

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL  
INSTITUTO DE FÍSICA – ESCOLA DE ENGENHARIA  
BACHARELADO EM ENGENHARIA FÍSICA

Maria Paula Coser Zamin

**REINFORCEMENT LEARNING APLICADO PARA OTIMIZAÇÃO DO  
CUSTO DA ENERGIA ELÉTRICA EM UM MODELO  
COMPUTACIONAL DE PLANTA DE CHUVEIRO COM  
AQUECIMENTO COM DOIS GRAUS DE LIBERDADE**

Porto Alegre  
2023

Maria Paula Coser Zamin

**REINFORCEMENT LEARNING APLICADO PARA OTIMIZAÇÃO DO  
CUSTO DA ENERGIA ELÉTRICA EM UM MODELO  
COMPUTACIONAL DE PLANTA DE CHUVEIRO COM  
AQUECIMENTO COM DOIS GRAUS DE LIBERDADE**

Trabalho de Conclusão do Curso II de Bacharelado em  
Engenharia Física da Universidade Federal do Rio  
Grande do Sul.

Orientador: Prof. Dr. Jorge Otávio Trierweiler  
Coorientador: Octávio B. Santiago

Porto Alegre  
2023

Maria Paula Coser Zamin

**REINFORCEMENT LEARNING APLICADO PARA OTIMIZAÇÃO DO  
CUSTO DA ENERGIA ELÉTRICA EM UM MODELO  
COMPUTACIONAL DE PLANTA DE CHUVEIRO COM  
AQUECIMENTO COM DOIS GRAUS DE LIBERDADE**

Este Trabalho de Conclusão de Curso foi julgado adequado para obtenção do Título de “Bacharel em Engenharia Física” e aprovado em sua forma final pela banca examinadora.

Porto Alegre, 25 de agosto de 2023.

---

Prof. Dr. Renato Vaz Linn  
Coordenador do Curso

**Banca Examinadora:**

---

Prof. Dr. Jorge Otávio Trierweiler  
Orientador  
Universidade Federal do Rio Grande do Sul

---

Prof. Dr. Marcelo Farenzena  
Universidade Federal do Rio Grande do Sul

---

Dr. Jônathan William Vergani Dambros  
Universidade Federal do Rio Grande do Sul

## AGRADECIMENTOS

Primeiramente, gostaria de agradecer aos meus pais, Marilene Coser Zamin e Paulo Gelindo Zamin, por terem me dado condições para que eu pudesse estudar em uma das melhores universidades do país e por sempre terem me incentivado nos estudos. Também agradeço por todo o apoio e suporte e por nunca terem deixado eu desistir dos meus sonhos e da minha carreira, mesmo com obstáculos no meu caminho.

Em seguida, gostaria de agradecer ao meu orientador, Jorge Otávio Trierweiler, por todo o apoio, suporte, orientação, dedicação e conselhos durante a construção do trabalho. Também gostaria de agradecer ao Octávio B. Santiago por ter atuado como meu coorientador e por toda ajuda fornecida.

Agradeço à UFRGS, que além de ter sido responsável pela minha formação profissional, preparou-me para encarar os desafios da vida. Também agradeço a todos os professores pelos conhecimentos e experiências compartilhados. Gostaria de agradecer especialmente ao meu orientador de Iniciação Científica, Rogério Luis Maltez, por ter me dado minha primeira oportunidade de trabalho e por ter me ensinado o gosto por pesquisa.

Também agradeço ao meu primeiro gerente, Ricardo Moraes, por ter me dado uma oportunidade que mudaria minha vida com meu primeiro estágio, pois me descobri como Cientista de Dados. Aos meus colegas de trabalho, obrigada pelos ensinamentos em Inteligência Artificial, que foram essenciais para que eu tivesse a base para desenvolver este trabalho.

Gostaria de agradecer também aos meus colegas de curso por todos os bons momentos compartilhados durante esta trajetória. Finalmente, agradeço de forma geral a todos que contribuíram, de alguma forma, para a realização deste trabalho.

## RESUMO

Com o crescente aumento da tarifa da energia elétrica no Brasil, é imprescindível que novas estratégias sejam exploradas para a redução do consumo de energia elétrica nos eletrodomésticos que mais encarecem a conta de luz do consumidor, a exemplo do chuveiro elétrico. Assim, este estudo aplicou *Reinforcement Learning* (RL) para otimização do custo da energia elétrica em um modelo computacional de planta de chuveiro com aquecimento com dois graus de liberdade. RL é um método de aprendizado de máquina que pode ser utilizado como uma alternativa ou complementação às estratégias clássicas de controle e otimização de processos industriais. Na primeira fase deste estudo, buscou-se compreender qual é a melhor estratégia de aplicação de RL no modelo de planta de chuveiro para maximizar a qualidade do banho: RL como ferramenta da camada supervisória ou RL como ferramenta de otimização em tempo real. Assim, três configurações de chuveiro foram treinadas com os algoritmos *Proximal Policy Optimization* (PPO) e *Soft Actor-Critic* (SAC) para diferentes temperaturas ambiente. O PPO obteve uma performance superior ao SAC em um tempo menor de treinamento, logo optou-se por utilizar somente o PPO. Além disso, o PPO alcançou recompensas médias maiores quando RL foi aplicado como ferramenta de otimização em tempo real, que incluía duas configurações de chuveiro, sendo que em uma delas foi implementada uma estratégia de *split-range* para a fração da resistência elétrica. Logo, essas duas configurações e o algoritmo PPO foram utilizados na segunda fase do estudo, cujo objetivo principal era verificar se é possível aplicar RL para otimizar os custos do banho, com ênfase no custo da energia elétrica, ao mesmo tempo que a qualidade é maximizada. Foram testados diversos modelos para ambas as configurações, variando os objetivos e os cenários de temperatura ambiente e tarifa da energia elétrica. O modelo que obteve em média os menores custos do banho e a maior qualidade foi o da configuração com *split-range* onde foi implementado um agente que construía uma política específica para cada cenário, e depois aprendia a selecionar qual a melhor política a ser aplicada dependendo das condições iniciais. Portanto, considera-se que este estudo obteve sucesso em aplicar RL para otimização da planta de chuveiro em questão e que o tópico apresenta diversas oportunidades e possibilidade para estudos futuros.

**Palavras-chave:** *Reinforcement Learning*. *Proximal Policy Optimization*. *Soft Actor-Critic*. Modelo computacional de planta de chuveiro. Camada supervisória. Camada de otimização em tempo real. Índice de qualidade do banho. Otimização de custos.

## LISTA DE FIGURAS

<b>Figura 1</b> - Iteração entre o agente e o ambiente em um problema MDP.....	16
<b>Figura 2</b> - Representação esquemática da arquitetura do algoritmo PPO. ....	20
<b>Figura 3</b> - Efeito do <i>clipping</i> na função objetivo do <i>actor</i> para o algoritmo PPO. ....	21
<b>Figura 4</b> - Representação esquemática dos símbolos utilizados no <i>brain design</i> . ....	25
<b>Figura 5</b> - Representação esquemática do modelo de planta de chuveiro.....	29
<b>Figura 6</b> - Representação esquemática da configuração A.....	32
<b>Figura 7</b> - Representação esquemática da configuração B.....	32
<b>Figura 8</b> - Representação esquemática da configuração C.....	33
<b>Figura 9</b> - Representação gráfica do Índice de Qualidade do Banho (IQB).....	34
<b>Figura 10</b> - Recompensa média obtida com os algoritmos PPO e SAC para temperatura ambiente de 25°C.....	38
<b>Figura 11</b> - Recompensa média obtida com o algoritmo PPO para diferentes temperaturas ambiente. ....	39
<b>Figura 12</b> - Temperatura de saída, vazão de saída e IQB para a configuração B treinada com o algoritmo PPO a 25°C.....	40
<b>Figura 13</b> - Comportamento das variáveis de aquecimento, das válvulas e do tanque para a configuração B treinada com o algoritmo PPO a 25°C. ....	40
<b>Figura 14</b> - Temperatura de saída, vazão de saída e IQB para a configuração C treinada com o algoritmo PPO a 25°C.....	41
<b>Figura 15</b> - Comportamento das variáveis de aquecimento, das válvulas e do tanque para a configuração C treinada com o algoritmo PPO a 25°C. ....	41
<b>Figura 16</b> - <i>Brain design</i> para o modelo 1.....	44
<b>Figura 17</b> - <i>Brain design</i> para o modelo 2.....	44
<b>Figura 18</b> - <i>Brain design</i> para os modelos 3 e 4. ....	44
<b>Figura 19</b> - <i>Brain design</i> para os modelos 5 e 6. ....	45
<b>Figura 20</b> - Temperatura de saída, vazão de saída e IQB para o modelo 3 da configuração B treinado com o algoritmo PPO para o cenário de temperatura ambiente de 15,5°C e tarifa da energia elétrica de R\$2,10. ....	51
<b>Figura 21</b> - Comportamento das variáveis de aquecimento, das válvulas e do tanque para o modelo 3 da configuração B treinado com o algoritmo PPO para o cenário de temperatura ambiente de 15,5°C e tarifa da energia elétrica de R\$2,10.....	52

## LISTA DE TABELAS

<b>Tabela 1</b> - Objetivos básicos utilizados para o <i>brain design</i> . .....	25
<b>Tabela 2</b> - Descrição das variáveis físicas do modelo computacional de chuveiro.....	29
<b>Tabela 3</b> - Espaço de ações para as configurações A, B e C. ....	36
<b>Tabela 4</b> - Número total de <i>steps</i> e tempo total de treinamento para cada modelo treinado com o algoritmo PPO na segunda fase do estudo. ....	46
<b>Tabela 5</b> - Soma do IQB durante um episódio, soma do IQB médio de um episódio e valores mínimo e máximo do IQB médio para todas as 96 combinações de cenários de temperatura ambiente e tarifa da energia elétrica. ....	47
<b>Tabela 6</b> - Soma dos custos totais do banho, elétrico, de gás e de água para todas as 96 combinações de cenários de temperatura ambiente e tarifa da energia elétrica. ....	48
<b>Tabela 7</b> – Resultados para validação dos melhores modelos para cenários com temperaturas ambiente de 15,5°C, 20,5°C e 25,5°C e tarifa da energia elétrica de R\$2,10. ....	50

## LISTA DE ABREVIATURAS E SIGLAS

A2C	– Advantage Actor-Critic
AC	– Actor-Critic
CRISP-DM	– Cross Industry Standard Process for Data Mining
DDPG	– Deep Deterministic Policy Gradient
DRL	– Deep Reinforcement Learning
DRTO	– Dynamic Real Time Optimization
DQN	– Deep Q-Learning
MDP	– Markov Decision Process
ML	– Machine Learning
MPC	– Model Predictive Control
PI	– Proportional Integral
PID	– Proportional Integral Derivative
PPO	– Proximal Policy Optimization
RL	– Reinforcement Learning
RTO	– Real Time Optimization
SAC	– Soft Actor-Critic
TR	– Trust Region
TRPO	– Trust Region Policy Optimization



## SUMÁRIO

<b>1</b>	<b>INTRODUÇÃO .....</b>	<b>9</b>
<b>2</b>	<b>REVISÃO BIBLIOGRÁFICA .....</b>	<b>13</b>
2.1	REINFORCEMENT LEARNING .....	13
2.1.1	Conceitos básicos .....	13
2.1.2	Markov Decision Process .....	15
2.1.3	Proximal Policy Optimization .....	18
2.1.4	Soft Actor-Critic .....	22
2.2	DESIGN DE SISTEMAS AUTÔNOMOS .....	23
<b>3</b>	<b>METODOLOGIA .....</b>	<b>26</b>
<b>4</b>	<b>ESTUDO DE CASO .....</b>	<b>28</b>
4.1	MODELO COMPUTACIONAL DE CHUVEIRO .....	28
4.1.1	Apresentação da modelagem da planta de chuveiro .....	28
4.1.2	Configurações do modelo de chuveiro.....	31
4.2	ÍNDICE DE QUALIDADE DO BANHO.....	33
4.3	CÁLCULO DO CUSTO DO BANHO .....	34
<b>5</b>	<b>RESULTADOS E DISCUSSÃO .....</b>	<b>36</b>
5.1	OTIMIZAÇÃO DE QUALIDADE.....	36
5.2	OTIMIZAÇÃO DE QUALIDADE E CUSTOS .....	42
5.3	VALIDAÇÃO DAS SOLUÇÕES PROPOSTAS .....	50
<b>6</b>	<b>CONCLUSÃO .....</b>	<b>53</b>
	<b>REFERÊNCIAS .....</b>	<b>55</b>
	<b>APÊNDICE A – Resultados complementares para a segunda fase do estudo com o algoritmo PPO .....</b>	<b>59</b>
	<b>APÊNDICE B – Implementação em Python do modelo 3 da configuração de chuveiro B com o algoritmo PPO.....</b>	<b>109</b>

## 1 INTRODUÇÃO

No Brasil, entre 2012 e 2020, a tarifa média da luz em reais por *MWh* aumentou em aproximadamente 105,8% (EPE, 2022). Apesar de a procura por fontes de energia própria estar sendo cada vez mais comum no país, a exemplo da energia solar, que entre 2020 e 2021, teve um aumento de 40,9% de potência instalada (EPE, 2022), apenas 0,18% da população utiliza uma fonte própria de energia em sua residência (Eletrobras, 2019). Portanto, a maioria dos brasileiros ainda depende da energia elétrica proveniente da rede geral de distribuição e, conseqüentemente, sofre com o aumento das tarifas da luz. Além disso, alguns eletrodomésticos que apresentam um consumo médio alto de energia acabam encarecendo ainda mais a conta de luz do consumidor.

Um exemplo é o chuveiro, que de acordo com (Ghisi; Gosch; Lamberts, 2007), é responsável por 20% dos gastos totais com a conta de luz. Conforme (Eletrobras, 2019), 37,52% dos domicílios brasileiros utilizam a energia proveniente da rede geral para o aquecimento da água do banho, o que é uma quantidade representativa, já que 60,23% das residências não utilizam nenhum tipo de aquecimento. Ou seja, o aumento da tarifa média da luz, combinado com o fato de grande parte da população brasileira ainda depender da energia elétrica da rede geral para o funcionamento do chuveiro, torna este um dos eletrodomésticos mais caros de uma residência. Uma alternativa mais viável seria utilizar o aquecimento a gás para o banho, já que o custo do gás geralmente é mais barato que a tarifa da energia elétrica. Entretanto, somente 0,59% da população brasileira aderiu ao aquecimento a gás no chuveiro (Eletrobras, 2019).

Outros fatores que também influenciam no custo do banho são o horário e a potência elétrica do chuveiro. No que tange ao primeiro deles, segundo (Eletrobras, 2019), 43,86% da população brasileira utiliza o chuveiro entre 18h00 e 20h00, horário em que a energia elétrica tende a ser mais cara devido à sobrecarga no sistema. Já em relação ao segundo fator, existe uma tendência à aquisição desses equipamentos com maior potência elétrica. De acordo com (EPE, 2016), em 2014, o consumo unitário de eletricidade do chuveiro foi de 493 *kWh*/equipamento/ano, enquanto o consumo unitário projetado para 2024 é de 515 *kWh*/equipamento/ano. Apesar de uma maior potência elétrica trazer mais conforto, a demanda de energia é maior, o que encarece a conta de luz.

Desse modo, de acordo com (Amasyali *et al.*, 2021), o número de estudos envolvendo estratégias de controle de sistemas de chuveiro e aquecimento de água para a redução do consumo de energia elétrica está crescendo cada vez mais. Existem várias alternativas disponíveis, desde as mais simples, como sistema baseado em regras, passando por controladores PIDs, até abordagens mais sofisticadas, tais como Controle Preditivo Multivariável (*Model Predictive Control* - MPC) e, mais recentemente, *Reinforcement Learning* (RL) (Wang; Hong, 2020). No caso da abordagem baseada em regras se poderia reduzir o *setpoint* da temperatura do banho quando a energia elétrica está mais cara.

Já na abordagem MPC, primeiramente, é necessário desenvolver um modelo que represente as características térmicas e dinâmicas do sistema. Depois, esse modelo é utilizado para prever possíveis distúrbios, como variações na temperatura ambiente da água utilizada para aquecer o banho. Por fim, as previsões dos distúrbios são integradas no modelo para resolver um problema de otimização com um ou mais objetivos. Entretanto, segundo (Wang; Hong, 2020) e (Amasyali *et al.*, 2021), o controle MPC também apresenta limitações, como a necessidade de desenvolver modelos individuais para sistemas com diferentes características, a quantidade de tempo exigida para construir modelos com acurácia, e a impossibilidade de realizar a otimização em tempo real.

As estratégias mencionadas nos últimos dois parágrafos também estão inseridas no contexto de controle de processos industriais, que possui cinco camadas, sendo que as três primeiras envolvem estratégias de controle aplicadas diretamente na planta (Trieweiler; Farenzena, 2007). A primeira delas, a regulatória, realiza o controle de inventário das plantas industriais através de controladores *Proportional Integral Derivative* (PID), garantindo a segurança e a estabilidade operacional. A próxima camada é a supervisória, que utiliza controladores PID e MPC para assegurar a qualidade do produto, diminuindo sua variabilidade e satisfazendo as restrições operacionais. Já a terceira camada é a de otimização em tempo real, cujo objetivo é maximizar a rentabilidade do processo através da otimização de um problema estacionário (RTO) ou dinâmico (DRTO) que define os *setpoints*, os *targets* e as especificações das variáveis manipuladas.

Assim, RL surge como uma alternativa às estratégias de controle clássicas, pois ele pode solucionar algumas das limitações do controle baseado em regra e do MPC e possibilitar o *tuning* contínuo e *online* dos parâmetros de controle (Spielberg; Gopaluni; Loewen, 2017).

RL é um método de aprendizado no qual através de um processo de tomada de decisão, busca-se desenvolver uma estratégia para maximizar o retorno total (Sutton; Barto, 2018). Os problemas de RL são divididos em episódios, sendo que um episódio pode conter uma iteração ou mais. A cada iteração, um agente interage com um ambiente incerto por um método de tentativa e erro, a fim de desenvolver uma política que realize um mapeamento entre as melhores ações. Este ambiente, representado por estados, pode ser construído na vida real ou em uma simulação, ou ser uma combinação de ambos (Winder, 2020).

Na literatura, existem alguns trabalhos que utilizam RL para otimização do custo da energia elétrica em sistemas de aquecimento elétrico de água. Amasyali *et al.*, 2021 propôs uma abordagem de RL para minimizar os custos de um aquecedor elétrico da água sob uma política de preços de eletricidade por tempo de uso e conseguiu reduzir os custos da energia elétrica na faixa de 19% a 35% se comparado ao MPC. No trabalho seguinte, Amasyali *et al.*, 2023 realizou o *deployment* do seu modelo sem nenhum treinamento prévio, e mesmo assim, o modelo de RL obteve uma economia entre 11% e 14% no custo da energia elétrica se comparado com a abordagem tradicional. Outros trabalhos, como os de (Al-Jabery *et al.*, 2015) e (Zsembinski *et al.*, 2021), também obtiver sucesso ao utilizar RL para redução dos custos da energia elétrica em sistemas de aquecimento de água.

Contudo, nenhum desses trabalhos aplica RL em um sistema de chuveiro com dois graus de liberdade, ou seja, com duas formas de aquecimento da água, elétrica ou a gás. Além disso, nenhum dos trabalhos citados implementa um modelo computacional que inclua o tempo morto na tubulação de saída do chuveiro ou do aquecedor de água, simulando a perda de calor do sistema para o ambiente e para a tubulação. Portanto, o objetivo geral desse trabalho é aplicar RL em um estudo de caso didático de um modelo computacional de planta de chuveiro com duas formas de aquecimento e com tempo morto. O sistema de RL projetado deverá ser capaz de maximizar a qualidade do banho e minimizar os custos, com ênfase no custo da energia elétrica, levando em conta cenários com diversas temperaturas ambiente e bandeiras da tarifa de energia elétrica.

Para isso, este trabalho foi dividido em dois estudos principais. O primeiro deles busca compreender qual é a melhor estratégia de aplicação de RL na planta de chuveiro: RL como ferramenta na camada supervisória ou RL como ferramenta na camada de otimização em tempo real (RTO). Para cada uma das estratégias serão implementados modelos computacionais para

a planta de chuveiro com diferentes configurações de controle. O treinamento será realizado com dois algoritmos, *Proximal Policy Optimization* (PPO) e *Soft Actor-Critic* (SAC), cujas implementações estão disponíveis na biblioteca *RLLib* do software *Python*. Em seguida, a estratégia e o algoritmo que resultaram na melhor performance serão utilizados no segundo estudo, que busca compreender se é possível aplicar RL para maximizar a qualidade do banho ao mesmo tempo em que o custo da energia elétrica é minimizado, levando em conta diferentes cenários iniciais.

O presente trabalho é dividido como segue. No Capítulo 2, é apresentada a revisão bibliográfica dos conceitos aplicados neste estudo, incluindo conceitos básicos de RL e o funcionamento dos algoritmos escolhidos. O Capítulo 3 detalha a metodologia utilizada, dando ênfase ao processo de formulação de um problema de RL. Em seguida, o Capítulo 4 descreve o estudo de caso, detalhando as equações utilizadas para simular a planta de chuveiro, as configurações das plantas projetadas e os índices desenvolvidos para compor as funções de recompensa do agente de RL. Finalmente, o Capítulo 5 apresenta os resultados encontrados, seguido pelas conclusões do Capítulo 6.

## 2 REVISÃO BIBLIOGRÁFICA

### 2.1 REINFORCEMENT LEARNING

#### 2.1.1 Conceitos básicos

*Reinforcement Learning* (RL) é um método de *Machine Learning* (ML) cujo objetivo é aprender, através de tentativa e erro, quais ações levam a uma maior recompensa numérica a longo prazo. RL é diferente dos outros métodos de ML, que podem ser supervisionados ou não-supervisionados. No aprendizado supervisionado, os algoritmos aprendem informações importantes e padrões através de um conjunto de dados de treinamento previamente rotulados. Já no aprendizado não-supervisionado, os dados utilizados para treinamento não são rotulados, e os algoritmos aprendem a tomar decisões através das características dos dados. Embora RL não dependa de exemplos com um comportamento correto, ele não pode ser pensado como um tipo de aprendizado não-supervisionado, pois o objetivo de RL é maximizar a recompensa total, e não descobrir um padrão escondido nos dados, por isso consiste em uma categoria por si diferente das demais (Sutton; Barto, 2018).

Existem quatro componentes principais em um problema de RL: *ação*, *ambiente*, *agente* e *recompensa* (Winder, 2020). As *ações* são as decisões a serem tomadas, como por exemplo, escolher entre aquecimento elétrico ou a gás no chuveiro e aumentar ou diminuir o *setpoint* da temperatura de saída do banho. O *ambiente* é um sistema ou simulação que retorna os estados obtidos após a interação com o agente. O *agente* é a entidade que toma as decisões e que será responsável por desenvolver uma política, ou estratégia, que mapeie as melhores ações. Em termos de controle, o agente equivale ao controlador, otimizador ou operador. Por fim, a *recompensa* é a forma de o agente saber quais ações obtiveram o melhor desempenho e quais levaram o sistema a falha, funcionando de forma análoga a uma função objetivo.

Dessa forma, a cada iteração no tempo  $t$ , a partir de um estado do ambiente  $s_t$ , onde  $s_t \in \mathcal{S}$ , e  $\mathcal{S}$  é um conjunto que contém todos os estados do ambiente, o agente seleciona uma ação  $a_t$ , onde  $a_t \in \mathcal{A}$ , e  $\mathcal{A}$  é um conjunto com todas as ações possíveis. Subsequentemente, o agente aplica esta ação ao ambiente e estima um valor para a recompensa. A interação entre o agente e o ambiente faz com que este sofra uma transição para um novo estado  $s_{t+1}$ , gerando

uma recompensa  $r_{t+1}$ , onde  $r_{t+1} \in \mathcal{R}$ , e  $\mathcal{R}$  é um conjunto que contém todas as recompensas. Comparando a recompensa obtida com a estimada, o agente é capaz de atualizar sua política na direção das melhores ações.

Para maximizar a recompensa cumulativa ao longo dos episódios, é necessário maximizar o valor esperado do retorno  $G_t$ , definido como uma função de sequência de recompensas (Winder, 2020). Um exemplo básico de uma função retorno seria a soma de todas as recompensas obtidas a cada iteração no tempo  $t$  dos episódios. Entretanto, se o problema de RL não for dividido em episódios, isso significa que as iterações podem ser infinitas, logo o retorno também pode ser infinito. Desse modo, para calcular a função retorno, utiliza-se um fator de desconto  $\gamma$ , onde  $0 \leq \gamma \leq 1$ , conforme a Equação 2.1. Se o valor de  $\gamma$  é 0, apenas a recompensa atual é considerada no cálculo do retorno. Quanto maior o valor de  $\gamma$ , maior é o peso dado para as recompensas em iterações futuras.

$$G_t = r_{t+1} + \gamma \times r_{t+2} + \gamma^2 \times r_{t+3} + \dots = \sum_{k=0}^{\infty} \gamma^k r_{t+k+1} \quad (2.1)$$

A política a ser desenvolvida pelo agente pode ser determinística ou estocástica. No primeiro caso, não há incerteza no ambiente, ou seja, a partir de um estado, retorna-se uma função com uma ação a ser tomada. No segundo caso, ao invés de obter apenas uma ação, o agente obtém uma distribuição de probabilidades para as ações, caracterizando um ambiente de incerteza. Para aprender políticas estocásticas, muitas vezes é necessário utilizar uma estratégia de *exploit/explore*, como a  $\epsilon$ -*greedy* (Winder, 2020). Quando o agente seleciona apenas as ações já conhecidas que resultaram uma melhor recompensa em episódios ou iterações passados, ele está realizando uma estratégia do tipo *exploit*. Já quando o agente seleciona ações aleatórias com o intuito de explorar o ambiente para obter recompensas que possam levar a uma maior recompensa a longo prazo, ele está aplicando uma estratégia do tipo *explore*.

Dito isto, os algoritmos de RL podem ser divididos em dois tipos principais: *model-based* e *model-free*. Os algoritmos *model-based* utilizam o conhecimento do ambiente para melhorar o aprendizado sem desenvolver políticas impróprias. Já os algoritmos *model-free* podem ser aplicados a qualquer problema, pois eles absorvem todas as informações sobre o ambiente para aprender a melhor política através de tentativa e erro. A forma de atualização da

política pode ser *online*, no qual o agente utiliza dados observados no presente e depois os descarta, ou *offline*, no qual o agente utiliza dados armazenados do passado. Por fim, a forma com que o agente seleciona uma ação através da política pode ser *off-policy* ou *on-policy*. Os agentes *off-policy* utilizam experiências coletadas de políticas anteriores para atualizar a política atual, a exemplo do algoritmo SAC. Enquanto isso, os agentes *on-policy* atualizam a política utilizando apenas as experiências coletadas pela política atual, o que garante mais estabilidade (ChatGPT, 2023a). Um exemplo de algoritmo *on-policy* é o PPO. Apesar disso, a abordagem *off-policy* pode ser mais eficiente, já que o aproveitamento de dados de experiências anteriores permite uma exploração mais eficaz do ambiente (ChatGPT, 2023a).

### 2.1.2 Markov Decision Process

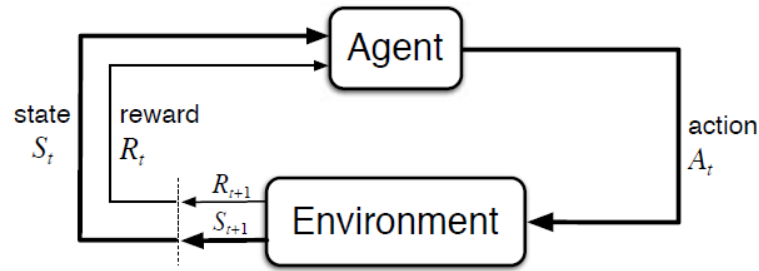
Para descrever a interação entre o agente e o ambiente, utiliza-se um formalismo chamado de *Markov Decision Process* (MDP). Em um problema MDP, o agente observa o estado do ambiente e sugere uma ação, enquanto o ambiente retorna um novo estado e um valor para a recompensa. Segundo (Winder, 2020), o agente nunca recebe toda a informação sobre o ambiente, mas sim um conjunto de observações que representem o estado atual do ambiente. A dinâmica de um problema MDP pode ser definida como a probabilidade de o ambiente sofrer uma transição para um novo estado e gerar uma recompensa dependendo apenas do estado anterior e de uma ação (Sutton; Barto, 2018):

$$p(\acute{s}, r | s, a) = \Pr\{s_t = \acute{s}, r_t = r \mid s_{t-1} = s, a_{t-1} = a\} \quad (2.2)$$

para todo  $\acute{s}$ ,  $s \in \mathcal{S}$ ,  $r \in \mathcal{R}$ , e  $a \in \mathcal{A}$ . Se o estado atual contiver toda informação relevante sobre as iterações passadas entre o agente e o ambiente, ou seja, somente ele precisa ser analisado para calcular a recompensa, é dito que o estado possui a propriedade de *Markov*. A iteração entre o agente e o ambiente é representada na Figura 1.



**Figura 1** - Iteração entre o agente e o ambiente em um problema MDP.



Fonte: (Sutton; Barto, 2018).

Em um MDP, se o agente seguir uma política  $\pi$ , então a probabilidade de ele selecionar uma ação  $a$  em um determinado estado  $s$  é dada por  $\pi(a | s)$  (Sutton; Barto, 2018). Quando o valor esperado do retorno é calculado através da política  $\pi$  a partir de um estado inicial  $s$ , conforme a Equação 2.3, obtêm-se a função estado valor  $v_\pi(s)$ , que estima o quão bom é para o agente estar no presente estado. Já se o cálculo é feito a partir da escolha de uma ação  $a$  em um estado inicial  $s$ , de acordo com a Equação 2.4, é possível obter a função ação valor  $q_\pi(s, a)$ , que estima o quão bom é para o agente realizar uma determinada ação no presente estado. A expressão que relaciona a função estado valor com a função ação valor é apresentada na Equação 2.5.

$$v_\pi(s) = \mathbb{E}_\pi[G_t | s_t = s] = \mathbb{E}_\pi \left[ \sum_{k=0}^{\infty} \gamma^k r_{t+k+1} | s_t = s \right] \quad (2.3)$$

$$q_\pi(s, a) = \mathbb{E}_\pi[G_t | s_t = s, a_t = a] = \mathbb{E}_\pi \left[ \sum_{k=0}^{\infty} \gamma^k r_{t+k+1} | s_t = s, a_t = a \right] \quad (2.4)$$

$$v_\pi(s) = \sum_a \pi(s | a) \times q_\pi(s | a) \quad (2.5)$$

Para relacionar o valor de um estado com os valores dos próximos estados, utilizam-se as equações de Bellman para a função estado valor e ação valor, representadas, respectivamente, pelas Equações 2.6 e 2.7 (Sutton; Barto, 2018). As equações de Bellman simplificam o cálculo do valor de um estado, pois elas decompõem as funções valores em duas

partes, a recompensa imediata e os valores futuros descontados. Com isso, é possível dividir um problema de RL entre subproblemas simples e recursivos e encontrar a solução ótima de cada um deles. A solução ótima global ocorre quando o agente maximiza o valor esperado do retorno a longo prazo, ou seja, quando ele encontra uma política ótima. Logo, ao maximizar as equações de Bellman, é possível encontrar as políticas ótimas para as funções estado valor e ação valor, dadas pelas Equações 2.8 e 2.9, respectivamente (Sutton; Barto, 2018).

$$v_{\pi}(s) = \sum_a \pi(a | s) \sum_{\acute{s}, r} p(\acute{s}, r | s, a) [r + \gamma v_{\pi}(\acute{s})] \quad (2.6)$$

$$q_{\pi}(s, a) = \sum_{\acute{s}} p(\acute{s}, r | s, a) [r(s | a) + \gamma \sum_{\acute{a}} \pi(\acute{a}' | \acute{s}') \times q_{\pi}(\acute{s}, \acute{a})] \quad (2.7)$$

$$v_{*}(s) = \max_a \sum_{\acute{s}, r} p(\acute{s}, r | s, a) [r + \gamma v_{*}(\acute{s})] \quad (2.8)$$

$$q_{*}(s, a) = \sum_{\acute{s}, r} p(\acute{s}, r | s, a) [r + \gamma \max_{\acute{a}} q_{*}(\acute{s}, \acute{a})] \quad (2.9)$$

A solução para as duas últimas equações acima pode ser encontrada de forma exata ou aproximada, dependendo da complexidade do problema e do algoritmo a ser utilizado. Por exemplo, se os espaços de estados e ações são finitos e discretos, é possível encontrar a solução exata para a política ótima com métodos tabulares, como *Monte Carlo*, *Dynamic Programming*, *Q-Learning* e *SARSA*. Entretanto, muitas vezes, o problema exige que os estados e ações sejam contínuos, sendo necessário utilizar funções de aproximação para encontrar as soluções das Equações 2.8 e 2.9. Para isso, utilizam-se as redes neurais, pois elas têm a capacidade de aproximar qualquer função (Anderson, 2022). Assim, surge a classe de algoritmos de *Deep Reinforcement Learning* (DRL) (Sutton; Barto, 2018).

Se o agente precisa consultar uma função valor para encontrar a política ótima, o algoritmo é classificado como *value-based*, a exemplo do algoritmo *Q-Learning*. Já se o agente não precisa consultar uma função valor para mapear a política ótima, o algoritmo é chamado de *policy-based*. Exemplos de algoritmos do tipo *policy-based* são Monte Carlo (REINFORCE) e

REINFORCE com *baseline*. Existem ainda algoritmos que combinam ambas as abordagens *value-based* e *policy-based*, como é o caso de *Actor-Critic* (AC), *Deep Deterministic Policy Gradient* (DDPG), *Trust Region Policy Optimization* (TRPO), *Proximal Policy Optimization* (PPO) e *Soft Actor-Critic* (SAC). Segundo (Winder, 2020), as vantagens dos algoritmos de RL que utilizam abordagens do tipo *policy-based* são:

- Maior efetividade em espaços de ações de alta dimensão;
- Aplicabilidade em ações contínuas;
- Habilidade de aprender políticas estocásticas sem a necessidade de aplicação de estratégias *exploit/explore*;
- Melhor convergência do que os algoritmos *value-based*, já que para estes uma mudança pequena nas funções ação valor pode alterar drasticamente a política do agente.

### 2.1.3 Proximal Policy Optimization

*Proximal Policy Optimization* (PPO) é um algoritmo de RL *model-free*, *on-policy* e *online* que é baseado na implementação dos métodos *Advantage Actor-Critic* (A2C) e *Trust-Region* (TR). A2C é uma das implementações possíveis dos algoritmos da família *Actor-Critic* (AC), que são capazes de aprender ao mesmo tempo a política ótima e a função valor estado através da aproximação por redes neurais. A rede neural que aproxima a política ótima é chamada de *actor*, e aquela que aproxima a função estado valor é chamada de *critic*. Assim, o *actor* é responsável por selecionar ações, que serão avaliadas pelo *critic*. Em seguida, o resultado é comparado com a recompensa obtida pelo ambiente, e os parâmetros das redes neurais são atualizados.

Em qualquer método com uma abordagem *policy-based*, define-se um modelo parametrizado de política com parâmetros  $\theta$ , de acordo com a Equação 2.10 (Winder, 2020), que é equivalente a uma função objetivo. Através do método de gradiente ascendente, conforme a Equação 2.11, é possível encontrar os parâmetros  $\theta$  da política que maximizam o retorno, ou seja, que fornecem a política ótima (Winder, 2020):

$$J(\theta) = \mathbb{E}_{\pi}[G \mid s, a] \quad (2.10)$$

$$\theta \leftarrow \theta + \alpha \times G \times \nabla \ln \pi(a | s, \theta) \quad (2.11)$$

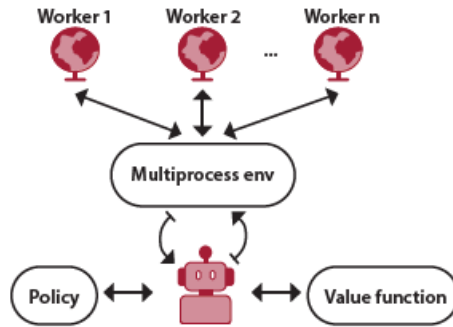
onde a quantidade  $\mathbb{E}_\pi[G \times \nabla \ln \pi(a | s, \theta)]$  representa o gradiente de  $J(\theta)$ , ou  $\nabla J(\theta)$ . Quando a função retorno é substituída pela função vantagem, apresentada na Equação 2.12, obtém-se o algoritmo A2C (Winder, 2020). A função vantagem pode ser considerada como o erro de predição do agente, pois ela é uma medida do quanto um estado é melhor ou pior do que o esperado.

$$\hat{A} = r + v_\pi(s') - v_\pi(s) \quad (2.12)$$

A presença da função vantagem na Equação 2.12 faz com que a política se direcione para a produção das ações de melhor performance (Winder, 2020). Ou seja, se a política atualizada for pior que a anterior, a função vantagem será negativa e, conseqüentemente, o gradiente também será negativo, direcionando a política para longe dessa região. Entretanto, como a função vantagem é uma aproximação, já que a função estado valor é aproximada por uma rede neural no algoritmo A2C, uma taxa de atualização da política muito rápida pode diminuir drasticamente a performance da mesma. Para evitar esse problema, os métodos *Trust Region* (TR) especificam uma região em que as atualizações da política podem ser confiáveis. O algoritmo TRPO utiliza a divergência KL, uma medida do quanto duas distribuições são diferentes, para limitar o quanto a política pode ser atualizada de acordo com a mudança da política antiga  $\pi_{\theta_{old}}$  para a política nova  $\pi_\theta$  (Winder, 2020).

Entretanto, segundo (Schulman *et al.*, 2017), o algoritmo TRPO é muito complicado e não é compatível com arquiteturas de ruído e compartilhamento de parâmetros entre a política e a função valor. O algoritmo PPO utiliza a mesma ideia de TRPO, mas com uma implementação simplificada em primeira ordem. Para isso, Schulman *et al.*, 2017 introduziu a técnica de *clipping* na função objetivo, que limita a atualização da política em uma quantidade pequena. Isso previne a política de colapsar em direção de ações ruins. Basicamente, o algoritmo PPO utiliza a mesma arquitetura que o algoritmo A2C, onde definem-se múltiplos ambientes ou *workers* com apenas um agente (Morales, 2020), conforme ilustra a Figura 2.

**Figura 2** - Representação esquemática da arquitetura do algoritmo PPO.



Fonte: (MORALES, 2020).

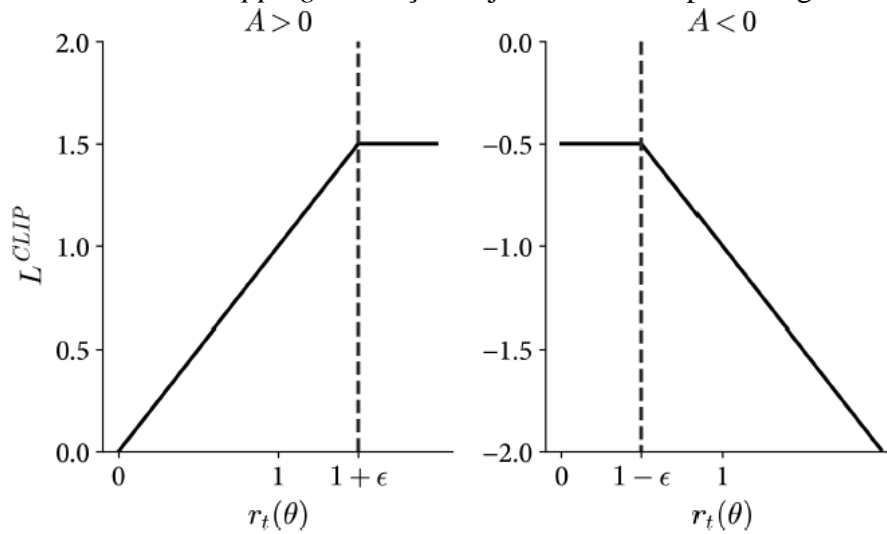
Cada um dos *workers* será responsável por coletar um número específico de episódios ou experiências, que é definido pelo parâmetro *batch size*. Os dados são armazenados em um vetor chamado de *replay buffer*. Em seguida, os índices das amostras do *replay buffer* são embaralhados para formar amostras com o tamanho definido pelo parâmetro *mini-batch size*. Então, durante cada *step*, o agente utiliza as amostras de cada *worker* para atualizar a sua política através do método de subida de gradiente. Isso causa um efeito estocástico em cada *mini-batch*, pois apesar de as amostras não serem sempre as mesmas, no término da otimização, a maioria delas terá sido reutilizada (Morales, 2020).

A função objetivo do PPO para o método de subida de gradiente é composta por três partes: a função objetivo do *actor*, a função objetivo do *critic* e um termo de entropia. Na linguagem de redes neurais, a função objetivo também é chamada de função perda. De acordo com (Schulman *et al.*, 2017), a função objetivo para o *actor* é definida como:

$$L^{CLIP}(\theta) = \hat{\mathbb{E}}_t[\min(r_t(\theta) \times \hat{A}_t, \text{clip}(r_t(\theta), 1 - \epsilon, 1 + \epsilon) \times \hat{A}_t)] \quad (2.13)$$

onde  $r_t(\theta)$  é o raio de probabilidade, dado por  $r_t(\theta) = \pi_\theta(a_t | s_t) / \pi_{\theta_{old}}(a_t | s_t)$ , e  $\epsilon$  é um valor entre 0 e 1 (Winder, 2020). A Figura 3 ilustra o efeito do *clipping* na Equação 2.13 de acordo com o sinal da função vantagem.

**Figura 3** - Efeito do *clipping* na função objetivo do *actor* para o algoritmo PPO.



Fonte: (WINDER, 2020).

Quando o valor do estado obtido for melhor que o valor estimado, a função vantagem será positiva, indicando que a política foi melhorada, logo seus parâmetros podem ser atualizados em direção a essas ações. Porém, se  $r_t(\theta)$  crescer a ponto de ultrapassar o valor  $1 + \epsilon$ , os parâmetros param de ser atualizados naquela direção para garantir que o agente não construa uma política de baixa performance. Já se o valor do estado obtido for pior que o valor estimado, a função vantagem será negativa, indicando que a política piorou. Nesse caso, se  $r_t(\theta)$  for grande, domina o valor  $r_t(\theta) \times \hat{A}_t$  da Equação 2.13, fazendo com que o agente volte a selecionar ações mais promissoras. Mas, se  $r_t(\theta)$  for pequeno, o *clipping* domina, e os parâmetros da política continuam a ser atualizados na mesma taxa, pois houve uma diferença significativa entre as políticas antiga e nova (Winder, 2020).

Schulman *et al.*, 2017 também descreve uma alternativa para o *clipping* da função objetivo do *actor*, que consiste em utilizar a divergência KL para penalizar a função objetivo caso a política se direcione para uma direção de baixa performance, conforme a Equação 2.14:

$$L^{KL PEN}(\theta) = \hat{\mathbb{E}}_t[r_t(\theta) \times \hat{A}_t - \beta \times KL[\pi_{\theta_{old}}(a_t | s_t), \pi_{\theta}(a_t | s_t)]] \quad (2.14)$$

onde  $\beta$  é o coeficiente de penalidade e é atualizado conforme o valor da divergência KL entre as duas políticas. As implementações do PPO disponíveis no *Python* variam, sendo que algumas utilizam o *clipping*, e outras a divergência KL na função objetivo do *actor*. As bibliotecas

*PyTorch*, *Tensorflow* e *RLLib* suportam as duas implementações, enquanto as bibliotecas *Keras*, *Stable-Baselines* e *OpenAI Baselines* suportam apenas a primeira.

De forma análoga, segundo (Morales, 2020), também é possível realizar um *clipping* na função estado valor, que será utilizada para construir a função objetivo para o *critic*, dada pela Equação 2.15 (Schulman *et al.*, 2017):

$$L_t^{VF}(\theta) = (V(s_t, \theta) - V_t^{targ})^2 \quad (2.15)$$

onde  $V(s_t, \theta)$  é a função estado valor atual e  $V_t^{targ}$  é a função estado valor com o *clipping*. Finalmente, define-se a função objetivo total para o PPO como uma combinação das funções objetivas do *actor* e do *critic* e um termo de entropia  $S[\pi_\theta](s_t)$ , que força o agente a explorar e desencoraja políticas determinísticas. Quando o *clipping* é realizado na função perda do *actor*, a função objetivo final torna-se a Equação 2.16. Já quando a divergência KL for utilizada na função perda do *actor*, obtém-se a Equação 2.17 para a função objetivo total do PPO. Em ambas as equações abaixo,  $c_1$  e  $c_2$  são coeficientes.

$$L_t^{CLIP+VF+S}(\theta) = \widehat{\mathbb{E}}_t[L_t^{CLIP}(\theta) - c_1 \times L_t^{VF}(\theta) + c_2 \times S[\pi_\theta](s_t)] \quad (2.16)$$

$$L_t^{KLPE+VF+S}(\theta) = \widehat{\mathbb{E}}_t[L_t^{KLPE}(\theta) - c_1 \times L_t^{VF}(\theta) + c_2 \times S[\pi_\theta](s_t)] \quad (2.17)$$

#### 2.1.4 Soft Actor-Critic

Os algoritmos de RL do tipo *on-policy*, como TRPO e PPO, exigem que novas amostras sejam coletadas para atualizar o gradiente. Porém, de acordo com (Haarnoja *et al.*, 2018), quanto mais complexa é a tarefa que o agente deve aprender, maior será o número de amostras necessárias para atualizar o gradiente, o que pode tornar o processo lento e ineficiente. Entretanto, ainda segundo (Haarnoja *et al.*, 2018), a combinação de aproximações de funções não-lineares e multidimensionais com redes neurais e algoritmos *off-policy* pode tornar o modelo de RL instável e dificultar sua convergência. Para resolver esses problemas, o algoritmo *Soft Actor-Critic* (SAC) introduz o conceito de entropia na função objetivo do gradiente da política (Winder, 2020).

Dessa forma, o agente é capaz de aprender políticas estocásticas e comportamentos probabilísticos. Para uma variável estocástica  $X = \{x_0, x_1, \dots, x_{n-1}\}$ , a entropia pode ser definida como:

$$S(X) = - \sum_{x \in X} p(x) \times \log_b p(x) \quad (2.18)$$

onde  $b$  é a base utilizada no logaritmo e  $p(x)$  é a distribuição de probabilidades. Em problemas de RL, ao maximizar-se a entropia, o agente é forçado a visitar todos os estados e ações. Portanto, ao invés de o agente do algoritmo SAC aprender uma política ótima que maximize a recompensa, ele é capaz de aprender todas as políticas (Winder, 2020), ou todas as maneiras de realizar uma tarefa. Assim como o PPO, o algoritmo SAC também se baseia na implementação dos algoritmos AC. Porém, ao contrário de muitos algoritmos AC, em que a entropia é utilizada como um regulador para atualizar o *actor*, na implementação SAC, a entropia é maximizada (Haarnoja *et. al*, 2018).

Para mais referências sobre o algoritmo SAC deve-se consultar (Haarnoja *et. al*, 2018) e (Morales, 2020). É importante ressaltar que as equações do SAC não serão detalhadas neste trabalho, pois como ele apresentou uma baixa velocidade de aprendizado dada a sua natureza probabilística, não necessária no estudo de caso de natureza puramente determinística, optou-se por focar somente no algoritmo PPO. Se verificou se essa premissa era válida através da simulação de banho para temperatura ambiente de 25°C e de fato, como esperado, os resultados foram inferiores ao PPO em um tempo de treinamento similar.

## 2.2 DESIGN DE SISTEMAS AUTÔNOMOS

Para realizar o *design* de um sistema autônomo, serão utilizados os conceitos de *Machine Teaching* e *brain design* descritos em (Anderson, 2022). De acordo com (Anderson, 2022), *Machine Teaching* é um ramo de pesquisa focado em incorporar conhecimento humano, experiências, critérios e objetivos para abstrair a complexidade de um problema a ser resolvido por um sistema autônomo, facilitando o aprendizado de tarefas. A incorporação de técnicas de *Machine Teaching* em sistemas autônomos é importante porque quando o sistema é muito



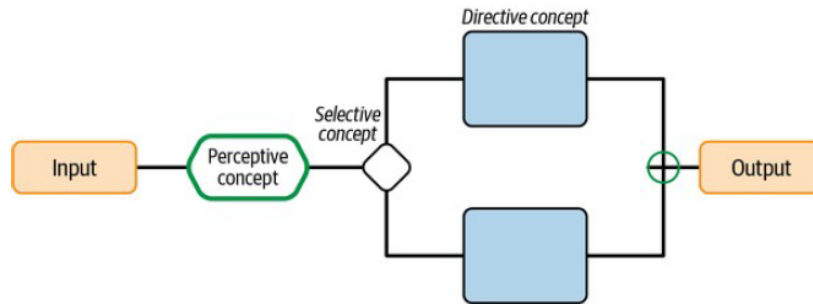
complexo e contém múltiplos objetivos ou condições iniciais, o aprendizado do agente pode ser ineficiente ou consumir muito tempo.

Por exemplo, ao aplicar RL em um sistema de chuveiro, o usuário pode desejar maximizar a qualidade do banho e ao mesmo tempo minimizar o custo da energia elétrica para diversas bandeiras em cenários com diferentes temperaturas ambientes. Entretanto, ao utilizar somente um agente para o treinamento, a política a ser desenvolvida por este agente em um cenário de dia frio e energia elétrica cara será diferente da política desenvolvida pelo mesmo agente em um cenário de dia quente e energia elétrica barata. Desse modo, o agente deveria ser capaz de aprender múltiplas políticas ótimas, o que pode não ser possível em problemas complexos onde somente um agente foi treinado.

Assim, define-se o *brain*, um componente que combina técnicas de RL, simulações e *Machine Teaching* e que é responsável pelo controle inteligente e pela otimização do sistema autônomo. O *brain* pode ser composto por um ou mais *concepts*. Segundo (Anderson, 2022), um *concept* é “uma noção ou uma ideia que compreende uma unidade componível de conhecimento”. Em outras palavras, é uma habilidade que deve ser aprendida. Quando existe apenas um *concept*, o *brain* pode ser entendido como um modelo *single concept*. Já quando há múltiplos *concepts*, o *brain* é definido como um modelo *multi concept*. Anderson, 2022 divide os *concepts* em três categorias: *perceptive*, *directive* e *selective*.

O primeiro *concept* é responsável por processar informações coletadas por sensores que serão utilizadas pelo *brain*. Já o segundo toma as decisões do sistema e age, sendo análogo ao agente ou operador. Por fim, o terceiro é um tipo especializado de *directive concept*, e pode ser entendido como um supervisor do *brain* ou o supervisor dos operadores. Ou seja, baseado no cenário atual do sistema, o *selective concept* é responsável por selecionar, entre vários *directive concepts*, qual deles deverá tomar as decisões sobre o sistema. O *selective concept* ainda pode ser dividido em duas categorias: *programmed* ou *learned* (Anderson, 2022). O primeiro é utilizado quando o engenheiro sabe em que condições cada *directive concept* deve ser acionado, construindo uma lógica para isso. Já o segundo é utilizado quando é difícil saber quando cada *directive concept* deve ser acionado. Para isso, treina-se um modelo específico para o *selective concept learned*. A Figura 4 ilustra os símbolos utilizados para realizar um *brain design* através de *concepts*.

**Figura 4** - Representação esquemática dos símbolos utilizados no *brain design*.



Fonte: (Anderson, 2022).

Além da definição dos *concepts*, ou de quais habilidades deverão ser aprendidas pelo agente, também deve-se listar os objetivos do projeto para realizar o *design* do sistema autônomo. A Tabela 1 apresenta os seis objetivos básicos utilizados durante o *brain design* para ensinar um *concept* a um sistema autônomo. Por exemplo, possíveis objetivos para um sistema autônomo de chuveiro poderiam ser: atingir a temperatura de 35°C quando a energia elétrica está cara, atingir a temperatura de 37°C quando a energia elétrica está barata e evitar temperaturas abaixo de 30°C.

**Tabela 1** - Objetivos básicos utilizados para o *brain design*.

<b>Objetivo</b>	<b>Descrição</b>
Maximizar	Aumentar um parâmetro o máximo possível
Minimizar	Diminuir um parâmetro o máximo possível
Atingir	Atingir um estado particular do processo
Levar	Levar um parâmetro para uma certa faixa de valores
Evitar	Evitar uma situação ou evitar que parâmetros assumam uma certa faixa de valores
Padronizar	Manter múltiplas variáveis de estado com valores próximos umas às outras

Fonte: adaptado de (Anderson, 2022).

### 3 METODOLOGIA

Existem metodologias já consolidadas para projetos de Ciência de Dados, sendo que uma das mais conhecidas é a CRISP-DM (do inglês, *Cross Industry Standard Process for Data Mining*), que divide um projeto em seis etapas principais: entendimento do negócio, exploração dos dados, preparação dos dados, modelagem, avaliação e *deployment*. Apesar disso, para este trabalho, foi elaborada a seguinte metodologia em forma de passo a passo com base nos conceitos explorados no Capítulo 2, dando mais ênfase ao processo de solução de um problema de RL. Esta metodologia considera que o agente de RL será treinado em um ambiente construído através de uma simulação computacional.

- **Passo 1:** Definir em qual linguagem de programação o problema será resolvido, considerando a facilidade de encontrar recursos e bibliotecas na área de RL e a eficiência computacional da linguagem.
- **Passo 2:** Desenvolver um modelo computacional utilizando a linguagem de programação definida no passo anterior. Este modelo deverá ser capaz de simular o comportamento de um sistema real, portanto, ele deve apresentar variáveis bem definidas. Além disso, ele servirá de base para a construção do ambiente de RL.
- **Passo 3:** Definir os *objetivos* para o agente de RL, e através deles, desenvolver uma expressão para a *recompensa*, que será a função objetivo do problema. Esta recompensa deverá conter as principais variáveis que influenciam no comportamento do sistema e no cumprimento dos objetivos.
- **Passo 4:** Escolher as variáveis para os espaços de estados e de ações, definindo se elas serão discretas ou contínuas e qual será a faixa de valores para cada variável.
- **Passo 5:** Escolher um ou mais algoritmos de RL para o treinamento dos modelos. Esta escolha deve levar em conta a complexidade do problema a ser resolvido e as variáveis presentes nos espaços de estados e de ações. Por exemplo, se todas as ações são contínuas, algoritmos como Monte Carlo e *Deep Q-Learning* (DQN) não podem ser utilizados, pois eles suportam somente ações discretas.

- **Passo 6:** Escolher uma biblioteca disponível na linguagem de programação do projeto para realizar a implementação dos algoritmos de RL. Preferencialmente, a biblioteca deve conter implementações que suportem os espaços de estados e de ações do problema. Caso contrário, o usuário poderá modificar os espaços ou customizar as implementações.
- **Passo 7:** Realizar o *brain design* para o sistema, considerando os objetivos, a recompensa e os espaços de estados e de ações do problema. Definir as seguintes variáveis: o tempo de duração de cada episódio; a quantidade de ações durante o episódio e o intervalo de tempo entre elas; e o número de *steps* ou de episódios para os algoritmos.
- **Passo 8:** Construir um ambiente para o agente de RL, utilizando uma biblioteca disponível na linguagem de programação escolhida no passo 1 que suporte os espaços de estados e de ações definidos para o problema e que seja compatível com as implementações dos algoritmos de RL.
- **Passo 9:** Definir as configurações para os parâmetros de cada algoritmo escolhido e realizar o treinamento para verificar a performance dos modelos. Caso a performance não seja satisfatória, modificar o número de episódios ou os parâmetros dos algoritmos e repetir os experimentos.
- **Passo 10:** Após o treinamento, validar os modelos desenvolvidos em cenários iniciais desconhecidos pelo agente de RL. Caso os modelos sejam aprovados no processo de validação, o *deployment* pode ser feito na vida real.

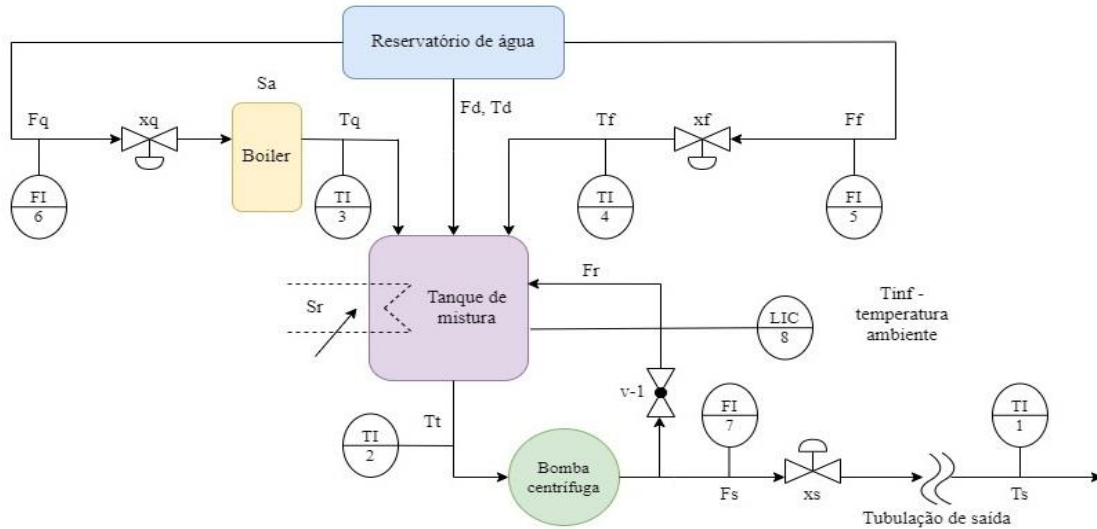
## 4 ESTUDO DE CASO

### 4.1 MODELO COMPUTACIONAL DE CHUVEIRO

#### 4.1.1 Apresentação da modelagem da planta de chuveiro

O modelo computacional de planta de chuveiro a ser utilizado neste trabalho é baseado no sistema apresentado durante as aulas da disciplina (ENG07044) Controle de Processos Industriais, ministradas pelo Prof. Dr. Jorge Otávio Trierweiler do Departamento de Engenharia Química da UFRGS. A explicação do sistema está disponível em (Trierweiler, 2021a) e (Trierweiler, 2021b). A planta de chuveiro consiste em um tanque de aquecimento cilíndrico, com área da base  $A$ , que mistura uma corrente fria e uma corrente quente, oriunda de um aquecimento a gás por um *boiler* com um controlador liga-desliga. O modelo do *boiler* é ideal, ou seja, não há perda de calor para o meio externo.

Além da corrente aquecida, uma resistência elétrica localizada no interior do tanque fornece energia para a mistura. No tanque, também é possível inserir uma corrente de distúrbio cuja temperatura e vazão não podem ser controladas. A alteração da temperatura ambiente também é considerada uma forma de distúrbio. Como a tubulação de descarga do tanque possui um comprimento significativo, a temperatura de saída da planta é afetada tanto pelo elevado tempo morto quanto pela perda de calor dependendo da temperatura ambiente. A Figura 5 contém um esquema da planta de chuveiro, e a Tabela 2 apresenta a descrição das variáveis do sistema presentes na Figura 5.

**Figura 5** - Representação esquemática do modelo de planta de chuveiro.

Fonte: adaptado de (Trierweiler, 2021a) e (Trierweiler, 2021b).

**Tabela 2** - Descrição das variáveis físicas do modelo computacional de chuveiro.

Variável	Descrição
$F_d$	Vazão da corrente de distúrbio
$T_d$	Temperatura da corrente de distúrbio
$T_{inf}$	Temperatura externa ou ambiente
$x_f$	Abertura da válvula de entrada da corrente fria
$x_q$	Abertura da válvula de entrada da corrente quente
$T_f$	Temperatura da corrente fria
$T_q$	Temperatura da corrente quente fornecida pelo aquecedor
$h$	Nível do tanque de mistura
$T_t$	Temperatura do fluido do tanque
$T_s$	Temperatura no final da tubulação
$S_a$	Seletor da fração de aquecimento pelo <i>boiler</i>
$S_r$	Seletor da fração de aquecimento pela resistência elétrica
$F_s$	Vazão de descarga
$F_r$	Vazão de reciclo da bomba para o tanque
$F_q$	Vazão de entrada do fluido aquecido
$F_f$	Vazão de entrada do fluido frio

Fonte: adaptado de (Trierweiler, 2021a) e (Trierweiler, 2021b).

A partir da equação da válvula se pode determinar a vazão de líquido em função da diferença de pressão, sendo descrita através da seguinte expressão (Luyben, 1996):

$$F = C_v \times f(x) \times \sqrt{\Delta P_v / G_f} \quad (4.1)$$

onde  $F$  é a vazão do líquido em  $gpm$ ,  $f(x)$  é uma função que relaciona a abertura da válvula com o tipo de obturador (linear, parabólico, igual percentagem, etc.),  $C_v$  é o coeficiente de vazão em  $gpm/\sqrt{psi}$ , ou o tamanho característico da válvula,  $\Delta P_v$  é a perda de carga na válvula em  $psi$ , e  $G_f$  é a razão entre a massa específica do líquido e a massa específica da água, ou gravidade específica.

Desse modo, através da Equação 4.1, é possível determinar as vazões das correntes frias e quentes. Para a vazão da válvula de saída, adota-se um procedimento similar, com exceção de que o valor para  $\Delta P_v$  será igual a diferença da perda de carga na válvula  $\Delta P_B$  fornecida pela bomba com a perda de carga na linha em função da vazão:

$$\Delta P_v = \Delta P_{B,MAX} \sqrt{1 - \left(\frac{F_s + F_r}{F_{MAX}}\right)^2} - \Delta P_n \left(\frac{F_s}{x_n}\right) \quad (4.2)$$

onde  $\Delta P_{B,MAX}$  é a perda máxima de carga na bomba,  $F_{MAX}$  é a vazão máxima da bomba,  $\Delta P_n$  é a perda de carga da vazão nominal, e  $x_n$  é vazão nominal com água.

A taxa de variação da energia interna, cinética e potencial de um sistema pode ser definida através da diferença entre o calor adicionado ao sistema por convecção, difusão ou reação com: o fluxo da energia interna, cinética e potencial que entra no sistema por convecção ou difusão; o fluxo da energia interna, cinética e potencial que sai do sistema por convecção ou difusão; e o trabalho feito pelo sistema na vizinhança. Já a taxa de variação da massa dentro do sistema é definida como a diferença entre o fluxo de massa que entra no sistema com o fluxo de massa que sai (Luyben, 1996).

Assim, definindo  $V_a$  como o volume de água dentro do aquecedor,  $V_t$  como o volume da tubulação,  $\rho$  como a massa específica da água,  $c$  como o calor específico da água,  $v$  como o volume do tanque de mistura, e  $\dot{Q}_a$  e  $\dot{Q}_r$  como as taxas de calor transferidas para a água no *boiler* e na tubulação, é possível escrever as Equações 4.3, 4.4, 4.5 e 4.6 para o balanço de energia no sistema. A Equação 4.6 representa o balanço de energia da tubulação de saída, que foi simulado como sendo uma série de tanques em série, tendo cada um uma perda de calor

para o meio ambiente. Em *Python*, o sistema é resolvido através do método *solve\_ivp* do módulo *integrate* da biblioteca *scipy*.

$$V_a \frac{dT_q}{dt} = F_q(T_f - T_q) + S_a \frac{\dot{Q}_a}{\rho c} \quad (4.3)$$

$$A \frac{dh}{dt} = F_f + F_q + F_d - F_s \quad (4.4)$$

$$\overbrace{(Ah)}^v \frac{dT_t}{dt} = F_f(T_f - T_t) + F_q(T_q - T_t) + F_d(T_d - T_t) + S_r \frac{\dot{Q}_r}{\rho c} \quad (4.5)$$

$$\frac{dT_s}{dt} = \frac{F_s}{V_t}(T_t - T_s) - \frac{\left( (T_t - T_{inf})(T_s - T_{inf}) \left( \frac{1}{2}(T_t - T_{inf}) + \frac{1}{2}(T_s - T_{inf}) \right) \right)^{\frac{1}{3}}}{\rho V_t c} \quad (4.6)$$

#### 4.1.2 Configurações do modelo de chuveiro

Através das equações descritas na seção anterior, três configurações de chuveiro foram implementadas em *Python*, a linguagem de programação escolhida para este projeto:

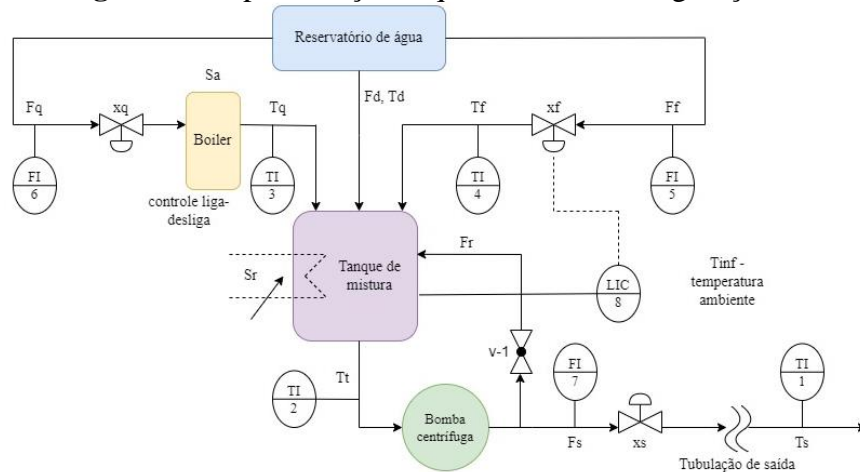
- *Configuração A*: Modelo de chuveiro com malha de inventário para o nível do tanque com controlador *Proportional Integral* (PI) e controle liga-desliga para a temperatura do *boiler*. Esta configuração, representada pela Figura 6, será utilizada para verificar o desempenho de RL como ferramenta na camada supervisória.
- *Configuração B*: Modelo de chuveiro com malha de inventário para o nível do tanque com controlador PI, controle liga-desliga para a temperatura do *boiler*, malha cascata para a temperatura de saída com controlador PID e com estratégia de *split-range* na fração da resistência elétrica. Esta configuração, representada na Figura 7, será utilizada para verificar o desempenho de RL



como ferramenta da otimização em tempo real. Nessa estratégia de *split range*, a potência elétrica é somente usada quando a válvula quente está totalmente aberta.

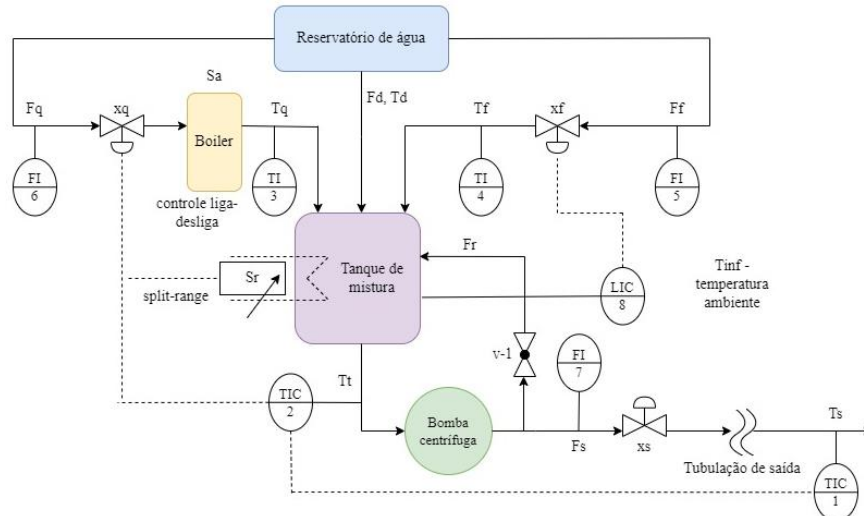
- **Configuração C:** Modelo de chuveiro com malha de inventário para o nível do tanque com controlador PI, controle liga-desliga para a temperatura do *boiler* e malha cascata para a temperatura de saída com controlador PID. A diferença para a configuração anterior é que nesta o agente poderá regular diretamente a fração da resistência elétrica. A configuração C é representada na Figura 8.

**Figura 6** - Representação esquemática da configuração A.

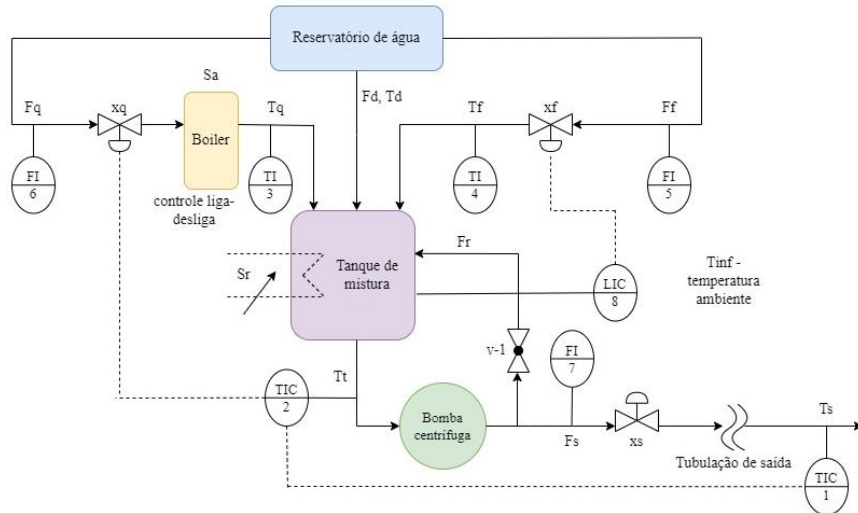


Fonte: elaborado pelo autor.

**Figura 7** - Representação esquemática da configuração B.



Fonte: elaborado pelo autor.

**Figura 8** - Representação esquemática da configuração C.

Fonte: elaborado pelo autor.

A implementação dos controladores PI e PID foi baseada na teoria de discretização do PID (Aström; Hägglund, 1995). Em ambas as três configurações, foi considerada uma vazão nominal de 5 *gpm* com água e uma perda de carga na tubulação de saída de 5 *psi*. Tanto o tanque quanto a saída da tubulação estão à mesma pressão. Já a alimentação do tanque está sujeita a uma pressão estática de 20 *psi* devido a coluna d'água. A perda de carga para vazão nominal de 5 *gpm* da caixa d'água até o ponto de bifurcação entre a corrente quente e fria é de 1 *psi*. A perda de carga no aquecedor de passagem para uma vazão nominal de 3 *gpm* é de 2 *psi*. A vazão máxima de cada planta deverá ser de 10 *gpm*. A bomba centrífuga disponível para operar a planta tem  $\Delta P_{B,max} = 30 \text{ psi}$  e  $F_{MAX} = 20 \text{ gpm}$ .

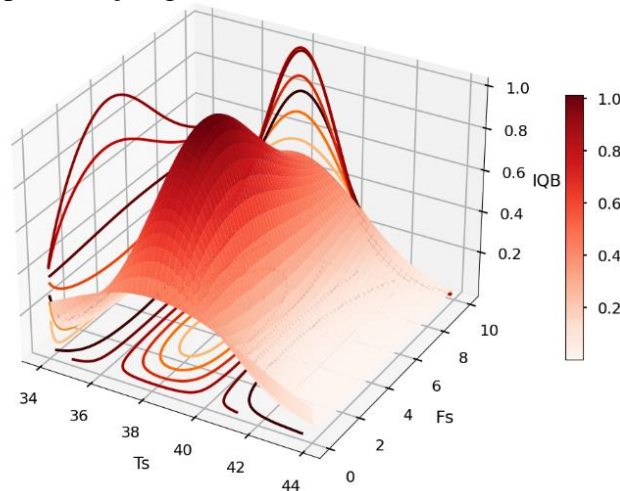
#### 4.2 ÍNDICE DE QUALIDADE DO BANHO

Para medir a qualidade do banho levando em conta a satisfação do consumidor, Giaquinto, 2017 desenvolveu uma relação empírica determinada Índice de Qualidade de Banho (IQB), conforme a Equação 4.7. O IQB será utilizado para construir as funções de recompensa para os agentes de RL.

$$IQB = \frac{1}{e} \times e^{\left[1 - \left(\frac{T_s - 38 + 0.02 \times F_s}{2}\right)^2\right]} \times \left[0.506 + \log\left(\log\left(\frac{10000 \sqrt{F_s}}{10 + F_s + 0.004 \times F_s^4}\right)\right)\right]^{20} \quad (4.7)$$

O valor do IQB é normalizado entre 0 e 1, e o valor máximo de 1 ocorre aproximadamente quando  $F_s = 3,8 \text{ L/min}$  e  $T_s = 37,5^\circ\text{C}$ . O IQB é penalizado quando a temperatura se mantém e a vazão aumenta ou diminui. O indicador também é afetado negativamente se a temperatura aumentar ou diminuir demasiadamente. A Figura 9 é uma representação gráfica do IQB para valores de  $F_s$  entre 0,1 a 10  $\text{L/min}$  e  $T_s$  entre  $32^\circ\text{C}$  a  $44^\circ\text{C}$ .

**Figura 9** - Representação gráfica do Índice de Qualidade do Banho (IQB).



Fonte: elaborado pelo autor.

### 4.3 CÁLCULO DO CUSTO DO BANHO

Para os modelos de RL que foram treinados com o objetivo de reduzir os custos do banho, com ênfase no custo da energia elétrica, foi necessário desenvolver expressões para calcular os custos elétrico, de gás e de água do banho durante cada iteração dos episódios. Em relação ao custo elétrico, multiplica-se a potência elétrica do chuveiro, definida como  $5500 \text{ W}$ , com a fração da resistência elétrica utilizada, a tarifa da energia elétrica em reais e o tempo de utilização da resistência elétrica em horas. A quantidade de fração da resistência elétrica utilizada é encontrada através do método de integração *trapz* da biblioteca *numpy* do Python. Para o custo da água, a vazão de saída em *litros/minutos* pode ser encontrada diretamente

através das equações do modelo de chuveiro. Assim, o custo da água será igual a vazão de saída em metros cúbicos multiplicada pelo tempo de cada ação e o custo da água em reais por metros cúbicos, definido como R\$4,00.

Por fim, para o custo do gás, primeiramente, calcula-se a quantidade utilizada da fração do aquecimento do *boiler*, que também é encontrada por integração com o método *trapz* do *numpy*. Para obter a potência útil do aquecedor, multiplica-se a potência do aquecedor, definida como  $29000 \text{ kcal/h}$ , pelo rendimento. Como o modelo de chuveiro não considera perda de calor no *boiler*, o rendimento é igual a 1. Em seguida, calculam-se quantas calorias foram fornecidas no banho pelo aquecedor. Depois, divide-se esse número pelo poder calorífico do gás, que é de aproximadamente  $11750 \text{ kcal/kg}$  para o gás GLP (Supergasbras, 2021), para encontrar a quantidade de gás utilizada no banho. Então, o custo de gás total é dado pela multiplicação da quantidade de gás que foi gasta com o valor em reais do *kg* do gás, definido como R\$3,00.

## 5 RESULTADOS E DISCUSSÃO

### 5.1 OTIMIZAÇÃO DE QUALIDADE

Inicialmente, as três configurações do modelo computacional de chuveiro descritas na Seção 4.1.1 foram implementadas utilizando *Python*, a linguagem de programação escolhida para este trabalho. Todos os modelos foram treinados em um computador com processador Intel(R) Core(TM) i7-10700, CPU @ 2,90GHz e memória RAM de 16GB. Na primeira fase do estudo, pretendia-se verificar qual estratégia para aplicação de RL no modelo de chuveiro resultaria no banho de maior qualidade. Para isso, considerou-se o sistema sem distúrbios, ou seja, a corrente e a vazão de distúrbio foram zeradas, e a temperatura ambiente assumiu valores fixos. Como o objetivo era maximizar a qualidade do banho, a recompensa foi definida como o IQB, descrito na Seção 4.2, no fim de cada iteração.

As variáveis que integraram o espaço de estados na configuração A foram: temperatura de saída ( $T_s$ ), temperatura do *boiler* ( $T_q$ ), temperatura do tanque ( $T_t$ ), nível do tanque ( $h$ ), vazão de saída ( $F_s$ ), abertura da válvula de corrente fria ( $x_f$ ), IQB e temperatura ambiente ( $T_{inf}$ ). Para as configurações B e C, adicionou-se a abertura da válvula de corrente quente ( $x_q$ ) ao espaço de estados especificado para a configuração A. Todas as variáveis dos espaços de estados são contínuas. Já o espaço de ações para cada configuração é apresentado na Tabela 3. Com exceção da ação de *split-range*, que é discreta, todas as outras ações são contínuas.

**Tabela 3** - Espaço de ações para as configurações A, B e C.

<b>Configuração A</b>	<b>Configuração B</b>	<b>Configuração C</b>
Abertura da válvula de corrente quente ( $x_q$ ): 0,01 a 0,99	<i>Setpoint</i> da temperatura de saída ( $SPT_s$ ): 30°C a 40°C	<i>Setpoint</i> da temperatura de saída ( $SPT_s$ ): 30°C a 40°C
<i>Setpoint</i> da temperatura do <i>boiler</i> ( $SPT_q$ ): 30°C a 70°C	<i>Setpoint</i> da temperatura do <i>boiler</i> ( $SPT_q$ ): 30°C a 70°C	<i>Setpoint</i> da temperatura do <i>boiler</i> ( $SPT_q$ ): 30°C a 70°C
Abertura da válvula de saída ( $x_s$ ): 0,01 a 0,99	Abertura da válvula de saída ( $x_s$ ): 0,01 a 0,99	Abertura da válvula de saída ( $x_s$ ): 0,01 a 0,99
Fração da resistência elétrica ( $S_r$ ): 0 a 1	Utilização do <i>split-range</i> : 0 ou 1	Fração da resistência elétrica ( $S_r$ ): 0 a 1

Fonte: elaborado pelo autor.

Em seguida, escolheram-se os algoritmos PPO e SAC para o treinamento. O primeiro deles foi escolhido pelas seguintes razões: é simples de implementar; a técnica de *clipping* nas funções objetivas aumenta a estabilidade do algoritmo; é capaz de convergir rapidamente; possui eficiência computacional; e é adequado para ambientes contínuos (ChatGPT, 2023b). Já o segundo foi escolhido pois possui excelente desempenho em ambientes contínuos, capacidade de exploração eficiente, e estabilidade de treinamento (ChatGPT, 2023c). Para as implementações dos algoritmos, foi utilizada a biblioteca *RLLib*, pois ela suporta uma implementação para o PPO que inclui um espaço de ações com variáveis contínuas e discretas ao mesmo tempo e uma implementação para o SAC com um espaço de ações com variáveis contínuas (Ray RLLib, 2023a).

Para verificar a estabilidade e convergência dos algoritmos selecionados, cada configuração foi treinada com um *brain* com apenas um *directive concept*, que aprenderá a controlar o banho para a temperatura ambiente de 25°C. Foram definidos 100 e 1000 *steps* para o PPO e o SAC, respectivamente. Como o SAC é um algoritmo que aprende todas as estratégias possíveis, espera-se que ele precise de mais *steps* para convergir. A duração de um episódio foi de 14 minutos devido ao tempo morto integrado no sistema, que faz com que a temperatura de saída demore mais para atingir valores ótimos. O tempo de cada iteração foi de 2 minutos, ou seja, a cada 2 minutos o agente executa uma ação, totalizando 7 ações por episódio. O episódio termina quando a duração do banho atingir 14 minutos ou quando o nível do tanque ultrapassar o valor máximo de 100.

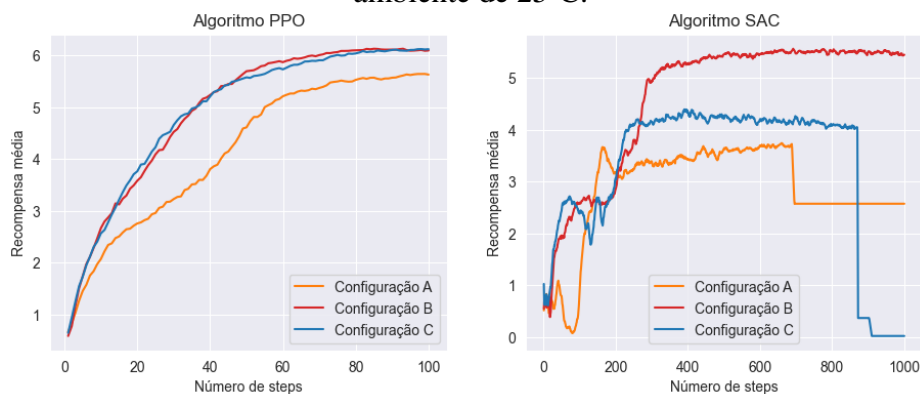
Para construir o ambiente e integrar o modelo computacional com o *brain design*, foi utilizada a biblioteca *Gymnasium*, que possibilita a representação geral de problemas de RL (Farama Foundation, 2023). Para ambos os algoritmos, utilizou-se o espaço *Box* do *Gymnasium* para representar os estados. Já para as ações, utilizou-se o espaço *Tuple* para o PPO e o espaço *Box* para o SAC. Apesar de o espaço *Tuple* suportar ações discretas e contínuas ao mesmo tempo, foi necessário utilizar o espaço *Box* para as ações do SAC pois a sua implementação na *RLLib* não suporta espaços de ações mistos. Logo, no caso da configuração B, a ação do *split-range* foi definida como contínua e depois arredondada para 0 ou 1.

Em relação aos parâmetros dos algoritmos, foram mantidos os valores padrão especificados em (Ray RLLib, 2023b) e (Ray RLLib, 2023c). A implementação padrão do PPO na *RLLib* aplica a técnica de *clipping* para a função do *actor* com coeficiente  $\epsilon = 0,3$  e para a

função estado valor com coeficiente de  $\epsilon = 10$ . Além disso, utilizam-se 2 *workers* com *batch size* de 4000 e *mini batch size* de 128. O termo de entropia na função objetivo final  $L_t^{CLIP+VF+S}(\theta)$  é zerado. A rede neural para ambos *actor* e *critic* é do tipo *Fully Connected Network*, configurada com duas camadas escondidas de 256 neurônios cada e com função de ativação tangente hiperbólica. O otimizador utilizado foi *Adaptive Moment Estimation* (Adam). Todos os códigos desenvolvidos podem ser encontrados no repositório do *GitHub* em (Zamin, 2023).

A Figura 10 ilustra a recompensa média obtida em cada *step* para os modelos treinados com ambos os algoritmos para a temperatura de 25°C. Para o PPO, o tempo total de treinamento dos modelos das configurações A, B e C foi de: 4 horas e 43 minutos, 4 horas e 31 minutos, e 4 horas e 28 minutos. Já para cada modelo do SAC, o tempo total de treinamento foi de: 5 horas, 5 horas e 6 minutos, e 7 horas e 40 minutos. Como o IQB possui o valor máximo de 1 e cada episódio possui 7 ações, é ideal que a recompensa média se aproxime de 7. Entretanto, devido ao tempo morto do sistema, dificilmente a recompensa média atingirá o valor de 7. Assim, na Figura 10, observa-se que a performance do PPO é superior ao SAC para todas as configurações, pois o primeiro obteve recompensas médias superiores. Ademais, nota-se que o SAC convergiu apenas para a configuração B, que utiliza a técnica de *split-range* na malha de controle da temperatura de saída, o que torna o controle do sistema mais estável.

**Figura 10** - Recompensa média obtida com os algoritmos PPO e SAC para temperatura ambiente de 25°C.



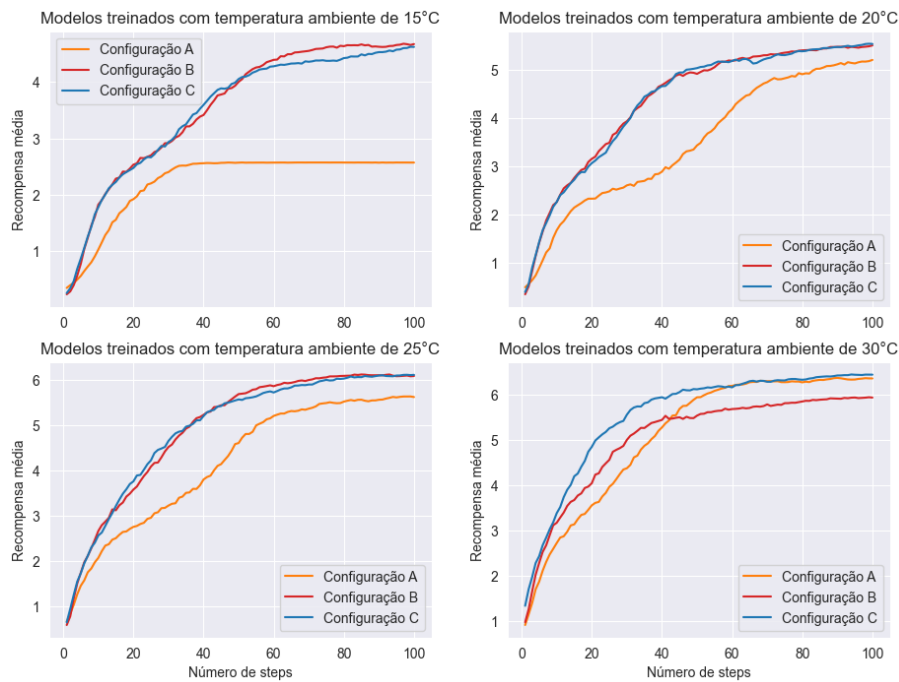
Fonte: elaborado pelo autor.

Existem duas possibilidades que podem explicar a falta de convergência do SAC. A primeira delas é o fato de o agente ser forçado a visitar todos os estados e ações devido a

presença do termo de entropia na função objetivo, o que pode aumentar o tempo de convergência do modelo. Além disso, segundo (Haarnoja *et al.*, 2018), o SAC é sensível ao parâmetro que controla a entropia, sendo necessário muitas vezes realizar a técnica de *hyperparameter tuning*, que consiste em testar o modelo com diversos parâmetros e escolher o conjunto que resultou na melhor performance. Entretanto, como o PPO obteve resultados satisfatórios, optou-se por não utilizar mais o SAC, focando apenas no PPO.

Em seguida, cada configuração foi treinada com o PPO para temperaturas ambiente de 15°C, 20°C, 25°C e 30°C, cujos gráficos com a recompensa média obtida durante cada *step* estão ilustrados na Figura 11. Isso foi feito para verificar o desempenho das estratégias de RL em temperaturas diferentes. As configurações B e C, em que o agente atua como ferramenta da camada de otimização em tempo real, obtiveram performances similares, com exceção dos modelos treinados com a temperatura de 30°C, em que a configuração C obteve uma recompensa média maior. Ou seja, em linhas gerais, a forma com que a fração da resistência elétrica é regulada, seja diretamente pelo agente ou através do *split-range*, não impacta consideravelmente na qualidade do banho quando a temperatura ambiente é fixa.

**Figura 11** - Recompensa média obtida com o algoritmo PPO para diferentes temperaturas ambiente.

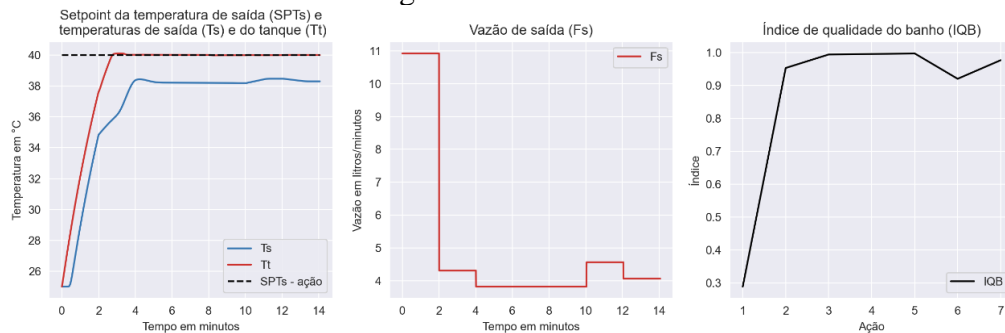


Fonte: elaborado pelo autor.

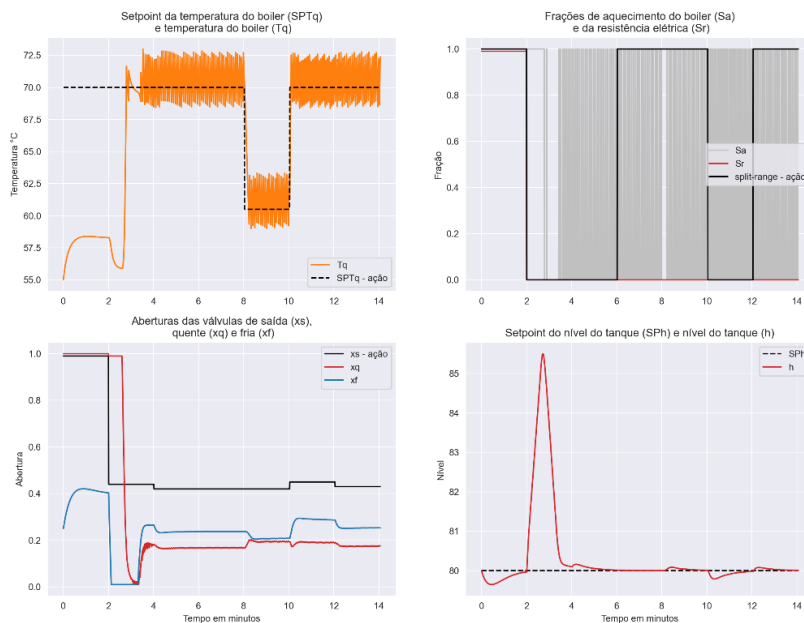


A configuração A, em que o agente atua na camada supervisória, obteve a pior performance para as temperaturas de 15°C, 20°C e 25°C, conforme ilustrado anteriormente na Figura 11. Desse modo, conclui-se que o agente é mais eficiente quando ele sugere o *setpoint* da temperatura de saída, atuando como otimizador em tempo real, e não quando ele regula diretamente a abertura da válvula de corrente quente, atuando na camada supervisória. As Figuras 12 e 13 ilustram o comportamento das principais variáveis do sistema obtidas pelo modelo treinado com a temperatura ambiente de 25°C para a configuração B. Já as Figuras 14 e 15 ilustram o mesmo para a configuração C.

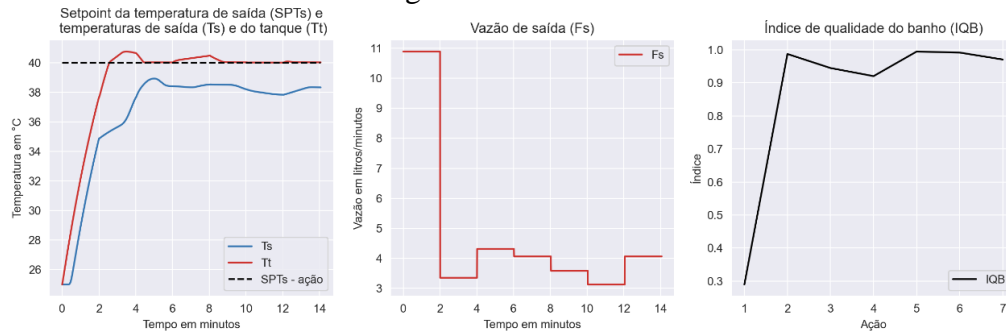
**Figura 12** - Temperatura de saída, vazão de saída e IQB para a configuração B treinada com o algoritmo PPO a 25°C.



**Figura 13** - Comportamento das variáveis de aquecimento, das válvulas e do tanque para a configuração B treinada com o algoritmo PPO a 25°C.

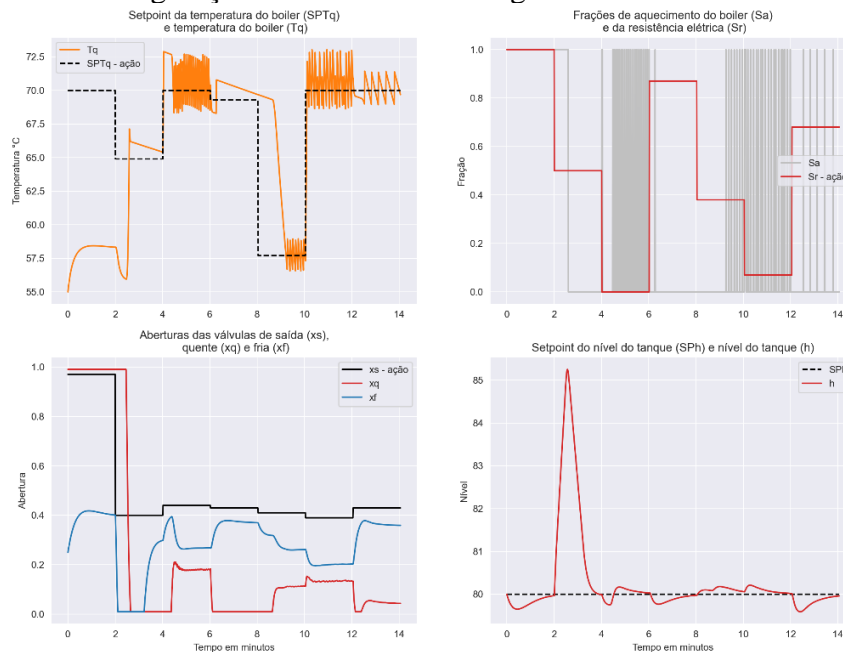


**Figura 14** - Temperatura de saída, vazão de saída e IQB para a configuração C treinada com o algoritmo PPO a 25°C.



Fonte: elaborado pelo autor.

**Figura 15** - Comportamento das variáveis de aquecimento, das válvulas e do tanque para a configuração C treinada com o algoritmo PPO a 25°C.



Fonte: elaborado pelo autor.

Tanto nas Figuras 12 e 14, observa-se que o pior IQB ocorre durante a primeira ação, pois no começo do episódio é preciso de mais água quente no *boiler* e no tanque, logo a vazão da corrente quente aumenta. Para compensar isso, o agente aumenta a vazão da válvula de saída. Como a temperatura e vazão de saída não estão nas faixas ideais, o IQB é baixo. Em seguida, o agente consegue identificar que não é mais necessária tanta corrente quente, e diminui a abertura da válvula de saída, mantendo seus valores aproximadamente constantes. A partir desse ponto, o IQB atinge valores ótimos e não sofre mais alterações significativas. Esse

comportamento repetiu-se para as outras temperaturas, sendo que quanto menor a temperatura, mais tempo o IQB demora para alcançar valores próximos a 1. Por exemplo, para a temperatura de 15°C, o IQB atinge valores ótimos somente após 6 minutos de banho.

Além disso, na Figura 13, o agente sugere ligar o *split-range* não somente no começo do episódio, mas também em momentos que a temperatura de saída poderia ser alterada da sua faixa ótima, mais especificamente nos minutos 6 e 12. Isso caracteriza-se como uma forma de prevenção para possíveis quedas da temperatura de saída. Porém, como não havia necessidade de acionar o *split-range*, a fração da resistência elétrica continuou zero. Essa forma de prevenção também ocorre com o agente treinado com a configuração C, conforme ilustrado na Figura 15, onde o agente aumenta o valor da fração da resistência elétrica nos minutos 6 e 12. O comportamento da temperatura de aquecimento do *boiler* também se apresentou similar em ambas as configurações B e C.

## 5.2 OTIMIZAÇÃO DE QUALIDADE E CUSTOS

Como as configurações B e C com o algoritmo PPO obtiveram o melhor desempenho na primeira fase do estudo, ambas as abordagens foram selecionadas para a segunda fase do estudo, cujo objetivo foi otimizar os custos do banho, com ênfase no custo da energia elétrica, ao mesmo tempo em que a qualidade do banho é maximizada. Todos os parâmetros do PPO, as configurações do episódio e o espaço de ações estabelecidos na primeira fase do estudo foram mantidos, com exceção do número de *steps* e do espaço de estados, onde neste último foram acrescentados os custos do banho (elétrico, gás e água). Para cada configuração, seis modelos foram treinados com os seguintes objetivos, considerando cenários em que a temperatura ambiente pode assumir valores inteiros entre 15°C e 30°C e a tarifa da energia elétrica pode assumir valores fixos entre R\$1,00 e R\$2,25 variando a cada R\$0,25:

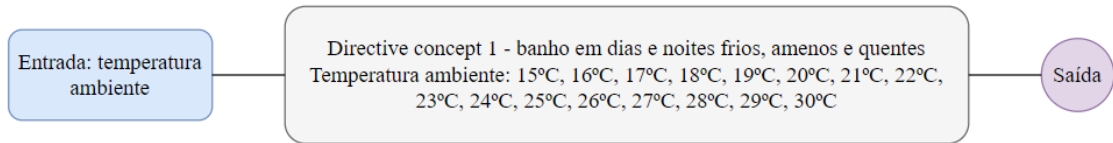
- *Modelo 1: Single concept* com um *directive concept* com o objetivo de maximizar a qualidade do banho;
- *Modelo 2: Single concept* com um *directive concept* com os objetivos de maximizar a qualidade do banho e minimizar os custos, com ênfase no custo da energia elétrica;

- *Modelo 3: Multi concept* com três *directive concepts* e um *selective concept* do tipo *learned* com o objetivo de maximizar a qualidade do banho;
- *Modelo 4: Multi concept* com três *directive concepts* e um *selective concept* do tipo *programmed* com o objetivo de maximizar a qualidade do banho;
- *Modelo 5: Multi concept* com seis *directive concepts* e um *selective concept* do tipo *learned* com os objetivos de maximizar a qualidade do banho e minimizar os custos, com ênfase no custo da energia elétrica;
- *Modelo 6: Multi concept* com seis *directive concepts* e um *selective concept* do tipo *programmed* com os objetivos de maximizar a qualidade do banho e minimizar os custos, com ênfase no custo da energia elétrica.

Os modelos 1, 3 e 4 são necessários para verificar se os modelos restantes realmente foram capazes de reduzir o custo da energia elétrica do banho. A comparação entre os métodos *single concept* e *multi concept* será realizada para compreender qual é a melhor abordagem de controle do sistema de chuveiro deste estudo. A desvantagem do método *single concept* é que ele pode obter uma performance pior do que o *multi concept*, pois o agente deverá aprender a controlar o banho de uma vez só com todos os cenários iniciais. Já o método *multi concept* pode demorar muito tempo para ser treinado, o que seria uma desvantagem caso ele fosse implementado em um chuveiro real e tivesse que ser retreinado constantemente. Ademais, para os modelos *multi concept*, decidiu-se comparar a performance dos dois tipos de *selective concept*, *programmed* e *learned*.

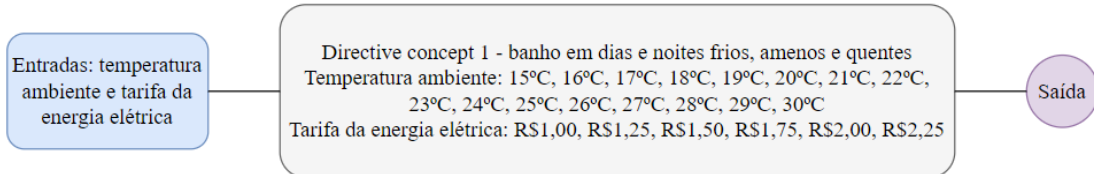
As Figuras 16 e 17 ilustram, respectivamente, o *brain design* para os modelos 1 e 2. Para todos os modelos, as variáveis de entrada são selecionadas aleatoriamente no *reset* de cada episódio, para que o agente possa ser treinado com todas as combinações. Os *designs* realizados consideram que a taxa da energia elétrica é mais cara durante a noite e mais barata durante o dia. Já a Figura 18 ilustra o *brain design* para os modelos 3 e 4. A diferença entre eles é o primeiro possui um *selective concept* do tipo *learned*, e o segundo do tipo *programmed*. Por fim, a Figura 19 ilustra o *brain design* para os modelos 5 e 6, que também diferem no tipo de *selective concept*.

**Figura 16 - Brain design para o modelo 1.**



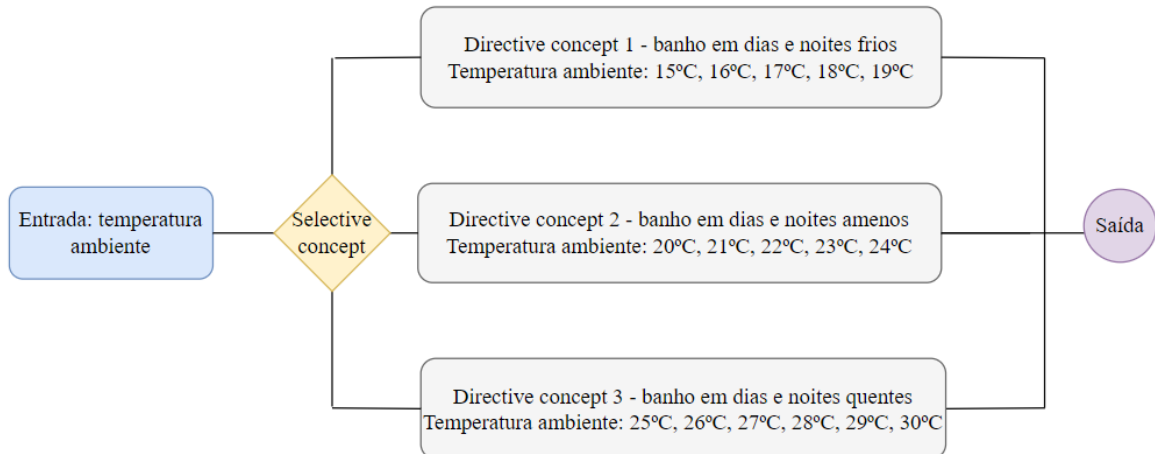
Fonte: elaborado pelo autor.

**Figura 17 - Brain design para o modelo 2.**



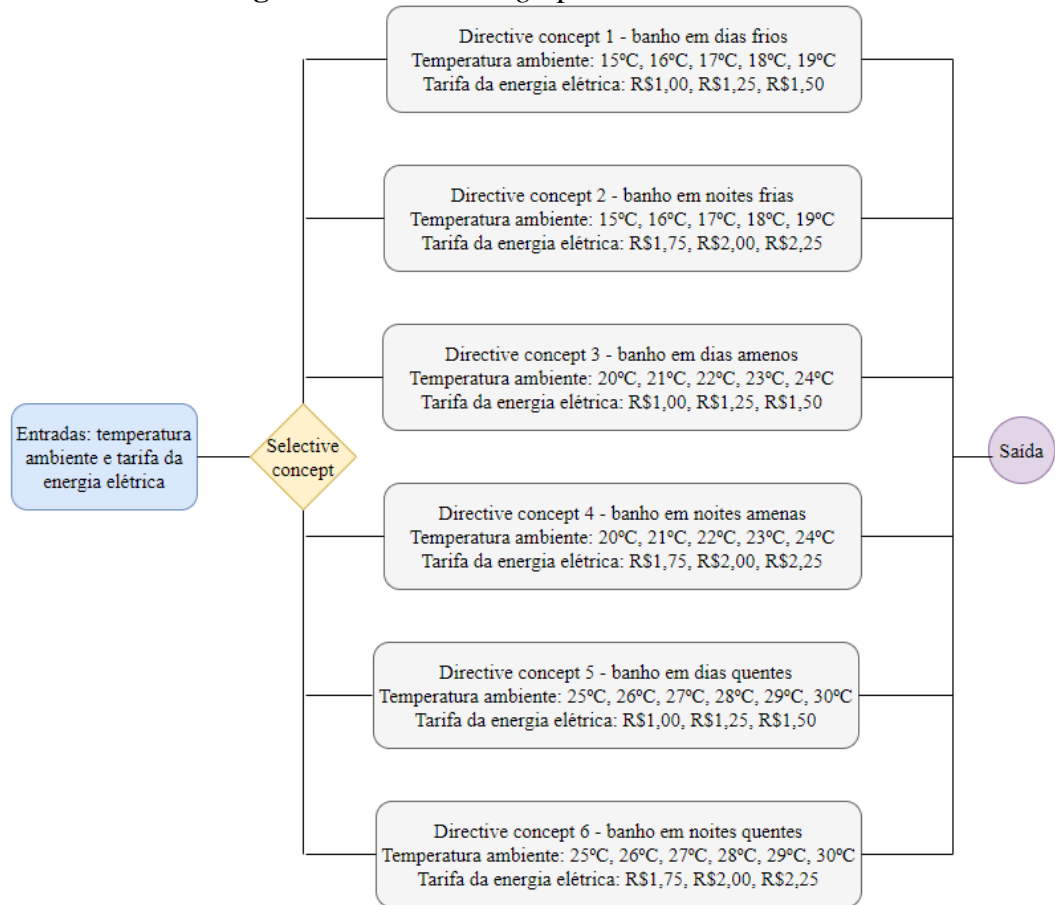
Fonte: elaborado pelo autor.

**Figura 18 - Brain design para os modelos 3 e 4.**



Fonte: elaborado pelo autor.

**Figura 19 - Brain design para os modelos 5 e 6.**



Fonte: elaborado pelo autor.

A recompensa para os modelos 1, 3 e 4 foi definida como o IQB obtido no final de cada iteração. Já a recompensa para os modelos restantes foi definida conforme a Equação 5.1. Como o objetivo principal continua sendo maximizar a qualidade do banho para todos os modelos, foi atribuído um peso maior ao IQB no cálculo da recompensa. Ademais, recompensas negativas podem induzir o agente a procurar o estado terminal, já que a recompensa não estaria sendo maximizada (Microsoft Project Bonsai, 2017). Em relação aos custos, foi atribuído um peso maior ao custo elétrico, pois ele é o custo mais importante a ser minimizado neste estudo. Os custos do banho foram calculados conforme a Seção 4.3.

$$\begin{aligned}
 \text{recompensa} = & 5 \times \text{IQB} - 2 \times \text{custo elétrico} - \text{custo do gás} \\
 & - \text{custo da água}
 \end{aligned}
 \tag{5.1}$$

A Tabela 4 mostra o número total de *steps* e o tempo total de treinamento para cada modelo. Como os modelos *single concept* precisavam aprender a controlar o banho em diversos cenários, aumentou-se o número de *steps*. Para os modelos *multi concept*, foi estabelecido 100 *steps* para treinar cada *concept*. Todos os códigos desenvolvidos na segunda fase do estudo também podem ser encontrados no repositório do *GitHub* em (Zamin, 2023). Após a finalização do treinamento, cada um dos 12 modelos foi avaliado em um episódio de 14 minutos de banho para todas as combinações de temperatura ambiente e tarifa da energia elétrica, totalizando 96 possíveis cenários.

**Tabela 4** - Número total de *steps* e tempo total de treinamento para cada modelo treinado com o algoritmo PPO na segunda fase do estudo.

	Número total de <i>steps</i>	Tempo total de treinamento
Configuração B – modelo 1	150	6 horas e 52 minutos
Configuração B – modelo 2	150	6 horas e 41 minutos
Configuração B – modelo 3	400	17 horas e 56 minutos
Configuração B – modelo 4	300	13 horas e 47 minutos
Configuração B – modelo 5	700	31 horas e 33 minutos
Configuração B – modelo 6	600	27 horas e 20 minutos
Configuração C – modelo 1	150	6 horas e 43 minutos
Configuração C – modelo 2	150	6 horas e 41 minutos
Configuração C – modelo 3	400	17 horas e 43 minutos
Configuração C – modelo 4	300	13 horas e 33 minutos
Configuração C – modelo 5	700	32 horas e 42 minutos
Configuração C – modelo 6	600	27 horas e 21 minutos

Fonte: elaborado pelo autor.

Para cada episódio avaliado, calculou-se o custo total do banho, o custo elétrico total, o custo de gás total, o custo de água total, a soma total do IQB durante um episódio e a média do IQB para todas as iterações do episódio. Estes resultados são apresentados no Apêndice A. A Tabela 5 mostra a soma do IQB durante um episódio, a soma do IQB médio em um episódio e os valores mínimo e máximo do IQB médio considerando todas as 96 combinações de cenários de cada modelo. Como o valor máximo do IQB é 1, espera-se que as somas do IQB e do IQB médio de um episódio para todos os cenários de um mesmo modelo atinjam idealmente os valores 672 e 96, respectivamente. Entretanto, de acordo com a Tabela 5, apesar de nenhum modelo atingir esses números, pois quanto menor a temperatura ambiente mais difícil é de obter

valores de IQB próximos a 1 no começo do episódio, a maioria dos modelos atingiu valores similares entre si, que estão próximos aos ideais.

**Tabela 5** - Soma do IQB durante um episódio, soma do IQB médio de um episódio e valores mínimo e máximo do IQB médio para todas as 96 combinações de cenários de temperatura ambiente e tarifa da energia elétrica.

	Soma do IQB de um episódio	Soma do IQB médio em um episódio	Valor mínimo do IQB médio	Valor máximo do IQB médio
Configuração B – modelo 1	557,32	79,61	0,65	0,92
Configuração B – modelo 2	552,45	78,94	0,67	0,90
Configuração B – modelo 3	560,57	79,95	0,68	0,91
Configuração B – modelo 4	546,35	78,06	0,65	0,90
Configuração B – modelo 5	553,73	79,13	0,67	0,91
Configuração B – modelo 6	541,52	77,33	0,66	0,90
Configuração C – modelo 1	557,32	79,61	0,68	0,92
Configuração C – modelo 2	532,13	76,01	0,61	0,89
Configuração C – modelo 3	564,09	80,66	0,68	0,93
Configuração C – modelo 4	549,11	78,43	0,62	0,93
Configuração C – modelo 5	551,32	78,80	0,68	0,91
Configuração C – modelo 6	542,03	77,43	0,64	0,92

Fonte: elaborado pelo autor.

Ou seja, o fato de os custos serem incluídos na recompensa de alguns modelos não influenciou na otimização da qualidade do banho. O modelo com os menores valores da Tabela 5 foi o número 2 da configuração C, com uma abordagem *single concept* para maximizar a qualidade e otimizar os custos do banho. Esse comportamento não se repetiu para o outro modelo *single concept* que também otimiza os custos do banho, o número 2 da configuração B, pois nesse caso, a estratégia de *split-range* ajuda a estabilizar a performance do agente. Logo, quando o agente é treinado em um chuveiro com uma estratégia que regula diretamente a fração da resistência elétrica, o *brain* do tipo *single concept* não é muito eficiente em lidar com objetivos opostos.

Ao comparar os custos totais para todos os 12 modelos, nenhum obteve valores significativamente menores. A Tabela 6 mostra a soma dos custos totais para todas as 96 combinações de cenários de cada modelo. Os modelos que obtiveram os maiores custos foram os de número 1, 3 e 4 da configuração C, pois como os custos não foram otimizados na recompensa, o agente preferiu utilizar mais a resistência elétrica ao invés do aquecimento do



*boiler*, aumentando consideravelmente o custo elétrico. Esse comportamento já havia sido verificado na primeira fase do estudo, conforme ilustrado na Figura 15.

**Tabela 6** - Soma dos custos totais do banho, elétrico, de gás e de água para todas as 96 combinações de cenários de temperatura ambiente e tarifa da energia elétrica.

	Soma do custo total do banho	Soma do custo elétrico total	Soma do custo de gás total	Soma do custo de água total
Configuração B – modelo 1	R\$ 143,20	R\$ 42,84	R\$ 72,37	R\$ 27,92
Configuração B – modelo 2	R\$ 143,81	R\$ 41,96	R\$ 73,53	R\$ 28,32
Configuração B – modelo 3	R\$ 139,56	R\$ 41,67	R\$ 71,06	R\$ 26,75
Configuração B – modelo 4	R\$ 143,34	R\$ 43,05	R\$ 72,24	R\$ 28,09
Configuração B – modelo 5	R\$ 142,92	R\$ 40,70	R\$ 74,16	R\$ 28,17
Configuração B – modelo 6	R\$ 143,77	R\$ 39,86	R\$ 74,90	R\$ 29,02
Configuração C – modelo 1	R\$ 217,69	R\$ 143,00	R\$ 46,29	R\$ 28,40
Configuração C – modelo 2	R\$ 160,90	R\$ 57,65	R\$ 72,78	R\$ 30,45
Configuração C – modelo 3	R\$ 213,90	R\$ 135,38	R\$ 49,12	R\$ 29,33
Configuração C – modelo 4	R\$ 217,24	R\$ 135,82	R\$ 51,21	R\$ 30,26
Configuração C – modelo 5	R\$ 145,60	R\$ 44,81	R\$ 72,37	R\$ 28,41
Configuração C – modelo 6	R\$ 156,20	R\$ 58,80	R\$ 68,35	R\$ 29,04

Fonte: elaborado pelo autor.

Em relação aos modelos 2, 5 e 6 da configuração C, verifica-se que apesar de o agente ter sido capaz de otimizar o custo elétrico do banho, não houve uma redução significativa dos custos de gás e de água, mesmo eles tendo sido incluídos na recompensa. O modelo 2, por ser *single concept*, resultou em uma soma de custo elétrico total um pouco maior pelo fato de o agente ter sido treinado com todos os cenários ao mesmo tempo. Já para todos os modelos da configuração B, os valores das somas dos custos foram similares, e o agente também não foi capaz de otimizar os custos de gás e de água nos modelos 2, 5 e 6. No caso do custo elétrico, este já havia sido otimizado devido a utilização do *split-range* no chuveiro.

Ou seja, quando o agente controla a fração da resistência elétrica através do *split-range*, a própria malha de controle faz o papel de otimizador do custo elétrico, porque ela utiliza a resistência elétrica somente quando necessário, que é no começo do episódio. Logo, o papel do agente torna-se apenas aprender a acionar o *split-range* no começo do episódio. Porém, se não for possível implementar uma estratégia de *split-range* no chuveiro, torna-se necessário incluir o custo elétrico na função recompensa do agente de RL caso deseje-se realizar sua otimização. Neste segundo caso, o agente de RL consegue reduzir o custo da energia elétrica

ao mesmo tempo em que maximiza a qualidade do banho. Além disso, a inclusão do *split-range* no chuveiro tornou os modelos 1 e 2 da configuração B, que são *single concept*, mais estáveis, de forma que sua performance foi similar aos modelos *multi concept* com a vantagem de terem sido treinados em menos tempo.

Em relação aos modelos treinados com o *selective concept* do tipo *learned*, em muitos casos, o *selective concept* selecionou *directive concepts* que não correspondiam ao cenário atual de temperatura ambiente e tarifa da energia elétrica. Essas constatações também podem ser verificadas nos resultados do Apêndice A. Em média, as somas do IQB e do custo total do banho foram 2.3% maior e 3.03% menor para os modelos com *selective concept* do tipo *learned* quando comparados com os que têm *selective concept* do tipo *programmed*. Assim, o fato de o agente ter escolhido *concepts* de outros cenários durante um episódio aumentou levemente a qualidade do banho e reduziu um pouco os custos. Portanto, seria viável utilizar os modelos com *selective concept* do tipo *learned* em produção, mesmo que o tempo de treinamento deles seja maior.

Por conseguinte, considerando que o usuário deseje dar ênfase na otimização do custo elétrico e que seja possível implementar a estratégia de *split-range* no chuveiro, os modelos mais indicados para uma implementação na vida real são os de número 1 e 3 da configuração B, pois com o *split-range* não é necessário incluir os custos na recompensa. O modelo 3 desta configuração foi o que obteve a melhor performance, sendo recomendado para situações em que as tarifas do gás e da água não variam muito. Entretanto, caso esses preços mudem constantemente e não estejam mais na faixa de valores em que o modelo inicial foi treinado, será necessário realizar novamente o treinamento. Nesse caso, o modelo 1 da configuração B é mais eficiente, pois apesar de ele ser *single concept*, o custo da energia elétrica foi otimizado sem perda de qualidade e o tempo de treinamento foi menor. Caso não seja possível adicionar o *split-range* ao sistema, o próximo modelo mais indicado é o número 5 da configuração C, pois ele atingiu valores similares de qualidade e custos quando comparado com os modelos 1 e 3 da configuração B.

### 5.3 VALIDAÇÃO DAS SOLUÇÕES PROPOSTAS

Conforme os resultados explicitados anteriormente, as soluções que geraram os menores custos do banho sem perda de qualidade foram aquelas com os modelos de número 1 e 3 da configuração B e de número 5 da configuração C. Para validar as soluções propostas, os três modelos foram avaliados em cenários desconhecidos pelo agente treinado, que incluíram temperaturas ambiente de 15,5°C, 20,5°C e 25,5°C e tarifa da energia elétrica de R\$2,10. Este valor da tarifa foi escolhido para validar os modelos no caso limite em que a energia elétrica é mais cara. Tabela 7 mostra estes resultados para os custos totais do banho, a soma do IQB e o IQB médio de cada episódio avaliado.

**Tabela 7** – Resultados para validação dos melhores modelos para cenários com temperaturas ambiente de 15,5°C, 20,5°C e 25,5°C e tarifa da energia elétrica de R\$2,10.

	Temperatura ambiente em °C	Custo total do banho (R\$)	Custo elétrico total (R\$)	Custo de gás total (R\$)	Custo de água total (R\$)	Soma do IQB de um episódio	IQB médio de um episódio
Configuração B – modelo 1	15,5	2,46	0,97	1,12	0,37	4,86	0,69
Configuração B – modelo 3	15,5	2,40	0,92	1,12	0,36	4,86	0,69
Configuração C – modelo 5	15,5	2,62	1,12	1,12	0,38	4,83	0,69
Configuração B – modelo 1	20,5	1,79	0,62	0,86	0,31	5,74	0,82
Configuração B – modelo 3	20,5	1,75	0,62	0,83	0,30	5,81	0,83
Configuração C – modelo 5	20,5	1,82	0,67	0,84	0,31	5,49	0,78
Configuração B – modelo 1	25,5	1,27	0,42	0,59	0,26	6,26	0,89
Configuração B – modelo 3	25,5	1,26	0,43	0,58	0,25	6,30	0,90
Configuração C – modelo 5	25,5	1,32	0,50	0,57	0,25	6,22	0,89

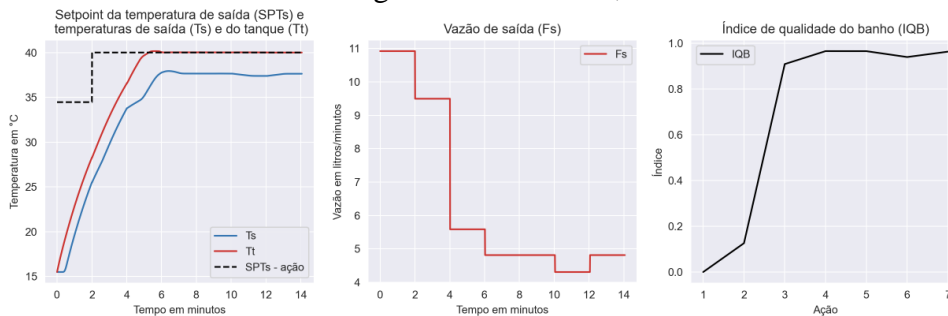
Fonte: elaborado pelo autor.

Ao comparar os valores obtidos da Tabela 7 com aqueles apresentados no Apêndice A em cenários de temperatura ambiente e tarifa de energia elétrica próximos, observa-se que os

valores da Tabela 7 ficam entre as faixas de valores dos cenários utilizados para o treinamento. Isso indica que os modelos de RL desenvolvidos funcionam para qualquer combinação de temperatura ambiente e tarifa da energia elétrica que esteja dentro do intervalo especificado no treinamento do agente. De acordo com a Tabela 7, o modelo que obteve os menores custos ainda continua sendo o de número 3 da configuração B, logo ele foi escolhido como a melhor solução. Sua implementação em *Python* pode ser encontrada no Apêndice B.

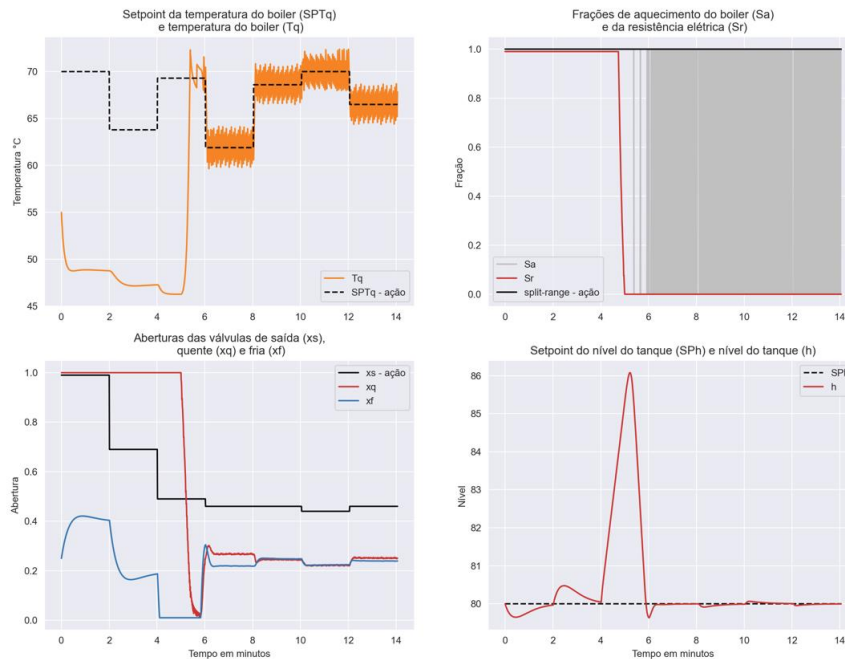
As Figuras 20 e 21 ilustram o comportamento das principais variáveis do sistema obtidas para este modelo com temperatura ambiente de 15,5°C e tarifa da energia elétrica de R\$2,10. Na Figura 20, observa-se novamente que o IQB assume valores próximos a zero no começo do episódio, conforme já havia sido explicado na primeira fase do estudo. De acordo com a Figura 21, apesar de a fração da resistência elétrica ser acionada somente no início do episódio, o agente de RL aprendeu a ligar o *split-range* em todas as ações, mesmo quando não é necessário. Esse comportamento também foi observado para as outras temperaturas. Ou seja, para o modelo 3 da configuração B, a utilização do *split-range* poderia ser removida do espaço de ações, deixando o *split-range* sempre ligado sem depender do agente de RL.

**Figura 20** - Temperatura de saída, vazão de saída e IQB para o modelo 3 da configuração B treinado com o algoritmo PPO para o cenário de temperatura ambiente de 15,5°C e tarifa da energia elétrica de R\$2,10.



Fonte: elaborado pelo autor.

**Figura 21** - Comportamento das variáveis de aquecimento, das válvulas e do tanque para o modelo 3 da configuração B treinado com o algoritmo PPO para o cenário de temperatura ambiente de 15,5°C e tarifa da energia elétrica de R\$2,10.



Fonte: elaborado pelo autor.

Considerando que o modelo 3 da configuração B represente fisicamente o comportamento de um chuveiro, para implementar a solução desenvolvida na vida real, poderia ser utilizado um microcontrolador do tipo *Raspberry Pi*. Seria necessário instalar sensores para medir as seguintes variáveis: temperatura de saída ( $T_s$ ), temperatura do *boiler* ( $T_q$ ), temperatura do tanque ( $T_t$ ), nível do tanque ( $h$ ), vazão de saída ( $F_s$ ), vazão da corrente quente ( $F_q$ ), vazão da corrente fria ( $F_f$ ) e temperatura ambiente ( $T_{inf}$ ). O *deployment* do agente de RL treinado seria realizado com o microcontrolador *Raspberry Pi* em *Python*. Os sensores fornecem a leitura das variáveis ao *Raspberry Pi*, que serão convertidas para valores analógicos e enviadas para o código com o modelo treinado. Em seguida, dependendo do estado recebido, o agente sugere as ações, que serão aplicadas no sistema através da comunicação com o *Raspberry Pi*.

## 6 CONCLUSÃO

Este trabalho aplicou a metodologia de *Reinforcement Learning* (RL) e *design* de sistemas autônomos para realizar a otimização dos custos, com ênfase no custo da energia elétrica, em um modelo computacional de planta de chuveiro com aquecimento com dois graus de liberdade, com tempo morto e sem corrente de distúrbio. Inicialmente, o estudo buscou compreender qual é a melhor estratégia de aplicação de RL na planta de chuveiro: RL como ferramenta da camada supervisória ou RL como ferramenta da camada de otimização em tempo real. Foram testadas três configurações de planta de chuveiro, com dois algoritmos de RL, *Proximal Policy Optimization* (PPO) e *Soft Actor-Critic* (SAC).

Ao treinar cada configuração com ambos os algoritmos para a temperatura ambiente de 25°C, constatou-se que o PPO obteve uma performance superior ao SAC em um tempo de treinamento similar. Em seguida, as três configurações de chuveiro foram treinadas com o PPO para as temperaturas ambiente de 15°C, 20°C, 25°C e 30°C. Ambas as configurações B e C, que representaram RL como ferramenta da camada RTO, obtiveram performances similares e alcançaram valores de recompensa média muito maiores que a configuração A, que representou RL como ferramenta da camada supervisória. Ou seja, o agente é mais eficiente quando ele atua como otimizador em tempo real, e não quando ele atua como ferramenta na camada supervisória. Além disso, quando a temperatura é fixa, a forma com que a fração da resistência elétrica é regulada, seja diretamente pelo agente ou através do *split-range*, não impacta consideravelmente na qualidade do banho.

Desse modo, ambas as configurações B e C e o algoritmo PPO foram utilizados na segunda fase do estudo, cujo objetivo principal era compreender se é possível aplicar RL para otimizar os custos do banho, com ênfase no custo elétrico, ao mesmo tempo em que a qualidade é maximizada. Seis modelos diferentes foram treinados para as configurações B e C com 96 combinações de cenários iniciais de temperatura ambiente e tarifa da energia elétrica. Os modelos 1 e 2 eram do tipo *single concept*, enquanto os modelos 3 a 6 eram do tipo *multi concept*, sendo que os de número 3 e 5 tinham um *selective concept* do tipo *learned*, e os de número 4 e 6 tinham um *selective concept* do tipo *programmed*. Ademais, os modelos 1, 3 e 4 tinham apenas o objetivo de maximizar a qualidade do banho, enquanto os outros também otimizavam os custos.

Todos os modelos da configuração B obtiveram resultados similares para a qualidade do banho e para os custos totais. Logo, a inclusão do *split-range* já torna o sistema otimizado, não sendo necessário incluir os custos na recompensa do agente. Para os modelos da configuração C, os de número 1, 3 e 4, que maximizaram somente a qualidade do banho, obtiveram um custo total do banho muito maior do que os de número 2, 5 e 6, que também realizaram a otimização dos custos. Logo, quando não é possível implementar uma estratégia de *split-range* no sistema, torna-se necessário incluir os custos na recompensa do agente para realizar a otimização dos custos. O modelo com a melhor performance foi o de número 3 da configuração B, seguido pelos modelos 1 da configuração B e 5 da configuração C.

Deste modo, entende-se que o estudo obteve sucesso em aplicar RL para otimização deste modelo de planta de chuveiro. Portanto, nesse caso, RL pode ser caracterizado como uma alternativa para as estratégias de controle e otimização tradicionais. Entretanto, para averiguar se RL apresenta um desempenho superior ao controle tradicional, seria necessário realizar comparações entre ambos os métodos, incluindo estudos com MPCs e ferramentas RTO. Novos experimentos variando os parâmetros do algoritmo SAC também poderiam ser realizados para tentar melhorar seu desempenho. Além disso, em trabalhos futuros, os coeficientes da expressão para a recompensa do agente poderiam ser alterados ou outras funções poderiam ser testadas para verificar se haveria diferenças nos desempenhos dos modelos. Em seguida, para realizar a otimização dos custos de gás e de água com variações nas tarifas, também seria possível treinar outros modelos com diferentes *brain designs*, incluindo mais *directive* e/ou *selective concepts*. Ademais, em todo o estudo realizado e nos trabalhos futuros poderia ser incluindo o efeito da corrente de distúrbio na simulação do banho.

Finalmente, outro estudo possível seria construir um modelo para que o cliente possa escolher qual é a temperatura de saída do banho. A solução que o agente de RL forneceu neste estudo converge sempre na mesma faixa de temperatura de saída, mas entende-se que alguns indivíduos possam preferir outras temperaturas de banho. Assim, o *setpoint* da temperatura de saída seria removido do espaço de ações e adicionado nas variáveis de entrada, junto com a temperatura ambiente e a tarifa da energia elétrica. Logo, seria necessário projetar mais *concepts* para englobar os novos cenários. O IQB poderia continuar sendo utilizado na recompensa, pois nesse caso, o trabalho do agente seria maximizar o IQB para a temperatura de banho específica que o usuário escolheu.

## REFERÊNCIAS

- AL-JABERY, Khalid *et al.* A novel grid load management technique using electric water heaters and Q-learning. **IEEE INTERNATIONAL CONFERENCE ON SMART GRID COMMUNICATIONS**, 2014, Venice. **Proceedings** [...]. Venice: IEEE, 2014. p. 776-781. Disponível em: <https://ieeexplore.ieee.org/document/7007742>. Acesso em: 29 abr. 2023.
- AMASYALI, Kadir *et al.* Deep reinforcement learning for autonomous water heater control. **Buildings**, Basel, v. 11, n. 11, p. 548, 2021. Disponível em: <https://www.mdpi.com/2075-5309/11/11/548>. Acesso em: 23 abr. 2023.
- AMASYALI, Kadir *et al.* Reinforcement-learning-based Smart Water Heater Control: An Actual Deployment. **IEEE POWER & ENERGY SOCIETY INNOVATIVE SMART GRID TECHNOLOGIES CONFERENCE**, 2023, Washington. **Proceedings** [...] Washington: IEEE, 2023. p. 1-5. Disponível em: <https://ieeexplore.ieee.org/document/10066373>. Disponível em: 29 abr. 2023.
- ANDERSON, Kence. **Designing Autonomous AI: A Guide for Machine Teaching**. 1st. ed. Sebastopol: O'Reilly Media, 2022.
- ASTRÖM, Karl J.; HÄGGLUND, Tore. **PID Controllers: Theory, Design, and Tuning**. 2nd. ed. Durham: Instrument Society of America, 1995.
- CHATGPT. Quais são as diferenças entre os algoritmos de Reinforcement Learning on-policy e off-policy? **OpenAI**, [s.l.], 2023a. Disponível em: <https://openai.com/blog/chatgpt>. Acesso em: 9 jun. 2023.
- CHATGPT. Quais são as vantagens do algoritmo Proximal Policy Optimization? **OpenAI**, [s.l.], 2023b. Disponível em: <https://openai.com/blog/chatgpt>. Acesso em: 29 abr. 2023.
- CHATGPT. Quais são as vantagens do algoritmo Soft Actor-Critic? **OpenAI**, [s.l.], 2023c. Disponível em: <https://openai.com/blog/chatgpt>. Acesso em: 25 maio 2023.
- ELETROBRAS. **Pesquisa de Posse e Hábitos de Uso de Equipamentos Elétricos na Classe Residencial**. Rio de Janeiro: ELETROBRAS; PROCEL, 2019. Disponível em: <https://eletrobras.com/pt/AreasdeAtuacao/BRASIL.pdf>. Acesso em: 10 jan. 2023.
- EPE. **Anuário Estatístico de Energia Elétrica 2022: Ano base 2021**. Rio de Janeiro: Ministério de Minas e Energia do Brasil, 2022. Disponível em: <https://www.epe.gov.br/pt/publicacoes-dados-abertos/publicacoes/anuario-estatistico-de-energia-eletrica>. Acesso em: 16 jan. 2023.
- EPE. **Nota Técnica DEA 12/16**. Avaliação da eficiência energética e geração distribuída para os próximos 10 anos (2015-2024). Rio de Janeiro: Ministério de Minas e Energia do Brasil, 2016. Disponível em: <https://www.epe.gov.br/sites-pt/publicacoes-dados->



abertos/publicacoes/PublicacoesArquivos/publicacao-245/topico-264/DEA%2012-16%20-%20Ef%20energetica%202015-2024%5B1%5D.pdf. Acesso em: 10 jan. 2023.

FARAMA FOUNDATION. Gymnasium Documentation. **Farama Foundation**, [s.l.], 2023. Disponível em: <https://gymnasium.farama.org/>. Acesso em: 24 abr. 2023.

GHSI, Eneidir; GOSCH, Samuel; LAMBERTS, Roberto. Electricity end-uses in the residential sector of Brazil. **Energy Policy**, Amsterdam, v. 35, n. 8, p. 4107–4120, 2007. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0301421507000559>. Acesso em: 11 jan. 2023.

GIAQUINTO, Cláudia Daniela Melo. **Metodologia para quantificação e acompanhamento de indicadores-chave de desempenho operacional**. 2017. Dissertação (Mestrado em Engenharia) – Escola de Engenharia, Universidade Federal do Rio Grande do Sul, Porto Alegre, 2017. Disponível em: <https://www.lume.ufrgs.br/handle/10183/172050>. Acesso em: 29 jul. 2023.

HAARNOJA, Tuomas *et al.* Soft Actor-Critic: Off-Policy Maximum Entropy Deep Reinforcement Learning with a Stochastic Actor. *In: INTERNATIONAL CONFERENCE ON MACHINE LEARNING*, 35., 2018, Stockholm. **Proceedings** [...]. Stockholm: PMLR, 2018. v. 5, p. 2976–2989. Disponível em: <https://arxiv.org/abs/1801.01290v2>. Acesso em: 4 jan. 2023.

HOTZ, Nick. What is CRISP DM?. **Data Science Process Alliance**, [s.l.], 2023. Disponível em: <https://www.datascience-pm.com/crisp-dm-2/>. Acesso em: 25 ago. 2023.

LUYBEN, William L. **Process Modelling, Stimulation and Control for Chemical Engineers**. 2th. ed. Singapore: McGraw-Hill, Inc., 1996.

MICROSOFT PROJECT BONSAI. Writing Great Reward Functions – Bonsai. [s.l.]: Bonsai, 2017. Disponível em: <https://www.youtube.com/watch?v=OR3PnJEisqk&t=189s>. Acesso em: 27 jul. 2023.

MORALES, Miguel. **Grokking Deep Reinforcement Learning**. 1st. ed. Shelter Island: Manning Publications Co., 2020.

NUMPY. Numpy API Reference Guide: numpy.trapz. **NumPy**, [s.l.], 2023. Disponível em: <https://numpy.org/doc/stable/reference/generated/numpy.trapz.html>. Acesso em: 30 jul. 2023.

RAY RLLIB. Industry-Grade Reinforcement Learning. **Ray**, [s.l.], 2023a. Disponível em: <https://docs.ray.io/en/latest/rllib/index.html>. Acesso em: 24 abr. 2023.

RAY RLLIB. Proximal Policy Optimization (PPO). **Ray**, [s.l.], 2023b. Disponível em: <https://docs.ray.io/en/latest/rllib/rllib-algorithms.html#ppo>. Acesso em: 2 jun. 2023.

RAY RLLIB. Soft Actor-Critic (SAC). **Ray**, [s.l.], 2023c. Disponível em: <https://docs.ray.io/en/latest/rllib/rllib-algorithms.html#sac>. Acesso em: 2 jun. 2023.

SCHULMAN, John *et al.* Proximal Policy Optimization Algorithms. **ArXiv**, Ithaca, 2017. Disponível em: <https://arxiv.org/abs/1707.06347v2>. Acesso em: 26 dez. 2022.

SCIPY. SciPy API Reference Guide: `scipy.integrate.solve_ivp`. **SciPy**, [s.l.], 2023. Disponível em: [https://docs.scipy.org/doc/scipy/reference/generated/scipy.integrate.solve\\_ivp.html](https://docs.scipy.org/doc/scipy/reference/generated/scipy.integrate.solve_ivp.html). Acesso em: 30 jul. 2023.

SPIELBERG, S. P.K.; GOPALUNI, R. B.; LOEWEN, P. D. Deep reinforcement learning approaches for process control. *In: INTERNATIONAL SYMPOSIUM ON ADVANCED CONTROL OF INDUSTRIAL PROCESSES*, 6th., 2017, Taipei. **Anais [...]**. Taipei: AdCONIP, 2017, p. 201–206. Disponível em: [https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=7983780&casa\\_token=Tn841KIYuYsAAAAA:nTAamOQE0i81heOU1Sy2PcT0kr7xSIKTrOM6EPEMU1VmfXMiwoZNV3WgZeQTKKUMxudAaC0cQ](https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=7983780&casa_token=Tn841KIYuYsAAAAA:nTAamOQE0i81heOU1Sy2PcT0kr7xSIKTrOM6EPEMU1VmfXMiwoZNV3WgZeQTKKUMxudAaC0cQ). Acesso em: 24 abr. 2023.

SUPERGASBRAS. Qual a vantagem do poder calorífico do GLP? **Supergasbras**, [s.l.], 2021. Disponível em: <https://www.supergasbras.com.br/super-blog/negocio/2021/qual-a-vantagem-do-poder-calorifico-do-glp#:~:text=GLP%3A%2011.750%20Kcal%2Fkg>. Acesso em: 13 maio 2023.

SUTTON, Richard S.; BARTO, Andrew G. **Reinforcement Learning: An Introduction**. 2th. ed. Cambridge: MIT Press, 2018.

TRIERWEILER, Jorge Otávio. 01\_5\_Resolução\_do\_Exercício\_01\_Primeira\_Parte. [s.l.]: Jorge Otávio Trierweiler, 2021a. Disponível em: [https://www.youtube.com/watch?v=KP9jhTENc6g&list=PLenMidpg34nvQskL5o\\_8DJ6StLc6NMArM](https://www.youtube.com/watch?v=KP9jhTENc6g&list=PLenMidpg34nvQskL5o_8DJ6StLc6NMArM). Acesso em: 25 abr. 2023.

TRIERWEILER, Jorge Otávio. 01\_6\_Continuação\_da\_Resolução\_Exercício\_1. [s.l.]: Jorge Otávio Trierweiler, 2021b. Disponível em: [https://www.youtube.com/watch?v=t-0neIqnnF0&list=PLenMidpg34nvQskL5o\\_8DJ6StLc6NMArM](https://www.youtube.com/watch?v=t-0neIqnnF0&list=PLenMidpg34nvQskL5o_8DJ6StLc6NMArM). Acesso em: 25 abr. 2023.

TRIERWEILER, Jorge O.; FARENZENA, Marcelo. Uma Visão Geral das Tecnologias Atualmente Empregadas em Controle Avançado de Processos Industriais. **Controle e Instrumentação**, São Paulo, v. 123, p. 65-78, 2007.

WANG, Zhe; HONG, Tianzhen. Reinforcement Learning for building controls: The opportunities and challenges. **Applied Energy**, Amsterdã, v. 269, p. 115036, 2020. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0306261920305481>. Acesso em: 24 abr. 2023.

WINDER, Phil. **Reinforcement Learning: Industrial Applications of Intelligent Agents**. 1th. ed. Sebastopol: O'Reilly Media, Inc., 2020.

ZAMIN, Maria Paula Coser. Repositório no GitHub com os códigos desenvolvidos. **GitHub**, [s.l.], 2023. Disponível em: <https://github.com/mpaulazamin/tcc-final-models>. Acesso em: 25 abr. 2023.

ZSEMBINSZKI, Gabriel *et al.* Deep Learning Optimal Control for a Complex Hybrid Energy Storage System. **Buildings**, Basel, v. 11, n. 5, p. 194, 2021. Disponível em: <https://www.mdpi.com/2075-5309/11/5/194/htm>. Acesso em: 19 abr. 2023.

## APÊNDICE A – Resultados complementares para a segunda fase do estudo com o algoritmo PPO

Das Tabelas A.1 até A.12, mostram-se os resultados principais para os modelos treinados com o algoritmo PPO para as configurações B e C de chuveiro na segunda fase do estudo, cujo objetivo era verificar se é possível aplicar RL no modelo de chuveiro para otimizar os custos do banho, com ênfase no custo da energia elétrica. Após a finalização do treinamento, todos os modelos foram avaliados para cada combinação de temperatura ambiente e tarifa de energia elétrica, totalizando 96 cenários. No episódio avaliado, calculou-se o custo total do banho, o custo elétrico total, o custo de gás total, o custo de água total, a soma do IQB obtido no episódio e a média do IQB do episódio. Os modelos treinados para cada configuração foram:

- *Modelo 1: Single concept* com um *directive concept* para maximizar a qualidade do banho;
- *Modelo 2: Single concept* com um *directive concept* para maximizar a qualidade do banho e minimizar os custos, com ênfase no custo da energia elétrica;
- *Modelo 3: Multi concept* com três *directive concepts* e um *selective concept* do tipo *learned* para maximizar a qualidade do banho;
- *Modelo 4: Multi concept* com três *directive concepts* e um *selective concept* do tipo *programmed* para maximizar a qualidade do banho;
- *Modelo 5: Multi concept* com seis *directive concepts* e um *selective concept* do tipo *learned* para maximizar a qualidade do banho e minimizar os custos, com ênfase no custo da energia elétrica;
- *Modelo 6: Multi concept* com seis *directive concepts* e um *selective concept* do tipo *programmed* para maximizar a qualidade do banho e minimizar os custos, com ênfase no custo da energia elétrica.

**Tabela A.1** – Custo total do banho em R\$ para os 6 modelos da configuração B treinados com o algoritmo PPO na segunda fase do estudo.

(continua)

Temperatura ambiente em °C	Tarifa da energia elétrica (R\$)	Modelo 1	Modelo 2	Modelo 3	Modelo 4	Modelo 5	Modelo 6	Maior diferença de valores
15	1,00	2,04	2,04	1,98	2,00	2,07	1,99	0,09
15	1,25	2,14	2,19	2,08	2,15	2,12	2,12	0,11
15	1,50	2,23	2,25	2,21	2,30	2,27	2,26	0,09
15	1,75	2,45	2,43	2,32	2,36	2,36	2,35	0,13
15	2,00	2,50	2,45	2,44	2,45	2,50	2,47	0,06
15	2,25	2,65	2,62	2,55	2,63	2,62	2,59	0,10
16	1,00	1,94	1,93	1,86	1,89	1,91	1,97	0,11
16	1,25	2,09	1,97	1,97	2,09	2,02	2,07	0,12
16	1,50	2,16	2,14	2,05	2,18	2,11	2,17	0,13
16	1,75	2,26	2,31	2,16	2,29	2,24	2,21	0,15
16	2,00	2,44	2,35	2,29	2,34	2,34	2,30	0,15
16	2,25	2,50	2,50	2,37	2,49	2,44	2,44	0,13
17	1,00	1,84	1,88	1,74	1,80	1,80	1,81	0,14
17	1,25	1,95	1,95	1,83	1,91	1,89	1,96	0,13
17	1,50	2,05	2,06	1,93	2,06	2,02	2,04	0,13
17	1,75	2,18	2,15	2,03	2,07	2,11	2,08	0,15
17	2,00	2,24	2,26	2,12	2,24	2,22	2,20	0,14
17	2,25	2,41	2,33	2,22	2,34	2,33	2,34	0,19
18	1,00	1,77	1,73	1,66	1,64	1,72	1,74	0,13
18	1,25	1,87	1,88	1,75	1,79	1,81	1,83	0,13
18	1,50	2,00	1,92	1,83	1,90	1,89	1,92	0,17
18	1,75	2,12	2,00	1,92	2,00	1,98	2,03	0,20
18	2,00	2,20	2,10	2,01	2,11	2,09	2,08	0,19
18	2,25	2,23	2,16	2,10	2,18	2,19	2,24	0,14
19	1,00	1,60	1,68	1,56	1,63	1,65	1,62	0,12
19	1,25	1,82	1,73	1,64	1,66	1,72	1,70	0,18
19	1,50	1,81	1,81	1,73	1,76	1,82	1,83	0,10
19	1,75	1,94	1,89	1,81	1,91	1,93	1,82	0,13
19	2,00	2,00	1,99	1,89	2,00	1,99	1,89	0,11
19	2,25	2,06	2,03	1,98	1,99	2,05	2,07	0,09
20	1,00	1,52	1,51	1,48	1,55	1,52	1,60	0,12
20	1,25	1,59	1,65	1,55	1,57	1,66	1,69	0,14
20	1,50	1,69	1,71	1,63	1,72	1,71	1,85	0,22
20	1,75	1,75	1,77	1,71	1,69	1,77	1,86	0,17

**Tabela A.1** – Custo total do banho em R\$ para os 6 modelos da configuração B treinados com o algoritmo PPO na segunda fase do estudo.

(continua)

Temperatura ambiente em °C	Tarifa da energia elétrica (R\$)	Modelo 1	Modelo 2	Modelo 3	Modelo 4	Modelo 5	Modelo 6	Maior diferença de valores
20	2,00	1,78	1,85	1,79	1,87	1,85	1,90	0,12
20	2,25	1,93	1,94	1,86	1,88	1,94	2,02	0,16
21	1,00	1,43	1,44	1,41	1,53	1,46	1,53	0,12
21	1,25	1,49	1,50	1,48	1,43	1,53	1,63	0,20
21	1,50	1,58	1,59	1,56	1,55	1,60	1,65	0,10
21	1,75	1,65	1,65	1,62	1,63	1,66	1,63	0,04
21	2,00	1,71	1,75	1,69	1,71	1,75	1,81	0,12
21	2,25	1,76	1,78	1,77	1,78	1,81	1,66	0,15
22	1,00	1,32	1,33	1,31	1,33	1,37	1,31	0,06
22	1,25	1,41	1,39	1,37	1,41	1,43	1,46	0,09
22	1,50	1,46	1,50	1,43	1,46	1,49	1,44	0,07
22	1,75	1,52	1,52	1,51	1,47	1,57	1,46	0,11
22	2,00	1,59	1,62	1,57	1,60	1,50	1,55	0,12
22	2,25	1,67	1,67	1,64	1,65	1,71	1,56	0,15
23	1,00	1,25	1,27	1,20	1,22	1,26	1,26	0,07
23	1,25	1,27	1,32	1,26	1,27	1,29	1,36	0,10
23	1,50	1,33	1,41	1,33	1,30	1,34	1,42	0,12
23	1,75	1,41	1,47	1,39	1,41	1,40	1,39	0,08
23	2,00	1,49	1,49	1,45	1,43	1,44	1,49	0,06
23	2,25	1,53	1,58	1,51	1,54	1,48	1,48	0,10
24	1,00	1,13	1,13	1,14	1,11	1,18	1,15	0,07
24	1,25	1,19	1,23	1,20	1,20	1,23	1,24	0,05
24	1,50	1,28	1,32	1,25	1,23	1,26	1,25	0,09
24	1,75	1,32	1,33	1,31	1,34	1,33	1,29	0,05
24	2,00	1,39	1,37	1,37	1,33	1,36	1,32	0,07
24	2,25	1,42	1,48	1,43	1,37	1,42	1,45	0,11
25	1,00	1,07	1,07	1,06	1,03	1,10	1,11	0,08
25	1,25	1,16	1,10	1,11	1,11	1,14	1,17	0,07
25	1,50	1,21	1,19	1,17	1,23	1,17	1,20	0,06
25	1,75	1,22	1,18	1,22	1,22	1,25	1,29	0,11
25	2,00	1,31	1,26	1,27	1,28	1,27	1,15	0,16
25	2,25	1,34	1,32	1,32	1,37	1,33	1,38	0,06
26	1,00	1,00	0,97	0,98	1,04	1,02	1,03	0,07
26	1,25	1,05	1,04	1,03	1,06	1,09	1,02	0,07

**Tabela A.1** – Custo total do banho em R\$ para os 6 modelos da configuração B treinados com o algoritmo PPO na segunda fase do estudo.

(conclusão)

Temperatura ambiente em °C	Tarifa da energia elétrica (R\$)	Modelo 1	Modelo 2	Modelo 3	Modelo 4	Modelo 5	Modelo 6	Maior diferença de valores
26	1,50	1,09	1,13	1,10	1,13	1,12	1,11	0,04
26	1,75	1,13	1,16	1,15	1,18	1,13	1,17	0,05
26	2,00	1,21	1,20	1,21	1,21	1,20	1,21	0,01
26	2,25	1,23	1,22	1,25	1,28	1,26	1,28	0,06
27	1,00	0,93	1,06	0,91	0,95	0,90	0,94	0,16
27	1,25	0,93	0,99	0,95	1,00	0,98	0,99	0,07
27	1,50	1,01	1,04	0,99	1,02	0,98	1,00	0,06
27	1,75	1,05	1,12	1,03	1,05	1,03	1,08	0,09
27	2,00	1,07	1,15	1,07	1,13	1,11	1,14	0,08
27	2,25	1,13	1,15	1,12	1,16	1,09	1,14	0,07
28	1,00	0,84	0,83	0,83	0,85	0,86	0,91	0,08
28	1,25	0,85	0,89	0,87	0,93	0,91	0,90	0,08
28	1,50	0,88	0,94	0,91	0,95	0,94	0,93	0,07
28	1,75	0,97	1,00	0,95	0,98	0,94	1,03	0,09
28	2,00	0,99	0,99	0,99	1,06	0,99	1,05	0,07
28	2,25	1,00	1,05	1,03	1,05	1,00	1,06	0,06
29	1,00	0,69	0,76	0,78	0,80	0,73	0,84	0,15
29	1,25	0,78	0,80	0,82	0,82	0,82	0,72	0,10
29	1,50	0,79	0,84	0,83	0,88	0,80	0,92	0,13
29	1,75	0,86	0,89	0,86	0,92	0,87	0,87	0,06
29	2,00	0,84	0,84	0,91	0,88	0,88	0,93	0,09
29	2,25	0,93	0,92	0,93	1,03	0,91	1,00	0,12
30	1,00	0,61	0,65	0,71	0,74	0,70	0,65	0,13
30	1,25	0,62	0,66	0,73	0,79	0,70	0,61	0,18
30	1,50	0,68	0,71	0,76	0,83	0,73	0,65	0,18
30	1,75	0,75	0,78	0,79	0,85	0,79	0,77	0,10
30	2,00	0,81	0,77	0,82	0,87	0,80	0,84	0,10
30	2,25	0,82	0,84	0,85	0,92	0,85	0,83	0,10

**Tabela A.2** – Custo total do banho em R\$ para os 6 modelos da configuração C treinados com o algoritmo PPO na segunda fase do estudo.

(continua)

Temperatura ambiente em °C	Tarifa da energia elétrica (R\$)	Modelo 1	Modelo 2	Modelo 3	Modelo 4	Modelo 5	Modelo 6	Maior diferença de valores
15	1,00	2,38	2,27	2,29	2,37	2,06	2,08	0,32
15	1,25	2,58	2,43	2,53	2,73	2,21	2,29	0,52
15	1,50	3,06	2,54	2,71	2,91	2,35	2,45	0,71
15	1,75	3,40	2,67	3,02	3,31	2,49	2,54	0,91
15	2,00	3,69	2,68	3,26	3,38	2,62	2,59	1,07
15	2,25	3,82	3,04	3,50	3,77	2,75	2,71	1,07
16	1,00	2,21	1,98	2,20	2,34	1,96	2,00	0,38
16	1,25	2,51	2,29	2,45	2,70	2,08	2,09	0,62
16	1,50	2,73	2,38	2,69	2,72	2,21	2,31	0,52
16	1,75	3,22	2,46	2,94	2,91	2,33	2,37	0,89
16	2,00	3,49	2,44	3,18	3,20	2,44	2,39	1,05
16	2,25	3,78	2,54	3,43	3,62	2,57	2,60	1,24
17	1,00	2,35	1,92	2,20	2,22	1,86	1,81	0,49
17	1,25	2,45	2,02	2,47	2,55	1,97	2,01	0,58
17	1,50	2,69	2,25	2,74	2,83	2,08	2,05	0,75
17	1,75	3,14	2,26	3,01	3,02	2,21	2,15	0,93
17	2,00	3,35	2,44	3,29	3,39	2,29	2,14	1,10
17	2,25	3,76	2,48	3,53	3,47	2,45	2,48	1,31
18	1,00	2,11	1,90	2,13	2,23	1,75	1,95	0,48
18	1,25	2,25	2,19	2,34	2,59	1,87	1,99	0,72
18	1,50	2,87	2,19	2,67	2,81	1,94	2,09	0,93
18	1,75	2,99	2,13	2,95	3,15	2,06	2,15	1,09
18	2,00	3,33	2,27	3,22	3,51	2,14	2,19	1,37
18	2,25	3,36	2,45	3,49	3,86	2,26	2,22	1,60
19	1,00	2,05	1,62	2,05	2,14	1,68	1,91	0,52
19	1,25	2,16	1,94	2,31	2,37	1,77	2,00	0,60
19	1,50	2,59	1,89	2,58	2,78	1,83	1,96	0,95
19	1,75	2,69	2,31	2,77	2,91	1,91	1,99	1,00
19	2,00	3,24	2,32	3,03	3,10	2,04	2,30	1,20
19	2,25	3,26	2,27	3,28	3,40	2,15	2,23	1,25
20	1,00	1,85	1,69	1,99	1,90	1,55	1,76	0,44
20	1,25	2,28	1,80	2,20	2,19	1,64	1,83	0,64
20	1,50	2,39	1,81	2,45	2,21	1,71	2,15	0,74
20	1,75	2,82	1,97	2,70	2,50	1,79	1,76	1,03



**Tabela A.2** – Custo total do banho em R\$ para os 6 modelos da configuração C treinados com o algoritmo PPO na segunda fase do estudo.

(continua)

Temperatura ambiente em °C	Tarifa da energia elétrica (R\$)	Modelo 1	Modelo 2	Modelo 3	Modelo 4	Modelo 5	Modelo 6	Maior diferença de valores
20	2,00	2,93	2,13	2,95	2,91	1,89	1,95	1,06
20	2,25	2,97	2,42	3,14	3,08	2,05	2,13	1,09
21	1,00	1,74	1,63	1,80	1,80	1,45	1,60	0,35
21	1,25	2,19	1,74	2,05	2,08	1,55	1,75	0,64
21	1,50	2,32	1,76	2,28	2,24	1,61	1,93	0,71
21	1,75	2,63	1,79	2,52	2,25	1,66	1,59	0,97
21	2,00	3,00	1,93	2,75	2,58	1,84	1,73	1,16
21	2,25	3,15	1,91	2,94	3,17	1,80	1,95	1,37
22	1,00	1,62	1,48	1,62	1,62	1,46	1,42	0,16
22	1,25	2,01	1,50	1,85	1,92	1,43	1,57	0,58
22	1,50	2,21	1,56	1,92	2,09	1,43	1,69	0,78
22	1,75	2,45	1,81	2,18	2,37	1,49	1,69	0,96
22	2,00	2,86	1,92	2,40	2,72	1,54	1,89	1,32
22	2,25	2,63	1,87	2,66	2,86	1,69	1,57	1,17
23	1,00	1,54	1,35	1,50	1,53	1,25	1,35	0,29
23	1,25	1,76	1,36	1,69	1,73	1,29	1,59	0,47
23	1,50	2,04	1,47	1,87	1,70	1,33	1,58	0,71
23	1,75	2,17	1,44	2,11	2,07	1,43	1,35	0,74
23	2,00	2,49	1,89	2,24	2,54	1,50	1,71	1,04
23	2,25	2,64	1,65	2,43	2,76	1,48	1,75	1,28
24	1,00	1,49	1,24	1,44	1,45	1,12	1,23	0,37
24	1,25	1,72	1,43	1,64	1,65	1,24	1,37	0,48
24	1,50	1,92	1,32	1,81	1,80	1,28	1,56	0,64
24	1,75	2,29	1,34	2,00	2,02	1,36	1,29	0,95
24	2,00	2,35	1,46	2,18	1,91	1,29	1,44	1,06
24	2,25	2,56	1,47	2,37	2,68	1,36	1,52	1,32
25	1,00	1,36	1,21	1,39	1,39	1,07	1,09	0,32
25	1,25	1,70	1,25	1,56	1,51	1,11	1,05	0,59
25	1,50	1,90	1,25	1,74	1,81	1,14	1,26	0,76
25	1,75	1,89	1,66	1,91	2,10	1,18	1,31	0,92
25	2,00	2,22	1,62	2,09	2,09	1,37	1,22	0,85
25	2,25	2,55	1,54	2,27	2,47	1,32	1,27	1,23
26	1,00	1,42	1,10	1,33	1,42	1,03	1,06	0,39
26	1,25	1,58	1,30	1,59	1,45	1,03	1,17	0,56

**Tabela A.2** – Custo total do banho em R\$ para os 6 modelos da configuração C treinados com o algoritmo PPO na segunda fase do estudo.

(conclusão)

Temperatura ambiente em °C	Tarifa da energia elétrica (R\$)	Modelo 1	Modelo 2	Modelo 3	Modelo 4	Modelo 5	Modelo 6	Maior diferença de valores
26	1,50	1,72	1,08	1,67	1,61	1,13	1,24	0,64
26	1,75	1,98	1,41	1,98	1,68	1,19	1,17	0,79
26	2,00	2,12	1,26	2,18	2,03	1,23	1,20	0,95
26	2,25	2,30	1,33	2,37	2,29	1,30	1,27	1,07
27	1,00	1,30	1,07	1,42	1,37	0,95	1,03	0,47
27	1,25	1,33	1,05	1,65	1,53	0,96	1,14	0,69
27	1,50	1,55	1,29	1,86	1,75	1,10	1,44	0,76
27	1,75	1,84	1,19	2,09	2,15	1,09	1,13	1,06
27	2,00	1,97	1,20	2,30	2,26	1,11	1,12	1,19
27	2,25	2,27	1,19	2,53	2,50	1,17	1,24	1,36
28	1,00	1,17	0,95	1,34	1,25	0,89	1,04	0,45
28	1,25	1,37	0,97	1,52	1,49	0,90	1,29	0,62
28	1,50	1,65	1,09	1,73	1,73	0,93	1,56	0,80
28	1,75	1,86	1,39	1,93	1,88	0,89	0,94	1,04
28	2,00	2,00	1,37	2,16	2,20	1,08	1,03	1,12
28	2,25	2,38	1,16	2,33	2,32	1,15	1,27	1,23
29	1,00	1,10	0,97	1,22	1,22	0,79	1,05	0,43
29	1,25	1,33	1,01	1,40	1,34	0,85	1,25	0,55
29	1,50	1,52	0,98	1,58	1,48	0,71	1,33	0,87
29	1,75	1,71	1,21	1,77	1,67	0,89	0,84	0,88
29	2,00	1,67	1,07	1,95	1,89	0,74	1,05	1,21
29	2,25	2,06	1,32	2,13	2,09	1,07	0,81	1,06
30	1,00	0,98	0,81	1,10	1,05	0,71	1,00	0,39
30	1,25	1,23	0,89	1,26	1,25	0,69	1,16	0,57
30	1,50	1,34	0,92	1,42	1,34	0,70	1,15	0,72
30	1,75	1,43	0,89	1,57	1,48	0,72	0,88	0,85
30	2,00	1,62	0,92	1,73	1,56	0,83	0,65	0,90
30	2,25	1,74	1,27	1,89	1,97	0,74	0,75	1,23

**Tabela A.3** – Custo elétrico total em R\$ para os 6 modelos da configuração B treinados com o algoritmo PPO na segunda fase do estudo.

(continua)

Temperatura ambiente em °C	Tarifa da energia elétrica (R\$)	Modelo 1	Modelo 2	Modelo 3	Modelo 4	Modelo 5	Modelo 6	Maior diferença de valores
15	1,00	0,48	0,47	0,45	0,48	0,47	0,47	0,03
15	1,25	0,59	0,60	0,57	0,60	0,59	0,59	0,03
15	1,50	0,70	0,72	0,68	0,72	0,70	0,72	0,04
15	1,75	0,85	0,83	0,80	0,83	0,82	0,82	0,05
15	2,00	0,95	0,91	0,91	0,94	0,94	0,95	0,04
15	2,25	1,07	1,04	1,02	1,06	1,05	1,07	0,05
16	1,00	0,45	0,44	0,42	0,45	0,44	0,45	0,03
16	1,25	0,56	0,50	0,53	0,56	0,55	0,55	0,06
16	1,50	0,67	0,65	0,63	0,67	0,66	0,67	0,04
16	1,75	0,79	0,77	0,74	0,78	0,76	0,77	0,05
16	2,00	0,90	0,86	0,84	0,85	0,87	0,87	0,06
16	2,25	1,01	0,99	0,95	1,01	0,98	0,95	0,06
17	1,00	0,42	0,40	0,39	0,42	0,40	0,41	0,03
17	1,25	0,53	0,51	0,48	0,51	0,51	0,51	0,05
17	1,50	0,63	0,60	0,58	0,62	0,61	0,62	0,05
17	1,75	0,74	0,71	0,67	0,67	0,71	0,71	0,07
17	2,00	0,83	0,81	0,77	0,84	0,81	0,81	0,07
17	2,25	0,95	0,90	0,87	0,95	0,91	0,91	0,08
18	1,00	0,39	0,38	0,35	0,37	0,37	0,39	0,04
18	1,25	0,48	0,47	0,44	0,47	0,46	0,45	0,04
18	1,50	0,58	0,56	0,53	0,57	0,56	0,57	0,05
18	1,75	0,69	0,64	0,62	0,67	0,65	0,69	0,07
18	2,00	0,78	0,71	0,71	0,76	0,74	0,73	0,07
18	2,25	0,86	0,80	0,80	0,86	0,83	0,86	0,06
19	1,00	0,33	0,34	0,33	0,35	0,33	0,33	0,02
19	1,25	0,45	0,42	0,42	0,43	0,42	0,43	0,03
19	1,50	0,50	0,50	0,50	0,50	0,50	0,52	0,02
19	1,75	0,58	0,58	0,58	0,61	0,58	0,58	0,03
19	2,00	0,68	0,66	0,66	0,70	0,66	0,67	0,04
19	2,25	0,76	0,75	0,75	0,77	0,75	0,80	0,05
20	1,00	0,31	0,31	0,31	0,31	0,31	0,31	0,00
20	1,25	0,39	0,39	0,39	0,38	0,39	0,42	0,04
20	1,50	0,47	0,46	0,47	0,46	0,46	0,49	0,03
20	1,75	0,54	0,54	0,54	0,55	0,54	0,58	0,04

**Tabela A.3** – Custo elétrico total em R\$ para os 6 modelos da configuração B treinados com o algoritmo PPO na segunda fase do estudo.

(continua)

Temperatura ambiente em °C	Tarifa da energia elétrica (R\$)	Modelo 1	Modelo 2	Modelo 3	Modelo 4	Modelo 5	Modelo 6	Maior diferença de valores
20	2,00	0,62	0,62	0,62	0,65	0,62	0,64	0,03
20	2,25	0,69	0,70	0,70	0,70	0,70	0,75	0,06
21	1,00	0,29	0,29	0,29	0,32	0,29	0,29	0,03
21	1,25	0,36	0,36	0,36	0,36	0,36	0,36	0,00
21	1,50	0,43	0,43	0,43	0,43	0,43	0,43	0,00
21	1,75	0,51	0,51	0,51	0,51	0,50	0,33	0,18
21	2,00	0,58	0,58	0,58	0,58	0,58	0,58	0,00
21	2,25	0,65	0,65	0,65	0,65	0,65	0,41	0,24
22	1,00	0,27	0,27	0,27	0,27	0,27	0,18	0,09
22	1,25	0,34	0,34	0,34	0,34	0,33	0,34	0,01
22	1,50	0,40	0,40	0,40	0,40	0,40	0,27	0,13
22	1,75	0,47	0,47	0,47	0,47	0,47	0,32	0,15
22	2,00	0,54	0,54	0,54	0,54	0,36	0,36	0,18
22	2,25	0,61	0,61	0,61	0,60	0,60	0,41	0,20
23	1,00	0,25	0,25	0,25	0,25	0,18	0,18	0,07
23	1,25	0,31	0,31	0,31	0,31	0,23	0,23	0,08
23	1,50	0,37	0,37	0,37	0,37	0,27	0,37	0,10
23	1,75	0,43	0,44	0,44	0,44	0,32	0,32	0,12
23	2,00	0,50	0,50	0,50	0,50	0,36	0,36	0,14
23	2,25	0,56	0,56	0,56	0,56	0,41	0,41	0,15
24	1,00	0,23	0,18	0,23	0,23	0,18	0,18	0,05
24	1,25	0,28	0,29	0,29	0,29	0,23	0,23	0,06
24	1,50	0,34	0,34	0,34	0,35	0,27	0,27	0,08
24	1,75	0,40	0,40	0,40	0,40	0,32	0,32	0,08
24	2,00	0,46	0,36	0,46	0,46	0,36	0,36	0,10
24	2,25	0,52	0,52	0,52	0,52	0,41	0,41	0,11
25	1,00	0,21	0,18	0,21	0,18	0,18	0,18	0,03
25	1,25	0,26	0,23	0,26	0,26	0,23	0,23	0,03
25	1,50	0,32	0,27	0,32	0,35	0,27	0,32	0,08
25	1,75	0,37	0,32	0,37	0,37	0,32	0,40	0,08
25	2,00	0,42	0,36	0,42	0,42	0,36	0,16	0,26
25	2,25	0,48	0,41	0,48	0,48	0,41	0,48	0,07
26	1,00	0,19	0,18	0,19	0,20	0,19	0,20	0,02
26	1,25	0,23	0,23	0,23	0,24	0,23	0,23	0,01

**Tabela A.3** – Custo elétrico total em R\$ para os 6 modelos da configuração B treinados com o algoritmo PPO na segunda fase do estudo.

(conclusão)

Temperatura ambiente em °C	Tarifa da energia elétrica (R\$)	Modelo 1	Modelo 2	Modelo 3	Modelo 4	Modelo 5	Modelo 6	Maior diferença de valores
26	1,50	0,29	0,27	0,29	0,27	0,28	0,29	0,02
26	1,75	0,34	0,32	0,34	0,34	0,33	0,34	0,02
26	2,00	0,38	0,36	0,39	0,39	0,38	0,40	0,04
26	2,25	0,43	0,41	0,44	0,41	0,42	0,41	0,03
27	1,00	0,18	0,25	0,17	0,18	0,17	0,18	0,08
27	1,25	0,22	0,22	0,21	0,22	0,21	0,22	0,01
27	1,50	0,26	0,27	0,25	0,27	0,25	0,27	0,02
27	1,75	0,31	0,34	0,29	0,31	0,29	0,30	0,05
27	2,00	0,35	0,35	0,33	0,36	0,33	0,35	0,03
27	2,25	0,40	0,40	0,37	0,40	0,37	0,39	0,03
28	1,00	0,15	0,16	0,15	0,16	0,15	0,16	0,01
28	1,25	0,19	0,20	0,19	0,20	0,19	0,22	0,03
28	1,50	0,23	0,23	0,23	0,24	0,23	0,24	0,01
28	1,75	0,25	0,27	0,26	0,27	0,27	0,29	0,04
28	2,00	0,31	0,32	0,30	0,32	0,31	0,32	0,02
28	2,25	0,34	0,35	0,34	0,36	0,34	0,36	0,02
29	1,00	0,14	0,14	0,14	0,14	0,14	0,14	0,00
29	1,25	0,17	0,17	0,17	0,18	0,17	0,00	0,18
29	1,50	0,20	0,21	0,20	0,21	0,21	0,23	0,03
29	1,75	0,24	0,24	0,24	0,25	0,24	0,24	0,01
29	2,00	0,27	0,27	0,27	0,28	0,28	0,28	0,01
29	2,25	0,31	0,31	0,31	0,32	0,31	0,32	0,01
30	1,00	0,13	0,12	0,12	0,12	0,12	0,00	0,13
30	1,25	0,01	0,16	0,15	0,16	0,15	0,00	0,16
30	1,50	0,19	0,19	0,18	0,22	0,19	0,00	0,22
30	1,75	0,22	0,21	0,21	0,22	0,22	0,22	0,01
30	2,00	0,25	0,26	0,24	0,25	0,25	0,24	0,02
30	2,25	0,28	0,27	0,27	0,28	0,28	0,27	0,01

**Tabela A.4** – Custo elétrico total em R\$ para os 6 modelos da configuração C treinados com o algoritmo PPO na segunda fase do estudo.

(continua)

Temperatura ambiente em °C	Tarifa da energia elétrica (R\$)	Modelo 1	Modelo 2	Modelo 3	Modelo 4	Modelo 5	Modelo 6	Maior diferença de valores
15	1,00	1,17	0,64	0,97	0,92	0,54	0,61	0,63
15	1,25	1,31	0,81	1,22	1,45	0,68	0,80	0,77
15	1,50	1,82	0,90	1,39	1,43	0,81	0,89	1,01
15	1,75	2,18	1,07	1,70	2,03	0,96	0,93	1,25
15	2,00	2,50	1,02	1,95	2,00	1,09	1,00	1,50
15	2,25	2,53	1,38	2,19	2,44	1,22	1,12	1,41
16	1,00	0,98	0,47	0,98	1,15	0,50	0,60	0,68
16	1,25	1,32	0,77	1,23	1,53	0,62	0,62	0,91
16	1,50	1,52	0,86	1,47	1,44	0,75	0,90	0,77
16	1,75	2,09	0,83	1,72	1,53	0,87	0,85	1,26
16	2,00	2,33	0,87	1,96	1,88	0,99	0,84	1,49
16	2,25	2,61	1,03	2,21	2,32	1,11	1,13	1,58
17	1,00	1,26	0,43	1,09	0,99	0,45	0,43	0,83
17	1,25	1,35	0,52	1,36	1,29	0,56	0,60	0,84
17	1,50	1,57	0,82	1,64	1,68	0,67	0,65	1,03
17	1,75	2,09	0,79	1,91	1,76	0,79	0,65	1,44
17	2,00	2,22	0,96	2,18	2,16	0,88	0,74	1,48
17	2,25	2,68	0,97	2,40	2,23	1,02	1,08	1,71
18	1,00	1,05	0,58	1,09	1,15	0,41	0,71	0,74
18	1,25	1,14	0,92	1,26	1,51	0,54	0,75	0,97
18	1,50	1,86	0,85	1,63	1,69	0,61	0,77	1,25
18	1,75	1,95	0,69	1,90	2,07	0,72	0,68	1,39
18	2,00	2,32	0,85	2,17	2,48	0,80	0,74	1,74
18	2,25	2,30	1,03	2,45	2,81	0,91	0,84	1,97
19	1,00	1,07	0,39	1,06	1,19	0,41	0,82	0,80
19	1,25	1,17	0,59	1,33	1,36	0,50	0,83	0,86
19	1,50	1,62	0,57	1,59	1,86	0,56	0,68	1,30
19	1,75	1,65	0,90	1,77	1,92	0,63	0,67	1,29
19	2,00	2,28	1,02	2,03	2,01	0,74	1,05	1,54
19	2,25	2,31	0,93	2,28	2,35	0,84	0,89	1,51
20	1,00	0,94	0,42	1,05	0,93	0,31	0,62	0,74
20	1,25	1,41	0,47	1,26	1,24	0,40	0,71	1,01
20	1,50	1,50	0,50	1,51	1,20	0,46	1,09	1,05
20	1,75	2,01	0,67	1,76	1,51	0,54	0,60	1,47

**Tabela A.4** – Custo elétrico total em R\$ para os 6 modelos da configuração C treinados com o algoritmo PPO na segunda fase do estudo.

(continua)

Temperatura ambiente em °C	Tarifa da energia elétrica (R\$)	Modelo 1	Modelo 2	Modelo 3	Modelo 4	Modelo 5	Modelo 6	Maior diferença de valores
20	2,00	1,99	0,82	2,01	1,98	0,64	0,66	1,37
20	2,25	2,04	1,21	2,20	2,13	0,85	0,89	1,35
21	1,00	0,93	0,47	0,93	0,96	0,31	0,59	0,65
21	1,25	1,41	0,60	1,17	1,26	0,36	0,68	1,05
21	1,50	1,54	0,52	1,41	1,39	0,41	0,95	1,13
21	1,75	1,85	0,51	1,64	1,28	0,52	0,46	1,39
21	2,00	2,26	0,73	1,88	1,68	0,69	0,52	1,74
21	2,25	2,39	0,75	2,07	2,27	0,61	0,84	1,78
22	1,00	0,86	0,38	0,80	0,74	0,37	0,37	0,49
22	1,25	1,32	0,43	1,02	1,08	0,36	0,50	0,96
22	1,50	1,46	0,45	1,12	1,27	0,36	0,65	1,10
22	1,75	1,67	0,75	1,38	1,57	0,42	0,56	1,25
22	2,00	2,19	0,88	1,59	1,88	0,44	0,83	1,75
22	2,25	1,85	0,80	1,86	2,00	0,62	0,46	1,54
23	1,00	0,78	0,38	0,74	0,76	0,19	0,36	0,59
23	1,25	1,00	0,32	0,92	0,87	0,24	0,61	0,76
23	1,50	1,32	0,43	1,11	0,74	0,32	0,64	1,00
23	1,75	1,45	0,40	1,38	1,26	0,43	0,38	1,07
23	2,00	1,81	0,97	1,51	1,80	0,48	0,73	1,33
23	2,25	1,90	0,59	1,70	2,00	0,45	0,76	1,55
24	1,00	0,84	0,30	0,74	0,70	0,22	0,34	0,62
24	1,25	1,08	0,51	0,92	0,94	0,30	0,54	0,78
24	1,50	1,25	0,38	1,11	1,02	0,35	0,67	0,90
24	1,75	1,64	0,33	1,30	1,23	0,45	0,34	1,31
24	2,00	1,66	0,49	1,48	1,10	0,41	0,55	1,25
24	2,25	1,91	0,49	1,65	2,00	0,44	0,59	1,56
25	1,00	0,72	0,39	0,70	0,65	0,21	0,28	0,51
25	1,25	1,08	0,38	0,87	0,78	0,25	0,24	0,84
25	1,50	1,28	0,39	1,06	1,15	0,29	0,48	0,99
25	1,75	1,27	0,73	1,23	1,43	0,34	0,49	1,09
25	2,00	1,61	0,75	1,41	1,42	0,52	0,39	1,22
25	2,25	1,92	0,64	1,58	1,87	0,43	0,43	1,49
26	1,00	0,86	0,35	0,68	0,82	0,25	0,27	0,61
26	1,25	1,02	0,44	0,98	0,71	0,24	0,36	0,78

**Tabela A.4** – Custo elétrico total em R\$ para os 6 modelos da configuração C treinados com o algoritmo PPO na segunda fase do estudo.

(conclusão)

Temperatura ambiente em °C	Tarifa da energia elétrica (R\$)	Modelo 1	Modelo 2	Modelo 3	Modelo 4	Modelo 5	Modelo 6	Maior diferença de valores
26	1,50	1,14	0,29	1,02	0,93	0,32	0,53	0,85
26	1,75	1,42	0,67	1,37	1,02	0,39	0,40	1,03
26	2,00	1,50	0,48	1,57	1,37	0,44	0,42	1,15
26	2,25	1,72	0,46	1,76	1,65	0,50	0,44	1,32
27	1,00	0,79	0,30	0,88	0,75	0,22	0,32	0,66
27	1,25	0,78	0,28	1,10	1,00	0,24	0,50	0,86
27	1,50	1,02	0,60	1,32	1,21	0,31	0,82	1,01
27	1,75	1,28	0,48	1,54	1,58	0,34	0,43	1,24
27	2,00	1,42	0,48	1,76	1,73	0,38	0,39	1,38
27	2,25	1,72	0,42	1,98	1,96	0,43	0,45	1,56
28	1,00	0,68	0,28	0,82	0,75	0,21	0,50	0,61
28	1,25	0,88	0,32	1,01	0,98	0,24	0,68	0,77
28	1,50	1,16	0,45	1,21	1,20	0,23	1,06	0,98
28	1,75	1,38	0,62	1,41	1,38	0,12	0,24	1,29
28	2,00	1,51	0,64	1,63	1,68	0,38	0,33	1,35
28	2,25	1,87	0,54	1,81	1,78	0,43	0,62	1,44
29	1,00	0,64	0,34	0,73	0,71	0,15	0,55	0,58
29	1,25	0,89	0,48	0,92	0,83	0,24	0,76	0,68
29	1,50	1,08	0,39	1,10	1,00	0,05	0,83	1,05
29	1,75	1,26	0,51	1,28	1,19	0,26	0,16	1,12
29	2,00	1,22	0,36	1,47	1,38	0,08	0,45	1,39
29	2,25	1,60	0,61	1,65	1,60	0,40	0,17	1,48
30	1,00	0,56	0,25	0,63	0,61	0,12	0,53	0,51
30	1,25	0,79	0,26	0,79	0,77	0,09	0,69	0,70
30	1,50	0,92	0,32	0,95	0,86	0,09	0,69	0,86
30	1,75	1,00	0,25	1,11	1,00	0,11	0,31	1,00
30	2,00	1,20	0,31	1,27	1,14	0,24	0,07	1,20
30	2,25	1,30	0,78	1,43	1,51	0,13	0,16	1,38



**Tabela A.5** – Custo de gás total em R\$ para os 6 modelos da configuração B treinados com o algoritmo PPO na segunda fase do estudo.

(continua)

Temperatura ambiente em °C	Tarifa da energia elétrica (R\$)	Modelo 1	Modelo 2	Modelo 3	Modelo 4	Modelo 5	Modelo 6	Maior diferença de valores
15	1,00	1,18	1,19	1,16	1,15	1,21	1,15	0,06
15	1,25	1,18	1,21	1,15	1,17	1,16	1,16	0,06
15	1,50	1,16	1,16	1,16	1,20	1,19	1,17	0,04
15	1,75	1,21	1,21	1,16	1,16	1,17	1,16	0,05
15	2,00	1,17	1,17	1,16	1,14	1,18	1,15	0,04
15	2,25	1,19	1,20	1,16	1,18	1,19	1,15	0,05
16	1,00	1,12	1,12	1,09	1,09	1,11	1,15	0,06
16	1,25	1,15	1,13	1,09	1,15	1,11	1,14	0,06
16	1,50	1,11	1,13	1,07	1,13	1,09	1,13	0,06
16	1,75	1,11	1,16	1,07	1,13	1,11	1,08	0,09
16	2,00	1,15	1,13	1,10	1,13	1,11	1,08	0,07
16	2,25	1,11	1,13	1,08	1,11	1,10	1,12	0,05
17	1,00	1,06	1,11	1,02	1,04	1,05	1,04	0,09
17	1,25	1,06	1,08	1,02	1,05	1,04	1,08	0,06
17	1,50	1,06	1,09	1,02	1,07	1,06	1,06	0,07
17	1,75	1,07	1,07	1,02	1,06	1,05	1,02	0,05
17	2,00	1,05	1,09	1,02	1,04	1,05	1,04	0,07
17	2,25	1,09	1,07	1,02	1,04	1,06	1,07	0,07
18	1,00	1,02	1,00	0,98	0,96	1,01	1,00	0,06
18	1,25	1,03	1,05	0,98	0,98	1,00	1,02	0,07
18	1,50	1,05	1,01	0,98	0,99	1,00	1,00	0,07
18	1,75	1,06	1,02	0,98	0,99	0,99	0,99	0,08
18	2,00	1,05	1,04	0,98	1,00	1,01	1,01	0,07
18	2,25	1,02	1,02	0,98	0,99	1,02	1,03	0,05
19	1,00	0,94	0,99	0,92	0,94	0,97	0,96	0,07
19	1,25	1,01	0,97	0,92	0,91	0,96	0,94	0,10
19	1,50	0,97	0,97	0,92	0,93	0,98	0,97	0,06
19	1,75	1,00	0,97	0,92	0,95	1,00	0,92	0,08
19	2,00	0,98	0,98	0,92	0,97	0,98	0,91	0,07
19	2,25	0,96	0,95	0,92	0,91	0,97	0,95	0,06
20	1,00	0,89	0,88	0,86	0,92	0,90	0,95	0,09
20	1,25	0,89	0,93	0,86	0,88	0,94	0,93	0,08
20	1,50	0,90	0,91	0,86	0,92	0,92	0,99	0,13
20	1,75	0,89	0,91	0,86	0,85	0,91	0,93	0,08

**Tabela A.5** – Custo de gás total em R\$ para os 6 modelos da configuração B treinados com o algoritmo PPO na segunda fase do estudo.

(continua)

Temperatura ambiente em °C	Tarifa da energia elétrica (R\$)	Modelo 1	Modelo 2	Modelo 3	Modelo 4	Modelo 5	Modelo 6	Maior diferença de valores
20	2,00	0,86	0,90	0,86	0,90	0,91	0,93	0,07
20	2,25	0,91	0,91	0,86	0,88	0,92	0,93	0,07
21	1,00	0,83	0,84	0,82	0,86	0,86	0,90	0,08
21	1,25	0,82	0,83	0,82	0,78	0,86	0,92	0,14
21	1,50	0,84	0,84	0,82	0,82	0,85	0,89	0,07
21	1,75	0,84	0,84	0,82	0,82	0,84	0,95	0,13
21	2,00	0,83	0,85	0,82	0,83	0,86	0,89	0,07
21	2,25	0,81	0,82	0,82	0,82	0,85	0,92	0,11
22	1,00	0,76	0,77	0,75	0,77	0,80	0,83	0,08
22	1,25	0,78	0,76	0,75	0,78	0,79	0,81	0,06
22	1,50	0,76	0,80	0,75	0,77	0,79	0,85	0,10
22	1,75	0,76	0,76	0,75	0,73	0,80	0,83	0,10
22	2,00	0,76	0,78	0,75	0,77	0,83	0,86	0,11
22	2,25	0,77	0,77	0,75	0,76	0,80	0,84	0,09
23	1,00	0,71	0,73	0,68	0,70	0,78	0,77	0,10
23	1,25	0,68	0,72	0,68	0,69	0,77	0,82	0,14
23	1,50	0,68	0,74	0,68	0,67	0,77	0,74	0,10
23	1,75	0,70	0,74	0,68	0,70	0,78	0,78	0,10
23	2,00	0,71	0,70	0,68	0,67	0,78	0,81	0,14
23	2,25	0,69	0,73	0,68	0,71	0,77	0,77	0,09
24	1,00	0,64	0,67	0,65	0,62	0,71	0,69	0,09
24	1,25	0,64	0,66	0,65	0,64	0,72	0,71	0,08
24	1,50	0,66	0,68	0,65	0,63	0,71	0,69	0,08
24	1,75	0,65	0,66	0,65	0,66	0,72	0,69	0,07
24	2,00	0,65	0,72	0,65	0,62	0,71	0,68	0,10
24	2,25	0,64	0,67	0,65	0,61	0,72	0,74	0,13
25	1,00	0,60	0,62	0,59	0,60	0,64	0,65	0,06
25	1,25	0,62	0,61	0,59	0,59	0,64	0,66	0,07
25	1,50	0,62	0,64	0,59	0,60	0,63	0,61	0,05
25	1,75	0,60	0,61	0,59	0,60	0,66	0,62	0,07
25	2,00	0,62	0,63	0,59	0,60	0,64	0,70	0,11
25	2,25	0,60	0,64	0,59	0,62	0,65	0,63	0,06
26	1,00	0,56	0,54	0,56	0,58	0,58	0,57	0,04
26	1,25	0,57	0,56	0,56	0,56	0,59	0,55	0,04

**Tabela A.5** – Custo de gás total em R\$ para os 6 modelos da configuração B treinados com o algoritmo PPO na segunda fase do estudo.

(conclusão)

Temperatura ambiente em °C	Tarifa da energia elétrica (R\$)	Modelo 1	Modelo 2	Modelo 3	Modelo 4	Modelo 5	Modelo 6	Maior diferença de valores
26	1,50	0,55	0,59	0,56	0,59	0,58	0,56	0,04
26	1,75	0,55	0,58	0,56	0,58	0,56	0,57	0,03
26	2,00	0,57	0,58	0,56	0,56	0,57	0,55	0,03
26	2,25	0,55	0,56	0,56	0,59	0,58	0,60	0,05
27	1,00	0,51	0,53	0,51	0,52	0,51	0,52	0,02
27	1,25	0,48	0,52	0,51	0,52	0,53	0,52	0,05
27	1,50	0,50	0,52	0,51	0,51	0,50	0,50	0,02
27	1,75	0,50	0,52	0,51	0,50	0,51	0,53	0,03
27	2,00	0,49	0,54	0,51	0,52	0,53	0,53	0,05
27	2,25	0,49	0,51	0,51	0,51	0,49	0,51	0,02
28	1,00	0,45	0,45	0,46	0,46	0,48	0,49	0,04
28	1,25	0,45	0,46	0,46	0,48	0,49	0,45	0,04
28	1,50	0,44	0,46	0,47	0,47	0,48	0,46	0,04
28	1,75	0,48	0,48	0,47	0,47	0,46	0,48	0,02
28	2,00	0,45	0,45	0,47	0,49	0,47	0,48	0,04
28	2,25	0,44	0,46	0,47	0,46	0,45	0,46	0,03
29	1,00	0,35	0,41	0,43	0,42	0,40	0,45	0,10
29	1,25	0,41	0,42	0,43	0,41	0,43	0,48	0,07
29	1,50	0,39	0,42	0,42	0,44	0,40	0,44	0,05
29	1,75	0,41	0,43	0,42	0,44	0,42	0,42	0,03
29	2,00	0,37	0,38	0,42	0,40	0,40	0,42	0,05
29	2,25	0,41	0,41	0,42	0,46	0,40	0,44	0,06
30	1,00	0,29	0,35	0,38	0,39	0,37	0,42	0,13
30	1,25	0,41	0,33	0,37	0,40	0,36	0,41	0,08
30	1,50	0,32	0,33	0,37	0,38	0,36	0,43	0,11
30	1,75	0,34	0,36	0,37	0,39	0,37	0,36	0,05
30	2,00	0,36	0,33	0,37	0,39	0,36	0,38	0,06
30	2,25	0,35	0,36	0,37	0,40	0,37	0,36	0,05

**Tabela A.6** – Custo de gás total em R\$ para os 6 modelos da configuração C treinados com o algoritmo PPO na segunda fase do estudo.

(continua)

Temperatura ambiente em °C	Tarifa da energia elétrica (R\$)	Modelo 1	Modelo 2	Modelo 3	Modelo 4	Modelo 5	Modelo 6	Maior diferença de valores
15	1,00	0,84	1,21	0,95	1,04	1,14	1,09	0,37
15	1,25	0,90	1,20	0,95	0,91	1,15	1,11	0,30
15	1,50	0,86	1,22	0,96	1,08	1,15	1,17	0,36
15	1,75	0,84	1,19	0,95	0,89	1,15	1,21	0,37
15	2,00	0,82	1,25	0,95	1,00	1,15	1,20	0,43
15	2,25	0,90	1,23	0,95	0,96	1,15	1,20	0,33
16	1,00	0,88	1,13	0,86	0,82	1,09	1,04	0,31
16	1,25	0,83	1,13	0,86	0,80	1,09	1,10	0,33
16	1,50	0,85	1,13	0,86	0,91	1,09	1,04	0,28
16	1,75	0,77	1,22	0,86	0,99	1,09	1,13	0,45
16	2,00	0,80	1,18	0,86	0,94	1,09	1,16	0,38
16	2,25	0,80	1,13	0,86	0,92	1,09	1,10	0,33
17	1,00	0,72	1,11	0,76	0,86	1,05	1,03	0,39
17	1,25	0,75	1,12	0,76	0,88	1,05	1,05	0,37
17	1,50	0,77	1,05	0,76	0,79	1,05	1,05	0,29
17	1,75	0,71	1,09	0,76	0,87	1,05	1,12	0,41
17	2,00	0,77	1,10	0,76	0,84	1,05	1,05	0,34
17	2,25	0,72	1,12	0,78	0,87	1,06	1,04	0,40
18	1,00	0,71	0,96	0,70	0,71	0,99	0,88	0,29
18	1,25	0,77	0,90	0,74	0,71	0,98	0,90	0,27
18	1,50	0,66	0,98	0,70	0,76	0,99	0,96	0,33
18	1,75	0,69	1,07	0,70	0,71	0,99	1,09	0,40
18	2,00	0,66	1,05	0,70	0,67	0,99	1,07	0,41
18	2,25	0,72	1,04	0,70	0,68	1,00	1,03	0,36
19	1,00	0,65	0,91	0,65	0,61	0,93	0,75	0,32
19	1,25	0,67	0,98	0,65	0,67	0,94	0,83	0,33
19	1,50	0,63	0,97	0,65	0,59	0,94	0,93	0,38
19	1,75	0,71	1,00	0,67	0,66	0,94	0,97	0,34
19	2,00	0,62	0,94	0,67	0,72	0,95	0,91	0,33
19	2,25	0,63	0,98	0,67	0,69	0,97	0,98	0,35
20	1,00	0,60	0,92	0,60	0,64	0,91	0,80	0,32
20	1,25	0,54	0,96	0,61	0,62	0,91	0,79	0,42
20	1,50	0,58	0,95	0,61	0,69	0,91	0,73	0,37
20	1,75	0,51	0,94	0,61	0,67	0,92	0,85	0,43

**Tabela A.6** – Custo de gás total em R\$ para os 6 modelos da configuração C treinados com o algoritmo PPO na segunda fase do estudo.

(continua)

Temperatura ambiente em °C	Tarifa da energia elétrica (R\$)	Modelo 1	Modelo 2	Modelo 3	Modelo 4	Modelo 5	Modelo 6	Maior diferença de valores
20	2,00	0,61	0,95	0,61	0,61	0,92	0,94	0,34
20	2,25	0,62	0,86	0,61	0,63	0,87	0,90	0,29
21	1,00	0,52	0,82	0,56	0,53	0,83	0,70	0,31
21	1,25	0,47	0,80	0,56	0,52	0,87	0,75	0,40
21	1,50	0,48	0,89	0,56	0,55	0,87	0,68	0,41
21	1,75	0,48	0,93	0,56	0,65	0,83	0,83	0,45
21	2,00	0,45	0,86	0,56	0,59	0,83	0,88	0,43
21	2,25	0,46	0,84	0,56	0,57	0,86	0,80	0,40
22	1,00	0,48	0,78	0,53	0,58	0,77	0,74	0,30
22	1,25	0,41	0,77	0,53	0,54	0,77	0,76	0,36
22	1,50	0,46	0,80	0,52	0,53	0,78	0,73	0,34
22	1,75	0,49	0,74	0,52	0,50	0,78	0,80	0,31
22	2,00	0,40	0,73	0,52	0,52	0,80	0,74	0,40
22	2,25	0,50	0,76	0,51	0,54	0,77	0,81	0,31
23	1,00	0,49	0,68	0,49	0,48	0,76	0,68	0,28
23	1,25	0,48	0,74	0,49	0,56	0,75	0,67	0,27
23	1,50	0,43	0,73	0,49	0,66	0,72	0,66	0,30
23	1,75	0,44	0,74	0,46	0,53	0,72	0,70	0,30
23	2,00	0,41	0,63	0,46	0,45	0,72	0,68	0,31
23	2,25	0,46	0,75	0,46	0,46	0,74	0,69	0,29
24	1,00	0,39	0,66	0,43	0,46	0,63	0,61	0,27
24	1,25	0,37	0,63	0,45	0,44	0,66	0,56	0,29
24	1,50	0,40	0,66	0,43	0,50	0,66	0,60	0,26
24	1,75	0,39	0,72	0,43	0,50	0,64	0,67	0,33
24	2,00	0,41	0,68	0,43	0,53	0,63	0,62	0,27
24	2,25	0,39	0,69	0,45	0,40	0,66	0,65	0,30
25	1,00	0,38	0,55	0,41	0,45	0,60	0,56	0,22
25	1,25	0,35	0,60	0,41	0,45	0,60	0,57	0,25
25	1,50	0,35	0,59	0,41	0,38	0,59	0,53	0,24
25	1,75	0,36	0,62	0,41	0,39	0,59	0,56	0,26
25	2,00	0,35	0,58	0,41	0,41	0,59	0,58	0,24
25	2,25	0,36	0,61	0,41	0,33	0,62	0,59	0,29
26	1,00	0,30	0,48	0,38	0,32	0,53	0,53	0,23
26	1,25	0,31	0,56	0,34	0,45	0,54	0,55	0,25

**Tabela A.6** – Custo de gás total em R\$ para os 6 modelos da configuração C treinados com o algoritmo PPO na segunda fase do estudo.

(conclusão)

Temperatura ambiente em °C	Tarifa da energia elétrica (R\$)	Modelo 1	Modelo 2	Modelo 3	Modelo 4	Modelo 5	Modelo 6	Maior diferença de valores
26	1,50	0,32	0,53	0,38	0,40	0,56	0,47	0,24
26	1,75	0,31	0,48	0,34	0,39	0,54	0,52	0,23
26	2,00	0,35	0,53	0,34	0,38	0,54	0,53	0,20
26	2,25	0,32	0,60	0,34	0,36	0,54	0,57	0,28
27	1,00	0,25	0,50	0,25	0,32	0,49	0,45	0,25
27	1,25	0,31	0,51	0,25	0,25	0,49	0,39	0,26
27	1,50	0,29	0,45	0,25	0,25	0,53	0,36	0,28
27	1,75	0,29	0,46	0,25	0,25	0,51	0,46	0,26
27	2,00	0,30	0,48	0,25	0,25	0,49	0,50	0,25
27	2,25	0,28	0,52	0,25	0,25	0,50	0,52	0,27
28	1,00	0,24	0,43	0,23	0,23	0,45	0,30	0,22
28	1,25	0,23	0,41	0,23	0,23	0,44	0,34	0,21
28	1,50	0,23	0,40	0,23	0,24	0,47	0,24	0,24
28	1,75	0,23	0,47	0,23	0,23	0,52	0,47	0,29
28	2,00	0,23	0,45	0,23	0,23	0,45	0,46	0,23
28	2,25	0,23	0,38	0,23	0,25	0,47	0,41	0,24
29	1,00	0,21	0,37	0,21	0,22	0,42	0,24	0,21
29	1,25	0,20	0,28	0,21	0,24	0,39	0,23	0,19
29	1,50	0,20	0,36	0,21	0,23	0,45	0,23	0,25
29	1,75	0,20	0,42	0,21	0,21	0,41	0,45	0,25
29	2,00	0,22	0,45	0,21	0,22	0,45	0,36	0,24
29	2,25	0,21	0,43	0,21	0,21	0,43	0,42	0,22
30	1,00	0,19	0,34	0,19	0,19	0,38	0,21	0,19
30	1,25	0,18	0,37	0,19	0,20	0,39	0,20	0,21
30	1,50	0,18	0,34	0,19	0,21	0,39	0,22	0,21
30	1,75	0,19	0,39	0,19	0,21	0,39	0,34	0,20
30	2,00	0,18	0,36	0,19	0,19	0,38	0,39	0,21
30	2,25	0,19	0,28	0,19	0,19	0,39	0,37	0,20

**Tabela A.7** – Custo de água total em R\$ para os 6 modelos da configuração B treinados com o algoritmo PPO na segunda fase do estudo.

(continua)

Temperatura ambiente em °C	Tarifa da energia elétrica (R\$)	Modelo 1	Modelo 2	Modelo 3	Modelo 4	Modelo 5	Modelo 6	Maior diferença de valores
15	1,00	0,38	0,38	0,37	0,37	0,39	0,37	0,02
15	1,25	0,37	0,39	0,36	0,38	0,37	0,37	0,03
15	1,50	0,37	0,38	0,37	0,39	0,38	0,38	0,02
15	1,75	0,39	0,39	0,37	0,37	0,37	0,37	0,02
15	2,00	0,38	0,37	0,37	0,37	0,38	0,37	0,01
15	2,25	0,39	0,38	0,37	0,38	0,38	0,37	0,02
16	1,00	0,37	0,37	0,35	0,36	0,36	0,38	0,03
16	1,25	0,38	0,34	0,35	0,38	0,36	0,38	0,04
16	1,50	0,37	0,36	0,34	0,37	0,36	0,37	0,03
16	1,75	0,37	0,38	0,34	0,38	0,37	0,36	0,04
16	2,00	0,38	0,37	0,35	0,36	0,36	0,35	0,03
16	2,25	0,37	0,37	0,35	0,37	0,36	0,37	0,02
17	1,00	0,36	0,37	0,33	0,35	0,35	0,35	0,04
17	1,25	0,36	0,36	0,33	0,35	0,35	0,36	0,03
17	1,50	0,36	0,37	0,33	0,36	0,35	0,36	0,04
17	1,75	0,37	0,36	0,33	0,35	0,35	0,34	0,04
17	2,00	0,36	0,37	0,33	0,36	0,35	0,35	0,04
17	2,25	0,37	0,36	0,33	0,35	0,36	0,36	0,04
18	1,00	0,36	0,35	0,33	0,32	0,34	0,35	0,04
18	1,25	0,36	0,36	0,33	0,34	0,34	0,35	0,03
18	1,50	0,37	0,35	0,33	0,34	0,34	0,35	0,04
18	1,75	0,37	0,35	0,33	0,34	0,34	0,35	0,04
18	2,00	0,37	0,35	0,33	0,35	0,34	0,35	0,04
18	2,25	0,34	0,34	0,33	0,34	0,35	0,36	0,03
19	1,00	0,33	0,35	0,31	0,33	0,34	0,33	0,04
19	1,25	0,36	0,34	0,31	0,32	0,34	0,33	0,05
19	1,50	0,34	0,34	0,31	0,32	0,34	0,33	0,03
19	1,75	0,35	0,34	0,31	0,34	0,35	0,32	0,04
19	2,00	0,34	0,34	0,31	0,34	0,34	0,31	0,03
19	2,25	0,34	0,33	0,31	0,32	0,34	0,32	0,03
20	1,00	0,32	0,31	0,30	0,33	0,32	0,34	0,04
20	1,25	0,32	0,33	0,30	0,31	0,34	0,35	0,05
20	1,50	0,32	0,33	0,30	0,33	0,33	0,36	0,06
20	1,75	0,32	0,33	0,30	0,30	0,32	0,35	0,05

**Tabela A.7** – Custo de água total em R\$ para os 6 modelos da configuração B treinados com o algoritmo PPO na segunda fase do estudo.

(continua)

Temperatura ambiente em °C	Tarifa da energia elétrica (R\$)	Modelo 1	Modelo 2	Modelo 3	Modelo 4	Modelo 5	Modelo 6	Maior diferença de valores
20	2,00	0,31	0,33	0,30	0,33	0,32	0,34	0,04
20	2,25	0,32	0,33	0,30	0,31	0,33	0,34	0,04
21	1,00	0,31	0,31	0,30	0,34	0,32	0,34	0,04
21	1,25	0,30	0,31	0,30	0,29	0,32	0,35	0,06
21	1,50	0,31	0,31	0,30	0,30	0,32	0,33	0,03
21	1,75	0,31	0,31	0,30	0,30	0,31	0,35	0,05
21	2,00	0,30	0,32	0,30	0,30	0,32	0,34	0,04
21	2,25	0,30	0,31	0,30	0,30	0,31	0,33	0,03
22	1,00	0,29	0,29	0,28	0,29	0,30	0,31	0,03
22	1,25	0,30	0,29	0,28	0,29	0,30	0,32	0,04
22	1,50	0,29	0,30	0,28	0,29	0,30	0,32	0,04
22	1,75	0,29	0,29	0,28	0,27	0,30	0,31	0,04
22	2,00	0,29	0,30	0,28	0,29	0,31	0,32	0,04
22	2,25	0,29	0,29	0,28	0,29	0,31	0,31	0,03
23	1,00	0,28	0,29	0,27	0,27	0,30	0,30	0,03
23	1,25	0,27	0,29	0,27	0,27	0,30	0,32	0,05
23	1,50	0,27	0,30	0,27	0,26	0,30	0,30	0,04
23	1,75	0,28	0,29	0,27	0,28	0,30	0,30	0,03
23	2,00	0,28	0,28	0,27	0,26	0,30	0,32	0,06
23	2,25	0,28	0,29	0,27	0,28	0,30	0,30	0,03
24	1,00	0,26	0,28	0,26	0,25	0,29	0,28	0,04
24	1,25	0,26	0,28	0,26	0,27	0,29	0,29	0,03
24	1,50	0,28	0,29	0,26	0,26	0,28	0,28	0,03
24	1,75	0,27	0,27	0,26	0,27	0,29	0,28	0,03
24	2,00	0,27	0,29	0,26	0,25	0,29	0,28	0,04
24	2,25	0,26	0,28	0,26	0,25	0,29	0,30	0,05
25	1,00	0,26	0,27	0,25	0,25	0,27	0,28	0,03
25	1,25	0,27	0,26	0,25	0,25	0,27	0,29	0,04
25	1,50	0,27	0,27	0,25	0,27	0,27	0,27	0,02
25	1,75	0,26	0,26	0,25	0,26	0,28	0,27	0,03
25	2,00	0,27	0,27	0,25	0,26	0,27	0,30	0,05
25	2,25	0,26	0,27	0,25	0,27	0,27	0,28	0,03
26	1,00	0,25	0,25	0,24	0,27	0,26	0,27	0,03
26	1,25	0,26	0,25	0,24	0,25	0,27	0,24	0,03



**Tabela A.7** – Custo de água total em R\$ para os 6 modelos da configuração B treinados com o algoritmo PPO na segunda fase do estudo.

(conclusão)

Temperatura ambiente em °C	Tarifa da energia elétrica (R\$)	Modelo 1	Modelo 2	Modelo 3	Modelo 4	Modelo 5	Modelo 6	Maior diferença de valores
26	1,50	0,25	0,27	0,25	0,27	0,26	0,26	0,02
26	1,75	0,25	0,27	0,25	0,26	0,24	0,26	0,03
26	2,00	0,26	0,26	0,26	0,26	0,25	0,25	0,01
26	2,25	0,25	0,25	0,25	0,28	0,26	0,27	0,03
27	1,00	0,25	0,27	0,23	0,25	0,23	0,25	0,04
27	1,25	0,23	0,25	0,23	0,25	0,25	0,25	0,02
27	1,50	0,24	0,25	0,23	0,24	0,23	0,24	0,02
27	1,75	0,24	0,26	0,23	0,24	0,23	0,25	0,03
27	2,00	0,23	0,26	0,23	0,26	0,24	0,26	0,03
27	2,25	0,24	0,24	0,23	0,25	0,22	0,24	0,03
28	1,00	0,23	0,23	0,22	0,24	0,23	0,26	0,04
28	1,25	0,22	0,23	0,22	0,25	0,23	0,23	0,03
28	1,50	0,21	0,24	0,22	0,24	0,23	0,23	0,03
28	1,75	0,24	0,25	0,22	0,24	0,22	0,25	0,03
28	2,00	0,22	0,22	0,22	0,25	0,22	0,25	0,03
28	2,25	0,22	0,24	0,22	0,23	0,21	0,24	0,03
29	1,00	0,19	0,22	0,22	0,23	0,20	0,24	0,05
29	1,25	0,20	0,21	0,22	0,22	0,22	0,24	0,04
29	1,50	0,20	0,22	0,21	0,23	0,20	0,25	0,05
29	1,75	0,21	0,22	0,21	0,24	0,21	0,22	0,03
29	2,00	0,20	0,19	0,21	0,20	0,20	0,22	0,03
29	2,25	0,21	0,21	0,21	0,25	0,20	0,24	0,05
30	1,00	0,19	0,18	0,21	0,23	0,20	0,23	0,05
30	1,25	0,20	0,17	0,20	0,24	0,19	0,20	0,07
30	1,50	0,18	0,19	0,20	0,23	0,19	0,23	0,05
30	1,75	0,19	0,21	0,20	0,23	0,20	0,19	0,04
30	2,00	0,20	0,18	0,20	0,23	0,19	0,21	0,05
30	2,25	0,19	0,20	0,20	0,24	0,20	0,19	0,05

**Tabela A.8** – Custo de água total em R\$ para os 6 modelos da configuração C treinados com o algoritmo PPO na segunda fase do estudo.

(continua)

Temperatura ambiente em °C	Tarifa da energia elétrica (R\$)	Modelo 1	Modelo 2	Modelo 3	Modelo 4	Modelo 5	Modelo 6	Maior diferença de valores
15	1,00	0,37	0,42	0,37	0,4	0,38	0,37	0,05
15	1,25	0,37	0,42	0,37	0,38	0,38	0,38	0,05
15	1,50	0,38	0,42	0,37	0,41	0,38	0,40	0,05
15	1,75	0,38	0,41	0,37	0,38	0,38	0,40	0,04
15	2,00	0,37	0,42	0,37	0,39	0,38	0,39	0,05
15	2,25	0,38	0,42	0,37	0,38	0,38	0,39	0,05
16	1,00	0,36	0,38	0,35	0,37	0,37	0,36	0,03
16	1,25	0,36	0,4	0,35	0,37	0,37	0,37	0,05
16	1,50	0,36	0,39	0,35	0,37	0,37	0,37	0,04
16	1,75	0,36	0,41	0,35	0,39	0,37	0,38	0,06
16	2,00	0,36	0,39	0,35	0,38	0,37	0,39	0,04
16	2,25	0,37	0,38	0,35	0,38	0,37	0,37	0,03
17	1,00	0,37	0,38	0,35	0,37	0,36	0,35	0,03
17	1,25	0,35	0,38	0,35	0,39	0,36	0,37	0,04
17	1,50	0,35	0,38	0,35	0,36	0,36	0,35	0,03
17	1,75	0,35	0,38	0,35	0,39	0,36	0,37	0,04
17	2,00	0,36	0,38	0,35	0,38	0,36	0,35	0,03
17	2,25	0,36	0,39	0,35	0,38	0,37	0,36	0,04
18	1,00	0,35	0,36	0,34	0,36	0,35	0,36	0,02
18	1,25	0,34	0,37	0,34	0,37	0,35	0,34	0,03
18	1,50	0,36	0,37	0,34	0,37	0,34	0,36	0,03
18	1,75	0,35	0,38	0,34	0,37	0,35	0,38	0,04
18	2,00	0,35	0,37	0,34	0,36	0,35	0,37	0,03
18	2,25	0,34	0,38	0,34	0,37	0,35	0,36	0,04
19	1,00	0,34	0,32	0,33	0,34	0,34	0,34	0,02
19	1,25	0,32	0,37	0,33	0,35	0,34	0,34	0,05
19	1,50	0,33	0,35	0,33	0,34	0,34	0,35	0,02
19	1,75	0,34	0,4	0,33	0,33	0,34	0,35	0,07
19	2,00	0,34	0,36	0,33	0,36	0,34	0,35	0,03
19	2,25	0,33	0,36	0,33	0,36	0,35	0,36	0,03
20	1,00	0,31	0,35	0,33	0,33	0,33	0,34	0,04
20	1,25	0,33	0,36	0,33	0,33	0,33	0,33	0,03
20	1,50	0,31	0,36	0,33	0,32	0,33	0,33	0,05
20	1,75	0,31	0,36	0,33	0,33	0,33	0,31	0,05

**Tabela A.8** – Custo de água total em R\$ para os 6 modelos da configuração C treinados com o algoritmo PPO na segunda fase do estudo.

(continua)

Temperatura ambiente em °C	Tarifa da energia elétrica (R\$)	Modelo 1	Modelo 2	Modelo 3	Modelo 4	Modelo 5	Modelo 6	Maior diferença de valores
20	2,00	0,32	0,36	0,33	0,33	0,34	0,35	0,04
20	2,25	0,31	0,35	0,32	0,33	0,33	0,34	0,04
21	1,00	0,29	0,34	0,31	0,3	0,31	0,31	0,05
21	1,25	0,3	0,33	0,31	0,31	0,33	0,33	0,03
21	1,50	0,29	0,35	0,31	0,31	0,33	0,31	0,06
21	1,75	0,3	0,35	0,31	0,32	0,31	0,30	0,05
21	2,00	0,29	0,34	0,31	0,31	0,32	0,33	0,05
21	2,25	0,29	0,32	0,3	0,33	0,32	0,31	0,04
22	1,00	0,28	0,32	0,29	0,3	0,32	0,31	0,04
22	1,25	0,29	0,31	0,29	0,31	0,30	0,32	0,03
22	1,50	0,29	0,31	0,28	0,3	0,30	0,31	0,03
22	1,75	0,3	0,31	0,28	0,3	0,29	0,32	0,04
22	2,00	0,28	0,31	0,29	0,31	0,30	0,31	0,03
22	2,25	0,27	0,31	0,29	0,31	0,30	0,30	0,04
23	1,00	0,28	0,29	0,28	0,28	0,30	0,30	0,02
23	1,25	0,28	0,3	0,28	0,3	0,29	0,32	0,04
23	1,50	0,28	0,3	0,28	0,31	0,28	0,29	0,03
23	1,75	0,27	0,29	0,27	0,29	0,29	0,26	0,03
23	2,00	0,27	0,29	0,27	0,29	0,29	0,30	0,03
23	2,25	0,28	0,3	0,27	0,3	0,29	0,29	0,03
24	1,00	0,27	0,29	0,27	0,28	0,26	0,28	0,03
24	1,25	0,26	0,29	0,28	0,27	0,28	0,27	0,03
24	1,50	0,27	0,28	0,27	0,28	0,27	0,29	0,02
24	1,75	0,27	0,29	0,27	0,29	0,27	0,27	0,02
24	2,00	0,27	0,29	0,27	0,28	0,25	0,27	0,04
24	2,25	0,27	0,29	0,28	0,29	0,26	0,27	0,03
25	1,00	0,26	0,27	0,27	0,28	0,26	0,25	0,03
25	1,25	0,27	0,28	0,27	0,28	0,26	0,24	0,04
25	1,50	0,27	0,27	0,27	0,28	0,25	0,25	0,03
25	1,75	0,25	0,32	0,27	0,28	0,25	0,25	0,07
25	2,00	0,26	0,29	0,27	0,27	0,27	0,25	0,04
25	2,25	0,27	0,28	0,27	0,27	0,27	0,25	0,03
26	1,00	0,26	0,26	0,27	0,29	0,25	0,26	0,04
26	1,25	0,25	0,3	0,28	0,29	0,25	0,26	0,05

**Tabela A.8** – Custo de água total em R\$ para os 6 modelos da configuração C treinados com o algoritmo PPO na segunda fase do estudo.

(conclusão)

Temperatura ambiente em °C	Tarifa da energia elétrica (R\$)	Modelo 1	Modelo 2	Modelo 3	Modelo 4	Modelo 5	Modelo 6	Maior diferença de valores
26	1,50	0,26	0,25	0,27	0,28	0,26	0,24	0,04
26	1,75	0,25	0,26	0,28	0,27	0,26	0,25	0,03
26	2,00	0,27	0,25	0,28	0,28	0,25	0,25	0,03
26	2,25	0,26	0,28	0,28	0,28	0,25	0,26	0,03
27	1,00	0,26	0,27	0,29	0,29	0,24	0,25	0,05
27	1,25	0,24	0,26	0,29	0,27	0,23	0,25	0,06
27	1,50	0,24	0,25	0,29	0,28	0,27	0,26	0,05
27	1,75	0,27	0,24	0,29	0,31	0,25	0,24	0,07
27	2,00	0,26	0,24	0,29	0,28	0,23	0,22	0,07
27	2,25	0,26	0,26	0,29	0,29	0,24	0,26	0,05
28	1,00	0,25	0,25	0,29	0,27	0,24	0,24	0,05
28	1,25	0,25	0,23	0,29	0,27	0,22	0,27	0,07
28	1,50	0,26	0,23	0,29	0,3	0,24	0,27	0,07
28	1,75	0,25	0,3	0,29	0,27	0,25	0,24	0,06
28	2,00	0,26	0,28	0,29	0,29	0,24	0,24	0,05
28	2,25	0,27	0,23	0,29	0,29	0,25	0,24	0,06
29	1,00	0,26	0,26	0,28	0,29	0,23	0,26	0,06
29	1,25	0,25	0,25	0,28	0,28	0,22	0,26	0,06
29	1,50	0,24	0,23	0,28	0,26	0,21	0,27	0,07
29	1,75	0,24	0,28	0,28	0,26	0,22	0,23	0,06
29	2,00	0,23	0,26	0,28	0,29	0,22	0,24	0,07
29	2,25	0,25	0,28	0,28	0,27	0,24	0,23	0,05
30	1,00	0,23	0,23	0,27	0,25	0,22	0,26	0,05
30	1,25	0,26	0,27	0,27	0,28	0,21	0,27	0,07
30	1,50	0,24	0,25	0,27	0,27	0,21	0,24	0,06
30	1,75	0,24	0,25	0,27	0,26	0,21	0,22	0,06
30	2,00	0,24	0,25	0,27	0,23	0,21	0,20	0,07
30	2,25	0,24	0,21	0,27	0,27	0,21	0,21	0,06

**Tabela A.9** – Soma total do IQB durante um episódio para os 6 modelos da configuração B treinados com o algoritmo PPO na segunda fase do estudo.

(continua)

Temperatura ambiente em °C	Tarifa da energia elétrica (R\$)	Modelo 1	Modelo 2	Modelo 3	Modelo 4	Modelo 5	Modelo 6	Maior diferença de valores
15	1,00	4,75	4,68	4,77	4,65	4,67	4,66	0,12
15	1,25	4,75	4,69	4,74	4,72	4,78	4,77	0,09
15	1,50	4,72	4,74	4,77	4,58	4,67	4,63	0,19
15	1,75	4,53	4,66	4,77	4,65	4,77	4,73	0,24
15	2,00	4,69	4,74	4,77	4,73	4,73	4,76	0,08
15	2,25	4,69	4,70	4,77	4,54	4,76	4,74	0,23
16	1,00	4,97	4,97	4,98	4,95	5,00	4,92	0,08
16	1,25	4,91	4,92	4,98	4,78	4,99	4,92	0,21
16	1,50	4,96	4,97	4,95	4,87	5,00	4,93	0,13
16	1,75	4,92	4,89	4,95	4,80	4,97	4,80	0,17
16	2,00	4,89	4,97	5,00	4,90	5,00	4,97	0,11
16	2,25	4,91	4,94	4,94	4,80	5,00	4,90	0,20
17	1,00	5,16	5,04	5,17	5,17	5,20	5,11	0,16
17	1,25	5,18	5,12	5,18	5,16	5,20	5,14	0,08
17	1,50	5,17	5,09	5,18	5,17	5,20	5,17	0,11
17	1,75	5,17	5,17	5,18	5,19	5,20	5,04	0,16
17	2,00	5,18	5,13	5,18	4,91	5,20	5,19	0,29
17	2,25	5,10	5,15	5,18	4,98	5,16	5,05	0,20
18	1,00	5,30	5,35	5,39	5,21	5,37	5,32	0,18
18	1,25	5,28	5,28	5,39	5,38	5,38	5,32	0,11
18	1,50	5,23	5,35	5,39	5,35	5,38	5,28	0,16
18	1,75	5,24	5,35	5,39	5,35	5,38	4,90	0,49
18	2,00	5,22	5,24	5,39	5,33	5,37	5,23	0,17
18	2,25	5,31	5,39	5,39	5,33	5,35	5,32	0,08
19	1,00	5,55	5,42	5,59	5,47	5,52	5,55	0,17
19	1,25	5,31	5,54	5,59	5,31	5,54	5,45	0,28
19	1,50	5,52	5,52	5,59	5,42	5,51	5,44	0,17
19	1,75	5,42	5,51	5,59	5,46	5,46	5,39	0,20
19	2,00	5,45	5,44	5,59	5,48	5,51	5,52	0,15
19	2,25	5,54	5,51	5,59	5,42	5,50	5,42	0,17
20	1,00	5,73	5,75	5,76	5,59	5,67	5,55	0,21
20	1,25	5,72	5,62	5,76	5,65	5,58	5,48	0,28
20	1,50	5,70	5,60	5,76	5,57	5,64	5,22	0,54
20	1,75	5,72	5,65	5,76	5,57	5,67	5,46	0,30

**Tabela A.9** – Soma total do IQB durante um episódio para os 6 modelos da configuração B treinados com o algoritmo PPO na segunda fase do estudo.

(continua)

Temperatura ambiente em °C	Tarifa da energia elétrica (R\$)	Modelo 1	Modelo 2	Modelo 3	Modelo 4	Modelo 5	Modelo 6	Maior diferença de valores
20	2,00	5,71	5,69	5,76	5,52	5,67	5,51	0,25
20	2,25	5,67	5,60	5,76	5,59	5,65	5,46	0,30
21	1,00	5,89	5,83	5,85	5,08	5,74	5,49	0,81
21	1,25	5,91	5,79	5,89	5,79	5,75	5,34	0,57
21	1,50	5,84	5,73	5,85	5,85	5,75	5,46	0,39
21	1,75	5,82	5,86	5,89	5,58	5,75	5,39	0,50
21	2,00	5,89	5,80	5,89	5,80	5,75	5,45	0,44
21	2,25	5,89	5,79	5,89	5,86	5,79	5,49	0,40
22	1,00	5,99	5,96	5,96	5,90	5,84	5,75	0,24
22	1,25	5,98	6,01	6,00	5,92	5,84	5,67	0,34
22	1,50	5,99	5,89	6,00	5,92	5,84	5,70	0,30
22	1,75	5,99	5,99	5,96	5,75	5,87	5,70	0,29
22	2,00	6,01	5,98	6,00	5,95	5,83	5,66	0,35
22	2,25	6,03	5,98	6,00	5,91	5,87	5,69	0,34
23	1,00	6,11	5,99	6,12	5,88	5,99	5,89	0,24
23	1,25	6,12	6,07	6,12	5,91	5,99	5,65	0,47
23	1,50	6,03	5,99	6,12	5,90	5,99	5,92	0,22
23	1,75	6,11	6,00	6,12	6,13	5,98	5,82	0,31
23	2,00	6,11	6,08	6,12	5,98	5,98	5,71	0,41
23	2,25	6,13	6,04	6,12	6,12	5,99	6,00	0,14
24	1,00	6,20	6,13	6,18	6,09	6,01	6,06	0,19
24	1,25	6,18	6,08	6,18	6,08	5,99	5,83	0,35
24	1,50	6,17	6,02	6,18	6,05	6,06	6,09	0,16
24	1,75	6,15	6,03	6,18	6,01	5,99	5,96	0,22
24	2,00	6,18	6,07	6,18	6,14	6,03	6,04	0,15
24	2,25	6,20	6,09	6,18	5,97	5,99	5,78	0,42
25	1,00	6,26	6,14	6,25	6,28	6,09	6,08	0,20
25	1,25	6,21	6,24	6,25	6,27	6,12	6,02	0,25
25	1,50	6,22	6,21	6,25	6,12	6,12	6,09	0,16
25	1,75	6,19	6,24	6,25	6,19	6,01	6,05	0,24
25	2,00	6,26	6,16	6,25	6,27	6,12	5,73	0,54
25	2,25	6,25	6,17	6,25	6,19	6,09	6,09	0,16
26	1,00	6,34	6,23	6,28	6,16	6,20	6,19	0,18
26	1,25	6,33	6,21	6,28	6,32	6,09	6,05	0,28

**Tabela A.9** – Soma total do IQB durante um episódio para os 6 modelos da configuração B treinados com o algoritmo PPO na segunda fase do estudo.

(conclusão)

Temperatura ambiente em °C	Tarifa da energia elétrica (R\$)	Modelo 1	Modelo 2	Modelo 3	Modelo 4	Modelo 5	Modelo 6	Maior diferença de valores
26	1,50	6,28	6,19	6,32	6,16	6,15	6,29	0,17
26	1,75	6,20	6,13	6,32	6,25	6,28	6,19	0,19
26	2,00	6,25	6,26	6,31	6,24	6,21	6,28	0,10
26	2,25	6,32	6,23	6,32	5,98	6,17	6,08	0,34
27	1,00	6,35	5,77	6,33	6,22	6,27	6,28	0,58
27	1,25	6,33	6,22	6,33	6,28	6,18	6,06	0,27
27	1,50	6,34	6,22	6,33	6,28	6,33	6,16	0,18
27	1,75	6,25	6,22	6,33	6,30	6,28	6,16	0,17
27	2,00	6,35	6,15	6,33	6,22	6,14	6,21	0,21
27	2,25	6,33	6,26	6,33	6,21	6,37	6,31	0,16
28	1,00	6,31	6,01	6,34	6,06	6,09	6,18	0,33
28	1,25	6,43	6,15	6,34	6,16	6,08	5,94	0,49
28	1,50	6,25	6,33	6,33	6,22	6,07	6,14	0,26
28	1,75	6,22	6,08	6,33	6,17	6,22	6,12	0,25
28	2,00	6,27	6,16	6,33	6,12	6,21	6,04	0,29
28	2,25	6,00	6,22	6,33	6,21	6,33	6,05	0,33
29	1,00	6,13	6,09	6,24	5,99	6,36	5,91	0,45
29	1,25	6,43	6,20	6,22	6,03	6,13	5,96	0,47
29	1,50	6,30	6,16	6,33	5,97	6,36	5,89	0,47
29	1,75	6,40	5,98	6,33	5,82	6,22	6,06	0,58
29	2,00	6,35	6,21	6,29	5,98	6,34	6,01	0,37
29	2,25	5,98	6,26	6,33	5,93	6,36	6,03	0,43
30	1,00	5,83	6,15	6,04	5,93	6,17	5,86	0,34
30	1,25	5,81	6,18	6,12	5,76	6,31	5,95	0,55
30	1,50	6,40	6,12	6,12	5,79	6,31	5,89	0,61
30	1,75	6,34	6,22	6,12	5,73	6,20	6,29	0,61
30	2,00	6,16	6,08	6,12	5,85	6,31	5,91	0,46
30	2,25	6,46	6,02	6,12	5,57	6,18	6,18	0,89

**Tabela A.10** – Soma total do IQB durante um episódio para os 6 modelos da configuração C treinados com o algoritmo PPO na segunda fase do estudo.

(continua)

Temperatura ambiente em °C	Tarifa da energia elétrica (R\$)	Modelo 1	Modelo 2	Modelo 3	Modelo 4	Modelo 5	Modelo 6	Maior diferença de valores
15	1,00	4,78	4,38	4,79	4,54	4,74	4,69	0,41
15	1,25	4,74	4,40	4,79	4,65	4,74	4,65	0,39
15	1,50	4,75	4,42	4,78	4,37	4,74	4,48	0,41
15	1,75	4,76	4,51	4,79	4,70	4,74	4,58	0,28
15	2,00	4,79	4,34	4,79	4,67	4,74	4,55	0,45
15	2,25	4,76	4,36	4,79	4,67	4,74	4,55	0,43
16	1,00	4,97	4,78	5,03	4,99	4,97	4,99	0,25
16	1,25	5,04	4,70	5,03	4,88	4,97	4,94	0,34
16	1,50	4,98	4,73	5,03	4,88	4,97	4,91	0,30
16	1,75	5,02	4,56	5,03	4,81	4,97	4,85	0,47
16	2,00	4,95	4,75	5,03	4,85	4,97	4,76	0,28
16	2,25	5,03	4,82	5,03	4,85	4,97	4,86	0,21
17	1,00	5,12	5,03	5,16	4,95	5,18	5,07	0,23
17	1,25	5,19	4,98	5,16	4,88	5,18	5,00	0,31
17	1,50	5,05	5,07	5,16	5,06	5,18	5,09	0,13
17	1,75	5,20	5,01	5,16	4,94	5,17	5,00	0,26
17	2,00	5,19	4,98	5,16	4,90	5,18	5,01	0,29
17	2,25	5,21	4,76	5,14	4,97	5,16	5,09	0,45
18	1,00	5,30	5,18	5,39	5,26	5,37	5,31	0,21
18	1,25	5,30	5,15	5,39	5,17	5,37	5,34	0,24
18	1,50	5,31	5,25	5,39	5,09	5,36	5,13	0,30
18	1,75	5,36	5,06	5,39	5,17	5,36	5,05	0,34
18	2,00	5,34	5,20	5,39	5,21	5,38	5,15	0,24
18	2,25	5,34	5,16	5,39	5,14	5,36	5,23	0,25
19	1,00	5,46	5,52	5,56	5,36	5,48	5,40	0,20
19	1,25	5,31	5,25	5,56	5,41	5,48	5,45	0,31
19	1,50	5,47	5,31	5,56	5,45	5,48	5,42	0,25
19	1,75	5,46	4,27	5,55	5,42	5,47	5,40	1,28
19	2,00	5,47	5,32	5,55	5,34	5,49	5,38	0,23
19	2,25	5,52	5,28	5,55	5,37	5,49	5,31	0,27
20	1,00	5,67	5,43	5,61	5,63	5,55	5,44	0,24
20	1,25	5,59	5,31	5,65	5,57	5,55	5,47	0,34
20	1,50	5,56	5,35	5,65	5,68	5,55	5,36	0,33
20	1,75	5,69	5,32	5,65	5,56	5,54	5,43	0,37



**Tabela A.10** – Soma total do IQB durante um episódio para os 6 modelos da configuração C treinados com o algoritmo PPO na segunda fase do estudo.

(continua)

Temperatura ambiente em °C	Tarifa da energia elétrica (R\$)	Modelo 1	Modelo 2	Modelo 3	Modelo 4	Modelo 5	Modelo 6	Maior diferença de valores
20	2,00	5,61	5,31	5,65	5,58	5,54	5,40	0,34
20	2,25	5,68	5,44	5,65	5,61	5,45	5,53	0,24
21	1,00	5,84	5,46	5,82	5,74	5,63	5,25	0,59
21	1,25	5,87	5,47	5,83	5,84	5,61	5,45	0,42
21	1,50	5,75	5,41	5,83	5,74	5,61	5,65	0,42
21	1,75	5,80	5,37	5,83	5,64	5,58	5,71	0,46
21	2,00	5,85	5,43	5,83	5,75	5,52	5,55	0,42
21	2,25	5,58	5,49	5,82	5,55	5,65	5,47	0,35
22	1,00	5,94	5,68	5,99	5,11	5,57	5,81	0,88
22	1,25	5,95	5,87	5,99	5,69	5,79	5,72	0,30
22	1,50	6,00	5,68	5,99	5,78	5,79	5,78	0,32
22	1,75	5,78	5,77	5,99	5,84	5,79	5,64	0,35
22	2,00	5,88	5,77	5,99	5,59	5,83	5,67	0,40
22	2,25	5,98	5,77	5,99	5,82	5,79	5,72	0,27
23	1,00	5,98	6,05	6,11	5,98	6,01	5,74	0,37
23	1,25	6,09	6,01	6,11	5,88	6,04	5,74	0,37
23	1,50	6,10	5,96	6,11	5,79	6,04	5,80	0,32
23	1,75	6,05	6,03	6,08	5,93	5,99	5,97	0,15
23	2,00	5,99	6,03	6,08	5,95	5,99	6,01	0,13
23	2,25	6,11	5,89	6,08	5,75	6,04	6,01	0,36
24	1,00	6,17	6,06	6,20	6,10	6,17	6,09	0,14
24	1,25	5,99	6,04	6,19	6,02	6,16	5,90	0,29
24	1,50	6,18	6,13	6,20	6,10	6,17	5,99	0,21
24	1,75	5,98	6,07	6,20	5,95	6,17	6,15	0,25
24	2,00	6,12	6,02	6,20	5,99	6,16	6,10	0,21
24	2,25	6,14	5,86	6,19	5,88	6,15	6,17	0,33
25	1,00	6,23	6,06	6,22	6,05	6,23	6,20	0,18
25	1,25	6,18	6,12	6,22	6,09	6,24	6,09	0,15
25	1,50	6,10	6,23	6,22	6,12	6,29	6,12	0,19
25	1,75	5,96	5,57	6,22	6,00	6,28	6,21	0,71
25	2,00	6,26	5,99	6,22	6,13	6,21	6,20	0,27
25	2,25	6,16	6,09	6,22	6,04	6,17	5,91	0,31
26	1,00	6,18	6,07	6,18	6,22	6,31	6,05	0,26
26	1,25	6,07	5,51	6,28	5,83	6,32	6,14	0,81

**Tabela A.10** – Soma total do IQB durante um episódio para os 6 modelos da configuração C treinados com o algoritmo PPO na segunda fase do estudo.

(continua)

Temperatura ambiente em °C	Tarifa da energia elétrica (R\$)	Modelo 1	Modelo 2	Modelo 3	Modelo 4	Modelo 5	Modelo 6	Maior diferença de valores
26	1,50	6,28	5,70	6,18	5,85	6,28	6,02	0,58
26	1,75	6,21	6,17	6,28	6,12	6,28	5,60	0,68
26	2,00	6,15	6,21	6,28	6,10	6,33	6,11	0,23
26	2,25	6,27	6,09	6,28	6,17	6,30	6,15	0,21
27	1,00	6,21	6,10	6,43	6,07	6,33	5,99	0,44
27	1,25	6,38	6,18	6,43	6,31	6,36	6,05	0,38
27	1,50	6,27	5,72	6,43	6,41	5,95	6,19	0,71
27	1,75	6,25	5,83	6,43	6,23	6,20	6,24	0,60
27	2,00	6,28	6,09	6,43	6,39	6,36	6,27	0,34
27	2,25	6,22	6,26	6,43	6,42	6,30	5,99	0,44
28	1,00	6,45	6,15	6,45	6,50	6,24	6,31	0,35
28	1,25	6,32	6,21	6,49	6,38	6,32	5,94	0,55
28	1,50	6,35	6,12	6,49	6,38	6,25	6,14	0,37
28	1,75	6,35	5,63	6,49	6,50	5,95	6,24	0,87
28	2,00	6,39	5,61	6,45	6,41	6,15	6,16	0,84
28	2,25	6,46	6,12	6,49	6,34	6,03	6,02	0,47
29	1,00	6,43	5,94	6,49	6,46	6,14	6,38	0,55
29	1,25	6,47	5,63	6,49	6,37	6,18	6,28	0,86
29	1,50	6,16	5,86	6,49	6,20	6,15	6,16	0,63
29	1,75	6,42	5,72	6,49	6,52	6,16	6,14	0,80
29	2,00	6,42	5,99	6,49	6,43	6,14	5,68	0,81
29	2,25	6,24	5,77	6,49	6,39	5,92	6,08	0,72
30	1,00	6,34	6,24	6,48	6,53	6,00	6,45	0,53
30	1,25	6,32	5,85	6,48	6,40	6,04	6,35	0,63
30	1,50	6,35	5,92	6,48	6,45	6,04	6,24	0,56
30	1,75	6,38	6,02	6,48	6,53	6,05	5,56	0,97
30	2,00	6,42	5,80	6,48	6,48	6,01	6,04	0,68
30	2,25	6,30	5,55	6,48	6,30	6,06	5,96	0,93

**Tabela A.11** – IQB médio para todas as iterações de um episódio para os 6 modelos da configuração B treinados com o algoritmo PPO na segunda fase do estudo.

(continua)

Temperatura ambiente em °C	Tarifa da energia elétrica (R\$)	Modelo 1	Modelo 2	Modelo 3	Modelo 4	Modelo 5	Modelo 6	Maior diferença de valores
15	1,00	0,68	0,67	0,68	0,66	0,67	0,67	0,02
15	1,25	0,68	0,67	0,68	0,67	0,68	0,68	0,01
15	1,50	0,67	0,68	0,68	0,65	0,67	0,66	0,03
15	1,75	0,65	0,67	0,68	0,66	0,68	0,68	0,03
15	2,00	0,67	0,68	0,68	0,68	0,68	0,68	0,01
15	2,25	0,67	0,67	0,68	0,65	0,68	0,68	0,03
16	1,00	0,71	0,71	0,71	0,71	0,71	0,70	0,01
16	1,25	0,70	0,70	0,71	0,68	0,71	0,70	0,03
16	1,50	0,71	0,71	0,71	0,70	0,71	0,70	0,01
16	1,75	0,70	0,70	0,71	0,69	0,71	0,69	0,02
16	2,00	0,70	0,71	0,71	0,70	0,71	0,71	0,01
16	2,25	0,70	0,71	0,71	0,69	0,71	0,70	0,02
17	1,00	0,74	0,72	0,74	0,74	0,74	0,73	0,02
17	1,25	0,74	0,73	0,74	0,74	0,74	0,73	0,01
17	1,50	0,74	0,73	0,74	0,74	0,74	0,74	0,01
17	1,75	0,74	0,74	0,74	0,74	0,74	0,72	0,02
17	2,00	0,74	0,73	0,74	0,70	0,74	0,74	0,04
17	2,25	0,73	0,74	0,74	0,71	0,74	0,72	0,03
18	1,00	0,76	0,76	0,77	0,74	0,77	0,76	0,03
18	1,25	0,75	0,75	0,77	0,77	0,77	0,76	0,02
18	1,50	0,75	0,76	0,77	0,76	0,77	0,75	0,02
18	1,75	0,75	0,76	0,77	0,76	0,77	0,70	0,07
18	2,00	0,75	0,75	0,77	0,76	0,77	0,75	0,02
18	2,25	0,76	0,77	0,77	0,76	0,76	0,76	0,01
19	1,00	0,79	0,77	0,80	0,78	0,79	0,79	0,03
19	1,25	0,76	0,79	0,80	0,76	0,79	0,78	0,04
19	1,50	0,79	0,79	0,80	0,77	0,79	0,78	0,03
19	1,75	0,77	0,79	0,80	0,78	0,78	0,77	0,03
19	2,00	0,78	0,78	0,80	0,78	0,79	0,79	0,02
19	2,25	0,79	0,79	0,80	0,77	0,79	0,77	0,03
20	1,00	0,82	0,82	0,82	0,80	0,81	0,79	0,03
20	1,25	0,82	0,80	0,82	0,81	0,80	0,78	0,04
20	1,50	0,81	0,80	0,82	0,80	0,81	0,75	0,07
20	1,75	0,82	0,81	0,82	0,80	0,81	0,78	0,04

**Tabela A.11** – IQB médio para todas as iterações de um episódio para os 6 modelos da configuração B treinados com o algoritmo PPO na segunda fase do estudo.

(continua)

Temperatura ambiente em °C	Tarifa da energia elétrica (R\$)	Modelo 1	Modelo 2	Modelo 3	Modelo 4	Modelo 5	Modelo 6	Maior diferença de valores
20	2,00	0,82	0,81	0,82	0,79	0,81	0,79	0,03
20	2,25	0,81	0,80	0,82	0,80	0,81	0,78	0,04
21	1,00	0,84	0,83	0,84	0,73	0,82	0,78	0,11
21	1,25	0,84	0,83	0,84	0,83	0,82	0,76	0,08
21	1,50	0,83	0,82	0,84	0,84	0,82	0,78	0,06
21	1,75	0,83	0,84	0,84	0,80	0,82	0,77	0,07
21	2,00	0,84	0,83	0,84	0,83	0,82	0,78	0,06
21	2,25	0,84	0,83	0,84	0,84	0,83	0,78	0,06
22	1,00	0,86	0,85	0,85	0,84	0,83	0,82	0,04
22	1,25	0,85	0,86	0,86	0,85	0,83	0,81	0,05
22	1,50	0,86	0,84	0,86	0,85	0,83	0,81	0,05
22	1,75	0,86	0,86	0,85	0,82	0,84	0,81	0,05
22	2,00	0,86	0,85	0,86	0,85	0,83	0,81	0,05
22	2,25	0,86	0,85	0,86	0,84	0,84	0,81	0,05
23	1,00	0,87	0,86	0,87	0,84	0,86	0,84	0,03
23	1,25	0,87	0,87	0,87	0,84	0,86	0,81	0,06
23	1,50	0,86	0,86	0,87	0,84	0,86	0,85	0,03
23	1,75	0,87	0,86	0,87	0,88	0,85	0,83	0,05
23	2,00	0,87	0,87	0,87	0,85	0,85	0,82	0,05
23	2,25	0,88	0,86	0,87	0,87	0,86	0,86	0,02
24	1,00	0,89	0,88	0,88	0,87	0,86	0,87	0,03
24	1,25	0,88	0,87	0,88	0,87	0,86	0,83	0,05
24	1,50	0,88	0,86	0,88	0,86	0,87	0,87	0,02
24	1,75	0,88	0,86	0,88	0,86	0,86	0,85	0,03
24	2,00	0,88	0,87	0,88	0,88	0,86	0,86	0,02
24	2,25	0,89	0,87	0,88	0,85	0,86	0,83	0,06
25	1,00	0,89	0,88	0,89	0,90	0,87	0,87	0,03
25	1,25	0,89	0,89	0,89	0,90	0,87	0,86	0,04
25	1,50	0,89	0,89	0,89	0,87	0,87	0,87	0,02
25	1,75	0,88	0,89	0,89	0,88	0,86	0,86	0,03
25	2,00	0,89	0,88	0,89	0,90	0,87	0,82	0,08
25	2,25	0,89	0,88	0,89	0,88	0,87	0,87	0,02
26	1,00	0,91	0,89	0,90	0,88	0,89	0,88	0,03
26	1,25	0,90	0,89	0,90	0,90	0,87	0,86	0,04

**Tabela A.11** – IQB médio para todas as iterações de um episódio para os 6 modelos da configuração B treinados com o algoritmo PPO na segunda fase do estudo.

(conclusão)

Temperatura ambiente em °C	Tarifa da energia elétrica (R\$)	Modelo 1	Modelo 2	Modelo 3	Modelo 4	Modelo 5	Modelo 6	Maior diferença de valores
26	1,50	0,90	0,88	0,90	0,88	0,88	0,90	0,02
26	1,75	0,89	0,88	0,90	0,89	0,90	0,88	0,02
26	2,00	0,89	0,89	0,90	0,89	0,89	0,90	0,01
26	2,25	0,90	0,89	0,90	0,85	0,88	0,87	0,05
27	1,00	0,91	0,82	0,90	0,89	0,90	0,90	0,09
27	1,25	0,90	0,89	0,90	0,90	0,88	0,87	0,03
27	1,50	0,91	0,89	0,90	0,90	0,90	0,88	0,03
27	1,75	0,89	0,89	0,90	0,90	0,90	0,88	0,02
27	2,00	0,91	0,88	0,90	0,89	0,88	0,89	0,03
27	2,25	0,90	0,89	0,90	0,89	0,91	0,90	0,02
28	1,00	0,90	0,86	0,91	0,87	0,87	0,88	0,05
28	1,25	0,92	0,88	0,91	0,88	0,87	0,85	0,07
28	1,50	0,89	0,90	0,90	0,89	0,87	0,88	0,03
28	1,75	0,89	0,87	0,90	0,88	0,89	0,87	0,03
28	2,00	0,90	0,88	0,90	0,87	0,89	0,86	0,04
28	2,25	0,86	0,89	0,90	0,89	0,90	0,86	0,04
29	1,00	0,88	0,87	0,89	0,86	0,91	0,84	0,07
29	1,25	0,92	0,89	0,89	0,86	0,88	0,85	0,07
29	1,50	0,90	0,88	0,90	0,85	0,91	0,84	0,07
29	1,75	0,91	0,85	0,90	0,83	0,89	0,87	0,08
29	2,00	0,91	0,89	0,90	0,85	0,91	0,86	0,06
29	2,25	0,85	0,89	0,90	0,85	0,91	0,86	0,06
30	1,00	0,83	0,88	0,86	0,85	0,88	0,84	0,05
30	1,25	0,83	0,88	0,87	0,82	0,90	0,85	0,08
30	1,50	0,91	0,87	0,87	0,83	0,90	0,84	0,08
30	1,75	0,91	0,89	0,87	0,82	0,89	0,90	0,09
30	2,00	0,88	0,87	0,87	0,84	0,90	0,84	0,06
30	2,25	0,92	0,86	0,87	0,80	0,88	0,88	0,12

**Tabela A.12** – IQB médio para todas as iterações de um episódio para os 6 modelos da configuração C treinados com o algoritmo PPO na segunda fase do estudo.

(continua)

Temperatura ambiente em °C	Tarifa da energia elétrica (R\$)	Modelo 1	Modelo 2	Modelo 3	Modelo 4	Modelo 5	Modelo 6	Maior diferença de valores
15	1,00	0,68	0,63	0,68	0,65	0,68	0,67	0,05
15	1,25	0,68	0,63	0,68	0,66	0,68	0,66	0,05
15	1,50	0,68	0,63	0,68	0,62	0,68	0,64	0,06
15	1,75	0,68	0,64	0,68	0,67	0,68	0,65	0,04
15	2,00	0,68	0,62	0,68	0,67	0,67	0,65	0,06
15	2,25	0,68	0,62	0,68	0,67	0,68	0,65	0,06
16	1,00	0,71	0,68	0,72	0,71	0,71	0,71	0,04
16	1,25	0,72	0,67	0,72	0,70	0,71	0,71	0,05
16	1,50	0,71	0,68	0,72	0,70	0,71	0,70	0,04
16	1,75	0,72	0,65	0,72	0,69	0,71	0,69	0,07
16	2,00	0,71	0,68	0,72	0,69	0,71	0,68	0,04
16	2,25	0,72	0,69	0,72	0,69	0,71	0,69	0,03
17	1,00	0,73	0,72	0,74	0,71	0,74	0,72	0,03
17	1,25	0,74	0,71	0,74	0,70	0,74	0,71	0,04
17	1,50	0,72	0,72	0,74	0,72	0,74	0,73	0,02
17	1,75	0,74	0,72	0,74	0,71	0,73	0,71	0,03
17	2,00	0,74	0,71	0,74	0,70	0,74	0,72	0,04
17	2,25	0,74	0,68	0,73	0,71	0,74	0,73	0,06
18	1,00	0,76	0,74	0,77	0,75	0,76	0,76	0,03
18	1,25	0,76	0,74	0,77	0,74	0,76	0,76	0,03
18	1,50	0,76	0,75	0,77	0,73	0,76	0,73	0,04
18	1,75	0,77	0,72	0,77	0,74	0,77	0,72	0,05
18	2,00	0,76	0,74	0,77	0,74	0,77	0,74	0,03
18	2,25	0,76	0,74	0,77	0,73	0,77	0,75	0,04
19	1,00	0,78	0,79	0,79	0,77	0,78	0,77	0,02
19	1,25	0,76	0,75	0,79	0,77	0,79	0,78	0,04
19	1,50	0,78	0,76	0,79	0,78	0,78	0,77	0,03
19	1,75	0,78	0,61	0,79	0,77	0,79	0,77	0,18
19	2,00	0,78	0,76	0,79	0,76	0,79	0,77	0,03
19	2,25	0,79	0,75	0,79	0,77	0,78	0,76	0,04
20	1,00	0,81	0,78	0,80	0,80	0,79	0,78	0,03
20	1,25	0,80	0,76	0,81	0,80	0,79	0,78	0,05
20	1,50	0,79	0,76	0,81	0,81	0,79	0,77	0,05
20	1,75	0,81	0,76	0,81	0,79	0,79	0,78	0,05

**Tabela A.12** – IQB médio para todas as iterações de um episódio para os 6 modelos da configuração C treinados com o algoritmo PPO na segunda fase do estudo.

(continua)

Temperatura ambiente em °C	Tarifa da energia elétrica (R\$)	Modelo 1	Modelo 2	Modelo 3	Modelo 4	Modelo 5	Modelo 6	Maior diferença de valores
20	2,00	0,80	0,76	0,81	0,80	0,79	0,77	0,05
20	2,25	0,81	0,78	0,81	0,80	0,79	0,79	0,03
21	1,00	0,83	0,78	0,83	0,82	0,78	0,75	0,08
21	1,25	0,84	0,78	0,83	0,83	0,80	0,78	0,06
21	1,50	0,82	0,77	0,83	0,82	0,80	0,81	0,06
21	1,75	0,83	0,77	0,83	0,81	0,81	0,82	0,06
21	2,00	0,84	0,78	0,83	0,82	0,80	0,79	0,06
21	2,25	0,80	0,78	0,83	0,79	0,80	0,78	0,05
22	1,00	0,85	0,81	0,86	0,73	0,83	0,83	0,13
22	1,25	0,85	0,84	0,86	0,81	0,83	0,82	0,05
22	1,50	0,86	0,81	0,86	0,83	0,80	0,83	0,05
22	1,75	0,83	0,82	0,86	0,83	0,83	0,81	0,05
22	2,00	0,84	0,82	0,86	0,80	0,83	0,81	0,06
22	2,25	0,85	0,82	0,86	0,83	0,83	0,82	0,04
23	1,00	0,85	0,86	0,87	0,85	0,86	0,82	0,05
23	1,25	0,87	0,86	0,87	0,84	0,86	0,82	0,05
23	1,50	0,87	0,85	0,87	0,83	0,86	0,83	0,04
23	1,75	0,86	0,86	0,87	0,85	0,86	0,85	0,02
23	2,00	0,86	0,86	0,87	0,85	0,85	0,86	0,02
23	2,25	0,87	0,84	0,87	0,82	0,86	0,86	0,05
24	1,00	0,88	0,87	0,89	0,87	0,88	0,87	0,02
24	1,25	0,86	0,86	0,88	0,86	0,88	0,84	0,04
24	1,50	0,88	0,88	0,89	0,87	0,88	0,86	0,03
24	1,75	0,85	0,87	0,89	0,85	0,88	0,88	0,04
24	2,00	0,87	0,86	0,89	0,86	0,88	0,87	0,03
24	2,25	0,88	0,84	0,88	0,84	0,87	0,88	0,04
25	1,00	0,89	0,87	0,89	0,86	0,88	0,89	0,03
25	1,25	0,88	0,87	0,89	0,87	0,89	0,87	0,02
25	1,50	0,87	0,89	0,89	0,87	0,89	0,87	0,03
25	1,75	0,85	0,80	0,89	0,86	0,87	0,89	0,10
25	2,00	0,89	0,86	0,89	0,88	0,89	0,89	0,03
25	2,25	0,88	0,87	0,89	0,86	0,89	0,84	0,05
26	1,00	0,88	0,87	0,88	0,89	0,90	0,86	0,04
26	1,25	0,87	0,79	0,90	0,83	0,90	0,88	0,11

**Tabela A.12** – IQB médio para todas as iterações de um episódio para os 6 modelos da configuração C treinados com o algoritmo PPO na segunda fase do estudo.

(conclusão)

Temperatura ambiente em °C	Tarifa da energia elétrica (R\$)	Modelo 1	Modelo 2	Modelo 3	Modelo 4	Modelo 5	Modelo 6	Maior diferença de valores
26	1,50	0,90	0,81	0,88	0,84	0,88	0,86	0,09
26	1,75	0,89	0,88	0,90	0,87	0,90	0,80	0,10
26	2,00	0,88	0,89	0,90	0,87	0,90	0,87	0,03
26	2,25	0,90	0,87	0,90	0,88	0,90	0,88	0,03
27	1,00	0,89	0,87	0,92	0,87	0,91	0,86	0,06
27	1,25	0,91	0,88	0,92	0,90	0,90	0,86	0,06
27	1,50	0,90	0,82	0,92	0,92	0,90	0,88	0,10
27	1,75	0,89	0,83	0,92	0,89	0,91	0,89	0,09
27	2,00	0,90	0,87	0,92	0,91	0,88	0,90	0,05
27	2,25	0,89	0,89	0,92	0,92	0,89	0,86	0,06
28	1,00	0,92	0,88	0,92	0,93	0,90	0,90	0,05
28	1,25	0,90	0,89	0,93	0,91	0,90	0,85	0,08
28	1,50	0,91	0,87	0,93	0,91	0,88	0,88	0,06
28	1,75	0,91	0,80	0,93	0,93	0,90	0,89	0,13
28	2,00	0,91	0,80	0,92	0,92	0,89	0,88	0,12
28	2,25	0,92	0,87	0,93	0,91	0,90	0,86	0,07
29	1,00	0,92	0,85	0,93	0,92	0,89	0,91	0,08
29	1,25	0,92	0,80	0,93	0,91	0,88	0,90	0,13
29	1,50	0,88	0,84	0,93	0,89	0,87	0,88	0,09
29	1,75	0,92	0,82	0,93	0,93	0,86	0,88	0,11
29	2,00	0,92	0,86	0,93	0,92	0,88	0,81	0,12
29	2,25	0,89	0,82	0,93	0,91	0,88	0,87	0,11
30	1,00	0,91	0,89	0,93	0,93	0,86	0,92	0,07
30	1,25	0,90	0,84	0,93	0,91	0,86	0,91	0,09
30	1,50	0,91	0,85	0,93	0,92	0,86	0,89	0,08
30	1,75	0,91	0,86	0,93	0,93	0,86	0,79	0,14
30	2,00	0,92	0,83	0,93	0,93	0,86	0,86	0,10
30	2,25	0,90	0,79	0,93	0,90	0,87	0,85	0,14



Por fim, as Tabelas A.13 a A.16 mostram quais foram os *directive concepts* selecionados pelo *selective concept* do tipo *learned* para os modelos 3 e 5 das configurações B e C durante cada ação dos episódios avaliados. Para os modelos de número 3, definem-se os seguintes *concepts*:

- *Concept 1*: Banho nas temperaturas de 15°C, 16°C, 17°C, 18°C e 19°C;
- *Concept 2*: Banho nas temperaturas de 20°C, 21°C, 22°C, 23°C e 24°C;
- *Concept 3*: Banho nas temperaturas de 25°C, 26°C, 27°C, 28°C, 29°C e 30°C.

Já para os modelos de número 5, os seguintes *concepts* foram utilizados:

- *Concept 1*: Banho nas temperaturas de 15°C, 16°C, 17°C, 18°C e 19°C e com tarifas de energia elétrica de R\$1,00, R\$1,25 e R\$1,50;
- *Concept 2*: Banho nas temperaturas de 15°C, 16°C, 17°C, 18°C e 19°C e com tarifas de energia elétrica de R\$1,75, R\$2,00 e R\$2,25;
- *Concept 3*: Banho nas temperaturas de 20°C, 21°C, 22°C, 23°C e 24°C e com tarifas de energia elétrica de R\$1,00, R\$1,25 e R\$1,50;
- *Concept 4*: Banho nas temperaturas de 20°C, 21°C, 22°C, 23°C e 24°C e com tarifas de energia elétrica de R\$1,75, R\$2,00 e R\$2,25;
- *Concept 5*: Banho nas temperaturas de 25°C, 26°C, 27°C, 28°C, 29°C e 30°C e com tarifas de energia elétrica de R\$1,00, R\$1,25 e R\$1,50;
- *Concept 6*: Banho nas temperaturas de 25°C, 26°C, 27°C, 28°C, 29°C e 30°C e com tarifas de energia elétrica de R\$1,75, R\$2,00 e R\$2,25.

**Tabela A.13** – Número dos concepts selecionados durante cada ação para o modelo 3 treinado com a configuração B.

(continua)

Temperatura ambiente em °C	Tarifa da energia elétrica (R\$)	Ação 1	Ação 2	Ação 3	Ação 4	Ação 5	Ação 6	Ação 7
15	1,00	1	3	1	1	1	1	1
15	1,25	1	3	1	2	1	1	1
15	1,50	1	3	1	1	1	1	1
15	1,75	1	3	1	1	1	1	1
15	2,00	1	3	1	1	1	1	1
15	2,25	1	3	1	1	1	1	1
16	1,00	1	3	1	1	1	2	1
16	1,25	1	3	1	1	1	2	1
16	1,50	1	3	1	2	2	2	1
16	1,75	1	3	1	2	1	2	2
16	2,00	1	3	1	1	1	1	1
16	2,25	1	3	1	1	2	2	1
17	1,00	1	3	1	2	2	1	2
17	1,25	1	3	1	2	2	2	2
17	1,50	1	3	1	2	2	2	2
17	1,75	1	3	1	2	2	2	2
17	2,00	1	3	1	2	2	2	2
17	2,25	1	3	1	2	2	2	2
18	1,00	1	3	1	2	2	2	2
18	1,25	1	3	1	2	2	2	2
18	1,50	1	3	1	2	2	2	2
18	1,75	1	3	1	2	2	2	2
18	2,00	1	3	1	2	2	1	2
18	2,25	1	3	1	2	2	2	2
19	1,00	1	3	1	2	2	2	2
19	1,25	1	3	2	2	2	2	2
19	1,50	3	3	2	2	2	2	2
19	1,75	1	3	1	2	2	2	2
19	2,00	1	3	2	2	2	2	2
19	2,25	1	3	1	2	2	2	2
20	1,00	1	3	2	2	2	2	2
20	1,25	1	3	2	2	2	2	2
20	1,50	3	3	2	2	2	2	2
20	1,75	1	3	2	2	2	2	2
20	2,00	3	3	2	2	2	2	2
20	2,25	3	3	2	2	2	2	2

**Tabela A.13** – Número dos concepts seleccionados durante cada ação para o modelo 3 treinado com a configuração B.

(continua)

Temperatura ambiente em °C	Tarifa da energia elétrica (R\$)	Ação 1	Ação 2	Ação 3	Ação 4	Ação 5	Ação 6	Ação 7
21	1,00	3	3	2	2	2	2	2
21	1,25	3	1	2	2	2	2	2
21	1,50	3	3	2	2	2	2	2
21	1,75	3	1	2	2	2	2	2
21	2,00	3	1	2	2	2	2	2
21	2,25	3	1	2	2	2	2	2
22	1,00	3	1	2	2	2	2	2
22	1,25	3	2	2	2	2	2	2
22	1,50	3	2	2	2	2	2	2
22	1,75	3	1	2	2	2	2	2
22	2,00	3	2	2	2	2	2	2
22	2,25	3	2	2	2	2	2	2
23	1,00	3	2	2	2	2	2	2
23	1,25	3	2	2	2	2	2	2
23	1,50	3	2	2	2	2	2	2
23	1,75	3	2	2	2	2	2	2
23	2,00	3	2	2	2	2	2	2
23	2,25	3	2	2	2	2	2	2
24	1,00	3	2	2	2	2	2	2
24	1,25	3	2	2	2	2	2	2
24	1,50	3	2	2	2	2	2	2
24	1,75	3	2	2	2	2	2	2
24	2,00	3	2	2	2	2	2	2
24	2,25	3	2	2	2	2	2	2
25	1,00	3	2	2	2	2	2	2
25	1,25	3	2	2	2	2	2	2
25	1,50	3	2	2	2	2	2	2
25	1,75	3	2	2	2	2	2	2
25	2,00	3	2	2	2	2	2	2
25	2,25	3	2	2	2	2	2	2
26	1,00	1	2	2	2	2	2	2
26	1,25	1	2	2	2	2	2	2
26	1,50	3	2	2	2	2	2	2
26	1,75	3	2	2	2	2	2	2
26	2,00	3	3	2	2	2	2	2
26	2,25	3	2	2	2	2	2	2

**Tabela A.13** – Número dos concepts seleccionados durante cada ação para o modelo 3 treinado com a configuração B.

(conclusão)

Temperatura ambiente em °C	Tarifa da energia elétrica (R\$)	Ação 1	Ação 2	Ação 3	Ação 4	Ação 5	Ação 6	Ação 7
27	1,00	1	2	2	2	2	2	2
27	1,25	1	3	3	2	2	2	2
27	1,50	1	2	2	2	2	2	2
27	1,75	1	3	3	2	2	2	2
27	2,00	1	2	2	2	2	2	2
27	2,25	1	2	2	2	2	2	2
28	1,00	1	3	3	2	2	2	3
28	1,25	1	3	3	2	2	2	3
28	1,50	1	3	3	2	2	2	2
28	1,75	1	3	3	2	2	2	2
28	2,00	1	3	3	2	2	2	2
28	2,25	1	3	3	2	2	2	2
29	1,00	1	3	2	3	2	2	3
29	1,25	1	3	2	3	2	3	2
29	1,50	1	3	3	3	3	3	3
29	1,75	1	3	3	3	3	3	3
29	2,00	1	3	3	3	3	2	3
29	2,25	1	3	3	3	3	3	3
30	1,00	1	3	3	3	2	3	3
30	1,25	1	3	3	3	3	3	3
30	1,50	1	3	3	3	3	3	3
30	1,75	1	3	3	3	3	3	3
30	2,00	1	3	3	3	3	3	3
30	2,25	1	3	3	3	3	3	3

**Tabela A.14** – Número dos concepts selecionados durante cada ação para o modelo 5 treinado com a configuração B.

(continua)

Temperatura ambiente em °C	Tarifa da energia elétrica (R\$)	Ação 1	Ação 2	Ação 3	Ação 4	Ação 5	Ação 6	Ação 7
15	1,00	1	2	2	1	3	1	3
15	1,25	1	2	2	2	4	3	1
15	1,50	1	2	2	2	1	4	5
15	1,75	1	2	2	2	1	4	3
15	2,00	1	2	2	2	2	1	3
15	2,25	1	2	2	1	1	1	1
16	1,00	1	2	2	6	1	1	2
16	1,25	1	2	2	1	6	3	1
16	1,50	1	2	2	3	2	2	6
16	1,75	1	2	2	4	1	1	6
16	2,00	1	2	2	2	3	4	4
16	2,25	1	2	2	4	1	2	3
17	1,00	1	2	2	2	1	4	2
17	1,25	1	2	2	2	1	1	1
17	1,50	1	2	2	2	4	4	2
17	1,75	1	2	2	2	2	4	6
17	2,00	1	2	2	4	1	4	4
17	2,25	1	2	2	1	1	1	3
18	1,00	1	2	2	4	3	6	6
18	1,25	1	2	2	1	4	1	2
18	1,50	1	2	2	2	6	6	3
18	1,75	1	2	2	1	2	2	6
18	2,00	1	2	2	4	3	2	4
18	2,25	1	2	2	1	6	1	2
19	1,00	1	2	6	4	3	2	4
19	1,25	1	2	6	1	2	1	1
19	1,50	1	2	6	2	4	4	6
19	1,75	1	2	6	1	3	6	1
19	2,00	1	2	6	3	4	4	4
19	2,25	1	2	6	6	2	1	2
20	1,00	1	2	6	4	2	2	1
20	1,25	1	2	6	1	1	5	4
20	1,50	1	2	6	6	2	6	3
20	1,75	1	2	6	4	6	4	2
20	2,00	1	2	6	2	1	2	1
20	2,25	1	2	6	4	1	1	3

**Tabela A.14** – Número dos concepts selecionados durante cada ação para o modelo 5 treinado com a configuração B.

(continua)

Temperatura ambiente em °C	Tarifa da energia elétrica (R\$)	Ação 1	Ação 2	Ação 3	Ação 4	Ação 5	Ação 6	Ação 7
21	1,00	1	2	6	6	6	6	6
21	1,25	1	2	6	2	6	2	1
21	1,50	1	2	6	2	4	4	4
21	1,75	1	2	6	4	6	1	2
21	2,00	1	2	6	6	2	6	2
21	2,25	1	2	6	6	2	2	3
22	1,00	1	2	6	6	2	2	1
22	1,25	1	2	6	2	2	6	5
22	1,50	1	2	6	3	2	4	2
22	1,75	1	2	6	5	6	6	2
22	2,00	1	6	2	4	6	2	6
22	2,25	1	2	6	6	6	3	3
23	1,00	1	6	6	6	6	2	1
23	1,25	1	6	6	2	1	2	1
23	1,50	1	6	6	2	5	6	3
23	1,75	1	6	6	6	6	2	1
23	2,00	1	6	6	2	5	6	3
23	2,25	1	6	6	2	4	2	1
24	1,00	1	6	6	6	2	1	2
24	1,25	1	6	6	6	2	1	1
24	1,50	1	6	6	6	4	3	6
24	1,75	1	6	6	1	1	2	1
24	2,00	1	6	6	6	6	2	2
24	2,25	1	6	2	1	1	1	6
25	1,00	1	6	5	6	2	1	5
25	1,25	1	6	4	5	6	2	2
25	1,50	1	6	6	2	2	6	5
25	1,75	1	6	6	4	1	1	2
25	2,00	1	6	6	2	2	2	4
25	2,25	1	6	6	6	6	2	5
26	1,00	1	6	5	2	3	6	4
26	1,25	1	6	4	1	2	3	5
26	1,50	1	6	6	2	5	5	2
26	1,75	1	6	3	5	5	5	5
26	2,00	1	6	5	2	6	5	2
26	2,25	1	6	1	6	6	3	5

**Tabela A.14** – Número dos concepts seleccionados durante cada ação para o modelo 5 treinado com a configuração B.

(conclusão)

Temperatura ambiente em °C	Tarifa da energia elétrica (R\$)	Ação 1	Ação 2	Ação 3	Ação 4	Ação 5	Ação 6	Ação 7
27	1,00	1	6	5	2	6	5	5
27	1,25	1	6	5	2	6	3	3
27	1,50	1	6	3	5	5	5	5
27	1,75	1	6	5	5	2	6	5
27	2,00	1	6	2	1	6	5	5
27	2,25	1	6	5	5	5	5	5
28	1,00	1	6	1	6	5	2	6
28	1,25	1	6	5	5	2	2	2
28	1,50	1	6	1	6	2	6	5
28	1,75	1	6	5	5	2	6	5
28	2,00	1	6	5	6	5	2	5
28	2,25	1	6	5	5	5	5	5
29	1,00	1	6	5	5	5	5	5
29	1,25	1	6	6	5	5	4	1
29	1,50	1	6	5	5	5	5	5
29	1,75	1	6	5	2	5	5	5
29	2,00	1	6	6	6	5	5	5
29	2,25	1	6	5	5	5	5	5
30	1,00	1	6	2	5	5	6	5
30	1,25	1	6	6	5	6	5	5
30	1,50	1	6	5	5	5	5	5
30	1,75	1	6	5	5	5	3	5
30	2,00	1	6	5	5	5	5	5
30	2,25	1	6	5	4	5	5	3

**Tabela A.15** – Número dos concepts selecionados durante cada ação para o modelo 3 treinado com a configuração C.

(continua)

Temperatura ambiente em °C	Tarifa da energia elétrica (R\$)	Ação 1	Ação 2	Ação 3	Ação 4	Ação 5	Ação 6	Ação 7
15	1,00	1	2	1	2	2	2	2
15	1,25	1	2	1	2	2	2	2
15	1,50	1	2	2	2	2	2	2
15	1,75	1	2	1	2	2	2	2
15	2,00	1	2	1	2	2	2	2
15	2,25	1	2	1	2	2	2	2
16	1,00	1	2	1	2	2	2	2
16	1,25	1	2	1	2	2	2	2
16	1,50	1	2	1	2	2	2	2
16	1,75	1	2	1	2	2	2	2
16	2,00	1	2	1	2	2	2	2
16	2,25	1	2	1	2	2	2	2
17	1,00	1	2	1	2	2	2	2
17	1,25	1	2	1	2	2	2	2
17	1,50	1	2	1	2	2	2	2
17	1,75	1	2	1	2	2	2	2
17	2,00	1	2	1	2	2	2	2
17	2,25	1	2	1	1	2	2	2
18	1,00	1	2	1	2	2	2	2
18	1,25	1	2	2	2	2	2	2
18	1,50	1	2	1	2	2	2	2
18	1,75	1	2	1	2	2	2	2
18	2,00	1	2	1	2	2	2	2
18	2,25	1	2	1	2	2	2	2
19	1,00	1	2	2	2	2	2	2
19	1,25	1	2	2	2	2	2	2
19	1,50	1	2	2	2	2	2	2
19	1,75	1	3	2	2	2	2	2
19	2,00	1	3	2	2	2	2	2
19	2,25	1	3	2	2	2	2	2
20	1,00	2	2	2	2	2	2	2
20	1,25	1	3	2	2	2	2	2
20	1,50	1	3	2	2	2	2	2
20	1,75	1	3	2	2	2	2	2
20	2,00	1	3	2	2	2	2	2
20	2,25	2	3	2	2	2	2	3



**Tabela A.15** – Número dos concepts selecionados durante cada ação para o modelo 3 treinado com a configuração C.

(continua)

Temperatura ambiente em °C	Tarifa da energia elétrica (R\$)	Ação 1	Ação 2	Ação 3	Ação 4	Ação 5	Ação 6	Ação 7
21	1,00	1	3	2	2	2	2	2
21	1,25	2	3	2	2	2	2	2
21	1,50	2	3	2	2	2	2	2
21	1,75	2	3	2	2	2	2	2
21	2,00	2	3	2	2	2	2	2
21	2,25	1	3	2	2	2	2	3
22	1,00	2	3	2	2	2	3	3
22	1,25	2	3	2	2	1	3	2
22	1,50	2	3	2	3	3	3	3
22	1,75	2	3	2	2	3	3	3
22	2,00	2	3	2	2	3	3	2
22	2,25	2	3	2	3	3	2	1
23	1,00	2	3	2	2	3	3	2
23	1,25	2	3	2	3	3	2	3
23	1,50	2	3	2	3	3	2	3
23	1,75	2	3	3	2	3	3	3
23	2,00	2	3	3	3	3	3	3
23	2,25	2	3	3	3	3	3	3
24	1,00	2	3	3	3	3	3	3
24	1,25	2	3	2	3	3	3	3
24	1,50	2	3	3	3	3	3	3
24	1,75	2	3	3	3	3	3	3
24	2,00	2	3	3	3	3	3	3
24	2,25	2	3	2	3	3	3	3
25	1,00	2	3	3	3	3	3	3
25	1,25	3	3	3	3	3	3	3
25	1,50	2	3	3	3	3	3	3
25	1,75	2	3	3	3	3	3	3
25	2,00	2	3	3	3	3	3	3
25	2,25	2	3	3	3	3	3	3
26	1,00	3	3	3	3	3	3	3
26	1,25	2	3	3	3	3	3	3
26	1,50	3	3	3	3	3	3	3
26	1,75	2	3	3	3	3	3	3
26	2,00	2	3	3	3	3	3	3
26	2,25	2	3	3	3	3	3	3

**Tabela A.15** – Número dos concepts seleccionados durante cada ação para o modelo 3 treinado com a configuração C.

(conclusão)

Temperatura ambiente em °C	Tarifa da energia elétrica (R\$)	Ação 1	Ação 2	Ação 3	Ação 4	Ação 5	Ação 6	Ação 7
27	1,00	3	3	3	3	3	3	3
27	1,25	2	3	3	3	3	3	3
27	1,50	3	3	3	3	3	3	3
27	1,75	2	3	3	3	3	3	3
27	2,00	3	3	3	3	3	3	3
27	2,25	2	3	3	3	3	3	3
28	1,00	2	3	3	3	3	3	3
28	1,25	3	3	3	3	3	3	3
28	1,50	3	3	3	3	3	3	3
28	1,75	3	3	3	3	3	3	3
28	2,00	2	3	3	3	3	3	3
28	2,25	3	3	3	3	3	3	3
29	1,00	3	3	3	3	3	3	3
29	1,25	3	3	3	3	3	3	3
29	1,50	3	3	3	3	3	3	3
29	1,75	3	3	3	3	3	3	3
29	2,00	3	3	3	3	3	3	3
29	2,25	3	3	3	3	3	3	3
30	1,00	3	3	3	3	3	3	3
30	1,25	3	3	3	3	3	3	3
30	1,50	3	3	3	3	3	3	3
30	1,75	3	3	3	3	3	3	3
30	2,00	3	3	3	3	3	3	3
30	2,25	3	3	3	3	3	3	3

**Tabela A.16** – Número dos concepts seleccionados durante cada ação para o modelo 5 treinado com a configuração C.

(continua)

Temperatura ambiente em °C	Tarifa da energia elétrica (R\$)	Ação 1	Ação 2	Ação 3	Ação 4	Ação 5	Ação 6	Ação 7
15	1,00	2	1	1	1	1	1	1
15	1,25	4	1	1	1	1	1	1
15	1,50	3	1	1	1	1	1	1
15	1,75	2	1	1	1	1	1	1
15	2,00	1	1	1	1	1	1	1
15	2,25	3	1	1	1	1	1	1
16	1,00	4	1	1	1	1	1	1
16	1,25	4	1	1	1	1	1	1
16	1,50	4	1	1	1	1	1	1
16	1,75	4	1	1	1	1	1	1
16	2,00	1	1	1	1	1	1	1
16	2,25	4	1	1	1	1	1	1
17	1,00	4	1	1	1	1	1	1
17	1,25	1	1	1	1	1	1	1
17	1,50	3	1	1	1	1	1	1
17	1,75	4	2	1	1	1	1	1
17	2,00	2	1	1	1	1	1	1
17	2,25	2	3	1	1	1	1	1
18	1,00	1	1	1	1	1	4	1
18	1,25	3	1	3	1	1	4	1
18	1,50	4	1	1	4	4	1	1
18	1,75	4	4	1	1	4	1	4
18	2,00	3	1	1	1	1	1	1
18	2,25	4	4	1	1	1	1	1
19	1,00	4	3	1	1	4	1	4
19	1,25	2	3	1	4	1	4	1
19	1,50	3	4	1	1	1	4	1
19	1,75	1	4	1	1	1	4	4
19	2,00	1	1	4	4	4	6	2
19	2,25	3	1	4	1	4	4	4
20	1,00	1	4	4	4	4	4	4
20	1,25	1	4	4	4	1	4	4
20	1,50	3	4	4	6	4	4	4
20	1,75	4	4	4	4	4	4	4
20	2,00	1	4	4	4	1	4	4
20	2,25	4	4	5	4	1	4	3

**Tabela A.16** – Número dos concepts selecionados durante cada ação para o modelo 5 treinado com a configuração C.

(continua)

Temperatura ambiente em °C	Tarifa da energia elétrica (R\$)	Ação 1	Ação 2	Ação 3	Ação 4	Ação 5	Ação 6	Ação 7
21	1,00	2	4	6	5	4	1	4
21	1,25	2	4	2	1	4	1	4
21	1,50	4	4	2	4	4	1	4
21	1,75	4	4	5	5	4	4	4
21	2,00	3	3	5	2	1	4	4
21	2,25	3	4	4	6	4	4	4
22	1,00	4	2	6	4	4	4	4
22	1,25	3	4	5	4	4	4	3
22	1,50	1	4	5	4	6	4	6
22	1,75	1	4	5	6	4	4	5
22	2,00	4	4	4	6	5	4	4
22	2,25	1	3	5	4	4	4	5
23	1,00	4	4	4	4	4	4	4
23	1,25	1	4	4	4	4	4	4
23	1,50	4	4	5	4	6	4	4
23	1,75	4	5	5	4	4	4	4
23	2,00	1	5	5	4	4	4	4
23	2,25	6	4	5	4	4	4	4
24	1,00	4	4	6	5	6	6	5
24	1,25	5	4	4	4	6	3	6
24	1,50	3	4	4	6	6	3	4
24	1,75	4	4	6	4	6	3	3
24	2,00	1	5	5	5	6	5	6
24	2,25	1	3	6	5	6	4	6
25	1,00	1	4	6	5	5	4	4
25	1,25	4	4	6	5	6	6	4
25	1,50	6	5	6	4	6	5	4
25	1,75	1	5	5	4	5	5	6
25	2,00	3	6	5	5	3	3	3
25	2,25	4	5	2	4	6	6	6
26	1,00	1	5	4	6	5	4	3
26	1,25	4	6	6	6	6	6	6
26	1,50	4	6	5	3	4	6	4
26	1,75	4	6	6	5	6	3	4
26	2,00	4	6	5	5	3	6	6
26	2,25	4	6	3	4	6	6	6

**Tabela A.16** – Número dos concepts seleccionados durante cada ação para o modelo 5 treinado com a configuração C.

(conclusão)

Temperatura ambiente em °C	Tarifa da energia elétrica (R\$)	Ação 1	Ação 2	Ação 3	Ação 4	Ação 5	Ação 6	Ação 7
27	1,00	1	5	5	5	3	6	6
27	1,25	1	6	6	6	6	6	6
27	1,50	1	6	2	4	3	5	2
27	1,75	1	5	2	6	6	6	6
27	2,00	1	6	6	6	6	6	6
27	2,25	1	4	6	5	6	6	6
28	1,00	3	6	3	6	6	6	6
28	1,25	1	6	6	6	6	6	6
28	1,50	6	6	6	6	6	6	6
28	1,75	2	6	6	6	6	2	4
28	2,00	1	6	6	6	6	6	2
28	2,25	4	2	6	6	6	6	6
29	1,00	3	6	6	6	6	6	6
29	1,25	1	6	6	6	6	6	6
29	1,50	2	6	6	6	6	6	6
29	1,75	3	6	6	6	6	6	6
29	2,00	2	6	6	6	6	6	6
29	2,25	4	6	6	6	2	3	6
30	1,00	3	6	6	6	6	6	6
30	1,25	6	6	6	6	6	6	6
30	1,50	6	6	6	6	6	6	6
30	1,75	6	6	6	6	6	6	6
30	2,00	3	6	6	6	6	6	6
30	2,25	6	6	6	6	6	6	6

## APÊNDICE B – Implementação em Python do modelo 3 da configuração de chuveiro B com o algoritmo PPO

Das Figuras B.1 a B.15 é apresentado o código utilizado na implementação do modelo computacional de chuveiro 3 para a configuração de planta B. Em seguida, as Figuras B.16 a B.36 mostram o código que realiza a construção do ambiente de RL e o treinamento e avaliação do agente. O código completo está disponível no *link* do *GitHub* em (Zamin, 2023).

**Figura B.1** – Pacotes importados para a simulação do modelo de chuveiro.

```
import math
import numpy as np
from math import sqrt
from scipy.integrate import solve_ivp
```

**Figura B.2** – Implementação do controlador PID.

```
def controle_pid(SP, PV, j, I_buffer, D_buffer, dt, Kp, Ti, Td, b,
c, N, UU_bias, UU_min, UU_max, metodo="backward"):
    """Calcula a saída do controlador PID e seus parâmetros.

    Argumentos:
        SP (float): Setpoint da variável.
        PV (float): Valor da variável medida.
        j (int): Buffer para armazenar o passo de tempo contado em
números inteiros.
        I_buffer: Buffer para armazenar os valores da ação integral.
        D_buffer: Buffer para armazenar os valores da ação
derivativa.
        dt (float): Passo de tempo.
        Kp (float): Ganho proporcional do controlador.
        Ti (float): Ganho integral do controlador.
        Td (float): Ganho derivativo do controlador:
        b (float): Peso do setpoint na ação proporcional do
controlador.
        c (float): Peso da derivada do setpoint na ação derivativa
do controlador.
        N (int): Fator de proporção da constante do filtro de 1ª
ordem para reduzir o efeito do ruído e tornar o controlador causal.
        UU_bias (float): Valor base (ou bias) da ação de controle.
Esse valor é somado ao valor de saída Uop calculado pelo
controlador para determinar a ação de controle total do controlador.
        UU_min (float): Valor mínimo que a variável pode atingir.
        UU_max (float): Valor máximo que a variável pode atingir.
```

**Figura B.3** – Implementação do controlador PID.

metodo (string): Método para resolver o controlador PID. As opções são: Backward, Forward, Tustin e Ramp.

Retorna:

numpy.array(): Array do numpy contendo o valor de saída do PID e os parâmetros calculados.

```

"""
# Define o método para o controlador PID:
if metodo == "backward":
    b1 = Kp * dt / Ti if Ti != 0 else 0.0
    b2 = 0.0
    ad = Td / (Td + N * dt)
    bd = Kp * Td * N / (Td + N * dt)

elif metodo == "forward":
    b1 = 0.0
    b2 = Kp * dt / Ti if Ti != 0 else 0.0
    ad = 1 - N * dt / Td if Td != 0 else 0.0
    bd = Kp * N

elif metodo == "tustin":
    b1 = Kp * dt / 2 / Ti if Ti != 0 else 0.0
    b2 = b1
    ad = (2 * Td - N * dt) / (2 * Td + N * dt)
    bd = 2 * Kp * Td * N / (2 * Td + N * dt)

elif metodo == "ramp":
    b1 = Kp * dt / 2 / Ti if Ti != 0 else 0.0
    b2 = b1
    ad = np.exp(-N * dt / Td) if Td != 0 else 0.0
    bd = Kp * Td * (1 - ad) / dt

# Ação derivativa:
D = ad * D_buffer + bd * ((c * SP[j] - PV[j]) - (c * SP[j-1] -
PV[j-1]))

# Ação integral:
II = b1 * (SP[j] - PV[j]) + b2 * (SP[j-1] - PV[j-1])
I = I_buffer + II

# Ação proporcional:
P = Kp * (b * SP[j] - PV[j])

# Saída do controlador PID:
Uop = UU_bias + P + I + D

```

**Figura B.4** – Implementação do controlador PID.

```

# Implementa anti-reset windup:
if Uop < UU_min:
    II = 0.0
    Uop = UU_min
if Uop > UU_max:
    II = 0.0
    Uop = UU_max

I = I_buffer + II

# Retorna a saída e os parâmetros do controlador:
return np.array([Uop, I, D])

```

**Figura B.5** – Implementação do modelo computacional de chuveiro.

```

def controle_boiler_on_off(Uop, medido, minimo, maximo, xq=0.2):
    """Controle liga e desliga do boiler.

    Argumentos:
        Uop (int): Variável de controle do seletor do boiler (Sa).
        medido (float): Valor medido para a temperatura do boiler
        (Tq).
        minimo (float): Valor mínimo para a temperatura do boiler.
        maximo (float): Valor máximo para a temperatura do boiler.
        xq (float): Abertura da válvula de corrente quente.

    Retorna:
        Uop (int): Variável de controle do seletor do boiler (Sa).
    """
    if medido < minimo:
        Uop = 1
    elif medido > maximo:
        Uop = 0

    # Desliga o sistema de aquecimento se não estiver passando água:
    if xq < 0.15:
        Uop = 0

    return Uop

def modelo_alimentacao_quente(xf, xq):
    """Modelo para a vazão válvula de alimentação da corrente quente
    (Fq).

    Argumentos:
        xf (float): Abertura da válvula de corrente fria.

```



**Figura B.6** – Implementação do modelo computacional de chuveiro.

```

xq (float): Abertura da válvula de corrente quente.

Retorna:
    Fq (float): Vazão da válvula da corrente quente.
    """

    t = 60 * (xq ** (8 / 5))
    n = -1 * (-180 - 45 * (xf ** 2) - 250 * (xq ** (16 / 5)) * (xf
** 2) - 1180 * (xq ** (16 / 5)) + 60 * sqrt(9 * (xq ** (16 / 5)) *
(xf ** 2) + 50 * (xq ** (32 / 5)) * (xf ** 2)))

    d = 6552 * (xq ** (16 / 5)) * (xf ** 2) + 648 * (xf ** 2) +
16400 * (xq ** (32 / 5)) * (xf ** 2) + 900 * (xq ** (16 / 5)) * (xf
** 4) + 2500 * (xq ** (32 / 5)) * (xf ** 4) + 81 * (xf ** 4) + 55696
* (xq ** (32 / 5)) + 16992 * (xq ** (16 / 5)) + 1296

    Fq = t * sqrt(n / d)
    return Fq

def modelo_alimentacao_fria(xf, Fq):
    """Modelo para a vazão da válvula de alimentação da corrente
fria (Ff).

    Argumentos:
        xf (float): Abertura da válvula de corrente fria.
        Fq (float): Vazão da válvula da corrente quente.

    Retorna:
        Ff (float): Abertura da válvula da corrente fria.
        """
    n = 2 * xf * sqrt(125 * xf * xf - Fq * Fq + 500) - xf * Fq
    d = (xf ** 2) + 4

    Ff = n / d
    return Ff

def modelo_valvula_saida(xs):
    """Modelo para a vazão válvula de saída (Fs).

    Argumentos:
        xs (float): Abertura da válvula de saída.

    Retorna:
        Fs (float): Vazão da válvula de saída.
        """
    n = 5 * (xs ** 3) * (sqrt(30)) * sqrt(-15 * (xs ** 6) +
sqrt(6625 * (xs ** 12) + 640 * (xs ** 6) + 16))

```

**Figura B.7** – Implementação do modelo computacional de chuveiro.

```

d = 20 * (xs ** 6) + 1

Fs = n / d
return Fs

def modelo_temperatura_boiler(t, Tq, Fq, Tf, Sa):
    """Modelo para a temperatura de aquecimento do boiler (Tq).

    Argumentos:
        t (float): Tempo para ser utilizado na solução da equação
diferencial ordinária pelo integrador do scipy.
        Tq (float): Temperatura de aquecimento do boiler.
        Fq (float): Vazão da válvula do boiler.
        Tf (float): Temperatura da corrente fria.
        Sa (float): Seletor do aquecimento do boiler.

    Retorna:
        Equação diferencial para o modelo do boiler.
    """
    return Fq * (Tf - Tq) + 250 * Sa

def modelo_nivel_tanque(t, h, Ff, Fq, Fd, Fs):
    """Modelo para o nível do tanque (h).

    Argumentos:
        t (float): Tempo para ser utilizado na solução da equação
diferencial ordinária pelo integrador do scipy.
        h (float): nível do tanque de aquecimento da mistura entre
corrente quente e fria.
        Fq (float): Vazão da válvula da corrente quente.
        Fd (float): Vazão da válvula do distúrbio no tanque.
        Fs (float): Vazão da válvula de saída.

    Retorna:
        Equação diferencial para o modelo do nível do tanque.
    """
    return (1 / 0.5) * (Ff + Fq + Fd - Fs)

```

**Figura B.8** – Implementação do modelo computacional de chuveiro.

```
def modelo_temperatura_tanque(t, Tt, Ff, Tf, Fq, Tq, Fd, Td, Sr, h):
    """Modelo para a temperatura de saída do tanque de aquecimento
    (Tt).

    Argumentos:
        t (float): Tempo para ser utilizado na solução da equação
        diferencial ordinária pelo integrador do scipy.

        Tt (float): Temperatura de saída do tanque de aquecimento da
        mistura entre corrente quente e fria.
        Ff (float): Vazão da válvula da corrente fria.
        Tf (float): Temperatura da corrente fria.
        Fq (float): Vazão da válvula da corrente quente.
        Tq (float): Temperatura de aquecimento do Boiler.

        Fd (float): Vazão da válvula do distúrbio no tanque.
        Td (float): Temperatura da corrente de distúrbio.
        Sr (float): Seletor da resistência elétrica do tanque de
        aquecimento.
        h (float): nível do tanque de aquecimento da mistura entre
        corrente quente e fria.

    Retorna:
        Equação diferencial para o modelo da temperatura de saída do
        tanque.
        """
    return (1 / (0.5 * h)) * (Ff * (Tf - Tt) + Fq * (Tq - Tt) + Fd *
    (Td - Tt) + 80 * Sr)

def modelo_temperatura_saida(t, Ts, Tt, Fs, Tinf):
    """Modelo para a temperatura de saída do chuveiro (Ts).

    Argumentos:
        t (float): Tempo para ser utilizado na solução da equação
        diferencial ordinária pelo integrador do scipy.
        Ts (float): Temperatura de saída do chuveiro.
        Fs (float): Vazão da válvula de saída.
        Tinf (float): Temperatura ambiente.

    Retorna:
        Equação diferencial para o modelo da temperatura de saída do
        chuveiro.
        """
    return (Fs / 2.5) * (Tt - Ts) - 0.8 * (((Tt - Tinf) * (Ts -
    Tinf) * (0.5 * (Tt - Tinf) + 0.5 * (Ts - Tinf))) ** (1 / 3))
```

**Figura B.9** – Implementação do modelo computacional de chuveiro.

```

def modelagem_sistema(t, Y, Sa, xf, xq, xs, Tf, Td, Tinf, Fd, Sr):
    """Reúne as equações diferenciais para a modelagem do sistema com
    tempo morto.

    Argumentos:
        t (float): Tempo para ser utilizado na solução da equação
        diferencial ordinária pelo integrador do scipy.
        Y (numpy.array): Condições iniciais para nível do tanque de
        aquecimento (h), temperatura do boiler (Tq), temperatura de saída do
        tanque de aquecimento (Tt), temperatura de saída do chuveiro (Ts).
        xf (float): Abertura da válvula de corrente fria.
        Sa (float): Seletor do aquecimento do boiler.
        xq (float): Abertura da válvula de corrente quente.
        xs (float): Abertura da válvula de saída.

        Tf (float): Temperatura da corrente fria.
        Td (float): Temperatura da corrente de distúrbio.
        Tinf (float): Temperatura ambiente.
        Fd (float): Vazão da válvula do distúrbio no tanque.
        Sr (float): Seletor da resistência elétrica do tanque de
        aquecimento.

    Retorna:
        numpy.array(): Equações diferenciais ordinárias para serem
        resolvidas pelo integrador do scipy.
    """
    N = 50
    Vt = 5
    UAt = 0.01
    Fq = modelo_alimentacao_quente(xf, xq)
    Ff = modelo_alimentacao_fria(xf, Fq)
    Fs = modelo_valvula_saida(xs)

    Tq = Y[0]
    h = Y[1]
    Tt = Y[2]

    EDOh = modelo_nivel_tanque(t, h, Ff, Fq, Fd, Fs)
    EDOTq = modelo_temperatura_boiler(t, Tq, Fq, Tf, Sa)
    EDOTt = modelo_temperatura_tanque(t, Tt, Ff, Tf, Fq, Tq, Fd, Td,
    Sr, h)

    return np.array([EDOTq, EDOh, EDOTt] + list((N - 1) / Vt * (Fs *
    (Y[2:N+1] - Y[3:N+2]) - UAt * (Y[3:N+2] - Tinf))))

```

**Figura B.10** – Implementação do modelo computacional de chuveiro.

```

def simulacao_malha_temperatura(SYS, Y0, UT, dt, I_buffer, D_buffer,
Tinf, split_range):
    """Modelagem do sistema com malha de inventário para nível do
tanque, controle liga-desliga (on-off) para a temperatura do boiler
(Tq) e malha cascata para temperatura de saída com split-range em
Sr.

    Argumentos:
        SYS: Função contendo as equações diferenciais para a solução
do problema.
        Y0 (numpy.array): Array contendo as condições iniciais para
nível do tanque de aquecimento (h), temperatura do boiler (Tq),
temperatura de saída do tanque de aquecimento (Tt), temperatura de
saída do chuveiro (Ts).

        UT (numpy.array): Array contendo a matriz de variáveis e
perturbações.
        dt (float): Intervalo de tempo.
        I_buffer (float): Buffer para armazenar os valores da ação
integral.
        D_buffer (float): Buffer para armazenar os valores da ação
proporcional.
        Tinf (float): Temperatura ambiente, utilizada para definir
valor de UU_bias para a temperatura de saída.
        split_range (int): Indica se deve ser utilizado o split-
range ou não. Um valor de 0 indica que não será utilizado, enquanto
que um valor de 1 indica que o split-range deve ser utilizado.

    Retorna:
        tuple: Tuple com intervalo de tempo, as variáveis após a
solução do sistema e o buffer do ajuste PID.
    """
    # Instantes de tempo:
    tempo_inicial = UT[0][0]
    tempo_final = UT[1][0]

    # Armazenamento dos dados resultantes da simulação:
    TT = np.arange(start=tempo_inicial, stop=tempo_final + dt,
step=dt, dtype="float")
    nt = np.size(TT)
    nu = np.size(UT, 1)
    ny = 4

```

**Figura B.11** – Implementação do modelo computacional de chuveiro.

```

# Número correspondente a cada malha de controle em relação a
posição do vetor Y0:
# 0 - malha boiler, 1 - malha nível, 2 - malha tanque, 3 - malha
saída:
id = [0, 1, 2, -1]

# Matriz para armazenar valores das variáveis de saída para cada
instante de tempo inicializando com os valores:
# Inicializada com valores relativos a condição inicial Y0:
YY = np.ones((nt, ny))@np.diag(Y0[id])
SP = np.ones((nt, ny))@np.diag(Y0[id])

# Matriz para armazenar valores das variáveis de entrada para
cada instante de tempo:
UU = np.zeros((nt, nu-1))

# Parâmetros das malhas de controle:
Kp = np.array([1, 0.3, 2.0, 0.51])

Ti = np.array([1, 0.8, 2.75, 0.6])
Td = np.array([1, 0, 0, 0.05])
b = np.array([1, 1, 1, 0.8])
c = np.array([0, 0, 0, 0.1])
UU_bias = np.array([0, 0.25, 0.25, Tinf])
UU_min = np.array([0, 0.01, 0.01, 20])
UU_max = np.array([1, 0.99, 0.99, 40])
N = 10

# Definindo o split-range:
if split_range == 1:
    UU_max[2] = 1.99

# Executando a simulação:
ii = 0

for k in np.arange(nt - 1):
    if TT[k] >= UT[ii+1, 0]:
        ii = ii + 1

    # Inicializando os valores:
    UU[k,:] = UT[ii, 1:nu]

    # Controle liga e desliga do boiler: malha 0
    SP[k, 0] = UU[k, 0]
    UU[k, 0] = controle_boiler_on_off(UU[k-1, 0], YY[k, 0],
SP[k, 0] - 1, SP[k, 0] + 1, xq=UU[k, 2])

```

**Figura B.12** – Implementação do modelo computacional de chuveiro.

```

# Malhas de controle PIDs:
for jj in [1, 2]:

    # Malha cascata:
    if jj == 2:
        # Setpoint da malha externa:
        SP[k, -1] = UU[k, jj]
        uu = controle_pid(SP[:, -1], YY[:, -1], k, I_buffer[-1],
            D_buffer[-1], dt, Kp[-1], Ti[-1], Td[-1], b[-1], c[-1],
            N, UU_bias = UU_bias[-1], UU_min =
UU_min[-1], UU_max = UU_max[-1])
        SP[k, jj] = uu[0]
        I_buffer[-1] = uu[1]
        D_buffer[-1] = uu[2]
    else:
        SP[k, jj] = UU[k, jj]

        uu = controle_pid(SP[:, jj], YY[:, jj], k, I_buffer[jj],
            D_buffer[jj], dt, Kp[jj], Ti[jj], Td[jj], b[jj], c[jj],
            N, UU_bias = UU_bias[jj], UU_min =
UU_min[jj], UU_max = UU_max[jj])

    # Split-range para controlar a temperatura do tanque:
    if jj == 2 and uu[0] > 1:
        UU[k, jj] = 1
        UU[k, -1] = uu[0] - 1
    else:
        UU[k, jj] = uu[0]

    I_buffer[jj] = uu[1]
    D_buffer[jj] = uu[2]

    # Integração do sistema:
    sol = solve_ivp(SYS, [TT[k], TT[k+1]], Y0,
args=tuple(UU[k]), atol = 1e-8, rtol = 1e-8)

    # Armazenamento dos valores calculados:
    Y0 = sol.y[:, -1]
    YY[k+1, :] = sol.y[id, -1]

UU[k+1, :] = UU[k, :]
SP[k+1, :] = SP[k, :]

return (TT, YY, UU, Y0, I_buffer, D_buffer)

```

**Figura B.13** – Cálculo do Índice de Qualidade do Banho (IQB).

```

def calculo_iqb(Ts, Fs):
    """Calcula o índice de qualidade do banho (iqb).

    Argumentos:
        Ts (float): Temperatura de saída do chuveiro.
        Fs (float): Vazão da válvula de saída.

    Retorna:
        iqb (float): Índice de qualidade do banho.
    """
    iqb = (1 / math.e) * math.exp((1 - ((Ts - 38 + 0.02 * Fs) / 2)
** 2) * np.power((0.506 + math.log10(math.log10((10000 *
np.sqrt(Fs)) / (10 + Fs + 0.004 * np.power(Fs, 4))))), 20))

    return iqb

```

**Figura B.14** – Cálculo dos custos do banho.

```

def custo_eletrico_banho(Sr, potencia_eletrica, custo_eletrico_kwh,
dt):
    """Calcula o custo da parte elétrica do banho.

    Argumentos:
        Sr (float): Seletor da resistência elétrica do tanque de
aquecimento.
        potencial_eletrica (float): Potência elétrica do tanque de
aquecimento (chuveiro) em kW.
        custo_eletrico_kwh (float): Custo do kWh da energia em reais
por hora.
        dt (float): Passo de tempo.

    Retorna:
        custo_eletrico_total (float): Custo da energia elétrica do
banho em reais.
    """
    Sr_utilizado = np.trapz(y=Sr, dx=dt)
    custo_eletrico_total = potencia_eletrica * Sr_utilizado *
custo_eletrico_kwh / 60

    return custo_eletrico_total

```



**Figura B.15** – Cálculo dos custos do banho.

```

def custo_gas_banho(Sa, potencia_aquecedor, custo_gas_kg, dt):
    """Calcula o custo do gás do banho.

    Argumentos:
        Sa (float): Seletor do aquecimento do boiler.
        potencia_aquecedor (float): Potência do aquecedor (boiler)
em kcal/h.
        custo_gas_kg (float): Custo do kg do gás em reais por kg.
        tempo (float): Tempo da ação em minutos.
        df (float): Passo de tempo da simulação.

    Retorna:
        custo_gas_total (float): Custo do gás do banho em reais.
    """
    Sa_utilizado = np.trapz(y=Sa, dx=dt)
    rendimento = 1
    potencia_util = potencia_aquecedor * rendimento
    kcal_fornecida_no_banho = potencia_util * Sa_utilizado / 60
    kg_equivalente_kcal = 11750
    quantidade_gas_kg = kcal_fornecida_no_banho / kg_equivalente_kcal
    custo_gas_total = custo_gas_kg * quantidade_gas_kg

    return custo_gas_total

def custo_agua_banho(Fs, custo_agua_m3, tempo):
    """Calcula o custo de água do banho.

    Argumentos:
        xs (float): Vazão da válvula de saída.
        custo_agua_m3 (float): Custo da água em m3 por reais.
        tempo (float): Tempo da ação em minutos.

    Retorna:
        custo_agua_total: Custo da água do banho em reais.
    """
    custo_agua_total = ((Fs * tempo) / 1000) * custo_agua_m3

    return custo_agua_total

```

**Figura B.16** – Pacotes importados para o treinamento de RL.

```

import ray
import ray.rllib.algorithms.ppo as ppo
import ray.rllib.algorithms.sac as sac
from ray.rllib.algorithms.algorithm import Algorithm
from ray.rllib.policy.policy import Policy

import argparse
import itertools
import random
import os
import glob
import gymnasium as gym
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
from itertools import accumulate

from controle_temperatura_saida import simulacao_malha_temperatura
from controle_temperatura_saida import modelagem_sistema
from controle_temperatura_saida import modelo_valvula_saida
from controle_temperatura_saida import calculo_iqb
from controle_temperatura_saida import custo_eletrico_banho
from controle_temperatura_saida import custo_gas_banho
from controle_temperatura_saida import custo_agua_banho

```

**Figura B.17** – Construção do ambiente de RL.

```

class ShowerEnv(gym.Env):
    """Ambiente para simulação do modelo de chuveiro."""

    def __init__(self, env_config):
        # Temperatura ambiente, algoritmo, custo da energia elétrica
        # em kWh, selective concept learned e modelo:
        self.Tinf_list = env_config["Tinf_list"]
        self.nome_algoritmo = env_config["nome_algoritmo"]
        self.custo_eletrico_kwh_list =
env_config["custo_eletrico_kwh_list"]
        self.selector = env_config["selector"]
        self.model = env_config["model"]

        # Tempo de simulação e de cada iteração:
        self.dt = 0.01
        self.tempo_iteracao = 2

```

**Figura B.18** – Construção do ambiente de RL.

```

# Utiliza split-range:
self.Sr = 0

# Potência da resistência elétrica em kW:
self.potencia_eletrica = 5.5

# Potência do aquecedor boiler em kcal/h:
self.potencia_aquecedor = 29000

# Custo do kg do gás e do m3 da água:
self.custo_gas_kg = 3
self.custo_agua_m3 = 4

# Selective concept learned seleciona qual concept treinado
será utilizado:
if self.selector == True:
    self.action_space = gym.spaces.Discrete(3)
    self.models = []

    for i in self.model:
        self.models.append(Policy.from_checkpoint(glob.glob(
i+"/*")[-1])['default_policy'])
        self.model = self.models

# Ações - SPTs, SPTq, xs, split-range:
else:
    self.action_space = gym.spaces.Tuple(
gym.spaces.Box(low=30, high=40, shape=(1,), dtype=np.float32),
gym.spaces.Box(low=30, high=70, shape=(1,), dtype=np.float32),
gym.spaces.Box(low=0.01, high=0.99, shape=(1,), dtype=np.float32),
gym.spaces.Discrete(2, start=0),,)

# Estados - Ts, Tq, Tt, h, Fs, xf, xq, iqb, Tinf:
self.observation_space = gym.spaces.Box(
    low=np.array([0, 0, 0, 0, 0, 0, 0, 0, 10]),
    high=np.array([100, 100, 100, 10000, 100, 1, 1, 1, 35]),
    dtype=np.float32,
)

def reset(self):

# Temperatura ambiente e custo da energia elétrica em kWh:
self.Tinf = random.choice(self.Tinf_list)
self.custo_eletrico_kwh =
random.choice(self.custo_eletrico_kwh_list)

```

**Figura B.19** – Construção do ambiente de RL.

```
# Distúrbios Fd e Td, temperatura da corrente fria Tf:
self.Fd = 0
self.Td = self.Tinf
self.Tf = self.Tinf

# Tempo inicial:
self.tempo_inicial = 0

# Nível do tanque de aquecimento e setpoint:
self.h = 80
self.SPh = 80

# Temperatura de saída:
self.Ts = self.Tinf

# Temperatura do boiler:
self.Tq = 55

# Temperatura do tanque:
self.Tt = self.Tinf

# Vazão de saída:
self.Fs = 0

# Abertura da válvula quente:
self.xq = 0

# Abertura da válvula fria:
self.xf = 0

# Índice de qualidade do banho:
self.iqb = 0

# Custo elétrico do banho:
self.custo_eletrico = 0

# Custo do gás do banho:
self.custo_gas = 0

# Custo da água do banho:
self.custo_agua = 0

# Condições iniciais - Tq, h, Tt, Ts:
self.Y0 = np.array([self.Tq, self.h] + 50 * [self.Tinf])
```

**Figura B.20** – Construção do ambiente de RL.

```

# Define o buffer para os ganhos integral e derivativo das malhas de
controle:
    # 0 - malha boiler, 1 - malha nível, 2 - malha tanque, 3 -
malha saída
    id = [0, 1, 2, -1]
    self.Kp = np.array([1, 0.3, 2.0, 0.51])
    self.b = np.array([1, 1, 1, 0.8])
    self.I_buffer = self.Kp * self.Y0[id] * (1 - self.b)
    self.D_buffer = np.array([0, 0, 0, 0])

    # Estados - Ts, Tq, Tt, h, Fs, xf, xq, iqb, Tinf:
    self.obs = np.array([self.Ts, self.Tq, self.Tt, self.h,
self.Fs, self.xf, self.xq, self.iqb, self.Tinf],
                        dtype=np.float32)

    return self.obs

def step(self, action):

    # Tempo de cada iteração:
    self.tempo_final = self.tempo_inicial + self.tempo_iteracao

    if self.selector == True:
        actions =
self.model[action].compute_single_action(self.obs, explore=False)[0]

        # Ambos os algoritmos realizam uma normalização interna
antes de o agente selecionar as ações.
        # Logo, é preciso reverter essa normalização, pois ela
não é feita automaticamente.

        # Setpoint da temperatura de saída:
        self.SPTs = round((actions[0][0] * np.std([30, 40])) +
np.mean([30, 40]), 2)
        if self.SPTs > 40:
            self.SPTs = 40
        if self.SPTs < 30:
            self.SPTs = 30

        # Setpoint da temperatura do boiler:
        self.SPTq = round((actions[1][0] * np.std([30, 70])) +
np.mean([30, 70]), 1)
        if self.SPTq > 70:
            self.SPTq = 70
        if self.SPTq < 30:
            self.SPTq = 30

```

**Figura B.21** – Construção do ambiente de RL.

```

# Abertura da válvula de saída:
self.xs = round((actions[2][0] * np.std([0.01, 0.99])) +
np.mean([0.01, 0.99]), 2)
if self.xs > 0.99:
    self.xs = 0.99
if self.xs < 0.01:
    self.xs = 0.01

# Split-range:
self.split_range = actions[3]

else
# Setpoint da temperatura de saída:
self.SPTs = round(action[0][0], 2)

# Fração de aquecimento do boiler:
self.SPTq = round(action[1][0], 1)

# Abertura da válvula de saída:
self.xs = round(action[2][0], 2)

# Split-range:
self.split_range = action[3]

# Variáveis para simulação - tempo, SPTq, SPh, xq, xs, Tf,
Td, Tinf, Fd, Sr:
self.UT = np.array([[self.tempo_inicial, self.SPTq,
self.SPh, self.SPTs, self.xs, self.Tf, self.Td, self.Tinf, self.Fd,
self.Sr],[self.tempo_final, self.SPTq, self.SPh, self.SPTs, self.xs,
self.Tf, self.Td, self.Tinf, self.Fd, self.Sr]])

# Solução do sistema:
self.TT, self.YY, self.UU, self.Y0, self.I_buffer,
self.D_buffer = simulacao_malha_temperatura(modelagem_sistema,
self.Y0, self.UT, self.dt, self.I_buffer, self.D_buffer, self.Tinf,
self.split_range)

# Valor final da temperatura do boiler:
self.Tq = self.YY[:,0][-1]

# Valor final do nível do tanque:
self.h = self.YY[:,1][-1]

# Valor final da temperatura do tanque:
self.Tt = self.YY[:,2][-1]

```

**Figura B.22** – Construção do ambiente de RL.

```

# Valor final da temperatura de saída:
self.Ts = self.YY[:,3][-1]

# Fração do aquecedor do boiler utilizada durante a
iteração:
self.Sa_total = self.UU[:,0]

# Fração da resistência elétrica utilizada durante a
iteração:
self.Sr_total = self.UU[:,8]

# Valor final da abertura de corrente fria:
self.xf = self.UU[:,1][-1]

# Valor final da abertura de corrente quente:
self.xq = self.UU[:,2][-1]

# Valor final da abertura da válvula de saída:
self.xs = self.UU[:,3][-1]

# Valor final da vazão de saída:
self.Fs = modelo_valvula_saida(self.xs)

# Cálculo do índice de qualidade do banho:
self.iqb = calculo_iqb(self.Ts, self.Fs)

# Cálculo do custo elétrico do banho:
self.custo_eletrico = custo_eletrico_banho(self.Sr_total,
self.potencia_eletrica, self.custo_eletrico_kwh, self.dt)

# Cálculo do custo de gás do banho:
self.custo_gas = custo_gas_banho(self.Sa_total,
self.potencia_aquecedor, self.custo_gas_kg, self.dt)

# Cálculo do custo da água:
self.custo_agua = custo_agua_banho(self.Fs,
self.custo_agua_m3, self.tempo_iteracao)

# Estados - Ts, Tq, Tt, h, Fs, xf, xq, iqb, Tinf:
self.obs = np.array([self.Ts, self.Tq, self.Tt, self.h,
self.Fs, self.xf, self.xq, self.iqb, self.Tinf], dtype=np.float32)

# Define a recompensa:
reward = self.iqb

```

**Figura B.23** – Construção do ambiente de RL.

```

# Incrementa tempo inicial:
self.tempo_inicial = self.tempo_inicial +
self.tempo_iteracao

# Para visualização:
self.SPTq_total = np.repeat(self.SPTq, 201)
self.Tq_total = self.YY[:,0]
self.SPh_total = np.repeat(self.SPh, 201)
self.h_total = self.YY[:,1]
self.Tt_total = self.YY[:,2]
self.SPTs_total = np.repeat(self.SPTs, 201)
self.Ts_total = self.YY[:,3]
self.xq_total = self.UU[:,2]
self.xf_total = self.UU[:,1]
self.xs_total = np.repeat(self.xs, 201)
self.Fs_total = np.repeat(self.Fs, 201)
self.Fd_total = np.repeat(self.Fd, 201)
self.Td_total = np.repeat(self.Td, 201)
self.Tf_total = np.repeat(self.Tf, 201)
self.Tinf_total = np.repeat(self.Tinf, 201)
self.split_range_total = np.repeat(self.split_range, 201)

info = {"SPTq": self.SPTq_total, "Tq": self.Tq_total, "SPh":
self.SPh_total, "h": self.h_total, "Tt": self.Tt_total, "SPTs":
self.SPTs_total, "Ts": self.Ts_total, "Sr": self.Sr_total, "Sa":
self.Sa_total, "xq": self.xq_total, "xf": self.xf_total, "xs":
self.xs_total, "Fs": self.Fs_total, "iqb": self.iqb,
"custo_eletrico": self.custo_eletrico, "custo_gas": self.custo_gas,
"custo_agua": self.custo_agua, "recompensa": reward,
"custo_eletrico_kwh": self.custo_eletrico_kwh, "Fd": self.Fd_total,
"Td": self.Td_total, "Tf": self.Tf_total, "Tinf": self.Tinf_total,
"split_range": self.split_range_total,}

# Termina o episódio se o tempo for maior que 14 ou se o
nível do tanque ultrapassar 100:
done = False
if self.tempo_final == 14 or self.h > 100:
    done = True

return self.obs, reward, done, info

def render(self):
    pass

```



**Figura B.24** – Treinamento do agente de RL.

```

def treina_agente(nome_algoritmo, n_iter_agente, n_iter_checkpoints,
concept, selector=False, model=None):

    # Define o local para salvar o modelo treinado e os checkpoints:
    path_root_models = "/models/"
    path_root = os.getcwd() + path_root_models
    path = path_root + "results_" + nome_algoritmo + "_concept_" +
concept

    # Define os concepts:
    if concept == "banho_dia_frio":
        Tinf_list = [15, 16, 17, 18, 19]
        custo_eletrico_kwh_list = [1, 1.25, 1.5, 1.75, 2, 2.25]

    if concept == "banho_dia_ameno":
        Tinf_list = [20, 21, 22, 23, 24]
        custo_eletrico_kwh_list = [1, 1.25, 1.5, 1.75, 2, 2.25]

    if concept == "banho_dia_quente":
        Tinf_list = [25, 26, 27, 28, 29, 30]
        custo_eletrico_kwh_list = [1, 1.25, 1.5, 1.75, 2, 2.25]

    if concept == "seleciona_banho":
        Tinf_list = [15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26,
27, 28, 29, 30]
        custo_eletrico_kwh_list = [1, 1.25, 1.5, 1.75, 2, 2.25]

    # Define as configurações para o algoritmo e constrói o agente:
    config = ppo.PPOConfig()

    # Constrói o agente:
    config.environment(env=ShowerEnv, env_config={"Tinf_list":
Tinf_list, "nome_algoritmo": nome_algoritmo,
"custo_eletrico_kwh_list": custo_eletrico_kwh_list, "selector":
selector, "model": model})
    agent = config.build()

    # Armazena resultados:
    results = []
    episode_data = []

    # Realiza o treinamento:
    for n in range(1, n_iter_agente):

        # Treina o agente:
        result = agent.train()

```

**Figura B.25** – Treinamento do agente de RL.

```

results.append(result)

# Armazena dados do episódio:
episode = {
    "n": n,
    "episode_reward_min": result["episode_reward_min"],
    "episode_reward_mean": result["episode_reward_mean"],
    "episode_reward_max": result["episode_reward_max"],
    "episode_len_mean": result["episode_len_mean"],
}
episode_data.append(episode)

# Salva checkpoint a cada n_iter_checkpoints iterações:
if n % n_iter_checkpoints == 0:
    file_name = agent.save(path)

df = pd.DataFrame(data=episode_data)
df.to_csv(path + "_episode_data" + ".csv")

return path

```

**Figura B.26** – Avaliação do agente de RL.

```

def avalia_agente(nome_algoritmo, Tinf_list,
custo_eletrico_kwh_list, selector=True):

    # Temperatura ambiente e custo da energia elétrica:
    Tinf_var = str(Tinf_list[0]).replace(".", "-")
    custo_eletrico_kwh_var =
str(custo_eletrico_kwh_list[0]).replace(".", "-")
    Tinf_num = Tinf_list[0]
    custo_eletrico_kwh_num = custo_eletrico_kwh_list[0]

    # Define o local do checkpoint salvo:
    path_root_models = "/models/"
    path_root = os.getcwd() + path_root_models
    path = path_root + "results_" + nome_algoritmo + "_concept_"

    banho_dia_frio = path + "banho_dia_frio"
    banho_dia_ameno = path + "banho_dia_ameno"
    banho_dia_quente = path + "banho_dia_quente"
    selector_path = path + "seleciona_banho"
    model = [banho_dia_frio, banho_dia_ameno, banho_dia_quente]

```

**Figura B.27** – Avaliação do agente de RL.

```

# Constrói o ambiente:
env = ShowerEnv({"Tinf_list": Tinf_list, "nome_algoritmo":
nome_algoritmo, "custo_eletrico_kwh_list": custo_eletrico_kwh_list,
"selector": selector, "model": model})
obs = env.reset()

# Define se será utilizado o selective concept learned
(selector=True) ou programmed (selector=False):
if selector == True:
    agent = Algorithm.from_checkpoint(glob.glob(selector_path
+ "/" + "*" ) [-1])

if selector == False:
    if Tinf_num < 20:
        agent =
Algorithm.from_checkpoint(glob.glob(banho_dia_frio + "/" + "*" ) [-1])
    elif Tinf_num >= 20 and Tinf_num < 25:
        agent =
Algorithm.from_checkpoint(glob.glob(banho_dia_ameno + "/" + "*" ) [-1])
    elif Tinf_num >= 25:
        agent =
Algorithm.from_checkpoint(glob.glob(banho_dia_quente + "/" + "*" ) [-1])

# Para visualização:
SPTq_list = []
Tq_list = []
SPh_list = []
h_list = []
Tt_list = []
SPTs_list = []
Ts_list = []
split_range_list = []
Sr_list = []
Sa_list = []
xq_list = []
xf_list = []
xs_list = []
Fs_list = []
iqb_list = []
custo_eletrico_list = []
custo_gas_list = []
custo_agua_list = []
recompensa_list = []
Fd_list = []
Td_list = []
Tf_list = []

```

**Figura B.28** – Avaliação do agente de RL.

```

Tinf_list = []
tempo_total = np.arange(start=0, stop=14 + 0.07, step=0.01,
dtype="float")
tempo_acoes = np.arange(start=1, stop=8, step=1, dtype="int")
concepts_selecionados_list = []

# Roda o episódio com as ações sugeridas pelo agente treinado:
for i in range(0, 1):

    episode_reward = 0
    print(f"Episódio {i}.")

    for i in range(1, 8):

        # Seleciona ações:
        action = agent.compute_single_action(obs)
        print(f"Iteração: {i}")
        print(f"Ação: {action}")
        concepts_selecionados_list.append(action)

        # Retorna os estados e a recompensa:
        obs, reward, done, info = env.step(action)
        print(f"Estados: {obs}")
        print(f"Temperatura ambiente:
{np.unique(info.get('Tinf'))[0]}")
        print(f"Custo elétrico do kWh:
{info.get('custo_eletrico_kwh')}")

        # Recompensa total:
        episode_reward += reward
        print(f"Recompensa: {reward}.")
        print("")

        # Para visualização:
        SPTq_list.append(info.get("SPTq"))
        Tq_list.append(info.get("Tq"))
        SPh_list.append(info.get("SPh"))
        h_list.append(info.get("h"))
        Tt_list.append(info.get("Tt"))
        SPTs_list.append(info.get("SPTs"))
        Ts_list.append(info.get("Ts"))
        Sr_list.append(info.get("Sr"))
        Sa_list.append(info.get("Sa"))
        xq_list.append(info.get("xq"))
        xf_list.append(info.get("xf"))
        xs_list.append(info.get("xs"))

```

**Figura B.29** – Avaliação do agente de RL.

```

Fs_list.append(info.get("Fs"))
iqb_list.append(info.get("iqb"))
custo_eletrico_list.append(info.get("custo_eletrico"))
custo_gas_list.append(info.get("custo_gas"))
custo_agua_list.append(info.get("custo_agua"))
recompensa_list.append(info.get("recompensa"))
Fd_list.append(info.get("Fd"))
Td_list.append(info.get("Td"))
Tf_list.append(info.get("Tf"))
Tinf_list.append(info.get("Tinf"))
split_range_list.append(info.get("split_range"))

print(f"Recompensa total: {episode_reward}")
print("")

# Custos cumulativos:
custo_eletrico_list_acumulado =
list(accumulate(custo_eletrico_list))
custo_gas_list_acumulado = list(accumulate(custo_gas_list))
custo_agua_list_acumulado = list(accumulate(custo_agua_list))

# Custos totais:
custo_eletrico_total = custo_eletrico_list_acumulado[-1]
custo_gas_total = custo_gas_list_acumulado[-1]
custo_agua_total = custo_agua_list_acumulado[-1]
custo_total_banho = custo_eletrico_total + custo_gas_total +
custo_agua_total
print(f"Custo elétrico total: {custo_eletrico_total}")
print(f"Custo de gás total: {custo_gas_total}")
print(f"Custo de água total: {custo_agua_total}")
print(f"Custo total do banho: {custo_total_banho}")

# Tabelas com resultados principais:
IQB_total_sum = sum(iqb_list)
IQB_mean = IQB_total_sum / len(iqb_list)

resultados_list = [Tinf_num, custo_eletrico_kwh_num, iqb_list[0],
iqb_list[1], iqb_list[2], iqb_list[3], iqb_list[4], iqb_list[5],
iqb_list[6], IQB_mean, IQB_total_sum, episode_reward,
custo_eletrico_total, custo_gas_total, custo_agua_total,
custo_total_banho,]

```

**Figura B.30** – Avaliação do agente de RL.

```

concepts_list = [Tinf_num, custo_eletrico_kwh_num,
concepts_selecionados_list[0], concepts_selecionados_list[1],
concepts_selecionados_list[2], concepts_selecionados_list[3],
concepts_selecionados_list[4], concepts_selecionados_list[5],
concepts_selecionados_list[6]]

# Para visualização:
SPTq = np.concatenate(SPTq_list, axis=0)
Tq = np.concatenate(Tq_list, axis=0)
SPh = np.concatenate(SPh_list, axis=0)
h = np.concatenate(h_list, axis=0)
Tt = np.concatenate(Tt_list, axis=0)
SPTs = np.concatenate(SPTs_list, axis=0)
Ts = np.concatenate(Ts_list, axis=0)
Sr = np.concatenate(Sr_list, axis=0)
Sa = np.concatenate(Sa_list, axis=0)
xq = np.concatenate(xq_list, axis=0)
xf = np.concatenate(xf_list, axis=0)
xs = np.concatenate(xs_list, axis=0)
Fs = np.concatenate(Fs_list, axis=0)
Fd = np.concatenate(Fd_list, axis=0)
Td = np.concatenate(Td_list, axis=0)
Tf = np.concatenate(Tf_list, axis=0)
Tinf = np.concatenate(Tinf_list, axis=0)
split_range = np.concatenate(split_range_list, axis=0)

# Gráficos:
sns.set_style("darkgrid")
path_imagens = os.getcwd() + "/imagens/"

fig, ax = plt.subplots(1, 3, figsize=(15, 4))
ax[0].plot(tempo_total, Ts, label="Ts", color="tab:blue",
linestyle="solid")
ax[0].plot(tempo_total, Tt, label="Tt", color="tab:red",
linestyle="solid")
ax[0].plot(tempo_total, SPTs, label="SPTs - ação",
color="black", linestyle="dashed")
ax[0].set_title("Setpoint da temperatura de saída (SPTs) e\n
temperaturas de saída (Ts) e do tanque (Tt)")
ax[0].set_xlabel("Tempo em minutos")
ax[0].set_ylabel("Temperatura em °C")
ax[0].legend()

ax[1].plot(tempo_total, Fs, label="Fs", color="tab:red",
linestyle="solid")
ax[1].set_title("Vazão de saída (Fs)")

```

**Figura B.31** – Avaliação do agente de RL.

```

ax[1].set_xlabel("Tempo em minutos")
ax[1].set_ylabel("Vazão em litros/minutos")
ax[1].legend()

ax[2].plot(tempo_acoes, iqb_list, label="IQB", color="black",
linestyle="solid")
ax[2].set_title("Índice de qualidade do banho (IQB)")
ax[2].set_xlabel("Ação")
ax[2].set_ylabel("Índice")
ax[2].legend()
plt.savefig(path_imagens + "resultadol_" + nome_algoritmo +
"_Tinf" + Tinf_var + "_tarifa" + custo_eletrico_kwh_var +
"_selector_" + str(selector) + ".png", dpi=200)
plt.cla()
plt.close(fig)

fig, ax = plt.subplots(2, 2, figsize=(15, 11))
ax[0, 0].plot(tempo_total, Tq, label="Tq", color="tab:orange",
linestyle="solid")
ax[0, 0].plot(tempo_total, SPTq, label="SPTq - ação",
color="black", linestyle="dashed")
ax[0, 0].set_title("Setpoint da temperatura do boiler (SPTq)\n e
temperatura do boiler (Tq)")
ax[0, 0].set_ylabel("Temperatura °C")
ax[0, 0].legend()

ax[0, 1].plot(tempo_total, Sa, label="Sa", color="silver",
linestyle="solid")
ax[0, 1].plot(tempo_total, Sr, label="Sr", color="tab:red",
linestyle="solid")
ax[0, 1].plot(tempo_total, split_range, label="split-range -
ação", color="black", linestyle="solid")
ax[0, 1].set_title("Frações de aquecimento do boiler (Sa)\n e da
resistência elétrica (Sr)")
ax[0, 1].set_ylabel("Fração")
ax[0, 1].legend()

ax[1, 0].plot(tempo_total, xs, label="xs - ação", color="black",
linestyle="solid")
ax[1, 0].plot(tempo_total, xq, label="xq", color="tab:red",
linestyle="solid")
ax[1, 0].plot(tempo_total, xf, label="xf", color="tab:blue",
linestyle="solid")
ax[1, 0].set_title("Aberturas das válvulas de saída (xs),\n
quente (xq) e fria (xf)")
ax[1, 0].set_xlabel("Tempo em minutos")

```

**Figura B.32** – Avaliação do agente de RL.

```

ax[1, 0].set_ylabel("Abertura")
ax[1, 0].legend()

ax[1, 1].plot(tempo_total, SPh, label="SPh", color="black",
linestyle="dashed")
ax[1, 1].plot(tempo_total, h, label="h", color="tab:red",
linestyle="solid")
ax[1, 1].set_title("Setpoint do nível do tanque (SPh) e nível do
tanque (h)")
ax[1, 1].set_xlabel("Tempo em minutos")
ax[1, 1].set_ylabel("Nível")
ax[1, 1].legend()
plt.savefig(path_imagens + "resultado2_" + nome_algoritmo +
"_Tinf" + Tinf_var + "_tarifa" + custo_eletrico_kwh_var +
"_selector_" + str(selector) + ".png", dpi=200)
plt.cla()
plt.close(fig)

fig, ax = plt.subplots(1, 3, figsize=(20, 4))
ax[0].plot(tempo_acoes, recompensa_list, label="Recompensa",
color="black", linestyle="solid")
ax[0].set_title("Recompensa do agente")
ax[0].set_xlabel("Ação")
ax[0].set_ylabel("Índice")
ax[0].legend()

ax[1].plot(tempo_acoes, custo_eletrico_list, label="Custo
elétrico", color="tab:blue", linestyle="solid")
ax[1].plot(tempo_acoes, custo_gas_list, label="Custo do gás",
color="tab:red", linestyle="solid")
ax[1].plot(tempo_acoes, custo_agua_list, label="Custo da água",
color="tab:orange", linestyle="solid")
ax[1].set_title("Custos do banho em cada ação")
ax[1].set_xlabel("Ação")
ax[1].set_ylabel("Custos em reais")
ax[1].legend()

ax[2].plot(tempo_acoes, custo_eletrico_list_acumulado,
label="Custo elétrico", color="tab:blue", linestyle="solid")
ax[2].plot(tempo_acoes, custo_gas_list_acumulado, label="Custo
do gás", color="tab:red", linestyle="solid")
ax[2].plot(tempo_acoes, custo_agua_list_acumulado, label="Custo
da água", color="tab:orange", linestyle="solid")
ax[2].set_title("Custos cumulativos do banho")
ax[2].set_xlabel("Ação")
ax[2].set_ylabel("Custos em reais")

```



**Figura B.33** – Avaliação do agente de RL.

```

ax[2].legend()
plt.savefig(path_imagens + "resultado3_" + nome_algoritmo +
"_Tinf" + Tinf_var + "_tarifa" + custo_eletrico_kwh_var +
"_selector_" + str(selector) + ".png", dpi=200)
plt.cla()
plt.close(fig)

fig, ax = plt.subplots(1, 1, figsize=(5, 4))
ax.plot(tempo_acoas, iqb_list, label="IQB", color="black",
linestyle="solid")
ax.set_title("Índice de qualidade do banho (IQB)")
ax.set_xlabel("Ação")
ax.set_ylabel("Índice")
ax.legend()
plt.savefig(path_imagens + "resultado4_" + nome_algoritmo +
"_Tinf" + Tinf_var + "_tarifa" + custo_eletrico_kwh_var +
"_selector_" + str(selector) + ".png", dpi=200)
plt.cla()
plt.close(fig)

return resultados_list, concepts_list

```

**Figura B.34** – Função *main*.

```

if __name__ == "__main__":

    # Argumentos:
    parser = argparse.ArgumentParser()
    parser.add_argument("treina", help="Treina o agente",
choices=("True", "False"))
    parser.add_argument("avalia", help="Avalia o agente",
choices=("True", "False"))
    parser.add_argument("selector", help="Avalia o agente",
choices=("True", "False"))
    args = vars(parser.parse_args())

    # Inicializa o Ray:
    ray.shutdown()
    ray.init()

    # Define os steps para o PPO:
    n_iter_agente = 101
    n_iter_checkpoints = 10

```

**Figura B.35** – Função *main*.

```

# Define a temperatura ambiente e o custo da energia elétrica:
Tinf_list = [15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27,
28, 29, 30]
custo_eletrico_kwh_list = [1, 1.25, 1.5, 1.75, 2, 2.25]

# Treina o agente:
if args["treina"] == "True":
    # Treina cada concept:
    banho_dia_frio = treina_agente(nome_algoritmo,
n_iter_agente, n_iter_checkpoints, "banho_dia_frio")
    banho_dia_ameno = treina_agente(nome_algoritmo,
n_iter_agente, n_iter_checkpoints, "banho_dia_ameno")
    banho_dia_quente = treina_agente(nome_algoritmo,
n_iter_agente, n_iter_checkpoints, "banho_dia_quente")
    model = [banho_dia_frio, banho_dia_ameno, banho_dia_quente]

# Treina o selector:
selector = treina_agente(nome_algoritmo,
    n_iter_agente,
    n_iter_checkpoints,
    "seleciona_banho",
    True,
    model)

# Avalia o agente:
if args["avalia"] == "True":
    # Define se será utilizado o selective concept
learned(selector=True) ou programmed (selector=False):
    if args["selector"] == "True":
        selector = True
    else:
        selector = False

# Tabelas com resultados principais:
df_resultados = pd.DataFrame(columns=["Temperatura
ambiente", "Tarifa da energia elétrica", "IQB 1", "IQB 2", "IQB 3",
"IQB 4", "IQB 5", "IQB 6", "IQB 7", "IQB médio", "IQB total",
"Recompensa total", "Custo elétrico total", "Custo de gás total",
"Custo de água total", "Custo total do banho"])
df_concepts = pd.DataFrame(columns=["Temperatura ambiente",
"Tarifa da energia elétrica", "Concept ação 1", "Concept ação 2",
"Concept ação 3", "Concept ação 4", "Concept ação 5", "Concept ação
6", "Concept ação 7"])

```

**Figura B.36** – Função *main*.

```

# Cria combinações com todas as temperaturas e tarifa:
combs = list(itertools.product(map(str, Tinf_list), map(str,
custo_eletrico_kwh_list)))
for j, k in combs:
    Tinf_val = float(j)
    custo_eletrico_kwh_val = float(k)
    resultados_list, concepts_list =
avalia_agente(nome_algoritmo, [Tinf_val], [custo_eletrico_kwh_val],
selector)

    df_resultados.loc[len(df_resultados)] = resultados_list
    df_concepts.loc[len(df_concepts)] = concepts_list

# Salva os resultados principais em um arquivo csv:
if selector:
    df_resultados.to_csv("./resultados_tabela_learned
/resultados_tabela_Tinf" + str(Tinf_val) + ".csv", index=False)
    df_concepts.to_csv("./resultados_concepts_learned
/resultados_concepts_Tinf" + str(Tinf_val) + ".csv", index=False)
else:
    df_resultados.to_csv("./resultados_tabela_programmed/res
ultados_tabela.csv", index=False)

# Reseta o Ray:
ray.shutdown()

```