UNIVERSIDAD DE CHILE
FACULTAD DE CIENCIAS FÍSICAS Y MATEMÁTICAS
DEPARTAMENTO DE INGENIERÍA ELÉCTRICA

DISEÑO E IMPLEMENTACIÓN DE UNA ESTRATEGIA DE CONTROL PARA SISTEMAS DE RIEGO UTILIZANDO APRENDIZAJE REFORZADO

MEMORIA PARA OPTAR AL TÍTULO DE INGENIERO CIVIL ELÉCTRICO

JOAQUÍN MERINO MACHUCA

PROFESOR GUÍA:
DIEGO MUÑOZ CARPINTERO

PROFESORA CO-GUÍA:
DORIS SÁEZ HUEICHAPAN

MIEMBRO DE LA COMISIÓN:
CARLOS FAÚNDEZ URBINA

SANTIAGO DE CHILE
2022

RESUMEN DE LA MEMORIA PARA OPTAR AL TÍTULO DE: Ingeniero Civil Eléctrico
POR: Joaquín Merino Machuca
FECHA: 2022
PROFESOR GUÍA: Diego Muñoz Carpintero

DISEÑO E IMPLEMENTACIÓN DE UNA ESTRATEGIA DE CONTROL PARA SISTEMAS DE RIEGO UTILIZANDO APRENDIZAJE REFORZADO

Tomando en cuenta las proyecciones de aumento poblacional y de escasez hídrica producto del calentamiento global, se hace urgente el desarrollo de técnicas agrícolas que permitan obtener el mayor rendimiento posible, para abastecer del alimento necesario a las nuevas generaciones, pero consumiendo la menor cantidad de agua que se requiera, para preservar por más tiempo las fuentes de agua.

Este escenario afecta particularmente a los países con una producción agrícola importante, entre los cuales destaca Chile, como uno de los principales proveedores de productos agrícolas a nivel mundial, donde parte importante de los productores realizan dicha actividad con fines de autosustento, muchos de los cuales pertenecen a pueblos indígenas y donde en los últimos años se ha experimentado una sequía que ha perjudicado significativamente el desarrollo de la actividad agrícola.

De acuerdo con lo anterior, en el presente trabajo se simuló un esquema de control de riego para los cultivos modelados del sistema hidrogeológico conformado por un conjunto de agricultores y pozos de la comunidad mapuche José Painecura, ubicada en la Región de la Araucanía, donde el modelo, adaptado a la arquitectura de entornos Gym, busca maximizar los rendimientos de los cultivos consumiendo la menor cantidad de agua posible.

Aquí, se estudiaron tres estratégias de aprendizaje reforzado para espacios de acción y de estado continuos: *Proximal Policy Optimization* (PPO), *Deep Determinstic Policy Gradient* (DDPG) y *Twin Delay Deep Deterministic Policy Gradient* (TD3), fundamentalmente por la capacidad de estas para encontrar soluciones de control óptimas (como en el control predictivo) y la de adecuar los parámetros del controlador ante variaciones de la planta controlada (como en el control adaptativo). Los resultados de las simulaciones permitieron comparar sus desempeños (optimización del agua y rendimientos obtenidos) con estrategias de riego basado en el balance hídrico de suelos y de control predictivo y donde los mejores desempeños entre los tres algoritmos de aprendizaje reforzado se obtuvieron con PPO, en la variante aquí llamada single.

Si bien los resultados obtenidos con PPO, DDPG y TD3 no constituyen una mejor solución al problema de maximizar rendimientos economizando agua, respecto a las estrategias de balance hídrico, los resultados presentados por sus desarrolladores en la resolución de otros problemas y los resultados en general, que muestran la capacidad del aprendizaje reforzado de encontrar mejores soluciones que las humanas, inducen a seguir explorando los controladores basados en aprendizaje reforzado para optimizar el riego en cultivos.

Tabla de Contenido

Índice de cuadros

Índice de figuras

Capítulo 1

Introducción

1.1. Motivación

Para el año 2050 se prevé un aumento poblacional mundial que alcance los 9700 millones de personas, lo que implicaría una mayor demanda alimentaria [3]. Esto plantea el desafío de desarrollar técnicas productivas más sustentables para la producción agrícola, sobre todo si se considera que este aumento poblacional ocurrirá dentro de un contexto de cambio climático acelerado, con limitaciones de agua [4].

En efecto, si se considera que el 80 % del agua utilizada globalmente es con fines agrícolas, donde el agua proveniente de acuíferos subterráneos constituye más del 50 % del agua consumida para riego [5], se hace evidente la necesidad de mejorar el manejo eficiente de este recurso para un desarrollo sustentable de la actividad agrícola.

Chile, se ve particularmente afectado por la situación descrita anteriormente, al ser un país cuya producción silvoagropecuaria constituye cerca del 3 % del PIB nacional [4] y donde las condiciones climáticas y antropogénicas que han derivado en la mega sequía que ha afectado al país durante la última década, con déficit de precipitaciones entre un 20 % y un 40 %, permiten proyectar la continuidad del fenómeno con un aumento en las condiciones de sequedad para los próximos años [6].

Para hacer frente a esta problemática, el Ministerio de Agricultura ha definido cuatro ejes en torno a los cuales estructurar sus políticas: Asociatividad (considerando que el 93 % de los agricultores son pequeños), desarrollo rural (mejorar la calidad de vida en los sectores rurales), sustentabilidad (mitigar el impacto medioambiental de la actividad silvoagropecuaria, enfatizando el consumo eficiente del recurso hídrico) y modernización de la actividad (aumentar la producción y diversidad).

En cuanto a la producción silvoagropecuaria nacional, esta se desarrolla principalmente en las áreas rurales, donde se concentra cerca de un cuarto de la población nacional [7]. Dentro de esta, la agricultura familiar campesina equivale a cerca del 90 % del total de unidades productivas agrícolas del país, donde un porcentaje importante de la misma está destinado al consumo doméstico. Por otro lado, se tiene que los agricultores individuales de pueblos indígenas constituyen el 17, 6 % del total de agricultores del país, donde los productores del pueblo mapuche conforman el grupo más significativo, con un 46 % aproximadamente [8].

Entonces, es dentro de este contexto que se desarrolla el presente trabajo, tomando como referencia a un grupo de diez agricultores de la comunidad mapuche José Painecura, ubicada en el sector de Hueñalihuen, comuna de Carahue en la Región de la Araucanía (ver Figura 1.1), la cual se dedica principalmente a la agricultura de subsistencia [9] y donde se propone un sistema de gestión del riego basado en aprendizaje reforzado, que sea capaz de maximizar la producción de los cultivos de los agricultores considerados, haciendo un uso eficiente del agua para riego obtenida desde pozos. Con esto, el trabajo queda enmarcado dentro de los ejes de sustentabilidad y modernización de la actividad agrícola contemplados por el Ministerio de Agricultura.

Figura 1.1: Comunidad mapuche José Painecura (Hueñalihuen, Región de la Araucanía).

La decisión de utilizar aprendizaje reforzado, se debe a que es posible su implementación como un esquema que explota los beneficios del control óptimo y del control adaptativo [10], lo que en teoría puede conducir a una gestión más eficiente de los sistemas agrícolas, maximizando rendimientos y minimizando costos de operación (fundamentalmente el consumo de agua), además de adaptarse a variaciones en los parámetros del sistema, como las características del suelo y las condiciones meteorológicas.

1.2. Objetivos

1.2.1. Objetivo general

El objetivo general consiste en evaluar y comparar diversos sistemas de control basados en aprendizaje reforzado, respecto a su eficiencia en la gestión del recurso hídrico utilizado en el riego de cultivos y a los rendimientos obtenidos.

1.2.2. Objetivos específicos

Los objetivos específicos del trabajo son:

- Construir un entorno de simulación para el sistema de cultivos y pozos considerados con el paquete Gym.
- Entrenar agentes basados en la estructura *Actor-Critic* de manera *offline*.
- Simular escenarios de escasez hídrica extrema.
- Determinar un agente que permita la obtención de altos rendimientos, economizando agua y sin secar los pozos.

1.3. Metodología

La metodología implementada consistió en:

- En primer lugar, dado que se trabajó sobre el mismo sistema considerado por Roje en su Tesis de Magister [1] y se tuvo acceso a sus códigos desarrollados para las simulaciones, se procedió a la adaptación de estos desde el lenguaje de programación Matlab a Python, verificándose que el comportamiento dinámico de ambos modelos fuera el mismo.

- En función de lo anterior, se efectuaron correcciones de los parámetros K_y, los cuales se encontraban sobredimensionados, lo que se traducía en una hipersensibilidad de los cultivos a la escasez de agua.
- Para aproximar de mejor forma la dinámica real de los cultivos en la comunidad, también se corrigieron los parámetros del suelo y se utlizaron coordenadas reales de familias dentro de la comunidad.
- Se desarrolló un modelo de crecimiento de los cultivos que permitiera dar cuenta del efecto del estrés hídrico en el desarrollo de estos, tanto en su follaje como en sus raíces.
- Formulación de recompensas en consideración de los objetivos de maximizar rendimientos y minimizar los consumos de agua.
- Adaptación del modelo completo a la estructura del entorno Gym.
- Generación de variantes del modelo completo para el entrenamiento de agentes sin consideración de pozos y agentes especializados por tipo de cultivo.
- Entrenamiento de los agentes mediante aprendizaje reforzado con las implementaciones de Stable Baselines 3.
- Desarrollo de una implementación del algoritmo PSO utilizado en el esquema de control predictivo, este último considerado como base comparativa.
- Simulación de la interacción de los agentes entrenados y de las estrategias base de comparación con el modelo.
- Análisis de los resultados y rediseño de los esquemas (arquitectura de las redes neuronales, recompensas, horizontes de predicción, etc.).

Capítulo 2

Estado del arte

En esta sección se presentan en forma resumida las estrategias utilizadas típicamente para programar los riegos aplicados a los cultivos y las soluciones más modernas que buscan mejorar el desempeño de estas con el fin de lograr mayores rendimientos y/o una administración de los recursos hídricos más eficiente.

2.1. Programación del riego

2.1.1. Balance hídrico

La programación de riegos basada en el balance hídrico [2], utiliza la ecuación del mismo nombre (ver sección 3.3.1), la que relaciona la humedad del suelo con los ingresos (precipitaciones, riego y ascenso capilar) y egresos (evapotranspiración, escurrimientos y percolación profunda) de agua del terreno cultivado, buscando mantenerla dentro de ciertos límites que permitan la mayor disponibilidad de agua en la zona de la raíz y así evitar el estrés hídrico.

Se trata de la estrategia más utilizada en la práctica [2] (aparte de los métodos tradicionales) gracias a su formulación matemática sencilla, ajustabilidad a distintos intervalos de operación (horas, días, semanas, etc.) y a que permite obtener buenos resultados, siempre que se tengan estimaciones adecuadas de las variables involucradas.

2.1.2. Control PID

Los controladores PID (Proporcional-Integral-Derivativo) constituyen una de las técnicas de control más básicas y su aplicación está limitada a sistemas SISO (*Single Input Single Output*) definidos por una ecuación de transferencia, con la cual se determinan los parámetros del controlador [11].

En [12], utilizan este tipo de controladores, utilizando como señal de control el porcentaje de humedad del suelo y una referencia del 31 %. El actuador opera sobre el tiempo de operación de una bomba con una tasa de bombeo constante y la componente integral del controlador está limitada para entregar valores de riego no negativos. Si bien manifiestan la obtención de resultados positivos (seguimiento de la referencia adecuado), no especifican la determinación de los parámetros del controlador y su aplicabilidad se ve limitada a monocultivos.

2.1.3. Control difuso

El control difuso, busca aproximar el proceso de toma de decisión humano haciendo uso de modelos basados en la lógica difusa [13]. Dichos modelos cuentan con un conjunto de reglas (base de conocimiento) de la forma "si $x \in A$ entonces $y \in B$", donde x es la variable antecedente (entrada), y es la variable consecuente (salida) y (A, B) son conjuntos difusos definidos por sendas funciones de pertenencia $\mu_A(\cdot)$ y $\mu_B(\cdot)$ [13].

En [14], proponen la utilización de un controlador de este tipo, utilizando como variables antecedentes la temperatura, la radiación solar y la humedad del suelo, para manipular el tiempo de riego (variable consecuente) y controlar el porcentaje de humedad del suelo, siguiendo una referencia de 30 %. De acuerdo a

los resultados presentados, si bien el sesguimiento de la referencia no es adecuado, los autores plantean que su controlador elige bien al regar cuando la temperatura y la radiación solar son menores, ya que esto reduce las pérdidas por evaporación y en consecuencia implica un menor consumo de agua.

2.1.4. Control predictivo basado en modelos

Las estrategias de control predictivo basado en modelos (MPC del inglés *Model Predictive Control*) corresponden a un tipo de controladores óptimos, donde se busca minimizar una función de costos que depende de un modelo, utilizado para predecir y optimizar las variables del sistema a controlar. La principal ventaja de este tipo de estrategias, es que permiten la incorporación de restricciones operacionales, evitando de esta forma forzar a los actuadores a operar en condiciones límite, que pudieran derivar en desgastes de sus materiales, o exigir al sistema fuera de sus capacidades físicas [15].

Dentro de la literatura, se pueden encontrar aplicaciones de este tipo de control en los trabajos de [16] y [1], donde el primero hace uso de la ecuación de balance hídrico para derivar un modelo lineal de la humedad del suelo, el que utiliza para resolver el problema de optimización del controlador, basado en el seguimiento de una referencia de humedad, aplicando restricciones respecto de la cantidad de agua diaria regada y respecto de la humedad misma. Por su parte, el segundo trabajo hace uso de un modelo no lineal, con múltiples cultivos (el anterior sólo considera uno) y que busca maximizar los rendimientos de estos, valiéndose de la ecuación que relaciona el rendimiento relativo con la evapotranspiración relativa (ver sección 3.4).

2.1.5. Control con aprendizaje reforzado

El aprendizaje reforzado es una rama de la inteligencia computacional que busca replicar el aprendizaje a través de la experiencia adquirida por un agente, al interactuar éste con un entorno [17]. Dicha interacción se establece mediante la ejecución de acciones por parte del agente sobre el entorno, las cuales son determinadas de acuerdo a un estado observado (parcial o completamente) y a una política de decisión. La idea central en este tipo de aprendizaje, es que el agente busca maximizar una cierta función de retorno esperado, que depende de las recompensas recibidas luego de ejecutar una serie de acciones (ver capítulo 4).

En [18], proponen un sistema de gestión de riego para cultivos de arroz utilizando el esquema de aprendizaje reforzado *Deep Q-Learning* (DQN) con predicciones de lluvias a siete días. Las variables de estado consideradas como entrada del controlador son las predicciones de lluvia de los próximos siete días, descenso máximo permitido después de lluvias H_p, el nivel de descenso de agua en el terreno h_t y sus límites (h_{min}, h_{max}) (donde H_p, h_{min} y h_{max} están tabulados para distintas etapas de crecimiento). Si bien DQN considera un conjunto discreto de acciones a_i, los autores proponen una formulación que permite aplicar riegos $m_t \in \mathbb{R}^+$, con tres funciones $f_i(h_t, h_{max})$ $i \in \{1, 2, 3\}$, es decir, la toma de decisión discreta por el controlador se reduce a elegir una de estas funciones en cada instante. El análisis de sus resultados manifiesta que la estrategia de control utilizada redujo el consumo de agua, redujo la cantidad de eventos de riego y no presentó disminución en los rendimientos, en comparación con los métodos tradicionales de inundación.

2.1.6. Discusión

Si bien los métodos expuestos presentan desempeños satisfactorios según sus autores, existen aspectos que motivan el desarrollo de nuevas técnicas que mejoren la eficiencia del riego y entreguen altos rendimientos.

Particularmente, en el caso del balance hídrico, si bien ofrece un buen compromiso entre consumo de agua y rendimiento de los cultivos (razón por la cual se lo considera como punto de comparación para el sistema desarrollado en el presente trabajo), no está comprobada la optimalidad de su aplicación, lo que induce a la búsqueda de posibles soluciones con un consumo hídrico menor y sin perjuicio en el rendimiento.

Respecto al control PID, su principal debilidad es que sólo se lo puede considerar para sistemas SISO, lo que requiere un modelo de desarrollo de los cultivos sobresimplificado. Además, al tratarse de un controlador basado en el seguimiento de una referencia, su desempeño se verá mermado si esta cambia demasiado o si las condiciones del sistema controlado divergen demasiado de su punto de operación.

En cuanto al control difuso, adolece de la necesidad de conocimiento experto para el ajuste de sus parámetros, particularmente para las funciones de pertenencia. Además, su aplicación como método de seguimiento de referencias, le confiere las debilidades mencionadas para el control PID.

Respecto a los sistemas que consideran MPC, su principal desventaja es la gran capacidad de cómputo requerida para poder resolver en cada instante el problema de optimización que entrega la acción de control necesaria. Los principales factores que acrecientan este problema son horizontes de predicción elevados, modelos complejos del sistema, alta cantidad de variables a tomar en cuenta, etc. En este sentido, se suele buscar un balance entre el nivel de complejidad del modelo y el tiempo de cómputo, sin embargo, esto deriva en la obtención de soluciones sub-óptimas y en consecuencia, se pierde la principal ventaja asociada a este tipo de controladores.

Por último, el control con aprendizaje reforzado, ofrece sacar partido de las ventajas de optimalidad del control predictivo y de adecuación a cambios en el punto de operación del sistema. Sin embargo, la determinación del riego para cultivos es un problema de naturaleza continua. Respecto a la implementación mencionada anterirmente, esta utiliza un esquema de aprendizaje discreto, modificado para la obtención de valores continuos de riego, lo cual se puede simplifar mediante la utilización de esquemas de aprendizaje reforzado desarrollados para espacios de acción y de estado continuos. Además, si bien se plantea una economización de agua respecto al método tradicional por inundación, queda abierta la discusión de su eficiencia respecto a otros métodos, por ejemplo, el balance hídrico.

Respecto a esto último, el presente trabajo también propone la utilización del aprendizaje reforzado para la implementación de un controlador que gestione los volúmenes de riego aplicados a un conjunto de cultivos, teniendo en cuenta sus características de optimalidad y adapatabilidad. Además, se debe destacar como argumento a favor de este tipo de estrategias, su capacidad de "aprendizaje" mediante una interacción directa con el sistema a controlar, es decir, la capcidad de desarrollar un controlador cuyos parámetros no están basados en un modelo, sino que en la planta misma y que además es posible adecuarlos a varaiciones de aquella. Es esta una de las principales ventajas y diferenciadores que tienen los controladores basados en aprendizaje reforzado.

Junto con lo anterior, se destaca que los tiempos de cómputo de la acción de control con este tipo de estrategias (aprendizaje reforzado), son significativamente menores, en comparación a lo observado con MPC, donde se deben ejecutar múltiples simulaciones del modelo para determinar la acción de control, razón por la cual el modelo tiene restricciones en cuanto su complejidad. Sin embargo, con aprendizaje reforzado, el modelo de la planta, en caso de que se utilice uno, puede ser tan detallado y realista como se desee, ya que dicha complejidad no perjudica los tiempos de cómputo del controlador y de hecho es beneficiosa durante el entrenamiento para obtener una mejor operación con la planta real.

Capítulo 3

Sistema hidrogeológico

Tomando en consideración la nomenclatura utilizada en [1], el presente trabajo considera como "sistema hidrogeológico" a aquel constituido por un pozo, una bomba, un estanque y un terreno con cultivos, donde cada uno de estos sistemas está asociado a un agricultor en particular. La interacción entre estos sistemas hidrogeológicos toma lugar en el acuífero desde el cual se abastecen de agua los pozos. A modo de ejemplo, la Figura 3.1 muestra un par de sistemas hidrogeológicos interactuando mediante un acuífero confinado, además de la superficie piezométrica máxima (cuando no hay extracción de agua) y real (línea segmentada y línea punteada respectivamente).

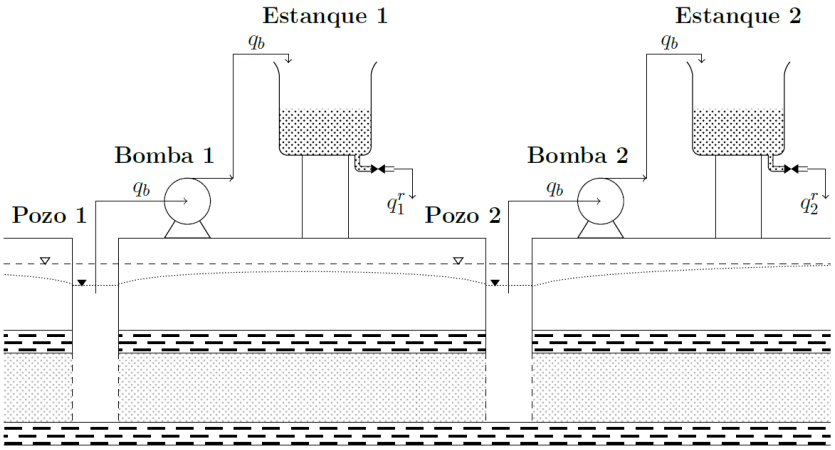


Figura 3.1: Dos sistemas hidrogeológicos con tasas de bombeo Q_1 y Q_2, interconectados mediante un acuífero confinado [1].

3.1. Profundidad del nivel de agua en un pozo

Considerando un pozo como el que se muestra en la Figura 3.2, desde el cual se extrae agua a una tasa de bombeo Q $[gal/min]$ constante, el descenso de la superficie piezométrica $d(r,t)$ $[ft]$, a una distancia r $[ft]$ (medida desde el centro del pozo) y habiendo transcurrido un tiempo t $[días]$ desde que comenzó el bombeo, se puede determinar según la ecuación de Theis [19]:

$$d(r,t) = \frac{Q}{4\pi T} W\left(\frac{r^2 S}{4Tt}\right), \tag{3.1}$$

con T la transmisividad del suelo donde se encuentra el acuífero en $[gpd/ft]$ (ver Apéndice A), S el coeficiente de almacenamiento del suelo (ver Apéndice A) y $W(\cdot)$ la función de pozos definida como:

$$W(u) = \int_u^\infty \frac{e^{-x}}{x} dx. \tag{3.2}$$

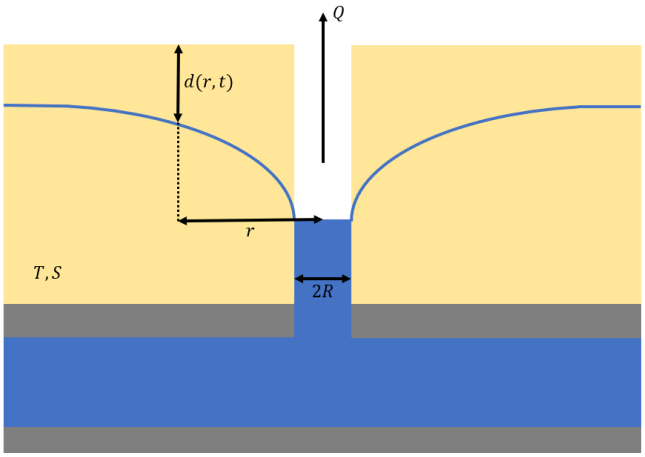


Figura 3.2: Descenso de la superficie piezométrica a una distancia r de un pozo con bombeo constante Q y radio R.

Luego, para una serie de n bombeos variables en el tiempo Q_k, como los presentados en la Figura 3.3, donde cada uno comienza en un instante t_k, se puede utilizar la ecuación de Theis y el principio de superposición para determinar la profundidad de la superficie piezométrica $d(r,t)$ como [20]:

$$d(r,t) = \frac{1}{4\pi T} \sum_{k=0}^{n-1} \Delta Q_k W\left(\frac{r^2 S}{4T(t-t_k)}\right), \tag{3.3}$$

con $\Delta Q_k = Q_k - Q_{k-1}$ y $\Delta Q_0 = Q_0$.

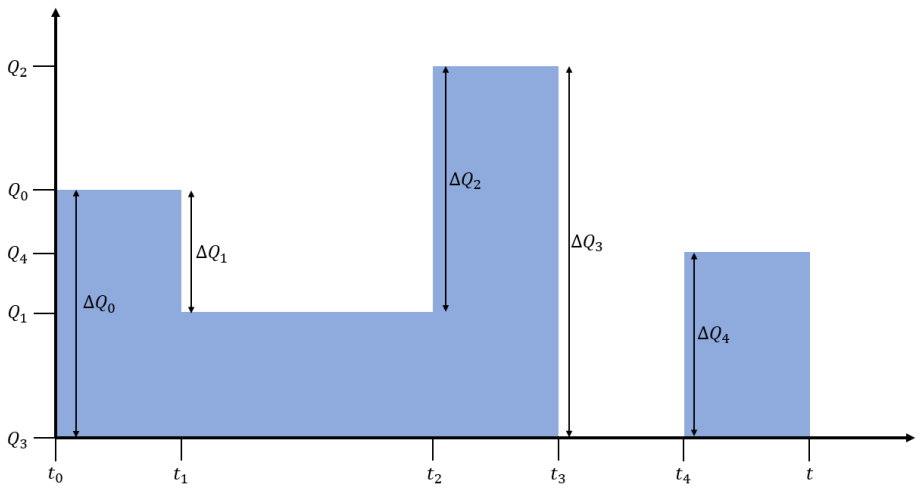


Figura 3.3: Bombeo variable en el tiempo.

Entonces, valiéndose de la ecuación 3.3 y del principio de superposición, para un sistema de m pozos, donde los bombeos varían diariamente, el nivel de descenso $d_{i,n}$ del i-ésimo pozo en el n-ésimo día se puede determinar como [21]:

$$d_{i,n} = \frac{1}{4\pi T} \sum_{j=1}^{m} \sum_{k=0}^{n-1} \Delta Q_{j,k} W\left(\frac{r_{i,j}^2 S}{4T(n-k)}\right), \tag{3.4}$$

con $r_{i,j}$ la distancia entre el i-ésimo y el j-ésimo pozo ($r_{i,i}$ corresponde al radio del i-ésimo pozo) y k el día en que inició el k-ésimo bombeo.

3.2. Recarga del acuífero y cota hídrica

La recarga del acuífero es la cantidad de agua que ingresa a este desde las capas superiores del suelo, principalmente por concepto de lluvias. Para modelar dicho fenómeno, se utiliza la fórmula de Chaturvedi [22], la que determina la cantidad de agua que ingresa al acuífero durante un año (medida en pulgadas) como:

$$R_{año} = \alpha(P_{año} - \beta)^{\gamma}, \tag{3.5}$$

donde P representa las precipitaciones anuales en $[in]$ y los parámetros α, β y γ dependen de las condiciones geográficas (en la formulación original $\alpha = 2$, $\beta = 15$ y $\gamma = 0{,}4$).

Ahora, dado que la fórmula original trabaja con valores anuales de precipitaciones, se la ajusta para operar con valores diarios y así determinar la recarga del acuífero durante un día [1]. También, dado que los datos de precipitaciones utilizados están medidos en milímetros, se utiliza el factor de conversión 25,4 ($25{,}4[mm] = 1[in]$) para aplicar adecuadamente la ecuación de Chaturvedi y poder estimar la recarga en milímetros como:

$$R_{día} = \frac{25{,}4\alpha}{365}\left(\frac{P_{día}}{25{,}4} - \beta\right)^{\gamma}, \tag{3.6}$$

con $P_{día}$ las precipitaciones diarias en $[mm]$.

Luego, el ascenso diario de la cota hídrica Δh se puede determinar como [23]:

$$\Delta h = \frac{R_{día}}{S_y}, \tag{3.7}$$

con S_y el rendimiento específico del suelo (ver Apéndice A).

3.3. Dinámica de los cultivos

3.3.1. Balance hídrico

El balance hídrico del terreno donde se encuentra un cultivo, corresponde a la evolución de la cantidad de agua almacenada en este, en función de los ingresos (precipitaciones, riegos y ascenso capilar) y egresos (evapotranspiración, percolación profunda y escurrimiento superficial) de agua. Dicha cantidad se mide en términos del agotamiento en la zona radicular D_r (*root zone depletion*). Luego, al final del k-ésimo día, el agotamiento $D_{r,k}$ se determina según la ecuación de balance hídrico [2]:

$$D_{r,k} = D_{r,k-1} + RO_k + ET_{a,k} + DP_k - P_k - CR_k - I_k, \tag{3.8}$$

con $D_{r,k-1}$ el agotamiento al comienzo del día evaluado (igual al agotamiento al final del día anterior), RO_k el escurrimiento superficial (*runoff*), $ET_{a,k}$ la evapotranspiración, DP_k la percolación profunda (*deep percolation*), P_k las precipitaciones, CR_k el ascenso capilar (*capillary rise*) e I_k el agua que ingresa por concepto de riego. Todas estas variables se miden en $[mm]$, ya que el balance volumétrico se realiza considerando bloques de $1[m^3]$ de suelo y asumiendo que el agua se distribuye de forma homogenea en el área del cultivo en consideración. La Figura 3.4 muestra una representación gráfica típica del balance hídrico en la zona de la raiz, además de presentar cotas hídricas asociadas a la cantidad de agua máxima que puede retener el suelo, el umbral de estrés hídrico y el punto de marchitez, que se explican más adelante.

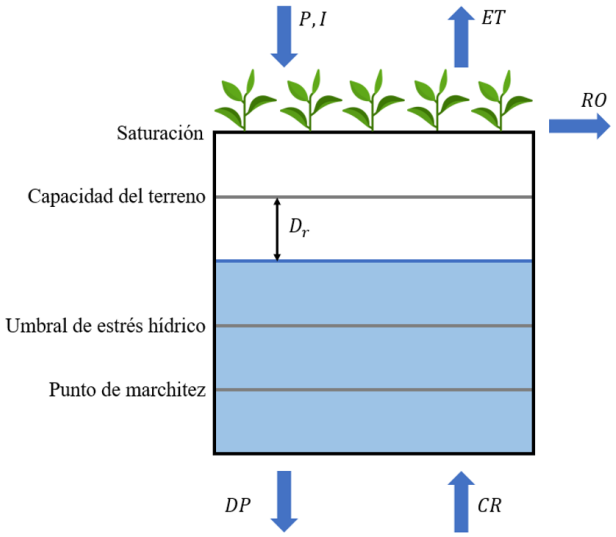


Figura 3.4: Balance de agua en la zona radicular.

A continuación, se desarrollan en mayor detalle las variables que intervienen en la ecuación 3.8.

Escurrimiento superficial

El escurrimiento superficial, corresponde al agua de lluvia que no logra permear hacia las capas inferiores del terreno y que fluye por la superficie de este. Para determinar esta variable, se utiliza la metodología del número de curva recomendada por [24], donde se considera el total de milímetros de agua precipitada durante un día. En primer lugar, se calcula la máxima profundidad de lluvia retenida durante una lluvia S (en $[mm]$) como:

$$S = 250\left(\frac{100}{CN} - 1\right), \tag{3.9}$$

con CN el número de curva, que depende del tipo de cultivo, características del suelo y factores medioambientales. El parámetro CN da cuenta de la impermeabilidad del terreno cultivado y va desde 0 (totalmente permeable) hasta 100 (totalmente impermeable).

Luego, si las precipitaciones P del día cumplen que $P > 0{,}2S$, se producirá escurrimiento en la superficie y se tiene que:

$$RO = \frac{(P - 0{,}2S)^2}{P + 0{,}8S} \tag{3.10}$$

Evapotranspiración

Dentro de las variables que determinan la evolución del nivel de agua en el terreno cultivado, la evapotranspiración es una de las más relevantes, ya que es la que permite determinar el rendimiento que se ha de obtener cuando se realice la cosecha [25] (ver sección 3.4).

La evapotranspiración hace referencia a las pérdidas de agua por concepto de evaporación en la superficie del terreno cultivado y de transpiración en las hojas de los cultivos. Al tratarse de dos fenómenos difícilmente medibles por separado, es que se los considera dentro de un mismo concepto. Ahora, si bien existe una metodolgía que permite estimar la evapotranspiración en función de la evaporación y la transpiración por separado, se recomienda su utilización para investigaciones especializadas en este fenómeno.

Ahora, para calcular la evapotranspiración en el transcurso de un día, se utiliza la metodología presentada en [2], partiendo por la evapotranspiración de referencia, luego la potencial o máxima y finalmente la real.

Las variables meteorológicas necesarias para la determinación de las distintas evapotranspiraciones son la radiación solar, la humedad relativa del aire, la velocidad del viento y la temperatura. Es importante destacar que la evapotranspiración también puede ser medida utilizando un lisímetro o estimada con el método del tanque de evaporación, sin embargo, estos elementos no están presentes en el área estudiada, razón por la cual se utiliza la metodología expuesta a continuación.

Entonces, comenzando con la evapotranspiración de referencia ET_o, esta se determina de acuerdo con la ecuación FAO Penman-Monteith [2], que es una versión estandarizada de la ecuación original de Penman-Monteith [26]:

$$ET_o = \frac{0{,}408\Delta(R_n - G) + \gamma \frac{900}{T+273} u_2 (e_s - e_a)}{\Delta + \gamma(1 + 0{,}34u_2)}, \tag{3.11}$$

con ET_o la evapotranspiración de referencia en $[mm]$, Δ la pendiente de la curva de presión de vapor respecto a la temperatura en $[kPa/°C]$, R_n la radiación solar diaria neta en la superficie del cultivo en $[MJ/(m^2día)]$, G la densidad de flujo de calor del suelo en $[MJ/(m^2día)]$, γ la constante psicrométrica en $[kPa/°C]$, T la temperatura media a dos metros de altura en $[°C]$, u_2 la velocidad del viento a dos metros de altura en $[m/s]$, e_s la presión de vapor de saturación en $[kPa]$ y e_a la presión de vapor real en $[kPa]$. La determinación de cada una de estas variables se encuentra en el Apéndice B. Es importante destacar que los valores numéricos presentes en la ecuación 3.11 ajustan las unidades de forma tal que ET_o se mide en milímetros [2].

Una vez se ha determinado la evapotranspiración de referencia ET_o, se procede a calcular la evapotranspiración máxima o potencial ET_p del cultivo en consideración. Para esto, [2] recomienda el método del coeficiente de cultivo K_c al trabajar con intervalos de tiempo diarios, donde la evapotranspiración máxima ET_p se determina como:

$$ET_p = K_c ET_o. \tag{3.12}$$

El coeficiente K_c depende de la etapa de crecimiento del cultivo y las características geográficas donde se desarrolla. Para determinar el valor de K_c correspondiente al día en consideración, se hace uso de valores tabulados de $K_{c,ini}$, $K_{c,mid}$ y $K_{c,end}$ para construir la curva del parámetro K_c como la que se muestra en la Figura 3.5 y desde allí extrapolar. Los valores $K_{c,ini}$, $K_{c,mid}$ y $K_{c,end}$ se pueden ajustar de mejor forma si se cuenta con datos históricos de frecuencia de riego, velocidad del viento, humedad relativa y altura del cultivo en la zona evaluada.

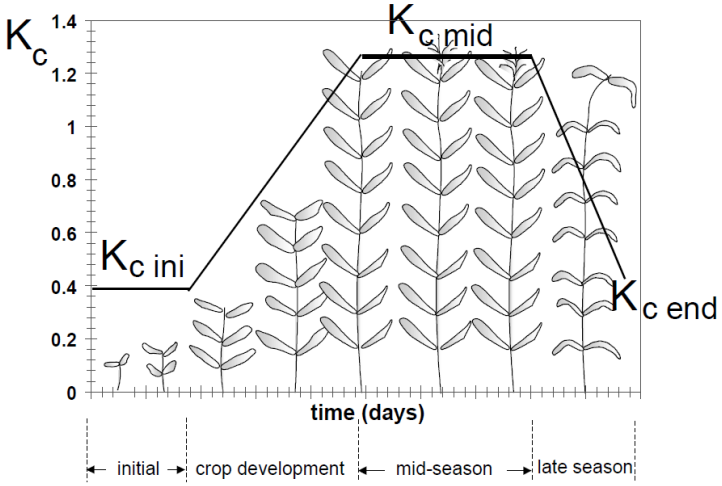


Figura 3.5: Curva teórica del parámetro K_c [2].

Ahora, la evapotranspiración real ET_a dependerá del grado de estrés hídrico al cual esté sometido el cultivo, el cual se representa mediante el coeficiente de estrés K_s y que da cuenta de la disponibilidad de agua para suplir los requerimientos hídricos del cultivo.

Para determinar K_s, primeramente se calcula el agua disponible total en la zona de la raíz TAW (*total available water*) como:

$$TAW = 1000(\theta_{fc} - \theta_{wp})Z_r, \tag{3.13}$$

con θ_{fc} el contenido de humedad a capacidad de campo en $[m^3/m^3]$, θ_{wp} el contenido de humedad en el punto de marchitez en $[m^3/m^3]$ y Z_r la profundidad de las raíces en $[m]$. Los parámetros θ_{fc} y θ_{wp} dependen del tipo de suelo y están tabulados. Por su parte, la profundidad de las raíces puede ser medida o estimada empíricamente, sin embargo, para efectos de simulación se requiere de un modelo de crecimiento acorde a los datos que se tengan (ver sección 5.3).

Una vez determinado TAW, se calcula el agua fácilmente aprovechable RAW (*readily available water*), que corresponde al porcentaje de agua que puede absorver la planta a través de su raíces sin sufrir estrés hídrico, según:

$$RAW = p \cdot TAW, \tag{3.14}$$

con p la fracción promedio del total de agua disponible en el suelo que puede ser agotada de la zona radicular antes de presentarse estrés hídrico ($p \in [0, 1]$). En general, p se encuentra tabulado según el tipo de cultivo y se lo puede ajustar en función de ET_p como:

$$p_{ajustado} = p + 0{,}04(5 - ET_p). \tag{3.15}$$

Con esto, el coeficiente de estrés K_s se determina como [2]:

$$K_s = \begin{cases} 1 & D_r \leq RAW \\ \frac{TAW - D_r}{TAW - RAW} & D_r > RAW \end{cases}. \tag{3.16}$$

Es importante notar que el agotamiento D_r está acotado por el intervalo $[0, TAW]$ con lo cual $K_s \geq 0$. La Figura 3.6 muestra las variables TAW, RAW, D_r, K_s y su interacción mediante la ecuación 3.16.

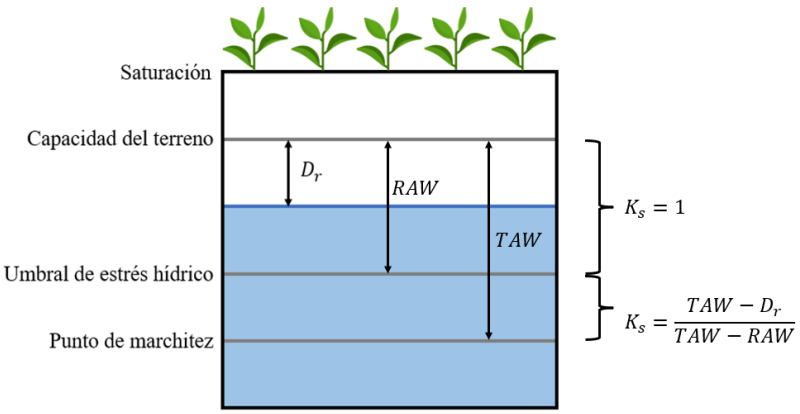


Figura 3.6: Interpretación gráfica de las variables TAW, RAW, D_r y K_s.

Finalmente, una vez determinado el valor de K_s, se tiene que la evapotranspiración real ET_a es de la forma:

$$ET_a = K_s ET_p. \tag{3.17}$$

Percolación profunda

La percolación profunda DP corresponde al agua que se pierde (deja de estar disponible para el cultivo) al descender hacia las capas inferiores del terreno cultivado, lo que ocurre cuando el agua que ingresa por concepto de lluvias sobrepasa la capacidad de campo. De acuerdo con [2], se puede determinar reordenando la ecuación del balance hídrico e imponiendo $D_{r,k} = 0$ como:

$$DP_k = P_k + CR_k + I_k - D_{r,k-1} - RO_k - ET_{a,k}. \tag{3.18}$$

Sin embargo, al utilizar DP_k, calculado con la ecuación 3.18, en el balance hídrico (ecuación 3.8), se obtiene $D_{r,k} = 0$, es decir, se genera una dependencia circular. Es por esto que en general se considera $DP_k = 0$ y se opera con $D_{r,k} \in [0, TAW]$.

Ascenso capilar

El ascenso capilar CR corresponde al agua que ingresa a la zona de la raíz proveniente desde las capas inferiores del terreno y depende del tipo de suelo, profundidad de las raíces, profundidad de la cota hídrica y la humedad del terreno. En general, se considera $CR = 0$, sobre todo cuando la cota hídrica está a más de un metro de profundidad respecto a las raíces [2].

3.3.2. Discusión

Si bien la dinámica de cultivos basada en el balance hídrico es uno de los modelos más utilizados, junto con la metodología fundada en la ecuación FAO Penman-Monteith para la estimación de la evapotranspiración, se destacan algunas dificultades en su implementación. En primer lugar, el modelo no considera el perjuicio provocado a los cultivos al existir ingresos de agua excesivos, lo que en la práctica puede llevar a la muerte de las plantas. Segundo, los parámetros definidos en función de la profundidad de las raíces (TAW y RAW) requieren de un modelo de crecimiento, sin embargo, si no se cuenta con los datos necesarios para la implementación de un modelo confiable, el crecimiento puede resultar estimado de forma inadecuada. Por último, existe poca información respecto a la estimación adecuada de la percolación profunda y el ascenso capilar, considerándoselos en general como variables despreciables.

3.4. Rendimiento de los cultivos

El rendimiento efectivo o real Y_a de un cultivo, es la cantidad cosechada por unidad de área al término del periodo de madurez. Esta variable, se puede calcular en función del rendimiento relativo Y_r, que no es más que el cuociente entre el rendimiento efectivo y el rendimiento máximo Y_{max} del cultivo, donde este último es un valor que se encuentra tabulado y que depende del cultivo en consideración, la estacionalidad y ubicación geográfica del mismo [25]. De esta forma se tiene:

$$Y_r = \frac{Y_a}{Y_{max}} \tag{3.19}$$

$$\Rightarrow Y_a = Y_r \cdot Y_{max}. \tag{3.20}$$

Por su parte, para determinar el rendimiento relativo Y_r, se hace uso de la ecuación introducida por Doorenbos y Kassam en 1979 [25], la que lo relaciona con el cuociente entre la evapotranspiración real ET_a y la potencial ET_p (evapotranspiración relativa):

$$Y_r = 1 - K_y \left(1 - \frac{ET_a}{ET_p}\right), \tag{3.21}$$

con K_y el factor de respuesta del rendimiento y que se encuentra tabulado según cultivo.

Es importante notar que la ecuación 3.21 considera las evapotranspiraciones de una etapa o de la temporada completa de desarrollo del cultivo, razón por la cual se propone utilizar una versión modificada [27], que permite estimar el rendimiento relativo Y_r en función de valores diarios de evapotranspiración como:

$$Y_r = \prod_{i=1}^{N} \prod_{k=1}^{T_i} \left(1 - K_{y,i} \left(1 - \frac{ET_{a,k}}{ET_{p,k}}\right)\right)^{\frac{1}{T_i}}, \tag{3.22}$$

con N la cantidad de etapas de crecimiento del cultivo, T_i la cantidad de días que dura la i-ésima etapa y $K_{y,i}$ el factor de respuesta del rendimiento de la i-ésima etapa.

Capítulo 4

Aprendizaje reforzado

El aprendizaje reforzado es una rama de la inteligencia computacional que busca replicar el aprendizaje a través de la experiencia adquirida por un agente, al interactuar éste con un entorno [17]. Dicha interacción se establece mediante la ejecución de acciones por parte del agente sobre el entorno, las cuales son determinadas de acuerdo a un estado observado (parcial o completamente) y a una política de decisión. La idea central en este tipo de aprendizaje, es que el agente busca maximizar una cierta función de retorno esperado, que depende de las recompensas recibidas luego de ejecutar una serie de acciones durante un episodio.

La Figura 4.1 muestra el esquema básico de aprendizaje reforzado. Aquí, en cada instante t, el agente observa un estado s_t (en general se considera que el estado es completamente observable) y actúa sobre el entorno ejecutando una acción a_t según la política de decisión π con la cual esté operando en dicho instante. Tras la acción, el entorno evoluciona hacia un nuevo estado s_{t+1} y el agente recibe una recompensa r_t.

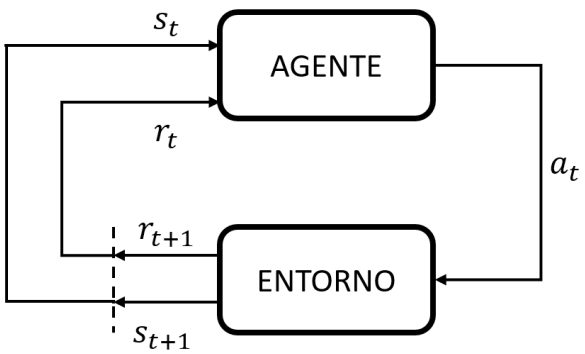


Figura 4.1: Diagrama típico de un algoritmo de aprendizaje reforzado.

Proceso de decisión de Markov

La formalización matemática para la dinámica descrita anteriormente corresponde a lo que se conoce como un proceso de decisión de Markov (MDP por su nombre en inglés *Markov Decision Process*), el que está definido por un espacio de estado $\mathcal{S}$, un espacio de acción $\mathcal{A}$, una probabilidad de transición $p(s'|s,a)$ con $s', s \in \mathcal{S}$ y $a \in \mathcal{A}$, una función de recompensa $r(a,s)$ y una política de decisión π [28], la cual puede ser determinística ($a = \pi(s)$) o estocástica ($\mathbb{P}(a) = \pi(a|s)$).

Luego, para una secuencia $s_t, a_t, s_{t+1}, a_{t+1}, \ldots$ se define el retorno R_t como:

$$R_t = \sum_{k=0}^{\infty} \gamma^k r_{t+k}, \tag{4.1}$$

con $\gamma \in [0,1]$ el factor de descuento, dando cuenta de que las recompensas futuras son menos relevantes que las actuales debido a su incertidumbre y $r_{t+k} = r(a_{t+k}, s_{t+k})$. Con esto, el principal objetivo de optimi-

zación en un MDP, consiste en encontrar una política óptima π^* que maximice el retorno esperado $\mathbb{E}_\pi[R_0]$ de una secuencia.

Funciones de valor y ecuación de Bellman

A partir de la ecuación 4.1, se pueden definir las funciones de valor para una acción $Q_\pi(s_t, a_t)$ y para un estado $V_\pi(s_t)$, además de la función de ventaja $A_\pi(s_t, a_t)$ como:

$$\begin{aligned} Q_\pi(s_t, a_t) &= \mathbb{E}_\pi[R_t|s_t, a_t], && (4.2)\\ V_\pi(s_t) &= \mathbb{E}_\pi[R_t|s_t], && (4.3)\\ A_\pi(s_t, a_t) &= Q_\pi(s_t, a_t) - V_\pi(s_t) && (4.4) \end{aligned}$$

donde la primera da cuenta de la conveniencia de ejecutar la acción a_t cuando se está en el estado s_t y la segunda evalúa la conveniencia de encontrarse en el estado s_t. Estas dos funciones son utilizadas en la mayoría de los algoritmos de aprendizaje reforzado (ya sea considerando una de las dos o ambas). Por su parte, la tercera función se puede interpretar igual que la primera y los algoritmos que la consideran, la utilizan en reemplazo de Q_π para reducir la varianza en el paso de actualización durante el entrenamiento [17].

Una propiedad importante de las funciones $Q_{\pi^*}(s_t, a_t)$, $V_{\pi^*}(s_t)$ y $A_{\pi^*}(s_t, a_t)$ es que satisfacen la ecuación de Bellman (esto es válido para π^*). Particularmente, para la función de valor de la acción se tiene:

$$Q_{\pi^*}(s_t, a_t) = \mathbb{E}_{\pi^*}[r_t + \gamma \mathbb{E}_{\pi^*}[Q_{\pi^*}(s_{t+1}, a_{t+1})]]. \quad (4.5)$$

La ecuación de Bellman se utiliza como un método de recursión para actualizar las estimaciones de estas funciones durante el entrenamiento y de la política óptima π^* del MDP.

Exploración vs explotación

Durante el proceso de entrenamiento del agente, periodo en el cual se busca encontrar la mejor política de decisión, se deben conciliar dos aspectos relevantes: exploración y explotación.

La exploración, consiste en probar la mayor cantidad posible de pares estado-acción, con el fin de contar con un conocimiento más amplio respecto a las consecuencias que tienen las acciones elegidas en un momento determinado. Esto ayuda en la determiandción de las acciones más convenientes, ya que permite explorar "rutas" que eventualmente retornen mayores recompensas totales.

Por otro lado, la explotación consiste en reforzar o potenciar la selección de las acciones que permitan maximizar el retorno total obtenido por el agente al término de un periodo. Esto implica aumentar la probabilidad de seguir "rutas" que entregan retornos totales mayores, de acuerdo a la experiencia que se ha adquirido mediante la interacción con el entorno.

Luego, es recomendable comenzar con un entrenamiento altamente exploratorio y transitar hacia uno más exploratorio.

Algoritmos *On-policy* y *Off-policy*

Los algoritmos de aprendizaje reforzado se pueden dividir en dos categorías: *on-policy* y *off-policy*.

La primera categoría, hace referencia a que durante el entrenamiento se cuenta con sólo una política π, la cual es utilizada por el agente para generar tuplas (s_t, a_t, r_t, s_{t+1}) y para la actualización de la misma (junto con sus funciones de valor).

Por otro lado, los algoritmos $off-policy$ cuentan con dos políticas durante el entrenamiento, una para la obtención de las tuplas (llamada política conductual) y otra para la actualización (llamada política objetivo), siendo esta última la que se utiliza una vez terminado el entrenamiento.

4.1. Métodos por gradiente de la política

Los métodos por gradiente de la política están basados en el teorema del mismo nombre [17] [29], el cual establece que:

$$\nabla_\theta J(\pi_\theta) = \mathbb{E}_{\pi_\theta}[\nabla_\theta \log(\pi_\theta(a|s)) Q_\pi(s,a)], \tag{4.6}$$

donde $J(\pi_\theta)$ es una función de desempeño que debe ser maximizada y que típicamente toma la forma de la esperanza de alguna función de valor (ecuaciones 4.2, 4.3 o 4.4), con lo que la maximización de $J(\pi_\theta)$ equivale a maximizar el retorno esperado. Por su parte, θ es una parametrización de la política π, generalmente una red neuronal artificial (ANN del inglés *Artificial Neural Network*), aunque también puede tomar otras formas.

Luego, estos algoritmos buscan aproximar la función de desempeño de la ecuación 4.6, mediante el muestreo de transiciones (s_t, a_t, r_t, s_{t+1}) durante el entrenamiento, promediando sus funciones de valor y actualizando los parámetros θ de la política π_θ, ejecutando un ascenso del gradiente de la forma:

$$\theta = \theta + \alpha \nabla_\theta \hat{J}(\pi_\theta), \tag{4.7}$$

donde α es la tasa de aprendizaje y $\hat{J}(\pi_\theta)$ corresponde a la función de desempeño estimada.

Los principales factores diferenciadores entre los distintos algoritmos de aprendizaje reforzado que utilizan el gradiente de la política son: la función de valor considerada como el desempeño $J(\pi_\theta)$ y la forma de estimarla.

Finalmente, es importante mencionar que una de las principales ventajas de este tipo de métodos, es su aplicabilidad a MDP con espacios de acción y de estado continuos.

4.1.1. Estructura actor-crítico

La estructura actor-crítico es un tipo de arquitectura utilizado por algunos métodos por gradiente de la política, donde el actor corresponde a una parametrización θ de la política π y el crítico corresponde a una parametrización w de la función de valor utilizada en la función de desempeño J. Típicamente, ambos elementos son implementados como ANN, donde los parámetros θ y w son los pesos y *biases* de las redes.

En este tipo de entrenamientos, el actor interactúa con el entorno siguiendo la política π_θ y generando tuplas de la forma $(s_t, a_t, r_t, s_{t+1}, a_{t+1})$, mientras que el crítico estima el valor o la calidad del estado-acción de acuerdo a los parámetros w (por ejemplo: $Q_w(s,a)$, $V_w(s)$ o $A_w(s,a)$). Luego, de acuerdo con las estimaciones realizadas por el crítico, se realiza la actualización de los parámetros θ mediante el ascenso del gradiente de la política (ecuación 4.6) y los parámetros w mediante el descenso del gradiente del promedio cuadrático (para aproximar el valor esperado) de la diferencia temporal δ_t, derivada de la ecuación de Bellman (ecuación 4.5):

$$\delta_t = r_t + \gamma v_w(s_{t+1}, a_{t+1}) - v_w(s_t, a_t), \tag{4.8}$$

con $v_w(s,t)$ la función de valor que corresponda.

Para finalizar, se debe destacar que el crítico se utiliza únicamente durante la etapa de entrenamiento y que al finalizar éste, la política de decisión queda completamente definida por el actor.

4.2. Proximal Policy Optimization

El algoritmo *Proximal Policy Optimization* (PPO) [30] es un algoritmo de aprendizaje reforzado *on-policy* basado en la estructura actor-critic, donde la política aprendida es estocástica de la forma $\mathbb{P}(a) = \pi(a|s)$. Su aplicabilidad es adecuada tanto para MDP con $(\mathcal{S}, \mathcal{A})$ discretos y/o continuos.

La función de desempeño $J(\pi_\theta)$ a maximizar se define como:

$$J(\pi_\theta) = \mathbb{E}\left[\text{mín}\left(\frac{\pi_\theta(a_t|s_t)}{\pi_{\theta_{old}}(a_t|s_t)}\hat{A}_t, \text{clip}\left(\frac{\pi_\theta(a_t|s_t)}{\pi_{\theta_{old}}(a_t|s_t)}, 1-\epsilon, 1+\epsilon\right)\hat{A}_t\right)\right], \tag{4.9}$$

con ϵ un hiperparámetro (típicamente $\epsilon = 0{,}2$) y $\hat{A}_t$ una estimación de la función de ventaja (ecuación 4.4) definida por el crítico V_w como:

$$\hat{A}_t = \sum_{k=0}^{T} \gamma^k \delta_{t+k}, \tag{4.10}$$

con $\delta_t = r_t + \gamma V_w(s_{t+1}) - V_w(s_t)$ y T la cantidad de transiciones realizadas.

La lógica detrás de esta formulación está relacionada con limitar el paso de actualización del gradiente de la política para evitar un colapso en el desempeño, es decir, la variación de los parámetros θ está limitada por un cierto intervalo definido por el parámetro ϵ. Para comprender este comportamiento, considerar la Figura 4.2 para una acción a_t. Si la ventaja $\hat{A}_t$ es positiva, el desempeño está limitado superiormente por $1+\epsilon$. Por otro lado, si la ventaja $\hat{A}_t$ es negativa, el desempeño está limitado superiormente por $1-\epsilon$. Esto significa que PPO considera un objetivo del tipo pesimista.

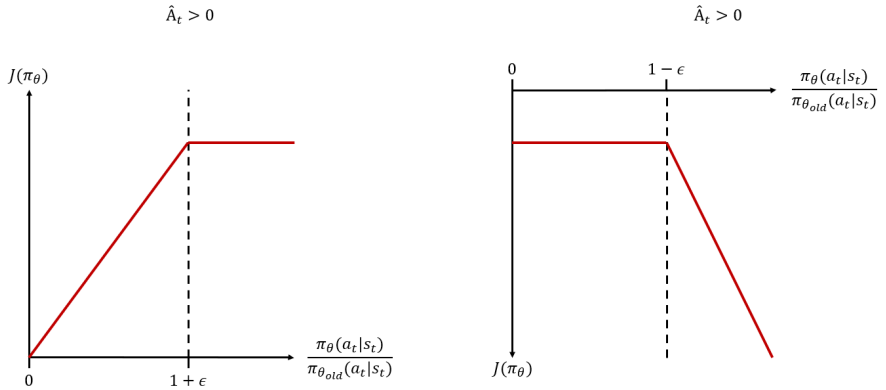


Figura 4.2: Función de desempeño $J(\pi_\theta)$ para una sola transición cuando $\hat{A}_t > 0$ y cuando $\hat{A}_t < 0$.

A continuación, se presenta el pseudocódigo del algoritmo PPO, donde se destaca la posibilidad de realizar un entrenamiento con múltiples actores en paralelo, lo que permite una mayor exploración y donde cada uno cuenta con la versión más actualizada de la política π_θ.

Algorithm 1: Entrenamiento PPO.

Inicializar aleatoriamente los parámetros del actor θ_0 y del crítico w_0
for $k = 0, 1, \ldots$ **do**
  for *actor*$= 1, 2, \ldots, N$ **do**
    Aplicar la política π_{θ_k} sobre el entorno durante T pasos de tiempo
    Estimar las ventajas $\hat{A}_t$ según la función de valor actual V_{w_k}
  end
  Actualizar la política actual según:

$$\theta_{k+1} = \arg \max_{\theta} J(\pi_{\theta_k})$$

  Actualizar la función de valor según:

$$w_{k+1} = \arg \min_{w} \mathbb{E}[\delta_t^2]$$

end

4.3. Deep Deterministic Policy Gradient

Deep Deterministic Policy Gradient (DDPG) [31] es un algoritmo de aprendizaje reforzado *off-policy* basado en la estructura actor-critic, donde como lo indica su nombre, la política aprendida es determinística de la forma $a = \pi(s)$. DDPG se caracteriza por la incorporación de la estructura actor-crítico por partida doble. Por un lado, se tiene al actor π_θ y al crítico Q_w exploratorios y por otro, al actor $\pi_{\theta'}$ y al crítico $Q_{w'}$ objetivos. La justificación detrás de esta configuración es que vuelve la actualización de los parámetros θ y w más estable durante el entrenamiento. Es importante mencionar que DDPG sólo puede ser utilizado en MDPs con espacios de acción continuos.

Otra característica de DDPG, es la incorporación de un buffer (típicamente llamado *replay buffer*) donde se van almacenando las tuplas (s_t, a_t, r_t, s_{t+1}) obtenidas durante el entrenamiento. La razón para utilizar un buffer, es que permite actualizar las ANN del sistema en función de subconjuntos de tuplas seleccionadas aleatoriamente en cada instante y con baja correlación entre sí, lo que ayuda a la estabilidad del entrenamiento.

Al tratarse de un algoritmo con una política determinística, DDPG considera la adición de ruido a las acciones determinadas por el actor durante el entrenamiento, con la finalidad de promover la exploración. Si bien en [31] recomiendan la utilización de un proceso Ornstein-Uhlenbeck, es posible utilizar otro tipo ruidos, por ejemplo ruido gaussiano blanco aditivo.

Ahora, la función de desempeño $J(\pi_\theta)$ a maximizar en DDPG es:

$$J(\pi_\theta) = \mathbb{E}[Q_w(s, a)], \tag{4.11}$$

para lo cual se utiliza el gradiente de la política.

A continuación, se presenta el pseudocódigo del algoritmo DDPG:

Algorithm 2: Entrenamiento DDPG

Inicializar aleatoriamente al crítico Q_w y al actor π_θ
Inicializar al crítico objetivo $Q_{w'}$ y al actor objetivo $\pi_{\theta'}$ con pesos $\theta' = \theta$ y $w' = w$
Inicializar el replay buffer $\mathcal{B}$
for *episodio* $\in (1, M)$ **do**
    Inicializar un proceso aleatorio $\mathcal{N}$ para exploración
    Observar el estado inicial s_0
    for $t \in (1, T)$ **do**
        Ejecutar la acción $a_t = \pi_\theta(s_t) + \mathcal{N}_t$
        Observar la recompensa r_t y el nuevo estado s_{t+1}
        Almacenar en $\mathcal{B}$ la tupla (s_t, a_t, r_t, s_{t+1})
        Seleccionar aleatoriamente N tuplas $(s_i, a_i, r_i, s_{i+1}) \in \mathcal{B}$
        Computar: $y_i = r_i + \gamma Q_{w'}(s_{i+1}, \pi_{\theta'}(s_{i+1}))$
        Actualizar el crítico minimizando la pérdida: $L = \frac{1}{N}\sum_i (y_i - Q_\theta(s_i, a_i))^2$
        Actualizar el actor maximizando el retorno muestreado $J(\pi_\theta)$ usando el gradiente:

$$\nabla_\theta J(\pi_\theta) \approx \frac{1}{N} \sum_i \nabla_a Q_w(s_i, a_i) \nabla_\theta \pi_\theta(s_i)$$

        Actualizar los parámetros del actor y el crítico objetivo según:

$$\begin{aligned} \theta' &= \tau\theta + (1-\tau)\theta' \\ w' &= \tau w + (1-\tau)w' \end{aligned}$$

    end
end

4.4. Twin Delayed Deep Deterministic Policy Gradient

Twin Delayed Deep Deterministic Policy Gradient (TD3) [32] es un algoritmo de aprendizaje reforzado *off-policy* basado en la estructura actor-critic, donde la política aprendida es determinística de la forma $a = \pi(s)$. Se trata de una versión modificada de DDPG, donde se busca subsanar tres fuentes de mal desempeño: alta varianza en las estimaciones, sobreajuste a *peaks* y sobreestimación de la función de valor.

Respecto a la varianza en las estimaciones, TD3 propone reducir la frecuencia con que se actualiza el actor π_θ y las ANNs objetivo respecto al crítico Q_w. Particularmente, considerando una actualización del crítico en cada paso de tiempo, se propone una actualización de los demás parámetros cada dos pasos de tiempo. Notar que esta modificación le otorga el nombre de *delay*.

Luego, respecto al sobreajuste a *peaks* de la función de valor estimada, TD3 utiliza una técnica de regularización de suavización de la política objetivo, donde se agrega un ruido truncado a la acción determinada por el actor objetivo en el paso de actualización, es decir, durante el entrenamiento, cuando se calcula y_i (línea 12 en el pseudocódigo de DDPG), se adiciona un ruido $\epsilon \sim \text{clip}(\mathcal{N}(0, \sigma), -c, c)$, con lo que se tiene:

$$y_i = r_i + \gamma Q_{w'}(s_{i+1}, \pi_{\theta'}(s_{i+1}) + \epsilon). \tag{4.12}$$

El truncamiento de ϵ es para mantener $\pi_{\theta'}(s_{i+1}) + \epsilon$ dentro del espacio de acción.

Por último, para reducir la sobreestimación de la función de valor, TD3 utiliza un segundo actor exploratorio (con su respectivo actor objetivo), es decir, TD3 cuenta con seis redes neuronales durante el entrenamiento: un actor π_θ, un crítico Q_{w_1}, un crítico Q_{w_2} y sendas redes neuronales objetivo $\pi_{\theta'}$, $Q_{w'_1}$ y $Q_{w'_2}$ (esta modificación le otorga el nombre *twin*). Luego, cuando se determinan los objetivos y_i, se elige el mínimo entre $Q_{w'_1}$ y $Q_{w'_2}$, con lo que se tiene:

$$y_i = r_i + \gamma \min_{k=1,2} Q_{w'_k}(s_{i+1}, \tilde{a}_{i+1}), \tag{4.13}$$

con $\tilde{a}_{i+1} = \pi_{\theta'}(s_{i+1}) + \epsilon$ ($\epsilon \sim \text{clip}(\mathcal{N}(0, \sigma), -c, c)$).

A continuación, se presenta el pseudocódigo del algoritmo TD3:

Algorithm 3: Entrenamiento TD3.

1 Inicializar al actor π_θ y los críticos Q_{w_1} y Q_{w_2} con parámetros aleatorios
2 Inicializar las redes objetivo: $\pi_{\theta'} = \pi_\theta$, $Q_{w'_1} = Q_{w_1}$ y $Q_{w'_2} = Q_{w_2}$
3 Inicializar el buffer $\mathcal{B}$
4 **for** $t = 1, \ldots, T$ **do**
5 Seleccionar acción con ruido de exploración $a_t = \pi_\theta(s_t) + \epsilon$ ($\epsilon \sim \mathcal{N}(0, \sigma)$)
6 Observar la recompensa r_t y estado nuevo s_{t+1}
7 Guardar transición (s_t, a_t, r_t, s_{t+1}) en $\mathcal{B}$
8 Seleccionar aleatoriamente N transiciones $(s_i, a_i, r_i, s_{i+1}) \in \mathcal{B}$
9 Calcular las acciones objetivo como: $\tilde{a}_{i+1} = \pi_{\theta'}(s_{i+1}) + \epsilon$ ($\epsilon \sim \text{clip}(\mathcal{N}(0, \tilde{\sigma}), -c, c)$)
10 Calcular los objetivos: $y_i = r_i + \gamma \min_{k=1,2} Q_{w'_k}(s_{i+1}, \tilde{a}_{i+1})$
11 Actualizar los críticos: $w_k = \arg\ \min_{w_k} \frac{1}{N} \sum_i (y_i - Q_{w_k}(s_i, a_i))^2$
12 **if** t mód d **then**
13 Actualizar el actor según: $\nabla_\theta J(\pi_\theta) = \frac{1}{N} \sum_i \nabla_a Q_{w_1}(s_i, a_i) \nabla_{\pi_\theta} \pi_\theta(s_i)$
14 Actualizar las redes objetivo:

$$\begin{aligned} w'_i &= \tau w_i + (1 - \tau) w'_i \\ \theta' &= \tau \theta + (1 - \tau) \theta' \end{aligned}$$

15 **end**
16 **end**

Capítulo 5

Modelo de la planta y sistema de control propuesto

En este capítulo, se presentan los parámetros de los agricultores, los cultivos , los pozos y del suelo que caracterizan al sistema agrícola-hidrogeológico estudiado. También, se presentan los datos meteorológicos utilizados en las simulaciones dinámicas del sistema y el modelo utilizado para las mismas, destacando el modelo de crecimiento propuesto y la arquitectura utilizada para la implementación de dicho modelo. Finalmente, se presentan las configuraciones del modelo utilizadas para el aprendizaje reforzado, además de las arquitecturas de las redes neuronales a entrenar y las recompensas estudiadas.

5.1. Sistema agrícola-hidrogeológico

El sistema considerado consiste en diez agricultores de la comunidad José Painecura (coordenadas $38°32'11{,}1''S$-$73°30'12{,}9''O$), dedicados al cultivo de arvejas, papas y tomates, donde cada uno de estos representa un sistema hidrogeológico (ver capítulo 3). Los parámetros que caracterizan al sistema se presentan a continuación.

Agricultores y cultivos

Las áreas cultivadas por cada agricultor según el tipo de cultivo se obtuvieron de [1] y se presentan en el Cuadro 5.1. Por su parte, los parámetros utilizados para caracterizar cada cultivo se muestran en el Cuadro 5.2.

Cuadro 5.1: Área cultivada por cada agricultor según cultivo (en $[m^2]$).

Agricultor	**Papas**	**Arvejas**	**Tomates**
A_1	500	250	0
A_2	1000	500	250
A_3	0	750	1000
A_4	0	0	125
A_5	250	125	125
A_6	250	250	0
A_7	250	500	125
A_8	0	750	500
A_9	0	0	125
A_{10}	250	500	125

Cuadro 5.2: Parámetros de los cultivos considerados en la simulación.

Parámetros	**Papas**	**Arvejas**	**Tomates**
Día inicial del cultivo	213 (1 de agosto) [a]	197 (16 de julio) [b]	214 (2 de agosto) [c]
Altura máxima $[m]$ [d]	0,6	0,5	0,6
Profundidad máxima de raíces $[m]$ [d]	0,4	0,8	1,1
Días etapa 1 [d]	25	20	30
Días etapa 2 [d]	30	30	40
Días etapa 3 [d]	45	35	40
Días etapa 4 [d]	30	15	25
Kc_1 [d]	0,4	0,5	0,6
Kc_2 [d]	1,15	1,15	1,15
Kc_3 [d]	0,75	1,1	0,8
p $(p = RAW/TAW)$ [d]	0,35	0,35	0,4
Ky_1 [e]	0,45	0,2	0,4
Ky_2 [e]	0,8	0,9	1,1
Ky_3 [e]	0,7	0,7	0,8
Ky_4 [e]	0,2	0,2	0,3
CN [f]	70	63	65

[a] Fuente: Inostroza, 2009 [33]

[b] Fuente: Mera *et al.*, 2015 [34]

[c] Fuente: ISF-Chile [35]

[d] Fuente: Allen *et al.*, 1998 [2]

[e] Fuente: Doorenbos & Kassam, 1979 [25]

[f] Fuente: Jensen & Allen, 2016 [24]

Pozos

Respecto a los pozos (uno por agricultor), se los consideró con un radio de 0,5 $[m]$ y una profundidad de 1,5 $[m]$ de acuerdo a lo expuesto en [36]. Las coordenadas geográficas de estos se obtuvieron de [37] y se presentan en el Cuadro 5.3. La Figura 5.1 grafica su distribución física dentro de la comunidad.

Cuadro 5.3: Coordenadas geográficas de los pozos.

Pozo	**Latitud**	**Longitud**
Pozo 1	$38°31'41{,}37''S$	$73°30'30{,}57''O$
Pozo 2	$38°31'59{,}98''S$	$73°30'30{,}81''O$
Pozo 3	$38°32'9{,}57''S$	$73°30'14{,}13''O$
Pozo 4	$38°31'45{,}75''S$	$73°30'28{,}7''O$
Pozo 5	$38°31'26{,}41''S$	$73°30'28{,}2''O$
Pozo 6	$38°31'13{,}7''S$	$73°30'0{,}63''O$
Pozo 7	$38°31'32{,}54''S$	$73°29'52{,}6''O$
Pozo 8	$38°31'42{,}7''S$	$73°30'2{,}67''O$
Pozo 9	$38°31'20{,}06''S$	$73°30'18{,}05''O$
Pozo 10	$38°31'24{,}24''S$	$73°30'7{,}2''O$

Figura 5.1: Localización de los pozos dentro de la comunidad José Painecura.

Parámetros del suelo

De acuerdo con lo mencionado en [9] y [38], el tipo de suelo que caracteriza el territorio donde se emplaza la comunidad es franco arcillo limoso (FAL), cuyos parámetros se resumen en el Cuadro 5.4 y se obtuvieron de [39].

Cuadro 5.4: Parámetros del suelo.

Parámetro	**Valor**
Conductividad hidráulica $[m/día]$	0,017
Almacenamiento específico $[m^{-1}]$	$7 \cdot 10^{-4}$
Rendimiento específico	0,0273
Espesor del acuífero $[m]$	10
θ_{fc}	0.35
θ_{wp}	0.22

5.2. Datos meteorológicos

Los datos meteorológicos utilizados para simular la dinámica del modelo implementado son la temperatura, la radiación solar y la velocidad del viento, obtenidos desde el Explorador Solar (`https://solar.minenergia.cl/exploracion`), y las precipitaciones, obtenidas desde el Explorador Climático (`https://explorador.cr2.cl/`). Se consideró un periodo de 155 días, que va desde el 14 de julio (día 195 del año) al 16 de diciembre (día 350 del año) del 2015. La Figura 5.2 muestra las temperaturas mínima y máxima, la radiación solar media, la velocidad del viento media y las precipitaciones para los días considerados.

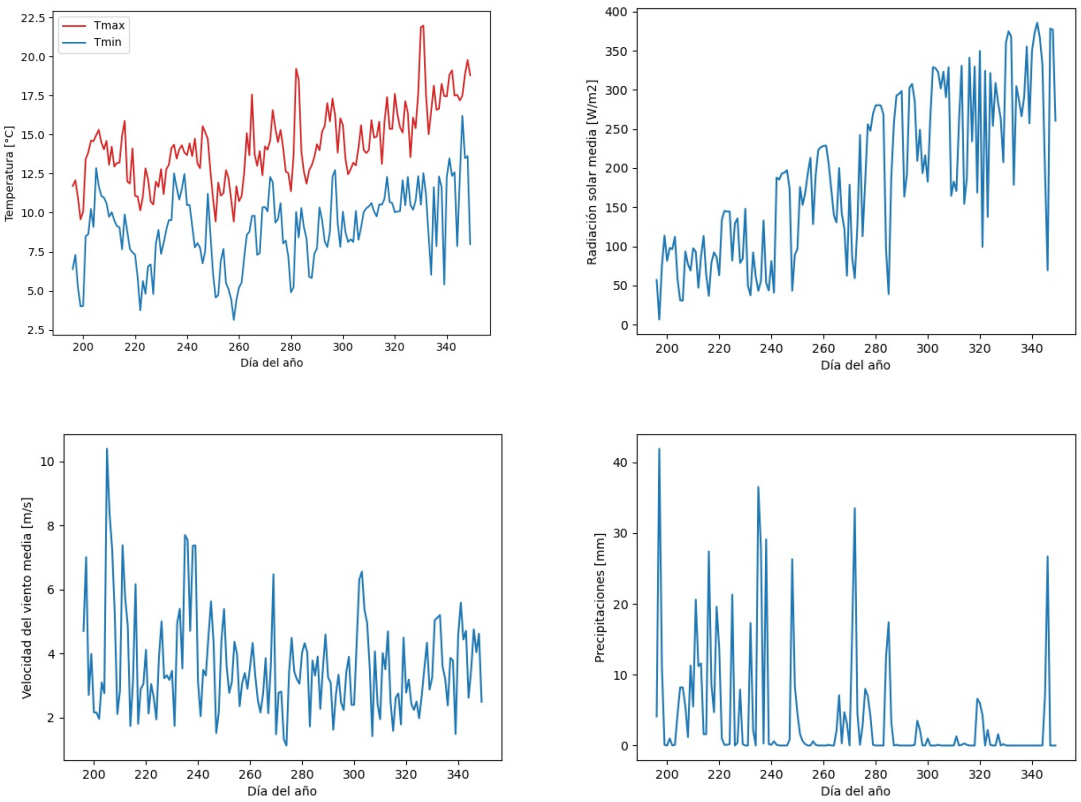


Figura 5.2: Datos meteorológicos utilizados en la simulación.

Es importante mencionar aquí, que la temperatura, la radiación solar y la velocidad del viento son necesarias para la determinación de la evapotranspiración de referencia ET_o, mientras que las precipitaciones se incorporan en el balance hídrico (ecuación 3.8).

5.3. Modelo del sistema agrícola-hidrogeológico

El modelo del sistema agrícola-hidrogeológico, es uno de tipo discreto con tiempo de muestreo diario, el cual utiliza principalmente tres ecuaciones para determinar la dinámica del sistema: el balance hídrico (ecuación 3.8), el rendimiento relativo (ecuación 3.22) diario de cada cultivo y el descenso de agua en cada pozo (ecuación 3.4). Dentro de la arquitectura de aprendizaje reforzado corresponde al entorno con el cual interactua el agente (controlador).

La entrada del modelo (acción), que también es la variable controlada, es el vector de riegos aplicados a los cultivos del sistema. Aquí, se asume que las tasas de bombeo durante el día son constantes y que el volumen de agua bombeado desde cada pozo durante un día, es igual a la suma de los volúmenes de agua suministrados a los cultivos del agricultor correspondiente. Este supuesto, deriva en que se considera un modelo sin pérdidas de agua en el sistema de cañerías y con eficiencia de 100 % del sistema de riego. Además, dado que diariamente el volumen de agua que ingresa desde los pozos es igual al que egresa hacia los cultivos, se ignora la participación de los estanques de acumulación.

Las variables de estado que se utilizaron como entrada de los controladores (agentes) fueron: los descensos radiculares D_r, los niveles RAW, los niveles de descenso de los pozos d, las temperaturas mínima y máxima, la radiación solar media, la velocidad del viento media y las precipitaciones. Es importante notar que si bien las variables meteorológicas corresponden más bien a perturbaciones del sistema y que no se utilizaron modelos para simular su evolución, dentro del esquema de aprendizaje reforzado, todo lo que se encuentre fuera del agente es parte del entorno, por lo que en este sentido, las variables meteorológicas son parte del entorno, es decir, el sistema mismo. También, se debe tener en consideración que para determinar la acción de control de determinado día, se utilizaron las variables de estado del día inmediatamente anterior, de acuerdo a la propiedad de Markov que caracteriza a los MDP (marco que engloba al aprendizaje reforzado), donde la probablidad de pasar a un cierto estado depende únicamente del estado actual [28].

La salida del modelo se definió como el vector de rendimientos relativos Y_r de los cultivos. Acá, es importante mencionar que no se lo consideró como entrada del controlador, ya que Y_r queda definido completamente por D_r y RAW (ver ecuaciones 3.16, 3.17 y 3.21), por lo que su incorporación sería redundante. Por otro lado, la ecuación 3.22, que entrega el rendimiento relativo al final del periodo productivo, permite formular la estimación del rendimiento relativo en cada día del proceso como:

$$Y_{r,k+1} = Y_{r,k} \left(1 - K_{y,i} \left(1 - K_{s,k+1}\right)\right)^{\frac{1}{T_i}} , \tag{5.1}$$

con $Y_{r,0} = 1$.

Dado que la metodología presentada en [2] no ofrece un modelo de crecimiento del follaje ni de las raíces, en [1] proponen un modelo crecimiento lineal durante las etapas 1 y 2, sin embargo, dicho modelo de desarrollo no considera los efectos del estrés hídrico sobre los cultivos. Luego, dado que no se logró ecntrar un modelo de crecimiento que se ajustara a los datos con los que se contaba, se propone el modelo de crecimiento expuesto a continuación.

Modelo de crecimiento

Utilizando como motivación la Figura 3.5, se propone un modelo de crecimiento lineal durante cada etapa cuando no existe estrés hídrico, pero que incluye un factor de penalización de la tasa de crecimiento diario en caso contrario. Esta formulación también se utilizó para el crecimiento de las raíces.

Ahora, para cada etapa de crecimiento i, se define la tasa de crecimiento Δh_i^* ideal (sin estrés hídrico) como:

$$\Delta h_i^* = \frac{h_{i,end} - h_{i-1,end}}{T_i}, \tag{5.2}$$

con T_i la duración en días de la i-ésima etapa y $h_{i,end}$ la altura ideal al final de la i-ésima etapa ($h_{0,end} = 0$) definida como:

$$h_{i,end} = \frac{K_{c,i}}{K_{c,3}} h_{max}, \tag{5.3}$$

donde h_{max} es la altura máxima del cultivo y particularmente para la etapa dos se fija $h_{2,end} = h_{3,end}$, ya que no existen valores tabulados de $K_{c,2}$.

Finalmente, para dar cuenta del impacto del estrés hídrico en el desarrollo del cultivo, la tasa de crecimiento se ajusta por un factor $\alpha = \frac{TAW - D_r}{TAW}$ (etapas 1,2 y 3) y $\alpha = 2 - \frac{TAW - D_r}{TAW}$ (etapa 4) cuando $K_s < 1$. Con esto, se logra que durante las primeras etapas de desarrollo, se reduzca la tasa de crecimiento si hay estrés y en la etapa final de marchitez, el estrés incrementa el decrecimiento.

Gym

Para ejecutar los entrenamientos de aprendizaje reforzado, el modelo se adaptó a la estructura de un entorno Gym (ver apéndice E), dentro del cual se definieron intervalos operacionales para las variables de estado (estado del entorno observado por el agente) y las de control (acción ejecutada por el agente sobre el entorno), en congruencia con las ecuaciones dinámicas y los datos meteorológicos. Con esto se tuvo que:

$$\begin{aligned}
D_r &\in [0, TAW_{max}] && (5.4)\\
RAW &\in [0, TAW_{max}] && (5.5)\\
d &\in [0, d_{max}] && (5.6)\\
T_{max} &\in [0, 100] && (5.7)\\
T_{min} &\in [0, 100] && (5.8)\\
R_s &\in [0, 1000] && (5.9)\\
u_z &\in [0, 10] && (5.10)\\
P &\in [0, 100] && (5.11)
\end{aligned}$$

con TAW_{max} definido según la ecuación 3.13 para $Z_r = Z_{max}$.

Respecto a la variable de control I, se limitó al intervalo $[0, 100]$, tomando en consideración los riegos máximos determinados por el método tradicional (ver capítulo 6), los que alcanzaban valores cercanos a los 70 $[mm]$. Sin embargo, siguiendo las recomendaciones de *Stable Baselines*, la acción (variable controlada) se limitó al intervalo $[-1, 1]$, de forma tal que el agente etrega un valor dentro de este último intervalo, el cual se debe proyectar a $[0, 100]$ como:

$$I_{real} = 50(I_{agente} + 1), \tag{5.12}$$

donde $I_{real} \in [0, 100]$ es el riego aplicado realmente e $I_{agente} \in [-1, 1]$ es la acción determinada por el agente.

Cabe destacar que *Stable Baselines* sí es capaz de operar con el intervalo $[0, 100]$ para la acción de control y se cree que la recomendación hace referencia al tipo de funciones de activación utilizadas.

5.4. Esquema de control

Considerando lo expuesto anteriormente, el esquema de control propuesto sigue la estructura presentada en la Figura 5.3, donde se destacan los bloques correspondientes a la estructura de una estrategia de control retroalimentado y los elementos propios de la estructura de aprendizaje reforzado.

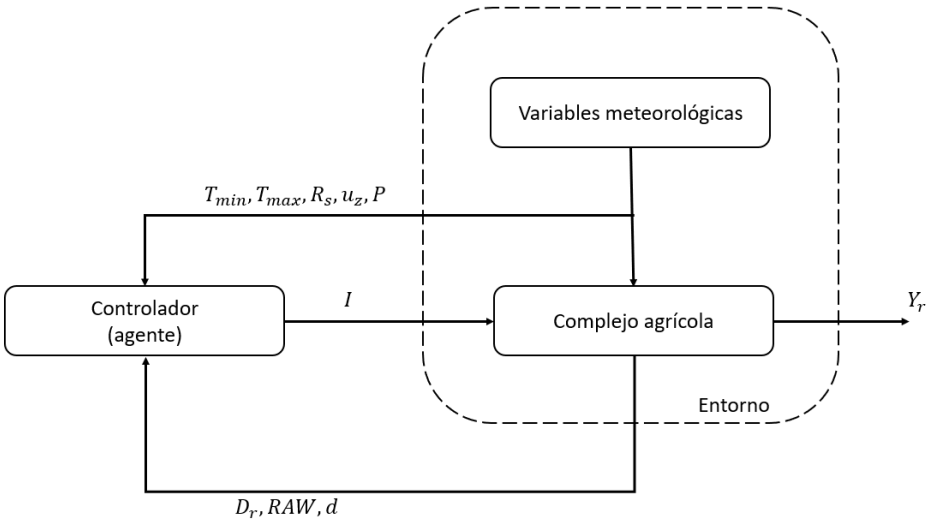


Figura 5.3: Esquema de control propuesto.

5.5. Aprendizaje reforzado

Dado que el sistema en consideración opera con variables continuas, se utilizaron los algoritmos de aprendizaje reforzado PPO, DDPG y TD3, los cuales están diseñados para trabajar con este tipo de entornos.

Para cada uno de estos algoritmos se evaluaron cuatro escenarios, definidos según la inclusión/exclusión de los pozos y la cantidad de cultivos controlada por un agente, teniéndose los siguientes casos:

- **Caso 1:** Modelo con pozos y un agente para los 22 cultivos
- **Caso 2:** Modelo con pozos y un agente por tipo de cultivo (3 agentes)
- **Caso 3:** Modelo sin pozos y un agente para los 22 cultivos
- **Caso 4:** Modelo sin pozos y un agente por tipo de cultivo (3 agentes)

Para el primer caso, las ANN del actor y del crítico eran del tipo *fully connected*, con dos capas ocultas de 400 neuronas y función de activación tanh(). El vector de estado estuvo compuesto por los descensos radiculares D_r, los niveles RAW, los niveles de descenso de los pozos d, las temperaturas mínima y máxima, la radiación solar media, la velocidad del viento media y las precipitaciones (59 variables). Se consideró un entrenamiento de 10^7 pasos.

Para el segundo caso, las ANN del actor y del crítico eran del tipo *fully connected*, con dos capas ocultas de 64 neuronas y función de activación tanh(). El vector de estado estuvo compuesto por el descenso radicular D_r del cultivo en consideración, el nivel RAW, las temperaturas mínima y máxima, la radiación solar media, la velocidad del viento media y las precipitaciones (7 variables). Notar que en este caso no se consideraron los pozos durante el entrenamiento, pero sí en la simulación del sistema completo. Se consideró un entrenamiento

de 10^6 pasos.

Para el tercer caso, las ANN del actor y del crítico eran del tipo *fully connected*, con dos capas ocultas de 400 neuronas y función de activación tanh(). El vector de estado estuvo compuesto por los descensos radiculares D_r, los niveles RAW, las temperaturas mínima y máxima, la radiación solar media, la velocidad del viento media y las precipitaciones (49 variables). Se consideró un entrenamiento de 10^7 pasos.

Finalmente, para el cuarto caso, las ANN del actor y del crítico eran del tipo *fully connected*, con dos capas ocultas de 64 neuronas y función de activación tanh(). El vector de estado estuvo compuesto por el descenso radicular D_r del cultivo en consideración, el nivel RAW, las temperaturas mínima y máxima, la radiación solar media, la velocidad del viento media y las precipitaciones (7 variables). Notar que el entrenamiento en este caso es el mismo que en el segundo. Se consideró un entrenamiento de 10^6 pasos.

Los parámetros de entrenamiento para los algoritmos PPO, DDPG y TD3 se consideraron con sus valores por defecto en *Stable Baselines* 3, que fue el paquete de Python utilizado para desarrollar los entrenamientos de aprendizaje reforzado y que cuenta con implementaciones validadas de los algoritmos evaluados.

Recompensas

Se realizaron pruebas con tres recompensas distintas, las cuales se definieron de forma tal que su maximización correspondiera a un bajo consumo de agua y a un alto rendimiento. Con esto, las recompensas evaluadas fueron:

$$r_1 = \frac{1}{R_{max}} \sum_{i=1}^{N_{cult}} c_1 Y_{r,i}^2 + \left(1 - \frac{I_i}{I_{max}}\right)^2, \tag{5.13}$$

$$r_2 = \begin{cases} \frac{1}{R_{max}} \left(c_1 Y_r^2 + \sum_{i=1}^{N_{cult}} \left(1 - \frac{I_i}{I_{max}}\right)^2 \right) & \text{día final} \\ \frac{1}{R_{max}} \sum_{i=1}^{N_{cult}} \left(1 - \frac{I_i}{I_{max}}\right)^2 & \text{demás días,} \end{cases} \tag{5.14}$$

$$r_3 = \begin{cases} \frac{1}{R_{max}} Y_r \sum_{i=1}^{N_{cult}} \left(1 - \frac{I_i}{I_{max}}\right)^2 & \text{día final} \\ \frac{1}{R_{max}} \sum_{i=1}^{N_{cult}} \left(1 - \frac{I_i}{I_{max}}\right)^2 & \text{demás días,} \end{cases} \tag{5.15}$$

donde $R_{max} = 155 N_{cult}(c_1 + 1)$ es el retorno máximo que se puede obtener durante un episodio de 155 días sin escalamiento (esto equivale a obtener rendimiento relativo 1 con riego 0 todos los días), N_{cult} es la cantidad de cultivos ($N_{cult} = 22$ para el sistema completo y $N_{cult} = 1$ durante el entrenamiento de los agentes single) y c_1 es un parámetro ajustable que da cuenta de la relevancia asignada al rendimiento. Cabe mencionar que la división por R_{max} se realiza para que los retornos alcanzados con cualquiera de las tres recompensas quede dentro del intervalo [0,1] y con ello facilitar la comparación de resultados entre recompensas diferentes.

Es importante notar que r_2 y r_3 corresponden al tipo de recompensas denominadas *sparse*, ya que dependen fuertemente del resultado alcanzado al final del episodio. Si bien suele evitarse este tipo de recompensas, en general permiten una mayor diferenciación entre un buen y un mal resultado.

Exploración

Como se mencionó en el capítulo 4, un aspecto relevante al momento de entrenar un agente es la exploración. Si bien esta está asociada a probar una gran cantidad de acciones, lo cual se puede obtener mediante selección aleatoria o adición de ruido, también se ve favorecida al enfrentar al agente a escenarios diversos. Con esto en mente y para evitar sobreajustes de los parámetros de las ANN que deriven en pérdida de robustez, se consideró la adición de un término aleatorio a las variables meteorológicas. Esto se puede entender como la simulación ante condiciones climatológicas diferentes, lo que se traduce en un agente con una "visión" más amplia respecto al comportamiento del entorno con el cual interactúa.

Las componentes aleatorias adicionadas a la temperatura mínima, temperatura máxima, velocidad de viento media y radiación solar media siguen una distribución uniforme $\mathcal{U}(-2, 2)$. Por su parte, a las precipitaciones se les adiciona una componente aleatoria con distribución uniforme $\mathcal{U}(-10, 10)$.

Es importante destacar que esta operación se efectuó sólo durante la etapa de entrenamiento y que los resultados presentados en el capítulo 6 consideran los datos meteorológicos en bruto.

Capítulo 6

Resultados

En este capítulo se presentan los resultados obtenidos con las diversas estrategias de control evaluadas, donde se hace distinción entre el modelo sin pozos y el modelo con pozos. Para cada uno de estos casos, se destaca el algoritmo de aprendizaje reforzado que mostró el mejor desempeño.

Los desempeños de los agentes entrenados mediante aprendizaje reforzado son comparados con otras cuatro estrategias: Sin riego, método tradicional, balance diario y PSO.

La primera, simplemente hace referencia a que se considera $I = 0$ para todos los cultivos durante todos los días. Esto permite tener una noción de cuales son las cotas inferiores reales de los rendimientos relativos de los cultivos y también, permite evaluar la pertinencia de la recompensa considerada.

El método tradicional y el balance diario, están basados en la ecuación de balance hídrico (ecuación 3.8), donde el primero corresponde a la recomendación de [2] y aplica riegos cuando se cumple que $D_r = RAW$, en cuyo caso la irrigación aplicada es:

$$I = D_r + RO + ET_p - P, \tag{6.1}$$

lo que deriva en $D_r = 0$, es decir, se riega hasta la saturación y se deja descender el nivel de agua justo antes de que ocurra estrés hídrico.

El segundo, en principio riega todos los días y la irrigación se determina como:

$$I = D_r + RO + ET_p - P - RAW, \tag{6.2}$$

lo que se traduce en regar solamente cuando $D_r > RAW$. En el caso de que las precipitaciones sean nulas, este método se encarga de mantener el descenso $D_r = RAW$, evitando de esta forma el estrés hídrico.

Por otro lado, PSO hace referencia a un controlador predictivo (MPC) cuya optimización se realiza mediante optimización por enjambre de partículas (ver apéndices C y D). El horizonte de predicción del MPC fue de 7 días, la función de costos a minimizar fue el opuesto del retorno de los 7 días, se consideraron 50 partículas y 500 iteraciones. También, se incluyeron como mejor solución global inicial los riegos determinados por el método tradicional cuando $P_{scale} = 0$ y los riegos determinados por el balance diario cuando $P_{scale} = 1$, de acuerdo a los resultados de los Cuadros 6.1 y 6.10. Por último, este controlador se evaluó sólo con la recompensa 1, ya que las recompensas 2 y 3 por su formulación *sparse*, llevarían siempre a riegos nulos.

Finalmente, las simulaciones consideran dos escenarios de precipitaciones determinados por el parámetro P_{scale}, el cual corresponde a un escalamiento de los datos de lluvias.

6.1. Simulación sin pozos

La simulación sin pozos corresponde a los Casos 3 y 4 (ver sección 5.5), donde se ignora la dinámica de los pozos y el acuífero, no existiendo las limitaciones en el riego debido a pozos secos, es decir, el hecho que desde un pozo seco no se puede extraer agua (riego nulo) no cuenta y en ese sentido se puede entender al acuífero como un contenedor infinito de agua.

6.1.1. Recompensas totales

Las recompensas totales (retornos) obtenidas para los distintos controladores evaluados se muestran en los Cuadros 6.1-6.3, donde se destaca el efecto diferenciador que tiene el parámetro c_1 entre una buena estrategia de riego en comparación con una mala, lo que en teoría facilita la distinción para el agente entrenado.

En particular, se observa que al incrementar c_1, la diferencia entre el esquema de no regar y los métodos tradicional y balance diario, también aumenta. Este efecto se hace más patente cuando se considera la ausencia total de precipitaciones ($P_{scale} = 0$), siendo esta una de las razones por la cual se decidió simular dicho escenario, ya que en éste, el rendimiento de los cultivos depende completamente del agua regada.

El principal argumento que justifica la diferenciación al aumentar c_1, es que su incremento implica una mayor relevancia del rendimiento relativo respecto a la economización de agua y en efecto, la estrategia de no regar siempre deriva en rendimientos inferiores a los de regar, cualquiera sea la política de riego considerada (ver Cuadros 6.7-6.9 y 6.16-6.18).

Ahora, respecto a las simulaciones con $P_{scale} = 1$, la distinción entre no regar y regar adecuadamente disminuye, ya que las precipitaciones favorecen el desarrollo de los cultivos, entregando un rendimiento relativo mínimo elevado (ver Cuadros 6.7-6.9). Luego, el no regar satisface ampliamente uno de los objetivos considerados, entiéndase, la economización de agua, lo que sumado a buenos rendimientos se traduce en un retoro elevado.

Recompensa 1				
Método	$P_{scale} = 1$		$P_{scale} = 0$	
	$c_1 = 1$	$c_1 = 10$	$c_1 = 1$	$c_1 = 10$
Sin riego	0,8505	0,7282	0,6571	0,3766
Método tradicional	0,8816	0,7985	0,8731	0,7965
Balance diario	0,8824	0,7985	0,87	0,7943
PSO	0,8577	0,7962	0,7231	0,7682
PPO	0,8519	0,7531	0,6826	0,5891
PPO single	0,8727	0,7458	0,7667	0,4786
DDPG	0,7545	0,7311	0,6145	0,4949
DDPG single	0,8505	0,7291	0,6571	0,7291
TD3	0,819	0,7419	0,6397	0,5775
TD3 single	0,8505	0,7291	0,6571	0,7291

Cuadro 6.1: Recompensas totales obtenidas para el modelo sin pozos con la recompensa 1.

Recompensa 2						
Método	$P_{scale} = 1$			$P_{scale} = 0$		
	$c_1 = 1$	$c_1 = 10$	$c_1 = 50$	$c_1 = 1$	$c_1 = 10$	$c_1 = 50$
Sin riego	0,9782	0,9605	0,9574	0,5262	0,1386	0,071
Método tradicional	0,9916	0,9985	0,9997	0,9827	0,9957	0,998
Balance diario	0,9925	0,9986	0,9997	0,9787	0,9923	0,9947
PPO	0,975	0,9657	0,9885	0,5891	0,7504	0,7252
PPO single	0,9769	0,9572	0,9965	0,5189	0,9618	0,993
TD3	0,899	0,9412	0,976	0,5506	0,4225	0,653
TD3 single	0,9782	0,9373	0,9574	0,5262	0,9419	0,71

Cuadro 6.2: Recompensas totales obtenidas para el modelo sin pozos con la recompensa 2.

De acuerdo a los resultados presentes en el Cuadro 6.3, la recompensa 3 es una mala recompensa, ya que la decisión de nunca regar siempre entrega un retorno mayor que los demás métodos, es decir, se trata de una recompensa que apunta a una economización excesiva del recurso hídrico, a expensas del rendimiento de los cultivos. Esto se traduce en que durante el entrenamiento, al tratar de maximizar el retorno esperado por los agentes, tenderán hacia la solución de no regar, independientemente de las precipitaciones.

Recompensa 3		
Método	$P_{scale} = 1$	$P_{scale} = 0$
Sin riego	0,9987	0,9937
Método tradicional	0,9832	0,9667
Balance diario	0,985	0,9618
PPO	0,9714	0,9438
PPO single	0,9647	0,9399
TD3	0,9987	0,9937
TD3 single	0,9987	0,9937

Cuadro 6.3: Recompensas totales obtenidas para el modelo sin pozos con la recompensa 3.

6.1.2. Riegos totales

Los volúmenes de riego totales se presentan en los Cuadros 6.4-6.6, donde se aprecia que en general los sistemas que consumen menores recursos hídricos son el no regar, el método tradicional y el balance diario, mientras que entre los algoritmos PPO, DDPG y TD3, el primero destaca como el más economizador, especialmente en su variante single.

Como es de esperar, el esquema determinado por la ausencia total de precipitaciones induce a un mayor consumo de agua en la mayoría de los casos evaluados, lo que delata el desarrollo de una sensibilidad por parte del agente frente a escenarios pluviométricos diferentes.

Por otro lado, se aprecia una gran variedad de volúmenes de agua aplicados, donde la tendencia es de aplicar más agua a medida que c_1 aumenta. La explicación para este fenómeno es análoga a lo referido anteriormente, en el sentido que al dar mayor relevancia al rendimiento de los cultivos, se presta menos atención al consumo de agua y en consecuencia se entrega mayor libertad para disponer de este recurso. Los

casos particulares donde esto no ocurre, se entienden como el resultado de un mal entrenamiento y requieren reajustar los parámetros de éste.

Recompensa 1				
Método	$P_{scale} = 1$		$P_{scale} = 0$	
	$c_1 = 1$	$c_1 = 10$	$c_1 = 1$	$c_1 = 10$
Sin riego	0	0	0	0
Método tradicional	1347,75	1347,75	2765,25	2765,25
Balance diario	944,5	944,5	2505,375	2505,375
PSO	209,735	947,0554	1019,5266	3125,2472
PPO	15071,8306	20091,5162	39200,7433	40161,9912
PPO single	1596,7979	8461,5189	3369,0411	1592,4757
DDPG	33350,0001	27187,5	33350,0069	27187,4992
DDPG single	0	100812,5	0	100812,5
TD3	6849,8468	51986,889	6687,5034	51936,5378
TD3 single	0,0266	100812,5	0	100811,7848

Cuadro 6.4: Riegos totales en $[m^3]$ para el modelo sin pozos con la recompensa 1.

Recompensa 2						
Método	$P_{scale} = 1$			$P_{scale} = 0$		
	$c_1 = 1$	$c_1 = 10$	$c_1 = 50$	$c_1 = 1$	$c_1 = 10$	$c_1 = 50$
Sin riego	0	0	0	0	0	0
Método tradicional	1347,75	1347,75	1347,75	2765,25	2765,25	2765,25
Balance diario	944,5	944,5	944,5	2505,375	2505,375	2505,375
PPO	17518,3409	66177,9204	76424,0429	22032,4786	69734,016	77542,4946
PPO single	195,7313	42254,5993	14321,429	1446,1243	36487,9649	16445,3159
TD3	36812,5	40687,5002	58125	36812,5	40687,5002	58125
TD3 single	0	87915,2783	0	0	81554,3652	0

Cuadro 6.5: Riegos totales en $[m^3]$ para el modelo sin pozos con la recompensa 2.

Recompensa 3		
Método	$P_{scale} = 1$	$P_{scale} = 0$
Sin riego	0	0
Método tradicional	1347,75	2765,25
Balance diario	944,5	2505,375
PPO	7217,4996	7549,8436
PPO single	2395,1591	4752,3196
TD3	37,5	37,4906
TD3 single	0	0

Cuadro 6.6: Riegos totales en $[m^3]$ para el modelo sin pozos con la recompensa 3.

6.1.3. Rendimientos

Los rendimientos relativos se presentan en los Cuadros 6.7-6.9, donde se observa que las técnicas de método tradicional y balance diario ofrecen los mejores rendimientos, seguidos por PPO single con las recompensas 2 y 3 (Cuadros 6.8 y 6.9), esto tomando en consideración el caso $P_{scale} = 0$.

Como se mencionó anteriormente, la presencia de lluvias en el esquema con $P_{scale} = 1$ establece un alto rendimiento mínimo, por lo que cualquier cantidad de agua extra que se suministre por medio del riego, contribuirá a aumentar dicho rendimiento. Esto explica el por qué de los altos rendimientos obtenidos bajo este esquema, para todas las estrategias de irrigación evaluadas. En este sentido, el esquema con $P_{scale} = 1$ no permite determinar adecuadamente la calidad de una estrategia de irrigación.

Recompensa 1				
Método	$P_{scale} = 1$		$P_{scale} = 0$	
	$c_1 = 1$	$c_1 = 10$	$c_1 = 1$	$c_1 = 10$
Sin riego	0,7679	0,7679	0,129	0,129
Método tradicional	1	1	0,999	0,999
Balance diario	0,9992	0,9992	0,9963	0,9963
PSO	0,8014	0,9864	0,3995	0,9492
PPO	0,8803	0,3792	0,9111	0,5864
PPO single	0,9608	0,8239	0,6936	0,2394
DDPG	0,86	0,8394	0,3686	0,4042
DDPG single	0,7679	1	0,129	1
TD3	0,7933	0,89	0,2048	0,5211
TD3 single	0,7679	1	0,129	1

Cuadro 6.7: Rendimientos relativos medios obtenidos para el modelo sin pozos con la recompensa 1.

Recompensa 2						
Método	$P_{scale} = 1$			$P_{scale} = 0$		
	$c_1 = 1$	$c_1 = 10$	$c_1 = 50$	$c_1 = 1$	$c_1 = 10$	$c_1 = 50$
Sin riego	0,7679	0,7679	0,7679	0,129	0,129	0,129
Método tradicional	1	1	1	0,999	0,999	0,999
Balance diario	0,9992	0,9992	0,9992	0,9963	0,9963	0,9963
PPO	0,796	0,884	0,9218	0,2886	0,7418	0,7374
PPO single	0,7679	1	0,9944	0,1482	1	0,9905
TD3	0,8016	0,8462	0,9177	0,3159	0,4428	0,6844
TD3 single	0,7679	1	0,7679	0,129	1	0,129

Cuadro 6.8: Rendimientos relativos medios obtenidos para el modelo sin pozos con la recompensa 2.

Recompensa 3		
Método	$P_{scale} = 1$	$P_{scale} = 0$
Sin riego	0,7679	0,129
Método tradicional	1	0,999
Balance diario	0,9992	0,9963
PPO	0,8698	0,4004
PPO single	0,9989	0,9521
TD3	0,7679	0,129
TD3 single	0,7679	0,129

Cuadro 6.9: Rendimientos relativos medios obtenidos para el modelo sin pozos con la recompensa 3.

Respecto a la influencia del parámetro c_1, como era de esperar, en la mayoría de los casos, los rendimientos fueron mayores al aumentar el valor de dicho parámetro, siguiendo el razonamiento expuesto anteriormente al respecto.

Por último, en cuanto a los esquemas $P_{scale} = 1$ y $P_{scale} = 0$, se hubiera esperado mayor similitud entre los rendimientos obtenidos, especialmente si se considera que siempre se puede regar el máximo, al no existir la limitante de los pozos. Sin embargo, al haber agentes que aprendieron a nunca regar o a regar cantidades insuficientes, se tienen las diferencias presentadas, definidas fundamentalmente por la cantidad de precipitaciones.

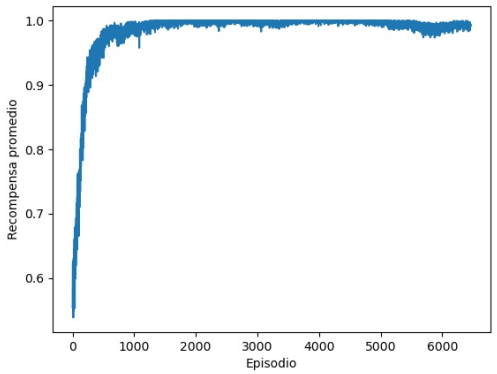
6.1.4. Mejor método: PPO single con la recompensa 3

De acuerdo a los resultados de los Cuadros 6.1-6.9, se considera que para el modelo del sistema sin pozos el agente que tuvo mejor desempeño fue el entrenado con PPO single y la recompensa 3. Las Figuras 6.1, 6.2 y 6.3 muestran las curvas de aprendizaje, los rendimientos con $P_{scale} = 1$ y los rendimientos con $P_{scale} = 0$ respectivamente para este agente.

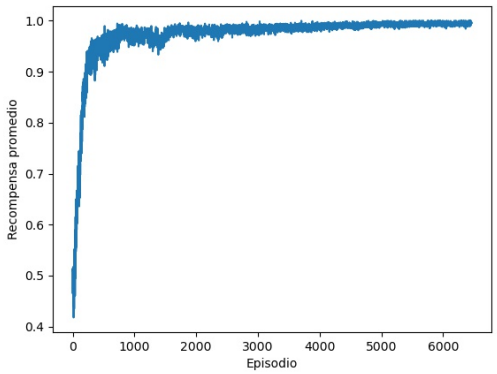
Se debe destacar que si bien PPO single con la recompensa 2 y $c_1 = 50$ presentó mejores rendimientos, estos se obtuvieron a expensas de un consumo significatimente mayor (aproximadamente 4,72 veces más, considerando ambos esquemas de lluvia), razón por la cual se optó por el agente presentado en esta sección.

Curvas de aprendizaje

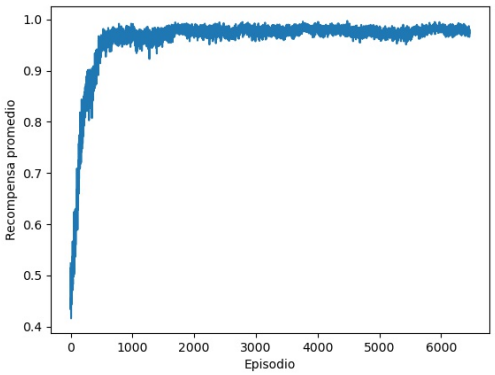
Las curvas de aprendizaje para cada cultivo demuestran una rápida convergencia a retornos cercanos al máximo teórico ($R_{max} = 1$) y poca varianza. Esto en general denota un entrenamiento adecuado, sin embargo, se debe tener en consideración que la recompensa 3 considerada aquí, no es una buena recompensa, por lo que en este caso, prolongar el entrenamiento podría derivar en la decisión de no regar nunca, ya que esta decisión es la que entrega mayor retorno (ver Cuadro 6.3).



(a) Arvejas.



(b) Papas.

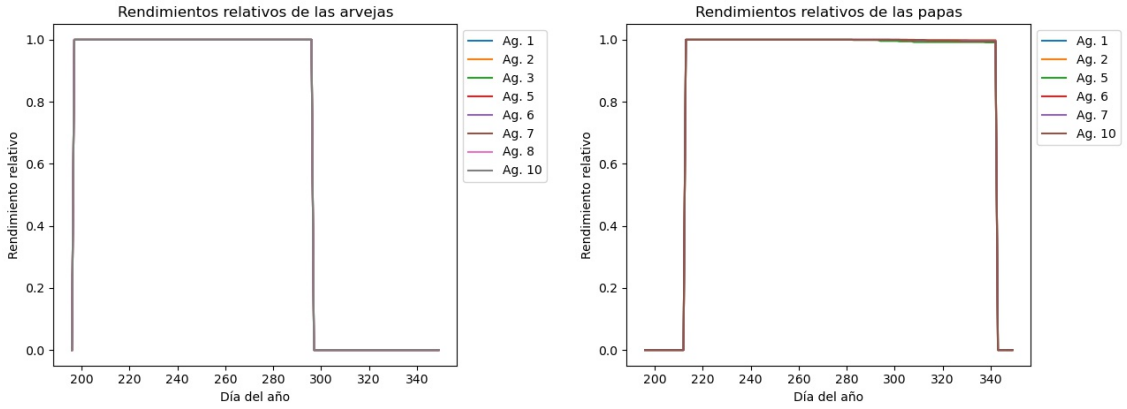


(c) Tomates.

Figura 6.1: Recompensa por episodio durante el entrenamiento para los agentes de arvejas, papas y tomates (agente PPO single con recompensa 3).

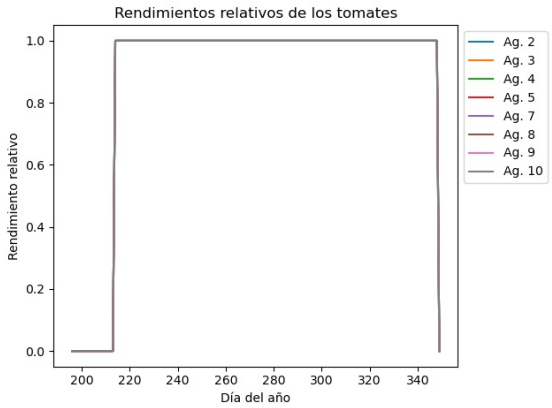
Rendimientos relativos con escala de precipitaciones $P_{scale} = 1$

Respecto a los rendimientos relativos, se aprecia que son iguales a 1 (rendimiento máximo) en el caso de las arvejas y los tomates, pero en el caso de las papas existen algunos levemente inferiores. Esto se debe al hecho de que las papas son un cultivo altamente sensible al estrés hídrico, por lo que se requieren déficits de agua menores para tener un impacto significativo en su rendimiento. En este sentido, el modelo logra asemejarse adecuadamente a la realidad.



(a) Arvejas.

(b) Papas.

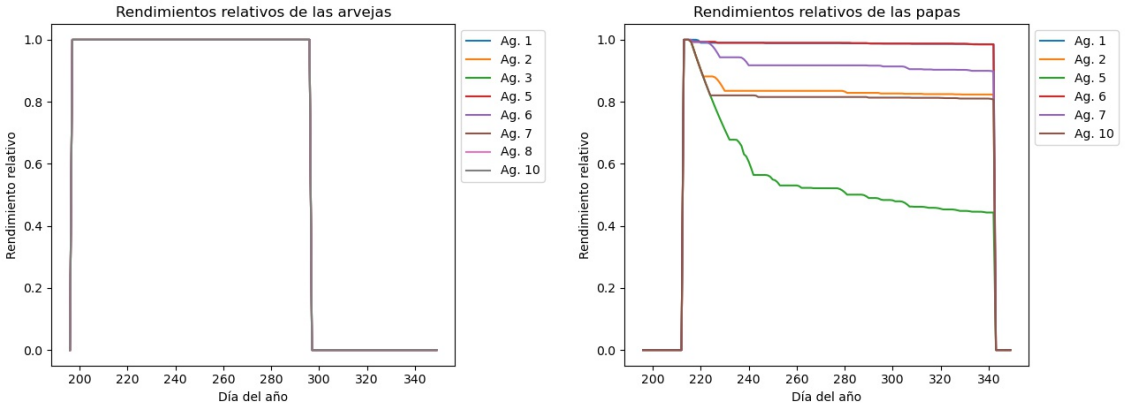


(c) Tomates.

Figura 6.2: Rendimientos relativos para los cultivos de arvejas, papas y tomates con $P_{scale} = 1$ (agente PPO single con recompensa 3).

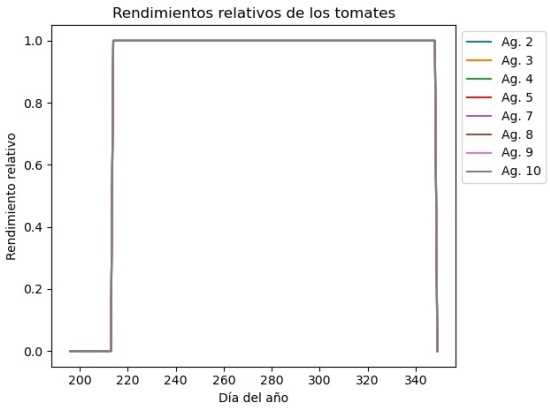
Rendimientos relativos con escala de precipitaciones $P_{scale} = 0$

En este caso, se hace más patente la sensibilidad de las papas al estrés hídrico, donde se ve particularmente perjudicado el rendimiento de papas del agricultor 5 ante el déficit de agua. Esta situación resulta llamativa, toda vez que el agente encargado de aplicar la acción (riego) a dicho cultivo, fue entrenado específicamente para este tipo de cultivo (configuración single). La única explicación para este fenómeno, considerando que los demás cultivos de papas no se ven tan afectados y que no existe el secado de pozos, es el carácter estocástico de PPO. En efecto, el agente de PPO aprende parámetros que definen una distribución de probabilidad sobre el espacio de acciones posibles, con lo cual no se puede garantizar la obtención de los mismos buenos resultados siempre, como tampoco la obtención de los mismos malos resultados siempre. Sin duda es este un punto débil en este tipo de algoritmos. Por lo tanto, la caída significativa del rendimiento de papas del agricultor 5 no es taxativa y tiene margen para mejorar.



(a) Arvejas.

(b) Papas.



(c) Tomates.

Figura 6.3: Rendimientos relativos Y_r para los cultivos de arvejas, papas y tomates con $P_{scale} = 0$ (agente PPO single con recompensa 3).

6.2. Simulación con pozos

La simulación con pozos corresponde a los Casos 1 y 2 (ver sección 5.5), donde la dinámica de los pozos y el acuífero ofrece límites al riego cuando hay pozos secos, es decir, el hecho que desde un pozo seco no se puede extraer agua se traduce en un riego nulo para los cultivos que dependen de él.

6.2.1. Recompensas totales

Las recompensas totales obtenidas para los distintos controladores evaluados se muestran en los Cuadros 6.10-6.12. En cuanto a estas, el análisis es similar al caso del modelo sin pozos, donde se observan altas recompensas en todos los casos cuando $P_{scale} = 1$, debido al aporte de las lluvias. En este mismo sentido, las distinciones entre un buen manejo del riego y uno deficiente se hacen más patentes cuando $P_{scale} = 0$, es decir, cuando el rendimiento depende completamente del riego. Esto también se observa al aumentar c_1.

En general, al comparar las recompensas del modelo sin pozos con las recompensas del modelo con pozos, se tiene que las primeras son mayores, lo que constituye un resultado esperado, ya que para una determinada secuencia de riegos que incurra en pozos secos, los riegos siguientes serán nulos para el modelo con pozos, lo que se traduce en pérdidas de rendimiento, lo que no ocurre para el modelo sin pozos, ya que, en este caso siempre hay disponibilidad de agua para alentar el desarrollo de los cultivos.

Recompensa 1				
Método	$P_{scale} = 1$		$P_{scale} = 0$	
	$c_1 = 1$	$c_1 = 10$	$c_1 = 1$	$c_1 = 10$
Sin riego	0,8505	0,7282	0,6571	0,3766
Método tradicional	0,8816	0,7985	0,8733	0,7962
Balance diario	0,8824	0,7985	0,8698	0,7943
PSO	0,8589	0,7975	0,73	0,7787
PPO	0,8481	0,7643	0,6793	0,467
PPO single	0,8723	0,7404	0,8517	0,4953
DDPG	0,7761	0,7343	0,6383	0,4911
DDPG single	0,8505	0,7331	0,6571	0,5358
TD3	0,7829	0,7304	0,65	0,441
TD3 single	0,8505	0,7331	0,6571	0,5358

Cuadro 6.10: Recompensas totales obtenidas para el modelo con pozos con la recompensa 1.

Recompensa 2						
Método	$P_{scale} = 1$			$P_{scale} = 0$		
	$c_1 = 1$	$c_1 = 10$	$c_1 = 50$	$c_1 = 1$	$c_1 = 10$	$c_1 = 50$
Sin riego	0,9782	0,9605	0,9574	0,5262	0,1386	0,071
Método tradicional	0,9916	0,9985	0,9997	0,9831	0,9958	0,998
Balance diario	0,9925	0,9986	0,9997	0,9787	0,9923	0,9947
PPO	0,9521	0,9609	0,969	0,6079	0,3011	0,3707
PPO single	0,9769	0,9579	0,9933	0,5189	0,5359	0,3468
TD3	0,8469	0,9415	0,9594	0,5365	0,3508	0,2371
TD3 single	0,9782	0,945	0,9573	0,5262	0,5131	0,071

Cuadro 6.11: Recompensas totales obtenidas para el modelo con pozos con la recompensa 2.

Recompensa 3		
Método	$P_{scale} = 1$	$P_{scale} = 0$
Sin riego	0,9987	0,9937
Método tradicional	0,9832	0,9675
Balance diario	0,985	0,9619
PPO	0,9412	0,9244
PPO single	0,9914	0,9891
TD3	0,8954	0,9024
TD3 single	0,9987	0,9937

Cuadro 6.12: Recompensas totales obtenidas para el modelo con pozos con la recompensa 3.

6.2.2. Riegos totales

Los volúmenes de riego totales para los distintos controladores evaluados se muestran en los Cuadros 6.13-6.15, donde nuevamente se destacan como las mejores estrategias el método tradicional y el balance diario, seguidos por PPO single. Igualmente, el esquema con $P_{scale} = 0$ induce a la utilización de volúmenes de agua mayores que con $P_{scale} = 1$.

Sin embargo, acá no se hace patente la tendencia de aumentar el riego junto con c_1. Esto se puede explicar por el hecho de que la aplicación de riegos elevados en etapas tempranas del desarrollo de los cultivos, conduce al secado de los pozos, luego de lo cual no es posible extraer más agua hasta que se recuperen, lo cual se dificulta si se tiene en consideración que el periodo del año correspondiente a las etapas finales de los cultivos, coincide con la primavera y el inicio del verano, estaciones caracterizadas por un descenso en las precipitaciones como se observa en la Figura 5.2.

Pese a lo anterior, no es posible definir una tendencia a la baja de los riegos para el modelo con pozos en comparación al modelo sin pozos, como se hubiera esperado. La incidencia de este fenómeno se manifiesta en casos puntuales.

Recompensa 1				
Método	$P_{scale} = 1$		$P_{scale} = 0$	
	$c_1 = 1$	$c_1 = 10$	$c_1 = 1$	$c_1 = 10$
Sin riego	0	0	0	0
Método tradicional	1347,75	1347,75	2652,75	2652,75
Balance diario	944,5	944,5	2491,875	2491,875
PSO	94,1752	787,5589	807,4834	2640,1907
PPO	14180,4917	34982,2863	11849,7715	51697,8836
PPO single	2290,0593	8575,7438	3267,4799	1572,6328
DDPG	18462,5	23537,5	12987,5	17850
DDPG single	0	43387,5	0	38975
TD3	21312,5	18875	7574,9999	10800,1771
TD3 single	0,0266	43387,5	0	38974,2848

Cuadro 6.13: Riegos totales en $[m^3]$ para el modelo con pozos con la recompensa 1.

Recompensa 2						
Método	$P_{scale} = 1$			$P_{scale} = 0$		
	$c_1 = 1$	$c_1 = 10$	$c_1 = 50$	$c_1 = 1$	$c_1 = 10$	$c_1 = 50$
Sin riego	0	0	0	0	0	0
Método tradicional	1347,75	1347,75	1347,75	2652,75	2652,75	2652,75
Balance diario	944,5	944,5	944,5	2491,875	2491,875	2491,875
PPO	19877,6182	74870,5013	83882,6401	22300,3159	72146,3788	94511,9184
PPO single	163,7202	23599,5942	9448,8991	1164,0926	14801,715	5854,6387
TD3	98812,5	25187,5	79437,5952	98812,5	25187,5	79438,6544
TD3 single	0	39751,3141	362,5	0	32535,7222	0

Cuadro 6.14: Riegos totales en $[m^3]$ para el modelo con pozos con la recompensa 2.

Recompensa 3		
Método	$P_{scale} = 1$	$P_{scale} = 0$
Sin riego	0	0
Método tradicional	1347,75	2652,75
Balance diario	944,5	2491,875
PPO	33250,163	35646,2869
PPO single	453,3194	510,8564
TD3	14250	14250,0006
TD3 single	0	0,0003

Cuadro 6.15: Riegos totales en $[m^3]$ para el modelo con pozos con la recompensa 3.

6.2.3. Rendimientos

Los rendimientos relativos se presentan en los Cuadros 6.16-6.18, donde se observa que las técnicas de método tradicional y balance diario ofrecen los mejores rendimientos, seguidos por PSO con la recompensa 1 y $c_1 = 10$, y PPO single con la recompensa 1 y $c_1 = 1$ (Cuadro 6.16), esto tomando en consideración el caso $P_{scale} = 0$.

Acá, también se tiene que para el caso $P_{scale} = 1$ los rendimientos son elevados debido al aporte de las precipitaciones y en general los rendimientos aumentan a mayor valor de c_1.

En comparación con los rendimientos obtenidos para el modelo sin pozos (Cuadros 6.7-6.9), la tendencia general es a la baja, lo que se debe al hecho ya mencionado de que pozos secos implican riegos nulos, por lo tanto, ante ciertas situaciones en que el modelo sin pozos logra inyectar un cierto volumen de agua a los cultivos, el modelo con pozos, particularmente si están secos, no lo puede hacer, lo que conduce a una caída en los rendimientos. Esta tendencia también se observa al comparar los escenarios $P_{scale} = 1$ y $P_{scale} = 0$, donde los rendimientos del segundo son, en general, significativamente inferiores a los del primero.

Recompensa 1				
Método	$P_{scale} = 1$		$P_{scale} = 0$	
	$c_1 = 1$	$c_1 = 10$	$c_1 = 1$	$c_1 = 10$
Sin riego	0,7679	0,7679	0,129	0,129
Método tradicional	1	1	0,9945	0,9945
Balance diario	0,9992	0,9992	0,9963	0,9963
PSO	0,8042	0,9935	0,3562	0,9644
PPO	0,8879	0,9206	0,3197	0,3187
PPO single	0,9955	0,8062	0,9534	0,2598
DDPG	0,8221	0,8607	0,2867	0,3511
DDPG single	0,7679	0,9079	0,129	0,5272
TD3	0,8078	0,8242	0,2363	0,2705
TD3 single	0,7679	0,9079	0,129	0,5272

Cuadro 6.16: Rendimientos relativos medios obtenidos para el modelo con pozos con la recompensa 1.

Recompensa 2						
Método	$P_{scale} = 1$			$P_{scale} = 0$		
	$c_1 = 1$	$c_1 = 10$	$c_1 = 50$	$c_1 = 1$	$c_1 = 10$	$c_1 = 50$
Sin riego	0,7679	0,7679	0,7679	0,129	0,129	0,129
Método tradicional	1	1	1	0,9945	0,9945	0,9945
Balance diario	0,9992	0,9992	0,9992	0,9963	0,9963	0,9963
PPO	0,8316	0,8945	0,8796	0,308	0,309	0,4188
PPO single	0,7679	0,9359	0,8377	0,1476	0,5571	0,3201
TD3	0,8686	0,8542	0,8285	0,2825	0,3606	0,2768
TD3 single	0,7679	0,9324	0,7679	0,129	0,5441	0,129

Cuadro 6.17: Rendimientos relativos medios obtenidos para el modelo con pozos con la recompensa 2.

Recompensa 3		
Método	$P_{scale} = 1$	$P_{scale} = 0$
Sin riego	0,7679	0,129
Método tradicional	1	0,9945
Balance diario	0,9992	0,9963
PPO	0,8745	0,3317
PPO single	0,7812	0,135
TD3	0,8034	0,2255
TD3 single	0,7679	0,129

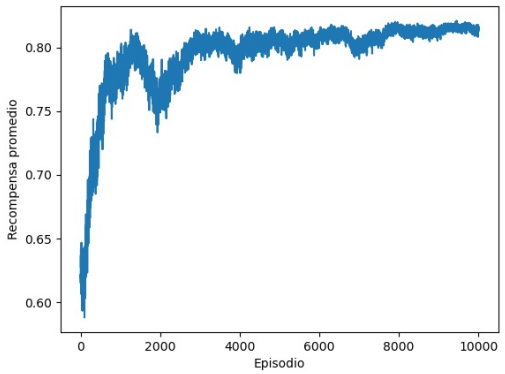
Cuadro 6.18: Rendimientos relativos medios obtenidos para el modelo con pozos con la recompensa 3.

6.2.4. PPO single con la recompensa 1 y $c_1 = 1$

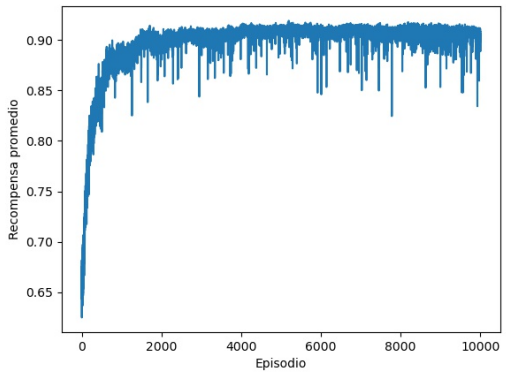
De acuerdo a los resultados de los Cuadros 6.10-6.18, se considera que para el modelo del sistema con pozos el agente que tuvo mejor desempeño fue el entrenado con PPO single y la recompensa 1 con $c_1 = 1$. Las Figuras 6.4, 6.5, 6.6, 6.7 y 6.8 muestran las curvas de aprendizaje, los rendimientos y descensos de los pozos con $P_{scale} = 1$, y los rendimientos y descensos de los pozos con $P_{scale} = 0$ respectivamente para este agente.

Curvas de aprendizaje

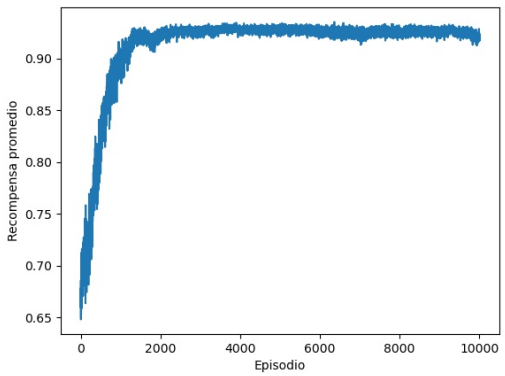
Las curvas de aprendizaje para cada cultivo, al igual que las presentadas anteriormente, demuestran una rápida convergencia, particularmente para las papas y los tomates. Sin embargo, respecto a las arvejas, la curva se muestra más oscilante, lo que se puede explicar en parte, por el hecho de que el periodo del cultivo de arvejas es menor, por lo que la exploración por episodio durante el entrenamiento se ve truncada tempranamente (en comparacación con los otros cultivos). También, el fenómeno de secado de pozos puede ocasionar inestabilidades que dificulten la búsqueda de soluciones óptimas, al agregar otro grado de complejidad al problema (en comparación al caso sin pozos). Esto último, también puede justificar la mayor varianza en la curva de aprendizaje de las papas, sobre todo recordando su alta sensibilidad al déficit de agua.



(a) Arvejas.



(b) Papas.

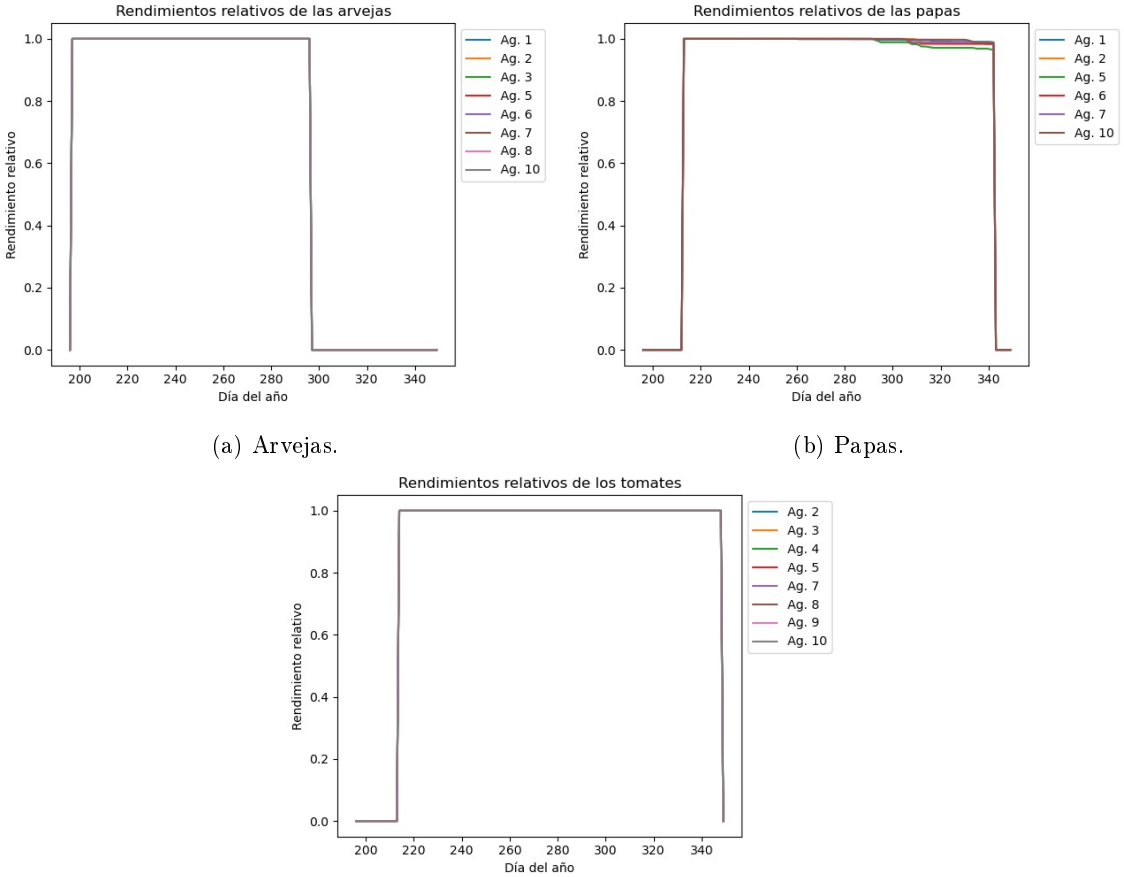


(c) Tomates.

Figura 6.4: Recompensa por episodio durante el entrenamiento para los agentes de arvejas, papas y tomates (agente PPO single con recompensa 1 y $c_1 = 1$).

Rendimientos relativos y descenso de pozos con $P_{scale} = 1$

En cuanto a los rendimientos relativos con $P_{scale} = 1$, al igual que para el modelo sin pozos, se aprecia que son iguales a 1 (rendimiento máximo) en el caso de las arvejas y los tomates, pero en el caso de las papas existen algunos levemente inferiores. La justificación para esto, es nuevamente el hecho de que las papas son un cultivo altamente sensible al estrés hídrico, por lo que se requieren déficits de agua menores para tener un impacto significativo en su rendimiento.



(a) Arvejas.

(b) Papas.

(c) Tomates.

Figura 6.5: Rendimientos relativos para los cultivos de arvejas, papas y tomates con $P_{scale} = 1$ (agente PPO single con recompensa 1 y $c_1 = 1$).

Ahora, respecto a los descensos en los pozos, se observa que para $P_{scale} = 1$ no se alcanza a secar ningún pozo, lo cual es un excelente resultado, ya que esto permite disponer de agua para otras necesidades humanas de los agricultores o, en el caso de extraer más agua, contar con una reserva hídrica para temporadas más calurosas y secas, donde se requiera aplicar mayores volúmenes de agua para suplir la ausencia de lluvias.

Es importante notar aquí, que la simulación entrega ascensos del agua en los pozos al nivel de suelo, lo cual resulta llamativo pero no totalmente irreal, si se considera que la comunidad José Painecura está ubicada en el sector costero, donde es posible encontar depósitos de agua a poca profundidad. De todas formas, plantea la interrogante respecto a las dimensiones consideradas para los pozos, obtenidos mediante encuestas a los habitantes de la comunidad de acuerdo con [36].

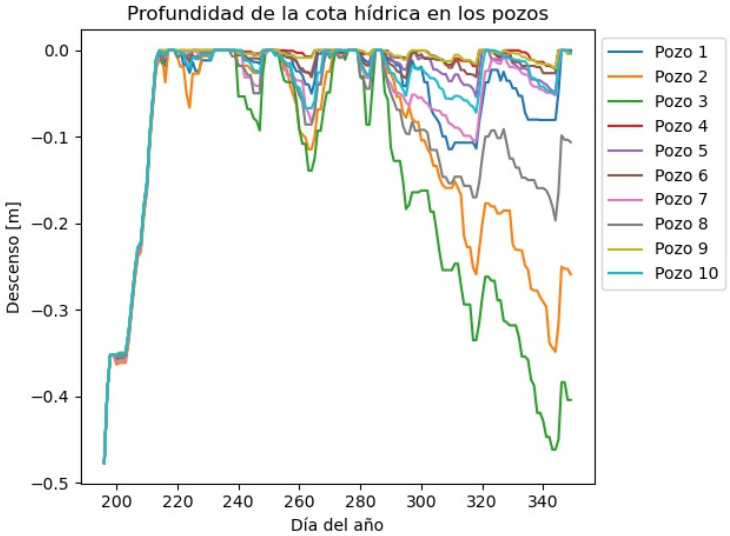
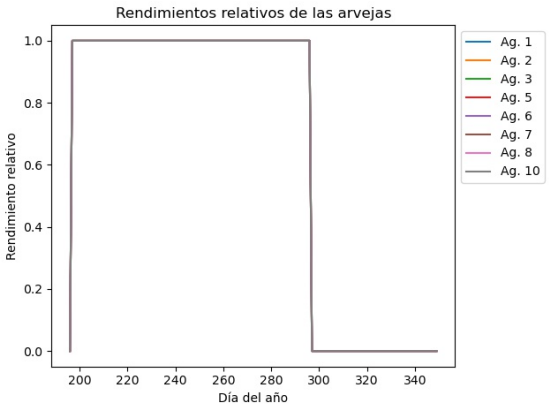


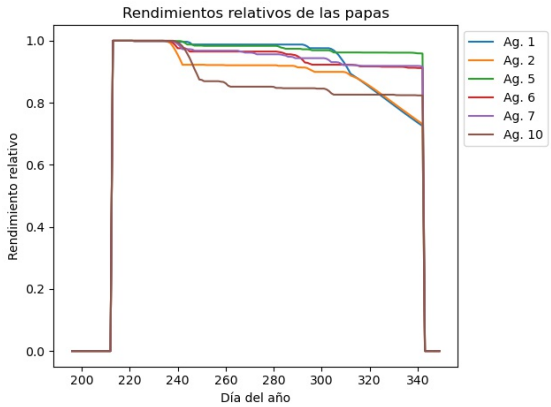
Figura 6.6: Descenso en el nivel de los pozos con $P_{scale} = 1$ (agente PPO single con recompensa 1 y $c_1 = 1$).

Rendimientos relativos y descenso de pozos con $P_{scale} = 0$

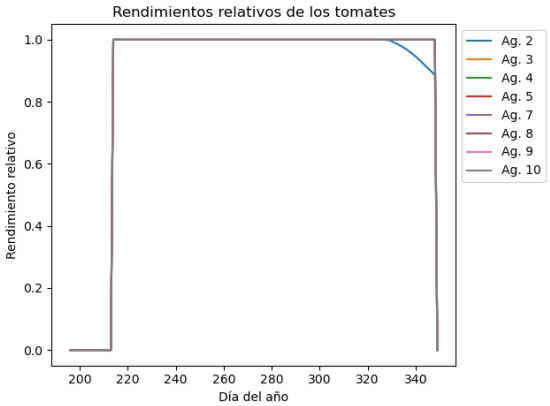
En cuanto a los rendimientos relativos con $P_{scale} = 0$, se observan mejores rendimientos para las papas que en el modelo sin pozos, lo que verífica el impacto de la estocasticidad del algoritmo PPO, es decir, las acciones aplicadas no son determinísticas y por lo tanto para las mismas observaciones existe la probabilidad de seguir otro curso de acción, derivando en resultados y recompensas distintas. Respecto a las arvejas, estas suelen presentar buenos rendimientos ya que su periodo de cultivo es el menor y está ubicado en un periodo del año caracterizado por altas precipitaciones (ver Figura 5.2). Por último, los tomates también presentan buenos rendimientos, donde el agricultor 2 obtiene la menor producción de estos debido a que el agente deja de regar hacia el final del periodo de cultivo (ver Figura 6.8 donde el nivel del pozo 2 se estanca y Figura F.8 donde el riego de tomates del agricultor 2 es nulo hacia el final). La justificación de esta última situación no es clara y da la impresión de que el agente considera el pozo de dicho agricultor como seco, cuando en realidad este no es el caso.



(a) Arvejas.



(b) Papas.



(c) Tomates.

Figura 6.7: Rendimientos relativos Y_r para los cultivos de arvejas, papas y tomates con $P_{scale} = 0$ (agente PPO single con recompensa 1 y $c_1 = 1$).

Respecto a los pozos, como era de esperar presentan una tendencia descendente, ya que el acuífero que les suministra agua, se modeló con la lluvia como única fuente de recarga. Pese a esto, es destacable el hecho de que sólo se alcancen a secar dos pozos, que están asociados a los agricultores con mayores superficies cultivadas y que por lo tanto, requieren mayores volúmenes de agua para desarrollar sus cultivos al máximo. En este sentido, el agente PPO single se destaca como una solución capaz de administrar el agua de forma eficiente, si bien no tan eficiente como el método tradicional y el balance diario.

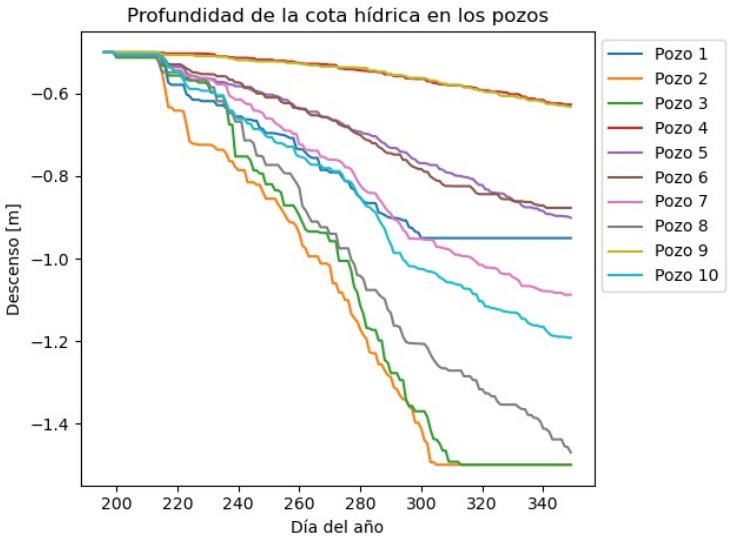


Figura 6.8: Descenso en el nivel de los pozos con $P_{scale} = 0$ (agente PPO single con recompensa 1 y $c_1 = 1$).

6.3. Análisis general de resultados

Respecto a las recompensas estudiadas, como ya se mencionó, resultan elevadas en el esquema con $P_{scale} = 1$ debido al aporte de las precipitaciones al desarrollo de cultivos de acuerdo con el modelo considerado, lo cual no permite analizarlas adecuadamente bajo dicho esquema. Es por esta razón que se consideró el esquema $P_{scale} = 0$, donde se manifiesta de mejor forma la diferenciación entre una buena y una mala estrategia de riego, al existir una completa dependencia entre los rendimientos y la gestión del agua suministrada a los cultivos.

También, se destaca que las recompensas 1 y 3 obtenidas con los esquemas con y sin pozos, son similares entre las distintas estrategias de aprendizaje reforzado estudiadas y ese sentido, manifiestan una cierta insensibilidad a la presencia de los pozos en el modelo. En cuanto a la recompensa 3, como ya se mencionó, se trata de una recompensa con poca capacidad diferenciadora entre una buena y una mala gestión, lo que explica este fenómeno. Por otro lado, la recompensa 1, considera los rendimientos parciales en cada día del proceso, donde los iniciales son elevados y van decayendo hacia el final, por lo tanto la evolución de estos es similar en ambos escenarios y su contribución a la recompensa final también.

En cuanto a los riegos totales, no existe una tendencia clara entre un mayor o menor consumo de agua al comparar los esquemas con pozos y sin pozos, donde se hubiera esperado que el esquema sin pozos entregara volúmenes de agua superiores (recordar que el esquema sin pozos hace referencia a despreciar el secado de pozos por extracción excesiva de agua). Esto se puede deber al hecho de no haber considerado en las recompensas los niveles de descenso de agua en los pozos, lo que repercute en que los agentes entrenados no aprenden a no secar los pozos y en este sentido los objetivos buscados en cada esquema son los mismos, sin el facto diferenciador de la ausencia de agua en los pozos que impacta en la no disponibilidad de agua para riego.

Por parte de los rendimientos, en general estos resultan superiores en el esquema sin pozos, lo que se justifica, como ya se ha mencionado, por la mayor libertad para disponer de grandes volúmenes de agua en todo momento al no existir la restricción de riegos nulos cuando se secan los pozos.

En cuanto a las mejores estrategias de aprendizaje reforzado, en ambos esquemas se destaca PPO en su versión *single* (un agente por tipo de cultivo), lo que se explica por el hecho de que durante el periodo de entrenamiento, se consideran secuencias de más de una transición de estado para la actualización de parámetros del agente, lo que permite estimar de mejor forma el impacto que tienen las acciones de control en las recompensas futuras. Esto no ocurren con DDPG y TD3, ya que estos muestrean aleatoriamente conjuntos de transiciones individuales sin estar estas relacionadas entre sí de forma secuencial durante el entrenamiento

y esto dificulta la estimación de la buena o mala calidad de las acciones aplicadas.

Para finalizar, se destaca que los resultados de riego diarios aplicados por los mejores agentes para cada modelo se presentan en el apéndice F.

Capítulo 7

Conclusiones

En el presente trabajo se estudiaron tres estrategias de aprendizaje reforzado para espacios de acción y de estado continuos (PPO, DDPG y TD3), sobre un modelo del sistema hidrogeológico conformado por un conjunto de agricultores y pozos de la comunidad José Painecura, comparando sus desempeños con estrategias de riego basadas en el balance hídrico de suelos y en control predictivo.

Los resultados obtenidos mostraron que los agentes entrenados no lograron superar el desempeño de los métodos de balance hídrico, ya que en las distintas configuraciones testeadas, cuando se alcanzaban altos rendimientos, esto ocurría a expensas de un mayor consumo de agua. Por otro lado, cuando los agentes determinaban volúmenes de agua menores que los obtenidos por el balance hídrico, esto impactaba negativamente en los rendimientos, los cuales también resultaban ser inferiores.

De los tres algoritmos de aprendizaje reforzado testeados, el que entregó mejores resultados fue PPO, en su versión single (un agente por tipo de cultivo), tanto para el modelo con pozos como para el modelo sin pozos, superando incluso al caso MPC con PSO. Esto se puede deber en primer lugar, al hecho de que durante el entrenamiento del agente con PPO, los parámetros de este se actualizan en función de trayectorias de más de una transición (s_t, a_t, r_t, s_{t+1}), lo que permite estimar de mejor forma el impacto (calidad) de las acciones ejecutadas (versus DDPG y TD3 que actualizan en función de transiciones a un paso). Además, el hecho de entrenar un agente por tipo de cultivo, implica la utilización de un sistema considerablemente más sencillo, en comparación al sistema agrícola completo con los 22 cultivos y los 10 pozos.

A pesar de lo anterior, PPO presenta la desventaja de que se trata de un actor estocástico, lo que impacta en la reproducibilidad de las acciones ejecutadas para una determinada serie de estados observados, donde es altamente probable que sean distintas entre una simulación y otra. Esto también afecta al hecho de alcanzar un controlador óptimo con PPO, ya que si bien, la serie de acciones ejecutadas durante distintos episodios entregarán recompensas totales similares, no se puede asegurar durante un episodio que la recompensa total al finalizar será la mejor posible.

Respecto a DDPG y TD3, los resultados muestran rendimientos deficientes en la mayoría de los casos y cuando no, se debe a la aplicación de altos volumenes de agua. Además, los agentes entrenados con estos algoritmos manifestaron una tendencia a la saturación, regando el máximo o el mínimo (ver Cuadros 6.4-6.6 y 6.13-6.15 donde hay casos en que el riego es 0 durante todo el periodo) y por lo tanto mostrándose insensibles al carácter continuo del problema estudiado.

En cuanto al/los simulador(es) desarrollado(s) con la estructura del entorno Gym, su implementación fue exitosa y permitió la utilización de las implementaciones de aprendizaje reforzado de *Stable Baselines 3* para el entrenamiento de los agentes de control y también, permite su utilizaciones con cualquier otro tipo de entornos compatibles con dicha estructura. En este sentido, se satifizo uno de los objetivos específicos del presente trabajo.

7.1. Trabajo futuro

Si bien los resultados obtenidos con PPO, DDPG y TD3 no constituyen una mejor solución al problema de maximizar rendimientos economizando agua respecto a las estrategias de balance hídrico, los resultados presentados por sus desarrolladores en la resolución de otros problemas y los resultados en general, que muestran la capacidad del aprendizaje reforzado de encontrar mejores soluciones que las humanas, inducen a seguir explorando los controladores basados en aprendizaje reforzado para optimizar el riego en cultivos.

Respecto a lo anterior, existen diversos caminos a seguir para mejorar los resultados presentados. En primer lugar, se propone la evaluación de otros algoritmos de aprendizaje reforzado, ya sea especializados en espacios de acción y de estado continuos o para espacios discretos, pero modificados adecuadamente como el presentado en [18].

Otra alternativa, especialmente tomando en cuenta la susceptibilidad de los algoritmos de aprendizaje reforzado a la elección de parámetros, es entrenar a los agentes con parámetros y arquitecturas de redes neuronales distintos a los presentados aquí, existiendo una amplia variedad de combinaciones posibles en este sentido. Particularmente, se recomienda experimentar con ANN con mayor cantidad de neuronas y tiempos de entrenamiento más largos. También, se plantea la posibilidad de considerar otras variables de estado, como se plantea en [18], donde utilizan predicciones de precipitaciones dentro del estado observado por el agente.

Un tercer camino propuesto, considera la definición de recompensas distintas a las evaluadas en el presente trabajo, ya que tal como se observa en los resultados tabulados de las recompensas totales y particularmente cuando se consideró $P_{scale} = 1$, desde el punto de vista de la recompensa, no existen grandes diferencias entre regar adecuadamente (método tradicional y balance diario) y no regar, lo que se traduce en que para el agente se vuelve mucho más compleja la tarea de discernir los volúmenes adecuados de riego a aplicar. En este sentido, una recompensa distinta puede ayudar tanto a acelerar el proceso de aprendizaje, como a alcanzar mejores políticas de toma de decisión.

Ahora, respecto a la modelación del sistema estudiado, se propone la realización de estudios de suelo en la misma comunidad José Painecura, para poder contar con parámetros que representen de forma más verídica la realidad del sistema, ya que los parámetros utilizados corresponden a valores típicos de acuerdo al tipo de suelo considerado, sin embargo pueden existir variaciones. Esto ayudaría al desarrollo de sistemas de gestión de riego mejor ajustados a la realidad local.

También, se propone la implementación de otros modelos dinámicos de los cultivos, especialmente aquellos que tengan en consideración los efectos asociados al riego excesivo (se sabe que esto es perjudicial para los cultivos, sin embargo el modelo utilizado aquí no considera este fenómeno), los niveles de salinidad del suelo y modelos de crecimiento de los cultivos que den cuenta del estrés hídrico (recordar que el modelo de crecimiento utilizado en este trabajo y su dependencia con el estrés hídrico no han sido validados empíricamente, sin embargo, su motivación esta fundada en curvas de crecimiento típicamente encontradas en la literatura). Es importante destacar aquí, que estos modelos pueden requerir mediciones de otros datos, por lo que se debe tener especial cuidado respecto a la información disponible para su implementación.

Por último, se propone la consideración de simulaciones con otros cultivos, como trigo, avena y manzanas, que también son desarrollados por las familias de la comunidad [37] y la consideración, en caso de contar con los datos, del agua utilizada para consumo humano, que agrega otra limitante a los volúmenes de agua a utilizar si es que no se quiere secar los pozos.

Bibliografía

[1] T. Roje, "Control predictivo aplicado al nexo "agua-energía-alimento" en comunidades rurales," Master's thesis, Universidad de Chile, 2020.

[2] R. Allen, L. Pereira, D. Raes, and M. Smith, "*Crop Evapotranspiration (guidelines for computing crop water requirements)*," No. 56 in FAO Irrigation and Drainage, (Rome), Food and Agriculture Organization of the United Nations, 1998.

[3] United Nations, Department of Economic and Social Affairs, Population Division, *World Population Prospects: The 2017 Revision, Key Findings and Advance Tables*, Working Paper No. ESA/P/WP/248, 2017.

[4] *Panorama de la agricultura chilena 2019*, (Santiago,Chile), Ministerio de Agricultura. Oficina de Estudios y Políticas Agrarias, 2019.

[5] J. Famiglietti, "The global groundwater crisis," *Nature Climate Change*, vol. 4, pp. 945–948, november 2014.

[6] R. D. Garreaud, J. P. Boisier, R. Rondanelli, A. Montecinos, H. H. Sepúlveda, and D. Veloso-Aguila, "The central chile mega drought (2010–2018): A climate dynamics perspective," *International Journal of Climatology*, 2019.

[7] OECD, *OECD Rural Policy Reviews: Chile 2014*. 2014.

[8] INE, "Productores agropecuarios y forestales individuales por pueblo originario según region provincia y comuna." Online: `https://www.ine.cl/estadisticas/economia/agricultura-agroindustria-y-pesca/censos-agropecuarios`, 2007.

[9] M. Casanova, "Propuesta para el desarrollo productivo agronómico de la comunidad lafquenche "josé painecura"," Master's thesis, Universidad de Chile, 2021.

[10] F. L. Lewis and D. L. Vrabie, "Reinforcement learning and adaptive dynamic programming for feedback control," *IEEE Circuits and Systems Magazine*, vol. 9, pp. 32–50, 2009.

[11] K. Ogata, *Modern Control Engineering*. New Jersey: Pearson, 5th ed., 2010.

[12] M. Goodchild, K. Kühn, A. Burek, M. Jenkins, and A. Dutton, "A method for precision closed-loop irrigation using a modified pid control algorithm," *Sensors and Transducers*, vol. 188, pp. 61–68, 05 2015.

[13] R. Babuška, *Fuzzy Modeling for Control*. New York: Springer, 1st ed., 1998.

[14] B. MMohammed, H. Bekkay, A. Migan-Dubois, M. Adel, and A. Rabhi, "An intelligent irrigation system based on fuzzy logic control: A case study for moroccan oriental climate region," (Morocco), 2nd international conference on Embedded Systems and Artificial Intelligence (ESAI'21), 2021.

[15] E. Camacho and C. Bordons, *Model Predictive Control*. London: Springer, 1999.

[16] S. Saleem, D. Delgoda, S. Ooi, K. Dassanayake, L. Liu, M. Halgamuge, and H. Malano, "Model predictive control for real-time irrigation scheduling," *4th IFAC Conference on Modelling and Control in Agriculture, Horticulture and Post Harvest Industry*, 2013.

[17] R. Sutton and A. Barto, *Reinforcement learning An introduction*. Cambridge, MA : The MIT Press, 2nd ed., 2018.

[18] M. Chen, Y. Cui, X. Wang, H. Xie, F. Liu, T. Luo, S. Zheng, and Y. Luo, "A reinforcement learning approach to irrigation decision-making for rice using weather forecasts," *Agricultural Water Management*, vol. 250, p. 106838, 2021.

[19] C. V. Theis, "The relation between the lowering of the piezometric surface and the rate and duration of discharge of a well using ground-water storage," *Eos, Transactions American Geophysical Union*, vol. 16, pp. 519–524, 1935.

[20] Y. K. Birsoy and W. K. Summers, "Determination of aquifer parameters from step tests and intermittent pumping data," *Ground Water*, vol. 18, pp. 137–146, 1980.

[21] N. Brozović, D. Sunding, and D. Zilberman, "Optimal management of groundwater over space and time," in *Frontiers in water resource economics*, vol. 29, ch. 6, pp. 109–135, Boston: Springer, 2006.

[22] T. G. Andualem, G. G. Demeke, I. Ahmed, M. A. Dar, and M. Yibeltal, "Groundwater recharge estimation using empirical methods from rainfall and streamflow records," *Journal of Hydrology: Regional Studies*, 2021.

[23] R. W. Healy and P. G. Cook, "Using groundwater levels to estimate recharge," *Hydrogeology Journal*, vol. 10, pp. 91–109, 2002.

[24] M. E. Jensen and R. G. Allen, "Evaporation, evapotranspiration, and irrigation water requirements," 2016.

[25] J. Doorenbos and A. H. Kassam, "*Yield response to water*," No. 33 in FAO Irrigation and Drainage, (Rome), Food and Agriculture Organization of the United Nations, 1979.

[26] H. L. Penman, "Natural evaporation from open water, bare soil and grass," *Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences*, vol. 193, no. 1032, pp. 120–145, 1948.

[27] D. Raes, S. Geerts, E. C. Kipkorir, J. Wellens, and A. Sahli, "Simulation of yield decline as a result of water stress with a robust soil water balance model," *Agricultural Water Management*, vol. 81, pp. 335–357, 2006.

[28] E. A. Feinberg and A. Shwartz, *Handbook of Markov Decision Processes : Methods and Applications*. International Series in Operations Research and Management Science, Boston: Springer, 1st ed., 2002.

[29] D. Silver, G. Lever, N. M. O. Heess, T. Degris, D. Wierstra, and M. A. Riedmiller, "Deterministic policy gradient algorithms," in *ICML*, 2014.

[30] J. Schulman, F. Wolski, P. Dhariwal, A. Radford, and O. Klimov, "Proximal policy optimization algorithms," *ArXiv*, vol. abs/1707.06347, 2017.

[31] T. P. Lillicrap, J. J. Hunt, A. Pritzel, N. M. O. Heess, T. Erez, Y. Tassa, D. Silver, and D. Wierstra, "Continuous control with deep reinforcement learning," *CoRR*, vol. abs/1509.02971, 2016.

[32] S. Fujimoto, H. van Hoof, and D. Meger, "Addressing function approximation error in actor-critic methods," *ArXiv*, vol. abs/1802.09477, 2018.

[33] J. Inostroza, "Manual de papa para la araucanía: manejo y plantación," No. 193 in Boletín INIA - Instituto de Investigaciones Agropecuarias, (Temuco), Instituto de Investigaciones Agropecuarias. Centro Regional de Investigación Carillanca, noviembre 2009.

[34] M. Mera, N. Espinoza, R. Galdames, and P. Navarro, "Producción de arveja para consumo fresco," No. 80 in Informativo INIA Carillanca, (Temuco), Instituto de Investigaciones Agropecuarias. Centro Regional de Investigación Carillanca, julio 2015.

[35] Ingeniería Sin Fronteras Chile. Región de la Araucanía, Chile, *Manual Básico para Cuidado y Cultivo de Plantas*.

[36] C. Ahumada, "Sistema de gestión de agua acoplado a una micro-red para comunidades mapuche," Master's thesis, Universidad de Chile, 2018.

[37] A. Fontanilla, "Estrategias de sustentabilidad de micro-redes/smart-farm en la comunidad mapuche josé painecura de hueñalihuen," Master's thesis, Universidad de Chile, 2018.

[38] Centro de Información de Recursos Naturales, *Estudio agrológico IX Región : descripciones de suelos, materiales y símbolos*, no. 122 in IREN-CIREN, (Santiago, Chile), 2002.

[39] S. P. Loheide, J. J. Butler, and S. M. Gorelick, "Estimation of groundwater consumption by phreatophytes using diurnal water table fluctuations : A saturated-unsaturated flow assessment," *Water Resources Research*, vol. 41, p. 07030, 2005.

[40] D. Todd and L. Mays, *Groudwater Hydrology*. USA: Wiley, 3rd ed., 2005.

[41] R. Heath, "Basic ground-water hydrology," No. 2220 in Water Suply Paper, (Reston, VA), U.S. Geological Survey, 1983.

[42] J. Kennedy, R. Eberhart, and Y. Shi, *Swarm Intelligence*. San Francisco: Morgan Kaufmann, 1st ed., 2001.

Anexos

Anexo A

Conceptos hidrogeológicos

Acá se definen algunos de los conceptos hidrogeológicos mencionados en el Capítulo 3. Siguiendo la nomenclatura típica, se utiliza la letra L para hacer referencia a unidades de medida de longitud y la letra T para las unidades de medida de tiempo. Todas estas definiciones se tomaron de [40], salvo el rendimiento específico, que se obtuvo de [41].

- **Acuífero no confinado:** Corresponde a acuíferos donde el nivel freático varía de acuerdo a las áreas de carga y recarga del acuífero, bombeo desde pozos y la permeabilidad del suelo. Estos cambios están asociados a variaciones en el volumen de agua almacenada en el acuífero.
- **Acuífero confinado:** Corresponde a acuíferos donde el agua subterránea está sometida a mayor presión que la atmosférica, ya que se encuentra confinada por capas de sustrato relativamente impermeables tanto superior como inferiormente.
- **Superficie piezométrica:** Definida para acuíferos confinados, corresponde a una superficie imaginaria, donde el nivel de presión hidrostática coincide con el nivel de presión del agua dentro del acuífero. Para un pozo penetrando un acuífero confinado, el nivel de agua en el pozo define la elevación de la superficie piezométrica.
- **Conductividad hidráulica:** Medida en $[L/T]$, representa la habilidad del agua para fluir a través del sedimento, es decir, caracteriza la porosidad de éste. Se la simboliza con la letra K.
- **Espesor:** Medido en $[L]$, corresponde a la distancia entre las capas impermeables inferior y superior de un acuífero confinado. Se lo simboliza con la letra b.
- **Transmisividad:** Medida en $[L^2/T]$, representa la tasa de transmisión del agua a través de un ancho unitario del acuífero con un gradiente hidráulico unitario. Se define como:

$$T = K \cdot b, \tag{A.1}$$

  con K la conductividad hidráulica y b el espesor del acuífero.
- **Almacenamiento específico:** Medido en $[L^{-1}]$, corresponde al volumen de agua liberado por volumen unitario de acuífero al ocurrir un descenso unitario en la cota hídrica. Se lo simboliza como S_s.
- **Rendimiento específico:** Adimensional, corresponde al volumen de agua liberado de un acuífero no confinado por unidad de área superficial por unidad de descenso de la cota hídrica, debido a la gravedad. Se lo simboliza como S_y.
- **Coeficiente de almacenamiento:** Adimensional, corresponde al volumen de agua liberado de un acuífero por unidad de área, al ocurrir un descenso de la cota hidráulica. Se define como:

$$S = S_y + S_s b, \tag{A.2}$$

  con S_y el rendimiento específico, S_s el almacenamiento específico y b el espesor del acuífero.

Anexo B

Cálculo de variables de la ecuación FAO Penman-Monteith

La ecuación FAO Penman-Monteith presentada en [2] permite calcular la evapotranspiración de referencia ET_o como:

$$ET_o = \frac{0{,}408\Delta(R_n - G) + \gamma\frac{900}{T+273}u_2(e_s - e_a)}{\Delta + \gamma(1 + 0{,}34u_2)}, \tag{B.1}$$

con ET_o la evapotranspiración de referencia en $[mm]$, Δ la pendiente de la curva de presión de vapor respecto a la temperatura en $[kPa/°C]$, R_n la radiación solar diaria neta en la superficie del cultivo en $[MJ/(m^2\,día)]$, G la densidad de flujo de calor del suelo en $[MJ/(m^2\,día)]$, γ la constante psicrométrica en $[kPa/°C]$, T la temperatura media a dos metros de altura en $[°C]$, u_2 la velocidad del viento a dos metros de altura en $[m/s]$, e_s la presión de vapor de saturación en $[kPa]$ y e_a la presión de vapor real en $[kPa]$.

A continuación se detalla el cálculo de cada una de las variables presentes en la ecuación B.1 de acuerdo a la metodología planteada en [2], considerando que se cuenta con mediciones de temperatura mínima T_{min} y máxima T_{max} en $[°C]$, radiación solar media R_s en $[W/m^2]$ y velocidad del viento u en $[m/s]$, además del número del día considerado $J \in [1, 365]$, latitud φ en $[rad]$, altura sobre el nivel del mar z en $[m]$ y la altura a la cual se mide la velocidad del viento h en $[m]$.

Temperatura media

La temperatura media T se define como:

$$T = \frac{T_{max} + T_{min}}{2}. \tag{B.2}$$

Constante psicrométrica

Para determinar la constante psicrométrica γ, primero se calcula la presión atmosférica P_{atm} en $[kPa]$ como:

$$P_{atm} = 101{,}3\left(\frac{293 - 0{,}0065z}{293}\right)^{5{,}26}, \tag{B.3}$$

con lo cual se tiene que:

$$\gamma = 0{,}665 \cdot 10^{-3} P_{atm}. \tag{B.4}$$

Presión de vapor de saturación media

La presión de vapor de saturación media se define como:

$$e_s = \frac{e_{s,min} + e_{s,max}}{2}, \tag{B.5}$$

donde $e_{s,min}$ y $e_{s,max}$ corresponden a las presiones de vapor de saturación mínima y máxima, las que se determinan como:

$$e_{s,min} = 0{,}6108 \exp\left(17{,}27 \frac{T_{min}}{T_{min} + 237{,}3}\right), \tag{B.6}$$

$$e_{s,max} = 0{,}6108 \exp\left(17{,}27 \frac{T_{max}}{T_{max} + 237{,}3}\right). \tag{B.7}$$

Presión de vapor real

Cuando no existen datos de la humedad relativa, la presión de vapor real e_a se puede estimar como:

$$e_a = e_{s,min}. \tag{B.8}$$

Pendiente de la curva de presión de vapor respecto a la temperatura

La pendiente de la curva de presión de vapor respecto a la temperatura Δ se define como:

$$\Delta = 4098 \frac{0{,}6108 \exp\left(\frac{17{,}27T}{T+237{,}3}\right)}{(T + 237{,}3)^2}. \tag{B.9}$$

Radiación solar diaria neta

Para determinar la radiación solar diaria neta, primero se determinan la distancia inversa relativa de la tierra al sol d_r y la declinación solar δ en $[rad]$ como:

$$d_r = 1 + 0{,}033 \cos\left(2\pi \frac{J}{365}\right), \tag{B.10}$$

$$\delta = 0{,}409 \sin\left(2\pi \frac{J}{365} - 1{,}39\right). \tag{B.11}$$

Luego, el ángulo a la puesta del sol ω_s en $[rad]$ se define como:

$$\omega_s = \arccos[-\tan(\varphi)\tan(\delta)]. \tag{B.12}$$

Con esto, se puede calcular la radiación solar extraterrestre diaria R_a en $[MJ/(m^2 día)]$ (notar que todas las radiaciones definidas a continuación están en estas mismas unidades) como:

$$R_a = \frac{1440}{\pi} G_{sc} d_r [\omega_s \sin(\varphi)\sin(\delta) + \cos(\varphi)\cos(\delta)\sin(\omega_s)], \tag{B.13}$$

con $G_{sc} = 0{,}082$ la constante solar en $[MJ/(m^2 día)]$.

Seguidamente, la radiación solar diaria de cielo despejado R_{so} se puede estimar como:

$$R_{so} = R_a \left(0{,}75 + 2 \cdot 10^{-5} z\right). \tag{B.14}$$

Por su parte, las radiaciones solares netas de onda corta R_{ns} y onda larga R_{nl} se calculan de acuerdo con:

$$R_{ns} = (1 - \alpha) R_s, \tag{B.15}$$

$$R_{nl} = \sigma(0{,}34 - 0{,}14\sqrt{e_a}) \left(\frac{(T_{min} + 273{,}16)^4 + (T_{max} + 273{,}16)^4}{2}\right) \left(1{,}35 \frac{R_s}{R_{so}} - 0{,}35\right), \tag{B.16}$$

con $\alpha = 0{,}23$ el albedo de referencia definido por la FAO en [2], $\sigma = 4{,}903 \cdot 10^{-9}$ $[MJ/(K^4 m^2 día)]$ la constante de Stefan-Boltzmann y R_s la radiación solar convertida desde el promedio en $[W/m^2]$ a $[MJ(m^2 día)]$ según:

$$R_s \left[\frac{MJ}{m^2 día}\right] = 0{,}0864 R_s \left[\frac{W}{m^2}\right]. \tag{B.17}$$

Finalmente, se tiene que la radiación solar neta sobre la superficie del cultivo está definida como:

$$R_n = R_{ns} - R_{nl}. \tag{B.18}$$

Densidad de flujo de calor del suelo

Respecto a la densidad de flujo de calor del suelo, para periodos diarios se asume que:

$$G \approx 0. \tag{B.19}$$

Velocidad del viento a dos metros de altura

La velocidad del viento a dos metros de altura u_2 se define como:

$$u_2 = \frac{4{,}87u}{\ln(67{,}8h - 5{,}42)}. \tag{B.20}$$

Anexo C

Control predictivo basado en modelos

El control predictivo basado en modelos (MPC, del inglés *Model Predictive Control*) corresponde a una estrategia de control óptimo, caracterizado por un modelo de tiempo discreto del sistema a controlar $x_{k+1} = f(x_k, u_k)$ (donde x_k representa las variables de estado y u_k las acciones de control en el instante k), un horizonte de predicción N, una función de costos $J(x_{k\to k+N}, u_{k\to k+N-1})$ y un conjunto de restricciones operacionales de la forma $(x_k, u_k) \in \mathbb{X} \times \mathbb{U}$ [15].

Luego, para cada instante k de operación, la acción de control u_k aplicada al sistema a controlar se determina resolviendo el siguiente problema de optimización:

$$\begin{aligned} \min_{x_{k\to k+N}, u_{k\to k+N-1}} \quad & J(x_{k\to k+N}, u_{k\to k+N-1}) \\ s.a. \quad & x_{k+j+1} = f(x_{k+j}, u_{k+j}) \qquad j = 0, \ldots, N-1 \\ & (x_{k+j}, u_{k+j}) \in \mathbb{X} \times \mathbb{U} \qquad j = 0, \ldots, N-1 \\ & x_{k+N} \in \mathbb{X}_f \end{aligned} \tag{C.1}$$

y donde u_k corresponde al primer término de $u^*_{k\to k+N-1}$.

Es importante notar que al tratarse de un controlador que actua en tiempos discretos, las acciones de control toman la forma de escalones. Si bien existen formulaciones de control predictivo para tiempo continuo, son más complejas de implementar y su utilización es limitada en la práctica.

Por último, dependiendo del tipo de modelo y función de costos considerados en el problema de optimización, se deberá considerar un método de resolución adecuado, que ofrezca un tiempo de computo razonable (generalmente se considera adecuado un tiempo de computo menor al 10 % del tiempo de muestreo) y una convergencia al óptimo rápida y estable (esto se debe a que en general se utilizan métodos numéricos para su resolución).

Anexo D

Optimización por enjambre de partículas

La optimización por enjambre de partículas (PSO, del inglés *Particle Swarm Optimization*) es un algoritmo de optimización estocástico basado en poblaciones, inspirado en el comportamiento de bandadas de aves, cardúmenes, enjambres de abejas, etc [42]. La idea básica consiste en que un conjunto de partículas, cada una de las cuales corresponde a una posible solución al problema de optimización que se pretenda resolver, exploran el dominio de la función objetivo, guiadas por la mejor solución histórica individual de cada una y la mejor solución histórica del conjunto, de forma tal, que con el transcurso del tiempo, las partículas se van acercando entre ellas y convergiendo a la solución buscada [42]. Pequeñas variaciones aleatorias se realizan al recorrido de cada partícula, para incentivar la exploración del espacio de soluciones y reducir el estancamiento en óptimos locales.

A continuación se presenta el pseudocódigo del algoritmo PSO:

Algorithm 4: PSO.

Inicializar aleatoriamente las posiciones de las partículas: $x_i \sim U(b_{lo}, b_{up})$
Inicializar aleatoriamente las velocidades de las partículas: $v_i \sim U(b_{lo} - b_{up}, b_{up} - b_{lo})$
Inicializar la mejor posición histórica de cada partícula: $p_i = x_i$
Inicializar la mejor posición global del enjambre: $g = \arg\min f(p_i)$
while *Condición de término no alcanzada* **do**
    for *Partícula* $i = 1, \ldots, S$ **do**
        for *Dimensión* $d = 1, \ldots, D$ **do**
            Elegir números aleatorios: $r_1, r_2 \sim U(0, 1)$
            Actualizar la velocidad de la partícula: $v_{i,d} = \omega v_{i,d} + \phi_1 r_1 (p_{i,d} - x_{i,d}) + \phi_2 r_2 (g_d - x_{i,d})$
        end
        Actualizar la posición actual de la partícula: $x_i = x_i + v_i$
        if $f(x_i) < f(p_i)$ **then**
            $p_i = x_i$
            if $f(p_i) < f(g)$ **then**
                $g = p_i$
            end
        end
    end
end
Devolver g como la mejor solución encontrada.

Anexo E

Entorno Gym

Gym es un paquete del lenguaje computacional Python, desarrollado para la experimentación de algoritmos de aprendizaje reforzado, ofreciendo implementaciones de entornos clásicos y la posiblidad de desarrollar entornos personalizados siguiendo la estructura presentada a continuación:

```
import gym
import numpy as np

class customEnv(gym.Env):

    def __init__(self,**kwargs):
        super(customEnv, self).__init__()
        # Definir parametros propios del modelo

        # Espacio de accion
        self.action_space = gym.spaces.Box(low   = action_lb,
                                           high  = action_hb,
                                           dtype = np.float32)

        # Espacio de estado
        self.observation_space = gym.spaces.Box(low   = state_lb,
                                                high  = state_hb,
                                                dtype = np.float32)

    def reset(self):
        '''Lleva al modelo a un estado inicial'''

        return obs

    def step(self,action):
        '''Aplica la accion y desarrolla la dinamica del modelo'''

        return obs,reward,done,{}

    def render(self):
        '''Visualizacion del modelo'''
```

Las razones principales para utilizar esta arquitectura son su implementación sencilla e intuitiva y su compatibilidad con paquetes que ofrecen implementaciones verificadas de algoritmos de aprendizaje reforzado (*Stable Baselines* y *Spinning Up*), convirtiendo a Gym y sus entornos en la plataforma estándar para este tipo de aplicaciones.

Anexo F

Riegos diarios por agricultor para los algoritmos con mejor desempeño

Aquí se presentan los riegos diarios aplicados a los 22 cultivos del sistema en consideración, agrupados por agricultor, de acuerdo a los agentes entrenados con aprendizaje reforzado que entregaron el mejor desempeño para los modelos con y sin pozos. Notar que estos riegos están en unidades de $[mm]$, de acuerdo con la ecuación del balance hídrico (ecuación 3.8).

F.1. PPO single con recompensa 3 (modelo sin pozos)

F.1.1. $P_{scale} = 1$

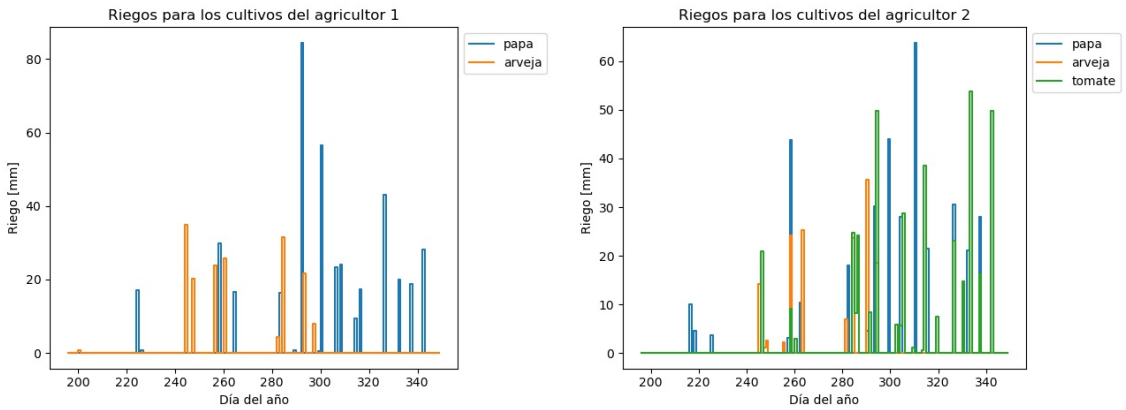


Figura F.1: Riegos aplicados a los cultivos de los agricultores 1 y 2, para el modelo sin pozos y con $P_{scale} = 1$ (agente PPO single con recompensa 3).

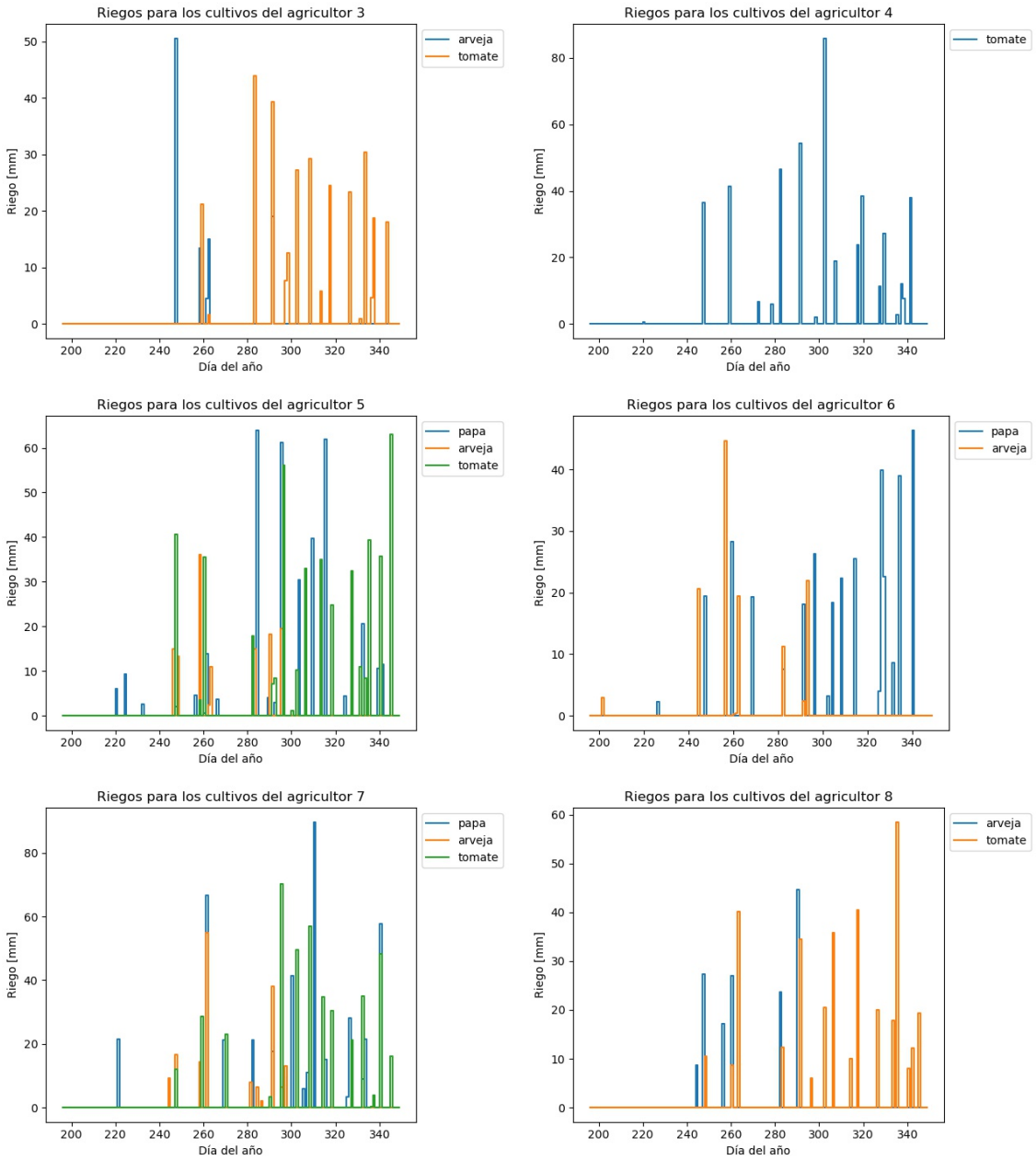


Figura F.2: Riegos aplicados a los cultivos de los agricultores 3 al 8, para el modelo sin pozos y con $P_{scale} = 1$ (agente PPO single con recompensa 3).

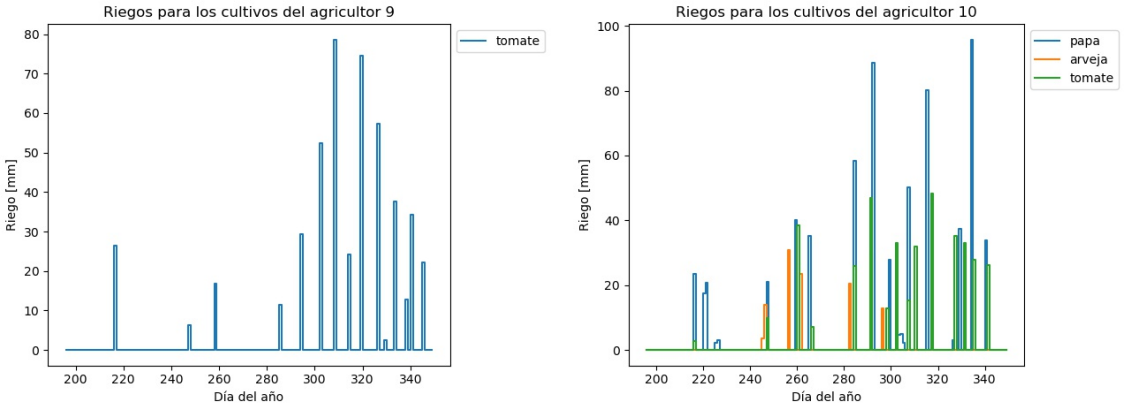


Figura F.3: Riegos aplicados a los cultivos de los agricultores 9 y 10, para el modelo sin pozos y con $P_{scale} = 1$ (agente PPO single con recompensa 3).

F.1.2. $P_{scale} = 0$

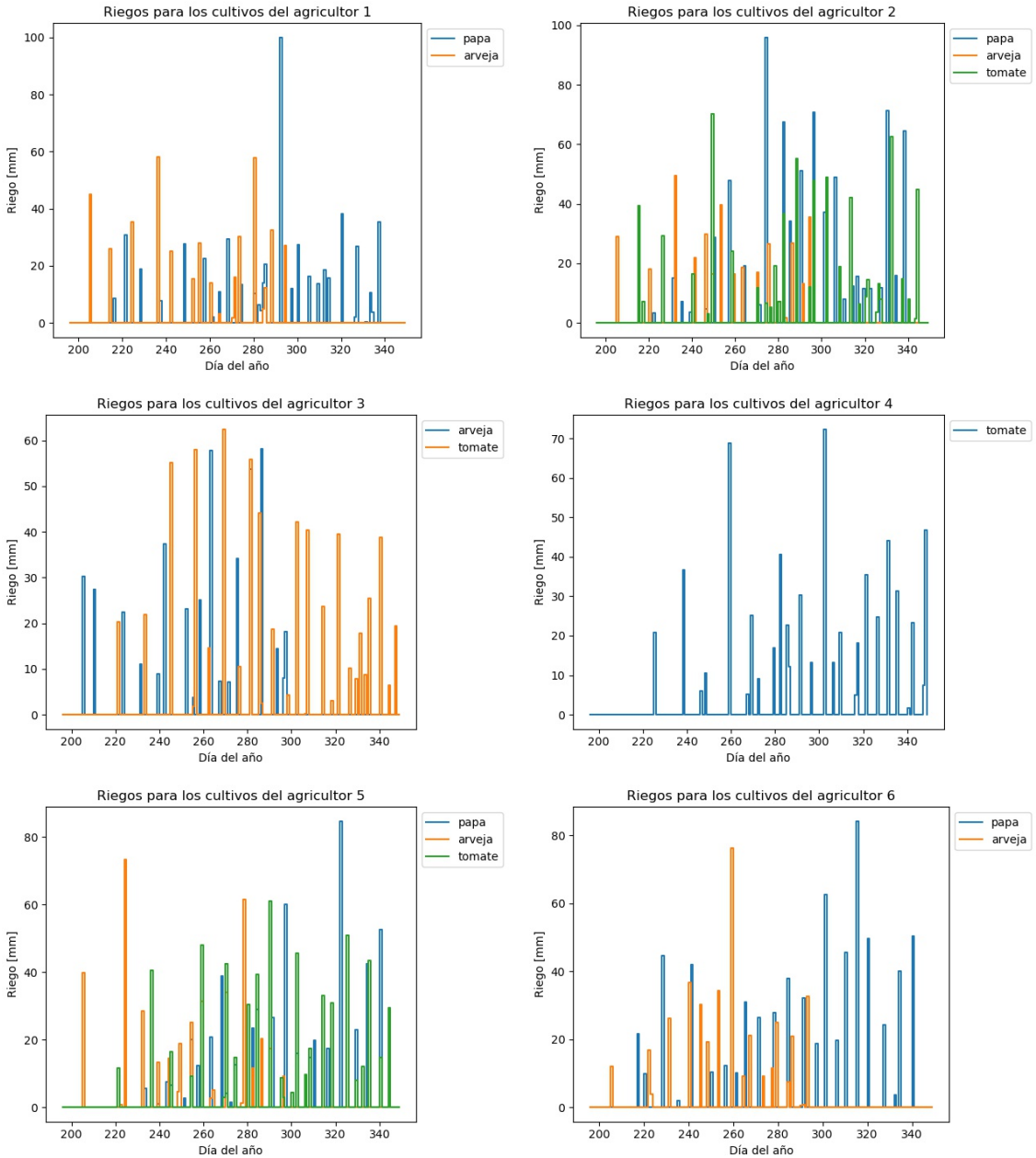


Figura F.4: Riegos aplicados a los cultivos de los agricultores 1 al 6, para el modelo sin pozos y con $P_{scale} = 0$ (agente PPO single con recompensa 3).

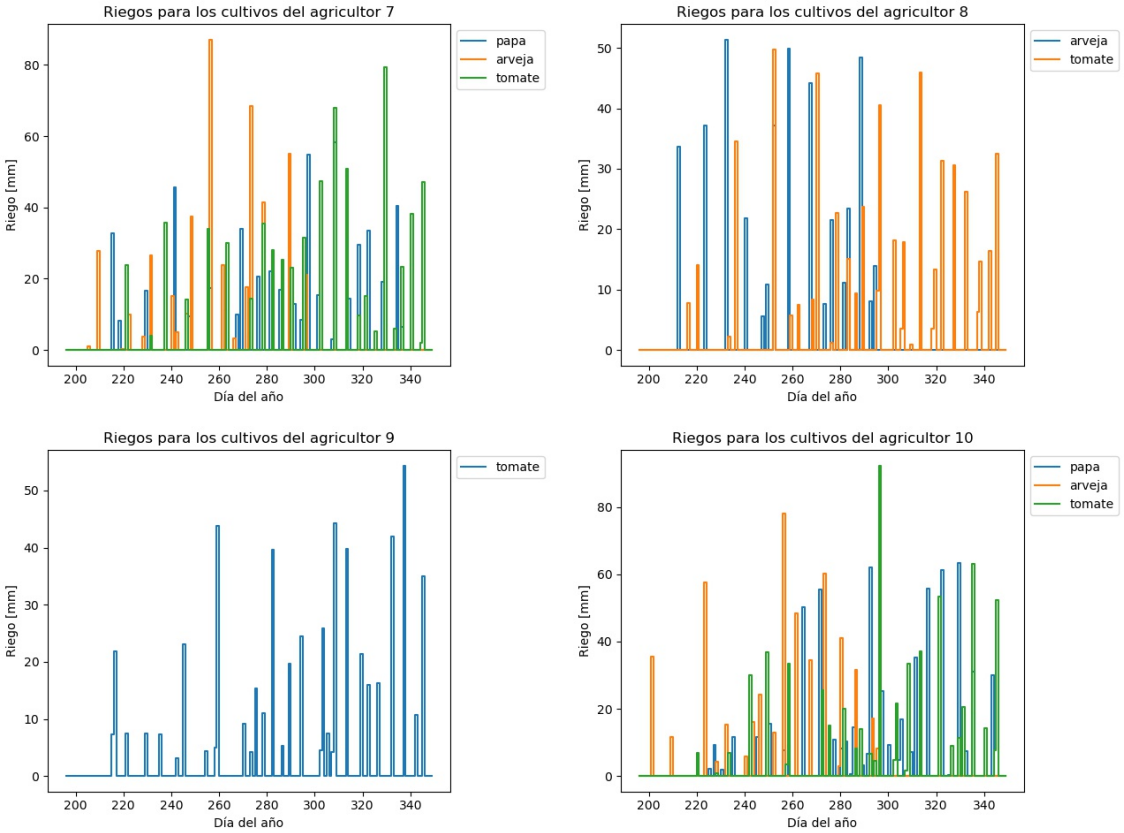


Figura F.5: Riegos aplicados a los cultivos de los agricultores 7 al 10, para el modelo sin pozos y con $P_{scale} = 0$ (agente PPO single con recompensa 3).

F.2. PPO single con recompensa 1 y $c_1 = 1$ (modelo con pozos)

F.2.1. $P_{scale} = 1$

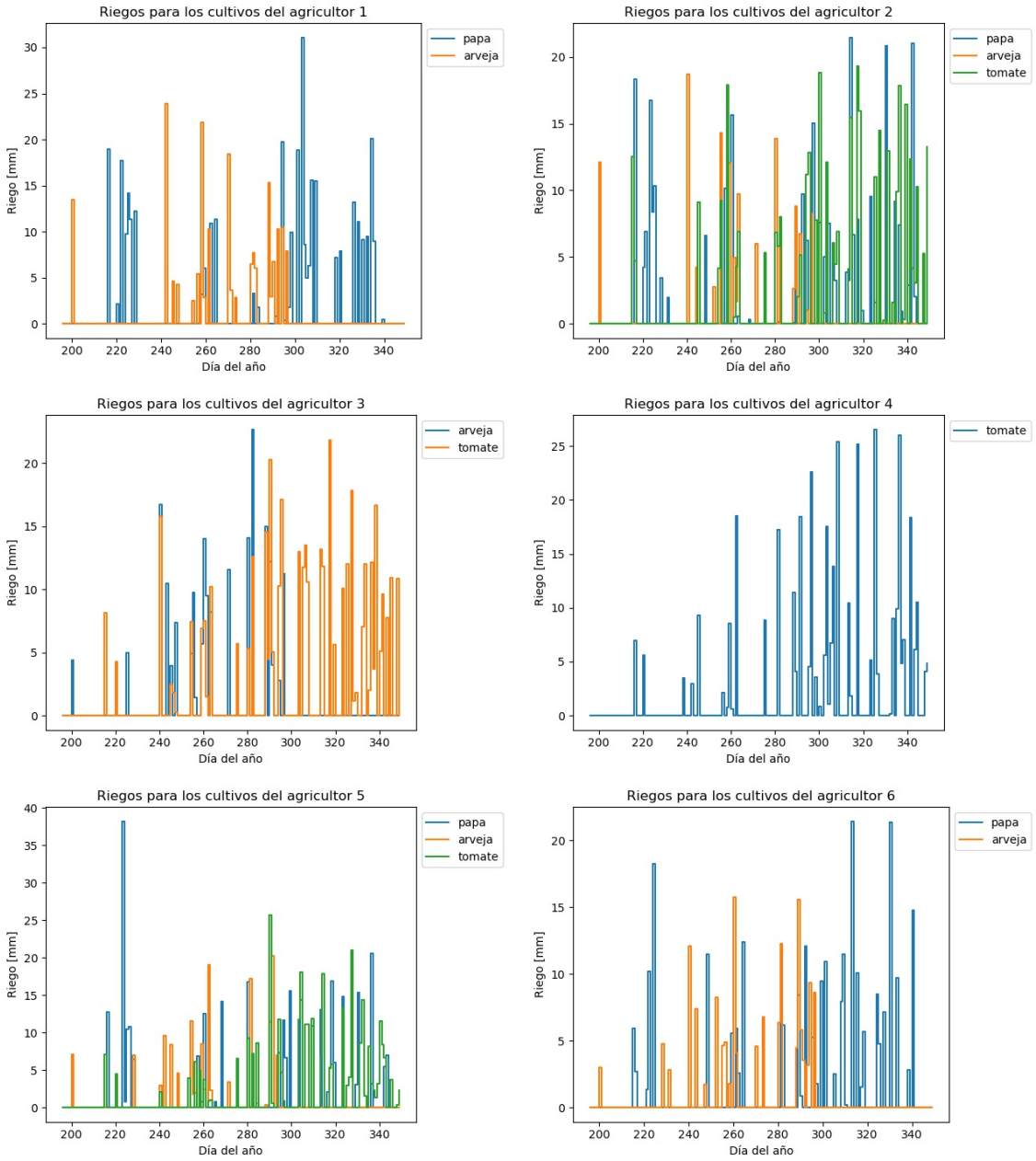


Figura F.6: Riegos aplicados a los cultivos de los agricultores 1 al 6, para el modelo con pozos y con $P_{scale} = 1$ (agente PPO single con recompensa 1 y $c_1 = 1$).

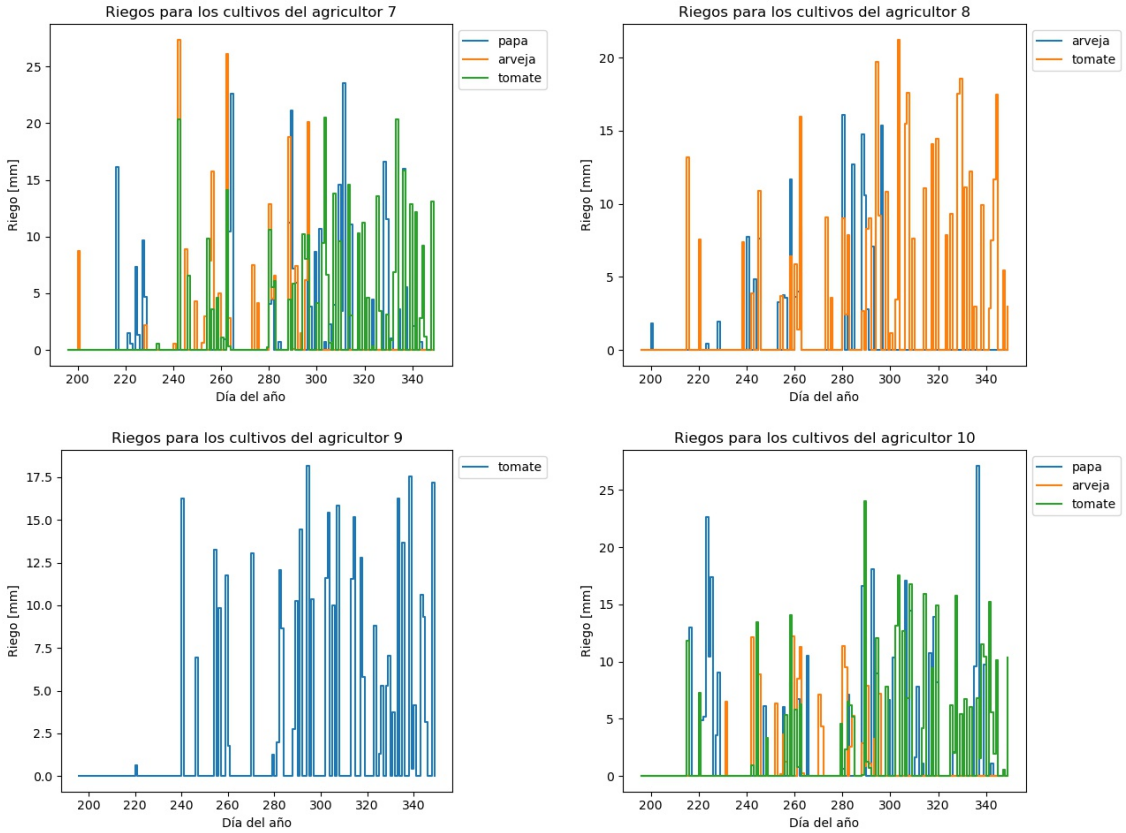


Figura F.7: Riegos aplicados a los cultivos de los agricultores 7 al 10, para el modelo con pozos y con $P_{scale} = 1$ (agente PPO single con recompensa 1 y $c_1 = 1$).

F.2.2. $P_{scale} = 0$

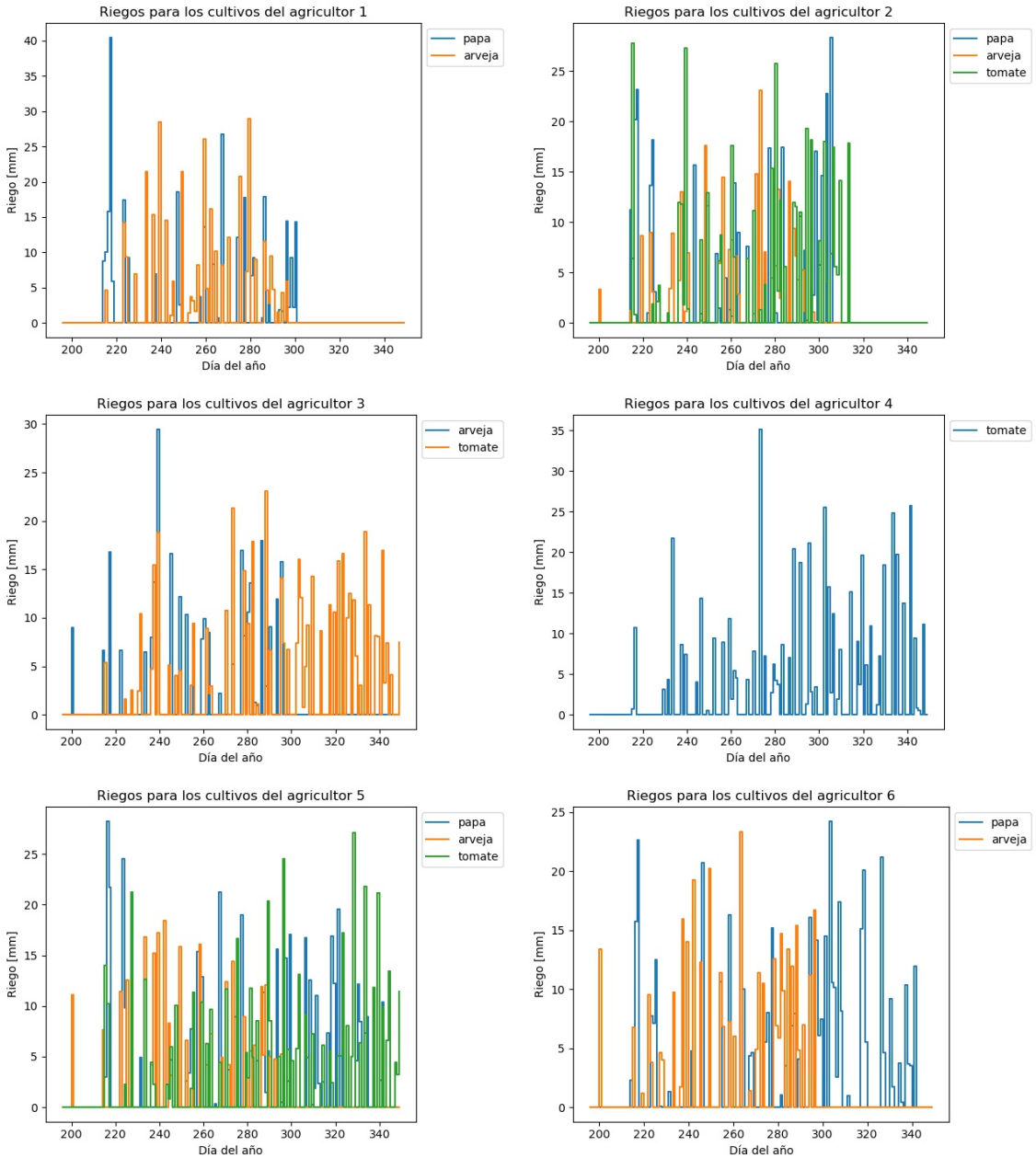


Figura F.8: Riegos aplicados a los cultivos de los agricultores 1 al 6, para el modelo con pozos y con $P_{scale} = 0$ (agente PPO single con recompensa 1 y $c_1 = 1$).

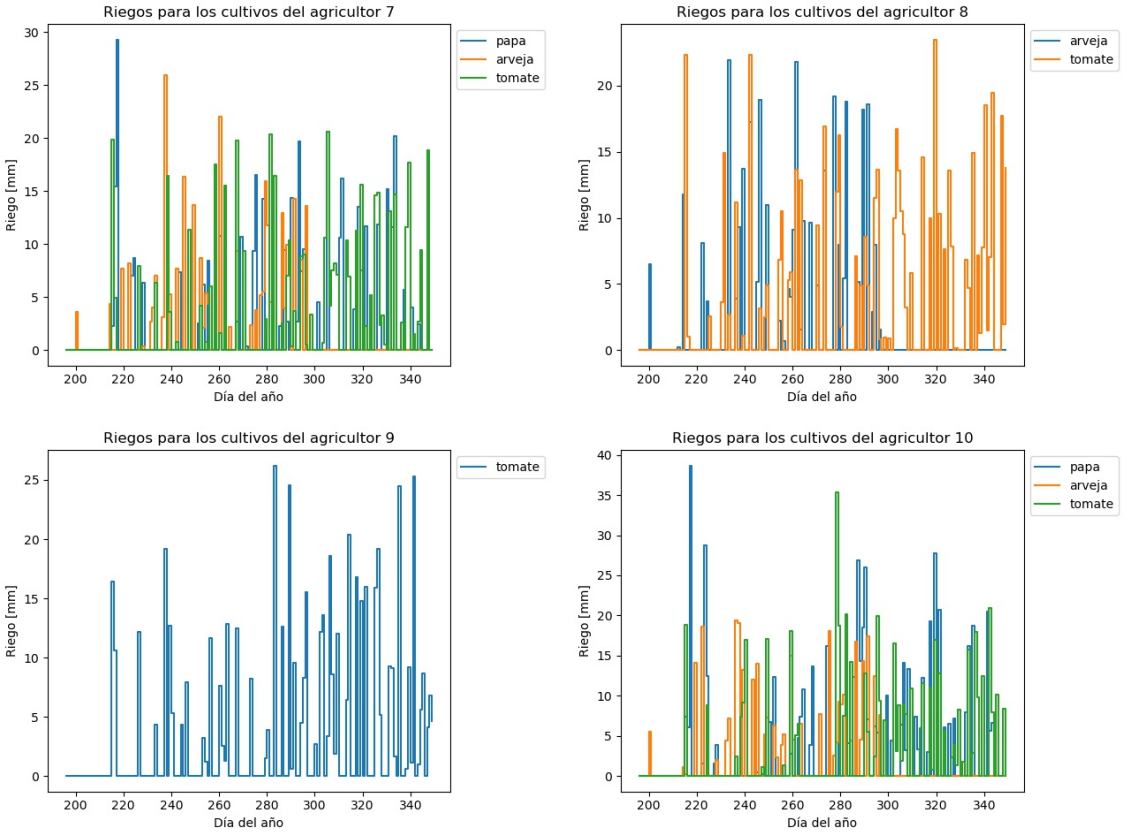


Figura F.9: Riegos aplicados a los cultivos de los agricultores 7 al 10, para el modelo con pozos y con $P_{scale} = 0$ (agente PPO single con recompensa 1 y $c_1 = 1$).