

EVALUACIÓN DE LA CAPACIDAD DE LA IA AGÉNTICA EN UN ESTUDIO DE CASO DE PROGRAMACIÓN HIDROTÉRMICA ESTOCÁSTICA

Julio Alberto Silva Dias y Mario Veiga Ferraz Pereira

Resumen

La inteligencia artificial agéntica está emergiendo como un paradigma en el que modelos con capacidad de razonamiento interactúan con herramientas externas, simulaciones y fuentes de datos para ejecutar tareas analíticas de múltiples pasos. Para las aplicaciones en sistemas de energía, esto plantea la cuestión de si los agentes de IA pueden operar de forma eficaz en entornos analíticos estructurados, utilizando capacidades específicas del dominio para explorar problemas de decisión complejos, en lugar de actuar meramente como envoltorios (wrappers) en torno a solucionadores (solvers) existentes.

Este artículo investiga esta cuestión mediante un estudio de caso de programación hidrotérmica estocástica, que combina incertidumbre, acoplamiento intertemporal, restricciones físicas y compromisos (trade-offs) operativos económicamente relevantes, a la vez que ofrece referencias confiables de optimización, como la programación dinámica dual estocástica (SDDP). Proponemos una arquitectura agéntica orientada a capacidades, en la que un agente de razonamiento interactúa con el modelo hidrotérmico únicamente a través de capacidades de dominio controladas y tipadas, sin acceso a métodos de programación dinámica preprogramados, valores del agua precomputados, variables duales o información interna del solucionador.

La arquitectura se evalúa sobre una representación simplificada, pero estructuralmente realista, del Sistema Interconectado Nacional brasileño. A lo largo de diez sesiones independientes, el agente infirió la lógica del valor del agua a partir de la retroalimentación de las simulaciones, construyó representaciones aproximadas del costo futuro y desplegó políticas de preservación del almacenamiento que mejoraron sustancialmente el comportamiento miope. En relación con un valor de referencia (benchmark) SDDP calculado de forma independiente, la mejor sesión alcanzó una brecha media de costo del 1,7%, mientras que la brecha media de las diez sesiones fue de aproximadamente el 6,0%.

Estos resultados indican que la orquestación estructurada de capacidades puede permitir que los agentes de razonamiento funcionen como capas analíticas integradas en entornos de modelado de sistemas de energía, siendo más valiosos para problemas en los que las formulaciones analíticas son incompletas, difíciles de especificar o insuficientes para capturar todo el proceso de exploración de la solución.

Palabras clave: IA agéntica, programación hidrotérmica, optimización estocástica, valor del agua, SDDP

Introducción

La inteligencia artificial agéntica está emergiendo como un paradigma en el que modelos de lenguaje con capacidad de razonamiento interactúan con herramientas externas, simulaciones, fuentes de datos

y entornos computacionales para ejecutar tareas analíticas de múltiples pasos. En lugar de producir únicamente salidas de texto estático, estos sistemas pueden seleccionar acciones estructuradas, invocar herramientas, interpretar las observaciones devueltas, revisar hipótesis y coordinar flujos de trabajo analíticos de múltiples pasos hacia un objetivo definido por el usuario [1]. Los marcos (frameworks) agénticos recientes suelen separar un núcleo de inferencia basado en un modelo de lenguaje — responsable del razonamiento, la planificación y la selección de acciones— de una capa de orquestación, o Harness, que gestiona las llamadas a herramientas, las observaciones, la memoria y el estado de ejecución a lo largo de los pasos [4, 2, 3]. Esta estructura general se ilustra en la Fig. 1, que muestra el núcleo de inferencia del LLM y el Harness que lo rodea.

Figura 1: Estructura general de un bucle de razonamiento agéntico basado en LLM.

La programación hidrotérmica proporciona un escenario adecuado para investigar esta cuestión. En los sistemas hidrotérmicos, las decisiones actuales de liberación de los embalses reducen los costos presentes de generación térmica, pero también afectan la exposición futura a la escasez, al despacho térmico costoso y a los déficits de energía. El agua almacenada tiene, por lo tanto, un costo de oportunidad, comúnmente representado mediante valores del agua o funciones de costo futuro. El problema combina afluencias estocásticas, dinámica de almacenamiento, restricciones de transmisión, sustitución térmica y compromisos operativos de largo plazo. Al mismo tiempo, cuenta con referencias bien establecidas de optimización estocástica, en particular la Programación Dinámica Dual Estocástica

(SDDP), que proporciona un valor de referencia riguroso para evaluar la calidad de las políticas operativas [5, 6].

La programación dinámica estocástica clásica proporciona la base conceptual para este problema de decisión secuencial, pero su aplicación directa está limitada por la maldición de la dimensionalidad. En los sistemas hidrotérmicos realistas, el vector de estado incluye los niveles de almacenamiento de los embalses, las condiciones hidrológicas y otras variables intertemporales, lo que hace que el número de estados posibles crezca exponencialmente con el tamaño del sistema y con la representación de la incertidumbre. Los métodos de descomposición como el SDDP evitan la enumeración completa de estados al aproximar las funciones de costo futuro y siguen siendo la referencia estándar para la planificación de la operación hidrotérmica de gran escala. Los métodos de aprendizaje automático, incluido el aprendizaje por refuerzo, también se han estudiado para la toma de decisiones secuenciales bajo incertidumbre. Sin embargo, en este artículo, estos métodos sirven principalmente como referencias conceptuales y valores de referencia: el objetivo no es proponer un nuevo solucionador de optimización hidrotérmica.

La cuestión central abordada en este artículo es, en cambio, arquitectónica y metodológica: ¿puede un agente de razonamiento, restringido a capacidades de dominio controladas, explorar un entorno analítico de sistema de energía y construir una política operativamente significativa sin acceso a un método de programación dinámica estocástica preprogramado, variables duales, valores del agua precomputados o aspectos internos del solucionador? Esta cuestión se refiere a cómo debe exponerse la funcionalidad analítica a un agente de IA de modo que el razonamiento, el cálculo, el control y la auditabilidad permanezcan separados. Está alineada con la noción emergente de habilidades o capacidades específicas del dominio: operaciones modulares que exponen funcionalidades seleccionadas de un entorno a un agente a través de interfaces estructuradas.

Para investigar esta cuestión, proponemos una arquitectura agéntica orientada a capacidades para la interacción asistida por IA con entornos analíticos de sistemas de energía. En esta arquitectura, el conocimiento del dominio y las operaciones del modelo se exponen al agente a través de capacidades estructuradas y tipadas, en lugar de mediante la ejecución irrestricta de código o el acceso directo a los aspectos internos del solucionador. El agente puede inspeccionar el sistema, consultar estadísticas de escenarios, ejecutar simulaciones, evaluar políticas candidatas y registrar artefactos relevantes para la política, pero todos los cálculos los ejecuta el entorno analítico controlado. El agente, por lo tanto, no reemplaza el modelo de dominio ni el solucionador de optimización; actúa como una capa de orquestación analítica que formula hipótesis, invoca capacidades, interpreta salidas y construye progresivamente una lógica operativa aproximada.

La arquitectura propuesta se instancia en un entorno de programación hidrotérmica estocástica basado en una representación simplificada, pero estructuralmente realista, del Sistema Interconectado Nacional brasileño. El agente interactúa con el modelo solo a través de capacidades controladas y no recibe acceso a una implementación del SDDP, rutinas de programación dinámica, información dual, valores del agua precomputados o estado interno del solucionador. La política resultante se evalúa entonces mediante simulación estocástica y se compara con dos referencias: una política miope que ignora el valor futuro del agua y un valor de referencia basado en SDDP calculado de forma independiente. Este escenario permite evaluar el comportamiento del agente en un problema cuya estructura de optimalidad es bien comprendida y para el cual se dispone de referencias confiables.

La motivación más amplia es evaluar si los agentes de razonamiento pueden funcionar como capas analíticas controladas en torno a modelos de sistemas de energía. No se espera que tales agentes reemplacen los métodos formales de optimización. En cambio, pueden apoyar la exploración, la interpretación, el análisis de sensibilidad, la interrogación de modelos y la construcción de políticas en escenarios en los que el proceso de decisión relevante va más allá de una única formulación de optimización totalmente especificada. La programación hidrotérmica se utiliza aquí como un caso de

validación con valor de referencia, antes de considerar problemas de sistemas de energía más abiertos, en los que las formulaciones analíticas pueden ser incompletas, difíciles de construir o insuficientes para capturar todo el proceso de exploración de la solución.

El artículo aborda las siguientes preguntas de investigación:

RQ1. ¿Puede un agente de razonamiento usar capacidades de dominio estructuradas para construir una política operativamente significativa en un problema estocástico de programación de sistemas de energía?

RQ2. En el caso de la programación hidrotérmica, ¿puede el agente inferir una lógica del valor del agua económicamente significativa a partir de la retroalimentación de las simulaciones, en lugar de información dual explícita o salidas de optimización precomputadas?

RQ3. ¿Cuán reproducible es el proceso de construcción de políticas a lo largo de sesiones independientes de agentes de razonamiento bajo la misma especificación de tarea y la misma interfaz de capacidades?

RQ4. ¿Qué propiedades arquitectónicas se requieren para apoyar la evaluación controlada, auditable y modular de agentes de razonamiento en entornos analíticos de sistemas de energía?

Las principales contribuciones del artículo son:

1. Una arquitectura agéntica orientada a capacidades para la interacción asistida por IA con entornos analíticos de sistemas de energía.
2. Una instanciación de esta arquitectura en un entorno de programación hidrotérmica estocástica, conectando el flujo de trabajo agéntico con la formulación clásica de optimización estocástica multietapa.
3. Una interfaz de capacidades estructurada que expone operaciones de dominio controladas para la inspección del modelo, la exploración basada en simulación, el registro de políticas y la evaluación estocástica, a la vez que oculta los aspectos internos del solucionador, las variables duales y los valores del agua precomputados.
4. Una evaluación experimental con valor de referencia que muestra que un agente de razonamiento puede construir políticas operativas hidrotérmicas de preservación del almacenamiento mediante la orquestación estructurada de capacidades, con evaluación comparativa frente a la operación miope y a referencias basadas en SDDP calculadas de forma independiente.

El resto del artículo se organiza de la siguiente manera. La Sección 2 revisa los trabajos relacionados. La Sección 3 define el problema de operación hidrotérmica estocástica multietapa. La Sección 4 presenta la arquitectura orientada a capacidades, el entorno analítico y el protocolo experimental. La Sección 5 presenta el caso de prueba y los resultados. La Sección 6 analiza las implicaciones y limitaciones. La Sección 7 concluye.

Trabajos Relacionados

Trabajos recientes sobre modelos de lenguaje que utilizan herramientas han demostrado que los grandes modelos de lenguaje pueden combinar el razonamiento mediado por el lenguaje con acciones computacionales externas, incluidas las llamadas a herramientas, el acceso a APIs y la interacción estructurada con entornos computacionales [1, 7, 8]. Las arquitecturas agénticas relacionadas organizan este proceso mediante módulos para la especificación de roles, la memoria, la planificación y la selección de acciones, y enfatizan cada vez más las habilidades o capacidades reutilizables como la

interfaz entre el modelo de razonamiento y su entorno [2, 3]. Los marcos industriales han seguido una dirección similar al proporcionar patrones de referencia para combinar modelos fundacionales, acceso a herramientas, orquestación de flujos de trabajo y mecanismos de salvaguarda (guardrails) [9, 10].

Aunque muchas demostraciones de agentes que usan herramientas se centran en la búsqueda web, la programación, la respuesta a preguntas o la recuperación de información, las aplicaciones de ingeniería de altas consecuencias requieren garantías más robustas de control, auditabilidad y reproducibilidad. En tales escenarios, el uso irrestricto de herramientas suele ser inadecuado: el agente debe interactuar con entornos computacionales validados a través de interfaces bien definidas. Los protocolos de interoperabilidad como el Model Context Protocol (MCP) ilustran esta dirección al exponer herramientas externas a través de interfaces tipadas [11]. Esto motiva el uso de interfaces de capacidades controladas para la interacción agéntica con modelos de sistemas de energía.

El aprendizaje automático también se ha aplicado ampliamente en sistemas de energía, incluida la previsión, la aproximación del flujo de potencia óptimo y el control [12, 13]. El aprendizaje por refuerzo se ha estudiado para la operación de embalses y el control de sistemas de potencia [14, 15]. Estos enfoques generalmente aprenden políticas o aproximaciones de funciones mediante datos, episodios de simulación, actualizaciones de gradiente o procedimientos explícitos de entrenamiento. El enfoque investigado aquí es diferente: el agente no se entrena como un controlador hidrotérmico y no aprende una política mediante episodios repetidos. En cambio, utiliza el razonamiento y la orquestación de capacidades para inspeccionar un entorno estructurado de sistema de energía, probar hipótesis y construir una aproximación relevante para la política durante una sesión analítica.

La planificación de la operación hidrotérmica se ha abordado desde hace tiempo mediante la programación dinámica estocástica y la descomposición. El SDDP [5] representa la función de costo futuro mediante cortes afines y se utiliza ampliamente en sistemas hidrotérmicos de gran escala. Trabajos posteriores extendieron el método al análisis de convergencia, la aversión al riesgo y los modelos intertemporales de afluencias [6, 16, 17]. Estos métodos proporcionan el valor de referencia de optimización para el presente trabajo y se utilizan como referencias externas, no como herramientas disponibles para el agente.

Problema de Operación Hidrotérmica Estocástica Multietapa

Considérese un horizonte finito $t = 1, \dots, T$. Sea V_t el almacenamiento de los embalses, H_t la información hidrológica y $X_t = (V_t, H_t)$ el estado del sistema. La incertidumbre se representa mediante ξ_t , incluidas las afluencias y, posiblemente, la demanda o la generación renovable. En cada etapa, las decisiones u_t incluyen la generación hidroeléctrica, la generación térmica, los flujos de interconexión, el vertimiento, el déficit y el almacenamiento de la etapa siguiente. El problema de la etapa es

$$\min_{u_t \in U_t(X_t, \xi_t)} c_t(u_t, \xi_t) + Q_{t+1}(X_{t+1}), \quad (1)$$

donde U_t contiene las restricciones de balance hídrico, generación, transmisión y demanda.

Para cada subsistema hidroeléctrico r , el almacenamiento evoluciona como

$$V_{t+1,r} = V_{t,r} + a_{t,r}(\xi_t) - q_{t,r} - s_{t,r}, \quad g_{t,r}^H = \eta_{t,r} q_{t,r}, \quad (2)$$

con afluencia $a_{t,r}$, caudal turbinado $q_{t,r}$, vertimiento $s_{t,r}$ y coeficiente de producción $\eta_{t,r}$. Para cada área eléctrica n ,

$$\sum_{r \in R_n} g_{t,r}^H + \sum_{g \in G_n} g_{t,g}^T + R_{t,n}(\xi_t) + \sum_{\ell \in I_n} f_{t,\ell} - \sum_{\ell \in O_n} f_{t,\ell} + d_{t,n} = D_{t,n}(\xi_t). \quad (3)$$

El conjunto factible también incluye límites de almacenamiento, generación hidroeléctrica, generación térmica, flujos de transmisión, vertimiento y déficits. El costo inmediato es

$$c_t(u_t, \xi_t) = \sum_n \sum_{g \in G_n} c_{t,g}^T g_{t,g}^T + \sum_n c_{t,n}^D d_{t,n}. \quad (4)$$

El problema estocástico multietapa es

$$\min_{\pi} \mathbb{E} \left[\sum_{t=1}^T c_t(u_t, \xi_t) + \Phi(V_{T+1}) \right], \quad u_t = \pi_t(X_t, \xi_t), \quad (5)$$

con políticas no anticipativas. La recursión de Bellman es

$$Q_t(X_t) = \mathbb{E}_{\xi_t|X_t} \left[\min_{u_t \in U_t(X_t, \xi_t)} \{c_t(u_t, \xi_t) + Q_{t+1}(X_{t+1})\} \right], \quad (6)$$

con la condición terminal $Q_{T+1}(X_{T+1}) = \Phi(V_{T+1})$. El valor marginal del agua almacenada es

$$\lambda_{t,r}(X_t) = \frac{\partial Q_t(X_t)}{\partial V_{t,r}}. \quad (7)$$

La discretización directa por SDP es intratable, ya que R embalses discretizados en K niveles y producen K^R estados de almacenamiento. El SDDP evita la enumeración completa de estados al aproximar los costos futuros mediante cortes afines:

$$Q_t(V) \approx \max_{k \in K_t} \{a_t^k + (\beta_t^k)^\top V\}. \quad (8)$$

En este artículo, el SDDP se utiliza como referencia conceptual y valor de referencia externo. El agente no recibe acceso a una implementación del SDDP.

Arquitectura Orientada a Capacidades

Principio arquitectónico

La arquitectura propuesta, orientada a capacidades (capability-driven architecture, CDA), separa el proceso de razonamiento del agente de la implementación computacional del modelo analítico de sistema de energía subyacente. El agente formula hipótesis, selecciona acciones analíticas, interpreta observaciones y refina su estrategia a lo largo de una sesión de múltiples pasos, mientras el entorno ejecuta operaciones de dominio validadas, como la inspección del modelo, la simulación, la evaluación de políticas y el registro de artefactos.

Todas las interacciones se producen a través de capacidades tipadas, con esquemas de entrada, esquemas de salida y semántica de ejecución definidos. El agente no tiene acceso a archivos del modelo, aspectos internos del solucionador, ejecución irrestricta de código, variables duales o salidas de optimización precomputadas. Esta separación permite que el agente opere como una capa de razonamiento y orquestación, a la vez que preserve el control, la auditabilidad y una frontera clara entre la inferencia mediada por el lenguaje y el cálculo validado.

Formalmente, una sesión analítica bajo la CDA puede representarse como una trayectoria de interacción

$$\tau = [(a_1, o_1), (a_2, o_2), \dots, (a_K, o_K)], \quad (9)$$

donde a_k denota la invocación de capacidad en el paso k y o_k la observación estructurada correspondiente devuelta por el entorno.

En cada paso, el agente selecciona la siguiente invocación de capacidad con base en todo el historial de interacción acumulado hasta el momento. Sea $h_k = (\mathcal{T}, a_1, o_1, \dots, a_{k-1}, o_{k-1})$ ese historial, donde $\mathcal{T}$ es la especificación de la tarea proporcionada en la inicialización de la sesión. El bucle de razonamiento del agente puede entonces escribirse como

$$a_k = \pi_A(h_k), \quad (10)$$

donde π_A es la función de orquestación del agente: el mecanismo por el cual el modelo de razonamiento interpreta el historial acumulado y selecciona la siguiente invocación de capacidad. Este objeto no debe confundirse con una política operativa hidrotérmica. Gobierna la secuenciación de las acciones analíticas durante la sesión agéntica, incluida la inspección, la simulación, la prueba de hipótesis y la posible modificación de artefactos relacionados con la política, como las aproximaciones de costo futuro. A diferencia de un controlador de aprendizaje por refuerzo entrenado, π_A no se aprende mediante interacción episódica con el sistema hidrotérmico; lo ejecuta el modelo de razonamiento mediante inferencia mediada por el lenguaje sobre el historial de la sesión.

El entorno, a su vez, mantiene un estado s_k compuesto por el modelo analítico cargado, las definiciones de capacidades disponibles y cualquier artefacto analítico con estado creado durante la sesión. En la instanciación hidrotérmica, el artefacto más importante de este tipo es la aproximación de costo futuro actualmente registrada.

Cada invocación de capacidad produce una observación y hace avanzar el estado del entorno de acuerdo con

$$(o_k, s_{k+1}) = \mathcal{E}(a_k, s_k), \quad (11)$$

donde $\mathcal{E}$ es una función de transición determinista. Las invocaciones de solo lectura (inspección, simulación y consultas de estadísticas de escenarios) dejan s_k inalterado. Las invocaciones que codifican políticas actualizan el conjunto de cortes registrado en s_{k+1} , lo que hace que el entorno dependa del estado con respecto al proceso de construcción de la política. La sesión termina en el paso K cuando el agente declara completada la tarea analítica. El estado terminal del entorno s_K contiene los artefactos producidos durante la sesión. En el caso hidrotérmico, estos artefactos incluyen la aproximación de costo futuro registrada. Esta aproximación induce posteriormente una política operativa hidrotérmica cuando se incorpora al problema de despacho de la etapa y se evalúa mediante simulación estocástica, como se describe en la Sección 5.

La trayectoria τ es el principal objeto del análisis post hoc: registra todo el camino analítico, desde la inspección del sistema hasta la validación de la política, y permite la reconstrucción y la auditabilidad del proceso de razonamiento del agente.

La Figura 2 ilustra la estructura general de los componentes de la arquitectura, resumiendo las relaciones formales descritas por las ecuaciones (9)–(11).

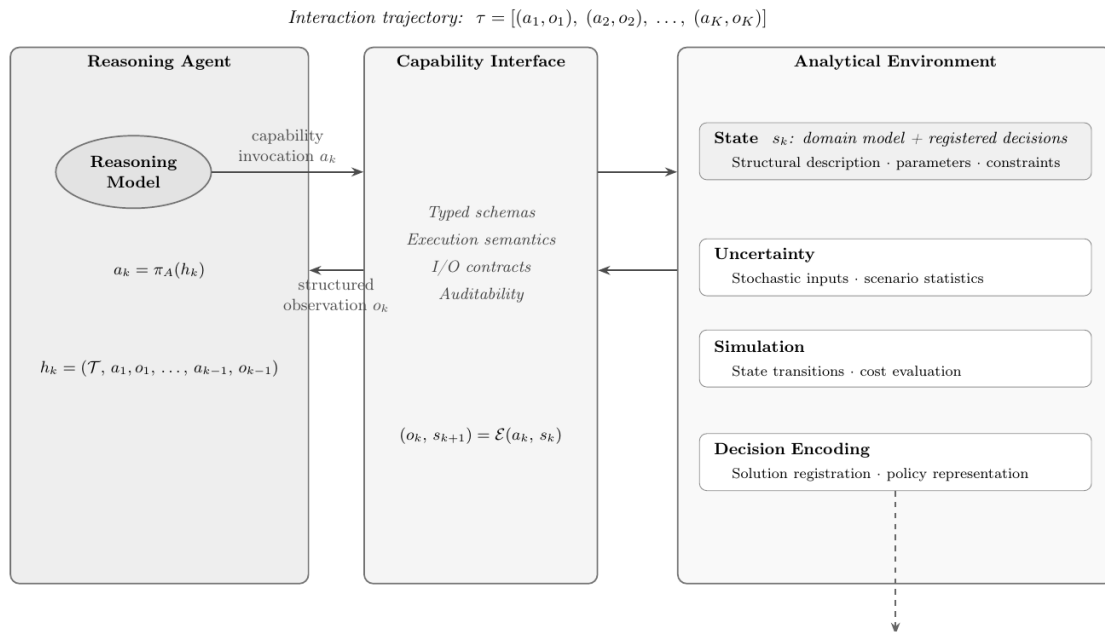


Figura 2: Arquitectura orientada a capacidades.

Instanciación hidrotérmica de la interfaz de capacidades

La instanciación hidrotérmica expone las operaciones de dominio al agente a través de ocho capacidades tipadas, organizadas en cuatro grupos analíticos. El diseño sigue un principio de delimitación deliberada de la información: cada capacidad proporciona la información necesaria para apoyar un paso analítico bien definido, a la vez que oculta los aspectos internos del solucionador, las plantillas algorítmicas, las variables duales y los flujos de datos en bruto que permitirían al agente eludir la tarea de razonamiento. Esta delimitación no es solo una conveniencia técnica; es el mecanismo por el cual el entorno restringe al agente a realizar un razonamiento relevante para el dominio, en lugar de un cálculo arbitrario o un mero envoltorio del solucionador.

Una propiedad arquitectónica adicional es la persistencia de estado de la sesión. El modelo se carga una sola vez en la inicialización de la sesión, fijando el embalse y estableciendo la convención de coordenadas del vector de almacenamiento utilizada en todas las llamadas posteriores de simulación y de codificación de política. Este contrato de datos explícito entre el entorno y el agente garantiza que cada invocación de capacidad opere sobre una representación compartida e inequívoca del sistema físico.

La Tabla 1 resume la interfaz en la práctica.

Tabla 1: Interfaz de capacidades hidrotérmicas: grupos de capacidades, entradas, salidas y roles analíticos.

Grupo	Entradas / salidas principales	Rol analítico
Inspección del sistema	datos del caso $\rightarrow$ embalses, etapas, escenarios	Inicializar la sesión; fijar la convención de coordenadas del vector de almacenamiento
	$\emptyset \rightarrow$ topología completa	Exponer el grafo de subsistemas, bloques de costo, transmisión y demanda
	$\emptyset \rightarrow T, S $	Proporcionar el horizonte de planificación y el número de escenarios

Escenarios hidrológicos	$\emptyset \rightarrow$ media, desviación estándar por nodo por etapa	Caracterizar la estructura distributiva de las afluencias; trayectorias individuales no accesibles
Simulación de la operación	etapa, escenario, $V \rightarrow$ costos, V'	Oráculo controlado de etapa única; solo resultados primales
	escenario $\rightarrow$ trayectorias de costo y almacenamiento	Trayectoria de horizonte completo para una serie de afluencias
	$\emptyset \rightarrow$ resultados agregados	Evaluación por lotes de todos los escenarios
Codificación de política	etapa, $\beta, \Delta \rightarrow$ confirmación	Registrar aproximaciones afines de costo futuro; reemplaza los cortes existentes en esa etapa

Las capacidades de inspección del sistema exponen la descripción estructural estática del modelo hidrotérmico. Una llamada de carga de datos inicializa la sesión y devuelve el nombre ordenado del embalse, el volumen inicial, la capacidad máxima y la unidad física de cada componente de almacenamiento. Este ordenamiento es relevante: define el sistema de coordenadas para todos los vectores de almacenamiento intercambiados en las llamadas posteriores, desde las entradas de simulación de etapa única hasta las matrices de coeficientes de los cortes de política. Una llamada de inspección de topología proporciona la descripción completa del sistema como un documento estructurado, que incluye nodos de subsistemas, arcos de interconexión direccionales con límites de capacidad, bloques de generación térmica con tramos de costo, perfiles de generación renovable y parámetros estacionales de demanda. La imagen resultante coincide con lo que un analista de planificación ensamblaría al revisar la documentación del modelo: topología del sistema, caracterización de los recursos y restricciones operativas.

Las capacidades de escenarios hidrológicos proporcionan una caracterización estocástica agregada de la incertidumbre de entrada. Para cada nodo de entrada —incluidas las afluencias a los embalses, la generación renovable y las perturbaciones de demanda— la capacidad devuelve estadísticas resumen por etapa a lo largo del conjunto de escenarios. Las realizaciones individuales de escenarios no son intencionalmente accesibles a través de esta capacidad. Esta elección tiene dos propósitos. Analíticamente, anima al agente a razonar a partir de la estructura distributiva, en lugar de memorizar trayectorias muestrales individuales. Operativamente, evita inyectar grandes tablas de escenarios en bruto en el historial de interacción, lo que expandiría el contexto del agente con detalles numéricos de bajo nivel y diluiría la señal analítica de alto nivel necesaria para los pasos de razonamiento posteriores. El conjunto completo de escenarios permanece disponible indirectamente a través de las capacidades de simulación y de evaluación por lotes.

Las capacidades de simulación de la operación constituyen el núcleo computacional de la interfaz y ofrecen tres niveles de granularidad analítica. Una llamada de simulación de etapa única evalúa una etapa a partir de un vector de almacenamiento inicial, un índice de etapa y un índice de escenario, devolviendo un vector de costos operativos por nodo y el vector de almacenamiento de fin de etapa resultante. Este oráculo apoya experimentos de perturbación dirigidos: al variar el vector de almacenamiento inicial en llamadas repetidas, manteniendo fijos la etapa y el escenario, el agente puede estimar la sensibilidad local del costo operativo respecto a los niveles de los embalses —el procedimiento numérico que produjo los valores marginales del agua reportados en la Sección 5—. Una llamada de simulación de secuencia ejecuta una trayectoria completa para un escenario de afluencias especificado, devolviendo perfiles de costo y almacenamiento etapa por etapa a lo largo de todo el horizonte de planificación. Una llamada de evaluación por lotes ejecuta el modelo sobre todos los escenarios y es apropiada para la validación final de la política.

Una propiedad definitoria de este grupo es que los costos operativos y las transiciones de estado son observables, pero las variables duales y los precios marginales no lo son. El solucionador de etapa se ejecuta internamente para cada llamada de simulación y devuelve solo resultados primales; los precios sombra no se exponen a través de la interfaz. Esta restricción tiene consecuencias directas para la estrategia analítica del agente. El costo de oportunidad del agua almacenada debe inferirse mediante experimentación activa, y no leerse a partir de la salida del solucionador.

Las capacidades de codificación de política permiten que el agente registre aproximaciones de costo futuro que se incorporan a las llamadas de simulación posteriores. Una llamada de registro de política acepta un índice de etapa, un vector de constantes del lado derecho y una matriz de coeficientes de almacenamiento, donde cada fila define una aproximación afín de costo futuro sobre la función de costo futuro. Esta llamada reemplaza todos los cortes existentes para la etapa objetivo: la acumulación incremental a lo largo de los pasos de iteración es responsabilidad del agente, no del entorno. Los coeficientes de almacenamiento siguen una convención de signos alineada con la economía hidrotérmica: un coeficiente negativo en el embalse i codifica la observación de que agua adicional en ese embalse reduce los costos operativos futuros esperados, es decir, que el agua almacenada tiene un valor marginal positivo. A la inversa, pasar vectores vacíos de coeficientes y de lado derecho borra los cortes de una etapa, lo que permite al agente revisar o reiniciar su aproximación de política durante la sesión analítica.

Esta taxonomía es específica de la aplicación hidrotérmica, pero refleja un principio arquitectónico más amplio: las capacidades deben corresponder a operaciones analíticas significativas en el dominio objetivo, agrupadas en un nivel de abstracción que coincida con la granularidad de razonamiento requerida por la tarea. En este caso, las capacidades expuestas reflejan el flujo de trabajo que siguen los analistas de planificación hidrotérmica: inspeccionar el sistema, caracterizar la estructura de incertidumbre, simular consecuencias operativas bajo condiciones controladas y codificar la lógica intertemporal de la política. La ocultación deliberada de la información dual y de las trayectorias individuales de escenarios garantiza que el agente deba involucrarse con el problema analítico en lugar de recuperar respuestas precomputadas a través de la interfaz.

La capacidad de codificación de política acepta aproximaciones afines de costo futuro de la forma:

$$Q_t(V) \geq \beta_t - \sum_{i \in S} \lambda_i V_i, \quad (12)$$

donde $Q_t(V)$ es la aproximación del costo futuro a partir de la etapa $t + 1$ en adelante, V_i es el almacenamiento de fin de etapa en el subsistema i , λ_i es el valor marginal del agua almacenada, β_t es un intercepto específico de la etapa y S es el conjunto de componentes de almacenamiento representados en el modelo.

Esta representación está estrechamente relacionada con los cortes afines utilizados en los métodos de descomposición como el SDDP. Sin embargo, al agente no se le proporcionó una implementación del SDDP, ningún otro método de programación dinámica estocástica preprogramado, ni una plantilla algorítmica para construir estos cortes. La capacidad solo definía el objeto de política admisible que podía registrarse en el entorno. El proceso de razonamiento utilizado para estimar los valores del agua, elegir los coeficientes y ensamblar la aproximación se dejó a cargo del agente.

Resultados

Tarea experimental y evaluación

Se instruyó al agente para que construyera una política operativa rentable para el sistema hidrotérmico. Se permitió que la sesión continuara hasta que el agente declarara que había completado la construcción

y la validación de la política. El registro de interacción se analizó entonces post hoc. A lo largo de la sesión, el agente tuvo acceso a las descripciones de las capacidades, a los esquemas de entrada-salida y a las salidas de sus propias invocaciones. No recibió una implementación de programación dinámica, como el SDDP, valores del agua precomputados o variables duales, acceso al estado interno del solucionador, ni orientación humana tras la inicialización de la tarea. Toda la información cuantitativa utilizada por el agente se obtuvo, por lo tanto, exclusivamente mediante invocaciones estructuradas de capacidades, lo que distingue este escenario experimental de uno en el que un agente simplemente llama a un solucionador de optimización existente.

Todas las sesiones se realizaron utilizando Claude Sonnet 4.6 (Anthropic), un gran modelo de lenguaje con capacidad de razonamiento en cadena de pensamiento extendida. El modelo se desplegó con el razonamiento extendido habilitado, lo que le permite realizar pasos de razonamiento intermedio explícitos antes de producir cada respuesta o invocación de capacidad. Las sesiones usaron un modo de orquestación automática en el que el modelo de razonamiento principal enruta de forma autónoma subtareas computacionalmente más ligeras —como el análisis de salidas estructuradas y el formateo de invocaciones— a Claude Haiku 4.5, un miembro más pequeño y de menor latencia de la misma familia de modelos, a la vez que conserva la inferencia completa con razonamiento extendido para los pasos analíticos que requieren razonamiento de múltiples pasos. Esta configuración refleja un patrón práctico de despliegue para flujos de trabajo agénticos de largo horizonte, en el que la profundidad del razonamiento y la eficiencia computacional se equilibran automáticamente a nivel de inferencia, sin intervención manual.

El experimento se diseñó para evaluar si el agente podía:

1. Identificar la estructura relevante del sistema hidrotérmico mediante las capacidades de inspección;
2. Inferir la lógica intertemporal del valor del agua mediante simulación;
3. Codificar una aproximación de costo futuro usando la representación de política disponible;
4. Validar la política resultante en escenarios estocásticos de afluencias.

La política resultante se evaluó utilizando el costo operativo de la media muestral, la variabilidad de costo entre escenarios, los eventos de déficit de energía, el uso de generación térmica de emergencia de alto costo, las trayectorias de almacenamiento de los embalses y la comparación a nivel de escenario con dos referencias:

- **Línea base miope:** una política de visión de corto plazo que prioriza la generación hidroeléctrica inmediata sin tener en cuenta el valor futuro del agua, correspondiente a resolver cada etapa de forma independiente con una aproximación de costo futuro nula.
- **Valor de referencia SDDP:** una política obtenida de forma independiente usando una implementación de programación dinámica dual estocástica construida para tal fin.

Dado que los agentes de razonamiento son sistemas estocásticos —cuyas salidas dependen de procesos internos de muestreo que varían entre sesiones incluso bajo prompts y condiciones iniciales idénticos—, el experimento se realizó diez veces de forma independiente, cada ejecución inicializada a partir del mismo modelo y la misma especificación de tarea. Las diez ejecuciones producen diferentes trayectorias de interacción $\tau^{(r)}$, $r = 1, \dots, 10$, lo que conduce potencialmente a diferentes objetos de política $\pi_r^{(r)}$ y a diferentes costos operativos. El análisis de la distribución de resultados entre ejecuciones evita que se extraigan conclusiones a partir de una única sesión, potencialmente atípica.

Caso de prueba hidrotérmico brasileño

El experimento utilizó una representación simplificada, pero estructuralmente realista, del Sistema Interconectado Nacional brasileño (SIN). El sistema se representa mediante cuatro subsistemas interconectados: Sudeste/Centro-Oeste (SECO), Sur (SUL), Nordeste (NE) y Norte (N). El horizonte de planificación abarca 24 etapas mensuales, de enero de 2025 a diciembre de 2026, con diez escenarios estocásticos de afluencias.

El caso de prueba incluye embalses hidroeléctricos agregados, bloques de generación térmica en capas, generación renovable, perfiles de demanda y enlaces de transmisión interregionales. No pretende reproducir el modelo oficial completo de operación brasileño. En cambio, proporciona un valor de referencia controlado con las principales características estructurales necesarias para el análisis de la operación hidrotérmica: acoplamiento espacial, dinámica de almacenamiento, incertidumbre hidrológica, sustitución térmica y penalizaciones por escasez.

La Tabla 2 resume los principales parámetros del caso.

Parámetro	Valor
Horizonte de planificación	Enero de 2025 – Diciembre de 2026
Etapas	24 etapas mensuales
Subsistemas	SECO, SUL, NE, N
Escenarios hidrológicos	10 escenarios estocásticos de afluencias
Representación de la política	Cortes afines de costo futuro

Tabla 2: Principales parámetros del caso de prueba.

La Tabla 3 resume las principales características de los subsistemas utilizadas en el caso de prueba.

Subsistema	Hidro máx. (MW)	Embalse (GWh)	Almacenamiento inicial	Rango de demanda (MW)
SECO	35.000	146.000	75%	42.800–50.000
SUL	17.000	14.600	70%	12.900–15.800
NE	10.000	37.960	50%	12.400–14.000
N	22.000	10.950	85%	7.400–8.400

Tabla 3: Características agregadas de los subsistemas en el caso de prueba hidrotérmico.

El caso de prueba también incluye enlaces de transmisión entre los subsistemas. Los principales límites direccionales son 15.600 MW de N a SECO, 2.000 MW de SECO a N, 6.200 MW de N a NE, 13.000 MW de NE a SECO y 8.000 MW entre SECO y SUL. La generación térmica se representa mediante bloques en capas con costos crecientes. El subsistema SECO desempeña un papel central porque su bloque térmico de alto costo a 600 \$/MWh actúa frecuentemente como recurso marginal de todo el sistema en condiciones de escasez.

La Figura 3 muestra el esquema del caso de prueba, ilustrando los cuatro subsistemas y los enlaces de transmisión direccionales que los conectan.

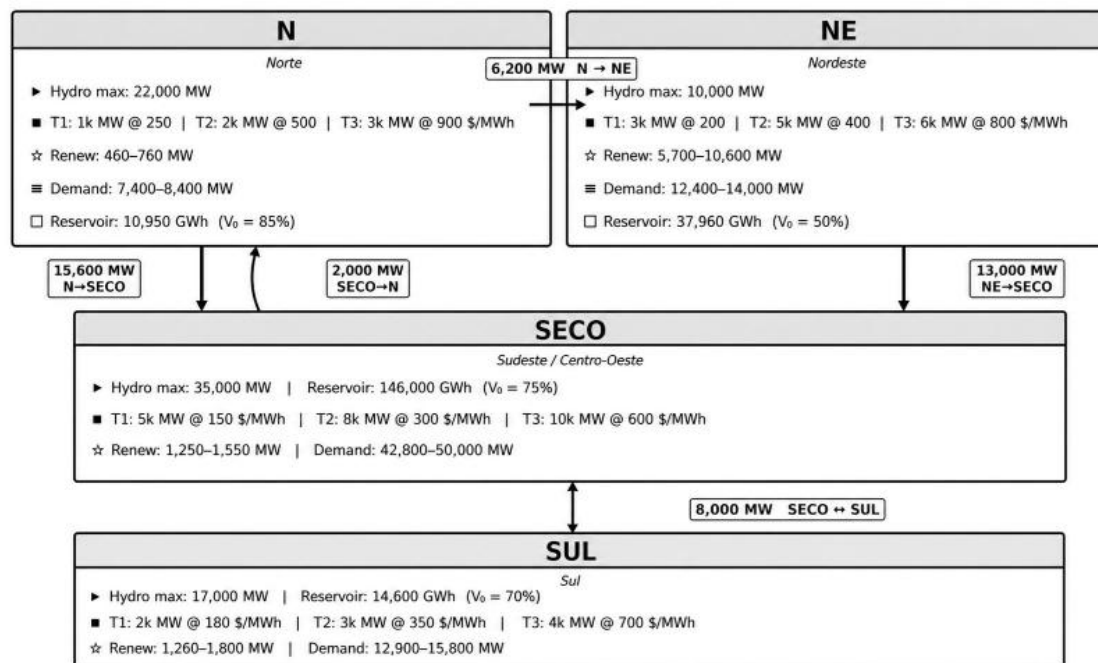


Figura 3: Esquema del caso de prueba hidrotérmico basado en el Sistema Interconectado Nacional brasileño (SIN).

Conjunto de escenarios hidrológicos

Los diez escenarios de afluencias se construyeron para representar regímenes hidrológicos estructuralmente distintos, en lugar de perturbaciones aleatorias en torno a una estimación central. La Tabla 4 reporta el tipo de cada escenario y la desviación anual de afluencias respecto a la media del conjunto, por subsistema y para el sistema en su conjunto.

Escenario	Tipo	SECO	SUL	NE	N	Sistema
1	Normal	+10	−4	+4	0	+3
2	Normal	0	+11	+4	+2	+3
3	Normal	+14	+2	+10	+9	+9
4	Húmedo	+31	+34	+30	+31	+31
5	Húmedo (La Niña)	+45	0	+36	+39	+35
6	Húmedo	+25	+33	+21	+29	+28
7	Seco	−25	−20	−19	−25	−24
8	Seco (El Niño)	−38	+15	−28	−23	−23
9	Seco	−14	−28	−12	−15	−16
10	Crítico	−47	−44	−45	−47	−46

Tabla 4: Conjunto de escenarios hidrológicos: tipo y desviación anual de afluencias respecto a la media del conjunto (%).

El conjunto abarca tres realizaciones cercanas a la media (escenarios 1–3), tres años húmedos (4–6, de +28% a +35% por encima de la media), tres años secos (7–9, de −16% a −24% por debajo de la media) y un escenario crítico (10) modelado a partir de la crisis energética brasileña de 2001, con todos los

subsistemas aproximadamente un 47% por debajo de sus afluencias medias. Las afluencias anuales totales del sistema varían de 233.300 GWh a 586.900 GWh, un factor de 2,5×, con razones individuales de subsistema que alcanzan 2,7× para el SECO.

Más allá de la severidad agregada, varios escenarios presentan una pronunciada heterogeneidad entre subsistemas. El escenario 5 reproduce un patrón del tipo La Niña, en el que SECO, NE y N están muy por encima de la media (+45%, +36%, +39%), mientras que el SUL recibe afluencias cercanas a la media. El escenario 8 invierte esto en una configuración El Niño: SECO, NE y N severamente por debajo de la media (−38%, −28%, −23%), mientras que el SUL está por encima de la media (+15%). En conjunto, los diez escenarios cubren un amplio espectro de severidad agregada y de estructura espacial de las afluencias, lo que proporciona una prueba exigente de robustez de la política a pesar del tamaño modesto del conjunto.

Estimación del valor marginal del agua

Conforme se estableció en el diseño experimental, cada una de las diez ejecuciones siguió una trayectoria de razonamiento independiente. A pesar de ello, la inspección de los registros de interacción reveló un patrón consistente entre las ejecuciones: el agente convergió recurrentemente hacia un enfoque de diferencias finitas para estimar el valor marginal del agua almacenada, independientemente de la secuencia específica de invocaciones de herramientas que la precedió. Esta convergencia sugiere que el enfoque es una consecuencia natural de las capacidades disponibles y de la estructura del problema, y no un artefacto de una sesión específica.

Para cada subsistema i , comparó el costo operativo simulado en un vector de almacenamiento de referencia con el costo obtenido tras perturbar el almacenamiento de ese subsistema en un pequeño incremento ΔV :

$$\hat{\lambda}_i = - \frac{C(V + \Delta V e_i) - C(V)}{\Delta V}. \quad (13)$$

Construcción regresiva de los cortes de costo futuro

Un segundo patrón que se repitió de forma consistente entre las ejecuciones fue la estrategia para la construcción de los cortes de costo futuro. Tras estimar los valores del agua por subsistema mediante diferencias finitas, el agente procedió sistemáticamente a codificarlos como aproximaciones afines de costo futuro sobre la función de costo de recurso (cost-to-go) por inducción regresiva (backward induction) —un procedimiento que emergió de forma independiente en cada sesión sin haber sido prescrito en la especificación de la tarea—.

Usando los coeficientes del valor del agua por subsistema $\hat{\lambda}_i$ estimados en el paso anterior, el agente construyó una secuencia de cortes de costo futuro por inducción regresiva. Comenzando con un valor terminal nulo, para cada etapa $t = T, \dots, 1$, simuló la operación a partir de un estado inicial y calculó:

$$\beta_t = c_t - \sum_i \hat{\lambda}_i V_{t,i}^{\text{fin}} + \beta_{t+1}, \quad (14)$$

donde c_t es el costo operativo inmediato devuelto por el simulador y $V_{t,i}^{\text{fin}}$ es el almacenamiento de fin de etapa resultante en el subsistema i .

La política resultante se representa mediante una aproximación afín de costo futuro sobre el costo futuro:

$$Q_t(V) \geq \beta_t - \sum_i \hat{\lambda}_i V_i. \quad (15)$$

El Algoritmo 1 resume el procedimiento reconstruido a partir del registro de interacción.

Algoritmo 1 Construcción regresiva, derivada por el agente, de cortes afines de costo futuro

Requiere: Coeficientes del valor del agua por subsistema $\hat{\lambda}_i$, intercepto terminal $\beta_{T+1} = 0$

1. **para** $t = T, T - 1, \dots, 1$ **hacer**
2. Simular la etapa t a partir de almacenamiento inicial nulo
3. Observar el costo inmediato c_t y el vector de almacenamiento final V_t^{fin}
4. Calcular $\beta_t = c_t - \sum_i \hat{\lambda}_i V_{t,i}^{\text{fin}} + \beta_{t+1}$
5. Definir el corte $Q_t(V) \geq \beta_t - \sum_i \hat{\lambda}_i V_i$
6. **fin para**
7. Registrar todos los cortes no terminales en el entorno de optimización

Este procedimiento se asemeja a un único paso regresivo simplificado del SDDP, pero no se le proporcionó al agente como una plantilla algorítmica. El perfil de interceptos decrece a medida que el horizonte se aproxima a su etapa terminal, reflejando el valor decreciente de los períodos operativos futuros.

Análisis de reproducibilidad entre ejecuciones

La Tabla 5 reporta el costo operativo medio obtenido en cada una de las diez ejecuciones independientes, listadas en el orden de las sesiones.

Ejecución	Costo medio (M\$)
R1	139,24
R2	148,33
R3	145,77
R4	153,26
R5	138,12
R6	153,94
R7	139,24
R8	139,97
R9	139,21
R10	137,72
Media general	143,48
Desviación estándar	6,02
Mín.	137,72
Máx.	153,94

Tabla 5: Costo operativo total medio para cada ejecución independiente (M\$).

Seis ejecuciones (R1, R5, R7–R10) produjeron políticas con costos medios en el rango de 137,7–140,0 M\$, una dispersión de apenas 2,3 M\$ entre seis sesiones independientes. Las cuatro ejecuciones restantes

(R2–R4, R6) arrojaron costos medios más altos, de 145,8–153,9 M\$. La media general entre las diez ejecuciones es de 143,5 M\$, con una desviación estándar de 6,0 M\$. Una observación notable es que las ejecuciones R1 y R7 produjeron costos por escenario idénticos hasta la precisión numérica, lo que indica que el agente llegó exactamente a la misma política mediante dos trayectorias de interacción independientes —una forma de convergencia que sugiere que el camino analítico hacia la estimación dominante del valor del agua está suficientemente restringido por la estructura de capacidades como para ser reproducible—.

Trayectoria de razonamiento de la ejecución de mejor desempeño

Entre las diez ejecuciones independientes, la sesión que alcanzó el menor costo operativo medio (137,7 M\$) dejó un registro de interacción suficientemente detallado para reconstruir su trayectoria analítica por completo. El agente comenzó con la inspección del sistema (Fase 1), cargando los datos del modelo, la estructura de costos y las estadísticas de los escenarios. En la Fase 2, simuló los diez escenarios sin ningún corte registrado, estableciendo una línea base en la que los embalses se vaciaban entre las etapas 6 y 8 en todos los escenarios y el sistema operaba enteramente con despacho térmico a partir de la etapa 8 —el clásico colapso miope—.

La Fase 3 consistió en experimentos dirigidos de perturbación de etapa única. Al perturbar cada embalse de forma aislada en una etapa y un escenario estresados, el agente estimó el valor marginal del agua almacenada: aproximadamente 600 \$/MWh para SECO, SUL y N, y 567 \$/MWh para NE. Una inferencia fundamental fue que los cuatro subsistemas estaban dominados por el mismo recurso marginal de todo el sistema —el tercer bloque térmico en SECO a 600 \$/MWh—, reflejando el alto grado de interconexión del sistema.

En la Fase 4, el agente registró un primer conjunto de cortes afines con coeficientes establecidos como ligeramente más negativos que el umbral estimado ($\alpha = [-700, -700, -600, -700]$ \$/MWh). La evaluación de esta política reveló una inconsistencia: el almacenamiento de SECO se estaba acumulando mientras el bloque térmico costoso se despachaba simultáneamente, lo que indicaba que los coeficientes estaban penalizando el uso del agua de forma demasiado agresiva. El agente identificó esta preservación excesiva y revisó los coeficientes a la baja en la Fase 5 ($\alpha_{\text{SECO}} = -590$, $\alpha_{\text{NE}} = -560$), alineándolos con los valores marginales estimados en la Fase 3. La política refinada produjo un costo operativo medio de aproximadamente 138 M\$ y se aceptó como salida final.

Esta trayectoria es notable por su estructura autocorrectiva. Sin orientación externa, el agente detectó una inconsistencia de la política mediante la retroalimentación de la simulación y revisó su representación del valor del agua en consecuencia.

Comparación con la operación miope

La Tabla 6 compara la política derivada por el agente con la línea base miope.

Indicador	Miope	Agente
Costo medio (M\$)	275,7	137,7
Emergencia T3 (etapas)	10	0
Eventos de déficit (etapas)	4	0
Ahorro medio (M\$)	–	138,0

Tabla 6: Comparación entre la operación miope y la política derivada por el agente.

La política miope agota rápidamente los embalses durante la primera estación húmeda, dejando un margen (buffer) hidráulico insuficiente para los períodos secos posteriores. La política derivada por el

agente, en cambio, preserva el almacenamiento, aceptando costos térmicos más altos a corto plazo para reducir el riesgo futuro de escasez. Este comportamiento es consistente con la interpretación económica de los valores del agua en la operación hidrotérmica estocástica: el agua debe utilizarse cuando su beneficio inmediato excede su costo de oportunidad futuro esperado.

La Figura 4 presenta la evolución estocástica del almacenamiento de los embalses a lo largo de los diez escenarios de afluencias bajo la política derivada por el agente y la línea base miope. Cada panel corresponde a un subsistema. Las líneas en negrita muestran la media de los escenarios; las bandas sombreadas reportan los intervalos P25–P75 y P10–P90 entre escenarios. Los niveles de almacenamiento se expresan como porcentaje de la capacidad máxima de cada subsistema.

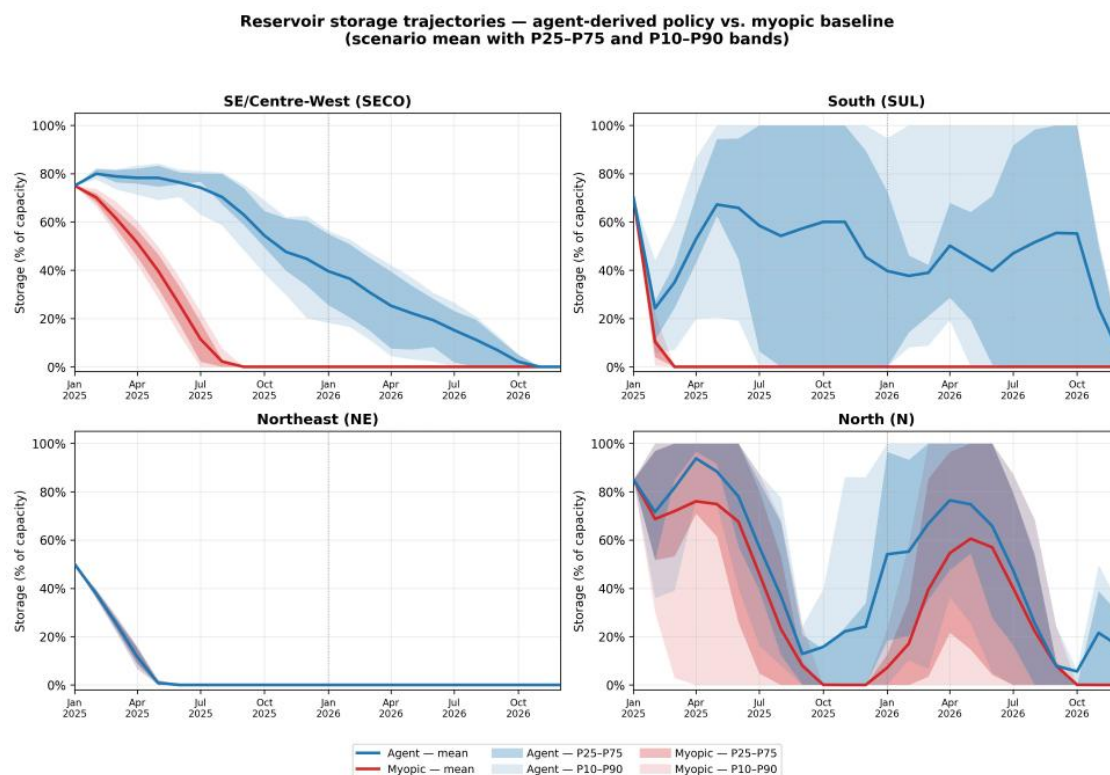


Figura 4: Trayectorias de almacenamiento de los embalses bajo la política derivada por el agente y la línea base miope. Las curvas rojas denotan la línea base miope.

El contraste entre las dos políticas es pronunciado. Bajo la línea base miope, todos los subsistemas se agotan a niveles cercanos a cero en las etapas 7–8 (julio–agosto de 2025) en la mayoría de los escenarios, y el sistema opera enteramente con despacho térmico durante el resto de ambas estaciones secas. Bajo la política derivada por el agente, el embalse de SECO, que concentra aproximadamente el 70% de la capacidad total de almacenamiento del sistema, se mantiene muy por encima de cero a lo largo de toda la primera estación seca y entra en 2026 con una reserva hidráulica suficiente para amortiguar el segundo período seco. Los subsistemas NE y N, que tienen capacidades absolutas menores, exhiben diferencias cualitativas similares, mientras que el SUL se recupera parcialmente por su propio régimen de afluencias.

Comparación con el SDDP

La Tabla 7 compara la ejecución del agente de mejor desempeño (R10, costo medio de 137,7 M\$) con el valor de referencia SDDP, usando realizaciones de afluencias coincidentes. Esta ejecución se selecciona aquí porque produjo el menor costo medio entre las diez sesiones independientes y, por lo tanto, representa el resultado más favorable del enfoque orientado a capacidades. Los resultados de las demás

ejecuciones se reportan en la Sección 5.5; entre las diez ejecuciones, la brecha media de costo respecto al SDDP es de aproximadamente el 6,0%, reflejando la variabilidad introducida por la naturaleza probabilística del agente de razonamiento. La referencia SDDP se obtuvo usando el solucionador comercial de PSR [18], aplicado a los mismos diez escenarios de afluencias con cinco escenarios regresivos por iteración.

Escenario	Agente (M\$)	SDDP (M\$)	Dif. (M\$)	Brecha (%)
1	113,5	114,5	-1,0	-0,9
2	111,6	107,2	4,4	4,1
3	90,0	81,1	8,9	11,0
4	70,0	64,0	6,0	9,4
5	73,7	66,3	7,4	11,2
6	128,8	126,3	2,5	2,0
7	193,4	191,8	1,6	0,8
8	178,7	178,5	0,2	0,1
9	231,9	237,9	-6,0	-2,5
10	185,7	186,6	-0,9	-0,5
Media	137,7	135,4	2,3	1,7
Desv. estándar	53,2	58,5	-	-

Tabla 7: Comparación a nivel de escenario entre la ejecución del agente de mejor desempeño (R10) y el valor de referencia SDDP.

Para la mejor ejecución, la brecha media de costo respecto al SDDP es del 1,7%. Este resultado debe interpretarse con cautela. La política SDDP se obtiene usando un algoritmo especializado de optimización estocástica con repetidas iteraciones hacia adelante y hacia atrás (forward-backward), mientras que la política derivada por el agente es una política afín simplificada, construida mediante interacción estructurada. La observación relevante, por lo tanto, no es que el agente compita con el SDDP, sino que el entorno orientado a capacidades permitió al agente construir una política con una lógica intertemporal coherente del valor del agua.

En tres escenarios, la política derivada por el agente logró un costo ex post menor que el valor de referencia SDDP. Esto no implica dominancia sobre el SDDP. En cambio, refleja la variación a nivel de escenario y la naturaleza conservadora de la política derivada por el agente, que tiende a preservar el almacenamiento de forma más agresiva. En promedio, la política SDDP sigue siendo el mejor valor de referencia de optimización estocástica.

Discusión

Orquestación estructurada de capacidades

El experimento muestra que un agente de razonamiento puede usar capacidades estructuradas para llevar a cabo operaciones analíticas relevantes para el dominio en un entorno estocástico de sistema de energía. Su trayectoria de interacción se asemejó al flujo de trabajo de un analista humano trabajando con un modelo de simulación: el agente no recibió una política óptima directamente, sino que desarrolló progresivamente una lógica operativa aproximada que guió las decisiones posteriores.



El punto importante no es que el agente descubriera un nuevo método de optimización. En cambio, la estructura de capacidades restringió la interacción de manera productiva. Al exponer capacidades en niveles apropiados de abstracción, el entorno permitió que el agente formulara y probara hipótesis operativas, infiriera la lógica intertemporal del valor del agua y construyera una representación aproximada de la estructura de costo de recurso (cost-to-go) subyacente al problema de programación dinámica estocástica.

Implicaciones para la analítica energética asistida por IA

La arquitectura propuesta sugiere un papel para los agentes de razonamiento como intermediarios analíticos integrados en entornos de modelado de sistemas de energía. En lugar de reemplazar simuladores, solucionadores de optimización o expertos humanos, los agentes pueden apoyar la capa exploratoria en torno a los modelos formales, seleccionando cálculos, interpretando salidas, probando hipótesis y coordinando análisis iterativos.

Este papel es especialmente relevante para problemas energéticos cuyo contexto de decisión es más amplio que una única formulación de optimización bien especificada. Los estudios de planificación y operación a menudo combinan supuestos inciertos, datos heterogéneos, múltiples modelos, restricciones regulatorias, consideraciones de diseño de mercado y compromisos difíciles de codificar completamente en forma analítica cerrada. En tales casos, un sistema agéntico puede ser útil no porque calcule directamente un óptimo, sino porque ayuda a estructurar la exploración de alternativas, sensibilidades, escenarios e insights relevantes para la política.

Las aplicaciones potenciales incluyen:

- Exploración de grandes espacios de escenarios estocásticos o de prueba de estrés;
- Análisis comparativo entre supuestos y configuraciones de modelo alternativos;
- Identificación de sensibilidades, cuellos de botella y factores operativos determinantes;
- Refinamiento iterativo de políticas candidatas o alternativas de planificación;
- Orquestación de múltiples herramientas de simulación, previsión y optimización;
- Explicación e interrogación de resultados de modelos para el apoyo a la decisión.

Desde esta perspectiva, el diseño de las capacidades se vuelve central. El agente solo puede razonar de forma eficaz sobre un modelo de dominio si las operaciones analíticas relevantes se exponen de manera controlada, semánticamente significativa y auditable. La programación hidrotérmica proporciona un escenario de validación con valor de referencia para esta arquitectura, porque se dispone de referencias de optimización confiables. Los trabajos futuros deberían evaluar si marcos agénticos similares pueden apoyar la exploración de soluciones en problemas en los que las formulaciones analíticas totalmente especificadas son incompletas, difíciles de construir o insuficientes para representar el contexto de decisión relevante.

Limitaciones

Deben reconocerse varias limitaciones. Algunas son específicas del estudio de caso de programación hidrotérmica utilizado para la validación, mientras que otras conciernen al uso más amplio de agentes de razonamiento en entornos analíticos estructurados de sistemas de energía.

En cuanto al estudio de caso específico de programación hidrotérmica, el experimento se realizó sobre una representación simplificada del SIN brasileño. Aunque estructuralmente realista, no captura toda la complejidad de la operación en el mundo real, incluido el compromiso de unidades (unit commitment) detallado, las restricciones de seguridad, las restricciones individuales de las centrales, las reglas de



mercado, las restricciones regulatorias y los requisitos de reserva operativa. El caso debe, por lo tanto, interpretarse como un valor de referencia controlado para la evaluación de la arquitectura, y no como una representación completa de la operación hidrotérmica de nivel de producción.

En segundo lugar, el procedimiento inducido por el agente no posee garantías de optimalidad. La calidad del comportamiento operativo resultante depende de las hipótesis formuladas por el agente, de la secuencia de invocaciones de capacidades y de las capacidades puestas a disposición por el entorno. Los métodos clásicos de optimización estocástica exploran sistemáticamente la formulación matemática y proporcionan garantías de convergencia bajo supuestos apropiados; el procedimiento derivado por el agente no lo hace.

Por último, la escalabilidad de la orquestación secuencial de capacidades sigue siendo una cuestión abierta. Los sistemas más grandes pueden requerir capacidades de simulación por lotes, evaluación paralela de escenarios, plantillas de política más robustas, mecanismos de gestión de contexto e integración más estrecha con solucionadores de nivel de producción.

Conclusiones

Este artículo evaluó si un agente de razonamiento puede operar como una capa analítica dentro de un entorno estructurado de modelado de sistemas de energía. Para ello, propusimos un marco orientado a capacidades en el que el agente interactúa con el entorno analítico únicamente a través de capacidades estructuradas y tipadas. Este diseño separa el razonamiento y la orquestación del cálculo de dominio validado, permitiendo que el agente inspeccione el modelo, ejecute simulaciones, pruebe hipótesis y registre artefactos relevantes para la política, sin ejecución irrestricta de código ni acceso a los aspectos internos del solucionador.

El marco se evaluó en un caso de prueba de programación hidrotérmica estocástica basado en una representación de cuatro subsistemas del SIN brasileño. Sin acceso a un método de programación dinámica estocástica preprogramado, variables duales, valores del agua precomputados o información interna del solucionador, el agente infirió la lógica del valor del agua a partir de la retroalimentación de la simulación y construyó aproximaciones afines de costo futuro. Cuando se incorporaron al modelo de despacho de la etapa, estas aproximaciones indujeron políticas operativas de preservación del almacenamiento que mejoraron sustancialmente la operación miope. A lo largo de diez sesiones independientes, la mejor política alcanzó una brecha media de costo del 1,7% respecto a un valor de referencia SDDP calculado de forma independiente, mientras que la brecha media entre las sesiones fue de aproximadamente el 6,0%.

Estos resultados no sugieren que los agentes de razonamiento deban reemplazar los métodos de optimización estocástica. En cambio, muestran que, cuando está restringida por una interfaz de capacidades apropiada, la IA agéntica puede apoyar la capa exploratoria en torno a los modelos de sistemas de energía: formulando hipótesis, seleccionando cálculos, interpretando salidas y construyendo artefactos relevantes para la política, mientras las herramientas de simulación y optimización siguen siendo el fundamento computacional.

Los trabajos futuros deberían evaluar marcos orientados a capacidades similares en problemas de sistemas de energía en los que el proceso de decisión relevante va más allá de una única formulación de optimización totalmente especificada, incluidos los escenarios que requieren exploración de escenarios, coordinación de múltiples modelos, análisis de sensibilidad y evaluación iterativa de políticas.

Referencias

[1] S. Yao, J. Zhao, D. Yu, et al., ReAct: Synergizing reasoning and acting in language models, in: International Conference on Learning Representations, 2023.

- 
- [2] L. Wang, C. Ma, X. Feng, et al., A survey on large language model based autonomous agents, *Frontiers of Computer Science* 18 (2024) 186345.
- [3] T. Masterman, S. Besen, M. Sawtell, A. Chandra, The landscape of emerging AI agent architectures for reasoning, planning, and tool calling: a survey, *arXiv preprint arXiv:2404.11584* (2024).
- [4] T.R. Sumers, S. Yao, K. Narasimhan, T.L. Griffiths, Cognitive architectures for language agents, *Transactions on Machine Learning Research* (2024).
- [5] M.V.F. Pereira, L.M.V.G. Pinto, Multi-stage stochastic optimization applied to energy planning, *Mathematical Programming* 52 (1991) 359–375.
- [6] A. Shapiro, W. Tekaya, J.P. da Costa, M.P. Soares, Risk neutral and risk averse stochastic dual dynamic programming method, *European Journal of Operational Research* 224 (2013) 375–391.
- [7] T. Schick, J. Dwivedi-Yu, R. Dessì, et al., Toolformer: Language models can teach themselves to use tools, *Advances in Neural Information Processing Systems* 36 (2023).
- [8] Y. Qin, S. Liang, Y. Ye, et al., ToolLLM: Facilitating large language models to master real-world APIs, in: *International Conference on Learning Representations*, 2024.
- [9] NVIDIA, NVIDIA AI Blueprints: reference workflows for building generative AI and agentic applications, Technical Documentation, 2024. URL: <https://www.nvidia.com/en-us/ai/blueprints/>.
- [10] NVIDIA, NeMo Guardrails: programmable guardrails for conversational AI applications, Technical Documentation, 2024. URL: <https://docs.nvidia.com/nemo/guardrails/>.
- [11] Anthropic, Model Context Protocol specification, Technical Documentation, 2024. URL: <https://modelcontextprotocol.io>.
- [12] S. Haben, S. Arora, G. Giasemidis, M. Voss, D. Vukadinović Greetham, Review of low voltage load forecasting: methods, applications, and recommendations, *Applied Energy* 304 (2021) 117798.
- [13] A. Venzke, G. Qu, S. Low, S. Chatzivasileiadis, Learning optimal power flow: worst-case guarantees for neural networks, in: *Proc. IEEE International Conference on Communications, Control, and Computing Technologies for Smart Grids*, IEEE, 2020.
- [14] A. Castelletti, S. Galelli, M. Restelli, R. Soncini-Sessa, Tree-based reinforcement learning for optimal water reservoir operation, *Water Resources Research* 46 (2010) W09507.
- [15] T. Yang, L. Zhao, W. Li, A.Y. Zomaya, Reinforcement learning in sustainable energy and electric systems: a survey, *Annual Reviews in Control* 49 (2020) 145–163.
- [16] A.B. Philpott, Z. Guan, On the convergence of stochastic dual dynamic programming and related methods, *Operations Research Letters* 36 (2008) 450–455.
- [17] P. Girardeau, V. Leclère, A.B. Philpott, On the convergence of decomposition methods for multistage stochastic convex programs, *Mathematics of Operations Research* 40 (2015) 130–145.
- [18] PSR Energy Consulting and Analytics, SDDP – Stochastic Dual Dynamic Programming software, Software Documentation, 2024. URL: <https://www.psr-inc.com/pt-br/software/sddp/>.