

# Propuesta de Proyecto Final

## Curso Taller de Aprendizaje por Refuerzo

Matías Rama

Florencia Roque

27 de junio de 2025

### 1. Introducción

En este proyecto proponemos una solución basada en métodos de aprendizaje por refuerzo para optimizar la operación de un sistema eléctrico con generación eólica, solar, térmica e hidráulica con embalse. El problema a resolver representa una instancia simplificada pero representativa del problema real de operación del sistema eléctrico uruguayo y permitirá evaluar la utilidad de métodos de aprendizaje por refuerzo en contextos energéticos con incertidumbre hidrológica. El objetivo es modelar el problema como un proceso de decisión de Markov (MDP), aplicar distintas estrategias de entrenamiento y evaluarlas mediante diferentes métricas.

### 2. Diseño de la solución

El presente proyecto aborda el problema de optimización de la operación de un sistema eléctrico básico con un horizonte temporal anual de paso semanal. Cabe destacar que el modelo se entrenará con un horizonte de dos años y luego se evaluará solo el primero, esto es con el objetivo de evitar tener en cuenta el erogado total del lago que se daría al final del horizonte temporal.

El sistema está compuesto por:

- Parque generador eólico.
- Parque generador solar.
- Central térmica de bajo costo.
- Central térmica de alto costo.
- Central hidráulica con embalse con aportes aleatorios.
- Demanda.

La generación eólica, solar y la demanda se consideran determinísticas, obtenidas a partir de un escenario históricos.

## 2.1. Espacio de estados

- Volumen del embalse.
- Variable hidrológica, discreta.
- Paso del tiempo (semana).

La variable hidrológica será modelada como una cadena de Markov no homogénea (una matriz por cada paso de tiempo).

## 2.2. Dinámica de los estados

- El volumen de la semana siguiente se calcula como  $v_{t+1} = v_t - q_t - d_t + a_t$
- La variable hidrológica de la semana siguiente se muestrea a partir de la matriz de Markov.
- El  $t$  siguiente es igual  $t + 1$ .

Donde  $v_t$  es el volumen del embalse en el paso  $t$ ,  $q_t$  el volumen turbinado,  $d_t$  el volumen vertido y  $a_t$  los aportes.

## 2.3. Espacio de acciones

El espacio de acciones está representado por el volumen de agua a turbinar en cada paso.

Tanto el volumen a turbinar como el volumen del embalse se definirán si son discretos o continuos, dependiendo de los métodos de reinforcement learning que se apliquen.

## 2.4. Función de recompensa

La recompensa está definida como el opuesto del costo incurrido por el despacho térmico total (el de las dos máquinas térmicas) necesario para satisfacer la demanda remanente, una vez considerada la generación eólica, solar e hidráulica.

## 3. Implementación

Se abordará el tema con algoritmos tabulares (por ejemplo, Value-Iteration o Q-learning) o con algoritmos de policy gradient (por ejemplo, Actor-Critic o Proximal Policy Optimization).

## 4. Desempeño

Para tener una métrica de la calidad de las políticas aprendidas, se utilizará como base una política óptima obtenida con MOP (modelo de operaciones usado para encontrar el despacho óptimo, entre otros usos). Adicionalmente, se podrían comparar las políticas aprendidas contra una política miope para tener idea del beneficio de optimizar el sistema hidrotérmico.