

Uso de Q-Learning para la estimación de niveles de presas.

Autor: Erika Rubí Jiménez Guzmán.



1. Introducción

La operación eficiente de presas es fundamental para gestionar recursos hídricos en contextos de múltiples demandas: riego, generación hidroeléctrica, control de inundaciones y conservación ambiental. Un desafío central es regular de forma óptima el nivel del embalse, lo cual depende de decisiones diarias de descarga ante condiciones hidrológicas inciertas y cambiantes.

Métodos clásicos como la Programación Dinámica Estocástica (**SDP**) requieren modelos detallados y discretización fina del espacio de estados, lo que limita su uso en sistemas complejos. El aprendizaje por refuerzo, especialmente el algoritmo Q-iteration ajustado con árboles de decisión, ofrece una solución flexible. Aprende directamente de datos o simulaciones simples, sin necesidad de modelar todo el sistema. Este enfoque permite regular niveles de embalse de forma dinámica, eficiente y adaptable. Este póster muestra cómo aplicar Q-learning basado en árboles para mejorar la operación de presas, superando las limitaciones de los métodos tradicionales.

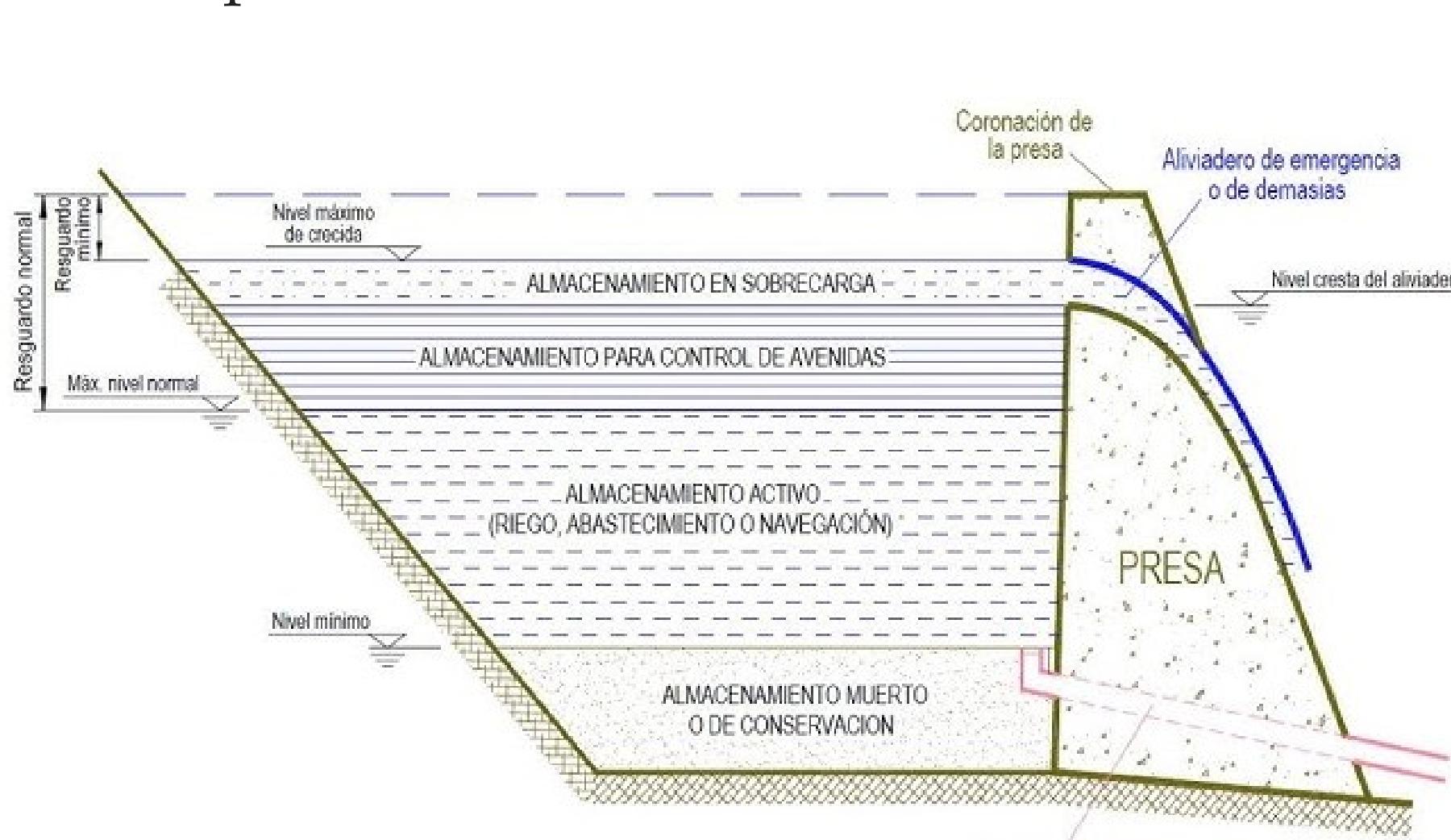
2. Formulación del problema

Programación Dinámica

La programación dinámica estocástica puede usarse para estimar y controlar los niveles de presas ante incertidumbre (por ejemplo, aportes variables por lluvia). En cada etapa se observa el nivel actual del embalse, se decide cuánto liberar, y se calcula cómo esa decisión (junto con el aporte incierto) modificará el nivel futuro. La política óptima minimiza los riesgos de desbordamientos o déficit de agua. Sin embargo, este método tiene dos grandes limitaciones, es computacionalmente costoso y exige un modelo completo.

Recientemente, el aprendizaje por refuerzo ha surgido como una alternativa prometedora. En particular, el algoritmo Q-iteration ajustado, una variante avanzada de Q-learning, permite aprender

políticas óptimas a partir de datos históricos o simulaciones simples, sin necesidad de un modelo completo del sistema.



Niveles de presas.

$$s' = \min \{ \max\{ s - a + \xi, 0 \}, S_{\max} \}$$

4. Fitted Q-Iteration

A diferencia del Q-learning clásico, que actualiza la tabla $Q(s, a)$ de manera incremental después de cada interacción, Q-iteration ajustado utiliza todos los datos disponibles en conjunto para aproximar la función Q. Esto se hace aplicando métodos de regresión (por ejemplo, árboles de decisión, redes neuronales u otros modelos) para estimar la relación entre estados, acciones y recompensas esperadas.

Este enfoque es útil para aprender políticas de operación óptimas a partir de registros hidrológicos, sin necesidad de construir un modelo hidrodinámico detallado. El algoritmo ajusta su política con base en la experiencia pasada, mejorando la toma de decisiones frente a escenarios inciertos como lluvias variables o demandas cambiantes.

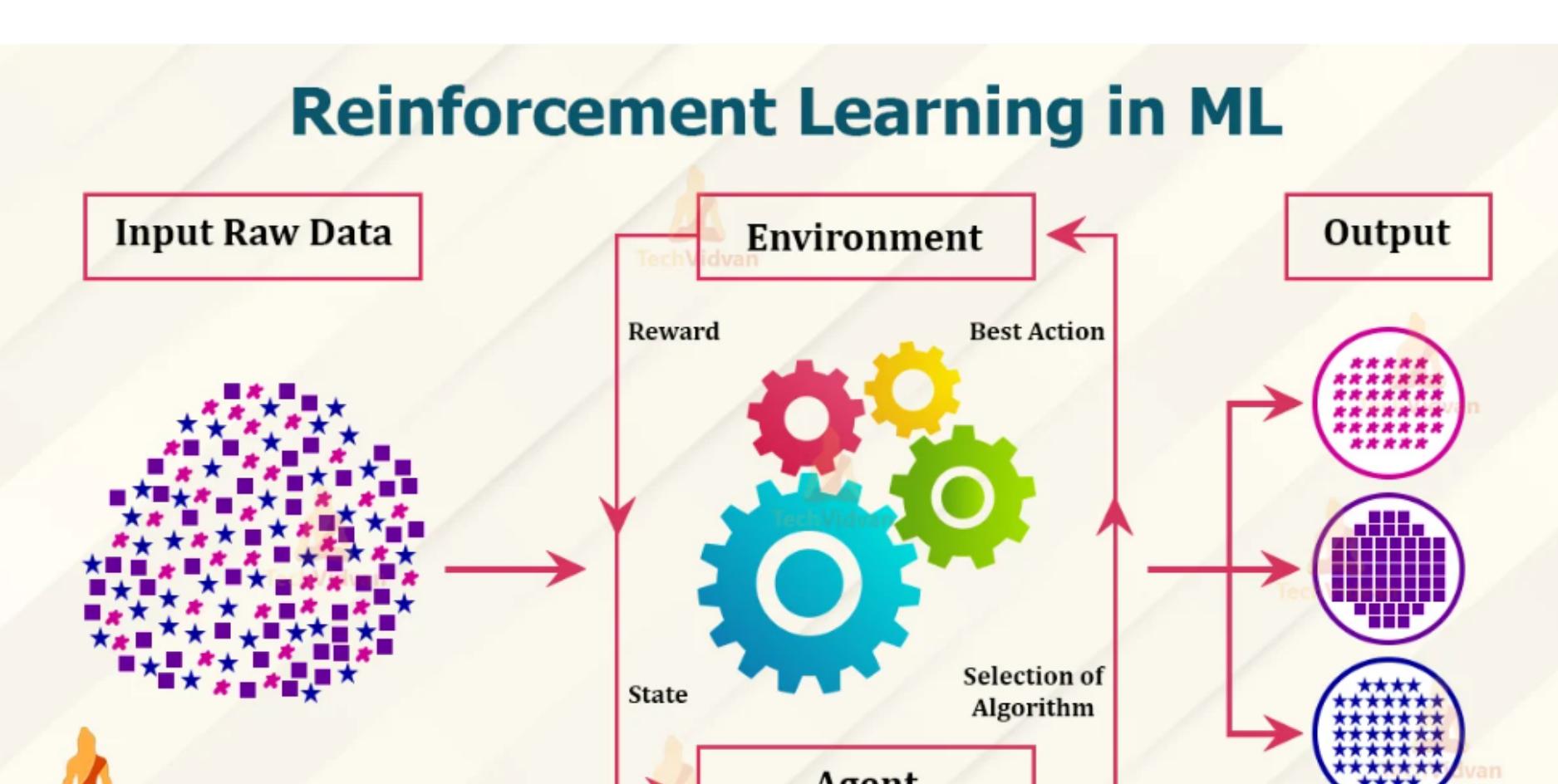


Ilustración de entrenamiento

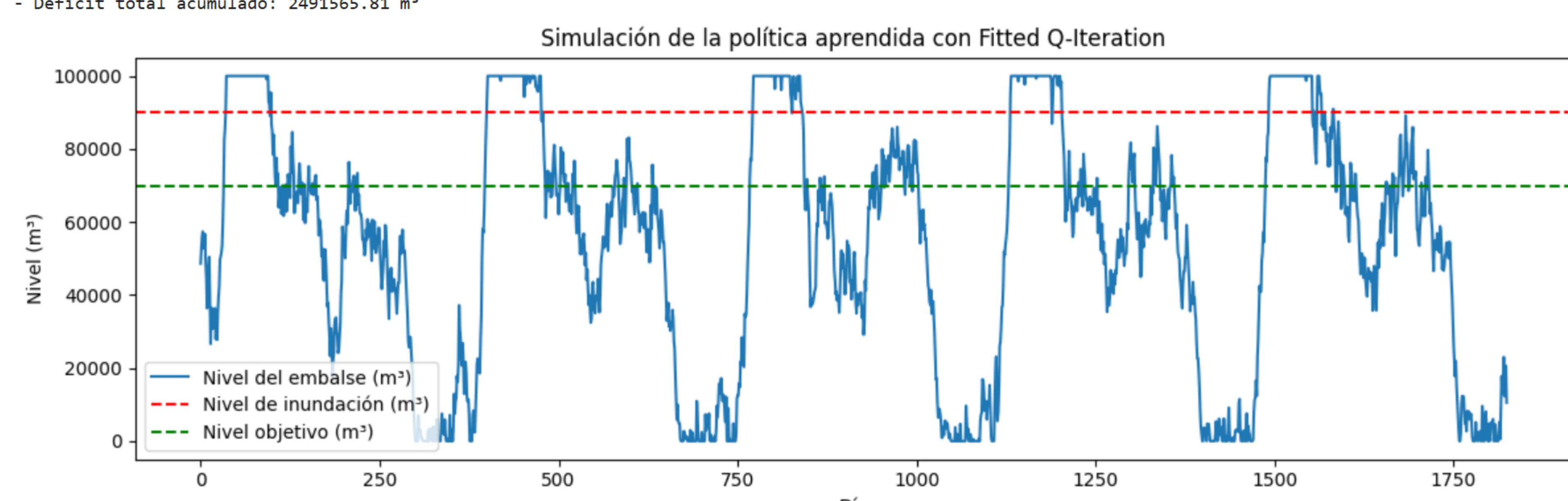
$$Q_h(s, a) = r(s, a) + \gamma \max_{a'} Q_{h-1}(s', a').$$

En cada iteración h , se construye $y_i = r_i + \gamma \max_{a'} Q_{h-1}(s'_i, a')$ y se ajusta un regresor f_{θ_h} para aproximar $Q_h(s, a)$.

6. Resultados

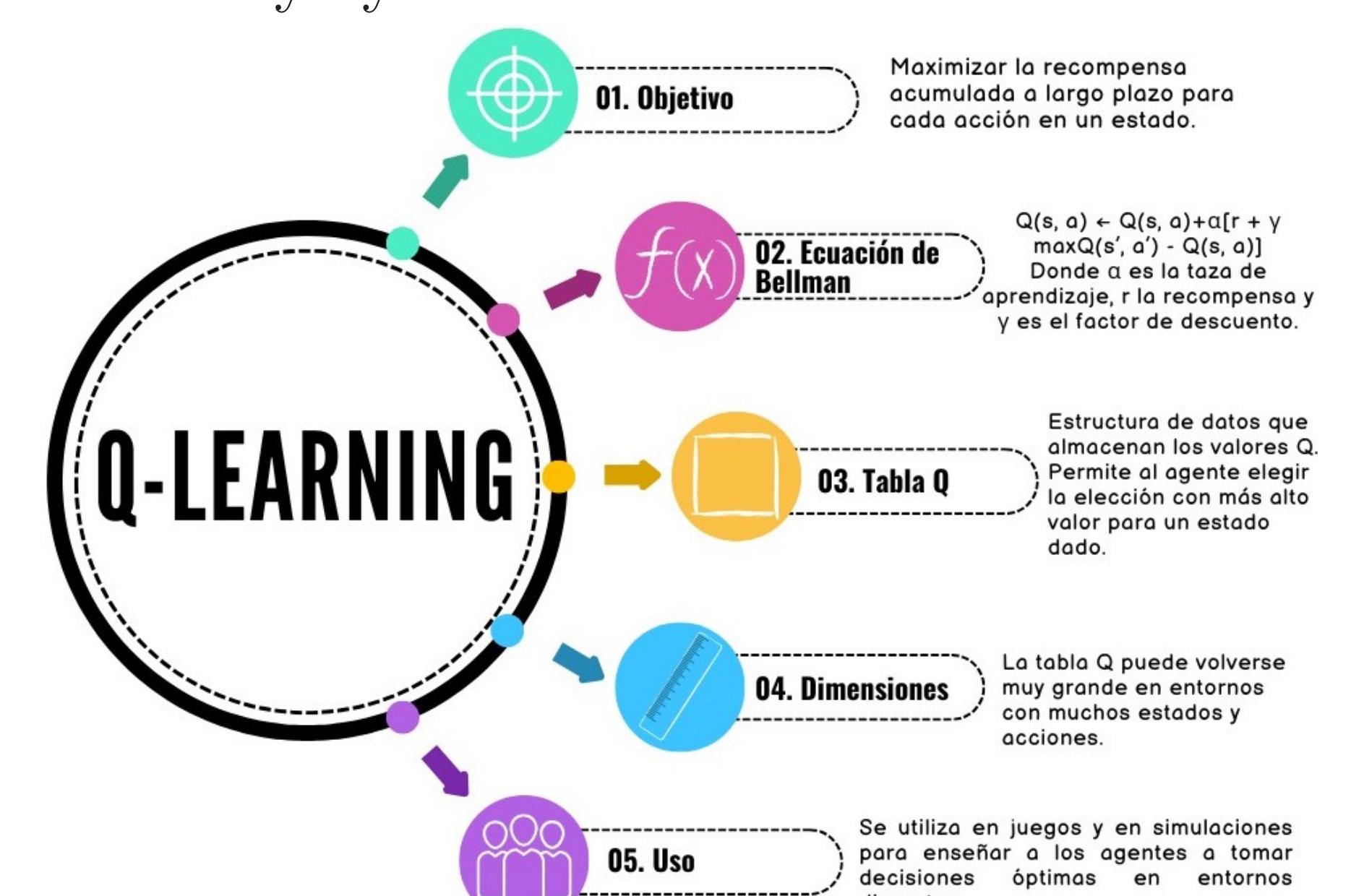
Durante cinco años de simulación, la política aprendida logró mantener el nivel del embalse cerca del objetivo, aunque se registraron 350 días con riesgo de inundación y un déficit acumulado de 2491565.81 m^3 . Estos resultados muestran que el algoritmo puede adaptarse a la estacionalidad del sistema, pero aún enfrenta desafíos ante eventos extremos.

Resultados de simulación (5 años):
- Días con riesgo de inundación: 350
- Déficit total acumulado: 2491565.81 m^3



3. Q-Learning

El Q-learning es un algoritmo de aprendizaje por refuerzo basado en la idea de aprendizaje mediante ensayo y error.



5. Aplicación

Se simuló una política de operación del embalse basada en aprendizaje por refuerzo. El algoritmo aprendió a decidir cuánto liberar diariamente, considerando el nivel del embalse y el día del año, con el objetivo de evitar inundaciones y minimizar déficit de agua. La política fue entrenada usando 10 años de datos sintéticos con estacionalidad y evaluada durante cinco años de operación virtual. Para la simulación se generaron datos sintéticos de las entradas de agua añadiendo un ruido (variable aleatoria que sigue una distribución normal) para reflejar la variabilidad climática lo cual permite entrenar la política en condiciones diversas.

7. Referencias

- Castelletti, A., Galelli, S., Restelli, M., Soncini-Sessa, R. (2010). Tree-based reinforcement learning for optimal water reservoir operation. *Water Resources Research*, 46(9), W09507. <https://doi.org/10.1029/2009WR008898>
- Lee, S., Labadie, J. W. (2007). Stochastic optimization of multireservoir systems via reinforcement learning. *Water Resources Research*, 43(10), W10410. <https://doi.org/10.1029/2006WR005627>
- Ris-Ala, R. (2023). Q-Learning Algorithm. In *Fundamentals of Reinforcement Learning* (pp. 31–55). https://doi.org/10.1007/978-3-031-37345-9_3
- Jiménez, R. (2025). Machine-Learning [Repositorio GitHub]. GitHub. <https://github.com/RubiJimenez929/Machine-Learning>