

Control de un colector solar cilindro-parabólico mediante aprendizaje por refuerzo

Leal, M.^a, Abad-Alcaraz, V.^a, Álvarez, J.D.^{a,*}, Castilla, M.M.^a, Torres, J.L.^b

^aDepartamento de Informática. Centro de Investigaciones en Energía Solar (CIESOL), Universidad de Almería - ceiA3, Almería, España.

^bDepartamento de Ingeniería. Centro de Investigaciones en Energía Solar (CIESOL), Universidad de Almería - ceiA3, Almería, España.

To cite this article: Leal, M., Abad-Alcaraz, V., Álvarez, J.D., Castilla, M.M. Torres, J.L., 2025. Reinforcement learning control for a parabolic trough solar collector. XX Simposio CEA de Control Inteligente, Huelva (Spain), 2025.

Resumen

Los colectores solares cilindro-parabólicos (CCP) se utilizan ampliamente para suministrar energía térmica en procesos industriales. El control eficaz de la temperatura del fluido de transferencia de calor (FTC) a un punto de consigna deseado, especialmente bajo la influencia de fluctuaciones externas como la irradiancia solar, sigue siendo un desafío operativo clave. Este trabajo presenta el diseño y la evaluación de una estrategia de control basada en Aprendizaje por Refuerzo (AR), diseñada para asegurar un seguimiento de alta fidelidad de la referencia de temperatura de salida en los CCP, a la vez que contrarresta eficazmente las perturbaciones de la irradiancia solar. Para cuantificar sus ventajas, el rendimiento del controlador AR se compara sistemáticamente con técnicas de control convencionales, como los controladores Proporcional-Integral-Derivativo (PID) y anticipativo, centrándose en las mejoras en el rechazo de perturbaciones y la eficiencia operativa general. Los resultados destacan el considerable potencial de los controladores basados en AR para gestionar con destreza las complejidades inherentes a dichos sistemas no lineales.

Palabras clave: Aprendizaje por refuerzo, Sistemas de energía solar, Control de temperatura, Control de sistemas no lineales, Rechazo de perturbaciones.

Reinforcement learning control for a parabolic trough solar collector

Abstract

Parabolic Trough Solar Collectors (PTCs) are widely adopted for supplying thermal energy to industrial processes. Effectively managing the Heat Transfer Fluid (HTF) temperature at a desired setpoint, especially under the influence of external fluctuations like solar irradiance, remains a key operational challenge. This work presents the design and evaluation of a control strategy rooted in Reinforcement Learning (RL), engineered to ensure high-fidelity tracking of the outlet temperature reference in PTCs while actively countering solar irradiance disturbances. To quantify its advantages, the RL controller's performance is systematically compared against conventional Proportional-Integral-Derivative (PID) and feedforward control techniques, focusing on improvements in disturbance rejection and overall operational efficiency. The outcomes highlight the considerable potential of RL-based controllers for adeptly managing the complexities inherent in such non-linear systems.

Keywords: Reinforcement learning, Solar energy systems, Temperature control, Nonlinear systems control, Disturbance rejection.

1. Introducción

El sector del calor es el mayor consumidor de energía final y una fuente importante de emisiones de CO₂. Las fuentes renovables, como las tecnologías termosolares, son cruciales y han alcanzado madurez para diversas aplicaciones International Energy Agency (2024). Los CCPs son una tecnología predominante para procesos industriales de media temperatura (100-

250 °C, hasta 400 °C), concentrando la radiación solar en un tubo receptor por donde fluye un FTC Horta and Fraunhofer Inst. Solar Energy Syst. (2015). El control busca regular la temperatura del FTC ante perturbaciones, especialmente la irradiancia. Las estrategias convencionales como el control PID y anticipativo (*feedforward*) tienen limitaciones para manejar la naturaleza distribuida y no lineal de los CCPs, y la variabilidad de la

*Autor para correspondencia: jhervas@ual.es

irradiancia Guzmán and Hägglund (2011). El Aprendizaje por Refuerzo (AR) ofrece una alternativa prometedora para mejorar la respuesta a perturbaciones dinámicas y manejar la no linealidad del sistema de forma más efectiva que los controladores tradicionales. Diversos estudios han explorado AR en sistemas solares Camacho et al. (2007); Carballo et al. (2025); Pataro et al. (2023); Díaz and Fernández (2022). El objetivo principal de este estudio es desarrollar un sistema de control basado en AR para el seguimiento preciso de la temperatura de referencia en la salida de un CCP, considerando las perturbaciones de la irradiancia, y comparar su rendimiento con estrategias clásicas (PID y anticipativo).

2. Estudio de Caso

El estudio de caso es un campo de colectores solares distribuidos con tecnología CCP. El modelo dinámico del campo se describe por un sistema de Ecuaciones Diferenciales Parciales (EDP) del balance energético. Para el FTC que circula por el tubo (Figura 1), la ecuación es:

$$A_i \rho C \frac{\partial T}{\partial t}(x, t) + \dot{q} \rho C \frac{T(L, t) - T(0, t)}{L} = \pi D_i h_i \cdot (T_\omega(x, t) - T(x, t)) \quad (1)$$

donde $T(x, t)$ es la temperatura del FTC y $T_\omega(x, t)$ la de la pared del tubo. Para la pared del tubo, el balance es:

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

donde $T_g(x, t)$ es la temperatura ambiente. Los parámetros se detallan en Camacho et al. (2007), tratándose de un modelo semi-físico. Las entradas son $T(0, t)$, $T_g(t)$, $I(t)$ y $v(t)$ (velocidad del FTC), y la salida es $T(L, t)$. Este estudio se centra en la relación dinámica entre irradiancia y velocidad del FTC.

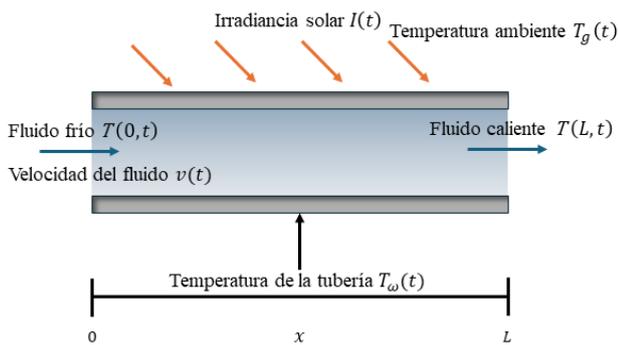


Figura 1: Diagrama del tubo absorbedor de un colector cilindro-parabólico

3. Arquitectura de Control

3.1. Control PID

Se comparan tres arquitecturas. La primera es un controlador Proporcional-Integral (PI). La variable de control es la velocidad del FTC, $v(t)$, utilizando en este trabajo su equivalente en flujo volumétrico, $\dot{q}(t)$. Para diseñar el PI, se obtiene

la función de transferencia entre $\dot{q}(s)$ y la temperatura de salida $T(L, s)$ linealizando el modelo no lineal:

$$\frac{T(L, s)}{\dot{q}(s)} = \frac{k_g}{\tau_g s + 1} \quad (3)$$

donde k_g es la ganancia estática y τ_g la constante de tiempo. Se usó el método de Cancelación Polo-Cero para el ajuste Åström and Hägglund (2009).

3.2. Control Anticipativo (Feedforward)

El segundo enfoque integra un componente anticipativo al PI, con una configuración adelanto-retraso:

$$F_{ff}(s) = \frac{k_d \tau_g s + 1}{k_g \tau_d s + 1} \quad (4)$$

donde k_d y τ_d provienen de la función de transferencia que relaciona la irradiancia con la temperatura de salida Ogunnaike and Ray (1994).

3.3. Aprendizaje por Refuerzo (AR)

El agente de AR observa el entorno, toma una acción y recibe una recompensa, aprendiendo a maximizarla. La Figura 2 muestra la arquitectura, que consta de (i) Observaciones: Integral del error, error instantáneo, derivada del error, temperatura de salida del colector e irradiancia solar incidente. (ii) Entorno: El colector solar no lineal (Sección 2). (iii) Agente: Se implementó un agente basado en Deep Deterministic Policy Gradient (DDPG) y se comparó con Twin Delayed DDPG (TD3). DDPG usa una arquitectura actor-crítico. (iv) Recompensa: Diseñada para guiar al agente, con un rango de temperatura seguro (100-240 °C):

$$RF(t) = \begin{cases} 10 + Bd & \text{si } |T_{ref}(t) - T(x, t)| \leq 0,15 \\ -1 + Bd & \text{si } |T_{ref}(t) - T(x, t)| > 0,15 \end{cases} \quad (5)$$

Es importante señalar en este punto que, Bd representa una penalización severa cuando la temperatura excede el rango seguro, $Bd = -300$, forzando el final del episodio.

4. Resultados

La implementación se realizó con RL Designer Toolbox en MATLAB MathWorks (2025). Se proponen dos experimentos: 1) Comparación entre agentes AR (DDPG y TD3). 2) Comparación del controlador DDPG con PI y PI con anticipativo. El entrenamiento usó un tiempo de muestreo de 10s, hasta 3000 episodios, con condiciones iniciales variables para T_{ref} (100-240 °C) e irradiancia $I(t)$ (400-800 W/m²). Otras perturbaciones se mantuvieron constantes. El criterio de parada fue una recompensa promedio umbral o el máximo de episodios. Hipérparámetros relevantes: varianza de ruido de exploración de 0.3 con decaimiento de 10^{-5} para ambos agentes.

Las Figuras 3 y 4 muestran los resultados de simulación. La Figura 3 compara los agentes DDPG y TD3. Se observa la evolución temporal de la temperatura de salida, la señal de control (caudal volumétrico) y el perfil de irradiancia. Ambos agentes buscan regular la temperatura a 200 °C. Visualmente, DDPG (azul) podría presentar una respuesta ligeramente más estable o con menor esfuerzo de control en ciertos tramos en comparación con TD3 (verde), aunque una evaluación cuantitativa es necesaria para una conclusión robusta.

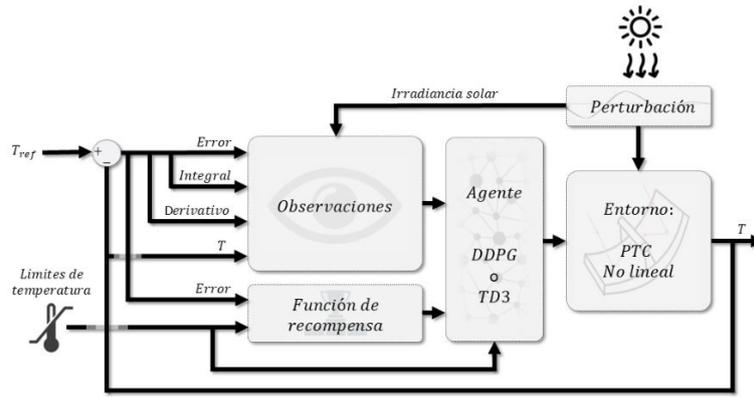


Figura 2: Arquitectura de control basada en AR para CCP.

La Figura 4 compara el agente DDPG con los controladores PID y PID con anticipativo (PID+FF). El agente DDPG demuestra una capacidad superior para seguir la referencia de temperatura y rechazar las perturbaciones de irradiancia en comparación con el PID clásico. Aunque el PID+FF mejora el rendimiento del PID, el DDPG tiende a ofrecer un control más preciso y adaptativo frente a las variaciones, lo que se refleja en menores errores de seguimiento y un esfuerzo de control diferente. Estos resultados sugieren que el enfoque AR puede manejar de manera más efectiva las no linealidades y perturbaciones del sistema CCP.

5. Conclusiones

Este trabajo presentó el desarrollo y evaluación de un sistema de control basado en AR para CCPs. El objetivo fue mejorar el seguimiento de referencia y el rechazo de perturbaciones en comparación con controladores PID y anticipativo. La configuración del agente AR (DDPG) y los resultados obtenidos sugieren un potencial significativo para manejar la no linealidad y las perturbaciones inherentes a estos sistemas, abriendo vías para futuras optimizaciones y aplicaciones industriales.

Agradecimientos

Esta publicación es parte de los proyectos de I+D+i PID2021-126889OB-I00 cofinanciada por el Ministerio de Ciencia e Innovación y fondos EIE y TED2021-131655B-I00 financiado por AEI/10.13039/501100011033/ y “Unión Europea NextGenerationEU”.

Referencias

Camacho, E. F., Rubio, F. R., Berenguel, M., Valenzuela, L., 2007. A survey on control schemes for distributed solar collector fields. part i: Modeling and basic control approaches. *Solar Energy* 81 (10), 1240–1251.

Carballo, J. A., et al., 2025. Reinforcement learning for heliostat aiming: Improving the performance of solar-tower plants. *Appl. Energy* 377.

Díaz, J. J., Fernández, J. A., 2022. The potential of control models based on reinforcement learning in the operation of solar thermal cooling systems. *Processes* 10 (8).

Guzmán, J. L., Hägglund, T., 2011. Simple tuning rules for feedforward compensators. *J. Process Control* 21 (1), 92–102.

Horta, P., Fraunhofer Inst. Solar Energy Syst., 2015. Process heat collectors: State of the art and available medium-temperature collectors. Tech. rep., IEA SHC Task 49/IV—Deliverable A.

International Energy Agency, 2024. Renewables 2024. <https://www.iea.org/reports/renewables-2024>, [Último acceso 30-Enero-2025].

MathWorks, 2025. Reinforcement learning toolbox. <https://es.mathworks.com/help/reinforcement-learning.html>, [Último acceso 30-Enero-2025].

Ogunnaik, B. A., Ray, W. H., 1994. *Process Dynamics, Modeling, and Control*. Oxford University Press, Oxford, UK.

Pataro, I. M. L., et al., 2023. Optimal model-free adaptive control based on reinforcement q-learning for solar thermal collector fields. *Eng. Appl. Artif. Intell.* 126.

Åström, K. J., Hägglund, T., 2009. *Advanced PID Control*. Pearson, Madrid, Spain.

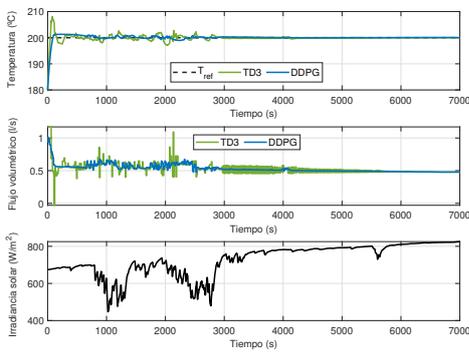


Figura 3: Comparativa de dos agentes: TD3 y DDPG

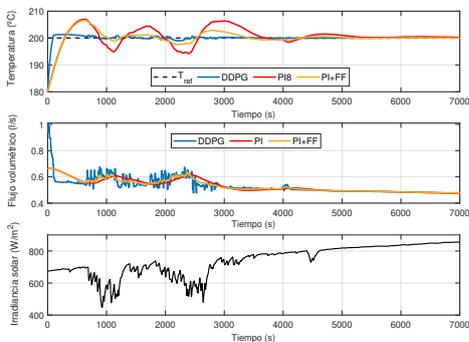


Figura 4: Comparativa de controladores DDPG, PID, y PID+Anticipativo