AGU ADVANCING EARTH AND SPACE SCIENCE

Water Resources Research®

RESEARCH ARTICLE
10.1029/2020WR029262

Key Points:

- We create a modeling framework that allows agriculture water users to learn and adapt to institutional and climatic changes
- This framework enables assessments of the irrigation water consumption impacts on regional water resources management
- The proposed reinforcement learning algorithm is generalizable for coupled human-nature systems

Supporting Information:

Supporting Information may be found in the online version of this article.

Correspondence to:

Y. C. E. Yang,
yey217@lehigh.edu

Citation:

Hung, F., & Yang, Y. C. E. (2021). Assessing adaptive irrigation impacts on water scarcity in nonstationary environments—A multi-agent reinforcement learning approach. *Water Resources Research*, *57*, e2020WR029262. https://doi.org/10.1029/2020WR029262

Received 17 NOV 2020
Accepted 3 SEP 2021

Assessing Adaptive Irrigation Impacts on Water Scarcity in Nonstationary Environments—A Multi-Agent Reinforcement Learning Approach

Fengwei Hung[1] **and Y. C. Ethan Yang**[1]

[1]Department of Civil and Environmental Engineering, Lehigh University, Bethlehem, PA, USA

Abstract One major challenge in water resource management is to balance the uncertain and nonstationary water demands and supplies caused by the changing anthropogenic and hydroclimate conditions. To address this issue, we developed a reinforcement learning agent-based modeling (RL-ABM) framework where agents (agriculture water users) are able to learn and adjust water demands based on their interactions with the water systems. The intelligent agents are created by a reinforcement learning algorithm adapted from the Q-learning algorithm. We illustrated this framework in a case study where the RL-ABM is two-way coupled with the Colorado River Simulation System (CRSS), a long-term planning model used for the administration of the Colorado River Basin, for assessing agriculture water uses impacts on water scarcity. Seventy-eight intelligent agents are simulated, which can be grouped into three categories based on their parameter values: the "aggressive" (swift actions; low regrets), the "forward-looking conservative" (mild actions; high regrets; fast learning), and the "myopic conservative" (mild actions; median regrets; slow learning). The ABM-CRSS results showed that the major reservoirs in the Upper Colorado Basin might experience more frequent water shortages due to the increasing water uses compared to the original CRSS results. If the drought continues, the case study also demonstrates that agents can learn and adjust their demands.

1. Introduction

Since the launch of the Harvard Water Program in 1955, the field of water resources management has been dedicated to providing practical solutions for complex water systems (Reuss, 2003). However, these efforts are hindered by the uncertainty and nonstationarity rooted in both the human (demand) and natural (supply) systems (Brown et al., 2015; Cosgrove & Loucks, 2015; Herman et al., 2020). Several planning approaches have emerged to address this issue, which can be grouped into three categories: dynamic planning, risk-based planning, and robust planning. The first category, dynamic planning, seeks a set of decision rules that yields the optimal outcomes (Castelletti et al., 2010; Harrison, 2007; Higgins et al., 2008). Risk-based planning aims at generating solutions under different risk levels and explores tradeoffs between risks and management objectives (Borgomeo et al., 2018; Hall et al., 2020; Lund, 2002), whereas robust planning looks for strategies that are robust (acceptable) across a variety of future conditions (Lempert & Collins, 2007). These categories are not mutually exclusive as several hybrid methods have been proposed. For instance, Dynamic Adaptive Policy Pathways incorporates robust decision-making into dynamic planning to determine the timing and actions for adaptation (Haasnoot et al., 2013; Kwakkel et al., 2015, 2016). Many-objective Robust Decision Making explores robust planning alternatives, plans that perform well in a set of future conditions, and the tradeoffs among management objectives (including risks) (Kasprzyk et al., 2013; Watson & Kasprzyk, 2017; Yan et al., 2017). Another example is risk-based stochastic programming which enhances dynamic planning by considering risk-aversion to avoid undesirable outcomes (Piantadosi et al., 2008) and learning by doing to improve decision-making adaptively (Hung & Hobbs, 2019). Nevertheless, the methods adopted in these studies often do not acknowledge the active interactions among human water users (stakeholders) and natural water supply. Water systems are inherently complex human-nature systems as stakeholders can interact with each other and the environment. Neglecting or simplifying the dynamics among stakeholders and stakeholders' adaptive behaviors may lead to biased conclusions.

Studies have indicated that incorporation of stakeholders objectives (Gold et al., 2019; Hadjimichael et al., 2020; Quinn et al., 2017) and consideration of stakeholders' cognitive beliefs and values (Glynn

et al., 2018; Moallemi et al., 2020), are critical for managing complex human-nature systems. Originating from the Artificial Intelligence community, agent-based modeling (ABM) was introduced to the water resources field in the late 2000s (Berglund, 2015). Since then, ABMs have emerged for assessing human systems dynamics and impacts on water systems (e.g., Al-Amin et al., 2018; Berglund, 2015; Ng et al., 2011; Yang et al., 2009). ABM is a distributed, bottom-up planning approach for understanding human impacts on system performance. An agent in an ABM is an object that interacts with other agents and the system of interest (a virtual environment, e.g., water resources models).

Berglund (2015) summarized earlier ABM development and developed case studies of two distinct agent types for water resources planning: the reactive agents respond to environmental signals based on behavioral rules, and the active agents pursue strategies to optimize their objectives. The former is also known as a descriptive model, and the latter a normative model (Smith, 1991). Recently, several ABM frameworks have been proposed to address the natural uncertainties and nonstationarity of complex human-natural systems. Giuliani et al. (2016) proposed a normative ABM framework to investigate the coevolution of agricultural water systems under climate change. Hyun et al. (2019) and Yang et al. (2020) applied psychological theories to simulate farmers' irrigation decision-making under uncertainty. Al-Amin et al. (2018) coupled a descriptive ABM with a groundwater model to evaluate water restriction programs in a watershed with multiple cities under future climate scenarios. Castilla-Rho et al. (2015) developed a similar ABM-groundwater model framework with a focus on participatory modeling. Generally, normative ABMs assume rationality and disregard human cognitive activities (e.g., learning and risk attitude), while descriptive ABMs only simulate encoded behavior rules that may misrepresent human response when the environment changes. We believe this research gap can be addressed, at least partially, by reinforcement learning (RL) algorithms.

RL is an area of machine learning where an intelligent agent can learn and improve its decision-making to optimize the long-term reward (Sutton & Barto, 1998). Depending on the context of the problems, RL methods can be either learning algorithms for searching optimal policy (a set of rules that guides an agent's actions in RL's terms) or decision-making models for simulating human adaptive behaviors (Seo & Lee, 2017; Sutton & Barto, 1998). The former is termed as *learning* methods, and the latter termed *planning* methods in the RL community (Sutton, 1992). RL has two essential characteristics: trial-and-error search and delayed reward, which enable agents to adapt their strategies by interacting with the environment. Since the introduction of RL to the water resources field, it has been extensively applied for optimal reservoir operations (Castelletti et al., 2010; Dariane & Moradi, 2016; Lee & Labadie, 2007; Madani & Hooshyar, 2014; Rieker & Labadie, 2012), with a few exceptions in dam sizing (Bertoni et al., 2020) and water and natural resource allocations (Bone & Dragićević, 2009; Ni et al., 2014). However, despite its increasing popularity, these applications only focus on RL's learning aspect for finding optimal policies in a stationary environment.

Focusing on the RL's planning aspect, this paper proposes an RL-ABM framework that equips agents (i.e., the agriculture water users in the case study) with the ability to adapt to a changing water system. The agents' decisions are the water quantity requests submitted to the water resources administration, and the water system is assumed under the impacts of climate change and growing water consumption. Rooted in psychology and cognitive science, the RL algorithm is advantageous, compared to the aforementioned methods, in simulating the constant changes in human beliefs and strategies through interactions with the environment. This feature is critical for the simulation of human reactions in nonstationary systems. Moreover, RL's parameters are relevant to human cognitive activities and can be used to characterize agents' behaviors (i.e., water diversion patterns in response to signals of environmental changes from the water system).

The RL-ABM framework is applied to the Colorado River Basin (CRB), United States (US), as an illustrative case study. The CRB is one of the most critical water sources in the Western US and Mexico and is facing increasing water stress due to recent droughts and the warming climate (US Bureau of Reclamation, 2012). The Colorado River Simulation System (CRSS), a water resources management model for CRB developed by the US Bureau of Reclamation (USBR) for reservoir operations and policy evaluations (US Bureau of Reclamation, 2007a; Zagona et al., 2001), is adopted in the case study as the virtual environment to be coupled with the RL-ABM for the assessment of water system response to dynamic agriculture water demands. The RL-ABM agents (agriculture water users in the CRSS) can be a single farm, an irrigation ditch, an irrigation district, American tribal water users, or a group of farming entities. Therefore, an agent represents only

the collective behavior of the water users in that group, but it may not represent the individual water users' decision-making.

In summary, this paper's contribution includes: (a) a modeling approach for studying human adaptation impacts on nonstationary water resources systems with a focus on the human cognitive aspect, and (b) an RL algorithm for simulating agriculture water users' adaptive water consumption decisions that incorporate additional water availability information. Moreover, RL-ABM parameters can better characterize an agent's cognitive processes comparing to Hyun et al. (2019), which can provide information about agents' reactions to environmental changes.

The remainder of this paper is organized as follows. Section 2 describes the modeling framework for agriculture water users' adaptive policy and the ABM-CRSS coupling. Section 3 introduces the case study. The results are shown in Section 4, followed by the discussion in Section 5. Finally, we present our conclusions and final remarks in Section 6.

2. RL-ABM Framework

The proposed RL-ABM framework builds on the widely applied Q-learning algorithm (C. Watkins, 1989; Watkins & Dayan, 1992). Q-learning is an algorithm that maximizes the expected long-term reward in a finite Markov Decision Process (MDP) environment by learning to improve the policy (a set of rules for selecting an action). Our method focuses on simulating an agriculture water user's decision-making and assumes that water users view the water system as a finite MDP. However, water users cannot directly observe the state of the environment (i.e., the maximum available water supply at the time and location where a decision is made) but simply accept their observations (i.e., water received) as the state of the environment. That is, the water users are facing a Partial Observable MDP (POMDP) problem (Kaelbling et al., 1998; Monahan, 1982).

This section introduces a variant of the Q-learning algorithm for a POMDP water system in which agriculture water users are modeled as agents. Specifically, our method incorporates realistic considerations of agents' diversion decisions: the inclusion of agents' perception about future water availability, the update of that perception in the form of prior transition probabilities, agents' decision uncertainty, the rate of learning, and parameters related to risk-perception and discounting.

2.1. Q-Learning Algorithm

The Q-learning algorithm simulates learning processes similar to the temporal differences method (Sutton, 1988), in which an agent takes an action a following a policy π based on the state s, evaluates the immediate reward $R(s,a)$ received by taking the action, and then updates the policy. In Q-learning, a policy π is in the form of a value function (called Q-function), which returns the expected value given a state-action pair. The definition of a generic Q-function is:

$$Q(s,a) = R(s,a) + \gamma \sum_{s' \in S} P(s,s') \max_{a'} Q(s',a') \quad (1)$$

where $Q(s,a)$ is the Q-function, γ is a discounting factor, $P(s, s')$ is the transition probability from current state s to the next state s', S is the set of states, and A is the set of actions. The discounting factor $\gamma \in (0,1)$ is a parameter that represents how the agent values future rewards. Therefore, the second term in Equation 1 is the expected future rewards discounted.

In a dynamic setting (i.e., multiple time steps involved in a decision process), the optimal action a^* is determined by the Q-function $Q_t(s,a)$ at that time t given the state s and the expected long-term rewards (i.e., the agent will choose an action that yields the highest Q value throughout the process). The Q-function is actively adjusted through an iterative process known as the Bellman equation, which is the optimality condition for a dynamic system:

$$Q_t(s,a^*) = R_t(s,a^*) + \gamma \sum_{s' \in S} P_t(s,s') \max_{a'} Q_{t-1}(s',a') \quad (2)$$

Building on the Bellman equation, Q-learning adapts a learning rate parameter α to control how quickly an agent responds to new information, as Equation 3 shows. A time index t is also added to the state and action variables (s_t and a_t) to distinguish their values at different times. The value of α is between 0 and 1. $\alpha = 0$ means the agent refuses to learn and rejects the new information, while $\alpha = 1$ means the agent fully accepts the new information and discards its prior belief.

$$Q_t(s_t, a_t) = (1-\alpha)Q_{t-1}(s_t, a_t) + \alpha(R_t(s_t, a_t) + \gamma \sum_{s' \in S}\left(P_t(s_t, s')\max_{a'} Q_{t-1}(s', a')\right) \quad (3)$$

Equation 3 is the update of the Q-function that accounts for the learning rate and delayed (future) rewards. Q-learning aims to balance exploration (actions for learning) and exploitation (actions for immediate rewards). Equation 3 alone only allows agents to take exploitation actions, except for radical changes in the environment, which would likely result in suboptimal solutions. Consequently, Q-learning introduces a trial-and-error mechanism to ensure sufficient exploration actions are taken. Among the exploration strategies, ϵ-greedy algorithm is the most widely applied for ease of use and tuning (Tijsma et al., 2017) and, thus, is applied in this study.

$$a_t = \begin{cases} a^* = \text{argmax}\, Q_t^*(s_t, a_t), \text{with probability}(1-\varepsilon) \\ a^* \neq a^*, \text{with a total probability } \varepsilon \end{cases} \quad (4)$$

where ϵ is the exploration rate. Essentially, ϵ-greedy agents adopt a randomized strategy: they take the optimal action with probability $(1-\epsilon)$ or other actions with probability ϵ (Equation 4). Therefore, a higher value of ϵ means that the agent is more likely to take exploration actions. Whereas $\epsilon = 0$ means agents always take actions to maximize the expected rewards and never risk learning new information.

2.2. Farmer's Q-Learning (FQL) Algorithm

To apply the Q-learning algorithm for agriculture water users' irrigation simulation, we first define a POMDP water system as a tuple (S, A, T, R, O, Ω). The POMDP components are explained below.

1. S is a finite set representing the states of the water system perceived by an agent. We defined the state of the water system (denoted s_t) the maximum available water for an agent at the location and time of a diversion decision (denoted $s_t \in S$; each agent has its own S)
2. A is a finite set of actions. The actions can be deterministic or stochastic. In this paper, we adopted a stochastic model to improve the representation of the variability in agents' diversions. The variability represents variables not included in this study and random errors. For simplicity, we consider only two actions in the case study: to request an increase (a^+) or a decrease (a^-) in the diversion. The actions (a^+ and a^-) indicate the direction of diversion change, and the amount of the change (ΔDiv) is assumed following a half-normal distribution ($\Delta Div \sim |X|$, X is a normal random variable). Thus, the decision includes two steps: (a) select a direction (a^+ and a^-) and (b) determine the requested increase or decrease quantity by drawing a random number from the half-normal distribution
3. T is the state-transition model that determines the agents' perceived state of the water system
4. R is the immediate reward function representing an agriculture water user's utility. In Equation 5, the *Reward* is the crop production as a function of water consumption (Div_t), the *Penalty* is the cost induced by prediction errors (e.g., the sunk costs of sowing and fertilizers), and the *regret* (a scaler) is a normalizing factor to combine reward and penalty into one utility function

$$R = \text{Reward} - \text{Penalty} = Div_t - \text{regret} * \left(Div_{\text{requested}} - Div_t\right) \quad (5)$$

where Div_t is the diversion received at time t, $\text{Div}_{\text{requested}}$ is the diversion received (Div_{t-1}) plus the change requested (ΔDiv) in the previous time step. Details about the reward function derivation are presented in the supporting information (Text S2).

1. O is the observation function, which generates observations for each action and state combination. An observation consists of two signals, the water availability information O_f and the diversion water received Div_t. The water availability information O_f is a reference for future water availability and external information that affects agents' decisions (e.g., precipitation forecasts and dam water levels)
2. Ω is a finite set of observations that may reveal to an agent

While an agent interacts with the POMDP, it makes observations, generates actions, and keeps updating an internal belief that summarizes its experience. From the human cognitive perspective, the state of the system $s_t \in S$ is an agent's perception about the available water quantity in the POMDP environment. Depending on the agents' locations in the river system and its interpretation of the observation, agents can have their own perceived state s and set S. Furthermore, since an agent cannot know the true (maximum) quantity of the available water, we need a function, called a state estimator (SE), to translate the observation Ointo the perceived state s. We assume that agents believe the water received (i.e., Div_t) are the maximum quantity available, so we can construct an SE that maps the diversion received (a continuous variable Div_t) to the state of the system (a discrete variable s).

The water quantity delivered to an agent (Div_t) and the maximum water quantify available (not directly observable) for the agent are simulated by a water resources management model (i.e., the CRSS in the case study) according to the agent's diversion request ($\mathrm{Div}_{\mathrm{requested}}$), the water budget at the agent's location, and the water allocation rules of the model. Thus, the water resources management model serves as the state-transition model T and the observation function O. More details of the CRSS are presented in the next section.

The transition probability P is an agent's prior belief, which is constructed based on the agent's experience. We assumed that agents do not fully understand the water allocation rules nor do they have the perfect information about future flows, so they view the state-transition process as a stochastic process. An agent's belief updating requires another SE to translate the water availability information O_f into water availability index $f_t \in \{0,1\}$ where $f_t = 1$ indicates an expectation of more water (e.g., a higher dam water level than the previous year) and $f_t = 0$ means the opposite. Upon received the observations f_t and Div_t, we can calculate the reward R_t, and update the Q-functions Q_t and transition probabilities P_t. The updating of Q_t and P_t can only happen at the start of each time step until the reward of a decision is known. Therefore, the update of the Q-function is revised as follows.

$$Q_t\left(f_{t-1}, a_{t-1}, s_{t-1}\right) = (1-\alpha)Q_{t-1}\left(f_{t-1}, a_{t-1}, s_{t-1}\right) + \alpha\left[R_t + \gamma\sum_{s'\in S}P_t\left(f_{t-1}, a_{t-1}, s_{t-1}, s'\right)\mathrm{Max}_{a'}\,Q_{t-1}\left(f_{t-1}, s', a'\right)\right] \quad (6)$$

Since the system is Markovian, updating a prior transition probability over the states requires only the knowledge of the previous time step and the current observation. The updating method here is similar to the replacing eligibility trace method in which the probability decays linearly every time a state is visited (Sutton & Singh, 1996). The idea is to generate a short-term memory process that reinforces the state visited and gradually decays the others over time. This belief update is computed using Equation 7.

$$P_t\left(f_{t-1}, a_{t-1}, s_{t-1}\right) = \begin{cases} \frac{\tau_s - 1}{\tau_s}P_{t-1}\left(f_{t-1}, a_{t-1}, s_{t-1}, s'\right) + \frac{1}{\tau_s}, s' = s_t \\ \frac{\tau_s - 1}{\tau_s}P_{t-1}\left(f_{t-1}, a_{t-1}, s_{t-1}, s'\right), \forall s' \neq s_t \end{cases} \quad (7)$$

where τ_s is the normalizing factor and $\frac{\tau_s - 1}{\tau_s}$ is the memory decay rate of state s. This formulation enables us to assign different memory decay rates ($\frac{\tau_s - 1}{\tau_s}$) to different states based on frequency so that the memory of rare events is stronger than frequent events. For example, a prior belief from state 1 to state 2 is 0.1 ($P_{1,2} = 0.1$), $\tau_1 = 2$ will result in a posterior belief $P_{1,2} = 0.55$ whereas, $\tau_1 = 10$ will yield a posterior belief $P_{1,2} = 0.19$. The underlying assumption is that people respond to rare extreme events (e.g., extreme and long-lasting droughts) more swiftly than frequent low-intensity events (e.g., small and frequent water deficits). This view is supported by human and organizational learning studies, but it is just one of the competing views of rare event learning (Lampel et al., 2009; Starbuck, 2009). For example, studies also indicate that people may not learn if they attribute rare events to unforeseeable external circumstances, or they have a strong prior belief and the impact of the event is insignificant.

We named this version of Q-learning the FQL algorithm. Figure 1 shows the flowchart of the FQL algorithm. At time t, an agent will receive a diversion Div_t and information about future water availability $O_{f,t}$. Then we convert the observations (Div_t and $O_{f,t}$) to discrete index (s_t, f_t), calculate the reward R_t, and update the transition probability P_t and the Q-function Q_t. The agent will decide whether to increase or decrease diversion (a_t) based on the updated Q-function and the ϵ-greedy algorithm: if a random number x is greater than or equal to the user-specified probability ϵ, the optimal action a^* is taken; otherwise, the alternative action $\overline{a^*}$ is chosen. The new ΔDiv is randomly generated based on the chosen action. Finally, the requested diversion $Div_{requested}$, which is the diversion at t (Div_t) plus the requested change in diversion (ΔDiv), is sent

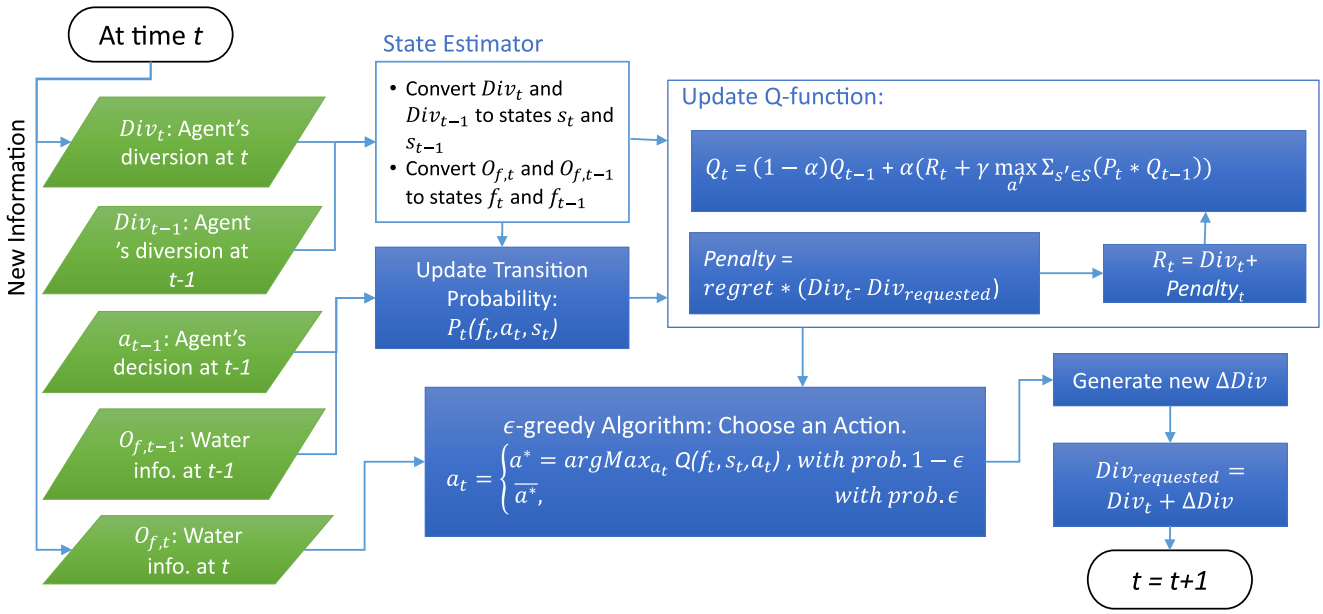

Figure 1. The computation flowchart of the Farmer's Q-Learning algorithm. Green parallelograms represent data, and blue rectangles are computation processes.

to the POMDP environment. A summary of the FQL parameters is presented in Table S1. The development of this FQL follows the standard ODD + D protocol for ABM (Müller et al., 2013), and more details of the model description are presented in Table S2

3. Colorado River Basin Case Study

3.1. Colorado River Basin (CRB) and the Colorado River Simulation System (CRSS)

The CRB generates approximately 20.23 billion m^3 (16.4 million acre-feet, MAF) of freshwater annually to support 40 million people and irrigate 2.2 million hectares of land in the Western US. However, due to the increasing water demand and the warming climate, water scarcity in CRB has become a pressing issue (Castle et al., 2014; Garrick et al., 2008). The water allocation in CRB is based on numerous compacts, federal laws, court decisions and decrees, contracts, and regulatory guidelines, collectively known as the "Law of the River" (Stern & Sheikh, 2019). USBR regulates the CRB water distribution in the US through operations of reservoirs, such as Glen Canyon Dam (Lake Powell) and Hoover Dam (Lake Mead). The management in CRB is divided into two areas: the Upper and Lower Basins (denoted UB and LB, respectively) bordered at Lee Ferry, Arizona, set by the 1922 Colorado River Compact. The 1922 Compact states that the UB States (comprising Colorado, New Mexico, Utah, Wyoming, and the upper part of Arizona) will not cause the flow to be depleted below an aggregate of 9.25 billion m^3 (7.5 MAF) in any period of 10 consecutive years for the consumptive uses of the LB States (Arizona, California, and Nevada).

The USBR develops the CRSS as a planning tool for analyses and discussions of water issues in the CRB, which comprises a set of water allocation rules and a river network with 12 reservoirs, 29 headwater tributaries, and 520 water users. For the future water supply of the basin, the CRSS provides multiple options, including resampling the historical flow or tree-ring records and simulation outputs from climate and hydrologic models. As for the future demands, the CRSS uses the values estimated by the USBR in coordination with Mexico and the Basin States. In addition, the CRSS has a set of rules that are designed to simulate the water allocation and dam operation based on "the Law of the River." For additional details of the CRSS, please refer to the USBR's website (https://www.usbr.gov/lc/region/g4000/riverops/model-info-APR2018.html) and the newly published whitepaper (Wheeler et al., 2019).

Coupling RL-ABM with CRSS enables agents to interact with the environment and adjust their policies based on the feedback from the CRSS. From the POMDP perspective, CRSS serves as both the state-transition

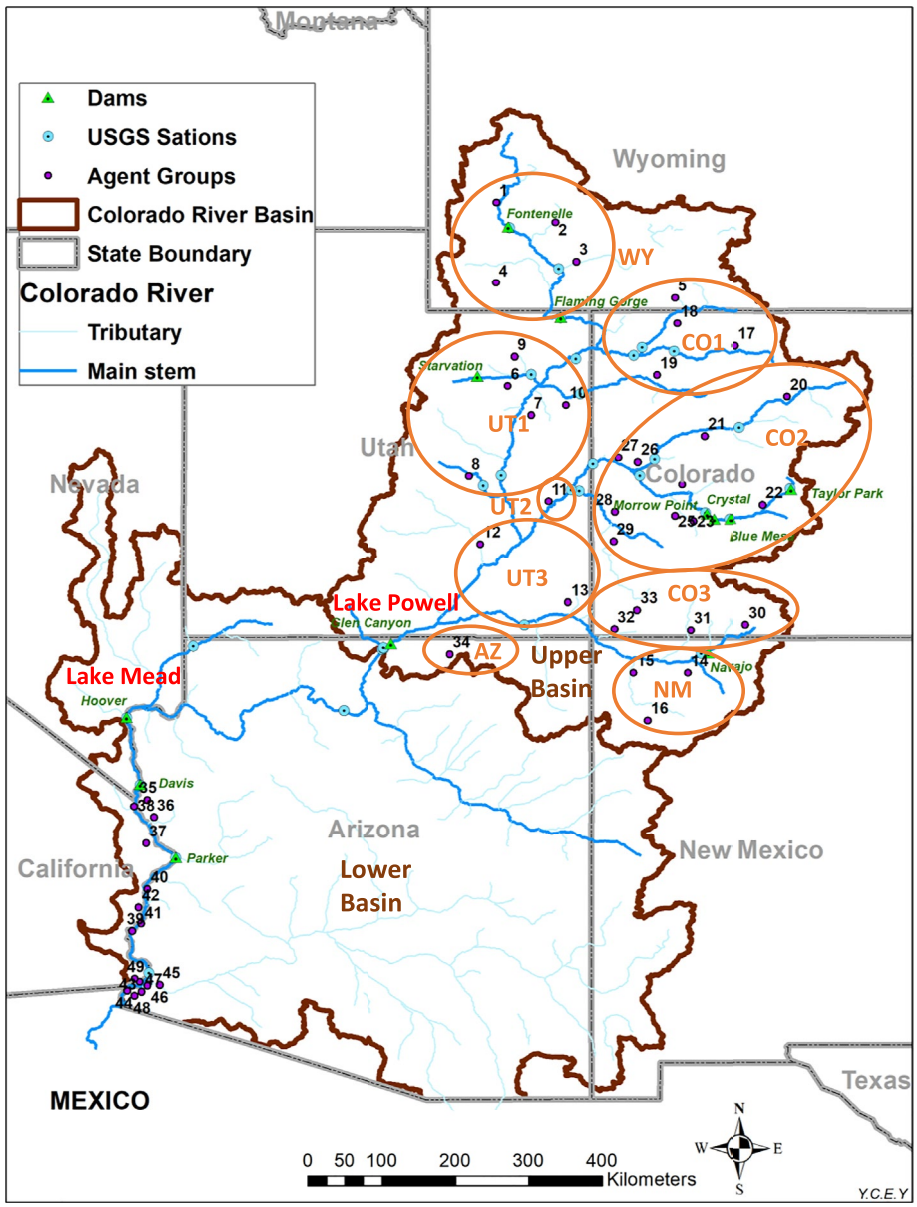

Figure 2. The map of the Colorado River Basin boundary (brown line; divided into Upper (UB) and Lower Basins at Glen Canyon Dam), the state boundary, the agent groups (points with numbers), the major reservoirs (triangles), and the sub-basins in the UB (orange circles).

function T and the observation function O. CRSS determines the water quantity available for each agent through water balance at the agent's location in the river network and the allocation rules. Since flows in the UB are from precipitation and in the LB are mainly regulated by reservoirs, we define the water information (O_f) in the UB and LB, the *winter precipitation* and the *Lake Mead water elevation*, respectively. Moreover, we used the CRSS's historical resampling method to generate future flow in the case study so that we can compile winter precipitation data using PRISM monthly precipitation products (Oregon State University, 2020) for ABM accordingly.

3.2. ABM Setup for the UB and LB

Figure 2 shows the map of the CRB, the major reservoirs, and the agriculture water groups in CRSS (49 groups; one group can have multiple agents). Among the water user groups, we identified 56 agents in 34

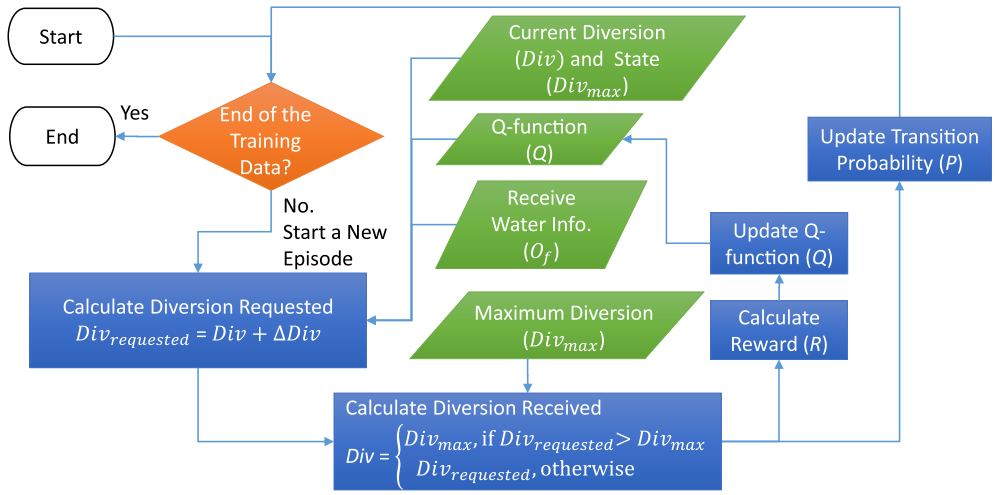

Figure 3. The schematic of the Farmer's Q-Learning training processes.

groups in the UB and another 22 agents in 15 groups in the LB. The locations of the water user groups show where the diversions happen but not where the water is consumed. The numbers in Figure 2 are the water user group IDs; the group and agent names are presented in Tables S3 and S4. Moreover, the orange circles in Figure 2 indicate the sub-basins in the UB, which will be applied for training UB agents and explained in the following section. Agents' state variables are discretized into 21 discrete states with the median (11th) state being the 2018 diversion rounded, so the increment of the state levels is approximately 10% of the 2018 diversion.

3.3. Agent Training and Testing

Before coupling ABM with CRSS, we need to estimate agents' initial transition probabilities (P) and Q-functions (Q). The initial P and Q are obtained by training agents in an environment constructed from historical data. Then the trained ABM is applied to simulate historical diversion, and the results are compared with the observations to assess ABM model performance.

Historically, the maximum available water ($Div_{\max}$) for an agriculture water user was jointly determined by the water regulations, dam operations, and the streamflow at the location and time of the diversion. We assumed the maximum available water for diversion is equal to or higher than the diversion recorded. Thus, we can apply a linear function to approximate the historical $Div_{\max}$: $Div_{\max} = (1+\rho) * Div$. The number $\rho \in (0,1)$ is a fraction and a threshold for penalty. $\rho = 0.2$ is chosen for training in the case study based on trial-and-error. Then, we constructed a joint cumulative probability density function (CDF) of the maximum available water for diversion and water information (O_f) using the historical data. With the CDF, we applied Monte Carlo (MC) methods to generate, for that agent, the system's state and water information $(Div_{\max}, O_f)$ as the training data set. Alternatively, historical streamflow data combined with statistical methods may provide more accurate $Div_{\max}$ estimation, which could be a future research direction.

In training, each agent is an independent sub-model trained alone in a training environment (i.e., the MC training data). Figure 3 shows the schematic of the training processes for one agent. The initial transition probabilities are uninformative (equal probability for each state), and the initial Q-functions are zero functions (a constant function that always generates zero). The agent training loops through all the training data, and each data point $(Div_{\max}, O_f)$ represents an episode (in machine learning term; a simulation). In each episode, an agent is provided with the water diversion (Div), the state (s), the Q-function (Q), and the water information (O_f). Based on the information received, the agent then takes an action a, determines the quantity of change ΔDiv, and submits a new diversion request ($Div_{\text{requested}}$). If the request does not exceed the maximum diversion ($Div_{\max}$) of this episode, the diversion (Div)is set to the requested quantity; otherwise,

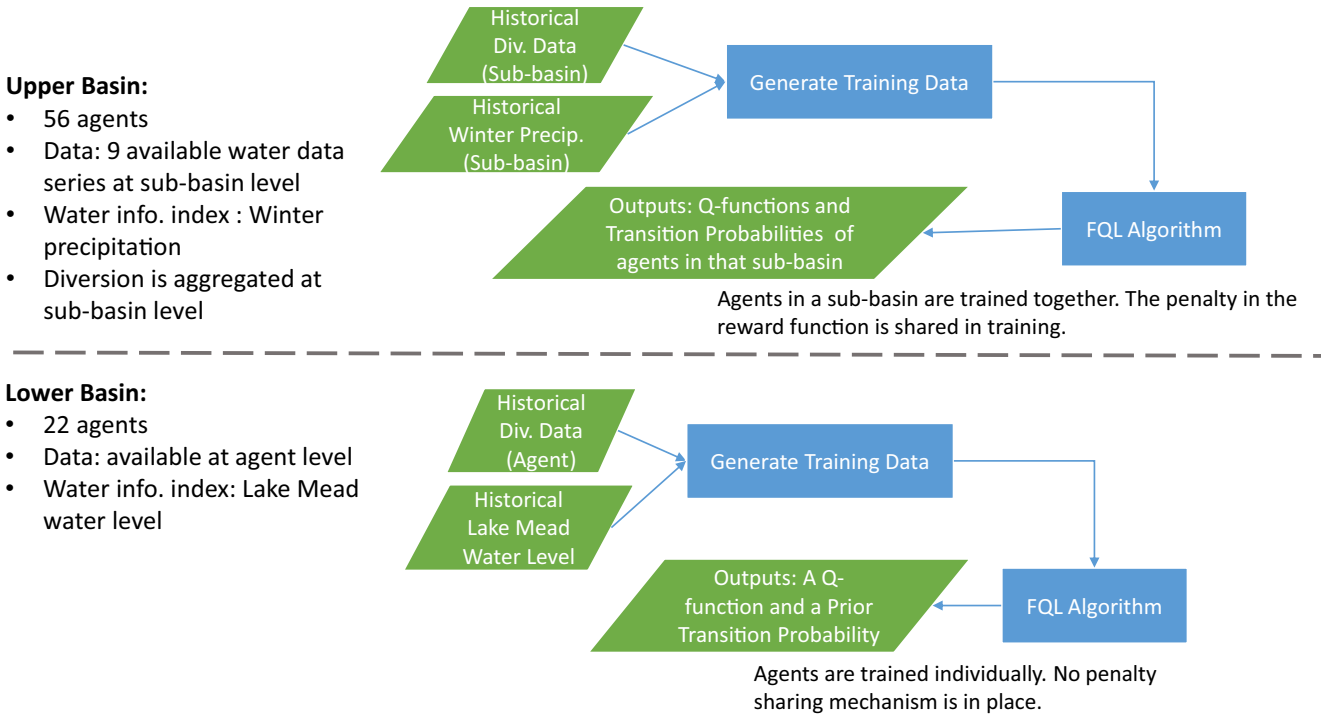

Figure 4. The training schematics for the Upper and Lower Basins agents and the differences.

the diversion is set to the maximum value. Then, the state (s), transition probability P, and Q-function Q are updated as well before entering the next episode.

The individual diversion data of LB agents are available. However, in the UB, complete (all UB States) and consistent (same temporal and spatial scales) agriculture depletion data can only be found at sub-basin scale (as indicated by the orange circles in Figure 2, US Bureau of Reclamation, 2015). To address the discrepancy between modeling structure and data, we modified the training procedure to enable UB agents in one sub-basin to be trained together, and the shortage is also shared based on the agent's diversion requests at the time. An alternative is to estimate agent historical diversions based on their current usage, but that could introduce errors to the model since, diversion partition in a sub-basin may not be stationary. The training can be improved if agent-level diversion information and shortage allocation rules become available. The water depletion data is converted to diversion by multiplying a scaler of two (assuming 50% of the diversion is depleted, which is the method used in CRSS). Figure 4 shows the training schematics for both UB and LB agents. Some states (such as Colorado) have detailed diversion data that is publicly available but is not included in this paper for data consistency.

Furthermore, the winter precipitation information is retrieved from PRISM at the US NOAA weather stations in that sub-basin. The historical dam water level, diversion, and depletion data applied in training are retrieved from the USBR website (https://www.usbr.gov/). Only data from 1971 to 2018 are applied for two reasons: (a) data availability and (b) no new treaty and major infrastructure installed after 1971. A summary of the training data used in the case study is provided in Table S5.

3.4. ABM Parameterization and Coupling With CRSS

The parameter values of the FQL algorithm are critical in modeling agents' behaviors and required calibration and validation before coupling with CRSS. Since the RL-ABM is a stochastic model, our parameterization focuses on reproducing the historical diversion trend and the variability. The Kling-Gupta efficiency (KGE) (Gupta et al., 2009) captures the similarity in probability distribution (mean and variance) and the Pearson correlation of two data series; thus, it is chosen as the performance criterion for training.

The parameters are associated with an agent's characteristics: the ability to learn (ϵ, α), the risk attitude ($\mu, \sigma, regret$), and the discounting of future rewards (γ). In the FQL algorithm, the parameter ϵ dictates how often an agent's decision would deviate from the optimal action; μ and σ are the mean and standard deviation of an agent's requested changes in diversion; the parameter *regret* is the unit penalty of an agent's unsatisfied demand.

The parameterization of ABM consists of five steps: (a) generate parameter sets as candidate ABMs, (b) use the training data set to train candidate ABMs and generate initial Q-functions and transition probabilities, (c) test the candidate ABMs in historical condition and calculate their KGE performance by comparing the results with historical data, and (d) repeat the testing multiple times, and (e) select a candidate ABM with the highest mean KGE value for each agent.

The training data is generated using the method described in Section 3.3. The size of the training data set is 2,000 so that agents have sufficient opportunities to explore and improve their Q-functions. For each agent (or sub-basin), we generated 100 candidate ABMs with parameters randomly sampled from the ranges listed below. UB agents in one sub-basin are assigned with the same parameter values, but they have their own Q-functions and transition probabilities. The parameter ranges are suggested in the literature (α, γ, ϵ) and by preliminary simulation results ($\mu, \sigma, regret$).

1. Mean diversion alteration μ : $(0, 0.5)$
2. Standard deviation of diversion alteration σ : $(0.5, 1.5)$
3. Learning rate α : $(0.5, 0.95)$
4. Discount rate γ : $(0.5, 0.95)$
5. Exploration rate ϵ : $(0.05, 0.3)$
6. Regret *regret* : $(0.5, 3)$

After training, the candidate ABMs are tested by simulating the diversion from 1971 to 2018 (48 years). This testing is repeated 30 times to minimize the influence of the stochasticity of the ABMs. The selected ABM for an agent consists of the best parameter set and the associated initial (trained) Q-function and transition probability. The choice of statistical metrics (e.g., mean and median of the KGE distribution) is context-dependent and can affect the parameters' selection. For example, if the focus of the study is the risk of low model performance, the infimum of the KGEs could be used.

Finally, the coupling of ABM (consists of the best models of all agents) and CRSS is accomplished by building communication between the two models, similar to the two-way coupling methods developed by Khan et al. (2017). When the CRSS finished flow simulation and water allocation, it outputs the results to ABM. Upon received the information from CRSS, ABM updates the Q-functions and transition probabilities and generates agents' diversion requests sent to the CRSS for the following year. Because the time scale in CRSS is monthly and ABM is running annually, agents' diversion requests generated from ABM are allocated to each month of a year based on individual agents' historical diversion patterns. For simplicity, we assume fixed monthly diversion distribution (based on the 2018's values in CRSS) and allocate the annual diversion requests accordingly.

4. Results

4.1. Agent Training Results

Figure 5 shows the results of the 30 test simulations (the gray lines), the historical diversion (the blue line) of the UB sub-basins, and the mean KGE values. Since UB agents in a sub-basin are trained collectively, the results are presented at sub-basin level. The mean KGE values of the UB sub-basins range from 0.28 (UT2) to 0.9 (AZ). Generally, the simulation results (Figure 5) can capture the historical diversion trends and transitional patterns except in the periods of high variability, for example, the peaks of the historical diversion in UT2 and UT3. Because the requested diversion change (ΔDiv) follows a half-normal distribution that tends to suggest a mild change (the higher values, the lower probability), it would take more time for the diversion to ramp up than the historical data. However, when a water shortage happens, the ABM will set an agent's diversion to the available diversion at the time (a much lower amount than their current diversion) so that

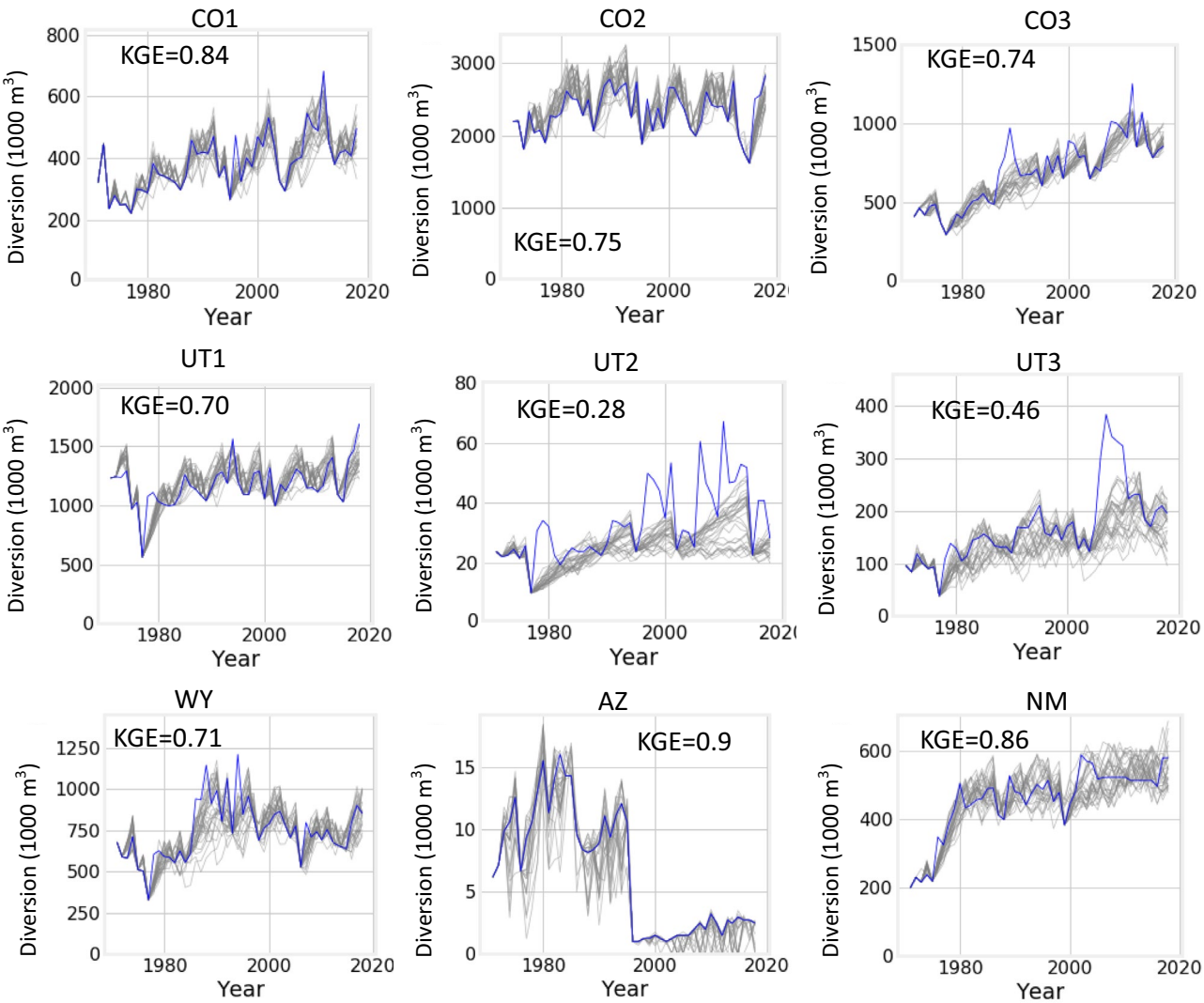

Figure 5. Training results of the Upper Basin agents grouped in sub-basins. The blue line is the historical diversion, and the gray lines are the 30 testing results.

the simulation results will always follow the diversion drops. This phenomenon is particularly evident in AZ's results in Figure 5 (bottom row, middle column). Additionally, UT2 has the lowest KGE value for the high variability in its historical diversion (diversion spans from 10,000 to 65,000 m^3). The variability may be caused by mechanisms absent in current RL-ABM, such as drought contingency plans and water conservation programs or large-scale new developments. Interestingly, the UT3's diversion increase in 2005–2011 can be partly explained by the dramatic changes in precipitation before and after that period. Utah had a four-year drought from 2001 to 2004 and another four-year drought after 2011 (National Integrated Drought Information System, 2021).

Among the agriculture water users in the LB, PVID, CRIR AZ, Wellton Mohawk IDD, Yuma County WUA, and Yuma Mesa IDD are the largest five in volume, which accounts for about 70% of the LB agriculture diversion. Therefore, we focus our discussion on these five agricultural water users here. In Figure 6, we can see that the mean KGE values of the five agents vary from 0.27 (PVID) to 0.67 (Wellton Mohawk IDD), and the results generally follow the patterns of the historical diversions. PVID has a low KGE value for a lower correlation (r =0.38) between the simulation and the historical data. One explanation of this result is that water information (the Lake Mead water level) is not the only consideration in PVID's diversion decisions. Other factors, such as crop selection and irrigation technology improvement, may be important but are

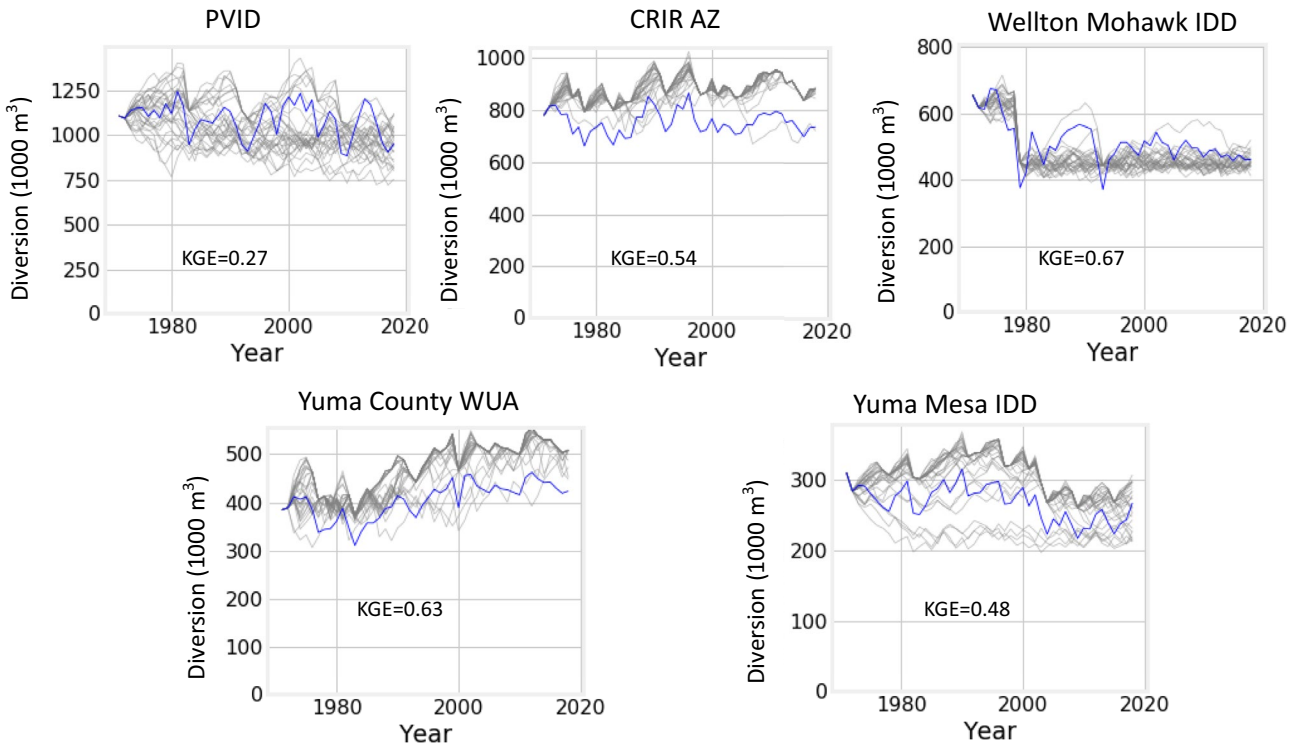

Figure 6. Training results of the largest five agents in the Lower Basins. The blue line is the historical diversion and the gray lines are the 30 testing results.

absent in our current model. Additionally, the results of CRIR AZ are constantly higher than the historical diversion due to the diversion prediction errors in the first 5 years (1971–1975), likely caused by factors absent in the current model. The prediction error then persists throughout the simulation.

The highest KGE values and the breakdown of KGE components (Pearson correlation "*r*", mean "*a*", and standard deviation "*b*" ratios) of the ABM testing are presented in Tables S6 (UB agents) and S8 (LB agents). The mean absolute error and root mean squared error of the best models are also presented in Tables S7 (UB agents) and S9 (LB agents) for reference.

4.2. Agent Characteristics and Clustering Analysis Under Historical Condition

The parameters of an agent determine how it reacts to environmental changes. By clustering parameter values, we can categorize agents into different types. The clustering method applied here is the K-means clustering. K-means clustering partitions the agents into k clusters in which each agent belongs to the cluster with the nearest mean (or centroid). Principal component analysis (PCA) can extract information from a high-dimensional space by projecting it into a lower-dimensional subspace. It is commonly combined with clustering methods, such as K-means, to help visualization and interpretation (Aubert et al., 2013).

Following the convention, we applied PCA to transform the parameters into uncorrelated principal components and reduce dimensions to 2D. Then we implemented K-means clustering on the agents with cluster numbers (k) set to 3. Table 1 shows the average parameter values of the three clusters. The three clusters are named by their characteristics: "Aggressive," "Forward-looking Conservative," and "Myopic Conservative." Agents in the Aggressive cluster have the highest distribution mean μ and standard deviation σ values of the action a and the lowest *regret*, which means the Aggressive agents tend to take more swift actions and do not regret much for the low penalty if a shortage happens. Conversely, agents in the Forward-looking Conservative cluster have the highest *regret*, learning rate α, and epsilon ϵ, which implies the agents are cautious in changing diversion but inclined to explore and accept new information. Agents in the Myopic Conservative cluster are also cautious in changing diversion (low μ and σ), but, unlike the Forward-looking

Table 1
The Average Parameter Values of the Three Clusters

	Mean: μ	Standard deviation: σ	Learning rate: α	Discount rate: γ	Epsilon: ϵ	*regret*
Aggressive	**0.36**	**1.22**	0.62	0.77	0.16	0.78
Forward-looking Conservative	0.20	0.60	**0.85**	**0.78**	**0.19**	**2.22**
Myopic Conservative	0.16	0.87	0.67	0.64	0.09	1.54

Note. The highest values of the parameters are highlighted in bold.

Conservative agents, they have the lowest learning rate α, discount rate γ, and epsilon ϵ. This implies that the Myopic Conservative agents tend to make decisions with existing knowledge and are reluctant to accept new information or even take exploration actions.

Figure 7 shows agents in the 2D principal component space. In the *x*-axis direction, the higher the *x* value, the more cautious the agent is toward changes. Whereas in the *y*-direction, the higher the *y* value, the less likely the agent is to take exploration actions and accept new information. Consequently, the Myopic Conservative agents are all located above the *x*-axis in Figure 7 except NorthBajaLLC, the Aggressive agents are all located on the left side of the *y*-axis, and the Forward-looking Conservative agents are only found on the right side of the *y*-axis. Geographically, UB agents and tribal agents in the LB are mostly Aggressive because of their increasing diversion patterns historically and underutilized water rights. The Forward-looking Conservative agents are all located below Parker Dam and close to the basin outlet. Being at the most downstream agents in the CRB under recent prolonged drought may have made farmers in the Forward-looking Conservative cluster more vigilant and willing to adapt to the system's changes. However, the analysis also shows that other LB agents in the Myopic Conservative cluster think the state of the system would not change much; therefore, they can continue relying on their prior knowledge for future diversion decision-making.

One caveat is that the cluster analysis results can only represent the agents' characteristics but not the actual water users (humans). The historical diversion records could be the results of local flow conditions and water regulation. One supporting evidence is the LB agents are all classified as conservative except three tribal water users (Fort Mohave Ind Res AZ, FtYumaReservation, and Chemehuevi Ind Res) due to the fully

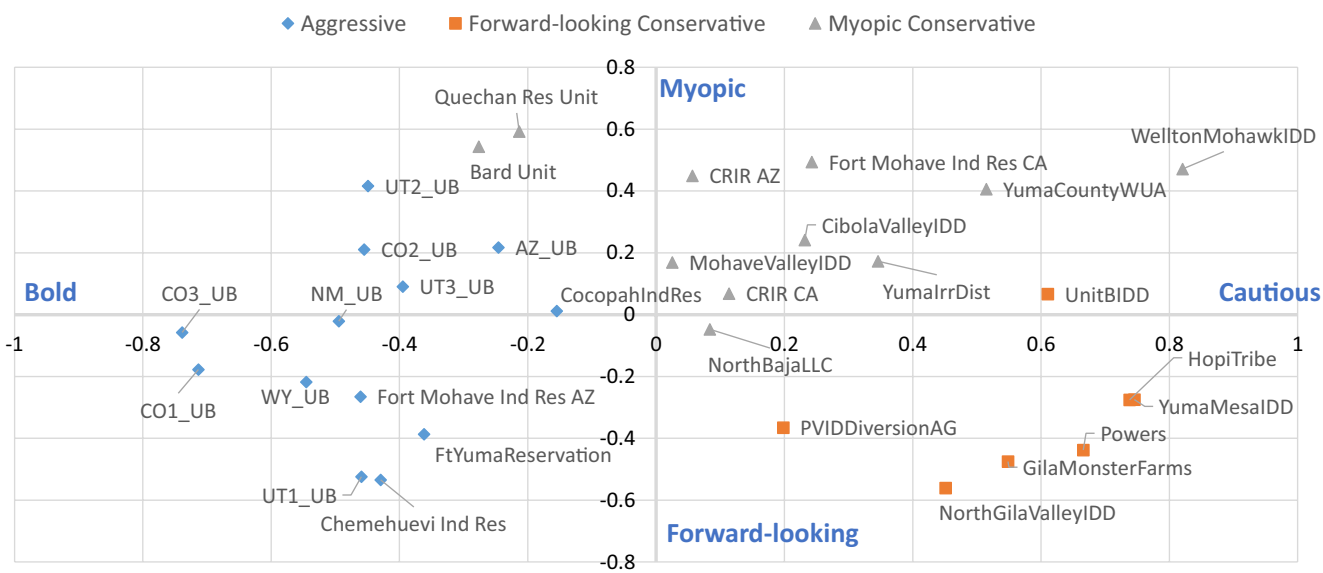

Figure 7. The clusters of agents in the principal component space. The x-component is an agent's attitude toward changes, and the y-component is the willingness to adapt. UB agents share the parameter values and, therefore, are labeled by sub-basins with "_UB." The Aggressive, Forward-looking Conservative, and Myopic Conservative clusters are marked by blue diamond, orange squares, and gray triangles, respectively.

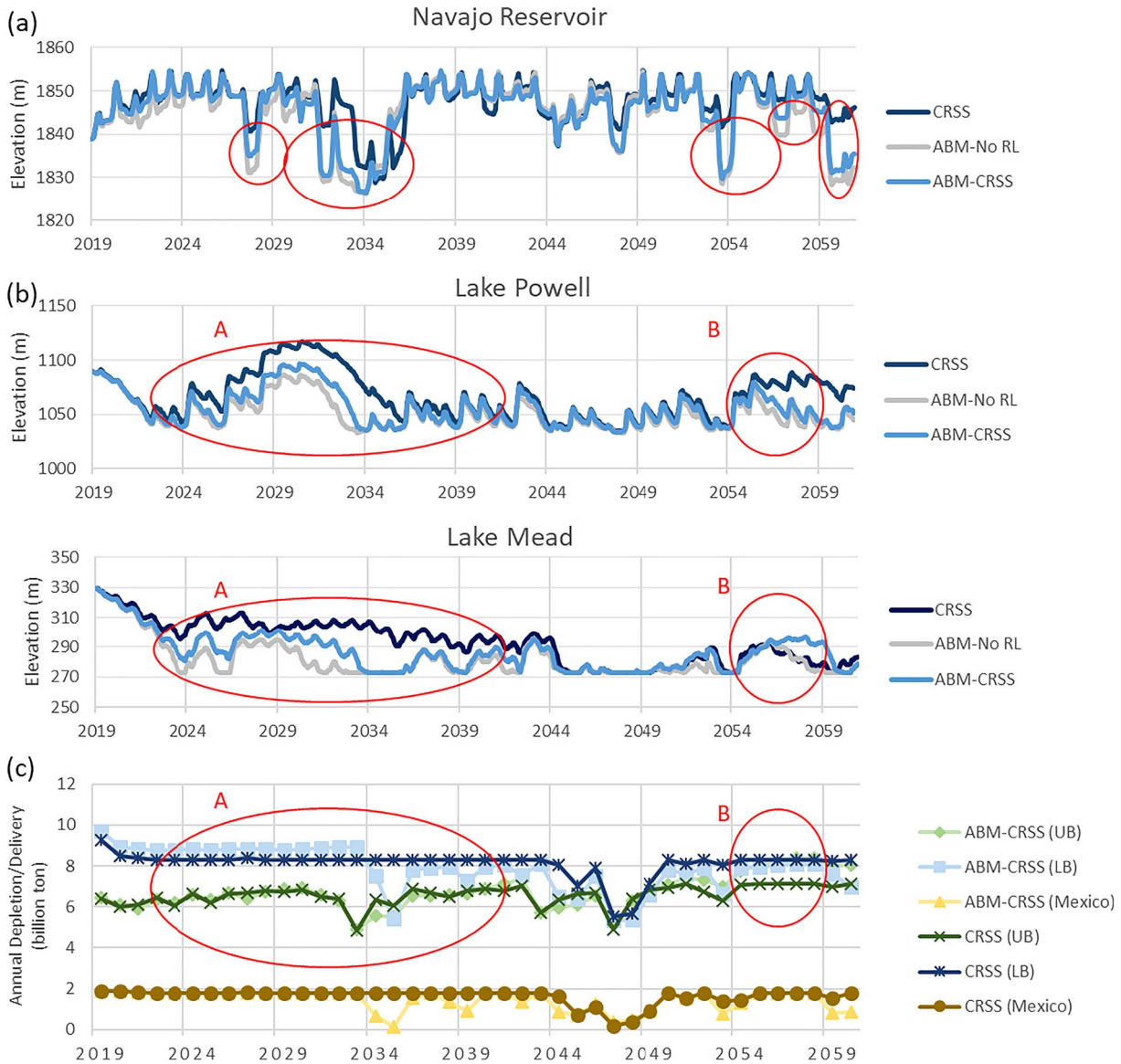

Figure 8. Comparison of the results from the agent-based modeling-Colorado River Simulation System (ABM-CRSS), the ABM-CRSS without learning (ABM-No RL), and the original CRSS. (a) The pool elevation of the Navajo Reservoir in which the red circles highlight the periods ABM-CRSS predicting a lower pool elevation. (b) The pool elevation in Lakes Powell and Mead. (c) The annual water consumption in the Upper Basin and Lower Basin, and the annual water delivery to Mexico. Circle A highlights the period ABM-CRSS predicting a lower pool elevation in both lakes, whereas Circle B indicates the period ABM-CRSS showing the opposite trend comparing to the original CRSS.

utilized water allocation (9.25 billion m^3 or 7.5 MAF). Supposing that the conditions have changed (e.g., more natural flows), the conservative agents may become more aggressive or vice versa. This means an agent's learning model (parameters values and model structure) should also be updated. This perspective will be a future direction.

4.3. Results of the Coupled ABM-CRSS Under a Drier-Than-Normal Future Condition

We investigate how agents affect the CRB's water resource management by comparing the results of the ABM-CRSS with the original CRSS and the ABM-CRSS with the learning disabled (denoted "ABM-No RL" in Figures 8a and 8b). In the ABM-No RL simulation, agents do not update their transition probability P and the Q-function and, thus, will take the optimal actions based on past experiences (i.e., the initial Q). The simulation period is set to 2019–2060. The future inflow condition can be described as drier-than-normal.

The inflow time series is constructed by using historical data of the periods 1988–2015 and 1934–1947 (Figure S1), which is the default setting of the original CRSS. Although the future flow shows an increasing trend, it remains lower than the historical average until the end of the simulation.

Figures 8a and 8b show the water level time series of the three reservoirs generated from the three simulations. The water levels of Lakes Powell and Mead are indicators of water curtailment for UB and LB water uses, respectively (US Bureau of Reclamation, 2007b). The Navajo Reservoir, located upstream of the San Juan River, is used as an example to illustrate human adaptation impacts on water resources in the upstream regions. In Figure 8a, the ABM-CRSS indicates that Navajo Reservoir could experience low water levels more frequently comparing to the CRSS (highlighted with red circles). This is because New Mexico agents are classified as Aggressive, who are slow learners (lower learning rate α) and able to tolerate short-term shortages (lower regret). Moreover, since the agents in the ABM-No RL do not learn, by comparing to the ABM-CRSS results, we can see that agents' learning capability can reduce water consumption in most years of significant water level decrease in Navajo Reservoir in Figure 8a.

The ABM-No RL and ABM-CRSS also predict significantly lower water levels in Lakes Mead and Powell than the CRSS prediction in the 2020–2040 period (period A in Figure 8b) because the agents are greedy in water uses by design. The agent's learning serves as a controller to prevent depleting reservoirs. Thus, when disabled, agents would continue withdrawing water despite the low water levels in the major dams, as the gray lines in Figure 8b show. Moreover, the ABM-CRSS shows that both reservoirs are depleted in 2034 but, during the 2054–2059 period, they show the opposite pattern: an increasing trend in Lake Mead and a decreasing trend in Lake Powell (period B in Figure 8b). By comparing the results of ABM-CRSS to ABM- No RL, we can see that the UB agents learned to request more water, while the LB agents learned to conserve water due to the persisting shortage conditions. In fact, LB agents may have learned to conserve water in period A, but we cannot see how that learning affects the system performance until the wet period arrives.

Figure 8c shows the UB and LB's total annual depletions and the annual water delivery to Mexico generated by the ABM-CRSS and the CRSS. The ABM-CRSS presents a higher depletion in the LB (the light blue line) from 2019 to 2033 and a significantly lower depletion afterward compared to the original CRSS (the dark blue line). The difference in the depletion pattern signals learning and adaptation behaviors of the LB agents in ABM-CRSS: the agents learn water scarcity in the drought period and reduce their water uses subsequently. For the UB annual depletion, the ABM-CRSS (the light green line) and the original CRSS (the dark green line) have similar patterns until 2051; after 2051, the ABM-CRSS shows a significantly higher depletion in the UB than the original CRSS. This is because the agents in Colorado and Utah continue to increase their diversions as the flow conditions return to normal. The annual depletion of all states generated from the original CRSS and ABM-CRSS are presented in Figures S2 and S3. The annual water consumption in the UB and LB, and the annual water delivery to Mexico of the ABM-No RL are also presented in the supporting information (Figure S4).

The 1944 Mexican Water Treaty committed to delivering 1.8 billion m^3 (1.5 MAF) freshwater from the US to Mexico on an annual basis. The original CRSS predicts a decade-long shortage in water delivered to Mexico from 2044 to 2055 and another mild shortage in 2059 (Figure 8c). However, the ABM-CRSS indicates that the shortage could happen as early as 2033 for significant storage declines in the major reservoirs due to the LB's excessive water uses from 2019 to 2033 (Figures 8b and 8c). In reality, it is unlikely for the Mexican delivery to be significantly dropped below 1.8 billion m^3. However, the model does not consider politics; the results merely show what could happen if the CRB's water storage is depleted.

4.4. Uncertainty and Nonstationarity of Agent Diversion Behavior and Basin Response

The previous section discusses the result of only one simulation to illustrate how the inclusion of human adaptive behaviors can affect the basin response. This section further investigates the uncertainty of system responses due to the dynamic agriculture water uses. The randomness in the RL-ABM represents the water use uncertainty in the coupled ABM-CRSS model, which can affect system response at different spatial scales (agent, State, sub-basin, and basin scales). To investigate agents' adaptation impacts on basin response, we repeated the coupled model simulation 100 times and summarized the results in Figures 9 and 10.

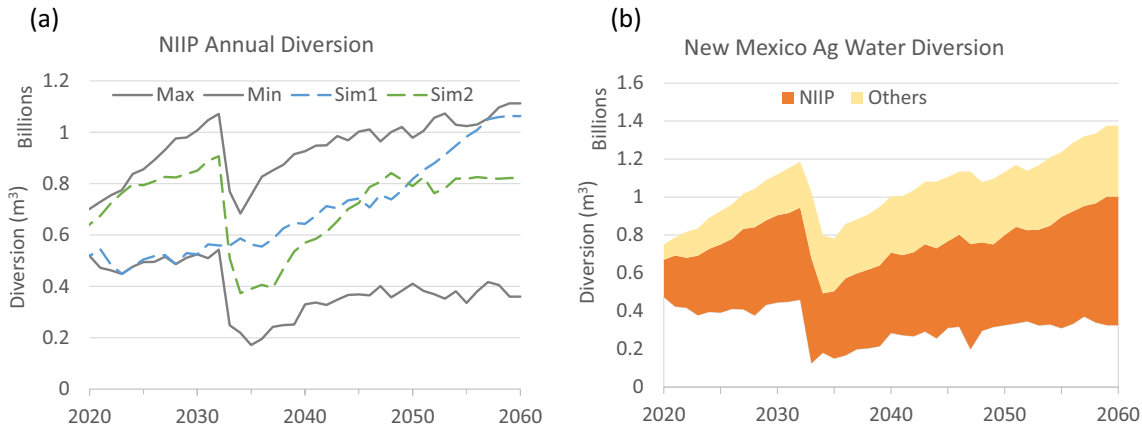

Figure 9. The New Mexico results of 100 agent-based modeling-Colorado River Simulation System simulations. (a) The annual diversion uncertainty range and two distinct diversion paths of agent the Navajo Indian Irrigation Project (NIIP) and (b) the New Mexico agriculture diversion uncertainty contributed from NIIP and other water users.

As an example of agents' adaptation, Figure 9a shows the predicted diversion ranges and two distinct diversion paths (Sim1 and Sim2) of the agent NIIP (the Navajo Indian Irrigation Project, the largest agriculture water user in New Mexico). Both the maximum and minimum diversion lines (maximum and minimum diversion in each year of the 100 simulations; gray lines in Figure 9a) indicate a severe water shortage from 2033 to 2035. However, this does not mean that the shortage is inevitable. For instance, the Sim1 line (blue dashed line in Figure 9a) shows a path that does not cause a water shortage. In Sim1 and Sim2, the NIIP agent experienced two very different environments, which significantly affect its diversion decisions later in the simulations. The NIIP agent in Sim2 suffered a multi-year water shortage (from 2033 to 2038) and changed its diversion policy at a high diversion level (0.8 billion m^3) after 2048 (green dashed line in Figure 9a). Contrarily, the NIIP agent in Sim1 survived the drought period without water curtailment and increased its diversion steadily throughout the simulation for the belief in water abundance.

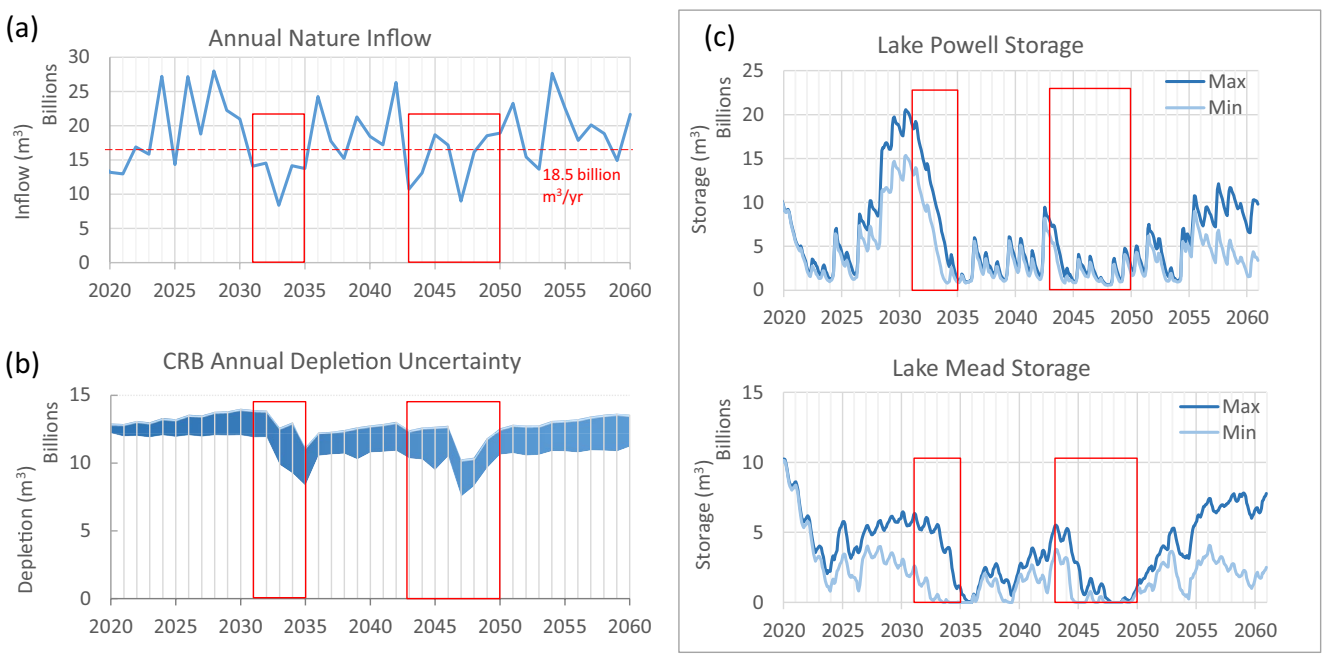

Figure 10. (a) The total water inflow to the Colorado River Basin (CRB); (b) the annual CRB water consumption (depletion) uncertainty range; and (c) the storage uncertainty ranges of Lake Powell and Lake Mead. The red boxes indicate the severe drought periods that average annual inflow is below 18.5 billion m^3 / yr (15 million acre-ft; the annual water consumption allowance to CRB in the 1922 Colorado River Compact).

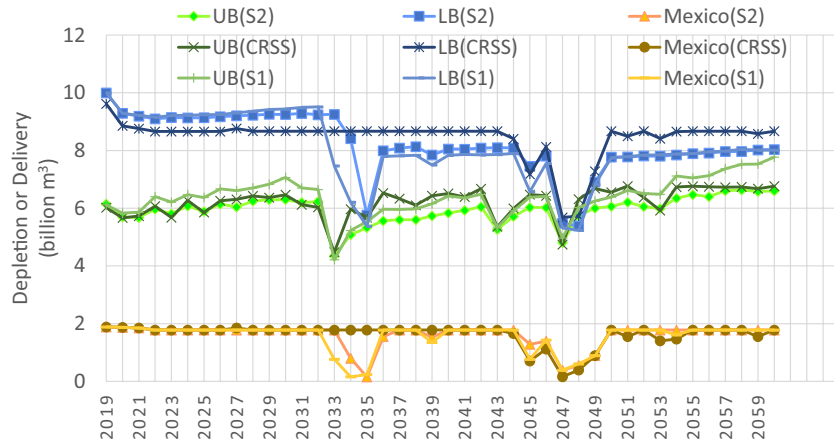

Figure 11. The annual depletions in the Upper and Lower Basins (UB and LB) and the water delivery to Mexico were generated from the original CRSS (in dark colors) and scenarios S1 and S2 from the agent-based modeling-Colorado River Simulation System (in light colors). Scenario S1 assumes the LB agents become more cautious in actions due to an education program that is extended to the whole CRB in Scenario S2.

As an example of uncertainty aggregation to the state level, we combined the uncertainty ranges of all New Mexico agents in Figure 9b, in which the contribution from NIIP is highlighted in orange for visualization. The lower bound of the uncertainty range (Figure 9b) is New Mexico's minimum diversion of each year of all simulations, and the contribution from the agents is the difference of their maximum and minimum annual diversions, as the gray lines show in Figure 9a. The New Mexico agriculture diversion shows a similar pattern as NIIP because agents in New Mexico are all categorized as aggressive agents.

Figure 10a shows the total inflow in the CRB, in which two severe drought periods are indicated by the red boxes (2031–2035 and 2043–2050). The droughts would deplete the storage in Lakes Powell and Mead (Figure 10c) first and then cause a shortage in CRB water consumption (Figure 10b). Interestingly, Figure 10b also indicates that the uncertainty range would shrink after a prolonged drought and gradually increase afterward. This is because agents learn to respond to the drought differently, and the diverse responses can reduce the uncertainty at basin scale. Another observation is that Lakes Powell and Mead are vulnerable to multi-year droughts, as all the simulations indicate extremely low storage in the two reservoirs during or after the drought events.

5. Discussion

5.1. Explore "Soft" Policy With Scenario Analysis

The RL-ABM framework simulates human adaptive behaviors based on their risk perception and willingness to learn. This feature enables us to assess and explore "soft" policies that affect people's decision-making, such as public education and stakeholder engagement programs, which have been highlighted as a broader impact at several funding agencies in the US. Here, we tested two scenarios solely for demonstration purposes. Future studies can include the most up-to-date management programs and insights drawn from social studies to generate more realistic results.

The first scenario (S1) assumes that a public education program is undergoing for LB agents, which deepens those agents' awareness of water scarcity. Consequently, these agents become more cautious, which means their actions follow a tight distribution and perceive a higher water shortage penalty. In terms of modeling, this means the LB agents' parameters μ, σ, and *regret* are changed to 0.1, 0.7, and 3, respectively. The second scenario (S2) assumes that the same program is extended to the whole CRB to alleviate water scarcity pressure further. Figure 11 shows the original CRSS results as the baseline and the results of these two scenarios. The S1 results in Figure 11 show a similar trend comparing to the results in Figure 8c until 2050. After 2050, the LB agents use significantly less water than the CRSS results, which helps restore the Mexico delivery to the 1.8 billion m^3. However, scenario S2 also predicts a prolonged shortage from 2032 to

2050 because the excessive water uses in the UB and LB continue to deplete water storage in the reservoirs, and no feedback mechanism is in place before 2032 to help maintain CRB storage in a health condition. In other words, agents would have a false perception of no water scarcity until a water shortage event occurs. The S2 scenario results can only delay the water shortage in Mexico delivery by one year, which suggests that soft policy alone may not be the solution for the water shortage problem in the CRB. However, the education program may help maintain Mexico delivery after the prolonged shortage after 2050, as the S2 scenario shows in Figure 11.

Both scenarios suggest that lacking a feedback mechanism before 2032 would deplete the CRB water storage and cause a water shortage for decades. This is evidence supporting the need for the 2019 Drought Contingency Plan (Stern & Sheikh, 2019), which is not yet in the CRSS model we applied in this paper, and other regulations to coordinate agriculture diversion.

5.2. Human System Modeling Improvement With ABM Parameterization and Limitations

The FQL algorithm can be generalized by including other physical and social-economic factors related to water resources management, such as temperature and crop prices, as state variables in the Q-function (Equation 5). In a more sophisticated application, agents could have different decision mechanisms, as the UB and LB agents do in our case study. Moreover, FQL's six parameters characterize an intelligent agent in terms of its risk attitude and willingness to adapt its policy. This feature improves the connection of modeling study to social science studies to interpret and communicate the parameters' meanings. For example, researchers can ask questions in an interview or survey like: "*how much more/less diversion in percentage would you like to request for next year, and how sure are you about your request* (a question to inform the action distribution parameters μ and σ)?" and "*how likely would you consider taking the other actions under given forecast* (a question help to quantify exploration rate ϵ)?"

Although the ABM-CRSS provides a bottom-up perspective for decision-making under uncertainty, the simulation results reflect only our current understanding of the human-nature system dynamics and assumptions and are subject to several limitations. The main limitations of this study lie in data availability and scarcity, as well as agriculture water users' decision mechanisms. We rely on historical observation data to train the intelligent agents, but consistent and complete data for the UB agents are only available at sub-basin level. Some agents in the LB only have a short period of record, which leads to low performance in training. The uncertainty in agent parameterization can significantly influence agent classification and coupled model results. Moreover, climate input and model structure uncertainty should also be included in the uncertainty analysis to provide a holistic view of system response uncertainty.

For agents' decision mechanisms, we assume that their decisions mainly depend on the current diversion level and the relevant water information, which could be an over-simplification. Other factors, such as production costs, crop demands, and water-saving technologies, may also be part of considerations. Also, the agent characterization should consider agents' physical and social-economic factors, as suggested in Pouladi et al. (2019), and the inconsistency in decision preference between a group and its members, known as the Arrow's Paradox (Franssen, 2005; Kasprzyk et al., 2016).

6. Summary and Conclusions

To address the uncertainty and nonstationarity issues in water resources management, we proposed a modeling framework with multiple intelligent agents capable of learning and adapting to environmental changes for the assessment of human impacts on regional water resources systems. The modeling framework consists of agents driven by the Farmer's Q-learning (FQL) developed in this paper and a water resources planning model (e.g., CRSS in our case study). The FQL algorithm, adapted from the Q-learning algorithm, incorporates agriculture water users' risk-perception, willingness to learn, and the prediction for future water availability to simulate human cognitive and decision processes that can be linked to social science studies to improve interpretation of dynamic human behaviors. The framework is applied to the CRB as an illustrative case study. Although the model coupling is explicitly tailored to the CRB, the proposed framework and algorithm are generalizable. They can be applied to other basins by modifying the definitions of the state variables, water information, and actions.

Through the clustering analysis of agents' parameters, we identified three types of intelligent agents based on the FQL parameters: the Aggressive, the Forward-looking Conservative, and the Myopic Conservative. In the case study, our analysis shows that the UB agents are all categorized in the Aggressive group because their water consumptions are still well below the 9.25 billion m^3 allocation stated in the Colorado River compact of 1922 (US Bureau of Reclamation, 2015). In contrast, the LB agents are either categorized as Forward-looking Conservative or Myopic Conservative because they rely on water release from upstream reservoirs and have fully utilized their water allocations.

The ABM-CRSS simulation results show that the CRB may experience more frequent shortages, and that the water levels in Lakes Powell and Mead may decline significantly due to the increasing water demands. The results also indicate that the LB agents may learn to reduce diversion from the persisting shortage condition, but the UB agents may not due to their underutilized water allocations. Moreover, the stochastic ABM-CRSS can be applied to assess system response uncertainties. Future research should also include other sources of uncertainty, such as climate inputs, parameters, and model structure uncertainties. Additionally, our ABM-CRSS framework can help water managers develop soft policies, such as public education programs and stakeholder engagement. The results shown in Section 5.1 support the 2019 Drought Contingency Plan and suggest that such programs are necessary to ensure sufficient water storage in CRB.

We acknowledge that our framework only represents the learning aspect of human behaviors. The model designs of other ABM studies that investigated human response to water conservation in stationary environments can enrich future RL-ABM research and applications by incorporating first-hand data from surveys for management program development. For example, Al-Amin et al. (2018) assessed the effectiveness of water conservation programs using an ABM with two types of agents, the municipals and households, to simulate the dynamic interactions between the agents. Our future work will consider incorporating multiple agent types, such as municipal and industrial water users. The immediate extension of this framework includes improving the FQL algorithm by accounting for crops' economic values and production functions. For the CRB, future research includes coupling RL-ABM with the latest version of CRSS to discuss the newly implemented 2019 Drought Contingency Plan, soft policy development, and uncertainty analysis.

Data Availability Statement

No new data are presented. The Upper Basin depletion data were retrieved from Colorado River Systems Consumptive Uses and Loses Reports downloaded from https://www.usbr.gov/uc/envdocs/plans.html#C-CULR and the Lower Basin diversion data are retrieved from Colorado River Systems Consumptive Uses and Loses Reports available at https://www.usbr.gov/lc/region/g4000/wtracct.html. Precipitation data are retrieved from Parameter-Elevation Regressions on Independent Slopes Model (PRISM) Data set (https://catalog.data.gov/dataset/parameter-elevation-regressions-on-independent-slopes-model-prism-dataset). Lake Mead water level data are retrieved from https://www.usbr.gov/lc/region/g4000/hourly/mead-elv.html. The scripts of the ABM in Python 3.6 are published in the GitHub repository: https://github.com/hfengwe1/RL-ABM-CRSS.

Acknowledgments
This research was supported by the Office of Science of the US Department of Energy as part of research in the Multi-Sector Dynamics, Earth and Environmental System Modeling Program and the US National Science Foundation (EAR #1804560). We want to thank Nathalie Voisin in Pacific Northwest National Laboratory and Vincent Tidwell in Sandia National Laboratories for commenting on the earlier version of the draft. The authors also want to thank the editors and three anonymous reviewers who helped us improve the quality of this paper.

References

Al-Amin, S., Berglund, E. Z., Mahinthakumar, G., & Larson, K. L. (2018). Assessing the effects of water restrictions on socio-hydrologic resilience for shared groundwater systems. *Journal of Hydrology*, *566*, 872–885. https://doi.org/10.1016/j.jhydrol.2018.08.045

Aubert, A. H., Tavenard, R., Emonet, R., De Lavenne, A., Malinowski, S., Guyet, T., et al. (2013). Clustering flood events from water quality time series using Latent Dirichlet Allocation model. *Water Resources Research*, *49*(12), 8187–8199. https://doi.org/10.1002/2013WR014086

Berglund, E. Z. (2015). Using agent-based modeling for water resources planning and management. *Journal of Water Resources Planning and Management*, *141*(11), 04015025. https://doi.org/10.1061/(asce)wr.1943-5452.0000544

Bertoni, F., Giuliani, M., & Castelletti, A. (2020). Integrated design of dam size and operations via reinforcement learning. *Journal of Water Resources Planning and Management*, *146*(4), 1–12. https://doi.org/10.1061/(ASCE)WR.1943-5452.0001182

Bone, C., & Dragićević, S. (2009). GIS and intelligent agents for multiobjective natural resource allocation: A reinforcement learning approach. *Transactions in GIS*, *13*(3), 253–272. https://doi.org/10.1111/j.1467-9671.2009.01151.x

Borgomeo, E., Mortazavi-Naeini, M., Hall, J. W., & Guillod, B. P. (2018). Risk, robustness and water resources planning under uncertainty. *Earth's Future*, *6*(3), 468–487. https://doi.org/10.1002/2017EF000730

Brown, C. M., Lund, J. R., Cai, X., Reed, P. M., Zagona, E. A., Ostfeld, A., et al. (2015). Scientific framework for sustainable water management. *Water Resources Research*, *51*, 6110–6124. https://doi.org/10.1002/2015WR017114.Received

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. https://doi.org/10.1029/2009WR008898

Castilla-Rho, J. C., Mariethoz, G., Rojas, R., Andersen, M. S., & Kelly, B. F. J. (2015). An agent-based platform for simulating complex human-aquifer interactions in managed groundwater systems. *Environmental Modelling & Software*, *73*, 305–323. https://doi.org/10.1016/j.envsoft.2015.08.018

Castle, S. L., Thomas, B. F., Reager, J. T., Rodell, M., Swenson, S. C., & Famiglietti, J. S. (2014). Groundwater depletion during drought threatens future water security of the Colorado River Basin. *Geophysical Research Letters*, *41*(16), 5904–5911. https://doi.org/10.1002/2014GL061055

Cosgrove, W. J., & Loucks, D. P. (2015). Water management: Current and future challenges and research directions. *Water Resources Research*, *51*(6), 4823–4839. https://doi.org/10.1002/2014WR016869

Dariane, A. B., & Moradi, A. M. (2016). Comparative analysis of evolving artificial neural network and reinforcement learning in stochastic optimization of multireservoir systems. *Hydrological Sciences Journal*, *61*(6), 1141–1156. https://doi.org/10.1080/02626667.2014.986485

Franssen, M. (2005). Arrow's theorem, multi-criteria decision problems and multi-attribute preferences in engineering design. *Research in Engineering Design*, *16*(1–2), 42–56. https://doi.org/10.1007/s00163-004-0057-5

Garrick, D., Jacobs, K., & Garfin, G. (2008). Models, assumptions, and stakeholders: Planning for water supply variability in the Colorado River Basin. *Journal of the American Water Resources Association*, *44*(2), 381–398. https://doi.org/10.1111/j.1752-1688.2007.00154.x

Giuliani, M., Li, Y., Castelletti, A., & Gandolfi, C. (2016). A coupled human-natural systems analysis of irrigated agriculture under changing climate. *Water Resources Research*, *52*(9), 6928–6947. https://doi.org/10.1002/2016WR019363

Glynn, P. D., Voinov, A. A., Shapiro, C. D., & White, P. A. (2018). Response to comment by Walker et al. on "From data to decisions: Processing information, biases, and beliefs for improved management of natural resources and environments. *Earth's Future*, *6*(5), 762–769. https://doi.org/10.1002/2018EF000819

Gold, D. F., Reed, P. M., Trindade, B. C., & Characklis, G. W. (2019). Identifying actionable compromises: Navigating multi-city robustness conflicts to discover cooperative safe operating spaces for regional water supply portfolios. *Water Resources Research*, *55*(11), 9024–9050. https://doi.org/10.1029/2019WR025462

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. https://doi.org/10.1016/j.jhydrol.2009.08.003

Haasnoot, M., Kwakkel, J. H., Walker, W. E., & ter Maat, J. (2013). Dynamic adaptive policy pathways: A method for crafting robust decisions for a deeply uncertain world. *Global Environmental Change*, *23*(2), 485–498. https://doi.org/10.1016/j.gloenvcha.2012.12.006

Hadjimichael, A., Quinn, J., Wilson, E., Reed, P., Basdekas, L., Yates, D., & Garrison, M. (2020). Defining robustness, vulnerabilities, and consequential scenarios for diverse stakeholder interests in institutionally complex river basins. *Earth's Future*, *8*(7), 1–22. https://doi.org/10.1029/2020EF001503

Hall, J. W., Mortazavi-Naeini, M., Borgomeo, E., Baker, B., Gavin, H., Gough, M., et al. (2020). Risk-based water resources planning in practice: A blueprint for the water industry in England. *Water and Environment Journal*, *34*(3), 441–454. https://doi.org/10.1111/wej.12479

Harrison, K. W. (2007). Test application of Bayesian Programming: Adaptive water quality management under uncertainty. *Advances in Water Resources*, *30*(3), 606–622. https://doi.org/10.1016/j.advwatres.2006.03.011

Herman, J. D., Quinn, J. D., Steinschneider, S., Giuliani, M., & Fletcher, S. (2020). Climate adaptation as a control problem: Review and perspectives on dynamic water resources planning under uncertainty. *Water Resources Research*, *56*(2), e24389. https://doi.org/10.1029/2019WR025502

Higgins, A., Archer, A., & Hajkowicz, S. (2008). A Stochastic non-linear programming model for a multi-period water resource allocation with multiple objectives. *Water Resources Management*, *22*(10), 1445–1460. https://doi.org/10.1007/s11269-007-9236-2

Hung, F., & Hobbs, B. F. (2019). How can learning-by-doing improve decisions in stormwater management? A Bayesian-based optimization model for planning urban green infrastructure investments. *Environmental Modelling & Software*, *113*, 59–72. https://doi.org/10.1016/j.envsoft.2018.12.005

Hyun, J. Y., Huang, S. Y., Yang, Y. C. E., Tidwell, V., & Macknick, J. (2019). Using a coupled agent-based modeling approach to analyze the role of risk perception in water management decisions. *Hydrology and Earth System Sciences*, *23*(5), 2261–2278. https://doi.org/10.5194/hess-23-2261-2019

Kaelbling, L. P., Littman, M. L., & Cassandra, A. R. (1998). Planning and acting in partially observable stochastic domains. *Artificial Intelligence*, *101*(1–2), 99–134. https://doi.org/10.1016/S0004-3702(98)00023-X

Kasprzyk, J. R., Nataraj, S., Reed, P. M., & Lempert, R. J. (2013). Many objective robust decision making for complex environmental systems undergoing change. *Environmental Modelling & Software*, *42*, 55–71. https://doi.org/10.1016/j.envsoft.2012.12.007

Kasprzyk, J. R., Reed, P. M., & Hadka, D. M. (2016). Battling Arrow's Paradox to discover robust water management alternatives. *Journal of Water Resources Planning and Management*, *142*(2), 04015053. https://doi.org/10.1061/(ASCE)WR.1943-5452.0000572

Khan, H. F., Yang, Y. C. E., Xie, H., & Ringler, C. (2017). A coupled modeling framework for sustainable watershed management in transboundary river basins. *Hydrology and Earth System Sciences*, *21*(12), 6275–6288. https://doi.org/10.5194/hess-21-6275-2017

Kwakkel, J. H., Haasnoot, M., & Walker, W. E. (2015). Developing dynamic adaptive policy pathways: A computer-assisted approach for developing adaptive strategies for a deeply uncertain world. *Climatic Change*, *132*(3), 373–386. https://doi.org/10.1007/s10584-014-1210-4

Kwakkel, J. H., Haasnoot, M., & Walker, W. E. (2016). Comparing Robust Decision-Making and Dynamic Adaptive Policy Pathways for model-based decision support under deep uncertainty. *Environmental Modelling & Software*, *86*, 168–183. https://doi.org/10.1016/j.envsoft.2016.09.017

Lampel, J., Lampel, J., Shamsie, J., & Shapira, Z. (2009). Experiencing the improbable: Rare events and organizational learning. *Organization Science*, *20*(5), 835–845. https://doi.org/10.1287/orsc.1090.0479

Lee, J. H., & Labadie, J. W. (2007). Stochastic optimization of multireservoir systems via reinforcement learning. *Water Resources Research*, *43*(11), 1–16. https://doi.org/10.1029/2006WR005627

Lempert, R. J., & Collins, M. T. (2007). Managing the risk of uncertain threshold responses: Comparison of robust, optimum, and precautionary approaches. *Risk Analysis*, *27*(4), 1009–1026. https://doi.org/10.1111/j.1539-6924.2007.00940.x

Lund, J. R. (2002). Floodplain planning with risk-based optimization. *Journal of Water Resources Planning and Management*, *128*(3), 202–207. https://doi.org/10.1061/(ASCE)0733-9496(2002)128:3(202)

Madani, K., & Hooshyar, M. (2014). A game theory-reinforcement learning (GT-RL) method to develop optimal operation policies for multi-operator reservoir systems. *Journal of Hydrology*, *519*(PA), 732–742. https://doi.org/10.1016/j.jhydrol.2014.07.061

Moallemi, E. A., Zare, F., Reed, P. M., Elsawah, S., Ryan, M. J., & Bryan, B. A. (2020). Structuring and evaluating decision support processes to enhance the robustness of complex human–natural systems. *Environmental Modelling & Software*, *123*, 104551. https://doi.org/10.1016/j.envsoft.2019.104551

Monahan, G. E. (1982). State of the art - A survey of partially observable Markov decision processes: Theory, models, and algorithms. *Management Science*, *28*(1), 1–16. https://doi.org/10.1287/mnsc.28.1.1

Müller, B., Bohn, F., Dreßler, G., Groeneveld, J., Klassert, C., Martin, R., et al. (2013). Describing human decisions in agent-based models-ODD+D, an extension of the ODD protocol. *Environmental Modelling & Software*, *48*, 37–48. https://doi.org/10.1016/j.envsoft.2013.06.003

National Integrated Drought Information System. (2021). *Drought in Utah from 2000-present*. 6 April 2021 Retrieved from https://www.drought.gov/states/utah

Ng, T. L., Eheart, J. W., Cai, X., & Braden, J. B. (2011). An agent-based model of farmer decision-making and water quality impacts at the watershed scale under markets for carbon allowances and a second-generation biofuel crop. *Water Resources Research*, *47*(9), 1–17. https://doi.org/10.1029/2011WR010399

Ni, J., Liu, M., Ren, L., & Yang, S. X. (2014). A multiagent Q-learning-based optimal allocation approach for urban water resource management system. *IEEE Transactions on Automation Science and Engineering*, *11*(1), 204–214. https://doi.org/10.1109/TASE.2012.2229978

Oregon State University. (2020). Parameter-Elevation Regressions on Independent Slopes Model (PRISM) Dataset. April 3, 2020 Retrieved from https://catalog.data.gov/dataset/parameter-elevation-regressions-on-independent-slopes-model-prism-dataset

Piantadosi, J., Metcalfe, A. V., & Howlett, P. G. (2008). Stochastic dynamic programming (SDP) with a conditional value-at-risk (CVaR) criterion for management of storm-water. *Journal of Hydrology*, *348*, 320–329. https://doi.org/10.1016/j.jhydrol.2007.10.007

Pouladi, P., Afshar, A., Afshar, M. H., Molajou, A., & Farahmand, H. (2019). Agent-based socio-hydrological modeling for restoration of Urmia Lake: Application of theory of planned behavior. *Journal of Hydrology*, *576*(July), 736–748. https://doi.org/10.1016/j.jhydrol.2019.06.080

Quinn, J. D., Reed, P. M., Giuliani, M., & Castelletti, A. (2017). Rival framings: A framework for discovering how problem formulation uncertainties shape risk management trade-offs in water resources systems. *Water Resources Research*, *53*(8), 7208–7233. https://doi.org/10.1002/2017WR020524

Reuss, M. (2003). Is it time to resurrect the harvard water program? *Journal of Water Resources Planning and Management*, *129*(5), 357–360. https://doi.org/10.1061/(ASCE)0733-9496(2003)129:5(357)

Rieker, J. D., & Labadie, J. W. (2012). An intelligent agent for optimal river-reservoir system management. *Water Resources Research*, *48*(9), 1–16. https://doi.org/10.1029/2012WR011958

Seo, H., & Lee, D. (2017). Reinforcement learning and strategic reasoning during social decision-making. In *Decision neuroscience: An integrative perspective* (pp. 225–231). Elsevier Inc. https://doi.org/10.1016/B978-0-12-805308-9.00018-X

Smith, V. L. (1991). Rational choice: The contrast between Economics and Psychology. *Journal of Political Economy*, *99*(4), 877–897. https://doi.org/10.1086/261782

Starbuck, W. H. (2009). Cognitive reactions to rare events: Perceptions, uncertainty, and learning. *Organization Science*, *20*(5), 925–937. https://doi.org/10.1287/orsc.1090.0440

Stern, C. V., & Sheikh, P. A. (2019). *Management of the Colorado River: Water allocations, drought, and the federal role*. Retrieved from https://crsreports.congress.gov

Sutton, R. S. (1988). Learning to predict by the methods of temporal differences. *Machine Learning*, *3*(1), 9–44. https://doi.org/10.1007/BF00115009

Sutton, R. S. (1992). Reinforcement Learning. In R. S. Sutton (Ed.), *Reinforcement learning*: Springer US. https://doi.org/10.1007/978-1-4615-3618-5

Sutton, R. S., & Barto, A. G. (1998). *Reinforcement learning: An introduction* (Second). The MIT Press.

Sutton, R. S., & Singh, S. P. (1996). *Reinforcement learning with replacing edibility traces*. Machine Learning.

Tijsma, A. D., Drugan, M. M., & Wiering, M. A. (2017). *Comparing exploration strategies for Q-learning in random stochastic mazes*. IEEE Symposium Series on Computational Intelligence. https://doi.org/10.1109/SSCI.2016.7849366

US Bureau of Reclamation. (2007a). *Colorado River interim guidelines for Lower Basin shortages and coordinated operations for Lake Powell and Lake Mead-Appendix A CRSS model documentation*.

US Bureau of Reclamation. (2007b). *Colorado River interim guidelines for lower basin shortages and coordinated operations for Lake Powell and Lake Mead-Final environmental impact statement*. Retrieved from https://www.usbr.gov/lc/region/programs/strategies/FEIS/index.html

US Bureau of Reclamation. (2012). *Colorado River Basin water supply and demand study-Executive summary. Reclamation: Managing water in the West*. Retrieved from https://www.usbr.gov/watersmart/bsp/docs/finalreport/ColoradoRiver/CRBS_Executive_Summary_FINAL.pdf

US Bureau of Reclamation. (2015). *Provisional Upper Ccolorado River Basin Consumptive Uses and Losses Report*.

Watkins, C. (1989). *Learning from delayed rewards (PhD thesis)*: University of Cambridge.

Watkins, C., & Dayan, P. (1992). Q-learning. *Machine Learning*, *8*(3–4), 279–292. https://doi.org/10.1007/BF00992698

Watson, A. A., & Kasprzyk, J. R. (2017). Incorporating deeply uncertain factors into the many objective search process. *Environmental Modelling & Software*, *89*, 159–171. https://doi.org/10.1016/j.envsoft.2016.12.001

Wheeler, K. G., Schmidt, J. C., Rosenberg, D. E., & Tarboton, D. G. (2019). *Water resource modelling of the Colorado river-Present and future strategies*. Center for Colorado River Studies.

Yan, D., Ludwig, F., Huang, H. Q., & Werners, S. E. (2017). Many-objective robust decision making for water allocation under climate change. *The Science of the Total Environment*, *607–608*, 294–303. https://doi.org/10.1016/j.scitotenv.2017.06.265

Yang, Y. C. E., Cai, X., & Stipanović, D. M. (2009). A decentralized optimization algorithm for multiagent system-based watershed management. *Water Resources Research*, *45*(8), 1–18. https://doi.org/10.1029/2008WR007634

Yang, Y. C. E., Son, K., Hung, F., & Tidwell, V. (2020). Impact of climate change on adaptive management decisions in the face of water scarcity. *Journal of Hydrology*, *588*(May), 125015. https://doi.org/10.1016/j.jhydrol.2020.125015

Zagona, E. A., Fulp, T. J., Shane, R., Magee, T., & Goranflo, H. M. (2001). Riverware: A generalized tool for complex reservoir system modeling. *JAWRA Journal of the American Water Resources Association*, *37*(4), 913–929. https://doi.org/10.1111/j.1752-1688.2001.tb05522.x