

UNIVERSIDAD TÉCNICA FEDERICO SANTA MARÍA
DEPARTAMENTO DE INGENIERÍA DE MINAS, METALURGIA Y MATERIALES
SANTIAGO – CHILE



PLANIFICACIÓN EN BLOCK/PANEL CAVING DESARROLLO E IMPLEMENTACIÓN DE UN MARCO DE APRENDIZAJE POR REFUERZO

Memoria para optar al título de Ingeniero Civil de Minas

Autor:
Martín Tomás Matus Saldías

Profesor Guía:
Dr. Gonzalo Nelis

Profesor Co-guía:
Dr. Sergio Palma

Julio, 2026



CONSTANCIA DE VALIDACIÓN Y CONFIDENCIALIDAD DE MONOGRAFÍA A REPOSITORIO ACADÉMICO

1.- IDENTIFICACIÓN DEL TRABAJO ACADÉMICO

Tipo de monografía (marcar una opción): ☒ Memoria o trabajo de título ☐ Tesis de Postgrado

Título del trabajo: Planificación en Block/Panel Caving: Desarrollo e Implementación de un Marco de Aprendizaje por Refuerzo

Nombre del candidato(a): Martín Tomás Matus Saldías

Carrera / Grado: Ingeniería Civil de Minas

Campus: San Joaquín, Santiago

Departamento: Ingeniería Minas, Metalurgia y Materiales

2.- VALIDACIÓN DEL PROFESOR GUÍA/DIRECTOR DE TESIS

Yo, Gonzalo Nelis Suazo, en mi calidad de profesor(a) guía/director(a) del trabajo académico mencionado anteriormente **DEJO CONSTANCIA** que:

- He revisado esta versión del documento y corresponde a la versión final aprobada del trabajo.
- El trabajo cumple con los requisitos académicos y de formato establecidos por la institución.

3.- EVALUACIÓN DE CONFIDENCIALIDAD POR PROPIEDAD INDUSTRIAL (marcar una opción)

☒ El trabajo **NO contiene** información que amerite confidencialidad y puede ser publicado de inmediato en repositorio con acceso abierto.

☐ El trabajo **CONTIENE** información con potenciales implicancias de propiedad industrial o intelectual y requiere un periodo de confidencialidad (**embargo**) por (marcar una opción):

☐ 6 meses ☐ 12 meses ☐ 2 años ☐ 3 años ☐ 5 años ☐ 10 años

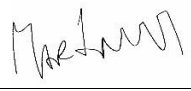
Fundamentación de la necesidad de confidencialidad (obligatorio si se solicita embargo):

4.- FIRMAS

Profesor(a) guía o director(a) de memoria o tesis:

Fecha: 04.08.2026 Firma: 

Estudiante o Candidato(a):

Fecha: 04/08/2026 Firma: 

Este formulario debe ser insertado como página 2 de la memoria o tesis, completado y firmado por estudiante y profesor(a) antes de la entrega en portal PRISMA de Biblioteca USM.

PLANIFICACIÓN EN BLOCK/PANEL CAVING: DESARROLLO E IMPLEMENTACIÓN DE UN MARCO DE APRENDIZAJE POR REFUERZO

Autor: Martín Tomás Matus Saldías

Resumen

Ante el creciente aumento en la demanda de minerales estratégicos, impulsado por la transición energética y la digitalización, Chile desempeña un papel fundamental en el abastecimiento mundial. El país lidera la producción global de cobre, ocupa el segundo lugar en molibdeno y el tercero en litio ([Consejo Minero, 2026](#)). Esto, sumado al agotamiento progresivo de las leyes minerales en superficie, ha obligado a la industria minera a una reevaluación estratégica de sus métodos de explotación, migrando hacia la minería subterránea.

En este escenario, los métodos de explotación masiva del tipo *caving* han adquirido un papel central, destacando por su alta productividad y bajo costo relativo. Lo anterior permite compensar las bajas leyes actuales con una alta tasa de extracción. El éxito de estos métodos depende de una planificación minera capaz de maximizar el VAN del proyecto, a través de una secuencia de extracción viable desde el punto de vista geomecánico y operacional, que garantice una operación segura.

La planificación en minería *caving* presenta desafíos debido a su naturaleza dinámica y a la interacción entre variables geotécnicas, productivas y económicas. En estos aspectos se evidencia una limitación del estado del arte actual para abordar dicha interrelación. Los modelos de optimización tradicionales, como la Programación Lineal Entera Mixta (MILP), presentan dificultades para capturar la evolución no lineal del hundimiento al requerir formulaciones que simplifiquen su dinámica.

Bajo este contexto, la presente memoria desarrolla un marco basado en Aprendizaje por Refuerzo (RL) aplicado a la planificación del hundimiento. Se evalúa la capacidad de este enfoque para determinar secuencias de extracción factibles, orientadas a maximizar el valor del proyecto, respetando las restricciones operacionales y geomecánicas del método. La validación se realiza sobre un caso de prueba mediante un currículum incremental, en el que cada etapa incorpora una restricción operacional para aislar y cuantificar su efecto. El agente valida el enfoque sobre el caso de prueba, recuperando la totalidad del valor económico disponible ($VRR = 1.0$). Al escalar la dimensión del perfil, la recuperación de valor desciende y depende de la inicialización del entrenamiento (0.643 ± 0.041 sobre tres semillas), lo que delimita el alcance del método tabular y motiva su extensión a esquemas de aproximación de función en un futuro. Este enfoque constituye una base metodológica para modelos de planificación que incorporen la naturaleza evolutiva del *block/panel caving*.

Palabras clave: planificación minera, *block/panel caving*, optimización, aprendizaje por refuerzo.

Agradecimientos

A mi padre, Marcos Matus. Gracias por la confianza y por cada oportunidad. Sé que me estarás acompañando y guiando, siempre.

A mi madre, Sandra Saldías. Su apoyo constante fue esencial para terminar esta etapa; nada de esto habría sido posible sin ella.

A mis hermanos, Benjamín y Agustín. Por el aguante, y porque su poca seriedad siempre me sacó una sonrisa. Recuerden que todo se puede, con trabajo y constancia. Los quiero.

A Fernanda, gracias por tu compañía y paciencia durante todo este proceso.

A los profesores Sergio Palma y Gonzalo Nelis, artífices de este tema: gracias por proponérmelo.

Al profesor Sergio, co-guía, además, por la oportunidad de la práctica y por acercarme a este proyecto.

Al profesor Gonzalo Nelis, profesor guía, por su tiempo, su buena disposición de siempre y su apoyo constante. Su confianza se la agradezco especialmente.

Tabla de Contenido

1. Introducción	6
1.1. Introducción	6
1.2. Descripción del problema	7
1.3. Preguntas de investigación e hipótesis	7
1.4. Objetivos	8
1.4.1. Objetivo general	8
1.4.2. Objetivos específicos	8
1.5. Alcances	8
2. Antecedentes	9
2.1. Selección del método de explotación	9
2.2. Métodos de explotación subterránea	10
3. Block/panel caving	12
3.1. Flujo gravitacional	13
3.2. Dilución	15
3.3. Estado de las columnas	17
3.4. Respuesta geomecánica	18
3.5. Estado del arte en planificación de <i>block caving</i>	20
3.5.1. Envolvente económica	20
3.5.2. Secuenciamiento	23
3.6. Implicancias para la planificación	26
4. Aprendizaje por refuerzo	28
4.1. Inteligencia artificial y machine learning	28
4.2. Aprendizaje por refuerzo	28
4.2.1. Proceso de decisión de Markov	28
4.2.2. Funciones de valor y Q-learning	30
4.2.3. Exploración y explotación	30
4.2.4. Q-learning tabular versus Deep Q-Network	30
4.3. Estado del arte en minería	31
4.3.1. Minería a cielo abierto	31
4.3.2. Minería subterránea	33
5. Metodología	35
5.1. Formulación del problema como proceso de decisión de Markov	35
5.2. Modelo económico y de bloques	37
5.3. Arquitectura del agente: Q-learning tabular	38
5.4. Diseño de restricciones: máscaras y penalizaciones	39

5.5. Currículum incremental de restricciones.....	40
5.5.1. E1 — Gravedad (precedencia).....	41
5.5.2. E2 — Temporalidad y descuento.....	41
5.5.3. E3 — Altura adyacente	42
5.5.4. E4 — Área mínima	42
5.5.5. E5 — Tasa de extracción por madurez	43
5.5.6. E6 — Dirección de propagación.....	43
5.5.7. E7 — Ángulo quiebre	43
5.6. Función de recompensa.....	45
5.7. Métricas de validación.....	46
5.8. Protocolo experimental.....	47
6. Resultados y análisis	48
6.1. Validación del núcleo de aprendizaje (E0–E2).....	48
6.2. Restricciones geométricas (E3–E4).....	51
6.3. Restricciones de propagación (E5–E6)	53
6.4. Control de ángulo (E7).....	55
6.4.1. Comportamiento sin penalización de ángulo	55
6.4.2. Comportamiento con penalización de ángulo.....	55
6.4.3. Sensibilidad al multiplicador de penalización	57
6.4.4. Convergencia y costo del enfoque tabular	58
6.5. Valorización económica del plan.....	59
6.5.1. Beneficio por columna y costo de desarrollo	59
6.5.2. Valorización dinámica de las columnas	60
7. Análisis de escalabilidad	62
7.1. Efecto del presupuesto de entrenamiento	62
7.2. Comparación con el caso base	63
7.3. Curva de escalamiento	64
8. Discusión	66
9. Conclusiones	67

Capítulo 1

Introducción

1.1. Introducción

La magnitud de la industria minera, pilar de la transición energética global, exige la mejora permanente de sus prácticas y procesos. Dado que las leyes minerales en los yacimientos superficiales disminuyen, la industria migra de forma decisiva hacia explotaciones subterráneas masivas ([Ranjith et al., 2017](#)). En faenas que agotan su explotación a cielo abierto, esta transición permite extender su vida operativa: un caso representativo es Chuquicamata, que pasó de ser una de las minas a cielo abierto más grandes del mundo a la explotación subterránea ([Flores and Catalan, 2019](#)), siendo los métodos de hundimiento la principal alternativa para mantener la escala productiva. Sin embargo, esta transición incrementa la complejidad. La minería subterránea es menos flexible que la minería a cielo abierto, estando sujeta a severas limitaciones geotécnicas, de equipos y de espacio ([Topal, 2008](#)). El *block caving*, si bien es atractivo por su bajo costo operacional, logrado al inducir el fracturamiento de la roca mediante la gravedad y los esfuerzos internos del macizo rocoso ([Gómez and Labbé, 2019](#)), presenta un considerable desafío de planificación. Las decisiones de cuándo y cuánto extraer de cada punto, si bien determinan el valor económico del proyecto, inducen cambios acoplados en la física del método. Por ejemplo, en el flujo gravitacional del material, que condiciona la recuperación de mineral y en el campo de esfuerzos in situ, cuya alteración puede desencadenar estallidos de roca en los niveles de producción. Los métodos de planificación tradicionales, a menudo basados en optimización matemática (MILP), presentan limitaciones para manejar esta realidad. Estos modelos, ampliamente utilizados en la industria, tienen dificultades para capturar la naturaleza evolutiva del proceso, ya que tratan el yacimiento como un sistema estático y no como uno que reacciona a las acciones del planificador. Esta limitación de los modelos estáticos motiva la exploración de nuevos paradigmas. Para abordar esta limitación, en la presente memoria se propone un marco metodológico basado en Aprendizaje por Refuerzo (RL). Este enfoque, se plantea con el fin de evaluar la opción de generar secuencias de extracción factibles, capaces de resguardar el valor económico del proyecto en entornos dinámicos. En lugar de resolver un plan estático y predefinido de principio a fin, el RL opera mediante un paradigma de decisión secuencial. El agente interactúa con un entorno simulado que transita hacia un nuevo estado espacial como consecuencia directa de las acciones ejecutadas en cada momento. Esta retroalimentación continua permite incorporar las reglas lógicas del método y generar programas de producción que se ajustan a la evolución física dictada por el avance del hundimiento.

1.2. Descripción del problema

La planificación de largo plazo en *block/panel caving* consiste en determinar la mejor secuencia para cada punto de extracción, definiendo cuánto y cuándo extraer. Esta secuencia debe respetar restricciones físicas y operacionales que no son independientes de la extracción, sino que evolucionan como consecuencia de ella. Cada decisión de extracción modifica el estado desde el cual se toman las decisiones posteriores. Las decisiones de extracción alteran la distribución de esfuerzos en el macizo, con riesgo de estallidos de roca o colapsos. Gobiernan el flujo del mineral fragmentado a través de la cavidad, y con ello la dilución y la ley que llega a planta. Mediante la tasa de extracción, se relacionan además con la respuesta sísmica de la mina. Cada decisión tiene, por lo tanto, consecuencias económicas, geomecánicas y de flujo acopladas entre sí.

El enfoque predominante para este problema es la optimización matemática, en particular la programación lineal entera mixta (MILP), que representa el yacimiento mediante un modelo cuyos parámetros geomecánicos y operacionales se fijan antes del proceso de optimización. Los efectos no lineales se simplifican mediante modelos empíricos y supuestos operacionales. El flujo gravitacional y la dilución, por ejemplo, se estiman mediante modelos de mezcla evaluados sobre una secuencia de extracción previamente definida. El plan resultante es óptimo respecto de esa representación, y su cercanía al óptimo del modelo puede cuantificarse mediante el *gap* de optimización. Al fijar previamente estos fenómenos, sin embargo, el modelo no captura la evolución del hundimiento, por lo que el plan obtenido habitualmente requiere un posprocesamiento manual para volverlo operacional.

Este trabajo aborda el problema como una decisión secuencial. La política de extracción se construye mediante la interacción con un entorno que evoluciona en respuesta a cada acción, y el estado del sistema se actualiza a medida que la secuencia avanza.

1.3. Preguntas de investigación e hipótesis

Pregunta de investigación

¿Es posible generar secuencias de extracción factibles para la minería de *block/panel caving* mediante un agente de Aprendizaje por Refuerzo (RL) que, respondiendo de forma adaptativa a la evolución espacial del yacimiento, preserve el valor económico del proyecto mientras respeta las restricciones operacionales y geomecánicas del método?

Hipótesis de trabajo

La implementación de un agente de Aprendizaje por Refuerzo (RL) para la planificación estratégica en *block/panel caving* permite identificar secuencias que preservan el valor económico del proyecto, manteniendo el cumplimiento de las restricciones operacionales y geomecánicas. De este modo, el enfoque tiende a producir planes de producción factibles, al capturar la interrelación de las variables operativas y representar cómo cada acción extractiva modifica el estado del entorno simulado.

1.4. Objetivos

1.4.1. Objetivo general

Desarrollar e implementar un marco de Aprendizaje por Refuerzo (RL) para la determinación de políticas de secuenciamiento de extracción en minería de *block/panel caving*, incorporando progresivamente las restricciones operacionales del método, para determinar la viabilidad de la herramienta en la generación de planes de producción factibles.

1.4.2. Objetivos específicos

- Modelar el problema de secuenciamiento extractivo en *block/panel caving* como un Proceso de Decisión de Markov (MDP), definiendo el espacio de estados, compuesto por la posición del bloque en el perfil y el período temporal, el conjunto de acciones permitidas y la función de recompensa basada en la valorización económica de los bloques.
- Desarrollar un entorno de simulación computacional que integre la dinámica evolutiva del hundimiento, incorporando progresivamente las restricciones operacionales y físicas del método.
- Evaluar el desempeño de las políticas obtenidas mediante indicadores económicos y operacionales, y analizar su comportamiento al escalar la dimensión del perfil de extracción.

1.5. Alcances

- El estudio se centra exclusivamente en el proceso de planificación minera a largo plazo para el método de explotación *block/panel caving*.
- Se asume una configuración geométrica donde el modelo de bloques ha sido rebloqueado conforme a la malla de extracción. De esta forma, el nivel base del modelo representa estrictamente la cota de las bateas de extracción.
- La implementación del algoritmo y el entorno de simulación se desarrolla sobre un perfil 2D compuesto por leyes sintéticas.
- El avance temporal y espacial de la extracción está discretizado en función del modelo de bloques. En consecuencia, la resolución vertical de extracción en cada columna opera en incrementos discretos equivalentes a la altura de un bloque.

Capítulo 2

Antecedentes

Determinar la mejor estrategia productiva, en función de los objetivos del dueño del negocio y sujeto a restricciones físicas, geológicas y operativas, corresponde a la planificación minera según [Johnson \(1968\)](#). Para cumplir con este fin, se debe responder tres preguntas: qué parte del yacimiento extraer, con qué método hacerlo, y en qué secuencia. Para ello, es necesario determinar primero qué parte del yacimiento resulta económicamente viable de extraer, lo cual se logra valorizando cada bloque del modelo.

2.1. Selección del método de explotación

La forma de determinar qué porción del yacimiento resulta explotable es a través de la valorización de bloque:

$$V_{\text{bloque}} = T_{\text{bloque}} [-(C_{\text{mina}} + C_{\text{planta}}) + L_{\text{bloque}} \cdot \text{Rec} \cdot k \cdot (P - C_{\text{RyV}})] \quad (2.1)$$

donde:

- T_{bloque} : tonelaje del bloque.
- C_{mina} , C_{planta} : costos de mina y planta.
- L_{bloque} : ley del bloque.
- Rec : recuperación metalúrgica.
- k : factor de conversión.
- P : precio.
- C_{RyV} : costo de refinación y venta.

Tradicionalmente, en minería a cielo abierto, esta valorización es la que utilizan [Lerchs and Grossmann \(1965\)](#) para generar *pits* anidados, a través de la multiplicación del precio P por un *revenue factor*, lo que permite definir a partir de ellos, el *pits* final y las fases de explotación.

Al despejar L_{bloque} de la Ecuación 2.1 e imponer la condición:

$$V_{\text{bloque}} = C_{\text{mina}} \cdot T_{\text{bloque}}$$

Se obtiene la Ecuación 2.2, correspondiente a la ley de corte marginal (LCM). Esta define el umbral inferior para el procesamiento del material. Los bloques con una ley inferior a la LCM se destinan a botadero, mientras que aquellos con una ley superior pueden ser enviados a planta. En particular, los bloques con leyes comprendidas entre la LCM y la ley de corte crítica (LCC) se procesan, aun cuando ello implique una pérdida económica, ya que esta es menor que la asociada a su envío a botadero.

$$L_{LCM} = \frac{C_{planta}}{Rec \cdot k \cdot (P - C_{RyV})} \quad (2.2)$$

Por otra parte, al imponer la condición de beneficio nulo del bloque,

$$V_{bloque} = 0,$$

Se obtiene la Ecuación 2.3, correspondiente a la ley de corte crítica (LCC). Esta representa el límite entre aquellos bloques cuyo procesamiento genera un beneficio positivo y aquellos cuya valorización resulta negativa.

$$L_{LCC} = \frac{C_{mina} + C_{planta}}{Rec \cdot k \cdot (P - C_{RyV})} \quad (2.3)$$

La minería a cielo abierto destaca por su capacidad para realizar remociones masivas de material cercano a la superficie, y la valorización de cada bloque permite determinar directamente su destino mediante estas leyes de corte. Sin embargo, este método se ve limitado por la profundidad del yacimiento: a medida que la explotación profundiza, el volumen de estéril que debe ser removido para acceder al mineral aumenta de forma considerable, y este aumento en la razón estéril/mineral puede volver el diseño económicamente inviable. Si además se consideran restricciones geométricas y ambientales que imposibilitan el desarrollo del rajo, la explotación superficial puede dejar de ser factible, lo que obliga a evaluar alternativas subterráneas.

En minería subterránea puede aplicarse un criterio análogo, exigiendo que $V_{bloque} > 0$. A diferencia del rajo abierto, donde el destino de cada bloque admite mayor flexibilidad, en este contexto se emplea únicamente la LCC y la planificación se orienta a enviar el material a planta. Al visualizar solo el material mineralizado que generaría un beneficio positivo al ser explotado y procesado, se transita desde el modelo geológico hacia el recurso económico.

2.2. Métodos de explotación subterránea

La elección del método de explotación subterráneo está condicionada por la forma espacial del recurso económico. Geométricamente, destacan cuatro tipos principales:

- Vetas: cuerpos tabulares con alta inclinación (subverticales).
- Mantos: cuerpos tabulares de baja inclinación (subhorizontales).
- Clavos: estructuras cilíndricas de gran desarrollo vertical.
- Yacimientos masivos: cuerpos de dimensiones extraordinarias en todas sus direcciones.

Considerando de manera conjunta la geometría del yacimiento y la calidad geotécnica tanto del mineral como de la roca de caja, es posible determinar la estrategia de extracción más adecuada. La competencia del macizo rocoso determina si una excavación puede mantenerse estable por sí sola o si colapsará. Bajo esta condición de estabilidad geomecánica, las operaciones subterráneas se clasifican en ([Hamrin, 1982](#)):

- Cavidades autosoportadas: la alta competencia de la roca permite mantener case-rones vacíos o soportados por pilares naturales.
- Cavidades artificialmente soportadas: se requiere la incorporación de relleno o fortificación sistemática para mantener la estabilidad.
- Cavidades temporalmente soportadas: el propio mineral fragmentado actúa como soporte transitorio durante la extracción.
- Cavidades no soportadas o hundidas: diseñadas para inducir el colapso controlado del macizo rocoso.

Capítulo 3

Block/panel caving

La Figura 3.1 presenta un esquema general del método de *block caving*. Este método está diseñado para la explotación de cuerpos mineralizados masivos susceptibles de ser extraídos mediante un proceso de hundimiento inducido. Su disposición se basa en la construcción de un nivel de hundimiento (*undercut level*), donde se realiza una socavación inicial mediante perforación y tronadura para eliminar el soporte basal del bloque mineralizado e inducir su fracturamiento. Cuando el área socavada alcanza el radio hidráulico crítico, el macizo rocoso comienza a ceder y a fluir por gravedad (Gómez and Labbé, 2019). Posteriormente, la extracción continua del mineral fragmentado favorece la propagación ascendente del hundimiento a través del macizo, pudiendo involucrar la roca suprayacente y generar subsidencia en superficie (Brown, 2003).

Al utilizar la energía gravitacional y los esfuerzos in situ para inducir la fragmentación continua de la roca, este método reduce las operaciones de perforación y tronadura de producción. Como consecuencia, presenta menores costos operacionales que otros métodos de explotación subterránea, lo que contribuye a la explotación económicamente viable de yacimientos masivos de baja ley.

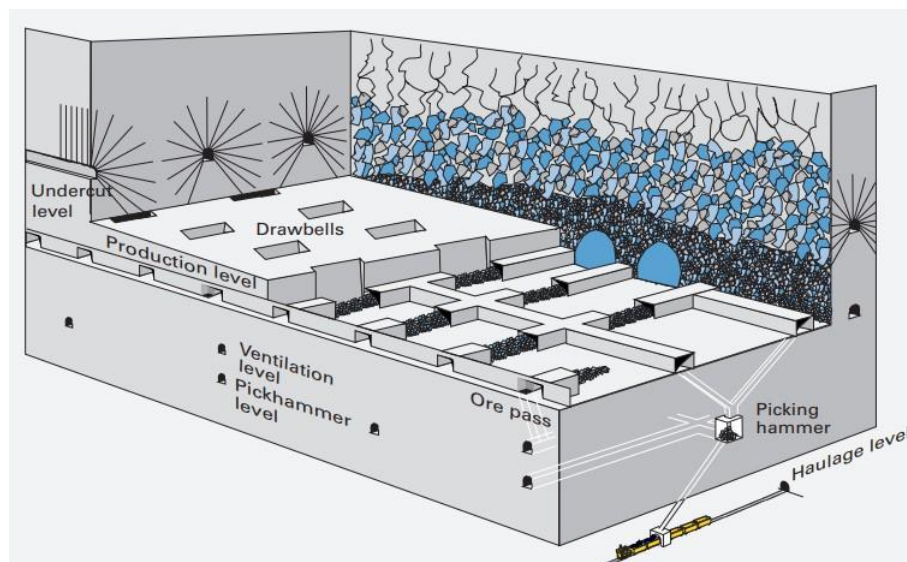


Figura 3.1: Esquema tradicional de *block caving*. (Atlas Copco, 2007).

La principal diferencia entre *block* y *panel caving* radica en la incorporación de nuevas áreas de producción. Mientras el *block caving* desarrolla la explotación por bloques discretos, el *panel caving* lo hace mediante paneles continuos, favoreciendo una explotación

más flexible y continua del yacimiento. Además, el primero se utiliza principalmente en mineral secundario y el segundo en mineral primario (Vienne, 2005).

3.1. Flujo gravitacional

Como se mencionó anteriormente, los métodos *caving* dependen de la fragmentación del macizo rocoso y del flujo gravitacional del material quebrado, lo que permite su recuperación mediante puntos de extracción ubicados en el nivel de producción. En este contexto, la teoría de flujo gravitacional describe el comportamiento del material granular durante su extracción, considerando la formación y evolución de una zona de movimiento generada desde un punto de extracción en la base del volumen fragmentado (Kvapil, 2008). Como se ilustra en la Figura 3.2, esta teoría distingue dos volúmenes concéntricos: el elipsoide de extracción (E_N), que corresponde al material que efectivamente es recuperado por el punto de extracción, y el elipsoide de aflojamiento (E_G), una región mayor en la que los fragmentos se desplazan y sueltan sin ser necesariamente extraídos. En la actualidad, los conceptos anteriores han sido remplazados por zona de extracción aislada (IEZ) y zona de movimiento aislada (IMZ).

El comportamiento del material quebrado está condicionado por la interacción de distintas fuerzas, entre ellas las desarrolladas entre fragmentos, las que surgen entre los fragmentos y la roca in situ en las proximidades de la batea, y la acción gravitacional. El movimiento del material corresponde, por tanto, a un balance dinámico de fuerzas que evoluciona a medida que progresa la extracción desde la base del bloque.

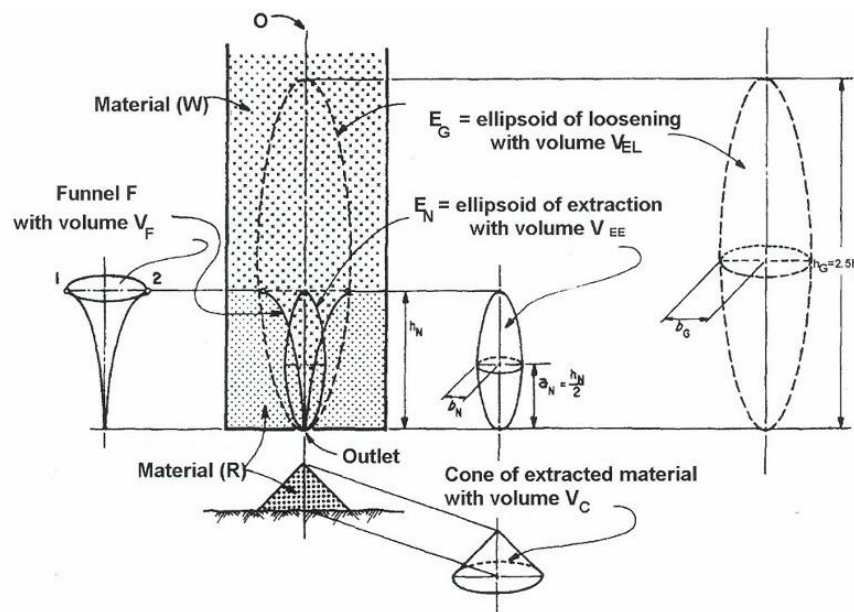


Figura 3.2: Teoría Kvapil. (Kvapil, 2008).

La base teórica del flujo gravitacional se estableció a partir del modelo de elipsoides descrito anteriormente. Su estudio evolucionó hacia métodos numéricos que se distinguen por cómo obtienen el movimiento del material.

El método de elementos discretos (DEM) resuelve las fuerzas de contacto entre partículas e integra su movimiento paso a paso. Captura la migración de finos y, mediante

leyes de contacto adicionales, el redondeamiento asociado a la fragmentación secundaria. Su costo computacional restringe su aplicación a volúmenes acotados.

Los autómatas celulares (AC) reproducen el flujo mediante la migración ascendente de vacíos, con dirección determinada por simulación de Monte Carlo. Operan a escala de mina con bajo costo computacional y son la base de herramientas tridimensionales como FlowSim (Castro and Whiten, 2007). El modelo requiere calibración con datos experimentales para reproducir las geometrías de las zonas de movimiento.

Los emuladores incrementales aplican reglas de crecimiento a escala de capa. REBOP simula la evolución de las zonas de movimiento aisladas (IMZ) mediante dilatación, colapso y erosión, a partir de patrones observados en DEM y operando sobre discos. La Figura 3.3 muestra la evolución del perfil de descenso para tonelajes crecientes: la IMZ se propaga en altura a medida que avanza la extracción, y el descenso del material se hace visible en la deformación progresiva de las capas horizontales. Pierce (2010) reformula estas reglas en términos de propiedades del material: diámetro medio de fragmento, porosidad y ángulo de fricción. Incorpora migración de finos y fragmentación secundaria, y entrega tonelajes y leyes a partir de la secuencia de extracción, con la dilución como resultado del flujo. El mismo autor documenta que el código no representa el doblado de los ejes de las IMZ bajo tiraje traslapado, y que sobrestima tanto el ancho de la IMZ en campo cercano para fragmentación gruesa como el grado de fragmentación secundaria.

El estudio del flujo continúa en desarrollo, con trabajos experimentales que caracterizan mecanismos específicos. Aravena et al. (2025) cuantifican la migración de finos en un modelo físico y muestran que el contenido de fino y la inclinación de la interfaz estéril/mineral son las variables dominantes, mientras que la extracción interactiva retrasa su ingreso respecto de la aislada.

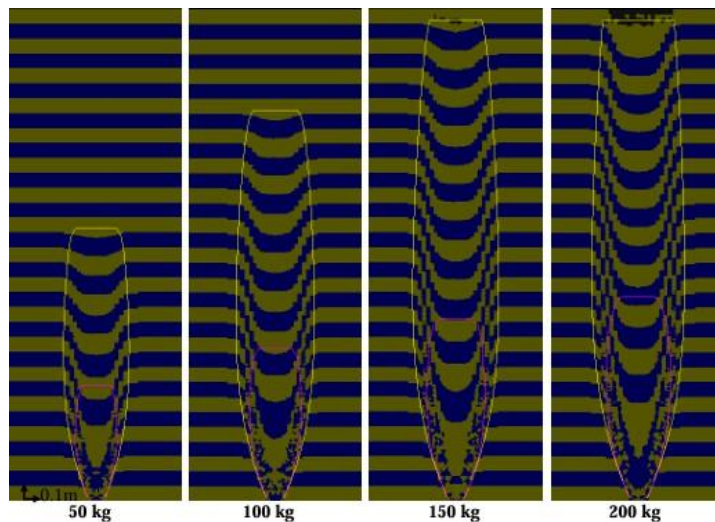


Figura 3.3: Evolución de la zona de movimiento aislada (IMZ, amarillo) y de la zona de extracción aislada (IEZ, magenta) según tonelaje extraído en simulación REBOP. Adaptado de Pierce (2010).

3.2. Dilución

La dilución corresponde al material sin valor económico que se mezcla con el mineral extraído y reduce su ley media. Su ingreso incrementa el tonelaje que debe procesarse, disminuye la ley promedio del material enviado a planta y altera sus propiedades geometa-lúrgicas. Por ello, para planificar la extracción es necesario considerar el comportamiento del material una vez fragmentado, dado que la ley que efectivamente llega a un punto de extracción depende del proceso de flujo y mezcla dentro de la columna quebrada y, en general, no coincide con la ley del bloque que ocupaba esa posición antes del quiebre (Laubscher, 1994).

Es posible distinguir entre dilución planificada y no planificada. La primera corresponde al material estéril que, encontrándose fuera de la definición económica de mineral, se incorpora previamente como parte del diseño minero. La dilución no planificada, u operativa, es aquella que se extrae por sobre las reservas definidas y cuya conveniencia económica depende de su contenido de material.

A partir de un *back-analysis* de datos operacionales de las divisiones El Salvador y Andina de Codelco, Castro and Paredes (2014) revisan los mecanismos de ingreso de dilución en *panel caving*, complementando el modelo de flujo gravitacional de Laubscher. Identifican tres mecanismos según la posición de la fuente de estéril y la forma en que esta alcanza la columna de tiraje (Figura 3.4). La dilución vertical ocurre cuando la fuente se encuentra sobre el panel en explotación e ingresa por descenso gravitacional, con percolación adicional del material fino; presenta punto de entrada tardío y recuperaciones altas. La dilución lateral se genera cuando la fuente es adyacente al panel: el *caveback* propaga preferentemente hacia la interfaz lateral con el material quebrado y, si el *air gap* es suficientemente grande, el estéril ingresa a las columnas y se desplaza horizontalmente según la pendiente de la pila, con punto de entrada temprano y migración de grandes distancias. La dilución por hundimiento ocurre cuando el radio hidráulico del área abierta es insuficiente para sostener la propagación continua: se produce un colgamiento, se acumula un *air gap* de gran extensión y, al incorporarse nueva área, el *caveback* propaga de forma súbita y el estéril ingresa a los primeros puntos abiertos tras un estallido de aire. En este caso el punto de entrada y la recuperación son bajos, como documentan en el sector Inca West de El Salvador pese a una uniformidad de tiraje de 84 %.

Los dos últimos mecanismos comparten origen operacional. Los autores obtienen una relación entre la tasa de extracción y la velocidad de propagación del hundimiento, evidenciando que el *air gap* se forma cuando la primera supera a la segunda, y una segunda relación entre la pendiente local de la pila entre bateas contiguas y el ángulo de fricción del material diluyente, que determina si el estéril desliza y se desplaza lateralmente. Concluyen que las condiciones de borde y la propagación del hundimiento dominan la planificación de largo plazo por sobre la uniformidad de tiraje de corto plazo, y que la estrategia de tiraje y el manejo del *caveback* factores para el control de dilución.

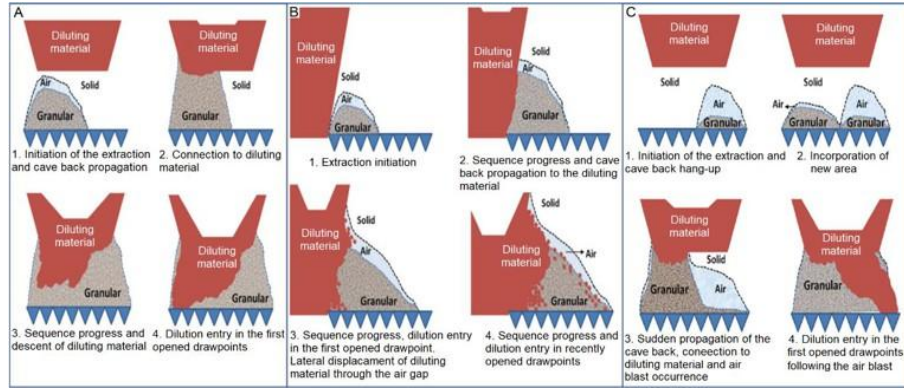


Figura 3.4: Representación esquemática de los mecanismos de dilución: (a) dilución gravitacional, (b) dilución lateral a través del espacio de aire y (c) dilución por *caving*. Adaptado de [Castro and Paredes \(2014\)](#).

La estimación de la dilución se apoya principalmente en modelos empíricos. El más utilizado es el propuesto por Laubscher, que define el punto de entrada de dilución (PED) como el porcentaje de extracción de la columna en el cual el material diluyente aparece por primera vez en el punto de extracción ([Laubscher, 1994](#)), como se muestra en la Ecuación (3.1).

$$PED = \frac{H_c - \frac{HIZ}{s}}{H_c} \times dcf \times 100 \quad (3.1)$$

donde:

H_c : altura de la columna de mineral.

HIZ : altura de la zona de interacción.

s : factor de esponjamiento.

dcf : factor de control de tiraje, que ajusta el resultado según la uniformidad de la extracción en los puntos vecinos: si el tiraje es homogéneo, se aproxima a uno y la entrada de dilución se retrasa.

El modelo es lineal, de cálculo rápido e incorporable en modelos de planificación. Asume, sin embargo, un PED fijo a lo largo de toda la extracción. No depende de la secuencia de explotación ni de la propagación del caving, y no captura el mecanismo de dilución lateral.

[Susaeta \(2004\)](#) caracteriza la dilución a partir de la condición de flujo de cada punto de extracción. Define el grado de interacción $G_i = v_{ti}/v_{ta}$, donde v_{ta} es la velocidad de descenso sobre el punto de extracción y v_{ti} la del material que sobreyace el pilar mayor. La Figura 3.5 identifica ambas velocidades y delimita la zona de flujo interactivo en la base de la columna.

Los modos de flujo son casos límite de G_i . La igualdad de ambas velocidades produce flujo interactivo. La anulación de v_{ti} produce flujo aislado. El caso intermedio define el flujo aislado-interactivo. Cada modo tiene una curva de dilución asociada, descrita por dos

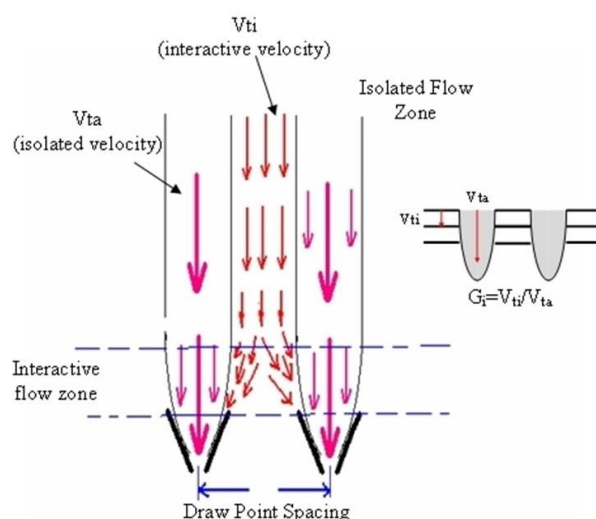


Figura 3.5: Modelo de flujo aislado-interactivo. [Susaeta \(2004\)](#).

puntos de entrada. El PEDZA es el porcentaje de extracción en que la dilución aparece. El PEDZI es el porcentaje en que la curva cambia de pendiente y la dilución se acelera.

La transición entre modos depende de la uniformidad del tiraje. [Susaeta \(2004\)](#) la mide mediante el índice de uniformidad (IU), que compara el tonelaje extraído de un punto con el de sus vecinos e incorpora los vecinos inactivos mediante un término aditivo, a diferencia del factor de control de tiraje de Laubscher. El IU clasifica la condición de tiraje de cada período a partir del número de vecinos inactivos y del índice de uniformidad específico. Agregando esa clasificación sobre la vida del punto se obtiene el CUI, correspondiente al cociente entre el tonelaje extraído en períodos uniformes o semi-uniformes y el tonelaje de la columna económica, medido hasta el 100 % de extracción.

[Susaeta et al. \(2008\)](#) cuantifican su efecto mediante un *back-analysis* de ocho sectores de Codelco. En Andina LHD, el PEDZA pasa de 32 % a 71 % de extracción cuando el CUI sube desde el rango 53–69 % a valores sobre 79 %, manteniendo constante la geometría.

El enfoque incorpora la uniformidad del tiraje como variable de control. Los puntos de entrada se obtienen de curvas observadas y no se predicen. La curva aplicable se selecciona con un indicador medido sobre la vida completa del punto. Su uso en planificación exige suponer la uniformidad que la propia secuencia determina.

Los modelos empíricos son lineales y de cálculo rápido, e incorporables al proceso de planificación, pero asumen un comportamiento fijo de la dilución. Los modelos de simulación representan la dependencia entre tiraje y dilución, pero su costo los vuelve poco prácticos para la toma de decisiones secuencial. Ambos estiman la dilución asociada a una secuencia dada.

3.3. Estado de las columnas

Una columna de extracción atraviesa distintos estados a medida que avanza el hundimiento. Antes del paso del nivel de socavación permanece sin socavar. Una vez socavada entra en quiebre, proceso durante el cual el macizo se fractura progresivamente hacia arriba. Cuando el fracturamiento alcanza el nivel de producción, la columna queda liberada y su punto de extracción puede abrirse para iniciar el tiraje. La transición entre estados no depende solo del tiempo: la extracción del material esponjado es condición para que el

quiebre progrese, de modo que el propio tiraje gobierna el avance de la columna hacia su estado siguiente.

El umbral de liberación se establece empíricamente mediante una relación 1:3 entre metros de columna extraídos y metros de columna quebrados, lo que sitúa la liberación aproximadamente el 30 %. Por ejemplo, los criterios de diseño del Proyecto Mina Chuquicamata Subterránea sitúan en 25 % de extracción la conexión del *caving* a superficie o al material quebrado del nivel superior (Codelco, 2009).

La velocidad de extracción, expresada en toneladas por día por metro cuadrado de área, determina cuánto material puede retirarse de un punto por unidad de área y tiempo. Está gobernada por la propagación del hundimiento, la condición de esfuerzos y el tiraje local, de modo que su límite admisible depende del estado alcanzado por la columna. Durante el quiebre la restricción la impone la propagación del *caving*. Una vez liberada la columna, la restricción pasa a ser la capacidad del sistema de manejo de materiales. Los criterios de diseño del Proyecto Mina Chuquicamata Subterránea ilustran esa diferenciación: 0.20 t/m²-día durante el período de hundimiento, 0.39 t/m²-día hasta alcanzar el 25 % de extracción de la columna, y 0.60 t/m²-día para columnas conectadas o liberada (Codelco, 2009).

El cambio de estado modifica, por tanto, el conjunto de decisiones disponibles. Extraer hasta el umbral de liberación habilita una velocidad de extracción mayor para los períodos siguientes. Las restricciones operacionales de cada período dependen así de las decisiones tomadas hasta ese momento.

3.4. Respuesta geomecánica

El ángulo de extracción (ϑ_1) se define como el ángulo entre la última batea incorporada al frente de extracción y el último punto en proceso de quiebre (Cornejo and Pardo, 2014) (Figura 3.6). Relaciona el avance vertical del quiebre (v_q) con el avance horizontal del frente (v_a) mediante:

$$\tan \vartheta_1 = \frac{v_q}{v_a}. \quad (3.2)$$

La relación es lineal. Todo aumento en la velocidad de extracción en quiebre debe acompañarse de un aumento en la velocidad de incorporación de área para mantener el ángulo estable (Cornejo and Pardo, 2014). Empíricamente el ángulo se mantiene entre 30° y 45°, rango que se asocia a una concentración de esfuerzos más favorable en el frente de hundimiento y a un menor riesgo sísmico. Las desviaciones en ambas direcciones son desfavorables. Una incorporación de área acelerada respecto de la extracción acuesta el frente y genera mayor concentración de esfuerzos en el frente de socavación. Una extracción que avanza más rápido que la incorporación eleva el ángulo, lo que puede adelantar la entrada de dilución o de agua-barro y forzar el cierre prematuro de puntos.

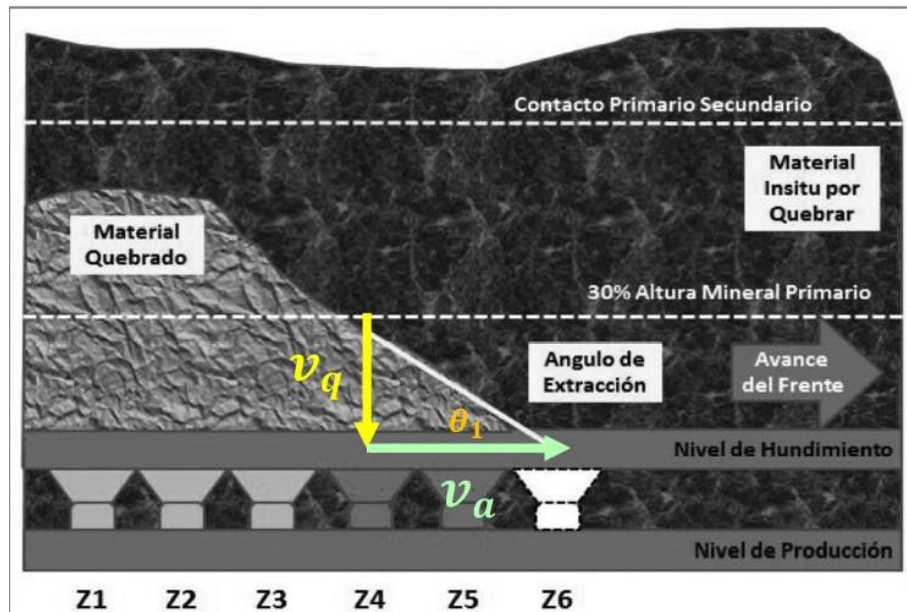


Figura 3.6: Definición geométrica del ángulo de extracción (ϑ_1) como relación entre el avance vertical del quiebre y el avance horizontal del frente. Adaptado de [Cornejo and Pardo \(2014\)](#).

El avance del frente modifica de forma continua el campo de esfuerzos del macizo, en magnitud y en orientación, con efectos directos sobre los puntos de extracción ([Oyanader Willson, 2015](#)). Los puntos construidos por delante del frente quedan expuestos al *abutment stress* y registran mayor deterioro en sus viseras, mientras que los ubicados bajo la sombra del frente se encuentran en una zona de relajación.

La respuesta sísmica acompaña a este proceso, [de Beer \(2022\)](#) cuantifica la correlación entre tasas de proceso y tasa de liberación sísmica en un panel caving, y reporta que la tasa de extracción actúa con correlación sostenida sobre un período largo y un desfase significativo respecto de la respuesta sísmica. La tasa de tronadura de producción presenta mayor correlación, sobre un período más corto y con menor desfase, en el socavón y bajo él. El autor atribuye esa diferencia a la condición de *abutment* esforzado de esa zona, asociados a pilares y a contrastes de rigidez entre litologías.

La relación entre tiraje y sismicidad no es directa. Un ángulo bajo mantiene un área extensa en quiebre y sin liberar, donde se concentra la sismicidad inducida por el proceso de fracturamiento. El ángulo opera así como variable de control: el tiraje afecta la respuesta sísmica a través de la geometría del frente que resulta de él. Los riesgos geomecánicos asociados a la operación de *caving* se resumen en la Figura 3.7.

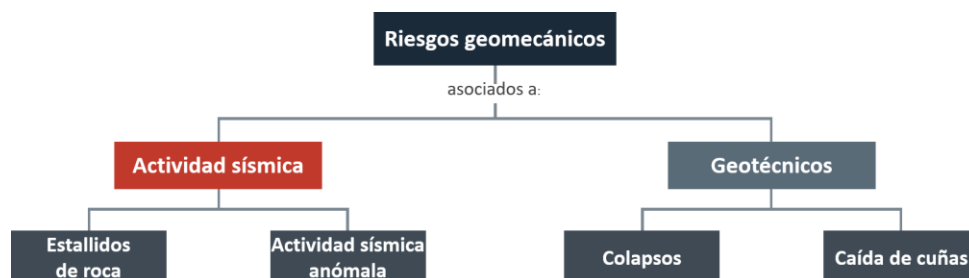


Figura 3.7: Riesgos geomecánicos en operaciones de *caving*. Adaptado de [Oyanader Willson \(2015\)](#).

La posición del frente en cada período resulta de las decisiones de extracción e incorporación de área acumuladas hasta ese momento. La concentración de esfuerzos sobre los puntos de extracción y la respuesta sísmica del sector quedan determinadas, por esa vía, por la secuencia seguida y no solo por las condiciones iniciales.

3.5. Estado del arte en planificación de *block caving*

La planificación en minería *caving* tiene como objetivo definir qué fracción del yacimiento será explotada y establecer la secuencia en que se realizará su extracción. Como se mencionó anteriormente, la configuración general de estos métodos considera distintos niveles de desarrollo, donde una vez generado el hundimiento, el mineral fragmentado es recuperado mediante los puntos de extracción. En este contexto, la planificación puede dividirse en dos etapas principales: la determinación de la envolvente económica del proyecto, y la definición de la secuencia de extracción.

3.5.1. Envolvente económica

Herramientas comerciales como GEOVIA PCBC abordan esta definición en dos etapas. La primera corresponde al módulo *Footprint Finder*, que opera sobre el modelo de bloques geológico junto con los costos mineros y factores de recuperación. El programa recorre cada nivel del modelo construyendo columnas verticales que acumulan valor económico, y aplica a cada columna una mezcla vertical basada en el método de Laubscher (1994). Repitiendo el procedimiento en todos los niveles se obtiene una curva de tonelaje y valor económico en función de la elevación (Figura 3.8), que entrega una primera aproximación a la cota del nivel de hundimiento y a la forma del polígono (Villa, 2014). Cuando la curva presenta más de un máximo local, la elección del nivel queda sujeta al criterio del planificador.

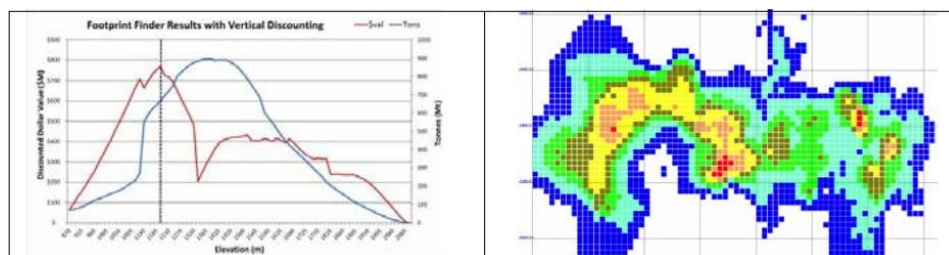


Figura 3.8: Resultados de Footprint Finder: valor económico y tonelaje por elevación (izquierda), y distribución del valor económico por columna (derecha). (Villa, 2014).

Una vez definida la elevación, se genera el arreglo de puntos de extracción y se construye el *slice file*: sobre cada punto se levanta una columna según una forma de cono definida, que se intersecta con el modelo de bloques resolviendo los solapamientos entre conos vecinos para no contabilizar dos veces el mismo material (Figura 3.9). Sobre estas columnas se recalcula la altura óptima de tiraje de cada punto, procedimiento conocido como *Best Height of Draw*, que refina el resultado anterior al incorporar las ubicaciones reales de los puntos y criterios operacionales adicionales (Diering et al., 2010).

El polígono resultante debe además cumplir requerimientos geométricos: un área mínima y una forma convexa que resulten favorables para la propagación del *caving* y la

redistribución de esfuerzos. El área mínima se estima habitualmente mediante el gráfico de estabilidad de Laubscher, que entrega el radio hidráulico requerido en función de la clasificación ajustada del macizo rocoso (Laubscher, 1994, 2000). Cumplir estos requerimientos puede implicar incorporar columnas de valor económico negativo o excluir columnas rentables, ajuste que se realiza de forma manual sobre el resultado del cálculo.

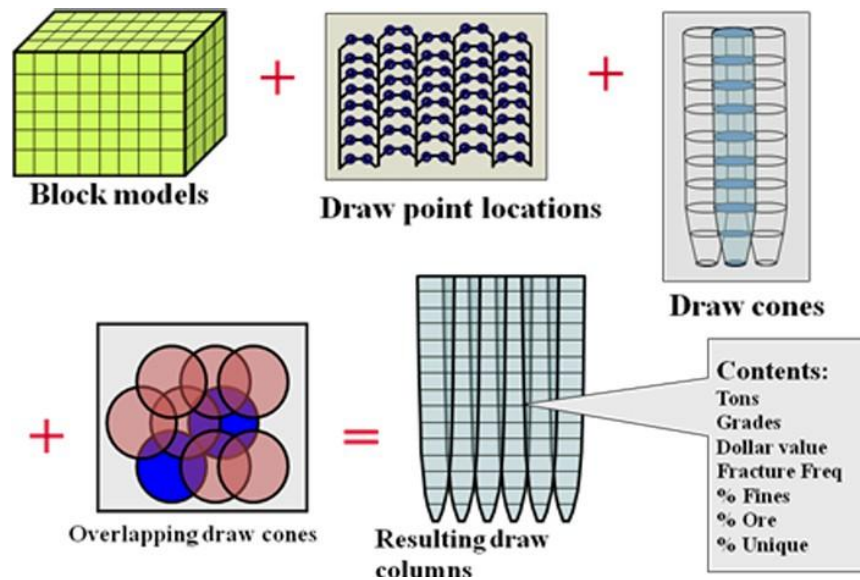


Figura 3.9: Esquema de conversión del modelo de bloques a slice file. (Diering et al., 2010).

Noriega et al. (2018) abordan una limitación distinta del procedimiento anterior. Dado que cada columna se abre en un período determinado según la dirección de avance y el punto de partida en el *footprint*, el beneficio percibido por su extracción depende del período en que la columna entra en producción. Los autores sostienen, en consecuencia, que la elevación de socavación no puede evaluarse sin incorporar ese descuento horizontal.

Su marco de programación entera opera en dos etapas sobre cada elevación candidata. La primera define el *layout* maximizando el contenido metálico mediante la agregación de columnas del modelo de bloques en unidades representativas de una zona de tiraje, sujeta a adyacencia, ausencia de islas y una altura mínima de columna. La segunda programa la extracción de esas unidades maximizando el beneficio descontado (Figura 3.10), con precedencias verticales dentro de cada columna y una precedencia horizontal que impone un frente de avance cóncavo, definido por el azimut de la dirección principal y dos ángulos diagonales de cara. El procedimiento se repite sobre las elevaciones candidatas y se selecciona la de mayor VPN.

Aplicado a un pórfido cuprífero de 58.5 Mt sobre dieciséis elevaciones, con azimut de 290° y ángulos de cara de 45°, el modelo identifica la elevación óptima con una brecha de 5 % y tiempos de 30 a 60 minutos por nivel. El *layout* resultante comprende 64 unidades de columna, equivalentes a 128 puntos de extracción. Un análisis de sensibilidad sobre el precio del metal muestra que la elevación seleccionada se mantiene en cuatro de los cinco escenarios evaluados, desplazándose un nivel únicamente ante una caída del 20 % en el precio. La forma del contorno, en cambio, no está sujeta a restricción: la primera etapa maximiza contenido metálico sin criterios de suavidad, la misma carencia que los autores atribuyen a los métodos que revisan.

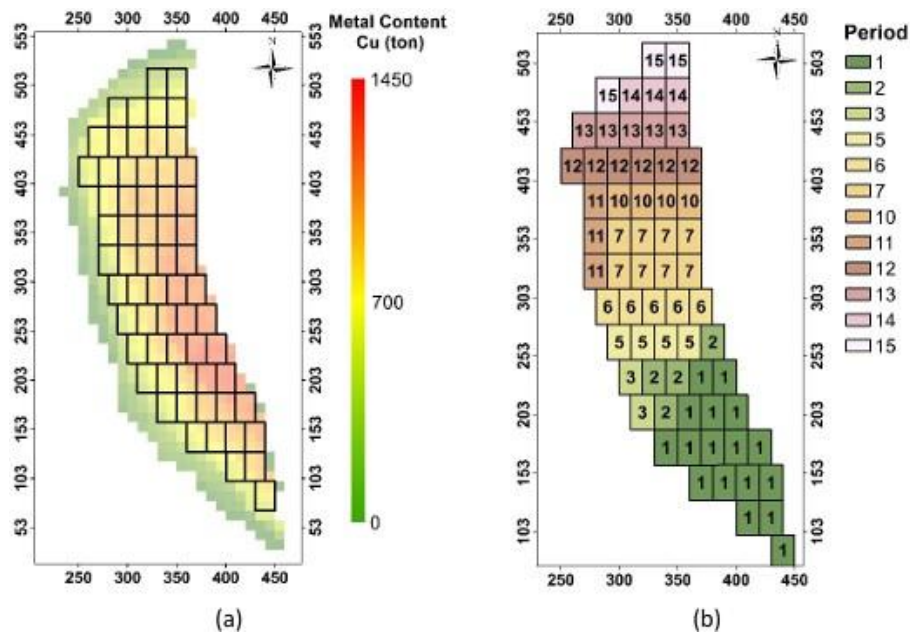


Figura 3.10: Selección del nivel de hundimiento y definición de la secuencia de extracción. (Noriega et al., 2018).

Saavedra et al. (2023) identifican esta práctica como una limitación. El algoritmo de PCBC evalúa cada columna de manera independiente, de modo que la envolvente resultante no incorpora restricción geométrica alguna sobre la forma del *footprint* ni sobre la regularidad de las alturas de columna. El resultado son frentes de extracción irregulares, columnas aisladas y zonas que no alcanzan el radio hidráulico crítico, características que condicionan la respuesta del macizo por cuanto la geometría del frente y del nivel de extracción opera como parámetro de control del esfuerzo inducido. La regularización de la envolvente queda entonces condicionada a un suavizado manual posterior.

Para resolverlo, los autores reformulan el problema como una instancia del *ultimate pit* invertido, el mismo problema que resuelven los algoritmos de Lerchs–Grossmann o *pseudoflow* en rajo abierto, e incorporan tres familias de arcos de precedencia. Los arcos verticales aseguran la conectividad de la columna y una altura mínima. Los horizontales, contruidos a partir de un conjunto de puntos iniciales y una función de forma, garantizan la conectividad del cuerpo y regulan la suavidad del contorno en planta. Los diagonales acotan la diferencia de altura entre columnas vecinas por sobre una altura de interacción. Esta última restricción incorpora dentro de la formulación una regularización que hasta entonces se ejecutaba como posprocesamiento sobre la solución ya calculada. La formulación conserva la estructura de cierre máximo y se resuelve hasta optimalidad exacta, sin recurrir a heurísticas ni a algoritmos aproximados.

En su caso de estudio sobre un pórfido cuprífero, los autores comparan la solución del PCBC, su suavizado manual y la envolvente generada por el método propuesto. Frente al suavizado manual, la envolvente automatizada alcanza 2402 MMUSD contra 2392 MMUSD, con cerca de un 9 % menos de tonelaje total y un 12 % menos de estéril. La Figura 3.11 contrasta ambas soluciones: la vista en planta muestra la diferencia en conectividad y regularidad del contorno, y la vista isométrica el efecto sobre la uniformidad de las alturas de columna.

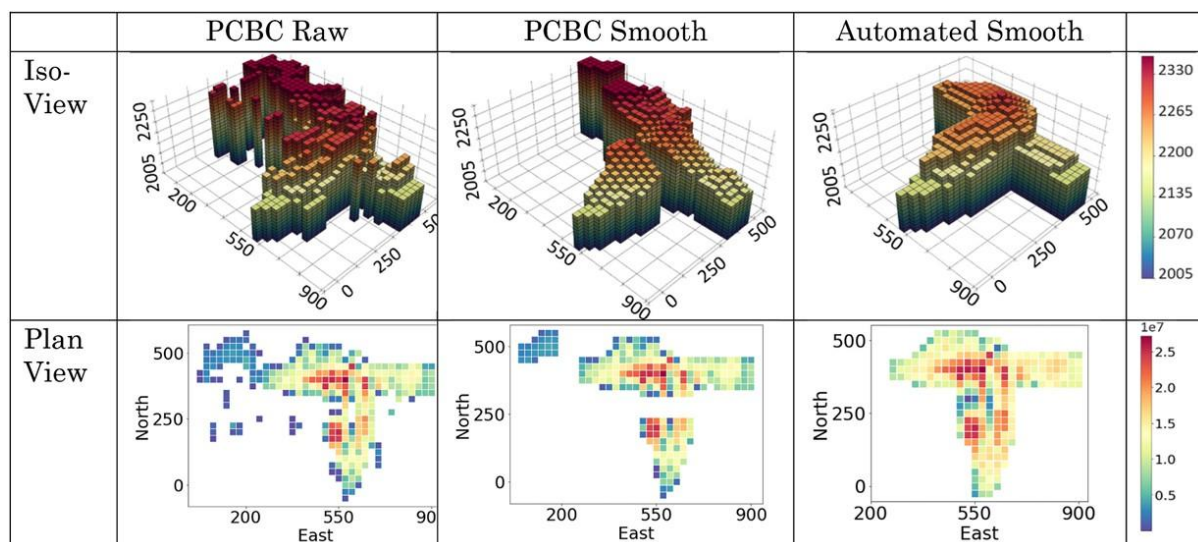


Figura 3.11: Comparación de envolventes económicas: PCBC Raw, PCBC Smooth y Automated Smooth. Vista isométrica (arriba) y en planta (abajo). (Saavedra et al., 2023).

3.5.2. Secuenciamiento

Una vez definida la envolvente económica, el siguiente paso consiste en planificar el orden y ritmo de extracción dentro de esa zona, considerando cuándo, dónde y cuánto material será extraído. En esta etapa se deben incorporar variables y restricciones operativas tales como las capacidades productivas, limitaciones de desarrollo y condiciones geomecánicas.

Enfoques heurísticos

GEOVIA PCBC adapta al *block caving* el concepto de *best* y *worst case* empleado en minería de superficie, con el propósito de acotar el valor alcanzable por una secuencia de extracción. Sobre la base de construidas por el *Footprint Finder*, las columnas se ordenan de mayor a menor valor económico para construir la secuencia *best case*, y en orden inverso para la *worst case*, incorporando el efecto del tiempo mediante una tasa de descuento anual (Villa, 2014).

Ambas secuencias son no operacionales: responden únicamente al orden de valor económico y no consideran la geometría del frente, las precedencias entre puntos vecinos ni el ritmo de apertura. Su función es servir de referencia para evaluar secuencias operacionales construidas por el planificador. En un caso de estudio sobre un pórfido cuprífero sintético, las secuencias operacionales evaluadas sobre el *footprint* completo, definidas por distintas formas de frente y puntos de inicio, alcanzan entre un 75 % y un 85 % del valor de la secuencia *best case*.

La herramienta permite además comparar estrategias de secuenciamiento estructuralmente distintas sobre un mismo yacimiento. La Figura 3.12 contrasta el secuenciamiento del *footprint* completo con su división en dos zonas, Este y Oeste, cada una con secuencia propia, evaluadas sobre la misma escala de valor (Villa, 2014).

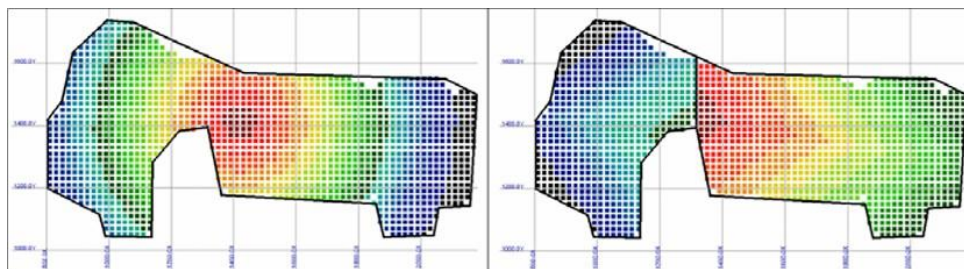


Figura 3.12: Comparación entre el secuenciamiento del *footprint* completo y su división en dos zonas, Este y Oeste. (Villa, 2014).

El procedimiento tiene bajo costo computacional y se aplica de forma iterativa, generando y evaluando secuencias candidatas contra las cotas. Su alcance es de evaluación. Las cotas delimitan el rango de valor, pero no indican en qué punto de ese rango se encuentra la mejor secuencia factible ni entregan un mecanismo para aproximarse a ella. A esto se suma que el valor económico de cada columna se calcula de forma independiente y previa al ordenamiento, de modo que la secuencia afecta el momento en que se percibe ese valor, pero no la ley ni el tonelaje que efectivamente reporta cada columna.

Programación matemática

La programación matemática, particularmente la Programación Lineal Entera Mixta (MILP), constituye el enfoque exacto más utilizado para la planificación de la producción en minería *caving*. En términos generales, estos modelos combinan dos tipos de variables de decisión principales:

- Variables binarias, que indican si el drawpoint i se activa en el período t .
- Variables continuas, que representan la cantidad de toneladas extraídas desde el drawpoint i en el período t .

Pourrahimian and Askari-Nasab (2012) proponen un marco MILP para la programación de producción en *block caving*. La formulación maximiza el VPN y emplea tres variables de decisión por unidad y período: una continua, que representa la porción de la columna a extraer en cada período, y dos binarias que indican si el punto de extracción está activo en el período y si su extracción se inicia en él. La unidad de decisión es el punto de extracción. La precedencia se impone únicamente en dirección horizontal: el conjunto de predecesores de cada punto se define geométricamente antes de optimizar, como los vecinos inmediatos que quedan detrás de una línea perpendicular a la dirección de avance. La variable binaria de inicio no puede activarse hasta que esos predecesores hayan alcanzado la tasa mínima de tiraje. La construcción admite ocho conjuntos distintos, uno por dirección de avance.

El modelo incorpora restricciones de capacidad de extracción, tasa de extracción, número máximo de puntos activos, número de puntos nuevos por período, continuidad de la extracción y agotamiento completo de reservas. Para controlar el tamaño del problema, los autores agregan los puntos de extracción en *clusters* mediante *fuzzy c-means*, agrupando por ubicación física del punto y tonelaje de su columna. La agregación supone que la porción programada de cada *cluster* se extrae de todos sus puntos en proporción al tonelaje de cada columna, de modo que dentro del *cluster* no existe decisión de secuencia. Sobre un caso real de 102 puntos y 15 períodos, las variables binarias se reducen de 3,060

a 510 y el tiempo de cómputo entre un 97 % y un 99.9 % según la dirección, con diferencias de VPN inferiores al 1 % en la dirección más favorable. Entre las cuatro direcciones de avance reportadas, la diferencia de VPN entre la más y la menos favorable alcanza un 2.5 % a nivel de punto de extracción.

No obstante, una limitación operativa de esta formulación, reconocida por los propios autores como trabajo futuro, era la gestión de la tasa de extracción mediante límites estáticos (tasas constantes como cotas superior e inferior). Si bien esto definía un rango de operación, generaba ambigüedad, ya que la optimización podía resultar en un perfil de extracción errático dentro de dichos límites, sin asegurar un sentido operacional realista. [Nezhadshahmohammad et al. \(2017\)](#) retoman esta limitación y mantienen el mismo esquema de optimización de la formulación anterior: maximización del VPN y las mismas tres variables de decisión por punto y período. La diferencia está en el tratamiento de la tasa extracción. En lugar de fijarla con un valor mínimo y uno máximo constantes, el modelo describe cómo debe variar a lo largo de la vida del punto. La tasa parte baja, sube hasta un régimen estable y luego decae (Figura 3.13). Ese comportamiento se traduce en restricciones que la optimización cumple período a período según el agotamiento acumulado de la columna, conservando la misma formulación lineal entera. La formulación mantiene la precedencia horizontal de inicio entre puntos adyacentes e incorpora restricciones de capacidad, ley, número de puntos activos y nuevos, y vida máxima del punto. La forma de la curva y sus quiebres corresponden a datos de entradas. La optimización resuelve cuándo y cuánto extraer sobre ese perfil. Sobre un caso de 298 puntos de extracción y quince períodos, los perfiles obtenidos por punto reproducen la secuencia de inicio a tasa mínima, ascenso gradual, permanencia en régimen estable y descenso posterior.

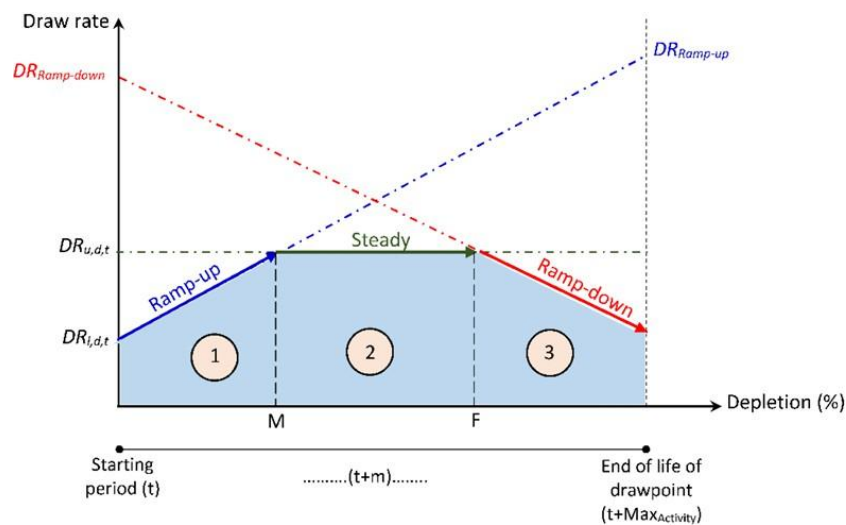


Figura 3.13: Estado de punto de extracción. ([Nezhadshahmohammad et al., 2017](#)).

[Smoljanovic et al. \(2011\)](#) desplazan el problema desde la programación de la extracción sobre una secuencia fija hacia la optimización de la secuencia de apertura misma. La formulación maximiza el VPN y emplea dos variables por punto y período: una binaria que indica si el punto está abierto y una continua que acumula la fracción de columna extraída. Sobre esas variables se definen dos incrementos por período, la apertura y la extracción, que alimentan las restricciones de capacidad de mina, de calle de producción y de punto individual, esta última operando como control de tasa de extracción. La geometría se incorpora representando el nivel de producción como un grafo, cuyos nodos son los puntos

de extracción y cuyas aristas conectan puntos vecinos según una tolerancia de distancia. Sobre ese grafo se imponen tres condiciones. La conectividad obliga a que cada punto solo pueda abrirse si su predecesor en el camino hacia un punto de acceso ya está abierto, de modo que la explotación crece conectada y sin puntos aislados. El avance máximo limita cuántas posiciones puede progresar el frente a lo largo de ese camino entre períodos. El requisito de vecinos mínimos exige que, al abrir un punto, un número dado de sus vecinos se abra dentro de una ventana de períodos. A diferencia de las formulaciones que precálculan la precedencia como listas estáticas por dirección de avance, aquí las relaciones entre puntos vecinos quedan situadas en el tiempo mediante la ventana temporal y el avance máximo, lo que permite que los perfiles de ramp-up y ramp-down emerjan de la geometría en lugar de imponerse como cotas. El modelo se aplica a un caso de 332 puntos distribuidos en 20 calles de producción sobre 12 períodos, evaluando cada punto como acceso posible para comparar escenarios. El resultado central es la magnitud de la sensibilidad del VPN a las decisiones de secuenciamiento: la diferencia entre el mejor y el peor punto de inicio bordea el 50 %, mientras que variar los componentes del sistema minero, número de chancadores y de equipos LHD por calle, mueve el VPN en torno a un 5 a 6 %. Los autores concluyen que la ubicación del inicio y la geometría de la secuencia dominan el valor del proyecto por sobre la capacidad del manejo de materiales, y que la incorporación explícita de las capacidades de cada componente como restricción revela diferencias de VPN entre sistemas mineros que un modelo sin esa desagregación no distinguiría.

3.6. Implicancias para la planificación

El flujo gravitacional se ha estudiado con modelos físicos, elementos discretos y emuladores incrementales, y la dilución con modelos empíricos calibrados en faena. Ninguno de esos desarrollos se incorpora dentro de los modelos de secuenciamiento revisados. En todas las formulaciones la mezcla se resuelve en el preprocesamiento y el modelo opera sobre leyes ya diluidas. [Saavedra et al. \(2023\)](#) corrigen las leyes con el método de [Laubscher \(1994\)](#), que estima la mezcla vertical en cada columna según el porcentaje de entrada de la primera dilución, y descartan los métodos de modelación del flujo gravitacional porque exigen una secuencia previa. [Noriega et al. \(2018\)](#) omiten la mezcla y advierten que incorporarla requeriría una formulación no lineal.

Ambas decisiones responden a una misma circularidad. Estimar la mezcla exige conocer la secuencia. Secuenciar exige conocer la envolvente. Determinar la envolvente exige valores que dependen de la mezcla. [Diering et al. \(2010\)](#) lo reconocen al distinguir los métodos de mezcla lineales, que permiten calcular reservas antes del programa de producción, de los no lineales, en los que la reserva explotable es función de la secuencia y de la estrategia de tiraje. En estos últimos la mezcla debe incorporarse dentro del programador.

Dentro de los modelos la dilución interviene de manera indirecta, mediante restricciones que la invocan sin cuantificarla. Los límites a la vida activa del punto se fundan en que su prolongación incrementa la probabilidad de recompactación. El perfil de tasa de tiraje debe ser rápido para evitar recompactación y lento para evitar la entrada de aire y de dilución ([Nezhadshahmohammad et al., 2017](#)). Las restricciones geométricas de uniformidad asumen la función del *draw ratio*, mecanismo tradicional de control de la entrada de dilución ([Smoljanovic et al., 2011](#)). Ninguna formulación revisada entrega la dilución como variable de salida.

En operación, El Teniente limita a 40 m la diferencia de altura extraída entre columnas

vecinas, de modo que un tiraje más homogéneo retrasa la entrada de dilución (Bustamante et al., 2021).

La respuesta del macizo recibe igual tratamiento. Las restricciones sobre el contorno del *footprint* y la forma del frente se justifican por su efecto sobre el esfuerzo inducido (Saavedra et al., 2023; Noriega et al., 2018), pero ninguna formulación estima el estado tensional de la solución obtenida ni la respuesta sísmica asociada. El control de la dilución y de la concentración de esfuerzos se ejerce sobre descriptores geométricos de la extracción, que operan como sustitutos de los fenómenos que representan.

El límite del enfoque es que los descriptores se fijan antes de conocer el estado del sistema sobre el que actúan. Los cuatro fenómenos revisados en este capítulo evolucionan con la secuencia. El régimen de flujo de cada vecindad determina la mezcla y la ley recuperable en períodos posteriores (Sección 3.1). El punto de entrada de dilución depende de la altura de columna remanente y del régimen de flujo (Sección 3.2). El estado de cada columna fija la velocidad de extracción admisible y cambia al alcanzar el umbral de liberación (Sección 3.3). La concentración de esfuerzos y la respuesta sísmica dependen de la posición del frente, resultado de las decisiones de extracción e incorporación de área acumuladas (Sección 3.4).

Extraer consume reservas y modifica las condiciones bajo las cuales se decide después. Planificar en *block caving* es escoger una secuencia de decisiones sobre un sistema cuyo estado cambia por efecto de esas mismas decisiones. Esto motiva el uso de metodologías que estimen la evolución del estado mientras la secuencia se construye.

Capítulo 4

Aprendizaje por refuerzo

4.1. Inteligencia artificial y machine learning

La Inteligencia Artificial (IA) es el campo de la informática dedicado al diseño de agentes que perciben su entorno y actúan sobre él para alcanzar objetivos determinados, involucrando razonamiento, aprendizaje y toma de decisiones. Un subcampo central de la IA es el Machine Learning (ML), que dota a los sistemas computacionales de la capacidad de mejorar su desempeño en una tarea a partir de la experiencia, sin ser programados mediante reglas explícitas para cada caso. De acuerdo con [Russell et al. \(2004\)](#), el ML se organiza en tres paradigmas principales, según el tipo de retroalimentación disponible durante el aprendizaje:

- Aprendizaje supervisado: el modelo aprende una función que mapea entradas a salidas conocidas a partir de un conjunto de datos etiquetados, minimizando el error respecto de la salida conocida.
- Aprendizaje no supervisado: el modelo identifica estructuras, agrupaciones o patrones latentes en datos sin etiquetas ni una señal explícita de desempeño.
- Aprendizaje por refuerzo (RL): un agente aprende a tomar decisiones interactuando secuencialmente con un entorno, recibiendo una señal de recompensa y ajustando su comportamiento para maximizar la recompensa acumulada a lo largo del tiempo.

A diferencia del aprendizaje supervisado, donde cada ejemplo de entrenamiento indica la acción correcta, en RL el agente solo recibe una evaluación parcial y retrasada de sus decisiones, la recompensa, y debe descubrir por prueba y error qué secuencia de acciones la maximiza. Esta característica hace de RL el paradigma natural para problemas de decisión secuencial bajo incertidumbre, donde cada decisión afecta qué decisiones son válidas y convenientes en el futuro.

4.2. Aprendizaje por refuerzo

4.2.1. Proceso de decisión de Markov

El marco formal estándar para representar problemas de decisión secuencial bajo incertidumbre es el Proceso de Decisión de Markov (MDP), introducido por [Bellman \(1957\)](#) en el contexto de la programación dinámica. Un MDP se define mediante la tupla (S, A, P, R, γ) , donde:

- S : conjunto de estados posibles del entorno.
- A : conjunto de acciones que el agente puede ejecutar.
- $P(s'|s, a)$: probabilidad de transitar al estado s' al ejecutar la acción a en el estado s .
- $R(s, a)$: recompensa esperada de esa transición.
- $\gamma \in [0, 1)$: factor de descuento.

La propiedad de Markov establece que la transición y la recompensa dependen únicamente del estado y la acción actuales, y no de la historia completa que llevó a ese estado. Esta propiedad es lo que permite resolver el problema de forma recursiva mediante la ecuación de Bellman, en lugar de tener que considerar explícitamente cada trayectoria pasada.

La función de valor-acción óptima $Q^*(s, a)$ satisface la ecuación de optimalidad de Bellman:

$$Q^*(s, a) = E \left[R(s, a) + \gamma \max_{a'} Q^*(s', a') \right] \quad (4.1)$$

la cual expresa que el valor óptimo de tomar la acción a en el estado s corresponde a la recompensa inmediata esperada más el valor descontado del mejor estado-acción alcanzable desde s' .

Siguiendo la formalización [Sutton and Barto \(2015\)](#), en su texto de referencia estándar en RL, el problema se representa como un ciclo de interacción agente-entorno, tal como se presenta en la Figura 4.1. En cada paso discreto de tiempo t , el agente observa el estado s_t , selecciona una acción a_t según su política π , recibe una recompensa $r_t = R(s_t, a_t)$ y observa el nuevo estado s_{t+1} . El objetivo del agente es aprender una política π^* que maximice el retorno esperado descontado:

$$\pi^* = \arg \max_{\pi} E_{\pi} \left[\sum_{t=0}^{\infty} \gamma^t R(s_t, a_t) \right] \quad (4.2)$$

El factor de descuento γ garantiza que el retorno converja cuando el horizonte es infinito, y regula cuánto peso da el agente a las recompensas futuras frente a las recompensas inmediatas.

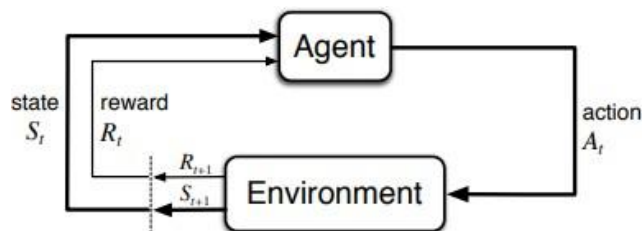


Figura 4.1: El ciclo de interacción agente-entorno en aprendizaje por refuerzo. ([Sutton and Barto, 2015](#)).

4.2.2. Funciones de valor y Q-learning

Para evaluar qué tan buena es una política, se define la función de valor-acción $Q^\pi(s, a)$ como el retorno esperado al ejecutar la acción a en el estado s y seguir la política π en adelante. La función de valor-acción óptima, $Q^*(s, a)$, satisface la ecuación de optimalidad de Bellman, y permite recuperar la política óptima simplemente eligiendo, en cada estado, la acción que maximiza Q^* .

Q-learning, propuesto originalmente por [Watkins \(1989\)](#) y formalizado con una prueba de convergencia por [Watkins and Dayan \(1992\)](#), es un método de diferencia temporal que aprende una aproximación de Q^* directamente a partir de la experiencia del agente, sin requerir un modelo explícito de $P(s'|s, a)$ (por eso se clasifica como un método *model-free*). Tras cada transición observada (s_t, a_t, r_t, s_{t+1}) , la tabla de valores se actualiza mediante la regla:

$$Q(s_t, a_t) \leftarrow Q(s_t, a_t) + \alpha [r_t + \gamma \max_{a'} Q(s_{t+1}, a') - Q(s_t, a_t)], \quad (4.3)$$

[Watkins and Dayan \(1992\)](#) demuestran que bajo el supuesto de que todo par estado-acción se visita un número infinito de veces y que la tasa de aprendizaje decae según un esquema apropiado (condición de Robbins-Monro: $\sum_t \alpha_t = \infty$ y $\sum_t \alpha_t^2 < \infty$), la Ecuación 4.3 converge a Q^* con probabilidad 1, independientemente de la política usada para explorar el entorno durante el entrenamiento (método *off-policy*).

4.2.3. Exploración y explotación

Dado que Q se aprende a partir de la experiencia, el agente enfrenta el dilema de exploración-explotación. Debe elegir entre ejecutar la acción que según su estimación actual, tiene mayor valor (explotación), o ejecutar una acción distinta para obtener información sobre partes del espacio de estados-acciones no haya visitado (exploración). [Sutton and Barto \(2015\)](#) describen la política ϵ -greedy como el mecanismo para resolver este dilema: con probabilidad ϵ el agente elige una acción al azar, y con probabilidad $(1 - \epsilon)$ elige la acción de mayor valor estimado según la tabla Q actual. Esta política de comportamiento (ϵ -greedy), utilizada para explorar el entorno, es distinta de la política objetivo (greedy respecto a Q) que el agente efectivamente está aprendiendo. Esto es consistente con la naturaleza *off-policy* de Q-learning descrita anteriormente, y es lo que permite que la exploración no comprometa la convergencia hacia Q^* . Es común decaer ϵ a lo largo del entrenamiento, comenzando de un valor alto que favorece la exploración inicial, hacia un valor bajo que busque la explotación una vez que la tabla Q se ha estimado con suficiente experiencia.

En conjunto, la tasa de aprendizaje α , el factor de descuento γ y el esquema de exploración ϵ constituyen los hiperparámetros centrales de cualquier implementación de Q-learning tabular.

4.2.4. Q-learning tabular versus Deep Q-Network

La implementación de RL puede realizarse mediante Q-learning tabular o mediante Deep Reinforcement Learning (DRL), donde la tabla de valores Q se reemplaza por una red neuronal que la aproxima. En este trabajo se utiliza el primero.

En términos generales, la diferencia principal radica en la representación del valor $Q(s, a)$. Mientras que Q-learning tabular almacena $Q(s, a)$ de forma exacta en una tabla con una entrada por cada par estado-acción, [Mnih et al. \(2015\)](#) propusieron el Deep

Q-Network (DQN), que lo aproxima mediante una red neuronal $Q(s, a; \theta)$. [Huang et al. \(2021\)](#) ilustran esta distinción al abordar la selección de relays en redes IoT con buffer (Figura 4.2). Dado que el espacio de estados crece exponencialmente con el tamaño del buffer y el número de relays, reemplazan la tabla Q por una red neuronal profunda, logrando reducir la complejidad espacial y temporal del método sin necesidad de almacenar explícitamente todos los pares estado-acción.

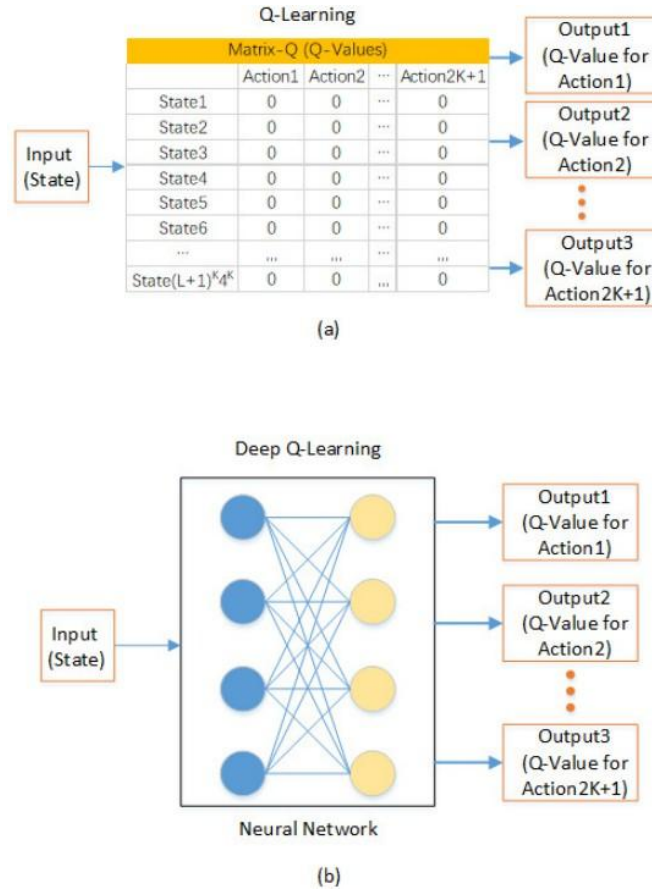


Figura 4.2: Comparación esquemática entre la tabla Q y la representación mediante una red neuronal profunda. ([Huang et al., 2021](#)).

4.3. Estado del arte en minería

La implementación del RL en la industria minera es de desarrollo reciente, y su adopción difiere considerablemente entre minería a cielo abierto y subterránea.

4.3.1. Minería a cielo abierto

En minería a cielo abierto las aplicaciones de RL se concentran en el despacho dinámico de flotas y en la optimización estocástica de complejos mineros. Los tres trabajos revisados difieren en el rol que asignan al agente: aprender la política de despacho de cada camión, seleccionar heurísticas dentro de un optimizador, o asignar la flota completa bajo incertidumbre.

Huo et al. (2023) formulan el despacho de flota en rajo abierto como un problema multi-agente resuelto con Q-learning tabular. Cada camión es un agente y el estado registra cinco variables: ubicación, estado de carga, ley del material cargado, largo de cola y necesidad de mantención. La acción es la próxima ubicación dentro de un conjunto discreto de destinos. La recompensa combina seis términos, entre ellos penalizaciones por regresar cargado a la pala, por cola larga y por quedar atrapado en ciclos entre dos ubicaciones. Los autores justifican el uso de una tabla de valores por el tamaño acotado del espacio de estados, que permite tasas de aprendizaje altas. La comparación es contra dos reglas de despacho fijas implementadas en el mismo simulador. El agente supera a la mejor de ellas en 43 % de producción por camión y reduce en 37 % las emisiones de CO₂ equivalente por tonelada, con una flota de cinco camiones.

Noriega et al. (2025) resuelven el mismo problema con Double Deep Q-Learning y una red neuronal como aproximador de la función de valor. El estado combina capacidades y distancias de los caminos pala-destino, avance de las metas de turno, productividad por pala y por destino, y ley del material en cada frente. La acción es un vector one-hot sobre los caminos disponibles. El entorno es una simulación de eventos discretos construida con un año de registros de equipos, cuyas distribuciones se seleccionan mediante pruebas de bondad de ajuste (Figura 4.3). La recompensa penaliza déficits de productividad en palas y destinos, y desviaciones de ley respecto al objetivo. Las líneas base son dos heurísticas de despacho: envío a la pala con menor cola y envío a la pala más rezagada respecto de su meta de turno. Los autores no reportan cifras porcentuales de mejora; la comparación es gráfica, con intervalos de confianza al 95 %. Declaran como limitación que la política aprendida puede requerir reentrenamiento para períodos de producción posteriores.

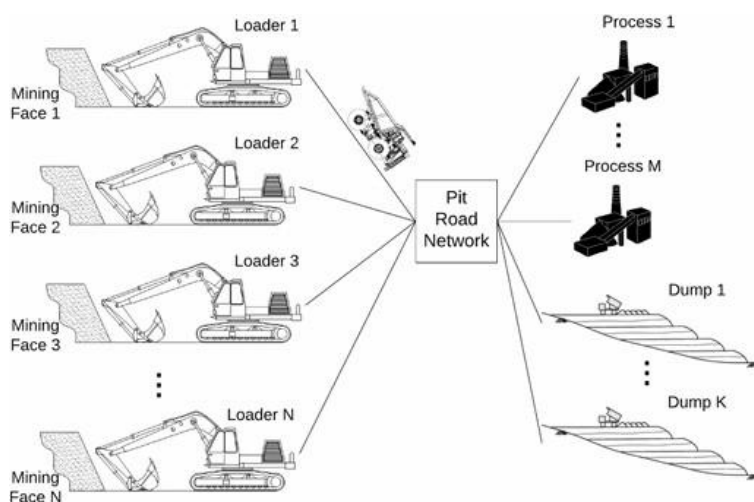


Figura 4.3: Problema despacho de equipos en mina cielo abierto. (Noriega et al., 2025).

La elección entre una tabla de valores y un aproximador funcional sigue siendo un tema de discusión. Mientras Noriega et al. (2025) cuestionan el uso de tablas por su limitada escalabilidad, Huo et al. (2023) justifican su empleo en problemas con espacios de estados acotados, aunque reconocen que el costo computacional aumenta al incorporar más dimensiones.

Yaakoubi and Dimitrakopoulos (2023) integran RL con la optimización estocástica simultánea de complejos mineros. Si bien la aplicación es minera, el estudio se centra en la selección de heurísticas dentro del optimizador: el agente decide, en cada iteración de un recocido simulado multi-vecindad, cuál de 38 heurísticas perturbativas aplicar. Es un

uso de RL como metaheurística, más que como controlador de la operación. La línea base es el mismo hiper-heurístico sin componente de aprendizaje, de modo que la comparación aísla el efecto del selector. Sobre dos complejos mineros industriales, reportan reducciones de 40 % a 75 % en tiempo de ejecución y de 50 % a 80 % en número de iteraciones, con un flujo de caja y un metal recuperado entre 2 % y 3 % superiores. La mejora depende, por tanto, del conjunto de heurísticas disponibles y no de una política de decisión construida desde cero.

4.3.2. Minería subterránea

Chimunhu et al. (2022) establecen que las aplicaciones de ML en planificación subterránea se ubican antes del modelo de optimización, prediciendo insumos como ley final, dilución y disponibilidad de equipos. Ninguna de las aplicaciones que documentan emplea RL. Su revisión alcanza hasta 2021. Desde entonces han aparecido aplicaciones de RL en minería subterránea, aunque ninguna aborda el secuenciamiento en hundimiento masivo. Noriega and Boisvert (2025) aplican aprendizaje por refuerzo profundo al diseño de caserones en minería subterránea bajo incertidumbre geológica, formulado como proceso de decisión de Markov. El estado registra, en tres escalas espaciales, los bloques ya extraídos junto con la media y la desviación estándar del valor económico sobre el conjunto de realizaciones. La acción define la ubicación, el largo y el ancho del caserón, con la altura fija en 40 m (Figura 4.4). La recompensa corresponde al valor esperado del caserón sobre las realizaciones, y toma valor cero cuando ese valor resulta negativo. Los autores emplean *proximal policy optimization* (PPO) y justifican la elección en que los métodos basados en valores Q resultan inestables con espacios de acción de alta dimensión. La línea base es Mineable Shape Optimizer (MSO) alimentado con el modelo promedio de las realizaciones. Sobre 100 realizaciones de un yacimiento de oro reportan una utilidad esperada 8.3 % superior y un 3.4 % más de oro extraído.

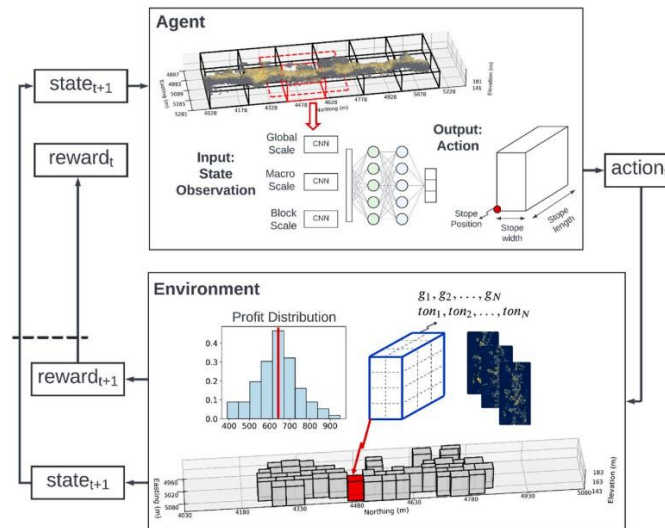


Figura 4.4: Esquema de aplicación de aprendizaje por refuerzo en el diseño de caserones bajo incertidumbre geológica y de leyes. (Noriega and Boisvert, 2025).

El alcance del trabajo es el diseño de la envolvente de extracción de un nivel, sin dimensión temporal: no asigna períodos ni define una secuencia de extracción. Los au-

tores declaran que, hasta donde conocen, el aprendizaje por refuerzo profundo no se ha explorado en problemas de diseño de minas subterráneas.

Li et al. (2026) abordan el secuenciamiento de corto plazo en minería (*strip-filling*), método en que el cuerpo mineralizado se divide en paneles, cada panel en franjas y cada franja en bloques de arranque. Formulan el problema como proceso de decisión de Markov y lo resuelven con aprendizaje por refuerzo profundo. El estado codifica atributos de bloques y de equipos sobre un grafo heterogéneo, y la acción escoge en un mismo paso el bloque a explotar y el equipo asignado. La recompensa es la reducción del tiempo de término entre pasos consecutivos, sin descuento. Las restricciones se imponen por construcción del conjunto de acciones factibles, no mediante penalización. La política se entrena con PPO sobre instancias de dos paneles y se evalúa hasta ocho paneles sin reentrenamiento. El agente optimiza el tiempo de término del programa mediante la asignación de bloques a equipos. El valor económico no interviene: la ley del bloque figura como atributo de estado pero no en la recompensa.

La comparación abarca cuatro grupos: otros algoritmos de RL profundo, cuatro reglas compuestas construidas por los autores a partir de la práctica de programación manual, cuatro metaheurísticas, y una formulación exacta resuelta con Gurobi. La ventaja depende del comparador. Frente a DQN el método reduce el tiempo de término en 13.6 %; frente al algoritmo genético, en 3.3 %; y en el caso real de una mina de fosfato, la diferencia con el genético baja a 0.7 %. La reducción de tiempo de cómputo respecto del genético alcanza un factor de 65 en la instancia mayor. Donde existe óptimo probado, el método queda 4.1 % y 5.7 % por sobre la solución exacta, y en las instancias mayores la formulación exacta no termina dentro de 24 h. El entorno es determinista, y los autores plantean como trabajo futuro incorporar fallas de equipo, cambios en las condiciones geológicas e incidentes de seguridad.

Ninguno de los trabajos revisados aborda específicamente la planificación y secuenciamiento de la producción en métodos de hundimiento masivo mediante RL. Frente a esta brecha, la presente memoria formula el problema de secuenciamiento en block caving como un MDP e implementa un agente de Q-learning tabular.

Capítulo 5

Metodología

Este capítulo describe la formulación del problema de secuenciamiento en *block caving* como un proceso de decisión de Markov, el modelo económico de bloques usado como caso de estudio, la arquitectura del agente de Q-learning tabular, el mecanismo de incorporación de restricciones operacionales, y el protocolo experimental empleado para evaluar el currículum incremental descrito en la Tabla 5.2. El orden de presentación sigue la secuencia en que estos elementos se construyen: primero el entorno (Secciones 5.1–5.2), luego el agente y su mecanismo de aprendizaje (Secciones 5.3–5.5), y finalmente la función de recompensa integrada y las métricas usadas para validar cada etapa (Secciones 5.6–5.8). La validación sigue un currículum incremental de ocho etapas, en el que cada una incorpora una restricción adicional sobre la anterior. La etapa E0 corresponde al entorno sin restricciones operacionales y fija la referencia de comportamiento del agente, cuyo criterio de ordenamiento es únicamente el valor de bloque. E1 impone la precedencia vertical por gravedad. E2 incorpora la temporalidad y con ella el efecto del descuento sobre el VAN. E3 acota la diferencia de altura entre columnas adyacentes. E4 exige el área mínima asociada al radio hidráulico. E5 limita la tasa de extracción. E6 y E7 incorporan la dirección de avance y el ángulo de extracción respectivamente.

5.1. Formulación del problema como proceso de decisión de Markov

El secuenciamiento de extracción se formula como un problema de decisión secuencial, en cada paso, el agente decide qué bloque extraer dado el estado actual del frente de extracción, y recibe una recompensa asociada a esa decisión. Esta formulación permite incorporar de forma natural el hecho de que una decisión de extracción actual restringe qué decisiones son físicamente válidas en el futuro (por ejemplo, la gravedad impone que un bloque solo puede extraerse si el bloque bajo él ya fue extraído). El problema se formula como un proceso de decisión de Markov definido por la tupla (S, A, P, R, γ) , donde:

- S : en las etapas E0 y E1, el estado corresponde a la posición del bloque candidato (i, j) dentro del modelo de bloques. No se incorpora una noción explícita de tiempo, por lo que la tabla de valores tiene dimensión 10×10 . Desde E2 en adelante, la planificación se organiza en períodos y el estado se extiende a (i, j, t) , con t el índice de período acotado a un máximo de 50 períodos (parámetro `max_t_estado`). La tabla de valores pasa entonces a dimensión $10 \times 10 \times 50$. Es importante notar que esta es una representación reducida del estado real del entorno: el estado verdadero

del tablero es la configuración completa de bloques ya extraídos (un espacio de 2^{100} configuraciones), intratable para una tabla exacta. La aproximación usada aquí evalúa cada bloque candidato de forma aislada (condicionado solo por su posición y, desde E2, por el período), delegando toda la lógica de restricciones, qué bloques son físicamente válidos en cada momento a funciones de validación externas a la tabla (`es_extraccion_valida`, `bloque_disponible_periodo`), en lugar de codificarla dentro del estado.

- **A**: el conjunto de acciones disponibles en un instante dado es el conjunto de bloques que son simultáneamente físicamente válidos (no extraídos, y que cumplen las restricciones duras activas en la etapa: gravedad, dirección, contigüidad, según corresponda) y disponibles según el período (dentro de la cuota de extracción por madurez de columna y del límite de aperturas nuevas por período, cuando la etapa las exige). A este conjunto se agrega explícitamente la acción de detenerse, en donde el agente puede optar por no extraer más bloques en la época en curso.
- $P(s'|s, a)$: la transición es determinista dado el estado del tablero, extraer el bloque (i, j) lo marca como extraído y, según la etapa, puede significar el avance de período. La única fuente de aleatoriedad del entorno es la posición inicial del agente en E0 (elegida uniformemente al azar sobre el modelo de bloques).
- $R(s, a)$: función de recompensa compuesta por el valor económico del bloque, un descuento temporal y las penalizaciones activas en la etapa. Esto se detalla en la Sección 5.6.
- $\gamma \in [0, 1)$: factor de descuento del algoritmo de Q-learning (distinto del descuento temporal económico f_t de la Sección 5.5, que pondera el valor de un bloque según el período en que se extrae). En los experimentos reportados, $\gamma = 0.97$

El objetivo del agente es encontrar una política π que maximice el retorno esperado descontado, a través de la Ecuación 4.2. Una vez definidos los estados, las acciones y la función objetivo, el proceso de interacción entre el agente y el entorno sigue el ciclo mostrado en la Figura 5.1. En primer lugar, las restricciones operacionales determinan las acciones factibles, correspondientes al bloque más profundo sin extraer de cada columna abierta y a la posible apertura de una columna contigua. A continuación, el agente selecciona una de estas acciones y el entorno actualiza su estado, marcando el bloque como extraído y habilitando el bloque inmediatamente superior como una nueva acción disponible. De este modo, cada extracción modifica el conjunto de decisiones posibles para el siguiente paso.

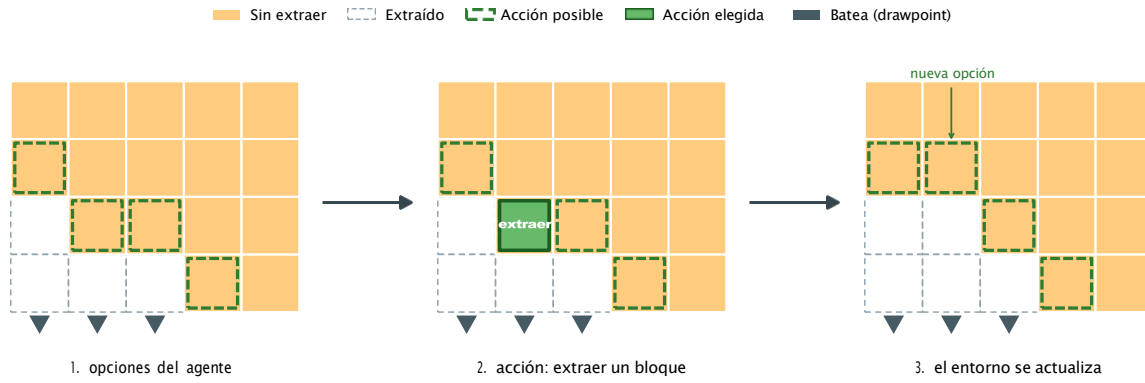


Figura 5.1: Ciclo de decisión sobre el modelo de bloques.

5.2. Modelo económico y de bloques

El caso de estudio corresponde a un modelo de bloques sintético de 10×10 (Figura 5.2a), que representa un yacimiento en el que el mineral de mayor valor se encuentra en profundidad. Cada bloque queda definido por su ubicación (x, y) , correspondiente a su columna y fila, respectivamente.

El valor económico de cada bloque se calcula a partir de su ley mineral mediante la Ecuación 2.1. El valor resultante se expresa en miles de USD (kUSD).

Los parámetros económicos empleados se presentan en la Tabla 5.1.

Tabla 5.1: Parámetros económicos del caso de estudio.

Parámetro	Símbolo	Valor
Precio (USD/lb)	P	4.5
Costo refinación y venta (USD/lb)	C_{RYV}	0.36
Tonelaje por bloque (t)	t	2 600
Recuperación metalúrgica (%)	R	0.80
Factor de conversión (lb/t)	k	2 204.6
Costo de mina (USD/t)	C_M	12.5
Costo de planta (USD/t)	C_P	15.0

Con los parámetros técnico-económicos definidos, es posible obtener una matriz de valorización, como se muestra en la Figura 5.2b, lo que define el entorno donde el agente realizará el entrenamiento. En ella, las filas superiores (0, 1 y 2) corresponden a material clasificado como estéril, mientras que el material inferior constituye mineral.

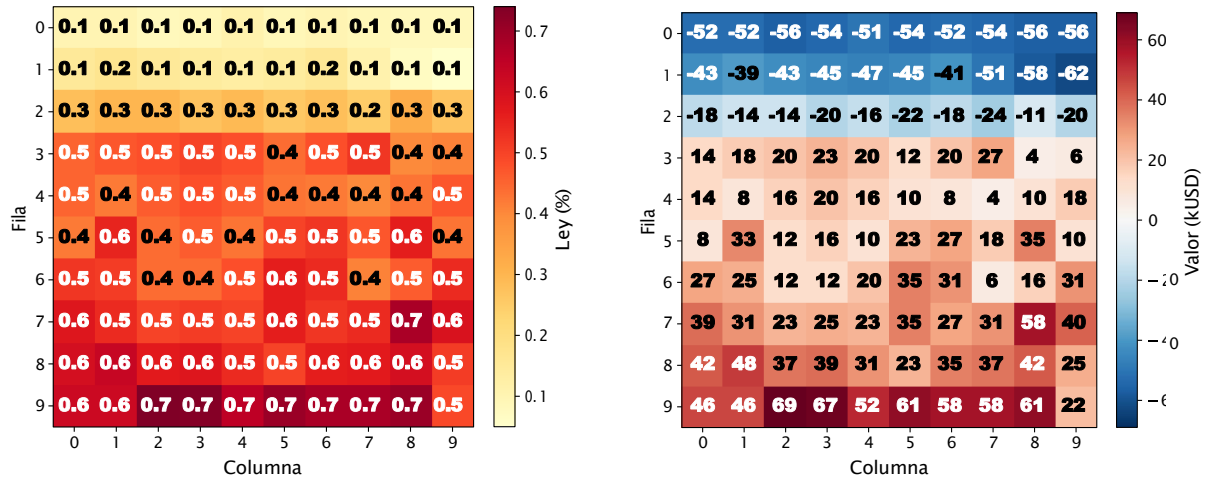


Figura 5.2: Caso de estudio: modelo de bloques sintético de 10×10 .

5.3. Arquitectura del agente: Q-learning tabular

El agente se modela en modo selector: en cada paso elige directamente cuál bloque extraer de entre los físicamente válidos y disponibles según el período, sin necesidad de navegar el tablero paso a paso. Esta elección evita que el agente deba aprender además una política de desplazamiento, aislando el aprendizaje al problema de secuenciamiento propiamente tal.

El agente aprende una función de valor Q almacenada en forma tabular, indexada por bloque. La regla de actualización depende de la etapa del currículum:

E0 y E1. La tabla se actualiza directamente hacia el valor económico inmediato del bloque,

$$Q(i, j) \leftarrow Q(i, j) + \alpha [V_{i,j} \cdot 40 - Q(i, j)]. \quad (5.1)$$

Dado que el entorno es determinista (el valor de un bloque no cambia entre episodios), esta regla converge a $Q(i, j) = V_{i,j} \cdot 40$: en la práctica, el agente aprende una puntuación de prioridad por bloque, y la política óptima resultante es extraer siempre el bloque válido de mayor valor. Esta propiedad permite usar el orden por valor (óptimo) como referencia de validación en E0 (Sección 5.7).

E2 en adelante — régimen de diferencia temporal. Desde que el estado incorpora el período t , la actualización sí propaga valor desde el mejor bloque disponible en el período siguiente:

$$Q(i, j, t) \leftarrow Q(i, j, t) + \alpha [r + \gamma \max_{(i', j') \text{ válidos}} Q(i', j', t') - Q(i, j, t)], \quad (5.2)$$

donde α es la tasa de aprendizaje y t' es el período vigente tras la transición. El máximo se acota inferiormente en 0 porque el agente siempre tiene la opción de detenerse (valor 0 en adelante). Este piso evita que una continuación de valor esperado negativo se propague hacia atrás y desincentive una extracción presente que en sí misma es correcta.

La selección de acciones emplea una política ε -greedy sobre el conjunto de bloques disponibles más la opción de detenerse: con probabilidad ε se elige uniformemente al azar entre ambas opciones, y en caso contrario se elige el bloque de mayor Q entre los disponibles, pero si ese máximo es negativo, el agente también decide detenerse en lugar de extraer. El programa de ε sigue una regla por tramos tal como está implementada:

$$\varepsilon_e = \begin{cases} 1,0 & \text{si } e < 0,3 E \\ \max(\varepsilon_{\min}, \varepsilon_{e-1} \cdot \exp(-\frac{e}{0,7 E})) & \text{si } e \geq 0,3 E \end{cases} \quad (5.3)$$

donde e es la época, E el número total de épocas de la etapa, y ε_{e-1} el valor de ε en la época anterior.

Hiperparámetros. El factor de descuento $\gamma = 0.97$ responde a la longitud del horizonte. Esto se debe a que, si se considera hasta 50 períodos, un bloque extraído al final conserva un peso 0.97^{50} aproximado de 0.22, mientras que con $\gamma = 0.90$ caería a 0.90^{50} aproximadamente a 0.005, y el agente ignoraría los bloques profundos de mayor ley, que son los que sostienen el valor del plan. La tasa de aprendizaje $\alpha = 0.1$ se fija baja porque el entorno es determinista, lo que genera que el valor de un bloque no varía entre episodios y no requiere promediar ruido. La función es estabilizar la propagación de Bellman en la tabla, evitando que cada visita sobrescriba el valor de estados poco visitados, como ocurriría con α cerca de 1. El programa de ε (Ecuación 5.3) mantiene exploración total ($\varepsilon = 1$) durante el primer 30 % de las épocas para poblar la tabla antes de explotar: fijar una política sobre una tabla vacía la anclaría en el orden de visita inicial. Conserva además un piso $\varepsilon_{\min} = 0.05$ para no quedar atrapado en un óptimo local. El límite de 2000 pasos por época es una cota de seguridad, ya que un plan completo en la grilla 10×10 requiere alrededor de 100 extracciones, de modo que el límite no restringe la política y solo acota episodios.

5.4. Diseño de restricciones: máscaras y penalizaciones

Las restricciones operacionales del método se incorporan mediante dos mecanismos de naturaleza distinta, elegidos según si la restricción corresponde a algo físicamente imposible de violar, o a algo que es ejecutable pero operacionalmente indeseable y que se prefiere que el agente aprenda a evitar.

Máscara (restricción dura). El espacio de acciones se restringe directamente. El agente no puede proponer una acción que viole la restricción, porque esta se excluye del conjunto de bloques válidos (`es_extraccion_valida`) o disponibles (`bloque_disponible_periodo`) antes de que el agente elija. Se usa para restricciones que representan imposibilidades físicas u operacionales estrictas, correspondiendo a la gravedad (E1) o la dirección de propagación (E6). Lo anterior permite evitar que el agente invierta entrenamiento en aprender restricciones propias del método de explotación, como la extracción de bloques de abajo hacia arriba, ya que estas se incorporan directamente mediante una máscara. En consecuencia, las acciones disponibles corresponden únicamente a los bloques que no poseen un bloque sin extraer en su parte inferior.

Penalización (restricción blanda). El agente puede ejecutar la acción, pero recibe un costo restado de la recompensa, por lo que debe aprender a evitarla a través de la señal de refuerzo. Se usa cuando conviene mantener el carácter de aprendizaje de la restricción en vez de imponerla de entrada. Así se penalizan la diferencia de altura entre columnas vecinas (E3), el área mínima antes de que esté formada (E4 a E6, antes de endurecerse a máscara en E7) y el ángulo de propagación fuera de la banda operacional (E7). Al incorporar las restricciones como penalizaciones, el agente conserva la posibilidad de incumplirlas, asumiendo el costo asociado. Esto puede ocurrir cuando estima que la penalización, descontada por el factor de descuento, es inferior al beneficio esperado de extraer un bloque de mayor ley, ya sea de forma inmediata o en etapas futuras.

Las penalizaciones se expresan como múltiplos del valor económico de referencia, de modo que están ancladas a la misma escala que el resto de la recompensa y no dependen de la magnitud absoluta de los parámetros económicos elegidos para el caso de estudio:

$$\text{pen}_k = \mu_k \cdot \bar{V}_+ \cdot 40, \quad (5.4)$$

donde μ_k es el multiplicador de la restricción k y $\bar{V}_+$ el valor económico promedio de los bloques positivos del modelo. Los multiplicadores usados en los experimentos son $\mu_{\text{altura}} = 1.0$, $\mu_{\text{área}} = 0.5$, $\mu_{\text{ángulo}} = 1.5$ y $\mu_{\text{bloque negativo}} = 0.8$.

Los valores de los multiplicadores fueron obtenidos de manera empírica. Como puede observarse, todos se encuentran expresados en un factor común de carácter económico, con el objetivo de que la penalización sea transversal a todos los bloques cuando se produzca el incumplimiento de alguna restricción.

5.5. Currículum incremental de restricciones

La estrategia de entrenamiento sigue un currículum incremental: cada etapa agrega una restricción operacional respecto de la etapa anterior, manteniendo todas las anteriores activas (Tabla 5.2). Esto permite aislar el efecto de cada restricción sobre el comportamiento del agente y atribuir cualquier cambio observado en las métricas a la restricción recién incorporada, en lugar de a una combinación no identificable de varios cambios simultáneos.

Tabla 5.2: Currículum incremental de restricciones.

Etapas	Restricción agregada	Modo	Efecto aislado
E0	—	—	RL puro (orden por valor)
E1	Gravedad	Máscara	Precedencia vertical
E2	Temporalidad	Estructural	Efecto del descuento (VAN)
E3	Altura adyacente	Penalización	Continuidad vertical del frente
E4	Área mínima (RH)	Penalización	Formación del radio hidráulico
E5	Tasa de extracción	Estructural	Ramp-up de producción
E6	Dirección	Máscara	Sentido de propagación
E7	Ángulo	Penalización	Ángulo operacional del frente

5.5.1. E1 — Gravedad (precedencia)

Un bloque (i, j) solo puede extraerse si todos los bloques por debajo en su columna ya fueron extraídos, reproduciendo la restricción de precedencia vertical propia de cualquier método de explotación por hundimiento:

$$\text{extraer}(i, j) \text{ válido} \Leftrightarrow \forall k > i: \text{extraído}(k, j) = \text{verdadero}. \quad (5.5)$$

La Figura 5.3 ilustra esta condición sobre una columna: los bloques inferiores ya extraídos habilitan un único bloque válido, mientras que los situados por sobre él permanecen bloqueados. Esta es la única restricción activa en E1, y se incorpora sin introducir aún noción de tiempo (el estado sigue siendo (i, j) , sin período), de modo que su efecto sobre el comportamiento del agente pueda observarse de forma aislada respecto de E0.



Figura 5.3: Restricción de precedencia vertical.

5.5.2. E2 — Temporalidad y descuento

A partir de esta etapa la extracción se organiza en períodos: cada período tiene una capacidad máxima de bloques (*capacidad_por_periodo*), y al alcanzarla, o al agotarse las opciones disponibles del período, se avanza al siguiente. El factor de descuento temporal por período se define como:

$$f_t = \theta^{t-1}, \quad \theta = 0,92, \quad (5.6)$$

y la recompensa de un bloque pasa a ser:

$$R_{i,j} = V_{i,j} \cdot 40 \cdot f_t \quad (5.7)$$

La Figura 5.4 muestra el efecto de este factor sobre una configuración de dos bateas: los bloques extraídos en el primer período conservan su valor, mientras que los de períodos posteriores lo ven reducido. La figura ilustra además el caso con capacidad de dos bloques por período, a diferencia de la configuración de un bloque por período del ejemplo anterior.

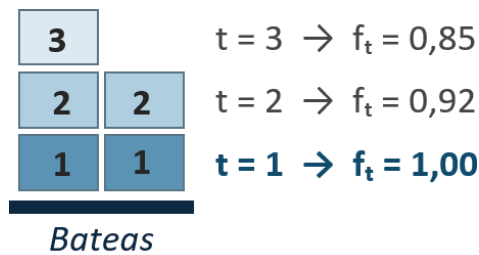


Figura 5.4: Restricción de temporalidad y capacidad.

5.5.3. E3 — Altura adyacente

Penalización blanda cuando la diferencia de altura con una columna vecina excede el umbral $\Delta_{m'ax}$:

$$\text{Restricción altura} \Leftrightarrow \exists v \in \{j-1, j+1\} : h_j - h_v > \Delta_{m'ax}, \quad (5.8)$$

donde h_j es la altura (bloques extraídos) de la columna j , con $\Delta_{m'ax} = 2$. La recompensa se reduce en $\text{pen}_{\text{altura}}$ por cada violación. Esta restricción representa, en términos geomecánicos, la necesidad de mantener un frente de hundimiento parejo: un escalón excesivo entre columnas vecinas introduce una discontinuidad en el frente que concentra esfuerzos y favorece condiciones de inestabilidad local. Un frente disparejo refleja, además, un tiraje no uniforme entre puntos de extracción vecinos, régimen bajo el cual los elipsoides de extracción no se traslapan de forma homogénea, lo que adelanta el descenso del contacto con el estéril y, con ello, la incorporación de dilución. La Figura 5.5 ilustra esta condición:



Figura 5.5: Restricción diferencia altura adyacente.

5.5.4. E4 — Área mínima

Se exige un segmento contiguo de al menos $A_{m'in}$ columnas abiertas antes de permitir seguir extrayendo bloques de una misma columna. Sea C el conjunto de columnas abiertas y c_i la columna ubicada en la posición i de la fila; la condición de área formada es:

$$\text{RH formado} \Leftrightarrow \exists i : \{c_i, c_{i+1}, \dots, c_{i+A_{m'in}-1}\} \subseteq C \quad (5.9)$$

con $A_{m'in} = 4$. Lo que se contabiliza es el tramo consecutivo más largo de columnas abiertas, que debe alcanzar $A_{m'in}$. Mientras el área no está formada, se penaliza tanto abrir una columna no adyacente a las ya abiertas (Figura 5.6) como extraer bloques adicionales de una columna sin haber formado el área. Esto representa la necesidad de generar primero el radio hidráulico crítico. En esta etapa la restricción opera en modo penalización; desde E7 se endurece a máscara de acciones.

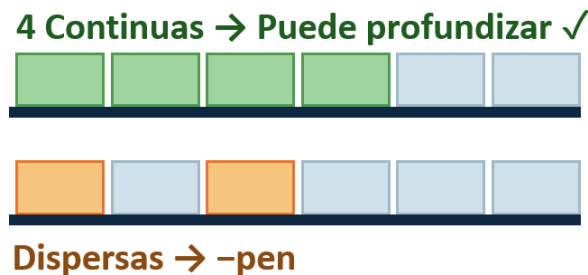


Figura 5.6: Restricción área mínima.

5.5.5. E5 — Tasa de extracción por madurez

El número máximo de bloques extraíbles por columna en un período depende de su estado de madurez:

$$\text{tope}(j) = \begin{cases} 1 & \text{si } b_j = 0 \text{ (apertura)} \\ 2 & \text{si } 1 \leq b_j \leq u \text{ (quiebre)} \\ 3 & \text{si } b_j > u \text{ (régimen)} \end{cases} \quad (5.10)$$

donde b_j son los bloques extraídos de la columna j al inicio del período y $u = \lceil F \cdot 0,30 \rceil$ es el umbral de transición ($F = \text{filas}$, $u = 3$ para $F = 10$). Lo anterior, se presenta en la Figura 5.7. Además, se limita a 2 el número de aperturas nuevas por período. Se implementa restringiendo directamente su disponibilidad en el período, es decir, el bloque es físicamente válido, pero no aparece entre las opciones que el agente puede elegir en ese período una vez agotada su cuota. Por lo anterior, se clasifica como restricción estructural en la Tabla 5.2.

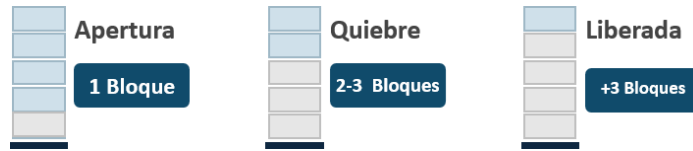


Figura 5.7: Restricción estructural según madurez de la columna.

5.5.6. E6 — Dirección de propagación

Máscara que fija el inicio en una esquina y el sentido de avance. Por ejemplo, para sentido izquierda a derecha, una apertura en la columna c es válida si:

$$\text{apertura}(c) \text{ válida} \Leftrightarrow c = 0 \text{ (si frente vacío), o } c > \text{máx}(C). \quad (5.11)$$

Como se presenta en la Figura 5.8, esta restricción impide retroceder en el sentido de propagación una vez iniciado el hundimiento, permitiendo reflejar que no se abren nuevos sectores en sentido contrario al del avance del frente ya establecido.

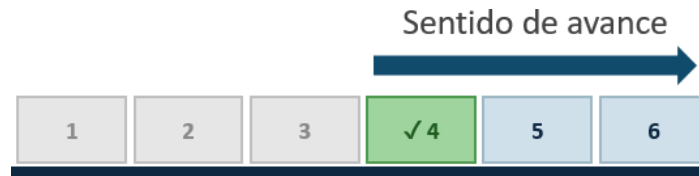


Figura 5.8: Restricción de dirección de propagación.

5.5.7. E7 — Ángulo quiebre

El ángulo del frente de hundimiento respecto de la vertical debe mantenerse dentro de la banda operacional $[\vartheta_{mín}, \vartheta_{máx}] = [27^\circ, 45^\circ]$. Si bien en los capítulos anteriores se discutió un umbral de entre 30° y 45° , en este trabajo dicho umbral se reduce con el fin de permitir una mayor diversidad de ángulos en la matriz de 10×10 . Esta decisión favorece el aprendizaje del agente, ya que incrementa la cantidad de configuraciones posibles durante el entrenamiento.

La Figura 5.9 muestra el procedimiento utilizado para medir el ángulo del frente. La medición se realiza sobre la zona de quiebre, definida entre la columna abierta más reciente y la última columna que aún permanece en quiebre ($b_j \leq u$). El ángulo se calcula entre el borde de dicha zona y la columna de mayor profundidad mediante $\vartheta = \arctan(\Delta z / \Delta x)$, excluyendo las columnas ya liberadas ($b_j > u$).

El ángulo es una propiedad que emerge de la secuencia completa de extracción y no de una acción aislada. En consecuencia, penalizar únicamente la extracción que revela el ángulo distribuye incorrectamente el crédito entre las acciones que realmente lo generaron. Por esta razón, el control se formula sobre cada extracción individual, en lugar de aplicarse sobre el ángulo agregado.

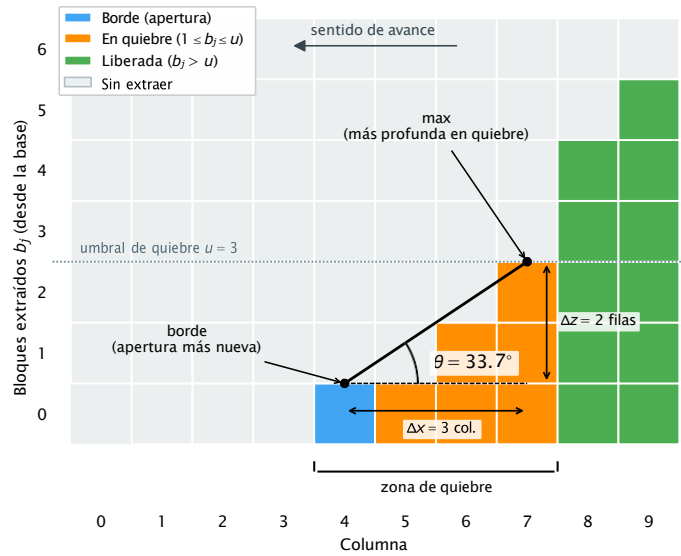


Figura 5.9: Medición del ángulo del frente.

Reglas locales de escalón. La banda angular se traduce a restricciones sobre diferencias de filas entre columnas vecinas, verificables acción por acción. Los umbrales se derivan de la banda y del umbral de quiebre, sin parámetros libres adicionales (Tabla 5.3).

Tabla 5.3: Derivación de los umbrales de E7 a partir de la banda angular y el umbral de quiebre (grilla 10×10 , umbral de transición 30 %).

Parámetro	Derivación	Valor
Umbral de quiebre u	$\lceil F \cdot 0,30 \rceil$	3 bloques
Desnivel de la zona d_z	$u - 1$	2 filas
Escalón máximo $e_{m'ax}$	$\lfloor \tan \vartheta_{m'ax} \rfloor \cdot 1 \text{ col}$	1 fila/col
Pendiente mínima $p_{m'in}$	$\tan \vartheta_{m'in} \approx 0,51 \rightarrow \text{discreta}$	0.5 filas/col
Ancho máximo de zona $w_{m'ax}$	$\lfloor d_z / \tan \vartheta_{m'in} \rfloor$	3 columnas

La zona de quiebre se define como el tramo contiguo de columnas desde el borde (columna abierta más nueva) hasta la última columna aún en quiebre ($0 < b_j \leq u$). Las columnas liberadas ($b_j > u$) quedan fuera de la zona y su profundización se rige únicamente por la restricción de altura adyacente (E3). Dentro de la zona operan cinco reglas,

evaluadas sobre cada extracción individual (Tabla 5.4), activas solo una vez formada el área mínima.

Tabla 5.4: Reglas de escalón de E7. b_j : bloques extraídos de la columna j tras la extracción evaluada; e : exceso en filas o columnas que escala la penalización.

Regla	Condición de violación	Desviación que previene
Profundización adelantada	$b_j - b_{\text{vecina más nueva}} > e_{m'ax}$	frente sobre 45°
Escalera invertida	$b_j > b_{\text{vecina más antigua}}$	perfil no monótono
Apertura tardía	escalón con el borde anterior $> e_{m'ax}$	$> 45^\circ$ local
Apertura temprana	zona resultante con $w > w_{m'ax}$ o pendiente $< p_{m'in}$	frente bajo 27°
Liberación sin apertura	el borde supera u sin columna siguiente abierta (si aún queda alguna por abrir)	zona de quiebre rota

La magnitud de la penalización es proporcional al exceso e y lleva el descuento temporal:

$$\text{pen}_\vartheta = \mu_\vartheta \cdot \bar{V}_+ \cdot 40 \cdot e \cdot f_t, \quad \mu_\vartheta = 1,5 \quad (5.12)$$

5.6. Función de recompensa

La función de recompensa acumula el valor económico descontado menos las penalizaciones activas en la etapa correspondiente. En su forma más general (E7, con todas las restricciones activas):

$$R = V_{i,j} \cdot 40 \cdot f_t - \text{pen}_{\text{neg}} \cdot f_t - \text{pen}_{\text{altura}} - \text{pen}_{\text{área}} - \text{pen}_\vartheta \quad (5.13)$$

donde cada término de penalización solo se resta si la etapa vigente lo tiene activo y la extracción efectivamente la viola (Tabla 5.2); el término pen_ϑ ya incorpora el descuento f_t y el exceso de escalón e (Ecuación 5.12). El término pen_{neg} (penalización adicional por extraer un bloque de valor negativo) es cero en E0–E2, y solo se activa desde E3, según lo justificado en la Sección 5.4.

Los multiplicadores de penalización (Ecuación 5.4) se calibraron de forma empírica bajo tres criterios. El primero fija la escala de cada penalización al valor económico promedio de un bloque positivo, de modo que su magnitud sea relativa e independiente de los parámetros económicos elegidos. El segundo exige que el número de violaciones de cada restricción tienda a cero durante el entrenamiento sin una caída sostenida en la razón de recuperación de valor (VRR, Sección 5.7): el agente debe aprender a respetar la restricción sin renunciar al valor capturable. El tercero fija los pesos relativos entre penalizaciones según el incentivo económico a violar cada una. El ángulo recibe el multiplicador más alto, 1.5, porque abrir una columna lejos del frente accede a bloques de alto valor y el incentivo a violarla es mayor.

5.7. Métricas de validación

Cada etapa se evalúa con un conjunto común de métricas, calculadas sobre la mejor secuencia obtenida durante el entrenamiento.

Razón de recuperación de valor (VRR).

$$VRR = \frac{\sum_{(i,j) \in \text{extraídos}} V_{i,j}}{\sum_{(i,j): V_{i,j} > 0} V_{i,j}}. \quad (5.14)$$

Corresponde a la fracción del valor económico positivo disponible en el modelo de bloques que el agente efectivamente logra capturar con su plan de extracción. Un VRR cercano a 1 indica que el agente extrae prácticamente todo el mineral económicamente valioso disponible.

Recompensa acumulada. Es la suma de la Ecuación 5.13 sobre toda la secuencia de extracción de la mejor época encontrada. Incluye tanto el descuento temporal como todas las penalizaciones vigentes en la etapa, por lo que actúa como un equivalente al valor presente neto del plan, penalizado por el incumplimiento de restricciones operacionales.

Correlación de Spearman. Mide la correlación entre el orden de extracción del agente y el orden que resultaría de extraer siempre el bloque válido de mayor valor económico (orden óptimo). Esta correlación es interpretable solo en E0, etapa en la que el orden efectivamente coincide con la política óptima (no hay restricciones que lo impidan).

Ángulo emergente por período. Al cierre de cada período se mide

$$\vartheta_p = \arctan \frac{b_{\max} - b_{\text{borde}}}{|c_{\max} - c_{\text{borde}}|}, \quad (5.15)$$

entre el borde y la columna más profunda aún en quiebre. Los períodos previos a la formación del área mínima no son medibles, y los períodos terminales —sin columnas restantes por abrir— se reportan, pero se excluyen del conteo de cumplimiento, ya que en ellos el agente ya no dispone de la decisión de apertura. Por la discretización en bloques enteros, la banda efectiva es $[\arctan(1/2), \arctan(1)] = [27^\circ, 45^\circ]$.

Violaciones de escalón por 100 extracciones. Número de reglas de la Tabla 5.4 disparadas en un episodio, normalizado por bloques extraídos.

VAN por columna. Para la secuencia de un determinado progreso de entrenamiento,

$$VAN_c = \sum_{b \in c} V_b \cdot f_{t(b)} - C_d \cdot f_{t_{\text{ap}}(c)}, \quad (5.16)$$

donde $t(b)$ es el período de extracción del bloque b , $t_{\text{ap}}(c)$ el período de apertura de la columna y C_d el costo de desarrollo por batea. Ambos términos usan el mismo factor f_t de la Ecuación 5.6.

5.8. Protocolo experimental

Cada etapa del currículum se entrena de forma independiente. Al iniciar una nueva etapa se reinician la tabla Q y ϵ , por lo que el agente aprende desde cero. Así, entre etapas solo se mantienen las restricciones incorporadas previamente, no la política aprendida.

El presupuesto de épocas crece con la complejidad de las restricciones activas: 1.000 épocas para converger en E0–E2, mientras que E3 requiere 5.000, E4 10.000 y E7 15.000 para recuperar el valor total (la progresión se reporta con cada etapa en el Capítulo 6). El máximo es de 2.000 pasos por época, suficiente para agotar el modelo de bloques de 10×10 en cualquier configuración de restricciones. Se registra un progreso del entrenamiento al 25 %, 50 % y 75 %, además del mejor resultado global, para observar la evolución del aprendizaje dentro de una misma corrida.

Semilla y la robustez estadística: el entorno es determinista y cada entrenamiento fija la semilla aleatoria al inicio, por lo que toda corrida es reproducible. El análisis detallado de cada etapa emplea la semilla 42. La robustez frente al azar de exploración se evalúa sobre la etapa final con tres semillas de entrenamiento (42, 1 y 7), reportando media y desviación estándar del VRR: para el caso base 10×10 (Tabla 6.2) y para los casos escalados del análisis de escalabilidad (Capítulo 7).

El entorno de cómputo corresponde a una implementación en Python (NumPy para el cálculo tabular, Matplotlib/Seaborn para las figuras, SciPy para el cálculo del coeficiente de Spearman cuando está disponible).

Capítulo 6

Resultados y análisis

6.1. Validación del núcleo de aprendizaje (E0-E2)

El primer caso corresponde a un agente que puede realizar la extracción sin restricciones, con libertad para extraer bloques sin seguir las reglas del método. Al 25 % del entrenamiento (Figura 6.1a), el agente extrae los bloques con un alto grado de aleatoriedad. Varios bloques de mineral permanecen sin extraer, mientras que se registran 24 extracciones en la zona de estéril. Este comportamiento cambia en la mejor secuencia obtenida (Figura 6.1b), en donde el agente concentra la extracción en la zona de mineral y evita el estéril.

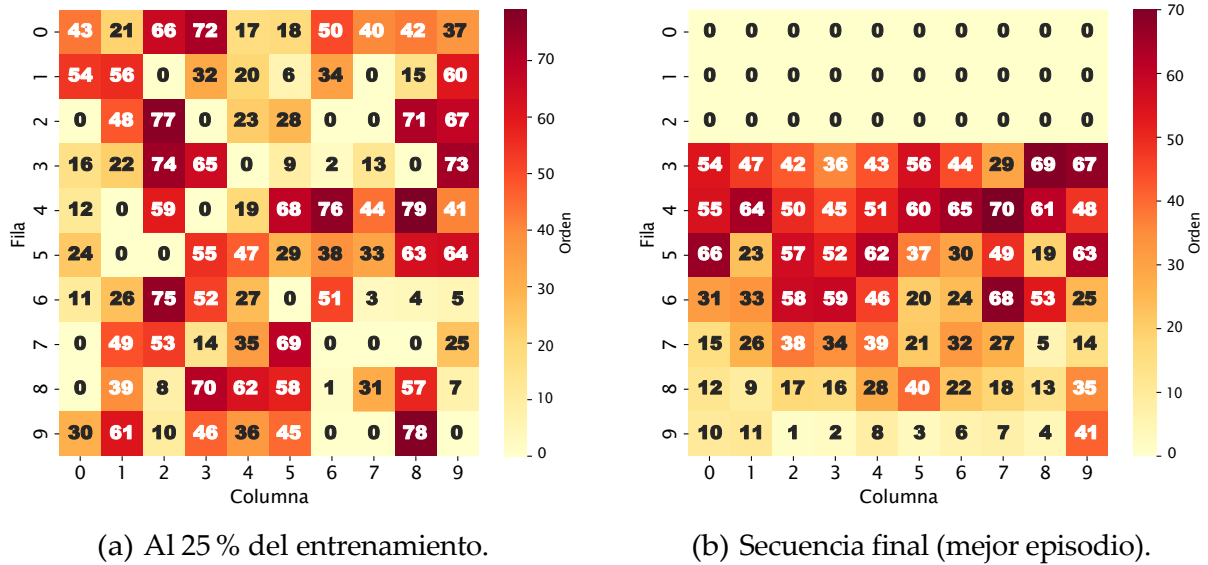


Figura 6.1: Evolución del aprendizaje en E0.

Se identifica la capacidad del agente de recuperar todos los bloques definidos como mineral ($VRR = 1.0$). En el gráfico de recompensa acumulada (Figura 6.2a) se evidencia el comportamiento de aprendizaje del agente. Inicialmente la recompensa es baja durante las primeras épocas (exploración), y se observa una transición en torno a las 250–300 épocas, donde la recompensa aumenta progresivamente. No obstante, es posible identificar el efecto del ϵ_{min} , correspondiente a 0.05, ya que si bien el agente explota la política aprendida, mantiene un grado de aleatoriedad residual.

En este primer caso es posible realizar la comparación con un orden óptimo, definido como el ordenamiento de los bloques de mayor a menor valor. Al comparar el resultado del agente con este orden de referencia, se obtiene, para 1000 épocas, un coeficiente de correlación de Spearman de $\rho = 0.997$ (Figura 6.2b), lo que evidencia la reproducción de un orden cercano al óptimo.

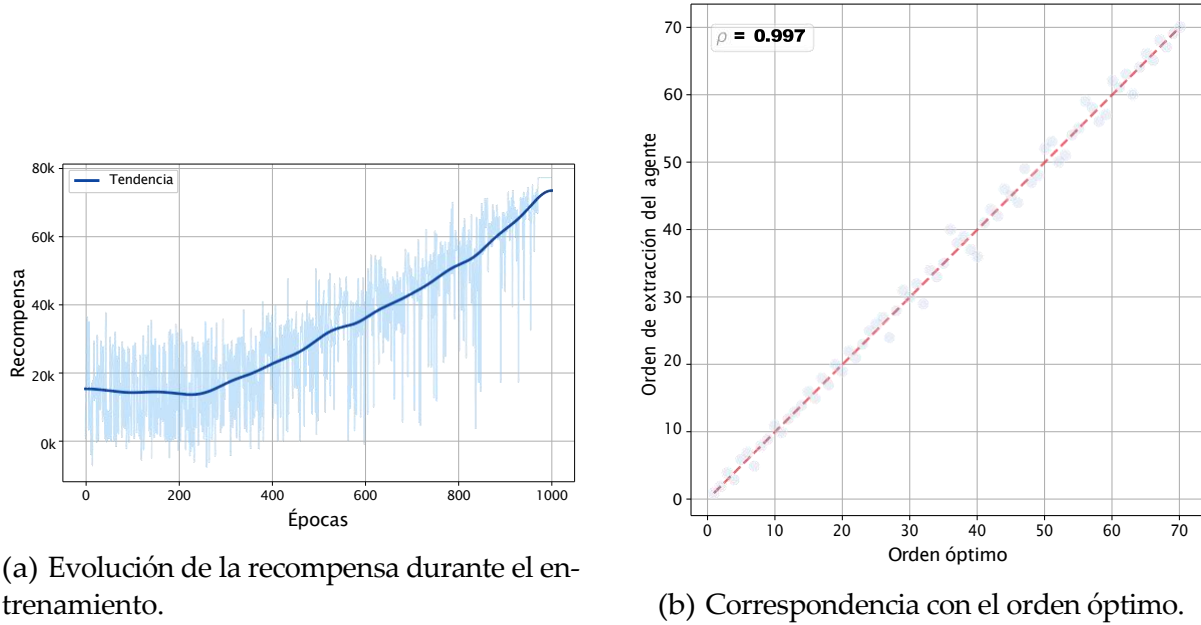
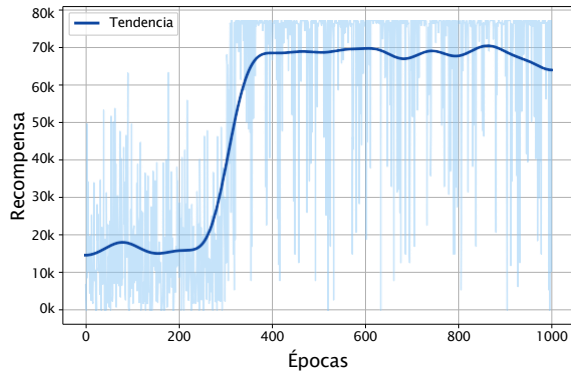
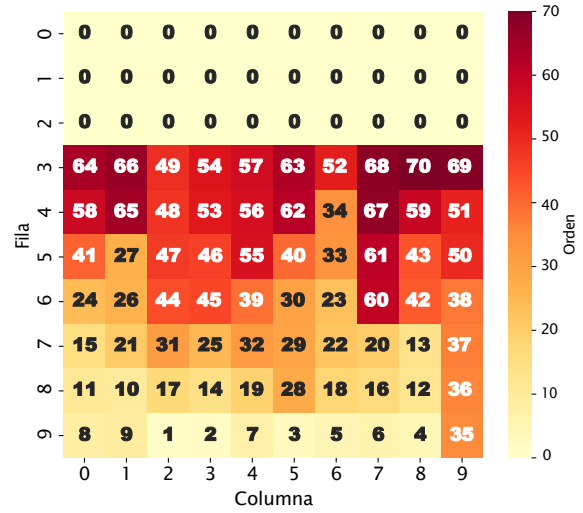


Figura 6.2: Validación del aprendizaje en E0.

Validado el comportamiento del agente sin restricciones, la siguiente etapa incorpora la precedencia gravitacional. Los resultados muestran que el agente obtiene nuevamente la recuperación total del beneficio ($VRR = 1.0$) y no extrae bloques negativos, alcanzando la misma recompensa que en E0. Este comportamiento se evidencia en el gráfico de recompensa acumulada (Figura 6.3a), que presenta un comportamiento inicial con épocas de baja recompensa al inicio y una estabilización hacia el final. Es posible observar en el gráfico de secuencia de extracción (Figura 6.3b) que el agente extrae primero los bloques de la fila más profunda y, una vez extraído este soporte, los cercanos a la superficie son los últimos, respetando la precedencia gravitacional del método.



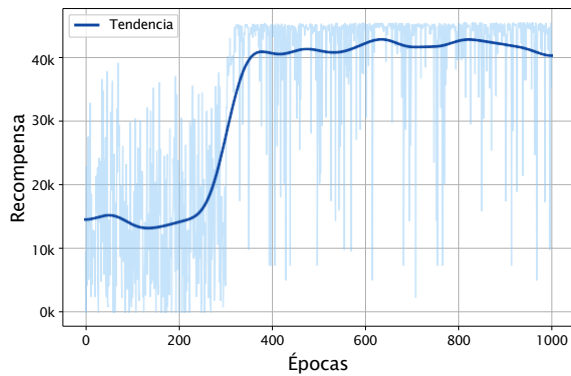
(a) Evolución de la recompensa.



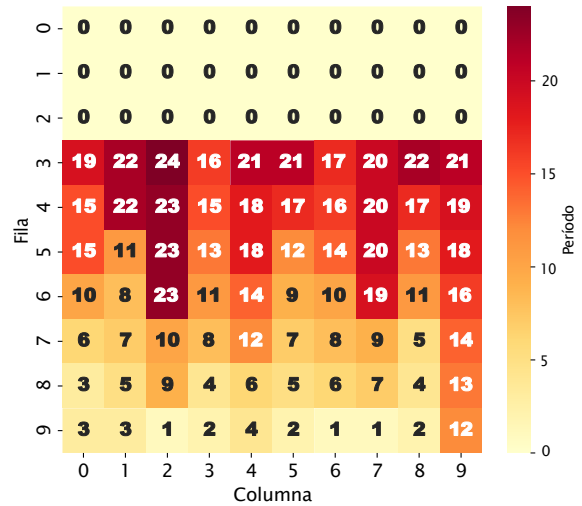
(b) Secuencia de extracción.

Figura 6.3: Resultados de E1 (precedencia gravitacional).

En las etapas anteriores, el agente extrae los bloques en un orden cercano al óptimo de manera deliberada, determinado por la propagación de valor de la ecuación de Bellman. Al comparar E0 y E1 se observa la misma recompensa total pese a que las secuencias de extracción difieren. La ausencia de un factor de descuento temporal impide que el orden alcanzado por la política se refleje en EL valor de la recompensa. La etapa E2 incorpora los períodos de extracción y el descuento temporal, dado por el factor $f_t = 0,92^{(t-1)}$, de modo que el valor de un bloque extraído disminuye conforme avanza el tiempo. El estado del agente pasa a incorporar la dimensión temporal, y la actualización de la tabla Q propaga el valor futuro descontado mediante la ecuación de Bellman. Como consecuencia, el impacto del ordenamiento por valor se vuelve observable en la recompensa, ya que el descuento penaliza la extracción tardía del material de mayor valor. La incorporación del descuento reduce el valor de la recompensa total. En E0 y E1 alcanzaba cerca de 70k; en E2 baja a 40k, aunque la curva de aprendizaje mantiene un comportamiento similar (Figura 6.4a). Esta etapa incorpora además la capacidad de extracción, que limita al agente a tres bloques por período (Figura 6.4b).



(a) Evolución de la recompensa.

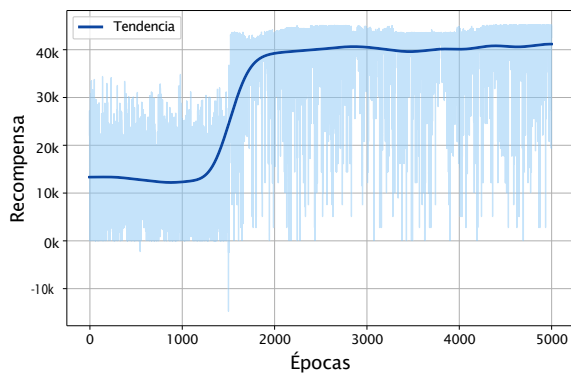


(b) Secuencia de extracción por período.

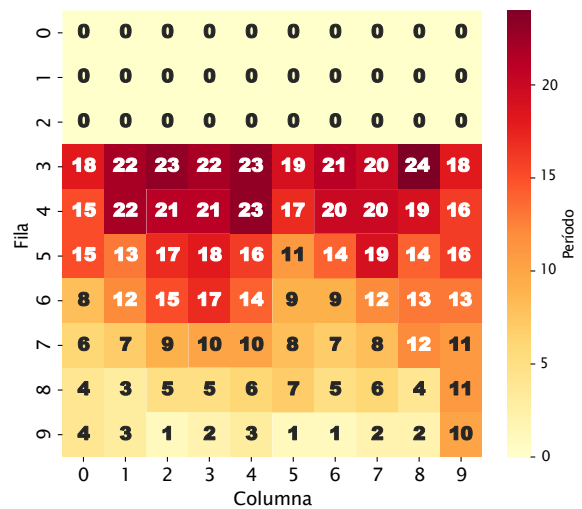
Figura 6.4: Resultados de E2 (temporalidad y capacidad).

6.2. Restricciones geométricas (E3-E4)

La etapa E3 incorpora la restricción de altura adyacente, que limita la diferencia de altura permitida entre columnas vecinas. Al introducir esta penalización, fue necesario aumentar el número de épocas de entrenamiento hasta 5.000 para alcanzar la recuperación total del valor ($VRR = 1.0$), como se observa en el gráfico de recompensa acumulada (Figura 6.5a), donde se aprecia la convergencia de la política de aprendizaje. La Figura 6.5b presenta la mejor secuencia de extracción, donde la diferencia de altura entre columnas vecinas se mantiene dentro del límite de dos bloques en todos los períodos.



(a) Evolución de la recompensa.



(b) Secuencia de extracción por período.

Figura 6.5: Resultados de E3 (Altura adyacente).

En las siguientes figuras se visualiza la diferencia máxima de altura entre columnas adyacentes, evaluada tanto en la etapa anterior (E2), sin control de altura, como al incorporar la penalización correspondiente (E3). En el primer caso (Figura 6.6a), se registran

violaciones del umbral en 15 períodos, mientras que al incorporar la penalización directamente en la función de recompensa (Figura 6.6b), el agente aprende secuencias que respetan el límite establecido, sin registrar violaciones (0 de un total de 24 períodos), manteniendo una diferencia máxima de dos bloques entre columnas adyacentes.

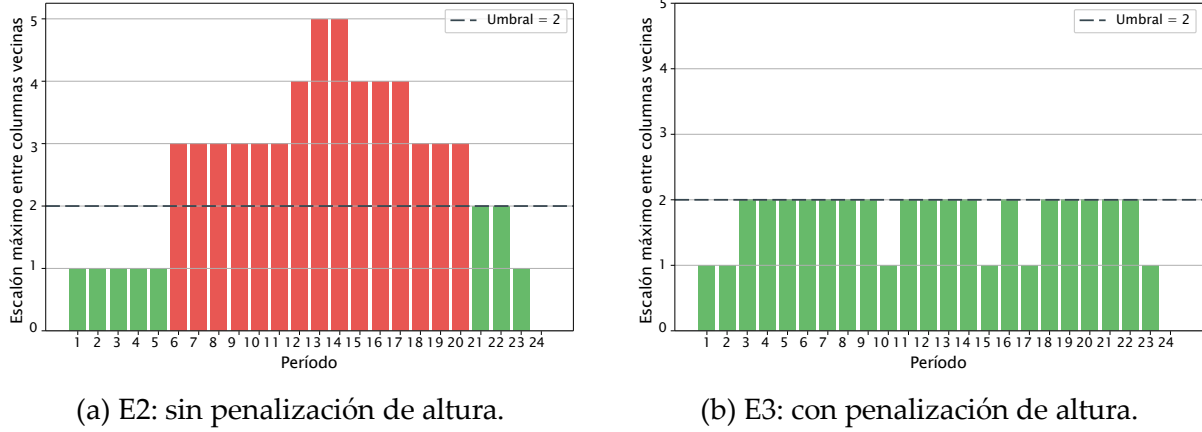


Figura 6.6: Efecto de la penalización de altura.

Tras validar el control de altura, se incorpora la restricción de área mínima, que representa el frente mínimo necesario para la propagación del *caving*. Esta restricción implementa un área mínima de cuatro columnas contiguas. Como se observa en el gráfico de violaciones (Figura 6.7), en las primeras épocas el agente presenta un elevado número de violaciones de esta restricción, llegando a cerca de 25. Este valor corresponde a la cantidad de veces que el agente no cumple la restricción en un episodio, ya sea extrayendo en altura antes de conformar el área o dejando columnas sin abrir entre las extraídas. El agente requiere nuevamente 10.000 épocas para recuperar el 100 % del valor en esta etapa, y en torno a las 3.000 épocas las violaciones comienzan a disminuir, tendiendo a cero.

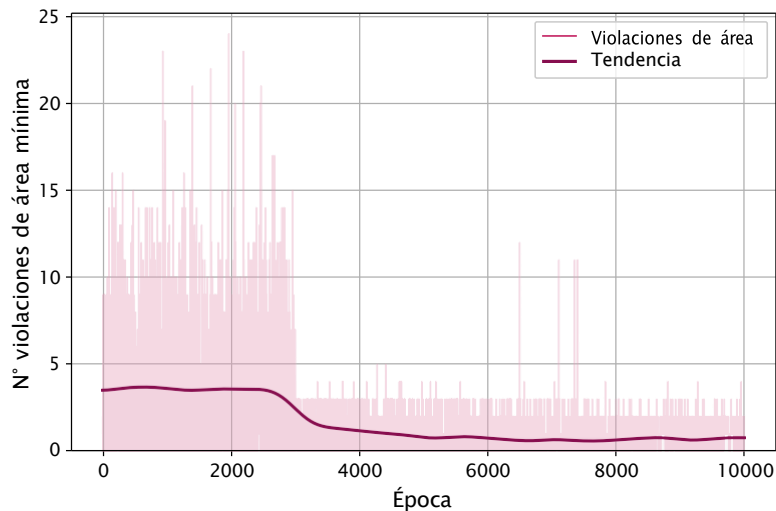


Figura 6.7: Evolución de las violaciones de área mínima durante el entrenamiento en E4.

Visualmente, es posible evidenciar el comportamiento en el gráfico de recompensa (Figura 6.8a), donde la disminución de las violaciones de área se relaciona con el aumento de la recompensa del agente en torno a las 3.000 épocas, punto en el que la recompensa

comienza a aumentar y estabilizarse. Este comportamiento también se aprecia en la secuencia de extracción (Figura 6.8b), donde el agente forma el área mínima abriendo las columnas 2, 3 y 4 en el período 1, para posteriormente abrir las columnas adyacentes en el período 2. De esta forma, se abren nueve columnas contiguas hasta el período 3, y recién en el período 4 el agente comienza a extraer bloques de la Fila 8.

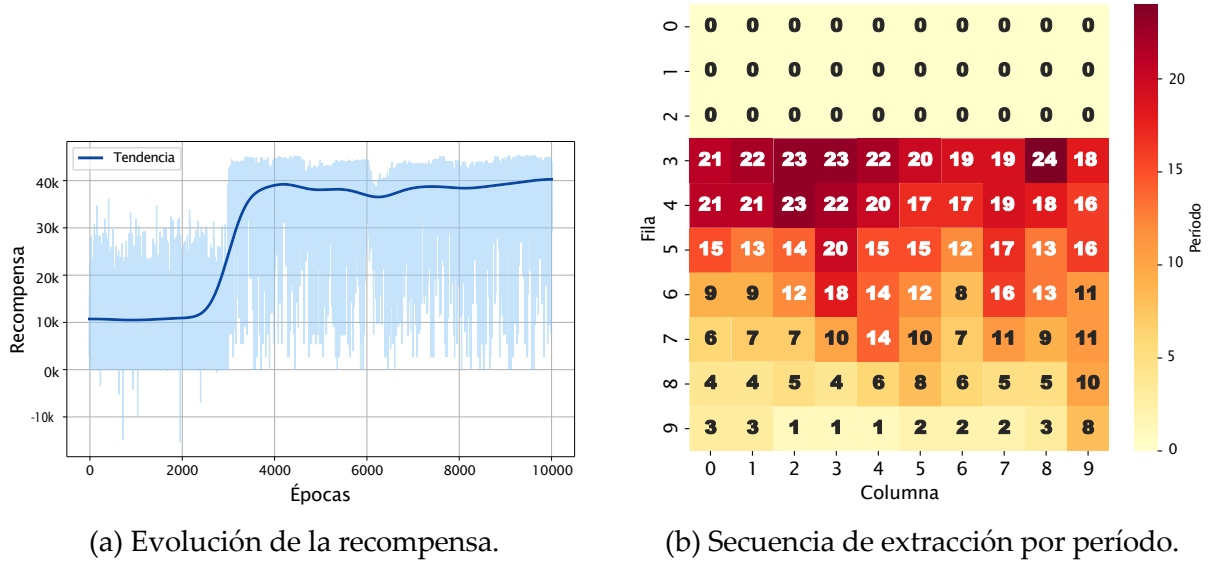


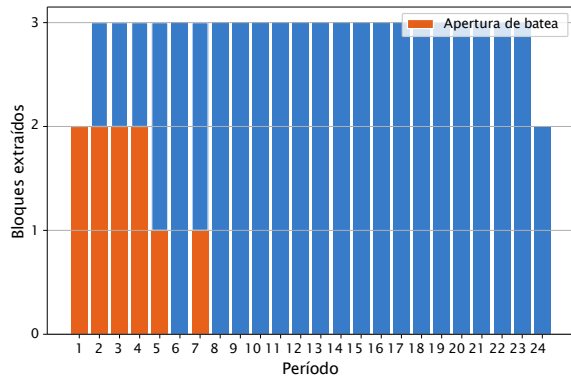
Figura 6.8: Resultados de E4 (área mínima contigua).

6.3. Restricciones de propagación (E5–E6)

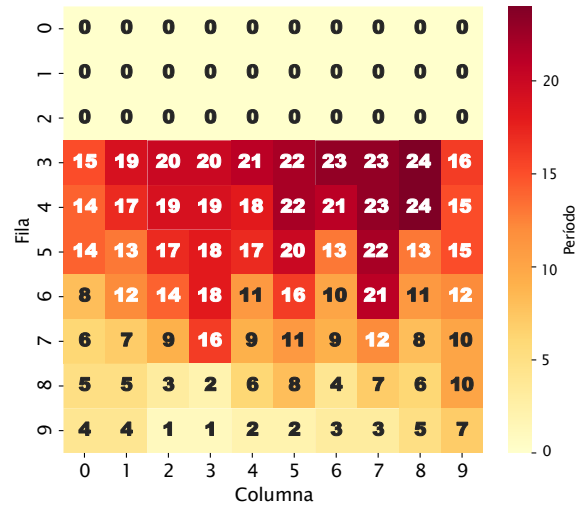
Validadas las etapas anteriores, se incorpora el estado de las columnas, con el fin de regular la cantidad de aperturas por período. Según la metodología, se configura un máximo de dos aperturas por período.

Este comportamiento se aprecia en la Figura 6.9a, donde las bateas en apertura se limitan a un máximo de dos por período. En la Figura 6.9b se evidencia que en el período 1, el agente extrae dos bloques, y al cumplir la restricción de apertura, no dispone de más acciones. En el período 2 abre las bateas adyacentes, conformando las cuatro columnas contiguas del área mínima necesaria para iniciar el *caving* (columnas 2, 3, 4 y 5). Recién con el área mínima alcanzada, y con una capacidad de un bloque adicional, profundiza la columna 3, abierta en el período 1. De esta forma, es posible identificar que el agente respeta tanto la capacidad de extracción como el máximo de aperturas, en función del estado de cada columna.

La restricción de columnas en quiebre, definida con un máximo de 2 bloques en el mismo período, se identifica en la columna 9, donde en el período 10 se extraen 2 bloques. Posterior al quiebre, el máximo permitido es de 3 bloques, al encontrarse la columna en estado liberado. Si bien no se observa la extracción de 3 bloques en un mismo período, producto de la restricción de diferencia máxima de 2 bloques entre columnas adyacentes, sí se identifican 2 extracciones en el mismo período, por ejemplo en la columna 0 (Filas 4 y 5) o en la columna 5 (Filas 3 y 4).



(a) Perfil de producción por período.

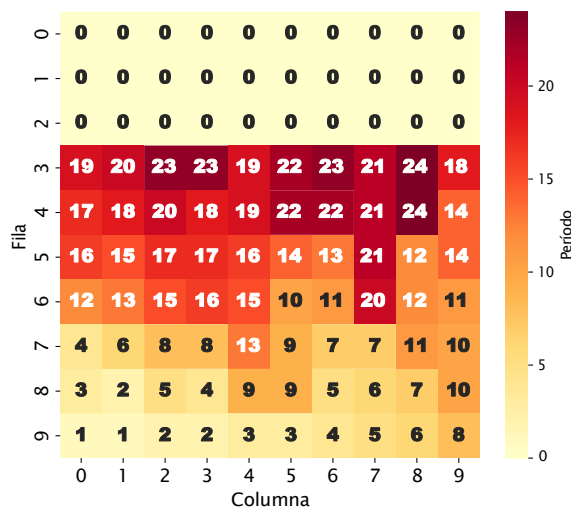


(b) Secuencia de extracción.

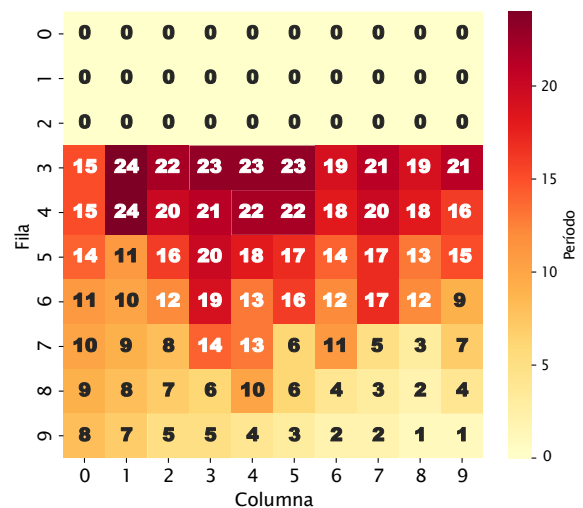
Figura 6.9: Resultados de E5 (tasa de extracción).

Una vez implementada la restricción anterior, y con el propósito de mantener frentes de propagación en una dirección determinada, se incorporaron las direcciones de propagación. Para ello, se evaluó la apertura de izquierda a derecha y viceversa, lo cual puede evidenciarse en la Figura 6.10a, que muestra la apertura de izquierda a derecha, y en la Figura 6.10b, que muestra el caso inverso. Lo anterior queda determinado por el sentido en que se van abriendo los bloques, siguiendo las restricciones descritas previamente.

En ambos casos, el agente aprende una política que le permite recuperar el valor total ($VRR=1.0$), y la propagación resulta contigua, sin retrocesos ni saltos. Esto evita lo que ocurría anteriormente, cuando bastaba con cumplir el área mínima para iniciar la extracción, lo que permitía avanzar en una dirección y luego retroceder, tal como se observa en la Figura 6.9b, donde el agente avanza hacia la derecha y, al llegar a la columna 7 en el periodo 3, vuelve a abrir las columnas 0 y 1 en el período 4.



(a) Propagación de izquierda a derecha.



(b) Propagación de derecha a izquierda.

Figura 6.10: Dirección de propagación en E6.

6.4. Control de ángulo (E7)

6.4.1. Comportamiento sin penalización de ángulo

La Figura 6.11 presenta el resultado de la etapa E7 sin penalización de ángulo. De los cinco períodos de frente activo, tres registran un ángulo dentro de la banda objetivo. La evolución no es uniforme a lo largo de los períodos. El ángulo se mantiene en la banda durante los períodos 2 y 3 (26.6° y 33.7°), desciende bajo el piso de 27° en los períodos 4 y 5 (21.8° y 18.4°), y retorna a la banda en el período 6 (26.6°). Este descenso coincide con la fase de mayor tasa de apertura, en la que el frente pasa de cinco a ocho columnas abiertas entre los períodos 3 y 5. En ausencia de una penalización que exija un desnivel mínimo, el agente prioriza la apertura de nuevas columnas, la cual permite el acceso a mineral adicional, por sobre la profundización de las ya abiertas, lo que reduce la pendiente del frente y genera ángulos bajo la banda. Los períodos 7 a 9 corresponden a la fase terminal: las diez columnas ya han sido abiertas y el agente solo extrae bloques de columnas existentes (Figura 6.11a). Por lo que su ángulo refleja el cierre del frente y no una decisión de propagación, y se excluyen del cómputo. Aunque la política captura la totalidad del valor económico disponible ($VRR = 1.0$), el ángulo permanece fuera de banda en dos de los cinco períodos activos, lo que indica que sin una señal explícita el agente satisface el objetivo económico sin regular la geometría de propagación.

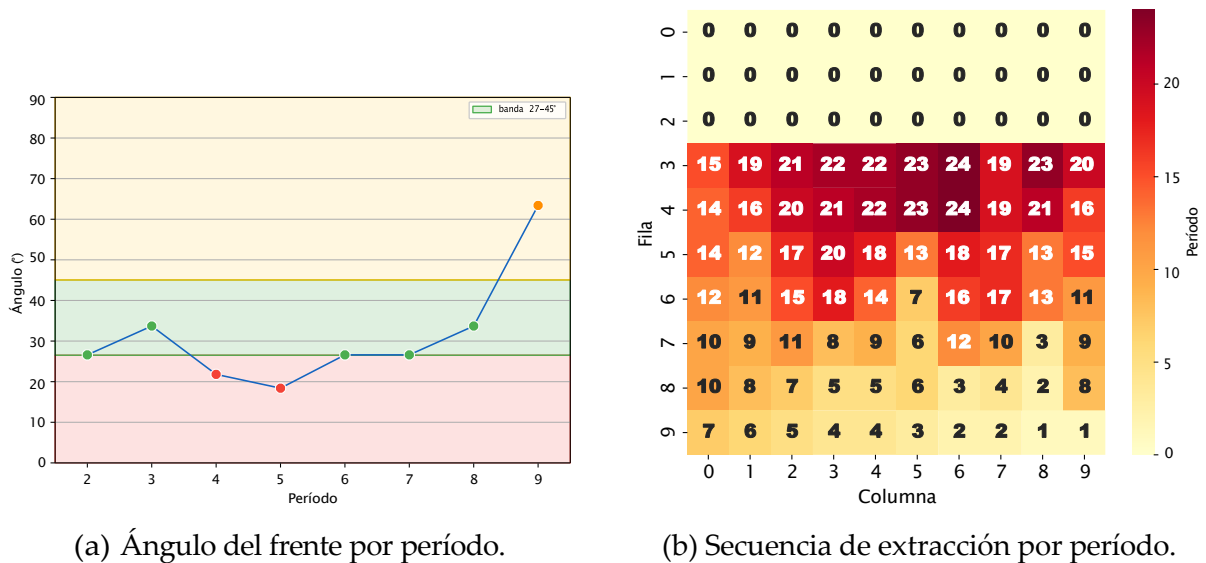


Figura 6.11: Etapa E7 sin penalización de ángulo ($\mu_\theta = 0$), 10,000 épocas: 3 de 5 períodos activos en banda.

6.4.2. Comportamiento con penalización de ángulo

La Figura 6.12 presenta el resultado de E7 con penalización ($\mu_\theta = 1,5$). De los nueve períodos de frente activo, ocho registran un ángulo dentro de la banda objetivo, frente a los tres de cinco del caso sin penalización. La trayectoria del ángulo es estable, salvo el período inicial de formación (18.4° , con solo cuatro columnas abiertas), el ángulo se mantiene en la banda a lo largo de todos los períodos, concentrándose en 33.7° . Este comportamiento se explica por un cambio en el ritmo de apertura, ya que sin penalización el frente completa la apertura de las diez columnas en el período 7, con penalización lo

hace recién en el período 12, profundizando las columnas ya abiertas antes de expandir el frente (Figura 6.12b). La penalización no restringe la velocidad a la que se abren, lo que preserva el desnivel entre columnas y sostiene el ángulo dentro de la banda. Este resultado, obtenido a 15,000 épocas de entrenamiento, se logra sin pérdida de valor económico (VRR = 1.0), lo que indica que la señal de refuerzo logra incorporar la restricción geométrica a la política sin comprometer el objetivo de recuperación.

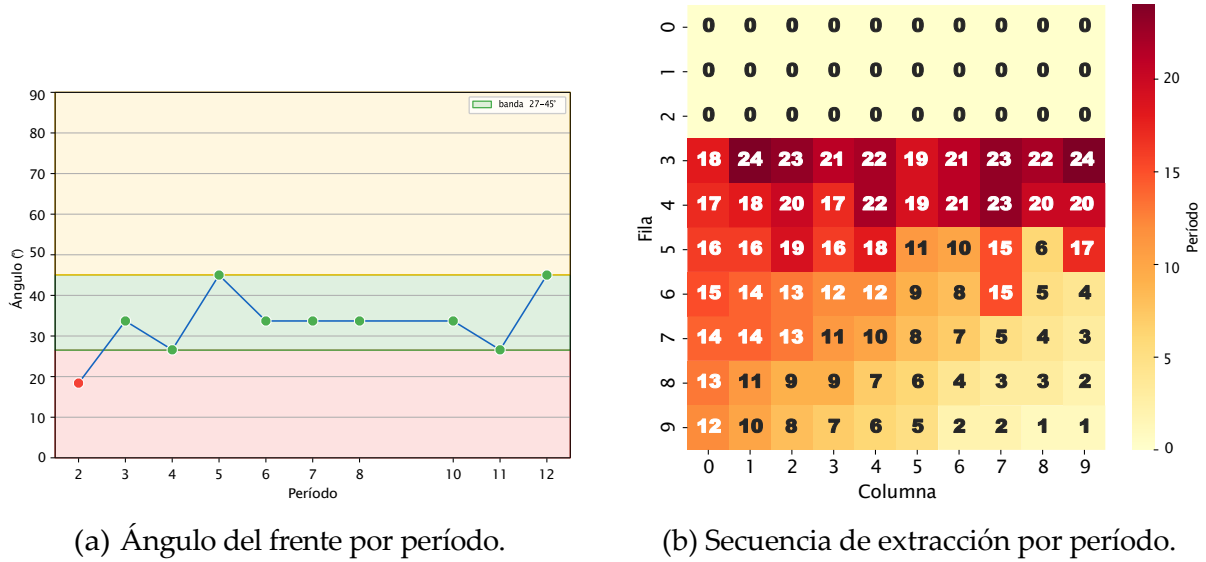


Figura 6.12: Etapa E7 con penalización de escalón ($\mu_{\theta} = 1,5$), 15 000 épocas: 8 de 9 períodos activos en banda, VRR = 1,000.

La Figura 6.13 presenta la evolución de las violaciones del ángulo durante el entrenamiento, expresadas por cada 100 extracciones, junto con el número de extracciones por episodio. A diferencia de una curva de aprendizaje convencional, la tasa de violaciones no decrece de forma monótona: parte en torno a 2–3 violaciones por 100 extracciones, aumenta hasta un máximo cercano a 13 alrededor de la época 10,000, y descende luego hasta estabilizarse en torno a 8. Este comportamiento se explica por el orden en que el agente adquiere las dos capacidades que exige la etapa. Durante las primeras 3,000 épocas el agente extrae aproximadamente un bloque por episodio, por lo que en la mayoría de los episodios no llega a conformar el frente ni a alcanzar la zona donde aplican las reglas de ángulo, las violaciones son escasas porque no opera en el régimen donde esta se evalúa. A partir de la época 3,000 el número de extracciones por episodio aumenta de forma sostenida y el agente aprende a desarrollar el frente en profundidad, pasando de un bloque a más de 40 por episodio, y las violaciones se incrementan de manera acoplada. El agente comienza a operar de forma consistente en la zona restringida antes de haber ajustado la geometría de propagación. El aumento de extracciones y el de violaciones indican que el incremento no corresponde a un deterioro de la política, sino a que la capacidad de extracción se adquiere antes que la regulación geométrica.

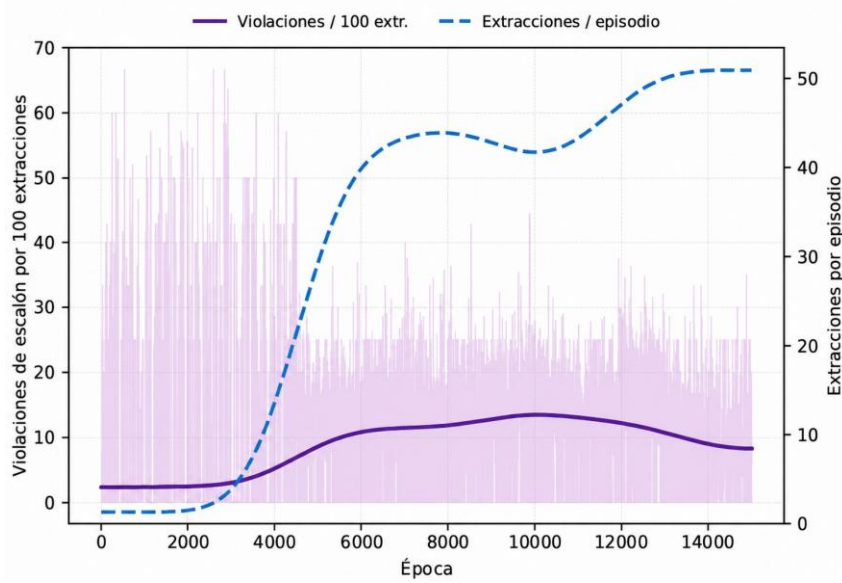


Figura 6.13: Violaciones de escalón por cada 100 extracciones durante el entrenamiento ($\mu_\theta = 1,5$), con las extracciones por episodio superpuestas.

En la fase final del entrenamiento las violaciones descienden mientras el número de extracciones por episodio se mantiene, lo que corresponde al ajuste de la geometría del frente sin pérdida de capacidad extractiva. No obstante, la tasa no converge a cero. Esta incurre entre 2 a 3 violaciones por episodio, lo que descarta que el remanente provenga de la aleatoriedad de exploración e indica que corresponde a una decisión de la política. En el mejor episodio, la penalización de ángulo asociada equivale al 5.24 % del valor extraído. El agente incorpora, por lo tanto, un número acotado de violaciones cuando la penalización que estas conllevan es inferior al valor que obtendría evitándolas. Este comportamiento es coherente con la elección de implementar la restricción como penalización y no como máscara, ya que permite el intercambio entre cumplir la geometría o absorber la penalización cuando la recompensa lo compensa.

6.4.3. Sensibilidad al multiplicador de penalización

La Tabla 6.1 presenta el efecto del multiplicador de penalización de ángulo sobre el comportamiento del agente, con cada configuración entrenada a 10.000 épocas. Al aumentar μ_θ , la tasa de violaciones de escalón de la política disminuye, de 26.17 por cada 100 extracciones sin penalización a 5.75 con $\mu_\theta = 1.5$, lo que confirma que la señal de penalización reduce la incidencia de la restricción geométrica. La recompensa del mejor episodio se mantiene en torno a 37,000 para valores de μ_θ entre 0.5 y 2.0. Con $\mu_\theta = 3.0$ la recompensa cae a 18,035 y el número de bloques extraídos se reduce a 24: la penalización supera el incentivo económico de desarrollar el frente, y el agente reduce la extracción para evitar incurrir en violaciones. Se adopta $\mu_\theta = 1.5$ por corresponder al valor que minimiza la tasa de violaciones dentro del rango en que la recompensa permanece estable, previo al colapso observado en $\mu_\theta = 3.0$. Dado que cada configuración corresponde a una única corrida, la tabla presenta variabilidad: la tasa de violaciones en $\mu_\theta = 2.0$ (8.78) supera la de $\mu_\theta = 1.5$ (5.75), rompiendo la tendencia monótona. En consecuencia, $\mu_\theta = 1.5$ se interpreta como un valor representativo del rango estable y no como un óptimo puntual.

Tabla 6.1: Ablación del multiplicador de penalización de ángulo μ_ϑ (E7, 10 000 épocas, semilla única). $\mu_\vartheta = 0$ es el control sin penalización.

μ_ϑ	Viol./100 extr.	R_{mejor}
0.0	26.17	43 270
0.5	19.04	37 635
1.0	12.44	37 295
1.5	5.75	37 532
2.0	8.78	36 870
3.0	1.55	18 035

Viol./100 extr.: violaciones de escalón por cada 100 extracciones en el último 10 % de las épocas (política convergida). R_{mejor} : recompensa del mejor episodio.

6.4.4. Convergencia y costo del enfoque tabular

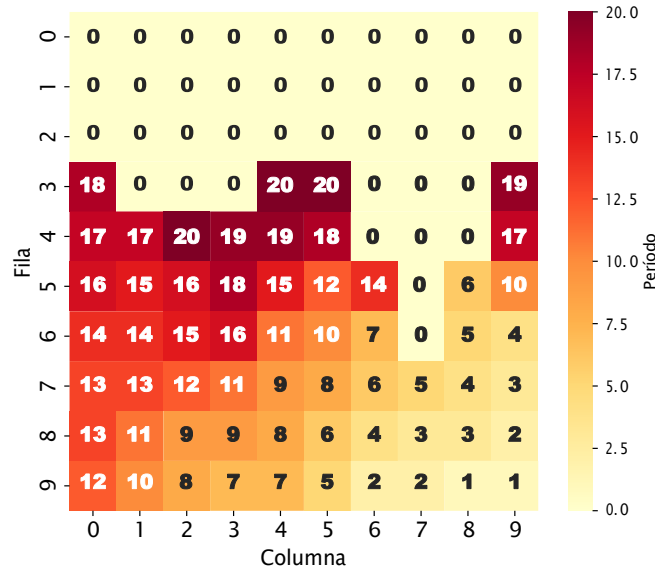


Figura 6.14: Secuencia de extracción con penalización ($\mu_\vartheta = 1.5$) a 10.000 épocas (VRR = 0.918).

Los resultados con penalización de la Sección 6.4.2 se obtuvieron con 15,000 épocas. Entrenada solo 10,000 épocas, la política con penalización de ángulo ($\mu_\vartheta = 1.5$) alcanza un VRR de 0.918 (Figura 6.14), donde las columnas centrales de la fila 3 permanecen sin extraer. Al extender el entrenamiento a 15,000 épocas, la política con penalización alcanza también un VRR de 1.00 (Figura 6.12b), manteniendo 8 de 9 periodos activos en banda. La diferencia observada se debe a un mayor número de épocas requerido para la convergencia. Lo que queda en evidencia al comparar 10,000 y 15,000 épocas, es que la recompensa del mejor episodio aumenta de 37,532 a 37,684, mientras el VRR aumenta de 0.918 a 1.000. Este incremento marginal de la recompensa frente al aumento del VRR indica que los bloques incorporados en la fase final del entrenamiento son de bajo valor económico, ubicados en la parte superior del mineral, próximos a la ley de corte, lo que

es consistente con una política que prioriza los bloques de mayor valor e incorpora los marginales en último lugar.

La incorporación de la restricción de ángulo incrementa el costo de entrenamiento requerido para alcanzarlo.

Tabla 6.2: Robustez del caso base 10×10 frente a la semilla de entrenamiento (15.000 épocas). Media sobre 3 semillas.

Semilla	VRR	Bloques de mineral	Época del mejor
42	1.000	70/70	14,957
1	0.994	68/70	11,503
7	0.998	69/70	9,499
Media $\pm$ desv.	0.997 $\pm$ 0.002	—	—

Sobre tres semillas de entrenamiento, el caso base recupera un VRR de 0.997 ± 0.002 (Tabla 6.2): las tres corridas convergen al máximo o muy cerca (0.994–1.000), y el resultado no depende de la semilla 42.

6.5. Valorización económica del plan

6.5.1. Beneficio por columna y costo de desarrollo

La Figura 6.15 presenta el beneficio bruto extraído por columna en el plan final, frente a un costo de desarrollo de 50 kUSD por batea. El beneficio de cada columna, entre 152 y 226 kUSD, supera el costo de desarrollo en un factor de al menos tres. Bajo un criterio de evaluación estático, en el que el beneficio de la columna se compara directamente contra el costo, las diez columnas resultarían aprobadas. Este criterio evalúa el beneficio sobre el plan final y sin considerar el momento en que se realiza la extracción.

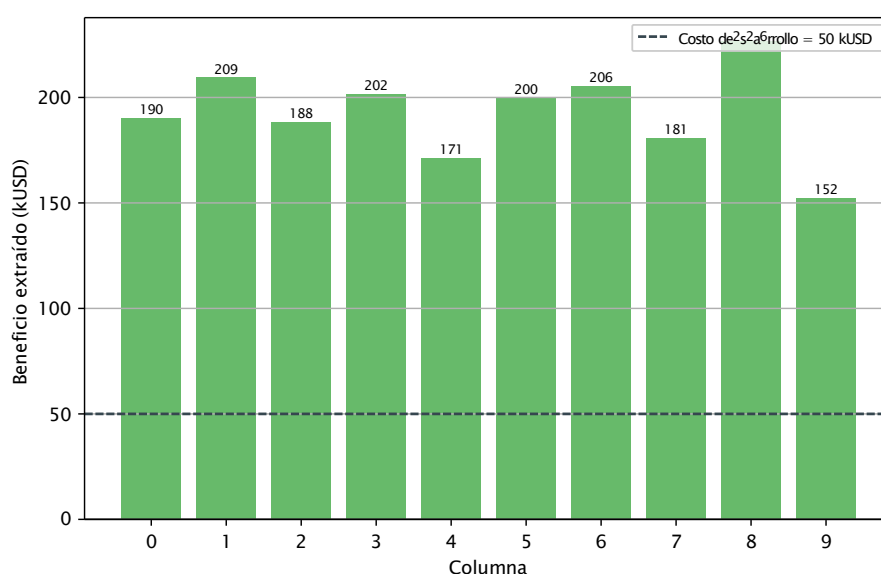


Figura 6.15: Beneficio bruto extraído por columna en el plan final de E7, frente a un costo de desarrollo de 50 kUSD por batea.

6.5.2. Valorización dinámica de las columnas

La Figura 6.16 muestra la evolución de la altura por columna a lo largo del entrenamiento. La altura no constituye una decisión que la política construye de forma progresiva: en el 25 % del entrenamiento únicamente la columna 9 ha sido abierta, con una altura de un bloque; en el del 50 % las alturas de las columnas son heterogéneas, variando entre cinco y siete bloques. En el mejor episodio, las diez columnas alcanzan una altura de siete bloques, lo que coincide con la altura máxima que delimita los bloques clasificados como mineral respecto de los de estéril.

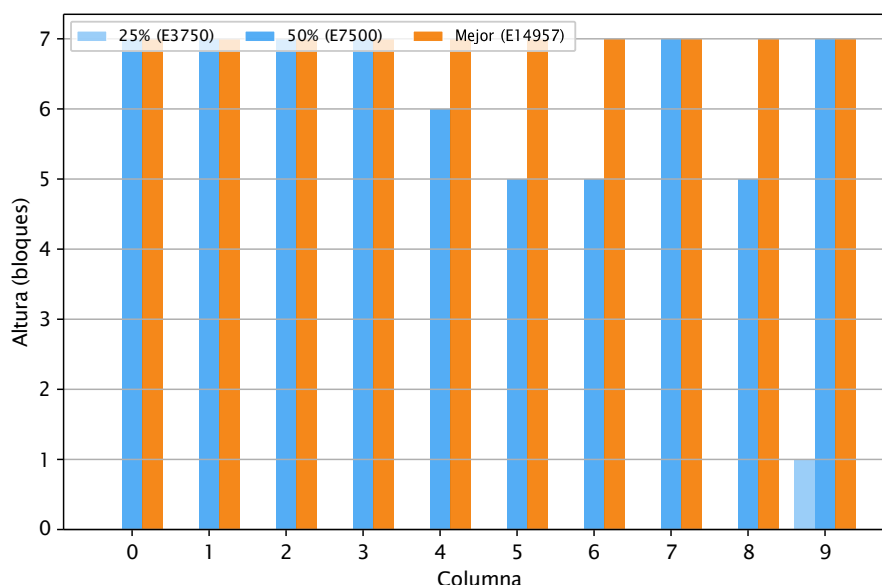


Figura 6.16: Evolución de la altura por columna a través del progreso del entrenamiento (25 %, 50 % y mejor episodio).

La Figura 6.17 presenta el VAN por columna en los mismos porcentajes de entrenamiento evaluados anteriormente, considerando el costo de desarrollo de 50 kUSD. Dado que el beneficio depende de la altura, y esta evoluciona con la política, el VAN de una columna varía a lo largo del entrenamiento. La columna 9 muestra este comportamiento en donde en el 25 % del entrenamiento, con una altura de un bloque, su ingreso descontado (aproximadamente 22 kUSD) no cubre el costo de desarrollo, resultando en un VAN de -28 kUSD, luego en el 50 % del entrenamiento, con una altura de siete bloques, su ingreso descontado (aproximadamente 115 kUSD) determina un VAN de $+65$ kUSD. La misma columna transita de no rentable a rentable a medida que la política la desarrolla. A diferencia del enfoque de delimitación de reservas (Sección 3.5.1), donde el beneficio de cada columna es un valor fijo evaluado de forma independiente a la secuencia, en el marco propuesto la rentabilidad de una columna es una consecuencia de la política de extracción. Un criterio de desarrollo aplicado sobre una altura parcial descartaría columnas que resultan rentables una vez desarrolladas completamente.

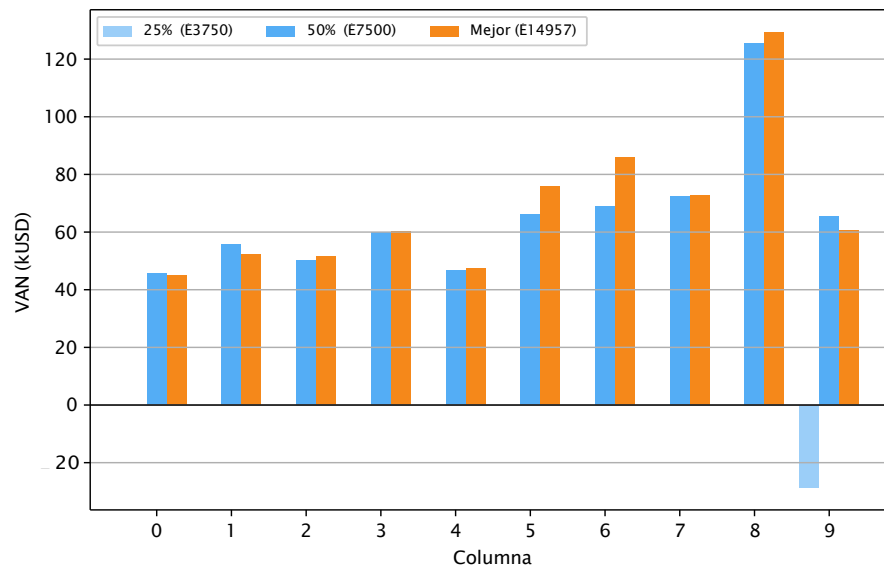


Figura 6.17: Evolución del VAN por columna a través del entrenamiento, considerando un costo de desarrollo de 50 kUSD por batea.

Capítulo 7

Análisis de escalabilidad

Este capítulo evalúa el comportamiento del marco al escalar el perfil de 10×10 a 20×20 , entrenando la etapa E7 ($\mu_{\vartheta} = 1.5$) con todas las restricciones activas y los mismos hiperparámetros. La única variable es el tamaño del yacimiento, de modo que cualquier diferencia de desempeño es atribuible a la escala.

7.1. Efecto del presupuesto de entrenamiento

El caso 20×20 se entrena primero con 10,000 épocas: la secuencia resultante (Figura 7.2) recupera un VRR de 0.570, una fracción baja del valor disponible. Para caracterizar el efecto de la escala se entrenan dos configuraciones adicionales, de 30,000 y 50,000 épocas (Tabla 7.1). Cada escenario es una corrida independiente. El decaimiento de ϵ se ajusta en función de las épocas totales, de modo que un escenario con menos épocas, realiza menos actualizaciones y, por el mismo esquema de ϵ , explora menos.

Tabla 7.1: Entrenamiento en el caso 20×20 .

Épocas	VRR	Bloques de mineral	Época del mejor	Tiempo [min]
10,000	0.570	88 / 213	3.148	18.2
30,000	0.668	93 / 213	28,536	54.5
50,000	0.701	101 / 213	33,112	92.8

El valor recuperado crece con el aumento de las épocas, comenzando con VRR 0.570 al evaluar 10,000 , 0.668 con 30,000 , y 0.701 con 50,000. Sin embargo, también se puede identificar el aumento del tiempo de cómputo.

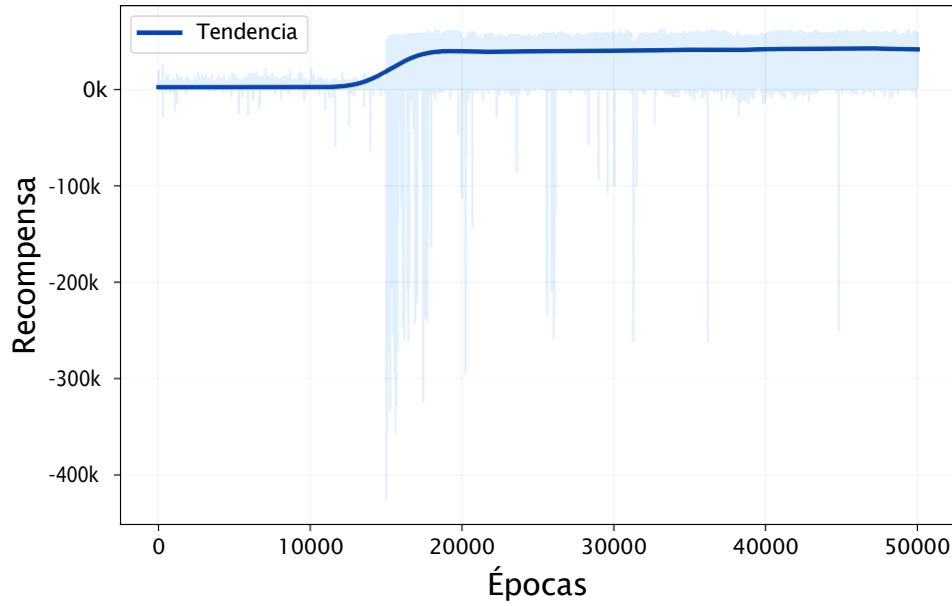


Figura 7.1: Evolución de la recompensa durante el entrenamiento del caso 20×20 (50,000 épocas).

Las secuencias de los tres escenarios (Figura 7.2) muestran que en todos los casos el agente respeta la dirección de propagación con la cual se evalúa, es decir, de derecha a izquierda, la formación del área mínima y una diferencia de altura menor a la de 2 bloques. El desarrollo del frente crece con las épocas, el caso de 50.000 épocas extrae 8 bloques más que el de 30,000 (101 frente a 93). En ninguno se extrae todo el mineral. El desarrollo se concentra en la esquina de inicio y las columnas más lejanas quedan sin abrir.

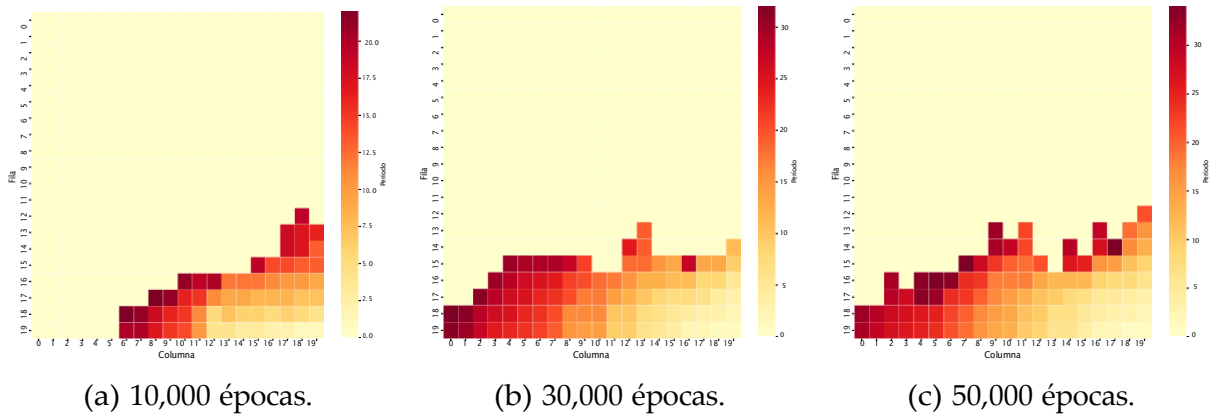


Figura 7.2: Secuencia del mejor episodio para tres escenarios de entrenamiento.

El estancamiento proviene del presupuesto de entrenamiento, ya que no alcanza para poblar las zonas alejadas del punto de inicio, y la política explota la esquina que sí desarrolló, dejando sin abrir mineral rentable.

7.2. Comparación con el caso base

Frente al caso base 10×10 , que recupera la totalidad del valor ($VRR = 1.000$) en 5 min de cómputo, el caso 20×20 alcanza un VRR de 0.701 tras 93 min (Tabla 7.2). El agente

extrae 101 de los 213 bloques de mineral, el cual corresponde al 47 % de los bloques, que concentran el 70 % del valor recuperable. Lo anterior, significa que es selectivo por ley, pero no completa el yacimiento. Ninguno de esos 101 bloques es estéril, por lo que el VRR de 0.701 corresponde a mineral capturado. La diferencia con el caso base proviene del tamaño del espacio de estados, cuya degradación se analiza en la Sección 7.3.

Tabla 7.2: Comparación entre el caso base y el caso escalado (E7).

	10×10	20×20
Bloques totales	100	400
Bloques de mineral	70	213
Épocas de entrenamiento	15,000	50,000
Época del mejor episodio	14,970	33,112
Tiempo de cómputo	5 min	93 min
Valor recuperado (VRR)	1.000	0.701
Bloques de mineral extraídos	70 / 70	101 / 213

Sobre tres semillas de entrenamiento, el VRR del caso 20×20 varía entre 0.607 y 0.701, con media 0.643 ± 0.041 (Tabla 7.3), en donde ninguna semilla extrae estéril. La semilla 42, empleada en el análisis detallado, corresponde al extremo superior del rango, por lo que 0.701 debe leerse como un caso favorable. La época del mejor episodio varía entre 18,562 y 42,920, de modo que tanto el valor recuperado como el momento de convergencia dependen de la inicialización. El método tabular presenta, por lo tanto, sensibilidad a la semilla al escalar en donde converge a un valor cuyo nivel es inestable entre corridas, lo que refuerza la limitación de eficiencia de aprendizaje.

Tabla 7.3: Robustez del caso escalado 20 × 20

Semilla	VRR	Bloques de mineral	Época del mejor
42	0.701	101/213	33,112
1	0.623	92/213	18,562
7	0.607	88/213	42,920
Media ± desv.	0.643 ± 0.041	—	—

7.3. Curva de escalamiento

Para aislar el efecto del tamaño, se entrena una serie de yacimientos con aproximadamente 53 % de mineral y las mismas restricciones, variando solo la dimensión: 10 × 10, 15 × 15 y 20 × 20, cada uno con tres semillas hasta converger. El 10 × 10 de esta serie es un modelo generado, distinto del caso de estudio desarrollado de los capítulos anteriores, por lo que al comparar ambos casos se identifica que se recupera casi la totalidad del valor, lo que confirma que el desempeño en el caso pequeño no depende del modelo de bloques particular.

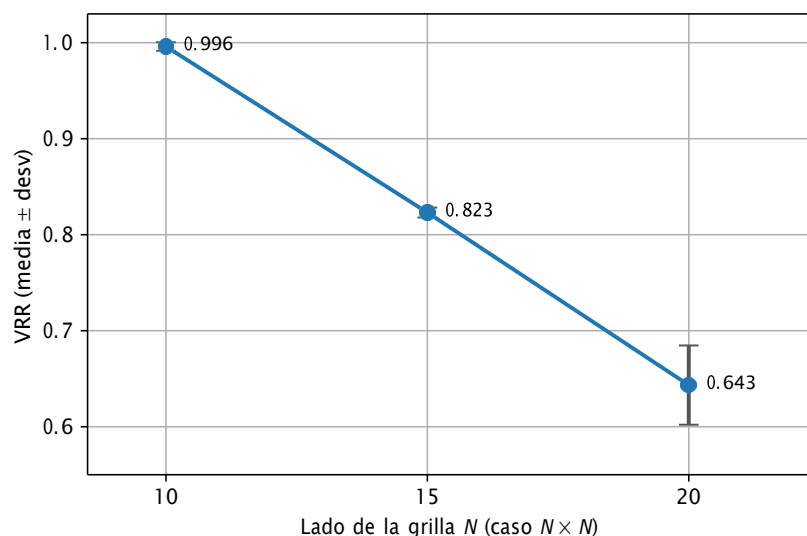


Figura 7.3: Curva de escalamiento: VRR (media $\pm$ desviación sobre 3 semillas) frente al tamaño de la grilla.

Tabla 7.4: Curva de escalamiento: VRR según el tamaño del perfil, etapa E7 ($\mu_\theta = 1.5$), media sobre 3 semillas.

Caso	Bloques de mineral	Casillas de la tabla Q	VRR (media $\pm$ desv)
10×10	53	5,300	0.996 ± 0.004
15×15	123	21,375	0.823 ± 0.005
20×20	213	61,200	0.643 ± 0.041

La curva (Figura 7.3, Tabla 7.4) muestra que el VRR decae de forma sostenida con el tamaño: 0.996 a 0.823 y finalmente 0.643. La caída es progresiva. La variabilidad entre semillas se mantiene baja hasta 15×15 ($\leq 0,005$) y se dispara en 20×20 (0.041). Al aumentar el tamaño, la degradación alcanza primero al valor recuperado y, en la escala mayor, también a la estabilidad.

La causa está en el tamaño del espacio de estados. La tabla Q indexa el estado por (*fila, columna, periodo*): $10 \times 10 \times 53 = 5,300$ casillas en el caso menor, frente a $20 \times 20 \times 153 = 61,200$ en el mayor. Esa indexación es una reducción del estado real del tablero. La configuración de bloques ya extraídos, cuyo tamaño pasa de 2^{100} a 2^{400} . La tabla no captura esa configuración ni transfiere valor entre estados parecidos, de modo que poblar el caso grande exige un muestreo que crece con el espacio real. Esta es la limitación que motiva reemplazar la tabla por una aproximación de función (Capítulo 8).

Capítulo 8

Discusión

La máscara y la penalización blanda son dos maneras de imponer una restricción, y la diferencia entre ellas define cómo el agente resuelve el choque entre geometría y economía. Con penalización, el control de ángulo admite un intercambio, donde el agente comete 2 o 3 violaciones por episodio (el 5.24 % del valor extraído en el mejor caso) cuando el valor que gana con ellas supera lo que le cuestan. Una máscara lo habría prohibido sin importar la economía. Elegir entre una y otra es una decisión de política, de acuerdo a qué restricciones son inviolables y cuáles se pueden transar por valor.

El control del ángulo permitió identificar cómo formular la recompensa de una propiedad que emerge de toda la secuencia. Una primera versión penalizaba el ángulo calculado sobre la extracción que lo revelaba. Como el ángulo depende de todas las aperturas anteriores, el castigo caía sobre una acción que no era la única responsable, y el agente no corregía la geometría. Al reformularlo como reglas de escalón, verificables en cada extracción, el problema se solucionó. Esto es replicable en el sentido de que, cuando la violación de una restricción aparece con retraso, conviene controlarla acción por acción.

En lo económico, la rentabilidad de una columna resultó depender de la secuencia y no ser un dato fijo del modelo de bloques. La columna 9 pasa de un VAN de -28 kUSD con altura parcial (25 % del entrenamiento) a $+65$ kUSD ya desarrollada (50 %). Esto la separa del criterio de envolvente estática, que fija la rentabilidad antes de conocer la secuencia.

El límite del método aparece al escalar, y está en la representación del estado. La tabla indexa cada estado por (*fila, columna, periodo*) y lo aprende por separado, sin trasladar lo aprendido a los estados vecinos, por lo que cubrir un yacimiento más grande exige muchos más episodios. En 10×10 el costo es solo tiempo: la penalización de ángulo retrasa la convergencia de 10,000 a 15,000 épocas para obtener el mismo VRR de 1.000. En 20×20 , en cambio, el presupuesto evaluado no alcanza para poblar la tabla y la política converge a un VRR de 0.643 ± 0.041 , inestable entre semillas. El techo no lo imponen las restricciones, sino la eficiencia de muestreo de la tabla; de ahí la aproximación de función que se propone en las conclusiones.

Capítulo 9

Conclusiones

La implementación y validación del agente de Q-learning demuestra que el aprendizaje por refuerzo es una opción viable para determinar secuencias de extracción factibles en *block/panel caving*, siempre que el entorno MDP modele las propiedades físicas del método y sus características operacionales. Sobre el caso de estudio, el agente recupera la totalidad del valor económico disponible ($VRR = 1.000$) respetando simultáneamente todas las restricciones. La factibilidad geomecánica no se impone sobre el plan: emerge de la dinámica del entorno con el que el agente interactúa.

Se desarrollaron tres elementos metodológicos. Primero, un currículum incremental que incorpora una restricción por etapa y permite aislar y cuantificar su efecto sobre la política. Segundo, una formulación del control de ángulo mediante reglas locales de escalón, que resuelve la asignación de crédito asociada a una propiedad geométrica emergente. Tercero, una valorización económica dinámica, en la que la rentabilidad de cada columna es consecuencia de la secuencia de extracción y no un dato fijo del modelo de bloques.

La política obtenida corresponde a una solución calibrada al conjunto de recompensas, penalizaciones y máscaras del entorno. Al modificar la tasa de descuento, un multiplicador de penalización o una restricción geotécnica, la función de recompensa y el espacio de acciones se actualizan, y el agente converge a una nueva política sin reprogramar la lógica de decisión. La ecuación de Bellman traduce el impacto diferido de cada decisión en una señal inmediata que guía la política. Extraer un bloque de bajo valor para habilitar el acceso a mineral rentable es un caso de ese mecanismo.

El alcance de este resultado está acotado por la representación tabular del estado. Al escalar el perfil de 10×10 a 20×20 , la política converge a un valor subóptimo ($VRR = 0.643 \pm 0.041$ sobre tres semillas, con 0.701 en el mejor caso) tras un cómputo cercano a 19 veces mayor. La tabla no generaliza entre estados vecinos y el costo de muestreo crece con el tamaño del problema. El límite identificado corresponde a la eficiencia de aprendizaje.

Es posible evaluar alternativas que reduzcan el costo computacional: acotar el espacio de estados descartando bloques de bajo valor esperado, modificar la exploración mediante decaimiento por estado o conteo de visitas.

De persistir la limitación, el paso siguiente es una función de valor aproximada mediante redes neuronales, combinada con enmascaramiento. Se plantea, asimismo, extender el marco a casos tridimensionales, donde la geometría del frente y el ángulo de quiebre operan en tres ejes que el modelo bidimensional no representa.

Referencias

- Aravena, M., Palma, S., Vivanco-Avaria, F. J., and Nelis, G. (2025). Quantification of fines migration in block caving mining through laboratory experiments. *Results in Engineering*, 25:103576.
- Atlas Copco (2007). *Mining Methods in Underground Mining*. Atlas Copco, 2 edition.
- Bellman, R. (1957). A markovian decision process. *Journal of Mathematics and Mechanics*, 6(5):679–684.
- Brown, E. T. (2003). *Block Caving Geomechanics*. Julius Kruttschnitt Mineral Research Centre, The University of Queensland, Brisbane, Australia.
- Bustamante, D., Vargas, E., Huenul, E., and Meléndez, M. (2021). Metodología cálculo de envolvente de recursos minerales mina el teniente. In *Proceedings of Geomin-Mineplanning*.
- Castro, R. L. and Paredes, P. S. (2014). Empirical observations of dilution in panel caving. *Journal of the Southern African Institute of Mining and Metallurgy*, 114(6):455–462.
- Castro, R. L. and Whiten, W. J. (2007). A new cellular automaton to model gravity flow in block caving based on physical modelling observations. In *APCOM 2007*, pages 233–241, Santiago, Chile. Gecamin Ltda.
- Chimunhu, P., Topal, E., Ajak, A. D., and Asad, W. (2022). A review of machine learning applications for underground mine planning and scheduling. *Resources Policy*, 77:102693.
- Codelco (2009). Criterios específicos de diseño, proyecto mina chuquicamata subterránea (msc-ico-vcpch-0000-gen-cdi-290, rev. p). Estudio de prefactibilidad, Vicepresidencia Corporativa de Proyectos, Codelco Chile, Santiago, Chile.
- Consejo Minero (2026). Cifras actualizadas de la minería en Chile. Consultado el 30 de junio de 2026.
- Cornejo, J. and Pardo, C. (2014). Management indicators for the cave geometry control, El Teniente mine. In Castro, R., editor, *Proceedings of the Third International Symposium on Block and Sublevel Caving (Caving 2014)*, pages 243–249, Santiago, Chile. Universidad de Chile.
- de Beer, W. (2022). Quantitative correlation of production and seismic response in block caves. *Caving 2022: Proceedings of the 5th International Symposium on Block and Sublevel Caving*, pages 473–486.

- Diering, T., Richter, O., and Villa, D. (2010). Block cave production scheduling using pcbc. In *Proceedings of the SME Conference, Phoenix, USA*.
- Flores, G. and Catalan, A. (2019). A transition from a large open pit into a novel “macroblock variant” block caving geometry at chuquicamata mine, codelco chile. *Journal of Rock Mechanics and Geotechnical Engineering*, 11(3):549–561.
- Gómez, R. and Labbé, E. (2019). An option for the change to massive underground mining: the current chilean methodology of block caving. *Boletín Geológico y Minero*, 130(1):181–198.
- Hamrin, H. (1982). Choosing an underground mining method. In Hustrulid, W. A., editor, *Underground Mining Methods Handbook*, chapter 1.6, pages 88–112. Society of Mining Engineers of AIME, New York.
- Huang, C., Chen, G., and Gong, Y. (2021). Delay-constrained buffer-aided relay selection in the internet of things with decision-assisted reinforcement learning. *IEEE Internet of Things Journal*, 8(13):10198–10208.
- Huo, D., Sari, Y. A., Kealey, R., and Zhang, Q. (2023). Reinforcement learning-based fleet dispatching for greenhouse gas emission reduction in open-pit mining operations. *Resources, Conservation and Recycling*, 188:106664.
- Johnson, T. B. (1968). *Optimum Open Pit Mine Production Scheduling*. PhD thesis, Operations Research Department, University of California, Berkeley.
- Kvapil, R. (2008). *Gravity Flow in Sublevel and Panel Caving - A Common Sense Approach*. Luleå University of Technology Press, Luleå, Sweden.
- Laubscher, D. (1994). Cave mining — the state of the art. *Journal of the Southern African Institute of Mining and Metallurgy*, 94:279–293.
- Laubscher, D. (2000). *Block Cave Manual*. Julius Kruttschnitt Mineral Research Centre.
- Lerchs, H. and Grossmann, I. F. (1965). Optimum design of open-pit mines. *CIM Bulletin*, 58(633):47–54.
- Li, N., Ding, Y., Dai, B., Wang, L., Chang, J., Ye, H., and Wang, Q. (2026). Deep reinforcement learning framework for the underground mine short-term production scheduling problem. *Swarm and Evolutionary Computation*, 100:102229.
- Mnih, V., Kavukcuoglu, K., Silver, D., Rusu, A. A., Veness, J., Bellemare, M. G., Graves, A., Riedmiller, M., Fidjeland, A. K., Ostrovski, G., Petersen, S., Beattie, C., Sadik, A., Antonoglou, I., King, H., Kumaran, D., Wierstra, D., Legg, S., and Hassabis, D. (2015). Human-level control through deep reinforcement learning. *Nature*, 518(7540):529–533.
- Nezhadshahmohammad, F., Khodayari, H., and Pourrahimian, Y. (2017). Draw rate optimisation in block cave production scheduling using mathematical programming. *Proceedings of the First International Conference on Underground Mining Technology, ACG*.
- Noriega, R. and Boisvert, J. (2025). Optimization of underground mining production layouts considering geological uncertainty using deep reinforcement learning. *Engineering Applications of Artificial Intelligence*, 139:109493.

- Noriega, R., Pourrahimian, Y., and Askari-Nasab, H. (2025). Deep reinforcement learning based real-time open-pit mining truck dispatching system. *Computers & Operations Research*, 173:106815.
- Noriega, R., Pourrahimian, Y., and Liu, W. V. (2018). Optimisation of the undercut level elevation in block caving mines using a mathematical programming framework. In Potvin, Y. and Jakubec, J., editors, *Caving 2018: Proceedings of the Fourth International Symposium on Block and Sublevel Caving*, pages 363–372, Perth. Australian Centre for Geomechanics.
- Oyanader Willson, J. A. (2015). Estudio de perfil para la explotación de recursos sur codelco división el teniente. Memoria para optar al título de ingeniero civil de minas, Universidad de Chile. Profesor guía: Rojas Gajardo, P.
- Pierce, M. E. (2010). *A Model for Gravity Flow of Fragmented Rock in Block Caving Mines*. PhD thesis, The University of Queensland, Brisbane, Australia. Summary of a thesis submitted for the degree of Doctor of Philosophy.
- Pourrahimian, Y. and Askari-Nasab, H. (2012). *An Application of Mixed Integer Linear Programming for Block Cave Production Scheduling*, volume 340. MOL Report Four.
- Ranjith, P. G., Zhao, J., Ju, M., De Silva, R. V. S., Rathnaweera, T. D., and Bandara, A. K. M. S. (2017). Opportunities and challenges in deep mining: A brief review. *Engineering*, 3(4):546–551.
- Russell, S. J., Norvig, P., Corchado Rodríguez, J. M., and Joyanes Aguilar, L. (2004). *Inteligencia artificial: un enfoque moderno*. Pearson Prentice Hall.
- Saavedra, F., Morales, N., Nelis, G., and Gómez, R. (2023). A fast method to find smooth economic envelopes for block and panel caving mines. *Resources Policy*, 83:103703.
- Smoljanovic, M., Rubio, E., and Morales, N. (2011). Panel caving scheduling under precedence constraints considering mining system. In *Proceedings of the 35th International Symposium on the Application of Computers and Operations Research in the Minerals Industry (APCOM)*, pages 407–417, Wollongong, Australia.
- Susaeta, A. (2004). Theory of gravity flow (part 1). In *Proceedings of MassMin 2004*, pages 167–172, Santiago, Chile. Instituto de Ingenieros de Chile.
- Susaeta, A., Rubio, E., Henríquez, J., and País, G. (2008). Dilution behaviour at Codelco panel cave mines. In Schunnesson, H. and Nordlund, E., editors, *MassMin 2008: Proceedings of the 5th International Conference and Exhibition on Mass Mining*, pages 167–178, Luleå, Sweden. Luleå University of Technology.
- Sutton, R. S. and Barto, A. G. (2015). Reinforcement learning: An introduction: Second edition, in progress. Draft.
- Topal, E. (2008). Early start and late start algorithms to improve the solution time for long-term underground mine production scheduling. *Journal of the South African Institute of Mining and Metallurgy*, 108(2):99–107.

- Vienne, J. (2005). Curso: Planificación minera subterránea. Material de curso, Departamento de Ingeniería en Minas, Universidad de Santiago de Chile y División Andina, Codelco-Chile.
- Villa, D. (2014). Mine sequence optimization for block caving using the concept of “best and worst case”. White paper, Dassault Systèmes GEOVIA.
- Watkins, C. J. C. H. (1989). *Learning from Delayed Rewards*. PhD thesis, King’s College, University of Cambridge.
- Watkins, C. J. C. H. and Dayan, P. (1992). Q-learning. *Machine Learning*, 8(3):279–292.
- Yaakoubi, Y. and Dimitrakopoulos, R. (2023). Learning to schedule heuristics for the simultaneous stochastic optimization of mining complexes. *Computers & Operations Research*, 159:106349.