

# **OPTIMIZACIÓN DE ESTRATEGIAS DE APUNTE DE HELIOSTATOS EN CENTRALES DE TORRE MEDIANTE APRENDIZAJE POR REFUERZO**

**Carballo J.A.\*, Bonilla J.\*, Cruz N.C.\*\*\*, Álvarez J.D.\*\*, Avila-Marín A.\*, Berenguel M.\*\***

\* CIEMAT - Plataforma Solar de Almería (PSA), Ctra. de Senés km 4,5, E-04200 Tabernas, Almería, España  
\*\* CIESOL, Centro de Investigaciones en Energía Solar, Universidad de Almería - CIEMAT, Departamento de

Informática, E-04120, Almería, España

\*\*\* Departamento de Ingeniería de Computadores, Automática y Robótica, Universidad de Granada, Granada, España

## **RESUMEN**

Las centrales solares de torre representan una pieza clave en los sistemas energéticos renovables por su alta eficiencia térmica y capacidad de almacenamiento de energía. Uno de los desafíos principales en este tipo de instalaciones es la optimización de las estrategias de apunte de heliostatos para maximizar la captación de energía solar, garantizando al mismo tiempo la seguridad operativa y la durabilidad del receptor. En este trabajo se presenta un enfoque avanzado basado en Aprendizaje por Refuerzo, implementado mediante el algoritmo Soft Actor-Critic y su versión distribuida Ray RLlib. Este nuevo enfoque mejora la eficiencia de entrenamiento y permite la optimización simultánea de múltiples configuraciones de agentes. Además, el modelo se entrena utilizando datos reales de irradiancia directa normal del conjunto de datos tipo de la Plataforma Solar de Almería, capturando efectos transitorios como nubosidad y variaciones estacionales. Los resultados preliminares muestran un incremento del 9,1 % en la potencia absorbida anual en comparación con las estrategias de apunte convencionales, así como una reducción significativa del tiempo de entrenamiento. Este avance proporciona una herramienta robusta y escalable para el control óptimo de heliostatos en condiciones reales, con gran potencial de aplicación industrial en centrales solares de torre.

**PALABRAS CLAVE:** Energía solar de concentración, Aprendizaje por refuerzo, Optimización distribuida, Centrales de torre solar

## **ABSTRACT**

Solar tower plants represent a key component in renewable energy systems due to their high thermal efficiency and energy storage capacity. One of the main challenges in this type of facility is the optimization of heliostat aiming strategies to maximize solar energy capture while ensuring operational safety and receiver durability. This work presents an advanced approach based on Reinforcement Learning, implemented through the Soft Actor-Critic algorithm and its distributed version, Ray RLlib. This new approach improves training efficiency and enables the simultaneous optimization of multiple agent configurations. Furthermore, the model is trained using real Direct Normal Irradiance data from the year-type dataset of the Plataforma Solar de Almería, capturing transient effects such as cloud cover and seasonal variations. Preliminary results show a 9.1% increase in annual absorbed power compared to conventional aiming strategies, as well as a significant reduction in training time. This development provides a robust and scalable tool for optimal heliostat control under real operating conditions, with great potential for industrial application in solar tower plants.

**KEYWORDS:** Concentrated solar power, Reinforcement learning, Distributed optimization, Solar tower plants

## INTRODUCCIÓN

Las plantas solares de torre concentran la radiación solar mediante un extenso campo de heliostatos que reflejan y concentran la radiación solar sobre un receptor central situado en la parte superior de una torre, donde la radiación solar se transforma en calor de alta temperatura para su posterior aprovechamiento en ciclos termoeléctricos o sistemas de almacenamiento térmico. Esta configuración permite alcanzar rendimientos superiores a otras tecnologías solares, pero también exige un control preciso del apuntado de cada heliostato para evitar pérdidas ópticas, puntos calientes o daños en el receptor (Lin et al., 2025).

El rendimiento global del sistema depende, por tanto, en gran medida de la estrategia de apunte y coordinación del campo de heliostatos (Kuhl et al., 2025). En trabajos previos (Carballo et al., 2025), se demostró la viabilidad del uso del Aprendizaje por Refuerzo (Reinforcement Learning, RL) como una alternativa inteligente frente a los métodos empíricos o geométricos tradicionales. Este enfoque permitió a los agentes de RL aprender de forma autónoma políticas de apunte que maximizan la potencia absorbida en función de la posición solar y las condiciones instantáneas del campo.

No obstante, se identificaron dos retos principales que limitaban su aplicabilidad industrial: por un lado, la eficiencia computacional del proceso de entrenamiento, que requería largos tiempos de simulación; y por otro, la necesidad de incorporar en el modelo la variabilidad real del recurso solar, incluyendo transitorios de nubosidad y fluctuaciones estacionales, con el fin de garantizar un comportamiento robusto y fiable bajo condiciones reales de operación.

## METODOLOGÍA

En este trabajo se ha implementado el entorno de aprendizaje utilizando Ray Rllib (Liang et al., 2025), una librería de código abierto diseñada para el entrenamiento distribuido y escalable en Aprendizaje por Refuerzo (Reinforcement Learning, RL). Esta herramienta permite desplegar entornos de simulación paralelos que ejecutan múltiples agentes de manera simultánea, acelerando significativamente tanto la recopilación de datos como el proceso de actualización de las políticas de control (Ray Train). Gracias a esta capacidad de paralelización masiva, el sistema puede explorar un espacio mucho más amplio de configuraciones, reduciendo el tiempo necesario para alcanzar estrategias óptimas y mejorando la estabilidad del aprendizaje.



Fig. 1. Software empleado

Además, se ha integrado un módulo de optimización automática de hiperparámetros, que ajusta dinámicamente variables críticas como las tasas de aprendizaje, las arquitecturas de las redes neuronales, los coeficientes de regularización o los parámetros de exploración y explotación (Ray Tune). Esta automatización permite obtener políticas más robustas sin necesidad de intervención manual.

Los agentes son entrenados utilizando datos reales de irradiancia directa normal (DNI) registrados en la Plataforma Solar de Almería (PSA), con una resolución temporal de un minuto. Este nivel de detalle permite simular con precisión en el entorno (SolarPilot) las condiciones dinámicas del recurso solar, incluyendo transitorios debidos a nubosidad, variaciones estacionales y fluctuaciones diarias del flujo solar. De este modo, el agente de RL adquiere la capacidad de responder eficazmente a escenarios cambiantes, garantizando una estrategia de apunte más realista, adaptable y aplicable a condiciones operativas del mundo real.

## RESULTADOS

La adopción del marco distribuido Ray Rllib ha supuesto un avance decisivo en términos de rendimiento computacional y eficiencia experimental. Gracias a su arquitectura paralela, fue posible reducir el tiempo total de entrenamiento, pasando de casi 9 días para un solo agente a 4 días para el entrenamiento de más de 500 agentes con

diferentes configuraciones de hiperparámetros, variando parámetros como la profundidad de las redes neuronales, los coeficientes de aprendizaje o las estrategias de exploración.

Este incremento en la capacidad de parallelización y búsqueda masiva de hiperparámetros permitió realizar un análisis exhaustivo del espacio de configuraciones posibles, identificando combinaciones más estables, eficientes y generalizables. Como resultado, las políticas obtenidas mostraron una mayor robustez frente a condiciones cambiantes de radiación solar, incluyendo períodos de nubosidad parcial o fluctuaciones estacionales.

Los agentes mejor entrenados lograron un incremento medio del 9,1 % en la energía anual absorbida respecto a las estrategias de apunte convencionales basadas en posiciones fijas o modelos geométricos simplificados. Además de esta mejora en el rendimiento energético, el sistema mostró una notable capacidad de adaptación a diferentes escenarios operativos sin necesidad de reentrenamientos extensivos.

Estos resultados confirman la superioridad técnica y la viabilidad práctica del enfoque propuesto, validando el uso del aprendizaje por refuerzo distribuido como herramienta eficaz para la optimización del apuntado de heliostatos en plantas solares de torre sometidas a condiciones reales de alta variabilidad solar.

## CONCLUSIONES

La integración del entorno Ray RLlib con datos reales de irradiancia solar constituye un avance sustancial en la aplicación del Aprendizaje por Refuerzo (RL) a la optimización del apunte de heliostatos en plantas solares de torre. Esta combinación permite entrenar agentes inteligentes capaces de aprender estrategias de control robustas directamente a partir de condiciones reales de operación, sin depender de modelos analíticos o simplificaciones geométricas.

La arquitectura distribuida de Ray RLlib aporta una gran escalabilidad y eficiencia computacional, posibilitando la ejecución simultánea de cientos de experimentos y reduciendo de forma drástica los tiempos de entrenamiento. Este enfoque no solo acelera el proceso de desarrollo de políticas óptimas, sino que también mejora la capacidad del sistema para generalizar frente a distintas condiciones ambientales, configuraciones de campo y perfiles de irradiancia.

Los resultados obtenidos demuestran la madurez y el potencial de esta metodología para su implantación a escala industrial, al ofrecer una solución automatizada, adaptable y fácilmente transferible a distintas plantas solares de torre.

Como líneas de trabajo futuro, se plantea la validación experimental en campo de las estrategias desarrolladas, particularmente en la planta CESA-1 de la Plataforma Solar de Almería, con el objetivo de evaluar su comportamiento bajo condiciones operativas reales. Asimismo, se propone explorar técnicas de aprendizaje transferido (transfer learning) que permitan adaptar políticas previamente entrenadas a nuevas instalaciones con diferentes configuraciones de campo o receptores, minimizando el tiempo de reentrenamiento y facilitando la replicabilidad del sistema en distintos emplazamientos solares.

## REFERENCIAS

Lin, X., Zhao, X., Liu, Z., Huang, W., Zhao, Y., & Feng, J. (2025). Real-time and high-accuracy radiative flux distribution simulation based on analytical model for solar power tower system. *Solar Energy*, 287, 113208. <https://doi.org/10.1016/j.solener.2024.113208>

Kuhl, M., Pargmann, M., Cherti, M., Jitsev, J., Maldonado, D., & Pitz-Paal, R. (2024). Flux density distribution forecasting in concentrated solar tower plants: A data-driven approach. *Solar Energy*, 282, 112894. <https://doi.org/10.1016/j.solener.2024.112894>

Carballo, J. A., Bonilla, J., Cruz, N. C., Fernández-Reche, J., Álvarez, J. D., Avila-Marín, A., & Berenguel, M. (2025). Reinforcement learning for heliostat aiming: Improving the performance of solar tower plants. *Applied Energy*, 377, 124574. <https://doi.org/10.1016/j.apenergy.2024.124574>

Liang, E., Liaw, R., Nishihara, R., Moritz, P., Fox, R., & Gonzalez, J. (2025). Ray RLlib: A composable and scalable reinforcement learning library. *arXiv preprint arXiv:1712.09381*. <https://arxiv.org/abs/1712.09381>