Crystals 2020

Optimal Trajectory Tracking Control of Batch Crystallization Processes Using Reinforcement Learning

Paul Danny Anandan, Chris Rielly, Brahim Benyahia
Department of Chemical Engineering, Loughborough University,
Epinal Way, Loughborough, Leicestershire, UK – LE11 3TU.

Loughborough University

1. Introduction and Motivation:

- The control of the particle size distribution, crystal habit and crystal purity are crucial in most crystallization processes to meet the targeted critical quality attributes of the final product [1].
- This work tries to address the increasing demand for more advanced, versatile, robust and cost-effective control technologies for crystallization processes [2].
- Reinforcement Learning (RL), has gained a lot of interest for process control and optimization, while positively impacting both research and industries [3].
- This work proposes a novel RL method for optimal trajectory tracking and control of batch crystallization processes, reducing quality variation and wastes.

3. Overview of the Reward Function:

$$R_t = \left\{ \overbrace{\overbrace{(10(|e| < 0.1)}^{\text{Reward if the error is lesser than 0.1}} - \overbrace{1(|e| \ge 0.1))}^{\text{Penalty if the error is greater than or equal to 0.1}}}^{\text{Error}} + \overbrace{\left(\overbrace{\left(\left|\int edt\right| < 60\right)}^{\text{Reward if the integrated error is lesser than 60}} - 5\overbrace{\left(\left|\int edt\right| \ge 60\right)}^{\text{Penalty if the integrated error is greater than or equal to 60}} \right)}^{\text{Integrated Error}} \right\}$$

$$+ \underbrace{\underbrace{\left\{25\left({}^{T_s}/_{T_f}\right)\right\}}_{\text{Simulation Time}}}_{\text{Reward for every incremental time step of the simulation period}} + \underbrace{\underbrace{\{-100(280 \ge T \ge 315)\}}_{\text{Temperature Bounds}}}_{\text{Penalty if the simulation is stopped in the middle by exceeding the temperature bounds}}$$

<ins>*Note:*</ins>

R_t - the reward given to the RL agent at time step 't' of the simulation
T_s - the time interval after which the RL agent takes a control action
T_f - the final time step at which the simulation ends

Figure 2: Overview of defining a reward function

4. Overview of a RL Training Outcome:

- The blue plot denotes the sum of individual rewards gained by the RL agent at the end of each simulation.
- The red plot indicates the moving average of 20 simulations, and this value can be used to stop the training on reaching a certain target value.
- The green plot refers to the final score gained by the agent for its control actions for each simulation.

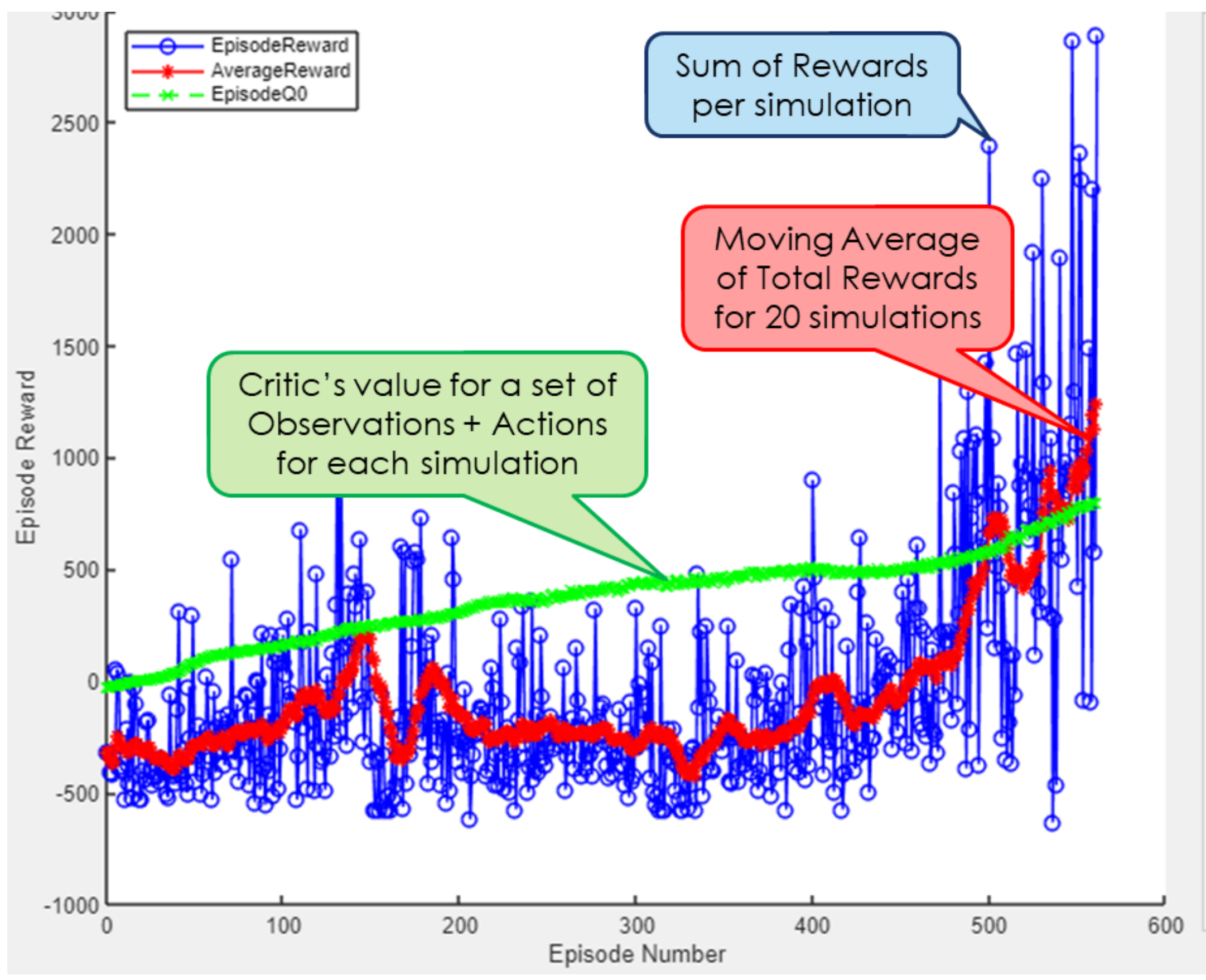

Figure 3: Overview of a RL Training Outcome

7. References:

1. D. Fysikopoulos, B. Benyahia, A. Borsos, Z. K. Nagy, and C. D. Rielly, "A framework for model reliability and estimability analysis of crystallization processes with multi-impurity multi-dimensional population balance models," *Computers and Chemical Engineering*, 2019.
2. R. Lakerveld and B. Benyahia, "Process Control," in *The Handbook of Continuous Crystallization*, Edited by Nima Yazdanpanah and Zoltan K Nagy. London: Royal Society of Chemistry, 2020, pp. 172–218.
3. P. Petsagkourakis, I. O. Sandoval, E. Bradford, D. Zhang, and E. A. del Rio-Chanona, "Reinforcement learning for batch bioprocess optimization," *Computers and Chemical Engineering*, vol. 133, p. 106649, 2020.
4. Z. K. Nagy, J. W. Chew, M. Fujiwara, and R. D. Braatz, "Comparative performance of concentration and temperature controlled batch crystallizations," *Journal of Process Control*, vol. 18, no. 3–4, pp. 399–407, 2008.

2. Problem Formulation:

- The objective is to train the RL Agent to achieve the targeted temperature, supersaturation and particle size dynamic profiles or trajectories by adjusting the cooling rate of a batch crystallizer.
- A mathematical model of the cooling crystallization of paracetamol in water, validated elsewhere [4], is used to train the agent, reduce the experimental burden and explore wider design and operating spaces.
- Several RL training strategies were implemented to enhance the performance of the reward functions and achieve robust and optimal training of the agent.

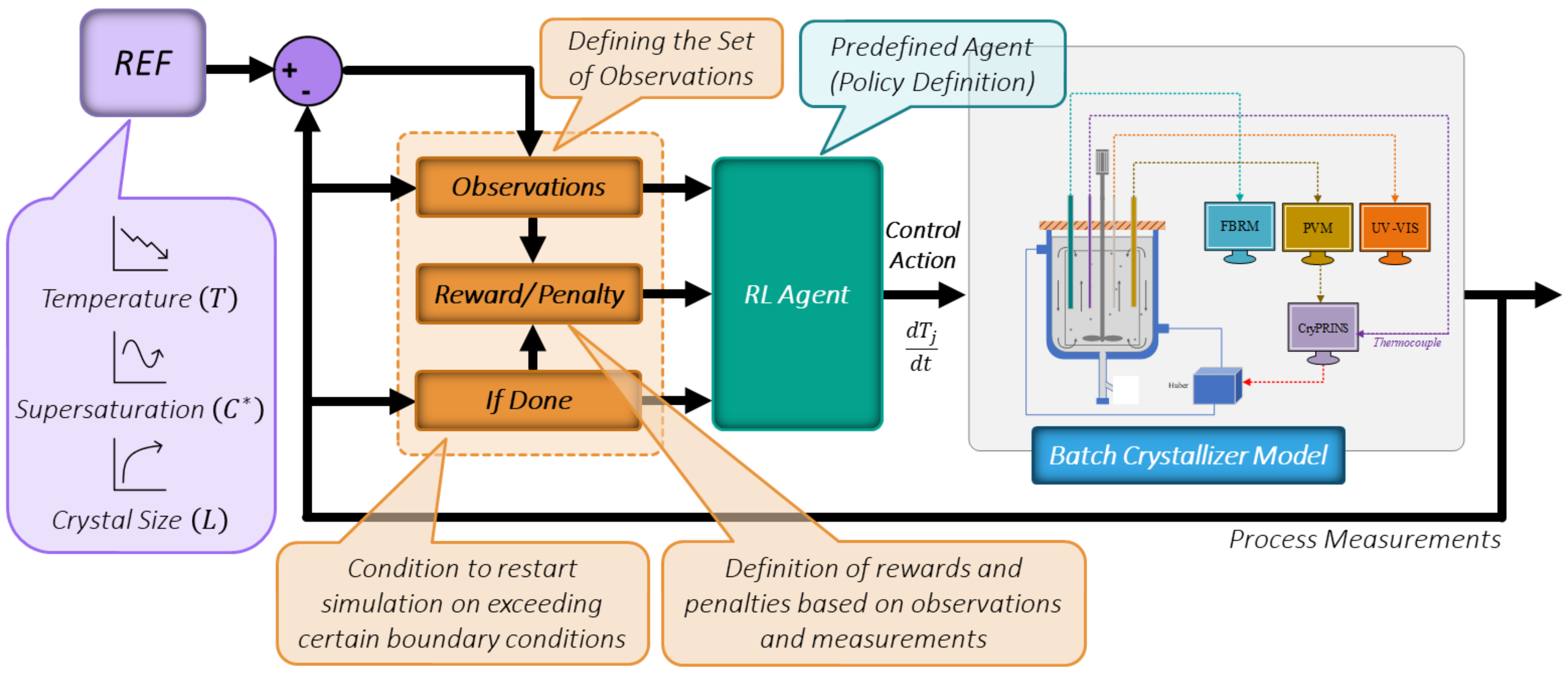

Figure 1: Overview of the RL Training Setup for the Batch Crystallisation Process

5. Results and Validation:

- Several RL training policies were implemented to continuously improve the efficiency of the reward functions and training outcome. The training performance is evaluated by the cost of the training (number of episodes) and the value of the attained maximum reward.
- The performance of the trained RL Agent was tested and compared against the traditional PID and MPC controllers. The results indicate that RL can provide equally competing and robust control performance.

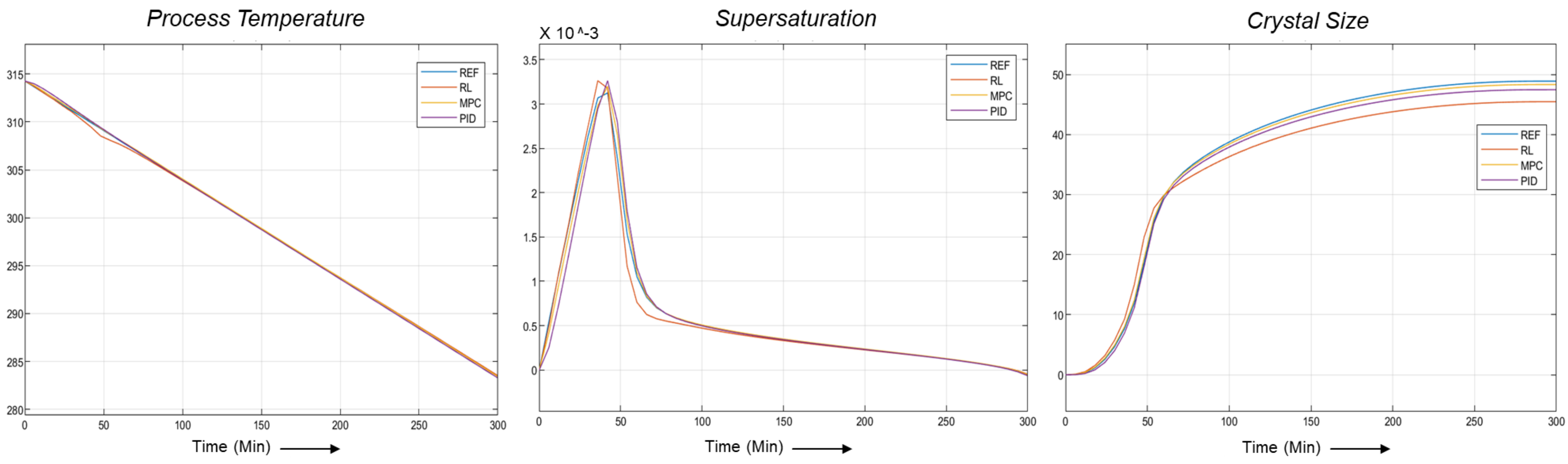

Figure 4: Performance comparison of the RL agent against the traditional PID and MPC Controllers

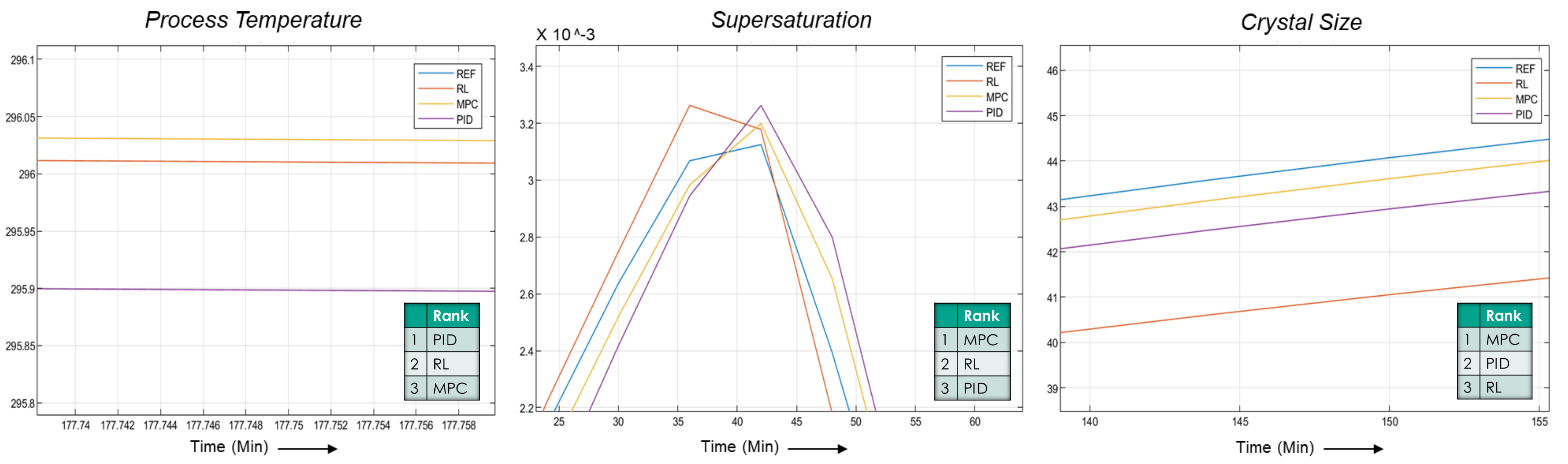

Figure 5: A closer view of the performance comparison and their rankings against each target reference profile

6. Conclusion and Future Work

- Model-based RL was achieved considering trajectory tracking control of the temperature, supersaturation and particle size in a batch (cooling) crystallization system.
- Various reward functions and training stargies were implemented to optimise the training performance and enhance robustness of the control strategy.
- Next step - Control the system against multiple uncertainties (Process/ Measurement noise) and compare the model-based RL against benchmark control strategies (e.g. NMPC). Develop Robust RL in presence of noisy measurements, control perturbation and model uncertainties.

Funding: The authors gratefully thank the financial support from the UK's EPSRC for funding (EP/R032858/1) the ARTICULAR Project.