Title:

Generalizing Reservoir Operations using a Piecewise Classification and Regression Approach

Authors:

Lucas Ford[a], A. Sankarasubramanian[a]

a. Department of Civil, Construction, and Environmental Engineering, North Carolina State University, Raleigh, NC, United States of America

Corresponding Author:

Lucas Ford, lcford2@ncsu.edu

Abstract:

Inflow anomalies at varying temporal scales, seasonally varying storage mandates, and multi-purpose allocation requirements contribute to reservoir operational decisions. The difficulty of capturing these constraints across many basins in a generalized framework has limited the accuracy of streamflow estimates in Land Surface Models for locations downstream of reservoirs. We develop a Piece Wise Linear Regression Tree to learn generalized daily operating policies from 76 reservoirs from four major basins across the coterminous US. Reservoir characteristics, such as residence time and maximum storage, and daily state variables, such as storage and inflow, are used to group similar observations across all reservoirs. Linear regression equations are then fit between daily state variables and release for each group. We recommend two models – Model 1 (M1) that performs the best when simulating untrained records but is complex, and Model 2 (M2) that is nearly as performant as M1 but more parsimonious. The simulated release median root mean squared error is 49.7% (53.2%) of mean daily release with a median Nash-Sutcliffe Efficiency of 0.62 (0.52) for M1 (M2). Long-term residence time is shown to be useful in grouping similar operating reservoirs. Release from low residence time reservoirs can be mostly described using inflow-based variables. Operations at higher residence time reservoirs are more related to previous release variables or storage variables, depending on the current inflow. The ability of the models presented to capture operational dynamics of many types of reservoirs indicates their potential to be used for untrained and limited data reservoirs.

Keywords: reservoir operation, generalized release policies, reservoir statistical modeling

Index Terms: 1834 Human impacts, 1847 Modeling, 1857 Reservoirs, 1816 Estimation and forecasting, 1884 Water supply

1. Introduction

More than half of the river systems across the globe are regulated by dams to provide water for human needs (Nilsson et al., 2005) and, globally, dams hold 1/6 of the global annual river discharge (Hanasaki et al., 2006). The role reservoirs and dams play in altering local and regional streamflow patterns cannot be understated, nor can their role in serving modern society for various purposes (Chalise et al., 2021). Despite the flow alteration induced by reservoirs, dams are essential to modern life and humans have been constructing dams to help ensure water availability for 5000 years (Tortajada et al., 2012). Today, dams help control floods, smooth natural variation in water supply by storing water for future use, generate carbon-neutral electricity by releasing water to turn turbines, create navigable waters for shipping, and provide recreation benefits to society (Binnie, 2004; Ford et al., 2022).

While dams provide many benefits and are critical for socio-economic development, they can have a significant effect on impacting in-lake and downstream water quality due to reduced transport (Biemans et al., 2011; McCartney, 2009; Pokhrel et al., 2016). Even though dams typically do not create large reductions in mean annual streamflow, except for decreases due to increased evaporation, streamflow modulation due to reservoir operations tends to take place predominantly over sub-annual time periods such as seasonally, monthly, or daily (Haddeland et al., 2014). But, in arid regions that experience significant interannual variability in streamflow, reservoirs (e.g., Hoover Dam) and change in management practices combine to regulate streamflow over multiple years (Kumar et al., 2022). In addition to evaporative losses directly from reservoirs, consumptive use, which is mostly from irrigation, also decreases the total amount of streamflow that is eventually discharged into the oceans globally by more than 4% an average (Haddeland et al., 2006a). Further, in arid regions, consumptive use can decrease river discharge by as much as 30% per month in the arid western U.S (Haddeland et al., 2006b).

Additionally, reservoirs can influence their local climate via the changes in available water and energy that result from large volumes of impounded water. Highly regulated basins in Mediterranean and arid climates experience more intense storms than unregulated basins due to increased lake evaporation (Degu et al., 2011). Despite these impacts on local and basin-wide impacts on land-surface response, most land-surface models (LSMs) do not consider sub-grid-scale reservoir storage and operational policies in estimating the streamflow and evapotranspiration from the LSMs.

Efforts to quantify reservoir influence on streamflow are traditionally based on basin-level reservoir system models such as RiverWare (Zagona et al., 2001), Water Evaluation And Planning System (WEAP) (Yates et al., 2005), HEC-ResSim (Klipsch et al., 2021), MODSIM (Labadie, 2005), and Generalized Multi-Reservoir Analyses using Probabilistic Streamflow forecasts (GRAPS) (Xuan et al., 2020). Though the above simulation and optimization of reservoir systems can be very accurate for individual reservoirs/basins, they do not scale to continental-scale LSMs due to the computational complexity in running the simulation-optimization models within the LSMs (Voisin et al., 2013). Recently, studies have focused on quantifying reservoir influence on local land-surface response – streamflow and evapotranspiration – in LSMs (Hanasaki et al., 2006). Efforts to improve the representation of reservoirs in LSMs can be broadly grouped into three categories: *inflow-demand characterization, optimization-simulation modeling,* and *data-driven modeling*. Due to the complexity of reservoir operations and lack of detailed release and inflow information and generalizable operating policies, initial efforts to capture reservoir operations accurately in LSMs have employed generic release policies largely based on inflow and demand (Hanasaki et al., 2006) or optimization schemes (Haddeland et al., 2006a). Haddeland et al. (2006a) use a priority-based optimization routine to estimate reservoir releases for modifying LSMs response. Recent

efforts have leveraged data-driven methods to learn reservoir releases policies from historical data (Chen et al., 2022; Coerver et al., 2018; Turner et al., 2020, 2021) or to derive generalized policies for reservoirs in a specific region (Yang et al., 2016, 2021; Zhao & Cai, 2020).

The basis for *inflow-demand characterization* methods can be found in Hanasaki et al. (2006) where reservoirs' monthly release is estimated using only information regarding storage capacity, purpose, inflow, and downstream demand. With respect to reservoir purpose, only two categories are considered (irrigation and non-irrigation), each of which being parameterized in slightly different ways. Though this parameterization is simple and requires little data, it reduces the error in streamflow simulation downstream of reservoirs when compared to simulations that do not consider reservoir operations. Voisin et al. (2013) modified the Hanasaki et al. (2006) model by including flood control and irrigation purposes for multipurpose reservoirs and using natural flow to derive releases rather than impounded flow. Voisin et al. (2013) also represent demand using a crop model instead of observed withdrawals and include reservoir storage targets to help estimate release. While these methods improve on conventional reservoir representations, which in many cases can be as simple as treating reservoirs as weirs (e.g., National Water Model, Barlage et al., 2018), they rely on accurate downstream demand estimates to accurately characterize release patterns. Additionally, they do not leverage the potential benefits of *optimization-simulation* or *data-driven modeling* to learn release patterns from historical data.

Haddeland et al. (2006a) implement an *optimization-simulation model* that determines the optimal daily release from a single reservoir given information regarding storage, inflow, and downstream demand. Each reservoir is optimized with respect to an objective function that is designed for its specific purpose. For example, release from an irrigation reservoir is set to minimize the difference between irrigation demand and reservoir release in each time step while constraining release to be less than the demand. For reservoirs with multiple purposes,

Haddeland et al. (2006a) meet irrigation demands first, then optimize for flood control, and then, when applicable, maximize hydropower generation. While this approach provides good agreement between observed and simulated streamflow, it is reliant on a modified Metropolis Markov Chain Monte Carlo optimization scheme (SCEM-UA, Vrugt et al., 2003) which adds additional computational costs (Voisin et al., 2013). The generalized and simplistic nature of *inflow-demand characterization* methods make them easier to integrate into LSMs and thus are more commonly used than *optimization-simulation* methods.

Data-driven approaches have recently become a popular tool for more accurate prediction of reservoir releases due to their ability to learn relationships between various reservoir variables and release from historical data. Machine learning methods such as neural-networks (Coerver et al., 2018) and hidden-Markov decision trees (Chen et al., 2022) have been successfully used to estimate historical release patterns for specific reservoirs; however, variants of neural-network models are often criticized for their "black-box" nature that prohibits interpretation of the drivers influencing reservoir operations. Further, this opaqueness limits the generalization of these methods to reservoirs outside the training set since there is no functional relationship to extend the knowledge based on a specific basin reservoir to other reservoirs in the region (Yassin et al., 2019). Several transparent data-driven methods have also been employed to learn release patterns for specific reservoirs. Yang et al. (2016) simulate reservoir operations for nine major reservoirs in California by fitting a Classification and Regression Tree (CART) model to each reservoir separately. As CART models must be grown fairly large to fully capture complex relationships (Loh, 2011), separate CART models for every reservoir could result in unwieldy and significant complexity while representing in LSMs. Turner et al. (2021) fit harmonic functions to define the normal storage levels of reservoirs and then parameterize release policies for when storage is above, below, and within the normal operating range. These

parameterizations are based on a combination of harmonic and linear functions and can easily be interpreted to understand the important variables and parameters that drive seasonal variations in releases for a given reservoir. Each of the *data-driven* methods discussed so far fit specific models for each individual reservoir considered for that study. Thus, even if the model form is generalized, the actual parameterizations and their estimates are only applicable to individual reservoirs. *While this approach can result in accurate predictions for the reservoirs included in each study, the inability to apply those methods to reservoirs not in the training set limits their practical application in LSMs.*

Turner et al. (2021) propose a solution to this limitation that relies on extrapolating parameterizations from "data-rich" reservoirs, those that have parameterizations fit to their data, to "data-scarce" reservoirs, those that do **not** have parameterizations fit to them. The underlying assumption of this extrapolation method is that reservoirs that are close in proximity, ideally within the same HUC4, and have similar operating purposes will be operated similarly. This extrapolation method is shown to be effective in many cases, but it does not perform well in regions where there are few "data-rich" reservoirs as the extrapolation procedure relies on having similar reservoirs close to the one being extrapolated for. Further, because there can be reservoirs that have different operating purposes and exist in different basins that still operate very similarly, the spatial proximity and similarity in operations assumptions made for this method may limit its effectiveness.

Zhao and Cai (2020) fit a Hidden-Markov decision tree model (HMM) to a subset of reservoirs in the Upper Colorado River basin, and then test the fitted model on a different subset of reservoirs also in the Upper Colorado River basin. This spatial-split-sample validation procedure ensures the common model is generalized to all reservoirs in the Upper Colorado, rather than relying on specific models for every reservoir in the basin. However, the exclusion of

154 reservoirs from other basins limits the model's applicability for CONUS scale studies or models
155 as reservoirs in various regions can have different operating characteristics (Turner et al., 2021).
156 Zhao and Cai (2020) also do not consider any reservoir characteristics, such as storage capacity,
157 for fitting the HMM. In addition to being easy to calculate or find for many reservoirs, these
158 characteristic variables could help identify similar operating reservoirs across and within basins.
159 To summarize, the generalized methods of Biemans et al. (2011), Haddeland et al.
160 (2006), Hanasaki et al. (2006), Voisin et al. (2013), and Yassin et al. (2019) rely on optimization
161 or downstream demand estimates rather than incorporating historical data to determine optimal
162 parameterizations. Chen et al. (2022), Coerver et al. (2018), Turner et al. (2021), and Yang et al.
163 (2016, 2021) use data-driven methods to learn the operation policies for specific reservoirs.
164 While these methods provide an improvement in model accuracy, they require different
165 parameter sets for every reservoir in the study area and thus cannot generalize to other reservoirs,
166 with the exception of Turner et al. (2021). Zhao & Cai (2020) develop a common release model
167 for reservoirs in the Upper Colorado River basin, but the exclusion of other basins prevents this
168 method from being widely applicable. Though these are recent significant advances in
169 parametrizing reservoir operation, no work to date has provided a generalized parameterization
170 that captures reservoir operation based on historical storage and release patterns and has potential
171 to be applied to reservoirs regardless of basin and type.
172 To address these limitations, we propose a generalized release framework that leverages
173 data-driven methods with an emphasis on providing interpretable parameterizations based on
174 publicly available time series associated with reservoir operation for four major basins with
175 contrasting reservoir characteristics and operational patterns. Towards this, we propose a
176 Piecewise Linear Regression Tree (PLRT) that provides both a classification tree, which is
177 similar to reservoir rule curves, and also a piecewise regression that estimates the release with

178 relevant predictors within the tree. The proposed PLRT is based on the model presented by
179 Alexander and Grimshaw (1996), with modifications in finding the predictors subspace for
180 classification, ensuring minimum sample size within the tree for developing the piecewise
181 regression, and allowing different sets of classification and regression variables. We intend to
182 leverage this PLRT framework to classify/group daily reservoir operations. The proposed PLRT
183 captures the non-linear relationship between the current states of a reservoir and reservoir
184 characteristics to estimate the release patterns using the time series of storage, inflow, and release
185 records. Our PLRT model provides reservoir-operating policies in the form of simple conditional
186 statements and linear equations that can be applied generally to any reservoir. The PLRT models
187 are trained on a wide variety of reservoirs from four major basins that vary in hydroclimatology
188 and operational characteristics over the CONUS to provide release parameterizations and
189 equations which are reservoir-agnostic and robust.

190 The rest of this paper is organized as follows. The study areas and methods (Section 2)
191 are presented along with a discussion regarding variable selection. Next, the performance of the
192 PLRT model is summarized (Section 3) along with the interpretation of the regression tree splits
193 and parameters. Finally, we discuss the strengths and weaknesses of our approach along with
194 potential improvements and future work.

195 **2. Methods and Data**

196 **2.1 Study Area and Data**

197 Daily storage and release records are obtained for 76 reservoirs from the Colorado (19
198 reservoirs), Columbia (11 reservoirs), Missouri (19 reservoirs), and Tennessee (27 reservoirs)
199 river basins (Figure 1). Operating data for the Colorado and Columbia River basins is obtained
200 from the HydroData and Hydromet data portals from the U.S. Bureau of Reclamation (USBR),
201 respectively. Data for the Missouri River reservoirs is collected from the USBRs Hydromet data

202 portal and from the Missouri River Basin Water Management Division of the US Army Corps of
203 Engineers (USACE). The Tennessee Valley Authority (TVA) provided daily records for 27
204 reservoirs they own and operate in the Tennessee River Basin. These basins provide a wide
205 variety of reservoirs for generalizing the reservoir operation as they are multi-purpose and
206 belong to different hydroclimatic regimes ranging from humid (Tennessee and Columbia) to arid
207 (Missouri and Colorado) with the runoff being driven in different proportions by rainfall and
208 snowmelt in each basin. As can be seen in Table 1, the reservoirs considered in this study range
209 from small (700 acre-ft of storage capacity) to very large (more than 29 million acre-ft of storage
210 capacity) with average residence times (Equation 1) ranging from less than a day to more than 4
211 years.

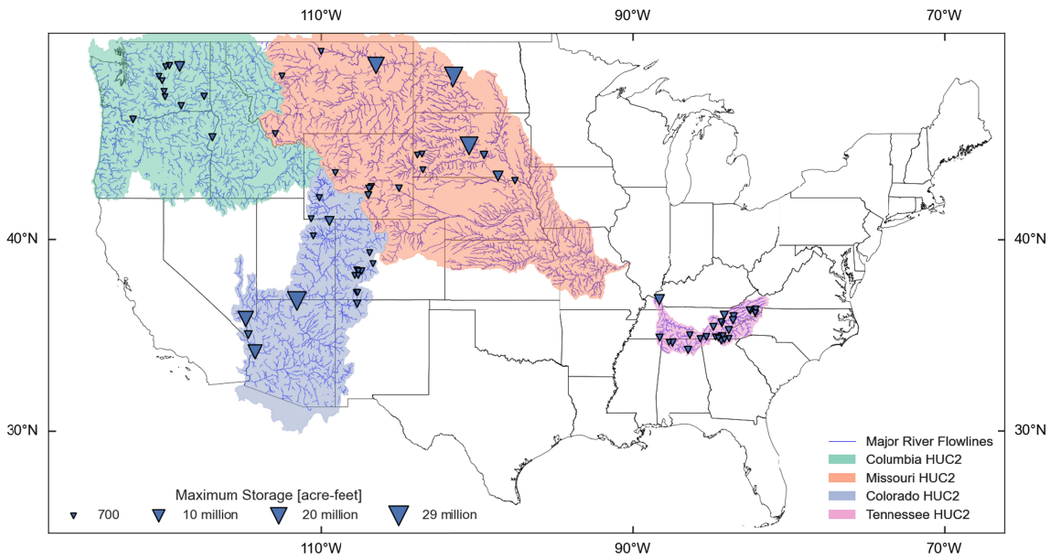

Figure 1 – Studied basins and reservoirs. River flow lines are retrieved from NHDPlus HR (U.S. Geological Survey, 2019) and are filtered to only include those with a stream order greater than 3. River basin boundaries are from the USGS Watershed Boundary (U.S. Geological Survey, 2013).

213

Table 1. Minimum, mean, and maximum values for storage capacity (S_{max}), residence time (RT), mean net inflow ($\overline{NI}$), average storage fraction ($\bar{S}/S_{max}$), and Pearson correlation ($r(R,I)$) between daily release and inflow for all four basins

Metric	Colorado	Missouri	Columbia	Tennessee
S_{max} [1000 acre-ft]				
Minimum	14	1.9	12	0.7
Mean	3,447	4,368	839	879
Maximum	25,695	29,269	5,186	5,649
RT [days]				
Minimum	11	1.2	0.1	0.2
Mean	470	475	13	42
Maximum	1,408	1,576	72	192
$\overline{NI}$ [1000 acre-ft / day]				
Minimum	0.1	0.0	8.4	0.7
Mean	1.5	3.8	51	30
Maximum	7.0	14	95	119
$\bar{S}/S_{max}$				
Minimum	0.0	0.4	0.7	0.5
Mean	0.7	0.7	0.9	0.7
Maximum	0.9	0.9	1.0	1.0
$r(R,NI)$				
Minimum	0.0	0.2	0.8	0.1
Mean	0.7	0.7	1.0	0.7
Maximum	1.0	1.0	1.0	1.0

$$RT = \bar{S}/\bar{R} \quad (1)$$

The data collected for each reservoir does not span the same temporal extents. Only reservoirs with at least 5 years of continuous daily records are included in the study and the longest continuous record is more than 58 years (Glen Canyon Dam). Across all basins, the average length of record is approximately 25 years. More than 83% of all daily records across all basins and reservoirs occur during or after 1990.

Many reservoir data sources report storage and release while not reporting total inflow or evaporation. To encompass as many reservoirs as possible, net inflow ($NI_{r,t}$) is calculated for each reservoir by rearranging the mass balance equation for a reservoir (Equation 2). This is done for all reservoirs to ensure consistency in the variables. Using net inflow also removes the need to calculate evaporation when simulating reservoirs. From here onwards, the term “inflow” implies “net inflow” from equation (2).

$$NI_{r,t} = S_{r,t} - S_{r,t-1} + D_{r,t} \; \forall \, r \in R \; t \in T_r \quad (2)$$

225

226 **2.2 Predictors Selection for PLRT**

227 Initial predictors' selection to estimate the release is driven by both a rearrangement of
228 the mass balance equation for reservoirs, where discharge ($D_{r,t}$) is a function of current and past
229 storage ($S_{r,t}$) and inflow, and variables used in past studies. Several studies (Chen et al., 2022;
230 Coerver et al., 2018; Yang et al., 2016, 2021; Zhao & Cai, 2020) use past (end of time step for
231 previous day) storage and current inflow as predictors and (Coerver et al., 2018; Yang et al.,
232 2021) also use lagged storage and inflow. Yang et al. (2016) also include variables like dry/wet
233 year indicators, runoff indicators, snow depth in upstream mountains, precipitation, and
234 downstream river stage but find that the importance of each variable varies greatly across the set
235 of reservoirs in their study. As we aim to develop the model mimicking the operational model as
236 opposed to inflow prediction, we limit the input variables to storage, inflow, past release, and
237 quantities derived from these (e.g., lagged storage, interaction terms, rolling means).

238 *Daily Variables*

239 Since estimating downstream demand projection (Biemans et al., 2011; Haddeland et al.,
240 2006; Hanasaki et al., 2006; Voisin et al., 2013), or downstream river stage (Yang et al. 2016) is
241 difficult over multiple locations for large basins, we leverage the strong autocorrelation patterns
242 of release to attempt to capture the same information but without requiring another variable to be
243 collected. Across the 76 reservoirs included in this study, the mean lag-1 correlation for release
244 is 0.932 and the minimum is 0.770. When accounting for this relationship, there still exists a
245 weekly seasonal relationship (Figure S1, Partial Autocorrelation Function (PACF)); however, the
246 release relationship with rolling weekly mean release (0.878 average Pearson's r) provides more
247 explanatory power than with weekly lagged release (0.604 average Pearson's r).

Further defining the state of the reservoir, the relationships between release and lagged inflow is calculated for each reservoir and their spatial variation under each basin is summarized for different lags (Figure 2a). There are distinct differences in the inflow-release relationship between basins with release from reservoirs in the Missouri and Colorado River basins generally being less related to inflow, or in some cases inversely related to inflow, than reservoirs in the Columbia and Tennessee River basins. This is partially because arid river basins (Missouri and Colorado) have higher inflow variability and have larger reservoirs compared to humid river basins (Columbia and Tennessee) which experience lesser inflow variability and have relatively smaller systems. This trend also holds true for the 14 lags considered. Additionally, the data show that even within a basin there is a range of release-inflow relationships. This clearly indicates that reservoir operators can respond to inflow in vastly different ways both in different basins and within the same basin.

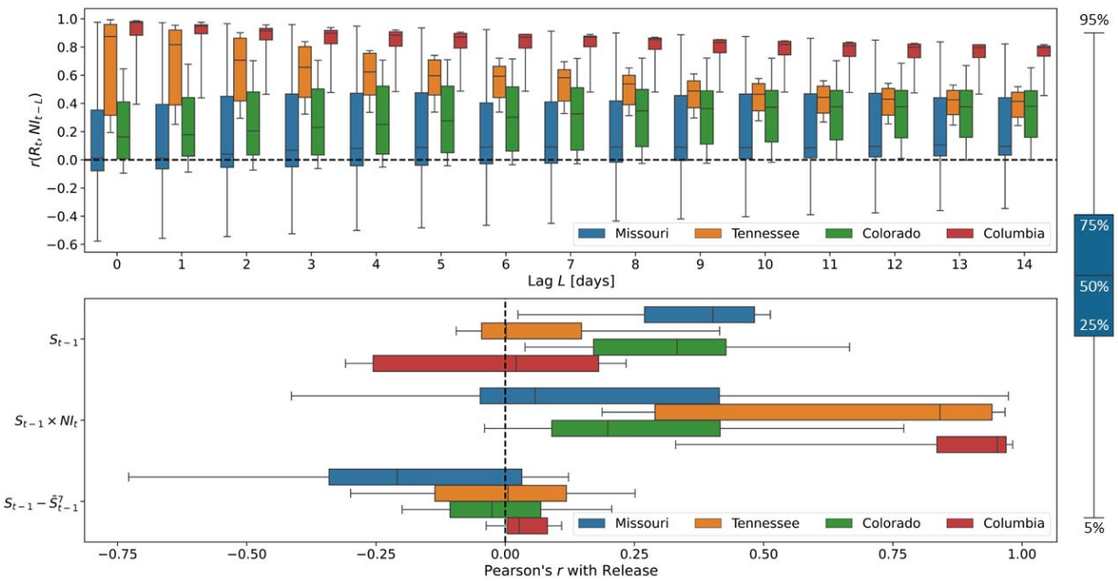

Figure 2 – Release and inflow correlations for lag 0 to 14 for each reservoir, colored by basin (a). Correlations between release and storage-based variables for each reservoir, colored by basin (b). Box plot whiskers represent the 5th and 95th percentiles, as noted by the example boxplot on the right. The dashed black line in each plot indicates the zero line.

To quantify the reservoir dynamics between storage, net-inflow and release under each basin, Figure 2b presents the Pearson correlation between release and several storage-based variables: end of time step storage for previous day S_{t-1}, storage and inflow interaction $S_{t-1} \times NI_t$, and the difference between previous storage and weekly mean storage $S_{t-1} - \bar{S}^7_{t-1}$ (i.e., weekly storage differential). In addition to previous studies indicating the usefulness of storage-based variables in modeling reservoir operations (Yang et al., 2016, 2021; Zhao & Cai, 2020), these variables are included because their relationships with release can vary greatly within basins but can be similar for reservoirs that otherwise are not similar. Similar to the inflow relationships in Figure 2a, there are clear inter- and intra-basin differences between the release response to current and past week storage levels with Colorado and Missouri River basins having mostly positive responses (release increases when storage increases) while release at Tennessee reservoirs has very little dependence on release and some Columbia River reservoirs show strong negative relationships with storage. Regardless of the basin, the relationships between release

and storage are nearly identical to those between release and weekly mean storage, thus weekly mean storage is not included in Figure 2b. This similarity indicates they are redundant variables for statistical learning models; therefore, only previous storage will be included as a predictor.

Interactions between storage and inflow ($S_{t-1} \times NI_t$) have the potential to explain interesting operational scenarios. As shown in Figure 2b, most of the reservoirs have a positive correlation between the storage and inflow interaction and release. This indicates that for situations of high inflow and high storage, which are potential flooding scenarios, release would be higher, and the opposite holds true for low inflow and low storage scenarios. There are some reservoirs in the Colorado and Missouri basins that exhibit relationships opposite of those just described (i.e., release decreases when inflow and storage is high). While these account for only five reservoirs, it highlights the operational differences that exist between reservoirs and emphasizes the need to include this variable as a predictor.

Additionally, the difference between the previous storage and the weekly mean storage may provide insight into reservoir storage trends. When this difference is positive, the reservoir could be building up storage for the summer season or receiving spring flows from snowmelt and when it is negative, the reservoir may be drawing down in anticipation of high flows. Regardless of the basin, when the correlation between weekly storage differential term ($S_{t-1} - \bar{S}^7_{t-1}$) and release is substantial, it is generally negative. This aligns well with the above-discussed dynamics – storage building up and drawing down – indicating the release would decrease when storage is increasing and vice versa. Including this variable as a predictor for release could help partially capture operational patterns that take place over weekly cycles.

In summary, the models developed here will use previous release ($D_{r,t-1}$), previous storage ($S_{r,t-1}$), and current net inflow ($NI_{r,t}$) as independent variables in trying to predict current release ($D_{r,t}$). Also included as independent variables are the past week rolling means of

297 release ($\overline{D}^7_{r,t-1}$) and inflow ($\overline{NI}^7_{r,t}$) to capture the weekly variability in release and inflow.
298 Interaction between storage and net inflow ($S_{r,t-1} \times NI_{r,t}$) and the difference between previous
299 storage and weekly mean storage ($S_{r,t-1} - \bar{S}^7_{r,t-1}$) are also included as predictors.

300 *Reservoir Characteristic Variables*

301 While daily varying values of storage, release, inflow, and their interaction terms explain
302 the variability in reservoir release patterns, using physical or categorical variables can help group
303 reservoirs into clusters with similar operating characteristics. In this study, we attempt to use
304 reservoir primary purpose (categorical), multipurpose nature (binary), maximum storage ($S_{r,max}$)
305 (positive, continuous), and residence time (RT_r) (positive, continuous) (Equation 1) as reservoir
306 characteristics to combine reservoirs into similar groups for modeling purpose.

307 **2.3 Variable Standardization**

308 As our goal is to fit a single, generalized model to all reservoirs, we standardize the
309 release and independent variables to ensure that reservoirs with large release values are not
310 implicitly given more weight during fitting. Three common standardization or normalization
311 methods are considered: 0 to 1 normalization, -1 to 1 normalization, and standardizing to a zero
312 mean, unit standard deviation distribution. Since the response generating functions in PLRTs are
313 multiple linear regression models, standardizing to a zero mean, unit standard deviation
314 distribution will match the assumptions of those models better than the other normalization
315 techniques. This is done for all daily varying variables, but not for any of the reservoir
316 characteristic variables, using Equation 3.

$$x_{r,t} = \left(X_{r,t} - \overline{X_r}\right)/s_r \;\; \forall\, r \in R \; t \in T_r \qquad 3$$

317

2.4 Piecewise Linear Regression Tree (PLRT)

When exploring model formulations, interpretability and parsimony were major considerations. The interpretability provides information on converting the release estimates into relevant operational rules/policies. Model parsimony ensures a simpler model form with no overfitting, facilitating application even for basins with limited data. Given the reservoir mass balance in Equation 2, a choice for a parsimonious model that can still provide satisfactory levels of accuracy and reliability is the multiple linear regression. This formulation is shown in Equation 4, where $\boldsymbol{Y}$ and $\boldsymbol{X}$ are the response vector and predictor matrix, respectively. $\boldsymbol{\beta}$ is the coefficient vector and $\boldsymbol{\epsilon}$ is the error vector. The length of $\boldsymbol{Y}$ and $\boldsymbol{\epsilon}$ is equal to the number of observations in the data set N and the length of $\boldsymbol{\beta}$ is equal to the number of predictors in the model P, plus 1 if an intercept is included. Therefore, the dimensions of $\boldsymbol{X}$ are N rows by $P + 1$ columns.

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon} \quad (4)$$

While this multiple linear regression exhibits many desirable qualities, studies show that a more complex and flexible model is better suited to capture the dynamics of reservoir operations (Yang et al., 2021) even for deriving operational policies for a single reservoir. According to Yang et al. (2021), regression tree-based models such as CART, Random Forest, or XGBoost can efficiently and accurately capture reservoir release policies. Regression trees are comprised of nodes and directed edges that connect those nodes (Loh, 2011). There are two types of nodes. The first type is splitting nodes, where the data set is split by grouping the records where a particular independent variable is less than or equal to some threshold value (τ) into one subset and the records where said independent variable is greater than the same threshold value (τ). The second type of node is leaf nodes, which occur at the end of a branch of the tree and are where the dependent variable is estimated. Each node begins as a splitting node

and only becomes a leaf node if the data cannot be split further, which can be due to limitations on tree depth, minimum samples required in each node, or error reduction requirements.

In traditional regression trees, like those used in CART, Random Forest, and XGBoost, the average of the dependent variable from the subset of the training set that is each leaf node becomes the response estimate (equation 5, where K_l is the set of all observations in leaf node *l*) (Yang et al., 2021). Due to this behavior, traditional regression trees can be called piecewise constant regression trees (PCRTs). Though regression trees provide more flexibility to capture complex relationships, to achieve accurate predictions these methods must grow very large trees or, in the case of XGBoost, add many additional weak trees, which significantly limits their interpretability (Loh, 2011).

$$\hat{y}_l = \frac{1}{N_l} \sum_{k \in K_l} y_k \qquad 5$$

To bridge this accuracy/interpretability gap between multiple linear regression and regression trees, we implement a Piecewise Linear Regression Tree (PLRT) that is built on the work from Alexander and Grimshaw (1996). Piecewise linear regression helps to estimate the non-linear relationship between release and the predictors through localized linear regression between release and the predictors within each leaf node. Similar to hidden states in Zhao and Cai (2020), the tree groups reservoirs that operate similarly under given conditions without being limited to specific basins or purposes. PLRT replaces the mean estimator in each leaf node in PCRTs with a linear regression (equation 4). An illustrative example of the developed PLRTs is provided in Figure 3 along with PCRTs.

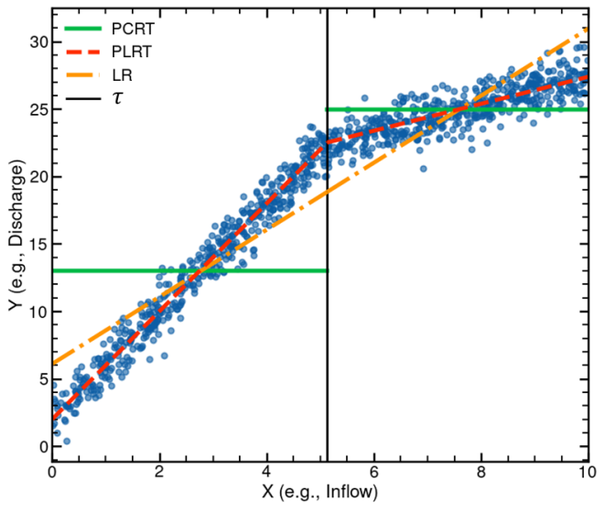

Figure 3 – Illustrative example of the difference between PCRT, PLRT, and Linear Regression (LR). X is representative of an independent variable and Y is a dependent variable. Though PCRT could approach the accuracy of PLRT as the tree is grown larger, PLRT can provide a high level of accuracy with a very shallow tree.

As noted by Alexander and Grimshaw (1996) and Loh (2014), limited computational power has prevented widespread usage of PLRTs despite their ability to accurately model systems with much shallower trees than other methods. However, as access to computational resources has become more ubiquitous (Thompson et al., 2020), the practical limits of such algorithms have been overcome to a reasonable extent.

Similar to Alexander and Grimshaw (1996), our PLRT definition uses multiple linear regression developing a local regression between the predictands and predictors within that tree node. These parameters for these regressions are fit using the matrix form of the least squares estimator (Equation 6).

$$\boldsymbol{\beta} = (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{Y} \qquad 6$$

PLRT – Classification and Parameter Estimation Algorithm

Since there are no readily available software packages to fit PLRT models, we developed a PLRT model in Python to develop reservoir operating policies based on the collected data from

the four major basins. Our PLRT software implementation can be found on GitHub (https://github.com/lcford2/py-plrt) or PyPI (https://pypi.org/project/py-plrt/). The developed methodology in the package is based on PLRT formulation from Alexander and Grimshaw (1996) but customized for the reservoir operation generalization. In this section, we provide the algorithmic details on developing classification and estimation of regression parameters along with the details on how our approach is different from the PLRT formulation of Alexander and Grimshaw (1996).

The parameter estimation for PLRT is implemented as a recursive depth-first growing of a binary tree where each splitting node stores information about the optimal splitting variable and threshold as well as the resulting child nodes and each leaf node stores the optimal parameters of the linear regression for the data in that node. Figure 4 provides the flowchart detailing the steps involved in the general parameter estimation of PLRT. The first step of fitting a PLRT is to fit a linear regression to the entire data set and calculate the MSE. This MSE is used to ensure that the regressions resulting from the next steps improve the model performance enough to be valid. For each candidate split, which consists of a splitting variable (x) and a threshold (τ), the data is split into two subgroups and regressions are fit for each. The candidate split resulting the largest reduction in MSE is chosen as optimal, and the process repeats for each of the data subgroups until termination conditions are met.

To implement a computationally efficient PLRT formulation, finding the optimal splitting threshold (τ) for each independent variable is a critical step. The method proposed by Alexander and Grimshaw (1996) enumerates all possible thresholds for a splitting variable (x), splits the data set for each threshold, then fits models for the resulting subsets and then selects the threshold that results in the lowest error. Rather than enumerating all possible thresholds, we discretize the space between the minimum and maximum x values into 1000 possible values and

then follow remaining steps of the procedure as outlined below using Mean Squared Error (MSE) (equation 7) as the error metric. This approximation limits the computational complexity of the model, especially for many observations, while still providing a value that is near-optimal or optimal.

$$MSE = \frac{1}{N}\sum_{i=0}^{N}(\hat{y}_i - y_i)^2 \quad (7)$$

Further modifying the original PLRT formulation by Alexander and Grimshaw (1996), we allow the model to use different types of independent variables, both continuous and categorical, for splitting the dataset for fitting the regression in each leaf node. In practice, this means that daily variables can be used in the regression equations while splitting variables can be those daily variables or reservoir characteristic variables. This allows categorical variables that are constant in time to be included in the model without the need to encode it for the regression equations.

In addition to the above-mentioned deviations from Alexander and Grimshaw (1996), our formulation facilitates considering multiple criteria for determining valid splits. In our model, a minimum sample size (mss) can be defined as a fraction of the number of observations in the full data set. This is enforced when evaluating potential splitting candidates to ensure that splits are not being made to fit small fractions of the original data set. Further, to encourage a more parsimonious model if one is available, each node can use the persistence model (Equation 8) instead of the multiple linear regression if its performance is near or better than that of the multiple linear regression.

$$\hat{y}_i = y_{i-1} \quad (8)$$

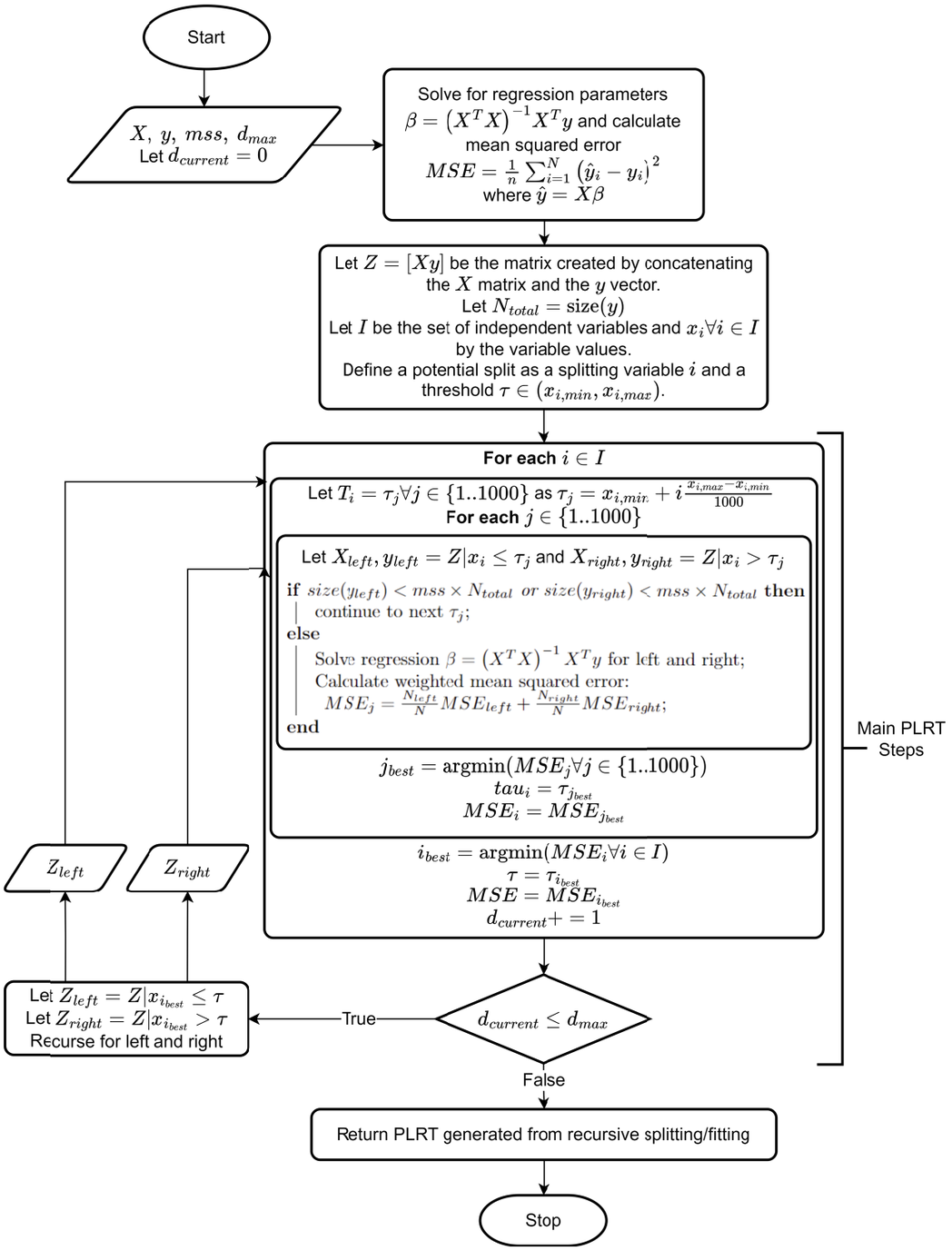

Figure 4 – PLRT implementation flowchart. Successive boxes inside other boxes indicate the body of loops defined by the bold "For each" statements. X is the matrix of independent variables, y is the dependent variable vector, mss is the minimum sample size as a fraction of the size of y, d_{max} is the maximum allowable tree depth. In practice, the Z_{left} subset is solved until the recursion termination condition is met, then the same is performed for Z_{right}. The "size" function provides the number of observations for its argument. The "argmin" function returns the index corresponding to the minimum value of its argument.

As with many data-driven statistical models, the fitting procedure of PLRTs can be tuned by changing a set of parameters. The two most important parameters for PLRTs are the minimum sample size (mss), as defined above, and maximum tree depth (d_{max}). Increasing mss results in less leaves by ensuring that each leaf is fit to a larger proportion of the data. d_{max} limits how many splits are allowed down any single branch of the PLRT. A low d_{max} combined with a high mss limits the flexibility of the model thus gives fewer total regression equations. Increasing d_{max} and decreasing mss too much can result in a tree that is overfit to the training data. We perform a grid search over many combinations of these two parameters to determine the optimal tree.

2.5 Model Training and Evaluation Procedure

Since our interest is a simpler formulation that can be represented in a LHM for simulating release over a period of time given only initial conditions and inflow, we evaluate the performance of PLRT by maintaining the temporal relativity of records. As the training and testing set for our generalization model should include samples from every reservoir, we construct these sets by selecting the first 80% of each reservoir's records for training and use the remaining to validate the model.

To test the model performance similar to how it may be used in an LHM, we simulate release over the testing set. Reservoirs are initialized with a week of storage, inflow, and release values and release is then predicted from the PLRT model. Based on the estimated release, end-of-the-day storage ($S_{r,t}$) calculated using the mass balance equation 2. The next time step uses these calculated quantities along with observed inflow to generate the predictor variables values for the next time step and the process repeats. Though in an LSM the observed inflow would not be known, the focus of this work is to model the release patterns of reservoirs under perfect

information on streamflow; therefore, this simulation scheme can be treated as a best-case scenario in terms of the model's ability to simulate release.

2.6 Model Selection

To determine what combination of parameters generate the best performing PLRT model, we fit the model with maximum depths (d_{max}) from 1 to 8 and minimum sample sizes (mss) from 0.01 to 0.1, as well as 0.15 and 0.2. The chosen limits for these parameters are chosen because higher values result in identical or worse performing models. To compare the simulation performance across reservoirs, the *nRMSE* is calculated by normalizing the *RMSE* by the daily mean release for each reservoir (Equation 9). The results from this parameter sweep are shown in Figure SI-1 and the mean, median, minimum, and maximum *nRMSE* values for the best 10 unique trees are shown in Table 2.

$$nRMSE_r = \frac{1}{\overline{D_r}} \sqrt{\frac{1}{N} \sum_{i=0}^{N} \left(\widehat{D}_{r,i} - D_{r,i}\right)^2} \quad (9)$$

Regardless of the MSS, models with a maximum d_{max} of 1, indicating there are only two groups that have a regression fit for them, are the worst performing. As the mss decreases, the variation in models for different d_{max} values increase. Often, if a model improves the median performance it comes at the cost of the worst performing reservoirs performing more poorly. An ideal model would result in min, mean, median, and maximum *nRMSE* values that are nearest to zero. Additionally, as our goal is to generalize reservoir operations, a more parsimonious model is preferred if the performances are comparable. With these considerations, a maximum d_{max} of 5 with an mss of 0.01 results in the most performant model (M1); however, these parameters result in 25 different regression equations, many of which are very similar. A slightly less performant parameter combination that results in a much more parsimonious model (M2), only 7

regression equations, is d_{max} of 4 and an mss of 0.10. The choice of M1 and M2 over the other models in Table 2 is driven by the relative performance to model complexity tradeoff.

The main difference between M1 and M2 is the maximum error is larger for M2 than M1; however, both maximum errors are greater than 250% of the daily mean release for that reservoir thus determining which one is best based on this difference is not an effective strategy. Therefore, the remainder of our analysis will present results from both models to determine if there are any differences in performance between the two models based on various criteria such as reservoir storage, seasonality, and other attributes.

Table 2. nRMSE performance metrics for model 1 (M1) and model 2 (M2).

Model	# Leaf Nodes	Mean	Median	Minimum	Maximum
M1 (d_{max} = 5, mss = 0.01)	25	0.512	0.498	0.014	2.725
M2 (d_{max} = 4, mss = 0.10)	7	0.518	0.509	0.034	3.011
d_{max} = 5, mss = 0.10	8	0.514	0.493	0.034	3.092
d_{max} = 5, mss = 0.15	5	0.533	0.518	0.035	3.012
d_{max} = 5, mss = 0.08	9	0.528	0.528	0.035	3.027
d_{max} = 6, mss = 0.04	17	0.527	0.512	0.036	2.718
d_{max} = 5, mss = 0.04	15	0.520	0.498	0.036	2.804
d_{max} = 5, mss = 0.05	13	0.528	0.510	0.036	2.796
d_{max} = 7, mss = 0.04	18	0.523	0.524	0.036	2.718
d_{max} = 8, mss = 0.04	19	0.525	0.522	0.036	2.718

To illustrate the trees generated by these two PLRT models, the optimal tree for the M2 parameters is presented in Figure 5. As the optimal tree for the M1 parameters is substantially larger than the M1 parameters (25 final groups rather than 7), that tree is not shown here but can be found in the supplementary information (SI) (Figure SI-2). Any variable in Figure 5 that is lowercased indicates that it has been standardized using Equation 3. Each table presents the optimal parameters for their corresponding independent variables for the regression equation

described by Equation 4 where β_0 is the intercept term. Each record is evaluated at the root node ($RT \leq 9.553$ days) first, and if evaluated to true (false) it will follow the green (orange) arrow to the next node.

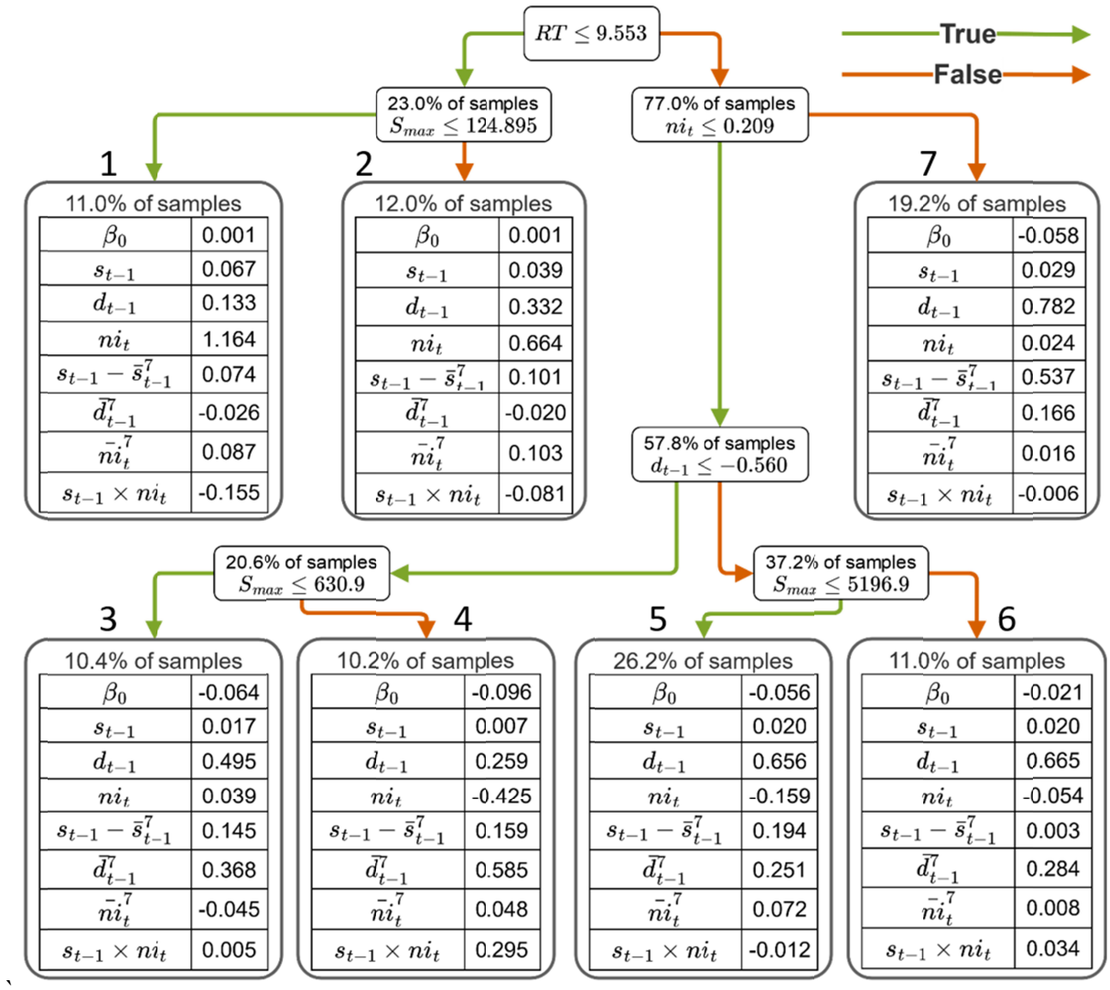

Figure 5 – Optimal PLRT for M2 parameter combination (d_{max} = 4, mss = 0.10). Tables provide the optimal parameters for the regression equation in Equation 4. Lower-case variables indicate standardized versions of the respective upper-case variable.

Examining the tree and parameters in Figure 5 provides insight into which variables have the most impact on release for different types of reservoirs. Release from reservoirs with low residence time can be nearly completely characterized by inflow if they have low storage capacity (Node 1) whereas larger reservoirs with similar residence times have less dependence on inflow and more on previous release (Node 2). For reservoirs with larger residence times,

there are different operating modes that are defined by current inflow and previous release. When current inflow is in the upper 25% of the distribution (Node 3), these reservoirs' release is characterized largely by previous release and the difference in current and past week mean storage.

When inflow is in the lower 75% of the distribution, there seems to be an operational distinction that can be made between situations when the previous release is the lower third of the distribution and when it is in the upper two thirds. For these records, when the previous release is lower the dependence on previous release is also lower but the dependence on past week mean release is larger. The storage and inflow interaction term positively affects release for large reservoirs when the previous release is low, but the storage and inflow interaction term tends to depress current release. When the previous release is higher, the parameters are similar for smaller and larger reservoirs with the major difference being that smaller reservoirs are more dependent on the difference between current and past week mean storage than larger reservoirs. Smaller reservoirs are also more dependent on release.

3. Results

3.1 Model Performance Across Reservoir Attributes

To determine under which circumstances, if any, M1 and M2 perform differently, we characterize their performance across six reservoir attributes in Figure 6. These attributes are release and storage seasonality, calculated following the procedure described by Markham (1970), maximum storage, mean daily release, daily release coefficient of variation (standard deviation divided by mean), and reservoir residence time. For each attribute, the reservoirs are divided into five groups based for each 20^{th} percentile of the attribute then the *nRMSE*s are averaged over those reservoirs. The values that define these percentile groups are presented in

Table 3. Across all 6 attributes, there are not substantial performance differences or trends that can be identified between M1 and M2.

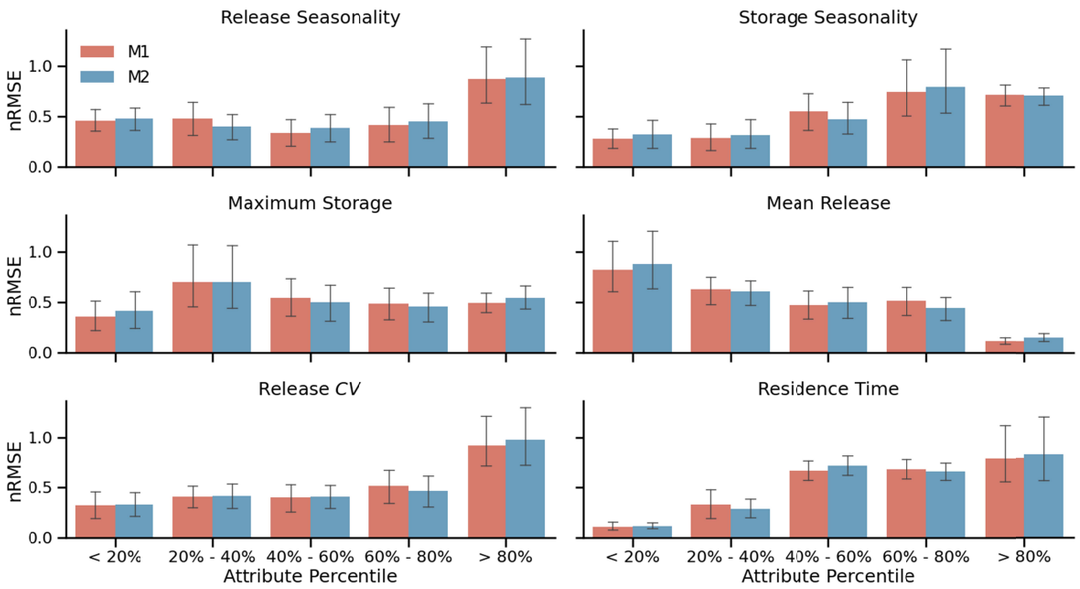

Figure 6 – Average nRMSE values for 20% ranges of reservoir Release Seasonality, Storage Seasonality, Maximum Storage, Mean Release, Release Coefficient of Variation (CV), and Residence Time. Error bars are the standard deviation of the nRMSE values for the reservoirs in each bin.

Table 3. Defining values for percentile groups for reservoir characteristics in Figure 6

Characteristic	Minimum	20%	40%	60%	80%	Maximum
Release Seasonality	0.008	0.126	0.171	0.224	0.396	0.823
Storage Seasonality	0.001	0.012	0.021	0.047	0.102	0.291
Maximum Storage [1000 acre-feet]	0.726	64.2	182	622	1,505	29,269
Mean Release [1000 acre-ft/day]	0.009	0.176	1.08	7.45	53.1	118.9
Release *CV*	0.289	0.583	0.708	0.821	0.959	3.50
Residence Time [days]	0.090	4.49	34.1	127	436	1,576

As there is no explicit representation of seasonal differences in release patterns in our formulation, it is important to examine the performance across varying release seasonality to understand the model limitations. For the reservoirs with a seasonality index less than 0.396 (80th percentile), there is no substantial difference in performance. However, the mean *nRMSE* for

reservoirs with a seasonality index greater than 0.396 is approximately twice as large as those with low seasonality. Though the error is larger or reservoirs with high seasonality, the *NSE*'s between high and low seasonality reservoirs is similar. This indicates that accounting for seasonal biases in release could be enough to improve the error in highly seasonal reservoirs. The storage seasonality displays similar trends to the release seasonality but with the upper 40% having slightly worse performance.

There is no trend in performance observed for reservoirs with different maximum storages; however, there is a distinct trend with mean release. Reservoirs with a high mean release perform substantially better than those with lower mean release. This occurs despite standardizing all the variables for the fitting and simulation process. However, as *nRMSE* is normalized with the mean release, this trend could simply be due to a smaller divisor for those reservoirs with a low mean release. The *NSE*s for this characteristic exhibit a slightly different pattern but with higher release reservoirs still performing the best. The difference is that the model captures the variability of the reservoirs with the lowest release almost as good as those with the highest release, but those in the middle are not as well captured. From Table 3, we can see that those in the lowest 20^{th} percentile release less than 176 acre-feet per day, thus leading to an *nRMSE* average close to 100% but an average *NSE* near 0.6. So, while the highest release reservoirs are modeled the best for both metrics, there is less to differentiate the reservoirs in the lower 80^{th} percentile groups.

Regarding the coefficient of variation (*CV*), the lower 80% of reservoirs, which encompass those that have a release standard deviation up to 96% of the release mean, all perform similarly with a mean *nRMSE* near 50%. When the release standard deviation is greater than the mean (upper 20%), the average *nRMSE* is double that for reservoirs where it is less than the mean. Similar to release *CV*, reservoirs with low residence time have lower *nRMSE* than high

residence time reservoirs. The *nRMSE* for reservoirs with a residence time greater than 34 days, or approximately a month, is very similar and is approximately 75%.

This relationship between performance and residence time is likely driven by the variation in the relationships between inflow and release. Since water is stored for a short period of time in low residence time reservoirs, the inflow and release can be highly correlated. An extreme example would be run-of-river reservoirs, where there is very little storage and thus little operational flexibility to hold water for future uses. For high residence times, especially those which hold water for more than a year on average, the relationship between inflow and release can be nearly zero. This can be seen by looking the regression parameters for high residence time reservoirs in Figure 5 that are essentially zero. After a week of simulation, the only observed information being fed to the model is the inflow and if there is no distinct relationship between inflow and release then the performance will deteriorate.

To summarize, there is very little to distinguish each model by examining these attributes. There are, however, several attributes that provide insight into how the models presented here can simulate reservoir releases. In general, attributes that are related to release (seasonality, daily mean, *CV*, and residence time) exhibit more relationships with performance than storage-based attributes. Very high seasonality, *CV*, and residence time are all indicators of poorer performance than their counterparts while reservoirs with high daily mean releases tend to have better performance.

In addition to the overall trends between reservoir attributes and model performance, the spatial variation in these attributes provides insight into *where* these models perform best. Figure 7 displays the locations of the reservoirs modeled in this work where the markers are colored according to the 20^{th} percentile groups their attributes fall in. The attribute maps in Figure 7 are

562 laid out in the same orientation as the attribute bar charts in Figure 6 and the 20th percentile

563 groups are the same across both figures.

563

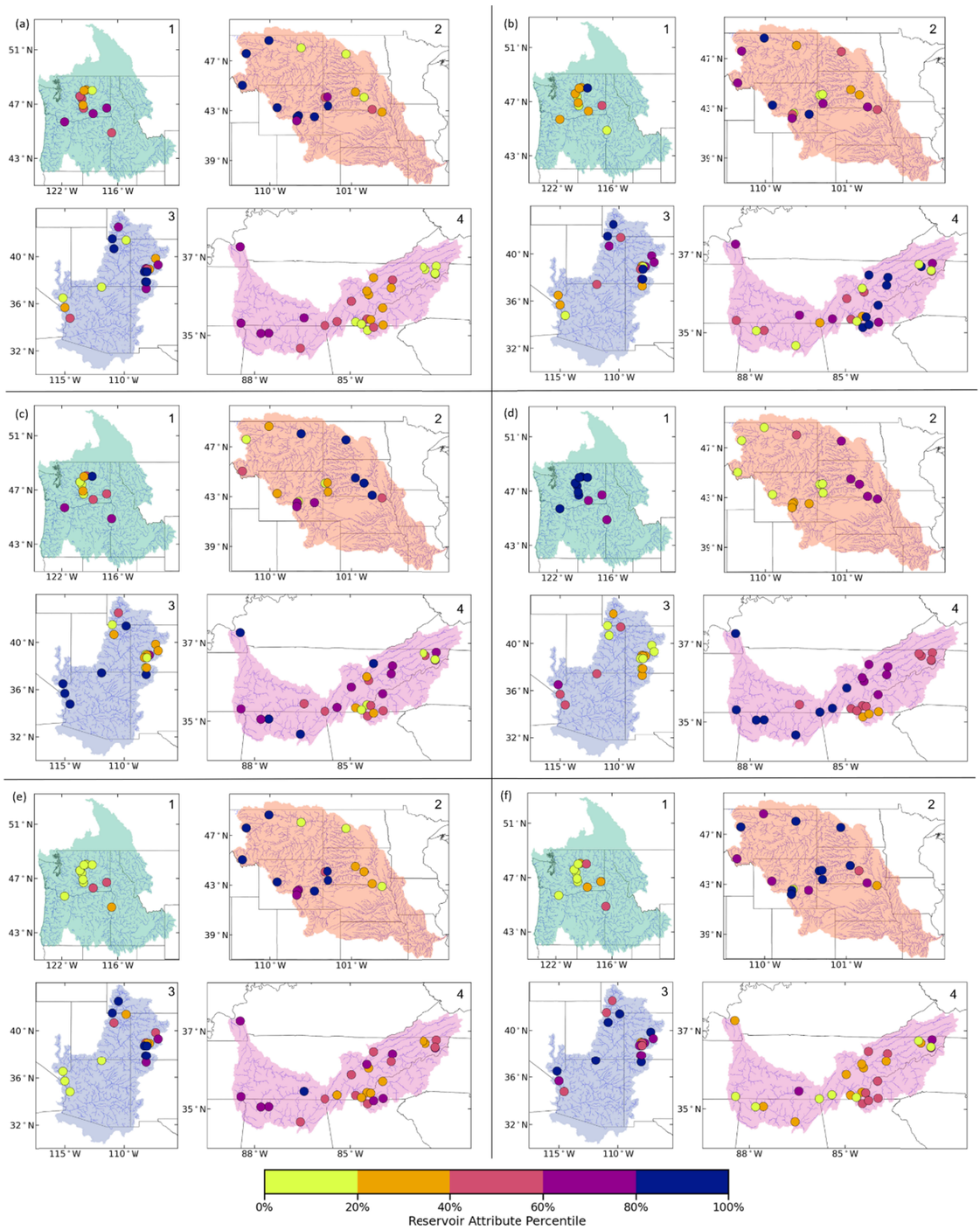

Figure 7 – Reservoir attributes 20th percentile group for each basin (1 – Columbia, 2 – Missouri, 3 – Colorado, 4 – Tennessee) (a) Release Seasonality, (b) Storage Seasonality, (c) Maximum Storage, (d) Mean Release, (e) Release CV, (f) Residence Time.

As depicted in Figure 7(a), the reservoirs with the highest release seasonality, and thus the worst performing, are mostly upstream tributary reservoirs in the Missouri and Colorado river basins. This indicates that though the error for these reservoirs can be relatively large, it likely does not contribute significantly to the total error in their respective basins. This contrasts with the Tennessee River basin, where the downstream reservoirs have more seasonality than upstream reservoirs. In this case, not accounting for the seasonal operational differences could contribute significantly to the total streamflow prediction error in the basin; however, since these reservoirs are only slightly seasonal as compared to those in the Missouri or Colorado River basins their performance is much better (Figure 6). Interestingly, there is very little overlap between reservoirs with high storage and release seasonality as only 20 of the 76 reservoirs are in the same 20th percentile group.

The distribution of maximum reservoir storages is well dispersed across four basins. The four to six largest reservoirs in the Missouri, Colorado, and Tennessee river basins are all modeled and make up most of the reservoirs in the top 20th percentile group. Similar to the seasonalities discussed in the previous paragraph, only 18 reservoirs fall in the same percentile group between maximum storage and mean release. This dispersion contrasts with the dispersion found for the mean release, where all 15 of the reservoirs in the top 20th percentile group are in the Tennessee of Columbia River basins. Since reservoirs with higher mean daily release are modeled better, this indicates that these basins are better modeled than the Missouri and Colorado River basins.

In the Columbia, Colorado, and Missouri basins, the coefficient of variation of release generally decreases as you move from upstream reservoirs to downstream reservoirs. In the Tennessee River basin, however, the release *CV* is much better distributed throughout the basin

586 with a slight trend towards increasing *CV* as you move downstream. As those with larger *CV*s are
587 generally modeled worse than those with smaller *CV*'s, these trends indicate that our models
588 predict release better for downstream reservoirs in the western basins than for the upstream
589 reservoirs. It also indicates that performance across the Tennessee River basin is relatively
590 consistent.

591 As shown in Figure 6, there is a distinct relationship between residence time and model
592 performance with lower residence time reservoirs performing better than those that hold water
593 for longer periods. The reservoirs in the top 20^{th} percentile in terms of residence time (greater
594 than 436 days) all fall in either the Colorado or Missouri river basin. In fact, 28 of the 30
595 reservoirs in this study with a residence time greater than 127 days (approximately one third of a
596 year) are in the Colorado and Missouri River basins with the other two in the Tennessee River
597 basin. As these basins are largely arid (Colorado) or semi-arid (Missouri) (Zomer & Trabucco,
598 2022), many of the reservoirs are designed to manage interannual variability in water supply thus
599 their operational policies are less affected by their current state.

600 The reservoirs in the Columbia River basin occupy the other side of the residence time
601 distribution with 9 of 11 reservoirs holding water for less than 34.1 days, on average. The
602 operations at these reservoirs are heavily dependent on the current and recent past state of the
603 reservoir and thus the variables included in this study provide enough information to accurately
604 characterize the release patterns. Similarly, most reservoirs in the Tennessee River basin have a
605 residence time less than 127 days, indicating their operations are also more dependent on current
606 and recent past data rather than over-year predictors.

607 **3.2 Model Accuracy for Varying Simulation Time Horizons**

608 The results thus far have been with respect to simulation over the testing set where only
609 the initial state of the reservoir and the inflow time series is given, the state of the reservoir for

each subsequent time step is calculated from the previous storage and the predicted release. As the minimum length of the time series for any given reservoir is 5 years, and 20% of the data is set aside for testing, the minimum length of the time series used for simulation is 1 year. Though the minimum is 1 year, more than 50% of the reservoirs tested are simulated for more than 5 years, since they have an overall time series of 25 years or more. The release parameterizations in the leaves of the models are dependent on past storage and release; therefore, there are two sources for error accumulation that, over long periods of time, can prevent the model from accurately predicting future releases. Further, in many cases it is not practical nor desirable to simulate at a daily level for periods over a year.

To understand how the model performs under varying time-horizons, we reinitialize the model with observed storage and release values at daily, weekly, monthly (30 days), seasonally (90 days), and semi-annual (180 days) frequencies. Reinitialization at these frequencies allows the simulation period to be split into several smaller periods, each of which represents a time horizon of interest, while still evaluating over the entire testing period. To reinitialize, the observed storage, release, and inflow over the past week are used to calculate independent variables, rather than using the calculated storage and release from previous time steps. The past week must be used as there are independent variables that rely on seven days of data. For each subsequent time step, the remaining observed values are used until a week of releases have been predicted. For example, the seven day mean release will be for the second time step after reinitialization will be calculated using the predicted release from the previous time step and observed releases for the six days before that. For each reservoir, the percent improvement in *RMSE* relative to simulation with no reinitialization is calculated for these reinitialization frequencies and depicted with boxplots in Figure 8. The box and whisker plots in Figure 8 are configured in the same manner as those in Figure 2.

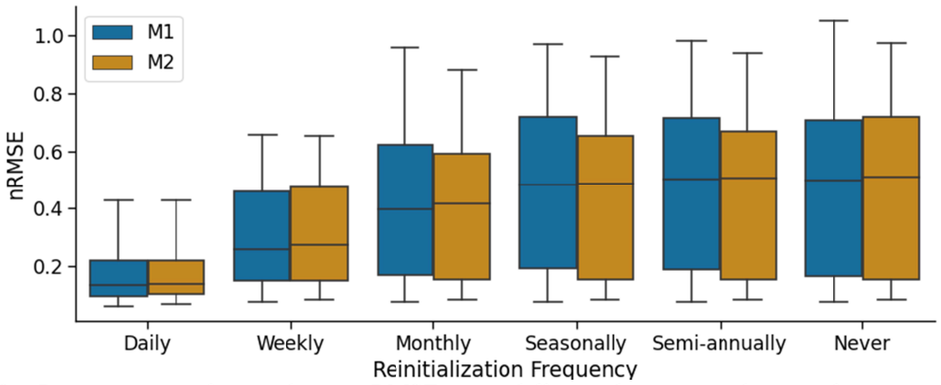

Figure 8 – Reservoir specific simulation nRMSE given different data reinitialization frequencies. Never is when the entire testing period is simulated with no reinitialization.

When reinitializing less frequently, such as seasonally or semi-annually, there is very little, if any, improvement over never simulating data. M2 sees more improvement than M2 as the 75th percentile *nRMSE* decreases more but most of the distribution remains the same as if there was no reinitialization. The monthly frequency is when the *nRMSE* distribution begins to shift away towards zero, indicating a more accurate model. At this frequency, the 50th and 75th percentiles are markedly lower than the cases with less frequent reinitialization. Daily and weekly frequencies result in the largest reductions in *nRMSE* over all reservoirs. At the daily level, the median *nRMSE* is lower than the 25th percentile for all other frequencies. Further, the inter-quartile range for daily reinitialization is approximately one quarter of the inter-quartile range for less frequent reinitializations, indicating a much tighter distribution of performance around the median.

Reinitializing daily is functionally the same as just predicting release values over the testing set as if all independent variable values are known. For the task of predicting daily release, the errors for daily reinitialization are indicative of the models' performances. Weekly reinitialization is also expected to perform well due to the reliance on seven-day information for the rolling mean variables, $\bar{D}^{7}{}_{r,t-1}$ and $\bar{S}^{7}{}_{r,t-1}$. Since all independent variables are calculated

using data that comes from at most the previous week, a weekly reinitialization results in some observed information being present for each simulation time step.

4. Discussion

We propose a Piece-wise Linear Regression Tree (PLRT) for generalizing reservoir operations across four major basins and for providing an interpretable, parsimonious, and accurate model that has potential of incorporating the developed tree-based regression equations in an LSM. The proposed PLRT is flexible in quantifying the non-linear relationship between the selected predictors and the release by assuming a local-linear form within the tree. Our approach here is to combine the available storage, release, and inflow data across all the reservoirs from the four basins and develop a tree model that learns from the similarity in operations across them. Given that the relationship is represented in a regression form within the tree, this has potential for implementation within the LSM in a simplified parameterized form that can improve streamflow estimation from LSMs for controlled basins. With trees smaller than 8 groups and regression equations built only on time-varying physical state variables, these models can simulate reservoir releases over long periods of times (mean of 5.5 years; longest simulation period is 16.7 years) with a median *RMSE* of approximately 51% of a reservoirs daily release values and a median *NSE* of 0.56. Further, as there is not a specific model for each individual reservoir, this approach can be applied to reservoirs untrained reservoirs with no modifications thus widely increasing their applicability.

An additional benefit of using PLRTs is the ability to extract general lessons from the regression equations in the leaves of the tree and the splits that generate those leaves. For example, release at reservoirs with residence times less than approximately 10 days are very strongly related to the current inflow whereas most higher residence time reservoirs do not operate based on current inflow and, when they do, release decreases with increased inflow

674 rather than increasing. Release at these higher residence time reservoirs is also better estimated
675 using the previous days release than lower residence time reservoirs. For reservoirs with
676 residence times greater than 10 days, the current storage minus the past week mean storage term
677 plays a significant role in explaining release patterns for high inflow periods. In this case, since
678 the coefficient is positive, this term works to bring the current storage back towards the past
679 weeks mean storage. When storage has been building up over the past week ($s_{t-1} - \bar{s}_{t-1}^{7} > 0$),
680 this term will increase release to attempt to level off this storage increase. The opposite is true
681 when storage has been declining over the past week.

682 One of the limitations of this approach is the inability to encode seasonal information into
683 the model without reducing its ability to be applied broadly to reservoirs in many different
684 basins. This harms performance for reservoirs with highly seasonal release patterns but could
685 largely be mitigated by post processing the model results for reservoirs with large swings in
686 seasonal operational patterns using the monthly/seasonal mean as the model captures the daily
687 variations well, just not the magnitude differences between seasons. Our approach seems to favor
688 reservoirs with large mean releases even though all values are standardized to mean zero before
689 fitting the models as both the *nRMSE* and *NSE* are significantly better for high release reservoirs
690 than low release reservoirs. This could be due to the relative amount of variation in release
691 patterns tends to decrease as mean release increases thus making these reservoirs slightly easier
692 to model. This is supported by the poor performance for reservoirs with high release *CV*. High
693 release reservoirs tend to be along the main stem of a river basin, thus modeling them well
694 indicates that the overall basin discharge can be significantly improved.

695 Finally, though the models developed here perform adequately when simulating over
696 long periods of time, there is much that can be gained by reinitializing at the highest frequency
697 possible or shortening the time horizon. For short term forecasts our approach with a daily data

reinitialization can model reservoirs with a median *nRMSE* less than 15% of a reservoirs mean release and a median *NSE* of 0.963. For medium range there is a slight drop off in performance to a median *nRMSE* of 27.4% of daily mean release and a median *NSE* of 0.871 when using a weekly reinitialization. There is little difference in model performance for sub-seasonal to seasonal forecasts, such as monthly and beyond. This results primarily from the error in release estimation leading towards initial storage. In reality, for long-range forecasts, given initial storage conditions are known (Li et al., 2014), one can estimate the releases with reduced error, which could be inferred from the assimilation frequency (Figure 8). Furthermore, in many cases the representation of reservoirs in hydrologic models results in release patterns that match reality substantially worse than the models developed here so it is possible that this generalized method is still an improvement.

Though many approaches to generalized reservoir operation models require more variables than this study, including those that are difficult to find or derive such as downstream demand, upstream snow depth, or climatological variables, it is important to discuss the main data requirements as a possible limitation. To accurately determine the residence time of a reservoir, good estimates of long term mean inflow or release and long term mean storage must be available. Additionally, as a zero-mean and unit-standard deviation standardization is used for all independent and dependent variables used in the regressions, reasonable estimates of the long term means and standard deviations must be calculated, usually requiring an observed time series of at least a year. For data-scarce reservoirs, it is possible to use time-series data at coarser time scales such as monthly or yearly and downscale the mean and standard deviation to daily levels. Another way to address this would be to normalize data between 0 and 1. This could be done by only knowing the maximum values of variables and assuming the minimum is zero. However,

this would require replacing the linear regression equations with generalized linear models which could significantly increase the computational costs of fitting the model.

Further, a recently published data set, ResOpsUS (Steyaert et al., 2022), could be used to increase the number of reservoirs used in the fitting process. This data set contains more than 600 reservoirs with nearly 500 of them having the required time series to be used by our approach (daily time series of storage and outflow). Incorporating the reservoirs from this data set could make this model significantly more robust as the model would be able to learn from the similarities and differences in the operational patterns from a wider array of reservoirs. It would also be possible to conduct a spatial evaluation of the model, in addition to a temporal evaluation as we have done here, by leaving out a percentage of reservoirs from each basin when fitting and then testing on those left out. We did not pursue the spatial validation in this study, as our goal is to develop a data-driven, interpretable, and parsimonious modeling approach that can facilitate developing a generalized set of reservoir release equations, which can be used in LHM for estimating the reservoir release in controlled basins.

5. Summary and Conclusions

We develop a Piece Wise Linear Regression Trees to learn generalized operating policies for daily release from 76 reservoirs from four major basins – Missouri, Colorado, Columbia, and Tennessee – across the coterminous US. Reservoir characteristics and daily state variables are used to group similar observations across all reservoirs, and then linear regression equations are fit to daily state variables by classifying the independent variables into different groups. Two models are identified: Model 1 (M1) that performs the best when simulating untrained records but is complex, and Model 2 (M2) that is nearly as performant as the more complex model but more parsimonious. Of the reservoir characteristics considered, long-term residence time is shown to be the most useful in grouping similar operating reservoirs, followed by reservoir

storage capacity. Release from low residence time reservoirs (less than 10 days) can be mostly described using inflow-based variables. Operations at higher residence time reservoirs are more related to previous release variables or storage variables, depending on the current inflow and storage capacity.

The generalized reservoir operation model developed here represents a deviation from the current body of literature by leveraging data-driven methods to develop a single model that can predict release from many reservoirs. Fitting on many reservoirs increases the robustness of this model by learning from the wide array of operational characteristics that the reservoirs in this study represent. The models accurately and reliably predict daily reservoir operations by grouping similar reservoirs and observations and then fitting linear regression equations between reservoir state and release. Overall, the best performing reservoirs are those with lower residence times or high daily mean release. A particular benefit of this approach is having the potential to apply the model to reservoirs not in the training set. The ability of these models to extract general operational characteristics from similar reservoirs and reservoir states should allow the models to accurately predict release from untrained reservoirs.

As the models developed here can be decomposed into a set of Boolean decisions and regression equations, it is possible to extract information from the trees that can be used to further improve this, or other, reservoir modeling method(s). General lessons from the optimal models can be applied without the need to use the full PLRT models developed here. Similarly, comparing the operational policies learned here with water supply manuals for certain reservoirs could provide insight on how to further develop generalized reservoir operation models and should be considered in future studies.

6. Availability Statement

The Piecewise Linear Regression Tree (PLRT) source code can be found on Zenodo at https://zenodo.org/record/7650071 or on GitHub at https://github.com/lcford2/py-plrt. Additionally, it can be installed using the Python package manager "pip" using the package name "py-plrt". The code used to perform all analysis, the data used, and the model results can be found on GitHub at https://github.com/lcford2/predict-release.

Acknowledgements

This research is supported in part by the of the Blue Waters sustained-petascale computing project, which is supported by the National Science Foundation (awards OCI-0725070 and ACI-1238993) the State of Illinois, and as of December, 2019, the National Geospatial-Intelligence Agency. Blue Waters is a joint effort of the University of Illinois at Urbana-Champaign and its National Center for Supercomputing Applications. Additional support for this research is provided by the National Science Foundation, United States of America under Grant No. CBET-1805293.

CRediT authorship contribution statement

Lucas Ford: Conceptualization, Methodology, Software, Investigation, Data curation, Writing, Visualization, Funding acquisition. **A. Sankarasubramanian:** Conceptualization, Methodology, Writing – review & editing, Supervision, Funding acquisition.

Declaration of competing interest

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

References:

Alexander, W. P., & Grimshaw, S. D. (1996). *Treed Regression*. *Source: Journal of Computational and Graphical Statistics* (Vol. 5).

Barlage, M., Dugger, A., FitzGerald, K., Karsten, L., McAllister, M., McCreight, J., et al. (2018). *The WRF-Hydro modeling system technical description, (Version 5.0).* Retrieved from https://ral.ucar.edu/sites/default/files/public/WRFHydroV5TechnicalDescription.pdf

Biemans, H., Haddeland, I., Kabat, P., Ludwig, F., Hutjes, R. W. A., Heinke, J., et al. (2011). Impact of reservoirs on river discharge and irrigation water supply during the 20th century. *Water Resources Research*, *47*(3), 1–15. https://doi.org/10.1029/2009WR008929

Binnie, C. J. A. (2004). The Benefits of Dams to Society. In *Long-term benefits and performance of dams: Proceedings of the 13th Conference of the Britich Dam Society and the ICOLD European Club meeting help at the University of Kent, Canterbury, UK from 22 to 26 June 2004*. Canterbury, UK: Thomas Telford Publishing.

Chalise, D. R., Sankarasubramanian, A., & Ruhi, A. (2021). Dams and Climate Interact to Alter River Flow Regimes Across the United States. *Earth's Future*, *9*(4). https://doi.org/10.1029/2020EF001816

Chen, Y., Li, D., Zhao, Q., & Cai, X. (2022). Developing a generic data-driven reservoir operation model. *Advances in Water Resources*, *167*, 104274. https://doi.org/10.1016/j.advwatres.2022.104274

Coerver, H. M., Rutten, M. M., & Van De Giesen, N. C. (2018). Deduction of reservoir operating rules for application in global hydrological models. *Hydrology and Earth System Sciences*, *22*(1), 831–851. https://doi.org/10.5194/hess-22-831-2018

Degu, A. M., Hossain, F., Niyogi, D., Pielke, R., Shepherd, J. M., Voisin, N., & Chronis, T. (2011). The influence of large dams on surrounding climate and precipitation patterns. *Geophysical Research Letters*, *38*(4), 1–7. https://doi.org/10.1029/2010GL046482

Ford, L., de Queiroz, A., DeCarolis, J., & Sankarasubramanian, A. (2022). Co-Optimization of Reservoir and Power Systems (COREGS) for seasonal planning and operation. *Energy Reports*, *8*, 8061–8078. https://doi.org/10.1016/j.egyr.2022.06.017

Haddeland, I., Skaugen, T., & Lettenmaier, D. P. (2006). Anthropogenic impacts on continental surface water fluxes. *Geophysical Research Letters*, *33*(8), 2–5. https://doi.org/10.1029/2006GL026047

Haddeland, I., Lettenmaier, D. P., & Skaugen, T. (2006). Effects of irrigation on the water and energy balances of the Colorado and Mekong river basins. *Journal of Hydrology*. https://doi.org/10.1016/j.jhydrol.2005.09.028

Haddeland, I., Heinke, J., Biemans, H., Eisner, S., Flörke, M., Hanasaki, N., et al. (2014). Global water resources affected by human interventions and climate change. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(9), 3251–3256. https://doi.org/10.1073/pnas.1222475110

Hanasaki, N., Kanae, S., & Oki, T. (2006). A reservoir operation scheme for global river routing models. *Journal of Hydrology*, *327*(1–2), 22–41. https://doi.org/10.1016/j.jhydrol.2005.11.011

Klipsch, J. D., Hurst, M. B., Modini, G. C., Black, D. L., & O'Connell, S. M. (2021). HEC-ResSim User Manual. *US Army Corps of Engineers Institute for Water Resources Hydrologic Engineering Center (HEC)*. U.S. Army Corps of Engineers.

Kumar, H., Hwang, J., Devineni, N., & Sankarasubramanian, A. (2022). Dynamic Flow Alteration Index for Complex River Networks With Cascading Reservoir Systems. *Water Resources Research*, *58*(1). https://doi.org/10.1029/2021WR030491

Labadie, J. (2005). MODSIM: River basin management decision support system. *Watershed Models. CRC Press, Boca Raton, Florida*.

Li, W., Sankarasubramanian, A., Ranjithan, R. S., & Brill, E. D. (2014). Improved regional water management utilizing climate forecasts: An interbasin transfer model with a risk management framework. *Water Resources Research*, *50*, 6810–6827. https://doi.org/10.1111/j.1752-1688.1969.tb04897.x

Loh, W. Y. (2011). Classification and regression trees. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, *1*(1), 14–23. https://doi.org/10.1002/widm.8

Loh, W. Y. (2014). Fifty years of classification and regression trees. *International Statistical Review*, *82*(3), 329–348. https://doi.org/10.1111/insr.12016

Markham, C. G. (1970). *Seasonality of Precipitation in the United States*. *Source: Annals of the Association of American Geographers* (Vol. 60).

McCartney, M. (2009). Living with dams: Managing the environmental impacts. *Water Policy*, *11*(SUPPL. 1), 121–139. https://doi.org/10.2166/wp.2009.108

Nilsson, C., Reidy, C. A., Dynesius, M., & Revenga, C. (2005). Fragmentation and flow regulation of the world's large river systems. *Science*, *308*(5720), 405–408. https://doi.org/10.1126/science.1107887

Pokhrel, Y. N., Hanasaki, N., Wada, Y., & Kim, H. (2016). Recent progresses in incorporating human land-water management into global land surface models toward their integration into Earth system models. *Wiley Interdisciplinary Reviews: Water*, *3*(4), 548–574. https://doi.org/10.1002/wat2.1150

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

Thompson, N. C., Greenewald, K., Lee, K., & Manso, G. F. (2020). The Computational Limits of Deep Learning. Retrieved from http://arxiv.org/abs/2007.05558

Tortajada, C., Altinbilek, D., & Biswas, A. K. (2012). *Impacs of Large Dams: A Global Assessment*. Springer International Publishing. https://doi.org/10.1007/978-3-642-23571-9

Turner, S. W. D., Doering, K., & Voisin, N. (2020). Data-Driven Reservoir Simulation in a Large-Scale Hydrological and Water Resource Model. *Water Resources Research*, *56*(10). https://doi.org/10.1029/2020WR027902

Turner, S. W. D., Steyaert, J. C., Condon, L., & Voisin, N. (2021). Water storage and release policies for all large reservoirs of conterminous United States. *Journal of Hydrology*, *603*. https://doi.org/10.1016/j.jhydrol.2021.126843

U.S. Geological Survey. (2013). Watershed Boundary Dataset. Retrieved from https://www.usgs.gov/national-hydrography/watershed-boundary-dataset#publications

U.S. Geological Survey. (2019). National Hydrography Dataset Plus High Resolution. Retrieved from https://pubs.er.usgs.gov/publication/ofr20191096

Voisin, N., Li, H., Ward, D., Huang, M., Wigmosta, M., & Leung, L. R. (2013). On an improved sub-regional water resources management representation for integration into earth system models. *Hydrology and Earth System Sciences*, *17*(9), 3605–3622. https://doi.org/10.5194/hess-17-3605-2013

Vrugt, J. A., Gupta, H. V., Bouten, W., & Sorooshian, S. (2003). A Shuffled Complex Evolution Metropolis algorithm for optimization and uncertainty assessment of hydrologic model parameters. *Water Resources Research*, *39*(8). https://doi.org/10.1029/2002WR001642

Xuan, L., Ford, L., Mahinthakumar, G., Souza Filho, F. A., Lall, U., & Arumugam, S. (2020). GRAPS: Generalized Multi-Reservoir Analyses using Probabilistic Streamflow. *Environmental Modelling & Software*, *133*(June), 104802. https://doi.org/10.1016/j.envsoft.2020.104802

Yang, T., Gao, X., Sorooshian, S., & Li, X. (2016). Simulating California reservoir operation using the classification and regression-tree algorithm combined with a shuffled cross-validation scheme. *Water Resources Research*, *52*, 1626–1651. https://doi.org/10.1002/2015WR017394

Yang, T., Zhang, L., Kim, T., Hong, Y., Zhang, D., & Peng, Q. (2021). A large-scale comparison of Artificial Intelligence and Data Mining (AI&DM) techniques in simulating reservoir releases over the Upper Colorado Region. *Journal of Hydrology*, *602*. https://doi.org/10.1016/j.jhydrol.2021.126723

Yassin, F., Razavi, S., Elshamy, M., Davison, B., Sapriza-azuri, G., & Wheater, H. (2019). Representation and improved parameterization of reservoir operation in hydrological and land-surface models, 1–30.

Yates, D., Sieber, J., Purkey, D., & Huber-Lee, A. (2005). WEAP21—A Demand-, Priority-, and Preference-Driven Water Planning Model. *Water International*, *30*(4), 487–500. https://doi.org/10.1080/02508060508691893

Zagona, E. A., Fulp, T. J., Shane, R., Magee, T., & Goranflo, H. M. (2001). RIVERWARE: A GENERALIZED TOOL FOR COMPLEX RESERVOIR SYSTEM MODELING1. *JAWRA Journal of the American Water Resources Association*, *37*(4), 913–929. https://doi.org/https://doi.org/10.1111/j.1752-1688.2001.tb05522.x

Zhao, Q., & Cai, X. (2020). Deriving representative reservoir operation rules using a hidden Markov-decision tree model. *Advances in Water Resources*, *146*(September). https://doi.org/10.1016/j.advwatres.2020.103753

Zomer, R. J., & Trabucco, A. (2022). *Global Aridity Index and Potential Evapo-Transpiration (ET 0) Database v3*. Retrieved from https://cgiarcsi.community/2019/01/24/global-