water

MDPI

*Article*

# State Selection and Cost Estimation for Deep Reinforcement Learning-Based Real-Time Control of Urban Drainage System

**Wenchong Tian, Kunlun Xin *, Zhiyu Zhang, Zhenliang Liao and Fei Li**

College of Environmental Science and Engineering, Tongji University, Shanghai 200092, China; wenchong@tongji.edu.cn (W.T.)
* Correspondence: xkl@tongji.edu.cn

**Abstract:** In recent years, a real-time control method based on deep reinforcement learning (DRL) has been developed for urban combined sewer overflow (CSO) and flooding mitigation and is more advantageous than traditional methods in the context of urban drainage systems (UDSs). Since current studies mainly focus on analyzing the feasibility of DRL methods and comparing them with traditional methods, there is still a need to optimize the design and cost of DRL methods. In this study, state selection and cost estimation are employed to analyze the influence of the different states on the performance of DRL methods and provide relevant suggestions for practical applications. A real-world combined UDS is used as an example to develop DRL models with different states. Their control effect and data monitoring costs are then compared. According to the results, the training process for DRL is difficult when using fewer nodes information or water level as the input state. Using both upstream and downstream nodes information as input improves the control effect of DRL. Also, using the information on upstream nodes as the input state is more effective than using downstream nodes; using flow as input is more likely to have a better control effect than using water level, while using both flow and water level cannot significantly further improve the control effect. Because the cost of flow monitoring is higher than water level monitoring, the number of monitoring nodes and the use of flow/water level need to be balanced based on cost-effectiveness.

**Keywords:** urban drainage system; flooding and overflow mitigation; deep reinforcement learning; cost estimation

check for updates

**Citation:** Tian, W.; Xin, K.; Zhang, Z.; Liao, Z.; Li, F. State Selection and Cost Estimation for Deep Reinforcement Learning-Based Real-Time Control of Urban Drainage System. *Water* **2023**, *15*, 1528. https://doi.org/10.3390/w15081528

Academic Editor: Stefan Kroll

Received: 10 March 2023
Revised: 6 April 2023
Accepted: 8 April 2023
Published: 13 April 2023

## 1. Introduction

Combined sewer overflow (CSO) and flooding have become an inevitable problem in many urban drainage systems (UDSs) around the world due to climate change and urbanization [1–6]. For example, the heavy rainfall in Zhengzhou in 2021 caused a serious loss of life and property, again highlighting the importance of urban water management [7,8]. Many studies show that real-time control (RTC) of UDS is an effective method to mitigate CSO and flooding in the context of smart cities [6,8,9]. It makes full use of the storage capacity of the pipe network and structures through controlling actuators such as pumps, gates, and valves to change the hydraulic conditions of the pipe network for CSO and flooding mitigation.

Traditional real-time control methods include: heuristic control (HC) [10–13], model predictive control (MPC) [13–15], and supervised learning control [16]. HC controls UDS through preset rules which cannot be adjusted in operation and has weak adaptability [10,17]. MPC combines rainfall prediction, a UDS model, and an optimization algorithm to optimize the control strategy through receding horizons at each control time point, but is limited by prediction accuracy and computing speed [17]. Supervised learning control trains an AI controller through a large amount of UDS operation data [16], thus its effectiveness depends on the quality and quantity of collected data and might lack feasibility in the context of UDS.

Recent studies show that deep reinforcement learning (DRL) can be used to establish a more advanced real-time control method for CSO and flooding mitigation and solve the problems of traditional RTC methods [18–22]. DRL trains an agent to flexibly control the pump and gate of UDS based on the real-time state, thereby adjusting the hydraulic conditions and water volume of the pipe network for CSO and flood mitigation, and therefore has better adaptability than HC. Also, a trained DRL agent directly provides the control strategy, and no longer needs to solve an optimization problem in real-time operation, leading to a faster computational speed compared with MPC. In addition, the training process of DRL does not require a large amount of monitoring data, but rather to obtain learning data through simulation and trial-and-error. Therefore, it requires only a well-established UDS model and sufficient rainfall data, and eliminates the need for a large amount of monitoring data compared to supervised learning.

Although the optimization of state selection has already existed in other fields [23,24], the optimizations of state selection and cost are still missing in the existing DRL in the context of UDSs. Different states will provide different information to DRL, resulting in different control effects. Therefore, it is beneficial to improve the operation of DRL by fully considering the state selection to help it obtain as much useful information as possible. Meanwhile, adding an input into a DRL system requires the establishment of a corresponding data monitoring system and a data transmission network system. Considering the price difference of data monitoring systems, the cost of a DRL using water level as input can be significantly different from a DRL using flow as input. Therefore, how to achieve a balance between a DRL method's control effects and its construction costs is an important issue that must be considered in the practical application. The optimal combination of the input features and their rationality were analyzed by cross-validation and mechanism analysis in related deep learning application studies [24,25]. In addition, there were also studies that provided preliminary accounting for construction costs of real-time control systems and data monitoring systems in UDS [10,26]. However, none of these studies have been conducted for DRL in the context of UDS.

The main objective of this paper is thus to understand the influence of different state selections on the performance of DRL in CSO and flood mitigation by state design based on monitoring locations (different node locations in UDS) and monitoring information (water level or flow). The impact of state selection on the construction price is also analyzed through cost estimation. A real-world UDS is used as an example for verification. Its DRL design and implementation suggestion is proposed by taking into account both control effects and costs.

## 2. Materials and Methods

Simulation, state selection, and cost estimation are applied to analyze the influence of different states on both the control effect and construction cost of DRL. Different state scenarios based on the location of pipeline network nodes and information type (water level, flow, or both) are designed according to the characteristics of UDSs and are used to train different DRL agents. The cost of each input scenario is calculated using the collected price information from different service providers in China. The following section provides details on the applied methods.

### *2.1. Case Study and Rainfall Data*

The case study is a real-world combined UDS located in eastern China, which has 139 nodes, 140 pipelines, and three pump stations. In this study, we focus on the area with a high level of flooding and CSO, thus only two pump stations are considered. Its storm water management model (SWMM) is shown in Figure 1. The version of SWMM is 5.1 [27]. In dry periods, the sewage pumps transfer water from the storage tanks downstream, while on rainy days, the storm pumps discharge excess water into a nearby river as the CSO. During heavy rainfalls, the maximum capacity of the pumping stations might exceed the maximum capacity of the pipeline network, causing both flooding and CSO in this area.

Therefore, its control system needs to smartly balance the water volume both upstream and downstream to minimize the total CSO and flooding. These pump stations have a rule-based heuristic control (HC) system that operates the pumps to start/stop working based on a sequence of water level threshold values [21]. The threshold values are shown in Table 1. The system is named HC in this study. More details of this case can be found in our previous studies [6,19–21,28].

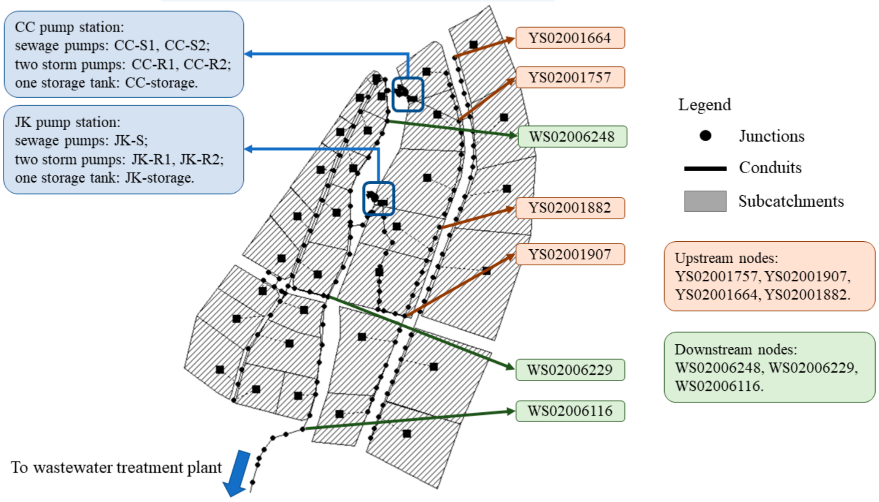

**Figure 1.** The SWMM model of the case study. The CC pump station has two sewage pumps (CC-S1, 2), two storm pumps (CC-R1, 2), and one storage tank (CC-storage). The JK pump station has one sewage pump (JK-S), two storm pumps (JK-R1, 2), and one storage tank JK-storage. The nodes in the upstream and downstream are used in the state design in Section 2.2.

**Table 1.** Rule-based heuristic control (HC) system in the case study.

| Pump Name | Type | Startup Depth of HC (m) | Shutoff Depth of HC (m) |
|---|---|---|---|
| CC-S1 | Sewage pump | 0.8 | 0.5 |
| CC-S2 | Sewage pump | 1.0 | 0.5 |
| CC-R1 | Stormwater pump | 1.2 | 0.5 |
| CC-R2 | Stormwater pump | 1.4 | 0.5 |
| JK-R1 | Stormwater pump | 4.2 | 1.2 |
| JK-R2 | Stormwater pump | 4.3 | 1.2 |
| JK-S | Sewage pump | 4.0 | 1.0 |

The rainfall data used in this study for DRL training and testing are designed rainfalls based on Chicago histograms (Equation (1)), where $i(t_b)$ is the storm intensity before the rainfall peak; $i(t_a)$ is the storm intensity after the rainfall peak; $t_a$ and $t_b$ are corresponding rainfall duration time; $A$ is the rainfall intensity with a recurrence period of one year; $C$ is an empirical constant; $P$ is the rainstorm return period; $K$ is the peak intensity position coefficient (or time-to-peak ratio); and $n$ and $b$ are the parameters related to the region. These parameters are randomly selected within the range provided by historical research on the rainstorm intensity in the study area. The detailed values are provided by previous studies [19,20] and are shown in Table 2. These rainfalls are shown in Figure 2.

$$\begin{aligned} \text{Before the peak}: \; i(t_b) &= \frac{[A(1+Clog(P))]\left[\frac{(1-n)t_b}{K}+b\right]}{\left(\frac{t_b}{K}+b\right)^{1+n}} \\ \text{After the peak}: \; i(t_a) &= \frac{[A(1+Clog(P))]\left[\frac{(1-n)t_a}{1-K}+b\right]}{\left(\frac{t_a}{1-K}+b\right)^{1+n}} \end{aligned} \tag{1}$$

**Table 2.** The range of parameters in Equation (3).

| *A* (mm) | *C* | *P* (Year) | *n* | *b* | *K* |
|---|---|---|---|---|---|
| 21~35 | 0.939~1.20 | 1~5 | 0.86~0.96 | 16~22 | 0.3~0.8 |

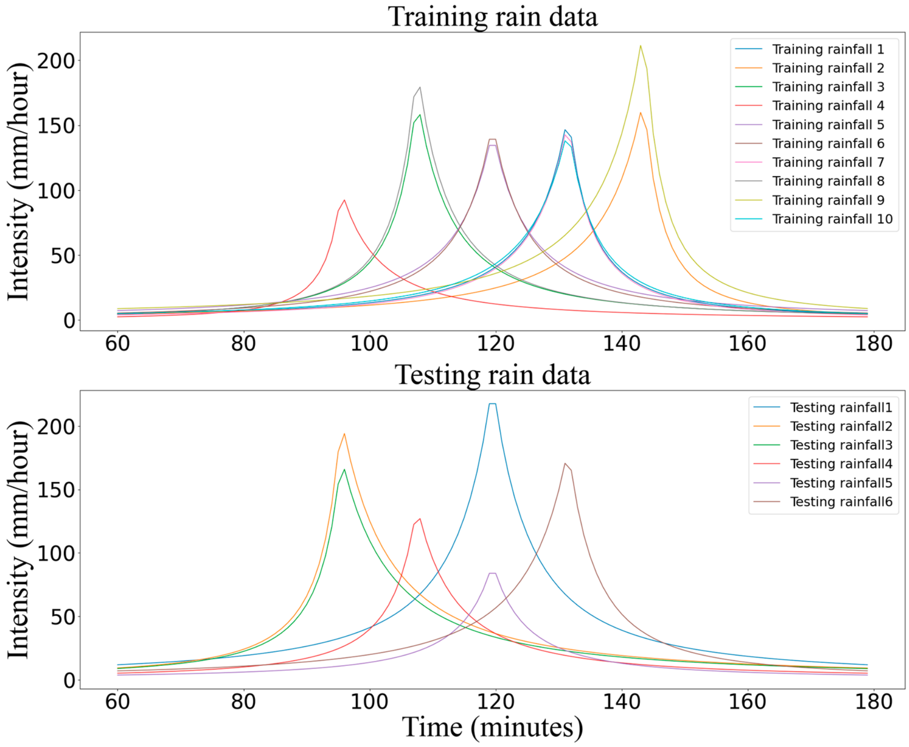

**Figure 2.** The rainfall data used for DRL training and testing.

### *2.2. Deep Reinforcement Learning for Flooding and Overflow Mitigation*

When establishing DRL for a UDS, its model, e.g., SWMM, is applied as an environment controlled by a DRL agent, while its agent is established by a deep neural network to input real-time state and output action. The state is real-time information on UDS, which includes the water level and flow of pipeline network nodes, rainfall intensity, and so on. In this study, the influence of different states on the control effect of DRL is mainly studied. Different state selection scenarios based on the node's location and information type (water level or flow) are designed and used to train different DRLs for comparison. Details of scenarios are provided in Section 2.3. The action refers to the control command in the next time period, which generally includes the startup/shutoff of a valve or a pump, etc. There are two pump stations in this case study and each of them has corresponding stormwater pumps and sewage pumps, thus the action is the control command of all these pumps (startup/shutoff). The system needs a reward to score the control effectiveness of action in real-time according to a control objective, e.g., flooding and CSO mitigation. The reward in this study is calculated using Equation (2), where $CSO_t$ is the total CSO of time $t$; $flooding_t$ is the total flooding of time $t$; and $total\ inflow_t$ is the total inflow of time $t$.

$$r_t = -\frac{CSO_t + flooding_t}{total\ inflow_t} \tag{2}$$

The DRL agent needs to be trained before practical application. Its training process consists of two steps: sampling and updating. The sampling process uses a DRL agent to control a UDS model through simulation and collects state, action, and reward data during control. The updating process selects a DRL update algorithm with respect to the type of DRL agent and uses it to train the DRL agent with the collected data to maximize the weighted total reward, i.e., the q-value function of Equation (3); $s_t$, $a_t$, and $r_t$ are the state, action, and reward of the DRL method; $\gamma$ is the discount factor between 0 and 1, used to guarantee the convergence of $q$ value and govern the temporal context of the reward; $k$ is the number of forward time-steps; $T$ is the total number of time-steps; $\pi$ is the policy function used by the agent to provide action with respect to state, and is a deep neural network in this study. After multiple epochs of sampling and updating, the trained DRL agent is able to find the action that obtains the highest $q$ value in play, and thus it becomes an expert in controlling UDS.

$$q_\pi(a_t, s_t) = \mathbb{E}\left[\sum_{k=0}^{T} \gamma^k r_{t+k+1} \middle| a_t, s_t\right] \tag{3}$$

DRLs can be classified as value-based or policy-based. The value-based DRL uses a deep neural network to predict the $q$ value of all the available actions and selects the action with the highest corresponding $q$ value for control (Mnih et al., 2015). This study is more concerned with the influence of different state selections on DRL control output than on the $q$ value prediction; therefore, proximal policy optimization (PPO) is used. It belongs to the policy-based models [29,30] which use a deep neural network to approximate policy function (input state and output action). Then, this deep neural network is trained by one of the gradient-based algorithms. In this study, Adam [31] is used as the training algorithm. One special feature of the policy-based method is that its gradient cannot be directly obtained by the derivatives of $q$ value function. Instead, an estimator called policy gradient (Equation (4)) [30] is used as the gradient of the deep neural network in the Adam algorithm, where $q$ is the $q$ value, $\pi_\theta(a_t|s_t)$ is the policy function based on a deep neural network with parameter $\theta$. The training process of PPO and its implementation of PPO in a simulated environment is described in Figure 3.

$$g = \mathbb{E}_t[\nabla_\theta log \pi_\theta(a_t|s_t) q] \tag{4}$$

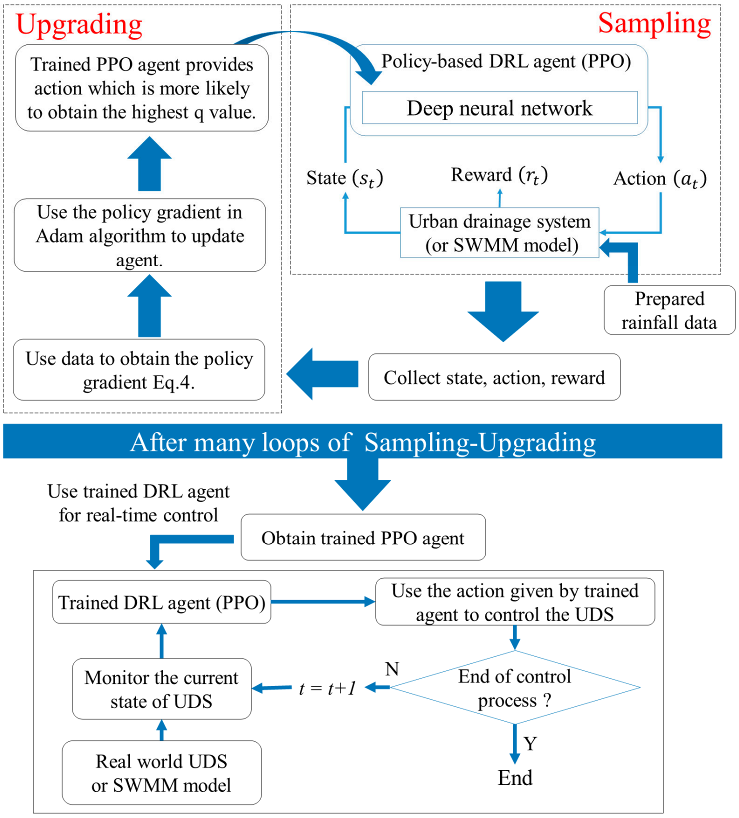
Upgrading
Sampling
Trained PPO agent provides action which is more likely to obtain the highest q value.
Policy-based DRL agent (PPO)
Deep neural network
State ($s_t$)
Reward ($r_t$)
Action ($a_t$)
Urban drainage system (or SWMM model)
Use the policy gradient in Adam algorithm to update agent.
Prepared rainfall data
Use data to obtain the policy gradient Eq.4.
Collect state, action, reward
After many loops of Sampling-Upgrading
Use trained DRL agent for real-time control
Obtain trained PPO agent
Trained DRL agent (PPO)
Use the action given by trained agent to control the UDS
Monitor the current state of UDS
t = t+1
N
End of control process ?
Y
Real world UDS or SWMM model
End

**Figure 3.** The training and implementation of PPO in a simulated environment.

In this study, the architecture of the deep neural network is M-30-30-30-1. M is the size of the input layer and is related to the selection of pipeline nodes and input type. There are 3 hidden layers and each of them has 30 nodes. The output layer has 1 node, which represents the action. Since there are a total of 7 pumps, each of them has two kinds of settings (startup/shutoff), thus there are $2^7$ combinations of all possible actions of pumps, that is, the output of the deep neural network is between 1 and 128 in this study. The trained system is named PPO in this study.

### *2.3. The Design of State Selection Scenarios*

As mentioned above, existing DRL studies lack optimal design for system inputs, so the influence of state selection on the DRL control effect is still unknown. This study investigates the influence of state selection on DRL through state design. Specifically, different state selection scenarios based on different node locations and the type of information (water level or flow) are provided to train different DRL agents for comparison.

The DRL agent needs to provide the action of pumps that can obtain the highest $q$ value in the next time period based on real-time states. Therefore, the data that reflects the real-time situation of the UDS and provides effective information for pump controlling can all be used as a state, and the most related one is the information on the forebay storage tanks of all the pump stations. Meanwhile, pump stations need to balance the water volume both upstream and downstream to avoid the situation in which a large amount of stormwater and sewage water run into a node simultaneously. Therefore, with the pump station as the center, the information on its upstream and downstream nodes can also support the control of the DRL agent. The distance between the node and the pump station can be used as a reference for node selection. Studies on control systems have provided analysis methods of controllability of networks [32,33] and important nodes on a network [34]. However, none of them is for black-box control systems of urban drainage pipeline networks, and thus they need to be adjusted to fulfill the hydrodynamic properties of pipeline systems. Therefore, they are not considered in this study.

Based on the above two points, the state selection scenarios designed in this study are divided into four categories: storage tank nodes only (state-1), storage tank nodes and upstream nodes (state-2), storage tank nodes and downstream nodes (state-3), and storage tank nodes and both upstream and downstream nodes (state-4).

In addition to node location, the flow or water level information provided by each node also affects DRL control performance. Therefore, the use of flow or water level is also considered in the state selection. In the previous study, the flooding value was used as a state of DRLs [19]. This is available in a simulated environment but is not considered in this study due to the difficulty of flooding real-time monitoring in practical applications.

The state selection scenarios based on the integration of node location and information types are shown in Table 3, and the location of each node is shown in Figure 2. Each scenario is used to construct a PPO-based DRL (named PPO) to analyze the effect of state selection on the control effectiveness. Since many studies have confirmed the necessity of rainfall information for real-time control [14,19], real-time rain intensity is used as one of the input states in all the scenarios.

**Table 3.** The scenarios of state selection.

| Scenario Name | Type of Nodes Info. | Selected State, Node Names, and RG (Real-Time Rain Intensity) |
|---|---|---|
| state-1-wl | Water-level | CC-storage, JK-storage, RG. |
| state-1-fl | Flow | |
| state-1-wfl | Both | |
| state-2-1-wl | Water-level | CC-storage, JK-storage, YS02001757, YS02001907, RG. |
| state-2-1-fl | Flow | |
| state-2-1-wfl | Both | |

**Table 3.** *Cont.*

| Scenario Name | Type of Nodes Info. | Selected State, Node Names, and RG (Real-Time Rain Intensity) |
|---|---|---|
| state-2-2-wl<br>state-2-2-fl<br>state-2-2-wfl | Water-level<br>Flow<br>Both | CC-storage, JK-storage, YS02001757, YS02001907, YS02001664, YS02001882, RG. |
| state-3-1-wl<br>state-3-1-fl<br>state-3-1-wfl | Water-level<br>Flow<br>Both | CC-storage, JK-storage, WS02006248, WS02006229, RG. |
| state-3-2-wl<br>state-3-2-fl<br>state-3-2-wfl | Water-level<br>Flow<br>Both | CC-storage, JK-storage, WS02006248, WS02006229, WS02006116, RG. |
| state-4-1-wl<br>state-4-1-fl<br>state-4-1-wfl | Water-level<br>Flow<br>Both | CC-storage, JK-storage, YS02001757, YS02001907, WS02006248, WS02006229, RG. |
| state-4-2-wl<br>state-4-2-fl<br>state-4-2-wfl | Water-level<br>Flow<br>Both | CC-storage, JK-storage, YS02001757, YS02001907, YS02001664, YS02001882, WS02006248, WS02006229, WS02006116, RG. |

### 2.4. *The Unit Cost of DRL Data Monitoring*

The construction of a DRL real-time control system requires an online data monitoring system, control actuators, a data and computing center, and a data transmission network. The control actuators in this study are the pumps in the two pump stations. A data and computing center is used to collect all the online monitoring data and run the DRL agent to provide real-time control actions. Since dispatching facilities and a control center are already available in this case area, the costs of these parts are the same for all the scenarios and are ignored in this study.

The data transmission in the case study mainly relies on wireless base stations for real-time communication. Since the case UDS is located in an urban area with well-developed network facilities, there is no need to establish a dedicated network and the cost of this part depends on the network communication service. The online data monitoring system is used to obtain real-time water flow and water level data. The selection of flow or water level leads to different costs of system construction as the price of their monitoring facilities is significantly different.

Accordingly, the construction cost difference of all the state selection scenarios mainly comes from the cost of the online data monitoring system. Thus, this cost is mainly analyzed in this study to reveal the cost-effectiveness of each state selection. Referring to the existing quotations of online monitoring service providers in five different regions of China (Shanghai, Wuhan, Shenzhen, Chongqing, and Fujian), the unit prices of different water level and flow monitoring, software support, and network communication services are obtained and shown in Table 4, and the final cost can be calculated by multiplying the unit price by the quantity.

**Table 4.** The unit cost of data monitoring and network communication service.

| | Unit Cost (CNY) | Type | Description |
|---|---|---|---|
| Water-level meter | 6500~9000 | Ultrasonic digital water level gauge, photoelectric multipurpose water-level gauge, radar water-level gauge. | This price includes power supply (lithium batteries or solar energy), data collection and transmission equipment, software support, and equipment installation service. |
| Flow meter | 10,000~40,000 | Doppler flow gauge, radar flow gauge | |
| Network communication service | 50~100 per month | Wireless network or cable broadband network | This price is related to the data transmission method and the cost of the network service in different areas. |

It is worth mentioning that the prices here are heavily influenced by the cost of installation labor and network services in different regions of China, as well as the principles of monitoring equipment; therefore, there is a difference among them and they can only be used as a reference.

## 3. Results and Discussion

Different PPOs are established based on different state selection scenarios, and they are trained by 10 training rainfall events (Figure 2) using the method provided in Section 2.1. Then, all the trained PPOs are used to control the UDS in 6 testing rainfall events (Figure 2) through simulation and are compared with the HC in the case study area. All the control processes have 8 h of time length, which includes 1 h before rainfall, 2 h of rainfall event (training and testing rainfalls), and 5 h after rainfall. The control interval is 5 min. After that, the construction costs of all the PPOs' online data monitoring systems are analyzed according to the price data in Section 2.4 as to the cost-effectiveness of different state selection scenarios.

### *3.1. Training Process of DRLs*

The reward sum of all the PPOs during the training process in the 10 rainfalls are shown in Figure 4. Each PPO is named after its input scenario's name in Table 3, e.g., state-1-1 water level means the input of this PPO is the water level of CC-storage, JK-storage. According to the training process, the total reward of different PPOs gradually increases and eventually converges to −50, indicating that the training process can gradually improve the control effect of PPOs in CSO and flooding mitigation. The training speed of PPOs using only water level or fewer nodes information (state-1-1) is slower compared to the others. This shows that training a PPO using only water level information or less node information as input is difficult. Since the final reward sum of all PPOs is close to −50, the use of water level has a limited impact on the final training effect.

### *3.2. Control Effect of DRLs*

All the trained PPOs were used to control the UDS model in six designed rainfalls. The sum of the CSO and flooding volume during these control processes are provided in Table 5. According to the results, all the PPOs achieved better control performance than HC, which indicates the effectiveness of DRL in CSO and flooding mitigation and is consistent with the existing studies. The results of PPO robustness can be found in our previous research [19,20].

**Table 5.** The sum of CSO and flooding volume ($10^3$ $m^3$) of PPO in the 10 test rainfalls.

| | **Rain1** | **Rain2** | **Rain3** | **Rain4** | **Rain5** | **Rain6** |
|---|---|---|---|---|---|---|
| HC | 42.75 | 33.36 | 29.23 | 17.86 | 10.76 | 28.14 |
| state-1-wl | 41.06 | 31.54 | 27.33 | 15.98 | 8.77 | 26.38 |
| state-1-fl | 41.28 | 31.76 | 27.58 | 16.49 | 8.87 | 26.17 |
| state-1-wfl | 41.01 | 31.44 | 27.41 | 16.09 | 8.81 | 26.28 |
| state-2-1-wl | 41.04 | 31.53 | 27.51 | 16.37 | 8.71 | 26.74 |
| state-2-1-fl | 40.89 | 31.47 | 27.50 | 17.12 | 9.85 | 26.87 |
| state-2-1-wfl | 41.01 | 31.55 | 27.62 | 15.95 | 8.78 | 26.21 |
| state-2-2-wl | 40.98 | 31.34 | 27.37 | 15.92 | 8.77 | 26.30 |
| state-2-2-fl | 41.12 | 31.48 | 27.52 | 15.92 | 9.14 | 26.11 |
| state-2-2-wfl | 40.98 | 31.58 | 27.42 | 15.93 | 8.79 | 26.12 |
| state-3-1-wl | 41.14 | 31.66 | 27.64 | 16.52 | 8.79 | 26.55 |
| state-3-1-fl | 40.90 | 31.52 | 27.37 | 15.98 | 8.87 | 26.15 |
| state-3-1-wfl | 40.98 | 31.54 | 27.27 | 15.94 | 8.93 | 26.09 |
| state-3-2-wl | 42.26 | 33.17 | 28.42 | 17.35 | 10.41 | 27.66 |
| state-3-2-fl | 40.98 | 31.50 | 27.29 | 15.95 | 8.79 | 26.32 |
| state-3-2-wfl | 40.97 | 31.75 | 27.42 | 16.34 | 9.09 | 26.16 |
| state-4-1-wl | 41.01 | 31.54 | 27.39 | 16.05 | 9.00 | 26.14 |
| state-4-1-fl | 40.80 | 31.52 | 27.28 | 15.87 | 8.73 | 26.29 |
| state-4-1-wfl | 40.99 | 31.55 | 27.36 | 16.07 | 8.82 | 26.33 |
| state-4-2-wl | 41.06 | 31.55 | 27.37 | 15.95 | 8.74 | 26.26 |
| state-4-2-fl | 40.99 | 31.65 | 27.53 | 15.84 | 8.73 | 26.15 |
| state-4-2-wfl | 41.00 | 31.56 | 27.35 | 15.96 | 8.79 | 26.29 |

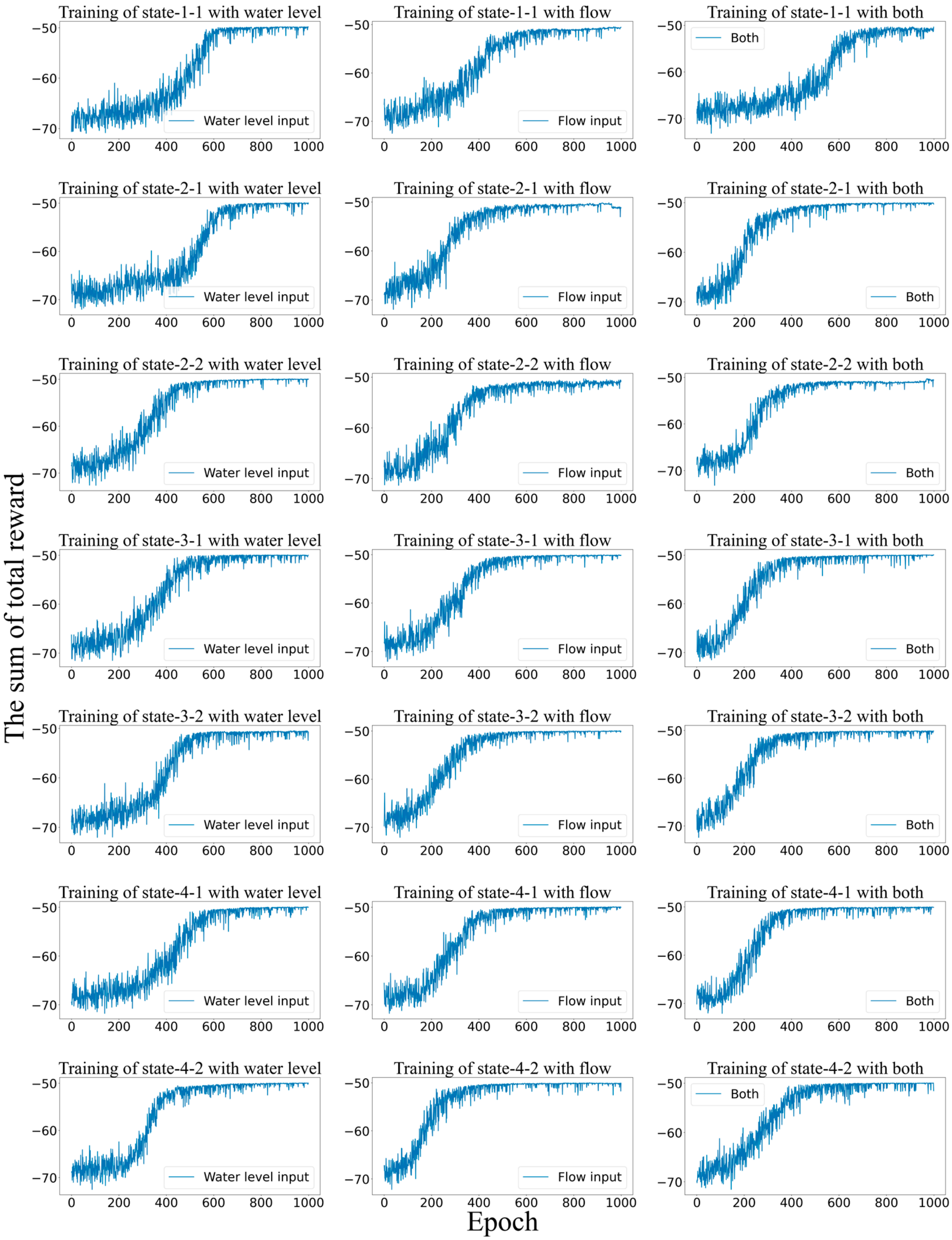

**Figure 4.** Training process. The reward sums of all the epochs in 10 training events.

Meanwhile, the performances of the PPOs were different. It can be found that simply adding monitoring nodes was not necessarily effective in improving the control effect. Also, flow-based PPO was more likely to achieve better performance than water-level-based PPO. To further elaborate on these findings, we summarized and classified the above results and obtained the following Tables 6 and 7 to illustrate the influence of node number, location, and input type on the control effect.

**Table 6.** Node number and location vs. average control effect.

| | **Rain1** | **Rain2** | **Rain3** | **Rain4** | **Rain5** | **Rain6** |
|---|---|---|---|---|---|---|
| state-1 | 41.17 | 31.65 | 27.46 | 16.24 | 8.82 | 26.28 |
| state-2 | 41.01 | 31.47 | 27.49 | 16.23 | 9.01 | 26.42 |
| state-3 | 41.21 | 31.83 | 27.57 | 16.28 | 9.10 | 26.48 |
| state-4 | 40.97 | 31.59 | 27.39 | 16.02 | 8.85 | 26.22 |

**Table 7.** Input type vs. average control effect.

| | **Rain1** | **Rain2** | **Rain3** | **Rain4** | **Rain5** | **Rain6** |
|---|---|---|---|---|---|---|
| Water level | 41.22 | 31.76 | 27.58 | 16.31 | 9.03 | 26.58 |
| Flow | 40.99 | 31.56 | 27.44 | 16.17 | 9.00 | 26.29 |
| Both | 40.99 | 31.57 | 27.41 | 16.04 | 8.86 | 26.21 |

The average values of the PPOs' control effects according to the number of input nodes were calculated and are provided in Table 6, in which it can be seen that the performance of PPO increased with the number of selected upstream nodes. Also, the worst case in each rainfall was state-3, which shows that the use of downstream nodes was not effective. Using upstream nodes as input (state-2) achieved better performance than using downstream nodes (state-3), and it was much better when both upstream and downstream were considered (state-4).

The average values of the PPOs' control effect according to the type of input were also calculated and are shown in Table 7. It can be seen that the use of water level as input was the worst case, while the control effect of the PPO using water flow as input was better than that of the PPO using water level as input. The PPOs using both have a certain level of improvement, but it is not significant in some cases (e.g., Rain1 and Rain2). If we only focus on the control effect, the PPO that considered the most nodes (state-4) and used water flow as input was relatively better than others in this case.

### 3.3. Cost-Effectiveness Analysis

The cost-effectiveness analysis of all the scenarios is provided based on the unit price in Table 4 and the performance in Table 5. In order to evaluate the control effect of PPOs more specifically, their performances are reflected by calculating the Ratio of PPOs' CSO and flooding volume to HC's in all the testing rainfall events (Equation (5), where $CSO_{HC,i}$ and $flooding_{HC,i}$ are the sum of CSO and flooding volumes of HC in $i$th test rainfall, and $CSO_{PPO,i}$ and $flooding_{PPO,i}$ are the sum of CSO and flooding volumes of PPO in $i$th test rainfall).

$$Ratio = \frac{1}{6}\sum_{i=1}^{6} \frac{\left(CSO_{HC,i} + flooding_{HC,i}\right) - \left(CSO_{PPO,i} + flooding_{PPO,i}\right)}{\left(CSO_{HC,i} + flooding_{HC,i}\right)} \tag{5}$$

The result of the cost-effectiveness analysis is shown in Table 8. It can be seen that the rise in the number of monitoring nodes gradually increases the system's spending, especially for the PPO based on flow input because the price of water flow monitoring is much higher than water level monitoring. Meanwhile, when the *Ratio* of the system is further considered, the state-4-1-fl works better than others because it achieves a good control effect with an acceptable cost, which is 60,300~240,600 CNY.

**Table 8.** Cost-effective analysis of different scenarios.

| Scenario Name | Quantity | | | Total Cost (CNY) | Ratio |
|---|---|---|---|---|---|
| | Water-Level Gauge | Flow Gauge | Communication Service (1 Month) | | |
| state-1-wl | 2 | 0 | 2 | 13,100~18,200 | 0.0853 |
| state-1-fl | 0 | 2 | | 20,100~80,200 | 0.0769 |
| state-1-wfl | 2 | 2 | | 33,100~98,200 | 0.0845 |
| state-2-1-wl | 4 | 0 | 4 | 26,200~36,400 | 0.0796 |
| state-2-1-fl | 0 | 4 | | 40,200~160,400 | 0.0551 |
| state-2-1-wfl | 4 | 4 | | 66,200~196,400 | 0.0849 |
| state-2-2-wl | 6 | 0 | 6 | 39,300~54,600 | 0.0874 |
| state-2-2-fl | 0 | 6 | | 60,300~240,600 | 0.0807 |
| state-2-2-wfl | 6 | 6 | | 99,300~294,600 | 0.0866 |
| state-3-1-wl | 4 | 0 | 4 | 26,200~36,400 | 0.0763 |
| state-3-1-fl | 0 | 4 | | 40,200~160,400 | 0.0856 |
| state-3-1-wfl | 4 | 4 | | 66,200~196,400 | 0.0856 |
| state-3-2-wl | 5 | 0 | 5 | 32,750~45,500 | 0.0205 |
| state-3-2-fl | 0 | 5 | | 50,250~200,500 | 0.0864 |
| state-3-2-wfl | 5 | 5 | | 82,750~245,500 | 0.0771 |
| state-4-1-wl | 6 | 0 | 6 | 39,300~54,600 | 0.0824 |
| state-4-1-fl | 0 | 6 | | 60,300~240,600 | 0.0889 |
| state-4-1-wfl | 6 | 6 | | 99,300~294,600 | 0.0840 |
| state-4-2-wl | 9 | 0 | 9 | 58,950~81,900 | 0.0865 |
| state-4-2-fl | 0 | 9 | | 90,450~360,900 | 0.0872 |
| state-4-2-wfl | 9 | 9 | | 148,950~441,900 | 0.0857 |

## 4. Conclusions

Although DRL has shown promise in real-time control of urban drainage systems, there is still a lack of analysis on the optimization of system design and cost-effectiveness. In this study, state design and cost estimation are employed to analyze the influence of the state selection on DRL's performance from the perspective of CSO and flooding mitigation and cost-effectiveness by taking PPO as an example. The significance of this article is to study the design and expenditure of DRL in urban drainage systems, providing support for promoting the implementation of this method. The following conclusions are obtained.

The choice of different states affects the training process of PPO. The training speed of PPO using only water level or a small number of nodes as input is relatively slower than using water flow and more nodes as input. This indicates that it is difficult to train a PPO based on water level and fewer nodes information.

Providing more upstream node information as input to PPO improves its control effect, and using upstream node information on UDS is more effective than using downstream information. The control effect is better when water flow is used as input than when water level is used under the same conditions, while PPOs using both of them have a certain level of improvement, but it is not significant in some cases. Therefore, purely from the perspective of the control effect, the PPO using water flow in both upstream and downstream nodes is relatively better in this case.

Although the performance of the PPOs using water flow as input is better compared to the PPOs using water level, their design based on cost-effective analysis still needs to be considered in the practical application because the cost of flow monitoring is higher than that of water level monitoring. In this case, the PPO state-4-1-fl is recommended because of its relatively low cost and acceptable control performance.

**Author Contributions:** Conceptualization, W.T., K.X., Z.Z. and Z.L.; methodology, W.T.; visualization, W.T. and Z.Z.; writing—original draft, W.T. and K.X.; writing—review and editing, W.T., K.X. and F.L.; project administration, K.X. and F.L.; supervision: K.X., Z.L. and F.L. All authors have read and agreed to the published version of the manuscript.

**Funding:** National Natural Science Foundation of China (grant no. 52270093, funder: Kunlun Xin; grant no. 52170102, funder: Zhenliang Liao; grant no. 52000142, funder: Fei Li).

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** The details of the SWMM in the case is available in these references: [19–21]. All the code and data can be found on Github (https://github.com/DantEzio/DRL_state_selection_cost (accessed on 7 April 2023)).

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Fu, G.; Jin, Y.; Sun, S.; Yuan, Z.; Butler, D. The role of deep learning in urban water management: A critical review. *Water Res.* **2022**, *223*, 118973. [CrossRef]
2. Liao, Z.; Gu, X.; Xie, J.; Wang, X.; Chen, J. An integrated assessment of drainage system reconstruction based on a drainage network model. *Environ. Sci. Pollut. Res. Int.* **2019**, *26*, 26563–26576. [CrossRef] [PubMed]
3. Ochoa, D.; Riano-Briceno, G.; Quijano, N.; Ocampo-Martinez, C. Control of urban drainage systems: Optimal flow control and deep learning in action. In Proceedings of the 2019 American Control Conference, Philadelphia, PA, USA, 10–12 July 2019; pp. 4826–4831.
4. Qi, W.; Ma, C.; Xu, H.; Chen, Z.; Zhao, K.; Han, H. A review on applications of urban flood models in flood mitigation strategies. *Nat. Hazards* **2021**, *108*, 31–62. [CrossRef]
5. Xie, J.; Chen, H.; Liao, Z.; Gu, X.; Zhu, D.; Zhang, J. An integrated assessment of urban flooding mitigation strategies for robust decision making. *Environ. Model. Softw.* **2017**, *95*, 143–155. [CrossRef]
6. Zhi, G.; Liao, Z.; Tian, W.; Wu, J. Urban flood risk assessment and analysis with a 3D visualization method coupling the PP-PSO algorithm and building data. *J. Environ. Manag.* **2020**, *268*, 110521. [CrossRef] [PubMed]
7. Normile, D. Zhengzhou Subway Flooding a Warning for Other Major Cities. AAAS Articles DO Group. 2021. Available online: https://www.science.org/content/article/zhengzhou-subway-flooding-warning-other-major-cities (accessed on 7 April 2023).
8. Huang, X.; Wang, D.; Li, L.; Li, Q.; Zeng, Z. Modeling Urban Impact on Zhengzhou Storm on July 20, 2021. *J. Geophys. Res. Atmos.* **2022**, *127*, e2022JD037387. [CrossRef]
9. Fu, G.; Butler, D.; Khu, S.T. Multiple objective optimal control of integrated urban wastewater systems. *Environ. Model. Softw.* **2008**, *23*, 225–234. [CrossRef]
10. Jean, M.; Morin, C.; Duchesne, S.; Pelletier, G.; Pleau, M. Optimization of Real-Time Control with Green and Gray Infrastructure Design for a Cost-Effective Mitigation of Combined Sewer Overflows. *Water Resour. Res.* **2021**, *57*, e2021WR030282. [CrossRef]
11. Lund, N.S.V.; Falk AK, V.; Borup, M.; Madsen, H.; Steen Mikkelsen, P. Model predictive control of urban drainage systems: A review and perspective towards smart real-time water management. *Crit. Rev. Environ. Sci. Technol.* **2018**, *48*, 279–339. [CrossRef]
12. Schütze, M.; Butler, D.; Beck, B.M. *Modelling, Simulation and Control of Urban Wastewater Systems*; Springer Science & Business Media: Berlin/Heidelberg, Germany, 2011.
13. Sun, C.; Puig, V.; Cembrano, G. Real-Time Control of Urban Water Cycle under Cyber-Physical Systems Framework. *Water* **2020**, *12*, 406. [CrossRef]
14. Lund NS, V.; Borup, M.; Madsen, H.; Mark, O.; Mikkelsen, P.S. CSO reduction by integrated model predictive control of stormwater inflows: A simulated proof-of-concept using linear surrogate models. *Water Resour. Res.* **2020**, *56*, e2019WR026272. [CrossRef]
15. Sun, C.; Lorenz Svensen, J.; Borup, M.; Puig, V.; Cembrano, G.; Vezzaro, L. An MPC-Enabled SWMM Implementation of the Astlingen RTC Benchmarking Network. *Water* **2020**, *12*, 1034. [CrossRef]
16. Garcia, L.; Barreiro-Gomez, J.; Escobar, E.; Tellez, D.; Quijano, N.; Ocampo-Martinez, C. Modeling and real-time control of urban drainage systems: A review. *Adv. Water Resour.* **2015**, *85*, 120–132. [CrossRef]
17. Liao, Z.; Zhang, Z.; Tian, W.; Gu, X.; Xie, J. Comparison of Real-time Control Methods for CSO Reduction with Two Evaluation Indices: Computing Load Rate and Double Baseline Normalized Distance. *Water Resour. Manag.* **2022**, *36*, 4469–4484. [CrossRef]
18. Saliba, S.M.; Bowes, B.D.; Adams, S.; Beling, P.A.; Goodall, J.L. Deep Reinforcement Learning with Uncertain Data for Real-Time Stormwater System Control and Flood Mitigation. *Water* **2020**, *12*, 3222. [CrossRef]
19. Tian, W.; Liao, Z.; Zhang, Z.; Wu, H.; Xin, K. Flooding and Overflow Mitigation Using Deep Reinforcement Learning Based on Koopman Operator of Urban Drainage Systems. *Water Resour. Res.* **2022**, *58*, e2021WR030939. [CrossRef]

20. Tian, W.; Liao, Z.; Zhi, G.; Zhang, Z.; Wang, X. Combined Sewer Overflow and Flooding Mitigation Through a Reliable Real-Time Control Based on Multi-Reinforcement Learning and Model Predictive Control. *Water Resour. Res.* **2022**, *58*, e2021WR030703. [CrossRef]
21. Zhang, Z.; Tian, W.; Liao, Z. Towards coordinated and robust real-time control: A decentralized approach for combined sewer overflow and urban flooding reduction based on multi-agent reinforcement learning. *Water Res.* **2023**, *229*, 119498. [CrossRef]
22. Zhang, M.; Xu, Z.; Wang, Y.; Zeng, S.; Dong, X. Evaluation of uncertain signals' impact on deep reinforcement learning-based real-time control strategy of urban drainage systems. *J. Environ. Manag.* **2022**, *324*, 116448. [CrossRef]
23. Gao, Y.; Sarker, S.; Sarker, T.; Leta, O.T. Analyzing the critical locations in response of constructed and planned dams on the Mekong River Basin for environmental integrity. *Environ. Res. Commun.* **2022**, *4*, 101001. [CrossRef]
24. Sarker, S. Investigating Topologic and Geometric Properties of Synthetic and Natural River Networks under Changing Climate. Ph.D. Thesis, University of Central Florida, Orlando, FL, USA, 2021; p. 965.
25. Li, L.; Qiao, J.; Yu, G.; Wang, L.; Li, H.-Y.; Liao, C.; Zhu, Z. Interpretable tree-based ensemble model for predicting beach water quality. *Water Res.* **2022**, *211*, 118078. [CrossRef] [PubMed]
26. Cabral, M.; Loureiro, D.; Almeida, M.d.C.; Covas, D. Estimation of costs for monitoring urban water and wastewater networks. *J. Water Supply: Res. Technol. -Aqua* **2019**, *68*, 87–97. [CrossRef]
27. Rossman, L. *SWMM 5.1 Storm Water Management Model User's Manual*; US Environmental Protection Agency: Cincinnati, OH, USA, 2015.
28. Lou. Low Impact Development Layout of Sponge City Construction Based on SWMM. Master's Thesis, Tongji University, Shanghai, China, 2018. (In Chinese).
29. Schulman, J.; Levine, S.; Abbeel, P.; Jordan, M.; Moritz, P. Trust region policy optimization. In Proceedings of the International Conference on Machine Learning, Pittsburgh, PA, USA, 23–29 July 2015; pp. 1889–1897.
30. Schulman, J.; Wolski, F.; Dhariwal, P.; Radford, A.; Klimov, O. Proximal policy optimization algorithms. *arXiv* **2017**, arXiv:1707.06347.
31. Kingma, D.P.; Ba, J. Adam: A Method for Stochastic Optimization. *arXiv* **2014**, arXiv:1412.6980. [CrossRef]
32. Liu, Y.-Y.; Slotine, J.-J.; Barabási, A.-L. Controllability of complex networks. *Nature* **2011**, *473*, 167–173. [CrossRef] [PubMed]
33. Yuan, Z.; Zhao, C.; Di, Z.; Wang, W.-X.; Lai, Y.-C. Exact controllability of complex networks. *Nat. Commun.* **2013**, *4*, 2447. [CrossRef] [PubMed]
34. Sarker, S.; Veremyev, A.; Boginski, V.; Singh, A. Critical Nodes in River Networks. *Sci. Rep.* **2019**, *9*, 11178. [CrossRef]

**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.