applied sciences

Article

Advanced Control by Reinforcement Learning for Wastewater Treatment Plants: A Comparison with Traditional Approaches

Félix Hernández-del-Olmo [1,*], Elena Gaudioso [1], Natividad Duro [2], Raquel Dormido [2] and Mikel Gorrotxategi [1]

[1] Department of Artificial Intelligence, National Distance Education University (UNED), Juan del Rosal 16, 28040 Madrid, Spain
[2] Department of Computer Sciences and Automatic Control, National Distance Education University (UNED), Juan del Rosal 16, 28040 Madrid, Spain
* Correspondence: felixh@dia.uned.es

Abstract: Control mechanisms for biological treatment of wastewater treatment plants are mostly based on PIDS. However, their performance is far from optimal due to the high non-linearity of the biological and changing processes involved. Therefore, more advanced control techniques are proposed in the literature (e.g., using artificial intelligence techniques). However, these new control techniques have not been compared to the traditional approaches that are actually being used in real plants. To this end, in this paper, we present a comparison of the PID control configurations currently applied to control the dissolved oxygen concentration (in the active sludge process) against a reinforcement learning agent. Our results show that it is possible to have a very competitive operating cost budget when these innovative techniques are applied.

Keywords: advanced control; reinforcement learning; wastewater system

check for updates

Citation: Hernández-del-Olmo, F.; Gaudioso, E.; Duro, N.; Dormido, R.; Gorrotxategi, M. Advanced Control by Reinforcement Learning for Wastewater Treatment Plants: A Comparison with Traditional Approaches. *Appl. Sci.* **2023**, *13*, 4752. https://doi.org/10.3390/app13084752

Academic Editor: Ramaraj Boopathy

Received: 27 February 2023
Revised: 30 March 2023
Accepted: 5 April 2023
Published: 10 April 2023

1. Introduction

Wastewater treatment plants (WWTPs) are a very complex process operated to obtain an effluent from the wastewater that can be returned to the water cycle with a minimal impact on the environment. An efficient operation of WWTPs should also guarantee minimization of the operational costs and the sludge production. The control systems used to achieve these goals have a strong impact on the efficiency and operation of the WWTP.

A conventional WWTP is composed of different treatment stages and processes linked to each other. Usually these treatment stages are as follows: pre-treatment, primary, secondary and tertiary treatment [1]. In the pre-treatment stage, large debris is removed from the raw wastewater. In the primary treatment stage, smaller particles such as the settable solids and floatable materials from the effluent are removed. Secondary treatment involves complex biological processes that are used to remove refractory solids, such as sludge, that were not removed during primary treatment. Some WWTPs provide only secondary treatment and when this treatment ends, the effluent is disinfected and released to the environment. If the WWTP provides tertiary treatment process, once the secondary effluent is transferred, additional unwanted constituents (for instance, phosphorus or nitrogen) are removed to meet some regulatory requirements.

Sludge removed in the secondary treatment is further treated in a separate sludge digestion process. This process is very important during the design and operation of all WWTPs. It is fundamental to reduce its volume and to stabilize the organic materials. Stabilized sludge can be better manipulated and a volume reduction implies a decrease in the costs of pumping and storage.

Different types of biological WWTPs treatments can be found [2]. However, each treatment can be classified as aerobic, anaerobic, or anoxic treatment depending on whether oxygen is used or not. The most typical secondary wastewater treatment plant is the activated sludge process (ASP). It uses microbial degradation for the digestion of soluble organic constituents within primarily treated effluent. In our paper, the BSM1 is used as the benchmark scenario to implement the control strategies [3], as it is one of the most popular in the literature for WWTPs. The plant layout represents the ASP, one of the most demanding process in a WWTP with nitrification/denitrification stages [4]. It is composed of activated sludge reactors in series, followed by a clarifier. The first two reactors are non-aerated and the other three are aerobic tanks [3].

WWTPs operate by controlling the values of certain variables that are returned by sensors located in the plant. In the ASP, several variables are considered [5–7]: dissolved oxygen concentration, ammonia concentration, internal recycle flow rate, sludge recycle flow rate or external carbon dosing. Among all these variables, dissolved oxygen concentration (DO) is the one in charge of the most expensive cost of the plant [5,8]. In aerobic treatment, blow oxygen is supplied in the aeration tank to speed up the degradation process, which implies an important energy demand. Therefore, the DO content in the aeration tank is an essential control test.

The most widely used approach to control different variables in WWTPs is PID controllers [9–12]. However, due to the complexity of WWTP processes and the big differences in external conditions, the performance of PIDs is not optimal [13], especially when operational cost is a major concern [14]. More advanced control techniques have also been proposed for WWTPs, such as model predictive control or fuzzy-logic control [15,16]. In fact, artificial intelligence approaches such as artificial neural networks [17–21] or reinforcement learning [22] have been applied in the last decade [23,24]. In addition, among artificial intelligence techniques for industrial applications, deep learning techniques are increasing in popularity. For example, recurrent neural networks are used to detect anomalies in influent conditions [25]. Additionally, deep learning algorithms, namely, recurrent neural network, long-short term memory and gated recurrent unit [26], are used to analyze and predict water quality. Besides the control of variables in the plant, machine learning algorithms are also used to develop soft sensors that predict values that are difficult to obtain using physical sensors [27–30]. In addition, we also proposed a Reinforcement Learning (RL) agent to control DO concentration in the active sludge process of a WWTP [8,31] (in Table 1 some of current control approaches in WWTPs are summarized). However, a recurrent finding in our search was a lack of comparison of novel techniques with current approaches already working. In fact, as a first approach in a previous work [8], we compared our RL agent against a PI cascade control, but we also proposed in that same work that a more exhaustive search would be interesting. In [8] we have improved the economic and environmental performance of the WWTPs using a RL approach. Using the BSM1 improvements in the operating costs of the N-ammonia removal process have been obtained. The proposed control scheme shows better performance than a manual plant operator when disturbances affect the plant and savings in a year of a working BSM1 plant are shown.

Table 1. Control approaches in WWTPs.

Reference	Techniques	Goal
[10,11]	PIDs	Control of DO concentration
[12]	PIDs and fuzzy logic techniques	Control of DO and ammonium/nitrate concentration
[15]	Genetic algorithm optimization	Control of nitrogen and ammonia concentration
[16]	Fuzzy control	Control of nitrogen and ammonia concentration
[17,21]	Artificial Neural Networks	Control of nitrogen and/or DO concentration

Table 1. *Cont.*

Reference	Techniques	Goal
[18]	Recurrent Neural Networks	Control of phosphorus concentration
[19]	Artificial Neural Networks	Prediction of effluent biochemical oxygen demand and the effluent total nitrogen
[22]	Reinforcement Learning	Optimal control of hydraulic retention time and internal recycling ratio in an naerobic–anoxic–aerobic system
[25]	Recurrent neural networks	detect anomalies in influent conditions
[26]	Recurrent neural network, long-short term memory	analyze and predict water quality
[28,29]	recurrent neural networks deep learning network	predict the ammonium, total nitrogen, and total phosphorus removal efficiency
[30]	Machine learning techniques (Support Vector Machine, Decision Trees, Random Forest and Gaussian Naive Bayes, k-nearest neighbors)	Predict weather conditions
[31]	Reinforcement learning	Control DO concentration

In [9] a review that covers automatic control of continuous aeration systems is presented. The paper focuses on recent published research that describes different control structures to control the DO concentration and the aerobic volume, with a special focus on plants with nitrogen removal.

The main goal of this paper is to compare the RL agent with more complex PI-based techniques; more specifically, those described in [9]. The advantages of using the RL agent approach against the PID structures are shown: (i) the RL agent operates without knowing the initial model, (ii) the RL agent outperforms all configurations of traditional PIDs, even when the changing environment is more significant (for example, on weekends or during storms) and (iii) a comparative study of the savings obtained is presented and it is shown how the RL agent minimizes the costs.

The rest of the paper is organized as follows. In Section 2, we briefly describe the BSM1 simulation model. The PI-based techniques considered for comparison and an overview of the proposed RL agent [8,31] are also presented in this section. In Section 3 we describe the results and in Section 4, the conclusions are presented.

2. Materials and Methods

In this section, we begin with a brief description of the WWTP simulation model. Next, we describe the control structures used in this paper for the comparison [9], as well as the RL agent used.

2.1. WWTP Simulation Model

The comparison is made in the well-known BSM1 [32] simulation model. It defines a plant layout incorporating an active sludge model, influent loads, test procedures and evaluation criteria. A more in-depth description of BSM1 can be found in [8].

The BSM1 model is initialized with a constant input into the influent for a period of at least 100 days to obtain an stable plant [3]. Three different influents are defined as initial conditions, each one representing a different atmospheric event, namely: *dry* (the dry period), *rain* (rain events) and *storm* (storm events). Each of these influents is composed of the data for two weeks of simulation. The data corresponding to the rain influent are the same as in dry weather, but adding an episode of rain, and in the case of the storm influent, adding two storm episodes, which are shorter, but of greater intensity than the rain episode.

Figure 1 shows the influent chart with the different weather conditions in BSM1.

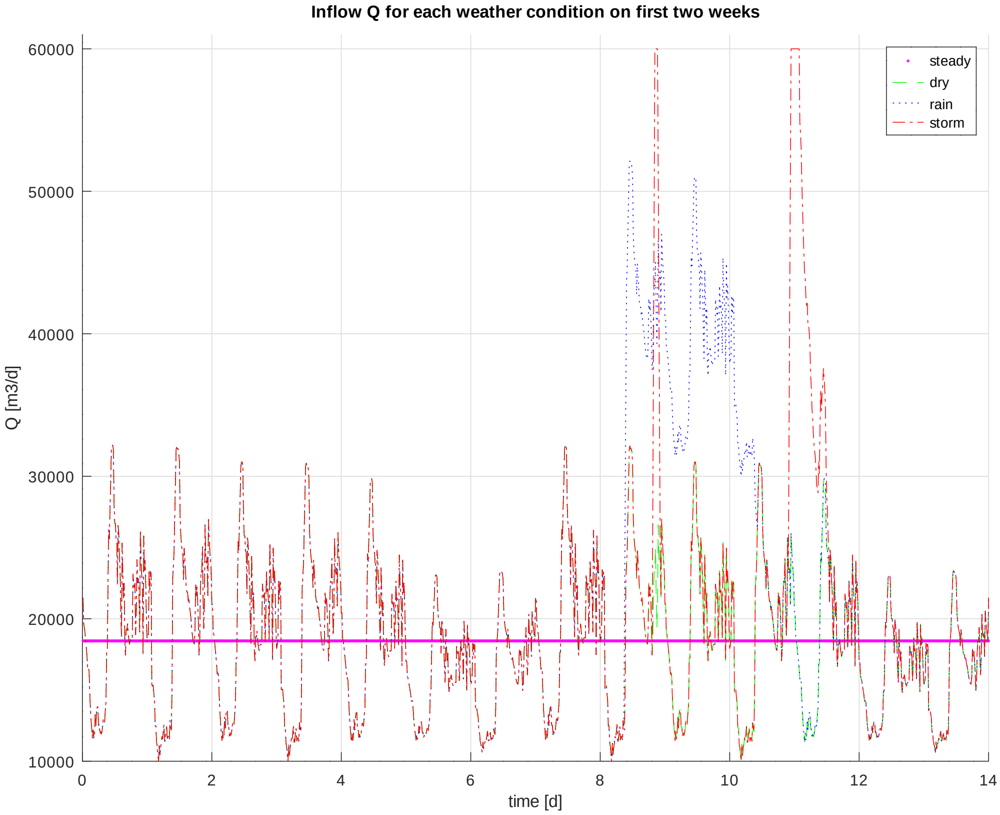

Figure 1. Different weather conditions in the influent defined in BSM1: dry (yellow), rain (blue) and storm (red).

2.2. Control Structures

Control structure design requires different decisions about the control system before the controller design can be accomplished. It includes the selection of manipulated and controlled variables, or the selection of control configuration and the control law. A critical step is the control configuration selection to determine how to obtain the structure of the overall controller that interconnects the variables; that is, how to pair the variables to form control loops. This is a problem of a combinatorial nature, as the number of control configurations increases quickly with the number of process variables. Normally, the control structure will heavily depend on the process to be controlled [33].

The selection of a proper aeration control structure is a particularly complex task due to the intrinsic complexity of the WWTPs systems. Its design and implementation must be carried out properly so as not to compromise its potential for optimization. In this paper, we consider the following five classic control structures for aeration control [9]: PI cascade control structure, ammonium-based control: feedback control, ammonium-based control: feedfordward–feedback control, advanced SISO and MIMO controllers and control of the aerobic volume. Following notation in [9], we will refer to these configurations as A, B1, B2, C and D, respectively.

Each control structure presents a different level of complexity both in programming and in sensors.

2.2.1. PI Cascade Control Structure

The most popular aeration control structure is a feedback controller, often set up in a cascade. Figure 2 shows a PI cascade control structure. It defines two control loops, outer and inner, in the aeration process of the tank and two PI controllers in its structure. The internal loop controls the air flow rate and the external loop controls the dissolved oxygen.

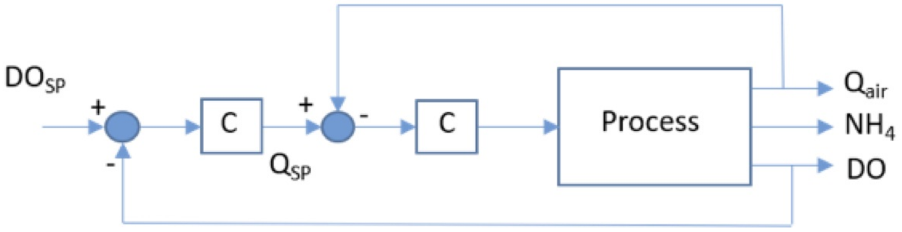

Figure 2. PI cascade control structure.

WWTP is a non-linear process with increasing response times and a measured controlled variable in the internal loop (air flow rate Q_{air}). Air flow rate control and dissolved oxygen (DO) control obtain benefit from a simple cascade control structure. The advantage of the cascade control is that the inner loop compensates for the sensitivity and nonlinearity of the plant in the closed loop, reducing the parameter perturbations in the outer loop. In this structure, the DO set-point is decided by the operator.

2.2.2. Ammonium-Based Control: Feedback Control

The main difference between this strategy and the previous one is that the DO set-point is calculated based on the measured ammonium concentrations in the outlet of the activated sludge process or from a sensor included in the structure (see Figure 3). Consequently, we now have a triple cascade controller. In this case, the operator must decide the set point of the ammonium concentrations.

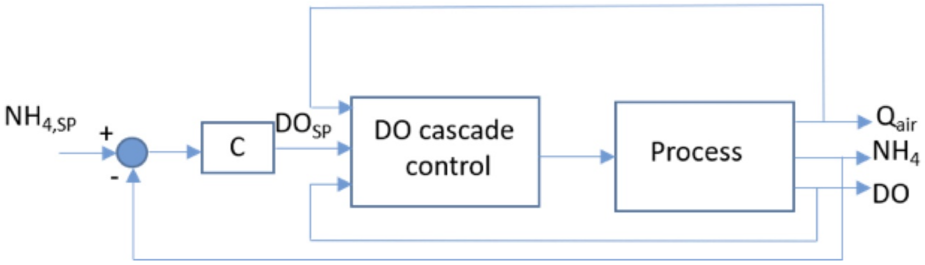

Figure 3. Ammonium-based control: feedback control.

2.2.3. Ammonium-Based Control: Feedforward–Feedback Control

An alternative structure with which to calculate the DO set-point is to add a feedforward control for improved disturbance rejection combined with feedback control, as shown in Figure 4. The influent ammonium concentration and influent flow rate are the main disturbances used by the feedforward control. This control allows to predict their behavior and consequently modify the set point. In parallel, a feedback control is added in the structure to tune the prediction made considering the true measurements.

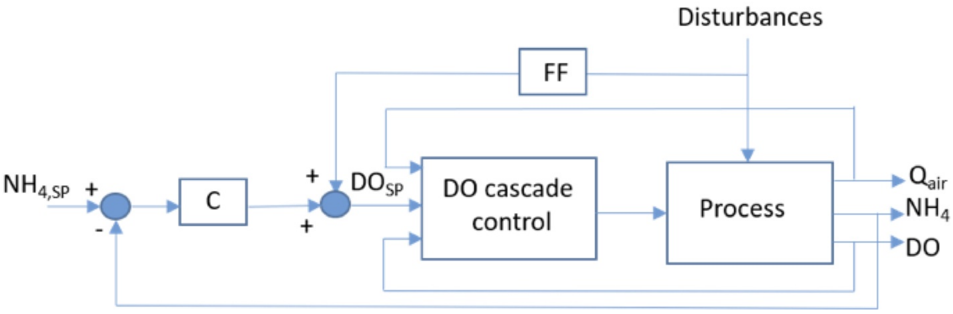

Figure 4. Ammonium-based control: feedforward-feedback control.

2.2.4. Advanced SISO and MIMO Controllers

Advanced control techniques use different model-based and optimal controllers. Model-based controllers refer to those control algorithms which include a process model in the control law. The model can be either based on a real simulated process or on a black-box. Usually, the model is used to identify the output of the controller to obtain optimal behavior. Figure 5 shows the control structure. There are two main differences from Figure 4. The first is the treatment that the controller makes with the disturbance. The second refers to the two new inputs in the first controller: a cost function to determine the optimal solution and the constraints on the system. This control structure uses a MIMO controller, that it is to say that each manipulated variable affects several controlled variables, causing loop interactions.

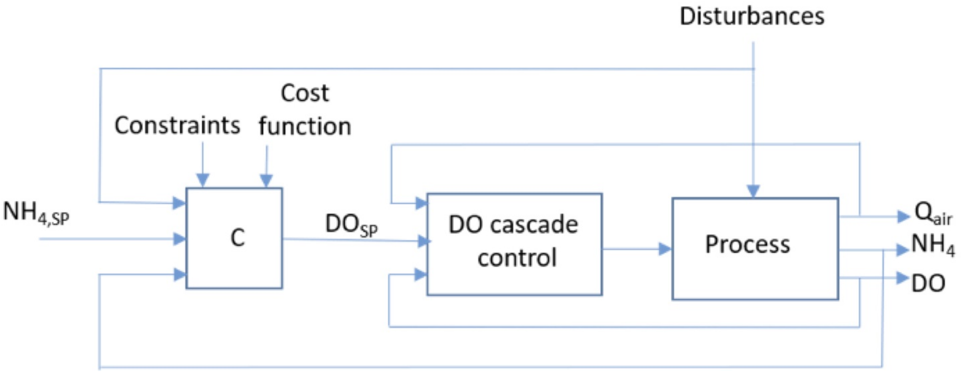

Figure 5. Advanced SISO and MIMO controllers.

2.2.5. Control of the Aerobic Volume

Another possibility for control is to use a control structure such as the one shown in Figure 6. This structure can operate in two different forms at the same time using a switch to adjust the aeration intensity. The set point of the N-ammonia is compared with the real one, and a rule is used as input to determine if aeration should be supplied in order to obtain an operation closer to the ammonium set-point. Moreover, it allows us to obtain energy savings and an improvement in the denitrification process.

Therefore, the volume of the plant is modified through the use of another tank due to a condition. The rule is as follows: if the aeration of the last (aerobic) tank is at 100% for 24 h, then a new tank is added to the circuit. This new added tank operates at 100% while activated. When the aeration of the last tank falls below 100%, the extra tank is removed from the circuit.

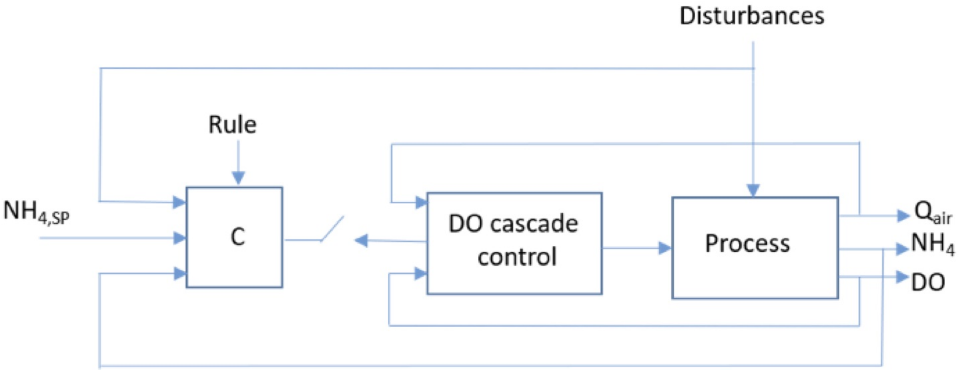

Figure 6. Control of the aerobic volume.

2.3. *Reinforcement Learning*

Reinforcement Learning (RL) is a machine learning area that trains an agent to perform a task in an uncertain environment. To this end, it completes certain actions in order to maximize reward over time.

2.3.1. Reinforcement Learning Elements

The main two elements in a RL system are the agents and the environment. The agent is the tool that, based on rewards and punishments, makes decisions about the environment. The world with which the agent interacts is called the environment. The RL process is usually modeled as an iterative loop. When the RL agent receives a state from the environment, it takes an action that moves the environment to a new state. Then, the environment rewards the RL agent, which makes a new decision, repeating the RL loop until the goal is reached or a maximized reward is achieved.

2.3.2. Reinforcement Learning Agent

A RL agent acts in an environment to change certain conditions. For each action in each time step t, the agent obtains a reward. The goal of the RL agent is to optimize the expected rewards, taking into account the conditions defined in the environment.

This environment is usually defined as a Markov decision process. Nevertheless, this model is not mandatory. RL also supports model-free algorithms [34]. In these methods, the agent has to obtain the model of the environment as well as the optimal policy. This is the case for the RL agent defined in the BSM1 model, where the main objective of the agent is to lower the operation cost (OC) of the plant as much as possible while keeping the N-ammonia under 4 mg/L.

The RL agent receives the value of S_{NH} and S_O as inputs and, as outputs, the DO set-point for tank 5 for each time step t. The control loop that sets the DO set-point with the RL agent is shown in Figure 7.

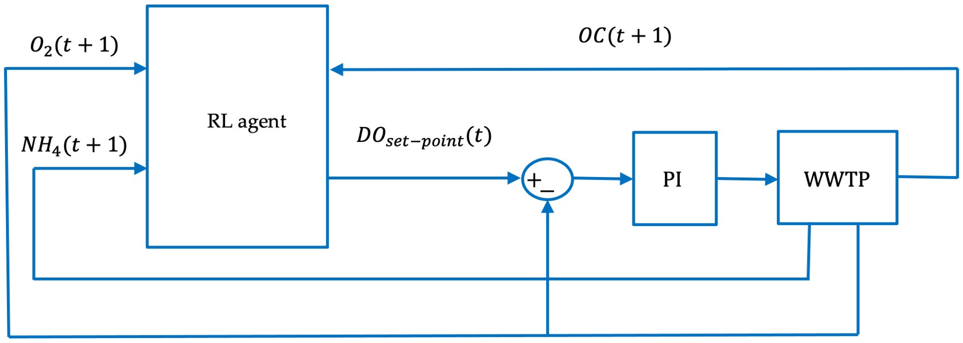

Figure 7. Control loop using the RL agent. Reprinted/adapted with permission from Ref. [8]. Copyright 2016, Hernández-del-Olmo et al.

The agent should optimize the value of OC selection between three possible DO set-points for tank 5: 1.2 mg/L, 1.5 mg/L 1.85 mg/L. The operation cost OC for each time step t is defined in [8] (see Equation (1)):

$$OC(t) = \gamma_1(AE(t) + ME(t) + PE(t)) + \gamma_2 SP(t) + EF(t) \tag{1}$$

where AE is the aeration energy (kWh), ME is the mixing energy (kWh), PE is the pumping energy (kWh), SP is the sludge production for disposal (kg) and EF stands for the effluent fines (€). Weights γ_1 and γ_2 are set in proportion to the weights in the operating cost index defined in the benchmark BSM1. In the same way as Stare et al. [35], we consider 0.1 €/kWh. Hence, γ_1 = 0.1 €/kWh, and γ_2= 0.5 €/kg following closely [32] that considers γ_2 as five times γ_1. OC considered by RL agent and how the values of AE, ME, PE and SP are obtained are defined in [8]. The algorithm of the RL agent is shown in Algorithm 1.

Algorithm 1: RL agent pseudocode.

```
Setup:
$\gamma$ = Time horizon
$max\_action = 2$ // actions: {0,1,2}
$DO_{max}$ = Set-point max
$DO_{min}$ = Set-point min
$DO_{step} = (DO_{max} - DO_{min})/(max\_action + 1)$
Input:
$s(t) = [NH_4(t), O_2(t)]$// state of the environment
$r(t) = -OC(t)$// reward
Output:
$DO$ : Real
Internal:
$Q(s, a)$ : random initialization
$a$ : action (0..max_action)
Algorithm Main:
    Initialize Q(s,a)
    while true do
        // execute every 15 minutes
        $s(t) = [NH_4(t), O_2(t)]$
        $a = next\_action(Q, s)$
        $DO = DO_{min} + a \times DO_{step}$
        $execute(DO)$// now the plant has the control
        $r(t) = -OC(t)$// the plant returns its reward
        $Q = update\_Q(Q, s, a, r, \gamma)$
    end
```

Even though there are several approaches with which to achieve an adaptive behavior using ML (see Section 1), most of them belong to a *non-interactive* way of learning. In fact, the use of past data is always necessary to build an accurate ML model of the plant and the influent. However, in our work, we focused on the RL approach because we wanted to show an ML system that updates its model in run-time while it interacts with the plant [31]. In fact, the most distinct characteristic is that the RL agent keeps *its own interaction with the plant* into the model. This approach has a lot of advantages, but the main one is that the agent adapts to drifting characteristics of the influent and the plant in an autonomous way.

This agent acts as a new block in the BSM1 model, as shown in Figure 8.

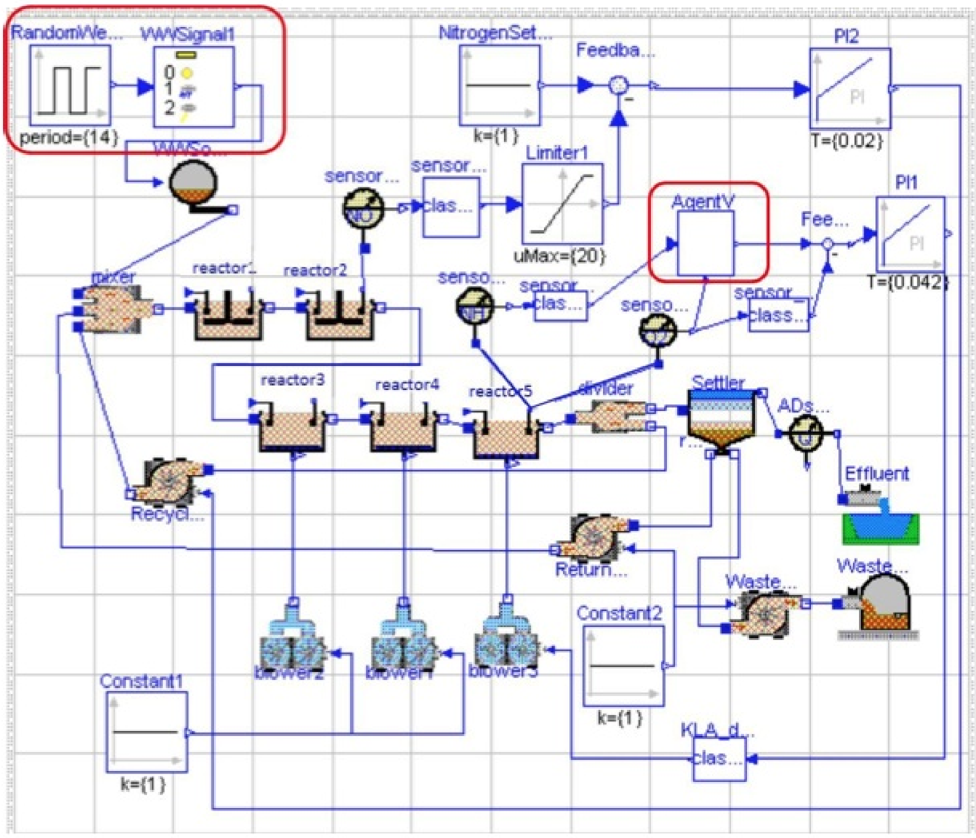

Figure 8. RL agent in BSM1.

The weather considered by the RL agent in the simulation is defined in the BSM1, where it rains 20% of the time, there is a storm 10% of the time and the remaining 70% the time is dry. The distribution is random. The climate is defined with a value of "0" for dry, "1" for rain and "2" for storm.

2.4. Conditions of the Simulation

In the previous section, we showed the simulated WWTP and the climate pattern, but here, we present this pattern again in more detail (see Figure 9). Figure 9 shows the most significant climate periods to obtain a deeper insight of what happens in each control approach. All simulations have been developed under the same weather conditions. So the comparison between the different control configurations are more realistic.

It is worth noting that PIDs are configured, pre-programmed and optimized from the very beginning. However, the RL agent starts with a blank model which is learned day after day while trying to control the WWTP to the best possible extent. Moreover, PIDs are optimized for each of the three possible weather conditions (dry, rainy and stormy), and their configuration is changed to the best one for each condition. However, the RL agent faces a somewhat more real situation because it learns from the WWTP without any consideration, nor any information about the current weather condition. To sum up, we will compare a *real* RL agent behavior against several *ideally* optimized PIDs.

In order to obtain insight into the PIDs and the RL agent behaviors, we will focus on two main variables [8]: N-ammonia concentration in tank 5 (S_{NH}) and total Operation Cost (OC) of the plant (see Equation (1)). We will also focus on two phases of the RL agent: the initial phase, in which the RL agent has learned something for the first time, and the final phase, when the RL agent has already had the opportunity to learn by facing different WWTP conditions. To this end, we will extract in Table 2 the six main periods we will consider from Figure 9.

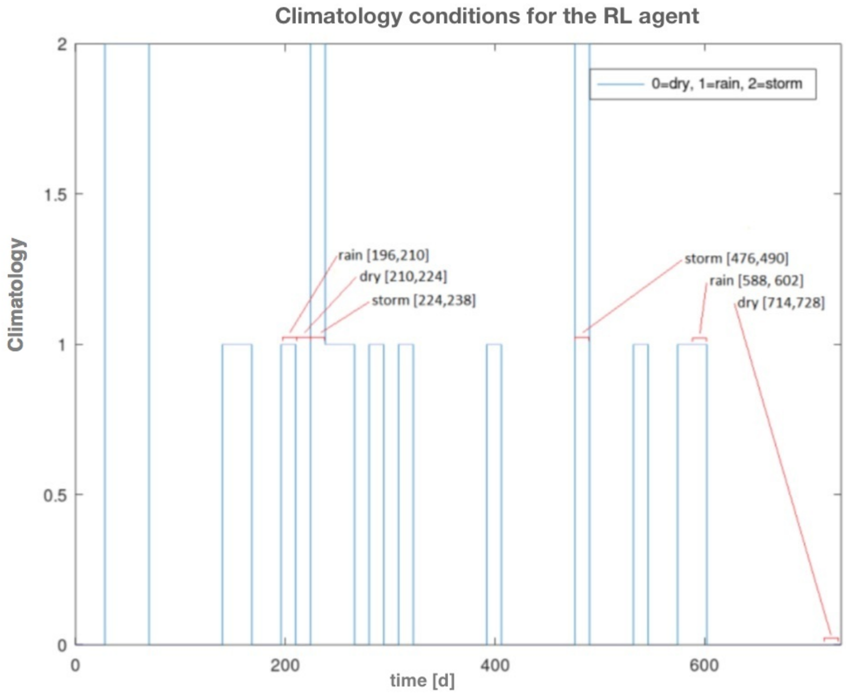

Figure 9. Days on which the RL agent and the PIDs will be compared.

Table 2. Comparison periods.

Weather Condition	Initial Phase Days	Final Phase Days
dry	210–224	714–728
rain	196–210	588–602
storm	224–238	476–490

In order to obtain an initial understanding of what improves in the RL agent against the PID behaviors, in the next section, we will first show of all the savings we achieved using the RL agent.

3. Results

In this section, we will compare PIDs against the RL agent: we will show the pros and cons of having an RL agent to control a WWTP, focusing on the biological stage, against the widely used and tested PID approaches.

3.1. Operation Cost Savings

Figure 10 shows the saving obtained during the initial and the final 2 weeks (second and third columns of Table 2) for each of the three weather conditions. We present three groups of comparisons as follows. In black, configuration A (see Section 2.2.1); B1 in red (see Section 2.2.2); [B2,C,D] in different degrees of blue (see Sections 2.2.3–2.2.5, respectively). We can also find a summary of the RL improvement against each method in Table 3.

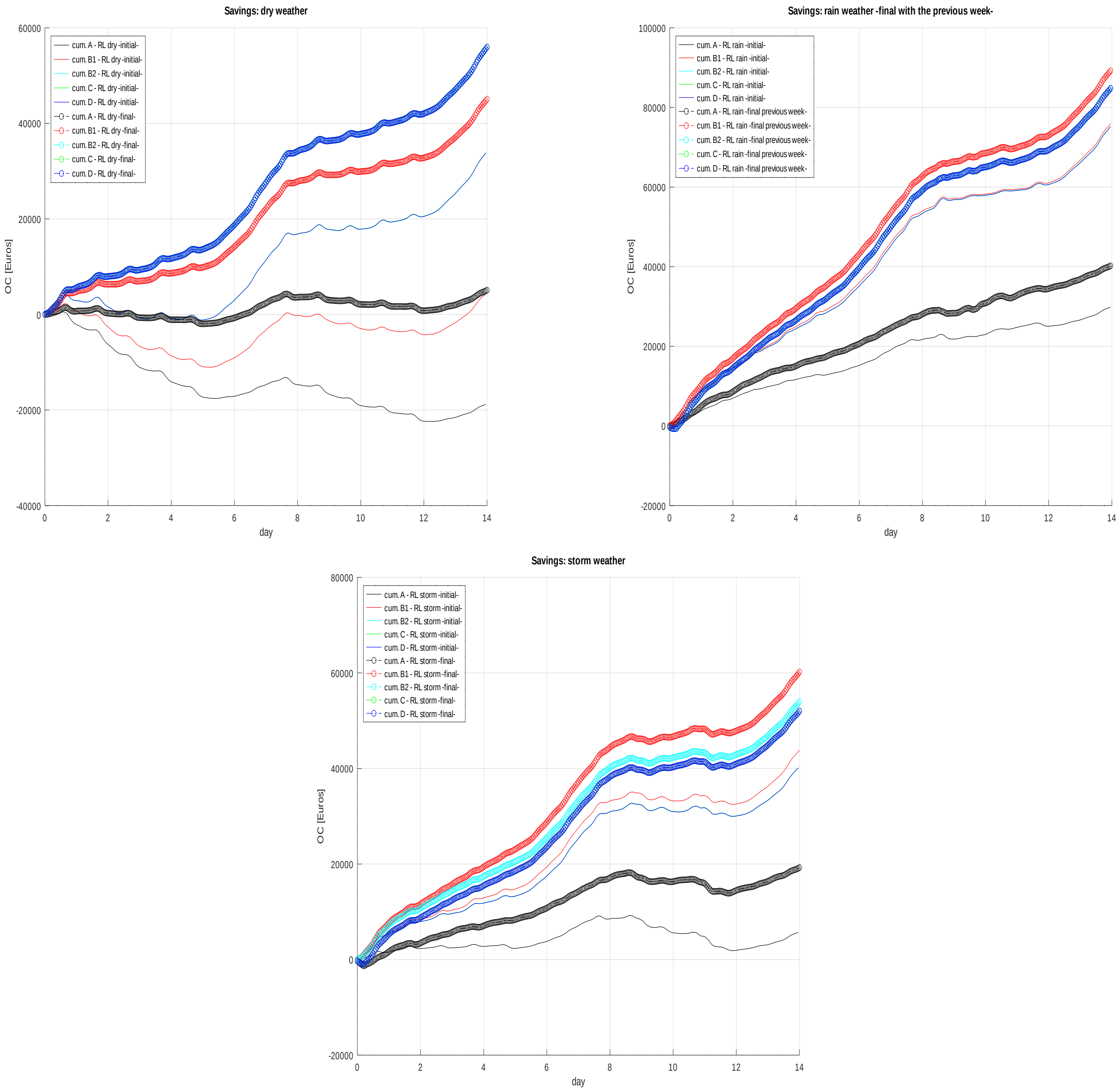

Figure 10. Savings obtained using the RL agent (cumulated difference of Operation Costs).

Table 3. OC improvement percentage ($\frac{final-initial}{initial} \times 100$) of the RL agent against each control method from the initial to the final stage (14 days cumulation sum).

Control Method	Dry	Rain	Storm
A	126.29%	34.61%	235.6%
B1	937.97%	17.50%	37.45%
B2	64.38%	12.80%	33.64%
C	64.18%	12.76%	28.68%
D	64.32%	12.84%	28.77%

In concrete, looking at Figure 10, it can be appreciated that the RL has the A configuration, the hardest one to overcome in every weather condition. On the opposite side, the blue group are clearly overcome by the RL agent since the onset of each weather condition. Lastly, B1 configuration is in the middle.

We will start the comparison with the A and B1 configurations on dry weather. Remarkably, this peaceful dry weather condition is the hardest one that the RL agent must face. Let us obtain some insights to identify why.

In Figure 10, dry weather, the RL agent shows that the initial phase has a hard time when it tries to control the plant with low OC, at least when compared to A and B1 configurations. However, notice that even though it behaves worse on weekdays, it is on weekends when the agent makes use of its flexible and adaptive behavior to improve the OC. This is a recurrent fact, as we will see in Sections 3.2 and 3.3. In fact, on the initial phase, it ends with the 14 days overcoming the savings in all configurations except for A. Nevertheless, once the RL agent has had the opportunity to learn (in the final phase), it ends up positive after 2 weeks of dry condition against each configuration, even compared with the most conservative A configuration.

Now, attending to Figure 10, for rainy and stormy weather conditions, the RL agent overcomes every PID configuration from the very first time. Once again, the flexible and adaptive behavior of the RL agent is noteworthy during these unmerciful weather conditions.

In order to gain more insight into the different behaviors and why the RL agent ends up saving that much, we will see in detail the two main variables of the process: S_{NH} and OC.

3.2. N-Ammonia Concentration

Although the whole Activated Sludge Process is measured by the Operation Cost (OC), first, we will look at the ammonia (S_{NH}) signal in order to deeply understand the OC measure (showed in the last Section 3.3). Notice first that OC (see Equation (1)) is a trade off between the energy used (mostly the energy for blowing oxygen into the aerobic tanks) and the Effluent Fines (mostly obtained because of high values of S_{NH} at the effluent). Thus, the greater the N-Ammonia concentration S_{NH} at the effluent, the higher the payments in fines at the WWTP.

In Figure 11, we show the S_{NH} initial phase for each weather condition: second column of Table 2. If we look at the dry weather condition in Figure 11, we see that the RL agent is the worst in fines (too high ammonia) for each top of the graph. Clearly, the higher the S_{NH} concentration, the worse the fine. On the contrary, the B2, C and D groups are the PIDs with the best behavior in N-ammonia. In the same Figure 11, for rainy weather, the RL is a little better: it is the worst for some tops, but not all. In addition, for rainy weather, some bottoms are better for RL. The results are similar for stormy weather.

Now, let us focus on the final phase, Figure 12, the third column of Table 2. Let us start with dry weather. Comparing it with Figure 11, we see lower tops and lower bottoms in the RL agent's behavior; thus, it behaves better than the rest of the controls (except for some tops that the agent cannot improve more). However, in rainy weather, the agent is one of the worst (at least concerning the S_{NH} behavior). Finally, we see an improvement in the stormy weather, in which we can see lower tops and lower bottoms as well. In summation, we see a light improvement for dry and stormy weather conditions in S_{NH} behavior in the final phase.

As a summary, let us say that N-ammonia concentration is only a part of the whole picture, because saving energy is also important. In this section, we wanted to show in detail that the RL agent can learn to sacrifice the effluent water cleaning in order to save energy. In fact, the agent's decision depends on the context, because its final objective is to reduce the global operation cost in the WWTP.

Let us focus on this measure OC in the next subsection.

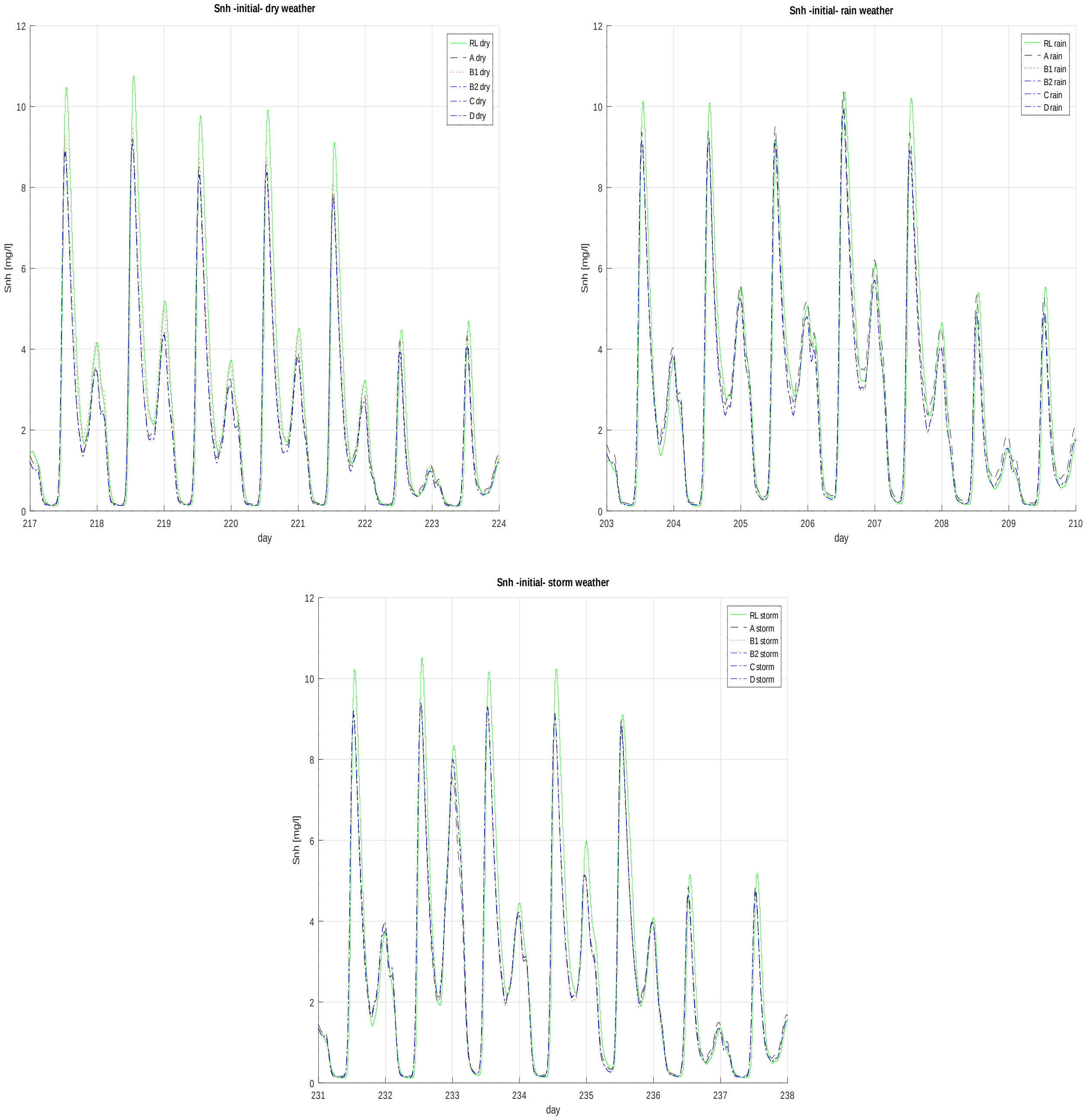

Figure 11. S_{NH} concentration for the **initial** phase.

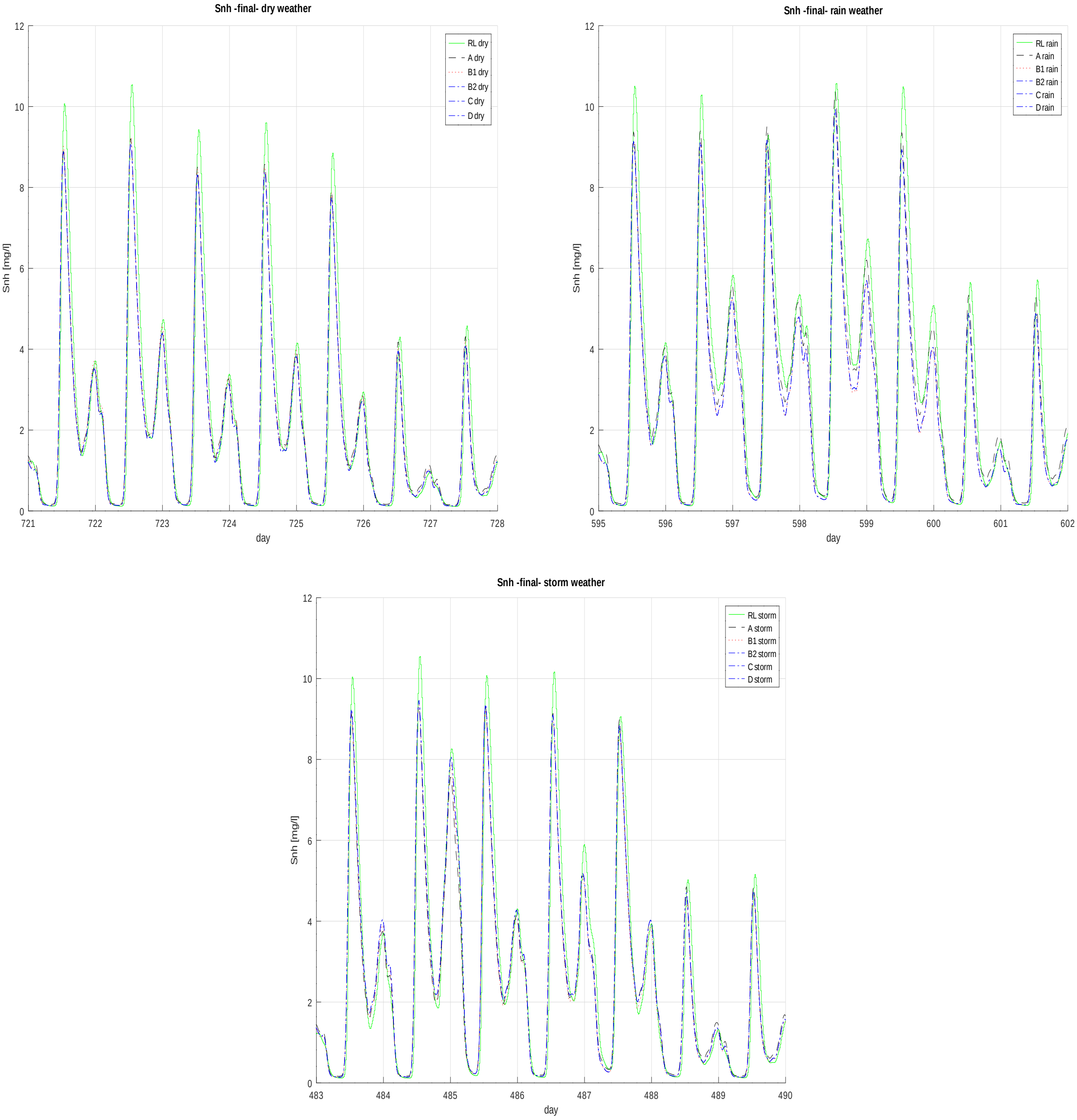

Figure 12. S_{NH} concentration for the **final** phase.

3.3. Operation Cost

A final target of the control is the minimization of the Operation Cost (OC). In Figure 13, we observe the dry weather condition in the initial phase (second column of Table 2). The RL agent is shown to be the worst OC for the most of tops, but also the best for the most of the bottoms (on many occasions, in combination with the A configuration). The excellent behavior of the RL agent on weekends, when it can save a lot of OC, is also relevant. If we look at Figure 11, dry weather, the RL agent minimally increases the S_{NH} without incurring high fines in order to save energy.

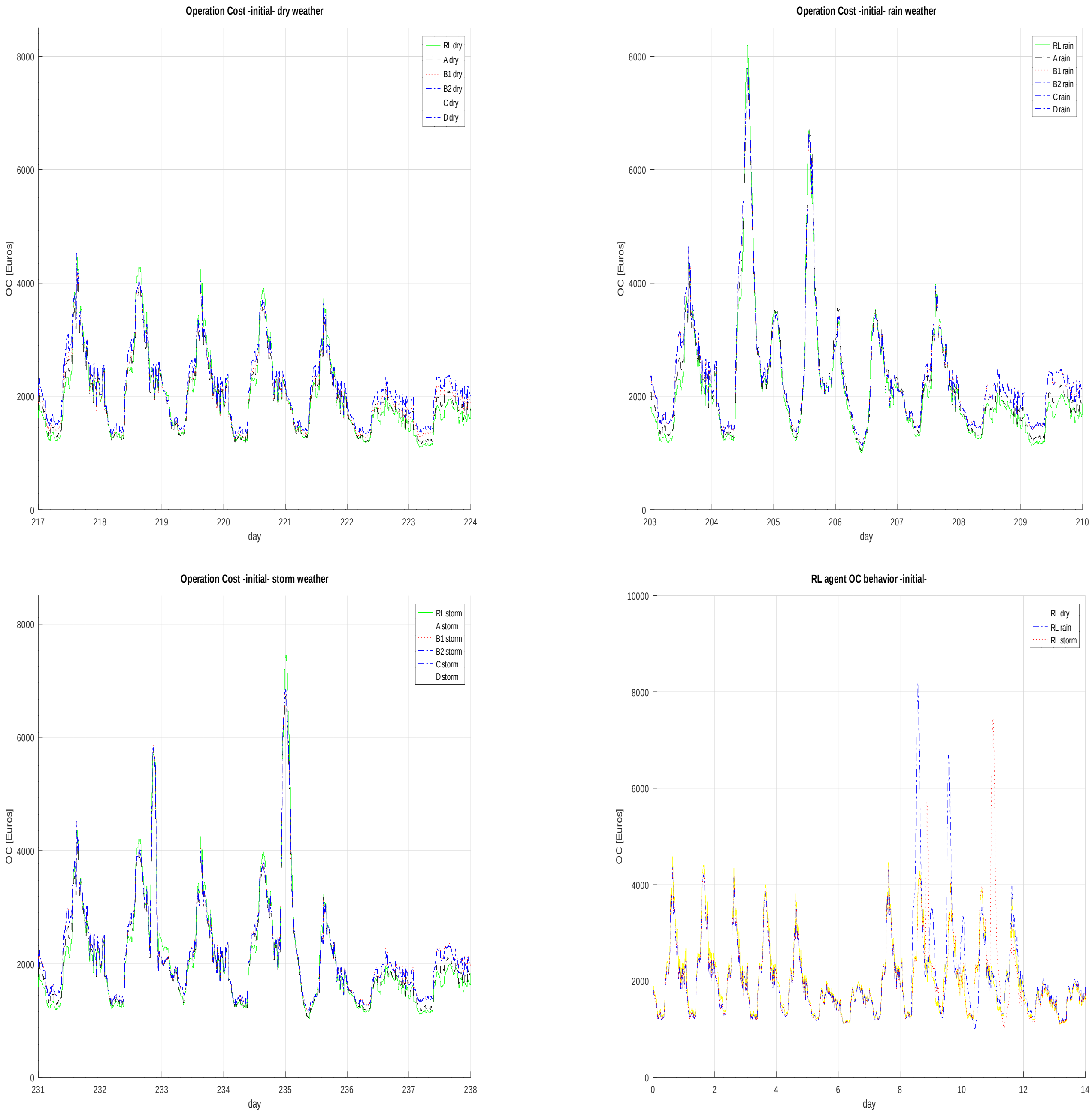

Figure 13. Operation Cost *OC* for the **initial** phase.

In Figure 13, dry weather, the control configurations are ordered from better to worse attending to the *OC* as follows: A, B1 and the set of B2, C and D. This order is the same if configurations are sorted from less to greater attending S_{NH} concentration (see Figure 11). The RL agent has not been included into this list because its behavior is different. RL agent behavior changes depending on tops and bottoms, weekday or weekend, etc.

Figure 13, rainy weather, shows better RL behavior in *OC* than we first expected just looking at the S_{NH} concentration (Figure 11). Although this fact is the most extreme for the rainy weather condition, it also happens for the rest of them. Thus, in the RL configuration, *OC* tops are higher than in the rest of the configurations and bottoms are always lower. What is more, the RL agent excels on weekends for every weather condition for the final phase (see Figure 14).

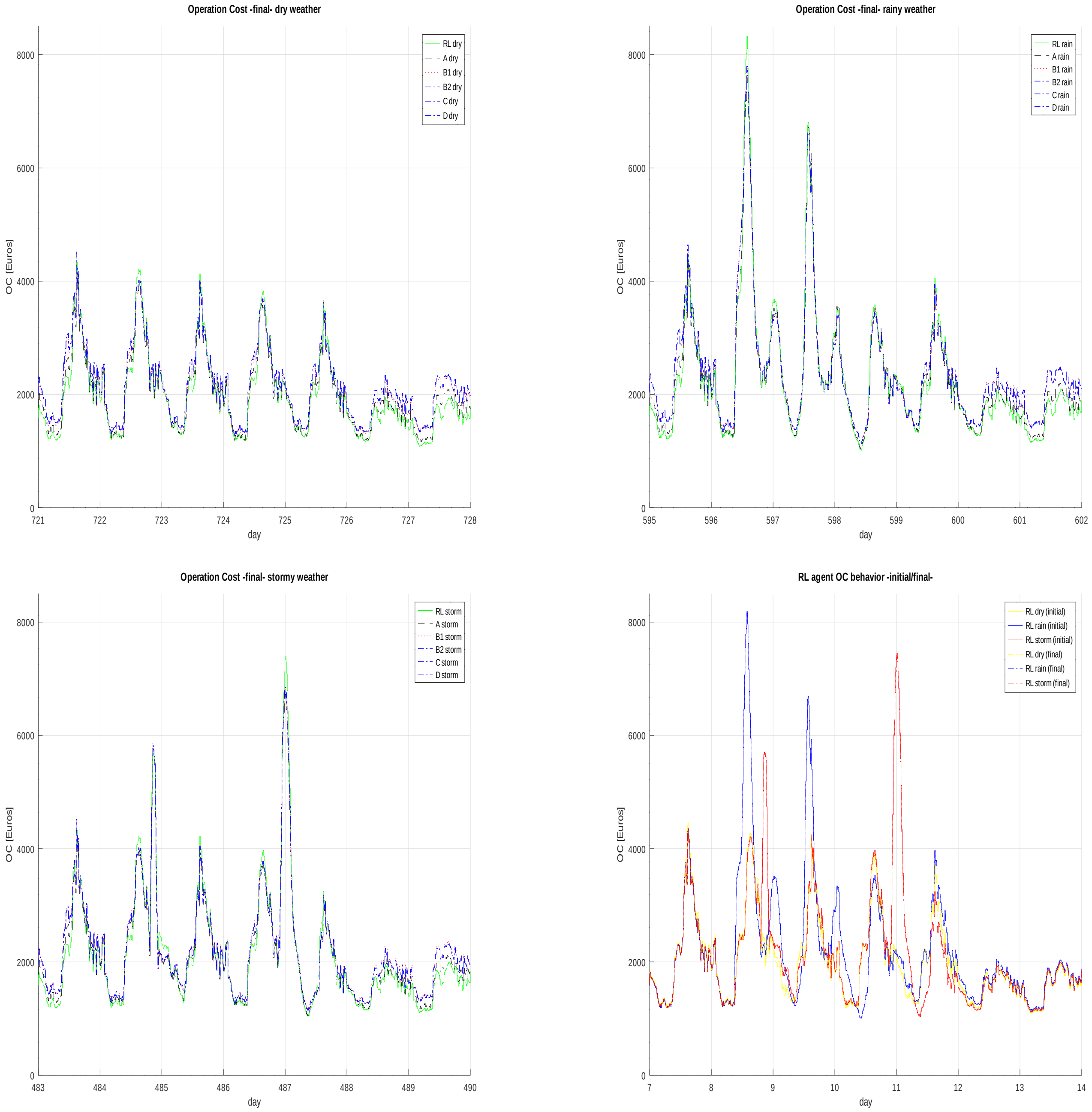

Figure 14. Operation Cost *OC* for the **final** phase.

4. Conclusions

The results of this paper show that it is possible to use an RL agent without an initial model in the control of WWTPs. In fact, a better performance is shown compared with the use of traditional PID configurations. This improvement is achieved because the RL agent can use adaptive control strategies to follow the changes in the influent. In this way, the behavior of the plant can be optimized to achieve its objective: minimizing the WWTP operation cost. The cost function considers two factors: the fines due to high concentrations of pollutants (mostly nitrogen from N-ammonia) and the energy consumption to supply oxygen to the plant.

In fact, the RL agent finds a balance between these opposing factors. In this way, it often allows the pollutants to exceed their limit to save the large expense of reducing their concentration. In addition, it can reduce the energy consumption of the blowers whenever it observes a low concentration of S_{NH}.

Although the so-called cold start problem is a challenge in the reinforcement learning area, in this paper, we did not focus on it because we just wanted to compare the evolution of the RL agent and compare different stages of the RL agent with the control methods. However, we are conscious of this problem and we have already worked on it [31]. As a first step, in order to observe the learning process of the agent, this work shows a simulation of the functioning of the RL agent in comparison with traditional PID approaches, in different weather conditions and how the agent can adapt to varying conditions. In general, the RL agent behaves differently on tops and bottoms of influent pollution concentration. In fact, on tops, the RL agent is more influenced by fines (because of the high concentration of pollutants to the effluent) while, on bottoms, the agent is more guided by the energy consumption. Finally, it is noteworthy that the RL agent performs especially well on weekends and in rainy and stormy weather conditions, where it greatly reduces consumption compared to the PIDs configurations. That is, due to the adaptative behavior of the RL agent, it can follow the changing environment better than the PIDs.

The results of this work show that RL technique can be applied with improvements as an alternative to classical WWTP control solutions. However, we consider that there is still a lack of application of this strategy in other areas, not only under simulation but also in the real plant.

Finally, we would like to note the BSM1 limitations. In future, we want to work on a more detailed model that, among other things, takes into account the inertia in turbine/motors of the blowers. In this case, we will measure aspects such as load rejection performance and see how this affects the OC, the different control methods and the RL agent's learning.

Author Contributions: Conceptualization, F.H.-d.-O.; Methodology, E.G. and R.D.; Software, F.H.-d.-O. and M.G.; Validation, F.H.-d.-O. and M.G.; Formal analysis, F.H.-d.-O. and E.G.; Investigation, F.H.-d.-O., N.D. and R.D.; Writing—original draft, F.H.-d.-O., E.G., N.D. and R.D.; Writing—review & editing, F.H.-d.-O., E.G. and N.D. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported, in part, by the Spanish Ministry of Economy and Competitiveness under Project CICYT PID2019-108377RB-C32, by the Spanish State Research Agency (AEI) under Project PID2020-112658RB-I00.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Influent data and more about BSM1 can be obtained at http://iwa-mia.org/benchmarking#BSM1 (accessed on 4 April 2023).

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

References

1. Dochain, D.; Vanrolleghem, P. *Dynamical Modelling & Estimation in Wastewater Treatment Processes*; IWA Publishing: London, UK, 2001.
2. Machineni, L. Review on biological wastewater treatment and resources recovery: Attached and suspended growth systems. *Water Sci. Technol.* **2020**, *80*, 2013–2026. [CrossRef] [PubMed]
3. Alex, J.; Benedetti, L.; Copp, J.; Gernaey, K.; Jeppsson, U.; Nopens, I.; Pons, M.; Rieger, L.; Rosen, C.; Steyer, J.; et al. Benchmark Simulation Model no. 1 (BSM1). Scientific and Technical Report, IWA Taskgroup on Benchmarking of Control Stategies for WWTPs, Department of Industrial Electrical Engineering and Automation. Lund University. 2008. Available online: https://www.iea.lth.se/publications/Reports/LTH-IEA-7229.pdf (accessed on 4 April 2023).
4. Metcalf, E.; Eddy, H. *Wastewater Engineering: Treatment and Reuse*, 4th ed.; McGraw-Hill Publishing: New York, NY, USA, 2003.

5. Olsson, G.; Nielsen, M.; Yuan, Z.; Lynggaard-Jensen, A.; Steyer, J. *Instrumentation, Control and Automation in Wastewater Systems*; IWA Publishing: London, UK, 2005.
6. Yang, T.; Qiu, W.; Ma, Y.; Chadli, M.; Zhang, L. Fuzzy model-based predictive control of dissolved oxygen in activated sludge processes. *Neurocomputing* **2014**, *136*, 88–95. [CrossRef]
7. Holenda, B.; Domokos, E.; Redey, A.; Fazakas, J. Dissolved oxygen control of the activated sludge wastewater treatment process using model predictive control. *Comput. Chem. Eng.* **2008**, *32*, 1270–1278. [CrossRef]
8. Hernández-del Olmo, F.; Gaudioso, E.; Dormido, R.; Duro, N. Energy and Environmental Efficiency for the N-Ammonia Removal Process in Wastewater Treatment Plants by Means of Reinforcement Learning. *Energies* **2016**, *9*, 755. [CrossRef]
9. Åmand, L.; Olsson, G.; Carlsson, B. Aeration control—A review. *Water Sci. Technol.* **2013**, *67*, 2374–2398. [CrossRef]
10. Du, S.; Yan, Q.; Qiao, J. Event-triggered PID control for wastewater treatment plants. *J. Water Process. Eng.* **2020**, *38*, 101659. [CrossRef]
11. Meneses, M.; Concepción, H.; Vilanova, R. Joint Environmental and Economical Analysis of Wastewater Treatment Plants Control Strategies: A Benchmark Scenario Analysis. *Sustainability* **2016**, *8*, 360. [CrossRef]
12. Lozano, A.B.; Del Cerro, F.; Lloréns, M. Methodology for Energy Optimization in Wastewater Treatment Plants. Phase III: Implementation of an Integral Control System for the Aeration Stage in the Biological Process of Activated Sludge and the Membrane Biological Reactor. *Sensors* **2020**, *20*, 4342. [CrossRef]
13. Iratni, A.; Chang, N. Advances in control technologies for wastewater treatment processes: Status, challenges, and perspectives. *IEEE/CAA J. Autom. Sin.* **2019**, *6*, 337–363. [CrossRef]
14. Wang, D.; Thunéll, S.; Lindberg, U.; Jiang, L.; Trygg, J.; Tysklind, M.; Souihi, N. A machine learning framework to improve effluent quality control in wastewater treatment plants. *Sci. Total Environ.* **2021**, *784*, 147138. [CrossRef]
15. Várhelyi, S.; Tomoiagă, C.; Brehar, M.; Cristea, V. Dairy wastewater processing and automatic control for waste recovery at the municipal wastewater treatment plant based on modelling investigations. *J. Environ. Manag.* **2021**, *287*, 112316. [CrossRef]
16. Santín, I.; Vilanova, R.; Pedret, C.; Barbu, M. New approach for regulation of the internal recirculation flow rate by fuzzy logic in biological wastewater treatments. *ISA Trans.* **2022**, *120*, 167–189. [CrossRef]
17. Pisa, I.; Morell, A.; Vilanova, R.; Vicario, J.L. Transfer Learning in Wastewater Treatment Plant Control Design: From Conventional to Long Short-Term Memory-Based Controllers. *Sensors* **2021**, *21*, 6315. [CrossRef]
18. Hansen, L.D.; Stokholm-Bjerregaard, M.; Durdevic, P. Modeling phosphorous dynamics in a wastewater treatment process using Bayesian optimized LSTM. *Comput. Chem. Eng.* **2022**, *160*, 107738. [CrossRef]
19. Meng, X.; Zhang, Y.; Qiao, J. An adaptive task-oriented RBF network for key water quality parameters prediction in wastewater treatment process. *Neural Comput. Appl.* **2021**, *33*, 11401–11414. [CrossRef]
20. Chen, Y.; Song, L.; Liu, Y.; Yang, L.; Li, D. A Review of the Artificial Neural Network Models for Water Quality Prediction. *Appl. Sci.* **2020**, *10*, 5776. [CrossRef]
21. Cao, W.; Yang, Q. Online sequential extreme learning machine based adaptive control for wastewater treatment plant. *Neurocomputing* **2020**, *408*, 169–175. [CrossRef]
22. Pang, J.; Yang, S.; He, L.; Chen, Y.; Ren, N. Intelligent Control/Operational Strategies in WWTPs through an Integrated Q-Learning Algorithm with ASM2d- Guided Reward. *Water* **2019**, *11*, 927. [CrossRef]
23. Bahramian, M.; Dereli, R.K.; Zhao, W.; Giberti, M.; Casey, E. Data to intelligence: The role of data-driven models in wastewater treatment. *Expert Syst. Appl.* **2023**, *217*, 119453. [CrossRef]
24. Qambar, A.S.; Al Khalidy, M.M. Optimizing dissolved oxygen requirement and energy consumption in wastewater treatment plant aeration tanks using machine learning. *J. Water Process. Eng.* **2022**, *50*, 103237. [CrossRef]
25. Dairi, A.; Cheng, T.; Harrou, F.; Sun, Y.; Leiknes, T. Deep learning approach for sustainable WWTP operation: A case study on data-driven influent conditions monitoring. *Sustain. Cities Soc.* **2019**, *50*, 101670. [CrossRef]
26. Jiang, Y.; Li, C.; Sun, L.; Guo, D.; Zhang, Y.; Wang, W. A deep learning algorithm for multi-source data fusion to predict water quality of urban sewer networks. *J. Clean. Prod.* **2021**, *318*, 128533. [CrossRef]
27. Ching, P.; So, R.H.; Morck, T. Advances in soft sensors for wastewater treatment plants: A systematic review. *J. Water Process. Eng.* **2021**, *44*, 102367. [CrossRef]
28. Yaqub, M.; Asif, H.; Kim, S.; Lee, W. Modeling of a full-scale sewage treatment plant to predict the nutrient removal efficiency using a long short-term memory (LSTM) neural network. *J. Water Process. Eng.* **2020**, *37*, 101388. [CrossRef]
29. Shi, S.; Xu, G. Novel performance prediction model of a biofilm system treating domestic wastewater based on stacked denoising auto-encoders deep learning network. *Chem. Eng. J.* **2018**, *347*, 280–290. [CrossRef]
30. Hernández-del Olmo, F.; Gaudioso, E.; Duro, N.; Dormido, R. Machine Learning Weather Soft-Sensor for Advanced Control of Wastewater Treatment Plants. *Sensors* **2019**, *19*, 3139. [CrossRef]
31. Hernández-del Olmo, F.; Gaudioso, E.; Dormido, R.; Duro, N. Tackling the start-up of a reinforcement learning agent for the control of wastewater treatment plants. *Knowl. Based Syst.* **2018**, *144*, 9–15. [CrossRef]
32. Copp, J. *The COST Simulation Benchmark: Description and Simulator Manual*; Scientific and Technical Report; Office for Official Publications of the European Community: Luxembourg, 2002.
33. Dorf, R.; Bishop, R.H. *Modern Control Systems*, 13th ed.; Pearson: London, UK, 2017.

34. Buşoniu, L.; Babuška, R.; De Schutter, B.; Ernst, D. *Reinforcement Learning and Dynamic Programming Using Function Approximators*; CRC Press: Boca Raton, FL, USA, 2010.
35. Stare, A.; Vrečko, D.; Hvala, N.; Strmčnik, S. Comparison of control strategies for nitrogen removal in an activated sludge process in terms of operating costs: A simulation study. *Water Res.* **2007**, *41*, 2004–2014. [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.