EuroSun2026

REINFORCEMENT LEARNING FOR ADAPTIVE CONTROL OF DECENTRALISED FEED-IN OF SOLAR HEAT INTO DISTRICT HEATING NETWORKS

Summary

This work presents the control optimisation for decentralised district heating feed-in stations using artificial intelligence. Reinforcement learning, a subcategory of machine learning (ML), is used for adaptive PID parameter optimisation. The learning algorithm was developed and pre-trained in a validated digital twin of a decentralised solar thermal feed-in station. Simulations performed with the digital twin show up to 7% higher solar heat feed-in and up to 8% lower set-point temperature deviation over one year, when using adaptive control instead of manually tuned PID parameters by experts. The pre-trained controller was successfully deployed in an experimental plant of the solar thermal feed-in station. A continuous test over 23 days confirms the controller's real-time capability and validates the potential to improve control performance through adaptive control.

Keywords: Machine Learning, Artificial Intelligence, Reinforcement Learning, Solar Thermal, Decentralised Feed-In Stations, District Heating, Control Optimisation, Digital Twin, Cloud-based Controller

Field of application for the findings: District Heating, Solar Thermal Systems, Renewable Energy

Addressed audience: Scientific Community, District Heating Operators, Solar Thermal System Suppliers

1. Introduction

District heating (DH) is crucial for decarbonising the heating sector. Large-scale solar thermal (ST) plants already offer a proven, economically viable, and flexible solution for transforming existing DH networks and supplying new systems with renewable heat. Most of the realised systems are integrated centrally, so the collector fields are usually conntected to the primary heat generation facility, offering the possibility to reheat and to readjust the solar delivered heat. In urban areas, the availability of suitable areas for ST systems near the network's heat generation facility is often limited. Therefore, decentralised ST systems installed on roofs of larger buildings or decentralised open spaces offer an attractive alternative to central integration.

Controlling decentralised feed-in of solar heat into DH networks is challenging due to heat production fluctuations, varying network conditions as well as feed-in temperature requirements. Experimental testing of the ST feed-in station in Düsseldorf highlighted the significant impact of PID control parameter tuning on feed-in duration and temperature deviation from set point (Schramedei et al. 2023). The tuning of PID parameters is usually performed manually applying heuristics proposed by Ziegler & Nichols (1942) or by trial and error. The manual tuning is time-consuming, requires expertise and can lead to unsatisfying control performance especially in highly dynamic systems. With adaptive PID control, the parameters are automatically tuned and adjusted to different operating conditions, overcoming the limitations of manually tuned static PID controllers.

Adaptive Reinforcement learning (RL)-based control outperformed state-of-the-art control strategies in several studies. Korupu und Manimozhi (2025) showed automatic and precise adaptation of PI parameters in an industrial water supply system based on RL, resulting in a significant improvement in control performance under varying operating conditions. In a pumped hydroelectric energy storage, a RL-based controller was able to efficiently adapt the turbine control to varying headwater pressure conditions, fully replacing the manual retuning of PID parameters (Enyekwe et al. 2023). Sabahi et al. (2024) demonstrated the online adaptation of PID parameters in microgrids to cope with volatile conditions, including fluctuating demand and weather conditions.

This work builds on previous studies to implement a RL-based adaptation of control parameters for a decentralised solar heating system. The RL algorithm is developed and pre-trained in a digital twin of the feed-in station in Düsseldorf and later applied in the experimental plant.

2. Methodology

2.1 Digital twin

A digital twin of the experimental plant was developed to implement and pre-train the adaptive controller in a safe and controlled learning environment. The digital twin accelerates training compared to learning in the real environment by a factor of approximately 340. It consists of a TRNSYS simulation model of the ST feed-in station coupled with a duplicate of the controller hard- and software used in the experimental plant. The controller receives the simulated temperatures and flow rates, and the simulation input data for time step t. Subsequently, the controller determines the control signals and the set rotation speed for the pumps in the primary and secondary circuit. The pump speed is controlled with PID controllers. The control variables for the primary and secondary pump are the set volume flow and the set feed-in temperature, respectively. The volume flows are then calculated in TRNSYS for $t+1$ based on the control signals and rotation speeds for t. The simulation time step is 60 s.

The digital twin is validated with historical measurement data applying static reference values for the PID parameters proportional gain K_P, integration time T_I and derivative time T_D. The reference values were determined during the testing and optimisation phase of the experimental plant being the parameter combination that achieved the most robust control behaviour among all evaluated combinations (Schramedei et al. 2023).

2.2 Reinforcement learning algorithm

The deep deterministic policy gradient (DDPG) algorithm is applied being suitable for adaptive PID control given its applicability in continuous state and action spaces, its ability to update actions in near real time and its online learning ability. Figure 1 shows a simplified scheme of the algorithm developed by Lillicrap et al. (2015). As usual in RL, DDPG consists of an agent operating within a learning environment by performing actions to achieve a defined goal. In order to learn from previous experiences and improve action selection, the agent receives a reward which is a numerical value that evaluates the performed action with regards to the defined goal.

The environment provides the agent with the state of the previous action update time step (s_t), the reward (r_{t+1}) for the executed action (a_t), and the resulting subsequent state (s_{t+1}). The agent provides a_t, in this case the PID controller parameters (K_P, T_I, T_D) for both pump speed controllers. The policy function μ describes the agent's decision-making strategy. DDPG uses two different artificial neural networks. The actor network applies μ to determine the vector a_t composed of continuous control values within a pre-defined range. The critic network evaluates the performed action and estimates the Q-value, which is defined as the expected cumulative reward of future states with the discount factor for future states Γ. The training in the digital twin starts with a "warm-up" phase with random actions. After this warm-up, learning is performed by activating the optimizers for both networks to minimize the loss with regards to the respective loss function.

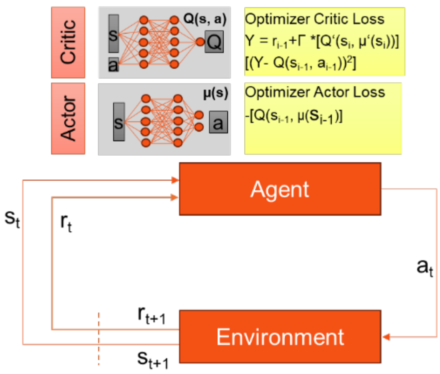

Figure 1: Simplified scheme of DDPG

2.3 Cloud-based control framework

The controller of the experimental plant is connected to a web-based cloud system enabling data exchange via the communication protocol Modbus. A web Application Programming Interface (API) is used to retrieve measurement data for each action update step needed to calculate the new state vector and the reward obtained for the previous action. The RL algorithm then updates the PID parameters as a function of the new state vector and sends them to the controller via the web API. The algorithm is retrained with data gathered in the real environment after each update step. The retraining is required to adjust the pre-trained RL models to discrepancies between the experience collected in the real environment and that generated within the digital twin. This framework allows for an adaptive PID control and online learning in near real time.

3. Results

3.1. Performance of the adaptive controller in the digital twin

The digital twin is applied for pre-training of the DDPG algorithm over 11 episodes. Each episode consists of a complete simulation run with input data from the entire year 2023. The reward increases significantly during the first 4 episodes, then the learning slows down. The best overall reward is obtained in episode 9 with a relative improvement of 22.7 % compared to episode 1. The trained DDPG is tested with pre-trained RL models from episode 9 and input data from the year 2024. Table 1 quantifies the relative change (%Δ) of simulation results obtained with the adaptive DDPG controller compared to the reference case, which is the simulation with static PID parameters. With DDPG, the simulated heat fed into the DH network within the set temperature range of 80 °C ±3 K increases and the simulated mean feed-in temperature deviation from the set point decreases, indicating an improvement in control performance. The adaptive controller shows a higher relative improvement in the training run than in the test run. This can be explained by the training process, where multiple episodes are executed using the exact same dataset, which implies the risk of overfitting to the training data.

Table 1: Training and test result of DDPG algorithm and reference in digital twin

Performance indicator	Training (2023)			Test (2024)		
	DDPG	Ref.	%Δ	DDPG	Ref.	%Δ
Feed-in heat in ±3 K range [GJ]	212.4	198.7	+6.9%	164.2	161.3	+1.8%
Mean temp. dev. from set point [K]	2.99	3.25	-8.0%	3.32	3.57	-7.0%

3.2. Performance of the adaptive controller in the experimental plant

The pre-trained DDPG algorithm was continuously tested in the real plant for 23 days in October 2025. Figure 2 shows the measured return (r) and supply (s) temperatures in both primary (prim) and secondary (sec) circuit, volume flow rates and irradiance for two test days. For most of the period with $\dot{V}_{sec} > 0$, T^s_{sec} remains within the set range (black-dotted lines). Temperature deviations and resulting pump switch-offs occur due to insufficient solar irradiance. They are not caused by the controller, as $\dot{V}_{sec}$ is at its lower limit during these periods.

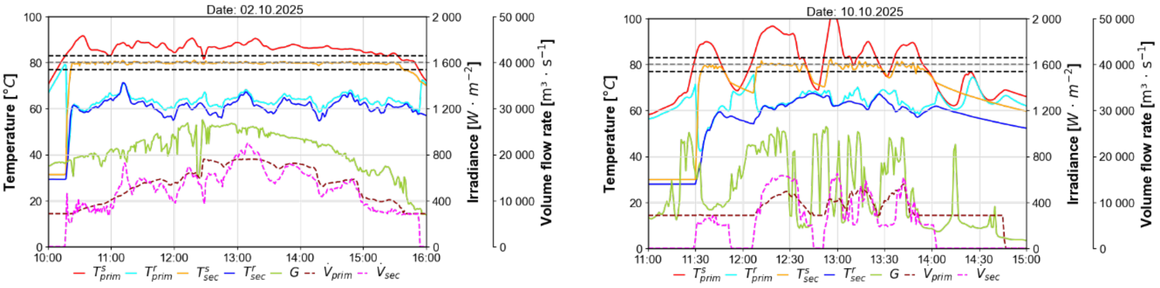

Figure 2: Measurement results for two exemplary days of adaptive controller testing

4. Conclusion

The results described in this work show a successful pre-training of an adaptive, ML-based control algorithm with a digital twin. The simulation demonstrates improved control performance compared to a static PID controller that was already optimized manually by experts. Deployment in the operational environment confirms the controller's real-time capability and further validates the potential to improve control performance further through continuous adaptation of PID parameters. Future work will focus on continuous testing in the real environment to evaluate the control performance for different seasons. Furthermore, a modification of the training process using datasets from different years will be examined aiming for reducing the risk of overfitting to a single training dataset. Finally, the transferability of the ML algorithm to a different hydraulic variant of the feed-in station will be analysed.

5. References

Enyekwe, I., Bai, W., Lee, K.Y., Nag, S., 2023. Speed Control of an Adjustable Speed Pumped Storage Hydropower Plant with a Deep Reinforcement Learning-Based Governor. DOI: 10.11159/jmids.2023.004

Korupu, V. L., Manimozhi, M., 2025. Leveraging Deep Reinforcement Learning for Effective PI Controller Tuning in Industrial Water Tank Systems. DOI: 10.48084/etasr.9602

Lillicrap, T.P., Hunt, J.J., Pritzel, A., Heess, N., Erez, T., Tassa, Y., Silver, D., Wierstra, D., 2015. Continuous control with deep reinforcement learning. arXiv:1509.02971.

Sabahi, K., Jamil, M., Shokri-Kalandaragh, Y., Tavan, M., Arya, Y., 2024. Deep Deterministic Policy Gradient Reinforcement Learning Based Adaptive PID Load Frequency Control of an AC Micro-Grid. DOI: 10.1109/ICJECE.2024.3353670

Ziegler, J. G.; Nichols, N. B., 1942. Optimum Settings for Automatic Controllers.

Schramedei, R., Schwutke, R., Ott, S., Walser, T., Klöck, M. & Schmidt, T., 2023. SWD.SOL II–Evaluierung der dezentralen Einbindung von solarer Wärme in das KWK-Fernwärmesystem der Stadtwerke Düsseldorf AG: Endbericht. DOI:10.2314/KXP:1856405486