



AVALIAÇÃO DA CAPACIDADE DA IA AGÊNTICA EM UM ESTUDO DE CASO DE PLANEJAMENTO HIDROTÉRMICO ESTOCÁSTICO

Julio Alberto Silva Dias e Mario Veiga Ferraz Pereira

Resumo

A inteligência artificial agêntica está emergindo como um paradigma no qual modelos com capacidade de raciocínio interagem com ferramentas externas, simulações e fontes de dados para executar tarefas analíticas de múltiplas etapas. Para aplicações em sistemas de energia, isso levanta a questão de se os agentes de IA podem operar de forma eficaz em ambientes analíticos estruturados, utilizando capacidades específicas do domínio para explorar problemas de decisão complexos, em vez de atuarem meramente como invólucros (wrappers) em torno de solvers existentes.

Este artigo investiga essa questão por meio de um estudo de caso de programação hidrotérmica estocástica, que combina incerteza, acoplamento intertemporal, restrições físicas e trade-offs operacionais economicamente relevantes, ao mesmo tempo em que oferece referências confiáveis de otimização, como a programação dinâmica dual estocástica (SDDP). Propomos uma arquitetura agêntica orientada a capacidades, na qual um agente de raciocínio interage com o modelo hidrotérmico somente por meio de capacidades de domínio controladas e tipadas, sem acesso a métodos de programação dinâmica pré-programados, valores da água pré-computados, variáveis duais ou informações internas do solver.

A arquitetura é avaliada em uma representação simplificada, mas estruturalmente realista, do Sistema Interligado Nacional brasileiro. Ao longo de dez sessões independentes, o agente inferiu a lógica do valor da água a partir do retorno das simulações, construiu representações aproximadas de custo futuro e implementou políticas de preservação de armazenamento que melhoraram substancialmente o comportamento míope. Em relação a um benchmark SDDP calculado de forma independente, a melhor sessão atingiu uma diferença média de custo de 1,7%, enquanto a diferença média das dez sessões foi de aproximadamente 6,0%.

Esses resultados indicam que a orquestração estruturada de capacidades pode permitir que agentes de raciocínio funcionem como camadas analíticas integradas a ambientes de modelagem de sistemas de energia, sendo mais valiosos para problemas em que as formulações analíticas são incompletas, difíceis de especificar ou insuficientes para capturar todo o processo de exploração da solução.

Palavras-chave: IA agêntica, programação hidrotérmica, otimização estocástica, valor da água, SDDP

Introdução

A inteligência artificial agêntica está emergindo como um paradigma no qual modelos de linguagem com capacidade de raciocínio interagem com ferramentas externas, simulações, fontes de dados e ambientes computacionais para executar tarefas analíticas de múltiplas etapas. Em vez de produzir apenas saídas

de texto estático, esses sistemas podem selecionar ações estruturadas, invocar ferramentas, interpretar observações retornadas, revisar hipóteses e coordenar fluxos de trabalho analíticos de múltiplas etapas rumo a um objetivo definido pelo usuário [1]. Frameworks agênticos recentes costumam separar um núcleo de inferência baseado em modelo de linguagem — responsável pelo raciocínio, planejamento e seleção de ações — de uma camada de orquestração, ou Harness, que gerencia chamadas de ferramentas, observações, memória e estado de execução ao longo das etapas [4, 2, 3]. Essa estrutura geral é ilustrada na Fig. 1, que mostra o núcleo de inferência do LLM e o Harness que o circunda.

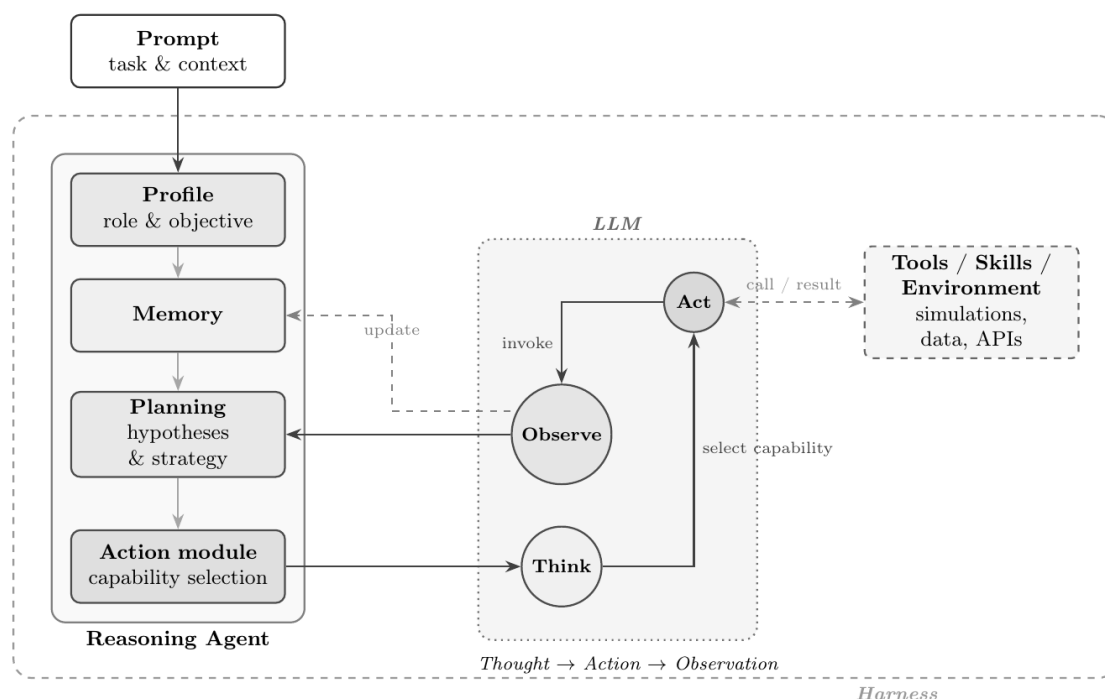


Figura 1: Estrutura geral de um laço de raciocínio agêntico baseado em LLM.

Embora esses avanços tenham sido demonstrados principalmente em aplicações genéricas de software corporativo, programação e recuperação de informações, eles levantam uma questão metodológica relevante para a análise de sistemas de energia: podem os agentes de raciocínio operar de forma eficaz dentro de ambientes analíticos estruturados em problemas de decisão de engenharia complexos, ou eles apenas fornecem interfaces em linguagem natural para ferramentas computacionais existentes? Essa questão é particularmente importante no planejamento e na operação de sistemas de energia, em que os processos de apoio à decisão dependem de modelos especializados que representam incerteza, restrições físicas, acoplamentos intertemporais e trade-offs economicamente relevantes.

A programação hidrotérmica fornece um cenário adequado para investigar essa questão. Em sistemas hidrotérmicos, as decisões atuais de deplecionamento dos reservatórios reduzem os custos presentes de geração térmica, mas também afetam a exposição futura à escassez, ao despacho térmico caro e aos déficits de energia. A água armazenada tem, portanto, um custo de oportunidade, comumente representado por meio de valores da água ou funções de custo futuro. O problema combina afluências estocásticas, dinâmica de armazenamento, restrições de transmissão, substituição térmica e trade-offs operacionais de longo prazo. Ao mesmo tempo, possui referências bem estabelecidas de otimização estocástica, em particular a Programação Dinâmica Dual Estocástica (SDDP), que fornece um benchmark rigoroso para avaliar a qualidade das políticas operativas [5, 6].

A programação dinâmica estocástica clássica fornece a base conceitual para esse problema de decisão sequencial, mas sua aplicação direta é limitada pela maldição da dimensionalidade. Em sistemas hidrotérmicos realistas, o vetor de estado inclui os níveis de armazenamento dos reservatórios, as

condições hidrológicas e outras variáveis intertemporais, fazendo com que o número de estados possíveis cresça exponencialmente com o tamanho do sistema e com a representação da incerteza. Métodos de decomposição como o SDDP evitam a enumeração completa de estados ao aproximar as funções de custo futuro e permanecem como a referência padrão para o planejamento da operação hidrotérmica de grande porte. Métodos de aprendizado de máquina, incluindo aprendizado por reforço, também têm sido estudados para a tomada de decisão sequencial sob incerteza. Neste artigo, no entanto, esses métodos servem principalmente como referências conceituais e benchmarks: o objetivo não é propor um novo solver de otimização hidrotérmica.

A questão central abordada neste artigo é, em vez disso, arquitetural e metodológica: pode um agente de raciocínio, restrito a capacidades de domínio controladas, explorar um ambiente analítico de sistema de energia e construir uma política operacionalmente significativa sem acesso a um método de programação dinâmica estocástica pré-programado, variáveis duais, valores da água pré-computados ou aspectos internos do solver? Essa questão diz respeito a como a funcionalidade analítica deve ser exposta a um agente de IA de modo que raciocínio, computação, controle e auditabilidade permaneçam separados. Ela está alinhada com a noção emergente de habilidades ou capacidades específicas de domínio: operações modulares que expõem funcionalidades selecionadas de um ambiente a um agente por meio de interfaces estruturadas.

Para investigar essa questão, propomos uma arquitetura agêntica orientada a capacidades para a interação assistida por IA com ambientes analíticos de sistemas de energia. Nessa arquitetura, o conhecimento de domínio e as operações do modelo são expostos ao agente por meio de capacidades estruturadas e tipadas, e não por meio de execução irrestrita de código ou acesso direto aos aspectos internos do solver. O agente pode inspecionar o sistema, consultar estatísticas de cenários, executar simulações, avaliar políticas candidatas e registrar artefatos relevantes para a política, mas todas as computações são executadas pelo ambiente analítico controlado. O agente, portanto, não substitui o modelo de domínio nem o solver de otimização; ele atua como uma camada de orquestração analítica que formula hipóteses, invoca capacidades, interpreta saídas e constrói progressivamente uma lógica operativa aproximada.

A arquitetura proposta é instanciada em um ambiente de programação hidrotérmica estocástica baseado em uma representação simplificada, mas estruturalmente realista, do Sistema Interligado Nacional brasileiro. O agente interage com o modelo apenas por meio de capacidades controladas e não recebe acesso a uma implementação do SDDP, rotinas de programação dinâmica, informações duais, valores da água pré-computados ou estado interno do solver. A política resultante é então avaliada por meio de simulação estocástica e comparada com duas referências: uma política míope que ignora o valor futuro da água e um benchmark baseado em SDDP calculado de forma independente. Esse cenário permite que o comportamento do agente seja avaliado em um problema cuja estrutura de otimalidade é bem compreendida e para o qual há benchmarks confiáveis disponíveis.

A motivação mais ampla é avaliar se os agentes de raciocínio podem funcionar como camadas analíticas controladas em torno de modelos de sistemas de energia. Não se espera que tais agentes substituam os métodos formais de otimização. Em vez disso, eles podem apoiar a exploração, a interpretação, a análise de sensibilidade, a interrogação de modelos e a construção de políticas em cenários em que o processo de decisão relevante vai além de uma única formulação de otimização totalmente especificada. A programação hidrotérmica é utilizada aqui como um caso de validação com benchmark, antes de se considerarem problemas de sistemas de energia mais abertos, nos quais as formulações analíticas podem ser incompletas, difíceis de construir ou insuficientes para capturar todo o processo de exploração da solução.

O artigo aborda as seguintes questões de pesquisa:

RQ1. Pode um agente de raciocínio usar capacidades de domínio estruturadas para construir uma política operacionalmente significativa em um problema estocástico de programação de sistemas de energia?

RQ2. No caso da programação hidrotérmica, pode o agente inferir uma lógica de valor da água economicamente significativa a partir do retorno das simulações, em vez de informações duais explícitas ou saídas de otimização pré-computadas?

RQ3. Quão reprodutível é o processo de construção de políticas em sessões independentes de agentes de raciocínio sob a mesma especificação de tarefa e a mesma interface de capacidades?

RQ4. Quais propriedades arquiteturais são necessárias para apoiar a avaliação controlada, auditável e modular de agentes de raciocínio em ambientes analíticos de sistemas de energia?

As principais contribuições do artigo são:

1. Uma arquitetura agêntica orientada a capacidades para a interação assistida por IA com ambientes analíticos de sistemas de energia.
2. Uma instanciação dessa arquitetura em um ambiente de programação hidrotérmica estocástica, conectando o fluxo de trabalho agêntico à formulação clássica de otimização estocástica multiestágio.
3. Uma interface de capacidades estruturada que expõe operações de domínio controladas para inspeção do modelo, exploração baseada em simulação, registro de políticas e avaliação estocástica, ao mesmo tempo em que oculta os aspectos internos do solver, as variáveis duais e os valores da água pré-computados.
4. Uma avaliação experimental com benchmark mostrando que um agente de raciocínio pode construir políticas operativas hidrotérmicas de preservação de armazenamento por meio de orquestração estruturada de capacidades, com avaliação comparativa em relação à operação míope e a referências baseadas em SDDP calculadas de forma independente.

O restante do artigo está organizado da seguinte forma. A Seção 2 revisa trabalhos relacionados. A Seção 3 define o problema de operação hidrotérmica estocástica multiestágio. A Seção 4 apresenta a arquitetura orientada a capacidades, o ambiente analítico e o protocolo experimental. A Seção 5 apresenta o estudo de caso e os resultados. A Seção 6 discute implicações e limitações. A Seção 7 conclui.

Trabalhos Relacionados

Trabalhos recentes sobre modelos de linguagem com uso de ferramentas mostraram que grandes modelos de linguagem podem combinar raciocínio mediado por linguagem com ações computacionais externas, incluindo chamadas de ferramentas, acesso a APIs e interação estruturada com ambientes computacionais [1, 7, 8]. Arquiteturas agênticas relacionadas organizam esse processo por meio de módulos para especificação de papéis, memória, planejamento e seleção de ações, e enfatizam cada vez mais habilidades ou capacidades reutilizáveis como a interface entre o modelo de raciocínio e seu ambiente [2, 3]. Frameworks industriais seguiram uma direção semelhante, fornecendo padrões de referência para combinar modelos de fundação, acesso a ferramentas, orquestração de fluxos de trabalho e mecanismos de salvaguarda (guardrails) [9, 10].

Embora muitas demonstrações de agentes que usam ferramentas se concentrem em busca na web, programação, resposta a perguntas ou recuperação de informações, aplicações de engenharia de alta

criticidade exigem garantias mais robustas de controle, auditabilidade e reprodutibilidade. Nesses contextos, o uso irrestrito de ferramentas é frequentemente inadequado: o agente deve interagir com ambientes computacionais validados por meio de interfaces bem definidas. Protocolos de interoperabilidade como o Model Context Protocol (MCP) ilustram essa direção ao expor ferramentas externas por meio de interfaces tipadas [11]. Isso motiva o uso de interfaces de capacidades controladas para a interação agêntica com modelos de sistemas de energia.

O aprendizado de máquina também tem sido amplamente aplicado em sistemas de energia, incluindo previsão, aproximação de fluxo de potência ótimo e controle [12, 13]. O aprendizado por reforço tem sido estudado para a operação de reservatórios e o controle de sistemas de potência [14, 15]. Essas abordagens geralmente aprendem políticas ou aproximações de funções por meio de dados, episódios de simulação, atualizações de gradiente ou procedimentos explícitos de treinamento. A abordagem investigada aqui é diferente: o agente não é treinado como um controlador hidrotérmico e não aprende uma política por meio de episódios repetidos. Em vez disso, ele utiliza raciocínio e orquestração de capacidades para inspecionar um ambiente estruturado de sistema de energia, testar hipóteses e construir uma aproximação relevante para a política durante uma sessão analítica.

O planejamento da operação hidrotérmica há muito tempo é tratado por meio de programação dinâmica estocástica e decomposição. O SDDP [5] representa a função de custo futuro por meio de cortes afins e é amplamente utilizado em sistemas hidrotérmicos de grande porte. Trabalhos subsequentes estenderam o método para a análise de convergência, a aversão a risco e os modelos intertemporais de aflúências [6, 16, 17]. Esses métodos fornecem o benchmark de otimização para o presente trabalho e são utilizados como referências externas, não como ferramentas disponíveis ao agente.

Problema de Operação Hidrotérmica Estocástica Multiestágio

Considere um horizonte finito $t = 1, \dots, T$. Seja V_t o armazenamento dos reservatórios, H_t a informação hidrológica e $X_t = (V_t, H_t)$ o estado do sistema. A incerteza é representada por ξ_t , incluindo aflúências e, possivelmente, demanda ou geração renovável. Em cada estágio, as decisões u_t incluem geração hidrelétrica, geração térmica, fluxos de intercâmbio, vertimento, déficit e armazenamento do estágio seguinte. O problema do estágio é

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

onde U_t contém as restrições de balanço hídrico, geração, transmissão e demanda.

Para cada subsistema hidrelétrico r , o armazenamento evolui como

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

com aflúência $a_{t,r}$, vazão turbinada $q_{t,r}$, vertimento $s_{t,r}$ e coeficiente de produção $\eta_{t,r}$. Para cada área elétrica n ,

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

O conjunto factível também inclui limites de armazenamento, geração hidrelétrica, geração térmica, fluxos de transmissão, vertimento e déficits. O custo imediato é

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

O problema estocástico multiestágio é

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

com políticas não antecipativas. A recursão de Bellman é

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

com a condição terminal $Q_{T+1}(X_{T+1}) = \Phi(V_{T+1})$. O valor marginal da água armazenada é

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

A discretização direta por SDP é intratável, pois R reservatórios discretizados em K níveis já produzem K^R estados de armazenamento. O SDDP evita a enumeração completa de estados ao aproximar os custos futuros por meio de cortes afins:

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

Neste artigo, o SDDP é utilizado como referência conceitual e benchmark externo. O agente não recebe acesso a uma implementação do SDDP.

Arquitetura Orientada a Capacidades

Princípio arquitetural

A arquitetura proposta, orientada a capacidades (capability-driven architecture, CDA), separa o processo de raciocínio do agente da implementação computacional do modelo analítico de sistema de energia subjacente. O agente formula hipóteses, seleciona ações analíticas, interpreta observações e refina sua estratégia ao longo de uma sessão de múltiplas etapas, enquanto o ambiente executa operações de domínio validadas, como inspeção do modelo, simulação, avaliação de políticas e registro de artefatos.

Todas as interações ocorrem por meio de capacidades tipadas, com esquemas de entrada, esquemas de saída e semântica de execução definidos. O agente não tem acesso a arquivos do modelo, aspectos internos do solver, execução irrestrita de código, variáveis duais ou saídas de otimização pré-computadas. Essa separação permite que o agente opere como uma camada de raciocínio e orquestração, ao mesmo tempo em que preserva o controle, a auditabilidade e uma fronteira clara entre a inferência mediada por linguagem e a computação validada.

Formalmente, uma sessão analítica sob a CDA pode ser representada como uma trajetória de interação

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

onde a_k denota a invocação de capacidade na etapa k e o_k a correspondente observação estruturada retornada pelo ambiente.

Em cada etapa, o agente seleciona a próxima invocação de capacidade com base em todo o histórico de interação acumulado até então. Seja $h_k = (\mathcal{T}, a_1, o_1, \dots, a_{k-1}, o_{k-1})$ esse histórico, onde $\mathcal{T}$ é a especificação da tarefa fornecida na inicialização da sessão. O laço de raciocínio do agente pode então ser escrito como

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

onde π_A é a função de orquestração do agente: o mecanismo pelo qual o modelo de raciocínio interpreta o histórico acumulado e seleciona a próxima invocação de capacidade. Esse objeto não deve ser confundido com uma política operativa hidrotérmica. Ele governa o sequenciamento das ações analíticas durante a sessão agêntica, incluindo inspeção, simulação, teste de hipóteses e possível modificação de artefatos relacionados à política, como as aproximações de custo futuro. Diferentemente de um controlador de aprendizado por reforço treinado, π_A não é aprendido por meio de interação episódica

com o sistema hidrotérmico; ele é executado pelo modelo de raciocínio por meio de inferência mediada por linguagem sobre o histórico da sessão.

O ambiente, por sua vez, mantém um estado s_k composto pelo modelo analítico carregado, pelas definições de capacidades disponíveis e por quaisquer artefatos analíticos com estado criados durante a sessão. Na instanciação hidrotérmica, o artefato mais importante desse tipo é a aproximação de custo futuro atualmente registrada.

Cada invocação de capacidade produz uma observação e avança o estado do ambiente de acordo com

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

onde $\mathcal{E}$ é uma função de transição determinística. Invocações somente de leitura (inspeção, simulação e consultas a estatísticas de cenários) deixam s_k inalterado. Invocações que codificam políticas atualizam o conjunto de cortes registrado em s_{k+1} , tornando o ambiente dependente de estado em relação ao processo de construção da política. A sessão termina na etapa K quando o agente declara a tarefa analítica concluída. O estado terminal do ambiente s_K contém os artefatos produzidos durante a sessão. No caso hidrotérmico, esses artefatos incluem a aproximação de custo futuro registrada. Essa aproximação induz subsequentemente uma política operativa hidrotérmica quando incorporada ao problema de despacho do estágio e avaliada por meio de simulação estocástica, conforme descrito na Seção 5.

A trajetória τ é o principal objeto de análise post hoc: ela registra todo o caminho analítico, desde a inspeção do sistema até a validação da política, e permite a reconstrução e a auditabilidade do processo de raciocínio do agente.

A Figura 2 ilustra a estrutura geral dos componentes da arquitetura, resumindo as relações formais descritas pelas equações (9)–(11).

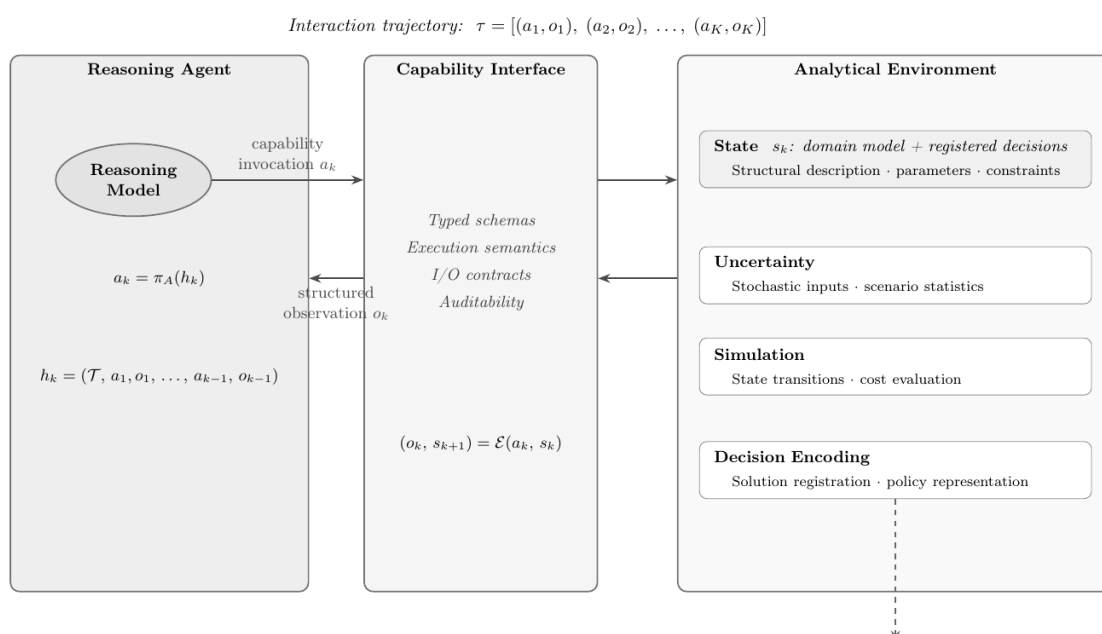


Figura 2: Arquitetura orientada a capacidades.

Instanciação hidrotérmica da interface de capacidades

A instanciação hidrotérmica expõe as operações de domínio ao agente por meio de oito capacidades tipadas, organizadas em quatro grupos analíticos. O projeto segue um princípio de delimitação deliberada de informações: cada capacidade fornece a informação necessária para apoiar uma etapa

analítica bem definida, ao mesmo tempo em que oculta aspectos internos do solver, modelos algorítmicos, variáveis duais e fluxos de dados brutos que permitiriam ao agente contornar a tarefa de raciocínio. Essa delimitação não é apenas uma conveniência técnica; é o mecanismo pelo qual o ambiente restringe o agente a realizar um raciocínio relevante para o domínio, em vez de computação arbitrária ou de mero envolvimento do solver.

Uma propriedade arquitetural adicional é a persistência de estado da sessão. O modelo é carregado uma única vez na inicialização da sessão, fixando o reservatório e estabelecendo a convenção de coordenadas do vetor de armazenamento utilizada em todas as chamadas subsequentes de simulação e de codificação de política. Esse contrato de dados explícito entre o ambiente e o agente garante que toda invocação de capacidade opere sobre uma representação compartilhada e inequívoca do sistema físico.

A Tabela 1 resume a interface na prática.

Tabela 1: Interface de capacidades hidrotérmicas: grupos de capacidades, entradas, saídas e papéis analíticos.

Grupo	Principais entradas / saídas	Papel analítico
Inspeção do sistema	dados do caso $\rightarrow$ reservatórios, estágios, cenários	Inicializar a sessão; fixar a convenção de coordenadas do vetor de armazenamento
	$\emptyset \rightarrow$ topologia completa	Expor o grafo de subsistemas, blocos de custo, transmissão e demanda
	$\emptyset \rightarrow T, S $	Fornecer o horizonte de planejamento e a contagem de cenários
Cenários hidrológicos	$\emptyset \rightarrow$ média, desvio-padrão por nó por estágio	Caracterizar a estrutura distributiva das afluições; trajetórias individuais não acessíveis
Simulação da operação	estágio, cenário, $V \rightarrow$ custos, V'	Oráculo controlado de estágio único; somente resultados primais
	cenário $\rightarrow$ trajetórias de custo e armazenamento	Trajetória de horizonte completo para uma série de afluições
	$\emptyset \rightarrow$ resultados agregados	Avaliação em lote de todos os cenários
Codificação de política	estágio, $\beta, \Lambda \rightarrow$ confirmação	Registrar aproximações afins de custo futuro; substitui os cortes existentes naquele estágio

As capacidades de inspeção do sistema expõem a descrição estrutural estática do modelo hidrotérmico. Uma chamada de carregamento de dados inicializa a sessão e retorna o nome ordenado do reservatório, o volume inicial, a capacidade máxima e a unidade física de cada componente de armazenamento. Essa ordenação é relevante: ela define o sistema de coordenadas para todos os vetores de armazenamento trocados em chamadas subsequentes, desde as entradas de simulação de estágio único até as matrizes de coeficientes dos cortes de política. Uma chamada de inspeção de topologia fornece a descrição completa do sistema como um documento estruturado, incluindo nós de subsistemas, arcos de intercâmbio direcionais com limites de capacidade, blocos de geração térmica com patamares de custo, perfis de geração renovável e parâmetros sazonais de demanda. A imagem resultante corresponde ao que um analista de planejamento montaria ao revisar a documentação do modelo: topologia do sistema, caracterização dos recursos e restrições operacionais.

As capacidades de cenários hidrológicos fornecem uma caracterização estocástica agregada da incerteza de entrada. Para cada nó de entrada — incluindo aflúências aos reservatórios, geração renovável e perturbações de demanda — a capacidade retorna estatísticas-resumo por estágio ao longo do conjunto de cenários. As realizações individuais de cenários não são intencionalmente acessíveis por meio dessa capacidade. Essa escolha tem dois propósitos. Analiticamente, ela incentiva o agente a raciocinar a partir da estrutura distributiva, em vez de memorizar trajetórias amostrais individuais. Operacionalmente, ela evita a injeção de grandes tabelas de cenários brutos no histórico de interação, o que expandiria o contexto do agente com detalhes numéricos de baixo nível e diluiria o sinal analítico de alto nível necessário para as etapas de raciocínio subsequentes. O conjunto completo de cenários permanece disponível indiretamente por meio das capacidades de simulação e de avaliação em lote.

As capacidades de simulação da operação constituem o núcleo computacional da interface e oferecem três níveis de granularidade analítica. Uma chamada de simulação de estágio único avalia um estágio a partir de um vetor de armazenamento inicial, um índice de estágio e um índice de cenário, retornando um vetor de custos operativos por nó e o vetor de armazenamento de fim de estágio resultante. Esse oráculo apoia experimentos de perturbação direcionados: ao variar o vetor de armazenamento inicial em chamadas repetidas, mantendo fixos o estágio e o cenário, o agente pode estimar a sensibilidade local do custo operativo em relação aos níveis dos reservatórios — o procedimento numérico que produziu os valores marginais da água relatados na Seção 5. Uma chamada de simulação de sequência executa uma trajetória completa para um cenário de aflúências especificado, retornando perfis de custo e armazenamento estágio a estágio ao longo de todo o horizonte de planejamento. Uma chamada de avaliação em lote executa o modelo em todos os cenários e é apropriada para a validação final da política.

Uma propriedade definidora desse grupo é que os custos operativos e as transições de estado são observáveis, mas as variáveis duais e os preços marginais não são. O solver de estágio é executado internamente para cada chamada de simulação e retorna somente resultados primais; os preços-sombra não são expostos pela interface. Essa restrição tem consequências diretas para a estratégia analítica do agente. O custo de oportunidade da água armazenada deve ser inferido por meio de experimentação ativa, e não lido a partir da saída do solver.

As capacidades de codificação de política permitem que o agente registre aproximações de custo futuro que são incorporadas às chamadas de simulação subsequentes. Uma chamada de registro de política aceita um índice de estágio, um vetor de constantes do lado direito e uma matriz de coeficientes de armazenamento, cada linha definindo uma aproximação afim de custo futuro sobre a função de custo futuro. Essa chamada substitui todos os cortes existentes para o estágio-alvo: a acumulação incremental ao longo das etapas de iteração é responsabilidade do agente, e não do ambiente. Os coeficientes de armazenamento seguem uma convenção de sinais alinhada com a economia hidrotérmica: um coeficiente negativo no reservatório i codifica a observação de que água adicional naquele reservatório reduz os custos operativos futuros esperados, ou seja, que a água armazenada tem valor marginal positivo. Inversamente, passar vetores vazios de coeficientes e de lado direito limpa os cortes de um estágio, permitindo que o agente revise ou reinicie sua aproximação de política durante a sessão analítica.

Essa taxonomia é específica da aplicação hidrotérmica, mas reflete um princípio arquitetural mais amplo: as capacidades devem corresponder a operações analíticas significativas no domínio-alvo, agrupadas em um nível de abstração que corresponda à granularidade de raciocínio exigida pela tarefa. Neste caso, as capacidades expostas espelham o fluxo de trabalho seguido pelos analistas de planejamento hidrotérmico: inspecionar o sistema, caracterizar a estrutura de incerteza, simular consequências operacionais sob condições controladas e codificar a lógica intertemporal da política. A ocultação deliberada das informações duais e das trajetórias individuais de cenários garante que o agente deva

engajar-se com o problema analítico em vez de recuperar respostas pré-computadas por meio da interface.

A capacidade de codificação de política aceita aproximações afins de custo futuro da forma:

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

onde $Q_t(V)$ é a aproximação do custo futuro a partir do estágio $t + 1$ em diante, V_i é o armazenamento de fim de estágio no subsistema i , λ_i é o valor marginal da água armazenada, β_t é um intercepto específico do estágio e S é o conjunto de componentes de armazenamento representados no modelo.

Essa representação está estreitamente relacionada aos cortes afins utilizados em métodos de decomposição como o SDDP. Entretanto, o agente não recebeu uma implementação do SDDP, qualquer outro método de programação dinâmica estocástica pré-programado, ou um modelo algorítmico para a construção desses cortes. A capacidade apenas definia o objeto de política admissível que poderia ser registrado no ambiente. O processo de raciocínio utilizado para estimar os valores da água, escolher os coeficientes e montar a aproximação foi deixado a cargo do agente.

Resultados

Tarefa experimental e avaliação

O agente foi instruído a construir uma política operativa custo-eficiente para o sistema hidrotérmico. A sessão teve permissão para prosseguir até que o agente declarasse ter concluído a construção e a validação da política. O log de interação foi então analisado post hoc. Ao longo da sessão, o agente teve acesso às descrições das capacidades, aos esquemas de entrada-saída e às saídas de suas próprias invocações. Ele não recebeu uma implementação de programação dinâmica, como o SDDP, valores da água pré-computados ou variáveis duais, acesso ao estado interno do solver, ou orientação humana após a inicialização da tarefa. Toda a informação quantitativa utilizada pelo agente foi, portanto, obtida exclusivamente por meio de invocações estruturadas de capacidades, distinguindo esse cenário experimental de um em que um agente simplesmente chama um solver de otimização existente.

Todas as sessões foram conduzidas utilizando o Claude Sonnet 4.6 (Anthropic), um grande modelo de linguagem com capacidade de raciocínio em cadeia de pensamento estendida. O modelo foi implantado com o raciocínio estendido habilitado, o que permite que o modelo realize etapas explícitas de raciocínio intermediário antes de produzir cada resposta ou invocação de capacidade. As sessões usaram um modo de orquestração automática no qual o modelo de raciocínio principal roteia autonomamente subtarefas computacionalmente mais leves — como a análise de saídas estruturadas e a formatação de invocações — para o Claude Haiku 4.5, um membro menor e de menor latência da mesma família de modelos, ao mesmo tempo em que mantém a inferência completa com raciocínio estendido para as etapas analíticas que exigem raciocínio de múltiplas etapas. Essa configuração reflete um padrão prático de implantação para fluxos de trabalho agênticos de longo horizonte, no qual a profundidade do raciocínio e a eficiência computacional são equilibradas automaticamente no nível da inferência, sem intervenção manual.

O experimento foi projetado para avaliar se o agente conseguia:

1. Identificar a estrutura relevante do sistema hidrotérmico por meio das capacidades de inspeção;
2. Inferir a lógica intertemporal do valor da água por meio de simulação;
3. Codificar uma aproximação de custo futuro usando a representação de política disponível;
4. Validar a política resultante em cenários estocásticos de afluências.

A política resultante foi avaliada usando o custo operativo da média amostral, a variabilidade de custo entre cenários, eventos de déficit de energia, o uso de geração térmica de emergência de alto custo, as trajetórias de armazenamento dos reservatórios e a comparação em nível de cenário com duas referências:

- **Baseline míope:** uma política de visão de curto prazo que prioriza a geração hidrelétrica imediata sem considerar o valor futuro da água, correspondendo à resolução de cada estágio de forma independente com uma aproximação de custo futuro nula.
- **Benchmark SDDP:** uma política obtida de forma independente usando uma implementação de programação dinâmica dual estocástica construída para esse fim.

Como os agentes de raciocínio são sistemas estocásticos — cujas saídas dependem de processos internos de amostragem que variam entre sessões mesmo sob prompts e condições iniciais idênticos —, o experimento foi conduzido dez vezes de forma independente, cada execução inicializada a partir do mesmo modelo e da mesma especificação de tarefa. As dez execuções produzem diferentes trajetórias de interação $\tau^{(r)}$, $r = 1, \dots, 10$, levando potencialmente a diferentes objetos de política $\pi_t^{(r)}$ e a diferentes custos operacionais. A análise da distribuição de resultados entre execuções evita que conclusões sejam tiradas com base em uma única sessão, potencialmente atípica.

Estudo de caso hidrotérmico brasileiro

O experimento utilizou uma representação simplificada, mas estruturalmente realista, do Sistema Interligado Nacional brasileiro (SIN). O sistema é representado por quatro subsistemas interligados: Sudeste/Centro-Oeste (SECO), Sul (SUL), Nordeste (NE) e Norte (N). O horizonte de planejamento abrange 24 estágios mensais, de janeiro de 2025 a dezembro de 2026, com dez cenários estocásticos de aflúências.

O estudo de caso inclui reservatórios hidrelétricos agregados, blocos de geração térmica em camadas, geração renovável, perfis de demanda e intercâmbios inter-regionais de transmissão. Não pretende reproduzir o modelo oficial completo de operação brasileiro. Em vez disso, fornece um benchmark controlado com as principais características estruturais necessárias para a análise da operação hidrotérmica: acoplamento espacial, dinâmica de armazenamento, incerteza hidrológica, substituição térmica e penalidades de escassez.

A Tabela 2 resume os principais parâmetros do caso.

Parâmetro	Valor
Horizonte de planejamento	Janeiro de 2025 – Dezembro de 2026
Estágios	24 estágios mensais
Subsistemas	SECO, SUL, NE, N
Cenários hidrológicos	10 cenários estocásticos de aflúências
Representação da política	Cortes afins de custo futuro

Tabela 2: Principais parâmetros do estudo de caso.

A Tabela 3 resume as principais características dos subsistemas utilizadas no estudo de caso.

Subsistema	Hidro máx. (MW)	Reservatório (GWh)	Armazenamento inicial	Faixa 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

Tabela 3: Características agregadas dos subsistemas no estudo de caso hidrotérmico.

O estudo de caso também inclui intercâmbios de transmissão entre os subsistemas. Os principais limites direcionais são 15.600 MW de N para SECO, 2.000 MW de SECO para N, 6.200 MW de N para NE, 13.000 MW de NE para SECO e 8.000 MW entre SECO e SUL. A geração térmica é representada por blocos em camadas com custos crescentes. O subsistema SECO desempenha um papel central porque seu bloco térmico de alto custo a 600 \$/MWh frequentemente atua como recurso marginal de todo o sistema em condições de escassez.

A Figura 3 mostra o esquema do estudo de caso, ilustrando os quatro subsistemas e os intercâmbios direcionais de transmissão que os conectam.

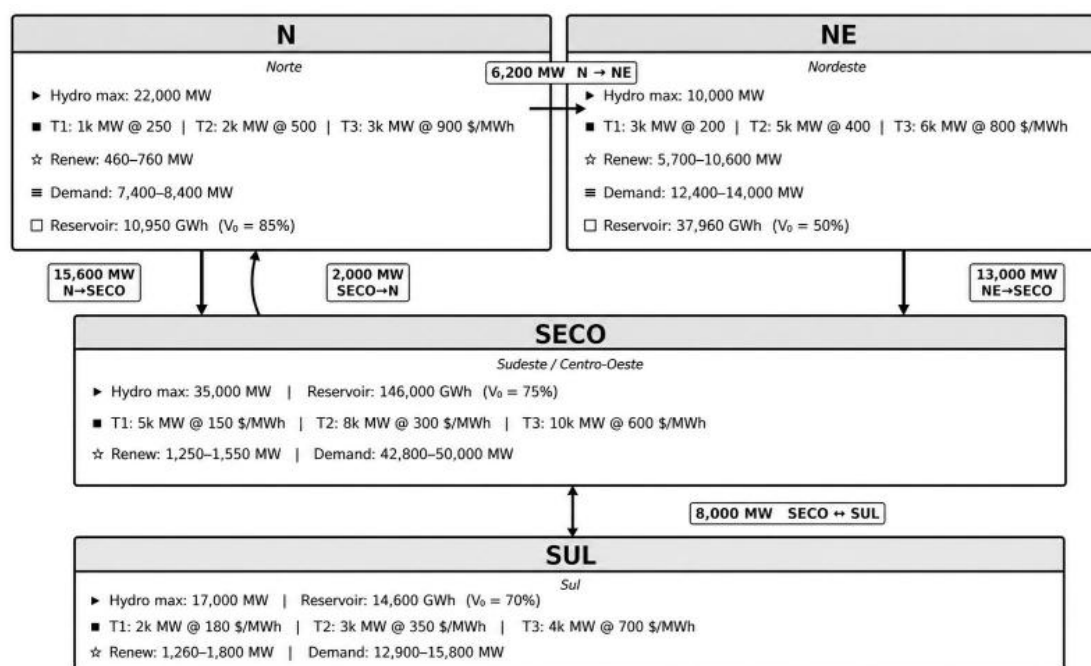


Figura 3: Esquema do estudo de caso hidrotérmico baseado no Sistema Interligado Nacional brasileiro (SIN).

Conjunto de cenários hidrológicos

Os dez cenários de afluências foram construídos para representar regimes hidrológicos estruturalmente distintos, e não perturbações aleatórias em torno de uma estimativa central. A Tabela 4 reporta o tipo de cada cenário e o desvio anual de afluências em relação à média do conjunto, por subsistema e para o sistema como um todo.

Cenário	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	Úmido	+31	+34	+30	+31	+31
5	Úmido (La Niña)	+45	0	+36	+39	+35
6	Úmido	+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

Tabela 4: Conjunto de cenários hidrológicos: tipo e desvio anual de afluências em relação à média do conjunto (%).

O conjunto abrange três realizações próximas da média (cenários 1–3), três anos úmidos (4–6, de +28% a +35% acima da média), três anos secos (7–9, de -16% a -24% abaixo da média) e um cenário crítico (10) modelado a partir da crise energética brasileira de 2001, com todos os subsistemas aproximadamente 47% abaixo de suas afluências médias. As afluências anuais totais do sistema variam de 233.300 GWh a 586.900 GWh, um fator de 2,5×, com razões individuais de subsistema chegando a 2,7× para o SECO.

Além da severidade agregada, vários cenários apresentam pronunciada heterogeneidade entre subsistemas. O cenário 5 reproduz um padrão do tipo La Niña, no qual SECO, NE e N estão bem acima da média (+45%, +36%, +39%), enquanto o SUL recebe afluências próximas da média. O cenário 8 inverte isso em uma configuração El Niño: SECO, NE e N severamente abaixo da média (-38%, -28%, -23%), enquanto o SUL está acima da média (+15%). Em conjunto, os dez cenários cobrem um amplo espectro de severidade agregada e de estrutura espacial das afluências, fornecendo um teste exigente de robustez da política, apesar do tamanho modesto do conjunto.

Estimativa do valor marginal da água

Conforme estabelecido no projeto experimental, cada uma das dez execuções seguiu uma trajetória de raciocínio independente. Apesar disso, a inspeção dos logs de interação revelou um padrão consistente entre as execuções: o agente convergiu recorrentemente para uma abordagem de diferenças finitas para estimar o valor marginal da água armazenada, independentemente da sequência específica de invocações de ferramentas que a precedeu. Essa convergência sugere que a abordagem é uma consequência natural das capacidades disponíveis e da estrutura do problema, e não um artefato de uma sessão específica.

Para cada subsistema i , ela comparou o custo operativo simulado em um vetor de armazenamento de referência com o custo obtido após perturbar o armazenamento daquele subsistema por um pequeno incremento ΔV :

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

Construção retroativa dos cortes de custo futuro

Um segundo padrão que se repetiu de forma consistente entre as execuções foi a estratégia para a construção dos cortes de custo futuro. Após estimar os valores da água por subsistema por meio de diferenças finitas, o agente prosseguiu sistematicamente para codificá-los como aproximações afins de custo futuro sobre a função de custo de recurso (cost-to-go) por indução retroativa — um procedimento que emergiu de forma independente em cada sessão sem ter sido prescrito na especificação da tarefa.

Usando os coeficientes de valor da água por subsistema $\hat{\lambda}_i$ estimados na etapa anterior, o agente construiu uma sequência de cortes de custo futuro por indução retroativa. Começando com um valor terminal nulo, para cada estágio $t = T, \dots, 1$, ele simulou a operação a partir de um estado inicial e calculou:

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

onde c_t é o custo operativo imediato retornado pelo simulador e $V_{t,i}^{\text{fin}}$ é o armazenamento de fim de estágio resultante no subsistema i .

A política resultante é representada por uma aproximação afim de custo futuro sobre o custo futuro:

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

O Algoritmo 1 resume o procedimento reconstruído a partir do log de interação.

Algoritmo 1 Construção retroativa, derivada pelo agente, de cortes afins de custo futuro

Requer: Coeficientes de valor da água por subsistema $\hat{\lambda}_i$, intercepto terminal $\beta_{T+1} = 0$

1. **para** $t = T, T - 1, \dots, 1$ **faça**
2. Simule o estágio t a partir de armazenamento inicial nulo
3. Observe o custo imediato c_t e o vetor de armazenamento final V_t^{fin}
4. Calcule $\beta_t = c_t - \sum_i \hat{\lambda}_i V_{t,i}^{\text{fin}} + \beta_{t+1}$
5. Defina o corte $Q_t(V) \geq \beta_t - \sum_i \hat{\lambda}_i V_i$
6. **fim para**
7. Registre todos os cortes não terminais no ambiente de otimização

Esse procedimento assemelha-se a um único passo retroativo simplificado do SDDP, mas não foi fornecido ao agente como um modelo algorítmico. O perfil de interceptos decresce à medida que o horizonte se aproxima do estágio terminal, refletindo o valor decrescente dos períodos operativos futuros.

Análise de reprodutibilidade entre execuções

A Tabela 5 reporta o custo operativo médio obtido em cada uma das dez execuções independentes, listadas na ordem das sessões.

Execução	Custo médio (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

Execução	Custo médio (M\$)
R9	139,21
R10	137,72
Média geral	143,48
Desvio-padrão	6,02
Mín.	137,72
Máx.	153,94

Tabela 5: Custo operativo total médio para cada execução independente (M\$).

Seis execuções (R1, R5, R7–R10) produziram políticas com custos médios na faixa de 137,7–140,0 M\$, uma dispersão de apenas 2,3 M\$ entre seis sessões independentes. As quatro execuções restantes (R2–R4, R6) renderam custos médios mais altos, de 145,8–153,9 M\$. A média geral entre todas as dez execuções é de 143,5 M\$, com desvio-padrão de 6,0 M\$. Uma observação notável é que as execuções R1 e R7 produziram custos por cenário idênticos até a precisão numérica, indicando que o agente chegou exatamente à mesma política por meio de duas trajetórias de interação independentes — uma forma de convergência que sugere que o caminho analítico até a estimativa dominante do valor da água é suficientemente restrito pela estrutura de capacidades para ser reproduzível.

Trajetória de raciocínio da execução de melhor desempenho

Entre as dez execuções independentes, a sessão que atingiu o menor custo operativo médio (137,7 M\$) deixou um log de interação suficientemente detalhado para reconstruir sua trajetória analítica por completo. O agente começou com a inspeção do sistema (Fase 1), carregando os dados do modelo, a estrutura de custos e as estatísticas dos cenários. Na Fase 2, ele simulou todos os dez cenários sem nenhum corte registrado, estabelecendo um baseline no qual os reservatórios esvaziavam entre os estágios 6 e 8 em todos os cenários e o sistema operava inteiramente com despacho térmico a partir do estágio 8 — o clássico colapso míope.

A Fase 3 consistiu em experimentos direcionados de perturbação de estágio único. Ao perturbar cada reservatório isoladamente em um estágio e cenário estressados, o agente estimou o valor marginal da água armazenada: aproximadamente 600 \$/MWh para SECO, SUL e N, e 567 \$/MWh para NE. Uma inferência fundamental foi a de que todos os quatro subsistemas eram dominados pelo mesmo recurso marginal de todo o sistema — o terceiro bloco térmico no SECO a 600 \$/MWh —, refletindo o alto grau de interligação do sistema.

Na Fase 4, o agente registrou um primeiro conjunto de cortes afins com coeficientes definidos como ligeiramente mais negativos do que o limiar estimado ($\alpha = [-700, -700, -600, -700]$ \$/MWh). A avaliação dessa política revelou uma inconsistência: o armazenamento do SECO estava se acumulando enquanto o bloco térmico caro estava sendo despachado simultaneamente, indicando que os coeficientes estavam penalizando o uso da água de forma agressiva demais. O agente identificou essa preservação excessiva e revisou os coeficientes para baixo na Fase 5 ($\alpha_{\text{SECO}} = -590$, $\alpha_{\text{NE}} = -560$), alinhando-os com os valores marginais estimados na Fase 3. A política refinada produziu um custo operativo médio de aproximadamente 138 M\$ e foi aceita como saída final.

Essa trajetória é notável por sua estrutura autocorretiva. Sem orientação externa, o agente detectou uma inconsistência da política por meio do retorno da simulação e revisou sua representação do valor da água de acordo.

Comparação com a operação míope

A Tabela 6 compara a política derivada pelo agente com o baseline míope.

Indicador	Míope	Agente
Custo médio (M\$)	275,7	137,7
Emergência T3 (estágios)	10	0
Eventos de déficit (estágios)	4	0
Economia média (M\$)	–	138,0

Tabela 6: Comparação entre a operação míope e a política derivada pelo agente.

A política míope deplecia rapidamente os reservatórios durante a primeira estação úmida, deixando um buffer hidráulico insuficiente para os períodos secos posteriores. A política derivada pelo agente, em vez disso, preserva o armazenamento, aceitando custos térmicos mais altos no curto prazo para reduzir o risco futuro de escassez. Esse comportamento é consistente com a interpretação econômica dos valores da água na operação hidrotérmica estocástica: a água deve ser utilizada quando seu benefício imediato excede seu custo de oportunidade futuro esperado.

A Figura 4 apresenta a evolução estocástica do armazenamento dos reservatórios ao longo dos dez cenários de afluências sob a política derivada pelo agente e o baseline míope. Cada painel corresponde a um subsistema. As linhas em negrito mostram a média dos cenários; as bandas sombreadas reportam os intervalos P25–P75 e P10–P90 entre cenários. Os níveis de armazenamento são expressos como percentual da capacidade máxima de cada subsistema.

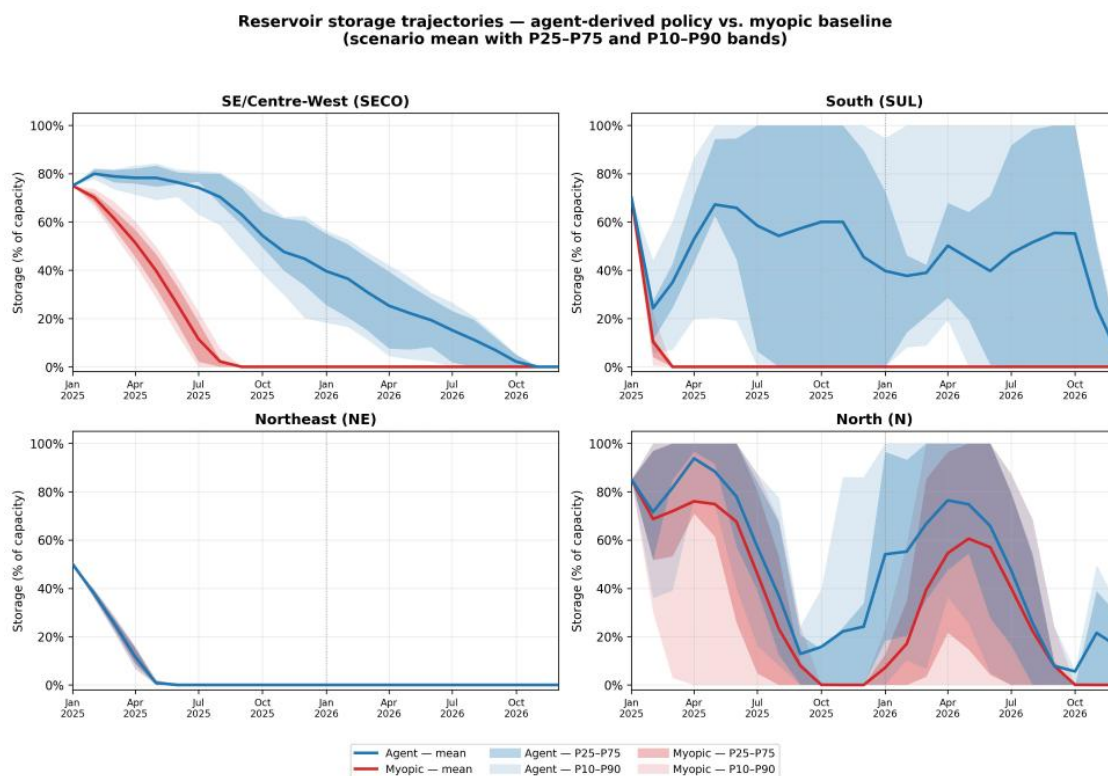


Figura 4: Trajetórias de armazenamento dos reservatórios sob a política derivada pelo agente e o baseline míope. As curvas vermelhas denotam o baseline míope.

O contraste entre as duas políticas é pronunciado. Sob o baseline míope, todos os subsistemas se depletam a níveis próximos de zero nos estágios 7–8 (julho–agosto de 2025) na maioria dos cenários, e o sistema opera inteiramente com despacho térmico pelo restante de ambas as estações secas. Sob a política derivada pelo agente, o reservatório do SECO, que concentra aproximadamente 70% da capacidade total de armazenamento do sistema, é mantido bem acima de zero ao longo de toda a primeira estação seca e entra em 2026 com reserva hidráulica suficiente para amortecer o segundo período seco. Os subsistemas NE e N, que possuem capacidades absolutas menores, exibem diferenças qualitativas semelhantes, enquanto o SUL é parcialmente recuperado por seu próprio regime de afluições.

Comparação com o SDDP

A Tabela 7 compara a execução do agente de melhor desempenho (R10, custo médio de 137,7 M\$) com o benchmark SDDP, usando realizações de afluições correspondentes. Essa execução é selecionada aqui porque produziu o menor custo médio entre as dez sessões independentes e, portanto, representa o resultado mais favorável da abordagem orientada a capacidades. Os resultados das demais execuções são reportados na Seção 5.5; entre todas as dez execuções, a diferença média de custo em relação ao SDDP é de aproximadamente 6,0%, refletindo a variabilidade introduzida pela natureza probabilística do agente de raciocínio. A referência SDDP foi obtida usando o solver comercial da PSR [18], aplicado aos mesmos dez cenários de afluições com cinco cenários retroativos por iteração.

Cenário	Agente (M\$)	SDDP (M\$)	Dif. (M\$)	Gap (%)
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
Média	137,7	135,4	2,3	1,7
Desvio-padrão	53,2	58,5	-	-

Tabela 7: Comparação em nível de cenário entre a execução do agente de melhor desempenho (R10) e o benchmark SDDP.

Para a melhor execução, a diferença média de custo em relação ao SDDP é de 1,7%. Esse resultado deve ser interpretado com cautela. A política SDDP é obtida usando um algoritmo especializado de otimização estocástica com repetidas iterações forward-backward, enquanto a política derivada pelo agente é uma política afim simplificada, construída por meio de interação estruturada. A observação relevante, portanto, não é que o agente compete com o SDDP, mas que o ambiente orientado a capacidades permitiu ao agente construir uma política com lógica intertemporal coerente de valor da água.

Em três cenários, a política derivada pelo agente alcançou um custo ex post menor do que o benchmark SDDP. Isso não implica dominância sobre o SDDP. Em vez disso, reflete a variação em nível de cenário e a natureza conservadora da política derivada pelo agente, que tende a preservar o armazenamento de forma mais agressiva. Em média, a política SDDP permanece o melhor benchmark de otimização estocástica.

Discussão

Orquestração estruturada de capacidades

O experimento mostra que um agente de raciocínio pode usar capacidades estruturadas para realizar operações analíticas relevantes para o domínio em um ambiente estocástico de sistema de energia. Sua trajetória de interação assemelhou-se ao fluxo de trabalho de um analista humano trabalhando com um modelo de simulação: o agente não recebeu uma política ótima diretamente, mas desenvolveu progressivamente uma lógica operativa aproximada que orientou as decisões subsequentes.

O ponto importante não é que o agente tenha descoberto um novo método de otimização. Em vez disso, a estrutura de capacidades restringiu a interação de maneira produtiva. Ao expor capacidades em níveis apropriados de abstração, o ambiente permitiu que o agente formulasse e testasse hipóteses operacionais, inferisse a lógica intertemporal do valor da água e construísse uma representação aproximada da estrutura de custo de recurso (cost-to-go) subjacente ao problema de programação dinâmica estocástica.

Implicações para a análise energética assistida por IA

A arquitetura proposta sugere um papel para os agentes de raciocínio como intermediários analíticos integrados a ambientes de modelagem de sistemas de energia. Em vez de substituir simuladores, solvers de otimização ou especialistas humanos, os agentes podem apoiar a camada exploratória em torno dos modelos formais, selecionando computações, interpretando saídas, testando hipóteses e coordenando análises iterativas.

Esse papel é especialmente relevante para problemas energéticos cujo contexto de decisão é mais amplo do que uma única formulação de otimização bem especificada. Os estudos de planejamento e operação frequentemente combinam premissas incertas, dados heterogêneos, múltiplos modelos, restrições regulatórias, considerações de desenho de mercado e trade-offs difíceis de codificar completamente em forma analítica fechada. Nesses casos, um sistema agêntico pode ser útil não porque calcula diretamente um ótimo, mas porque ajuda a estruturar a exploração de alternativas, sensibilidades, cenários e insights relevantes para a política.

As aplicações potenciais incluem:

- Exploração de grandes espaços de cenários estocásticos ou de teste de estresse;
- Análise comparativa entre premissas e configurações de modelo alternativas;
- Identificação de sensibilidades, gargalos e fatores operacionais determinantes;
- Refinamento iterativo de políticas candidatas ou alternativas de planejamento;
- Orquestração de múltiplas ferramentas de simulação, previsão e otimização;
- Explicação e interrogação de resultados de modelos para apoio à decisão.

Nessa visão, o projeto das capacidades torna-se central. O agente só pode raciocinar de forma eficaz sobre um modelo de domínio se as operações analíticas relevantes forem expostas de maneira controlada, semanticamente significativa e auditável. A programação hidrotérmica fornece um cenário

de validação com benchmark para essa arquitetura, porque há referências de otimização confiáveis disponíveis. Trabalhos futuros devem avaliar se frameworks agênticos semelhantes podem apoiar a exploração de soluções em problemas nos quais as formulações analíticas totalmente especificadas são incompletas, difíceis de construir ou insuficientes para representar o contexto de decisão relevante.

Limitações

Várias limitações devem ser reconhecidas. Algumas são específicas do estudo de caso de programação hidrotérmica utilizado para validação, enquanto outras dizem respeito ao uso mais amplo de agentes de raciocínio em ambientes analíticos estruturados de sistemas de energia.

Quanto ao estudo de caso específico de programação hidrotérmica, o experimento foi conduzido sobre uma representação simplificada do SIN brasileiro. Embora estruturalmente realista, ela não captura toda a complexidade da operação no mundo real, incluindo o despacho de unidades (unit commitment) detalhado, restrições de segurança, restrições individuais de usinas, regras de mercado, restrições regulatórias e requisitos de reserva operativa. O caso deve, portanto, ser interpretado como um benchmark controlado para a avaliação da arquitetura, e não como uma representação completa da operação hidrotérmica de nível de produção.

Em segundo lugar, o procedimento induzido pelo agente não possui garantias de otimalidade. A qualidade do comportamento operativo resultante depende das hipóteses formuladas pelo agente, da sequência de invocações de capacidades e das capacidades disponibilizadas pelo ambiente. Os métodos clássicos de otimização estocástica exploram sistematicamente a formulação matemática e fornecem garantias de convergência sob premissas apropriadas; o procedimento derivado pelo agente não o faz.

Por fim, a escalabilidade da orquestração sequencial de capacidades permanece uma questão em aberto. Sistemas maiores podem exigir capacidades de simulação em lote, avaliação paralela de cenários, modelos de política mais robustos, mecanismos de gerenciamento de contexto e integração mais estreita com solvers de nível de produção.

Conclusões

Este artigo avaliou se um agente de raciocínio pode operar como uma camada analítica dentro de um ambiente estruturado de modelagem de sistemas de energia. Para tanto, propusemos um framework orientado a capacidades no qual o agente interage com o ambiente analítico somente por meio de capacidades estruturadas e tipadas. Esse projeto separa o raciocínio e a orquestração da computação de domínio validada, permitindo que o agente inspecione o modelo, execute simulações, teste hipóteses e registre artefatos relevantes para a política, sem execução irrestrita de código ou acesso aos aspectos internos do solver.

O framework foi avaliado em um estudo de caso de programação hidrotérmica estocástica baseado em uma representação de quatro subsistemas do SIN brasileiro. Sem acesso a um método de programação dinâmica estocástica pré-programado, variáveis duais, valores da água pré-computados ou informações internas do solver, o agente inferiu a lógica do valor da água a partir do retorno da simulação e construiu aproximações afins de custo futuro. Quando incorporadas ao modelo de despacho do estágio, essas aproximações induziram políticas operativas de preservação de armazenamento que melhoraram substancialmente a operação míope. Ao longo de dez sessões independentes, a melhor política atingiu uma diferença média de custo de 1,7% em relação a um benchmark SDDP calculado de forma independente, enquanto a diferença média entre as sessões foi de aproximadamente 6,0%.

Esses resultados não sugerem que os agentes de raciocínio devam substituir os métodos de otimização estocástica. Em vez disso, mostram que, quando restrita por uma interface de capacidades apropriada, a IA agêntica pode apoiar a camada exploratória em torno dos modelos de sistemas de energia:

formulando hipóteses, selecionando computações, interpretando saídas e construindo artefatos relevantes para a política, enquanto as ferramentas de simulação e otimização permanecem como o alicerce computacional.

Trabalhos futuros devem avaliar frameworks orientados a capacidades semelhantes em problemas de sistemas de energia nos quais o processo de decisão relevante vai além de uma única formulação de otimização totalmente especificada, incluindo cenários que exijam exploração de cenários, coordenação de múltiplos modelos, análise de sensibilidade e avaliação iterativa de políticas.

Referências

- [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.