Reinforcement Learning-based Real-time Control of Coastal Urban Stormwater Systems to Mitigate Flooding and Improve Water Quality

Benjamin D. Bowes[1], Cheng Wang[1], Mehmet B. Ercan[1,2], Teresa B. Culver[1], Peter A. Beling[3], and Jonathan L. Goodall[1,4*]

[1] *Dept. of Engineering Systems and Environment, Univ. of Virginia, 151 Engineer's Way, P.O. Box 400747, Charlottesville, VA 22904, USA*

[2] *Xylem, South Bend, IN, USA*

[3] *Dept. of Industrial and Systems Engineering, Virginia Polytechnic Institute and State University, 250 Perry St, Blacksburg, VA 24061, USA*

[4] *Dept. of Computer Science, Univ. of Virginia, Rice Hall, 85 Engineer's Way, PO Box 400740, Charlottesville, VA 22904, USA*

[*] *Corresponding Author: goodall@virginia.edu*

This is a preprint for the published paper:

Bowes, B.D., Wang, C., Ercan, M.B., Culver, T.B., Beling, P.A. and Goodall, J.L., 2022. Reinforcement learning-based real-time control of coastal urban stormwater systems to mitigate flooding and improve water quality. Environmental Science: Water Research & Technology. https://doi.org/10.1039/D1EW00582K

ABSTRACT

Real-time control of stormwater systems can reduce flooding and improve water quality. Current industry real-time control strategies use simple rules based on water quantity parameters at a local scale. However, system-level control methods that also incorporate observations of water quality could provide improved control and performance. Therefore, the objective of this research, is to evaluate the impact of local and system-level control approaches on flooding and sediment-related water quality in a stormwater system within the flood-prone coastal city of Norfolk, Virginia, USA. Deep reinforcement learning (RL), an emerging machine learning technique, is used to learn system-level control policies that attempt to balance flood mitigation and treatment of sediment. RL is compared to the conventional stormwater system and two methods of local-scale rule-based control: (i) industry standard predictive rule-based control with a fixed detention time and (ii) rules based on water quality observations. For the studied system, both methods of rule-based control improved water quality compared to the passive system, but increased total system flooding due to uncoordinated releases of stormwater. An RL agent learned controls that maintained target pond levels while reducing total system flooding by 4% compared to the passive system. When pre-trained from the RL agent that learned to reduce flooding, another RL agent was able to learn to decrease TSS export by an average of 52% compared to the passive system and with an average of 5% less flooding than the rule-based control methods. As the complexity of stormwater RTC implementations grows and climate change continues, system-level control approaches such as the RL used here will be needed to help mitigate flooding and protect water quality.

Keywords: Real-time Control, Reinforcement Learning, Smart Stormwater Systems, Urban Flooding, Water Quality

Water Impact Statement

Advances in smart and connected technologies can reduce flooding and improve water quality through real-time stormwater system control. Currently, real-time stormwater control operates at local-scales with fixed rules. We present a method for learning system-level control strategies that balance competing flood mitigation and pollutant treatment goals. With continued adoption of stormwater real-time control, these system-level control approaches can improve flood and pollutant mitigation.

1 INTRODUCTION

Communities rely on stormwater systems to mitigate flooding and treat polluted runoff from urban areas. However, as urbanization increases and climate change continues to alter precipitation, temperature, and sea levels, communities will be faced with increased stormwater runoff causing greater flooding and water pollution[1–4]. Conventional stormwater systems are designed based on historic data assuming stationarity of future conditions. They are largely static systems, unable to dynamically adapt to unanticipated conditions. Increasing the resilience of stormwater systems to these unanticipated and changing land use and climate conditions will require new approaches to dynamically control both flood mitigation and pollutant treatment.

The adoption of smart cities approaches is allowing stormwater managers to begin to monitor and control individual components of conventional stormwater systems, which are gravity-driven and behave statically, in real-time[5]. While the use of real-time control (RTC) is fairly established in combined sewer systems[6–8], recent research has shown that retro-fitting conventional stormwater components (e.g., a retention pond) for RTC can allow more efficient local operation, mitigating flooding from storms[9,10] and preventing erosive, high velocity flows[11]. RTC can also provide more efficient treatment of pollutants such as sediment and nutrients, primarily through increased detention time[12,13]. For instance, RTC of a retention pond increased removal of total suspended solids (TSS) and nitrate (NO_3) by roughly 40%, compared to passive pond operation[14].

In practice, stormwater RTC is generally performed using local rule-based control (RBC), which is almost exclusively based on volumetric data (e.g., depth, current and forecast rainfall)[14–16]. For instance, a rule may open a valve when the water level in a storage pond reaches a certain height or proactively drain water from a pond based on a rainfall forecast to create additional storage capacity before a large storm. In most studies using RBC, water quality is not considered or is inferred through hydraulic retention time, rather than directly observed or used in control rules. However, pollutant characteristics are highly variable between sites and storms and there is a need for more generalizable RTC methods for enhancing pollutant treatment. Toward this end, the benefits of using real-time water quality observations in control rules has recently been explored in simulation. For example, using the concentration of TSS to trigger a valve controlling outflow from a storage pond can improve TSS capture in the pond compared to the passive system and other volumetric control rules[17]. Given the effectiveness of RTC-enabled individual infrastructure components to adapt to different storm events, system-level RTC (i.e., control of multiple infrastructure components based on information from locations throughout the system) has the potential to more holistically enhance flood and pollution mitigation through coordinated control of multiple components[18].

As the complexity of controlled stormwater systems increases, the task of creating rules to (i) mitigate flooding, (ii) protect the quality of receiving waters, or (iii) balance both flooding and water quality, becomes nontrivial. Further, controlling for flooding and water quality can be competing goals. For example, draining a stormwater pond is the simplest way to prevent it from flooding. However, treatment of pollutants can require holding more water; TSS requires still conditions for settling, but stormwater inflow could resuspend sediment if a pond is drawndown to shallow depths. Maintaining more submerged (anaerobic) areas can increase denitrification, but reduces capacity to capture additional stormwater without flooding. Control rules or thresholds can be set to attempt to balance these goals, but they may only perform well under a limited range of conditions. Instead of attempting to create rules that cover all possible interactions between stormwater system components, pollutants, and environmental conditions, recent research has explored system-level methods of optimizing stormwater RTC. For instance, system-level control of a coastal urban stormwater system reduced total system flooding, even under sea level rise conditions[9]. In terms of water quality, flow from a system of ponds to a treatment wetland has been controlled to increase the efficiency of nitrate removal by 46%[18]. A study using system-level RTC for both water quantity and quality used linear optimization to control retention basin outflows. However, water quality control still relied on fixed rules to extend detention time (i.e., hold water after a storm for a set amount of time) and system control based on either observed or simulated real-time water quality measurements was not included. Continuing improvements in real-time water quality sensors, could allow more direct observation and control of water quantity in conjunction with some water quality parameters[19,20]. Making the best use of these growing sensing capabilities requires new methods of creating stormwater RTC policies that balance flood mitigation and water quality improvement.

Recent advances in Reinforcement Learning (RL), a type of machine learning, provide an alternate approach to system-level stormwater RTC where control policies can be learned , instead of using predetermined rules[21]. In RL, an agent (i.e., algorithm) does not have known answers to learn from, which is the standard supervised machine learning paradigm, but instead is rewarded based on how well its control actions meet specified stormwater system goals (e.g., flood mitigation, improved water quality). The reward signal is used to guide the agent's learning towards actions that maximize the return from areward function. Classical tabular RL is closely related to Dynamic Programming and has been explored for multi-objective reservoir management[22–26]. However, because tabular RL is limited to systems with relatively small numbers of possible states and actions, Deep Reinforcement Learning (also widely referred to as RL), which uses neural networks as function approximators instead of using lookup tables, has been used for control of more complex systems[27,28]. This approach to learning allows RL increased flexibility to optimize control actions, balance competing objectives based on the formulation of the reward function, and has the potential to continually adapt system controls to evolving environmental conditions (e.g., increased runoff from urbanization or climate change).

Initial research with RL for stormwater systems demonstrated control policies that reduced peak flows could be learned using water quantity observations from a complex system[29]. Flood mitigation improvements have also been achieved using RL-based RTC to learn system-level policies with water quantity data[10,30], while being robust to uncertainty in observations and forecasts[31].Despite the fact that many stormwater systems are used for pollutant treatment as well as flood mitigation,previous RL research has not considered using water quality observations to inform RTC methods. Given real-time water quality observations, RL may be able to learn to balance competing water quantity and quality goals throughout a stormwater system over a large range of conditions and could outperform rule-based methods. This paper is the first to incorporate water quality observations into RL-based control policies and, therefore, aims to illustrate RL's ability to learn system-level control policies considering the competing goals of flood mitigation and water quality protection.

2 METHODS

This research compares RL and RBC for their ability to both mitigate flooding and improve water quality compared to conventional static stormwater infrastructure. A simulation of Norfolk, Virginia's stormwater system including water quantity and quality processes is used as the controlled system. Two methods of local-scale, rule-based control are implemented: (i) predictive RBC with a fixed detention time and (ii) RBC based on water quality observations. RL is implemented for system-level control that incorporates measures of water quality and flood mitigation. After comparing the performance of these methods, their robustness to changes in system behavior is evaluated by simulating groundwater exchange within the controlled ponds.

2.1 Study Area

The City of Norfolk, Virginia, specifically its Hague neighborhood, is used as the study area for this research. Norfolk is situated near the mouth of the Chesapeake Bay on the eastern coast of the U.S. (Fig. 1, A and B). The city has a high rate of relative sea level rise partly due to regional land subsidence[32] and its low elevation, flat topography, and regular hurricane season contribute to increasingly frequent and severe recurrent flooding[1]. Additionally, Norfolk has a high groundwater table that responds quickly to storm events[33] and could contribute significant amounts of water to retention ponds that are being actively controlled[10]. The Hague neighborhood is a historic part of Norfolk and is adjacent to many city government buildings and the region's main hospital; the Hague also experiences some of the most frequent flooding in the city[9,34].

The quality of stormwater runoff from the city contributes to the health of the Chesapeake Bay, which has a long history of impairments such as hypoxia caused by eutrophication[35,36]. Pollutants carried by the city's stormwater (such as TSS, nitrogen, and phosphorous) are regulated to meet the Total Maximum Daily Loads (TMDLs) set for the Bay. As outlined in the City's Chesapeake Bay TMDL Action Plan, the City is required to reduce pollutant loadings by 5.75%, 35%, and 60% by 2021, 2026, and 2031, respectively[37].

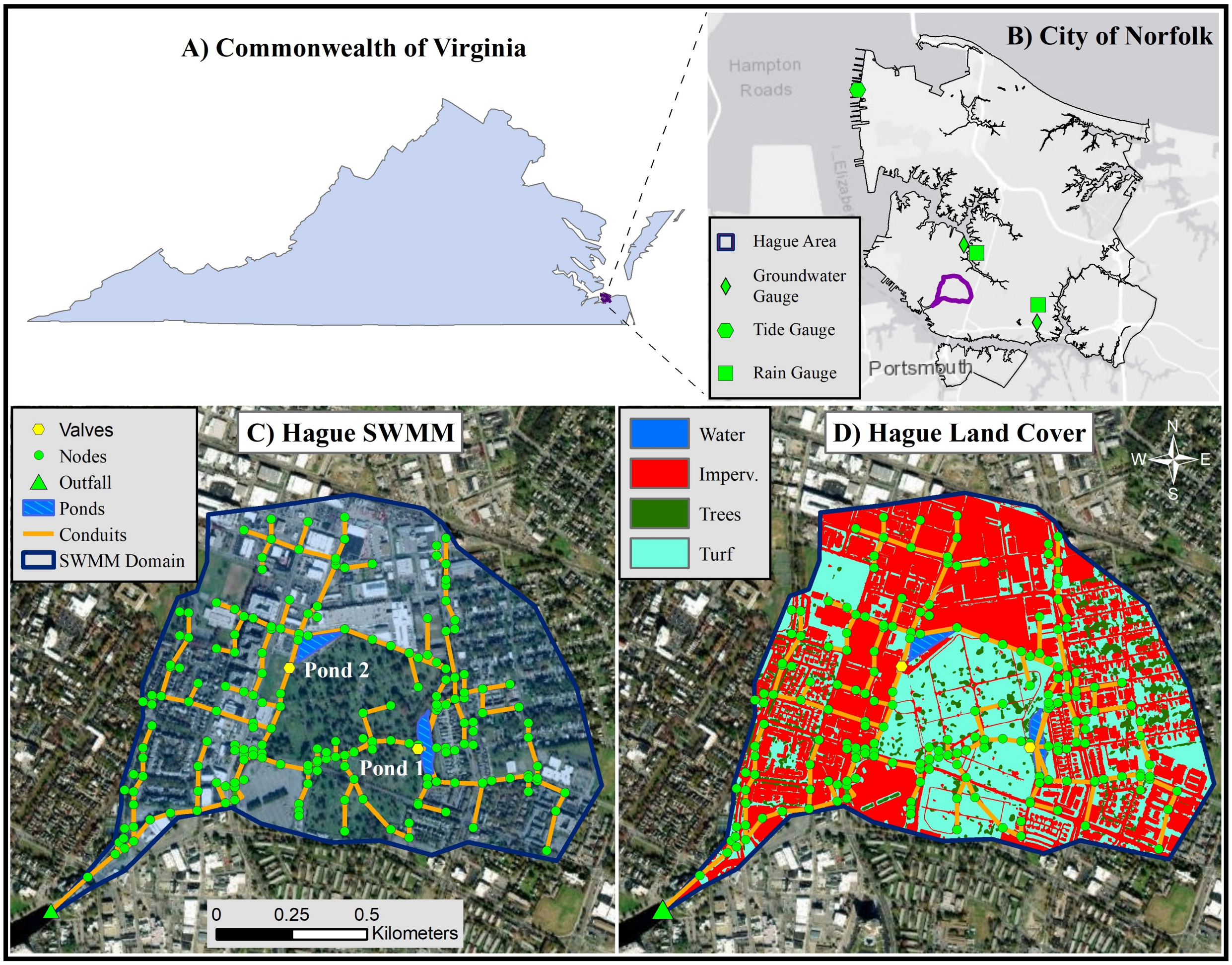

Figure 1. Study area - Hague area of Norfolk, Virginia USA with (C) the SWMM model and (D) land cover data.

2.2 SWMM Model

The Hague's recurrent flooding prompted Norfolk to build a model of the existing conventional stormwater system using the U.S. Environmental Protection Agency's (EPA) Stormwater Management Model (SWMM) (Fig. 1, C). The city verified that the SWMM model behavior sufficiently represented the physical system and calibrated it to match observed flooding in the Hague from Hurricane Matthew, which caused wide-spread flooding in October, 2016. The Hague SWMM model was updated by Sadler et al.[9] to simulate real-time control infrastructure (i.e., an additional retention pond and a valve, pump, and inflatable dam). In the current study, the SWMM simulation from Sadler et al.[9] is driven by long-term observed rainfall with a tidal boundary condition and has been enhanced to include groundwater and water quality processes. Land cover within the Hague SWMM model domain was extracted for each subwatershed from a 1m resolution dataset[38] and included three pollutant generating land covers: impervious, turf grass, and trees (Fig. 1, D). Impervious cover represents 56% of the model domain, while turf grass and trees cover 37% and 6%, respectively; the remainder of the land cover is water. SWMM input files with full configuration details can be found in the open source code repository (see Section 5).

2.2.1 Input Data

Observed rainfall, tide, and groundwater data were collected from gauges in Norfolk for the period between 1 January, 2010 and 6 November, 2019 (Fig. 1, B). Fifteen minute rainfall data came from two stations near the Hague that are operated by the Hampton Roads Sanitation District (HRSD). Rainfall data is processed by first removing any values over the 1000-year 15-minute value for Norfolk (59.2mm); these large values represented less than 0.01% of the rainfall datasets. Any missing values from one rain gauge are filled with the value from the other gauge if available; there were no periods where both rain gauges were missing data. Finally, the mean of the two rain gauges is taken to create a single time series for the SWMM model. Observed 6-minute tide data came from the Sewells Point gauge operated by the National Oceanic and Atmospheric Administration (NOAA). Tide data are referenced to the North American Vertical Datum of 1988 (NAVD88) and were resampled to an hourly interval for use as a SWMM boundary at the stormwater system outfall.

Forecasts for use in the RTC control methods were created from the observed rainfall and tide data.These forecasts are a rolling window of values over the next n time steps. In this work, a 24 hour forecast of 15 minute rainfall contains n=96 values. Because the focus of this work is on comparison of the RTC scenarios, the forecasts were assumed to represent perfect knowledge.

2.2.2 Groundwater Exchange Simulation

Because Norfolk has a high groundwater table and is already experiencing impacts from a high rate of relative sea level rise, the robustness of stormwater RTC methods to groundwater exchange will be increasingly important. While groundwater interactions with the retention ponds in Norfolk have not been studied specifically, it has been demonstrated that increased groundwater table levels due to sea level rise could contribute to retention ponds in coastal areas, decreasing their ability to appropriately manage consecutive storm events[39]. To address this need, groundwater exchange with controlled ponds is simulated in a number of the scenarios in this research.

Groundwater data was collected from two shallow monitoring wells operated by HRSD and referenced to NAVD88. Outliers from these data were removed with a Hampel filter (as in[33]) to remove large erroneous values and replace them with the median of a one-day rolling window. Groundwater observations are then aggregated to an hourly time step. A single time series for the Hague area was interpolated using inverse distance weighting between Pond 1, the two groundwater monitoring wells, and the tidal level at the stormwater system outfall (assumed to be equal to the groundwater table level at the land/water interface). From 2010-2019, the groundwater table is higher than the water level in Pond 1 and lower than the water level in Pond 2, 93.7% and 73.8% of the time, respectively. This indicates that Pond 1 may be gaining water from groundwater flow while Pond 2 may be losing water to groundwater. The groundwater table level is only below the bottom elevation of either pond 0.09% of the time.

The Hague SWMM model provided by the City of Norfolk did not originally simulate groundwater processes and was not configured to easily allow simulation of groundwater exchange with the controlled ponds using SWMM's aquifer components. To address this, a conceptual model of the unconfined aquifer surrounding the existing Hague pond (Pond 1) was developed. Groundwater exchange was calculated externally from the SWMM simulation using the Dupuit equation and added (or subtracted, in the case of infiltration) to the pond as an inflow using pyswmm functionality[40]. The Dupuit equation is commonly used to calculate exchange between a water body and an unconfined aquifer[41] and is written as

$$Q = \frac{K}{2L}(h_1^2 - h_2^2) \cdot A \tag{1}$$

where Q is the seepage rate into or out of the pond, K is the saturated hydraulic conductivity of soil surrounding the pond, h_1 and h_2 are the heights above a fixed datum for the pond water level and groundwater table level, respectively. L is the horizontal distance between h_1 and h_2, and A is the surface area over which seepage can occur (a function of pond water level).

Saturated hydraulic conductivity of the soil surrounding the existing pond (Pond 1) was estimated from the National Resource Conservation Service (NRCS) Web Soil Survey as 0.60m/day. This soil is classified as a fine sandy loam with 61% sand, 22% clay, and 17% silt. Values for h_1 were based on SWMM's simulation of pond water level and h_2 was the observed groundwater table level. Because L controls the hydraulic gradient (when the other variables are held constant), smaller values of L should increase exchange between the ponds and the simulated aquifer. The sensitivity of pond depth and inflow to the distance between measured water levels (L), was tested for $L = 7.62$, 3.0, 1.5, and 0.3m using the passive (i.e., uncontrolled) SWMM model. A single value of L was chosen and used to demonstrate the impact of groundwater exchange on flooding and water quality with the control methods.

The impact of groundwater exchange with the controlled ponds was evaluated for the month of September, 2016. This month had two hurricanes and one tropical storm, which caused the groundwater table level to reach a height of 1.08m (compared to the mean of 0.61m). Because groundwater exchange may increase infiltration and reduce flooding and TSS outflow from the controlled ponds, a direct comparison of a single RTC method's performance with and without groundwater exchange may not be fair. To account for this, the percent difference between the passive system and each RTC method's total flooding and TSS loads (with or without groundwater exchange) will be compared.

2.2.3 Water Quality Simulation

Water quality processes, specifically for TSS, were modelled using SWMM's buildup, washoff, and treatment equations[42]. TSS was chosen for this study to allow comparison with previous RTC literature, and because it is straight-forward to simulate (through gravitational settling) and known to carry other sorbed pollutants[43]. Pollutant buildup within each subcatchment is modelled as a power function

$$B = min(C_1, C_2 \cdot t^{C_3}) \tag{2}$$

where B is the buildup of TSS (mass per unit area), C_1 is the maximum buildup possible, C_2 is the buildup rate (buildup per day), t is the antecedent dry period, and C_3 is a dimensionless buildup time exponent. Washoff of accumulated TSS from subcatchments is modelled with an exponential function

$$W = E_1 \cdot q^{E_2} \cdot B \tag{3}$$

where W is the washoff rate (mass per area per hr), E_1 is the washoff coefficient (per unit of rain), q is the runoff rate (per hr), E_2 is the washoff exponent, and B is the amount of built-up pollutant remaining. Treatment of TSS occurs in the retention ponds and is modelled as a first order decay based on a generalized settling velocity (similar to[17]) with resuspension as a factor of depth and inflow velocity (inspired by[6])

$$C = \begin{cases} TSS \cdot exp(-v_s/DEPTH \cdot DT/3600)) & FLOW \leq \tau \\ TSS & FLOW > \tau \\ TSS \cdot (1 - exp(-v_s/DEPTH \cdot DT/3600)) & FLOW > \tau,\ DEPTH \leq \delta \end{cases} \tag{4}$$

where C is the TSS concentration (mg/L) in the pond after treatment, TSS is the inflow concentration, v_s is the generalized settling velocity (m/hr), $DEPTH$ is the pond water depth (m), DT is the SWMM routing time step (seconds), $FLOW$ is the total inflow rate (m^3/s) (including groundwater, when simulated), τ is a flow threshold to distinguish when settling occurs, and δ is a depth threshold to distinguish when resuspension occurs (one quarter of the maximum pond depth in this implementation). The first case in Eq. (4) allows settling over the simulation time step when inflow is low and reduces TSS concentration. When the inflow rate is above the threshold, no settling occurs (i.e., no TSS treatment). The final case in Eq. (4) simulates resuspension when inflow is high and the pond depth is low by increasing the TSS concentration by the amount that would have been settled according to v_s. Resuspension is included because RTC creates the potential for low water depths in retention ponds; if a pond is drawndown before high storm inflows, sediment may be resuspended and carried downstream.

Each land cover category within the SWMM model domain is given individual characteristics for the buildup and washoff processes (starting values were taken from[44]). With no observed pond water quality data available, the SWMM pollutant processes were calibrated to the annual loading and treatment percent of TSS in Pond 1 (the existing pond) (Table 1). TSS loading was estimated using the loading rates provided in Norfolk's Virginia Stormwater Management Permit[45]. The treatment efficiencies of the passive retention ponds were assumed to be 60% as specified in the Virginia Department of Environmental Quality's Chesapeake Bay TMDL Special Condition Guidance[46]. The load into Pond 1 was calibrated using the buildup coefficient C_2 so that the mean annual load over 2010-2019 was within 2% of the estimated value. The treatment was calibrated using the flow threshold (τ) and the settling velocity (v_s) so that the mean annual reduction was within 5% of the estimated value for the passive simulation. While calibrating this SWMM model to observed values would be desirable, the scope of this paper is on comparison of the RTC methods and not exact quantification of TSS.

Table 1. Calibrated buildup, washoff, and treatment parameters used in the Hague SWMM model. Note that treatment occurs in the stormwater ponds and is not dependent on land cover.

	Buildup			Washoff		Treatment	
Land Use	C_1	C_2	C_3	E_1	E_2	v_s	τ
Impervious	150.16	0.364	1.54	6.97	1.57		
Turf Grass	62.0	0.325	1.26	5.91	1.46	0.105	5.66
Trees	9.22	0.133	0.87	2.11	1.02		

2.3 Real-time Control Scenarios

Real-time control of the Hague stormwater system was simulated with three strategies and compared to the passive system. The three control strategies are (i) predictive RBC with a fixed detention time, (ii) TSS concentration-based RBC, and (iii) RL approaches that includes simulated real-time measurement of TSS concentration in the system state and/or reward function. In the passive system scenario, weirs control flow out of the retention ponds and maintain a permanent pool of approximately half capacity. In the RTC scenarios, the passive weirs are replaced with valves. The valve on Pond 1 is at the same elevation of the passive weir (i.e., halfway up the pond's side) due to pipe configuration constraints). The valve on Pond 2 is at the bottom of the pond side, which allows Pond 2 to be fully emptied or filled. Both RBC scenarios represent local (i.e., individual) control of the retention ponds, while RL can coordinate its control actions based on system-level information. The pyswmm Python package[40] is used to implement all RTC scenarios on a standard PC with 8 cores, 16GB RAM, and an NVIDIA Quadro P2000 Graphical Processing Unit (GPU).

2.3.1 Detention Rule-based Control

In this scenario, RBC is based on industry standard methods that use rainfall forecasts for predictive control of stored water to mitigate flooding, while controlling water quality with a fixed detention time[12,14,47]. The general process of this RBC (RBC-DTN) is shown in Figure 2 and detailed in[10]. Briefly, if a forecast storm is expected to flood the pond, the valve will open to drain an equivalent volume of water (plus a safety factor). When the pond is drawndown sufficiently, the valve will close to retain the incoming runoff for a fixed time (24hr in this case). At the end of the retention period, the valve opens to the minimum setting to bring the water level back to the target operating depth within a fixed time (24hr). Outside of storm events, the valve operates based on the observed water level in order to maintain a target depth in the pond. A fail-safe rule overrides any previous rules by completely opening the valve if the pond is flooding. A decision diagram detailing these rules is shown in Appendix A (Fig. 1).

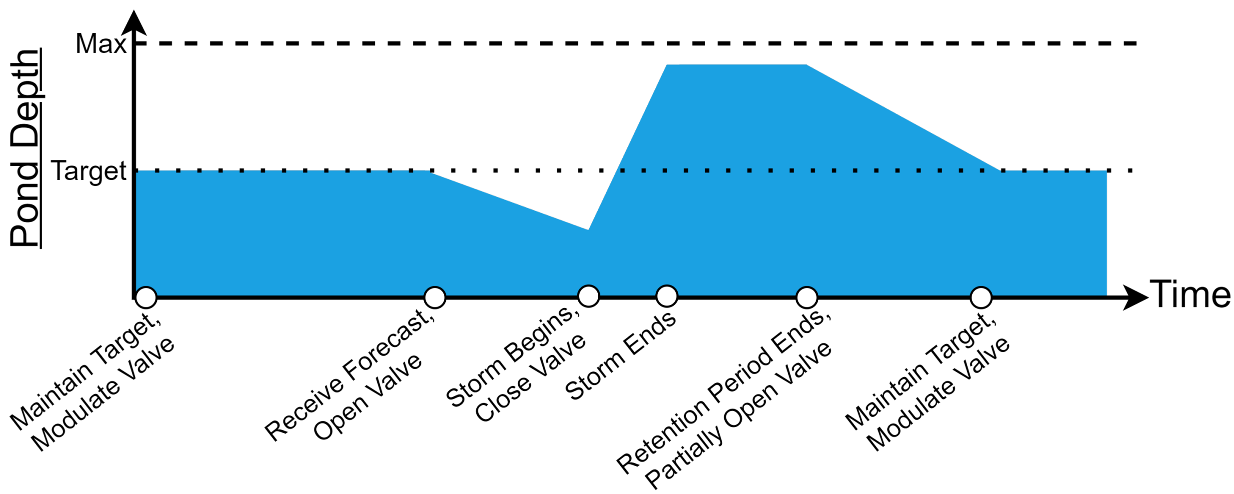

Figure 2. General schema of the Detention Rule-based Control (RBC-DTN) scenario. Forecasts allow predictive control of the pond water level to mitigate flooding while a fixed detention time after storm events helps improve water quality.

2.3.2 TSS Rule-based Control

The TSS RBC (RBC-TSS) scenario was inspired by Sharior et al.[17]. Instead of using a fixed detention time, this RBC is innovative because it uses the real-time concentration of TSS in a retention pond to trigger valve operation (Fig. 3). For example, when the TSS concentration is above a threshold, the valve can be closed to retain stormwater and allow treatment by settling. Otherwise, the valve is open and acts as a weir to maintain a permanent pool of water. In this study, the TSS threshold was set to 1 mg/L because observed data from the ponds were not available for a more realistic threshold; in Sharior et al.[17], the threshold is 15 mg/L based on regulatory constraints for their study area and calibrated model. A contingency rule limits flooding of the pond by opening the valve if a threshold depth is reached. A decision diagram detailing these rules is shown in Appendix A (Fig. 2).

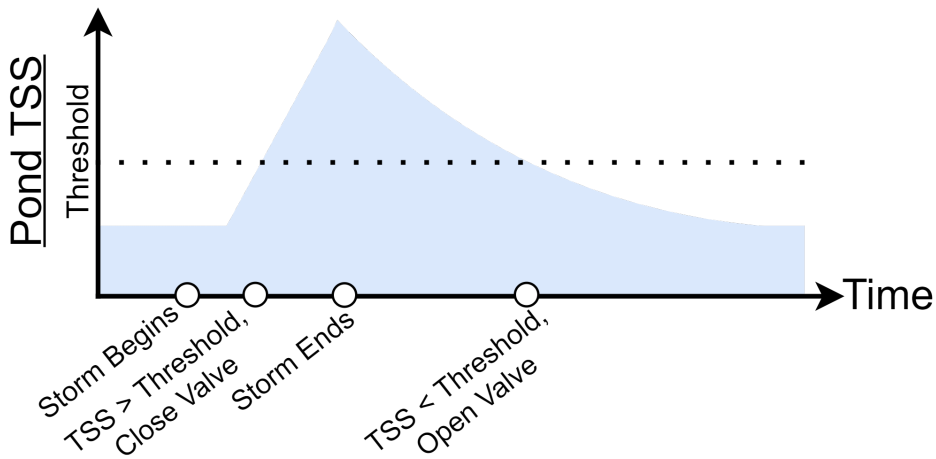

Figure 3. General schema of the TSS Rule-based Control (RBC-TSS) scenario. Detention is based on observed TSS concentration, not a fixed length of time, making it adaptive to individual storm events.

2.3.3 Reinforcement Learning

Reinforcement learning can be visualized as an agent that interacts with an environment (Fig. 4). The RL agent learns through sequential interactions with the environment. At each step in the learning process, the RL agent receives information about the state (s) of the environment and can take actions (a). The next state (s'), therefore, depends on the agent's actions and the agent is rewarded (positively or negatively) based on how well its actions meet user-specified objectives in a reward function (r). The agent's ultimate goal is to find a policy ($\pi(a|s)$) that maximizes the expected return

$$G_t = r_t + \gamma r_{t+1} + \gamma^2 r_{t+2} + \cdots = \sum_{k=0}^{\infty} \gamma^k r_{t+k} \tag{5}$$

where $r_t = r(s_t, a_t, s_{t+1})$ and $\gamma \in [0, 1]$ is a discount factor weighting the importance of short-term and long-term reward.

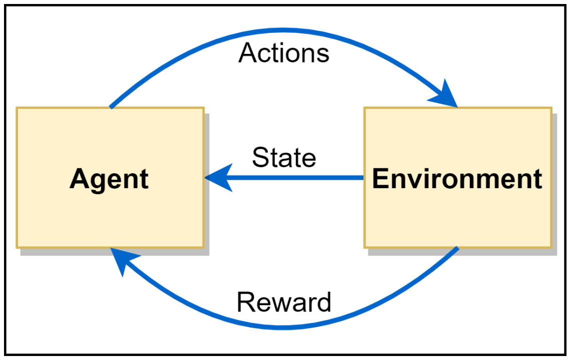

Figure 4. Reinforcement learning paradigm.

In this case the environment is the SWMM model described in section 2.2, which provides state information at a 15-minute simulation time step. The state space (S) is defined as: the current depth (m) and outflow (m^3/s) of the two retention ponds, the load of TSS (mg) in each pond's outflow, the current position of each controllable valve, the sum of the 24 hr rainfall forecast (mm), and the mean value of the 24 hr tide forecast (m). The action space (A) of the agent is to open or close either valve to any degree. The reward (r) is based on how well the agent meets user-specified objectives such as flood and pollutant reduction.

The deep reinforcement learning algorithm used in this research, Deep Deterministic Policy Gradients (DDPG), is an actor-critic RL agent using deep neural networks as function approximators[28]. DDPG allows controls (i.e., valve positions) over a continuous action state and has been used in previous research to learn control policies that mitigate flooding[10,30,31]. The actor in DDPG is a deep feed-forward neural network that learns a policy ($\pi(a|s)$); the critic is a deep feed-forward neural network that approximates the value of being in a specific state and taking specific actions called the Q-value

$$Q^{\pi}(s,a) = r(s,a,s') + \gamma \sum_{s' \in S} P^{a}_{s,s'} \sum_{a' \in A} \pi(a'|s') Q^{\pi}(s',a') \tag{6}$$

where $P^{a}_{s,s'}$ is the probability of transitioning between two states. This equation is known as the Bellman equation and is a key component of RL[21]. By approximating the Q-value, the critic can reduce the variance of policy gradients from the actor, which helps speed the learning process. During training, the actor receives the state of the stormwater system and outputs the actions to be taken based on its learned policy. The critic then receives the actions and states and outputs an estimated Q-value. The actions and Q-value estimates output from the critic are used to update the agent. An in-depth description of the DDPG algorithm can be found in Lillicrap et al.[28].

When training RL agents, more explicit reward functions can improve the ability to learn appropriate policies[29]. The reward functions used here have a conditional format based on the rainfall forecast that guide agent learning under different conditions; this structure has been used to improve RL agent policies for flood mitigation[10]. The rainfall forecast can signal to the agent that flooding may occur and control actions are rewarded differently because the pond level may need to be altered from the target. When no rainfall is forecast, a different set of rewards is triggered that incentivize actions for goals like maintaining target depths or increasing retention time for additional TSS treatment.

In this research, three RL agents are trained and tested. The first agent (RL-FD) is rewarded for reducing total flooding throughout the stormwater system and maintaining target pond depths

$$r = \begin{cases} -\Sigma Flooding[system,\ Pond1 * 1000,\ Pond2] & F \geq \delta \\ -(|Pond1_{depth} - \tau| + |Pond2_{depth} - \tau|) & F < \delta \end{cases} \tag{7}$$

where $Flooding[system]$ is the incremental system flood volume, $Flooding[Pond1]$ is the flooding rate at Pond 1, and $Flooding[Pond2]$ is a binary reward (0 or 1000). F is the sum of rainfall in a 24hr forecast, δ is the rainfall threshold (12.7mm in this research), and τ is the target depth (1.8m and 1.1m for Ponds 1 and 2, respectively). Several of the nodes upstream of Pond 2 are at lower elevations than the top of the pond and flood before Pond 2 will; therefore instead of the Pond 2 flooding rate, the binary reward acts as a penalty in cases where the pond is above the depth that causes flooding upstream (1.75m).

The second RL agent (RL-FDTSS) is rewarded for reducing total flooding throughout the stormwater system, maintaining target pond depths, and minimizing the export of TSS from the ponds

$$r = \begin{cases} -\Sigma Flooding[system,\ Pond1 * 1000,\ Pond2] & \\ \qquad +TSS[Valve1,\ Valve2] & F \geq \delta \\ -(|Pond1_{depth} - \tau| + |Pond2_{depth} - \tau| & \\ \qquad +TSS[Valve1,\ Valve2] + Flooding[system/35000]) & F < \delta \end{cases} \tag{8}$$

where $TSS[Valve1, Valve2]$ is the incremental TSS load of the controlled valves.

The third RL agent (RL-FD+FDTSS) aims to balance RL-FD and RL-FDTSS by initializing the trained neural network weights and memory from RL-FD and training for 50,000 additional time steps using the reward for RL-FDTSS (Eq. 8). This can be considered as pre-training for RL-FD+FDTSS, a common practice in deep machine learning to provide appropriate initial conditions and reduce computational time (for examples in hydrology see[48] or[49]).

The RL agents are trained on one month of data (August, 2019), which has the fifth highest monthly total rainfall (256.5mm) of the dataset distributed across 7 storm events. The mean tide level in this month is 0.16m, with a maximum value of 1.0m from Tropical Storm Erin late in the month. In previous research, this month of data was found to provide a representative range of state information that allowed an RL agent to learn effective flood mitigation policies[10], while also keeping computational costs reasonable. A visualization of the rainfall and sea level training data, as well as the TSS concentration in each pond is given in Figure 5. RL-FD is trained for 100,000 steps of the training data with a discount factor (weighting of current and future rewards) of 0.5. RL-FDTSS and RL-FD+FDTSS are both trained for 150,000 steps, when the pre-training from RL-FD is considered for RL-FD+FDTSS, with a discount factor of 0.99. RL agents are tested on the remaining data (2010-2019). Each RL agent has the same neural neural network architecture; these and the shared RL hyperparameters are documented in the open source code repository linked in section 5. The DDPG algorithm is implemented with the keras-rl[50], openai gym[51], and Tensorflow[52] python packages; the wandb[53] python package was used for tracking training progress and comparing agents during the hyperparameter tuning process.

2.3.4 RTC Comparisons

The RTC scenarios are evaluated in three main comparisons as shown in Table 2; each comparison has a different control scale and prioritization of flooding and water quality improvement. First, a baseline for flood mitigation is established by comparing the passive system and RL-FD, which does not consider water quality in its control policy. While the design of the passive system does consider pollutant treatment, the main focus is on flood mitigation. Second, trade-offs between the RBC methods, which focus on flooding and TSS at the pond scale, are compared to the passive system. In this comparison, the RBC controls prioritize enhancing pollutant treatment as this is one of the largest benefits these systems have had in practice. Third, system-level control trade-offs with RL-FDTSS and RL-FD+FDTSS, which considered both flooding and TSS in their training, are compared to the passive system and RL-FD. These three comparisons are made without simulating groundwater exchange to keep the focus on control actions and reduce computational expense. The impact of groundwater exchange is then examined on a subset of the data to evaluate its potential impact on RTC of the stormwater system.

Table 2. Comparisons of stormwater control scenarios

Comparison	Control Method	
Baseline	Passive	RL-FD
Local	RBC-DTN	RBC-TSS
System	RL-FDTSS	RL-FD+FDTSS

3 RESULTS

3.1 Baseline Flood and TSS Control

Figure 5 illustrates how the passive system and RL-FD respond to the storm events in August, 2019. Operation of Pond 1 is similar between these two methods because the controllable valve is at the same elevation as the fixed weir; water is released as soon as depth increases from a storm event. However, RL-FD learned to close the valve when high tide levels caused backflow into the pond to prevent water level fluctuations (e.g., Aug. 26-27). RL-FD learned to lower Pond 2's depth, which is fully controllable, before certain storm events (e.g., the Aug. 4 storm) while remaining close to the target depth during dry periods.

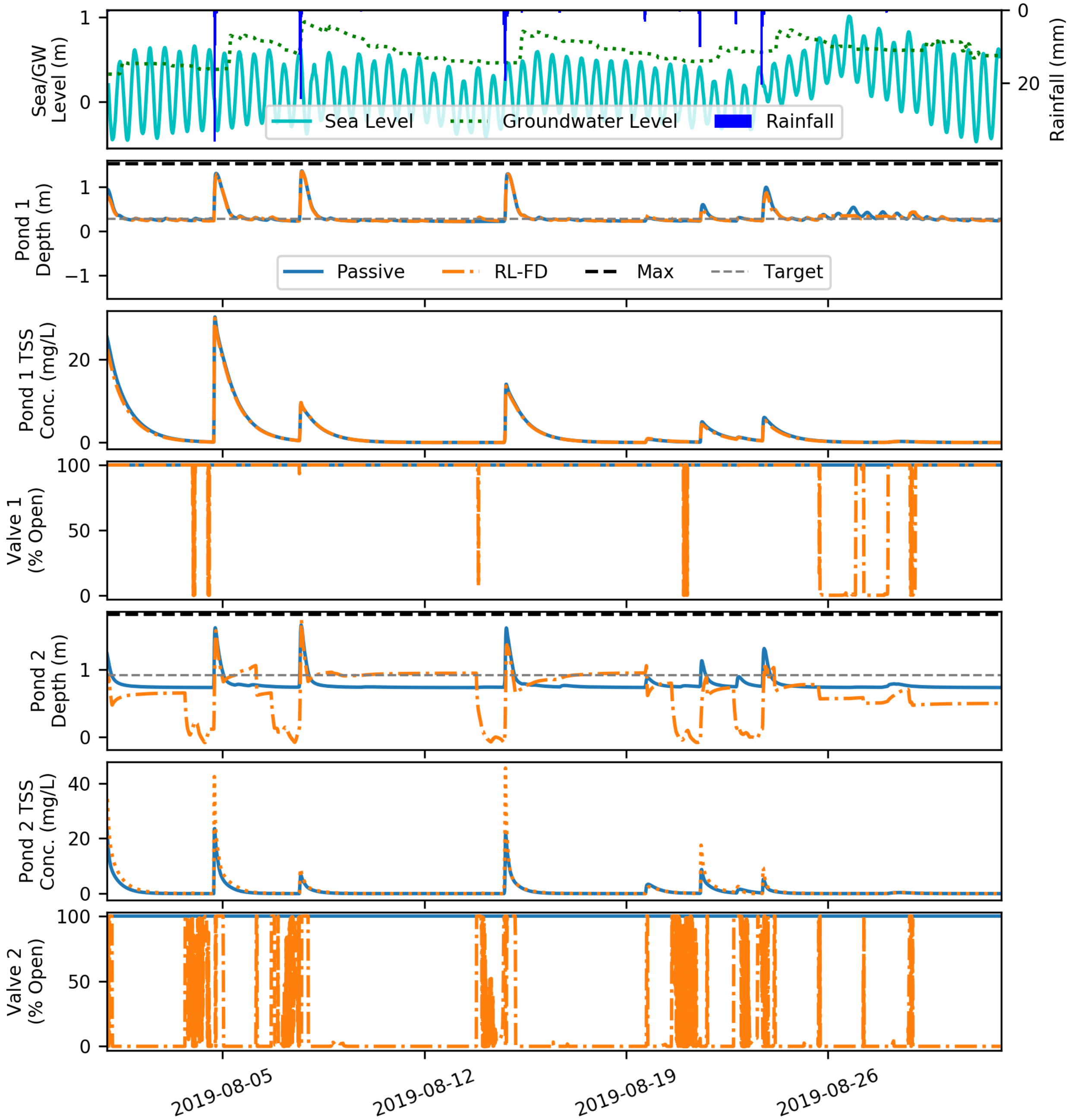

Figure 5. Comparison of passive and RL-FD system operation for August, 2019. From top to bottom, these plots illustrate the hydrological model drivers (rainfall, sea level, and groundwater level) and the depth, TSS concentration, and valve position for Ponds 1 and 2, respectively. In this case, the passive system cannot alter its behavior, while RL-FD can control the valves in response to observed and forecast water quantity conditions.

The system-level control policy learned by RL-FD allowed it to reduce the total volume of flooding by 4.0% (72301m^3) compared to the passive system (Fig. 6, A). While RL-FD's training did not include any water quality information, it's policy does provide improved TSS capture at both ponds (i.e., lower loads at the valves). Compared to the passive system, RL-FD reduced TSS by 15.1% (16436kg) and 14.8% (14074kg) at Valves 1 and 2, respectively (Fig. 6, B).

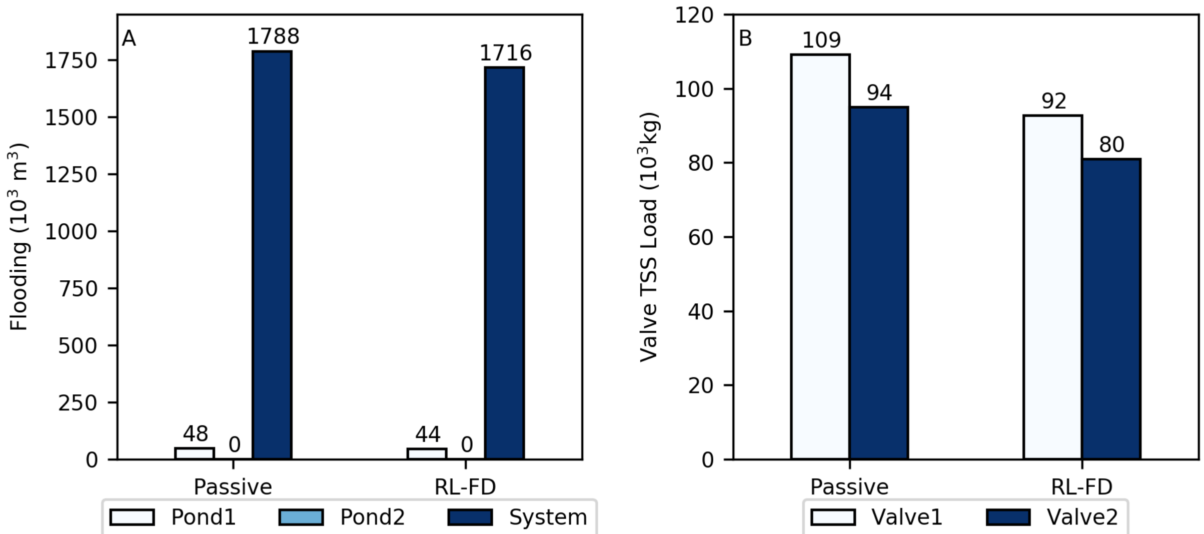

Figure 6. Total flood volumes (A) and TSS loads (B) for the passive and RL-FD baseline scenarios, 2010-2019.

3.2 Local Control with RBC

An example of the RBC methodologies compared to the passive system is shown in Figure 7. Both RBC methods operate the ponds individually (i.e., rules are not coordinated between the ponds) to mitigate flooding of the pond by releasing water or to improve water quality by retaining runoff after a storm event. RBC-DTN has a fixed detention time, while RBC-TSS adapts detention time based on the concentration of TSS in the pond. For example, after the Aug. 8 storm both methods retain water for similar amounts of time. This indicates that the fixed 24hr retention time of RBC-DTN was adequate to treat the TSS washed into the ponds after the short buildup period following the Aug. 4 storm. However, after longer periods of TSS buildup, RBC-TSS retains stormwater longer than RBC-DTN until TSS is sufficiently treated and the concentrations drop below the threshold (e.g., following the Aug. 15 and 22 storms).

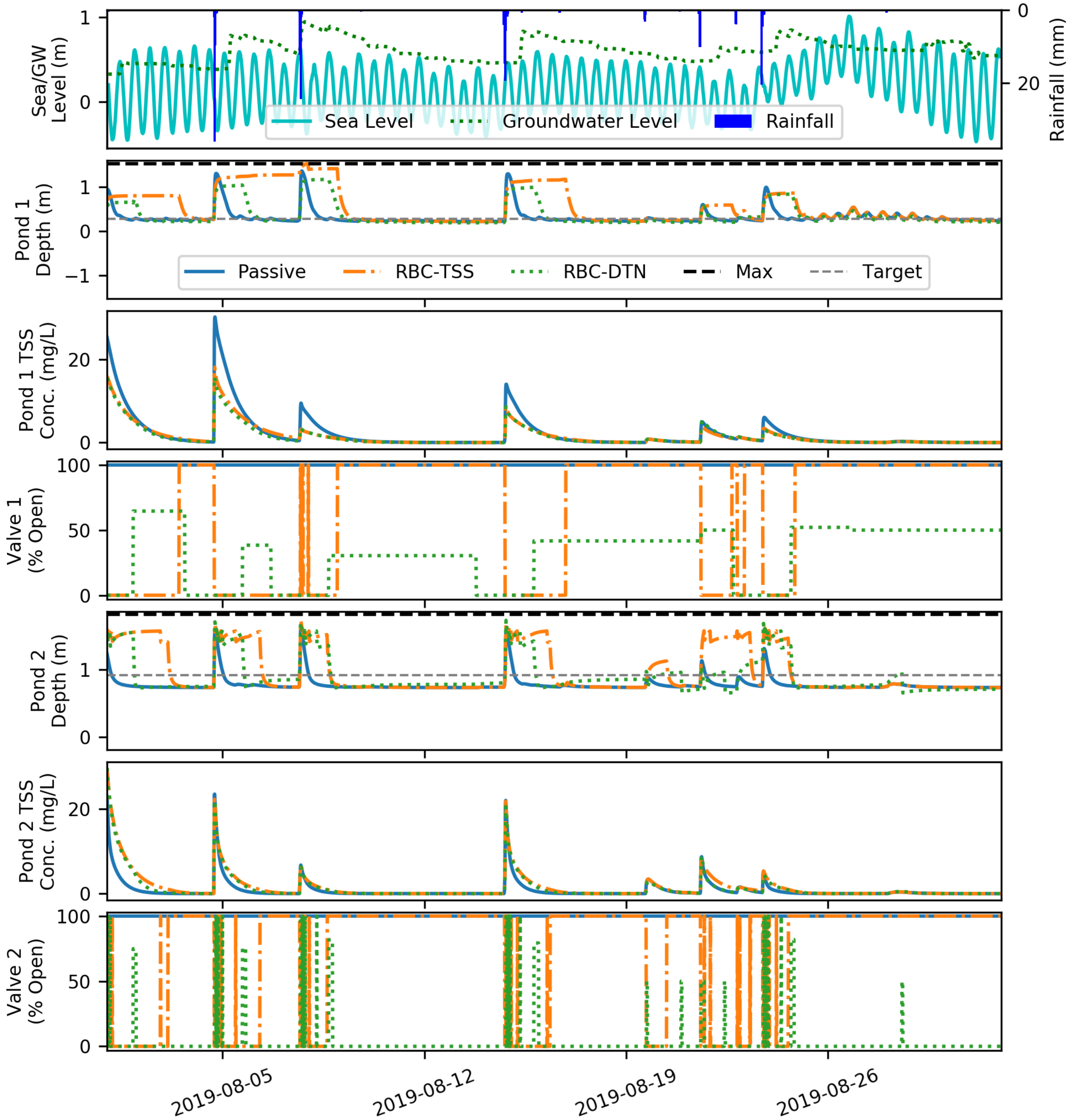

Figure 7. Comparison of local RTC methods (RBC-TSS, RBC-DTN) and passive system operation for August, 2019. From top to bottom, these plots illustrate the hydrological model drivers (rainfall, sea level, and groundwater level) and the depth, TSS concentration, and valve position for Ponds 1 and 2, respectively. In this case, the passive system cannot alter its behavior, RBC-DTN retains water for a fixed period after storm events to allow settling of TSS, and RBC-TSS adaptively retains water when TSS concentrations are above a threshold.

Both rule-based control methods provide reductions in TSS export from the controlled ponds compared to the passive system. However, this is at the expense of increased flooding because operation of the two valves is not coordinated and does not consider flooding in other parts of the stormwater system (Fig. 8). Compared to the passive system, RBC-TSS increased total system flood volume by 12.0% (215011m^3), while decreasing TSS by 95.5% (104222kg) and 32.8% (31116kg) at Valves 1 and 2, respectively. RBC-DTN increased flooding by 9.0% (161259kg) and decreased TSS for Valves 1 and 2 by 49.2% (53710kg) and 4.5% (4227kg) compared to the passive system. RBC for Pond 2 does not treat TSS as efficiently as Pond 1 due in part to the system configuration. Water needs to be released if the Pond 2 depth exceeds 1.75m; this is necessary to alleviate upstream flooding due to this SWMM model's specific pipe configuration. Further, valve 1 is approximately halfway up the side of Pond 1 (i.e., Pond 1 cannot be fully drained), which increases the detention time compared to Pond 2.

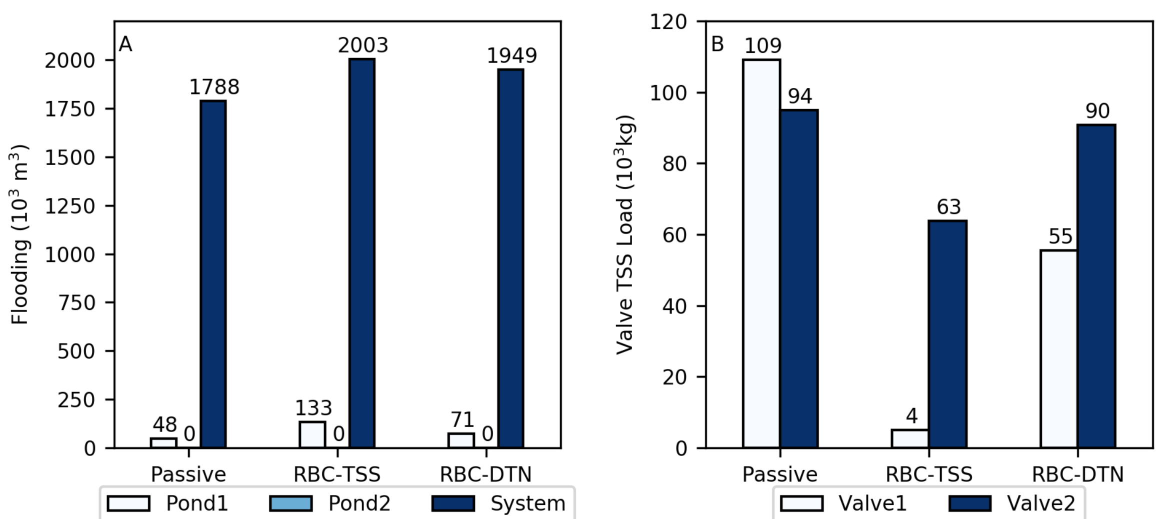

Figure 8. Total flood volumes (A) and TSS loads (B) for local RTC methods (RBC-TSS, RBC-DTN) and passive system operation, 2010-2019.

3.3 System-level Control with RL

Both RL-FDTSS and RL-FD+FDTSS learned policies with multiple objectives of flood mitigation, TSS reduction, and target pond depths. When tested on the training data (Fig. 9), these agents generally kept valve 1 open to maintain the target depth before and between storms (neither agent can lower the water level in Pond 1 below the target, due to the valve placement) and closed valve 1 during storms to capture TSS. After storm events, RL-FDTSS held water to improve TSS treatment; in contrast RL-FD+FDTSS closed valve 1 long enough to capture initial TSS inflow, but quickly released water to return the pond to its target depth. The agents have similar policies for valve 2 that favor holding water above the target depth to treat TSS while draining the pond before storm events to prevent flooding. However, RL-FDTSS tended to release water more gradually and hold it at high levels between storms than RL-FD+FDTSS, increasing TSS treatment.

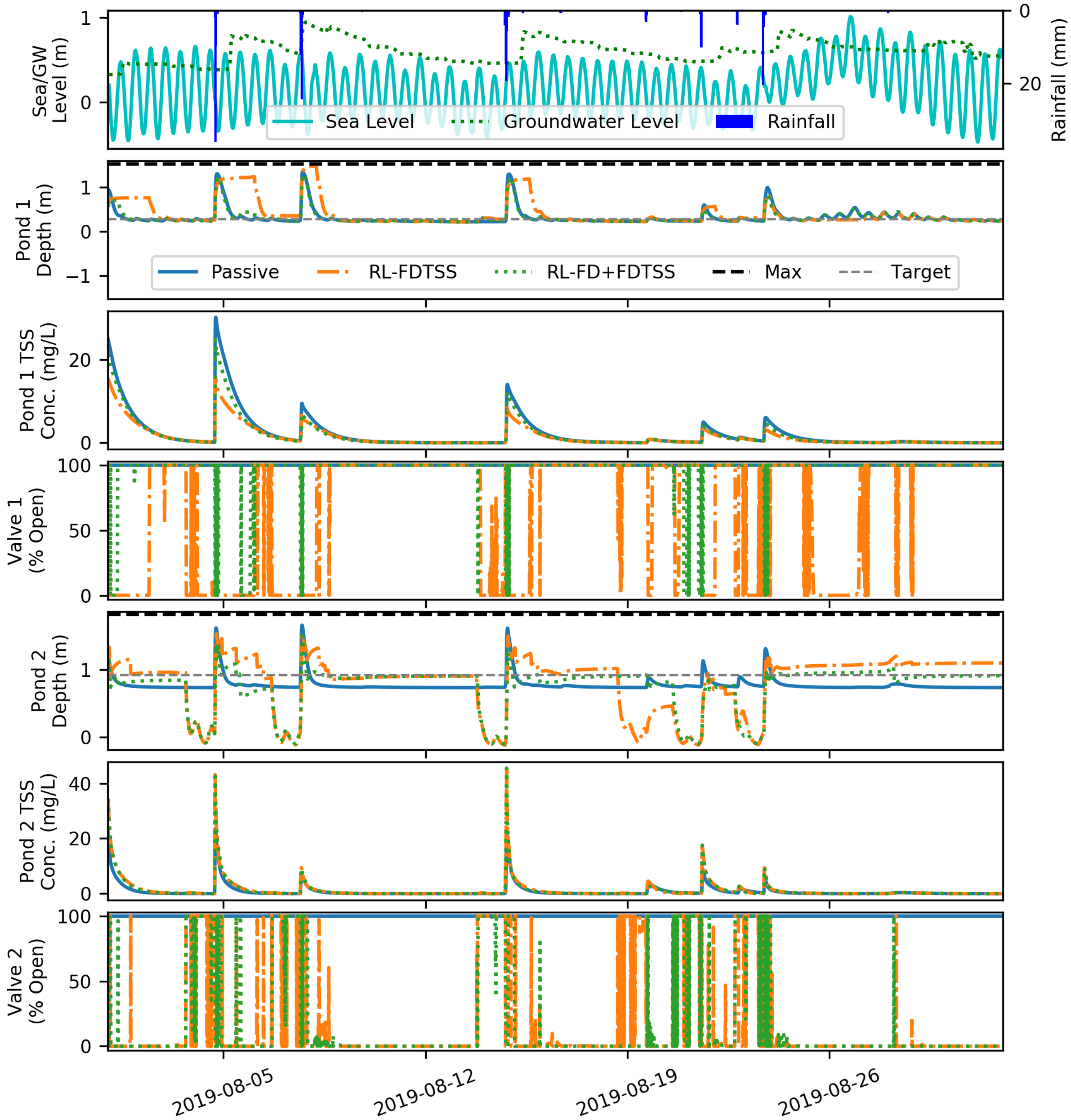

Figure 9. Comparison of RL-FDTSS, RL-FD+FDTSS, and passive system operation for August, 2019. From top to bottom, these plots illustrate the hydrological model drivers (rainfall, sea level, and groundwater level) and the depth, TSS concentration, and valve position for Ponds 1 and 2, respectively. The passive system cannot alter its behavior; RL-FDTSS and RL-FD+FDTSS use water quantity and quality (i.e., TSS observations) information to make control decisions. RL-FD+FDTSS was pre-trained from RL-FD and learned a different balance of flood and TSS control than RL-FDTSS.

On the test dataset (2010-2019), RL-FDTSS had 11.3% (212740m^3) more total system flooding and 74.6% (179429m^3) more flooding at Pond 1 than RL-FD+FDTSS (Fig. 10). Both RL-FDTSS and RL-FD+FDTSS increased system-wide flooding compared to the passive system by 16.8% (300183m^3) and 4.9% (87443m^3), respectively. In terms of TSS reduction, both of these agents provide improvements compared to the passive system. RL-FDTSS reduced TSS by 95.1% (103816kg) and 81.3% (77185kg) at valves 1 and 2, while RL-FD+FDTSS reduced TSS by 39.5% (43129kg) and 65.0% (61701kg).

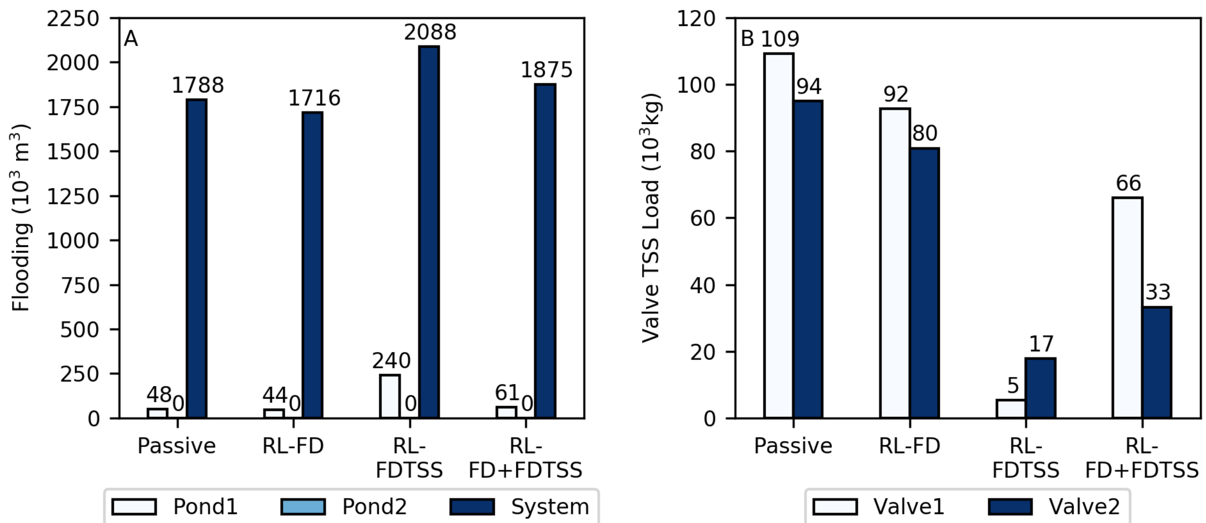

Figure 10. Total flood volumes (A) and TSS loads (B) for RL-FDTSS and RL-FD+FDTSS, 2010-2019.

3.4 Multi-objective Comparison of RTC Methods

A comparison of performance trade-offs for each stormwater control method is shown in Figure 11. In terms of flood volume, only RL-FD reduced flooding compared to the passive system at both the system-level and at Pond 1. RL-FD+FDTSS outperformed the local-scale RBC methods and RL-FDTSS. Pond 2 did not flood in any of the scenarios because of the configuration of this SWMM model; several nodes upstream of Pond 2 have lower maximum depths and flood with any rainfall when the pond is above a certain level.

All RTC methods reduced TSS loads at both valves compared to the passive system. TSS load reduction at valve 1 was greatest for RBC-TSS and RL-FDTSS; RBC-TSS used water quality

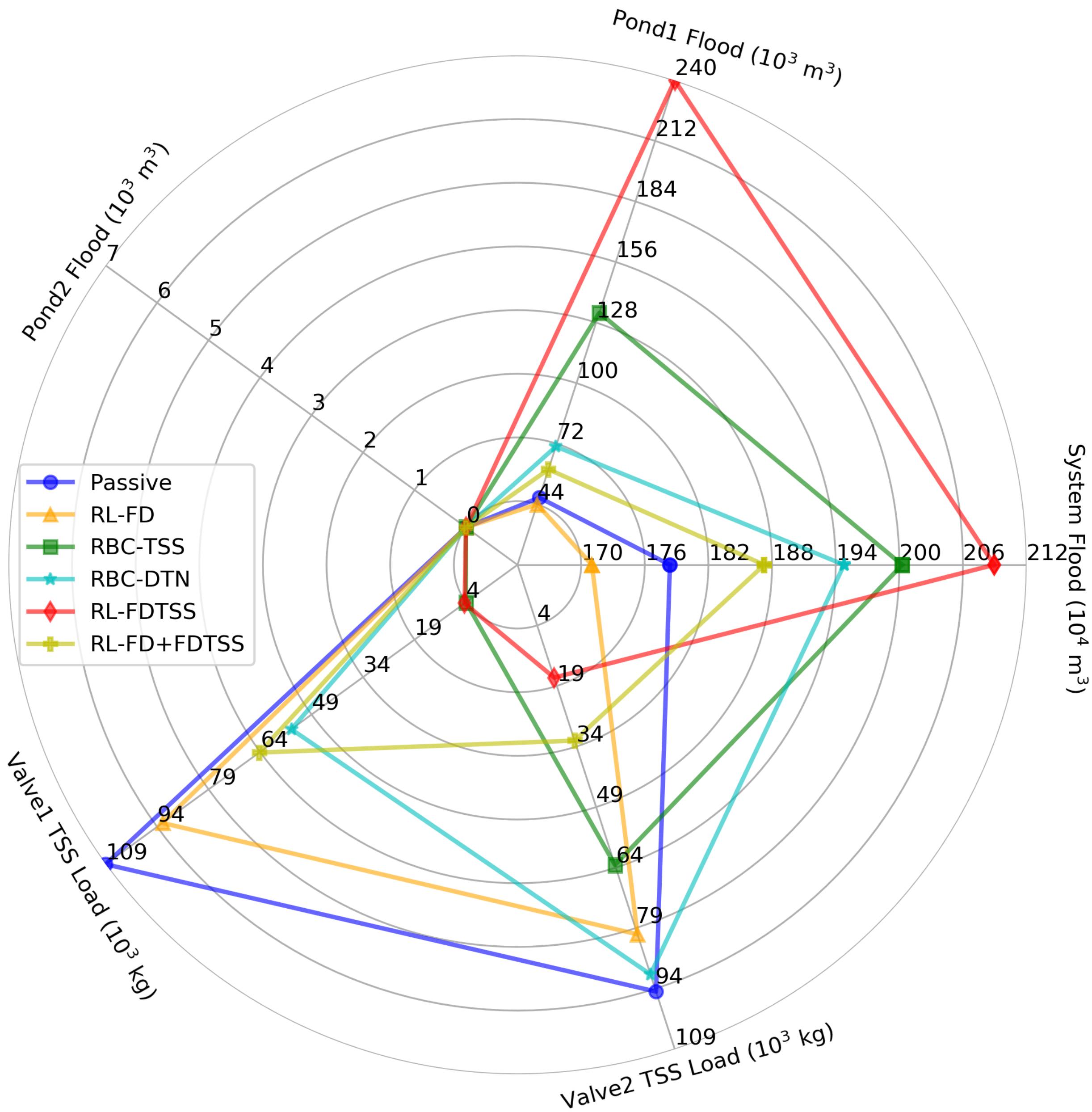

Figure 11. Comparison of flood volume and TSS load trade-offs for each control method, 2010-2019.

observations to inform control, while RL-FDTSS learned a control policy from scratch that included penalties for high TSS loads. At valve 2, the local-scale RBC methods had fixed rules to release water when Pond 2's depth reached the threshold for upstream flooding. This limited their ability to capture the first flush of TSS during large storm events. The system-level RL agents outperformed the passive system and had similar trends in performance for both valves. RL-FD did not consider

TSS in its policy and had the smallest reduction; RL-FD+FDTSS, which had some training with the reward function for RL-FDTSS, had more TSS reduction than RL-FD. RL-FDTSS was trained from start to finish with a reward function that penalized TSS export from the ponds and had the greatest reductions in TSS.

The RTC methods made varying degrees of progress towards meeting the city's TMDL TSS reduction goals of 5.75, 35, and 60% by 2021, 2026, and 2031, respectively. The percent reductions achieved by the RTC methods compared to the passive system are given in Table 3. All RTC methods exceeded the 5.75% reduction goal. RL-FD+FDTSS exceeded the 35% goal and both RBC-TSS and RL-FDTSS exceeded the 60% goal.

Table 3. Percent reduction in total pond TSS export for each RTC method compared to the Passive system.

Control Method	RL-FD	RBC-DTN	RBC-TSS	RL-FDTSS	RL-FD+FDTSS
Reduction (%)	14.94	28.38	66.29	88.66	51.35

In terms of maintaining the target depth at Pond 2, RBC-TSS was most similar to the passive system because the valve was at the same height as the target depth (Fig. 12). However, RBC-TSS was able to close the valve to treat TSS and therefore had a greater percentage of time above the target compared to the passive system. RBC-DTN and the RL agents could fully drain or fill Pond 2 and had a greater percentage of time at lower depths. This helped prevent the pond from flooding, but long periods of time at low depths are undesirable in reality. The target depth comparison also illustrates differences in policy learned by RL-FDTSS and RL-FD+FDTSS. Across the entire test set, RL-FDTSS had a tendency to keep Pond 2 at very low water levels. In contrast, RL-FD+FDTSS's policy kept the water level at or above the target depth approximately 90% of the time, indicating that it learned a policy to only drain the ponds when needed (a benefit of pretraining RL-FD+FDTSS from RL-FD).

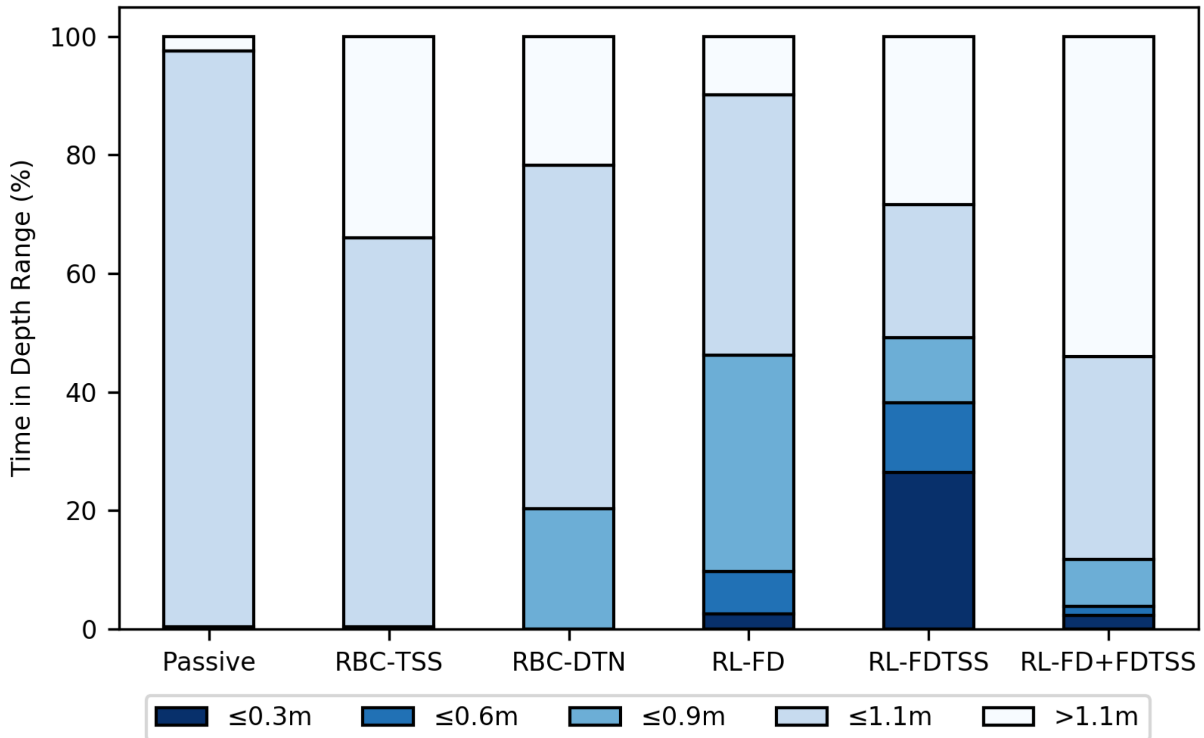

Figure 12. Comparison of time below or above the Pond 2 target depth (1.1m) for each control method, 2010-2019.

3.5 Impact of Groundwater Exchange on RTC Methods

In comparing the sensitivity of pond-aquifer flow to the Dupuit fitting parameter L, it was found that L=7.6m and L=3.0m had no noticeable impact on the mean depth of Pond 1, while the mean depth at Pond 2 increased by 14% (Table 4). When L=1.5m, Pond 1 tends to gain a small amount of water, while Pond 2 gains slightly less water compared to the larger values of L. As an example, during the dry period without groundwater exchange between Sept. 9 and 19, the water level at both ponds is slightly elevated compared to the simulation without groundwater exchange (Fig. 13, No GW). When L=0.3m, the mean depth at Ponds 1 and 2 increased by 10% and 7%, respectively. This value of L caused total monthly inflow volume to increase at Pond 1 by 24%. At Pond 2, however, total monthly inflow volume decreased by 1% and the pond lost water between Sept. 9 and 19 (Fig. 13). Because L=0.3m had the largest increase in depth at Pond 1 and altered Pond 2's behavior during dry weather, it was chosen for use in the RTC simulation with groundwater exchange.

Because groundwater exchange also allows increased infiltration, all of the RTC methods have a

Table 4. Percent difference in mean pond depth for groundwater exchange simulated with varying values of L compared to the simulation without groundwater exchange over the month of Sept. 2016.

L (m)	7.6	3.0	1.5	0.3
Pond 1	0.0	0.0	+2.8	+10.3
Pond 2	+13.9	+13.9	+12.8	+6.9

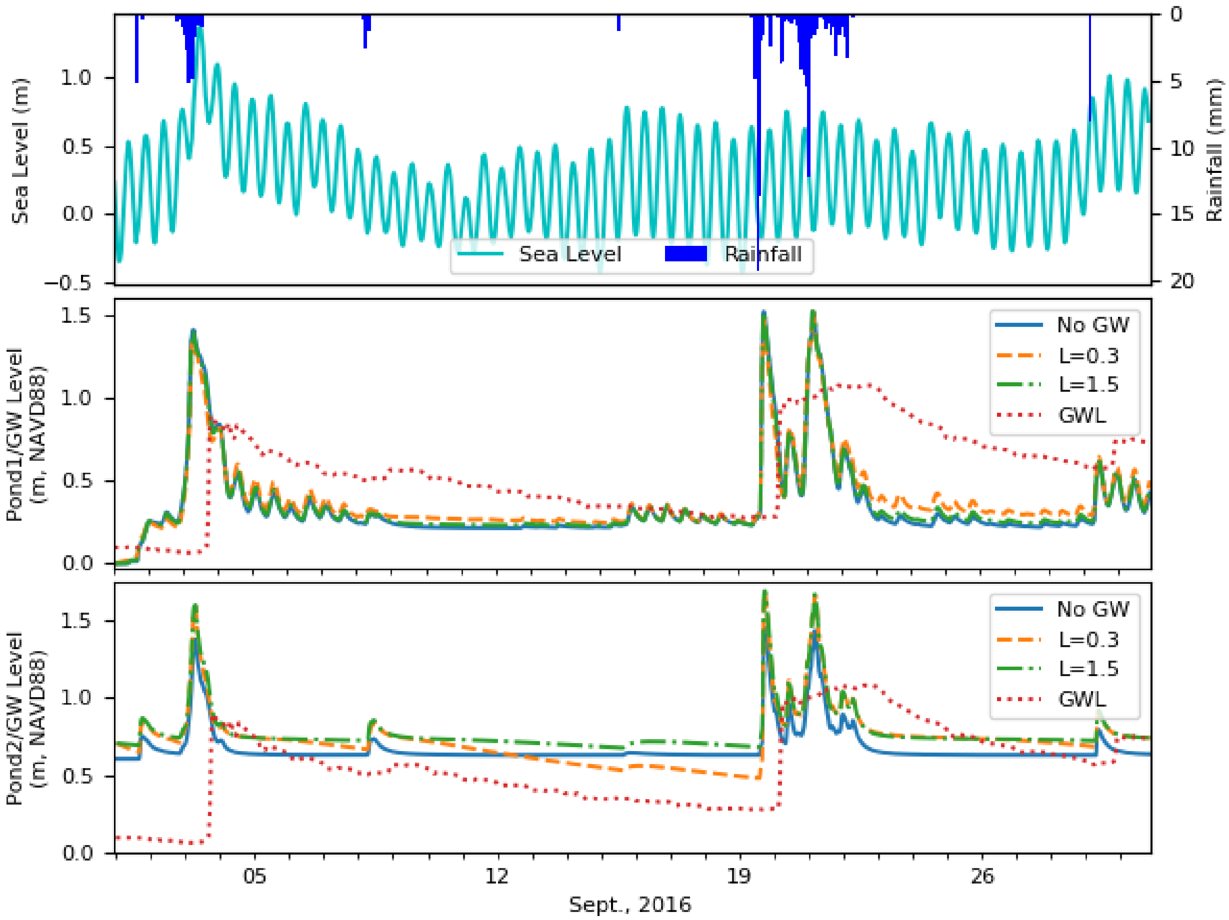

Figure 13. Comparison of passive pond operation for simulations without groundwater exchange (No GW) and with $L = 1.5m$ or $L = 0.3m$ in the Dupuit equation, September, 2016.

smaller change in total flood volume compared to the passive system when groundwater exchange is included (with the exception of RBC-DTN, which had a larger percent change and reduced flooding, instead of increasing it) (Fig. 14, A). All RTC methods were still effective at reducing TSS loads for valves 1 and 2 (Fig. 14, B and C, respectively). Of the RBC methods, RBC-DTN had a smaller decrease in Valve 1 TSS load with groundwater exchange than without. RL-FDTSS was the only RL method to perform worse for TSS reduction when groundwater exchange was added to the simulation. This may indicate overfitting to the training data (which did not include groundwater exchange), limiting RL-FDTSS's ability to control new pond behaviors. An example time series visualization and statistics of valve operation by the RTC methods is available in Appendix A, Figs. 3 and 4).

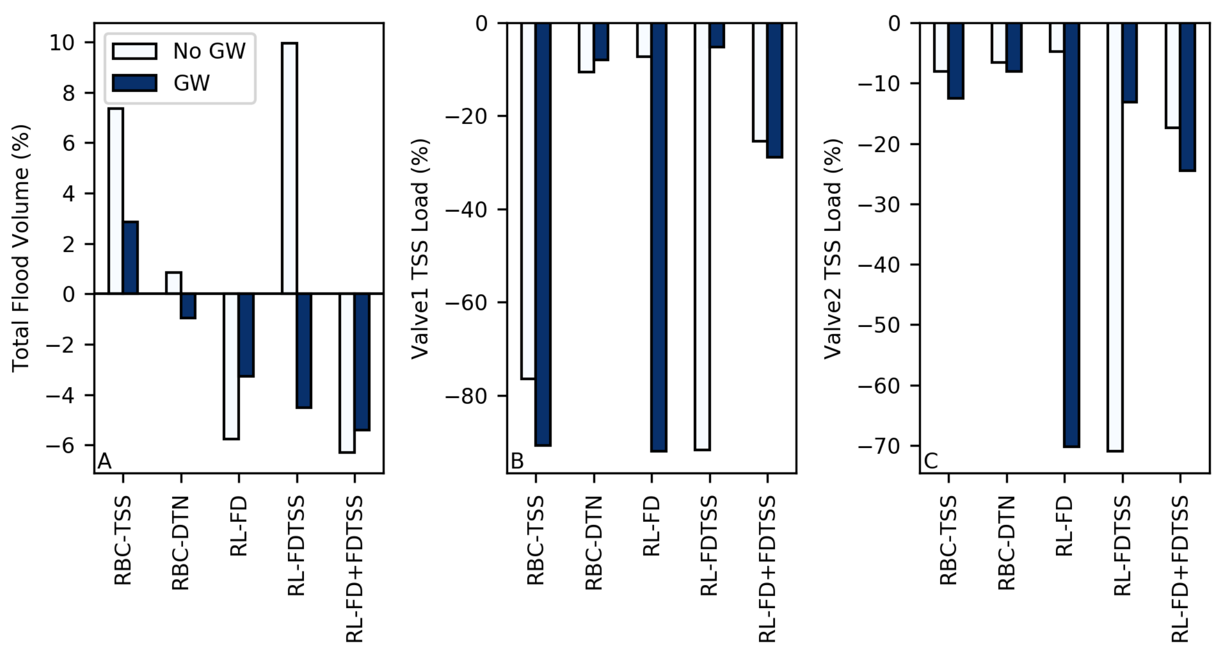

Figure 14. Comparison of percent difference from the passive system for each RTC method's total flood volume (A) and TSS loads (B and C) for simulations with and without groundwater (GW) exchange for September, 2016.

4 DISCUSSION

4.1 Towards System-level Control

As the complexity of an environment and control objectives increases, it becomes much harder for a single RL agent to learn an effective control policy. This can be seen in the performance of RL-FD and RL-FDTSS. RL-FD had fewer goals and a simpler reward function that allowed it to learn an effective policy. In contrast, RL-FDTSS had a more complicated reward function and more goals. While it learned an effective policy for minimizing TSS, that was at the expense of both increasing system flooding and allowing Pond 2 to remain at undesirably low depths for long periods of time. As demonstrated by RL-FD+FDTSS, pretraining from an agent that performs well on simpler, but related, goals is one way to approach this challenge. This pretraining allowed RL-FD+FDTSS to outperform RL-FDTSS for flood mitigation, but at the expense of somewhat reduced TSS treatment. Other methods such as Multi Agent RL (MARL), Multi-Objective RL (MORL), and boosting/ensemble methods may also be beneficial. In MARL, each pond could be controlled by an individual agent tuned to that pond's specific goals, while also operating cooperatively towards system-level goals[54,55]. In MORL, sets of policies are learned to approximate a Pareto frontier[56]; this is especially valuable for comparing trade-offs among agents. Similar multi-objective optimization is well studied for reservoir operation and could provide an alternative to MORL[57]. Boosting and other ensemble methods attempt to combine agent policies or neural network outputs to increase performance[58,59]. In the context of RL for stormwater systems, this maybe beneficial for combining agents that are trained for different purposes (e.g., an agent for extreme events, an agent for average events, an agent for dry periods).

Of the RTC methods implemented here, both RBC-DTN and the RL agents use current observations and forecasts to inform control decisions ahead of storm events. Perfect forecast data were used in this research to keep the focus on the control methodology, however, forecasts can contain a significant amount of uncertainty in reality. As an example specific to coastal systems, tide forecasts are based on the astronomical tide cycle which does not account for storm tides. In practice, RBC implementations have handled forecast uncertainty by using a probability threshold (e.g., take an action if the rainfall forecast probability is greater than 50%), as well as other fail-safes[14]. Stormwater RTC research using linear optimization and water quality control rules found that errors in rainfall prediction (i.e., an unforeseen storm event) could cause flooding of stormwater ponds, but that the system-level control could quickly adapt and recover based on observations of current conditions[60]. Recent work with RL (specifically the DDPG algorithm used in this study) for stormwater RTC has indicated that this algorithm is robust to uncertainty in sensed and forecast data in both training and testing[31]. While quality-controlled observations could be used in off-line training, doing so could limit an agent's performance when deployed and using noisy data to inform control actions. In the current research, the RL agents were robust to altered pond behavior when groundwater exchange was simulated (groundwater exchange was not included in the RL training process). However, as stormwater RTC continues to move towards system-level control to accommodate the increasing density of controlled infrastructure components, changing environmental conditions, and more stringent environmental regulations, understanding the impact of sensed and forecast data uncertainty on RTC methods will be essential.

While sensed and forecast data can be a source of uncertainty, the formulation of RTC methods can also introduce uncertainty in their performance. For example, both RBC methods use thresholds to trigger control actions. RBC-DTN uses a time threshold (24 hours) for retaining runoff after storm events. RBC-TSS uses a TSS concentration threshold to either retain or release water from the ponds. Changing these thresholds would change the performance of the RBC methods (e.g., increased detention time can be expected to increase TSS treatment to a certain extent), however the exact impact on the performance of the RBC methods used in this study is uncertain. The RL implementations in this research also include user defined thresholds in their reward functions and the agents' performance can be very sensitive to these values. In addition, the RL agents benefit from system-level information when learning their control policies. In practice, sensor networks are subject to accuracy limitations, communication interruptions, and cyber-cognitive vulnerabilities (i.e., automated control algorithms, like the RL agents trained here, being used in unexpected situations that they may not have been trained or tested for)[61], to name a few sources of uncertainty.

RL is known to suffer from issues including reward gaming, where the agent learns to exploit its environment in unintended ways to gain reward[62]. In the context of stormwater RTC, reward gaming was observed in early attempts at training RL agents related to simulation processes within the SWMM model. For example, flood water in the Hague SWMM model does not pond and reenter the stormwater system as it would in reality, but is simply recorded as flooding and lost from the simulation. One consequence of this model process is that any TSS within flood water is also lost from the system. If rewards are poorly shaped (i.e., TSS much more heavily weighted than flooding), the RL agent can learn policies that induce flooding because the rewards gained by the corresponding TSS reduction outweigh penalties for flooding. This highlights the need for domain specific knowledge when crafting reward functions and careful consideration of simplifications within simulated environments.

Beyond water level, flooding, and water quality, more direct monetary costs could be included in the RL reward functions. Some costs of RTC are long term (e.g., the purchase and installation of sensors and valves, as well as their maintenance), and may not be useful when learning real-time control policies. A small cost could be incurred for every valve adjustment, which could be considered in optimization. However, a city may not want to limit valve movements based on a small cost if it also limits system efficiency. As with trade-offs between flooding and TSS capture, the balance between limiting and allowing valve movements could be difficult to find. Another potential cost that could be included is that of dredging retention ponds to remove accumulated sediment and maintain appropriate capture volumes. Such maintenance may have to become more frequent with RTC methods that capture sediment, but may still be too long of a time scale when developing sub-hourly control policies. Additionally, improved water quality and flood mitigation could offset costs associated with RTC (see, for example, [14,47]).

4.2 Trade-offs of Local-scale RBC

Both RBC methods used in this research performed RTC at the local-scale (i.e., operating each pond individually) and reduced TSS loads, but at the expense of increased system-level flooding. RBC-DTN showed similar TSS reductions for Pond 1 (49%) as previous studies in other locations (approximately 40% reported by Marchese et al. [12]. However, as water quality sensor technology becomes less expensive and more robust [63–66], control based on water quality observations, such as the RBC-TSS implemented here, may provide a more adaptive solution. RBC-TSS reduced TSS by 96% for Pond 1 compared to the passive system, similar to the value found by Sharior et al. [17] for a different site. The RBC methods did not perform as well for Pond 2 in this study due to the configuration of the upstream pipes. Specifically, when water reached 1.75m (which is less than the maximum depth), the contingency rules to prevent upstream flooding would open valve 2. Without this rule, the RBC methods greatly increased upstream flooding, but it also releases stormwater with high concentrations of TSS during large storm events.

The results of RBC demonstrate that fixed rules, like those used in RBC-DTN, may not provide the most efficient treatment because pollutants are highly variable between sites and storms [19]. One solution could be the combination of the two RBC methods used here (e.g., the predictive drawdown capability of RBC-DTN coupled with adaptive detention time based on observed water quality as in RBC-TSS), but this is still limited as a local-control scheme. While adapting rules based on water quality may be fairly straight-forward for a single pollutant at a single site, controlling a stormwater system for multiple pollutants with different treatment processes (e.g., nitrogen species) will require system-level control[18]. As an example, consider two ponds in series; the upstream pond is controlled to optimize TSS removal, while the downstream pond is controlled to maintain anaerobic conditions for denitrification. If the upstream pond retains water to allow settling of sediment, it might deprive the downstream pond of inflow needed to maintain anaerobic conditions unless these ponds are operated as a system.

4.3 Groundwater Exchange Limitations and Impact

Due to the specific configuration of the studied SWMM model, groundwater exchange was calculated externally from the SWMM model and added (or subtracted in the case of infiltration) to the ponds' inflow at each control time step. While this process is based on in-situ soil properties for Pond 1 in Norfolk's Hague region, the Dupuit equation (which is intended for systems at a steady state) may not provide the most accurate representation of groundwater exchange. Under real-time control, ponds can be rapidly drained and refilled before and during a storm event. The Boussinesq equation for transient unconfined aquifer flow would provide a more realistic representation and is commonly implemented as a simpler alternative to Richards equation (see, for example,[67]). Coupling such a model with the SWMM model used here would allow for more precision, but as an initial demonstration of groundwater impact on ponds controlled in real-time, the Dupuit equation was quick to implement and run.

In the simulated RTC scenarios set up in this research, groundwater exchange with controlled ponds decreased flooding through infiltration; TSS loads were also reduced because less water was exiting the ponds through the valves. It should be noted that neither of the RBC methods were recalibrated to account for groundwater exchange nor were the RL agents retrained with groundwater exchange being simulated. Retraining with groundwater exchange simulated and including groundwater observations or forecasts as part of the RL agents' state may allow the agents to learn more effective policies. However, with the limited impact of groundwater exchange in this specific simulation, it was not necessary; the RL agents' learned policies and the local scale RBC methods were robust toward altered pond behavior when groundwater exchange was simulated.

5 CONCLUSIONS

In this research, real-time control (RTC) methods are applied to a coastal stormwater infrastructure system and evaluated on their ability to mitigate flooding and improve water quality by capturing TSS in controlled retention ponds. The RTC methods used include local control with rules (RBC) and system-level control with deep reinforcement learning (RL). The impact of groundwater exchange on the performance of the controlled ponds was evaluated as a condition that may be important in coastal areas. This research contributes to the growing field of stormwater RTC by being the first to evaluate the ability of RL to learn system-level control policies considering both water quantity and water quality goals, as well as being the first to consider the impact of groundwater on the performance on controlled ponds in a coastal city.

Two methods of RBC were used (i) RBC-DTN, which is based on industry standard stormwater RTC and predictively manages ponds to prevent flooding while retaining runoff for a fixed detention time to improve water quality and (ii) RBC-TSS, which uses observations of water quality to inform valve operation in order to improve TSS capture. Both RBC methods are transparent and provide water quality benefits compared to the passive system. RBC-TSS provided more adaptive operation and demonstrates the potential for water quality observations to be incorporated with RTC as sensor technology improves. However, the local operation of both RBC methods caused increased total system flooding.

Three RL agents were trained and tested for their ability to learn effective system-level control policies. The goal of the first agent (RL-FD) was to mitigate flooding and maintain target pond depths; it reduced flooding compared to the passive system, but did not consider water quality in its control policy. The second and third RL agents (RL-FDTSS and RL-FD+FDTSS) attempted to learn policies for more objectives: mitigate flooding, maintain target pond depths, and reduce TSS load at the controlled valves. RL-FDTSS learned a policy from scratch, while RL-FD+FDTSS was pretrained by using the neural network weights and memory from RL-FD, but was trained to consider water quality as well using the reward function from RL-FDTSS. Both RL-FDTSS and RL-FD+FDTSS provided water quality benefits but increased flooding compared to the passive system. RL-FDTSS decreased TSS loads by an average of 88%, but increased system-wide flooding by 17%. RL-FD+FDTSS's pretraining was effective at reducing training time and allowed it to learn a policy that reduced TSS by an average of 52%, with only a 5% increase in total flood volume, compared to the passive system.

Given the growing adoption of rule-based stormwater RTC and the ability of RL to learn system-level control policies, future work could investigate control of more complex stormwater systems and integrations of RL and RBC. More complex stormwater systems could include retention ponds in series, pollutants that are treated through chemical and biological processes (e.g., nitrogen)/multiple pollutants, and different controllable assets such as pumps. Integration of RL and RBC could include using RL to better parameterize variables within an existing control rule (see Likmeta, et al.,[68] for an example in autonomous vehicles), as well as adding or removing rules from a set of rules. These avenues for future research could allow stormwater RTC providers to increase the complexity of controlled networks, improving flood mitigation and water quality, while maintaining the operational transparency needed for critical stormwater infrastructure systems.

DATA, MODEL, AND CODE AVAILABILITY

The data, models, and code used in this study are available on GitHub at https://github.com/UVAdMIST/swmm_wq_rl.

CONFLICTS OF INTEREST

There are no conflicts of interest to declare.

ACKNOWLEDGMENTS

This work was funded as part of two National Science Foundation grants: Award #1735587 (CRISP-Critical, Resilient Interdependent Infrastructure Systems and Processes) and Award #1737432 (SCC-IRG Track 1: Overcoming Social and Technical Barriers for the Broad Adoption of Smart Stormwater Systems). We acknowledge HRSD for continued access to their high quality data and the City of Norfolk for information regarding their stormwater retention ponds.

A ADDITIONAL FIGURES

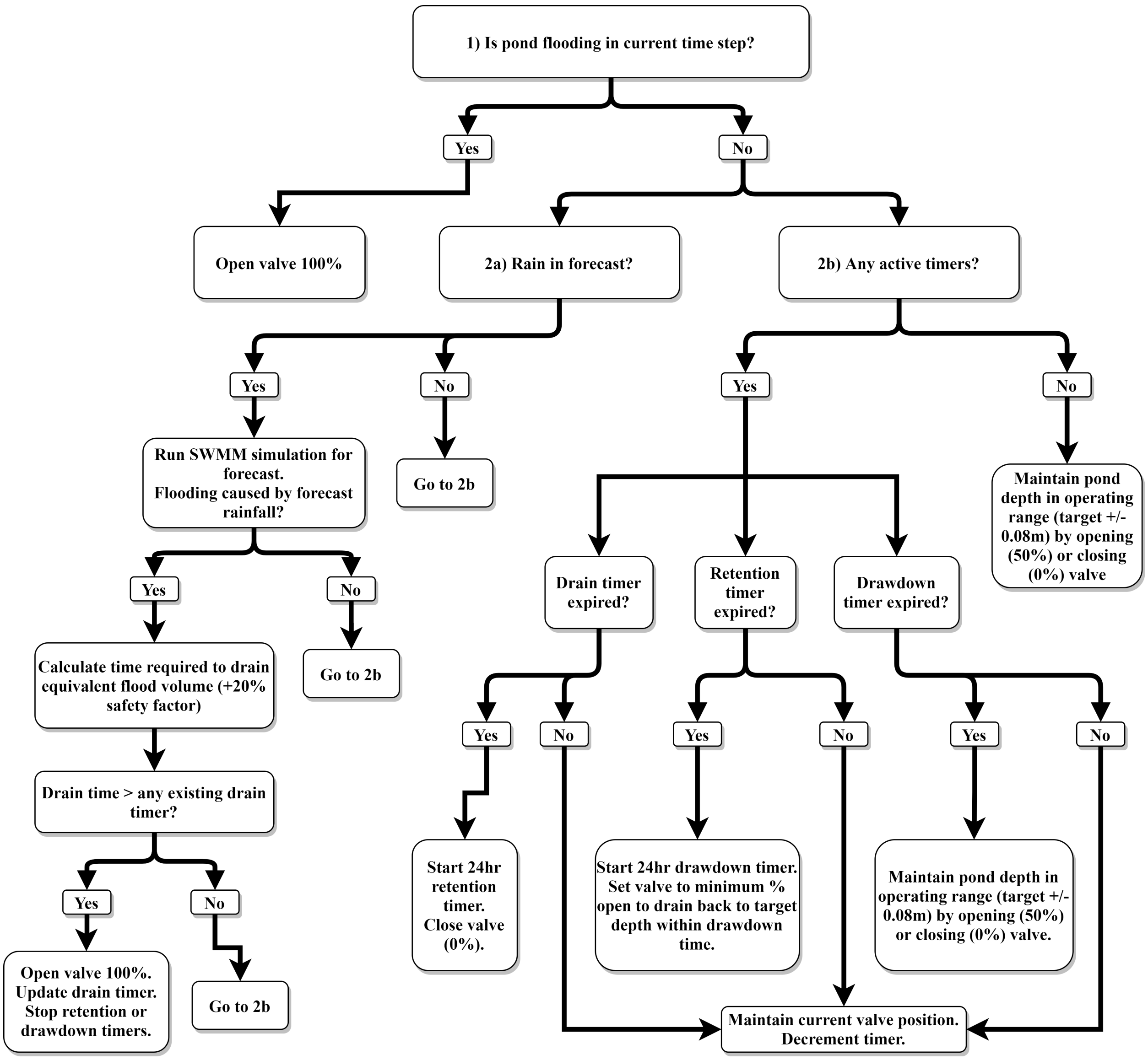

Figure 1. RBC-DTN decision tree (adapted from[10]).

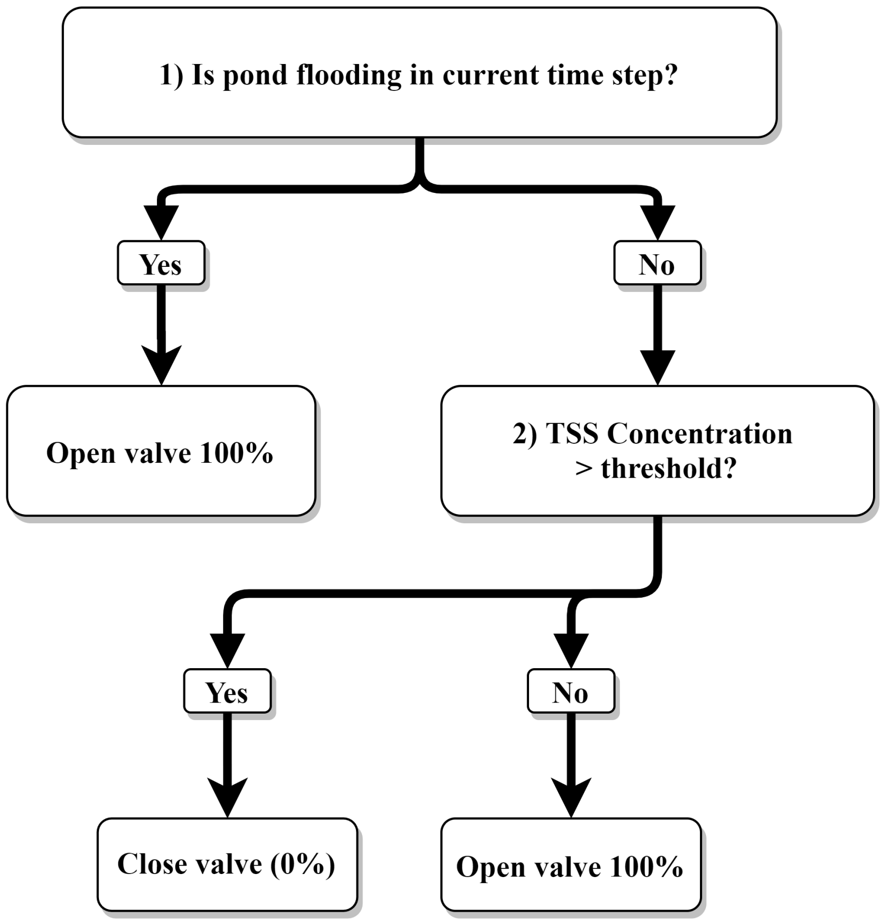

Figure 2. RBC-TSS decision tree.

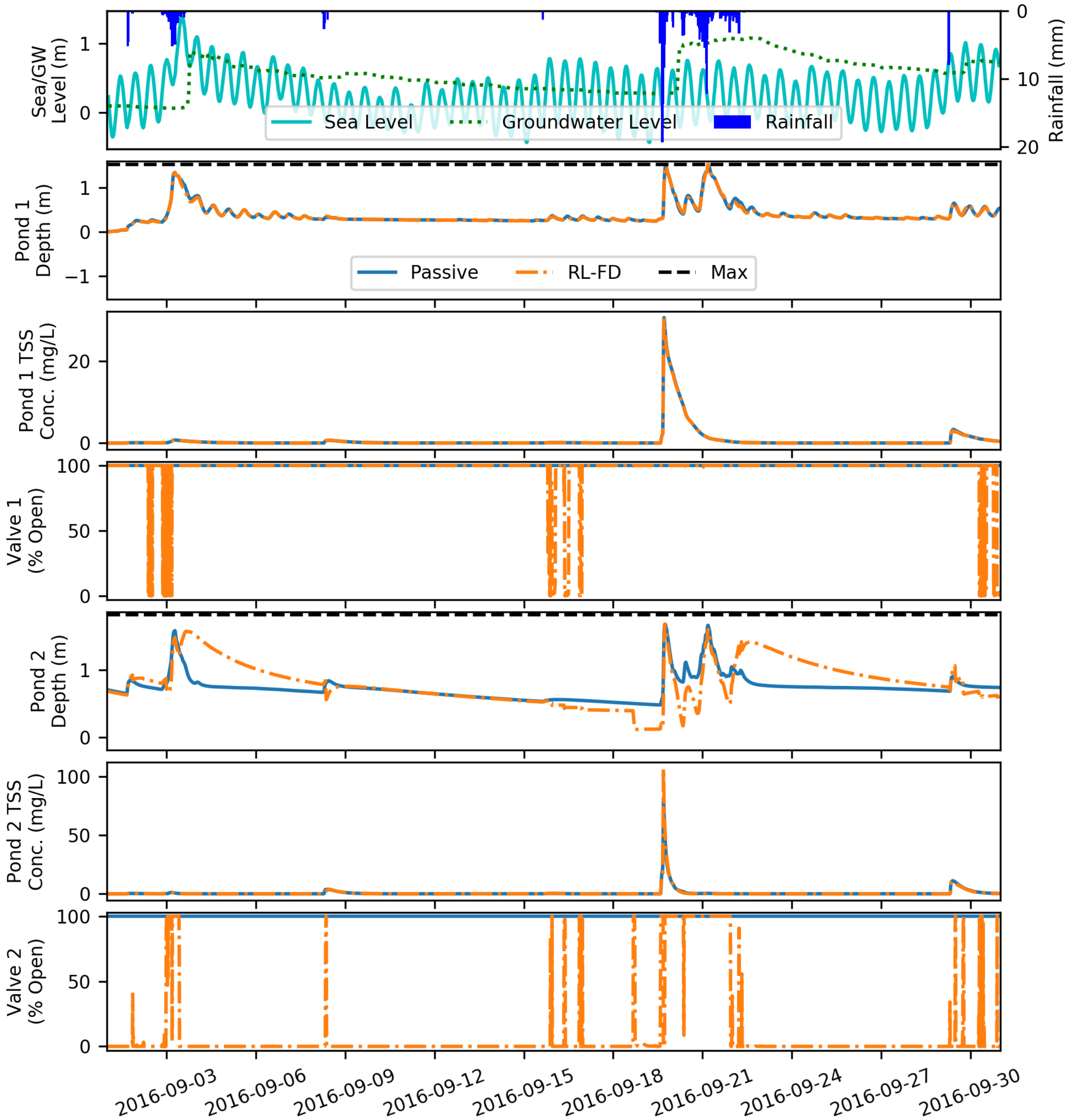

Figure 3. Comparison of RL-FD and passive system operation for September, 2016 with groundwater exchange at the controlled ponds.

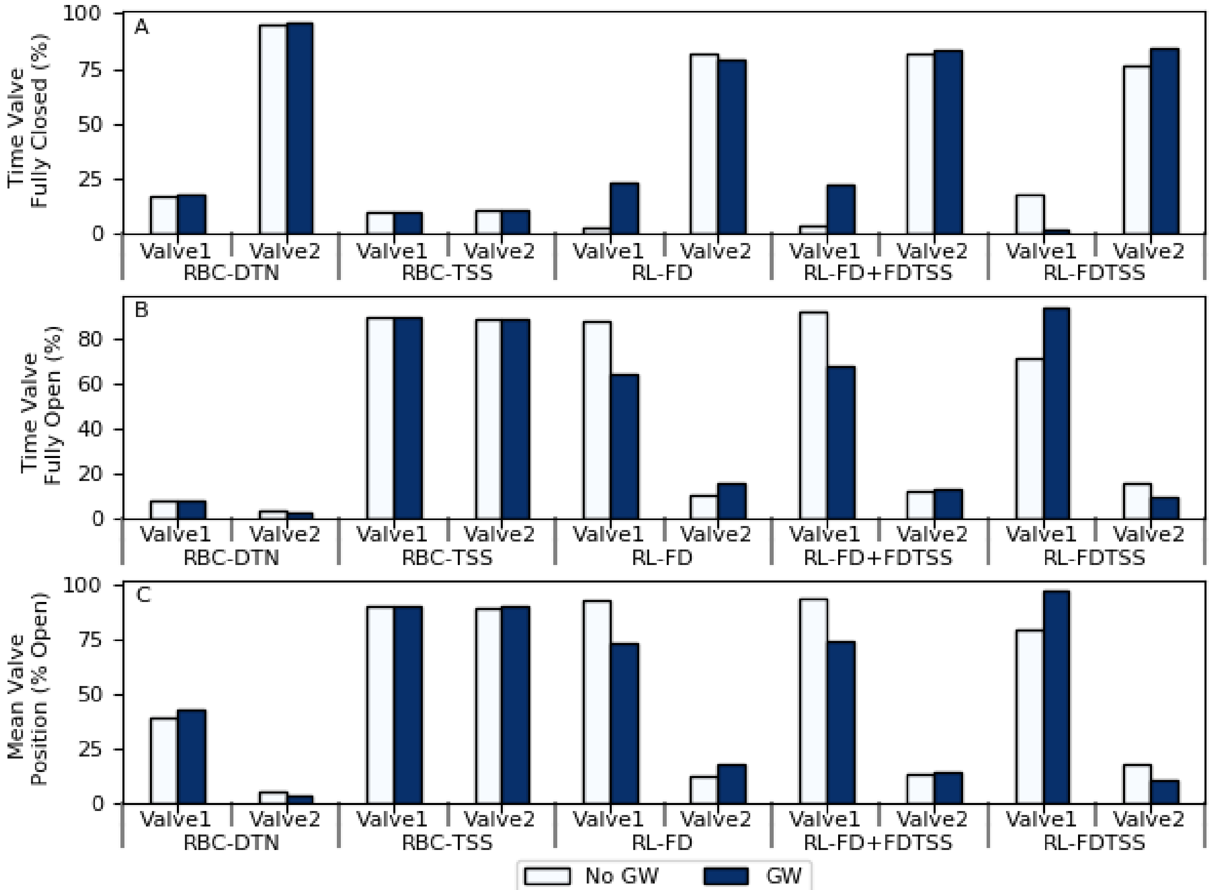

Figure 4. Comparison of control policies (% of time a valve is fully closed (A), fully open (B), and the mean valve position (C)) for simulations with and without groundwater (GW) exchange for September, 2016.

REFERENCES

1 Sweet WV, Park J. From the extreme to the mean: Acceleration and tipping points of coastal inundation from sea level rise. Earth's Future. 2014 12;2(12):579-600. Available from: http://doi.wiley.com/10.1002/2014EF000272.

2 Moftakhari HR, AghaKouchak A, Sanders BF, Feldman DL, Sweet W, Matthew RA, et al. Increased nuisance flooding along the coasts of the United States due to sea level rise: Past and future. Geophysical Research Letters. 2015 11;42(22):9846-52. Available from: http://doi.wiley.com/10.1002/2015GL066072.

3 Moftakhari HR, AghaKouchak A, Sanders BF, Matthew RA. Cumulative hazard: The case of nuisance flooding. Earth's Future. 2017 2;5(2):214-23. Available from: https://onlinelibrary.wiley.com/doi/abs/10.1002/2016EF000494.

4 Alamdari N, Sample DJ, Ross AC, Easton ZM. Evaluating the Impact of Climate Change on Water Quality and Quantity in an Urban Watershed Using an Ensemble Approach. Estuaries and Coasts. 2020 1;43(1):56-72. Available from: https://doi.org/10.1007/s12237-019-00649-4.

5 Kerkez B, Gruden C, Lewis M, Montestruque L, Quigley M, Wong B, et al. Smarter Stormwater Systems. Environmental Science and Technology. 2016;50:72677273. Available from: https://pubs.acs.org/doi/10.1021/acs.est.5b05870.

6 Troutman SC, Love NG, Kerkez B. Balancing water quality and flows in combined sewer systems using real-time control. Environmental Science: Water Research and Technology. 2020 5;6(5):1357-69. Available from: https://pubs.rsc.org/en/content/articlehtml/2020/ew/c9ew00882a.

7 Kroll S, Weemaes M, Van Impe J, Willems P. A Methodology for the Design of RTC Strategies for Combined Sewer Networks. Water. 2018 11;10(11):1675. Available from: http://www.mdpi.com/2073-4441/10/11/1675.

8 Montestruque L, Lemmon MD. Globally Coordinated Distributed Storm Water Management System. In: Proceedings of the 1st ACM International Workshop on Cyber-Physical Systems for Smart Water Networks. New York, NY, USA: ACM; 2015. p. 1-6. Available from: https://dl.acm.org/doi/10.1145/2738935.2738948.

9 Sadler JM, Goodall JL, Behl M, Bowes BD, Morsy MM. Exploring real-time control of

stormwater systems for mitigating flood risk due to sea level rise. Journal of Hydrology. 2020 4;583(124571):124571. Available from: https://linkinghub.elsevier.com/retrieve/pii/S0022169420300317.

10 Bowes BD, Tavakoli A, Wang C, Heydarian A, Behl M, Beling PA, et al. Flood mitigation in coastal urban catchments using real-time stormwater infrastructure control and reinforcement learning. Journal of Hydroinformatics. 2021 5;23(3):529-47. Available from: https://iwaponline.com/jh/article/23/3/529/77759/Flood-mitigation-in-coastal-urban-catchments-using.

11 Wong BP, Kerkez B. Real-Time Control of Urban Headwater Catchments Through Linear Feedback: Performance, Analysis, and Site Selection. Water Resources Research. 2018;54(10):7309-30. Available from: https://onlinelibrary.wiley.com/doi/abs/10.1029/2018WR022657.

12 Marchese D, Johnson J, Akers N, Huffman M, Hlas V. Quantitative Comparison of Active and Passive Stormwater Infrastructure: Case Study in Beckley, West Virginia. Proceedings of the Water Environment Federation. 2018 1;2018(9):4298-311. Available from: https://accesswater.org/publications/-300096/quantitative-comparison-of-active-and-passive-stormwater-infrastructure--case-study-in-beckley--west-virginia.

13 Shishegar S, Duchesne S, Pelletier G. An integrated optimization and rule-based approach for predictive real time control of urban stormwater management systems. Journal of Hydrology. 2019;577:124000.

14 OptiRTC, Geosyntec Consultants Inc . Water Quality Summary Report National Fish and Wildlife Foundation Smart, Integrated Stormwater Management Systems Anacostia River Watershed Water Quality Study; 2017. Available from: www.optirtc.com.

15 Muschalla D, Vallet B, Anctil F, Lessard P, Pelletier G, Vanrolleghem PA. Ecohydraulic-driven real-time control of stormwater basins. Journal of Hydrology. 2014 4;511:82-91.

16 Gaborit E, Muschalla D, Vallet B, Vanrolleghem PA, Anctil F. Improving the performance of stormwater detention basins by real-time control using rainfall forecasts. Urban Water Journal. 2013 8;10(4):230-46.

17 Sharior S, McDonald W, Parolari AJ. Improved reliability of stormwater detention basin performance through water quality data-informed real-time control. Journal of Hydrology. 2019;573:422-31. Available from: https://www.sciencedirect.com/science/article/pii/S0022169419302598.

18 Mullapudi A, Wong BP, Kerkez B. Emerging investigators series: building a theory for smart stormwater systems. Environmental Science: Water Research & Technology. 2017 1;3(1):66-77. Available from: http://xlink.rsc.org/?DOI=C6EW00211K.

19 Wong BP, Kerkez B. Adaptive measurements of urban runoff quality. Water Resources Research. 2016 11;52(11):8986-9000. Available from: http://doi.wiley.com/10.1002/2015WR018013.

20 Chen Y, Han D. Water quality monitoring in smart city: A pilot project. Automation in Construction. 2018 5;89:307-16.

21 Sutton RS, Barto AG. Reinforcement Learning: An Introduction. 2nd ed. Cambridge, Massachusetts: The MIT Press; 2018.

22 Lee JH, Labadie JW. Stochastic optimization of multireservoir systems via reinforcement learning. Water Resources Research. 2007;43(11). Available from: http://doi.wiley.com/10.1029/2006WR005627.

23 Castelletti A, Yajima H, Giuliani M, Soncini-Sessa R, Weber E. Planning the Optimal Operation of a Multioutlet Water Reservoir with Water Quality and Quantity Targets. Journal of Water Resources Planning and Management. 2014;140(4):496-510. Available from: https://ascelibrary.org/doi/pdf/10.1061/%28ASCE%29WR.1943-5452.0000348.

24 Castelletti A, Pianosi F, Restelli M. A multiobjective reinforcement learning approach to water resources systems operation: Pareto frontier approximation in a single run. Water Resources Research. 2013;49:3476-86. Available from: https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1002/wrcr.20295.

25 Pianosi F, Castelletti A, Restelli M. Tree-based fitted Q-iteration for multi-objective Markov decision processes in water resource management. Journal of Hydroinformatics. 2013;15(2):258-70. Available from: https://iwaponline.com/jh/article-pdf/15/2/258/386917/258.pdf.

26 Delipetrev B, Jonoski A, Solomatine DP. A novel nested stochastic dynamic programming (nSDP) and nested reinforcement learning (nRL) algorithm for multipurpose reservoir optimization. Journal of Hydroinformatics. 2017;19(1):47-61. Available from: https://iwaponline.com/jh/article-pdf/19/1/47/390803/jh0190047.pdf.

27 Mnih V, Kavukcuoglu K, Silver D, Rusu AA, Veness J, Bellemare MG, et al. Human-level control through deep reinforcement learning. Nature. 2015;518:529-43. Available from: https://www.nature.com/articles/nature14236.pdf.

28 Lillicrap TP, Hunt JJ, Pritzel A, Heess N, Erez T, Tassa Y, et al. Continuous control with deep reinforcement learning. International Conference on Learning Representations. 2015 9:14. Available from: https://goo.gl/J4PIAzhttp://arxiv.org/abs/1509.02971.

29 Mullapudi A, Lewis MJ, Gruden CL, Kerkez B. Deep reinforcement learning for the real time control of stormwater systems. Advances in Water Resources. 2020 6;140:103600. Available from: https://linkinghub.elsevier.com/retrieve/pii/S0309170820302499.

30 Wang C, Bowes B, Tavakoli A, Adams S, Goodall J, Beling P. Smart Stormwater Control Systems: A Reinforcement Learning Approach. In: Hughes AL, McNeill F, Zobel C, editors. ISCRAM 2020 Conference Proceedings - 17th International Conference on Information Systems for Crisis Response and Management. Blacksburg, VA; 2020. p. 2-13.

31 Saliba SM, Bowes BD, Adams S, Beling PA, Goodall JL. Deep Reinforcement Learning with Uncertain Data for Real-Time Stormwater System Control and Flood Mitigation. Water. 2020 11;12(11):3222. Available from: https://www.mdpi.com/2073-4441/12/11/3222.

32 Eggleston J, Pope J. Land Subsidence and Relative Sea-Level Rise in the Southern Chesapeake Bay Region. Reston, Virginia: U.S. Geological Survey; 2013. Available from: https://pubs.usgs.gov/circ/1392/pdf/circ1392.pdf.

33 Bowes BD, Sadler JM, Morsy MM, Behl M, Goodall JL. Forecasting Groundwater Table in a Flood Prone Coastal City with Long Short-term Memory and Recurrent Neural Networks. Water. 2019;11(5):1098. Available from: https://www.mdpi.com/2073-4441/11/5/1098.

34 Sadler JM, Goodall JL, Morsy MM, Spencer K. Modeling urban coastal flood severity from crowd-sourced flood reports using Poisson regression and Random Forest. Journal of Hydrology. 2018 4;559:43-55. Available from: http://linkinghub.elsevier.com/retrieve/pii/S0022169418300519.

35 Chesapeake Bay Foundation. State of the Bay. Chesapeake Bay Foundation; 2018. Available from: https://www.cbf.org/document-library/cbf-reports/2018-state-of-the-bay-report.pdf.

36 Murphy RR, Kemp WM, Ball WP. Long-Term Trends in Chesapeake Bay Seasonal Hypoxia, Stratification, and Nutrient Loading. Estuaries and Coasts. 2011 11;34(6):1293-309. Available from: https://link.springer.com/article/10.1007/s12237-011-9413-7.

37 CHESAPEAKE BAY TMDL ACTION PLAN VSMP MS4 Permit No. VA0088650. Norfolk: City of Norfolk; 2018. Available from: https://www.norfolk.gov/DocumentCenter/View/38025/

Final-Report---Chesapeake-Bay-TMDL-Action-Plan---06_28_2018_FINAL?bidId=.

38 Virginia Geographic Information Network. Virginia Land Cover Dataset; 2016. Available from: https://vgin.maps.arcgis.com/home/item.html?id=d3d51bb5431a4d26a313f586c7c2c848.

39 Davtalab R, Mirchi A, Harris RJ, Troilo MX, Madani K. Sea Level Rise Effect on Groundwater Rise and Stormwater Retention Pond Reliability. Water. 2020 4;12(4):1129. Available from: https://www.mdpi.com/2073-4441/12/4/1129.

40 McDonnell B, Ratliff K, Tryby M, Wu J, Mullapudi A. PySWMM: The Python Interface to Stormwater Management Model (SWMM). Journal of Open Source Software. 2020 8;5(52):2292. Available from: https://joss.theoj.org/papers/10.21105/joss.02292.

41 Pells SE, N Pells PJ. Application of Dupuit's Equation in SWMM to Simulate Baseflow. Journal of Hydrologic Engineering. 2016 1;21(1):06015009. Available from: https://ascelibrary.org/doi/abs/10.1061/%28ASCE%29HE.1943-5584.0001245.

42 Rossman LA, Huber WC. Storm Water Management Model Reference Manual Volume III – Water Quality. Cincinnati: USEPA; 2016.

43 Guan M, Ahilan S, Yu D, Peng Y, Wright N. Numerical modelling of hydro-morphological processes dominated by fine suspended sediment in a stormwater pond. Journal of Hydrology. 2018 1;556:87-99.

44 Tetra Tech. Stormwater Best Management Practices (BMP) Performance Analysis. USEPA; 2010. Available from: https://www3.epa.gov/region1/npdes/stormwater/assets/pdfs/BMP-Performance-Analysis-Report.pdf.

45 of Norfolk V. CHESAPEAKE BAY TMDL ACTION PLAN VSMP MS4 Permit No. VA0088650. Norfolk; 2018. Available from: https://www.norfolk.gov/DocumentCenter/View/38025/Final-Report---Chesapeake-Bay-TMDL-Action-Plan---06_28_2018_FINAL?bidId=.

46 Virginia Department of Environmental Quality. Chesapeake Bay TMDL Action Plan Guidance; 2015.

47 Wright J, Marchese D. Briefing: Continuous monitoring and adaptive control: the 'smart' storm water management solution. Proceedings of the Institution of Civil Engineers - Smart Infrastructure and Construction. 2017 12;170(4):86-9. Available from: https://www.icevirtuallibrary.com/doi/10.1680/jsmic.17.00017.

48 Read JS, Jia X, Willard J, Appling AP, Zwart JA, Oliver SK, et al. Process-Guided Deep Learning

Predictions of Lake Water Temperature. Water Resources Research. 2019 11;55(11):9173-90. Available from: https://agupubs.onlinelibrary.wiley.com/doi/full/10.1029/2019WR024922.

49 Jia X, Willard J, Karpatne A, Read J, Zwart J, Steinbach M, et al. Physics Guided RNNs for Modeling Dynamical Systems: A Case Study in Simulating Lake Temperature Profiles. In: Proceedings of the 2019 SIAM International Conference on Data Mining. Philadelphia, PA: Society for Industrial and Applied Mathematics; 2019. p. 558-66. Available from: https://epubs.siam.org/doi/10.1137/1.9781611975673.63.

50 Plappert M. keras-rl; 2016. Available from: https://github.com/keras-rl/keras-rl.

51 Brockman G, Cheung V, Pettersson L, Schneider J, Schulman J, Tang J, et al.. OpenAI Gym. arXiv; 2016. Available from: http://arxiv.org/abs/1606.01540.

52 Abadi M, Agarwal A, Barham P, Brevdo E, Chen Z, Citro C, et al. TensorFlow: Large-Scale Machine Learning on Heterogeneous Distributed Systems. arXiv preprint arXiv:160304467. 2016. Available from: https://arxiv.org/pdf/1603.04467.pdf.

53 Biewald L. Experiment Tracking with Weights and Biases; 2020. Available from: https://www.wandb.com/.

54 Su J, Adams SC, Beling PA. Value-Decomposition Multi-Agent Actor-Critics. CoRR. 2020;abs/2007.1. Available from: https://arxiv.org/abs/2007.12306.

55 Baldazo D, Parras J, Zazo S. Decentralized Multi-Agent Deep Reinforcement Learning in Swarms of Drones for Flood Monitoring. In: 27th European Signal Processing Conference (EUSIPCO); 2019. Available from: https://www.eurasip.org/Proceedings/Eusipco/eusipco2019/Proceedings/papers/1570533953.pdf.

56 Parisi S, Pirotta M, Restelli M. Multi-objective reinforcement learning through continuous pareto manifold approximation. Journal of Artificial Intelligence Research. 2016 10;57:187-227. Available from: https://jair.org/index.php/jair/article/view/11026.

57 Quinn JD, Reed PM, Giuliani M, Castelletti A. What Is Controlling Our Control Rules? Opening the Black Box of Multireservoir Operating Policies Using Time-Varying Sensitivity Analysis. Water Resources Research. 2019 7;55(7):5962-84. Available from: https://onlinelibrary.wiley.com/doi/abs/10.1029/2018WR024177.

58 Wiering MA, van Hasselt H. Ensemble algorithms in reinforcement learning. IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics. 2008 8;38(4):930-6. Available from:

https://ieeexplore.ieee.org/document/4509588.

59 Wang Y, Jin H. A Boosting-based Deep Neural Networks Algorithm for Reinforcement Learning. In: 2018 Annual American Control Conference (ACC). IEEE; 2018. p. 1065-71. Available from: https://ieeexplore.ieee.org/document/8431647/.

60 Shishegar S, Duchesne S, Pelletier G, Ghorbani R. A smart predictive framework for system-level stormwater management optimization. Journal of Environmental Management. 2021 1;278:111505.

61 Marchese D, Jin A, Fox-Lent C, Linkov I. Resilience for Smart Water Systems. Journal of Water Resources Planning and Management. 2020;146(1):2519002. Available from: https://doi.org/10.1061/http://ascelibrary.org/doi/10.1061/%28ASCE%29WR.1943-5452.0001130.

62 Amodei D, Olah C, Steinhardt J, Christiano P, Schulman J, Dan M. Concrete Problems in AI Safety. CoRR. 2016. Available from: http://arxiv.org/abs/1606.06565.

63 Miller MP, Tesoriero AJ, Capel PD, Pellerin BA, Hyer KE, Burns DA. Quantifying watershed-scale groundwater loading and in-stream fate of nitrate using high-frequency water quality data. Water Resources Research. 2016 1;52(1):330-47. Available from: https://onlinelibrary.wiley.com/doi/full/10.1002/2015WR017753https://onlinelibrary.wiley.com/doi/abs/10.1002/2015WR017753https://agupubs.onlinelibrary.wiley.com/doi/10.1002/2015WR017753.

64 Hensley RT, Cohen MJ, Korhnak LV. Hydraulic effects on nitrogen removal in a tidal spring-fed river. Water Resources Research. 2015 3;51(3):1443-56. Available from: https://onlinelibrary.wiley.com/doi/full/10.1002/2014WR016178https://onlinelibrary.wiley.com/doi/abs/10.1002/2014WR016178https://agupubs.onlinelibrary.wiley.com/doi/10.1002/2014WR016178.

65 United States Geological Survey. Next Generation Water Observing System (NGWOS);. Available from: https://www.usgs.gov/mission-areas/water-resources/science/next-generation-water-observing-system-ngwos?qt-science_center_objects=0#qt-science_center_objects.

66 United States Geological Survey. WaterQualityWatch – Continuous Real-Time Water Quality of Surface Water in the United;. Available from: https://waterwatch.usgs.gov/wqwatch/faq?faq_id=1.

67 Litwin D, Tucker G, Barnhart K, Harman C. GroundwaterDupuitPercolator: A Landlab component for groundwater flow. Journal of Open Source Software. 2020 2;5(46):1935. Available from: https://joss.theoj.org/papers/10.21105/joss.01935.

68 Likmeta A, Metelli AM, Tirinzoni A, Giol R, Restelli M, Romano D. Combining reinforcement learning with rule-based controllers for transparent and general decision-making in autonomous driving. Robotics and Autonomous Systems. 2020;131:103568.