# Welsh Water Innovation Event 2025

# Multi-Objective Reinforcement Learning for the Control of Storm-water Systems under Distributional Shift

Daisy Welham[1], Sara Sharifzadeh[1], Chedly Tizaoui[2], Liam Butler[3]

[1]Swansea University Department of Computer Science, [2]Swansea University Department of Engineering, [3]Dwr Cymru

## Technical Challenge

Urban Water Systems (UWS) often require trade-offs between heterogeneous objectives:- for example, minimizing pumping costs subject to not causing a tank to overflow. Existing programmable logic controllers (PLC) represent an acceptable way to trade-off the different objectives, but these rely on knowledge from domain experts and may not necessarily find the optimum trade-off.

Reinforcement learning (RL) is known to perform well at optimizing simple objectives in simple environments, but can encounter issues when either the environment or the objective is too complex or inconsistent.

This work focuses initially on training reinforcement learning agents to control a simple pump in a stable environment with multiple heterogeneous objectives. Then, once the agent is trained in some simple environment, we deploy it in a more complex environment where shifts between inflow patterns are frequent.

We train Q-learning and actor-critic agents, as well as a majority-voting ensemble method that uses both the other agents as well as a PLC controller. We show that the ensemble method reduces pumping costs by 30% compared to the PLC controller and also reduces overflow rates from 2.1% to 0.27% in high-rain conditions compared to the use of actor-critic alone.

## Objective Function Design

We define objective functions and combine them. We can define the risk-avoidance reward for picking an action which turns the system from state $s_m$ to $s_n$ as:-

$$R_r = L_m - L_n \tag{1}$$

-:where $R_r$ is the risk-avoidance reward and $L_i$ is the water level in state $s_i$. The financial cost of pumping is given by:-

$$R_c = (L - H_P) \cdot P \cdot C_1 \tag{2}$$

-:where $R_c$ is the reward for reducing operational costs and $P$ is 6 if the pump is "on" and 0 if it is 'off'. So we needed a reward function which maximises $R_r$ and $R_c$. One natural way to do this is to maximise:-

$$R_k = R_r + \lambda R_c \tag{3}$$

-:where $\lambda$ represents the appropriate trade-off between risk avoidance and financial savings. If $\lambda$ is too high then the agents will overflow the tank to save financial cost, and if it's too low the agents will always keep the pump on to minimise risk of overflow. So we make the modification that $\lambda$ scales with the water level, i.e.:-

$$\lambda(L) = k_\lambda(H - L) \tag{4}$$

## Main Contributions

The main contributions of this work are:-

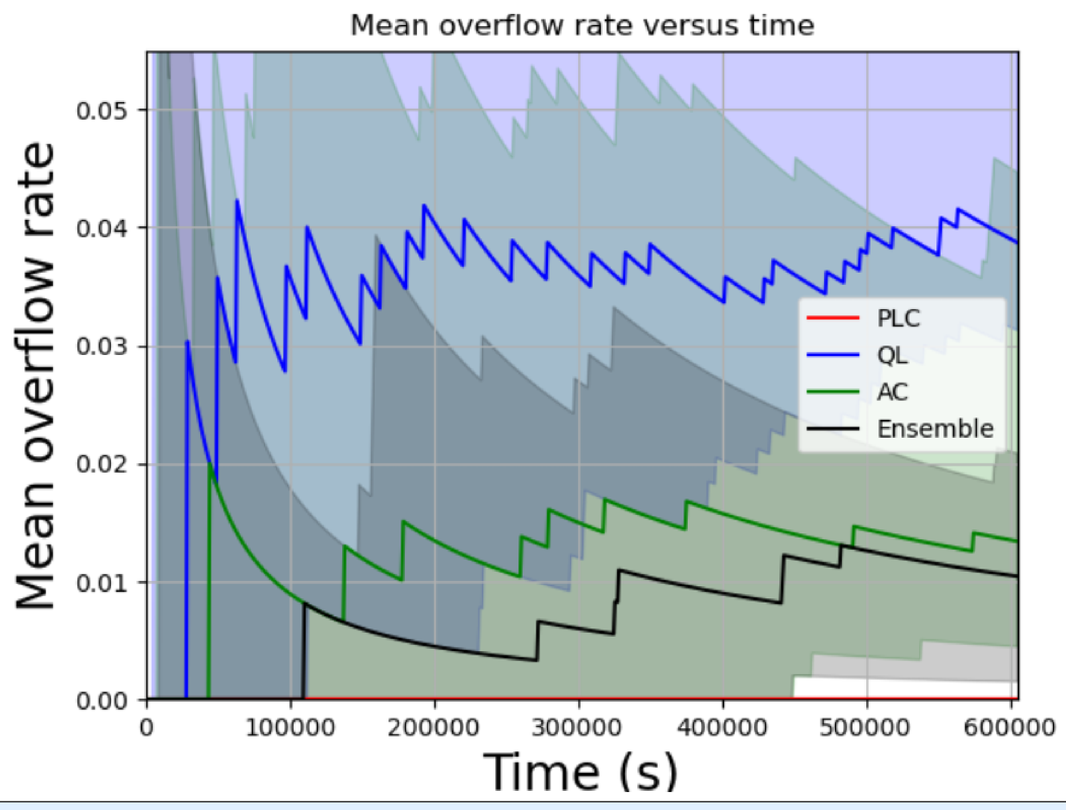
- **Heterogeneous objectives:** We apply the RL agents to a multi-objective optimization problem with heterogeneous objectives. It is hoped that progress in one domain (e.g. storm-water systems control) may have applications to similar domains (e.g. sewage treatment) due to the heterogeneous nature of their objectives.
- **Ensemble Methods** The application of a majority-voting ensemble method in this domain is novel and succeeds in reducing both operational costs and overflow rates.
- **Distributional Shift** We highlight the issue of capability robustness failure under distributional shift and propose the use of an ensemble method as a potential way to reduce (though not entirely eliminate) this problem.

## Results

| Agent | Values | Normal | Low Rain | High Rain |
|---|---|---|---|---|
| PLC | Water Level (m) | 14.9 ± 6.3 | 14.0 ± 5.7 | 15.7 ± 6.7 |
| | Cost | 5.2 ± 0.9 | 2.9 ± 0.6 | 7.3 ± 1.1 |
| | Overflow Rate | 0% | 0% | 0% |
| QL | Water Level (m) | 23.4 ± 3.6 | 21.9 ± 3.2 | 24.3 ± 4.0 |
| | Cost | 3.6 ± 0.6 | 1.8 ± 0.4 | 4.4 ± 0.7 |
| | Overflow Rate | 0% | 0% | 0.36% |
| AC | Water Level (m) | 23.8 ± 3.6 | 22.3 ± 3.2 | 24.9 ± 4.2 |
| | Cost | 3.6 ± 0.6 | 1.8 ± 0.4 | 4.3 ± 0.7 |
| | Overflow Rate | 0% | 0% | 2.1% |
| Ensemble | Water Level (m) | 23.3 ± 3.6 | 21.9 ± 3.2 | 24.3 ± 4.0 |
| | Cost | 3.7 ± 0.6 | 1.8 ± 0.4 | 4.5 ± 0.7 |
| | Overflow Rate | 0% | 0% | 0.27% |

TABLE I

RESULTS OF THE TESTS REPORTING THE MEAN WATER LEVEL (IN METRES) ± STANDARD DEVIATION, THE MEAN COST, AND THE MEDIAN OVERFLOW RATE.

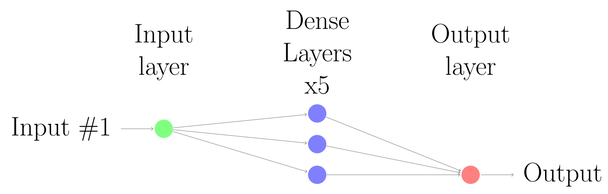



## Architectures

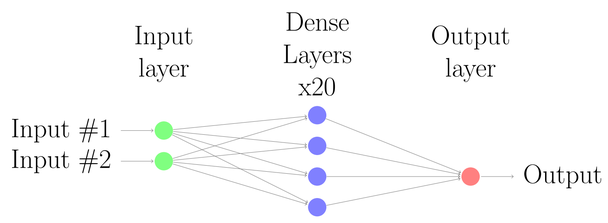
We implement the PLC controller as an if-then program:- if the water level in a tank is below a $10m$ the pump turns fully off, if it is above $25m$ the pump turns fully on, and otherwise the pump remains in whatever state it is currently in.

The Q-Learning agent discretizes pump activations to the nearest 10% and the water level to the nearest $0.1m$. Q-values are initialised to 0 and updated using Bellman.

The actor is a dense neural network with 5 dense hidden layers. It takes a water level as input and gives a pump activation as output. The critic takes a pair composed of a water level with an action, has 20 dense hidden layers, and outputs an evaluation for how good the chosen action is.




The critic has an input layer of size 2, 20 dense layers, and an output layer of size 1. The inputs are the water level and a possible action, and the critic is trained to output what the reward model (see "Reward Model" section) will give as a reward for taking that action from that level




The ensemble method uses majority-voting of PLC, Q-learrning, and the actor to pick an action.