ICA2025OSLO
14th IWA Specialised Conference on
Instrumentation, Control and Automation

# Reinforcement learning boosting aeration control

Wambecq Tom *, Amerlinck Youri **, De Bock Birgit ***

* Dijkstraat 8, 2630 Aartselaar, Belgium, Tom.wambecq@aquafin.be

** Dijkstraat 8, 2630 Aartselaar, Belgium, Youri.amerlinck@aquafin.be

*** Dijkstraat 8, 2630 Aartselaar, Belgium, Birgit.debock@aquafin.be

**Abstract:** Aeration is a key element in a wastewater treatment plant (WWTP). As such, aeration control is essential for meeting the treatment targets, on one hand guaranteeing a good effluent quality but on the other hand reducing the energy consumption and related costs as low as possible. Although many control algorithms exist, in practice it still proves to be challenging to fine tune the algorithm to keep it stable under all conditions. In that sense, AI and in particular Reinforcement Learning are promising technologies. Within this project a Reinforcement Learning (RL) algorithm was implemented in the Python package Gymnasium, creating a custom environment for optimizing aeration in a wastewater treatment plant. This custom environment simulates, in a full-fledged dynamic WWTP model, a two-hour cycle where the aeration time can be adjusted to stabilize ammonium concentration in the effluent. The RL algorithm was trained using Stable-Baselines3, particularly the Soft Actor-Critic (SAC) policy method. The RL leverages a neural network with two layers of 64 nodes each. Key measurements include influent flow, ammonium and nitrate concentrations, temperature, and aeration time (expressed as percentage of the cycle time). The goal is to maintain ammonium between 0.7 mg/L and 1.3 mg/L, earning discrete rewards for meeting this range, with penalties for exceeding 15 mg/L. Training begins with random weights and iteratively refines strategies based on rewards obtained from simulations. Over 12 hours and 100,000 time steps (simulating 45 years), the AI learns system dynamics and optimizes aeration control, significantly reducing ammonium variations. This allows the RL to adapt to system deviations dynamically, producing a more robust and realistic control strategy. A feedback loop ensures better model fit and minimizes errors in real-world implementation, enabling the RL to handle system variability more effectively. This dynamic approach enhances both accuracy and practicality in optimizing wastewater treatment processes. Results proof that the chosen method outperforms traditional calibration approaches by incorporating a feedback loop and as such is the method ready for application on a real-world WWTP as next step.

**Keywords**: AI; Control; Wastewater treatment plant; Reinforcement learning, activated sludge model

Aeration control is essential at a well-managed wastewater treatment plant (WWTP). It provides the mean to meet the treatment objectives under all circumstances while minimizing the energy consumption and costs. Many control strategies have been developed and implemented over the years, from simple on-off controls, over PID to model predictive control. Although many of these control strategies have proven to improve the WWTP results, fine tuning those controllers still proves to be a challenge in practice [ref 1].
In this project, the python package “gymnasium” [ref 2] is used to implement a reinforcement learning (RL) algorithm. In this package a certain environment can be created by the user. This environment, here a WWTP model including the aeration cycle, will have outputs (measurements) and the user can influence it by means of actions. For example: in a moon lander environment we get the speed and location as measurements and as actions : do we want to give propulsion, go left or right,… In an environment we want to achieve a certain goal (like landing the moon lander) and rewards will be given depending on how close we get to this goal. When we get out of the boundaries of our environment (like a moon lander crashing), the environment crashes and no extra rewards can be obtained anymore. For each time step, we can adjust the actions and run the entry and try again. However, an option is to let an AI learn an optimum strategy to achieve the highest reward.

For our case, the environment will be a wastewater treatment plant model (for schematic overview see figure 1). The goal is to optimize the aeration to get a stable ammonium concentration in the effluent. In the plant, there is an intermittent aeration with a fixed 2-hour cycle time. The action which is taken at each step (every 2 hours) is to determine the percentage of the total cycle time, which will be used for aeration during the next cycle. E.g. a percentage of 60%, represents an

aeration phase lasting 72 minutes (60%*2h) and consecutively an unaerated (denitrification) phase lasting 48 minutes (2h – 72 min). During the aeration phase, within the aeration basin, the oxygen setpoint is reached using a fixed PID controller. The measurement values utilized are the past 2 day (i.e. 24 2-hour-average) simulated values of the influent flow, ammonium concentration, nitrate concentration, temperature and the aeration percentage.

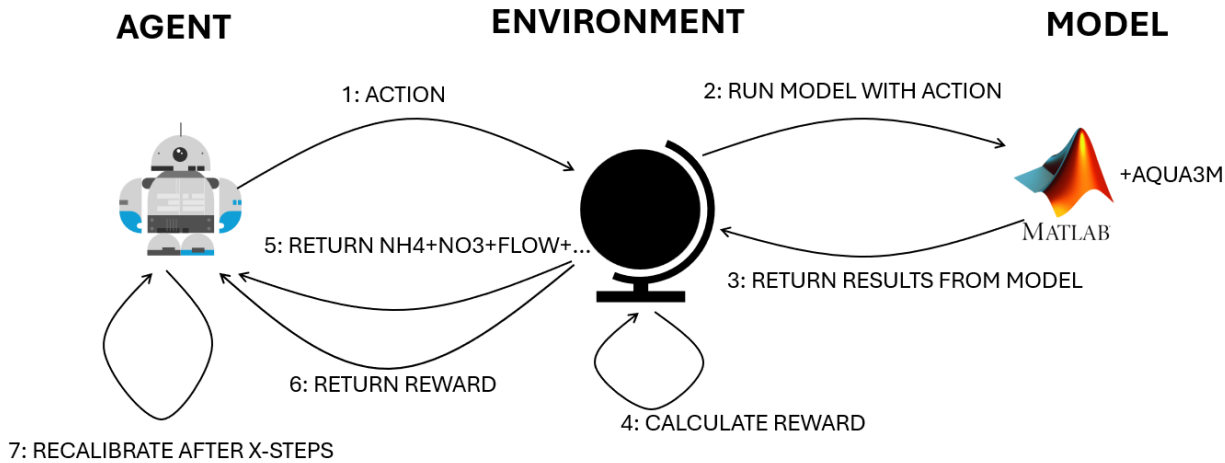


**Figure 1** Schematic overview of the interaction between the agent (stable-baseline 3), the environment (gymnasium) and Matlab

The target of the control was to keep the 2-hour average ammonium concentration between 0.7mg/l and 1.3mg/l. In view of the RL, a reward of 1 was given for every previous timestep where this condition is met. In addition, instantaneous ammonium concentration is not allowed to surpass 15mg/l, which receives a penalty in view of the RL. A system with continuous rewards would also be possible but the best results were obtained using this discrete reward system.

For the optimization, the RL was trained along with the model. This RL was developed in stable_baseline3 [ref 3]. The best results were obtained from the soft actor critic (SAC) policy method, however different policies were also tested (Deep Q Learning, Proximal policy optimization,…). The RL algorithm contains a machine learning neural network with 2 layers of each 64 nodes. The algorithm starts with random weights and by running the WWTP model within the environment, it receives rewards. From these rewards a new updated model can be calibrated. This new model is tested and small perturbations are added to see if an even more optimal strategy could be devised. The calibration step is repeated and the whole process starts again. The training of the algorithm takes 12 hours real time in which 100,000 time steps (45 years of simulation time) are simulated. In the beginning, up to about 10,000 time steps (Figure 2), a big variation in (the 2h average) ammonium concentration is seen, while after the full calibration and training, an almost stable signal is observed (Figure 3). The RL learned the act upon the characteristics and the variations of the system.

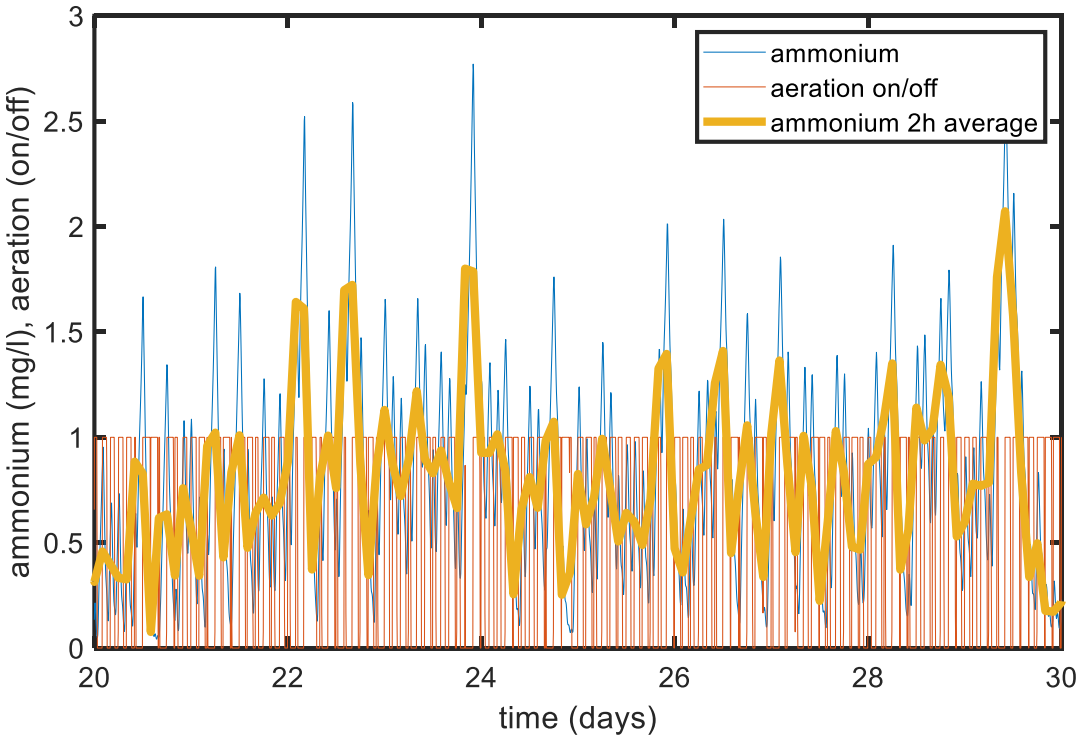


**Figure 2** The ammonium, 2-hour average ammonium and aeration phase as function of the time after 10.000 time steps of calibration

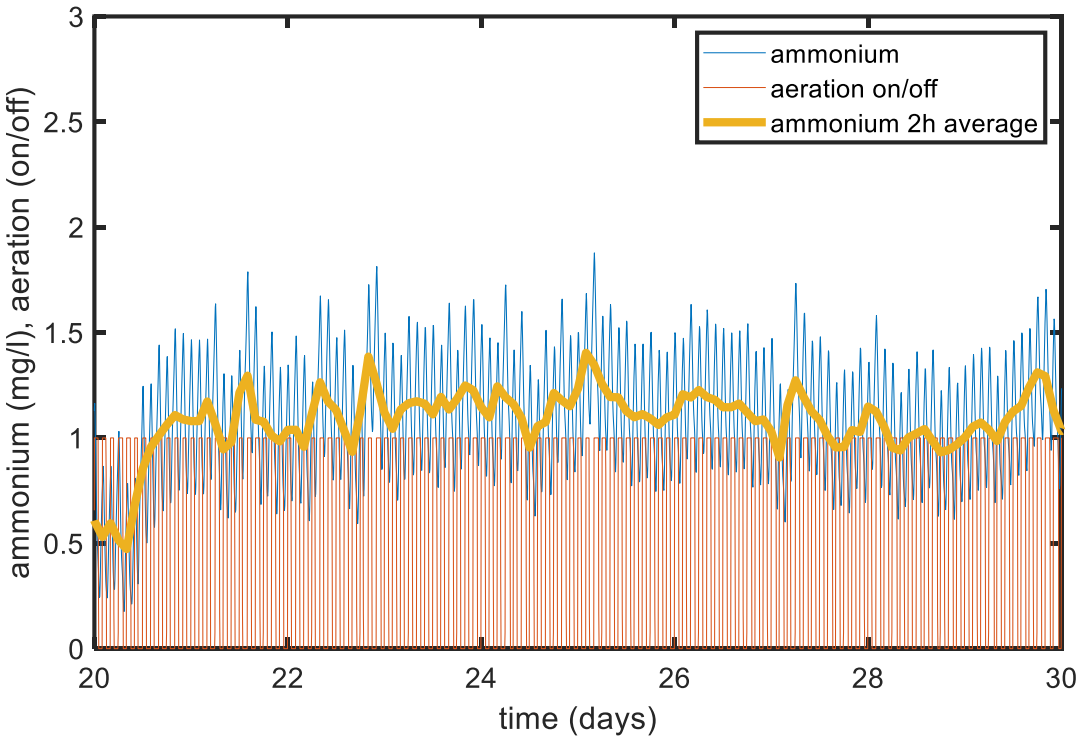


**Figure 3** The ammonium, 2-hour average ammonium and aeration phase as function of the time after 100.000 time steps of calibration

For calibrating a neural network to control the aeration, another option would be to use the WWTP model once and calibrating a 'perfect' neural network on these data. The disadvantage of this

approach would be that a small deviation could cause the system to behave differently than how the data were created and therefore create a non-optimal system. This disadvantage is circumvented in the presented approach, thanks to the feedback loop, where a small deviation worsening the effluent results, leads to less rewards. With the same amount of data generation/simulated time by the WWTP model, the presented approach incorporates extra information to train the model more accurately and understand the underlying system. The feedback loop results in a better fit and mimics the reaction in reality to a greater extent. As such, it will thus aid to implement this control on the real WWTP as less error is expected.
The project demonstrated that applying Reinforcement Learning in combination with a full-fledged dynamic WWTP model, delivers superior results to prior implemented (fine tuning of) control strategies. In a future project, this method will be applied to model and control a real wastewater treatment plant.

## REFERENCES

**[1]** De Gussem K., Wambecq T., et al 2011 Cost Optimization and Minimisation of the Environmental Impact through Life Cycle Analysis of the Waste Water Treatment Plant of Bree (Belgium). *Water Science & Technology* ***63****(1), 164-70*.

**[2]** https://gymnasium.farama.org/index.html

**[3]** https://pypi.org/project/stable-baselines3/