Computer Modeling in Engineering & Sciences

Tech Science Press

https://doi.org/10.32604/cmes.2026.080261

ARTICLE

Intelligent Control of Parabolic Trough Collectors via Deep Reinforcement Learning

Marta Leal, Verónica Abad-Alcaraz, María del Mar Castilla and José Domingo Álvarez*

Department of Informatics, CIESOL—ceiA3, Ctra. Sacramento s/n, La Cañada de San Urbano, University of Almería, Almería, Spain

*Corresponding Author: José Domingo Álvarez. Email: jhervas@ual.es

Received: 05 February 2026; Accepted: 12 May 2026; Published: 27 July 2026

ABSTRACT: The effective control of parabolic trough collectors (PTCs) remains a significant challenge due to the inherent non-linearities of the system and the continuous impact of environmental disturbances. Although PTCs are a key technology for industrial process heat and large-scale electricity generation, classical control strategies often struggle to maintain optimal performance under fluctuating conditions. To address these limitations, this paper presents a novel reinforcement learning (RL)-based controller, designed specifically for solar thermal systems. The proposed RL agent is designed to learn directly from operational data, enabling it to adapt its control policy in real time to mitigate external disturbances. Experimental results demonstrate that the RL controller achieves a fast and well-damped closed-loop response, significantly outperforming traditional control benchmarks. Specifically, the RL controller is compared with a Proportional-Integral controller combined with a feedforward controller, and with a Model-based Predictive Controller. In all simulation-based comparisons, the RL controller outperforms the aforementioned controllers in terms of setpoint tracking or disturbance rejection. These results highlight the potential of machine learning to improve the operational reliability and efficiency of complex renewable energy systems.

KEYWORDS: Reinforcement learning control; control systems; parabolic trough collectors; solar thermal collectors

1 Introduction

The global solar thermal energy landscape presents a dual paradoxical scenario. On the one hand, the technology has consolidated a mature position, and its installed capacity and energy production consistently grow year after year. On the other hand, this progress is mitigated by a notable contraction in new installations, a dynamic that reflects the complexity of economic, political, and competitive factors at play in the market. Solar thermal energy has established a solid and growing base in the global energy mix. At the end of 2022, the cumulative operating solar thermal capacity was estimated at 542 gigawatt-thermal (GWth), which corresponds to 774 million square meters of collector area. This figure continued to expand, reaching a cumulative growth of 3% in 2023, for a total of approximately 560 GWth [1].

Historically, the residential sector has been the pillar of solar thermal energy globally. Domestic hot water heating accounts for the largest share of installed capacity, with small systems making up approximately 60% of new installations. This technology has been established as a simple and cost-effective solution for providing heat in homes, dominating the market for hot water systems. However, this residential segment is showing signs of maturation in some markets, and in regions like China and Europe, small systems have lost market share in recent years. This shift suggests that the residential market, while still important, is no longer the sole growth engine, forcing the industry to seek expansion in larger and higher-value applications to sustain its trajectory. In contrast, demand for large-scale projects is increasing, with several multi-megawatt plants under construction in 2022 for commercial and industrial clients, signalling a new era for big solar in those regions. Besides that, in some countries, interest also is rising in hybrid systems, particularly combined solar thermal and heat pump systems in district heating networks [2].

The Solar Heat for Industrial Processes (SHIP) segment is the clear indicator of the solar thermal industry's strategic evolution. The industrial sector accounts for approximately one-third of CO_2 emissions in the United States, and more than half of its heat needs operate in temperature ranges that solar thermal energy can provide efficiently (<400°C) [3]. This sector is one of the most difficult to decarbonize due to its high dependence on fossil fuels. The deployment of new SHIP projects has grown significantly, with a record of 114 projects in 2022 and 116 in 2023 [4]. The technology's ability to convert solar energy directly into heat, instead of electricity, gives it a unique competitive advantage in this sector.

The value of solar thermal energy lies in its ability to provide direct heat and dispatchable energy through its inherent thermal storage. This fundamental role distinguishes it from solar photovoltaic (PV) energy, which focuses on electricity production. As energy policies mature, moving from widespread subsidies to more selective incentives and stable regulatory frameworks, solar thermal energy is positioned as a key component for decarbonizing the heating sector, an often-overlooked but critical area for the global energy transition. Therefore, there is not competition between PV energy and solar thermal energy since are designed for different purposes. While PV converts solar radiation directly into electricity, solar thermal converts it directly into heat. The strategic value of solar thermal is not in competing in the electricity market dominated by PV, but in its ability to decarbonize the heating sector, which accounts for almost two-thirds of energy use for heating and is a crucial challenge for the energy transition [5].

Parabolic Trough Collectors (PTCs) are a cornerstone of the solar thermal landscape, playing a crucial role in both industrial process heat and electricity generation. Their importance stems from a combination of their technological maturity, high efficiency, and the ability to integrate with thermal energy storage. PTCs represent the most widely used Concentrated Solar Power (CSP) technology [6], specifically are line-focusing systems that use long, curved mirrors to concentrate solar radiation onto a receiver tube positioned along the focal line. The receiver tube contains a Heat Transfer Fluid (HTF) that is heated up by concentrated solar radiation, and the resulting thermal energy can be used for a thermal industry process or used to generate steam and supply a turbine to produce electricity [6]. PTCs are used for medium- and high- heating applications, and research is being conducted on developing the structure, material, shape, and size of PTCs to enhance their efficiency [7–9]. However, controlling these types of systems to improve their performance is also an active field of research. One of the main control objectives in PTC systems is to maintain the average temperature of the solar field around a setpoint set by the operators. Thus, the controller manipulates the HTF flow to regulate its outlet temperature considering some disturbances, as solar irradiance or HTF inlet temperature. In adition, PTCs are distributed parameter systems with strong non-linearities that must be taken into account by the controller [10].

During the last decades several control system architectures have been tested in PTCs systems. Focusing on recent years, it is possible to find in literature works that develop classical Proportional Integral (PI) controllers for temperature reference traking plus a FeedForward (FF) controller with the aim to reject the disturbances dynamics [11], whereas other works are focused on the use of more advanced controllers as a novel Model Predictive Control (MPC) algorithm based on the Fast Fourier Transform (FFT) to control the temperature of the collector field [12]. Besides that, some works take into account the distributed nature of the PTC, that can be modelled by Partial Diferential Ecuations (PDE), to develop a robust bilinear setpoint tracking through a constrained controller based on Lyapunov stability theory [13]. Focusing on disturbances and taking into account that any type of solar system is constantly subject to disturbances caused by solar irradiance, some controllers focus on forecasting solar irradiance using some type of predictor, such as a Kalman filter [14]. In the last years, nonlinear MPC (NMPC) has become a dominant framework for PTC systems due to its ability to explicitly handle nonlinearities and constraints. For instance, reference [15] proposed an MPC scheme complemented with an advanced meteorological disturbance model, explicitly incorporating cloud cover and weather dynamics into the prediction horizon. This approach improves robustness in the face of changing environmental conditions, which are often the main cause of reduced performance in solar thermal systems, demonstrating that the integration of disturbance prediction improves setpoint tracking and the operational stability of the MPC. Similarly, reference [16] develop an optimal MPC-based disturbance rejection controller for nonlinear PTC systems with lumped uncertainties. Their formulation focuses on compensating unmeasured disturbances and model mismatch, achieving improved regulation performance. The study highlights the importance of robust NMPC formulations, particularly when simplified models are used for real-time feasibility. To address the computational burden and modeling limitations of classical NMPC, reference [17] introduced a learning-based practical NMPC, where data-driven elements are used to approximate system behavior and improve prediction accuracy. The approach maintains NMPC structure while reducing computational complexity, enabling real-time applicability in solar flat collectors. This study shows that learning-based enhancements can achieve performance comparable to full NMPC while mitigating the need for highly detailed first-principles models.

Finally, it is possible to find in literature some examples of intelligent control and Machine Learning (ML) in this kind of solar plants. For instance, reference [18] presented an adaptive controller for the temperature of a concentrated solar thermal plant, specifically a Fresnel plant, based on deep Reinforcement Learning (RL). The authors use a deep deterministic policy gradient algorithm to continuously adjust the parameters of a PI controller. The effectiveness of the controller is validated through simulations using a verified linear model of the Fresnel plant. On the other hand, reference [19] developed a hybrid fuzzy convolution model (HFCM) that takes full advantages of data, models (dynamic and steady-state), and prior knowledge on industrial distributetd solar field. The HFCM is then extended for use in deep deterministic policy gradient (DDPG) algorithm to learn the control task. The DDPG agent is compared in simulation with other controllers obtaining best results in temperature tracking and energy gain. At last, reference [20] introduced a hybrid control strategy for PTCs that combines artificial neural networks (ANNs) in a feedforward structure with a conventional feedback controller. The main objective is to address the strong nonlinearities and disturbances—particularly solar irradiance fluctuations—that affect outlet temperature regulation through the feedforward controller based on ANNs. This reduces the workload on the feedback controller, which is responsible for ensuring stability and correcting residual errors. The simulation results show that the proposed approach achieves better tracking performance, a faster response and greater disturbance rejection compared to strategies based solely on models or feedback.

In this work, a RL controller based on a DDPG agent is developed to regulate the outlet temperature of a PTC. A key distinction of this research is the use of a high-fidelity non-linear distributed model of the PTC to train the agent, allowing it to directly account for the distributed nature and dynamic non-linearities of this kind of solar plant. Unlike previous studies that utilize the RL agent as an adaptive tuner for classical PI parameters or rely on supervised ANN for feedforward-feedback mapping, our approach employs the DDPG agent as a direct controller for the volumetric flow rate. Furthermore, while other advanced RL frameworks integrate hybrid fuzzy convolution models to address data scarcity and multi-objective energy-gain optimization, this work prioritizes setpoint tracking, i.e., thermal regulation, and robust disturbance rejection including the disturbances in the reward function. Simulation results show that the DDPG agent trained with data achieves a faster, better-damped closed-loop response and superior tracking performance compared to classical PID controllers, and modern MPC strategies.

The rest of the paper is organized as follows: Section 2 presents the detailed modeling of the parabolic trough collector, including its thermal dynamics, linearization, and transfer function representation. Section 3 describes the control architectures employed to regulate the outlet temperature, covering classical PI and feedforward strategies, a MPC approach, and an RL framework based on the DDPG algorithm. Section 4 presents the results, comparing the performance of the proposed controllers under different disturbance scenarios, and Section 5 provides a discussion of the findings and concluding remarks.

2 Model of the System

The case study analysed in this work focuses on a distributed solar collector field that employs PTC technology. A comprehensive description of this type of plant, along with detailed information on the physical parameters, can be found in [21]. A parabolic trough solar power plant can be regarded as a tubular heat exchanger, a configuration that is widely employed in the process industry. Therefore, the knowledge gained from operating such systems can also be transferred to various conventional industrial applications [22].

Fig. 1 presents a schematic diagram of the absorber tube of a PTC, illustrating the temperatures and all the relevant variables considered in its energy balance and modeling. Under standard assumptions and hypotheses, the distributed solar collector field can be modelled using a temperature-based distributed parameter approach. In this formulation, the system dynamics are governed by a set of PDEs that describe the energy balance, as expressed in Eq. (1). By applying the principle of energy conservation to a control volume of length Δx over a time interval dt, the governing equation for the fluid flowing through the absorber tube is derived.

$$A_i \rho C \frac{\partial T}{\partial t}(x,t) + \dot{q} \rho C \frac{\partial T}{\partial x}(x,t) = \pi D_i h_i (T_\omega(x,t) - T(x,t)) \tag{1}$$

where $\dot{q} = A_i \cdot v$, $T(x,t)$ is the HTF temperature and $T_\omega(x,t)$ is the pipe wall temperature.

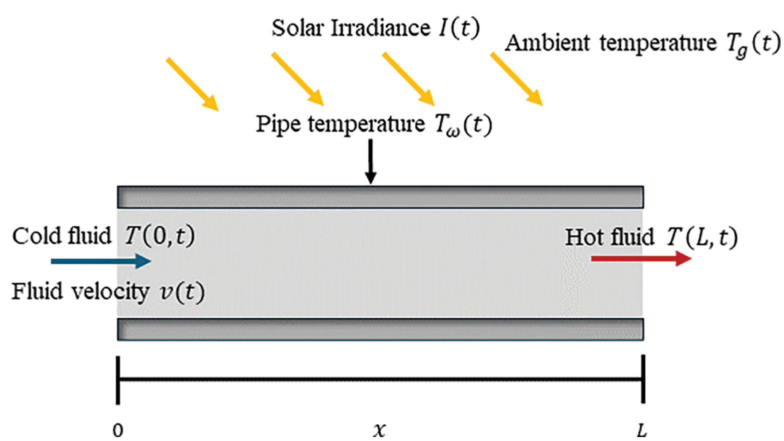


Figure 1: Absorber tube diagram of a parabolic trough collector.

Similarly, an energy balance can be applied to the pipe wall of the collector, resulting in the expression shown in Eq. (2).

$$\rho_\omega C_\omega A_o \frac{\partial T_\omega}{\partial t}(x,t) = I(t)\eta_o G - \pi D_o h_o \left(T_\omega(x,t) - T_g(x,t)\right) - \pi D_i h_i \left(T_\omega(x,t)\right) - T(x,t)) \quad (2)$$

where $T_g(x,t)$ indicates the ambient temperature.

The definitions of the parameters used in Eqs. (1) and (2), as well as in subsequent equations of this mathematical development, are provided in Table 1.

Table 1: Model parameters of a tubular solar heat exchanger.

Parameter	Description	Unit
γ	Auxiliary parameter	$^\circ C \cdot m^2/J$
η_0	Collector optical efficiency	–
ρ	HTF density	kg/m^3
ρ_ω	Pipe density	kg/m^3
τ_1	T-related fundamental time constant against T_ω variations	s
τ_{12}	T_ω-related fundamental time constant against T variations	s
τ_2	T_ω-related fundamental time constant against T_g variations	s
A_i	Inner pipe area	m^2
A_o	Outer pipe area	m^2
C	Specific heat capacity of the HTF	$J/(kg^\circ C)$
C_ω	Specific heat capacity of the pipe	$J/(kg^\circ C)$
D_i	Inner pipe diameter	m
D_o	Outer pipe diameter	m
G	Collector aperture	m
h_i	Pipe convective HTC	$W/(m^{2\circ}C)$
h_o	External pipe convective HTC	$W/(m^{2\circ}C)$
$I(t)$	Solar irradiance	W/m^2
L	Pipe length	m
$\dot{q}(t)$	Volumetric flow rate	m^3/s
$T(x,t)$	HTF temperature	$^\circ C$
$T_g(t)$	Ambient temperature	$^\circ C$
$T_\omega(x,t)$	Pipe temperature	$^\circ C$
t	Time	s
$v(t)$	HTF velocity	m/s
x	Space	m

Note: HTC: Heat transfer coefficient; HTF: Heat Transfer fluid.

The model is considered semi-physical as it combines pre-existing insights into system dynamics. It adopts a hybrid approach, merging theoretical concepts with empirical data and includes adjustable parameters that enable significant physical interpretation. The simulations carried out to evaluate the performance of the proposed control strategies are based on the non-linear collector model defined in Eqs. (1) and (2). However, to implement classical linear control techniques, it is necessary to derive a linear model. For this purpose, a system linearization was performed by means of a Taylor series expansion of the non-linear terms. Subsequently, the resulting expressions were simplified, leading to Eqs. (3) and (4).

$$\frac{\partial T}{\partial t} = -(v - v_s)\frac{dT_s}{dx} - v_s\frac{\partial T}{\partial x} + \frac{1}{\tau_1}(T_w - T) \tag{3}$$

$$\frac{\partial T_w}{\partial t} = I\gamma - \frac{1}{\tau_2}(T_w - T_g) - \frac{1}{\tau_{12}}(T_w - T) \tag{4}$$

where τ_1, τ_2 and τ_{12} are time constants relating the temperatures and γ is an auxiliary parameter.

In this study, the analysis is focused on the linearized transfer functions that relate the main system inputs to the HTF outlet temperature, $T(L,s)$. Specifically, the influence of the HTF inlet temperature, $T(0,s)$, solar irradiance, $I(s)$, and HTF velocity, $v(s)$, on the outlet temperature is considered. The ambient temperature, $T_g(s)$, is assumed to remain approximately constant during normal operation and is therefore not explicitly addressed in the control-oriented analysis. Under these assumptions, the system can be represented by individual Single Input Single Output (SISO) transfer functions linking each input variable to the outlet temperature.

Eqs. (3) and (4) are linear but still in partial derivative terms, through Laplace transform it is possible to find the equivalent irrational transfer functions. The detailed mathematical procedure to find the transfer functions is presented in [22]. As a result, the transfer functions relating the HTF outlet temperature to the HTF inlet temperature, solar irradiance, and HTF velocity are obtained and reported in Eqs. (5)–(7), respectively. These transfer functions provide a suitable basis for the design of feedback or feedforward control strategies aimed at regulating the HTF outlet temperature under varying inlet, solar, and flow conditions.

$$\frac{T(L,s)}{T(0,s)} = e^{-\frac{L}{v}s}\left(K_D\frac{-\beta s + 1}{\tau s + 1}\right) \tag{5}$$

$$\frac{T(L,s)}{I(s)} = \frac{k_I}{s^3 + a_1 s^2 + a_2 s + a_3}\left(1 - e^{-\frac{L}{v}s}\left(K_D\frac{-\beta s + 1}{\tau s + 1}\right)\right) \tag{6}$$

$$\frac{T(L,s)}{v(s)} = \frac{k_v(b_0 s + 1)}{s^2 + a_1 s}\left(1 - e^{-\frac{L}{v}s}\left(\frac{-\beta s + 1}{\tau s + 1}\right)\right) \tag{7}$$

where K_D, k_I, k_v refer to static gains, the coefficient b_0 is a zero that appear in Eq. (7), whereas a_1, a_2, a_3 denote the coefficients of the polynomial in the denominator appearing in Eqs. (6) and (7). Note that a_1, β and τ in Eqs. (5)–(7) are general parameters and do not have the same value.

The main control objective in a solar collector field is to maintain the HTF outlet temperature at a defined reference value, denoted as T_{ref}, which is manually set during plant operation. This outlet temperature can be regulated by adjusting the HTF velocity, which directly determines the volumetric flow rate of the HTF, $\dot{q}(s)$. The flow is driven by a pump. In this work, it is assumed that the pump's operating range is between 2 and 12 l/s. The plant consists of ten parallel collector loops, each operating within a flow range of approximately 0.2 to 1.2 l/s, assuming a perfectly balanced distribution. The volumetric flow rate is obtained by multiplying the HTF velocity by the internal cross-sectional area of the pipe, A_i, $\dot{q} = A_i \cdot v$, making the HTF velocity the only variable that can be directly manipulated. The non-linear characteristics of the system, together with the influence of external disturbances such as solar irradiance, HTF inlet temperature, and ambient temperature, increase the complexity of the control task and must be carefully considered when designing effective control strategies.

3 Control Architecture

This section describes the different control architectures used to regulate the outlet HTF temperature, $T(L,t)$, ranging from classical ones to artificial intelligence based ones. Firstly, the PTC has been controlled using a conventional PI controller. In addition, this controller has been improved by adding a feedforward to the proposed architecture in order to counteract undesirable behaviour caused by disturbances. Subsequently, a more advanced controller, an MPC approach, has been implemented. Finally, an RL based control architecture has been developed. This last technique is not based on a model, such as the MPC or PI controller, but on data, that is, it is configured with a large amount of data to learn what the best control action is in each situation.

3.1 Classical Control Strategies

In this section, two classical control schemes are presented. The first one is a conventional PI controller, widely used due to its simplicity and effectiveness in a variety of applications. The second approach enhances the classical PI controller by incorporating a feedforward component into its design, aiming to improve the system's dynamic response and disturbance rejection capability.

3.1.1 PID Control

The Proportional–Integral–Derivative (PID) controller represents the most widely employed feedback control strategy in industrial practice. Developed and refined over many decades, it has remained a fundamental element of classical control theory. The control signal generated by a PID controller can be mathematically expressed as shown in Eq. (8):

$$u(t) = K\left[e(t) + \frac{1}{T_i}\int_0^t e(\tau)\,d\tau + T_d\frac{de(t)}{dt}\right] \tag{8}$$

The control action is composed of three terms, each corresponding to a different temporal aspect of the error signal. The proportional term (P) reacts to the present error, the integral term (I) accounts for the accumulated past error, and the derivative term (D) provides a prediction of the future error by estimating its rate of change. Specifically, the derivative term, $T_d\frac{de(t)}{dt}$, can be interpreted as a linear extrapolation of the error T_d time units ahead. The controller parameters are the proportional gain K, the integral time constant T_i, and the derivative time constant T_d. While this structure provides flexibility and accurate control, the derivative term is often sensitive to noise and measurement fluctuations. For this reason, many practical implementations adopt a simplified PI configuration, which offers an effective trade-off between performance and robustness [23].

In order to design the PI controller for the PTC, it is essential to obtain a first or second transfer function that relates the manipulated variable, the volumetric flow rate $\dot{q}(s)$, to the controlled variable, the outlet temperature of the absorber tube, $T(L,s)$. The parameters of this transfer functions have been identified through reaction curve tests conducted in simulations of the non linear model, as shown in Eq. (9).

$$G(s) = \frac{T(L,s)}{\dot{q}(s)} = \frac{k_g}{\tau_g s + 1} \tag{9}$$

where k_g refers to the static gain in $\mathrm{m}^3/(\mathrm{s}^\circ\mathrm{C})$, whereas τ_g is the time constant in s.

The PI controller to be designed is a dynamic system with the transfer function given in Eq. (10):

$$C(s) = \frac{U(s)}{E(s)} = k_p\left(1 + \frac{1}{T_i s}\right) = k_p\left(\frac{1 + T_i s}{T_i s}\right) \tag{10}$$

The tuning method selected, given that the process is modeled as a first-order system without time delay, is the Pole–Zero Cancellation Method, which is a classical analytical tuning technique. This method is based on cancelling the process pole with the zero introduced by the PI controller. To achieve this, the integral time constant of the controller, T_i, is set equal to the time constant of the first-order process, τ_g, as shown in Eq. (11):

$$T_i = \tau_g \tag{11}$$

This condition ensures that the pole of the process $(1 + \tau_g s)$ is cancelled by the zero of the controller $\left(1 + \frac{1}{T_i s}\right)$, resulting in a simplified closed-loop dynamics.

The open-loop transfer function is given by Eq. (12):

$$G_{cd}(s) = C(s)G(s) = k_p\left(1 + \frac{1}{T_i s}\right)\frac{k_g}{\tau_g s + 1} \tag{12}$$

After applying the pole-zero cancellation ($T_i = \tau_g$), the closed-loop transfer function becomes Eq. (13):

$$G_{bc}(s) = \frac{G_{cd}(s)}{1 + G_{cd}(s)} = \frac{1}{1 + s\frac{T_i}{k_p k_g}} \tag{13}$$

This corresponds to a first-order system with unit steady-state gain and a closed-loop time constant defined in Eq. (14):

$$\tau_{bc} = \frac{T_i}{k_p k_g} \tag{14}$$

Finally, by selecting the desired closed-loop time constant τ_{bc}, the proportional gain k_p can be calculated as in Eq. (15):

$$k_p = \frac{T_i}{k_g \tau_{bc}} \tag{15}$$

3.1.2 Feedforward Control

Although feedback control is effective in many applications, it cannot by itself eliminate the effects of disturbances on the process output. Feedforward control complements feedback action by anticipating the impact of measurable disturbances and compensating for them before they affect system behaviour. This well-established technique, widely implemented in industrial processes, is often combined with PID controllers to enhance overall performance [24].

To design the feedforward controller, it is essential to obtain the transfer functions that relate the manipulated variable, the volumetric flow rate $\dot{q}(s)$, to the disturbances: the solar irradiance $I(s)$ and the HTF inlet temperature $T(0,s)$. These transfer functions were identified through reaction curve tests conducted on simulations of the non-linear model, as shown in Eqs. (16) and (17):

$$D_1(s) = \frac{T(L,s)}{I(s)} = \frac{k_{d_1}}{\tau_{d_1} s + 1} \tag{16}$$

$$D_2(s) = \frac{T(L,s)}{T(0,s)} = \frac{k_{d_2}}{\tau_{d_2}s+1}e^{-t_r s} \tag{17}$$

where k_{d_1} and k_{d_2} denote the static gains, expressed in °C m^2/W and dimensionless [−], respectively; τ_{d_1} and τ_{d_2} are the time constants, given in s; and t_r represents the transport time delay, also expressed in s.

For each disturbance independently, the proposed feedforward structure adopts a lead–lag configuration, as shown in Eqs. (18) and (19), respectively.

$$F_{ff}(s) = \frac{k_{d_1}}{k_g}\frac{\tau_g s+1}{\tau_{d_1}s+1} \tag{18}$$

$$F_{ff}(s) = \frac{k_{d_2}}{k_g}\frac{\tau_g s+1}{\tau_{d_2}s+1}e^{-t_r s} \tag{19}$$

Fig. 2 illustrates the classical feedforward control architecture, combining a PI controller with the feedforward action to improve disturbance rejection.

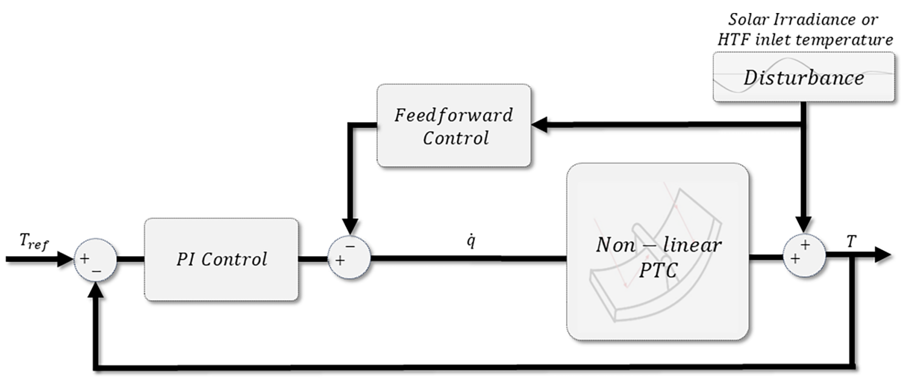


Figure 2: Classical feedforward control architecture for PTC.

3.2 Model Predictive Control

As a more advanced control strategy, a constrained MPC strategy was designed using the linear model of the parabolic trough collector described in Section 2, following the classical formulation presented [25]. The controller uses a discrete-time prediction model of the outlet temperature based on the identified transfer functions between the manipulated variable (HTF flow rate) and the plant outputs, together with the main measured disturbances. This model is internally represented in state-space form and is updated every T_p seconds, where T_p denotes the sample time of the controller.

At each control interval, MPC computes a finite sequence of future control moves that minimises a quadratic performance index over a prediction horizon N_p. Only the first element of this sequence is applied to the plant, and the optimisation is repeated at the next sample time following the standard sliding-horizon principle. The formulation includes explicit constraints on the range of the manipulated variable and its rate of change, as well as an output disturbance model to guarantee offset-free tracking and good disturbance-rejection properties. The controller was configured and tuned and a Kalman filter is used for internal state estimator.

3.2.1 Cost Function

The objective of the MPC algorithm is to keep the outlet temperature close to a reference trajectory while avoiding excessive variations in the HTF flow rate. This trade-off is expressed through the following quadratic cost function, as defined in Eq. (20):

$$J(t) = \sum_{i=1}^{N_p} \lambda_y \left(\hat{y}(t+i|t) - r(t+i)\right)^2 + \lambda_u \sum_{i=0}^{N_c-1} u(t+i|t)^2 + \lambda_{\Delta u} \sum_{i=0}^{N_c-1} \Delta u(t+i|t)^2, \tag{20}$$

where $\hat{y}(t+i|t)$ is the predicted outlet temperature at future step $t+i$ based on the information available at time t, and $r(t+i)$ is the corresponding reference trajectory. The term $u(t+i|t)$ denotes the future values of the manipulated variable and $\Delta u(t+i|t) = u(t+i|t) - u(t+i-1|t)$ is the increment of the control signal. The horizons N_p and N_c represent the prediction and control horizons, respectively, with $N_c \le N_p$.

The weighting coefficients λ_y, λ_u and $\lambda_{\Delta u}$ adjust the relative importance of each contribution to the cost. The first term penalises tracking error and drives the outlet temperature towards its reference. The second term penalises large absolute values of the HTF flow rate, indirectly accounting for actuator limitations and energy consumption. The third term penalises rapid changes in the control signal, promoting smoother actuator motion and improving robustness with respect to model mismatch.

The prediction $\hat{y}(t+i|t)$ is obtained by propagating the internal state-space model over the horizon, taking into account the current measured disturbance values and, when available, preview information of future reference and disturbance profiles, as can be seen in Fig. 3. This configuration is closely related to the classical Generalised Predictive Control (GPC) formulation, where the plant model is augmented with integrating dynamics and the optimisation is performed over predicted output trajectories. As in GPC, the use of an incremental control parametrisation and an explicit disturbance model enhances steady-state accuracy and disturbance rejection.

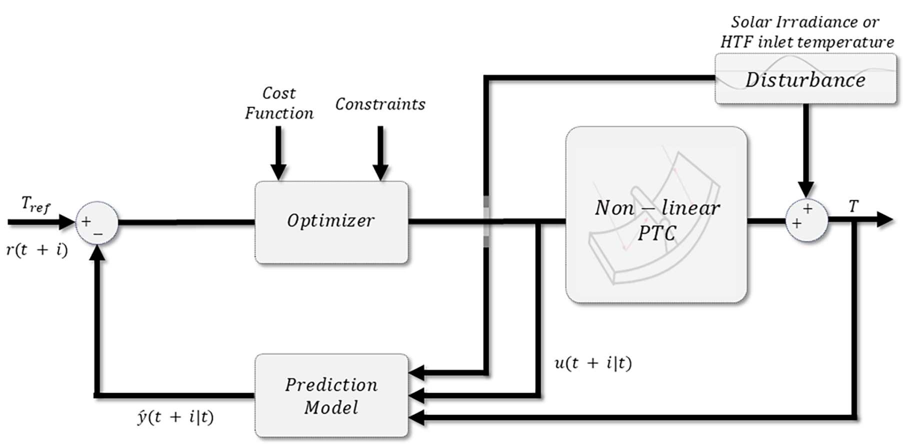


Figure 3: MPC architecture for PTC.

3.2.2 Constraints and Disturbance Modelling

The MPC formulation explicitly incorporates physical and operational constraints of the PTC. The manipulated variable is bounded within an admissible interval, as expressed in Eq. (21), which reflects the minimum and maximum HTF flow rates deliverable by the circulation pump.

$$u_{\min} \le u(t+i|t) \le u_{\max}. \tag{21}$$

In addition, rate constraints of the form given in Eq. (22), are imposed to limit how quickly the valve opening or pump speed can change between successive sampling instants, protecting the actuator and avoiding abrupt transients. In general, output constraints can be specified to keep the outlet temperature within a safe operating envelope. However, in the present implementation no explicit bounds were imposed on the output, so that all the active constraints in the optimisation problem are hard constraints applied only to the manipulated variable and its rate of change.

$$\Delta u_{\min} \leq \Delta u(t+i|t) \leq \Delta u_{\max}. \tag{22}$$

To achieve offset-free tracking in the presence of slowly varying disturbances, an unmeasured output disturbance model was included for the controlled temperature. It is implemented as a low-order stochastic model driven by white noise of ramp type. The estimator reconstructs this disturbance state from past measurements and compensates its effect over the prediction horizon, significantly reducing steady-state errors and improving rejection of external thermal perturbations.

Besides unmeasured disturbances, the controller also exploits preview information on measured noise that can be forecast or scheduled in advance, such as variations in solar input or inlet-fluid temperature. These measured disturbances are treated as exogenous signals entering the prediction model and are assumed known over the horizon. Providing this preview allows the optimiser to anticipate their effect on the outlet temperature and to generate proactive control actions, effectively combining feedback and feedforward behaviour within a single predictive framework.

Overall, the MPC scheme represents a constrained GPC-type controller based on a linear model of the PTC, with explicit handling of actuator limits, rate restrictions and disturbance dynamics. This architecture serves as a strong baseline for comparison with the proposed RL controller in terms of setpoint tracking, disturbance rejection and control effort.

3.3 Reinforcement Learning

The operation of the RL control strategy can be defined as a continuous interaction between the agent and its environment. At each control interval, the environment provides the agent with a set of observations that describe the current state of the PTC. Based on this information, the agent evaluates the situation, decides the most appropriate control action, and applies it to the system—modifying, in this case, the volumetric flow rate $\dot{q}(s)$ of the HTF.

After the action is executed, the environment evolves to a new state according to its non-linear dynamics, and a numerical reward is returned to the agent, reflecting the quality of the chosen action with respect to the control objective. Through this iterative process of observation, action, and feedback, the agent progressively improves its decision-making policy, aiming to maximize the cumulative reward over time.

Fig. 4 depicts the control architecture designed to regulate the outlet temperature of the absorber tube, $T(L, s)$, through the HTF flow rate. The following paragraphs describe in detail the main components of this architecture (environment, observations, agent and reward function), as well as the training procedure adopted to obtain the final RL based controller.

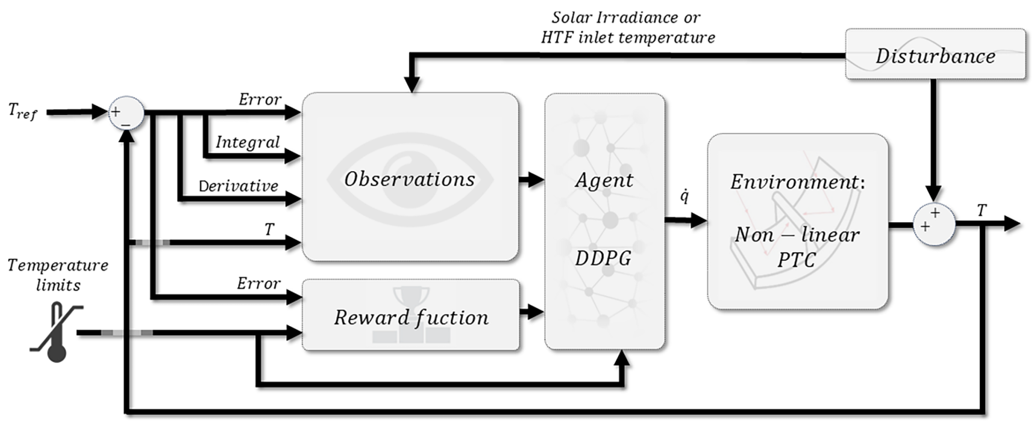


Figure 4: RL based control architecture for PTC.

3.3.1 Environment

The environment corresponds to the non-linear PTC model described in Section 2, which serves as the physical and dynamic system with which the RL agent interacts. Through this environment, the agent observes the system state, applies a control action—modifying the volumetric flow rate of the HTF—and receives the corresponding reward.

During the training process, a customised reset function randomises the initial operating conditions at the start of each episode, varying the reference temperature, HTF inlet temperature and solar irradiation within realistic ranges. This strategy exposes the agent to different disturbance scenarios, allowing it to develop a more robust and generalised control policy.

3.3.2 Observations

The state of the environment is described through a set of measurable variables that encapsulate the most relevant information for the control task. These variables, known as observations, are the inputs provided to the agent at each decision step and represent both the internal dynamics of the PTC and the external disturbances acting on it. Their correct selection is crucial, as they determine the agent's ability to perceive the system state and to learn an effective control policy.

In this implementation, five key elements are defined as observations, combining information about the tracking error, system dynamics, and external conditions:

- Current error: The direct difference between the desired outlet temperature and the current measured value. It represents the main feedback signal for the agent, guiding its corrective actions at each time step.
- Integral of the error: A cumulative measure that quantifies the long-term deviation between the outlet temperature of the PTC and the reference value. This term allows the agent to evaluate the persistence of the error over time and contributes to minimizing steady-state deviations.
- Derivative of the error: The temporal rate of change of the error, which provides information about the trend and speed of response of the system. By incorporating this term, the agent can anticipate the evolution of the process and mitigate oscillatory behavior.
- PTC outlet temperature: The controlled variable of interest and a direct indicator of process performance. Monitoring this temperature enables the agent to continuously assess the effectiveness of its control actions.

- Disturbances: In addition to the internal state variables, the agent also receives information related to the main external disturbances that affect the PTC's thermal behavior. Considering these perturbations is essential for achieving a realistic representation of the operating conditions and for developing a control policy capable of active disturbance rejection. This study has taken into account the two most influential disturbances in the PTC:
  - Solar irradiance: One of the primary disturbances acting on the process. Fluctuations in irradiance directly affect the thermal input received by the collector and, therefore, the outlet temperature. Including this variable among the observations allows the agent to recognize and compensate for the dynamic effects of solar variability.
  - HTF inlet temperature: A secondary disturbance that significantly influences the thermal balance of the system. Variations in this variable change the initial energy level of the working fluid entering the collector, modifying its outlet temperature after a delay, even under constant solar conditions.

It is important to highlight that the first three observations are similar to the actions performed by a PID controller, while the disturbance observations perform the function of a feedforward controller. All observations are continuously measured and updated during the simulation, forming the state vector that feeds the agent. This real-time processing ensures that the RL controller can adapt its decisions to the instantaneous dynamics of the PTC, achieving robust performance even in the presence of non-linearities and rapid environmental changes.

The observations provided to the RL agent are directly obtained from the simulation environment without applying explicit normalization or scaling techniques. Instead, all variables are inherently bounded within physically meaningful ranges through the environment definition. In particular, the observation space is constrained using predefined lower and upper limits, and all process variables (e.g., outlet temperature, disturbances, and error-related signals) evolve within realistic operating intervals. To enhance robustness and improve generalization, a stochastic reset strategy is applied at the beginning of each training episode. The reference temperature, inlet temperature, and solar irradiance are randomly sampled within predefined ranges that reflect typical operating conditions of the plant. Specifically, the reference temperature is drawn from a truncated normal distribution, while inlet temperature and irradiance are sampled within bounded intervals. This domain randomization approach exposes the agent to a wide range of operating scenarios and disturbance conditions during training, allowing it to learn a generalized control policy without requiring explicit input normalization. No additional measurement noise filtering or signal smoothing is applied, as the training is conducted in a simulation environment. However, variability in operating conditions is introduced through the randomized initialization of disturbances, which provides a degree of robustness against uncertainties.

3.3.3 Agent

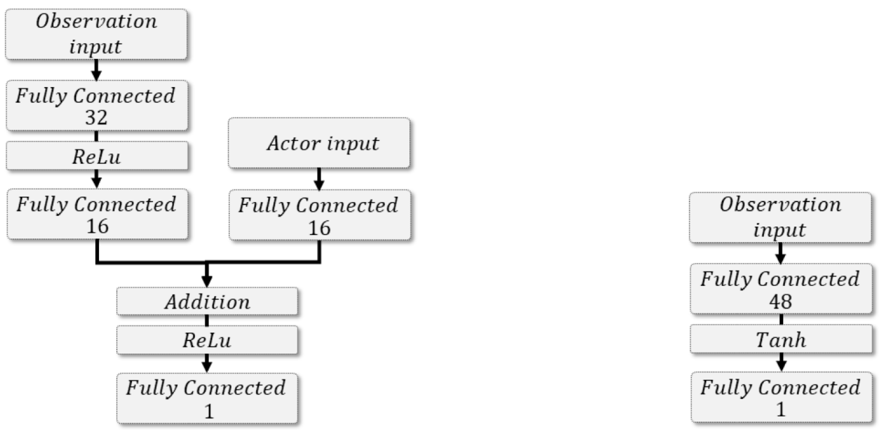
The agent constitutes the decision-making core of the control strategy and is responsible for learning the optimal policy that determines the appropriate volumetric flow rate at each time step. Its objective is to maximize the cumulative reward obtained from the environment by maintaining the collector outlet temperature as close as possible to its reference value, even in the presence of external disturbances.

Among the different agents with continuous action space, the DDPG agent has been chosen for this work due to its training speed and robustness. The DDPG agent has an actor-critic architecture, in which two neural networks are trained simultaneously:

- The actor network represents the policy function and generates a continuous control action based on the current observations. In this case, the action corresponds to the HTF volumetric flow rate within the operating limits of the pump (0%–100%) which correspond to the lower and upper limits of the HTF volumetric flow (2–12 l/s).

- The critic network estimates the state–action value function $Q(s, a)$, which assesses how good a given action is for a particular system state. The critic is trained to minimize the error between predicted and target Q-values, using gradient-based optimization to provide accurate feedback to the actor.

Both networks were implemented as fully connected feedforward architectures, whose structures are shown in Fig. 5. On one hand, the critic network (Fig. 5a) consists of two independent input branches—one for the observations vector and another for the action signal—that are later merged through an addition layer. This configuration allows the network to jointly process state and action information before estimating the scalar Q-value at the output neuron. On the other hand, the actor network (Fig. 5b) maps the observation input to a continuous control action through a sequence of dense layers with non-linear activation functions (ReLU and tanh), ensuring smooth and bounded responses suitable for continuous control tasks.



(a) Critic network. (b) Actor network.

Figure 5: Architectures of the deep neural networks used in the DDPG agent.

Other RL algorithms such as Twin Delayed Deep Deterministic Policy Gradient (TD3) were also evaluated in a previous work on this same plant configuration, showing inferior performance under the tested disturbance scenarios [26]. Therefore, in the present study only the DDPG agent was retained, given its better balance between control precision, learning stability, and implementation simplicity.

3.3.4 Reward

A reward function was defined to drive the learning process toward the desired control objectives while ensuring safe operation of the PTC. The design of the reward takes into account both the tracking performance and the thermal limits of the system, penalizing actions that lead to unsafe conditions or poor regulation of the outlet temperature.

From the standpoint of stability, safety and training efficiency, a valid operating range for the PTC outlet temperature has been defined between 180°C and 300°C. This interval guarantees that the HTF remains within safe operating limits, avoiding excessive overheating or underheating of the system. If at any time the outlet temperature $T_{\text{out}}(t)$ exceeds these limits, a strong penalty of $B_d = -300$ is applied and the episode is terminated to prevent further deviation. Besides that, the stopping condition for the reward function does not involve reaching a certain value, but rather achieving an average value over the last N executions. It allows the DDPG agent learning a policy able to maintain the controlled system among some stable limits in order to avoid penalties.

Within the admissible range, the instantaneous reward depends on the absolute tracking error, defined as the difference between the outlet temperature and its reference value $T_{\text{ref}}(t)$. When this error is smaller than 0.95°C, the agent receives a positive reward of 10, which reinforces precise tracking behaviour. Otherwise, a penalty of −1 is assigned, discouraging actions that increase the deviation from the target temperature. The overall reward function can be expressed as shown in Eq. (23).

$$\text{RF}(t) = \begin{cases} 10 + B_d, & \text{if } |T_{\text{ref}}(t) - T_{\text{out}}(t)| < 0.95, \\ -1 + B_d, & \text{if } |T_{\text{ref}}(t) - T_{\text{out}}(t)| \geq 0.95, \end{cases} \tag{23}$$

where the boundary penalty term B_d is defined as:

$$B_d = \begin{cases} 0, & \text{if } T_{\text{out}}(t) \in [180, 300], \\ -300, & \text{if } T_{\text{out}}(t) \notin [180, 300]. \end{cases} \tag{24}$$

This formulation encourages the agent to maintain accurate temperature tracking while strictly enforcing safety limits, see Eq. (24). The use of a large negative penalty for out-of-bound states accelerates learning convergence by clearly distinguishing desirable actions from unsafe ones.

4 Results

After defining the different control architectures, the implementation was carried out in MATLAB [27]. The performance of the proposed control strategies was assessed through two case studies. Firstly, the behaviour of a DDPG-based RL agent was compared with classical control strategies, namely a PI controller and a feedforward controller with standard compensation. Secondly, the performance of the DDPG-based RL controller was compared with that of an MPC strategy.

In both case studies are considered two independent disturbance scenarios. In the first scenario, the system was subjected to variations in solar irradiance, $I(t)$, while the HTF inlet temperature, $T(0,t)$, and the ambient temperature, $T_g(t)$, were kept constant at their reference values. In the second scenario, the disturbance was applied to the HTF inlet temperature, $T(0,t)$, while maintaining constant values of irradiance, $I(t)$, and ambient temperature, $T_g(t)$. These comparisons aim to evaluate the ability of the RL agent to adapt to different operating conditions in contrast to classical control approaches. It is important to highlight that, this study did not take ambient temperature into account, as it changes slowly and within a narrow range during the operation of this type of plant, and its influence on the dynamics of the HTF outlet temperature is minimal.

For all cases, the agents were trained with a sampling time of 10 s for reward calculation, simulating up to 5000 episodes to ensure sufficient exploration and convergence of the learning algorithm. Each episode was initialised with different starting conditions. Specifically, in the first scenario, the irradiance $I(t)$ varied within the range of 400–800 W/m^2, while $T(0,t)$ and $T_g(t)$ were kept constant. Conversely, in the second scenario, the HTF inlet temperature $T(0,t)$ varied between 100°C and 300°C, whereas $I(t)$ and $T_g(t)$ remained fixed. In both cases, the desired HTF outlet temperature T_{ref} ranged from 100°C to 240°C.

The training process for the agents was terminated either when the average reward over the last N episodes exceeded a predefined threshold, or when the maximum number of episodes was reached, ensuring robustness since the trained agent is able to achieve consistently high reward values across different scenarios. The most relevant hyperparameters employed for training the RL agent are summarised in Table 2.

Table 2: Main hyperparameters employed for training the DDPG agent.

Parameter	Description	Value/Unit
T_p	Sample time for reward calculation	10 s
γ	Discount factor	1.0
τ	Target smooth factor	1×10^{-3}
Mini-batch size	Number of samples per learning iteration	64
Experience buffer length	Size of replay memory	1×10^{6}
Actor learning rate	Learning rate of the policy network	1×10^{-4}
Critic learning rate	Learning rate of the value network	1×10^{-3}
Gradient threshold	Limit for gradient clipping	1
Noise variance	Initial variance of exploration noise	0.3
Noise decay rate	Decay rate of exploration noise	1×10^{-5}
Actor network structure	Fully connected layers: 48 (tanh activation)	–
Critic network structure	Observation path: 32, 16; Action path: 16 (ReLU activation)	–
Training episodes	Maximum number of episodes	5000

The hyperparameters reported in Table 2 were selected through a systematic trial-and-error process based on multiple training experiments. The default values from the *MATLAB Reinforcement Learning Toolbox* were retained when they proved adequate, and only those parameters with a significant impact on learning performance were adjusted. Specifically, the default DDPG configuration uses a discount factor of $\gamma = 0.99$, actor and critic learning rates of 0.01, an experience buffer of 10^4 samples, and 256 hidden units per layer. Several of these defaults were modified after empirical testing. The discount factor was set to $\gamma = 1.0$ to avoid penalising long-horizon rewards in a continuous regulation task. The actor and critic learning rates were reduced to 1×10^{-4} and 1×10^{-3}, respectively, to improve training stability and prevent oscillatory updates. In addition, the experience buffer was expanded to 1×10^{6} in order to retain a richer and more diverse set of past transitions.

Regarding the neural network architecture, the number of nodes per layer was deliberately kept moderate, with 48 units in the actor and 32 and 16 units in the critic. This choice was made after observing that larger networks did not consistently improve performance and, in several experiments, prevented the agent from learning altogether, likely due to over-parameterisation relative to the complexity of the control task. The remaining hyperparameters, including the mini-batch size (64), gradient threshold (1), noise variance (0.3), and noise decay rate (1×10^{-5}), were also tuned empirically and set to values that provided a good balance between exploration and convergence speed.

Finally, to compare the performance of the proposed controllers in each case, the results were analyzed using several well-established performance criteria. These quantitative criteria provide a comprehensive evaluation of the controller response in terms of tracking accuracy and control effort. The first index considered is the Integral of Absolute Error (IAE), defined in Eq. (25):

$$\mathrm{IAE} = \int_0^{t_f} |e(t)|\, dt \tag{25}$$

where $e(t)$ represents the error between the reference and the output, defined in Eq. (26):

$$e(t) = T_{\mathrm{ref}}(t) - T(x,t) \tag{26}$$

The other error-based indices include the Integral of Squared Error (ISE) and the Integral of Time-weighted Absolute Error (ITAE), expressed in Eqs. (27) and (28), respectively:

$$\text{ISE} = \int_0^{t_f} e^2(t)\, dt \tag{27}$$

$$\text{ITAE} = \int_0^{t_f} |t\, e(t)|\, dt \tag{28}$$

In addition, the Control Effort Index (CEI), defined in Eq. (29), was employed to quantify the magnitude of the control signal variation required to achieve the desired response:

$$\text{CEI} = \int_0^{t_f} |\Delta u(t)|\, dt \tag{29}$$

To provide a more complete characterization of the transient response, three additional time-domain performance indicators are also considered: the percentage overshoot (OS), the rise time (t_r), and the settling time (t_s). These metrics are widely used in control engineering to evaluate the dynamic behaviour of closed-loop systems.

The overshoot quantifies the maximum deviation of the system output with respect to the steady-state reference value. The rise time is defined as the time elapsed from the onset of the system response until the output first reaches the steady-state reference value. The settling time is defined as the time required for the system output to enter and remain within a ±2% band around the steady-state reference value.

Both t_r and t_s are computed from simulations in which a step change is applied to the reference while the disturbance inputs (irradiance and inlet temperature) remain constant, ensuring that the measured transient reflects the closed-loop dynamics exclusively, without distortion introduced by time-varying disturbances.

These additional indices are included in the analysis to complement the integral-based performance measures, allowing a more detailed evaluation of both tracking quality and transient dynamics under different disturbance scenarios.

During the simulation phase, the disturbance signals are defined using time-varying profiles obtained from predefined datasets, rather than artificially generated signals. In particular, solar irradiance and inlet temperature are introduced as discrete time series, where the first column represents time and the second column the corresponding variable value. These profiles capture realistic operating conditions, including both smooth variations and abrupt changes.

When solar irradiance is considered as the varying disturbance, the remaining variables are kept constant ($T_{\text{ref}} = 225°\text{C}$, $T_g = 28.02°\text{C}$, and $T_{\text{in}} = 183.03°\text{C}$). Conversely, when inlet temperature is varied, it is defined through an analogous time-dependent profile while the other variables remain fixed.

Before being applied to the environment, the disturbance signals are preprocessed to ensure physically consistent behaviour. Specifically, the signals are adjusted with respect to a baseline operating point and passed through a first-order low-pass filter with the transfer function defined in Eq. (30).

$$G(s) = \frac{1}{40s + 1}. \tag{30}$$

This filtering stage smooths high-frequency variations and introduces realistic dynamics consistent with the thermal response of the plant.

The use of experimentally inspired disturbance profiles, together with the filtering stage, ensures that the input signals remain representative of real operating conditions while avoiding unrealistically fast transients that could affect the stability of the controller.

4.1 Case 1. Comparison of DDPG Based RL Control Architecture with Classical Control Strategies

First, the results obtained under solar irradiance disturbances, $I(t)$, are presented. To tune the PI controller, a low-order model of the plant was derived following the methodology described in Section 3.1. This approach led to the identification of low-order transfer functions relating both the irradiance, $I(s)$, and the volumetric flow rate, $\dot{q}(s)$, to the outlet temperature, $T(L,s)$, as shown in Eqs. (31) and (32).

$$G(s) = \frac{T(L,s)}{\dot{q}(s)} = \frac{-7.16 \cdot 10^4}{222.404s + 1} \tag{31}$$

$$D_1(s) = \frac{T(L,s)}{I(s)} = \frac{0.1179}{254.062s + 1} \tag{32}$$

The remaining disturbances of the system, $T_g(s)$ and $T(0,s)$, were assumed to be constant for the simulations performed in this work. The PI controller was tuned following a pole-zero cancellation approach. An aggressive design was adopted, using a closed-loop time constant $\tau_{bc} = 0.7\tau$, which resulted in a proportional gain of $k_p = -1.9952 \times 10^{-5}\,^\circ\mathrm{C\,s\,m^{-3}}$ and an integral time of $T_i = 222.404$ s. The low numerical value of k_p is due to the use of volumetric flow rate expressed in m^3/s instead of l/s. Additionally, the gain is divided by the number of control loops (ten), which further reduces its magnitude. This clarification has been included to avoid misunderstandings. This closed-loop configuration ensures a fast response while maintaining stability.

The transfer function obtained by applying the classical feedforward structure defined in Eq. (18) is presented in Eq. (33).

$$F_{ff_1}(s) = -1.646 \times 10^{-6} \frac{222.404s + 1}{254.062s + 1} \tag{33}$$

Fig. 6 presents the simulation results comparing the three control strategies considered. The control goal is maintain the outlet temperature as close as possible from a desired reference of 225°C starting from a temperature of 235°C. The upper plot illustrates the temporal evolution of the outlet temperature under the DDPG-based RL agent (in blue), the classical PI controller (in orange) and the combined PI plus feedforward (PI+FF) controller (in purple), alongside the reference temperature (dash black line). The middle plot shows the corresponding control signals, highlighting how the volumetric flow rate varies over time to regulate the outlet temperature at the reference value of 225°C. The lower plot displays the solar irradiance profile for the selected day of operation, which ranges from 448 to 855 $\mathrm{W/m^2}$.

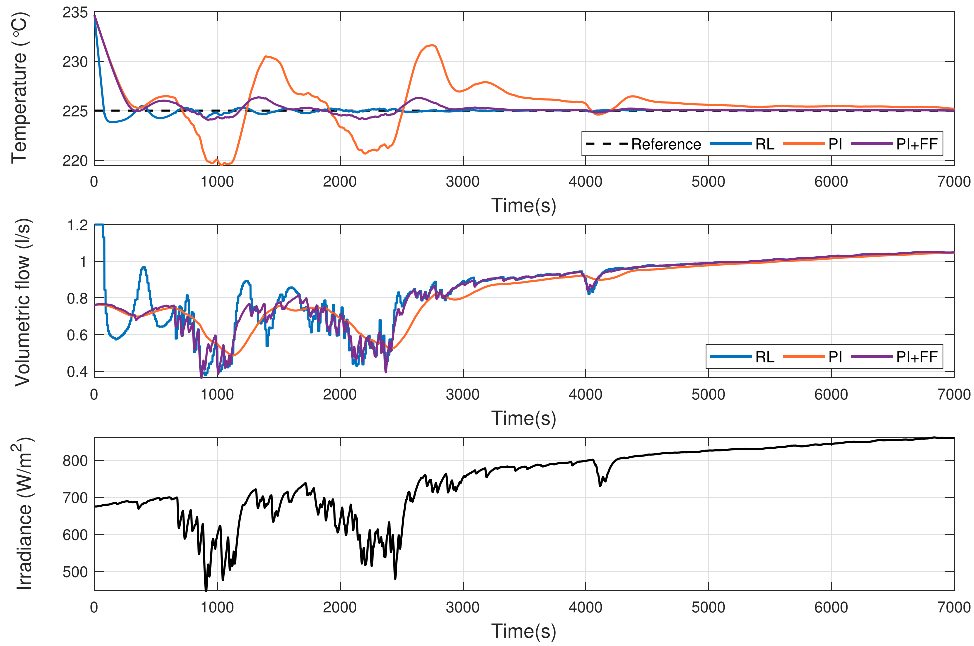


Figure 6: Simulation results of DDPG-based RL, PI and PI+FF controllers under solar irradiance disturbance.

The comparison reveals clear differences in tracking performance and control effort among the three strategies. The DDPG-based RL agent achieves the best reference tracking, with minimal overshoot and fast settling, reflected in the lowest IAE, ISE, and ITAE values (see Table 3), which are 123.12, 267.28, and 1.46×10^5, respectively. As expected, the PI+FF controller improves the disturbance rejection and tracking performance of the classical PI controller by reducing overshoot and oscillations, which is confirmed by its intermediate IAE, ISE, and ITAE values (305.01, 922.51, and 3.71×10^5; see Table 3). However, this improvement comes at the cost of increased control effort, as indicated by the higher CEI value (7.60) compared to the classical PI controller (1.55). The classical PI controller shows significant deviations from the reference, particularly during periods of high irradiance variability, resulting in the highest IAE, ISE, and ITAE values and the lowest CEI, indicating minimal but less effective control action. It is important to note that, in this type of system, in order to effectively reject the undesirable behaviour caused by solar radiation on the outlet temperature, the controlled variable must 'copy' the shape of the solar irradiance, see the central graph in Fig. 6, which causes high CEI index values but, on the other hand, is able of maintaining the outlet temperature close to its reference even against cloud transients.

Table 3: Performance criteria of different controllers under solar irradiance disturbance.

Controller	IAE	ISE	ITAE	CEI	OS (%)	t_r (s)	t_s (s)
DDPG	123.12	267.28	1.46×10^5	10.06	11.17	78	41
PID Controller	1233.79	4.41×10^3	2.81×10^6	1.55	–	715	229.39
PID Controller + FF	305.01	922.51	3.71×10^5	7.60	0.5	319	142

As observed in Table 3, there exists a clear trade-off between transient speed and overshoot: controllers exhibiting a higher percentage overshoot tend to achieve a faster rise time and settling time. This is consistent with classical control theory, where more aggressive closed-loop responses typically reduce the time to reach the reference value at the cost of increased overshoot.

The DDPG controller achieves the fastest response (t_r = 78 s, t_s = 41 s) with a moderate overshoot of 11.17%, while the PID+FF controller reaches the reference more slowly (t_r = 319 s, t_s = 142 s) but with a negligible overshoot of 0.5%. The classical PID controller, despite having no measurable overshoot, presents the slowest transient response (t_r = 715 s, t_s = 229 s), consistent with its poor integral-based performance metrics.

Additional scenario: aggressive irradiance profile with reference change

To further assess the robustness of the proposed controllers under more demanding operating conditions, an additional simulation scenario was considered using a real measured irradiance profile characterised by strong variability and a severe drop to approximately 300 W/m^2 around t = 700 s, followed by a gradual recovery. This scenario is significantly more challenging than those presented above, as the large irradiance transient drives the outlet temperature well below the operating reference, reaching physically unattainable conditions at the minimum irradiance point. To account for this, a reference change from 225°C to 200°C was imposed at t = 650 s, coinciding with the onset of the irradiance drop, and reverted once the irradiance recovered above approximately 1000 W/m^2.

Fig. 7 shows the comparison between the DDPG-based RL controller, the PI, and the PI+FF controller. Under these highly demanding conditions, the DDPG-based RL controller achieves the lowest IAE value (429.29), outperforming all classical baseline strategies. The PI controller exhibits the poorest tracking performance, with large and sustained deviations from the reference throughout the simulation (IAE = 2800.18). The PI+FF controller improves substantially upon the classical PI (IAE = 570.87), but still cannot match the RL performance. The quantitative performance criteria are reported in Table 4.

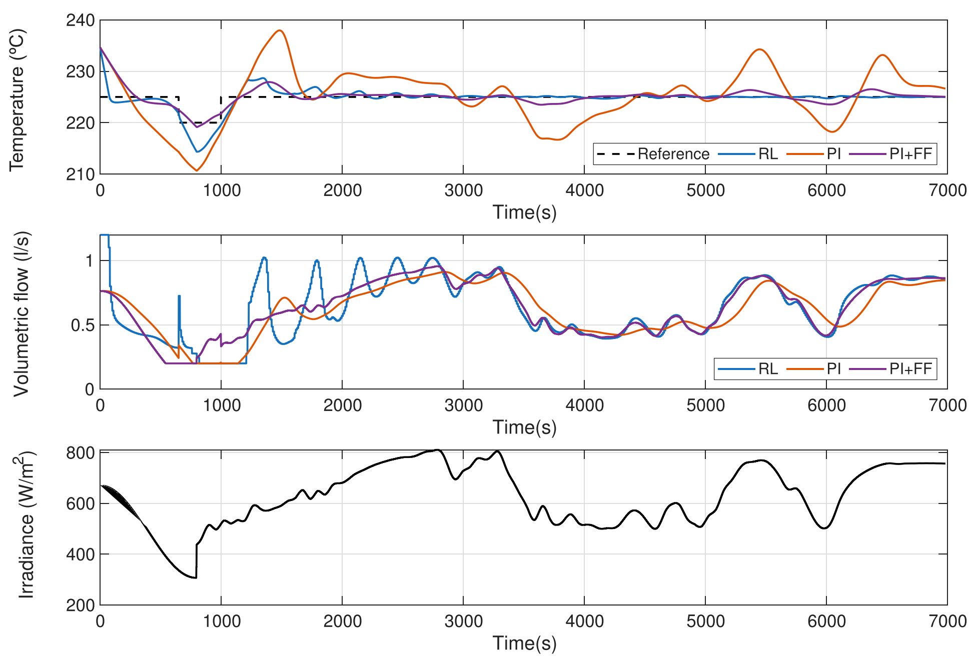


Figure 7: Simulation results of DDPG-based RL, PI and PI+FF controllers under aggressive solar irradiance disturbance with reference change.

Table 4: Performance criteria of different controllers under aggressive solar irradiance disturbance.

Controller	IAE	ISE	ITAE	CEI	OS (%)
DDPG	429.29	1194	5.97×10^5	9.68	73.42
PID Controller	2800.18	1.72×10^4	8.94×10^6	3.79	247.1
PID Controller + FF	570.87	1348	1.43×10^6	4.87	58

Secondly, the results obtained under HTF inlet temperature perturbations, $T(0,t)$, are presented. To tune the PI controller, a low-order transfer function relating $T(0, t)$ to the outlet temperature, $T(L,t)$, as presented in Eq. (34), is used.

$$D_2(s) = \frac{T(L,s)}{T(0,s)} = \frac{0.8115}{192.38s+1}e^{-250s} \tag{34}$$

The transfer function obtained by applying the classical feedforward is presented in Eq. (35).

$$F_{ff_2}(s) = -1.13 \times 10^{-5}\frac{222.404s+1}{192.38s+1}e^{-250s} \tag{35}$$

Fig. 8 presents the simulation results comparing the three control strategies considered. The upper plot illustrates the temporal evolution of the outlet temperature under the DDPG-based RL agent (in blue), the classical PI controller (in orange) and the combined PI plus feedforward (PI+FF) controller (in purple), alongside the reference temperature (dash black line). The middle plot depicts the corresponding control signals, showing how the volumetric flow rate varies over time to regulate the outlet temperature at the reference value of 225°C. Finally, the lower plot displays the HTF inlet temperature profile for the selected day of operation, which ranges from 180°C to 190°C.

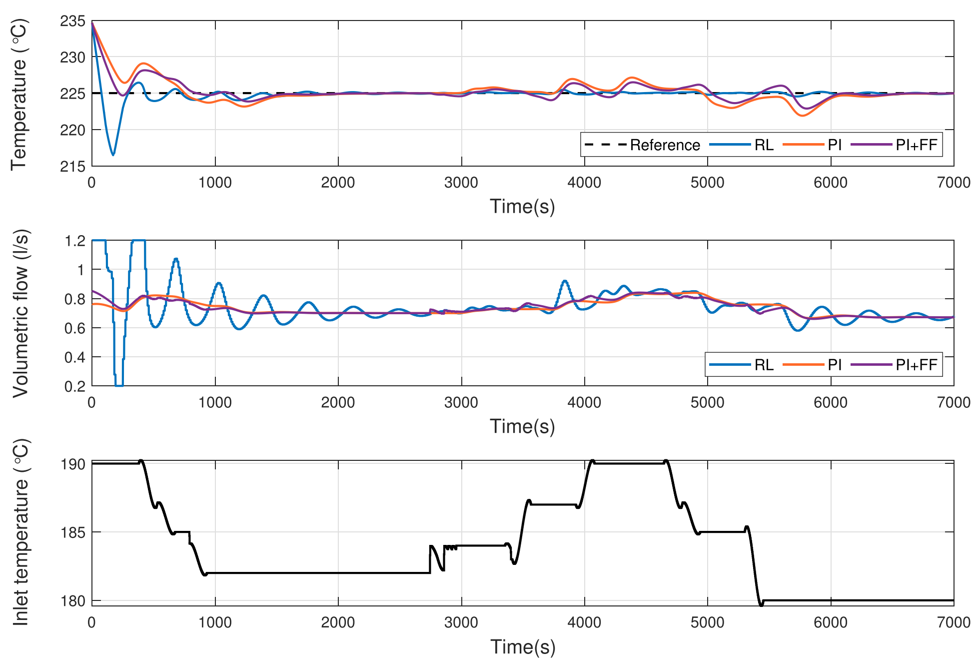


Figure 8: Simulation results of DDPG-based RL, PI and PI+FF controllers under HTF inlet temperature disturbance.

The comparison highlights notable differences in tracking performance and control effort under inlet temperature disturbances. The DDPG-based RL agent, after several oscillations in the control signal, achieves the best reference tracking, maintaining the outlet temperature closest to the desired setpoint, with reduced overshoot and no oscillations in the controlled variable once the desired reference has been reached, as confirmed by the lowest IAE, ISE, and ITAE values (see Table 5). The classical PI controller shows the poorest tracking performance, with large deviations from the reference during periods of rapid variation in the inlet temperature, resulting in the highest IAE, ISE, and ITAE values. The PI+FF controller exhibits intermediate tracking performance, improving on the classical PI controller by reducing overshoot and settling time, though it still cannot match the RL agent, as reflected in its IAE, ISE, and ITAE values. It is important to note that, both the PI and the PI+FF controller exhibit moderate oscillations at the middle of the simulation due to changes in the inlet temperature, whereas the DDPG-based RL agent is able to maintain the HTF outlet temperature in its reference value.

Table 5: Performance criteria of different controllers under HTF inlet temperature disturbance.

Controller	IAE	ISE	ITAE	CEI	OS (%)
DDPG	261.50	892.12	3.37×10^5	7.35	85.31
PID Controller	680.01	1.75×10^3	1.82×10^6	0.69	
PID Controller + FF	481.54	1.08×10^3	1.27×10^6	1.25	3.24

Regarding control effort, it is possible to find the same trade-off between performance and control effort, the DDPG-based RL agent requires a moderately high CEI (7.35) to achieve its superior performance, whereas both classical PI-based strategies operate with significantly lower CEI values (0.69 for PI and 1.25 for PI+FF), indicating more conservative but less effective control actions.

These results highlight the capability of the RL controller to handle the non-linear dynamics and external disturbances inherent to the PTC system, outperforming classical control strategies in terms of tracking accuracy. The addition of a feedforward term improves PI performance at the expense of increased control effort, but even so, the RL controller's results are better. It should be noted that, although classical PI controllers can be tuned immediately, RL-based controllers require a training period before achieving consistent performance.

4.2 Case 2. Comparison of DDPG Based RL Control Architecture with MPC

Before analysing the simulation results, Table 6 summarises the MPC configurations employed in the two disturbance scenarios. The controller was independently tuned for each case to achieve an appropriate trade-off between tracking performance and control effort. As a result, only the input increment weight, $\lambda_{\Delta u}$, and the disturbance and noise models differ between the irradiance-disturbance case and the HTF inlet-temperature–disturbance case. All remaining parameters, including the prediction and control horizons, sample time, weighting coefficients, λ_u and λ_y, as well as the actuator and rate constraints, are identical in both configurations.

Table 6: MPC configuration parameters for the two disturbance scenarios.

Parameter	Symbol	Unit	Irradiance Case	HTF Inlet-T Case
Prediction horizon	N_p	–	30	30
Control horizon	N_c	–	10	10
Sample time	T_l	s	7	7
Input weight	λ_u	–	0	0
Increment weight	$\lambda_{\Delta u}$	–	0.68	5.58
Output weight	λ_y	–	1.47×10^{-4}	1.47×10^{-4}
Min. flow rate	$u_{\min}$	l/s	−0.37	−0.37
Max. flow rate	$u_{\max}$	l/s	0.63	0.63
Min. rate change	$\Delta u_{\min}$	l/s	−0.1	−0.1
Max. rate change	$\Delta u_{\max}$	l/s	0.1	0.1
Disturbance model	–	–	$182.4/s^2$	$1.26/s^2$
Noise model	–	–	0.11	1.58

First, the results obtained when considering only solar irradiance as the measurable disturbance are presented. Fig. 9 shows the comparison between the two control strategies: DDPG-based RL agent (in blue) and MPC (in green). The upper plot illustrates the temporal evolution of the outlet temperature together with the reference signal (in black). Both controllers accurately track the 225°C set point with small deviations; however, DDPG-based RL agent achieves a faster and smoother settling response even with a smoother oscilations in the control signal, whereas MPC exhibits slightly larger oscillations during the transient period.

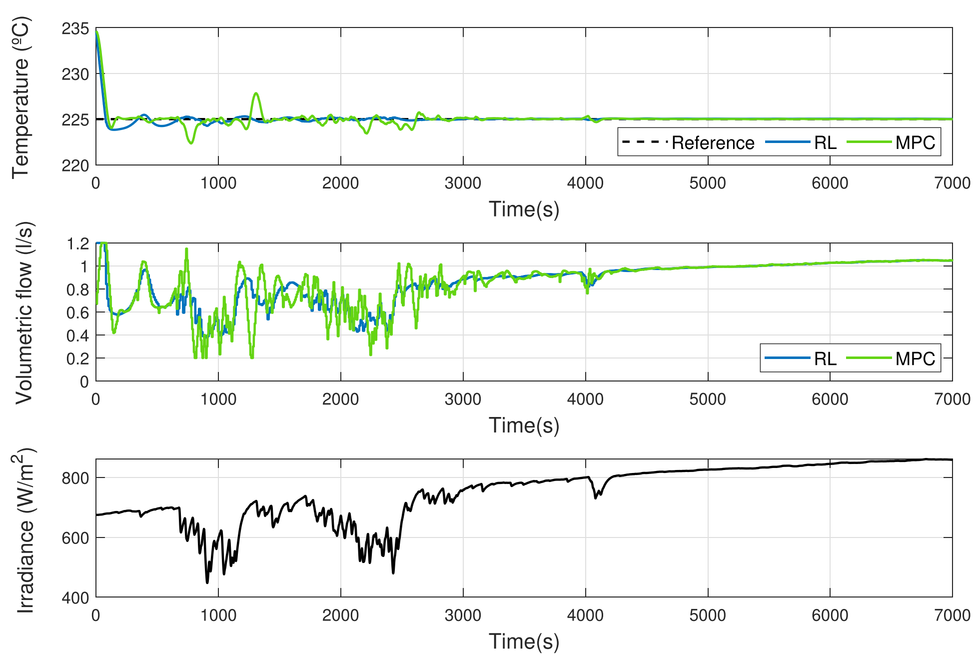


Figure 9: Simulation results of DDPG-based RL and MPC controllers under solar irradiance disturbance.

The middle plot shows the corresponding control signals, representing the volumetric flow rate variations over time. It can be observed that MPC produces more aggressive control actions, especially at the beginning of the simulation, whereas DDPG-based RL agent displays smoother actuation and lower fluctuations once steady-state conditions are reached.

The lower plot displays the solar irradiance profile for the considered operating period, which acts as the measurable disturbance. Despite the irradiance variations, both controllers manage to maintain the outlet temperature close to the reference value, demonstrating satisfactory disturbance rejection and robustness.

The quantitative comparison of control performance, expressed in terms of the IAE, ISE, ITAE, and CEI indices, is provided in Table 7, where only the DDPG-based RL control architecture and MPC controllers are compared. The comparison highlights that DDPG-based RL agent provides a slightly better transient response and lower control effort than MPC, which is consistent with the performance indices reported in Table 7.

Table 7: Comparative performance criteria between DDPG and MPC controllers under solar irradiance disturbance.

Controller	IAE	ISE	ITAE	CEI	OS (%)	t_r (s)
DDPG	123.12	267.28	1.46×10^5	10.06	11.17	100
MPC	180.09	1474.3	2.37×10^5	28.19	9.23	86

The numerical results confirm the observations from the figure: DDPG-based RL agent outperforms MPC in terms of tracking accuracy and transient behaviour, achieving lower IAE, ISE, and ITAE values. Although MPC ensures satisfactory regulation, its higher CEI value indicates a significantly greater control effort compared to the DDPG-based RL controller.

As observed in Table 7, the DDPG controller presents a slightly higher overshoot (11.17%) compared to MPC (9.23%), yet achieves a faster rise time (t_r = 78 s vs. t_r = 100 s) and settling time (t_s = 41 s vs. t_s = 60 s). This is again consistent with the trade-off between transient speed and overshoot discussed previously: a marginally more aggressive response allows the DDPG controller to reach and settle around the reference value faster, while maintaining a moderate and acceptable overshoot level.

Fig. 10 presents the comparison between the DDPG-based RL controller and the MPC under the same aggressive irradiance profile described above. The MPC achieves an IAE of 537.05, but at the cost of a substantially higher control effort (CEI = 22.33 vs. 9.68 for the RL controller). Furthermore, the MPC produces significantly more oscillatory control actions, particularly during the transient recovery phase, whereas the DDPG-based RL controller maintains smoother actuation throughout the simulation. The quantitative performance criteria are summarised in Table 8.

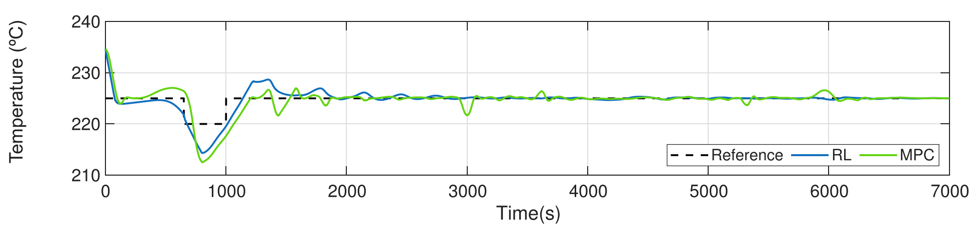


Figure 10: (Continued)

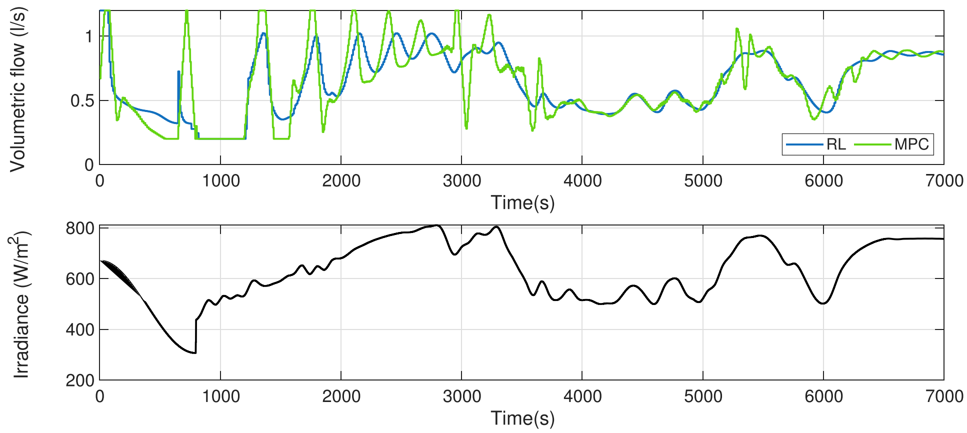


Figure 10: Simulation results of DDPG-based RL and MPC controllers under aggressive solar irradiance disturbance with reference change.

Table 8: Comparative performance criteria between DDPG and MPC controllers under aggressive solar irradiance disturbance.

Controller	IAE	ISE	ITAE	CEI	OS (%)
DDPG	429.29	1194	5.97×10^5	9.68	73.42
MPC	537.05	1999	8.70×10^5	22.33	33.4

Fig. 11 shows the system response when the inlet temperature is considered as the measurable disturbance. In this scenario, the disturbance exhibits several abrupt variations throughout the simulation period, making disturbance rejection more demanding than in the solar-irradiance case.

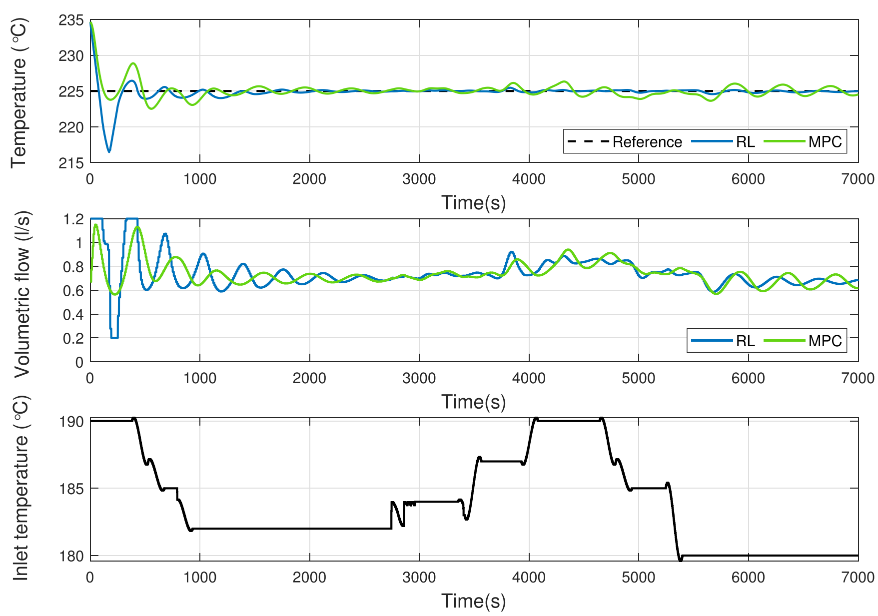


Figure 11: Simulation results of DDPG-based RL and MPC controllers under HTF inlet temperature disturbance.

In the upper plot, both controllers manage to track the 225°C reference, although the fluctuations introduced by the inlet-temperature disturbance are clearly reflected in the outlet response. The MPC controller (in green) presents larger deviations during the most abrupt disturbance changes, while the DDPG-based RL controller (in blue) achieves a smoother trajectory with reduced overshoot and quicker recovery after each disturbance.

The middle plot displays the corresponding control actions. The DDPG-based RL controller continues to apply smoother and less oscillatory variations in the volumetric flow rate, whereas MPC introduces sharper adjustments in an attempt to counteract the disturbance. Due to the stronger effect of inlet-temperature changes on the thermal dynamics, the DDPG-based RL controller requires a slightly higher overall control activity compared with the irradiance case, though still with significantly less oscillatory behaviour than MPC. The lower plot illustrates the inlet-temperature variations, which contain step-like increases and decreases between approximately 180°C and 190°C. These abrupt changes directly influence the outlet-temperature dynamics, explaining the stronger transient behaviour observed in both control responses.

The performance criteria in Table 9 support these observations. The RL controller achieves considerably lower IAE and ITAE values (261.50 and 3.37×10^5, respectively), confirming its superior tracking accuracy and transient performance under inlet-temperature disturbances. However, MPC yields a slightly lower CEI (5.87 vs. 7.35), indicating slightly reduced total control effort. Despite this, t, the significantly higher ITAE value obtained for the MPC (1.38×10^6) reveals slower recovery and significantly poorer long-term performance.

Table 9: Comparative performance criteria between DDPG and MPC controllers under HTF inlet temperature disturbance.

Controller	IAE	ISE	ITAE	CEI	OS (%)
DDPG	261.50	892.12	3.37×10^5	7.35	85.31
MPC	449.91	855.87	1.38×10^6	5.87	12.16

In conclusion, the proposed RL controller has been compared with two classical controllers, a PI plus FF controller and a MPC. These classical approaches were selected as baselines due to their prevalence in PTC systems. Overall, the obtained simulation results indicate that the DDPG-based RL controller provides better disturbance rejection, improved robustness, and a smoother temperature regulation in the presence of inlet-temperature variations, while the MPC controller prioritizes reduced control effort at the expense of tracking performance. It is important to note that, part of the improvement of the DDPG-based RL controller over the MPC is due to the training of the former with the non-linear model, which enables the DDPG-based RL controller to handle the non-linear dynamics of this type of system. The same non-linear model could have been used with the MPC; however, this would lead to more complex problems in terms of robustness and stability.

The work presented in this paper establishes a comparison against classical control architectures, but the advantages of the RL controller over other advanced control approaches should be addressed properly. Consequently, future research will focus on extending these comparisons and validating the proposed approach in real-world environments to ensure its practical applicability and robust performance across a broader range of state-of-the-art control strategies.

4.3 Real-Time Implementation Feasibility

Although the results presented in this work are obtained in simulation, the proposed control strategy is consistent with real-time industrial implementation. The case study is aligned with the CIESOL research facilities, which comprise a fully instrumented solar thermal plant operating under a SCADA-based supervision system. The installation includes distributed sensing and actuation across the solar collector field and associated subsystems, providing continuous measurements of key variables such as temperatures, flow rates and solar irradiance, which are required for feedback control.

Within this framework, the RL-based controller can be directly integrated as an external control module, using real-time measurements to construct the observation vector and applying the control action through the existing actuation system. In contrast to model-based approaches such as MPC, which require online optimisation at each sampling instant, the DDPG-based controller only involves the evaluation of neural networks once trained, resulting in low computational cost and fast execution compatible with typical industrial sampling times.

To further assess the practical applicability of the proposed controller, the computational cost and sample efficiency of the RL approach were analysed. The training process of the DDPG agent was performed offline on a high-performance workstation equipped with an AMD Ryzen 7 9800X3D processor, an NVIDIA GeForce RTX 5080 GPU (used for training), and a Samsung 9100 Pro 2 TB SSD. Two disturbance scenarios were considered. For the solar irradiance case, the average training time was approximately 28 min, whereas for the HTF inlet temperature disturbance the training required approximately 3 h and 50 min. The longer training time in the latter case is attributed to the increased complexity of the disturbance dynamics.

It is important to highlight that the training process does not necessarily reach the maximum limit of 5000 episodes. Instead, an early stopping criterion based on the average reward is employed. Training is terminated when the moving average reward exceeds a predefined threshold, indicating convergence of the learned policy. As a result, convergence was achieved after only 288 episodes in the irradiance disturbance scenario, whereas 2675 episodes were required for the HTF inlet temperature case. This demonstrates that, in practice, the required number of training samples can be significantly lower than the predefined maximum, depending on the disturbance complexity.

Despite this offline computational effort, the execution phase (inference) of the RL controller is computationally efficient. In Simulink simulations with a stop time of 7000 s, the RL controller required 21.55 s under irradiance disturbances and 25 s under inlet temperature disturbances. In comparison, the PI controller required 46.36 s and 142 s, respectively, while the PI+FF controller required 46.74 s and 137 s. The MPC controller showed higher computational demand in the irradiance case (186 s), while requiring 29.65 s in the inlet temperature scenario.

Regarding sample efficiency, although RL methods are typically data-intensive, the use of an early stopping criterion substantially reduces the effective number of training episodes required for convergence, particularly in the irradiance scenario. Furthermore, all training is performed offline, and once trained, the RL controller only requires the evaluation of neural networks, resulting in a low computational burden suitable for real-time implementation.

Furthermore, the use of a non-linear model during training, together with domain randomisation of operating conditions, allows the controller to capture the inherent non-linearities of the PTC and to improve its robustness under varying disturbances. This supports the applicability of the proposed control strategy beyond the simulated environment, facilitating its transfer to real operating conditions.

5 Conclusions

PTC is a key technology in the solar thermal energy landscape, as it plays a crucial role in both industrial process heat and electricity generation. Furthermore, its ability to convert solar energy directly into heat gives it a unique competitive advantage in the industrial sector. However, its distributed nature makes it difficult to control due to its non-linearities, meaning that classic controllers do not achieve the desired performance when it comes to maintaining the outlet temperature at a reference value.

In this work, an RL controller, specifically based on a DDPG agent, is compared with two classical controllers, a PI plus FF controller and an MPC, to maintain the outlet temperature of a PTC as close as possible to a given reference. In addition, the performance of the RL controller is compared when rejecting undesirable transient behaviour caused by disturbances. All case studies were carried out in simulation using a non-linear model of a PTC that takes into account the distributed nature of this type of solar plant. In all cases, the DDPG-based RL controller obtained better results than the PI+FF and MPC controllers in all statistical indices used in this work to measure controller performance, IAE (123.12), ISE (267.28), and ITAE (1.46×10^5) at time to reject solar irradiance disturbances and IAE (261.50), ISE (892.12), and ITAE (3.37×10^5) rejecting inlet temperature disturbances. This improvement in performance is achieved at the expense of a more aggressive control signal, which is logical considering that the controller has to reject disturbances and manage the non-linearities of the system. A key advantage of the DDPG-based RL controller is that it was trained using the non-linear model of the system, allowing it to handle complex dynamics that linearised models used in PI and MPC struggle to capture. Although a non-linear MPC could be implemented, it would involve a much more complex formulation in terms of stability and robustness compared to the RL framework.

Due to these promising results, this work has certain limitations. Specifically, the results were obtained exclusively through numerical simulations without validation in a real plant. Furthermore, the agent exhibits a high control effort that could compromise the actuators' lifespan, since the current reward function does not explicitly penalize mechanical wear. Finally, the comparison is restricted to classical and linear strategies, omitting an analysis against a nonlinear MPC. Therefore, future works are aimed to expand this work in several directions: (i) to test the controllers in an real PTC plant, including hardware-in-the-loop validation, in order to replicate these results. The objective is to verify that the advantage observed in simulation for the RL controller over the others is also maintained in a real plant. (ii) to compare the RL controller against a non-linear MPC to further analyze the trade-offs between computational complexity, stability, and control agility and, (iii) further refinement of the reward function to explore better ways to penalise aggressive control actions, seeking an optimal balance between tracking the desired setpoint and mechanical wear on the actuators due to aggressive control signals.

Acknowledgement: Not applicable.

Funding Statement: This work has been funded by the National R+D+i Plan Project PID2021-126889OB-I00 of the Spanish Ministry of Science and Innovation and EIE funds, and TED2021-131655B-I00 research project funded by AEI/10.13039/501100011033 and by the "European Union NextGenerationEU".

Author Contributions: The authors confirm contribution to the paper as follows: conceptualization, María del Mar Castilla and José Domingo Álvarez; methodology, Marta Leal, Verónica Abad-Alcaraz and José Domingo Álvarez; software, Marta Leal, Verónica Abad-Alcaraz and José Domingo Álvarez; validation, Marta Leal, Verónica Abad-Alcaraz, María del Mar Castilla and José Domingo Álvarez; formal analysis, Marta Leal and Verónica Abad-Alcaraz; resources, María del Mar Castilla and José Domingo Álvarez; writing—original draft preparation, Marta Leal, Verónica Abad-Alcaraz and José Domingo Álvarez; writing—review and editing, María del Mar Castilla and José Domingo Álvarez; visualization, Marta Leal and Verónica Abad-Alcaraz; supervision, María del Mar Castilla and José Domingo

Álvarez; project administration, María del Mar Castilla and José Domingo Álvarez; funding acquisition, María del Mar Castilla and José Domingo Álvarez. All authors reviewed and approved the final version of the manuscript.

Availability of Data and Materials: The data that support the findings of this study are available from the Corresponding Author, Jose Domingo Alvarez, upon reasonable request.

Ethics Approval: Not applicable.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. International Energy Agency Solar Heating & Cooling programme. Solar heat worlwide [Internet]. 2024 [cited 2025 Sep 11]. Available from: https://www.iea-shc.org/data/sites/1/publications/Solar-Heat-Worldwide-2024.pdf.
2. REN21. Renewables 2023 global status report collection, renewables in energy supply [Internet]. 2023 [cited 2025 Sep 11]. Available from: https://www.ren21.net/gsr-2023/modules/energy_supply/01_energy_supply.
3. US Department of Energy. Solar-thermal power and industrial processes basics [Internet]. 2025 [cited 2025 Sep 11]. Available from: https://www.energy.gov/eere/solar/solar-thermal-power-and-industrial-processes-basics.
4. REN21. Renewables 2024 global status report collection, renewables in energy supply [Internet]. 2024 [cited 2025 Sep 11]. Available from: https://www.ren21.net/gsr-2024/modules/energy_supply/02_market_and_industry_trends/08_solarthermal/.
5. International Energy Agency. Solar energy: mapping the road ahead [Internet]. 2019 [cited 2025 Sep 11]. Available from: https://www.iea.org/reports/solar-energy-mapping-the-road-ahead.
6. Palladino V, Di Somma M, Cancro C, Gaggioli W, De Lucia M, D'Auria M, et al. Innovative industrial solutions for improving the technical/economic competitiveness of concentrated solar power. Energies. 2024;17(2):360. doi:10.3390/en17020360.
7. Bortolato M, Dugaria S, Del Col D. Experimental study of a parabolic trough solar collector with flat bar-and-plate absorber during direct steam generation. Energy. 2016;116:1039–50. doi:10.1016/j.energy.2016.10.021.
8. Jaramillo O, Borunda M, Velazquez-Lucho K, Robles M. Parabolic trough solar collector for low enthalpy processes: an analysis of the efficiency enhancement by using twisted tape inserts. Renew Energy. 2016;93:125–41.
9. Zou B, Dong J, Yao Y, Jiang Y. An experimental investigation on a small-sized parabolic trough solar collector for water heating in cold areas. Appl Energy. 2016;163:396–407. doi:10.1016/j.apenergy.2015.10.186.
10. Camacho EF, Ruiz-Moreno S, Aguilar-López JM, Gallego AJ, García RA. The future of parabolic trough thermosolar plants: classic control and beyond. In: Nonlinear and constrained control: applications, synergies, challenges and opportunities. Cham, Switzerland: Springer; 2025. p. 443–70.
11. Leal M, Álvarez JD, del Mar Castilla M, Berenguel M. Feedforward PID controller design for a parabolic trough solar collector. IFAC-PapersOnLine. 2024;58(7):49–54. doi:10.1016/j.ifacol.2024.08.009.
12. Song Y, Xia C, Zhang J, Wang J, Li Y. Disturbance rejection model predictive control for solar collector fields based on fast Fourier transform. IFAC-PapersOnLine. 2024;58(13):284–9. doi:10.1016/j.ifacol.2024.07.496.
13. Mechhoud S, Belkhatir Z. Robust bilinear tracking control of a parabolic trough solar collector via saturation. J Process Control. 2024;143(7):103321. doi:10.1016/j.jprocont.2024.103321.
14. Song Y, Wang J, Zhang J, Li Y. Temperature homogenization control of parabolic trough solar collector field based on hydraulic calculation and extended Kalman filter. Renew Energy. 2024;226(10):120346. doi:10.1016/j.renene.2024.120346.
15. Kalanithi K, Samuel GG, Malathi M, Venkatesan R. Development and implementation of a model predictive control system for a solar parabolic trough plant influenced by an advanced meteorological disturbance model. Sci Rep. 2025;15(1):36650. doi:10.1038/s41598-025-20459-y.
16. Gao XH, Wei S, Su ZG. Optimal model predictive rejection control for nonlinear parabolic trough collector with lumped disturbances. Trans Inst Measur Cont. 2021;43(9):1903–14. doi:10.1177/0142331220983651.
17. Pataro IM, Gil JD, Álvarez JD, Guzmán JL, Lemos JM, Berenguel M. Learning-based practical nonlinear predictive controller for solar thermal collector fields. IEEE Trans Control Syst Technol. 2025;33(6):2108–20.

18. Meligy R, Mekid S, Montenon A. Deep reinforcement learning-based adaptive control of a concentrated solar thermal plant. Appl Therm Eng. 2025;275(3):126851. doi:10.1016/j.applthermaleng.2025.126851.
19. Tilahun FB. Fuzzy-based predictive deep reinforcement learning for robust and constrained optimal control of industrial solar thermal plants. Appl Soft Comput. 2024;159(4):111432. doi:10.1016/j.asoc.2024.111432.
20. An B, Zhang Q, Li L, Gao F, Wang K, Yang J. Artificial neural network-based feedforward-feedback control for parabolic trough concentrated solar field. Sustainability. 2025;17(8):3334. doi:10.3390/su17083334.
21. Camacho E, Rubio F, Berenguel M, Valenzuela L. A survey on control schemes for distributed solar collector fields. Part I: modeling and basic control approaches. Sol Energy. 2007;81(10):1240–51. doi:10.1016/j.solener.2007.01.002.
22. Álvarez J, Yebra L, Berenguel M. Adaptive repetitive control for resonance cancellation of a distributed solar collector field. Int J Adapt Control Signal Process. 2009;23(4):331–52. doi:10.1002/acs.1045.
23. Åström KJ, Hägglund T. Advanced PID. New York, NY, USA: ISA; 2006.
24. Guzmán JL, Hägglund T. Tuning rules for feedforward control from measurable disturbances combined with PID control: a review. Int J Cont. 2024;97(1):2–15. doi:10.1080/00207179.2021.1978537.
25. Camacho EF, Bordons C. Model predictive control. In: Advanced textbooks in control and signal processing. 2nd ed. London, UK: Springer; 2007.
26. Leal M, Abad-Alcaraz V, Álvarez JD, Del Mar Castilla M. Reinforcement learning for a parabolic trough solar collector. In: 2025 11th International Conference on Control, Decision and Information Technologies (CoDIT). Piscataway, NJ, USA: IEEE; 2025. p. 260–5. doi:10.1109/codit66093.2025.11321400.
27. MathWorks. Reinforcement learning toolbox [Internet]. 2025 [cited 2025 Oct 7]. Available from: https://es.mathworks.com/help/reinforcement-learning.html.