

## Artificial intelligence for decentralized water systems: A smart planning agent based on reinforcement learning for off-grid camp water infrastructures

Christos Makropoulos \* and Dimitrios Bouziotas 

Urban Water Management and Hydroinformatics Group, Department of Water Resources and Environmental Engineering, School of Civil Engineering, National Technical University of Athens (NTUA), Iroon Politechniou 5, 157 80 Zografou, Athens, Greece

\*Corresponding author. E-mail: cmakro@mail.ntua.gr

 CM, 0000-0003-0308-4265; DB, 0000-0003-2172-350X

### ABSTRACT

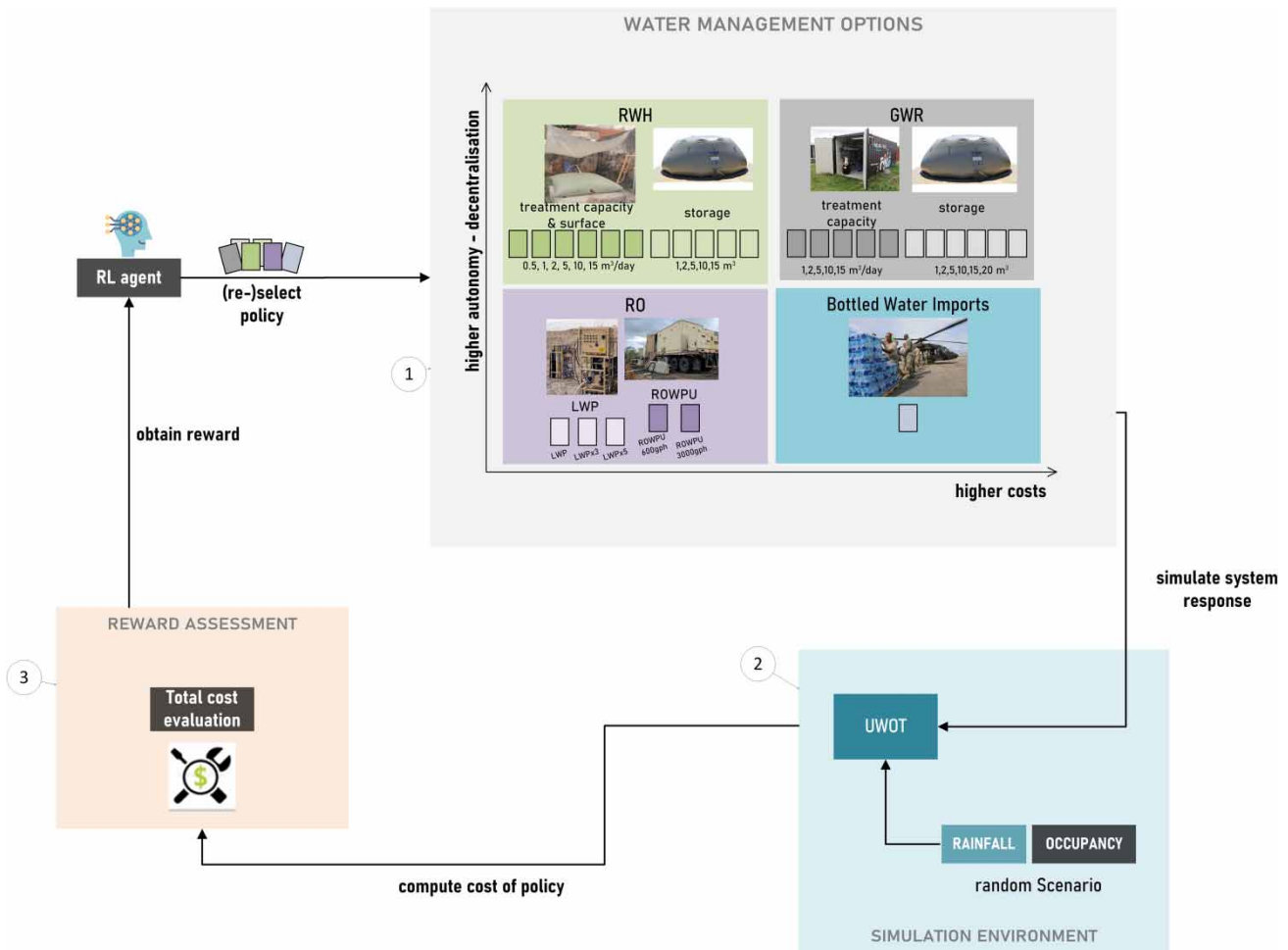
The planning and management of decentralized technologies in water systems is one of the promising, yet overlooked, domains where artificial intelligence (AI) can be successfully applied. In this study, we develop and deploy a reinforcement learning (RL)-based 'smart planning agent' capable of designing alternative decentralized water systems under demanding operational contexts. The agent's aim is to identify optimal water infrastructure configurations (i.e., proposed decisions on water management options and interventions) for different conditions with regard to climate, occupancy and water technology availability in a demanding, off-grid setting, i.e., a water system with high requirements of independence from centralized infrastructure. The agent is coupled with a source-to-tap water cycle simulation model capable of assessing and stress-testing the proposed configurations under different conditions. The approach is demonstrated in the case of a military camp deployed abroad for peacekeeping operations. The agent is tasked with selecting optimal interventions from an array of real-world camp water management technologies and evaluating their efficiency under highly variable, operational conditions explored through simulation. The results show that RL can be a useful addition to the arsenal of decision support systems (DSS) for distributed water system planning and management, especially under challenging, highly variable conditions.

**Key words:** artificial intelligence, decentralized technologies, deployed camps, reinforcement learning, simulation, water infrastructure

### HIGHLIGHTS

- We present a design framework coupling a water cycle simulation model with a reinforcement learning (RL) agent to support the planning of distributed water infrastructure.
- The framework is demonstrated in an off-grid camp and is tested against a highly variable environment with multiple external drivers, including climate, occupancy and technology availability.
- The results show that RL agents can be a useful planning aid for water infrastructures under deep uncertainty.

## GRAPHICAL ABSTRACT



## 1. INTRODUCTION

Artificial intelligence (AI) has triggered a paradigm shift in multiple scientific fields, water included, with numerous applications ranging from hydrology to water quality management (Doorn 2021; Tyrallis *et al.* 2021). AI applications in the water domain have seen rapid development in recent years (Savic 2019), with most cases focusing on exploratory data analyses and operational water management, including, for example, forecasting and water system control (Solomatine & Ostfeld 2008). In this context, tactical and strategic levels of water resource management, such as water systems planning, have received less attention to date with even fewer applications exploring the role of AI in decision-making practice (Hadjimichael *et al.* 2016). In this study, we demonstrate an AI application at the tactical and strategic level by supporting the selection of alternative, distributed water management technologies for a closed community (a deployed camp). The specific AI algorithm demonstrated is reinforcement learning (RL), a promising type of agent known for its efficiency and adaptability in highly variable environments, coupled with a previously developed water cycle simulation model, Urban Water Optioneering Tool (UWOT), that has been successfully employed in several strategic water resource management studies (e.g., Rozos & Makropoulos 2013; Bouziotas *et al.* 2019). The novelty of this work is that UWOT, in the present context, is not run as a user-driven planning model but acts as a simulation testbed, helping the AI 'learn' what constitutes optimal selection policies for distributed water infrastructures. While UWOT has been applied in diverse water systems before, this study constitutes the first case of this model being coupled with RL agency in a combined simulation-optimisation problem, as well as one of the first applications of RL agency in water planning under uncertainty in general.

Moreover, the study presents an application of RL algorithms for real-case tactical and strategic planning that is novel in literature, as most RL applications for water focus on operational contexts and simple topologies (Wang *et al.* 2021). As RL algorithms have not always converged or led to optimal policies for complex water systems (Mullapudi *et al.* 2020), the applicability of this AI technology to real planning cases remains to be explored. The approach is demonstrated in a particularly challenging case: that of a stand-alone (off-grid) water system with a high degree of decentralization and increased requirements of autonomy. This is a context found, for example, in isolated communities and military or refugee camps and disaster relief facilities, where decentralized interventions can be used to increase the robustness of water supply, reduce dependence on external sources that can be costly, sparse or simply unavailable, and increase overall system resilience (Bouziotas *et al.* 2019; Van de Walle *et al.* 2022). Such an off-grid system is of particular interest, not only because it allows the application of multiple decentralized supply technologies in a standardized (modular) format, but also because it must be deployed anywhere, under varying climates and operational characteristics, thus posing a challenge for efficient water supply management under uncertainty. The goal of the application is to increase water efficiency and reduce relevant supply costs of said deployable camp, given the constraints of:

- A set of demand patterns and corresponding potable and non-potable water requests dictated by the camp characteristics and operations.
- An array of different available supply technologies and relevant costs, including reverse osmosis (RO) units, imports of bottled water, and circular, decentralized onsite options such as rainwater harvesting (RWH) and greywater recycling (GWR). These technologies reflect real supply options available to a camp water planner when designing applications.

To explore multiple conditions of deployment in the off-grid system, we train the RL algorithm using the simulation capabilities of UWOT and assume a simulation environment characterized by high variability in terms of climate and rainfall characteristics, occupancy, and the availability of water infrastructure options.

## 2. MATERIALS AND METHODS

### 2.1. RL agents

Traditionally, research for optimal water supply infrastructure management has focused on operational and tactical settings (Mala-Jetmarova *et al.* 2017) using mathematical methods such as linear and non-linear programming (Price & Ostfeld 2014) and, in recent decades, evolutionary algorithms (Maier *et al.* 2014). Most applications focus on specific issues related to central supply systems, such as pump scheduling and leakage control or water quality control (Mala-Jetmarova *et al.* 2017). Applications for off-grid settings (e.g., deployed camps) are nascent for water management, with most optimization works in literature focusing on off-grid energy system design (Twaha & Ramli 2018); to the authors' attention, only one study recently employed optimization in off-grid water supply using linear programming (Abdullah & Gunal 2022). Moreover, fewer applications handle optimization in cases with high complexity, as many of the conventional optimization approaches require the strict formulation of the problem at hand and cannot take into account inherent uncertainties in the water system, e.g., in future water demands or natural weather variability (Maier *et al.* 2014; Mala-Jetmarova *et al.* 2017). There is therefore strong potential for the application of more recent advances in optimization (including AI algorithms) in highly variable settings such as off-grid systems. Among several AI algorithms that have been developed in recent years, RL has attracted significant attention for its success in solving high-complexity problems, becoming a popular choice in fields such as industrial process control, real-time scheduling and optimization, and autonomous vehicle guidance (Nian *et al.* 2020). In water, RL is typically applied in operational contexts, for instance, in optimal reservoir operations and water system control (Castelletti *et al.* 2002; Bhattacharya *et al.* 2003; Wang *et al.* 2021), with more limited exposure in tactical and strategical water management (Mason *et al.* 2016). The field is open to more complex applications, as most of the discussed applications with RL methods remain relatively simple and small, with just a few control assets (Wang *et al.* 2021). The fundamental idea in RL is to design an agent that learns via experiential interaction with a (real or simulated) environment. The agent, upon being exposed to the environment, is called to take an action or a sequence of actions that yield (cumulative) reward. Through repeated environmental exposure, the agent has the goal of maximizing his reward by learning which actions are optimal. A key difference with other AI optimization algorithms is that RL learns by repeated experiential (self-)learning, thrives in settings without prior guidance, knowledge or exposure to environmental information and adapts efficiently to randomized and noisy environments (Silver *et al.* 2018). Over the years, different families of RL algorithms have been developed and are

applied in different contexts, including for instance (deep) neural network-assisted Q-learning, associative search algorithms (also known as multi-armed bandits) and dynamic programming methods (Van Hasselt *et al.* 2016; Nian *et al.* 2020).

To choose the appropriate RL agent for policy planning and management, several potential RL algorithms were tested using a simplified version of the topology described in this work. Specifically, three algorithmic families were tested: deep Q-learning, multi-armed bandits and Proximal Policy Optimization (PPO) algorithms (Kuleshov & Precup 2014; Van Hasselt *et al.* 2016; Schulman *et al.* 2017). Specifics on these initial tests with regard to the used software and hardware, optimization time and overall convergence are shown in Supplementary Material, Appendix A. Following test runs, it was observed that deep Q-learning exhibited problems of convergence and instability, while multi-armed bandits, although more consistent, possessed a very slow convergence speed (see Supplementary Material, Appendix A) to be efficient enough for practical applications. Hence, it was observed that PPO algorithms were the most efficient: while relatively slow (with RL learning taking 18–24 h to run on average), these algorithms displayed robustness and a steady convergence towards optima across multiple simulations and within randomized environments. These instabilities in some algorithmic families are in line with reports on RL limitations discussed in literature (Nian *et al.* 2020; Wang *et al.* 2021), indicating that there is no ‘silver bullet’ with regard to RL algorithms and that the usefulness and stability of RL agents may vary depending on the application.

## 2.2. Water cycle simulation

RL algorithms can be trained in real settings or using synthetic datasets and simulation environments. In this study and to allow the agent to adapt to multiple, variable environments, we employ simulation as an AI testbed and use a water cycle model, the UWOT (Rozos & Makropoulos 2013), to mimic the response of the water system in terms of demands, water volumes supplied and runoff. The UWOT is a water cycle model of the metabolism modelling type, capable of simulating the complete water cycle from tap to source by modelling individual water uses and technologies/options for managing them at multiple scales with a bottom-up approach, starting from the micro-scale (individual water appliances) and going up to a neighbourhood, city or regional scale (Bouziotas *et al.* 2019). By using a topology of inter-connected demand and supply components, the UWOT can simulate multiple water cycle flows at a daily timescale, i.e., potable water demand and supply, generated wastewater and runoff, as well as their integration in terms of harvesting, reuse and recycling at different scales. The model follows a signal-based approach that builds demands from the bottom up, propagating them towards the source of water demands, i.e., the central drinking water network or any other central sources. The UWOT then matches demands with supply from different source types (Rozos & Makropoulos 2013) and logs time steps when a failure occurs, i.e., the demand cannot be met by provided supply.

In the case of a deployable camp, a simulation model is created that includes aspects of camp water demands at a fine level, i.e., at the level of camp appliances (lavatories, sinks, and washing machines) and nodal water needs (water usage by individuals, vehicle washing, etc.). To make results realistic and relevant to actual deployed camps, we rely on relevant NATO documentation describing demand planning specifics for deployable camps (USACE 2008), assuming a standard mechanized infantry company with a unit of 150 troops and 15 armoured vehicles. This is a tactical-sized unit which can also be used by the model as a baseline to simulate larger units or to introduce arbitrary extra demands, for instance, the need of aircraft washing, decontamination water, etc. For the purposes of this work, the camp is assumed to have, as its primary goal, a supportive (non-combat), training and advisory mission, and its water demands include, as per USACE (2008): (a) personnel needs, including drinking demands, personal hygiene and centralized hygiene, (b) food preparation demands, requiring potable water for ingredients washing and general food provision, (c) medical unit demands, (d) a common laundering unit, (e) vehicle washing and maintenance and (f) required demands for camp construction and maintenance. Based on these categories, total daily demands per person and per vehicle are calculated (see Table 1).

Based on the different demand categories and the planning guidelines on water supply quality characteristics, we assume that the camp relies on two main water quality standards for its planning, which also define the available sources (as per the NATO standard):

- potable quality standards, which describe uses such as drinking water, hygiene and food preparation. These standards have a strict source policy – they can be either from RO units, secured central supply or from bottled water imports for safety reasons.

**Table 1** | Overview of the unit water demands for the case study

		Per capita demand		Total camp demand	
Persons $n$	150	Personnel (L/person/day)	177.4	Personnel (m <sup>3</sup> /day)	26.6
Vehicles $m$	15	Vehicle (L/vehicle/day)	162.2	Vehicle (m <sup>3</sup> /day)	2.4

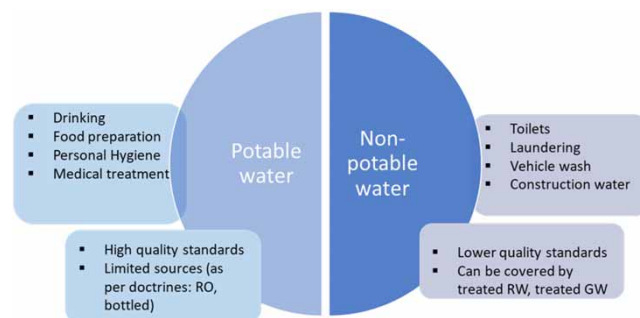
- non-potable quality standards, which describe non-consumptive uses such as toilet, laundering, vehicle wash and construction uses. These uses have a more relaxed policy, allowing water from (light) greywater sources, such as locally collected, recycled or reclaimed water, to be used as well.

A schematic overview of these two quality types and the associated demands per type is given in [Figure 1](#).

Besides demands, the UWOT also takes into account aspects of the camp water supply, which are selected based on both relevant NATO documentation ([NATO 1994](#); [USACE 2008](#)) and from the wider technology market for deployable, modular systems. The options considered include bottled water imports (standard in many real camp settings), standardized RO units used by NATO members, access to central infrastructure or onsite groundwater abstractions (if any), and decentralized interventions such as RWH and GWR. To standardize the latter options and present them as modular units to be used by the RL agent, we assume that RWH is available as a modular unit, similar to the one presented by [Nguyen et al. \(2013\)](#), comprising a rainwater collection surface that is linked to a storage unit (onion tank) underneath. This modularity allows the agents to deploy one or many of these units at every simulation run. Similarly, we assume a containerized GWR unit with set capacity, bearing similarity to solutions available in the market ([Jotem 2020](#)). Besides sources, storage units are included as part of the supply system as self-supporting, deployable (onion) tanks with a set capacity.

As a final pre-processing step, the UWOT also needs time-series of rainfall and occupancy patterns to produce demands. We randomized rainfall regimes and occupancy patterns using stochastically generated climate and variable occupancy settings, so that every simulation run represents unique camp deployment conditions for the agent. This stochastic nature exposes the agent to multiple, highly variable environments at each run, leading to more resilient options. To enable the exploration of scenarios based on realistic climatic conditions, historical rainfall data were obtained from the KNMI data platform ([KNMI 2022](#)) for two different climates: (i) temperate climate conditions and (ii) dry climate conditions, and a random subset of the rainfall time-series  $t$  equal to 10 days is used at every deployment, with the length value being selected as a trade-off between steady convergence rates and long simulation times (seen when  $t > 10$  days). A key research question of interest in this context is if and how these different climates lead the RL agent to select different water infrastructure options.

Occupancy variability is treated by assuming two occupancy scenarios: (i) a steady occupancy scenario, where the camp is filled with its planned occupancy (see [Table 1](#)), and remains constant across the simulation period, (ii) a variable occupancy scenario with randomly generated patterns. This reflects camp missions where personnel and vehicles need to leave the camp ( $n_{out}$  and  $m_{out}$ , respectively), following a set probability of absence per day ( $p_p$  and  $p_v$ , respectively). Troops and vehicles outside the camp are assumed not to contribute to the water demands of the camp (with the troop potable water demands being covered by bottled water). In real deployments, the randomness factors for the variable occupancy scenario ( $n_{out}$ ,  $m_{out}$ ,  $p_p$ ,  $p_v$ )

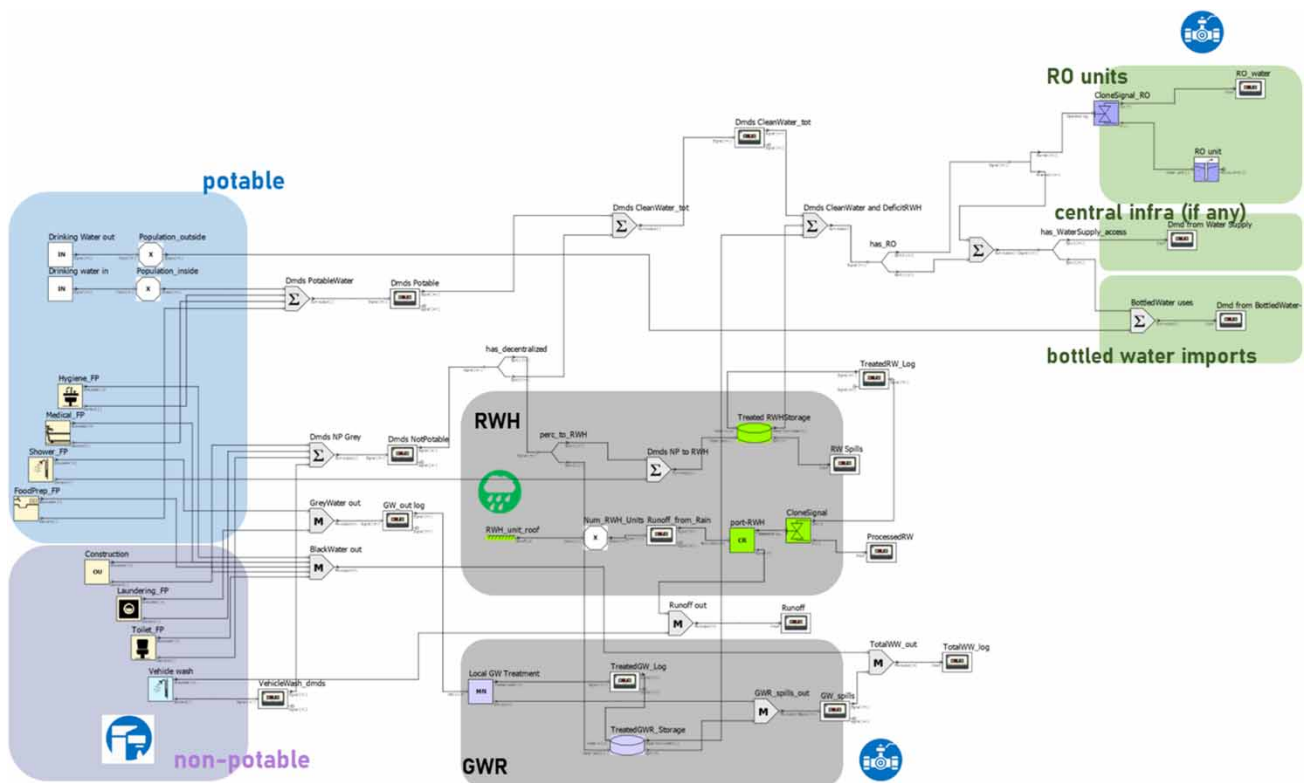
**Figure 1** | Different water quality categories and the corresponding mapped water needs assumed for the case study.

could be derived from the operational characteristics of the camp; for this case, it is empirically assumed, following feedback from real (undisclosed) cases, that 30% of personnel and 20% of the vehicles need to be outside the camp with an independent daily absence probability of  $p_p = p_v = 40\%$ .

### 2.3. Coupling of the water cycle model with the RL agent

Following the definition of different demand and supply options and variable conditions, the corresponding (signal-based) UWOT topology that represents the entire source-to-tap model of the deployable camp is schematized (Figure 2). The model generates demands from the tap level (left part of Figure 2) and proceeds to match them with different supply options (right part of Figure 2) coupled with potential decentralized (RWH/GWR) interventions. Two demand streams based on different water quality characteristics (potable and non-potable) are generated, with the latter being exposed also to decentralized supply options (according to demand coverage principles shown in Figure 1). To simplify the coupling with the RL agent, one topology including all possible supply options is generated; the availability of different supply options can be controlled by special ‘switch’ components (noted as splitting signals in Figure 2), which receive a binary input {0, 1}, indicating whether a particular supply option is available (1) to meet demands or not (0). This input can either (a) be set a priori as a parameter controlling the specific environment (e.g., the availability of RO units set before training to 0, reflecting deployment settings where the water manager does not have access to RO), or (b) can be controlled by the RL agent as part of its decision space, so that the RL agent can choose if a particular supply option is deployed or not as part of its learning.

Based on the source-to-tap model topology and prior to model simulation, a selection of supply options is exposed as a decision (action) space to the RL agent, so that the agent can choose combinations of supply options (turning certain supply options on or off) as a water planning and management policy for a particular deployment. This means that the agent acts as a camp water ‘planner and designer’, choosing a specific policy and, iteratively through its learning process, explores combinations of different policies under specific camp deployment conditions and ‘learns’ from this experience.

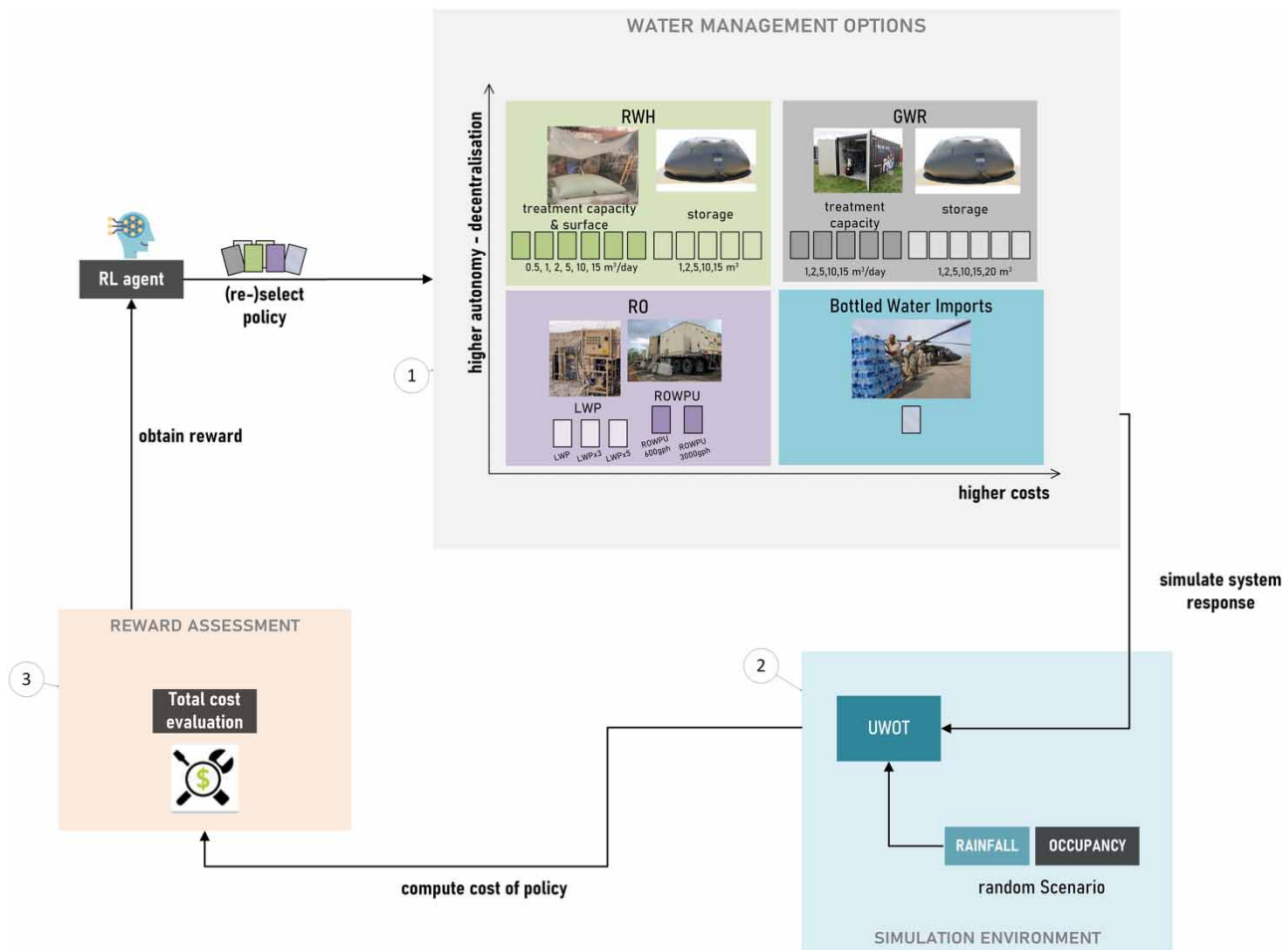


**Figure 2** | UWOT source-to-tap generic model for the off-grid system.

More specifically, the RL agent may:

1. **select** from different operationally available RO units (USACE 2008), if RO is available as an option, including the Light-weight Water Purifier (LWP) and two variants of the Reverse Osmosis Water Purification Unit (ROWPU), with a capacity of 600 gallons per hour ( $2.27 \text{ m}^3/\text{h}$ ) and 3,000 gallons per hour ( $11.35 \text{ m}^3/\text{h}$ ), respectively. For the LWP option, combinations of multiple units can be also selected (in this case, an array of 1, 3 or 5 units).
2. **activate** and **design** a containerized, modular GWR unit using a combination of different treatment capacities (1, 2, 5, 10, and  $15 \text{ m}^3/\text{day}$ ) and storage, i.e., onion tanks with different volumes (in this case, 1, 2, 5, 10, 15, and  $20 \text{ m}^3$ ).
3. **activate** and **design** a modular, portable RWH unit, similar to the one presented in Nguyen *et al.* (2013) that includes a preset collection surface ( $20 \text{ m}^2$ ), a set treatment capacity (in this case, 0.5, 1, 2, 5, 10, and  $15 \text{ m}^3/\text{day}$ ) and a selection of onion tanks for collected water storage (in this case, 1, 2, 5, 10, 15, and  $20 \text{ m}^3$ ). The RWH and GWR unit is designed as a hybrid system (Leong *et al.* 2017), so that the agent can choose which non-potable demands to allocate in each subsystem (including solutions where one of the two systems is deactivated).
4. **choose** to cover demands by bottled water imports. This is considered standard practice but comes with high costs in logistics and exposure to transport risks. Other imports or central infrastructure options, while modelled in the general topology (Figure 2), are deactivated and not considered in the RL agent learning process, so as to emulate off-grid deployments that rely exclusively on decentralized options, RO and bottled water imports.

A schematic of the policy selection process can be seen as the first step in Figure 3. Once an action (i.e., selection of supply options) is completed, an environment (model forcing of rainfall and occupancy) is generated randomly, representing the



**Figure 3** | Schematic of the coupling framework between the RL agent and the water cycle model.

deployment of a camp under specific climate and occupancy conditions. After this step, a simulation using the UWOT is executed (step 2). The model output, which takes the form of daily time-series of (un)met demands, as well as camp outflows, supplied water and water stored in tanks, can then be used to assess the performance of the agent's actions (i.e., to calculate the RL agent's reward). This constitutes the third step of the process. The reward assessment used in this work is based on a total cost evaluation of the chosen supply options, including the main capital expenditures (CAPEX) of every technology. It should be noted that, with the exception of bottled water, whose cost is documented from past camp deployments, other technologies lack consistent cost estimations (Klie & Rome 2005; Kershaw 2013), with data being limited for RO units and not available for decentralized technologies in military contexts. We therefore assume indicative costs based on civilian applications (Ricardo 2020), which are proportional to the complexity of deployment and maintenance of every technology (so that, for example, setting up a GWR is more expensive than RWH, or setting up an LWP is less expensive than deploying a larger ROWPU requiring higher maintenance costs). To guide storage sizing as part of the RL agent reward and to limit unwanted overflows, a (low) environmental overflow cost is also factored in, quantifying the impact of extensive overflows according to the polluter-pays principle (Ruiz-Rosa *et al.* 2020). An overview of the cost factors driving the RL agent reward evaluation is given in Table 2. It is noted that reliability for these systems is factored in through spill costs and higher CAPEX for larger tanks, as the RL agent is guided to avoid under- or over-designing these systems and minimizes unnecessary spillage.

Upon calculating the relevant cost of a given supply policy, the RL agent training process is then set as a negative cost (reward) maximization problem and the RL agent iteratively selects different supply options as policies and explores the action space. The initial agent actions (supply options) are fully randomized at every application. Considering that the RL agent may (de-)activate supply options and also choose how to divide non-potable demands between RWH and GWR, the order of magnitude of all possible combinations for technologies (i.e., all possible water management policies) is  $O(5)$  and as such cannot be exhaustively explored manually. In reality, the RL agent faces even larger complexity, given the stochasticity inherent in the scenarios of rainfall and occupancy. The framework is coded using python 3.x, with RL agents being operated through the respective AI python libraries of TensorFlow and Scikit-Learn.

### 3. RESULTS

Following the approach described in the previous section, the RL agent self-learns and selects optimal supply options at each run (i.e., through repeated exposure to a stochastically generated environment), thus creating technology bundles driven by the most cost-efficient solutions. Clearly, this selection is not the same across simulations, as stochasticity is introduced by:

- considering different climates in the area of deployment (specifically either temperate or dry climatic conditions). Relevant time-series for each climate type are selected and a random subsample of them is used each time to create the rainfall regime for the deployed camp;
- considering different occupancy in the camp, with the introduction of two occupancy settings: steady, i.e., constant, maximum occupancy at the camp, and variable, i.e., changing stochastically each day, according to the assumptions seen in Section 2.2. The results for the variable occupancy setting are discussed separately in the section that follows;
- allowing access to different supply options, with the introduction of two availability scenarios for RO units. In the first scenario, RO units are available to the planning agents who can use them in their array of options. In the second scenario, there is no RO unit availability, reflecting deployments where there is no locally treatable source with RO or the RO units

**Table 2** | Overview of the considered unit costs

Cost $C_i$	Process	€/gallon	€/L	Proportionality to reference RO (ROWPU) cost $C_i/C_{ROWPU}$
$C_B$	Bottled water	5.00	1.32	5.0
$C_{ROWPU}$	RO – ROWPU	1.00	0.26	1.0
$C_{LWP}$	RO – LWP	0.70	0.18	0.7
$C_{RWH}$	RWH	0.50	0.13	0.5
$C_{spill}$	Spill	0.10	0.03	0.1
$C_{GWR}$	GWR production	2.50	0.66	2.5

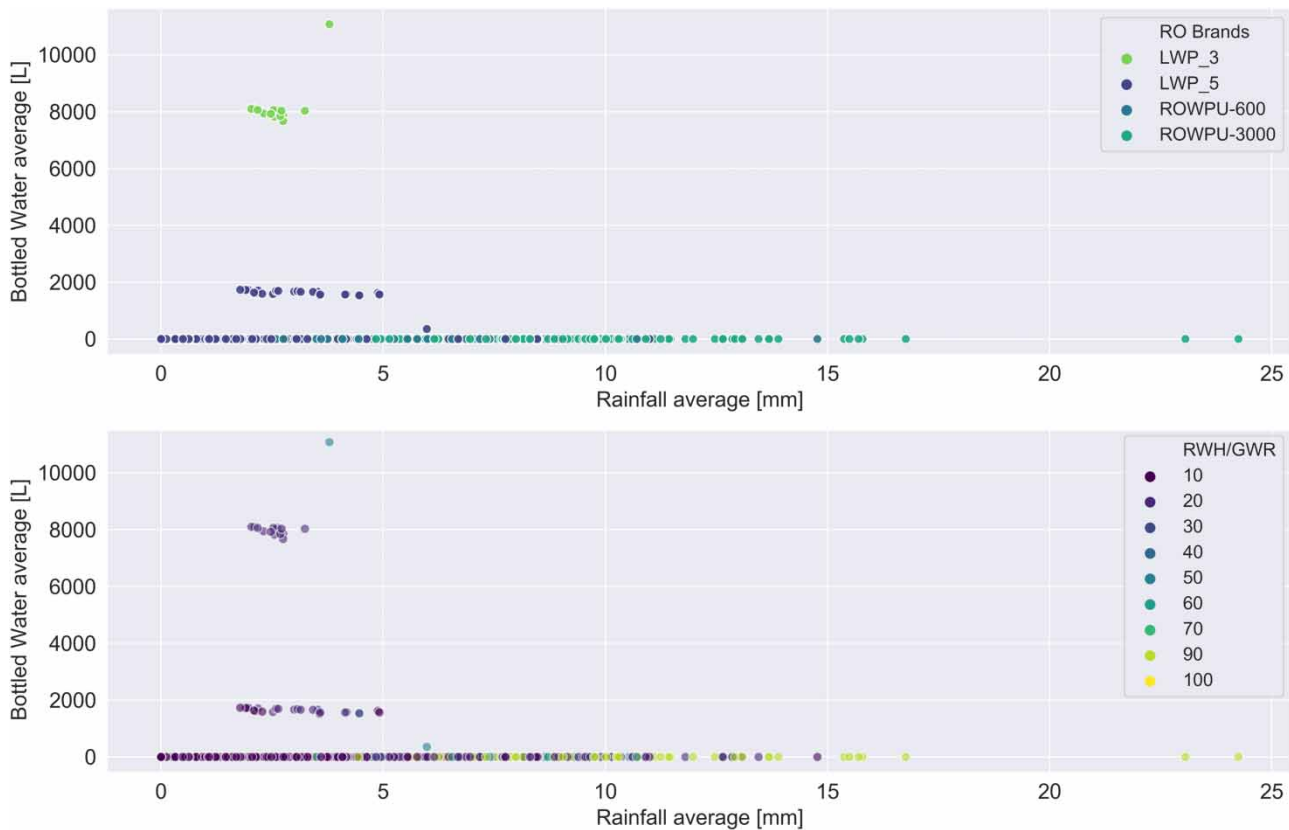


are unavailable due to high logistics and deployment costs. The RO supply options were chosen as the basis for option availability scenarios, as they proved to be highly influential in managing costs and securing supplied water against bottled water imports (see also Figure 4).

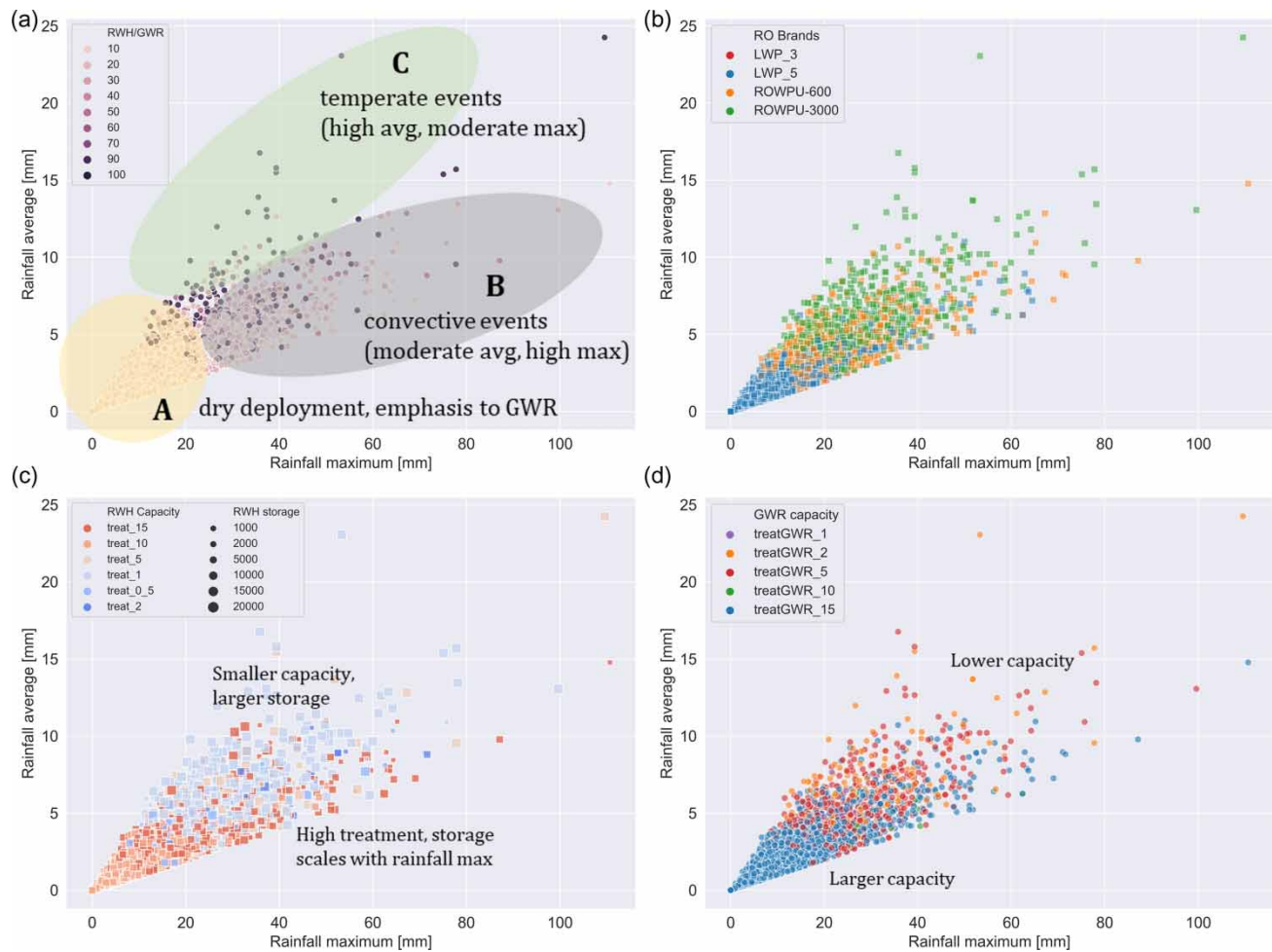
### 3.1. Supply policies vis-à-vis different climate regimes

The optimal supply policies found by the RL agent are best viewed and evaluated against factors such as the rainfall regime and overall camp occupancy. To visualize results, average bottled water imports (in L/day) are chosen as an indicator of management efficiency as a more intuitive metric for camp water management than total cost (Kershaw 2013). Using this metric, larger daily averages of bottled water indicate solutions that are dependent on imports and are thus significantly more costly due to the high cost of bottled water (Table 2). The decision space is shown as scatter plots against different environment dimensions, where every point reflects the optimal policy chosen by the agent for a particular camp deployment. As such, every point has a specific selected RO technology (indicated by RO units), it has a specific design for RWH or GWR units, and so on. Every point is thus a unique result of the agent's learning process, corresponding to the agent's response to a stochastically generated environment.

The effect of different rainfall regimes on the decision space is shown in Figure 5, which shows the impacts that different climates have on the selection of optimal RO technologies (panel a) and the design of the RWH/GWR system (panel b). The rainfall regimes that were synthetically generated range from very dry settings (zero or low rainfall average, in mm/day) to temperate settings with 10–15 mm of average daily rainfall. For the RO selection (Figure 5(b)), it can be observed that a group of policies exist, leading to high average bottled water import (of around 8 m<sup>3</sup>/day) utilizing lighter (and less productive, but also less energy dependent) LWP unit, in combinations of three units. In other runs, the agent opted to increase the RO units to 5 LPW units, reaching a local optimum of ~2 m<sup>3</sup>/day for bottled water imports. However, in most cases, the agent has chosen to increase the technology and capacity of the RO units further, utilizing the heavy duty units of ROWPU-500 and ROWPU-3000, to achieve zero bottled water imports. This space of solutions with ROWPUs



**Figure 4** | Visualization of the results in terms of the optimal RO technology selection and RWH/GWR system design.



**Figure 5** | Overview of optimal supply management policies, plotted against the rainfall variability of environments.

occurs regardless of rainfall averages, i.e., both in dry spells and wet spells, indicating that a suitable selection of RO technologies to minimize water imports under variable hydroclimatic conditions exists, thus posing a resilient solution. Interestingly, the selection of the specific ROWPU unit type does depend on the area's dryness/wetness, as the agent utilizes higher capacity units in combination with RWH systems for wetter environments, while the policy in drier settings relies on solutions that contain GWR.

The results indicate that GWR can substitute a large part of the demands and that higher capacity RO units are needed only when an intermittent source – such as RWH – is selected, introducing an element of uncertainty. It is also noted that there is no optimal policy that includes a single LWP RO unit; this is a sign that a single LWP is too small for the scale of the studied camp.

With regard to the decentralized elements of the system design (namely RWH/GWR), the agent can choose to utilize both of them, or to select one of them to cover part of non-potable demand. The agent then proceeds with the design of the hybrid system (defining capacity for treatment and storage for either GWR, RWH or both). The first design factor is included in the UWOT as the percentage of non-potable water demands split between the two subsystems; this percentage takes values between [0,100] with an interval of 10%; a value of 0% means that no demand is directed to the RWH system (and this system is inactive/closed), while a value of 100% means that all demands are directed to the RWH system (so that the GWR system is inactive/closed). Intermediate percentages mean that both units are active, and demands are split between the two units. Figure 5(b) shows the selection of the dominant RWH/GWR division for different runs, depending on the rainfall regime of the area. From these results, it is clear that rainfall has a strong effect on the design of the hybrid decentralized system: in dry deployments and under general water unavailability, the agent chooses to prioritize the GWR system and

deactivate the RWH system for non-potable demands. However, in temperate climate deployments, RWH becomes increasingly important as an alternative with a lower overall cost.

Figure 5 presents another overview of how the agent's plans are affected by the variability of rainfall in each camp deployment. Both axes show rainfall characteristics (maximum and average), so this space can be divided into three groups (see upper left panel of Figure 5(a)): firstly, a group of 'dry' deployments, meaning low average rainfall amounts and low maximum values. Secondly, a group of deployments in climates dominated by convective events: low to moderate rainfall average values, but high maximum values, and thirdly, cases with a temperate climate, with moderate rainfall maximum but high rainfall average values (i.e., multiple days of steady rain). Figure 5 demonstrates how these groups lead to different decisions by the RL agent, as the agent begins by 'closing off' RWH and prioritizing high-capacity GWR for dry deployments and then increasing the percentage/importance of RWH in temperate climates. These different climate zones also affect the sizing of the RWH and GWR units (Figure 5(c) and (d)) and the selection of RO technology (Figure 5(b)). In temperate climates, the RWH designed by the RL agent has a smaller treatment capacity but larger storage, while for convective events GWR continues to be the dominant decentralized option, so storage in RWH does not seem to play a role. The availability of RO options (panel (b)) is correlated with the dominance of the GWR system over RWH, in accordance with the findings of Figure 5.

### 3.2. Supply policies vis-à-vis different occupancy scenarios

Further to climate variability, it is also interesting to explore how variable occupancy affects water policies. Figure 6 compares the results of the (stochastically) variable occupancy scenario, plotted against the total camp occupancy (cumulative number of personnel at camp), which varies according to a preset pattern as described in Section 2.2 (reflecting conditions where personnel and vehicles are absent from the camp, leading to variability in camp occupancy and demands). The environment is a combined stochastic space now, with both rainfall and occupancy deployments varying randomly. It can be observed that, for the RO options (Figure 6(a)), the agent found multiple supply management policies of interest; for instance, a group of solutions with 'light RO' – LWP units, that, however, leads to a dependence on imported water if the occupancy exceeds a certain threshold. Higher capacity RO options are also utilized as more robust and lead to a smaller dependence on bottled water regardless of occupancy. Focusing on autonomy, i.e., on the use of RWH/GWR systems, results for variable occupancy can be seen in two variants: with (Figure 6(b)) and without (Figure 6(c)) access to RO solutions. When the agent is deprived of RO options (Figure 6(c)), one sees a smoother stratification of the design parameters of the RWH and GWR system. Two groups of optimal policies are now evident: one that places importance on RWH (40–50% of demands are covered by RWH) is more dependent on bottled water as the camp populations increase, and one that places more importance on GWR secures a steadier source of water and thus incurred less costs. Both policies have a much larger dependence on bottled water imports when RO options are removed.

### 3.3. Results vis-à-vis the availability of supply options

An interesting case, highlighted in the previous paragraph in light of variable occupancy, is the outcomes selected by the agent under the stricter constraint of no access to RO units, e.g., due to the unavailability of a local treatable surface

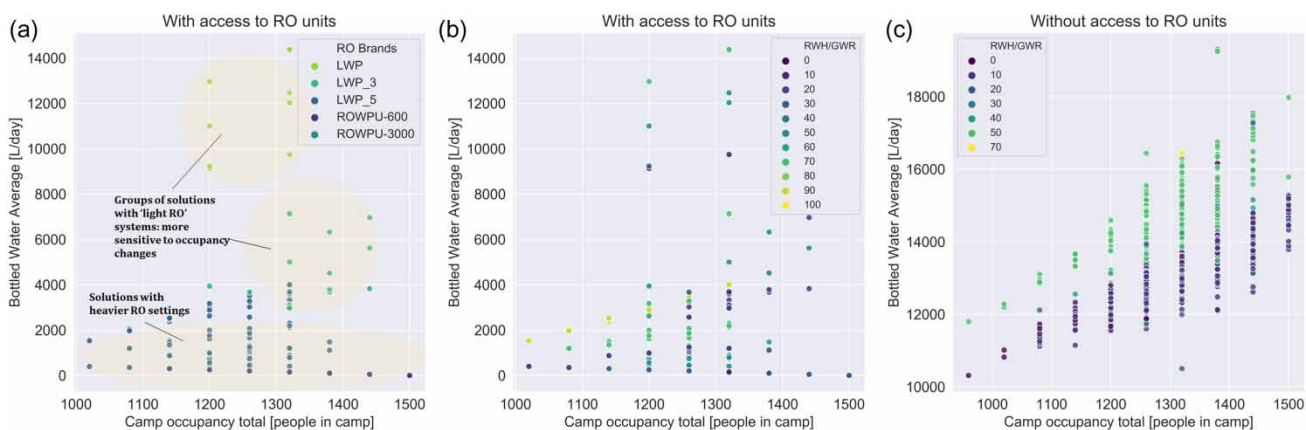
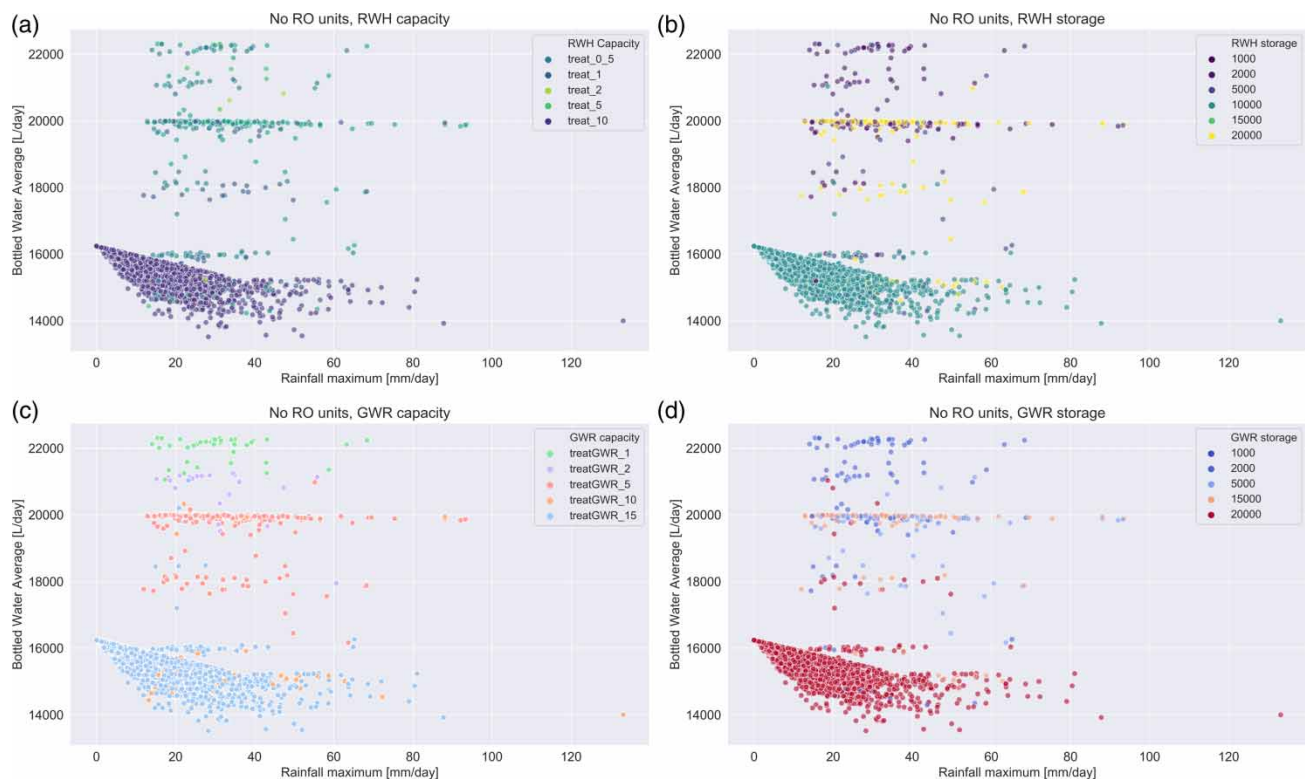


Figure 6 | Overview of optimal policy results for the scenario of variable occupancy.

water source. To explore this further, simulations have been repeated in a setting where deployments did not have access to RO units. In these more extreme cases, the agent is more limited in options; solutions can rely on autonomous, hybrid RWH/GWR units as before but necessarily also rely more heavily on bottled water imports. Figure 7 shows the results of these simulations. Unlike previous results here, there is always dependence on bottled water imports, reaching more than  $15 \text{ m}^3/\text{day}$  in most cases. The agent tries multiple combinations and finds different ‘plateaus’ of supply policies, exploring different storages for RWH and GWR (even very large ones), as well as different capacities. The lowest bottled water average comes for a large group of uniform policies (i.e., with similar values for RWH/GWR treatment and capacity) that lead to a smaller bottled water import footprint of around  $14\text{--}16 \text{ m}^3/\text{day}$ , depending on rainfall intensity and thus on the efficiency of the RWH system. The uniformity of these values across different rainfall regimes shows that this policy is resilient, pointing towards a robust hybrid RWH/GWR system design that can be used in different deployment settings, combining the ‘steady’ GWR stream of non-potable water with an RWH tank. Notice that this solution is not oversized: the GWR system is maximized, but for the RWH unit, the optimal capacity is  $10 \text{ m}^3/\text{day}$ , while the optimal storage is  $15 \text{ m}^3$  (and not  $15 \text{ m}^3/\text{day}$  and  $20 \text{ m}^3$ , which are the largest options).

#### 4. DISCUSSION

In this work, we demonstrated an application of RL algorithms to assist in planning decisions for off-grid water systems employing decentralized technologies. The approach is demonstrated in a hypothetical but realistic deployed camp context based on real camp deployment supply options and doctrines and offers insights on how RL agents can aid complex water management planning at a proof-of-concept level. In particular, the PPO algorithm has been employed as a planning agent to design an off-grid camp under uncertainty both in terms of climate (rainfall) and in terms of occupancy patterns. The results demonstrate that the RL agent can self-learn and systematically identify optimal supply policies, despite the high-complexity stemming from the multitude of supply option choices paired with the stochastic nature of the environments on climate and occupancy.



**Figure 7** | Results of bottled water average against rainfall regimes for the case where RO units are not available.

Similarly to other recent applications using RL methods for water (Wang *et al.* 2021), we find that, while RL generally suffers from poor convergence and stability, PPO demonstrates stability and consistent convergence in a complex application while achieving good performance in terms of runtime. While the PPO family of RL algorithms proved to be the more stable technology for the problem in question, more in-depth research is needed in the field of RL to identify appropriate algorithms for water management challenges with due consideration on issues of convergence and stability, as these are well-known limitations of RL approaches (Nian *et al.* 2020). This study is limited by data availability, as it relies on a cost-based reward function and thus requires accurate information on costs associated with different supply options; while our work was based on available data, the assumed costs of Table 2 are indicative and suitable at a proof-of-concept level only. More reliable cost data for planning interventions – particularly with regard to decentralized systems – would be beneficial for applications of this approach to real cases, possibly coming from *in situ* pilots. Moreover, the algorithmic dependence on (and relative importance of) costs could be further investigated, for instance, with a sensitivity analysis, e.g., in the form of batch runs using different cost ratios.

Another limitation of this study is that the effect of different parameters on the performance of the RL agents remains unexplored. There are several environment variables (such as daily absence probabilities ( $p_p, p_v$ ) or the time of deployment per run  $t$ ) that have been preselected in this study empirically, which potentially affect the convergence rate or the selection of technologies of the PPO algorithm. A more rigorous analysis is thus needed to estimate how robust RL agents are when faced with structural (i.e., parameter) uncertainty as well. Such an analysis is outside the scope of the work at this proof-of-concept level as it comes with high computational burden and the need of multiple recursive runs. We envision this analysis as stand-alone work, similar to the one seen, e.g., in Pasha & Lansey (2010), focusing on the sensitivity and robustness of PPO algorithms, and RL agents in general, for water applications, potentially using upgraded hardware to diminish runtimes and optimize RL applications for the tactical and strategic scale of planning.

The application used to demonstrate the approach, while simplified, is not significantly different from real-world camp deployments. Moreover, the coupling methodology shown in Figure 3 holds for different reward function formulations, for instance, non-cost-related ones. Examples could include rewards derived from minimizing demand deficits and thus bottled water imports (or equivalent transport trips, for instance). Similar reward functions can focus on energy or carbon emissions instead of costs, if data on energy consumption or carbon footprint of water supply options are available. In such cases, supply policies may differ considerably, as some of the supply options (e.g., RO and GWR) come with high related energy and carbon footprints. In these cases, the RL agent could be adapted to look for energy-efficient or carbon-neutral deployments for off-grid systems. We argue that this approach also holds for other applications of deployable systems in water management that share the need for autonomy and resilience, such as, for example, refugee camp designs or disaster risk mitigation installations (e.g., deployed to temporarily host the affected populations after extreme events such as earthquakes or hurricanes). Moreover, adaptations to address water quality issues in relevant systems could be explored (Eissa *et al.* 2022), if the RL agent is paired with a model capable of simulating water quality at a fine resolution.

## 5. CONCLUSIONS

In this work, a water infrastructure planning AI agent, based on RL, is integrated with a water cycle simulation model to explore different water supply policies for different camp deployment conditions. Uncertainty is factored in the studied case with regard to climate, occupancy and water infrastructure option availability, increasing the complexity of the application. The RL agent is supplied with data on the availability and cost of different water options and, at the same time, confronted with highly variable, randomized environments. The resulting framework is demonstrated in a hypothetical camp deployed off grid, with the agent being capable of selecting from a set of modular technologies that include conventional management options (i.e., RO units), as well as circular technologies (RWH/GWR).

The choice of the RL algorithm seems to be dependent on the complexity of the system at hand, with some methods generally demonstrating instability and the lack of convergence when facing higher complexity (Mullapudi *et al.* 2020; Wang *et al.* 2021). In our case, and in line with past literature (Wang *et al.* 2021), we find that some methods are unstable and that PPO outperforms other RL algorithm families with regard to stability, convergence and performance runtime in the planning task in hand. While not all solutions seem to be globally optimal, most simulations suggest that the agent can steadily find plateaus of optimality with shared design attributes, with most solutions having low potable water imports and overall cost. Interestingly, the results also indicate that, when confronted with less options (RO treatment unit unavailability), the

agent can find robust, hybrid, autonomous RWH/GWR designs, which have similar design (sizing) parameters across different environments. This is encouraging for water planning and management in such settings, as it suggests the existence of transferable/reusable but still resilient water system designs for off-grid deployments. The results are significant, as most RL applications for water systems so far have focused on relatively simple applications in controlled systems not exposed to multiple dimensions of uncertainty (Wang *et al.* 2021). Moreover, they are encouraging the use of RL algorithms in general, as some of these algorithms have been found to be non-converging or unstable when faced with complex water applications (Mullapudi *et al.* 2020).

Based on the findings of this work, it is suggested that PPO RL algorithms can be a valuable addition to the water management decision support ‘toolkit’, which is also relevant for longer-term planning – beyond current applications in real-time control. In such contexts, these algorithms can scan decision spaces and explore considerably higher numbers of policies than planners can do through manual labour and conventional scenario analysis. Moreover, such applications are useful as management instruments that support the rationalization of decisions on complex water management problems, typically performed based on the prior operational experience of individuals (van Winden & Dekker 1998). The demonstrated application, although at a proof-of-concept stage, can be readily expanded to real-world settings provided real cases and data can be identified, collected and analysed to enrich training.

On a more general note, it is further suggested that the proposed coupling between RL agents and physically based models can be useful in cases of strategic planning and management for complex infrastructures (such as in climate change adaptation or critical infrastructure protection contexts), or to explore how different hydrological, terrestrial, socioeconomic and policy-related determinants affect decision-making (Khan *et al.* 2021; Cacal & Taboada 2022), thus offering a promising field of application for RL in hydroinformatics.

## ACKNOWLEDGEMENTS

The research leading to these results has received funding from the European Defence Agency under Contract Number: 19.RTI.OP.373 for the Operational Budget Study ‘ARTENET: Artificial Intelligence for Energy and Environmental Technologies’. The research and its conclusions reflect only the views of the authors, and the European Defence Agency is not liable for any use that may be made of the information contained herein. The authors would like to acknowledge the help of Dr Kostas Eftaxias who provided valuable support with the RL algorithms and the setup of the smart agent. The authors would also like to thank Dr Ifigenia Koutiva for her help during preliminary runs of the coupled agent-simulation system.

## DATA AVAILABILITY STATEMENT

Data cannot be made publicly available; readers should contact the corresponding author for details.

## CONFLICT OF INTEREST

The authors declare there is no conflict.

## REFERENCES

- Abdullah, A. & Gunal, A. Y. 2022 *Optimization of Water Distribution System Within Tented Camps*. Çukurova Üniversitesi Mühendislik Fakültesi Dergisi, pp. 23–31. Available from: <https://dergipark.org.tr/en/doi/10.21605/cukurovaumfd.1094936>.
- Bhattacharya, B., Lobbrecht, A. H. & Solomatine, D. P. 2003 *Neural networks and reinforcement learning in control of water systems*. *Journal of Water Resources Planning and Management* **129** (6), 458–465.
- Bouziotas, D., van Duuren, D., van Alphen, H.-J., Frijns, J., Nikolopoulos, D. & Makropoulos, C. 2019 *Towards circular water neighborhoods: Simulation-based decision support for integrated decentralized urban water systems*. *Water* **11** (6), 1227.
- Cacal, J. C. & Taboada, E. B. 2022 *Assessment and evaluation of IWRM implementation in Palawan, Philippines*. *Civil Engineering Journal* **8** (2), 290–307.
- Castelletti, A., Corani, G., Rizzoli, A., Soncinie-Sessa, R. & Weber, E. 2002 *Reinforcement learning in the operational management of a water system*. In *IFAC Workshop on Modeling and Control in Environmental Issues*. 2002, Citeseer, pp. 325–330.
- Doorn, N. 2021 *Artificial intelligence in the water domain: Opportunities for responsible use*. *Science of the Total Environment* **755**, 142561.
- Eissa, M. E., Rashed, E. R. & Eissa, D. E. 2022 *Dendrogram analysis and statistical examination for total microbiological mesophilic aerobic count of municipal water distribution network system*. *HighTech and Innovation Journal* **3** (1), 28–36.
- Hadjimichael, A., Comas, J. & Corominas, L. 2016 *Do machine learning methods used in data mining enhance the potential of decision support systems? A review for the urban water sector* K. Gibert (ed.). *AI Communications* **29** (6), 747–756.

- Jotem 2020 *Black to Grey Water Conversion on Military Camp*. Available from: <https://www.dutchwatersector.com/news/black-to-grey-water-conversion-on-military-camp>.
- Kershaw, B. 2013 Drinking water. *Marine Corps Gazette* **8** (97), 59–61.
- Khan, A. U., Rahman, H. U., Ali, L., Khan, M. I., Khan, H. M., Khan, A. U., Khan, F. A., Khan, J., Shah, L. A., Haleem, K., Abbas, A. & Ahmad, I. 2021 *Complex linkage between watershed attributes and surface water quality: Gaining insight via path analysis*. *Civil Engineering Journal* **7** (4), 701–712.
- Klie, J. H. & Rome, S. 2005 *U.S. Army Reduces Water Costs with Mobile Purifier Units*. WaterWorld. Available from: <https://www.waterworld.com/home/article/16201606/us-army-reduces-water-costs-with-mobile-purifier-units> (accessed: 8 May 2022).
- KNMI 2022 *KNMI Data Platform*. 2022. Available from: <https://dataplatfom.knmi.nl/>.
- Kuleshov, V. & Precup, D. 2014 Algorithms for multi-armed bandit problems. *arXiv preprint arXiv:1402.6028*.
- Leong, J. Y. C., Oh, K. S., Poh, P. E. & Chong, M. N. 2017 *Prospects of hybrid rainwater-greywater decentralised system for water recycling and reuse: A review*. *Journal of Cleaner Production* **142**, 3014–3027.
- Maier, H. R., Kapelan, Z., Kasprzyk, J., Kollat, J., Matott, L. S., Cunha, M. C., Dandy, G. C., Gibbs, M. S., Keedwell, E., Marchi, A., Ostfeld, A., Savic, D., Solomatine, D. P., Vrugt, J. A., Zecchin, A. C., Minsker, B. S., Barbour, E. J., Kuczera, G., Pasha, F., Castelletti, A., Giuliani, M. & Reed, P. M. 2014 *Evolutionary algorithms and other metaheuristics in water resources: Current status, research challenges and future directions*. *Environmental Modelling & Software* **62**, 271–299.
- Mala-Jetmarova, H., Sultanova, N. & Savic, D. 2017 *Lost in optimisation of water distribution systems? A literature review of system operation*. *Environmental Modelling & Software* **93**, 209–254.
- Mason, K., Mannion, P., Duggan, J. & Howley, E. 2016 *Applying multi-agent reinforcement learning to watershed management*. In: *Proceedings of the Adaptive and Learning Agents Workshop (at AAMAS 2016)*.
- Mullapudi, A., Lewis, M. J., Gruden, C. L. & Kerkez, B. 2020 *Deep reinforcement learning for the real time control of stormwater systems*. *Advances in Water Resources* **140**, 103600.
- NATO 1994 *NATO STANAG 2136 – Requirements for Water Potability During Field Operations and In Emergency Situations*.
- Nguyen, D. C., Dao, A. D., Kim, T.-I. & Han, M. 2013 *A sustainability assessment of the rainwater harvesting system for drinking water supply: A case study of Cukhe Village, Hanoi, Vietnam*. *Environmental Engineering Research* **18** (2), 109–114.
- Nian, R., Liu, J. & Huang, B. 2020 *A review on reinforcement learning: introduction and applications in industrial process control*. *Computers & Chemical Engineering* **139**, 106886.
- Pasha, M. F. K. & Lansey, K. 2010 *Effect of parameter uncertainty on water quality predictions in distribution systems-case study*. *Journal of Hydroinformatics* **12** (1), 1–21.
- Price, E. & Ostfeld, A. 2014 *Discrete pump scheduling and leakage control using linear programming for optimal operation of water distribution systems*. *Journal of Hydraulic Engineering* **140** (6).
- Ricardo 2020 *Independent Review of Costs and Benefits of RWH and GWR Options in the UK*.
- Rozos, E. & Makropoulos, C. 2013 *Source to tap urban water cycle modelling*. *Environmental Modelling & Software* **41**, 139–150.
- Ruiz-Rosa, I., García-Rodríguez, F. J. & Antonova, N. 2020 *Developing a methodology to recover the cost of wastewater reuse: A proposal based on the polluter pays principle*. *Utilities Policy* **65**, 101067.
- Savic, D. A. 2019 *Artificial Intelligence: How Can Water Planning and Management Benefit From it?*
- Schulman, J., Wolski, F., Dhariwal, P., Radford, A. & Klimov, O. 2017 *Proximal policy optimization algorithms*. *arXiv preprint arXiv:1707.06347*.
- Silver, D., Hubert, T., Schrittwieser, J., Antonoglou, I., Lai, M., Guez, A., Lanctot, M., Sifre, L., Kumaran, D., Graepel, T., Lillicrap, T., Simonyan, K. & Hassabis, D. 2018 *A general reinforcement learning algorithm that masters chess, shogi, and Go through self-play*. *Science* **362** (6419), 1140–1144.
- Solomatine, D. P. & Ostfeld, A. 2008 *Data-driven modelling: Some past experiences and new approaches*. *Journal of Hydroinformatics* **10** (1), 3–22.
- Twaha, S. & Ramli, M. A. M. 2018 *A review of optimization approaches for hybrid distributed energy generation systems: Off-grid and grid-connected systems*. *Sustainable Cities and Society* **41**, 320–331.
- Tyralis, H., Papacharalampous, G. & Langousis, A. 2021 *Super ensemble learning for daily streamflow forecasting: Large-scale demonstration and comparison with multiple machine learning algorithms*. *Neural Computing and Applications* **33** (8), 3053–3068.
- USACE 2008 *Water Planning Guide: Potable Water Consumption Planning Factors by Environmental Region and Command Level*.
- Van de Walle, A., Torfs, E., Gaublomme, D. & Rabaey, K. 2022 *In silico assessment of household level closed water cycles: Towards extreme decentralization*. *Environmental Science and Ecotechnology* **10**, 100148.
- Van Hasselt, H., Guez, A. & Silver, D. 2016 *Deep reinforcement learning with double q-learning*. In *Thirtieth AAAI Conference on Artificial Intelligence*, 2016.
- van Winden, C. & Dekker, R. 1998 *Rationalisation of building maintenance by Markov decision models: A pilot case study*. *Journal of the Operational Research Society* **49** (9), 928–935.
- Wang, C., Bowes, B. D., Beling, P. A. & Goodall, J. L. 2021 *Reinforcement Learning for Flooding Mitigation in Complex Stormwater Systems during Large Storms*. In: *IEEE EUROCON 2021 – 19th International Conference on Smart Technologies*, 6 July 2021. IEEE, pp. 274–279. Available from: <https://ieeexplore.ieee.org/document/9535587/>.