









## Data-driven modeling of municipal water system responses to hydroclimate extremes

Ryan Johnson <sup>a,\*</sup>, Steven John Burian <sup>a,b</sup>, Carlos Anthony Oroza <sup>c</sup>, James Halgren<sup>a</sup>, Trevor Irons <sup>c,d</sup>, Danyal Aziz <sup>b</sup>, Daniyal Hassan <sup>c</sup>, Jiada Li <sup>e</sup>, Carly Hansen <sup>f</sup>, Tracie Kirkham<sup>g</sup>, Jesse Stewart<sup>g</sup> and Laura Briefer<sup>g</sup>

<sup>a</sup> Alabama Water Institute, University of Alabama, Tuscaloosa, Alabama, USA

<sup>b</sup> Civil, Construction and Environmental Engineering, University of Alabama, Tuscaloosa, Alabama, USA

<sup>c</sup> Civil and Environmental Engineering, University of Utah, Salt Lake City, Utah, USA

<sup>d</sup> Montana Technical University, Butte, Montana, USA

<sup>e</sup> Civil and Environmental Engineering, Colorado State University, Fort Collins, Colorado, USA

<sup>f</sup> Oak Ridge National Laboratory, Oak Ridge, Tennessee, USA

<sup>g</sup> Salt Lake City Department of Public Utilities, Salt Lake City, Utah, USA

\*Corresponding author. E-mail: rjohnson18@ua.edu

 RJ, 0000-0003-2195-739X; SJB, 0000-0002-5278-3188; CAO, 0000-0001-5522-7665; TI, 0000-0003-3547-9341; DA, 0000-0002-0734-6894; DH, 0000-0001-8812-7230; JL, 0000-0002-7238-5101; CH, 0000-0002-1288-2632

### ABSTRACT

Sustainable western US municipal water system (MWS) management depends on quantifying the impacts of supply and demand dynamics on system infrastructure reliability and vulnerability. Systems modeling can replicate the interactions but extensive parameterization, high complexity, and long development cycles present barriers to widespread adoption. To address these challenges, we develop the Machine Learning Water Systems Model (ML-WSM) – a novel application of data-driven modeling for MWS management. We apply the ML-WSM framework to the Salt Lake City, Utah water system, where we benchmark prediction performance on the seasonal response of reservoir levels, groundwater withdrawal, and imported water requests to climate anomalies at a daily resolution against an existing systems model. The ML-WSM accurately predicts the seasonal dynamics of all components; especially during supply-limiting conditions ( $KGE > 0.88$ ,  $PBias < \pm 3\%$ ). Extreme wet conditions challenged model skill but the ML-WSM communicated the appropriate seasonal trends and relationships to component thresholds (e.g., reservoir dead pool). The model correctly classified nearly all instances of vulnerability (83%) and peak severity (100%), encouraging its use as a guidance tool that complements systems models for evaluating the influences of climate on MWS performance.

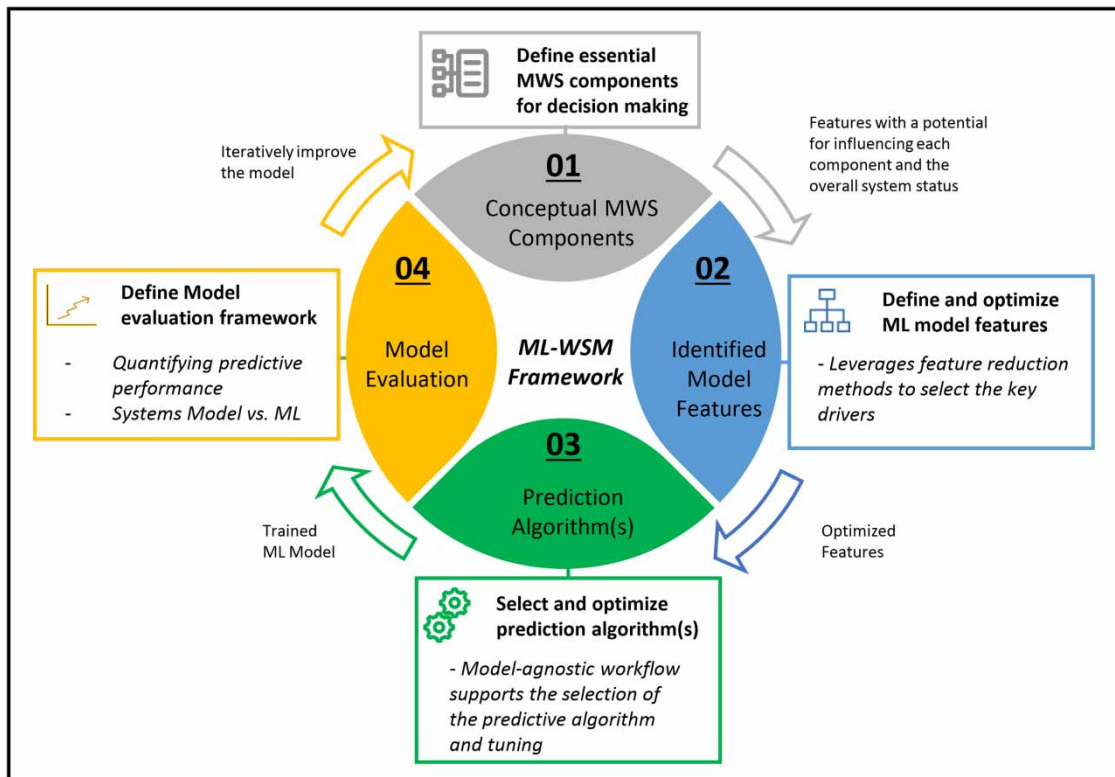
**Key words:** data-driven modeling, machine learning, municipal water system, water system climate vulnerability, XGBoost

### HIGHLIGHTS

- Machine learning models capture water system response to climate-driven supply and municipal demand.
- Machine learning can bypass the high parameterization and development challenges associated with systems models.
- Predictions of water system component status from the XGBoost algorithm can produce representative estimates of reliability, vulnerability, and severity to support management decision-making.

This is an Open Access article distributed under the terms of the Creative Commons Attribution Licence (CC BY-NC-ND 4.0), which permits copying and redistribution for non-commercial purposes with no derivatives, provided the original work is properly cited (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

## GRAPHICAL ABSTRACT



## 1. INTRODUCTION

Western US municipal water system (MWS) management makes critical operational decisions based on the estimated volume of winter water storage (i.e., snowpack), projected surface water yields, and the anticipated timing and magnitude of water demands. The accumulation of winter precipitation (e.g., snow-water-equivalent (SWE)) functions as a high-elevation storage reservoir, driving groundwater recharge through hydrological mechanisms and providing the greatest surface water yield during spring snowmelt (April through early July). Municipal water demand exhibits an opposing seasonal pattern with minimal demand during spring and a high peak during the mid-to late-summer as outdoor water use for irrigation increases to mitigate greater evapotranspiration (Monteith 1965; Shuttleworth *et al.* 2009; Lhomme *et al.* 2015). The seasonal variability in surface water supply, year-to-year fluctuations in annual yield, persistent summer drought, and dynamic climate-demand interactions require active monitoring and reliable projections of key MWS components (e.g., reservoir levels) to inform management decisions, particularly, in a changing climate (Purkey *et al.* 2007; Harpold *et al.* 2012; MacDonald *et al.* 2012; Barnett *et al.* 2019; Johnson *et al.* 2021; Jennings *et al.* 2022; Wlostowski *et al.* 2022).

Building climate resilience in MWS planning and management requires a standardized platform that is capable of characterizing system performance, highlighting potential vulnerabilities, and establishing a framework to gauge the degree of potential improvements from different strategies. Hashimoto *et al.* (1982) applied the concepts of reliability, resilience, and vulnerability (RRV), based on predetermined thresholds, to serve as a standardized protocol for evaluating reservoir performance to various conditions and assisting in the evaluation and selection of alternative design. The framework defines reliability as the probability of the non-exceedance of the threshold, resilience as the speed of recovery from an exceedance event (e.g., the number of days to return to a specified reservoir level once exceeded), and vulnerability as the severity of an exceedance event (i.e., the magnitude of a failure). The introduction and evolution of the RRV framework establish a platform to evaluate many aspects of the MWS to a range of alternative futures, including seasonal to decadal projections of surface water supplies, estimates of demand, and the operations and development of infrastructure (Makropoulos *et al.* 2018; Nikolopoulos *et al.* 2019). By modeling the MWS with a system modeling approach and evaluating the outcomes with an

RRV assessment, researchers, planners, and managers can characterize the projected system performance and explore solutions to mitigate vulnerabilities (Wang & Blackmore 2009; Füssel 2010; Goharian *et al.* 2017; Goharian & Burian 2018).

Models reflecting water system operations play a significant role in the planning, management, and design of water resource systems (Reuss 2003). Among water systems models, the systems modeling framework establishes a foundation to create a digital representation of the MWS, aiding in the understanding of the feedbacks and interactions between components as well as external influences. The application of a systems model framework to an MWS needs to include all key components that influence the overall performance, including but not limited to infrastructure connectivity, institutional and policy actions, and component capacity limitations to fit the conceptual model (Gastelum *et al.* 2008). Systems models use the continuity equation to define the intrasystem interactions, accounting for system-wide mass balance changes that describe the observed cause–effect relationships (Winz *et al.* 2008; Gastelum *et al.* 2009; Madani & Mariño 2009). While complex in model size and development, a systems model simplifies the hydrological system to the primary physical drivers, which must be explicitly defined to reflect real-world operations (Antunes *et al.* 2018; Jaiswal *et al.* 2020).

Given the systems modeling framework is versatile and applicable to a suite of water resource applications, MWS modeling brings a unique set of challenges. Fu *et al.* (2022) identifies multiple obstacles in the development of systems models for urban water system applications: (1) the complexity of MWS and interactions with ecosystems and climate systems (behaviors and cascading impacts) is particularly difficult to accurately capture, (2) there remains great difficulty in determining modeling assumptions, various processes and model structures, and calibrating a large number of model parameters, (3) the human resources and skills required challenges model development, and (4) the models are system-specific preventing the transferability from one MWS to another. Parameters include but are not limited to, the capacities of water treatment facilities, maximum flow rates of water transfer infrastructure, storage and operations of reservoirs, and overall connectivity of the system (Goharian *et al.* 2017). Calibration refers to the manual adjustment of parameters to reflect a test condition from observations. The calibration process may adjust aqueduct lengths, diameters, slope, or Manning values to reflect observations surrounding lag time, velocity, and/or volume. In a complex system, although simplified, there are often several sources, a variety of users (e.g., domestic, commercial, industrial), multiple reservoirs, different pressure zones, source prioritization schemes, and connectivity. Accurately integrating all of the key features to describe the real-world system interactions, high model complexity, an extensive period of development, and intrinsic limitations present obstacles to model development and have led to a stall in research capabilities (Marçais & de Dreuzy 2017; Jaiswal *et al.* 2020).

The effective use of extensive and semantically connected data describing systems demonstrates the potential to transform the modeling paradigm (Jadidoleslam *et al.* 2019; Fu *et al.* 2022). The exploration of data-driven models stems from the ability to simulate feedbacks and interactions between MWS components without an a priori understanding of the dominant driving mechanisms or interconnections (Kalin *et al.* 2010; Sarkar & Pandey 2015). McCuen (2016) found that data-driven machine learning (ML) can model hydrological change without explicit knowledge of the system, accelerating the development of models targeting water quality compared to systems models. Rather than parameterizing and calibrating each unique process or component (e.g., inputting flow capacities, reservoir volumes, travel time), ML elicits useful criteria and trends derived directly from data during training to determine and optimize internal parameters (Mahmoudi *et al.* 2016; Noori *et al.* 2020). Bypassing the need to define every component and interaction within the MWS makes ML approaches appropriate where the objective is to model behavior or outcomes of a system rather than to explicitly characterize the interconnected physical processes (Shen 2018). The demonstrated success in model skill results in the transdisciplinary application of ML to serve as a highly performant decision-making tool and an alternative to systems models (Ma *et al.* 2019; Haskins *et al.* 2020).

Applications using data-driven approaches to model hydrological systems indicate accurate predictions that benefit decision-making despite reduced computational and developmental complexities compared to systems models. Aghelpour & Varshavian (2020) used multilayered perceptron (MLP) networks to model daily Zilakirud River flows in northern Iran with high levels of accuracy during wet and dry years. The model operates as a flood warning system to trigger evacuation measures and mitigation actions, reducing the cost of life and damages. Mohammadi *et al.* (2020) applied support vector regression, random forest (RF), principal component analysis (PCA), and a grey wolf optimization algorithm to forecast monthly Lake Titicaca fluctuations in water level with low error. The framework supports the optimization of water storage for drinking, the production of hydroelectric power, and the balancing of beneficial water use practices concerning environmental, agricultural, and industrial users. Using an artificial neural network (ANN) and fuzzy analytic hierarchy process, Imani *et al.* (2021) predicted the resilience of water quality in São Pablo, Brazil to identify basins with urgent needs for

remediation. Rozos (2019) developed a reservoir optimization model built using a feedforward neural network, providing an array of options to mitigate contemporaneous system-compromising externalities to enhance the management of urban water resources. While previous research explores the expansion of ML approaches throughout many aspects of water resources, the application of ML to inform MWS decision-making to climate anomalies has not been studied to date.

The interactions between infrastructure, operations, and climate strongly influence the performance and vulnerabilities of an MWS. While systems modeling can capture MWS interactions and feedbacks, high parameterization, an immense period of development, and the simplification of complex hydrological processes hinder its widespread adoption. ML approaches demonstrate the capacity to address the limitations of systems models for MWS applications but have not been applied or investigated as a tool to support decision-making. We develop the Machine Learning Water Systems Model (ML-WSM) as a novel application of ML to address the research gap, investigating the achievable level of performance by ML approaches for modeling the responses of key MWS components to variations in supplies and demands driven by climate anomalies. Recognizing the heterogeneity of MWS, we design the ML-WSM as a modular and model-agnostic workflow to extend its application to any MWS with ample data, predict key components of the MWS, and coupled with a vulnerability assessment to support decision-making. We apply the ML-WSM framework to the Salt Lake City Department of Public Utilities (SLCDPU) water system in Utah, where the ML-WSM predicts the response of reservoir levels, groundwater withdrawal, and imported water use at a daily temporal resolution to extreme dry through wet climate conditions, benchmarking model performance to an existing systems model.

## 2. METHODS

We develop the ML-WSM as a generalizable ML framework to reduce the barriers to entry for evaluating the MWS response to externalities compared to a systems model. The framework consists of outlining the conceptual workflow (Section 2.1); model inputs, incorporation of system connectivity, and methods for feature optimization (Section 2.2); algorithm selection (Section 2.3), and evaluating the model (Section 2.4). Section 2.5 describes the coupled vulnerability assessment to gauge the projected reliability and vulnerability of the key MWS components.

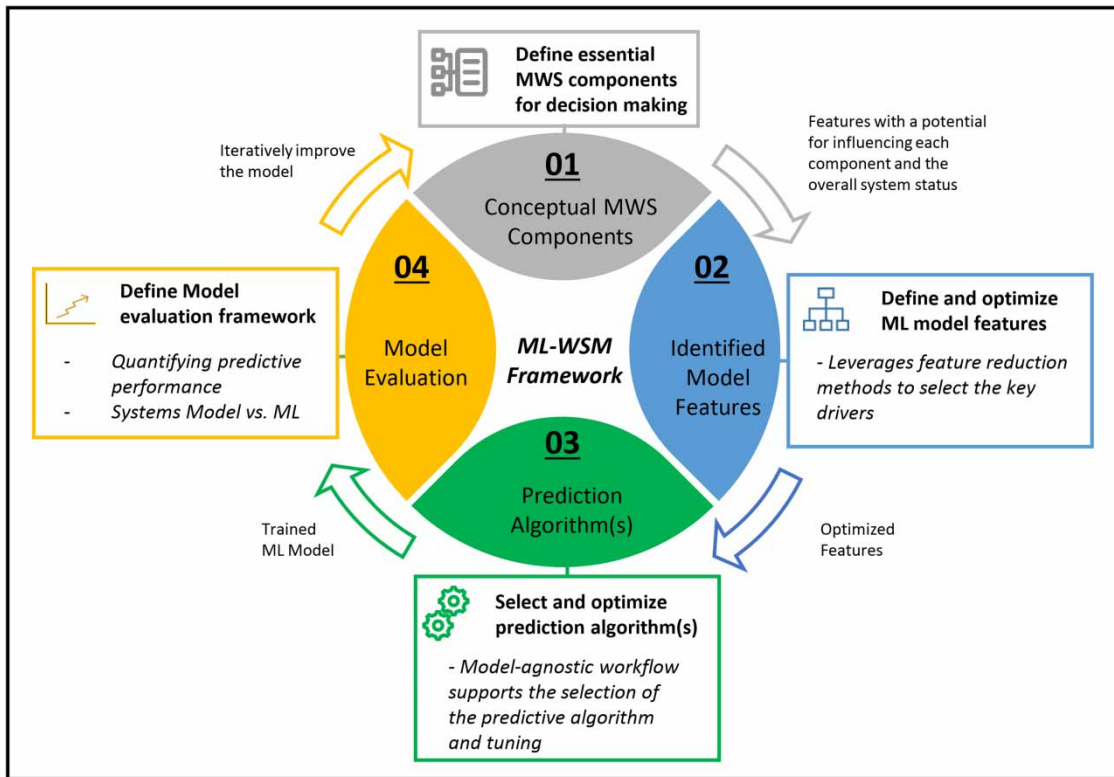
### 2.1. Conceptual workflow of the ML-WSM

The motivation of the ML-WSM is to streamline the development of an MWS model using a broadly adaptable system-agnostic workflow, assuming that system-specific features, supplies, demands, and infrastructure, while unique to each system, remain generic drivers of performance. The conceptual workflow of the ML-WSM initiates by (1) the user determining conceptual components of the MWS to model (e.g., reservoir levels, imported water requests, groundwater withdrawal), (2) identifying components (features) that influence the components and the overall system (e.g., other components, supply availability, demands), (3) investigating ML algorithms, and (4) evaluating model performance (Figure 1). The workflow encourages the iterative evaluation of different formulations to optimize ML-WSM performance.

Defining the goals of the ML-WSM will guide the development process and assist in identifying key MWS components to model. The goals can consider several forecast horizons and temporal resolutions that uniquely aid in planning and management guidance, such as sub-daily to support daily operations (e.g., peak daily MWS performance), daily to guide monthly to seasonal supply and demand management (e.g., drought contingency planning), or annual to inform long-term infrastructure development, and/or prepare for growth. The extent of the forecast horizon depends on the availability of model inputs (e.g., estimates of streamflow or climate) to develop an effective model. If the goal is to model sub-daily water system performance but only daily data is available, a data-driven model may not be appropriate because the available data does not match the modeling goals. If daily resolution data is available and the goal is to assess water system performance to seasonal climate variations, the ML-WSM could meet the expected modeling goals. We further recommend connecting the effective forecast horizon to prescribed levels of decision reliability quantified by the error bound, i.e., the largest difference between the optimal decisions made under any two climate scenarios (Zhao *et al.* 2019).

Aligning with the modeling goals, the user needs to identify specific MWS components of interest. Examples include reservoirs for storage, sustainable groundwater yields, imported water allotments, and multiple sectors of water use that define MWS performance and are of operational importance to management. If a component is important to decision-making, then it is an essential component of the conceptual framework.

Developing a representative ML-WSM depends on identifying the optimal features of each modeled MWS component to leverage the power of ML to bypass the manual calibration procedures of systems models. Selecting features containing the



**Figure 1** | The conceptual workflow (1) determines the key municipal water system (MWS) components, (2) identifies features that influence the key components, (3) selects ML algorithm(s) for evaluation, and (4) evaluates model performance to iteratively improve the model.

embedded intrasystem relationships for each MWS component minimizes prediction errors, reduces model complexity, and removes features that can be detrimental to performance (e.g., collinearity) and increase prediction uncertainty (Dormann *et al.* 2013; Sit *et al.* 2020). For example, a systems model requires the inflow to reservoir A, reservoir A levels, reservoir A level-capacity rating curve, reservoir A release rates, reservoir B levels, reservoir B level-capacity rating curve, and reservoir B release rates to model the reservoir interactions. A data-driven model may only require reservoir B levels or reservoir A inflow as these components may have all other system components embedded within the data. Section 2.2 provides a deeper perspective into feature considerations, including the availability of data and the temporal resolution to support the intended use of the model.

Modeling the MWS with ML can provide flexibility in model architecture that can adapt to the system of interest and/or preference of the developer. The ML-WSM can be as simple as a single ML model to predict many outputs (e.g., reservoir levels and groundwater withdrawal) or as complex as several interconnected submodules to predict individual components of a system (e.g., one for reservoir A, one for reservoir B, and one for groundwater withdrawal). A multimodel approach supports one-to-one and one-to-many relationships between MWS components, provides a platform to evaluate a variety of input features, and explores different algorithm types (Section 2.3) to optimize each submodule. Model configuration and development can leverage existing ML pipelines for algorithm optimization (e.g., grid search parameter optimization) and training (e.g., training/testing splits, hold-one-out) (Garreta *et al.* 2017).

Model evaluation forms the final step of the conceptual workflow, where Section 2.4 describes a sample of evaluation metrics that the developer can tailor to the respective problem. The ML-WSM workflow encourages iterative model development, where the developer engineers and tests new features, assesses system performance to different levels of feature collinearity, investigates different methods of feature selection, and explores a variety of ML algorithms to improve and adapt the ML-WSM framework to decision-making goals.

## 2.2. Feature engineering and selection

The modeled MWS components and the overall objectives will guide the feature development processes. Data availability is a strong determinate and guiding factor because there must be ample data at the appropriate temporal resolution to support model training, and by proxy, predictions (Ficchi *et al.* 2016; Sunkara & Singh 2022). Temporally relevant features are critical for developing ML models as the models exhibit a direct link between feature optimization and prediction skill (Chandrashekar & Sahin 2014; Li *et al.* 2017), and exhibit a preference for modeling at higher temporal resolutions because of the overall increase in training data quantity (Eggimann *et al.* 2017). Developing a model at a higher temporal resolution supports the upscaling of prediction (e.g., daily to monthly) as opposed to downscaling (e.g., monthly to daily). The optimal resolution will be a balance between the availability of data, dimensionality, and computational efficiency. The feature development process consists of two generalizable steps: (1) feature engineering and (2) feature selection that prepare, transform, construct, and filter features to optimize model performance (Sun *et al.* 2020; Wang *et al.* 2022).

Feature engineering should develop features describing water system component feedbacks and interactions. With the performance and resilience of arid MWS subject to reservoir level(s) (Goharian *et al.* 2017), groundwater withdrawal (Moghaddasi *et al.* 2022), and the volume of imported water requests (Mukheibir 2008) as well as influencing factors such as the availability of surface water (streamflow if used for an MWS), municipal water demand (water use across all sectors), hydroclimate conditions (temperature, precipitation, evapotranspiration), socioeconomic factors (population, number of households), and/or the time of year (day of year or month), it is essential to develop dynamic features at the respective time step representing these influences on the water system (Sun & Scanlon 2019). Feature engineering may require application-specific data processing methods, such as gap filling in time series data or scaling observations to match the desired temporal resolution (Rebora *et al.* 2016; Dembélé *et al.* 2017; Arriagada *et al.* 2021). Features describing the connectivity of the MWS can improve model performance and we recommend exploring MWS components from the previous timestep as features for the prediction of MWS components, as they can support the memory of initial conditions and interactions into the model (Långkvist *et al.* 2014; Hu *et al.* 2018; Moishin *et al.* 2021). For example, reservoir A levels from the preceding timestep (e.g., July 1) could be a feature of reservoir B levels (e.g., July 2). While ML-WSM development encourages the exploration of many water system components as influencing features, the feature space can become increasingly large and subject to the curse of high model dimensionality (Castelletti *et al.* 2010).

The goal of feature selection is to optimize the features of each model, as it has a direct connection to the performance of the model, reduces model dimensionality, improves learning accuracy, and facilitates conceptual model understanding (Cai *et al.* 2018; Sit *et al.* 2020). A general recommendation is to address any collinearity among features and then apply a feature optimization algorithm (Yan & Zhang 2015). Collinearity describes a condition where two or more features exhibit a linear relationship with another (or multicollinearity) and can become more significant with an increasingly larger feature space (Alin 2010). The removal of collinearity increases the variance within the features to increase the strength of the prediction-target relationships (Guyon & Elisseeff 2003; Cai *et al.* 2018). We recommend a variable inflation factor (VIF) test to address collinearity among features where the VIF determines the strength of the correlation by regressing a feature against all other features (Akinwande *et al.* 2015)

$$\text{VIF} = \frac{1}{(1 - R^2)} \quad (1)$$

where  $R$  is the Pearson correlation coefficient of two variables defined by

$$R = \frac{\sum_i (f_i - \bar{f})(y_i - \bar{y})}{\sqrt{\sum_i (f_i - \bar{f})^2} \sqrt{\sum_i (y_i - \bar{y})^2}} \quad (2)$$

where  $f_i$  and  $y_i$  are features within the feature space. From Equation (1),  $R^2$  values closer to 1 between two features will result in a higher VIF value. VIF values exceeding 10 indicate high levels of collinearity among features and we recommend removing the features displaying the lower correlation to the target (Menard 2002; Chatterjee & Simonoff 2013). A VIF of 10 should

not be viewed as a critical threshold (O'Brien 2007) but as a development recommendation as there is value in features describing MWS connectivity.

Feature selection methods aim to reduce model dimensionality to ultimately improve model performance, whether through a reduction in the total number of features or combining the information embedded between many features into fewer model inputs (Keogh & Mueen 2017; Jia *et al.* 2022). Common methods include PCA, LASSO regularization, recursive feature elimination (RFE), and auto-encoders. PCA functions as a statistical analysis method that transforms several features into a few integrated features reflecting the information contained in the original set of features (Moore 1981). LASSO regularization penalizes features of a model to a coefficient value of zero that are of minimal modeling significance, with the non-zero coefficient features being key model predictors (Muthukrishnan & Rohini 2016). The RFE algorithm prioritizes dimensionality reduction through the identification of strong predictors from the complete feature space to improve both model skill and minimize model complexity, removing noisy and non-informational features (Chen & Jeong 2007; Toloşi & Lengauer 2011). An autoencoder is a type of neural network that learns a compressed representation of the original feature space, commonly referred to as a bottleneck, where the autoencoder ingests the original feature space and the output of the model at the bottleneck functions as the input into the modeling algorithm (Wang *et al.* 2017; Han *et al.* 2018). We summarize the benefits and limitations of each method in the Supplementary Material. There are many dimensionality reduction methods available and we encourage examining their impact on model skill during development.

### 2.3. ML algorithm selection

ML algorithms determine patterns and relationships embedded within the data between inputs and outputs during the training process rather than the explicit instructions of static programming algorithms. Leveraging the power of ML requires the correct use of and selection of the algorithm (s) for the respective tasks (Raschka 2018; Lee & Shin 2020). ML algorithms can be fit into two main categories: unsupervised and supervised learning. The basis of unsupervised algorithms is that a machine can learn patterns without human guidance, useful for clustering and dimensionality reduction (Hofmann 2001) but does not support regression tasks. Supervised ML algorithms are flexible, comprehensive, and support both classification and regression modeling tasks by identifying general patterns that support predictions from a given set of inputs (Choudhary & Gianey 2017). Supervised learning connects the inputs to a labeled set of outputs, or targets, through extensive data processing (e.g., cleaning, randomizing, and structuring the input and target data) and model training procedures that align with the goals.

Algorithm selection can be challenging as there are many choices and the transferability of the optimal algorithm for one water system may not be ideal for another. Common ML algorithms for water resources modeling include ANNs (Kouziokas *et al.* 2018; Raj & David 2020; Xu *et al.* 2020), recurrent neural networks (RNN) (Kratzert *et al.* 2018; Gangrade *et al.* 2022; Krishnan *et al.* 2022), and decision tree algorithms (Li *et al.* 2022; Wu *et al.* 2022; Yusri *et al.* 2022). ANNs consist of a feed-forward network utilizing three types of layers: an input layer, middle hidden layers that perform the computational tasks, and an output layer with the prediction. Commonly used ANNs for water resources include MLP and Extreme Learning Machines. RNNs are a type of ANN that excel at time series modeling applications, demonstrating a memory-like capability by using prior inputs of a sequence to influence the predictions. The Long Short-Term Memory algorithm is a popular RNN. Decision tree learning imitates the human decision-making process with the model prediction pathway emulating the appearance of a tree (e.g., if-else statements), supporting greater interpretability of model architecture compared to other ML algorithms. Xtreme Gradient Boost (XGBoost), RF, and Light Gradient Boosted Machine are common decision tree algorithms within water resources. We summarize the benefits and limitations of these ML algorithms in Supplementary Material, Table S1. We recommend a review of the contemporary applications of supervised ML algorithms throughout water resources management (Choudhary & Gianey 2017; Tyrallis *et al.* 2019; Ghobadi & Kang 2023) and exploring multiple ML algorithms during the development process.

### 2.4. Model evaluation

Evaluation provides the foundation for benchmarking the performance of a model, with a recommendation to use 10–30% of the training data to evaluate model performance (training the ML model on 70–90% of available data) (Dao *et al.* 2020; Pham *et al.* 2020). We use standard metrics within water resources for determining model performance, including the root-mean-square error (RMSE), Kling-Gupta Efficiency (KGE), and Percent Bias (PBias) (Equations (3)–(5)). RMSE is the default measure of the physical differences between predicted and observed as a function of the square root of the second sample

moment or quadratic mean.

$$\text{RMSE} = \sqrt{\text{MSE}} = \sqrt{\frac{\sum_{i=1}^n (y_i - f_i)^2}{n}} \quad (3)$$

where  $y_t$  are the observed values,  $f_t$  are the daily forecasted values, and  $n$  are the number of predictions (i.e., timesteps) in the observed time series. The *KGE* is an expression of the distance between the point of ideal model performance in the space described by its three components, (1) correlation, (2) variability, and (3) bias (Gupta *et al.* 2009). Values approaching 1 indicate a perfect model fit between predictions and observations, and a benchmark for performance is  $-0.41$ , as values greater than this exhibit performance greater than the mean (Knoben *et al.* 2019).

$$\text{KGE} = 1 - \sqrt{(\rho - 1)^2 + (\beta' - 1)^2 + (\alpha - 1)^2} \quad (4)$$

The *PBias* computes the average amount that the observed is different than predicted as a percentage of the observed, communicating if, on average, the model favors predictions above ( $-$  bias) or below ( $+$  bias) the observed. Ideal *PBias* values are close to zero, with positive (underpredicting) or negative (overpredicting) values indicating the model prediction tendency.

$$\text{PBias} = \frac{1}{n} \sum_{i=1}^n \frac{(y_i - f_i)}{|y_i|} \quad (5)$$

Each evaluation metric characterizes model performance differently. The RMSE conveys error in component units; KGE expresses in a single metric the similarity between observed and simulated from correlation ( $r$ ), Bias ratio ( $\beta$ ), and the variability ratio ( $\gamma$ ); and PBias measures the average tendency ( $\pm\%$ ) of the predicted values relative to the observed.

## 2.5. Vulnerability metrics

We couple MWS performance evaluation tools to model outputs for assessing vulnerabilities and gauging performance relative to historical component observations. While the vulnerability assessment can be on any component, the results are most meaningful when these components are strong indicators of overall MWS performance (Goharian *et al.* 2017; Nikolopoulos *et al.* 2019). The simulated time series of the key components form the initial input.

$$X_t; \quad t = 1, 2, \dots, T \quad (6)$$

where  $X_t$  is the status of each component at time  $t$ , and  $T$  is the time period of the analysis. Using the time series of predictions, we calculate the system performance index (SPI)

$$\text{SPI} = f(X_t); \quad t = 1, 2, \dots, T \quad (7)$$

where the SPI value is either one or zero depending on whether the prediction exceeds a threshold at each timestep ( $f(Z_t)$ ). A threshold establishes a measure of comparison that defines and differentiates satisfactory (S) and unsatisfactory (U) states of the component for each timestep (e.g., the half-full capacity of a reservoir). Using the threshold, the calculation of SPI is

$$\text{SPI} = f(Z_t) \quad t = 1, 2, \dots, T \quad \text{and} \quad \begin{cases} Z_t = 1 & X_t \in S \\ Z_t = 0 & X_t \in U \end{cases} \quad (8)$$

From the SPI, we can calculate the RRV of the MWS.



### 2.5.1. Reliability

Reliability describes the frequency of a component operating in a satisfactory state

$$\alpha = \frac{\sum_{t=1}^T Z_t}{T} = 1 - \left(\frac{n_f}{T}\right) \quad (9)$$

where  $\alpha$  is the estimate of reliability, and  $n_f$  is the number of unsatisfactory timesteps in the simulation ( $T$ ). Reliability is at the temporal resolution of the simulation, with values close to 1 indicating optimal levels and values close to 0 indicating compromised levels.

### 2.5.2. Vulnerability

Reliability cannot fully describe the behavior of the MWS, as the degree of unsatisfactory conditions is a critical decision-making component (Sandoval-Solis *et al.* 2011; Asefa *et al.* 2014). Vulnerability communicates the magnitude of unsatisfactory conditions for a component as a function of exposure and severity.

$$\text{Vulnerability} = f(\text{exposure, severity}) \quad (10)$$

Exposure is the state of the unsatisfactory condition of each component due to externalities ( $e$ ).

$$\text{WSCI}_e = 1 - \frac{\text{WSC}_e}{\text{WSC}_H} \quad (11)$$

allowing for the calculation of the MWS component (WSC) index to externalities ( $\text{WSCI}_e$ ) by comparing the magnitude of the severity of the component during the simulation ( $\text{WSC}_e$ ) to the historical record observations ( $\text{WSC}_H$ ). The  $\text{WSCI}_e$  varies from 0 to 1, with values closer to 1 representing increased vulnerability and 0 conveying the opposite with respect to historical conditions. The historical period of observation or simulation determines each  $\text{WSC}_H$  while each scenario of various externalities has a unique  $\text{WSC}_e$ .

Severity describes the average magnitude of unsatisfactory conditions of each component under scenarios of external influences. The calculation of this metric is

$$S = \sum s_t * e_t \quad X_t \in U \quad (12)$$

where  $s_t$  quantifies the severity of unsatisfactory conditions at time  $t$ , and  $e_t$  is the occurrence probability of  $X_t$  (in the form of  $s_t$ ) as the most severe result from a set of unsatisfactory states. The average vulnerability of the system is a function of exposure and severity

$$\text{Vulnerability} = \text{WSCI}_c * \beta_{\text{WSC}} + S * \beta_S \quad (13)$$

where  $\beta_{\text{WSC}}$  and  $\beta_S$  weights communicate differences in the importance of exposure and severity on the overall vulnerability of a component.

### 2.5.3. Peak severity

Peak severity characterizes the maximum magnitude of unsatisfactory conditions of each component during a simulation. The calculation of this metric is

$$S = \max(s_t) \quad X_t \in U \quad (14)$$

where  $s_t$  is the severity of unsatisfactory conditions at time  $t$  when  $X_t$  (in the form of  $s_t$ ) is unsatisfactory. The metric communicates the most severe result of unsatisfactory states for a simulation, an indicator illustrating the severity of the system compared to historical observations.

#### 2.5.4. Improving the interpretability of the vulnerability assessment

The RRV values of 0–1 provide minimal operational guidance and to provide a useful tool for system management, we apply the Jenks classification algorithm to categorize the level of vulnerability and severity. Jenks classification minimizes the average deviation within each category while maximizing the deviation from the means of other categories (Jenks 1967). We suggest three categories ranging from Category 1 (Low) to Category 3 (High), connecting the simulated performance of each component to historical levels of vulnerability and severity. The categorized vulnerability results relate model predictions to the historical record.

### 3. APPLICATION OF THE ML-WSM TO A REAL-WORLD MWS

We investigate the utility of the ML-WSM by applying it to the SLCDPU water system and evaluate the performance of the model in three climate conditions. The MWS has an extensive data record to train the ML model and a refined systems model to serve as a performance benchmark (Section 3.1). We follow the conceptual ML-WSM framework to determine the key MWS components and corresponding model input features that are sensitive to changes in surface water supply availability and municipal demands (Section 3.2). Model development describes the selection and evaluation of different ML algorithms and assesses the influence of different features on model performance, exemplifying the iterative development process (Section 3.3). The evaluation scenarios (Section 3.4) describing the three climate scenarios complete the section.

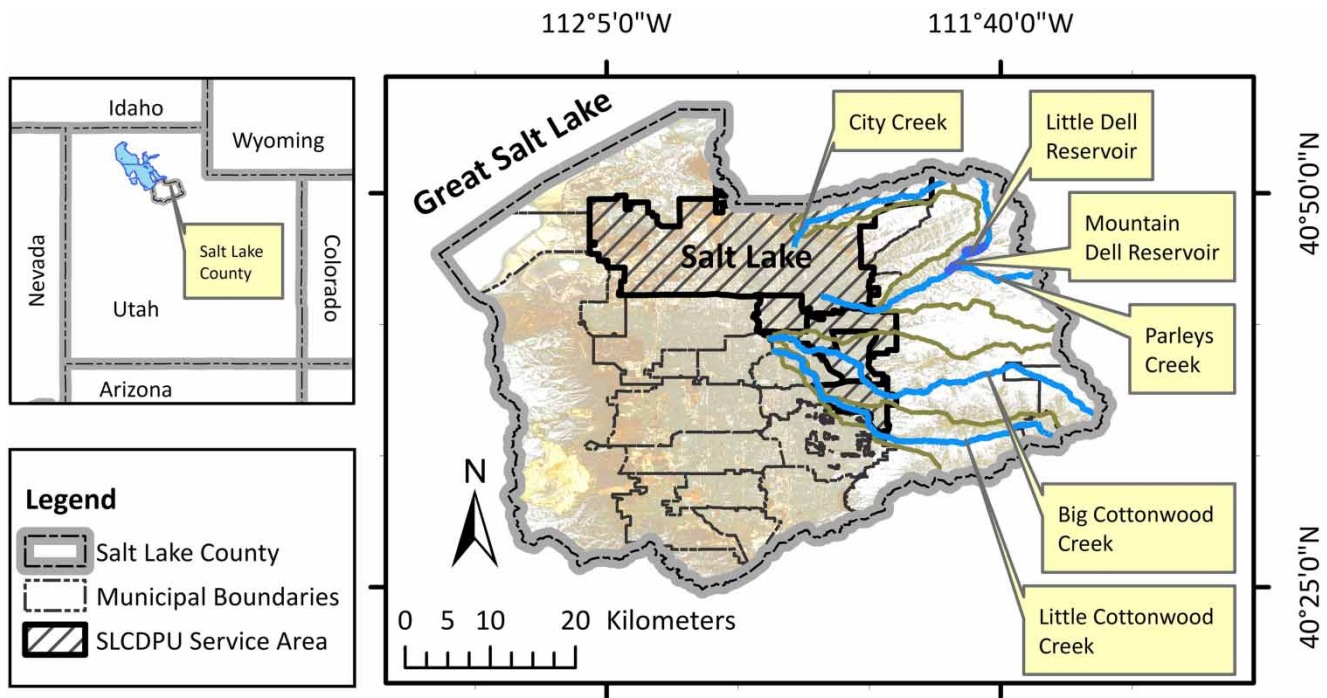
#### 3.1. Study area and systems model

The location of the SLCDPU in Utah shares many similarities with other western US water utilities in growing metropolitan areas. The municipality serves approximately 350,000 people across residential, institutional, and commercial sectors in four cities: Salt Lake City, Mill Creek, Holladay, and Cottonwood Heights. The interannual climate variability and seasonality of the region strongly influence winter snowpack extent and duration, the primary mechanism controlling surface water supplies (Scalzitti *et al.* 2016). The region experiences a cold semi-arid (BSk) climate that determines seasonal water use (Peel *et al.* 2007), outdoor water use can approach 1,000 mm for commercial and residential landscape irrigation from April to October but is negligible from November to March (Collins & Associates 2019). High seasonal water use places Utah as the second or third highest per-capita water use state depending on the year (Dieter 2018). The SLCDPU reports its monthly treated water releases into the distribution system, including leakage and unaccounted system losses, to the Utah Division of Water Rights (UDWR 2023).

The adjacent Wasatch mountain surface water, underlying valley groundwater, and imported water supplies from Deer Creek Reservoir satisfy municipal demands. From the Wasatch mountains, City Creek, Parleys Creek, Big Cottonwood Creek, and Little Cottonwood Creek contribute over 60% of the annual supply, with groundwater and imported water satisfying the remaining ~40% of water demands when surface supplies cannot (Collins & Associates 2019). The Parleys Creek watershed contains Mountain Dell Reservoir and Little Dell Reservoir with a storage capacity of  $3.2 \times 10^6 \text{m}^3$  and  $25 \times 10^6 \text{m}^3$ , respectively, and are the only long-term storage sources owned by the utility. When surface water supplies cannot meet demand, the SLCDPU has access to up to  $22 \times 10^6 \text{m}^3$  per year of sustainable groundwater. If surface and groundwater supplies cannot satisfy demands, imported water from Deer Creek Reservoir supports up to  $61 \times 10^6 \text{m}^3$  per year of supply. Figure 2 illustrates the location of the utility, the proximity of the surface water supplies, and the Dell reservoir system to the service area.

Changing climate conditions and the need for system resilience prompted the SLCDPU to develop a water system model. The Salt Lake City Water Systems Model (SLC-WSM) uses GoldSim modeling software (Goldsim 2013) to examine the impact of changes in surface water availability on system performance and investigate actions to build system resilience (Goharian *et al.* 2017). GoldSim supports submodels and linear programming to replicate the interconnections between different MWS components, demonstrated through investigations of water system response determined by reliability and cost (Lillywhite 2008), water system management decision-making to optimize reservoir operations (Alemu *et al.* 2011), MWS vulnerabilities to climate and population changes (Goharian *et al.* 2017), and the impact of modeled water demand accuracy on water system vulnerabilities during drought conditions (Johnson *et al.* 2021).

The SLC-WSM models the interactions between water system components using the conservation of mass and three governing modules to track water through the MWS at a daily time step: a reservoir operations module, a water supply module, and a water allocation module (Goharian *et al.* 2016, 2017). The reservoir operations module defines the rules regulating the Dell reservoir system, e.g., Little Dell Reservoir releases into Mountain Dell Reservoir and from Mountain Dell Reservoir into

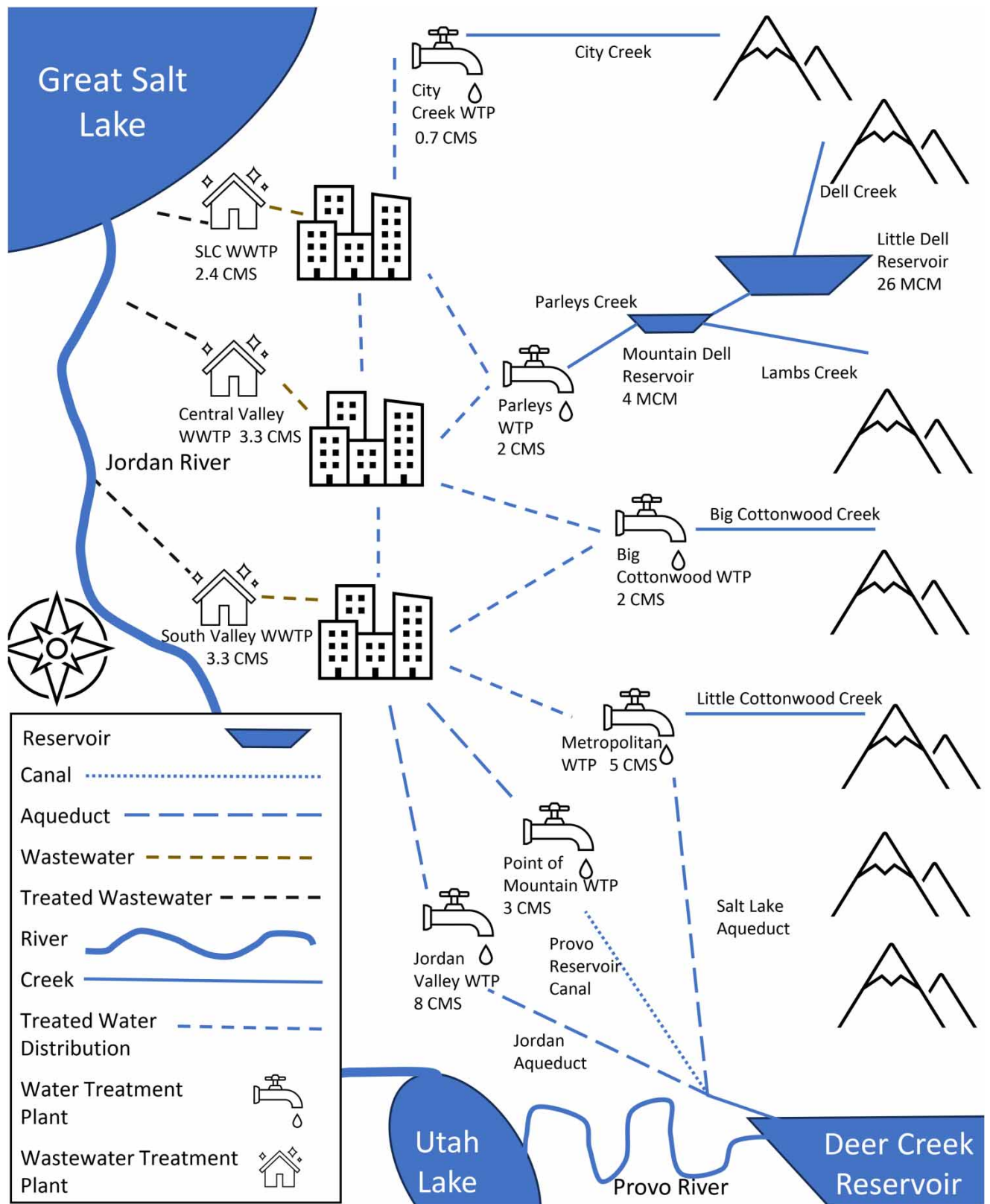


**Figure 2** | The Salt Lake City Department of Public Utilities (SLCDPU) water system depends on the winter snowpack in the adjacent Wasatch mountains for surface water supplies, filling the Dell reservoir storage system, and replenishing valley aquifers.

Parleys Creek and the water treatment plant (WTP). The water supply module uses the conservation of mass to track the movement of water at a given timestep through the water system, including surface water supplies, WTPs, water users and uses (indoor or outdoor), water transfer infrastructure, and wastewater treatment plants (WWTPs). Each component has multiple calibrated parameters to reflect the observed MWS connectivity and operations, such as WTP and WWTP capacities and efficiencies, water transfer infrastructure capacities and efficiencies (e.g., aqueduct capacities), access and limitations to imported water, groundwater pumping rates and sustainable withdrawal limitations, and rules defining sub-service area water users (e.g., annual demand patterns, number of wells per sub-area). [Figure 3](#) illustrates the connectivity of the water infrastructure and the respective capacities.

The water allocation module models the gravity-centric design and the operational structure of the water system ([Goharian et al. 2017](#); [Strong et al. 2020](#)). The gravity-centric architecture governs the allocation of water throughout the service area. For example, Cottonwood Heights, UT in the southwest corner of the service area in [Figure 3](#) has the highest elevation and access to surface water supplies from Little and Big Cottonwood Creeks, a select number of wells, and imported water. In contrast, Salt Lake City, UT in the northern portion of the service area has access to all sources due to its geographical location having the lowest elevation. The water allocation module defines source prioritization, i.e., surface water sources before groundwater withdrawal and imported water from Deer Creek Reservoir. The module initiates groundwater withdrawal when surface water supplies cannot satisfy demands. Imported water requests occur when surface and groundwater supplies (i.e., limited by the number of wells, extraction rates, and annual withdrawal limitations) cannot satisfy demands. The source prioritization scheme optimizes supply sources based on water quality, storage, and cost ([Strong et al. 2020](#)). Imported water from Deer Creek Reservoir is the least prioritized because it is a shared resource among other users (e.g., municipalities and irrigation districts) and is susceptible to harmful algae blooms ([Malmfeldt 2021](#)), requiring additional treatment to achieve acceptable water quality.

The three modules model the movement of water into, through, and out of the SLCDPU water system. [Goharian et al. \(2016, 2017\)](#) provide additional details on the iterative development, calibration, and validation of the SLC-WSM. While the SLC-WSM can replicate the feedbacks and interactions between water system components, there are nearly 4,300 elements that required manual calibration over a 10-year period.



**Figure 3** | The Salt Lake City Water Systems Model (SLC-WSM) models the interconnections, operations, and capacity limitations observed in the SLCDPU water system, including the gravity-centric distribution system and source prioritization structure tracking as water from the sources to demands to discharge.

### 3.2. Key components of the MWS and model inputs

Conceptualization of the ML-WSM begins by identifying key water system components and potentially influential features (Figure 1). The utility is primarily concerned with system vulnerabilities driven by the differential timing of surface water availability and peak municipal demands from April to October. The SLCDPU identified Mountain and Little Dell Reservoir levels, groundwater withdrawal, and imported water from Deer Creek Reservoir as indicators of system performance as part of the Salt Lake City Climate Vulnerability project (Strong *et al.* 2021). The Dell reservoir system is the only long-term water storage within the system, thus monitoring and forecasting levels support management decision-making. Daily pumping rates and sustainable annual yield limit the amount of groundwater the utility can use and water system management will benefit from the projected timing of the sustainable withdrawal threshold. The volume and timing of imported water from Deer Creek reservoir is the most critical indicator of system vulnerability, with estimates supporting proactive vs. reactive management.

Feature identification is the next step in the ML-WSM workflow. We include surface water supply features of City Creek, Parleys Creek, Big Cottonwood Creek, and Little Cottonwood Creek to represent water supply availability and total municipal demand as the primary system influencing features. To represent system connectivity and a temporal connection, we include the previous state ( $t-1$ ) of reservoir levels (% of full capacity), groundwater withdrawal, and imported water requests. If Mountain Dell Reservoir is 90% full today, the proceeding day's prediction should be within a few percentage points of 90% capacity. We include daily surface water supply (combined), day of the year, month, and population to complete the feature development phase.

Feature data comes from multiple sources. The utility provided a near-continuous long-term record of daily streamflow observations (1910–present) from the canyon mouths prior to extensive diversion. We create the total surface water supply feature by combining the flow rates of each supply creek from the streamflow observations. The Utah Department of Water Rights provides the total volume of water entering the distribution system, including all connected demands, leaks, and unaccounted-for losses (UDWR 2023). The Kem C. Gardner Policy Institute provides population estimates for each city within the service area. We use the simulated reservoir levels, groundwater withdrawal, and imported water requests from the SLC-WSM. While observed data would support a comparison between the SLC-WSM and ML-WSM, the motivation is to train the ML-WSM to replicate the SLC-WSM programming and use the SLC-WSM simulations to benchmark the prediction performance of the ML-WSM.

### 3.3. Model development

We develop the ML-WSM using Python v3.10.1 to take advantage of open-source libraries throughout the ML pipeline (e.g., Pandas, Numpy, Scikit-Learn), including feature selection and modeling algorithms. Given the need to identify the optimal drivers for the key MWS components, we begin by removing collinearity among the potential feature inputs. We use the Python Collinearity package (v0.6.1) to streamline the process with a VIF of 10 as recommended in the literature (Menard 2002; Chatterjee & Simonoff 2013), Equation (1). We use RFE within the Scikit-Learn package (v1.0.2) to identify the optimal features for each MWS component (Pedregosa *et al.* 2011). The Scikit-Learn RFE algorithm applies an exhaustive grid search to assign feature importance weights and recursively prunes the number of features via five-fold cross-validation. Preliminary model development investigated the use of PCA for dimensionality reduction but model preference was for interpretable features compared to components, favoring RFE.

The research and development process indicated that a multimodel approach, through the use of submodules for the modeling of each key MWS component, best replicates the interactions and feedbacks present in a complex MWS and we explore MLP and XGBoost algorithms. The MLP algorithm demonstrates functionality in handling large datasets, quickly converging to a solution, and successful application for water resources modeling activities (Kouziokas *et al.* 2018; Raj & David 2020; Xu *et al.* 2020). For each MWS component, MLP development uses mean squared error as the loss function, Adam optimizer ( $1e-4$ ), batch size of 100, 2,000 epochs, and a standardized architecture consisting of 6 hidden layers with the following number of nodes: 128, 128, 64, 64, 32, and 16, respectively. Algorithm training uses 17 years of daily data spanning from 2000 to 2020, omitting the three testing years described in Section 3.4. The MLP algorithm uses a random selection of 75% of the training data for training and performs cross-validation on the remaining 25% of training data to determine model performance.

The XGBoost algorithm optimizes the use of computational hardware and supports the training of large models (Chen & Guestrin 2016), demonstrating high performance in many water resource applications (Xenochristou & Kapelan 2020; Wu

*et al.* 2022; Yusri *et al.* 2022). Hyperparameter optimization is critical for tree-based algorithms, as hyperparameters cannot be estimated from data inputs and influence the performance and speed of prediction (Putatunda & Rama 2018; Probst *et al.* 2019). We use the Scikit-Learn GridSearchCV package to perform an exhaustive grid search across hyperparameter combinations to identify the optimal set for each submodule: objective: [reg:squarederror], learning rate: [0.01–1.0 by 0.05], max tree depth: [3–15, by 5], subsample: [0.6–0.9, by 0.1], column sample by tree: [0.6–0.9, by 0.1], lambda: [0.0–3.0, 0.1], alpha: [0.0–3.0, 0.5], minimum child weight: [2–10, by 1], and the number of estimators: [200–20,000, by 500]. T. Chen & Guestrin (2016) provides a comprehensive description of all hyperparameters. XGBoost model training uses the same 17 years of training data as the MLP model and undergoes three-fold cross-validation.

Iterative development is an essential step for optimizing ML-WSM performance. The initial MLP and XGBoost models used the features identified by the RFE but model performance was less than satisfactory. Feature selection is known to improve model performance (Cai *et al.* 2018) and using the RFE features as a benchmark, we iteratively add and remove features to reflect system connectivity. While diverting from the RFE-identified features could introduce collinearity into the model, we found that adding features that enhanced the conceptual understanding of water system components improved model performance. For example, the RFE algorithm identified the Deer Creek request<sup>-1</sup> and Mountain Dell Reservoir<sup>-1</sup> as the features for modeling Deer Creek request but the predictions did not capture the timing or magnitude of the requests. We iteratively examined several feature sets to reflect the physical influences on Deer Creek requests, with the optimal features for Deer Creek requests, and other MWS components, in Table 1. Integrating features related to the physical influences on Deer Creek requests improved the prediction performance, displayed by a reduction in *RMSE* from  $8.4 \times 10^4 \text{m}^3$  to  $2.5 \times 10^4 \text{m}^3$ . Comparing the final XGBoost and MLP models, the XGBoost models demonstrated greater model performance, exemplified by a reduction in *RMSE* from  $2.5 \times 10^4 \text{m}^3$  to  $1.4 \times 10^4 \text{m}^3$  for the Deer Creek requests model. Figure 4 illustrates the performance differences between the first and final MLP models and the final MLP and XGBoost models of Deer Creek water requests.

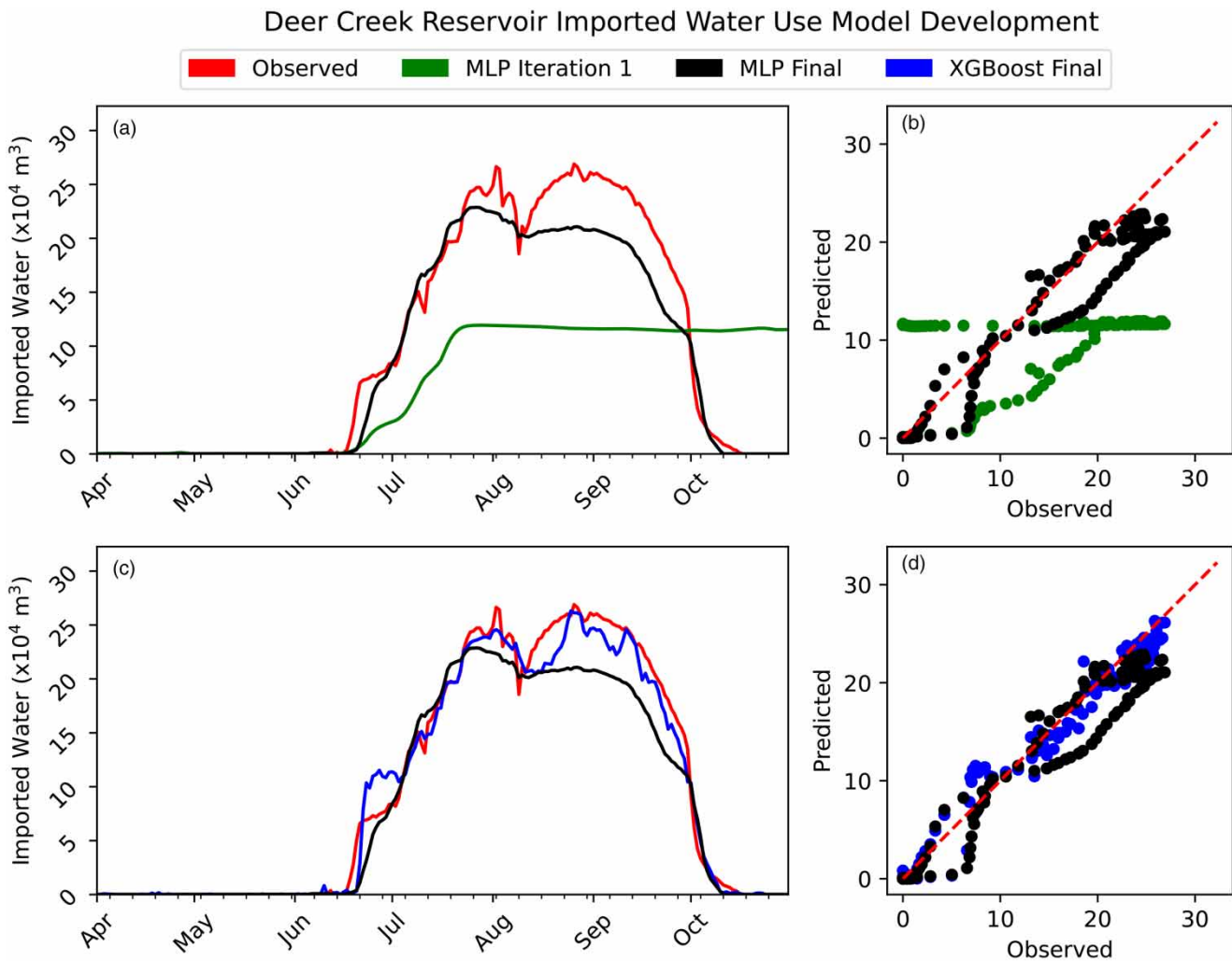
A comprehensive evaluation of each model completes the development of each component model. For two of the models, we discovered physically impossible prediction values. Predictions of groundwater withdrawal exceeded the maximum allowable rate ( $11 \times 10^4 \text{m}^3/\text{day}$ ) and predictions of Deer Creek requests were below  $7.3 \text{m}^3/\text{s}$ , which is the minimum combined flow rate for the Salt Lake Aqueduct, Provo Reservoir Canal, and Jordan Aqueduct (Figure 3). To address the physically impossible values, we constrained the model predictions to the above values if the predictions were above (groundwater withdrawal) or below (Deer Creek requests).

**Table 1** | The iterative ML-WSM development process identified the following features for use in modeling the SLCDPU water system

Features	Little Dell Reservoir	Mountain Dell Reservoir	Groundwater withdrawal	Deer Creek requests
Day				
Month	X			
Population				
Deer Creek Request <sup>-1**</sup>				X
Groundwater Withdrawal <sup>-1***</sup>		X	X	
Little Dell Reservoir <sup>-1***</sup>	X			X
Mountain Dell Reservoir <sup>-1***</sup>	X	X	X	X
Total Municipal Demands**			X	X
Total Surface Supply*		X		
City Creek*			X	
Dell Creek*	X	X	X	
Lambs Creek*		X		
Big Cottonwood Creek*				X
Little Cottonwood Creek*				

We considered all features in the features column for each MWS submodel. Superscript <sup>-1</sup> indicates observations or predictions from previous timesteps.

\*Cubic meters per second (CMS), \*\*Cubic meters per day, \*\*\*Percentage of full capacity (%).



**Figure 4** | The recursive feature engineering (RFE) features in the multilayered perceptron (MLP) model (MLP Iteration 1) miss the timing of and peak Deer Creek imported water use (a) and exhibits a strong underprediction bias (b). The MLP final model incorporates additional features to reflect system connectivity and significantly improves model performance (a and b). The final XGBoost model captures the observed peaks in Deer Creek imported water use (c) with increased accuracy compared to the final MLP model (d).

The final phase in model development is connecting each submodule, necessary for running time series simulations with the ML-WSM. Modeling predicts one time step at a time and by including system components from the previous time step ( $t-1$ ), the predictions of the previous time step influence the predictions for the following time step (e.g., the prediction of Little Dell Reservoir for July 1 forms an input for the prediction of Deer Creek requests for July 2).

All model development activities took place on a 6-core personal computer. The XGBoost models took  $\sim 5.5$  h per component while the MLP models converged in less than 10 min. The long training time for the XGBoost models is due to the comprehensive list of hyperparameters input into the GridSearchCV function. While using GPUs and/or high-performance computers will support faster model development, a personal computer benchmarks the development time for many users. The development times above do not include the time required to process the data into a model training ready format and the duration will depend on the number of training observations. Model development time will likely increase with increases in the data size.

### 3.4. Evaluation scenarios

Model evaluation design should align with the overall modeling goals identified in the conceptualization process. Aligning with the goals of the SLCDPU, we examine the ML-WSM performance to three different snowpack-driven scenarios

describing the climate variability of the region and use the SLC-WSM simulations as the observed. Using the Alta Guard MesoWest weather station located at the headwaters of Little Cottonwood Creek, we identify the most recent dry (2015, 680 cm), average (2017, 1,270 cm), and wet (2008, 1,660 cm) conditions to define the simulations (NOAA 2021). A Log-Pearson Type III analysis of the long-term (i.e., 1945–present) snowfall record indicates the dry year and wet year having a 150-year and 15-year return interval, respectively (Bobee 1975). The three water years determine the respective streamflow for supply and influence on per-capita water use. The 85%/15% training–testing split (17 years for training, 3 years for testing) fits within the recommended testing partitioning of 10–30% (Dao *et al.* 2020; Pham *et al.* 2020). Other system factors such as conservation, policy, and initial reservoir levels remain constant between all simulations to establish a baseline relative to the climate anomaly. We begin the ML-WSM and SLC-WSM simulations with the same status of each component on March 31 to initiate a model run, with the first day of prediction on April 1.

The SLC-WSM and ML-WSM operate at a daily time step and require the downscaling of monthly demand to a daily resolution. We use cubic spline interpolation to iteratively reduce the residual difference between the monthly per-capita demand and the monthly interpolated daily values (Supplementary Material, Figure S1). While the downscaling algorithm does not capture spikes in daily use (e.g., unexpected pipe breaks), seasonal assessments emphasize the long-term MWS performance compared to unexpected short-term events (Goharian *et al.* 2017; Goharian & Burian 2018).

The evaluation procedure includes the methods defining the vulnerability calculations for each MWS component. We use Equation (8) to calculate reservoir, total groundwater withdrawal, and imported water Deer Creek request SPI ( $X_t$ ) as a function of each respective component ( $Z_t$ ) at time  $t$  according to the component thresholds (Table 2). We derive the daily values from SLC-WSM simulations of the observed water demand, supply, and systems operations from 2000–2020 (omitting testing scenarios) to determine the thresholds for groundwater withdrawal and imported water requests. We use the dead pool level of 15% of capacity to form the vulnerability threshold for the Little Dell Reservoir. While the dead pool for Mountain Dell occurs at 25% capacity, reservoir rules adjust outflow rates to prevent a dead pool scenario, even during extreme supply-limited conditions. Given the reservoir operations, we set the unsatisfactory conditions threshold to be 45% of full capacity to allow for a greater degree of Mountain Dell Reservoir vulnerability between scenarios and models, emphasizing the differences related to the timing of snowmelt and late-season drawdown. Equation 13 determines vulnerability as a function of severity and exposure. Goharian *et al.* (2017) analyzed the relative importance of the contributing factors based on judgment, stakeholder surveys, management, and sensitivity analysis to determine that equal weighting was appropriate. We run the Jenks classification algorithm on the historical simulations to identify the natural breaks and categorize the levels as Low, Medium, and High. Supplementary Material, Tables S3 and S4 display the categorical ranges of vulnerability and peak severity.

## 4. RESULTS/DISCUSSION

Using dry, average, and wet climate scenarios, we assess the performance of the ML-WSM on the April–October daily predictions of reservoir levels, groundwater withdrawal, and imported water requests. We investigate the capabilities of the ML-WSM by calculating the predictive performance, comparing the produced measures of vulnerability, and critically evaluating the April to October predictions of each climate scenario. We conclude with a discussion of the observed benefits and limitations taken from the application of the ML-WSM to the SLCDPU water system, and overall ML, for the planning and management of water resources.

### 4.1. Performance of the ML-WSM

The ML-WSM accurately predicts the seasonal dynamics of each component well in all climate scenarios and we observe the greatest prediction skill during average and dry conditions, a scenario where supply limitations can occur and require attentive management. Table 3 displays the performance metrics for all climate scenarios and we include an in-depth evaluation for the dry climate scenario (Figure 5) to illustrate the predictive capacity of the ML-WSM.

**Table 2** | Unsatisfactory condition thresholds of the SLCDPU water system components

	Deer Creek requests	Groundwater	Little Dell reservoir	Mountain Dell reservoir
Threshold	>Historical daily mean	>Historical daily mean	<15%	<45% capacity



**Table 3** | The ML-WSM exhibits high performance across all metrics during average and dry climate conditions but demonstrates limitations during wet climate conditions for imported water requests

Component	Climate conditions	RMSE	KGE	PBias
Mountain Dell Reservoir level	Dry	3.25*	0.88	2.06
	Average	2.57*	0.91	-1.15
	Wet	6.59*	0.82	8.18
Little Dell Reservoir level	Dry	2.10*	0.97	2.85
	Average	1.95*	0.98	-0.32
	Wet	3.86*	0.82	4.47
Groundwater withdrawal	Dry	0.88**	0.89	-2.08
	Average	1.12**	0.91	-2.93
	Wet	1.09**	0.94	1.6
Deer Creek request	Dry	2.19**	0.91	3.45
	Average	2.14**	0.89	-3.03
	Wet	1.78**	-0.33	87.16

\*Units in % of full reservoir level, \*\*Units in  $\times 10^4 \text{ m}^3$ .

#### 4.1.1. Dry hydroclimate

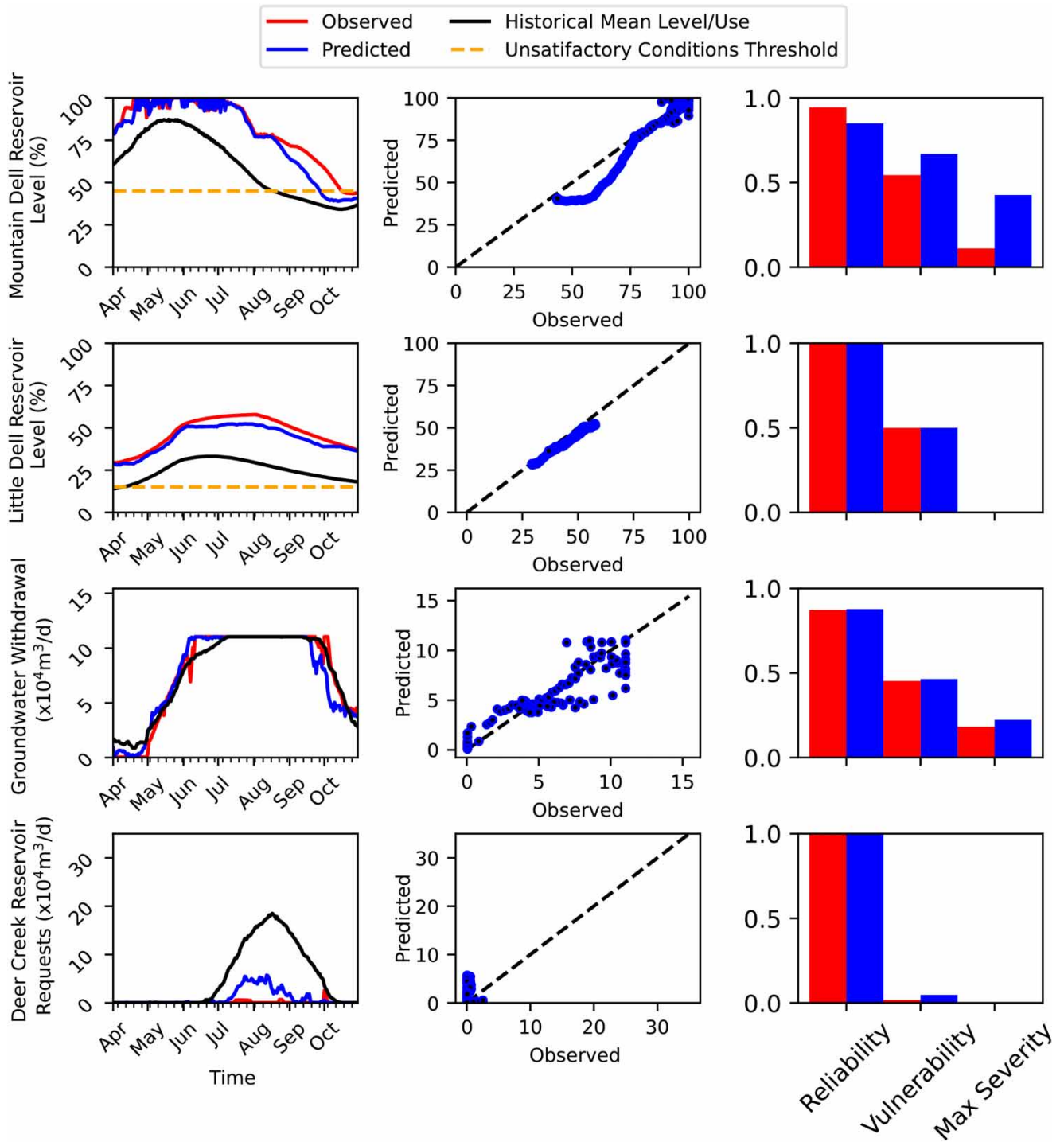
The ML-WSM accurately predicts the seasonal time series of each MWS component during dry climate conditions. For the Dell reservoir system, the predictions mirror the observed from the rise in storage from spring snowmelt to peak reservoir level (i.e., ~75% Mountain Dell and ~35% Little Dell of full capacity) as well as the drawdown timing and rate, resulting in a low *RMSE*, a high *KGE*, and a small positive *PBias* that indicates a slight underprediction in reservoir level (Table 3). Figure 5 illustrates the Mountain Dell Reservoir predictions surpassing the 45% of capacity threshold within 2 days of the observed and Little Dell Reservoir predictions nearing dead pool about a month early. While earlier predictions of a dead pool could be a critical error, the observed is only 5% higher and would likely support similar management actions.

The groundwater withdrawal and imported Deer Creek water requests predictions demonstrate system connectivity that resembles the observed difference in timing (i.e., early and late seasons), the magnitude of withdrawal/request, and the duration of use compared to the historical mean (Figure 5). The *RMSE*, *KGE*, and *PBias* metrics illustrate the high model performance. The model captures the increased early-season groundwater withdrawal rates driven by low surface water supplies and near-critical reservoir levels. Nearing the end of the season, the ML-WSM captures the groundwater withdrawal response to below-average reservoir levels, high demand, and increased rates of Deer Creek requests with the appropriate increase in daily withdrawal. Evaluating Deer Creek water requests, the model predicts the correct timing, rapidly increasing rate of request, bimodal seasonal request peaks, and peak magnitude. The accurate predictions result in the reliability, vulnerability, and peak severity mirroring the observed.

#### 4.2. Vulnerability assessment

We determine the error of the ML-WSM vulnerability assessment for each climate scenario to complete model evaluation. The simulations of historical MWS interactions and feedbacks from the SLC-WSM, coupled with unsatisfactory/satisfactory thresholds defining the status of each MWS component, support the categorization of vulnerability and peak severity with the Jenks natural breaks algorithm. We compare the classification of the system vulnerabilities suggested by the ML-WSM to those provided by the SLC-WSM (Table 4).

The ML-WSM demonstrates high skill in estimating the reliability, vulnerability, and peak severity of all MWS components for all climate scenarios. The reliability estimates for the Dell reservoir system exhibit the greatest difference from the observed with an underprediction of 0.09 during wet (Mountain Dell, 0.85 vs. 0.94) and dry (Little Dell, 0.78 vs. 0.88) conditions. For Mountain Dell, the response is because the modeled drawdown rate surpassed the critical threshold approximately 2 weeks before the simulation from the SLC-WSM (Supplementary Material, Figure S4). The ML-WSM correctly estimates the vulnerability classification in ten out of twelve scenarios/components. The misclassifications align with the dry scenario and groundwater withdrawal (i.e., high predicted vs. medium observed) and during the wet scenario for Deer Creek requests (i.e., medium predicted vs. low observed). While the ML-WSM does misclassify groundwater withdrawal during the dry scenario, the natural break between Medium and High from the Jenks algorithm (i.e., 0.53) splits the difference



**Figure 5** | The ML-WSM produces accurate MWS component predictions and vulnerability estimates during dry climate conditions, emphasized by the minimal prediction bias in the prediction vs. observed plots. The model captures the snowmelt-storage increase in the Dell reservoir system, increased initial groundwater withdrawal, and bimodal seasonal imported Deer Creek water requests, demonstrating the system response to reduced surface water supply and low reservoir levels.

between the actual (i.e., 0.51) and predicted (i.e., 0.55) values. The peak severity of each component predicted by the ML-WSM matches the observed for each climate scenario. While the model correctly classifies the peak severity of each component, peak severity values exhibit the greatest deviations from the observed compared to other metrics.

**Table 4** | The high prediction accuracy of the ML-WSM produced reliability, vulnerability, and peak severity estimates with minimal differences from the observed (in parenthesis) for each MWS component, resulting in the correct classification of 10 out of 12 vulnerability classes and all peak severity classes

Metric	Climate scenario (snowpack)	Mountain Dell	Little Dell	Groundwater withdrawal	Deer Creek request
Reliability	Dry	0.39 (0.41)	0.79 (0.88)	0.58 (0.55)	0.47 (0.43)
	Average	0.91 (0.90)	1.0 (1.0)	0.47 (0.55)	0.75 (0.76)
	Wet	0.85 (0.94)	1.0 (1.0)	0.64 (0.66)	0.99 (1.0)
Vulnerability	Dry	0.79 (0.78)	0.48 (0.49)	0.55 (0.51)	0.50 (0.45)
	Average	0.57 (0.62)	0.50 (0.50)	0.55 (0.53)	0.47 (0.39)
	Wet	0.67 (0.54)	0.50 (0.50)	0.48 (0.46)	0.13 (0.02)
Peak severity	Dry	1.0 (0.91)	0.31 (0.23)	0.57 (0.60)	0.60 (0.52)
	Average	0.19 (0.32)	0.0 (0.0)	0.30 (0.27)	0.73 (0.65)
	Wet	0.43 (0.11)	0.0 (0.0)	0.20 (0.18)	0.01 (0.0)
Vulnerability level	Dry	M (M)	L (L)	H (M)	H (H)
	Average	M (M)	L (L)	H (H)	H (H)
	Wet	M (M)	L (L)	M (M)	M (L)
Peak severity level	Dry	H (H)	M (M)	M (M)	M (M)
	Average	M (M)	L (L)	L (L)	H (H)
	Wet	M (M)	L (L)	L (L)	L (L)

The vulnerability and peak severity values and classifications highlight the information differences between each metric. For example, the supply and demand factors influencing the MWS during the dry climate conditions lead to a Deer Creek water request pattern similar to historical conditions but at a greater magnitude. While there is a greater overall quantity of water requested at the seasonal scale (High vulnerability), the conditions result in a Medium level of peak severity because there are no large differences between the modeled and the historical mean (i.e., unsatisfactory/satisfactory threshold). During the average climate conditions, the increased municipal water demand and seasonal supply limitations drive the Deer Creek requests to a peak in July compared to the historical mean peak in mid-August. While the spike in July imported water use leads to a High level of peak severity, the model estimates vulnerability to be medium because the seasonal request volume is minimally greater than the historically observed. We find peak severity communicates the most extreme state of the system while vulnerability refers to the average state of the system for the season, with the differences in information exemplifying the benefits of a multi-metric assessment to inform MWS planning and guidance activities.

### 4.3. Data-driven water system planning and management

Applying the ML-WSM to the SLCDFU water system demonstrates the benefits of incorporating ML methods into the planning and management tool kit. The ML model delivers minimal prediction error and captures the overall MWS dynamics and relationships without the need for an extensive system understanding or high parameterization. While we train the ML-WSM on the data from the SLC-WSM simulations, the model demonstrates the skill to model reservoir operations, imported water requests, and groundwater withdrawal as a function of their physical operations, limitations, and system connectivity. We foresee no significant deviations in model performance compared to a systems model for real-world application and view the use of physical observations as an opportunity to compare the performance between the ML-WSM and systems models. Although not a direct measure of prediction accuracy, the ML-WSM predicts 60× faster than the systems model (2 s vs. 120 s, respectively), allowing for more rapid simulations and investigations of MWS response. With respect to model development, algorithm training requires minimal developer input compared to systems models and collectively has 36 parameters (e.g., nine hyperparameters per MWS component model) compared to the 4,300 parameters in the SLC-WSM. Even though the XGBoost training took approximately 5.5 h/MWS component, model training can occur outside of business hours or in the background while the developer addresses other tasks. In comparison, the systems model required constant attention and iteration by the developers over many years.

The demonstration of system connectivity, high prediction performance, and open-source nature support the ML-WSM as a tool to advance water system management. Future research applications using a probability of inputs (e.g., stochastic modeling) or developing the ML-WSM with probabilistic algorithms (e.g., Gaussian Process Regression) could provide a probable

range of predictions (Dhara *et al.* 2018; Castellani *et al.* 2021; Fang *et al.* 2022) that reflect intrasystem variability and support risk-tolerance-based decision-making (Towler *et al.* 2013). While not modeling the components of an MWS, Sun *et al.* (2014) and Bonakdari *et al.* (2019) probabilistic algorithms successfully forecast streamflow and lake water levels, respectively.

The open-source concept of the ML-WSM leverages and contributes to advancements throughout the community modeling enterprise. The open-source Scikit-Learn package supports data processing, train/test partitioning, variable selection, model algorithms, and evaluation tools at no charge to the developer. Community modeling supports the latest advances in ML, where platforms like GitHub provide a virtual arena to share tools, discuss problems, and create solutions that support the transition of research to operations as other tools now support hydrological hazard awareness (Khattar *et al.* 2021), such as forecast-informed reservoir operations (Delaney *et al.* 2020) or the Next-Generation water resources framework of the National Water Model (Bartel *et al.* 2021). Open-source software not only supports end-users and stakeholders but can have broad-reaching impacts related to user engagement and learning opportunities. Researchers can adapt the ML-WSM for their systems modeling objective, use it as an educational tool, and/or contribute and enhance model functionality. Compared to a licensed software platform, the limitations surrounding the quantity and quality of available external resources can create roadblocks that hinder development and often only provide access to a few select individuals.

#### 4.4. Opportunities to advance the ML-WSM

Applying the ML-WSM to the SLCDPU system highlighted opportunities to advance the modeling framework ranging from general ML challenges to enhancing transferability among systems. A needed advancement is methods implementing physical constraints on MWS components to prevent impossible results, a known limitation of data-driven models (Qian *et al.* 2020). We encountered the prediction of impossible results during the preliminary phases of development with negative estimates of groundwater withdrawal and Deer Creek imported water requests less than the minimum flow rate. While we implemented model constraints, it implies that physically impossible errors can occur, such as reservoir levels exceeding 100%, below the dead pool, or even negative values. Future applications of the ML-WSM could explore physics-informed ML models which can address the physical limitations of the system, allowing network architectures to automatically satisfy some of the physical sets of assumptions before performing any computation (Qian *et al.* 2020). Physics-informed ML could account for the mass balance of reservoir level change, groundwater withdrawal, and imported water use from municipal water demand to ensure no net gain or loss of water at each time step of a simulation. Integrating methods to physically constrain model predictions can aid in model interpretability, addressing the black box and explaining the conceptual path to prediction (Molnar *et al.* 2020).

Another opportunity to advance ML models is to develop frameworks to improve the reliability of dynamical system predictions outside of the training bounds. Alterations of the statistical behavior between feature-target responses, such as from a changing climate or a change in service area composition, can challenge ML model prediction skill (Feudel *et al.* 2018; Kaszás *et al.* 2019). The parameterization and physical connections within a systems modeling framework can model interactions entering a state of non-stationarity, addressing a limitation of data-driven methods (Chantry *et al.* 2021; Shi *et al.* 2021). Evolving ML methods utilizing reservoir computing RNN models show promise in the prediction of long-term dynamical systems behavior influenced by non-stationarity, including sudden changes due to regime transitions (Patel *et al.* 2021; Patel & Ott 2023). While the use of the extreme wet and extreme dry testing scenarios exhibit conditions outside of the bounds of model training and the ML-WSM demonstrated high prediction skill, the high performance may be attributed to the selected forecasting horizon. To assess the long-term impacts of non-stationarity on MWS performance with ML, we recommend exploring reservoir computing RNN and/or physics-informed ML methods.

#### 4.5. Model synergy

We see a synergistic approach for the integration of ML into MWS management and decision-making, as the results indicate that the ML-WSM does not need to replace the systems model to be beneficial. The systems modeling methods will likely provide the most physically representative estimates of MWS performance but the fast and accurate representation of the MWS from the ML-WSM can function as a first approximation. Fast simulation speeds can complement a systems model analysis by quickly evaluating many system scenarios, assessing the respective responses of each key MWS component, and developing preliminary scenario-driven system management actions in a short period of time. For example, the process could quickly inform system management on the level of conservation needed to maintain a specific reliability threshold

during supply-limited conditions. For MWS without an existing model but with ample data, the ML-WSM can reduce the barriers to modeling the water system.

The ML-WSM framework demonstrates the potential to benefit MWS planning and management by establishing a pillar of increasing engagement throughout the community. The concept builds on the open-source nature of the ML community to interact with a larger audience than systems models. Open-source products like Tethys provide a platform to increase community awareness concerning system response to externalities (Swain *et al.* 2016). Maturing the ML-WSM into a web application targeting public water use awareness could engage a broad audience, as an interactive environment using a trained ML-WSM and multiple scenarios could present information within a game-like setting that encourages experimentation and serves as an education tool (Savic *et al.* 2016; Laucelli *et al.* 2019). A synergistic approach to water system modeling (ML and systems models) brings the benefits of physical and data-driven models to inform decision-making and address high-impact real-world water management challenges.

## 5. CONCLUSION

We apply the ML-WSM to the SLCDPU water system to identify the strengths and limitations of ML in modeling water system interactions between key MWS components that are critical to decision-making. Using the existing systems model as the observed baseline and three different climate scenarios affecting supply and demand, the ML-WSM demonstrates high prediction skill. Using the Xtreme Gradient Boosted (XGBoost) algorithm, the ML-WSM captures the defined feedbacks and interactions of seasonal reservoir level dynamics, groundwater withdrawal, and imported water requests with minimal error and without the high parameterization, high computational requirements, or the long development period of systems models. We couple the predictions to a vulnerability assessment that categorizes peak severity and vulnerability for each key component to improve the interpretability of the results, aiding in decision-making by providing a platform to compare simulations with the historically observed values. While a different measure of performance, the ML-WSM consistently predicted 60× faster than the systems model (2 s vs. 120 s, respectively) and model development took days compared to years, owing to minimal developer input compared to systems models. The novel application of ML for modeling MWS components demonstrates a key contribution to the water resources community by prototyping a complete data-driven MWS model that reduces the development barriers to entry, systematically bypassing the extensive parameterization of a systems model.

Even with the successful implementation of the ML-WSM on the SLCDPU water system, there is potential for improvement. The model evaluation identified impossible values that required physical model constraints to reflect system limitations. We recommend future work exploring physics-informed ML to address the identified limitations and reservoir pool algorithms to examine the feasibility of ML for modeling non-stationarity in water systems. Given the identified limitations, we foresee the ML-WSM to have broad-reaching impacts such as supporting interactive open-source tools, strengthening the understanding of the MWS for utilities without an existing systems model, and establishing a synergistic approach combining ML with physically based modeling systems to engage the greater community and contribute to a wide range of high-impact real-world water management challenges.

## OPEN RESEARCH

All Python v3.10.1-based models are available on Github: <https://github.com/whitelightning450/Machine-Learning-Water-Systems-Model>. The repository contains all data to train and run the ML-WSM, providing a framework guiding the adaptation of the code to another system of interest. The SLC-WSM is not provided for review due to security reasons specified by the SLCDPU. Permission for this model requires direct consent from the SLCDPU.

## ACKNOWLEDGEMENTS

Funding for this project was provided by the National Oceanic and Atmospheric Administration (NOAA), and awarded to the Cooperative Institute for Research to Operations in Hydrology (CIROH) through the NOAA Cooperative Agreement with The University of Alabama, NA22NWS4320003. The authors would like to thank all the SLCDPU staff for their time and commitment to the Salt Lake City Vulnerability Project. The provided funding and enthusiasm for science accelerated this research. This research would also like to thank Margaret Wolf, Logan Jamison, Dr Paul Brooks, and Dr Courtney Strong for their collaborative work on this project.

## DATA AVAILABILITY STATEMENT

All relevant data are available from an online repository or repositories: <https://github.com/whitelightning450/Machine-Learning-Water-Systems-Model>.

## CONFLICT OF INTEREST

The authors declare there is no conflict.

## REFERENCES

- Aghelpour, P. & Varshavian, V. 2020 Evaluation of stochastic and artificial intelligence models in modeling and predicting of river daily flow time series. *Stochastic Environmental Research and Risk Assessment* **34**. doi:10.1007/s00477-019-01761-4.
- Akinwande, M. O., Dikko, H. G. & Samson, A. 2015 Variance inflation factor: as a condition for the inclusion of suppressor variable (s) in regression analysis. *Open Journal of Statistics* **5** (7), 754.
- Alemu, E. T., Palmer, R. N., Polebitski, A. & Meaker, B. 2011 Decision support system for optimizing reservoir operations using ensemble streamflow predictions. *Journal of Water Resources Planning and Management* **137** (1), 72–82.
- Alin, A. 2010 Multicollinearity. *Wiley Interdisciplinary Reviews: Computational Statistics* **2** (3), 370–374.
- Antunes, A., Andrade-Campos, A., Sardinha-Lourenço, A. & Oliveira, M. 2018 Short-term water demand forecasting using machine learning techniques. *Journal of Hydroinformatics* **20**. doi:10.2166/hydro.2018.165.
- Arriagada, P., Karelovic, B. & Link, O. 2021 Automatic gap-filling of daily streamflow time series in data-scarce regions using a machine learning algorithm. *Journal of Hydrology* **598**, 126454.
- Asefa, T., Clayton, J., Adams, A. & Anderson, D. 2014 Performance evaluation of a water resources system under varying climatic conditions: reliability, resilience, vulnerability and beyond. *Journal of Hydrology* **508**, 53–65.
- Barnett, M. J., Jackson-Smith, D. & Endter-Wada, J. 2019 Implications of nontraditional housing arrangements for urban water management in the United States intermountain west. *Society & Natural Resources* **32** (5), 508–529.
- Bartel, R., Williamson, M. & Tubbs, C. 2021 *Next gen Water Modeling Framework Prototype*. GitHub. Available from: <https://github.com/NOAA-OWP/ngen>.
- Bobee, B. 1975 The log Pearson type 3 distribution and its application in hydrology. *Water Resources Research* **11** (5), 681–689.
- Bonakdari, H., Ebtehaj, I., Samui, P. & Gharabaghi, B. 2019 Lake water-level fluctuations forecasting using minimax probability machine regression, relevance vector machine, Gaussian process regression, and extreme learning machine. *Water Resources Management* **33**, 1–20. doi:10.1007/s11269-019-02346-0.
- Cai, J., Luo, J., Wang, S. & Yang, S. 2018 Feature selection in machine learning: a new perspective. *Neurocomputing* **300**, 70–79.
- Castelletti, A., Galelli, S., Restelli, M. & Soncini-Sessa, R. 2010 Tree-based reinforcement learning for optimal water reservoir operation. *Water Resources Research* **46** (9), 1–19.
- Castellani, G., Fiore, U., Marino, Z., Passalacqua, L., Perla, F., Scognamiglio, S. & Zanetti, P. 2021 Machine learning techniques in nested stochastic simulations for life insurance. *Applied Stochastic Models in Business and Industry* **37** (2), 159–181.
- Chandrashekar, G. & Sahin, F. 2014 A survey on feature selection methods. *Computers & Electrical Engineering* **40** (1), 16–28.
- Chanry, M., Christensen, H., Dueben, P. & Palmer, T. 2021 Opportunities and challenges for machine learning in weather and climate modelling: hard, medium and softai. *Philosophical Transactions of the Royal Society A* **379** (2194), 20200083.
- Chatterjee, S. & Simonoff, J. S. 2013 *Handbook of Regression Analysis*, Vol. 5. John Wiley & Sons, Hoboken.
- Chen, T. & Guestrin, C. 2016 *Xgboost: A Scalable Tree Boosting System*. pp. 785–794. doi:10.1145/2939672.2939785.
- Chen, X.-W. & Jeong, J. C. 2007 Enhanced recursive feature elimination. In: *Sixth International Conference on Machine Learning and Applications (ICMLA 2007)*, Cincinnati, OH, 13–15 December 2007. IEEE, New York, pp. 429–435.
- Choudhary, R. & Gianey, H. K. 2017 Comprehensive review on supervised machine learning algorithms. In: *2017 International Conference on Machine Learning and Data Science (MLDS)*. 14–15 December 2017. IEEE, New York, pp. 37–43.
- Collins, B. & Associates 2019 *2018 Supply and Demand Master Plan (Tech. Rep.)*. BCA, Salt Lake City, Utah.
- Dao, D. V., Adeli, H., Ly, H.-B., Le, L. M., Le, V. M., Le, T.-T. & Pham, B. T. 2020 A sensitivity and robustness analysis of gpr and ann for high-performance concrete compressive strength prediction using a monte carlo simulation. *Sustainability* **12** (3), 830.
- Delaney, C. J., Hartman, R. K., Mendoza, J., Dettinger, M., Delle Monache, L., Jasperse, J., Talbot, C., Brown, J., Evett, S. & Reynolds, D. 2020 Forecast informed reservoir operations using ensemble streamflow predictions for a multipurpose reservoir in northern California. *Water Resources Research* **56** (9), e2019WR026604.
- Dembéle, M., Mariéthoz, G. & Schaeffli, B. 2017 Gap filling of streamflow time series using direct sampling in data scarce regions. In *Swiss Geoscience Meeting, Edited*, Davos, Switzerland.
- Dhara, A., Trainor-Guitton, W. & Tura, A. 2018 Machine-learning-based methods for estimation and stochastic simulation. In *2018 seg International Exposition and Annual Meeting*.
- Dieter, C. A. 2018 *Water Availability and use Science Program: Estimated use of Water in the United States in 2015*. Geological Survey

- Dormann, C. F., Elith, J., Bacher, S., Buchmann, C., Carl, G., Carré, G., García Marquéz, J. R., Gruber, B., Lafourcade, B., Leitão, P. J., Münkemüller, T., McClean, C., Osborne, P. E., Reineking, B., Schröder, B., Skidmore, A. K., Zurell, D. & Lautenbach, S. 2013 [Collinearity: a review of methods to deal with it and a simulation study evaluating their performance](#). *Ecography* **36** (1), 27–46.
- Eggimann, S., Mutzner, L., Wani, O., Schneider, M. Y., Spuhler, D., Moy de Vitry, M., Beutler, P. & Maurer, M. 2017 [The potential of knowing more: a review of data-driven urban water management](#). *Environmental Science & Technology* **51** (5), 2538–2553.
- Fang, Y., Sun, Y., Zhang, L., Chen, G., Du, M. & Guo, Y. 2022 [Stochastic simulation of typhoon in northwest pacific basin based on machine learning](#). *Computational Intelligence and Neuroscience* **2022**, 6760944.
- Feudel, U., Pisarchik, A. N. & Showalter, K. 2018 [Multistability and tipping: from mathematics and physics to climate and brain – minireview and preface to the focus issue](#). *Chaos: An Interdisciplinary Journal of Nonlinear Science* **28** (3), 033501.
- Ficchi, A., Perrin, C. & Andréassian, V. 2016 [Impact of temporal resolution of inputs on hydrological model performance: an analysis based on 2400 flood events](#). *Journal of Hydrology* **538**, 454–470.
- Fu, G., Jin, Y., Sun, S., Yuan, Z. & Butler, D. 2022 [The role of deep learning in urban water management: a critical review](#). *Water Research* **223**, 118973.
- Füssel, H.-M. 2010 [Review and Quantitative Analysis of Indices of Climate Change Exposure, Adaptive Capacity, Sensitivity, and Impacts](#).
- Gangrade, S., Lu, D., Kao, S.-C. & Painter, S. L. 2022 [Machine learning assisted reservoir operation model for long-term water management simulation](#). *JAWRA Journal of the American Water Resources Association* **58** (6), 1592–1603.
- Garreta, R., Moncecchi, G., Hauck, T. & Hackeling, G. 2017 [Scikit-learn: Machine Learning Simplified: Implement Scikit-Learn Into Every Step of the Data Science Pipeline](#). Packt Publishing Ltd, Birmingham, UK.
- Gastelum, J., Valdes, J. & Stewart, S. 2008 [A decision support system to improve water resources management in the Conchos basin](#). *Water Resources Management* **23**, 1519–1548. doi:10.1007/s11269-008-9339-4.
- Gastelum, J., Valdes, J. & Stewart, S. 2009 [A system dynamics model to evaluate temporary water transfers in the Mexican Conchos basin](#). *Water Resources Management* **24**, 1285–1311. doi:10.1007/s11269-009-9497-z.
- Ghobadi, F. & Kang, D. 2023 [Application of machine learning in water resources management: a systematic literature review](#). *Water* **15** (4), 620.
- Goharian, E. & Burian, S. 2018 [Developing an integrated framework to build a decision support tool for urban water management](#). *Journal of Hydroinformatics* **20**, jh2018088. doi:10.2166/hydro.2018.088.
- Goharian, E., Burian, S. J., Bardsley, T. & Strong, C. 2016 [Incorporating potential severity into vulnerability assessment of water supply systems under climate change conditions](#). *Journal of Water Resources Planning and Management* **142** (2), 04015051.
- Goharian, E., Burian, S. J., Lillywhite, J. & Hile, R. 2017 [Vulnerability assessment to support integrated water resources management of metropolitan water supply systems](#). *Journal of Water Resources Planning and Management* **143** (3). doi:10.1061/(asce)wr.1943-5452.0000738.
- GoldSim. 2013 [GoldSim Probabilistic Simulation Environment](#). GoldSim Technological Group LLC, Issaquah, Washington.
- Gupta, H. V., Kling, H., Yilmaz, K. K. & Martinez, G. F. 2009 [Decomposition of the mean squared error and nse performance criteria: implications for improving hydrological modelling](#). *Journal of Hydrology* **377** (1–2), 80–91.
- Guyon, I. & Elisseeff, A. 2003 [An introduction to variable and feature selection](#). *Journal of Machine Learning Research* **3**, 1157–1182.
- Han, K., Wang, Y., Zhang, C., Li, C. & Xu, C. 2018 [Autoencoder inspired unsupervised feature selection](#). In: *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 15–20 April 2018. IEEE, New York, pp. 2941–2945.
- Harpold, A., Brooks, P., Rajagopal, S., Heidbuchel, I., Jardine, A. & Stielstra, C. 2012 [Changes in snowpack accumulation and ablation in the intermountain west](#). *Water Resources Research* **48** (11), 1–11.
- Hashimoto, T., Stedinger, J. & Loucks, P. 1982 [Reliability, resiliency, and vulnerability criteria for water resource system performance evaluation](#). *Water Resources Research* **18**. doi:10.1029/WR018i001p00014.
- Haskins, G., Kruger, U. & Yan, P. 2020 [Deep learning in medical image registration: a survey](#). *Machine Vision and Applications* **31** (1), 1–18.
- Hofmann, T. 2001 [Unsupervised learning by probabilistic latent semantic analysis](#). *Machine Learning* **42**, 177–196.
- Hu, C., Wu, Q., Li, H., Jian, S., Li, N. & Lou, Z. 2018 [Deep learning with a long short-term memory networks approach for rainfall-runoff simulation](#). *Water* **10** (11), 1543.
- Imani, M., Hasan, M. M., Bittencourt, L. F., McClymont, K. & Kapelan, Z. 2021 [A novel machine learning application: water quality resilience prediction model](#). *Science of The Total Environment* **768**, 144459. doi:10.1016/j.scitotenv.2020.144459.
- Jadidoleslam, N., Mantilla, R., Krajewski, W. F. & Cosh, M. H. 2019 [Data-driven stochastic model for basin and sub-grid variability of SMAP satellite soil moisture](#). *Journal of Hydrology* **576**, 85–97.
- Jaiswal, R., Ali, S. & Bharti, B. 2020 [Comparative evaluation of conceptual and physical rainfall–runoff models](#). *Applied Water Science* **10** (1), 1–14.
- Jenks, G. 1967 [The data model concept in statistical mapping](#). In: *International Yearbook of Cartography* (Frenezal, K., ed.). George Philip, London. pp. 186–190.
- Jennings, K. S., Wlostowski, A. N., Bash, R. E., Burkhardt, J., Wobus, C. W. & Aggett, G. 2022 [Data availability and sector-specific frameworks restrict drought impact quantification in the intermountain west](#). *Wiley Interdisciplinary Reviews: Water* **9** (3), e1586.
- Jia, W., Sun, M., Lian, J. & Hou, S. 2022 [Feature dimensionality reduction: a review](#). *Complex & Intelligent Systems* **8** (3), 2663–2693.
- Johnson, R. C., Wolf, M., Jamison, L., Burian, S., Oroza, C. A., Brooks, P. D., Strong, C., Stewart, J. & Kirkham, T. 2021 [Drought in the west: embedded water demand stationarity compromises system vulnerability analysis](#). *Open Water Journal* **7** (1), 6.
- Kalin, L., Isik, S., Schoonover, J. & Lockaby, G. 2010 [Predicting water quality in unmonitored watersheds using artificial neural networks](#). *Journal of Environmental Quality* **39**, 1429–1440. doi:10.2134/jeq2009.0441.

- Kaszás, B., Feudel, U. & Tél, T. 2019 [Tipping phenomena in typical dynamical systems subjected to parameter drift](#). *Scientific Reports* **9** (1), 8654.
- Keogh, E. J. & Mueen, A. 2017 [Curse of dimensionality](#). In: *Encyclopedia of Machine Learning and Data Mining*. Vol. 2017, (Sammut, C. & Webb, G.I., eds.). Springer, Cham, pp. 314–315.
- Khattar, R., Hales, R., Ames, D. P., Nelson, E. J., Jones, N. L. & Williams, G. 2021 [Tethys app store: simplifying deployment of web applications for the international geoglows initiative](#). *Environmental Modelling & Software* **146**, 105227.
- Knoben, W. J., Freer, J. E. & Woods, R. A. 2019 [Inherent benchmark or not? Comparing Nash–Sutcliffe and Kling–Gupta efficiency scores](#). *Hydrology and Earth System Sciences* **23** (10), 4323–4331.
- Kouziokas, G. N., Chatzigeorgiou, A. & Perakis, K. 2018 [Multilayer feed forward models in groundwater level forecasting using meteorological data in public management](#). *Water Resources Management* **32**, 5041–5052.
- Kratzert, F., Klotz, D., Brenner, C., Schulz, K. & Herrnegger, M. 2018 [Rainfall–runoff modelling using long short-term memory \(LSTM\) networks](#). *Hydrology and Earth System Sciences* **22** (11), 6005–6022.
- Krishnan, S. R., Nallakaruppan, M., Chengoden, R., Koppu, S., Iyapparaja, M., Sadhasivam, J. & Sethuraman, S. 2022 [Smart water resource management using artificial intelligence – a review](#). *Sustainability* **14** (20), 13384.
- Långkvist, M., Karlsson, L. & Loutfi, A. 2014 [A review of unsupervised feature learning and deep learning for time-series modeling](#). *Pattern Recognition Letters* **42**, 11–24.
- Laucelli, D. B., Berardi, L., Simone, A. & Giustolisi, O. 2019 [Towards serious gaming for water distribution networks sizing: a teaching experiment](#). *Journal of Hydroinformatics* **21** (2), 207–222.
- Lee, I. & Shin, Y. J. 2020 [Machine learning for enterprises: applications, algorithm selection, and challenges](#). *Business Horizons* **63** (2), 157–170.
- Lhomme, J.-P., Boudhina, N., Masmoudi, M. M. & Chehbouni, A. 2015 [Estimation of crop water requirements: extending the one-step approach to dual crop coefficients](#). *Hydrology and Earth System Sciences* **19** (7), 3287–3299.
- Li, Y., Li, T. & Liu, H. 2017 [Recent advances in feature selection and its applications](#). *Knowledge and Information Systems* **53**, 551–577.
- Li, X., Wu, X., Sun, M., Yang, S. & Song, W. 2022 [A novel intelligent leakage monitoring-warning system for sustainable rural drinking water supply](#). *Sustainability* **14** (10), 6079.
- Lillywhite, J. R. 2008 [Performance of Water Supply Operations Measured by Reliability and Marginal Cost](#). Unpublished doctoral dissertation, Department of Civil and Environmental Engineering, University of Utah.
- Ma, L., Liu, Y., Zhang, X., Ye, Y., Yin, G. & Johnson, B. A. 2019 [Deep learning in remote sensing applications: a meta-analysis and review](#). *ISPRS Journal of Photogrammetry and Remote Sensing* **152**, 166–177.
- MacDonald, R. J., Byrne, J. M., Boon, S. & Kienzle, S. W. 2012 [Modelling the potential impacts of climate change on snowpack in the north Saskatchewan river watershed, Alberta](#). *Water Resources Management* **26** (11), 3053–3076.
- Madani, K. & Mariño, M. 2009 [System dynamics analysis for managing Iran's Zayandeh-Rud river basin](#). *Water Resources Management* **23**, 2163–2187. doi:10.1007/s11269-008-9376-z.
- Mahmoudi, N., Orouji, H. & Fallah-Mehdipour, E. 2016 [Integration of shuffled frog leaping algorithm and support vector regression for prediction of water quality parameters](#). *Water Resources Management* **30**. doi:10.1007/s11269-016-1280-3.
- Makropoulos, C., Nikolopoulos, D., Palmen, L., Kools, S., Segrave, A., Vries, D., Koop, S., van Alphen, H. J., Vonk, E., Rozos, E., Medema, G. & Van Thienen, P. 2018 [A resilience assessment method for urban water systems](#). *Urban Water Journal* **15** (4), 316–328.
- Malmfeldt, M. P. 2021 [Cyanobacteria and Phytoplankton Responses to Nutrients in Deep-Water Montane Reservoirs](#). Unpublished doctoral dissertation, Brigham Young University.
- Marçais, J. & de Dreuzy, J.-R. 2017 [Prospective interest of deep learning for hydrological inference](#). *Groundwater* **55** (5), 688–692.
- McCuen, R. H. 2016 [Modeling Hydrologic Change: Statistical Methods](#). CRC press, Boca Raton, FL.
- Menard, S. 2002 [Applied Logistic Regression Analysis](#). No. 106. Sage, London.
- Moghaddasi, P., Kerachian, R. & Sharghi, S. 2022 [A stakeholder-based framework for improving the resilience of groundwater resources in arid regions](#). *Journal of Hydrology* **609**, 127737.
- Mohammadi, B., Guan, Y., Aghelpour, P., Emamgholizadeh, S., Zolá, R. & Zhang, D. 2020 [Simulation of Titicaca lake water level fluctuations using hybrid machine learning technique integrated with grey wolf optimizer algorithm](#). *Water* **12**. doi:10.3390/w12113015.
- Moishin, M., Deo, R. C., Prasad, R., Raj, N. & Abdulla, S. 2021 [Designing deep-based learning flood forecast model with convLSTM hybrid algorithm](#). *IEEE Access* **9**, 50982–50993.
- Molnar, C., Casalicchio, G. & Bischl, B. 2020 [Interpretable Machine Learning – A Brief History, State-of-the-art and Challenges](#).
- Monteith, J. L. 1965 [Evaporation and environment](#). *Symposia of the Society for Experimental Biology* **19**, 205–234.
- Moore, B. 1981 [Principal component analysis in linear systems: controllability, observability, and model reduction](#). *IEEE Transactions on Automatic Control* **26** (1), 17–32.
- Mukheibir, P. 2008 [Water resources management strategies for adaptation to climate-induced impacts in South Africa](#). *Water Resources Management* **22**, 1259–1276.
- Muthukrishnan, R. & Rohini, R. 2016 [Lasso: A feature selection technique in predictive modeling for machine learning](#). In: *2016 IEEE International Conference on Advances in Computer Applications (ICACA)*, 24–24 October 2016. IEEE, New York, pp. 18–20.
- Nikolopoulos, D., van Alphen, H.-J., Vries, D., Palmen, L., Koop, S., van Thienen, P., Medema, G. & Makropoulos, C. 2019 [Tackling the 'new normal': a resilience assessment method applied to real-world urban water systems](#). *Water* **11** (2), 330.
- NOAA. 2021 [Mesowest Alta Guard Current Conditions](#). Available from: <https://www.wrh.noaa.gov/mesowest/getobext.php>.



- Noori, N., Kalin, L. & Isik, S. 2020 Water quality prediction using SWAT-ANN coupled approach. *Journal of Hydrology* **590**, 125220. doi:10.1016/j.jhydrol.2020.125220.
- O'Brien, R. M. 2007 A caution regarding rules of thumb for variance inflation factors. *Quality & Quantity* **41**, 673–690.
- Patel, D. & Ott, E. 2023 Using machine learning to anticipate tipping points and extrapolate to post-tipping dynamics of non-stationary dynamical systems. *Chaos: An Interdisciplinary Journal of Nonlinear Science* **33** (2), 023143.
- Patel, D., Canaday, D., Girvan, M., Pomerance, A. & Ott, E. 2021 Using machine learning to predict statistical properties of non-stationary dynamical processes: system climate, regime transitions, and the effect of stochasticity. *Chaos: An Interdisciplinary Journal of Nonlinear Science* **31** (3), 033149.
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Müller, A., Nothman, J., Louppe, G., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M. & Duchesnay, E. 2011 Scikit-learn: Machine learning in python. *Journal of Machine Learning Research* **12** (10), 2825–2830.
- Peel, M. C., Finlayson, B. L. & McMahon, T. A. 2007 Updated world map of the köppen-geiger climate classification. *Hydrology and Earth System Sciences* **11** (5), 1633–1644. doi:10.5194/hess-11-1633-2007. Available from: <https://www.hydrol-earth-syst-sci.net/11/1633/2007/>
- Pham, B. T., Qi, C., Ho, L. S., Nguyen-Thoi, T., Al-Ansari, N., Nguyen, M. D., Ly, H.-B., Le, H. V. & Prakash, I. 2020 A novel hybrid soft computing model using random forest and particle swarm optimization for estimation of undrained shear strength of soil. *Sustainability* **12** (6), 2218.
- Probst, P., Wright, M. N. & Boulesteix, A.-L. 2019 Hyperparameters and tuning strategies for random forest. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery* **9** (3), e1301.
- Purkey, D. R., Huber-Lee, A., Yates, D. N., Hanemann, M. & Herrod-Julius, S. 2007 Integrating a climate change assessment tool into stakeholder-driven water management decision-making processes in California. *Water Resources Management* **21** (1), 315–329.
- Putatunda, S. & Rama, K. 2018 A comparative analysis of hyperopt as against other approaches for hyper-parameter optimization of xgboost. In: *Proceedings of the 2018 International Conference on Signal Processing and Machine Learning*, Shanghai, 28-30 November 2018. ACM, New York, pp. 6–10.
- Qian, E., Krämer, B., Peherstorfer, B. & Willcox, K. 2020 Lift & learn: physics-informed machine learning for large-scale nonlinear dynamical systems. *Physica D: Nonlinear Phenomena* **406**, 132401. doi:10.1016/j.physd.2020.132401.
- Raj, P. & David, P. E. 2020 *The Digital Twin Paradigm for Smarter Systems and Environments: The Industry use Cases*. Academic Press, Salt Lake City, UT.
- Raschka, S. 2018 *Model Evaluation, Model Selection, and Algorithm Selection in Machine Learning*. arXiv preprint arXiv:1811.12808.
- Rebora, N., Silvestro, F., Rudari, R., Herold, C. & Ferraris, L. 2016 Downscaling stream flow time series from monthly to daily scales using an auto-regressive stochastic algorithm: streamfarm. *Journal of Hydrology* **537**, 297–310.
- Reuss, M. 2003 Is it time to resurrect the Harvard water program? *Journal of Water Resources Planning and Management* **129** (5), 357–360.
- Rozos, E. 2019 Machine learning, urban water resources management and operating policy. *Resources* **8**, 173. doi:10.3390/resources8040173.
- Sandoval-Solis, S., McKinney, D. & Loucks, D. P. 2011 Sustainability index for water resources planning and management. *Journal of Water Resources Planning and Management* **137** (5), 381–390.
- Sarkar, A. & Pandey, P. 2015 River water quality modelling using artificial neural network technique. *Aquatic Procedia* **4**, 1070–1077. doi:10.1016/j.aqpro.2015.02.135.
- Savic, D. A., Morley, M. S. & Khoury, M. 2016 Serious gaming for water systems planning and management. *Water* **8** (10), 456.
- Scalzitti, J., Strong, C. & Kochanski, A. K. 2016 A 26 year high-resolution dynamical downscaling over the wasatch mountains: synoptic effects on winter precipitation performance. *Journal of Geophysical Research: Atmospheres* **121** (7), 3224–3240. doi:10.1002/2015JD024497.
- Shen, C. 2018 A transdisciplinary review of deep learning research and its relevance for water resources scientists. *Water Resources Research* **54** (11), 8558–8593.
- Shi, Z., Bai, Y., Jin, X., Wang, X., Su, T. & Kong, J. 2021 Parallel deep prediction with covariance intersection fusion on non-stationary time series. *Knowledge-Based Systems* **211**, 106523.
- Shuttleworth, W. J., Serrat-Capdevila, A., Roderick, M. L. & Scott, R. L. 2009 On the theory relating changes in area-average and pan evaporation. *Quarterly Journal of the Royal Meteorological Society: A Journal of the Atmospheric Sciences, Applied Meteorology and Physical Oceanography* **135** (642), 1230–1247.
- Sit, M., Demiray, B. Z., Xiang, Z., Ewing, G. J., Sermet, Y. & Demir, I. 2020 A comprehensive review of deep learning applications in hydrology and water resources. *Water Science and Technology* **82** (12), 2635–2670.
- Strong, C., Stong, L., Brooks, P., Wolf, M., Jamison, L., Burian, S. & Aziz, D. 2020 *Climate Vulnerability Assessment of Salt Lake City's Water Systems: Year 5 Report (Tech. Rep.)*. University of Utah, Salt Lake City, UT.
- Strong, C., Stong, L., Brooks, P., Wolf, M., Jamison, L., Burian, S. & Aziz, D. 2021 *Climate Vulnerability Assessment of Salt Lake City's Water Systems: Year 5 Report (Tech. Rep.)*. University of Utah, Salt Lake City, UT.
- Sun, A. Y. & Scanlon, B. R. 2019 How can big data and machine learning benefit environment and water management: a survey of methods, applications, and future directions. *Environmental Research Letters* **14** (7), 073001.
- Sun, A., Wang, D. & Xu, X. 2014 Monthly streamflow forecasting using Gaussian process regression. *Journal of Hydrology* **511**, 72–81. doi:10.1016/j.jhydrol.2014.01.023.

- Sun, Y., Haghighat, F. & Fung, B. C. 2020 A review of the-state-of-the-art in data-driven approaches for building energy prediction. *Energy and Buildings* **221**, 110022.
- Sunkara, S. V. & Singh, R. 2022 Assessing the impact of the temporal resolution of performance indicators on optimal decisions of a water resources system. *Journal of Hydrology* **612**, 128185.
- Swain, N. R., Christensen, S. D., Snow, A. D., Dolder, H., Espinoza-Dávalos, G., Goharian, E., Jones, N. L., Nelson, E. J., Ames, D. P. & Burian, S. J. 2016 A new open source platform for lowering the barrier for environmental web app development. *Environmental Modelling & Software* **85**, 11–26.
- Toloşi, L. & Lengauer, T. 2011 Classification with correlated features: unreliability of feature ranking and solutions. *Bioinformatics* **27** (14), 1986–1994.
- Towler, E., Roberts, M., Rajagopalan, B. & Sojda, R. 2013 Incorporating probabilistic seasonal climate forecasts into river management using a risk-based framework. *Water Resources Research* **49**. doi:10.1002/wrcr.20378.
- Tyralis, H., Papacharalampous, G. & Langousis, A. 2019 A brief review of random forests for water scientists and practitioners and their recent history in water resources. *Water* **11** (5). doi:10.3390/w11050910. Available from: <https://www.mdpi.com/2073-4441/11/5/910>
- UDWR. 2023 *Water Records/use Information (Tech. Rep.)*. Utah Division of Water Rights, Salt Lake City, Utah. Available from: <https://waterrights.utah.gov/wateruse/WaterUseList.asp>
- Wang, C.-H. & Blackmore, J. 2009 Resilience concepts for water resource systems. *Journal of Water Resources Planning and Management-ASCE* **135**. doi:10.1061/(ASCE)0733-9496(2009)135:6(528).
- Wang, S., Ding, Z. & Fu, Y. 2017 Feature selection guided auto-encoder. In *Proceedings of the AAAI Conference on Artificial Intelligence*. Vol. 31.
- Wang, Z., Xia, L., Yuan, H., Srinivasan, R. S. & Song, X. 2022 Principles, research status, and prospects of feature engineering for data-driven building energy prediction: a comprehensive review. *Journal of Building Engineering* **58**, 105028.
- Winz, I., Brierley, G. & Trowsdale, S. 2008 The use of system dynamics simulation in water resources management. *Water Resources Management* **23**, 1301–1323. doi:10.1007/s11269-008-9328-7.
- Wlostowski, A. N., Jennings, K. S., Bash, R. E., Burkhardt, J., Wobus, C. W. & Aggett, G. 2022 Dry landscapes and parched economies: a review of how drought impacts nonagricultural socioeconomic sectors in the US intermountain west. *Wiley Interdisciplinary Reviews: Water* **9** (1), e1571.
- Wu, J., Ma, D. & Wang, W. 2022 Leakage identification in water distribution networks based on xgboost algorithm. *Journal of Water Resources Planning and Management* **148** (3), 04021107.
- Xenochristou, M. & Kapelan, Z. 2020 An ensemble stacked model with bias correction for improved water demand forecasting. *Urban Water Journal* **17** (3), 212–223.
- Xu, T., Coco, G. & Neale, M. 2020 A predictive model of recreational water quality based on adaptive synthetic sampling algorithms and machine learning. *Water Research* **177**, 115788.
- Yan, K. & Zhang, D. 2015 Feature selection and analysis on correlated gas sensor data with recursive feature elimination. *Sensors and Actuators B: Chemical* **212**, 353–363.
- Yusri, H. I. H., Hassan, S. L. M., Halim, I. S. A. & Abdullah, N. E. 2022 Water quality classification using svm and xgboost method. In: *2022 IEEE 13th Control and System Graduate Research Colloquium (ICSGRC)*, 23-23 July 2022. IEEE, New York, pp. 231–236.
- Zhao, Q., Cai, X. & Li, Y. 2019 Determining inflow forecast horizon for reservoir operation. *Water Resources Research* **55** (5), 4066–4081.

First received 7 October 2022; accepted in revised form 14 August 2023. Available online 1 September 2023