Measuring the Performance of Multi-Objective Reinforcement Learning algorithms - Nile River Case Study

Jakub Kontak[1], Zuzanna Osika[1] and Pradeep Murukannaiah[1]

[1]Faculty of Electrical Engineering, Mathematics and Computer Science, Delft University of Technology, Netherlands

1. Introduction

- Simulations of real-world multi-objective problems are developed in water management.
- These simulations are not compatible with Gymnasium and don't allow in any other way for simple change of optimization algorithm. Thus, (Reinforcement Learning) RL algorithms cannot easily be used for these problems.
- Connecting two disciplines can have following benefits: potentially better solutions to the water management problems and a chance for benchmarking RL algorithms on real-world problems.
- In this research, we want to bridge the gap between those two disciplines by rewriting the simulation to make it compatible with Gymnasium and benchmarking multi-objective RL algorithm (MONES [2]) with water management algorithm (EMODPS [4]).

2. Simulation

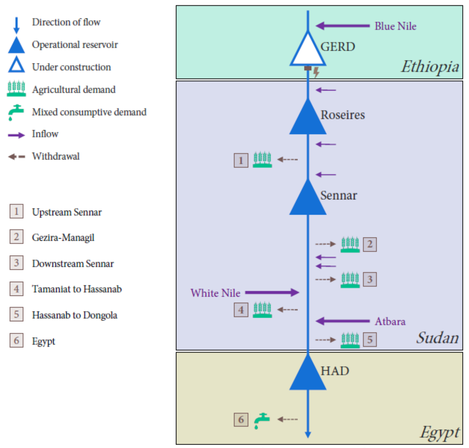


Image source: [1]

The simulation, adapted from [1], consists of four reservoirs (dams) that decide how much water they release at each timestep. The water releases have downstream impact on irrigation, power production and stored water in the next timestep. This influences objectives of the involved countries.

Country	Objective	Direction
Egypt	Irrigation demand deficit	Minimisation
Egypt	Minimum HAD water level	Maximisation
Sudan	Irrigation demand deficit	Minimisation
Ethiopia	Hydropower production	Maximisation

Developed simulation framework is Gymnasium compatible. It means that an agent can communicate with environment by providing actions and receiving observations and rewards. This standard API allows for simple changes of optimisation algorithm.

3. Methods – performance metrics

Min-max normalisation – normalisation is applied to the objectives to scale them to the 0, 1 range. This makes each objective contribute to the indicator in the comparable manner.

Hypervolume – a multi-dimensional volume spanned by the non-dominated points of a solution set with respect to the reference point [3].

Additive epsilon-indicator – a smallest factor by which the objective values for the Pareto front must be decreased such that each point in the Pareto front is weakly dominated by at least one solution from the solution set. The formula follows:

$$I_{\varepsilon^+} = \inf_{\varepsilon \in R}\{\forall V^{\pi} \in PF, \exists V^{\pi'} \in S : V_i^{\pi} \leq V_i^{\pi'} + \varepsilon, \forall i \in \{1, \dots, n\}\},$$

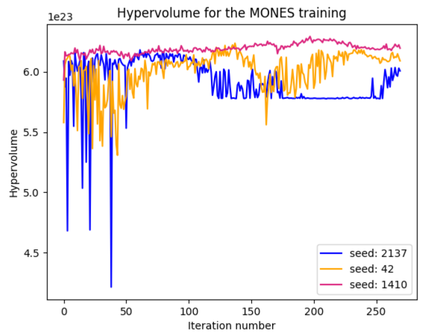
where V^{π} is a value of policy π, PF is the Pareto front, and S is the compared solution set.

Inverted Generational Distance Plus – an average modified distance from each point of the Pareto front to the closest point in solution set S [3]. It is given by the following formula:

$$IGD^{+}(S) = \frac{1}{|PF|}\left(\sum_{i=1}^{|PF|} d_i^{+2}\right)^{1/2},$$

where PF is the Pareto front, $d_i^{+} = \max(0, a_i - z_i)$ and S is the compared solution set.

4. MONES agent training



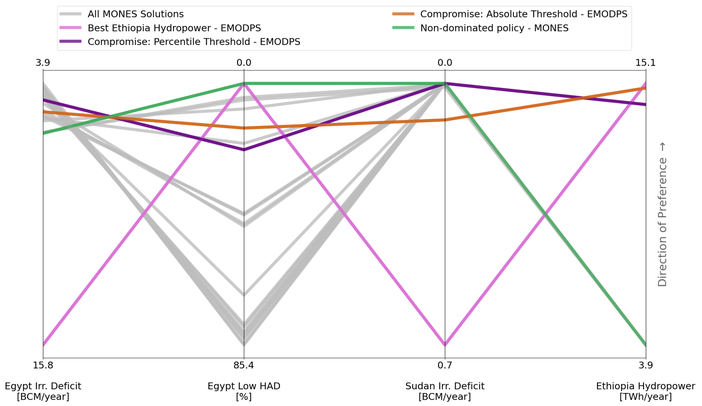
Hypervolume progress is used for training a MONES agent. Each training run takes around 10 hours using 32 CPUs. Hyperparameters are as follows:

- N iterations: 270
- Population size: 128
- N runs per individual: 1
- Indicator: Hypervolume

An agent is trained with three different random seeds and the hypervolume progress is plotted above. Each run has around 35 000 number of function evaluations (NFEs), which is comparable to 50 000 NFEs used for the EMODPS training [1].

5. Results – found policies

The results of the run with highest hypervolume (seed 1410) are analysed. Below the objectives for different MONES policies and some chosen EMODPS policies from [1] are plotted.



MONES produces feasible results for all objectives except the hydropower production. However, its results are in very small range for Egypt deficit, Sudan deficit and Ethiopia hydropower. The variability and more significant trade-offs can only be seen for the Egypt low HAD level objective. Low solution diversity can be seen even better when compared with EMODPS policies like Best Ethiopia Hydropower. A potential reason for that can be too low MONES exploration.

6. Results – performance metrics

The comparison of performance metrics between MONES and EMODPS can be seen in the table below, where the arrow indicates whether larger or smaller outcome is more desirable.

Metric	MONES	EMODPS
Hypervolume ↑	0.32	2.03
Additive ε-indicator ↓	1.0	0.005
IGD+ ↓	0.84	0.00002

EMODPS outperforms MONES significantly in every metric. However, the metrics are highly influenced by the poor performance of MONES in generating hydropower, despite its ability to find feasible solutions in the other objectives.

When creating non-dominated solution set from both algorithms it consists of 222 points from EMODPS and 1 point from MONES. Thus, EMODPS almost dominates the MONES solution set.

For MONES the number of non-dominated points in the solution set rises with higher hypervolumes. Seeds 2137, 42 and 1410 achieved 11, 14 and 18 non-dominated points, respectively. This order of seeds corresponds to the increasing order of hypervolume in the final solution set.

7. Conclusions

Contributions

- Implemented Nile River simulation compatible with the Gymnasium framework.
- Benchmarked MONES against EMODPS in the Nile River case study.

Findings

- EMODPS performs better than MONES in the Nile River simulation. Its results are more diverse and dominate most of the solutions found by MONES.
- MONES produces feasible solutions on three out of four objectives.
- MONES lacks some exploration to find more diverse policies and optimise the hydropower production.

Limitations

- Data used in this simulation comes from [1]. However, we cannot be sure of its correctness as it is a 20 years prediction, which can vary depending on many factors. Since data influences training, as well as outcomes this produces an uncertainty.
- The number of iterations is limited due to constrained computational resources, maybe much longer training time could achieve different results.

Future work

- Implement **EMODPS compatible with the Gymnasium framework**. Measure its performance on different RL benchmarks.

[1] Y Sari. Exploring trade-offs in reservoir operations through many objective optimisation: Case of Nile River basin. 2022.
[2] C. F. Hayes, R. Radaulescu, and E. Bargiacchi et. al. A practical guide to multi-objective reinforcement learning and planning. Auton Agent Multi-Agent Syst 36, 26. 2022.
[3] J. Blank and K. Deb, pymoo: Multi-Objective Optimization in Python, in IEEE Access, 8, 89497-89509. 2020.
[4] J. Z. Salazar, P. M. Reed, J. D. Herman, M. Giuliani, A. Castelletti. A diagnostic assessment of evolutionary algorithms for multi-objective surface water reservoir control. Advances in Water Resources, 92, 172-185. 2016.