Received: 20 September 2022 | Accepted: 12 August 2023

DOI: 10.1111/1752-1688.13161

AMERICAN WATER RESOURCES ASSOCIATION
JAWRA
Journal of the American Water Resources Association

RESEARCH ARTICLE

The consequences of neglecting reservoir storage in national-scale hydrologic models: An appraisal of key streamflow statistics

Glenn A. Hodgkins[1] | Thomas M. Over[2] | Robert W. Dudley[3] | Amy M. Russell[2] | Jacob H. LaFontaine[4]

[1]U.S. Geological Survey New England Water Science Center, Augusta, Maine, USA

[2]U.S. Geological Survey Central Midwest Water Science Center, Urbana, Illinois, USA

[3]U.S. Geological Survey New England Water Science Center, Pembroke, New Hampshire, USA

[4]U.S. Geological Survey South Atlantic Water Science Center, Norcross, Georgia, USA

Correspondence
Glenn A. Hodgkins, U.S. Geological Survey New England Water Science Center, Augusta, ME, USA.
Email: gahodgki@usgs.gov

Abstract

A better understanding of modeled streamflow errors related to basin reservoir storage is needed for large regions, which normally have many ungaged basins with reservoirs. We quantified the difference between modeled and observed streamflows for one process-based and three statistical-transfer hydrologic models, none of which explicitly accounted for reservoir storage. Streamflow statistics representing low to high flows, seasonality, annual variability, and daily autocorrelation were examined at 1082 study basins across the conterminous USA. All models increasingly overpredict (or decreasingly underpredict) observed annual maximum flows with increasing storage. Correlations between absolute values of errors for low-flow statistics and storage are often larger in magnitude than those for signed errors—additional storage is associated with increases in model errors in both directions even when its overall effect in one direction is weak. The rate of increase in absolute values of model errors was nonlinear for most statistics. For low flows, model errors had a change point to larger errors at 48 days of reservoir storage (relative to long-term mean daily flow); mean and high flows had change points at 147 to 176 days. We present predicted-to-observed errors for nine streamflow statistics over a large range of reservoir storage to help modelers and users of modeled streamflow understand the amount of storage for which explicit reservoir modeling is needed.

KEYWORDS
streamflow, hydrologic models, accuracy, reservoir storage, United States

1 | INTRODUCTION

Knowledge of the amount of water in streams and rivers is needed for reliable water supply (e.g., Miller et al., 2021), safe bridge design (McCuen et al., 2002), and aquatic habitat requirements (Hain et al., 2018), to list a few critical issues. Long-term data are needed to accurately quantify various flow statistics of interest. According to one estimate, there are over 4,800,000 km of streams and rivers in the United States

Paper No. JAWR-22-0115-P of the *Journal of the American Water Resources Association* (JAWR).

Discussions are open until six months from publication.

Research Impact Statement

We quantify changes in the accuracy of modeled streamflows across the United States as reservoir storage increases; errors increase in a nonlinear way for multiple streamflow statistics.

(Leopold, 1962). However, it is only feasible to collect long-term streamflow data on a very small percentage of these stream reaches, and there are many spatial and temporal gaps in observed streamflow records (e.g., Kiang et al., 2013). For ungaged reaches of interest, methods are needed to estimate streamflow; this can be done with either process-based models or statistical-transfer models (Hodgkins et al., 2020).

Evaluation of modeled streamflows against observed streamflows is crucial for determining how well models perform but relatively few studies have evaluated models for a range of streamflows at many basins over large, hydrologically diverse areas (e.g., Farmer et al., 2014; Gudmundsson et al., 2012; Hodgkins et al., 2020; Stahl et al., 2011). All of these studies looked at the accuracy of models for multiple flow statistics for hundreds of basins that were minimally influenced by anthropogenic effects, which is a common approach to basin modeling (Archfield et al., 2015). Low-streamflow model errors were larger than errors for other flow statistics in all of these European and North American studies, highlighting the importance of considering a range of flow statistics in model evaluations. In addition, there was considerable variation among the results for the various basins in these large, diverse areas, which points to the importance of a large-sample approach in hydrological modeling (Gupta et al., 2014).

Because human alteration of basins is common, it is important to test model accuracy at basins affected by such alteration, in addition to minimally altered ones. The 75,000 dams in the conterminous United States can store an average of about a year's worth of mean annual streamflow (Graf, 1999). Though storage reservoirs in a basin typically reduce flood peaks, low flows can decrease or increase depending on operational flow releases (Carlisle et al., 2019; Eng et al., 2013, 2019; FitzHugh & Vogel, 2011; Graf, 2006; Magilligan & Nislow, 2005; McManamay et al., 2012; Poff et al., 2007; Smakhtin, 2001; Wang et al., 2017; Williams & Wolman, 1984; Yang et al., 2021). Flood-control reservoirs reduce high-flow magnitudes by design, while reservoirs designed for other purposes such as drinking-water supply also can reduce high-streamflow magnitude (Carlisle et al., 2019; Graf, 2006). Water losses through evaporation, seepage, and water use associated with dams and reservoirs reduce the total amount of water released by dams (Graf, 1999; Wang & Hejazi, 2011; Zhao & Gao, 2019).

Based on analyses of 3355 basins in the US, more altered basins (combining the effects of flow regulation from storage reservoirs, irrigation, and urbanization), Eng et al. (2019) observed more decreases in low-flow magnitude than increases in the western deserts, western mountains, western plains, and coastal plains of the United States, whereas more altered basins had increased low flows in all other regions of the United States. Flood-control reservoirs, which are common in areas with wetter climates such as the Northeastern United States, tend to release stored water into streams during low-flow periods (Carlisle et al., 2019). The annual variability of low and high flows in streams in altered basins across the United States has generally decreased relative to natural conditions (Carlisle et al., 2019; Eng et al., 2019). Eng et al. (2019) found a low-flow seasonality shift for about 25% of regulated basins in the United States with the occurrence of low flows shifting from fall to other seasons. About 30% of regulated basins had a high-flow seasonality shift from spring and winter to other seasons. In a study that included 755 streamgages in the continental US having regulation indicated by their peak-data codes, Villarini (2016) found that the concentration in the seasonality of peak flows decreased after the onset of regulation but that the mean timing of seasonality generally was not significantly different.

Several recent studies focus on improving hydrologic models at the large scale by including reservoir storage (Chawanda et al., 2020; Dang et al., 2020; Gochis et al., 2018; Ouyang et al., 2021; Shin et al., 2019; Tefs et al., 2021; Yassin et al., 2019; Zajac et al., 2017), showing the recognized importance of modeling reservoirs to more accurately estimate streamflows for ungaged rivers. Zajac et al. (2017) compared modeled (LISFLOOD distributed hydrologic model) to observed daily flows for 390 large rivers worldwide. The skill of models for predicting daily streamflows, based on the Nash–Sutcliffe Efficiency coefficient (NSE), was better for a majority of study basins when lakes and reservoirs were modeled. They also looked at model accuracy for peak flows with 5- and 20-year return periods and found that including reservoirs and lakes in the models generally led to a smaller range of model errors but also to overprediction of peak flows. Ouyang et al. (2021) examined how well long short-term memory (LSTM) models predicted observed daily streamflows for 3557 basins across the conterminous US (CONUS) with a large range of reservoir storage. They found that the groups of basins with (1) no reservoir storage, (2) small amounts of storage (greater than zero to about 1 month of storage of the mean annual runoff), and (3) large amounts of storage (greater than about 1 month) had different overall NSE values when trained on all CONUS basins. Models for the group with small amounts of storage had the highest overall NSEs, followed by the group with no storage; the group with large amounts of storage had the lowest overall NSEs.

There has been limited work on how the amount of reservoir storage in a basin affects errors in the modeling of various streamflow statistics such as low flows, high flows, and statistics representing flow seasonality. Previous studies have looked at the influence of basin alteration by humans on various streamflow statistics by estimating natural flows at basins (e.g., Carlisle et al., 2019; Eng et al., 2019; Hailegeorgis &

Alfredsen, 2017). We are unaware of studies that have quantified the effects of ignoring reservoir storage in large-scale hydrologic models across a range of reservoir storage, in terms of typical model-error magnitudes (predicted to observed) for a variety of streamflow statistics.

1.1 | Study design

We analyze the effects of reservoir storage on hydrologic model accuracy for one process-based model and three statistical transfer models at 1082 basins across the CONUS that have a wide range of reservoir storage. We compare modeled to observed streamflows with increasing reservoir storage, for nine streamflow statistics that include a range of flow magnitudes and measures of flow variability and seasonality. The importance of these flow statistics to various problems in water management, including low-flow regulation, daily streamflow prediction, design of water supply systems, and characterization of floods is described in Section 2.6. We analyze whether increasing amounts of reservoir storage results in decreased model accuracy and whether the direction of the errors makes sense based on previous knowledge of the effects of reservoir storage on various streamflow statistics.

The results of this study will help modelers understand the level of reservoir storage that substantially affects a variety of streamflow statistics across a range of hydrologically diverse basins. If errors are larger than considered acceptable for various modeling applications, modeling capability for the effects of reservoir storage on streamflow can be added. It will further help users of the model predictions to know what levels of storage may cause unacceptably large errors.

2 | DATA AND METHODS

Modeled streamflows used in our study were developed using models based on reference (minimally altered) basins, and therefore the model predictions are estimates of natural flows. These models have been set up and applied to predict natural daily streamflows at a high spatial resolution across the CONUS. Model estimates are compared to observed flows that include variable amounts of basin reservoir storage.

2.1 | Observed streamflow data

Observed daily streamflow data from water year 1984 (beginning October 1, 1983) through water year 2016 (ending September 30, 2016; a water year is named for the year in which it ends) were downloaded from the USGS National Water Information System (NWIS) (US Geological Survey, 2019) for all non-reference streamgages (Falcone, 2011) in the CONUS. Approved streamflow data were downloaded using the R package 'dataRetrieval' (Hirsch & De Cicco, 2015).

Streamflow records were subjected to length and completeness testing. Complete water years were required to have a daily value for every day of the water year. Each decade (e.g., 1990–1999) was required to have a minimum of 8 complete years and partial decadal periods of 1984–1989 and 2010–2016 were required to have 5 of 6 and 6 of 7 years complete, respectively. Observed streamflow data used in this study are available in Russell et al. (2021).

2.2 | Streamgage selection and classification

Candidate streamgages were screened and classified on the basis of several measures of human influence tabulated in the GAGES-II database (Falcone, 2011) to focus on the effects of reservoir storage and minimize other direct basin effects. All candidate streamgages were required to have 2 percent or less impervious area (IMPNLCD06 field in Falcone, 2011) to minimize urban-related effects on streamflow (e.g., impervious area and stormwater drainage systems). Streamgages with any evidence of interbasin transfers of water in the USGS water-report remarks regarding hydrologic modifications (WR_REPORT_REMARKS) also were removed from the study. Basins with 1 percent or more irrigated agricultural land (PCT_IRRIG_AG), or any mention of irrigation in the USGS water-report remarks regarding hydrologic modifications, were also removed from the analysis. This screening resulted in 1082 study basins (Figure 1) with a large range of basin reservoir storage but very little other human basin impacts.

We computed days of reservoir storage, for each study basin, as our metric of the degree of reservoir regulation on streamflow. This variable was computed by dividing the GAGES-II normalized storage characteristic (STOR_NID_2009; Falcone, 2011) in megaliters per square kilometer (equivalent to millimeters) by observed mean daily flow (1984–2016) in millimeters per day. STOR_NID_2009 was computed for all dams based on an enhanced version of 2009 National Inventory of Dams storage which is generally equal to the maximum storage of the reservoir. We used STOR_NID_2009 because of inconsistencies noted by Falcone (2011) in the normal storage (STOR_NOR_2009); specifically,

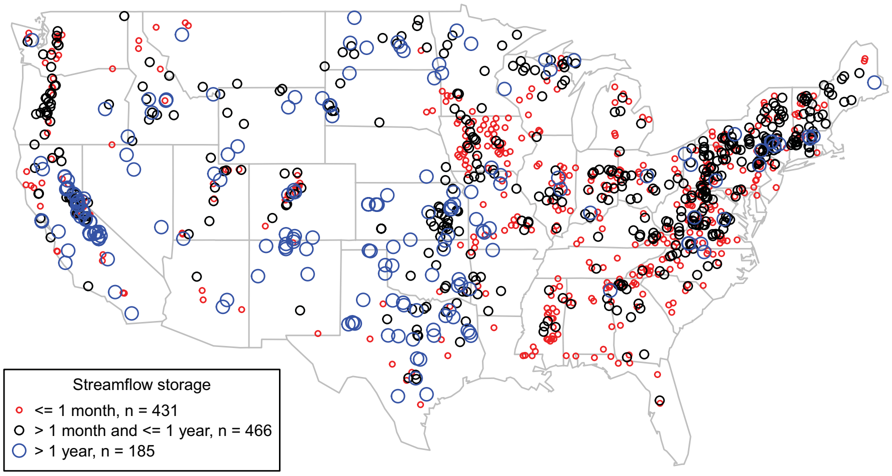

FIGURE 1 Study basins for relations between model errors and reservoir storage. Symbols represent the amount of reservoir storage in each of 1082 study basins, in relation to the long-term mean daily streamflow for each basin.

dams which obviously had large storage but had normal storage of zero. The NID normal storage represents the total storage in a reservoir below the normal retention level, including dead storage. Maximum storage also includes storage between the normal storage level and the maximum attainable level (National Inventory of Dams, 2022).

The basin reservoir storage in our study ranges from 0.0 to 67,400 days with a median of 55.7 days. A majority of basins have reservoir storage between 1 and 1000 days. Note that the study basins include many basins with very little regulation. We retain these basins to allow comparison of model errors across a wide range of basin storage. Basins qualifying for our study are located across the country; basins with the largest amounts of normalized storage (>1 year) are predominantly located throughout the western United States (Figure 1). Attributes and classifications of streamgages used in this study are available in Dudley and Hodgkins (2021).

2.3 | Streamflow estimates from a process-based model

The National Hydrologic Model Infrastructure application of the Precipitation-Runoff Modeling System (NHM-PRMS; LaFontaine et al., 2019; Leavesley et al., 1983; Markstrom et al., 2015; Regan et al., 2018, 2019) is a daily timestep, deterministic, distributed-parameter, process-based model application that simulates the effects of precipitation and air temperature on basin hydrology. We used modeled streamflow estimates from the NHM-PRMS stream segments that coincide with the streamgage locations of the basins used in this study; basins that have limited to substantial amounts of basin storage. The NHM-PRMS model application for these basins did not, by design, model basin reservoir regulation. The NHM-PRMS simulation outputs are available in a USGS data release (Hay & LaFontaine, 2020).

Daily values of precipitation and air temperature from the Daily Surface Weather and Climatological Summaries (Daymet) version 3 (Thornton et al., 1997, 2000, 2017) are used in the NHM-PRMS modeling to compute potential evapotranspiration, actual evapotranspiration, snowmelt, streamflow, infiltration, and groundwater recharge (LaFontaine et al., 2019). A stream network for the CONUS was used to route streamflow components of surface runoff, shallow subsurface runoff, and groundwater flow to basin outlets. For basin delineation and connectivity, the NHM-PRMS application uses the geospatial fabric for National Hydrologic Modeling (Viger & Bock, 2014). The land surface in the geospatial fabric is divided into hydrologic response units (HRUs): there are 109,951 HRUs across the CONUS, and they have a median size of 33.2 km^2. The flow from the HRUs is routed through a stream network having 56,549 segments with Muskingum routing. Although only modeled streamflows at selected streamgages are used in the present analysis, the published data (Hay & LaFontaine, 2020) include daily flows for all segments and climate forcing data and hydrologic fluxes and states for all HRUs and stream segments across the CONUS.

A static model parameterization was estimated for all HRUs and stream segments using physical characteristics, initial water content, and climate as described in LaFontaine et al. (2019). The parameterization of the NHM-PRMS used in this study was calibrated in two steps: (1) "by-hydrologic-response-unit," where each of the HRUs were calibrated individually (LaFontaine et al., 2019) and (2) "by-headwater," where headwater subbasins 3000 km^2 or smaller were calibrated (independently from one another but all HRUs and segments in a given headwater were calibrated together) (LaFontaine et al., 2019). For the first calibration step, the hydrologic simulation for each HRU in the CONUS was calibrated to monthly runoff values from a water balance model, actual evapotranspiration, and snow-water equivalent (a single final objective function was computed which was comprised of 3 objective functions, one for each water budget component) (LaFontaine et al., 2019, appendix 1, equations 1-1 through 1-5). For the second calibration step, streamflow simulations constructed using ordinary kriging (Farmer, 2016) were used to optimize NHM-PRMS-derived streamflows at headwater subbasin outlets (not at streamgage locations), to calibrate for streamflow timing. See LaFontaine et al. (2019) for details on the NHM-PRMS modeling and Hodgkins et al. (2020) for a more extensive summary of modeling details than is presented here. The typical third calibration step for NHM-PRMS, "by-observations", which calibrates directly to observed daily streamflows at basins of interest, was not used. This step was not used because the objective of the current study is to compare modeled and observed streamflows at gaged locations such that the modeled streamflows estimate streamflows as if each stream was ungaged. The calibration steps were evaluated by computing NSE and percent bias at all streamgages available as points of interest in the NHM.

2.4 | Streamflow estimates from statistical-transfer models

Three statistical-transfer models were used to simulate natural streamflow at all streamgage locations. The three models were nearest-neighbor drainage area ratio (NNDAR), nearest-neighbor nonlinear spatial interpolation using flow-duration curves (NNQPPQ), and ordinary kriging of the logarithms of discharge per unit area (OKDAR). These models were fit using data from minimally affected, that is, reference, basins (Russell et al., 2020) and applied to make daily streamflow estimates at 5439 non-reference streamgages (Russell et al., 2021); the streamgages considered in this study are a subset of these streamgages. The NNDAR and OKDAR models used for Russell et al. (2020) have also been used to compute estimates of natural flow at the more than 80,000 outlets of the HUC-12 basins covering the CONUS (Russell et al., 2018).

Introductions to and a prior application of the NNDAR and NNQPPQ models can be found in Farmer et al. (2014). The flow-duration curve modeling used in the NNQPPQ model follows methods described by Over et al. (2018). The OKDAR model was introduced by Farmer (2016).

These models can be described with the following equations:

$$Q_{\text{predNNDAR}}(t) = \left(A_{\text{pred}} / A_{\text{obsNN}}\right) Q_{\text{obsNN}}(t), \tag{1}$$

where $Q_{\text{predNNDAR}}(t)$ is the NNDAR-predicted streamflow on day t, A_{pred} is the drainage area at the prediction location, A_{obsNN} is the drainage area at the nearest neighbor streamgage, and $Q_{\text{obsNN}}(t)$ is the observed streamflow at the nearest neighbor streamgage on day t;

$$Q_{\text{predNNQPPQ}}(t) = Q_{\text{predFDC}}\left(p_{\text{obsNN}}(t)\right), \tag{2}$$

where $Q_{\text{predNNQPPQ}}(t)$ is the NNQPPQ-predicted streamflow on day t, $Q_{\text{predFDC}}(p_{\text{obsNN}}(t))$ is the predicted flow-duration curve quantile for exceedance probability $p_{\text{obsNN}}(t)$ at the prediction location, where $p_{\text{obsNN}}(t)$ is the exceedance probability at the nearest neighbor streamgage on day t;

and

$$Q_{\text{predOKDAR}}(t) = A_{\text{pred}} \exp\left(\frac{1}{n}\sum_{i=1}^{n} w_i \log\left(Q_{\text{obs\_}i}(t) / A_i\right)\right), \tag{3}$$

where $Q_{\text{predOKDAR}}(t)$ is the OKDAR-predicted streamflow on day t, A_{pred} is the drainage area at the prediction location, w_i is the ordinary kriging weight of the ith neighboring streamgage, $Q_{\text{obs\_}i}(t)$ is the discharge at the ith neighboring streamgage on day t, and A_i is the drainage area of the ith neighboring streamgage. In the ordinary kriging, following Farmer (2016), a single, pooled spherical variogram was fit for each study region.

Daily streamflow estimation was conducted by study region (hydrologic unit code level-2 regions as defined by Falcone (2011)) and began by building statistical models using GAGES-II reference streamgages from mostly undisturbed basins as index streamgages (Russell et al., 2020). Using location information and basin characteristics provided by Falcone (2011), estimates were then made using those statistical models at GAGES-II non-reference streamgages by treating those locations as if they were ungaged. The sets of streamflow estimates from the three models represent expected natural streamflow conditions with minimal disturbance by human activities; in other words, with minimal effects of reservoir storage, diversion, irrigation, channelization, land development, or other anthropogenic activities (Falcone et al., 2010).

The statistical-transfer models that use nearest-neighbor (NN) use a single index gage, selected from among the reference streamgages in the region, to transfer information to the "ungaged" locations in the current study. The distances between gaged and ungaged locations are defined by the geographic distance (computed using Vincenty's, 1975, formula for distance on the ellipsoid) between their locations. Within each study region, the NN approach uses the index streamgage closest to the ungaged location. Instead of using a single index gage, OKDAR uses pooled OKDAR to weight all the reference streamgages for use in estimating streamflow at the ungaged location (Equation 3). Streamgages with zero streamflows on a given day were ignored when computing the logarithms of discharge per unit area in the application of OKDAR in Russell et al. (2020) and thus in this study.

Using the drainage-area-ratio method with NN (equation 1) assumes that the streamflow per unit area is the same between the index streamgage and the ungaged location. Flows are then estimated by using the ratio of the drainage areas of the two sites. To apply the QPPQ method (Equation 2), daily streamflow at the index streamgage is converted to a time-series of daily exceedance probabilities. The QPPQ method assumes that the exceedance probability of streamflow on a given day is the same between the index streamgage and the ungaged location. To estimate streamflows and their associated exceedance probabilities at ungaged locations, Russell et al. (2020) developed regional regression equations for 27 flow quantiles with exceedance probabilities ranging from 0.02 to 99.98 percent in each study region using GAGES-II basin characteristics describing reference conditions only, that is, no basin characteristics describing reservoir storage or human effects on land cover were considered. Continuous flow-duration curves were constructed at each ungaged location using these regression equations by using interpolation between the 27 flow quantiles and extrapolation using the two nearest non-equal flow quantiles. The interpolation and extrapolation were done linearly between the logarithms of the streamflow and Gaussian quantiles. Estimates of daily streamflow at the ungaged location were then constructed using the index gage's time-series of daily exceedance probabilities in conjunction with the ungaged location's continuous flow-duration curve.

2.5 | Model accuracy for unregulated basins

The accuracy of the models used in the current study can be seen in the results of Hodgkins et al. (2020) for unregulated basins in the CONUS. Hodgkins et al. (2020) compared modeled to observed streamflows for statistical transfer models and NHM-PRMS for a range of flows from low to high, using volumetric efficiency (VE) as a measure of overall model fit (Criss & Winston, 2008). VE is defined as

$$\mathrm{VE} = 1 - \mathrm{mean}\left(\left| Q_{\mathrm{model}}(t) - Q_{\mathrm{obs}}(t)\right|\right) / \mathrm{mean}\left(Q_{\mathrm{obs}}(t)\right), \tag{4}$$

where $Q_{\mathrm{model}}(t)$ is the modeled flow at time t and $Q_{\mathrm{obs}}(t)$ is the observed flow at time t. VE scales the mean absolute error of the modeled flow by the mean flow and subtracts that ratio from 1. Therefore, a VE of 1.0 indicates perfect agreement between modeled and observed values, and a VE of 0.5 indicates that the magnitude of the average error for a time series of flows is 50% of the average observed value. By comparison, the more common NSE coefficient (Nash & Sutcliffe, 1970) scales the mean-squared error of the modeled and observed flows by the variance of the observed flows. Criss and Winston (2008) selected the mean absolute error measure so that large values were not emphasized as much as when using the mean-squared error, and they argued that normalizing by the mean of the observations is more intuitively understandable than the variance. Following that line of argument, we report the VE results from Hodgkins et al. (2020) as evidence of the models' performance for unregulated basins. For the statistical transfer models in Hodgkins et al. (2020), for annual maximum flows, about 60% of the 502 basins had $\mathrm{VE} > 0.5$; that is, 60% of basins had an average error magnitude of 50% or less of average observed values. For mean flows, about 80% of basins had VE greater than 0.5, while for annual 7-day low flows, about 45% to 65% of basins had VEs that were greater than this level. NHM-PRMS VE were >0.5 for about 50%, 70%, and 30% of basins, for these same high, mean, and low flows, respectively.

2.6 | Period of record streamflow statistics

Daily streamflows from October 1, 1983 through September 30, 2016 were used to compute nine period-of-record (POR) statistics for observed flows and modeled flows. Errors between modeled and observed POR statistics were then computed for all gages. Names, definitions, units, and other information for the POR statistics are given in Table 1, and the values of the statistics are available in Over et al. (2021). We show errors for 10 additional flow statistics in Supporting Information (Figures S1–S4).

The POR statistics in this article quantify a range of long-term flows, from very low to very high. This includes the 10% quantile of annual 7-day low flows (variable name: 7Q10), 5% quantile (Q0.05), mean flow (Qmean), 95% quantile (Q0.95), the 50% quantile of annual maxima (Qmax0.5), the coefficient of variation of annual mean streamflow (annCV), the lag-1 autocorrelation of daily streamflow (AR1), and the amplitude and phase of the annual streamflow cycle (seasAmp and SeasPhs). The statistics were computed using R version 3.6.2 (R Core Team, 2019); details for each POR statistic are given below. Only complete climate years (beginning April 1 and ending March 31; a climate year is named for the year in which it ends) were used for computation of the low-flow statistics; for all the other statistics, only complete water years were used.

The selected POR statistics characterize, at least approximately, the signatures of runoff variability as introduced by Sivapalan (2005), Wagener et al. (2007), and Wagener et al. (2013), which are: mean annual discharge and its variability, seasonal runoff, the flow-duration curve, low flows, and floods. In addition, three of the POR statistics, that is, the mean, AR1, seasAmp, and SeasPhs, are among the seven “fundamental daily streamflow statistics” useful in hydro-ecological stream classification (Archfield et al., 2014).

The mean streamflow is perhaps the most important streamflow statistic because of its connection to the water balance of a basin. The coefficient of variation of water-year average streamflow (annCV) characterizes the interannual variability of the streamflow, and with the mean, characterizes the overall variability of the hydrologic regime (e.g., McMahon et al., 2013); together these are used for practical purposes such as sizing water-supply systems (e.g., McMahon et al., 2007; Vogel et al., 1999).

The low-flow frequency statistic 7Q10 is commonly used for regulation in the United States in the context of water quality criteria, such as in setting total maximum daily loads (Ames, 2006; Kroll et al., 2004). 7Q10 was computed as the $p = 0.1$ (10%) quantile, obtained by fitting the log-Pearson type-3 (LP3) distribution to the non-zero climate-year-minimum 7-day-average flows for each year, with the effect of zero values

TABLE 1 Definitions of period-of-record (POR; October 1, 1983 through September 30, 2016) streamflow statistics measured at 1082 basins across the conterminous USA that have a wide range of reservoir storage. We compared modeled and observed streamflows with increasing reservoir storage for nine streamflow statistics (in bold). The observed flows reflect the effects of varying levels of storage while the modeled flows do not.

Flow statistic	Definition	Transformation	Units	Censoring level	Rounding	Name of statistic in data releases
7Q10	7-day climate year-minimum streamflow, 10% log-Pearson type 3-fitted quantile	log10	ft^3/s	0.005	0.01	l710.LP3
7Q2	7-day climate year-minimum streamflow, 50% log-Pearson type 3-fitted quantile	log10	ft^3/s	0.005	0.01	l750.LP3
Q0.01	Daily streamflow, 1% quantile	log10	ft^3/s	0.005	0.01	percs0.01
Q0.05	Daily streamflow, 5% quantile	log10	ft^3/s	0.005	0.01	percs0.05
Q0.1	Daily streamflow, 10% quantile	log10	ft^3/s	0.005	0.01	percs0.1
Q0.25	Daily streamflow, 25% quantile	log10	ft^3/s	0.005	0.01	percs0.25
Q0.5	Daily streamflow, 50% quantile (median)	log10	ft^3/s	0.005	0.01	dailymedian
Qmean	Mean daily streamflow	log10	ft^3/s	1.37E-06	1.00E-06	dailymean
Q0.75	Daily streamflow, 75% quantile	log10	ft^3/s	0.005	0.01	percs0.75
Q0.9	Daily streamflow, 90% quantile	log10	ft^3/s	0.005	0.01	percs0.9
Q0.95	Daily streamflow, 95% quantile	log10	ft^3/s	0.005	0.01	percs0.95
Q0.99	Daily streamflow, 99% quantile	log10	ft^3/s	0.005	0.01	percs0.99
Qmax0.5	Water-year maximum daily streamflow, 50% quantile (median)	log10	ft^3/s	0.005	0.01	ann1daymaxMedian
Qmax0.9	Water-year maximum daily streamflow, 90% quantile	log10	ft^3/s	0.005	0.01	annmax90
annCV	Coefficient of variation (standard deviation/mean) of water-year mean streamflow	log10	Dimensionless	None	None	annualCV
LCV	Coefficient of L-variation (L-scale/mean) of daily streamflow	None	Dimensionless	None	None	LCV
AR1	lag-1 autocorrelation of log10-transformed deseasonalized daily streamflow	None	Dimensionless	None	None	logAR1
seasAmp	Amplitude of annual cycle of normalized log10-transformed daily streamflow	None	Dimensionless	None	None	log.seasAmp
seasPhs	Phase of annual cycle of normalized log10-transformed daily streamflow	None	Days	None	1	log.seasPhs

being addressed using the conditional-probability adjustment (e.g., Kiang et al., 2018, p. 3). When the percentage of years of 7-day low flows that equal 0 was greater than 10%, then 7Q10 was set to 0. The LP3 distribution was fit with the function *qlpearsonIII* from the R package 'smwrBase' (Lorenz, 2015), which takes the mean, standard deviation, and skewness of the log-transformed data as input. Here, the sample mean, standard deviation, and skewness of the logarithms of the positive 7-day-average flows, computed as per equations 5–7 of England et al. (2019), were used as the input to *qlpearsonIII*.

Daily flow-duration curves characterize the distribution of daily streamflow, which is used directly for a range of management purposes (e.g., Castellarin et al., 2013; Vogel & Fennessey, 1995) and for distributional information in the flow-duration curve transfer method (also called QPPQ) of daily streamflow estimation (Archfield et al., 2010; Farmer et al., 2014; Fennessey, 1994; Mohamoud, 2008). The flow-duration statistics were computed using the *quantile* function from the R 'stats' package with the default quantile type of 7. We report results on two flow-duration statistics representing typical low flows (Q0.05) and high flows (Q0.95).

The 50th percentile of water-year daily maxima (Qmax0.5) characterizes the typical maximum annual flow of the basin. Although flood frequency analyses are typically performed with instantaneous maxima, when only daily data are available, the annual daily maxima are sometimes used as a stand-in (e.g., Perdios & Langousis, 2020); this statistic was used by FitzHugh and Vogel (2011) to characterize flood flows in their analysis of the hydrologic effects of dams. Furthermore, the instantaneous maximum can be estimated from the daily maximum (e.g., Chen et al., 2017). This high-flow statistic was computed as the 50% quantile of the series of annual one-day maxima, using the *quantile* function from the R 'stats' package with default quantile type of 7.

AR1 is the autoregressive lag-1 correlation coefficient of the log10-transformed, then deseasonalized, and then normalized daily streamflow time series. The deseasonalization was performed by subtraction of the seasonal monthly mean and the normalization by subtracting the mean and dividing by the standard deviation. The streamflow values were log10-transformed to adjust for the variation in autocorrelation with flow magnitude, which is smaller, on average, for high flows compared to low flows, because of the differences in rates of variation with magnitude. Like any linear correlation coefficient, AR1 values necessarily lie in the range of −1 to 1. The correlation value was computed using the *ar* function of the R 'stats' package.

seasAmp and seasPhs characterize the seasonal cycle of streamflow in a parametrically simpler way than the more common method of using monthly values (e.g., Weingartner et al., 2013). seasAmp and seasPhs are the amplitude (dimensionless) and phase (the day of the calendar year when the fitted harmonic is at its maximum), respectively, of a single harmonic with a period of 1 year fitted to the log10-transformed then normalized daily streamflow time series, where the normalization was performed by subtracting the mean and dividing by the standard deviation. Because of normalization, seasAmp values range from 0.04 to 1.21 for the observed streamflow among the 1257 basins in the Over et al. (2021) dataset from which the errors analyzed here were computed. The harmonic was fitted with the method outlined by Wilks (2006, section 8.4.3).

After the observed and modeled statistics were computed, model errors were computed. Many of the streamflow statistics used in this study have distributions that are highly skewed to the right, and errors in predicting them typically are proportional to the magnitude of the observed value. It is natural to characterize such errors in terms of ratios, in which case, for example, the magnitude of the error of a prediction that is twice as large as the observation is the same as one that is half the magnitude. A convenient side effect of this approach is that such errors are dimensionless. The errors were computed using ratios, in particular the $\log_{10}$-transformation of the ratio of the predicted value to the observed value, for all the POR statistics in this study except AR1, seasAmp, and seasPhs (Table 1). For these three statistics, which are not highly skewed and whose errors are not proportional to the magnitude of the observed values, the errors are untransformed differences, predicted minus observed. The differences in seasPhs were adjusted to account for the circular nature of this statistic, so that, for example, a predicted phase of 365 compared to an observed phase of 1 is assigned an error of −1.25 days (using a 365.25-day year) rather than an error of 364 days.

For the model errors computed as $\log_{10}$-transformed ratios of predicted to observed values, the resulting errors are dimensionless powers of 10. For example, if the error is 1, then the predicted value is 10 times the observed, whereas if the error is −1, the predicted value is one tenth of the observed. It is important to note that the use of such ratio-based errors avoids flow-magnitude effects on error magnitudes, for example, a 10-percent error relative to an observed statistic of 10 ft^3/s has the same error value ($\log_{10}(1.1)=0.0414$) as a 10 percent error relative to an observed statistic of 1000 ft^3/s. For the three streamflow statistics where errors were computed as differences, the units of predicted and observed values are retained; for these three statistics, only seasPhs is not unitless; it has units of days (Table 1).

All values of the statistics having streamflow (ft^3/s) units (Table 1) were rounded for consistency with the treatment of the observed and modeled time series data before computing the errors. The rounding was to two decimal places in their original ft^3/s units, except Qmean, which because it has higher precision being the average of many daily values, was rounded to six decimal places in ft^3/s units. In addition, seasPhs was rounded to the nearest whole day.

All the statistics having streamflow units have a censoring level of 0.005 ft^3/s, except for Qmean, which has a censoring level of 1.37 e-6 $ft^3/s = 0.005/(10 \times 365.25)$, because it is an average of at least 10 years of daily discharges with censoring levels of 0.005 ft^3/s. These censoring levels were used as the minimum value when log-transforming as part of the error computation, that is, all values less than the censoring level were set to the censoring level. If both the predicted and observed values were below the censoring levels, the error (computed as the log of the ratio of the predicted and observed values) is zero.

2.7 | Relations between modeled and observed values

The strength of correlation between modeled and observed streamflow statistics for each model, across all streamgages, was computed using Kendall's τ. Kendall's τ measures the strength of the monotonic relation between two variables and is a rank-based method (Helsel et al., 2020; Kendall, 1938). Linear (Pearson) correlations of $r=0.9$ typically correspond to τ values of about 0.7 (Helsel et al., 2020). The magnitude and significance for each streamflow statistic was computed using *cor.test* (x, y, method = "Kendall") in R (R Core Team, 2019). The significance of the Kendall τ correlations refers to the likelihood of the validity of a null hypothesis that the correlation is zero, a quantity that is called here the "*p* value."

The model errors as a function of reservoir storage were characterized with three conditional quantiles: the median and the first and third quartiles (that is, the 50th percentile and the 25th and 75th percentiles). The conditional median gives the central tendency of the errors as a function of the amount of storage, while the interquartile range indicates the spread of the errors around the median. These quantiles were computed using a nonparametric quantile regression method, additive quantile regression smoothing (Koenker et al., 1994), which was implemented using the function *rqss* from the R 'quantreg' package (Koenker, 2021). The quantiles were fit using storage values between 0.1 and 10,000 days (which utilizes all but 13 of the 1027 non-zero storage values) but are reported only for storage values between 1 and 1000 days. This limitation removes end segments that were determined to be unreliable because of their substantial sampling variability. The values of the predicted and observed streamflows for each basin in the study are assumed to be deterministic, in the sense that their uncertainties are not explicitly addressed.

To test for change points in relationships between the absolute values of model error versus reservoir storage, we used the Pettitt (1979) test as implemented in the function *pettitt.test* from the R package 'trend' (Pohlert, 2020). This test is a non-parametric method that tests for a shift in the central tendency of a series. It is usually applied to time series, and in that context was determined by Ryberg et al. (2020) to be the preferred method among several parametric and non-parametric alternatives to find change points in annual peak streamflows. Although the Pettitt test is usually applied to time series, we found its results more stable than an alternative change-point analysis that seeks break points in bent-line models in general x-y relationships (Muggeo, 2003).

2.8 | Effects of aridity on reservoir storage and model error

Both reservoir storage and model error magnitudes may increase with increasing aridity, the former because of the need for such reservoir storage and the latter because arid basins can be more difficult to model accurately. This could result in observed correlations in model error with reservoir storage in the current study being caused, at least in part, by this dual correlation rather than the direct effect of the neglect of the reservoir storage in the models considered here.

Evaluation of hydrologic models using reference (minimally altered) basins across the CONUS (e.g., Hodgkins et al., 2020; Salinas et al., 2013) have found that model goodness-of-fit statistics are poorer in the more arid western US. An aridity index, defined as the ratio of basin average precipitation to basin average potential evapotranspiration (UNEP, 1992), was computed for our study basins using values of these quantities provided by Falcone (2011). The Kendall's τ correlation for the relation between normalized reservoir storage (in days) and aridity was 0.096 ($p = 2 \times 10^{-6}$).

To investigate the effect of basin aridity on the relations between model errors and basin reservoir storage, Kendall semi-partial correlations were computed using the function *spcor* from the R package 'ppcor' (Kim, 2015a, 2015b). Correlations were computed between the absolute values of model error and reservoir storage, with the effect of aridity on reservoir storage removed, and compared to correlations that do not consider basin aridity.

3 | RESULTS

3.1 | Correlation between model error and basin reservoir storage

Errors in nine modeled flow statistics (Table 2) from three statistical-transfer models (NNDAR, NNQPPQ, OKDAR) and one process-based model (NHM-PRMS) were found to be significantly correlated with the amount of basin reservoir storage for many flow statistics. The significance and strength of correlations between model errors and basin storage varied, depending on the model and the statistic.

Kendall's τ correlations between errors in mean- and high-flow statistics (Qmean, Q0.95, Qmax0.5) and basin storage were positive for all models and significant ($p < 0.001$) for all but Qmean and Q0.95 for the NNQPPQ model (Table 2). The positive correlations indicate increasing model overprediction and/or decreasing model underprediction of the mean or high-flow magnitude with increasing storage in basins. For models with decreasing underprediction, the model underpredicts a flow statistic for basins with low storage amounts, and this underprediction becomes smaller with increasing basin storage (e.g., Figure 6). Magnitudes of the Kendall's τ correlations are greater than 0.3 for the NNDAR and OKDAR models for the high-flow statistic Qmax0.5 (the median POR annual water-year maximum daily streamflow; Table 2). The NHM-PRMS significant correlations for Qmean and Q0.95 in Table 2 are low, with Kendall's τ correlations <0.10. NNQPPQ has substantially lower correlations for these mean- and high-flow statistics than the other three models.

Streamflows for the low-flow statistics 7Q10 and Q0.05 have negative correlations for the NNDAR and OKDAR models (Table 2), indicating increasing underprediction and/or decreasing overprediction of low-flow magnitude with increasing storage in basins. However, these

TABLE 2 Kendall's τ correlations between model errors and reservoir storage amount for nine period-of-record (POR) streamflow statistics for one process-based model and three statistical transfer models. Units of storage are days of the long-term mean daily flow. The observed flows reflect the effects of varying levels of storage while the modeled flows do not. Correlations in bold are significant at $p<0.001$. Blue shades represent positive correlations and brown shades represent negative correlations; darker shades are stronger correlations. (Streamflow statistic definitions are in Table 1. Process-based model: NHM-PRMS, National Hydrologic Model Infrastructure application of the Precipitation-Runoff Modeling System. Statistical-transfer models: NNDAR, nearest-neighbor drainage area ratio; NNQPPQ, nearest-neighbor nonlinear spatial interpolation using flow-duration curves; OKDAR, ordinary kriging of the logarithms of discharge per unit area.)

Model	7Q10	Q0.05	Qmean	Q0.95	Qmax0.5	annCV	AR1	seasAmp	seasPhs
NNDAR	**−0.085**	**−0.080**	**0.207**	**0.172**	**0.336**	−0.037	−0.032	**0.221**	**−0.153**
NNQPPQ	0.008	0.025	0.054	0.044	**0.148**	**−0.137**	**−0.078**	**0.199**	**−0.119**
OKDAR	−0.043	−0.019	**0.150**	**0.139**	**0.328**	−0.030	**−0.132**	**0.189**	**−0.152**
NHM-PRMS	0.022	0.021	**0.083**	**0.096**	**0.245**	**−0.088**	**0.112**	**0.156**	0.007

correlations are not significant ($p<0.001$) except for NNDAR. The NNQPPQ and NHM-PRMS models, on the other hand, have low positive correlations that are not significant.

The coefficient of variation of annual water-year mean flows (annCV) statistic indicates the variability in interannual flows. The correlations of its modeled to observed errors versus reservoir storage are negative for all models (Table 2), indicating increasing underprediction and/or decreasing overprediction of annCV magnitude with increasing storage in basins. However, these correlations are significant only for the NNQPPQ and NHM-PRMS models.

The lag-1 autocorrelation (AR1) statistic, which measures day-to-day similarity of streamflows, is significantly correlated with storage for all models except NNDAR. For the NNQPPQ and OKDAR models, these errors are negative (Table 2), indicating increasing underprediction with increasing storage, whereas for the NHM-PRMS model, errors are positive. Kendall's τ correlations were low: −0.08, −0.13, and 0.11, respectively.

The seasAmp and seasPhs statistics characterize the amplitude and phase, respectively, of the seasonal streamflow cycle. All the models have significant positive correlations ($p<0.001$) for seasAmp errors, indicating increasing overprediction and/or decreasing underprediction of seasonal amplitude magnitude with increasing storage in basins (Table 2). Kendall's τ values ranged from 0.16 to 0.22. All models but NHM-PRMS have significant negative correlations with model errors for seasPhs, indicating the models are more likely to locate the maximum of the seasonal cycle earlier in the year relative to the observed flows as storage increases. Kendall's τ values range from −0.12 to −0.15 for the statistical transfer models.

3.2 | Correlation between model error and basin reservoir storage for absolute errors

Although the correlations between model errors and amount of reservoir storage are sometimes positive and sometimes negative, the corresponding correlations between absolute errors and the amount of storage are almost universally positive (being insignificantly negative only once) (Table 3). The absolute values of errors were computed with the log-10 logarithms of the ratio of the predicted to the observed flows for the low- and high-flow statistics and the annual CV. Positive correlations between absolute errors and the amount of storage indicate decreasing model accuracy with increasing amounts of reservoir storage. The errors for basins with little or no reservoir storage represent the variability of model errors for relatively natural basins (Figure 2; Figures S1–S4). Note that basins with substantial amounts of urban or irrigated agricultural land have been screened from our study. Changes in the absolute values of median model errors and model-error variability can

TABLE 3 Kendall's τ correlations between the absolute values of model errors and reservoir storage amount for nine POR streamflow statistics for one process-based model and three statistical transfer models. Units of storage are days of storage of the long-term mean daily flow. Correlations in bold are significant at $p<0.001$. Darker green shades represent stronger absolute correlations while lighter shades represent weaker correlations. (Streamflow statistic definitions are in Table 1. Model acronyms are defined in the Table 2 headnote.)

Model	7Q10	Q0.05	Qmean	Q0.95	Qmax0.5	annCV	AR1	seasAmp	seasPhs
NNDAR	**0.110**	**0.149**	**0.184**	**0.199**	**0.312**	**0.109**	0.066	**0.187**	**0.264**
NNQPPQ	0.026	0.044	**0.153**	**0.171**	**0.146**	**0.071**	**0.102**	**0.155**	**0.225**
OKDAR	**0.181**	**0.161**	**0.093**	**0.113**	**0.153**	**0.112**	**0.119**	**0.255**	**0.272**
NHM-PRMS	**0.073**	**0.085**	**0.070**	0.064	**0.074**	0.061	−0.014	**0.088**	**0.158**

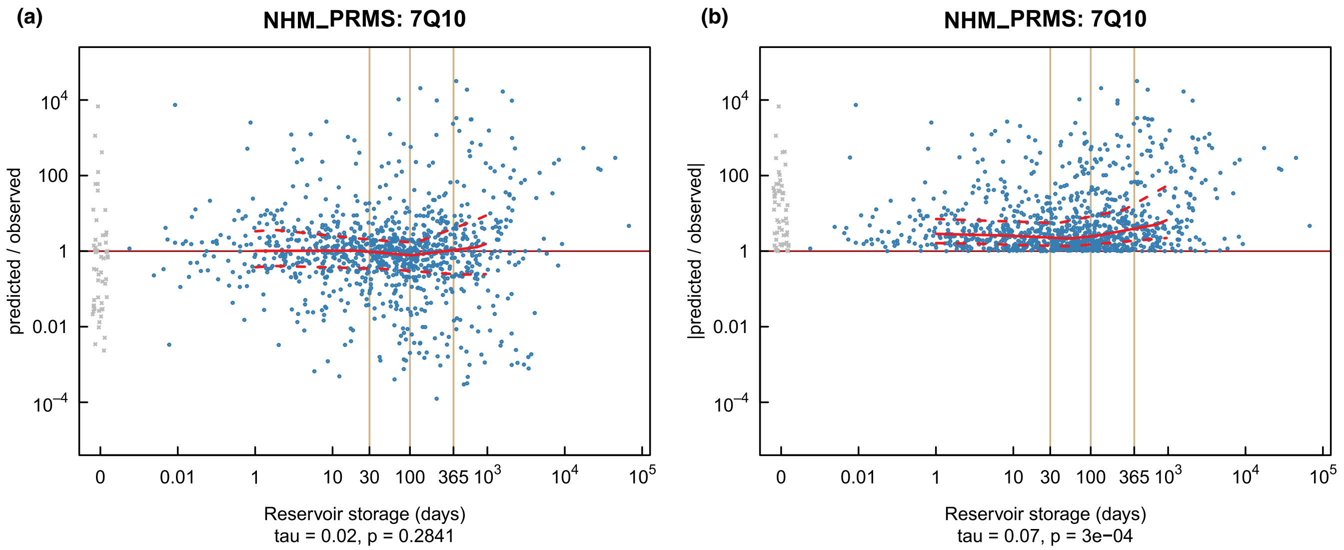

FIGURE 2 Relation between (a) NHM-PRMS model errors and basin reservoir storage for 10%-quantile, 7-day low flows (7Q10) and (b) the absolute values of NHM-PRMS 7Q10 errors and basin storage. The solid red lines indicate the 50th percentile of model error with increasing storage, the lower dashed red lines indicate the 25th percentile, and the upper dashed line the 75th percentile. All lines were computed with non-parametric quantile regression.

be seen as storage increases, particularly for larger amounts of storage. The magnitude and nature of these changes with increasing storage are discussed below.

Positive correlations between the absolute values of model errors and the amount of storage are significant ($p < 0.001$) for low-, mean-, and high-flow statistics, except for low flows for the NNQPPQ model and Q0.95 for NHM-PRMS (Table 3). While low-flow statistics were significant for signed errors only for NNDAR (Table 2), the absolute values of model errors are significantly correlated with basin storage for NNDAR, OKDAR, and NHM-PRMS (Table 3). As an example, the relation between signed NHM-PRMS 7Q10 errors and basin storage is insignificant while the relation with absolute errors shows significant increases (Figure 2). Correlations between model errors for mean- and high-flow statistics are generally significant for both signed and absolute errors.

The coefficient of variation of annual water-year mean flows statistic (annCV) has significant correlations ($p < 0.001$) between the absolute values of model errors for the three statistical transfer models (NNDAR, NNQPPQ, OKDAR) and basin reservoir storage amount; however, NHM-PRMS errors are not significant. For the autocorrelation statistic (AR1), NNQPPQ and OKDAR models have significant correlations between absolute model errors and reservoir storage. Both seasonal amplitude (seasAmp) and phase (seasPhs) absolute model errors have significant correlations with basin storage for all models.

To see whether basin aridity affects the relation between model error and reservoir storage, we used semi-partial correlation to remove the relation between the absolute values of model error and aridity. This resulted in very small reductions in the Kendall's τ between model errors and reservoir storage; the median reduction for the nine study variables for the four study models was 0.01 and the maximum reduction was 0.02.

3.3 | Change points for absolute model error with increasing basin reservoir storage

The relation between the absolute values of model error and basin reservoir storage is nonlinear for many streamflow statistics. To test whether study models had a change point (step trend) in absolute model error with increasing storage for streamflow statistics, we used the Pettitt test (Pettitt, 1979; Ryberg et al., 2020). Model errors for most streamflow statistics for most models had significant ($p < 0.001$) change points with increasing storage (Table 4). For example, the absolute NNQPPQ model errors for seasonal amplitude (seasAmp) have a change point to larger errors at 125 days of reservoir storage (Table 4; Figure 3).

The typical value of the change points, in days, was smaller for low flows than for mean and high flows. For the four models, the median significant change point for both the 10% quantile of annual 7-day low flows (7Q10) and the 5% quantile of daily flows (Q0.05) was 48 days of reservoir storage while the median value for Qmean, Q0.95, and Qmax0.5 was 176, 173, and 147 days, respectively. The median values for the seasonal amplitude and phase (seasPhs) of the annual streamflow cycle of 126 and 104 days fell between these values. The annual coefficient of variation (annCV) had the highest median change point for the models of 214 days. In contrast, the lag-1 autocorrelation of daily streamflow (AR1) had the lowest median significant change point of 30 days, though only two models had significant change points. The change points had

TABLE 4 Pettitt change points (step trends), in days, for the absolute values of model errors and reservoir storage amounts for nine POR streamflow statistics for one process-based model and three statistical transfer models. Units of storage are days of storage of the long-term mean daily flow. Change points in orange shading and bold text are significant at $p < 0.001$. (Streamflow statistic definitions are in Table 1. Model acronyms are defined in the Table 2 headnote.)

Model	7Q10	Q0.05	Qmean	Q0.95	Qmax0.5	annCV	AR1	seasAmp	seasPhs
NNDAR	**48**	**48**	**163**	**125**	**119**	**301**	32	**220**	**83**
NNQPPQ	2	301	**36**	**36**	**36**	**189**	**35**	**125**	**87**
OKDAR	**48**	**48**	**301**	**221**	**174**	**151**	**25**	**61**	**122**
NHM-PRMS	**135**	**135**	**189**	**227**	**176**	**238**	353	**127**	**199**

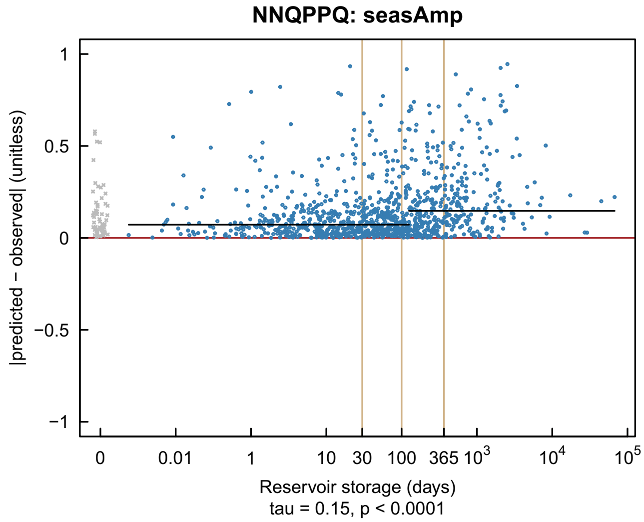

FIGURE 3 Relation between NNQPPQ model errors and basin reservoir storage for seasonal amplitude (seasAmp). The horizontal lines represent the median values before and after a significant change point (step trend).

some substantial variation by model. For the low-flow statistics, the NHM-PRMS model had a larger change point than the other models with significant change points. For the mean and high flow statistics, the NNQPPQ model had much smaller change points than the other models.

3.4 | Model error magnitude and variability with increasing basin reservoir storage

To quantify changes in model error magnitude and variability for the four study models with increasing amounts of normalized reservoir storage, we used non-parametric quantile regression. In the following paragraphs, we describe changes in median model errors with increasing storage at basins, and we use the difference between the 75th and 25th quantile errors to quantify changes in the variability of model errors with increasing storage. We discuss changes in model errors for nine streamflow statistics. The supplemental material contains plots of model errors versus storage amount for a more complete suite of 19 variables for both signed and absolute model errors (Figures S1–S4, one file per model). Reservoir storage amounts in study basins are quantified by days of storage, that is, the days of long-term mean daily flow stored by reservoirs in each basin.

Correlations between errors in mean- and high-flow statistics and basin storage amount were positive for all models and significant ($p < 0.001$) for most (Table 2). The correlations were strongest for Qmax0.5 (the median annual maximum flow). The magnitude of model errors, based on non-parametric quantile regression median values, increased substantially for Qmax0.5 for all four models with increasing storage (Table 5). For example, the median model error (predicted divided by observed values) for NNDAR increased from a median overprediction of 1.31 at 10 days of storage to a median overprediction of 2.48 at 365 days of storage (Table 5; Figure 4). The median model error for NHM-PRMS increased from an underprediction (0.73) to an overprediction (1.22) for these same storage amounts. In agreement with their signed correlation results (Table 2), the median errors of Qmean and Q0.95 for the NNDAR, OKDAR, and NHM-PRMS models generally increased with increasing reservoir storage (Table 5).

The high-flow statistic Qmax0.5 had a significant positive relation between the absolute values of model errors and storage amounts for all four models (Table 3). This is consistent with increases in the variability of Qmax0.5 error (predicted divided by observed) based on

TABLE 5 Magnitude of median model errors with increasing reservoir storage amounts for nine POR streamflow statistics for one process-based model and three statistical transfer models. Magnitudes are based on nonparametric quantile regression. Units of storage are days of storage of the long-term mean daily flow. The observed flows reflect the effects of varying levels of storage while the modeled flows do not. Blue shades represent larger predicted than observed streamflow statistics and brown shades represent the opposite; darker shades are larger differences. [Streamflow statistic definitions are in Table 1. Model acronyms are defined in the Table 2 headnote. 7Q10, Q0.05, Qmean, Q0.95, Qmax0.5, and annCV errors quantified as predicted divided by observed; AR1, seasAmp, and seasPhs errors quantified as predicted minus observed.]

Regulation (days)	7Q10	Q0.05	Qmean	Q0.95	Qmax0.5	annCV	AR1	seasAmp	seasPhs
NNDAR									
1	0.644	0.861	0.998	1.002	1.172	1.028	−0.0213	−0.0330	−2.30
10	0.640	0.695	0.992	1.016	1.314	1.065	−0.0298	−0.0453	−2.61
30	0.387	0.560	1.024	1.067	1.477	1.077	−0.0346	−0.0376	−2.63
100	0.281	0.440	1.008	1.098	1.755	1.079	−0.0374	0.0210	−1.55
365	0.340	0.430	1.170	1.127	2.481	1.019	−0.0310	0.0918	−12.87
1000	0.395	0.493	1.571	1.539	3.514	0.975	−0.0247	0.1319	−25.03
NNQPPQ									
1	0.376	0.515	0.991	1.096	1.248	1.072	−0.0389	−0.0407	−3.46
10	0.378	0.568	1.252	1.347	1.611	1.075	−0.0502	−0.0447	−2.99
30	0.480	0.705	1.411	1.526	1.873	1.036	−0.0561	−0.0331	−2.67
100	0.486	0.717	1.203	1.337	1.798	0.994	−0.0615	0.0092	−2.28
365	0.392	0.507	1.126	1.173	1.846	0.916	−0.0601	0.0723	−13.84
1000	0.425	0.393	1.353	1.333	2.376	0.818	−0.0586	0.1052	−19.26
OKDAR									
1	0.850	0.846	0.819	0.790	0.698	0.964	−0.0055	0.0608	−2.16
10	0.723	0.745	0.792	0.821	0.876	0.978	−0.0177	0.0484	−3.85
30	0.568	0.681	0.853	0.873	1.027	0.997	−0.0227	0.0565	−4.61
100	0.434	0.567	0.878	0.893	1.212	1.009	−0.0273	0.0909	−2.73
365	0.470	0.647	0.863	0.853	1.418	0.977	−0.0315	0.1593	−12.80
1000	1.072	1.074	1.172	1.074	1.749	0.906	−0.0348	0.1992	−21.80
NHM-PRMS									
1	1.013	0.957	0.686	0.581	0.653	1.022	−0.0865	−0.1217	−3.67
10	1.030	0.988	0.670	0.634	0.728	0.988	−0.0625	−0.1075	−4.88
30	0.962	0.961	0.747	0.738	0.827	1.011	−0.0531	−0.1005	−2.50
100	0.783	0.843	0.800	0.788	0.992	1.025	−0.0461	−0.0647	1.20
365	1.069	0.909	0.711	0.747	1.224	0.937	−0.0354	0.0042	−1.55
1000	1.606	0.974	0.829	0.811	1.582	0.852	−0.0254	0.0579	−3.69

non-parametric quantile regression (3rd quartile minus 1st quartile, Table 6). Model errors generally became more variable as storage increased (e.g., Figure 4). The Qmean and Q0.95 error variability also generally increased (Table 6), in agreement with the significant positive correlations in Table 3.

Correlations between model error and basin storage amount were generally insignificant and weak for low-flow statistics (Table 2). The median model error for the low streamflow statistics 7Q10 and Q0.05 varied by model (Table 5). For NNDAR, which had a significant but weak negative correlation between model error and storage amount, the median ratio of predicted to observed 7Q10 decreased from 0.64 at 10 days of storage to 0.34 at 365 days (Figure 5; Table 5).

The positive correlation between the absolute values of low-flow errors and basin reservoir storage was significant for the NNDAR, OKDAR, and NHM-PRMS models (Table 3). Increases in the variability of low-flow errors for these models can be seen in Table 6 and Figure 5. The rate of increase in the error variability in these models is generally nonlinear (Figure 5; Figures S1, S3, and S4).

As suggested by the moderately significant negative correlations between annCV errors and reservoir storage (Table 2), the median errors in annCV generally decrease somewhat with increasing reservoir storage, with lowest values for the largest storage tabulated (1000 days, Table 5). Similarly, in agreement with the positive correlations between absolute errors in annCV and basin storage (Table 3), the error widths generally increase with days of storage and reach their largest values at 1000 days of storage (Table 6).

AMERICAN WATER RESOURCES ASSOCIATION JAWRA JOURNAL OF THE AMERICAN WATER RESOURCES ASSOCIATION

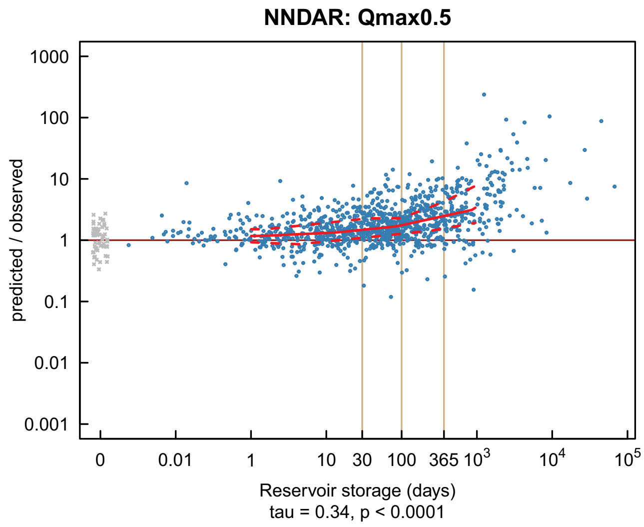

FIGURE 4 Relation between NNDAR model errors and basin reservoir storage for annual maximum flows (Qmax0.5). See Figure 2 caption for an explanation of the red lines.

The seasonal flow statistics seasAmp and seasPhs generally had significant relations with storage amount for the four models in our study (Table 2); seasAmp correlations were positive and significant for all models. The median model errors based on non-parametric quantile regression (Table 5) reflect the correlation results. The NNDAR, NNQPPQ, and NHM-PRMS models go from underpredictions of the seasonal amplitude to overpredictions with increasing reservoir storage in basins, whereas the OKDAR results indicate increasing overpredictions. Changes in seasAmp errors with storage are nonlinear for all four models: the slope of the relation is steeper for larger storage amounts (Table 5; Figure 6; Figures S1–S4).

All three statistical transfer models (NNDAR, NNQPPQ, OKDAR) had significant negative relations between model error and storage amount for seasPhs (Table 2). Median model errors for seasonal phase show substantial negative deviations from zero starting between 100 and 200 days of storage (Table 5; Figure 7; Figures S1–S4). Median errors indicate that seasonal phase is about 13 days early at 365 days of storage for these models.

The correlation between the absolute value of seasAmp and seasPhs errors and basin storage was positive and significant for all models (Table 3). Increases in the variability of model errors for these seasonality statistics are quantified in Table 6 and examples shown in Figures 6 and 7. The variability of model errors for both seasonality statistics increases in a nonlinear way with increasing storage. At 10 days of storage, seasPhs variability (3rd quartile minus 1st quartile) ranges from 7.6 to 22.4 days for the four models; however, at 365 days, seasPhs ranges from 27.3 to 55.4 days.

4 | DISCUSSION

Modeled streamflows in our study, by design, did not include effects of basin reservoir storage. When comparing modeled to observed flows, the observed flows reflect the effects of varying levels of storage while the modeled flows do not. Therefore, if a model increasingly over- or under-predicts a flow statistic with increasing storage (e.g., Figure 4), this is likely related to the omission of flow storage in the models (basins with substantial amounts of impervious area and irrigation were screened).

We analyzed changes in model errors with increasing reservoir storage at the national scale for multiple models and multiple flow statistics using >1000 basins. We normalized reservoir storage across the CONUS using the long-term mean daily flow at each basin. Results from smaller or drier basins will therefore be more comparable to larger or wetter basins. While beyond the scope of this present study, it would be useful to follow upon this national level analysis to determine whether there is substantial variation of model errors across regions. It would also be useful to analyze model errors by the primary use of reservoirs (such as flood control and hydropower) and to consider the specific management of each reservoir.

4.1 | Changes in model errors with increasing basin reservoir storage

The significant positive correlations of model errors and the amount of basin reservoir storage for high-flow statistics (Table 2) indicates increasing model overprediction and/or decreasing model underprediction of high-flow magnitude at basins with increasing storage (e.g.,

TABLE 6 Magnitude of model error variability (75th quantile minus 25th quantile) with increasing reservoir storage amounts for nine POR streamflow statistics for one process-based model and three statistical transfer models. Units of storage are days of storage of the long-term mean daily flow. Magnitudes are based on nonparametric quantile regression. Darker green shades represent larger model error variability while lighter shades represent smaller variability. (Streamflow statistic definitions are in Table 1. Model acronyms are defined in the Table 2 headnote. 7Q10, Q0.05, Qmean, Q0.95, Qmax0.5, and annCV errors quantified as predicted divided by observed; AR1, seasAmp, and seasPhs errors quantified as predicted minus observed.)

Regulation (days)	7Q10	Q0.05	Qmean	Q0.95	Qmax0.5	annCV	AR1	seasAmp	seasPhs
NNDAR									
1	8.17	4.10	1.263	1.300	1.633	1.248	0.056	0.118	8.68
10	9.55	3.89	1.274	1.370	2.011	1.611	0.081	0.130	8.82
30	10.14	4.31	1.262	1.363	2.026	1.873	0.078	0.115	11.24
100	11.44	5.51	1.248	1.436	1.807	1.798	0.067	0.161	17.80
365	63.15	19.57	1.536	1.781	2.714	1.846	0.068	0.244	29.11
1000	105.04	48.30	2.426	3.645	4.067	2.376	0.068	0.357	46.39
NNQPPQ									
1	35.89	21.01	2.164	1.793	2.100	1.462	0.082	0.114	10.92
10	314.73	82.05	2.508	2.513	3.662	1.586	0.102	0.147	11.08
30	143.67	33.56	2.967	3.383	5.327	1.513	0.102	0.161	11.71
100	39.51	16.30	3.165	3.735	5.224	1.438	0.094	0.175	20.01
365	32.46	19.34	3.013	3.234	3.328	1.601	0.089	0.272	34.07
1000	37.49	37.07	4.521	6.029	4.813	1.966	0.084	0.357	51.52
OKDAR									
1	4.32	2.90	1.309	1.285	1.557	1.177	0.052	0.101	8.11
10	3.92	2.66	1.316	1.447	1.829	1.213	0.066	0.111	7.56
30	4.12	2.58	1.338	1.399	1.789	1.212	0.067	0.111	9.53
100	4.32	3.26	1.365	1.405	1.977	1.214	0.061	0.130	15.79
365	12.58	6.33	1.453	1.700	2.488	1.368	0.074	0.249	27.27
1000	20.47	9.42	2.456	3.118	2.984	1.668	0.076	0.383	45.67
NHM-PRMS									
1	8.93	5.40	1.777	2.006	1.901	1.331	0.108	0.223	25.94
10	6.87	4.09	1.788	1.818	1.877	1.335	0.100	0.184	22.45
30	5.78	3.67	1.578	1.702	1.870	1.328	0.110	0.204	24.95
100	5.73	3.86	1.722	1.836	1.938	1.303	0.124	0.244	27.83
365	16.17	8.16	2.075	2.486	2.512	1.435	0.142	0.353	55.36
1000	36.81	17.38	3.477	3.807	3.555	1.600	0.166	0.448	77.88

Figure 4). For models with decreasing underprediction, the model underpredicts a flow statistic at basins with low storage amounts; this underprediction becomes smaller with increasing basin storage (e.g., Figure 6). These high-flow results are to be expected since reservoirs are generally built to store water at times of high flow and thus reduce high flows, for a combination of flood control and water supply reasons (e.g., Carlisle et al., 2019; Eng et al., 2013, 2019; FitzHugh & Vogel, 2011; Graf, 2006; Magilligan & Nislow, 2005; Poff et al., 2007; Smakhtin, 2001; Williams & Wolman, 1984; Yang et al., 2021).

There were large and significant increases in the ratios of predicted to observed high flows with increasing storage for maximum annual flow for all four study models (Tables 2 and 5). This result indicates that both process-based and statistical transfer models substantially overpredict or decreasingly underpredict maximum annual flow magnitude with increasing reservoir storage, and they do so more than for smaller high-flow statistics such as the 95th quantile daily flow. The NNQPPQ statistical transfer model had a lower correlation than the other models between model errors and basin storage high-flow statistics; this may be due to generally larger magnitude errors for these statistics for NNQPPQ than the other models (Figures S1–S4).

The significant positive correlations of model errors and the amount of basin reservoir storage for mean flow (except for NNQPPQ; Table 2) also indicates increasing model overprediction and/or decreasing model underprediction of mean-flow magnitude with increasing storage. Dams are not intended to reduce mean flows but our results indicate that they have this effect. This is likely due to reservoir evaporation and seepage losses (Graf, 1999; Wang & Hejazi, 2011; Zhao & Gao, 2019) because we controlled for diversions and irrigation in our data screening.

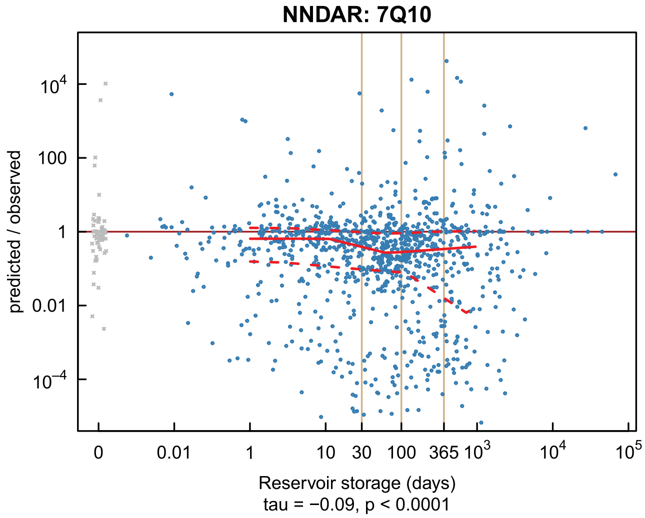

FIGURE 5 Relation between NNDAR model errors and basin reservoir storage for 10%-quantile, 7-day low flows (7Q10). See Figure 2 caption for an explanation of the red lines.

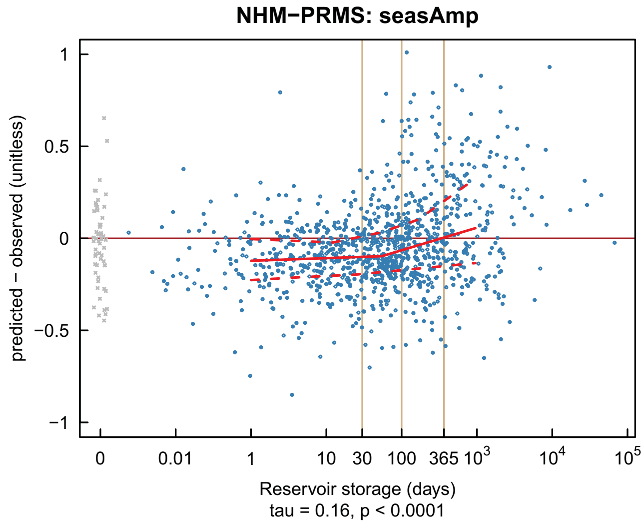

FIGURE 6 Relation between NHM-PRMS model errors and basin reservoir storage for seasonal amplitude (seasAmp). See Figure 2 caption for an explanation of the red lines.

The significant positive correlations and increases in median errors of seasonal amplitude errors with storage for all four models (Tables 2 and 5) are likely explained by the smoothing and lagging effects of reservoir storage. These effects also likely explain the significant negative correlations and decreases in median model errors of seasonal phase with storage for the three statistical-transfer models. The sign of the correlation with seasonal amplitude errors and the direction of the change in error indicates that, for larger amounts of storage, modeled flows (which do not include the effects of storage) have larger seasonal amplitude than observed flows (which include storage). This difference means that observed reservoir-affected flows have either smaller high flows or larger low flows or both. Earlier seasonal maxima for modeled flows relative to observed flows, as indicated by the negative correlations with seasonal phase errors, also makes sense as reservoir storage is expected to result in delayed release of streamflows. The reason for the insignificant correlation of model errors and seasonal phase for NHM-PRMS is not known; a potential explanation is the effect of river (not reservoir) routing, which is present in NHM-PRMS and not explicitly included in the statistical models.

The sign of correlations between model errors and basin reservoir storage for the annual coefficient of variation (annCV) is negative (and significantly so at $p < 0.001$ for the NNQPPQ and NHM-PRMS models; Table 2). In other words, the models increasingly underpredict annCV with increasing storage. This result indicates that the effect of storage on the observed interannual variability of mean annual flows, at least for the basins analyzed in our study, is generally to increase it relative to the model predictions and thus indirectly relative to the unregulated basins to which the models were calibrated. It would be surprising if this effect is the direct result of regulation because the effect of

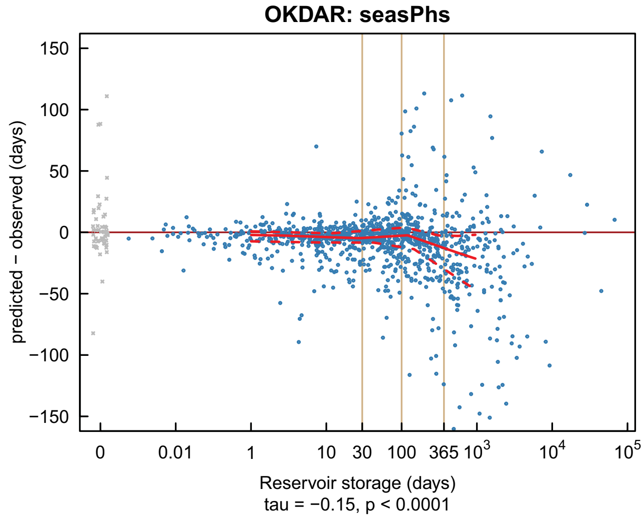

FIGURE 7 Relation between OKDAR model errors and basin reservoir storage for seasonal phase (seasPhs). See Figure 2 caption for an explanation of the red lines.

substantial amounts of storage is expected to smooth annual flows and thus reduce the variability. There may be scenarios where storage has this effect on interannual variability, but it is beyond the scope of this study to investigate them. It seems more likely that this result has arisen from sample effects, that is, that the natural variability at the interannual time scale of the collection of basins used in calibration is less than the natural variability in the regulated basins to which the models are being applied.

4.2 | Changes in model-error variability with increasing basin reservoir storage

Correlations in the absolute values of model errors for low-flow statistics are often larger in magnitude and more significant than those in the signed errors (Tables 2 and 3), indicating that more reservoir storage can cause increases in model errors in both directions even when its effect on errors in a particular direction is weak. We found that basin aridity did not have a substantial influence on the correlation of absolute values of model errors with normalized basin storage; this was true for the nine streamflow statistics for the four study models.

Conclusions based on the differences in results between absolute and signed errors for low flows is supported by the variability of model errors with increasing storage, as represented by the difference between the 75th and 25th quantiles using non-parametric quantile regression (Table 6). The presence of larger correlations for absolute errors than signed errors for low flows is consistent with previous work that showed that streamflow regulation associated with basin reservoir storage can cause these flows to decrease or increase (e.g., Carlisle et al., 2019; Eng et al., 2013, 2019; Smakhtin, 2001; Williams & Wolman, 1984).

A consistent result for multiple variables is the nonlinear response of model error variability to increases in flow storage (Figures S1–S4). The typical values of change points (step trends) to larger absolute values of model errors with increased days of reservoir storage (of the long-term mean daily flow of basins) depends on the variable considered (Table 4). For example, median change points for the study models were substantially smaller for low flows (48 days) than for mean and high flows (147 to 176 days). This makes physical sense, as less reservoir storage is required to affect low streamflows than high streamflows. Similarly, it makes physical sense that the typical values of the change points in annual CV (annCV), which is the ratio of the standard deviation to the mean of annual averages, are also large. That these values are typically larger than those of the mean apparently arises from the effects of the standard deviation of annual values. Likewise, the change point values for the 1-day autocorrelation errors, for the two models whose change points are significant, are the smallest of any statistic considered, which makes physical sense because this statistic focuses on fluctuations in streamflow at a 1-day time scale.

Correlations between seasonal-phase errors and reservoir storage are stronger for absolute errors (Table 3) than for signed errors (Table 2) for all models, as the increase in error distribution width is larger than the change in median error. These results indicate that model errors reflect streamflow being shifted both earlier and later. However, the dominant effect is to delay streamflows, as indicated by the significant negative correlations for all models but NHM-PRMS.

Changes in the absolute values of model errors for the seasonality statistics with increasing reservoir storage are nonlinear (Table 4; Figures S1–S4) as reflected in significant change points in the absolute values of model errors. The median change-point values for the study models toward increasing model error for seasonal amplitude and phase were 126 and 104 days, respectively. In other words, model error increases abruptly for basins with 3 to 4 months or more of reservoir storage of the long-term mean daily flows.

4.3 | Comparison to previous research

The most similar of prior studies to the present one is Ouyang et al. (2021), who quantified overall machine-learning model errors for daily streamflows at over 3500 basins in the CONUS for three groups of basins, using the NSE coefficient. When models were trained on basins with no reservoir storage, applying them to basins with small amounts of storage (greater than zero to about 1 month of storage of the mean annual runoff) resulted in dramatically reduced NSEs; when applying these models to basins with large amounts of storage (greater than about 1 month), median NSE values were even lower. Although Ouyang et al. (2021) characterized model errors for groups of basins using the NSE, which contrasts with our focus on predicted to observed errors for selected streamflow statistics across a continuous range of reservoir storage, their results agree with our computed absolute errors for one process-based deterministic model and three statistical transfer models in a general way. We found that median errors generally increase with increasing reservoir storage for low-flow through high-flow statistics, annual coefficient of variation, and for seasonality statistics.

5 | SUMMARY AND CONCLUSIONS

A large majority of streams are not gaged and require models to estimate streamflows. Evaluation of modeled flows against observed flows is crucial for determining how well they perform. Because reservoir storage of streamflow is common, it is important to quantify model accuracy for basins affected by such alteration. Furthermore, explicit modeling of reservoir storage and release at the scale of the CONUS, which is considered here, requires a substantial amount of information that is not readily available at such a scale. We are unaware of previous studies that have quantified the effects of ignoring reservoir storage in large-scale hydrologic models for a variety of streamflow statistics (predicted vs. observed) across a continuous range of reservoir storage amounts.

We compared modeled to observed streamflows for one deterministic, distributed-parameter, process-based model (the NHM-PRMS) and three statistical-transfer models, for 1082 basins across the CONUS that are affected by variable amounts of reservoir storage but have minimal other anthropogenic influences, isolating the effects of storage. The statistical-transfer models, which estimate flows from nearby gaged basins, are the NNDAR, NNQPPQ, and OKDAR models. We analyzed nine period-of-record streamflow statistics from 1984 to 2016: a range of streamflow magnitudes from very low to very high flows, annual flow variability, daily autocorrelation, and seasonal phase and amplitude. Modeled streamflows in our study, by design, did not include effects of basin reservoir storage. When comparing modeled to observed flows, the observed flows reflect the effects of varying levels of storage while the modeled flows do not. Therefore, if a model increasingly overpredicts a flow statistic with increasing storage, this is because the lack of considering flow storage in the models caused increasing model errors. Basin reservoir storage in this study was normalized and reflects the number of days of storage of the long-term mean daily flow for each basin. A majority of study basins had reservoir storage between 1 and 1000 days.

Kendall's τ correlations between errors in high-flow statistics and basin storage amount were positive for all models and significant ($p < 0.001$) for most. Both process-based and statistical transfer models increasingly overpredict or decreasingly underpredict the maximum annual flow magnitude with increasing reservoir storage, and more so than for smaller high flows such as the 95th quantile daily flow. Increases in median model errors with increasing storage were computed with non-parametric quantile regression for all four study models. As an example of model-error quantification, the median ratio of predicted-to-observed model error for maximum annual flow for the NNDAR model increased from 1.31 at 10 days of storage to 2.48 at 365 days.

There were significant positive correlations between mean-flow model errors and the amount of basin storage for most study models, indicating increasing model overprediction and/or decreasing model underprediction of mean-flow magnitude with increasing storage. Dams are not intended to reduce mean flows but our results indicate that they have this effect. This is likely due to reservoir evaporation and seepage losses since we did not use basins with known substantial irrigation or diversions.

All of the study models have significant positive correlations for seasonal amplitude errors, indicating increasing overprediction or decreasing underprediction of seasonal amplitude magnitude with increasing reservoir storage. Models have greater errors because of more flow smoothing (smaller observed high flows and/or larger low flows) with greater storage. Changes in seasonal amplitude errors with storage are nonlinear for all four models; the slope of the relation is steeper for larger storage amounts. All of the statistical transfer models have significant negative correlations with model errors and seasonal phase, indicating the models are more likely to locate the maximum of the seasonal cycle earlier in the year relative to the observed flows as storage increases. Median errors for all the statistical transfer models indicate that seasonal phase is about 13 days early at 365 days of storage.

Correlations between the absolute values of errors for low-flow statistics and reservoir storage are often larger in magnitude and more significant than those in the signed errors, indicating that additional storage can cause increases in model errors in both directions even when its effect on errors in a particular direction is weak. Positive correlations with absolute errors indicate decreasing model accuracy with increasing amounts of storage.

The rate of increase in model error variability is nonlinear for most streamflow statistics for most study models. The typical value of change points (step trends) toward increases in the absolute values of model errors with increased days of reservoir storage depends on the variable considered. For example, median change points for the study models were substantially smaller for low flows (48 days) than for mean and high flows (147 to 176 days). Median change-point values for seasonal amplitude and phase were 126 and 104 days, respectively.

Since there are so many ungaged streams that need to be modeled in the US and other parts of the world, providing stakeholders and water resource managers with appropriate tools and datasets that can help quantify human effects on streamflow is a critical need in the hydrologic community. This is particularly important for efforts at the regional, national, or continental scale where hundreds or thousands of basins may be modeled. The quantification of the magnitude of median model errors and model error variability with increasing reservoir storage for key streamflow statistics allows model developers to better understand when it is important to include methods of additional complexity for modeling reservoir storage in the models. The quantification of model errors also allows model users to know the level of reservoir storage for which the model predictions become unacceptably biased or uncertain, for specific streamflow statistics of interest.

AUTHOR CONTRIBUTIONS

Glenn A. Hodgkins: Conceptualization; data curation; formal analysis; investigation; methodology; project administration; resources; supervision; visualization; writing – original draft; writing – review and editing. **Thomas M. Over:** Conceptualization; data curation; formal analysis; investigation; methodology; project administration; resources; software; validation; visualization; writing – original draft; writing – review and editing. **Robert W. Dudley:** Conceptualization; data curation; formal analysis; investigation; methodology; project administration; resources; supervision; visualization; writing – original draft; writing – review and editing. **Amy M. Russell:** Conceptualization; data curation; formal analysis; investigation; methodology; project administration; resources; software; validation; writing – original draft; writing – review and editing. **Jacob H. LaFontaine:** Data curation; investigation; methodology; resources; software; validation; writing – original draft; writing – review and editing.

ACKNOWLEDGMENTS

Some of the R code used to compute the period-of-record statistics was originally developed by William Farmer (USGS) based in part on code from Stacey Archfield (USGS). Any use of trade, firm, or product names is for descriptive purposes only and does not imply endorsement by the US Government.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflict of interest.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are openly available in U.S. Geological Survey ScienceBase at https://doi.org/10.5066/P9FS37YQ, https://doi.org/10.5066/P9PGZE0S, https://doi.org/10.5066/P9Z5D4ZT, https://doi.org/10.5066/P9DPSY6G, https://doi.org/10.5066/P9XT4WSP, and https://doi.org/10.5066/P9PA9PKM.

ORCID

Robert W. Dudley https://orcid.org/0000-0002-0934-0568

REFERENCES

Ames, D.P. 2006. "Estimating 7Q10 Confidence Limits from Data: A Bootstrap Approach." *Journal of Water Resources Planning and Management* 132(3): 204–08. https://doi.org/10.1061/(ASCE)0733-9496(2006)132:3(204).

Archfield, S.A., S.L. Brandt, S.P. Garabedian, P.A. Steeves, R.M. Vogel, and P.K. Weiskel. 2010. "The Massachusetts Sustainable-Yield Estimator—A Decision-Support Tool to Assess Water Availability at Ungaged Stream Locations in Massachusetts." U.S. Geological Survey Scientific Investigations Report 2009-5227.

Archfield, S.A., D.M. Carlisle, and D.M. Wolock. 2014. "An Objective and Parsimonious Approach for Classifying Natural Flow Regimes at a Continental Scale." *River Research and Applications* 30(9): 1166–83. https://doi.org/10.1002/rra.2710.

Archfield, S.A., M. Clark, B. Arheimer, L.E. Hay, H. McMillan, J.E. Kiang, J. Seibert, et al. 2015. "Accelerating Advances in Continental Domain Hydrologic Modeling." *Water Resources Research* 51(12): 10078–91. https://doi.org/10.1002/2015WR017498.

Carlisle, D.M., D.M. Wolock, C.P. Konrad, G.J. McCabe, K. Eng, T.E. Grantham, and B. Mahler. 2019. "Flow Modification in the Nation's Streams and Rivers." U.S. Geological Survey Circular 1461. https://pubs.er.usgs.gov/publication/cir1461.

Castellarin, A., G. Botter, D.A. Hughes, S. Liu, T.B.M.J. Ouarda, J. Parajka, D.A. Post, et al. 2013. "Prediction of Flow Duration Curves in Ungauged Basins." In *Runoff Prediction in Ungauged Basins—Synthesis across Processes, Places and Scales*, edited by G. Blöschl, 135–62. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9781139235761.010.

Chawanda, C.J., J. Arnold, W. Thiery, and A. van Griensven. 2020. "Mass Balance Calibration and Reservoir Representations for Large-Scale Hydrological Impact Studies Using SWAT+." *Climatic Change* 163: 1307–27. https://doi.org/10.1007/s10584-020-02924-x.

Chen, B., W. Fang, W.F. Krajewski, F. Liu, and Z. Xu. 2017. "Estimating Instantaneous Peak Flow from Mean Daily Flow." *Hydrology Research* 48(6): 1474–88. https://doi.org/10.2166/nh.2017.200.

Criss, R., and W.E. Winston. 2008. "Do Nash Values Have Value? Discussion and Alternate Proposals." *Hydrological Processes* 22: 2723–25.

Dang, T.D., A.K. Chowdhury, and S. Galelli. 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://hess.copernicus.org/articles/24/397/2020/.

Dudley, R.W., and G.A. Hodgkins. 2021. "Modeled and Observed Trends at Managed Basins in the Conterminous U.S. from October 1, 1983 through September 30, 2016." U.S. Geological Survey Data Release. https://doi.org/10.5066/P9FS37YQ.

Eng, K., D.M. Carlisle, T.E. Grantham, D.M. Wolock, and R.L. Eng. 2019. "Severity and Extent of Alterations to Natural Streamflow Regimes Based on Hydrologic Metrics in the Conterminous United States, 1980–2014." U.S. Geological Survey Scientific Investigations Report 2019-5001. https://doi.org/10.3133/sir20195001.

Eng, K., D.M. Carlisle, D.M. Wolock, and J.A. Falcone. 2013. "Predicting the Likelihood of Altered Streamflows at Ungauged Rivers across the Conterminous United States." *River Research and Applications* 29(6): 781–91.

England, J.F., Jr., T.A. Cohn, B.A. Faber, J.R. Stedinger, W.O. Thomas, Jr., A.G. Veilleux, J.E. Kiang, and R.R. Mason, Jr. 2019. "Guidelines for determining flood flow frequency—Bulletin 17C (Ver. 1.1, May 2019)." U.S. Geological Survey Techniques and Methods 4-B5. https://doi.org/10.3133/tm4B5.

Falcone, J.A. 2011. "GAGES-II: Geospatial Attributes of Gages for Evaluating Streamflow [Digital Spatial Dataset]." U.S. Geological Survey Water Resources NSDI Node Web Page. https://doi.org/10.3133/70046617.

Falcone, J.A., D.M. Carlisle, D.M. Wolock, and M.R. Meador. 2010. "GAGES: A Stream Gage Database for Evaluating Natural and Altered Flow Conditions in the Conterminous United States. Data Paper in Ecological Archives E091-045-D1." *Ecology* 91(2): 621. https://doi.org/10.1890/09-0889.1.

Farmer, W.H. 2016. "Ordinary Kriging as a Tool to Estimate Historical Daily Streamflow Records." *Hydrology and Earth System Sciences* 20: 2721–35. https://doi.org/10.5194/hess-20-2721-2016.

Farmer, W.H., S.A. Archfield, T.M. Over, L.E. Hay, J.H. LaFontaine, and J.E. Kiang. 2014. "A Comparison of Methods to Predict Historical Daily Streamflow Time Series in the Southeastern United States." U.S. Geological Survey Scientific Investigations Report 2014-5231. https://doi.org/10.3133/sir20145231.

Fennessey, N.M. 1994. "A Hydro-Climatological Model of Daily Streamflow for the Northeast United States." Unpublished Ph.D. diss., Dept. of Civil and Environmental Engineering, Tufts University, Medford, MA.

FitzHugh, T.W., and R.M. Vogel. 2011. "The Impact of Dams on Flood Flows in the United States." *River Research and Applications* 27(10): 1192–215. https://doi.org/10.1002/rra.1417.

Gochis, D.J., M. Barlage, A. Dugger, K. FitzGerald, L. Karsten, M. McAllister, J. McCreight, et al. 2018. "The WRF-Hydro Modeling System Technical Description, Version 5.0." NCAR Technical Note.

Graf, W.L. 1999. "Dam Nation: A Geographic Census of American Dams and their Large-Scale Hydrologic Impacts." *Water Resources Research* 35(4): 1305–11.

Graf, W.L. 2006. "Downstream Hydrologic and Geomorphic Effects of Large Dams on American Rivers." *Geomorphology* 79(3–4): 336–60.

Gudmundsson, L., T. Wagener, L.M. Tallaksen, and K. Engeland. 2012. "Evaluation of Nine Large-Scale Hydrological Models with Respect to the Seasonal Runoff Climatology in Europe." *Water Resources Research* 48(11): W11504.

Gupta, H.V., C. Perrin, G. Blöschl, A. Montanari, R. Kumar, M. Clark, and V. Andréassian. 2014. "Large-Sample Hydrology: A Need to Balance Depth with Breadth." *Hydrology and Earth System Sciences* 18: 463–77. https://doi.org/10.5194/hess-18-463-2014.

Hailegeorgis, T.T., and K. Alfredsen. 2017. "Regional Statistical and Precipitation–Runoff Modelling for Ecological Applications: Prediction of Hourly Streamflow in Regulated Rivers and Ungauged Basins." *River Research and Applications* 33: 233–48. https://doi.org/10.1002/rra.3006.

Hain, E.F., J.G. Kennen, P.V. Caldwell, S.A. Nelson, G. Sun, and S.G. McNulty. 2018. "Using Regional Scale Flow–Ecology Modeling to Identify Catchments Where Fish Assemblages Are most Vulnerable to Changes in Water Availability." *Freshwater Biology* 63: 928–45. https://doi.org/10.1111/fwb.13048.

Hay, L.E., and J.H. LaFontaine. 2020. "Application of the National Hydrologic Model Infrastructure with the Precipitation-Runoff Modeling System (NHM-PRMS), 1980–2016, Daymet Version 3 Calibration." U.S. Geological Survey Data Release. https://doi.org/10.5066/P9PGZE0S.

Helsel, D.R., R.M. Hirsch, K.R. Ryberg, S.A. Archfield, and E.J. Gilroy. 2020. "Statistical Methods in Water Resources." U.S. Geological Survey Techniques and Methods 4-A3. https://doi.org/10.3133/tm4a3.

Hirsch, R.M., and L.A. De Cicco. 2015. "User guide to Exploration and Graphics for RivEr Trends (EGRET) and dataRetrieval: R Packages for Hydrologic Data (Version 2.0, February 2015).". U.S. Geological Survey Techniques and Methods 4-A10. https://doi.org/10.3133/tm4A10.

Hodgkins, G.A., R.W. Dudley, A.M. Russell, and J.H. LaFontaine. 2020. "Comparing Trends in Modeled and Observed Streamflows at Minimally Altered Basins in the United States." *Water* 12(6): 1728. https://doi.org/10.3390/w12061728.

Kendall, M.G. 1938. "A New Measure of Rank Correlation." *Biometrika* 30: 81–93.

Kiang, J.E., K.M. Flynn, T. Zhai, P. Hummel, and G. Granato. 2018. "SWToolbox: A Surface-Water Tool-Box for Statistical Analysis of Streamflow Time Series." U.S. Geological Survey Techniques and Methods 4-A11. https://doi.org/10.3133/tm4A11.

Kiang, J.E., D.W. Stewart, S.A. Archfield, E.B. Osborne, and K. Eng. 2013. "A National Streamflow Network Gap Analysis." U.S. Geological Survey Scientific Investigations Report 2013-5013.

Kim, S. 2015a. "ppcor: An R Package for a Fast Calculation to Semi-Partial Correlation Coefficients." *Communications for Statistical Applications and Methods* 22(6): 665–74. https://doi.org/10.5351/CSAM.2015.22.6.665.

Kim, S. 2015b. *ppcor: Partial and Semi-Partial (Part) Correlation. R Package Version 1.1*. https://CRAN.R-project.org/package=ppcor

Koenker, R. 2021. "quantreg: Quantile Regression, R Package Version 5.86." https://CRAN.R-project.org/package=quantreg.

Koenker, R., P. Ng, and S. Portnoy. 1994. "Quantile Smoothing Splines." *Biometrika* 81(4): 673–80. https://doi.org/10.1093/biomet/81.4.673.

Kroll, C., J. Luz, B. Allen, and R.M. Vogel. 2004. "Developing a Watershed Characteristics Database to Improve Low Streamflow Prediction." *Journal of Hydrologic Engineering* 9(2): 116–25. https://doi.org/10.1061/(ASCE)1084-0699(2004)9:2(116.

LaFontaine, J.H., R.M. Hart, L.E. Hay, W. Farmer, A. Bock, R.J. Viger, S.L. Markstrom, R.S. Regan, and J. Driscoll. 2019. "Simulation of Water Availability in the Southeastern United States for Historical and Potential Future Climate and Land-Cover Conditions." U.S. Geological Survey Scientific Investigations Report 2019-5039. https://pubs.er.usgs.gov/publication/sir20195039.

Leavesley, G.H., R.W. Lichty, B.M. Troutman, and L.G. Saindon. 1983. "Precipitation-Runoff Modeling System—User's Manual." U.S. Geological Survey Water-Resources Investigations Report 83–4238.

Leopold, L.B. 1962. "Rivers." *American Scientist* 50(4): 511–37.

Lorenz, D.L. 2015. "smwrBase—An R Package for Managing Hydrologic Data, Version 1.1.1." U.S. Geological Survey Open-File Report 2015–1202, 7 pp. https://www.usgs.gov/publications/smwrbase-r-package-managing-hydrologic-data-version-111.

Magilligan, F.J., and K.H. Nislow. 2005. "Changes in Hydrologic Regime by Dams." *Geomorphology* 71(1–2): 61–78.

Markstrom, S.L., R.S. Regan, L.E. Hay, R.J. Viger, R.M.T. Webb, R.A. Payn, and J.H. LaFontaine. 2015. "PRMS-IV, the Precipitation-Runoff Modeling System, Version 4." U.S. Geological Survey Techniques and Methods 6-B7. https://doi.org/10.3133/tm6B7.

McCuen, R.H., P.A. Johnson, and R.M. Ragan. 2002. "Highway Hydrology: Hydraulic Design Series Number 2." National Highway Institute, Federal Highway Administration. FHWA-NHI-02-001.

McMahon, T.A., G. Laaha, J. Parajka, M.C. Peel, H.H.G. Savenije, M. Sivapalan, J. Szolgay, et al. 2013. "Prediction of Annual Runoff in Ungauged Basins." In *Runoff Prediction in Ungauged Basins—Synthesis across Processes, Places and Scales*, edited by G. Blöschl, 70–101. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9781139235761.010.

McMahon, T.A., G.G.S. Pegram, R.M. Vogel, and M.C. Peel. 2007. "Revisiting Reservoir Storage–Yield Relationships Using a Global Streamflow Database." *Advances in Water Resources* 30(8): 1858–72. https://doi.org/10.1016/j.advwatres.2007.02.003.

McManamay, R.A., D.J. Orth, and C.A. Dolloff. 2012. "Revisiting the Homogenization of Dammed Rivers in the Southeastern US." *Journal of Hydrology* 424: 217–37.

Miller, O.L., A.L. Putman, J. Alder, M. Miller, D.K. Jones, and D.R. Wise. 2021. "Changing Climate Drives Future Streamflow Declines and Challenges in Meeting Water Demand across the Southwestern United States." *Journal of Hydrology X* 11: 100074. https://www.sciencedirect.com/science/article/pii/S2589915521000018.

Mohamoud, Y.M. 2008. "Prediction of Daily Flow Duration Curves and Streamflow for Ungauged Catchments Using Regional Flow Duration Curves." *Hydrological Sciences Journal* 53(4): 706–24. https://doi.org/10.1623/hysj.53.4.706.

Muggeo, V.M.R. 2003. "Estimating Regression Models with Unknown Break-Points." *Statistics in Medicine* 22: 3055–71.

Nash, J.E., and J.V. Sutcliffe. 1970. "River Flow Forecasting through Conceptual Models Part I—A Discussion of Principles." *Journal of Hydrology* 10(3): 282–90.

National Inventory of Dams (NID). 2022. "Data dictionary." https://nid.sec.usace.army.mil/#/documents.

Ouyang, W., K. Lawson, D. Feng, L. Ye, C. Zhang, and C. Shen. 2021. "Continental-Scale Streamflow Modeling of Basins with Reservoirs: Towards a Coherent Deep-Learning-Based Strategy." *Journal of Hydrology* 599: 126455. https://www.sciencedirect.com/science/article/pii/S0022169421005023.

Over, T.M., W.H. Farmer, and A.M. Russell. 2018. "Refinement of a Regression-Based Method for Prediction of Flow-Duration Curves of Daily Streamflow in the Conterminous United States." U.S. Geological Survey Scientific Investigations Report 2018-5072. https://doi.org/10.3133/sir20185072.

Over, T.M., J.R. Griffin, T.O. Hodson, and K.J. Miles. 2021. "Modeled and Observed Streamflow Statistics at Managed Basins in the Conterminous U.S. from October 1, 1983 through September 30, 2016." U.S. Geological Survey Data Release. https://doi.org/10.5066/P9Z5D4ZT.

Perdios, A., and A. Langousis. 2020. "Revisiting the Statistical Scaling of Annual Discharge Maxima at Daily Resolution with Respect to the Basin Size in the Light of Rainfall Climatology." *Water* 12(2): 610. https://www.mdpi.com/2073-4441/12/2/610.

Pettitt, A.N. 1979. "A Non-Parametric Approach to the Change-Point Problem." *Journal of the Royal Statistical Society: Series C (Applied Statistics)* 28: 126–35.

Poff, N.L., J.D. Olden, D.M. Merritt, and D.M. Pepin. 2007. "Homogenization of Regional River Dynamics by Dams and Global Biodiversity Implications." *Proceedings of the National Academy of Sciences of the United States of America* 104(14): 5732–37.

Pohlert, T. 2020. "trend: Non-Parametric Trend Tests and Change-Point Detection. R Package Version 1.1.4.". https://CRAN.R-project.org/package=trend.

R Core Team. 2019. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Regan, R.S., K.E. Juracek, L.E. Hay, S. Markstrom, R. Viger, J. Driscoll, J. LaFontaine, and P.A. Norton. 2019. "The U. S. Geological Survey National Hydrologic Model Infrastructure: Rationale, Description, and Application of a Watershed-Scale Model for the Conterminous United States." *Environmental Modelling & Software* 111: 192–203. https://doi.org/10.1016/j.envsoft.2018.09.023.

Regan, R.S., S.L. Markstrom, L.E. Hay, R.J. Viger, P.A. Norton, J. Driscoll, and J.H. LaFontaine. 2018. "Description of the National Hydrologic Model for Use with the Precipitation-Runoff Modeling System (PRMS)." In: Techniques and Methods 6-B9; U.S. Geological Survey. https://doi.org/10.3133/tm6B9.

Russell, A.M., T.M. Over, and W.H. Farmer. 2018. "Statistical Daily Streamflow Estimates at HUC12 Outlets in the Conterminous United States, Water Years 1981–2017." U.S. Geological Survey Data Release. https://doi.org/10.5066/P9DPSY6G.

Russell, A.M., T.M. Over, and W.H. Farmer. 2020. "Cross-Validation Results for Five Statistical Methods of Daily Streamflow Estimation at 1,385 Reference Streamgages in the Conterminous United States, Water Years 1981–2017." U.S. Geological Survey Data Release. https://doi.org/10.5066/P9XT4WSP.

Russell, A.M., T.M. Over, W.H. Farmer, and K.J. Miles. 2021. "Statistical Daily Streamflow Estimates at GAGES-II Non-Reference Streamgages in the Conterminous United States, Water Years 1981–2017." U.S. Geological Survey Data Release. https://doi.org/10.5066/P9PA9PKM.

Ryberg, K.R., G.A. Hodgkins, and R.W. Dudley. 2020. "Change Points in Annual Peak Streamflows: Method Comparisons and Historical Change Points in the United States." *Journal of Hydrology* 583: 124307. https://www.sciencedirect.com/science/article/pii/S002216941931042X.

Salinas, J.L., G. Laaha, M. Rogger, J. Parajka, A. Viglione, M. Sivapalan, and G. Blöschl. 2013. "Comparative Assessment of Predictions in Ungauged Basins—Part 2: Flood and Low Flow Studies." *Hydrology and Earth System Sciences* 17: 2637–52. https://doi.org/10.5194/hess-17-2637-2013.

Shin, S., Y. Pokhrel, and G. Miguez-Macho. 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.

Sivapalan, M. 2005. "Pattern, Process and Function: Elements of a New Unified Hydrologic Theory at the Catchment Scale." In *Encyclopaedia of Hydrologic Sciences*, edited by M.G. Anderson, 193–219. Hoboken, NJ: John Wiley.

Smakhtin, V.U. 2001. "Low Flow Hydrology: A Review." *Journal of Hydrology* 240(3–4): 147–86.

Stahl, K., L.M. Tallaksen, L. Gudmundsson, and J.H. Christensen. 2011. "Streamflow Data from Small Basins: A Challenging Test to High-Resolution Regional Climate Modeling." *Journal of Hydrometeorology* 12: 900–12. https://doi.org/10.1175/2011jhm1356.1.

Tefs, A.A.G., T.A. Stadnyk, K.A. Koenig, S.J. Dery, M.K. Macdonald, P. Slota, J. Crawford, and M. Hamilton. 2021. "Simulating River Regulation and Reservoir Performance in a Continental-Scale Hydrologic Model." *Environmental Modelling & Software* 141: 105025. https://www.sciencedirect.com/science/article/pii/S1364815221000682.

Thornton, P.E., H. Hasenauer, and M. White. 2000. "Simultaneous Estimation of Daily Solar Radiation and Humidity from Observed Temperature and Precipitation: An Application over Complex Terrain in Austria." *Agricultural and Forest Meteorology* 104: 255–71. https://doi.org/10.1016/s0168-1923(00)00170-2.

Thornton, P.E., S.W. Running, and M.A. White. 1997. "Generating Surfaces of Daily Meteorological Variables over Large Regions of Complex Terrain." *Journal of Hydrology* 190: 214–51.

Thornton, P.E., M.M. Thornton, B.W. Mayer, Y. Wei, R. Devarakonda, R.S. Vose, and R.B. Cook. 2017. *Daymet: Daily Surface Weather Data on a 1-Km Grid for North America, Version 3*. Oak Ridge, TN: ORNL DAAC. https://doi.org/10.3334/ORNLDAAC/1328.

U.S. Geological Survey. 2019. "National Water Information System—Web interface." https://waterdata.usgs.gov/nwis.

UNEP. 1992. *World Atlas of Desertification* (Edited by N. Middleton and D.S.G. Thomas). UNEP, Edward Arnold: London.

Viger, R.J., and A. Bock. 2014. "GIS Features of the Geospatial Fabric for National Hydrologic Modeling." U.S. Geological Survey. https://doi.org/10.5066/F7542KMD.

Villarini, G. 2016. "On the Seasonality of Flooding across the Continental United States." *Advances in Water Resources* 87: 80–91. https://doi.org/10.1016/j.advwatres.2015.11.009.

Vincenty, T. 1975. "Direct and Inverse Solutions of Geodesics on the Ellipsoid with Application of Nested Equations." *Survey Review* 23(176): 88–93.

Vogel, R.M., and N.M. Fennessey. 1995. "Flow Duration Curves II—A Review of Applications in Water Resources Planning." *Water Resources Bulletin* 31(6): 1029–39.

Vogel, R.M., M. Lane, R.S. Ravindiran, and P. Kirshen. 1999. "Storage Reservoir Behavior in the United States." *Journal of Water Resources Planning and Management* 125(5): 145–54. https://doi.org/10.1061/(ASCE)0733-9496(1999)125:5(245.

Wagener, T., G. Blöschl, D.C. Goodrich, H.V. Gupta, M. Sivapalan, Y. Tachikawa, P.A. Troch, and M. Weiler. 2013. "A Synthesis Framework for Runoff Prediction in Ungauged Basins." In *Runoff Prediction in Ungauged Basins—Synthesis across Processes, Places and Scales*, edited by G. Blöschl, 11–28. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9781139235761.010.

Wagener, T., M. Sivapalan, P.A. Troch, and R.A. Woods. 2007. "Catchment Classification and Hydrologic Similarity." *Geography Compass* 1(4): 901–31.

Wang, D., and M. Hejazi. 2011. "Quantifying the Relative Contribution of the Climate and Direct Human Impacts on Mean Annual Streamflow in the Contiguous United States." *Water Resources Research* 47: W00J12. https://doi.org/10.1029/2010WR010283.

Wang, W., H.Y. Li, L.R. Leung, W. Yigzaw, J. Zhao, H. Lu, Z. Deng, Y. Demisie, and G. Blöschl. 2017. "Nonlinear Filtering Effects of Reservoirs on Flood Frequency Curves at the Regional Scale." *Water Resources Research* 53(10): 8277–92.

Weingartner, R., G. Blöschl, D.M. Hannah, D.G. Marks, J. Parajka, C.S. Pearson, M. Rogger, et al. 2013. "Prediction of Seasonal Runoff in Ungauged Basins." In *Runoff Prediction in Ungauged Basins—Synthesis across Processes, Places and Scales*, edited by G. Blöschl, 102–34. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9781139235761.010.

Wilks, D.S. 2006. *Statistical Methods in the Atmospheric Sciences*, 2nd ed. Amsterdam: Elsevier.

Williams, G.P., and M.G. Wolman. 1984. "Downstream Effects of Dams on Alluvial Rivers." U.S. Geological Survey Professional Paper 1286.

Yang, W., H. Yang, D. Yang, and A. Hou. 2021. "Causal Effects of Dams and Land Cover Changes on Flood Changes in Mainland China." *Hydrology and Earth System Sciences* 25: 2705–20 https://hess.copernicus.org/articles/25/2705/2021/.

Yassin, F., S. Razavi, M. Elshamy, B. Davison, G. Sapriza-Azuri, and H. Wheater. 2019. "Representation and Improved Parameterization of Reservoir Operation in Hydrological and Land-Surface Models." *Hydrology and Earth System Sciences* 23: 3735–64 https://hess.copernicus.org/articles/23/3735/2019/.

Zajac, Z., B. Revilla-Romero, P. Salamon, P. Burek, F.A. Hirpa, and H. Beck. 2017. "The Impact of Lake and Reservoir Parameterization on Global Streamflow Simulation." *Journal of Hydrology* 548: 552–68 https://www.sciencedirect.com/science/article/pii/S0022169417301671.

Zhao, G., and H. Gao. 2019. "Estimating Reservoir Evaporation Losses for the United States: Fusing Remote Sensing and Modeling Approaches." *Remote Sensing of Environment* 226: 109–24. https://doi.org/10.1016/j.rse.2019.03.015.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Hodgkins, Glenn A., Thomas M. Over, Robert W. Dudley, Amy M. Russell and Jacob H. LaFontaine 2024. "The Consequences of Neglecting Reservoir Storage in National-scale Hydrologic Models: An Appraisal of Key Streamflow Statistics." *JAWRA Journal of the American Water Resources Association* 60 (1): 110–131. https://doi.org/10.1111/1752-1688.13161.