benefit from storage data sufficient to fit NOR but lack release or inflow data sufficient for release function identification. These cases may be fitted with a NOR and can then adopt a neighboring, similarly purposed reservoir's standardized release function as described in 2.3.

ResOpsUS provides adequate data for 449 reservoirs with storage targets and release functions, and 146 reservoirs with storage targets only. A total of 595 reservoir policies in ISTARF-CONUS may be considered fitted with data, while the remaining 1,335 reservoir policies are extrapolated. Of the 449 fully fitted reservoirs, 111 (~25%) have release adjustment parameters ($c, p_1, p_2$) discarded due to weak performance in model fitting (see 2.1.2). Discarding of the release adjustment parameters indicates that releases during normal operating conditions are driven more by seasonal timing (modeled using the harmonic) than by prevailing water availability. This does not necessarily indicate problematic data or weaker reservoir policy in general. Reservoirs with a discarded release adjustment function remain in the sample as candidates for extrapolation to data-scare reservoirs.

### 2.3. Extrapolation of rules to data-scarce reservoirs

For 1,335 data-scarce reservoirs for which we lack observed operations sufficient to employ the STARFIT algorithm, we extrapolate inferred policies from the 595 data-rich reservoirs. The premise for the extrapolation procedure is that reservoirs with similar operating purposes and which face similar seasonal hydrological conditions will be operated in a similar manner. For instance, many reservoirs in the Pacific Northwest draw down storage levels to capture flood waters coming from the spring freshet. While no two reservoirs will be operated in exactly the same way, this feature results in a reliable winter downshift in the normal operating range for flood control reservoirs of the Columbia river basin (illustrated by the example of Libby Dam in Fig. 2). Since the storage targets are defined in terms of % capacity and release functions as % deviation from long-term inflow rate, such rules can be simply copied from one reservoir to be imposed on another, with storage capacity and long-term inflow updated for the new reservoir. The validity of this procedure may be examined using the data-rich reservoirs. One can test the performance of the extrapolation scheme by assuming that each data-rich reservoir is data-scarce, copying its policy from a neighboring, similarly purposed reservoir, and simulating its storage and release behavior with observed inflow forcing (see section 2.5.3).

Extrapolation to data-scarce reservoirs is conducted using the United States Geological Survey's standardized river basins. These basins are known as hydrologic units and are defined using Hydrologic Unit Codes (HUCs). Conterminous United States is divided into 18 major hydrological regions known as HUC2s (e.g., California, Pacific Northwest, Great Lakes, etc), which are further subdivided into approximately 200 river basins, known as HUC4s. For a given data-scarce reservoir, data-rich reservoirs within the same HUC4 are first identified. These reservoirs are then ranked according to the degree to which their operating purposes align, based on four objectives recorded as a Boolean variable in GRanD (namely, flood, hydropower, irrigation, and water supply). To compute the operating alignment between reservoirs, we sum the number of matches. Four matches would represent a perfectly aligned reservoir and 0 matches would mean that the reservoirs share no operating objectives. If the HUC4 features no reservoirs with operating alignment>1 match, the larger HUC2 basin is searched for a better match. If there are multiple reservoirs with an equal number of matches, the closest dam distance-wise is selected. One exception is in the Great Lakes HUC2 basin where ResOpsUS has no suitable records for policy inference. Reservoirs in this region copy from best aligned reservoirs from the closest neighboring HUC2 region (depending on reservoir) on the periphery of the Great Lakes region. Data-scarce reservoirs located regions where data are limited in general, such as the Great Lakes, New England, Mid Atlantic, and South Atlantic Gulf, tend to rely heavily on small number of data-rich reservoirs within the same HUC2. In contrast, data-scarce reservoirs located in data-rich regions may copy rules from a much larger sample of nearby, similarly purposed reservoirs (Fig. 3).

### 2.4. Performance testing

#### 2.4.1. Daily simulation of reservoir policies with observed inflow

The performance of the STARFIT scheme is tested at a daily time step, which is a common temporal resolution adopted in LHMs. We use daily observed inflow to force each reservoir model, allowing for evaluation of simulated release and storage relative to observations. Weekly-resolution policies can be simulated and tested at daily resolution following an approach similar to Turner et al. (2020a). On each day of simulation, the normal operating range is determined based on the current operating week. Storage level then determines the mode of operation for that day (normal, above-normal, below-normal) as described above. During normal operations, the fitted release function provides weekly release as a function of current-period weekly inflow and water in storage. In daily simulation the full week's inflow is unknown but can be forecasted. We simulate with a week-ahead persistence forecast, meaning the current day's total mean inflow rate is assumed constant over the following seven days to give a weekly flow volume. The weekly release volume may then be determined, and a day's worth of that release volume (weekly release divided by seven) is

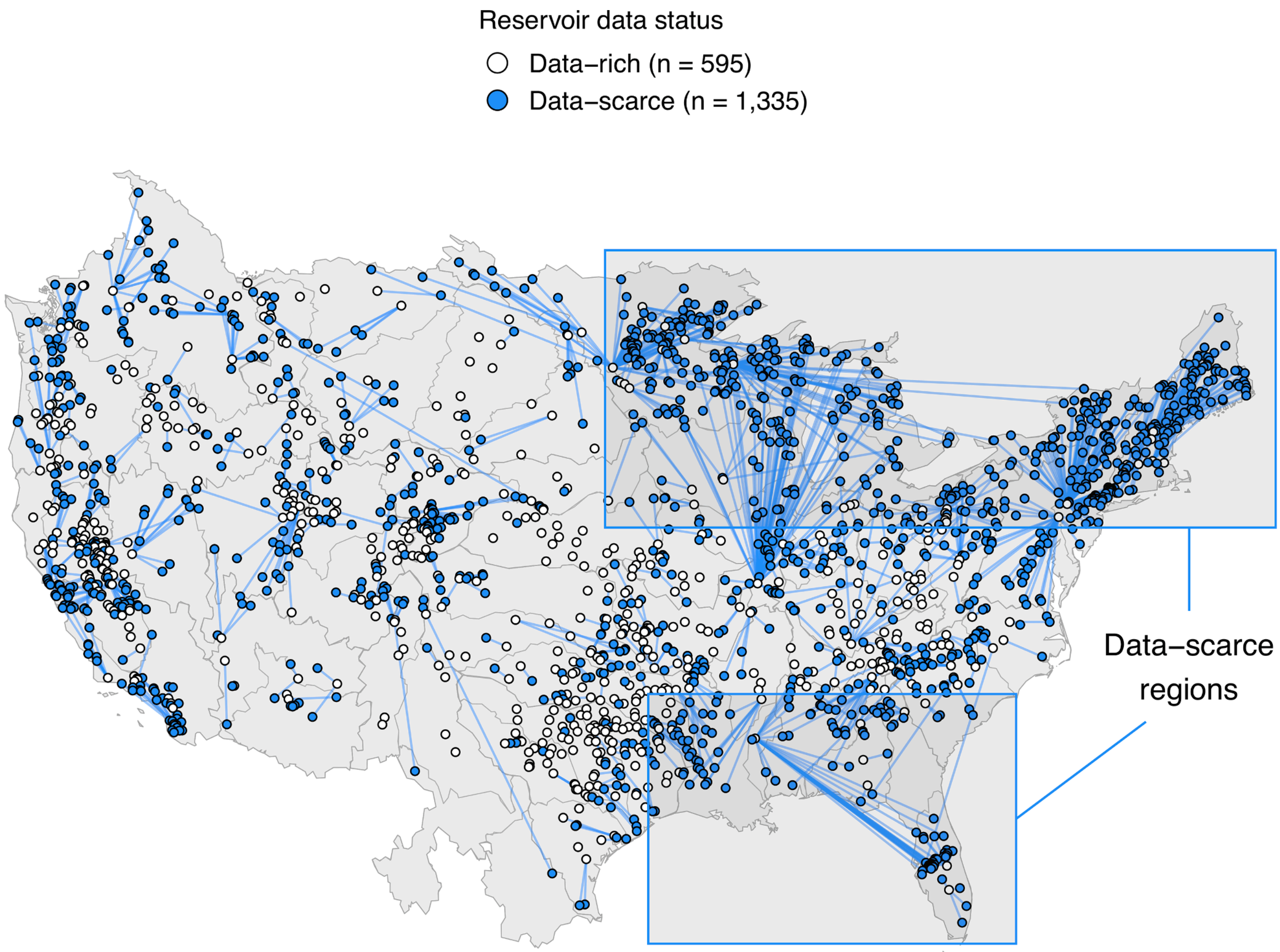


**Fig. 3.** Extrapolation from data-rich reservoirs (n = 595) to data-scarce reservoirs (n = 1,335) based on operational alignment, hydrological unit code, and distance.

implemented in simulation. As the simulation steps forward to the next day a new weekly flow forecast and associated release are generated.

Two daily simulations are evaluated. First, daily release is simulated with storage fixed to observed values. This simulation isolates the performance of the storage targets and release functions independent of storage simulation performance. Second, daily release and storage are simulated together. This simulation and resulting release and storage behaviors indicate the performance of the model with error in release allowed to accumulate through storage errors (thus affecting the next day's release decision). This latter simulation provides a more accurate indication of model performance when exposed to an actual LHM simulation, in which storages are simulated. Although STARFIT can fit policies to reservoirs with gaps in the observational record or with either inflow or release missing, the performance test requires reservoirs replete with lengthy (>5 years), continuous observed inflow to drive the simulation. Results are reported for 273 reservoirs with sufficient inflow data to conduct daily performance evaluation.

### 2.4.2. Benchmark

Simulation performances are benchmarked against Han-06, which is conceptually similar to a variety of other generic reservoir schemes (e.g., Biemans et al., 2011; Voisin et al., 2013) and is considered the benchmark to improve upon for reservoir operations in LHMs (Boulange et al., 2021). Han-06 sets provisional annual and monthly release targets to represent interannual and seasonal release patterns for each reservoir. Provisional annual release is modeled as a linear adjustment to mean annual inflow depending on the volume of water in storage at the beginning of the operational year. Specifically, the provisional annual release is equal to mean annual flow multiplied by the fractional deviation of storage from a predefined "alpha" constant. Monthly provisional release depends on reservoir purpose. Han-06 categorizes a reservoir as either "irrigation" or "non-irrigation." For the irrigation reservoirs, within-year variation in the release is determined by seasonality in downstream irrigation water demand allocated to that reservoir (downstream demand is allocated among multiple upstream reservoirs proportional to their mean annual inflow). For non-irrigation reservoirs, the monthly provisional release is set to the mean annual flow rate. Finally, Han-06 adjusts the provisional monthly release to obtain actual monthly release. The mode of adjustment depends on the ratio of reservoir storage volume to annual inflow volume, termed $c$. For reservoirs with $c > 0.5$, monthly release is assumed to be independent of monthly inflow, and for reservoirs with $c < 0.5$ monthly inflow is factored into the release to avoid over-frequent fill and spill cycles. Daily releases are updated for minimum and maximum storage, as well as a minimum flow release.

In this study we simulate Han-06 with "alpha" storage constant parameter set to 85% of capacity, as recommended in Hanasaki et al. (2006). We set downstream demand using the MOSART-WM large scale hydrological and water management model as configured in Voisin et al. (2017).

### 2.4.3. Metrics

We compute three metrics to compare performances of STARFIT to Han-06. These are:

1. Normalized RMSE (nRMSE – where the RMSE is normalized by dividing by standard deviation of the observations).
2. Normalized RMSE of above-standard-operating level periods (nRMSEˆ)—meaning the data are filtered for periods in which the observed reservoir storage is above the normal operating range. This is test of operations performance during high inflow periods and spill.
3. Normalized Transformed RMSE (nTRMSE) where the simulated and observed releases are first transformed using a box-cox transform with lamba = 0.3 (van Werkhoven et al., 2009). This transformation puts greater weight on low flow periods.

The nRMSE is closely related to the commonly used Nash-Sutcliffe Efficiency (NSE) while providing strictly non-negative error values (useful in our assessment for log scale plotting) that are interpretable and easily compared to the other nRMSE metrics. All three metrics are computed for each of the two release variables, namely daily release simulated with observed storage ($R_{S\text{-}OBS}$) (indicating performance of the the release decisions independent of storage error), daily release simulated with simulated storage ($R_{S\text{-}SIM}$) (indicating performance of combined storage and release simulation). We also compute these metrics on daily simulated storage (S) versus observed storage for each reservoir.

#### 2.4.4. Leave-one-out test for extrapolation approach

Reservoirs with extrapolated rules cannot be evaluated using the approach described above, since they lack observed data to drive the simulation and benchmark outputs. However, the general extrapolation approach may be evaluated with a leave-one-out method, wherein a data-rich reservoir is assumed to be data-scarce. The target reservoir copies the standardized NOR and release function from a neighboring, similarly purposed reservoir (as described in 2.3) and is then evaluated using the performance evaluation approach described above. Results derived from this extrapolation test are reported separately with the label “STARFITx.”

We also evaluate two alternative extrapolation settings, which are designed to reveal the relative importance of reservoir purpose and distance to the performance of the extrapolation. The first of these alternatives, setting “X1”, targets similarly-purposed reservoirs within the same river basin, but choses the farthest reservoir rather than the closest in cases where multiple reservoirs of equally good operational alignment are found. This setting reveals the importance of physical distance to extrapolation performance. Setting “X2” disregards the reservoir purpose entirely and instead copies the closest reservoir by distance. This setting reveals the importance of reservoir purpose in the extrapolation scheme.

## 3. Results

### 3.1. Performance of STARFIT and STARFITx

STARFIT provides robust performance improvements relative to Han-06 across all variables and metrics examined (Fig. 4). For data-rich reservoirs, STARFIT reduces error of simulated daily storage (S), release ($R_{S\text{-}SIM}$), and release assuming observed storage ($R_{S\text{-}OBS}$) for ~ 87% and

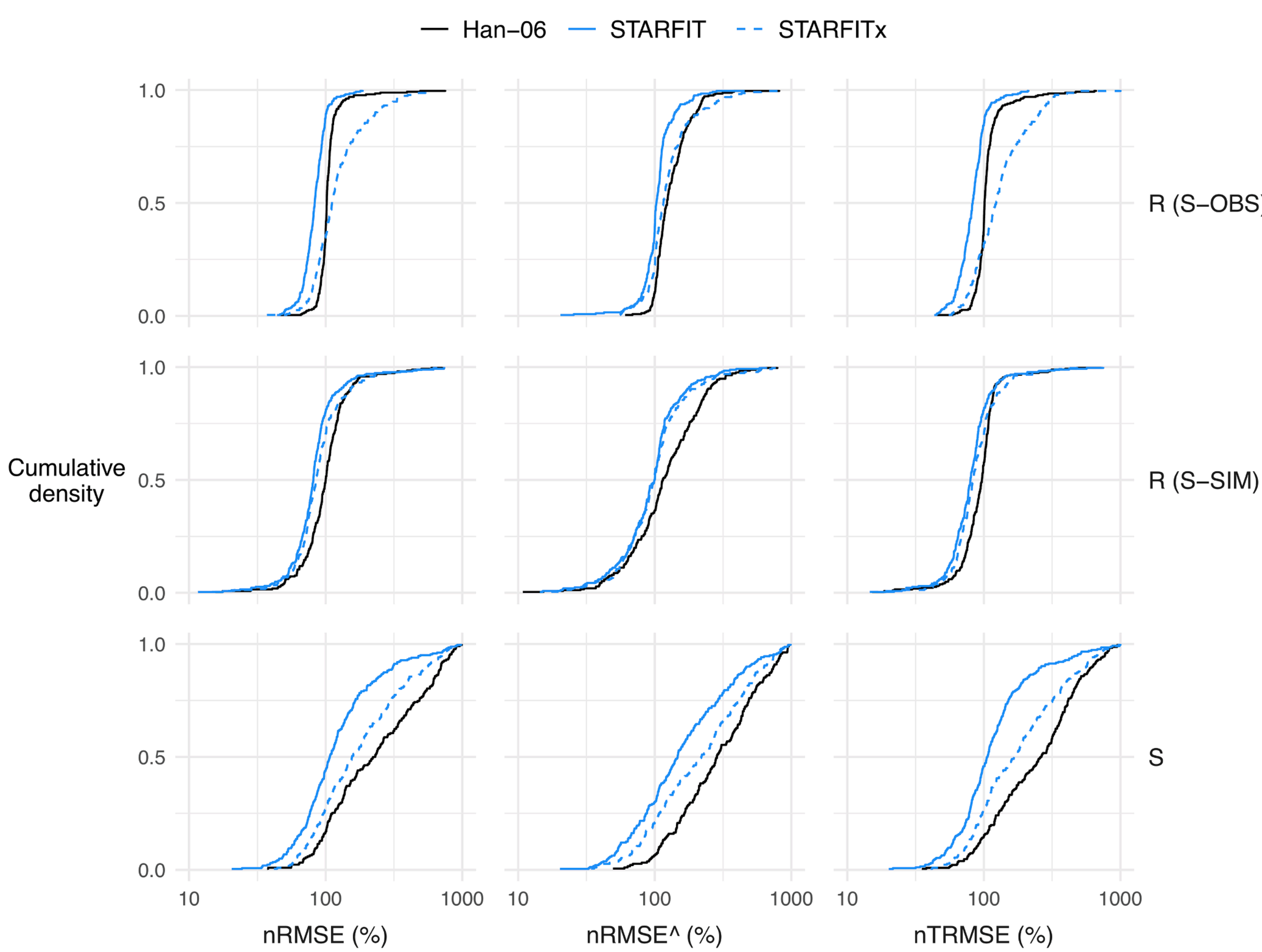


**Fig. 4.** Empirical cumulative distribution functions for each variable (Release with observed storage $R_{S\text{-}OBS}$, release with simulated storage $R_{S\text{-}SIM}$, and storage S) and metric across all reservoirs with sufficient daily storage, inflow, and release data for simulation and goodness-of-fit testing.

71–78%, and 85–91% of reservoirs, respectively (depending on performance metric assessed). Average magnitude of improvement with STARFIT is significant, with 64–67% reduction in median nRMSE for simulated storage (S) and 15–20% reduction in median nRMSE for release (both $R_{S\text{-}OBS}$ and $R_{S\text{-}SIM}$) relative to Han-06 (Table 2).

Reservoirs that adopt policies from similarly purposed, neighboring reservoirs are also improved, on average, over Han-06. STARFITx reduces storage error relative to Han-06 for 61% of reservoirs, while associated simulated releases ($R_{S\text{-}SIM}$) are improved for 65–71% of reservoirs. Unsurprisingly, STARFITx underperforms STARFIT. We find no obvious common features leading to greatest losses in performance in the STARFITx relative to STARFIT. Neither reservoir specifications (storage size, purpose, etc.) nor reservoir parameter values are correlated with performance deterioration in extrapolation mode.

STARFITx requires both storage targets and release functions to be copied from neighboring, similarly purposed reservoirs. Results for $R_{S\text{-}OBS}$ (storage set of observed rather than simulated) with STARFITx show decreased performance relative to Han-06 on average as well as decreased performance relative to $R_{S\text{-}SIM}$. This occurs because observed storage values often deviate outside an over-narrow NOR copied from another reservoir, causing frequent use of maximum and minimum release constraints at inappropriate times. The issue is less prominent for nRMSEˆ (flood performance) because the upper bound of NOR is often close to reservoir capacity, making it more amenable to successful copying. Overuse of minimum release is avoided if storage is simulated ($R_{S\text{-}SIM}$) and guided back inside the Normal Operating Range. Simulated storage in turn suffers a performance reduction when the copied NOR is unrepresentative of actual typical storages. Although these finding highlight fallibility in copying the NOR from another reservoir, performance of S and $R_{S\text{-}SIM}$ storage suggest that this is still preferable to Han-06 in most cases.

Performances of STARFIT and STARFITx vary widely across reservoirs We observe some very substantial gains in performance over Han-06 in some cases (Fig. 5; see Supplementary Materials for similar plot with absolute values instead of performance relative to benchmark); for storage simulation, reductions in nRMSE of $>$ 75% are observed for $\sim$ 35% of reservoirs with STARFIT and $\sim$ 15% of reservoirs with STARFITx. We find no particular category of reservoir for which performance improvements are greatest or most robust in general. For example, improvements for reservoirs with large storage capacity (>3,000 MCM) are distributed similarly the whole sample, as highlighted in Fig. 5. Reservoirs that lack the release adjustment parameters (which are discarded in one quarter of reservoirs due to poor fit of release with respect to water availability) do not perform significantly better or worse than those fitted with these functions either.

STARFIT underperforms relative to Han-06 in 13% of dams and 25% of dams for simulated storage and release, respectively. This is usually caused by over-constrained maximum or minimum release; operations that are difficult to guide with data (e.g., release decisions outside of the NOR) often underperform Han-06. Other instances of relative underperformance arise in cases with short periods of continuous release data (less than1 year) to perform the daily evaluation. We require five years' inflow data to drive the simulations, but for a small number of cases the observations of release or storage are incomplete during this period. In many of these cases we would expect STARFIT to exceed Han-06 performance with a lengthier sample. In general, stronger performance gains could be harnessed with further data extensions and refinements to ResOpsUS as well a more nuanced approach to defining maximum and minimum release constraints, which is out of scope for the present study.

STARFITx is evaluated by assuming that each data-rich reservoir is data-scarce, and then copying from a similarly purposed, neighboring reservoir. Assessing STARFITx using data-rich reservoirs raises the possibility of bias in overall performance, since most of these reservoirs are located in regions where other reservoirs are situated. In other words, a data-scarce reservoir in California is more likely to find a similarly purposed, nearby reservoir to copy than a data-scarce reservoir in New England, where reservoirs with data are less prevalent. To demonstrate that this potential bias is absent, we show distributions of performance improvement in both STARFIT and STARFITx (relative to Han-06) across individual HUC2 regions of CONUS (Fig. 6, see Supplementary Materials for similar plot with absolute values instead of performance relative to benchmark). Robust improvement of STARFIT and STARFITx relative to Han-06 in both simulated release and storage is found for all almost all regions. There is no apparent correlation between deterioration of performance in STARFITx (relative to STARFIT) and data prevalence of region. One region where STARFITx performances drop dramatically relative to STARFIT is the Pacific Northwest—a region with excellent reservoir operations data coverage. Complex or highly tailored operating arrangements for environmental flows, hydropower production, and flood control may make operating rules for these reservoirs less amenable for transfer to other reservoirs. Despite this effect, STARFITx still offers improvements for this region relative to Han-06 for storage simulations on average.

### 3.2. Alternative extrapolation settings

Evaluation of alternative extrapolation settings "X1" (farthest similarly purposed reservoir within basin) and "X2" (closest reservoir by distance, disregarding operational purpose) show that the chosen extrapolation strategy ("X") is prudent. Both settings X1 and X2 result in a modest increase the median error score and decrease in number of reservoirs improved relative to Han-06 (Table 3). These reported impacts are nullified somewhat by the fact that many reservoirs are copying from the same target reservoirs irrespective of the extrapolation scheme. For example, often there is only one reservoir in a HUC4 basin that has close operational alignment to the target. In those cases, X and X1 will be result in the same target reservoir. In some cases the best aligned reservoir operationally will also be the closest by distance, and in those cases settings X and X2 will copy from the same reservoir. For setting X1, approximately 37% of evaluated reservoirs change the target reservoir, and approximately 60% of these underperform STARFITx. For

**Table 2**

Summary of model error metric scores across all reservoirs (median scores reported). Performances are given for simulated release (assuming observed storage, $R_{S\text{-}OBS}$, and assuming simulated storage, $R_{S\text{-}SIM}$), and simulated storage (S).

| Variable | Metric | Han-06 | STARFIT | *% reservoirs improved* | STARFITx | *% reservoirs improved* |
|---|---|---|---|---|---|---|
| $R_{S\text{-}OBS}$ | nRMSE (%) | 101.9 | 83.1 | *91 %* | 112.8 | *39 %* |
| | nTRMSE (%) | 101.4 | 82.8 | *88 %* | 124.8 | *34 %* |
| | nRMSEˆ (%) | 125.8 | 102.9 | *85 %* | 116.1 | *62 %* |
| $R_{S\text{-}SIM}$ | nRMSE (%) | 100.1 | 81.3 | *75 %* | 85.7 | *66 %* |
| | nTRMSE (%) | 96.9 | 79.9 | *78 %* | 83.4 | *65 %* |
| | nRMSEˆ (%) | 117.6 | 100.3 | *71 %* | 98.7 | *71 %* |
| S | nRMSE (%) | 327.8 | 110.8 | *87 %* | 181.6 | *61 %* |
| | nTRMSE (%) | 316.8 | 108.1 | *88 %* | 185.0 | *61 %* |
| | nRMSEˆ (%) | 438.6 | 156.0 | *87 %* | 279.7 | *61 %* |

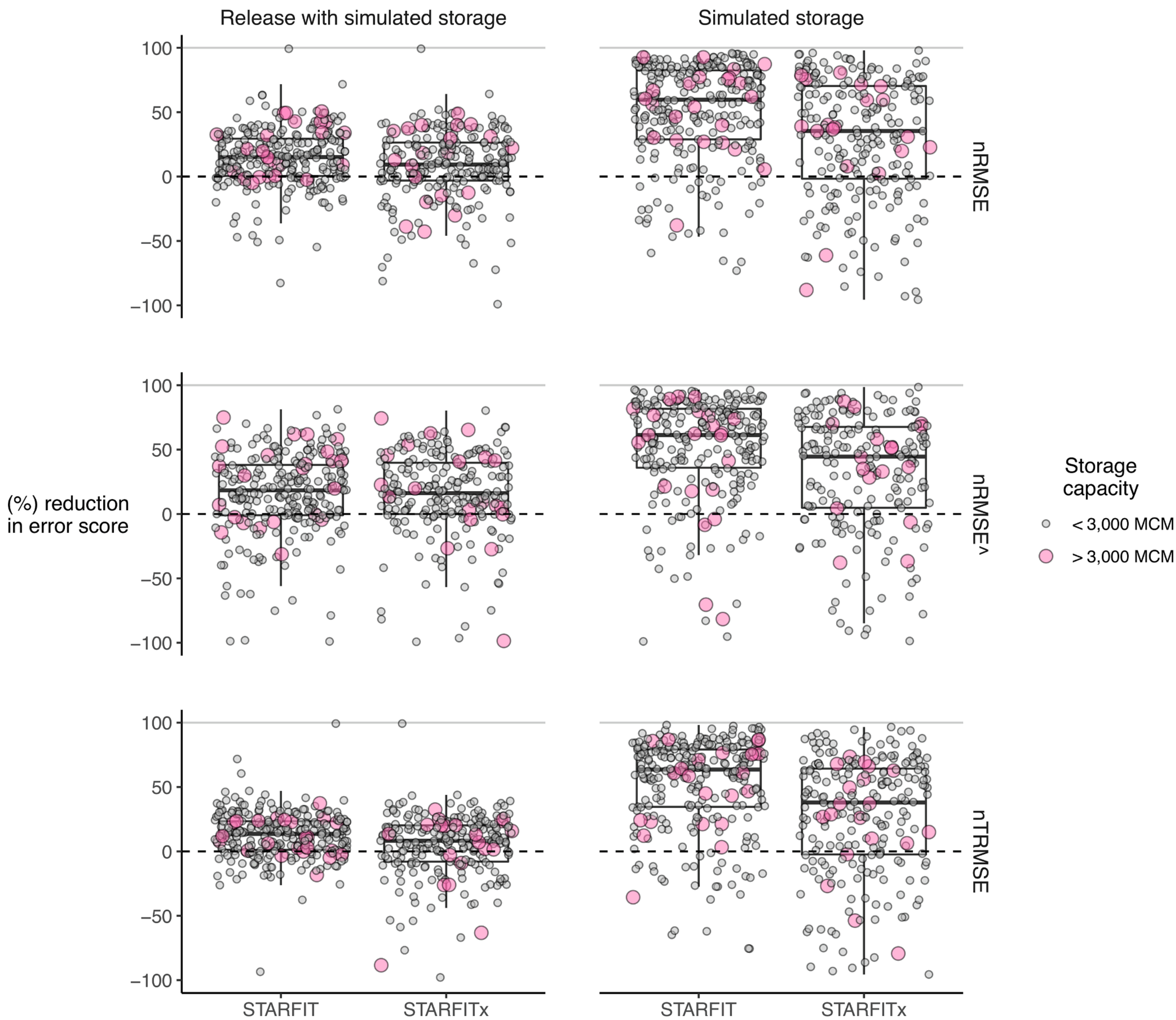


**Fig. 5.** Error reduction for STARFIT and STARFITx relative to Han-06 for simulated storage (S) and release ($R_{S\text{-}SIM}$) variables and three performance metrics. Each point represents a simulated reservoir. Boxplots give the interquartile range with whiskers extending to 1.5 times the interquartile range.

setting X2, about half of reservoirs change the target, and of these 65% underperform STARFITx. These results show indicate that copying the operating policy from a remote reservoir within good operational alignment and within the same hydrological unit can be a legitimate strategy in many cases.

## 4. Discussion

The performance enhancements in both simulated storage and release reported above indicate that use to ISTARF-CONUS would markedly improve on Han-06 for reservoirs where sufficient training data are available (n = 595) and would in most cases also outperform Han-06 when operating rules are copied from a similarly purposed reservoir in the same river basin (n = 1,335). These results indicate that ISTARF-CONUS can be readily adopted in CONUS-scale LHMs or global hydrological models that simulate large CONUS reservoirs. Performance improvements in an LHM simulation may differ to what we found in this study, in which observed inflow forcing was applied. LHMs may feature significant inflow bias due to error in climatic forcing as well as representation of physical catchment processes, flow routing, and upstream water consumption and regulation. However, ISTARF-CONUS will retain strong advantages over generic schemes irrespective of inflow conditions—particularly storage representation guided by the targeted operating range and, by extension, a reservoir's ability to buffer extreme floods or sustain supply during prolonged drought.

The significant difference in performance between STARFIT and STARFITx highlights what could be gained with increased access to reservoir operations records of storage, inflow, and release. Our operations are trained using ResOpsUS—the most comprehensive multi-agency inventory of multi-decadal reservoir operations available for the United States. Many of the reservoirs we label "data-scarce" are in fact measured at regular intervals and have documented records that could perhaps be collected through further agency outreach and survey. Although only covering a few years of operations, new satellite-based reconstructions of reservoir storage levels may also be introduced to extend and supplement reservoir observational records (Cooley et al., 2021). These data could be used to estimate the normal operating range for each reservoir. Combined with reanalysis climate to drive historical river flow simulations, they could also be used to back calculate weekly release and thus develop storage and release schemes for the United States or globally.

By computing all release decisions relative to mean long-term inflow,

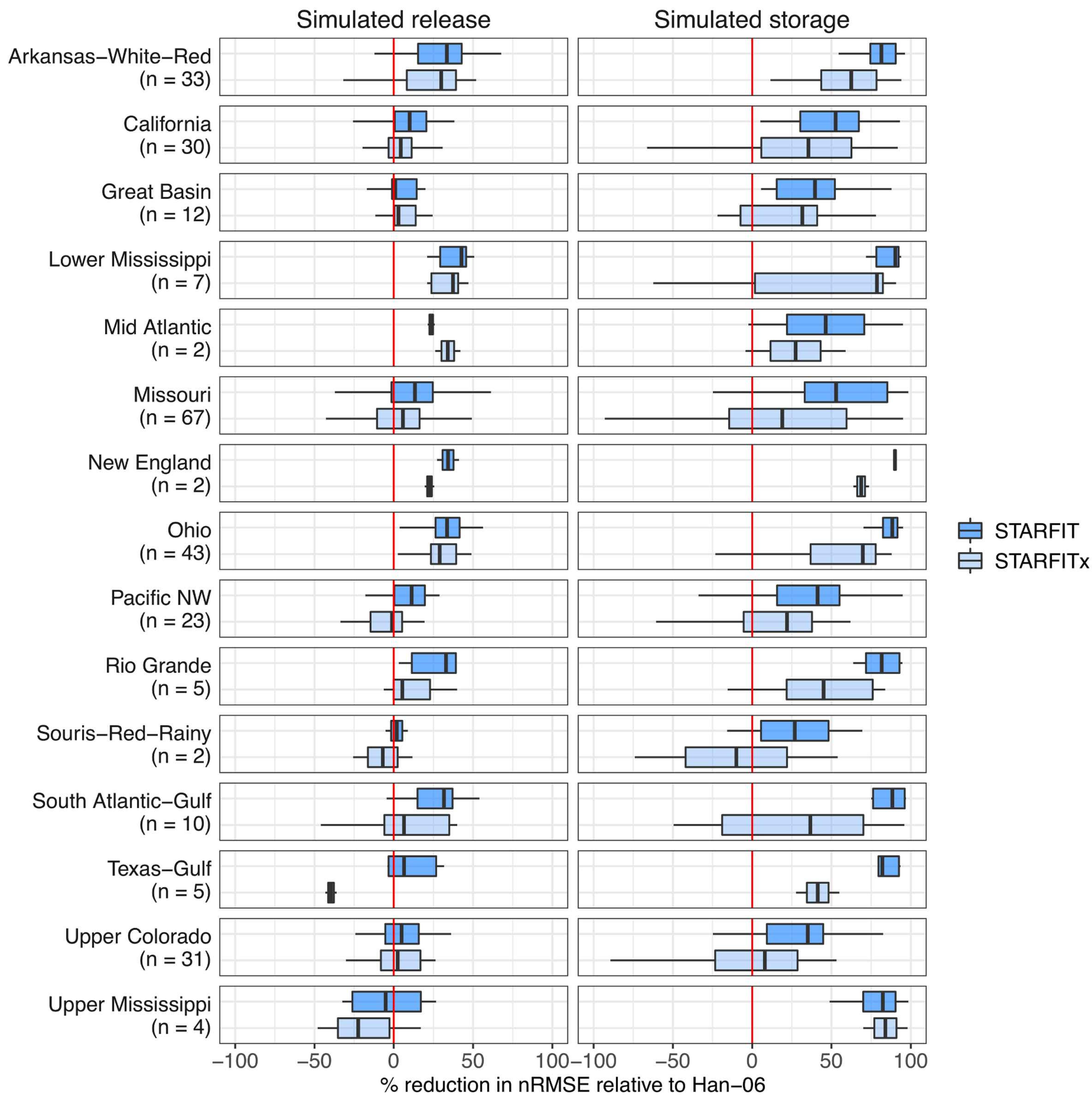


**Fig. 6.** Performance improvements in nRMSE relative to Han-06 for simulated release ($R_{S\text{-}SIM}$) and storage (S) with observed inflows. Performances cannot be evaluated for Great Lakes (no reservoirs represented ResOpsUS) or for Tennessee and Lower Colorado HUC2 regions (for which weekly policies are developed but cannot be assessed here due to lack observed daily inflow to drive daily-resolution simulation).

**Table 3**

Summary of model error metric scores across all reservoirs (median scores reported) for Han-06 and three STARFIT extrapolation settings (X: same purpose and shortest distance; X1: same purpose and longest distance within basin; X2: shortest distance only). Data for % of reservoir simulations improved is relative to Han-06.

| Variable | Metric | Han-06 | X | *% reservoirs improved* | X1 | *% reservoirs improved* | X2 | *% reservoirs improved* |
|---|---|---|---|---|---|---|---|---|
| $R_{S\text{-}SIM}$ | nRMSE (%) | 100.1 | 85.7 | *66 %* | 87.5 | *61 %* | *86.7* | *61 %* |
| | nTRMSE (%) | 96.9 | 83.4 | *65 %* | 86.4 | *61 %* | *86.4* | *61 %* |
| | nRMSEˆ (%) | 117.6 | 98.7 | *71 %* | 103.8 | *65 %* | *99.2* | *69 %* |
| S | nRMSE (%) | 327.8 | 181.6 | *61 %* | 202.3 | *57 %* | *190.9* | *57 %* |
| | nTRMSE (%) | 316.8 | 185.0 | *61 %* | 205.3 | *55 %* | *197.0* | *57 %* |
| | nRMSEˆ (%) | 438.6 | 279.7 | *61 %* | 291.9 | *56 %* | *310.9* | *60 %* |

ISTARF-CONUS gives the user the ability to adjust operations for reservoirs subject to inflow bias. This may be achieved without redefining the release model parameters, but instead by substituting mean inflow ($\bar{I}$) applied in Eq. (7) with average long-term inflow determined in a prior simulation of the LHM. Accounting for inflow bias in this way would result in reservoir release propagating the bias downstream while maintaining realistic storage levels to accommodate flood or drought. This feature also provides the capability to adjust operations for a changing climate dynamically within the simulation, for example by updating the mean annual inflow every ten years of simulation as climate conditions evolve.

STARFIT is one of many possible approaches one could follow to infer reservoir operating policies. One alternative is to identify the operating objectives of the reservoir, describe those objectives in a single mathematical function, and optimize a policy using stochastic dynamic programming (Castelletti et al., 2008). This approach could allow for more realistic operations under changing conditions (Giudici et al., 2021), although determining representative operating objectives across a large sample of reservoirs presents a significant challenge. Machine learning methods aimed reproducing historical decisions could be another avenue to explore, although these models require much longer sequences of data than are available for the problem of identifying a reservoir policy. Irrespective of the method employed, there is a limit to how well actual reservoir operations can be represented in a typical large-scale hydrological simulation in which reservoir operations lack forecast information and are uncoordinated. Actual reservoir operators often use medium to long-range forecasts to guide operations (with varying lead times throughout the typical operating year), and realistic representation of such forecasts in an LHM simulation remains a significant challenge—even if forecast use is well described in the operating policy (Turner et al., 2020a). Coordination of reservoir releases within a river basin is another important driver of decision making not well accommodated in LHMs (Rouge et al., 2021). A dataset of observations like ResOpsUS could be applied to study the important effects of reservoir system coordination and develop models to represent this within an LHM. This would impose a significant computational burden that may make such advances impractical at this time.

## 5. Conclusions

This study presents ISTARF-CONUS—a new dataset of water storage and release policies for all 1,930 large reservoirs of CONUS—and its associated policy inference algorithm, STARFIT. Performance evaluation of this scheme shows that it offers substantial and robust simulation improvement over a state-of-the-art approach adopted in many LHMs. Performance enhancements are most significant for simulated storage, which is guided by a Normal Operating Range identified by the inference algorithm. Improved storage representation also benefits release, since release is a function of storage. Even though most CONUS reservoirs lack data sufficient to train these operating schemes, we show that a simple extrapolation of policies from similarly purposed, neighboring reservoirs that are replete with operating data will still yield modest enhancements over conventional practice on average. The overall performance of the scheme will improve as and when more observational data are added. ISTARF-CONUS can be readily adopted into any LHM that represents large CONUS reservoirs or may be analyzed for insight into the diverse range of operating strategies employed in United States dam and reservoir operations.

## 6. Data statement

Links to data and code used to perform this experiment are available via a supporting meta-repository: https://immm-sfa.github.io/Turner-etal_2021_JoH/. This includes a link to the inventory of observed reservoir operations, ResOpsUS. The ISTARF-CONUS dataset of inferred reservoir operating policies is freely available at https://doi.org/10.5281/zenodo.4602277 and includes an implementation demo using STARFIT. The STARFIT algorithm is available as an open-source R library, available at Github (https://github.com/IMMM-SFA/starfit).

## Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Acknowledgements

This research was supported by the U.S. Department of Energy, Office of Science, as part of research in the MultiSector Dynamics, Earth and Environmental System Modeling Program.

## Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jhydrol.2021.126843.

## References


Biemans, H., Haddeland, I., Kabat, P., Ludwig, F., Hutjes, R.W.A., Heinke, J., Von Bloh, W. & Gerten, D. 2011. Impact of reservoirs on river discharge and irrigation water supply during the 20th century. Water Resources Research, 47.

Boulange, J., Hanasaki, N., Yamazaki, D., Pokhrel, Y., 2021. Role of dams in reducing global flood exposure under climate change. Nat. Commun. 12, 417.

Castelletti, A., Pianosi, F., Soncini-Sessa, R., 2008. Water reservoir control under economic, social and environmental constraints. Automatica 44, 1595–1607.

Coerver, H.M., Rutten, M.M., van de Giesen, N.C., 2018. Deduction of reservoir operating rules for application in global hydrological models. Hydrol. Earth Syst. Sci. 22, 831–851.

Cooley, S.W., Ryan, J.C., Smith, L.C., 2021. Human alteration of global surface water storage variability. Nature 591 (7848), 78–81.

Fleischmann, A.S., Brêda, J.P.F., Passaia, O.A., Wongchuig, S.C., Fan, F.M., Paiva, R.C.D., Marques, G.F., Collischonn, W., 2021. Regional scale hydrodynamic modeling of the river-floodplain-reservoir continuum. J. Hydrol. 596, 126114.

Giudici, F., Anghileri, D., Castelletti, A., Burlando, P., 2021. Descriptive or normative: How does reservoir operations modeling influence hydrological simulations under climate change? J. Hydrol. 595, 125996.

Haddeland, I., Skaugen, T., Lettenmaier, D.P., 2006. Anthropogenic impacts on continental surface water fluxes. Geophys. Res. Lett. 33.

Hanasaki, N., Kanae, S., Oki, T., 2006. A reservoir operation scheme for global river routing models. J. Hydrol. 327, 22–41.

Lehner, B., Liermann, C.R., Revenga, C., Vorosmarty, C., Fekete, B., Crouzet, P., Doll, P., Endejan, M., Frenken, K., Magome, J., Nilsson, C., Robertson, J.C., Rodel, R., Sindorf, N., Wisser, D., 2011. High-resolution mapping of the world's reservoirs and dams for sustainable river-flow management. Front. Ecol. Environ. 9, 494–502.

Macian-Sorribes, H., Pulido-Velazquez, M., 2017. Integrating Historical Operating Decisions and Expert Criteria into a DSS for the Management of a Multireservoir System. J. Water Resour. Plann. Manage. 143.

Masaki, Y., Hanasaki, N., Biemans, H., Schmied, H.M., Tang, Q., Wada, Y., Gosling, S.N., Takahashi, K., Hijioka, Y., 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. Environ. Res. Lett. 12 (5), 055002. https://doi.org/10.1088/1748-9326/aa57a8.

Masaki, Y., Hanasaki, N., Takahashi, K., Hijioka, Y., 2018. Consequences of implementing a reservoir operation algorithm in a global hydrological model under multiple meteorological forcing. Hydrological Sciences Journal-Journal Des Sciences Hydrologiques 63, 1047–1061.

Nazemi, A., Wheater, H.S., 2015. On inclusion of water resource management in Earth system models - Part 2: Representation of water supply and allocation and opportunities for improved modeling. Hydrol. Earth Syst. Sci. 19, 63–90.

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., Wada, Y., 2021. Global terrestrial water storage and drought severity under climate change. *Nature*. Clim. Change 11 (3), 226–233.

Powell, M. J. 1994. A direct search optimization method that models the objective and constraint functions by linear interpolation. Advances in optimization and numerical analysis. Springer.

Rouge, C., Reed, P.M., Grogan, D.S., Zuidema, S., Prusevich, A., Glidden, S., Lamontagne, J.R., Lammers, R.B., 2021. Coordination and control - limits in standard representations of multi-reservoir operations in hydrological modeling. Hydrol. Earth Syst. Sci. 25, 1365–1388.

Turner, S.W.D., Doering, K., Voisin, N., 2020a. Data-Driven Reservoir Simulation in a Large-Scale Hydrological and Water Resource Model. Water Resour. Res. 56.

Turner, S.W.D., Xu, W.W., Voisin, N., 2020b. Inferred inflow forecast horizons guiding reservoir release decisions across the United States. Hydrol. Earth Syst. Sci. 24, 1275–1291.

van Werkhoven, K., Wagener, T., Reed, P., Tang, Y., 2009. Sensitivity-guided reduction of parametric dimensionality for multi-objective calibration of watershed models. Adv. Water Resour. 32, 1154–1169.

Veldkamp, T.I.E., Wada, Y., Aerts, J.C.J.H., Doll, P., Gosling, S.N., Liu, J., Masaki, Y., Oki, T., Ostberg, S., Pokhrel, Y., Satoh, Y., Kim, H. & Ward, P.J. 2017. Water scarcity hotspots travel downstream due to human interventions in the 20th and 21st century. Nature Communications, 8.

Veldkamp, T.I.E., Zhao, F., Ward, P.J., de Moel, H., Aerts, J.C.J.H., Schmied, H.M., Portmann, F.T., Masaki, Y., Pokhrel, Y., Liu, X., Satoh, Y., Gerten, D., Gosling, S.N., Zaherpour, J., Wada, Y., 2018. Human impact parameterizations in global hydrological models improve estimates of monthly discharges and hydrological extremes: a multi-model validation study. Environ. Res. Lett. 13 (5), 055008. https://doi.org/10.1088/1748-9326/aab96f.

Voisin, N., Dyreson, A., Fu, T., O'Connell, M., Turner, S.W.D., Zhou, T., Macknick, J., 2020. Impact of climate change on water availability and its propagation through the Western US power grid. Appl. Energy 276.

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 Resour. Res. 53, 4253–4270.

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. Hydrol. Earth Syst. Sci. 17, 3605–3622.

Wada, Y., Bierkens, M.F.P., de Roo, A., Dirmeyer, P.A., Famiglietti, J.S., Hanasaki, N., Konar, M., Liu, J.G., Schmied, H.M., Oki, T., Pokhrel, Y., Sivapalan, M., Troy, T.J., van Dijk, A.I.J.M., van Emmerik, T., van Huijgevoort, M.H.J., van Lanen, H.A.J., Vorosmarty, C.J., Wanders, N., Wheater, H., 2017. Human-water interface in hydrological modelling: current status and future directions. Hydrol. Earth Syst. Sci. 21, 4169–4193.

Wada, Y., Wisser, D., Bierkens, M.F.P., 2014. Global modeling of withdrawal, allocation and consumptive use of surface water and groundwater resources. Earth Syst. Dyn. 5, 15–40.

Yassin, F., Razavi, S., Elshamy, M., Davison, B., Sapriza-Azuri, G., Wheater, H., 2019. Representation and improved parameterization of reservoir operation in hydrological and land-surface models. Hydrol. Earth Syst. Sci. 23, 3735–3764.