



# A B A K Ó S

Instituto de Ciências Exatas e Informática



Licença Creative Commons Attribution 4.0 International

## Caelus: ferramenta de inteligência artificial na predição de parâmetros da equação de chuvas intensas\*

Caelus: Artificial Intelligence Tool for Predicting Parameters of Extreme Rainfall Equation

Alex Tavares Silva<sup>1</sup>  
Wagner Rambaldi Telles<sup>2</sup>  
Jader Lugon Junior<sup>3</sup>

### Resumo

As precipitações intensas são fenômenos climáticos de grande impacto, especialmente em regiões com características geográficas e climáticas diversificadas, como o Brasil, onde uma boa predição dos parâmetros das equações de chuvas intensas é essencial para o planejamento urbano e a prevenção de desastres naturais. Nesse contexto, o Caelus surge como uma ferramenta computacional integrada que combina algoritmos de otimização e *Machine Learning* (ML) para estimar esses parâmetros. O estudo teve como objetivo o desenvolvimento e validação do sistema Caelus mediante as seguintes etapas metodológicas: (i) leitura e pré-processamento de dados climáticos e hidrológicos; (ii) avaliação de desempenho de algoritmos ML; e (iii) análise comparativa com modelos ML da literatura, demonstrando a eficácia do Caelus na predição de parâmetros com precisão compatível com resultados consolidados. A ferramenta forneceu estimativas confiáveis com MAE de 7,1 mm/h em ambas as fases e tempo de recorrência (TR) alinhado à literatura. Esses resultados destacam seu potencial para aplicação em todo o Brasil, bem como sua utilidade como modelo para estudos hidrológicos e meteorológicos em escala global. Conclui-se que o Caelus apresenta resultados promissores que podem apoiar no processo de decisão da gestão de riscos ambientais, contribuindo para o planejamento urbano eficiente e a melhoria da qualidade de vida em comunidades afetadas por eventos climáticos extremos, com perspectivas de aprimoramento contínuo e aplicações amplas.

**Palavras-chave:** Caelus. Inteligência artificial. *Machine Learning*. Luus-Jaakola. Parâmetros de chuvas intensas.

\*Submetido em 05/11/2023 - Aceito em 27/08/2025

<sup>1</sup>Instituto Federal Fluminense, Brasil - altasilva@gmail.com

<sup>2</sup>Universidade Federal Fluminense, Brasil - wtelles@id.uff.br

<sup>3</sup>Instituto Federal Fluminense, Brasil - jlugonjunior@gmail.com

### Abstract

Heavy rainfall events are high-impact climatic phenomena, especially in regions with diverse geographic and climatic characteristics, such as Brazil, where an accurate prediction of the parameters for intense rainfall equations is essential for urban planning and natural disaster prevention. In this context, Caelus emerges as an integrated computational tool that combines optimization algorithms and Machine Learning (ML) to estimate these parameters. This study aimed to develop and validate the Caelus system through the following methodological steps: (i) reading and preprocessing climatic and hydrological data; (ii) evaluating the performance of ML algorithms; and (iii) conducting a comparative analysis with ML models from the literature, demonstrating Caelus' effectiveness in predicting parameters with accuracy comparable to established results. The tool provided reliable estimates with a 7.1 mm/h MAE during training and validation, and a Recurrence Time (TR) consistent with literature. These results underscore its potential for nationwide application in Brazil, as well as its utility as a model for hydrological and meteorological studies on a global scale. In conclusion, Caelus delivers promising results that can support environmental risk management decision-making, contributing to efficient urban planning and improved quality of life in communities affected by extreme weather events—with prospects for continuous enhancement and broad applications.

**Keywords:** Caelus. Artificial Intelligence. Machine Learning. Luus-Jaakola. Extreme rainfalls parameters.

## 1 INTRODUÇÃO

Toda água proveniente do meio atmosférico que atinge a superfície terrestre, em hidrologia, é entendida como precipitação (Tucci, 2015). O mais importante componente do ciclo hidrológico é a precipitação pluvial e, de acordo com Oliveira (2019), forma o elo entre a água da atmosfera e a água do solo. O orvalho, a geada, a neve, a saraiva, o granizo, o chuvisco, a neblina, a garoa e a chuva são as formas mais comuns de precipitação.

Em particular, as precipitações em forma de chuva, ou seja, as precipitações em forma líquida, do ponto de vista da hidrologia, são as de maior interesse. Segundo Tucci (2015), a chuva, por sua capacidade de produzir escoamento, é o tipo mais importante para a hidrologia.

Especial atenção deve ser dada àquelas precipitações em forma de chuva mais intensas. As chuvas intensas, de acordo com Tucci (2015), são precipitações em grandes volumes geradas em pequenos intervalos de tempo, provocando riscos a vida humana e prejuízos ambientais (Farias; Coelho; Silva, 2013). Muitas vidas são perdidas por causa das chuvas intensas todos os anos, quando não, geram prejuízos materiais (Groisman; Knight; Karl, 2012; Pristo *et al.*, 2018).

As chuvas intensas podem ser representadas tanto por equações locais como por uma equação geral que relaciona parâmetros locais e adimensionais. Para se chegar a uma equação local de chuvas intensas, uma série de cálculos são necessários. Esses cálculos, por sua vez, necessitam de dados de chuvas que podem ser retirados de órgãos oficiais que efetuam registros diários, bem como de imagem de satélites através de algoritmos. No entanto, nem todas as informações necessárias para a construção das equações locais que representam as chuvas intensas estão disponíveis ou são de fácil obtenção (Senna *et al.*, 2010).

Alternativamente, métodos de otimização podem solucionar problemas quando se deseja estimar parâmetros considerados ótimos que maximizem ou minimizem uma dada função, denominada função objetivo. Para isso, são utilizados métodos determinísticos ou estocásticos. Métodos determinísticos geralmente convergem mais rápido para a solução, porém, podem ficar presos em ótimos locais, o que pode ser evitado utilizando fatores aleatórios (randômicos). Já os métodos estocásticos possuem características aleatórias, fornecendo uma garantia probabilística de encontrarem mínimo ou máximo da função (Telles, 2014).

Neste contexto, o objetivo deste artigo é avaliar a ferramenta Caelus baseada em Inteligência Artificial, especificamente a técnica de *Random Forest* em conjunto com o método de otimização Luus-Jaakola, para estimar os parâmetros da equação geral de chuvas intensas. Essa ferramenta se destaca por permitir a atualização ágil dos dados de precipitação, incorporando eventos de chuva recentes de forma mais eficiente do que outras metodologias encontradas na literatura. A abordagem proposta é avaliada com base na métrica de desempenho de Erro Médio Absoluto (do inglês *Mean Absolute Error* - MAE), garantindo precisão e confiabilidade nas estimativas. Além disso, a ferramenta possibilita a inclusão de dados atuais de precipitação, o que a torna particularmente relevante para aplicações em planejamento urbano e gestão de

riscos de desastres naturais.

## 2 FUNDAMENTAÇÃO TEÓRICA

Para a compreensão do leitor sobre o desenvolvimento da ferramenta Caelus, esta seção apresenta o embasamento teórico sobre as equações de chuvas intensas, bem como a manipulação e preparação preliminar dos dados para uma posterior estimativa dos parâmetros de interesse, utilizando para essa finalidade, métodos de otimização estocástica e inteligência computacional.

### 2.1 Equação geral de chuvas intensas

Através da equação geral de chuvas intensas é possível analisar intensidades de chuvas ocorridas em determinados tempos de recorrência e com determinada duração. Essa equação possui parâmetros que são sensíveis a cada região, o que torna mais difícil realizar projeções de forma automatizada. Na Equação 1 é apresentada a equação geral de chuvas intensas.

$$i = \frac{K \times TR^a}{(t + b)^c} \quad (1)$$

onde:  $i$  (mm/h) é a intensidade;  $TR$  (anos) é o tempo de retorno;  $t$  (minutos) é a duração da chuva e  $K$ ,  $a$ ,  $b$  e  $c$  são parâmetros adimensionais que devem ser determinados de acordo com a região.

### 2.2 Determinação de parâmetros por anamorfose

Segundo Oliveira (2019), em hidrologia, procura-se fazer inferências ou projeções futuras, tomando-se como base séries históricas, associando uma distribuição teórica a frequência de ocorrência de eventos. Essas séries históricas podem ser obtidas por meio de órgãos governamentais. Em seguida, utiliza-se a distribuição de Gumbel, a qual emprega parâmetros estatísticos de média e desvio padrão desses dados para a determinação da precipitação de um dia, correspondentes a diferentes tempos de recorrência (Pereira; Duarte; Sarmiento, 2017).

Para explicar o comportamento temporal do evento estudado, nesse caso as chuvas, deve-se verificar a aderência da distribuição teórica à frequência de ocorrência (Oliveira, 2019). Sendo assim, com o intuito de verificar se a distribuição de Gumbel se comporta de forma correta e coerente aos valores de precipitação máxima anual, os ajustes dos dados à distribuição estatística precisam ser avaliados no que se refere a qualidade. Utiliza-se então algum teste de aderência e, segundo Silva, Junior e Telles (2021), Kolmogorov-Smirnov é um teste bastante utilizado nesse tipo de estudo.

Para a obtenção de chuvas de 24 horas se faz necessário desagregá-la, multiplicando a precipitação diária por coeficientes encontrados na literatura. De acordo com Oliveira (2019), ao se utilizar operadores funcionais, transforma-se uma função não linear em uma função linear.

Esse procedimento é denominado anamorfose. Tanto modelos lineares como modelos lineares por anamorfose podem se ajustar aos dados observados das intensidades através do método da minimização da soma dos quadrados dos desvios. Um modelo geral proposto por Sherman (1931) para descrever a relação intensidade-duração-frequência (IDF) e reescrito por Wagesho e Claire (2016), tomando-se  $K \times TR^a$  igual a  $C$ , transforma a Equação (1) na Equação (2).

$$i = \frac{C}{(t + b)^c} \quad (2)$$

Aplicando logaritmo, transforma-se a Equação (2) na Equação (3), que é linear do tipo  $Y = AX + B$ , proposta por Farias, Coelho e Silva (2013).

$$\log(i) = \log(C) - c \times \log(t + b) \quad (3)$$

onde:

$$Y = \log(i);$$

$$A = \log(C);$$

$$B = C;$$

$$X = \log(t + b).$$

Pelo método dos mínimos quadrados tem-se, conforme Equação (4), que:

$$B = \frac{\sum XY - \frac{\sum X \sum Y}{n}}{\sum X^2 - \frac{(\sum X)^2}{n}} \quad (4)$$

onde:

$$A = \bar{Y} - B\bar{X};$$

$\bar{Y}$  é a média de  $Y$ ;

$\bar{X}$  é a média de  $X$ ;

$n$  é a quantidade de dados;

$\sum Y$  é o somatório de  $Y$ ;

$\sum X$  é o somatório de  $X$ .

### 2.3 Estimativa de parâmetros através do método de otimização estocástico Método Luus-Jaakola

O método Luus-Jaakola é um procedimento estocástico de busca aleatória, onde o domínio de busca inicial é o domínio de cada variável do problema e o intervalo de busca é reduzido (Luus; Jaakola, 1973).

Basicamente, o método gera soluções aleatórias considerando uma região de busca ampla, de acordo com possíveis valores das variáveis, enquanto a região de busca vai diminuindo cada vez mais ao longo das iterações (Telles, 2014).

Deve-se determinar os intervalos de busca para as variáveis que se deseja estimar, as quais dependem do problema a ser tratado, bem como a amplitude dos referidos intervalos  $r^{(k)} = (r_1^{(k)}, \dots, r_{num}^{(k)})$  onde  $num$  é o número de variáveis do problema.

Um fator de contração  $\epsilon$  determina a porcentagem de contração do intervalo e um *loop* externo,  $n_{out}$ , determina quantas vezes essa contração irá ocorrer. Por fim, um *loop* interno,  $n_{int}$ , determina o número de estimativas geradas a cada *loop*.

Após a configuração dos parâmetros iniciais, executa-se o algoritmo, que por sua vez gerará uma solução inicial,  $x^{(0)}$ , a qual passa a ser considerada, até o momento, a solução ótima  $x^*$ . Através da fórmula de recorrência,  $x^{(k)} = x^* + R^{(k)}r^{(k-1)}$ ,  $k = 1, \dots, n_{out}$ . Nas primeiras versões do Luus-Jaakola,  $R^{(k)}$  era uma matriz diagonal onde os elementos da diagonal variavam entre  $-0,5$  e  $0,5$ .

No Algoritmo 1 é apresentada a versão do método Luus-Jaakola, em linguagem descritiva.

---

**Algorithm 1: Luus-Jaakola**

---

- 1 Defina o número de loops externos:  $n_{out}$
- 2 Defina o número de loops internos:  $n_{int}$
- 3 Defina o fator de contração:  $\epsilon$
- 4 Defina o intervalo de busca inicial:  $r^{(0)}$
- 5 Gere uma solução inicial:  $x^{(0)}$
- 6 Faça:  $x^* = x^{(0)}$ 
  - 1: **for**  $k = 1$  até  $n_{out}$  **do**
  - 2:   **for**  $j = 1$  até  $n_{int}$  **do**
  - 3:      $x^{(k)} = x^* + R^{(k)}r^{(k-1)}$
  - 4:     **if**  $f(x^{(k)}) < f(x^*)$  **then**
  - 5:        $x^* = x^{(k)}$
  - 6:     **end if**
  - 7:   **end for**
  - 8:    $r^{(k)} = (1 - \epsilon)r^{(k-1)}$
  - 9: **end for**

---

**Nota:**  $R$  é uma matriz diagonal de números aleatórios entre  $-0,5$  e  $0,5$ .

---

No caso do problema proposto, têm-se quatro variáveis, a saber:  $K$ ,  $a$ ,  $b$  e  $c$ , que são os parâmetros adimensionais da equação de chuva, conforme Equação (1).

Primeiramente aplica-se o problema direto, obtendo as intensidades das chuvas por meio da equação geral. Em seguida, aplica-se o problema inverso, utilizando o método Luus-Jaakola que vai obtendo de forma aleatória os parâmetros, aplicando-os na equação geral (Equação (1)) e comparando-os com os resultados obtidos no problema direto.

Além da estimativa dos parâmetros por meio de métodos de otimização, uma abordagem igualmente interessante é o uso da inteligência artificial para a sua predição.

## 2.4 Estimativa de parâmetros através de inteligência computacional

A Inteligência Computacional, de acordo com Goldschmidt (2010), procura construir entidades artificiais inteligentes que vão além da perspectiva de compreensão do pensamento humano.

O computador desenvolve o reconhecimento de padrões ou a capacidade de aprender continuamente ou, ainda, fazer previsões com base em dados realizando ajustes sem serem especificamente programados para isso, processo pelo qual é conhecido como *Machine Learning* (ML).

ML se trata de algoritmos matemáticos, estatísticos e computacionais que são capazes de realizar um processo de inferência por meio de aprendizado baseado em exemplos, sendo considerado uma subárea da Inteligência Artificial e têm apresentado um crescimento enorme nas últimas décadas (Allende-cid, 2019).

Em situações reais, dados armazenados em forma digital são passíveis de falhas e não necessariamente estão organizados como o usuário precisa, sendo assim, para Netto e Maciel (2021), esses dados precisam passar por uma etapa de pré-processamento para eliminar dados faltantes, assim como outros.

No que se refere a ML, algumas etapas de um projeto de análise de dados são necessárias, a começar pela definição do problema, preparação dos dados, escolha dos algoritmos e visualização/interpretação das informações. Para Netto e Maciel (2021), embora não haja um único método para definição dessas etapas, são passos que, de certa maneira, se tornaram padrão para os projetos de análise de dados. Além disso, para Santos *et al.* (2019) podem ocorrer diferenças na performance dos resultados, a depender de algumas etapas como a de treinamento, e a de definição dos bancos de treinamento e teste.

Existem dois grandes grupos de algoritmos de ML quanto a intervenção de aprendizagem: os supervisionados e não supervisionados. Os algoritmos de aprendizagem supervisionada são aqueles em que os atributos dos dados se relacionam e se deseja prever quais atributos influenciam a variável independente (Netto; Maciel, 2021). Farias, Coelho e Silva (2013), por exemplo, experimentaram os classificadores não supervisionados e supervisionados para um estudo de uso e cobertura do solo.

Ainda para Netto e Maciel (2021), algoritmos de ML supervisionados podem se dividir em algoritmos de classificação ou algoritmos de regressão. Algoritmos de regressão resultam em valores que tentam prever alguma característica numérica. Regressão é utilizada quando um ou mais atributos são variáveis contínuas e dependentes.

Já os algoritmos de aprendizagem não supervisionada não possuem variável independente, tão pouco uma classificação prévia de atributos, tendo como objetivo descobrir alguma relação entre esses dados.

### 2.4.1 *Random Forest*

O *Random Forest* é um algoritmo de ML conhecido por sua eficácia e versatilidade em problemas de regressão e classificação (Liaw; Wiener, 2002).

No caso de regressão, um modelo de *Random Forest* é treinado para prever um valor numérico como saída. As árvores de decisão no *Random Forest* são ajustadas para estimar um valor contínuo como saída, e a saída final do *Random Forest* é obtida como uma média (ou mediana) das saídas de todas as árvores individuais (Breiman, 2002).

O *Random Forest* é uma técnica poderosa que pode ser aplicada em uma ampla gama de problemas de ML, incluindo regressão, dependendo do tipo de saída que se deseja prever. Um estudo realizado por Pereira, Beker e Schmoeller (2024), por exemplo, o qual se baseou em dados coletados entre 2008 e 2022 de 47 municípios do oeste do Paraná, determinou que dentre alguns modelos, o *Random Forest* foi o que teve o melhor desempenho.

Uma das principais vantagens do *Random Forest* é sua capacidade de lidar com dados de alta dimensionalidade e enfrentar o *overfitting*, um problema comum em muitos modelos de ML (Breiman, 2002).

## 3 MATERIAIS E MÉTODOS

Nesta seção é apresentada a ferramenta Caelus para a previsão de parâmetros da equação de chuvas intensas.

### 3.1 Caelus

O Caelus é uma ferramenta computacional em desenvolvimento, com a proposta de obter parâmetros da equação de chuvas intensas e apoiar a tomada de decisão, utilizando método de otimização e inteligência computacional, além de possuir uma Interface de Programação de Aplicações (do inglês *Application Programming Interface* - API) para verificar dados climáticos e realizar balanço hídrico climatológico desenvolvida por Silva (2024).

Para o desenvolvimento da ferramenta Caelus, foi utilizada a linguagem de programação Python 3.11 no sistema operacional GNU/Linux Ubuntu. Para interface utilizou-se as bibliotecas PyQt5, folium, ssl, webbrowser, pandas, rasterio, hydrobr, random, numpy, math, geopandas, requests, matplotlib e scikit-learn (Carvalho, 2020; Caria, 2021).

Imagens de satélite são obtidas por meio do projeto criado em 1999 pelo Serviço Geológico dos Estados Unidos (do inglês *United States Geological Survey* - USGS) e cientistas do Centro de Riscos Climáticos (do inglês *Climate Hazards Center* - CHC), suportado e financiado pela Agência dos Estados Unidos para o Desenvolvimento Internacional (do inglês *United States Agency for International Development* - USAID), pela Administração Nacional da Aeronáutica e Espaço (do inglês *National Aeronautics and Space Administration* - NASA) e pela Administração Oceânica e Atmosférica Nacional (do inglês *National Oceanic and Atmosphere*

*ric Administration* - NOAA), denominado Estimativas de Precipitação a Partir de Observações de Pluviômetros e Satélites (do inglês *Rainfall Estimates from Rain Gauge and Satellite Observations* - CHIRPS).

Após obtê-las, o Caelus as analisa e gera arquivos *Comma-Separated Values* (CSV), contendo informações de precipitação para cada região que possuir estações da Agência Nacional de Águas e Saneamento Básico (ANA) utilizando o pacote *opensource* HydroBr (Carvalho, 2020). Após a análise das imagens de satélite e a geração dos arquivos CSV, é possível estimar os parâmetros por meio do método de otimização Luus-Jaakola. Os parâmetros são obtidos automaticamente para cada região selecionada. Geralmente, deve-se selecionar a cidade ou o estado que se deseja obter os parâmetros. Com a obtenção dos parâmetros estimados pelo método de otimização Luus-Jaakola é possível aplicar um algoritmo de ML, capaz de realizar a previsão de parâmetros em regiões onde há escassez ou até mesmo a ausência de dados.

Caelus permite ser alimentado de forma mais automatizada e não depende de terceiros, embora possa haver esse apoio. Pode-se estimar parâmetros para uma cidade ou até mesmo um estado com apenas um clique, após se obter os dados de precipitação convertidos da imagem de satélite.

Para gerar o executável foi utilizada a ferramenta PyInstaller que tem como objetivo agrupar um aplicativo Python e todas as suas dependências em um único pacote, o que permite que o usuário execute a ferramenta empacotada sem instalar um interpretador Python ou qualquer módulo.

O PyInstaller permite gerar "executáveis" para Linux, MacOS X e Windows. No entanto, não é um compilador cruzado, o que significa que o executável destinado a um sistema operacional específico deve ser gerado a partir desse mesmo sistema. Até o presente momento, a ferramenta foi testada em sistemas Linux Ubuntu 22.04.2 LTS e Mint 21.1 por meio de uma máquina virtual com configuração de 4GB de memória RAM, 30GB de HD e processador de 4 núcleos.

Na Figura 1 é apresentada a tela inicial da ferramenta. Dois módulos foram implementados: *Rainfall* e *Weather*. O Módulo *Rainfall* destina-se a realizar a previsão de parâmetros da equação de chuvas intensas, descrita na Equação (1), enquanto o módulo *Weather* é voltado à realização da previsão do tempo e o balanço hídrico climatológico de determinada região.

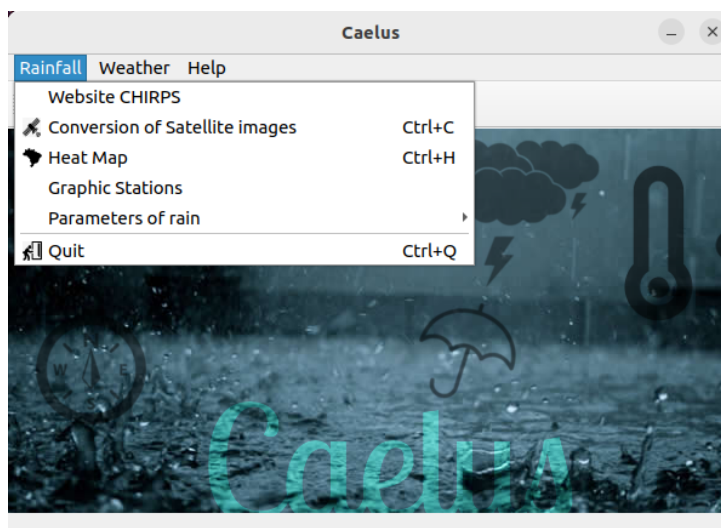
**Figura 1 – Janela principal da ferramenta Caelus**



Fonte: Elaborado pelos autores (2025).

O primeiro módulo (*Rainfall*), conforme Figura 2, possui uma opção (Website CHIRPS) para abrir a ferramenta web do projeto CHIRPS, onde serão obtidas as imagens de satélite. Já o segundo módulo utiliza uma API do OpenWeatherMap que permite predições do tempo atual ou com previsão de até 5 dias a cada 3 horas da localidade selecionada. Ainda no segundo módulo existem opções para previsão de tempo e para realização do balanço hídrico climatológico, conforme Tormnthwaite-Mather (1955).

**Figura 2 – Menu do módulo Rainfall**



Fonte: Elaborado pelos autores (2025).

O Caelus utiliza a equação de chuvas intensas, Equação (1), para o cálculo da intensidade de precipitação que ocorrerá em uma determinada região do território nacional, de acordo com o tempo de retorno e os quatro parâmetros presentes na referida equação. Estes quatro parâmetros são obtidos por meio do método de otimização LJ e, posteriormente, são previstos

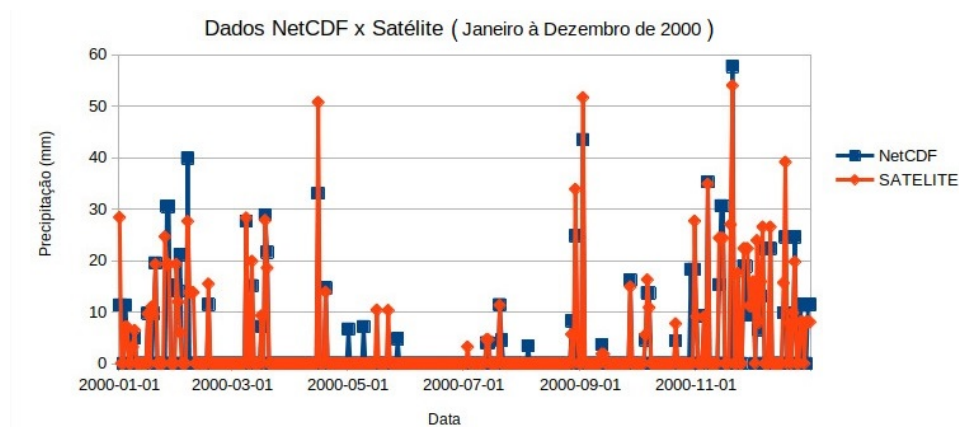
em regiões sem dados, utilizando-se, para essa finalidade, uma técnica de inteligência artificial.

Para a execução dos processos no Caelus, primeiramente, foram obtidas as imagens de satélite a serem convertidas para dados de precipitação, oriundas do projeto CHIRPS. Para isso tomou-se como referência o período de janeiro de 1981 a dezembro de 2020. Ressalta-se que Costa *et al.* (2019) realizam uma validação dos dados do projeto, tendo como referência geográfica, as estações de monitoramento da ANA.

Um ponto positivo na obtenção de imagens de satélite em vez da obtenção de anotações é que a conversão das imagens de satélite, pelo algoritmo, não deixam falhas de dados no arquivo convertido, sendo desnecessário se preocupar em tratar os dados antes de continuar o processo de obtenção dos parâmetros.

Na sequência, as imagens são convertidas em dados de precipitação, por meio de um algoritmo, com posterior validação dos dados. Para validação da conversão de imagem de satélite, comparou-se a conversão das imagens realizada pelo algoritmo com os dados registrados em formato NetCDF. Como exemplo, na Figura 3, é apresentada comparação dos dados para o município de Itaperuna para o período de janeiro a dezembro de 2000, com o intuito de mostrar que o algoritmo realiza as conversões de forma satisfatória.

**Figura 3 – Conversão das imagens de satélite em comparação aos dados obtidos por formato NetCDF**



Fonte: Elaborado pelos autores (2025).

Com as imagens convertidas em dados de precipitação, foi obtida a série histórica da precipitação máxima diária anual e aplicada a distribuição de Gumbel.

Em seguida, foi aplicado o método de otimização Luus-Jaakola para obtenção dos parâmetros  $K$ ,  $a$ ,  $b$ ,  $c$  da equação geral da chuva, Equação (1), considerando um total de 448 estações no Estado do Rio de Janeiro. Para essa etapa, o método Luus-Jaakola foi configurado com o número de iterações externas,  $n_{out}$ , igual a 100, o número de iterações internas,  $n_{int}$ , igual a 160, e o fator de contração do intervalo de busca,  $\epsilon$ , igual 0,1.

Essa etapa de estimativa dos parâmetros utilizando o método Luus-Jaakola é importante para que o algoritmo de inteligência computacional possa realizar a predição em locais onde não há estações.

Para essa finalidade, realizou-se o treinamento e validação das técnicas de ML (*Random Forest*) utilizando os parâmetros estimados pelo método Luus-Jaakola ( $K, a, b, c$ ) para duas situações (considerando apenas oito estações e com 447 estações), sendo que vinte por cento dos dados foram utilizados para validação.

Por fim, o processo termina com a previsão dos parâmetros de interesse ( $K, a, b, c$ ) utilizando a técnica de ML (*Random Forest*) para regiões onde não há informações para tais parâmetros.

Apesar da ferramenta Caelus realizar a estimativa e previsão dos parâmetros da equação de chuvas intensas para qualquer região do território nacional, neste trabalho, optou-se por apresentar os resultados para a cidade de Itaperuna, devido à sua importância para o Noroeste Fluminense.

## 4 RESULTADOS

Nesta seção são abordados os resultados obtidos com a estimativa dos parâmetros por meio de análise de imagem de satélites, utilizando o método Luus-Jaakola, bem como a previsão desses parâmetros previstos por inteligência artificial.

Além disso, os resultados gerados pelas simulações realizadas na região para diferentes tempos de recorrência e duração são discutidos.

### 4.1 Estimativa de parâmetros

Foram estimados os parâmetros  $K, a, b, c$  da equação geral da chuva, Equação (1), utilizando o método Luus-Jaakola. Na Tabela 1 são apresentados os resultados dos parâmetros para a cidade de Itaperuna, obtidos pela ferramenta Caelus, bem como aqueles obtidos por Oliveira (2019).

**Tabela 1 – Parâmetros da cidade de Itaperuna, obtidos pelo método Luus-Jaakola e por Oliveira (2019)**

|                        | $K$       | $a$    | $b$     | $c$    |
|------------------------|-----------|--------|---------|--------|
| <b>Média</b>           | 803,1889  | 0,2688 | 12,4888 | 0,7897 |
| <b>Desvio Padrão</b>   | 852,9547  | 0,1609 | 3,8936  | 0,0945 |
| <b>Menor Valor</b>     | 500,0000  | 0,1000 | 9,7758  | 0,7240 |
| <b>Maior Valor</b>     | 5237,1765 | 0,7478 | 26,7472 | 1,1426 |
| <b>Oliveira (2019)</b> | 808,4799  | 0,1531 | 9,7838  | 0,7242 |

Fonte: Elaborado pelos autores (2025).

Embora os parâmetros estimados pelo Caelus tenham se mostrado próximos aos obtidos por Oliveira (2019), as diferenças observadas podem ser atribuídas, em parte, às distintas abordagens metodológicas empregadas. Enquanto Oliveira (2019) utilizou métodos tradicionais de

calibração hidrológica, baseados em otimização determinística e ajuste manual de parâmetros, no Caelus foram adotadas técnicas de ML e algoritmo estocástico (Luus-Jaakola), que exploram automaticamente o espaço de parâmetros.

Essa diferença metodológica pode levar a variações nos resultados, uma vez que o Caelus prioriza a eficiência computacional e a busca por soluções globalmente ótimas, enquanto métodos convencionais podem depender mais da interpretação subjetiva do modelador.

Além disso, fatores como a seleção de dados de entrada, o período de calibração e as restrições impostas aos parâmetros em cada estudo também contribuem para as discrepâncias. Apesar disso, a proximidade entre os valores reforça a robustez do Caelus, sugerindo que, mesmo com abordagens distintas, a ferramenta é capaz de produzir estimativas consistentes com a literatura existente.

#### 4.2 Predição de parâmetros da equação geral de chuvas intensas

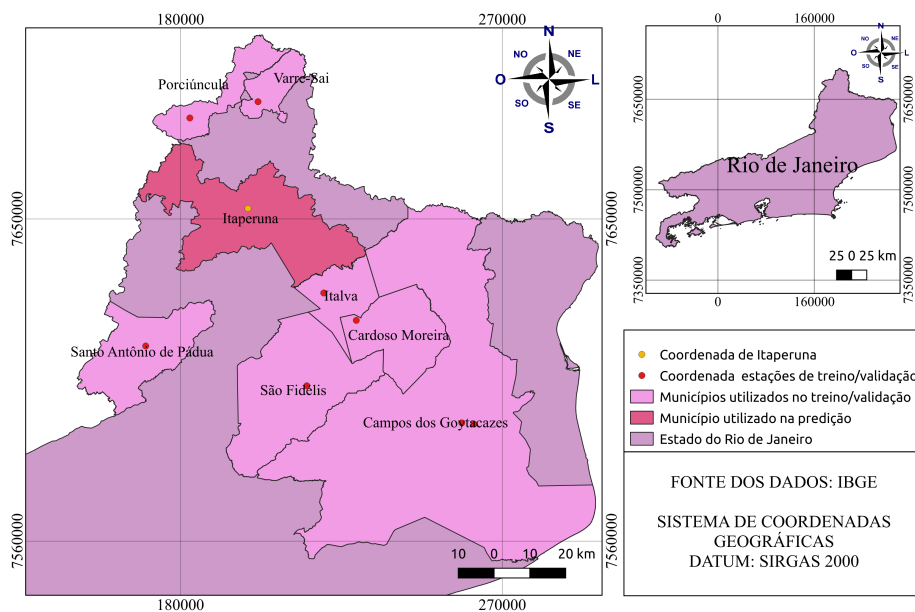
Na Tabela 2 estão contidas as informações de oito estações próximas do município de Itaperuna, utilizadas para treino e validação dos parâmetros, enquanto na Figura 4 é apresentado o mapa da região de estudo com a localização tanto dos municípios utilizados no treino e validação quanto o município com intenção de predição dos parâmetros. Ainda na referida tabela (Tabela 2), são apresentados os resultados do total de 100 corridas, obtidos pelo Luus-Jaakola, contendo a média encontrada na predição.

**Tabela 2 – Estações do estado do Rio de Janeiro utilizadas no treinamento e validação dos parâmetros**

| Local                  | Latitude | Longitude | $K$      | $a$    | $b$     | $c$    |
|------------------------|----------|-----------|----------|--------|---------|--------|
| Italva                 | -21,4167 | -41,7000  | 691,8813 | 0,2624 | 12,9625 | 0,8010 |
| Cardoso Moreira        | -21,4872 | -41,6136  | 809,2010 | 0,2561 | 13,3735 | 0,8113 |
| Campos                 | -21,7500 | -41,3333  | 950,9156 | 0,2538 | 12,2973 | 0,7852 |
| Campos Ponte Municipal | -21,7533 | -41,3003  | 862,3152 | 0,2709 | 12,2058 | 0,7830 |
| São Fidélis            | -21,6500 | -41,7500  | 900,0860 | 0,2415 | 13,5720 | 0,8161 |
| Santo Antônio de Pádua | -21,5419 | -42,1825  | 781,8607 | 0,2591 | 12,4480 | 0,7889 |
| Varre Sai              | -20,9314 | -41,8678  | 802,8832 | 0,2293 | 12,3206 | 0,7856 |
| Porciúncula            | -20,9692 | -42,0517  | 790,2010 | 0,2870 | 12,5759 | 0,7922 |

Fonte: Elaborado pelos autores (2025).

**Figura 4 – Localização dos municípios da região de estudo**



**Fonte: Elaborado pelos autores (2025).**

Na Tabela 3 foram utilizadas oito estações no treino e validação para obter os parâmetros listados. Por outro lado, na Tabela 4 foram utilizadas 447 estações, com o objetivo de comparar a diferença na previsão de parâmetros, por meio de aprendizado de máquina, utilizando tanto um número reduzido quando uma quantidade maior de estações. Pela referida tabela, percebe-se uma melhora na previsão dos parâmetros, o que deixa claro que com mais dados para treino, o algoritmo de ML têm um melhor desempenho.

**Tabela 3 – Resultados obtidos na previsão dos parâmetros considerando oito estações para treino e validação**

|                      | <i>K</i> | <i>a</i> | <i>b</i> | <i>c</i> |
|----------------------|----------|----------|----------|----------|
| <b>Média</b>         | 758,4676 | 0,2630   | 12,9229  | 0,8004   |
| <b>Desvio Padrão</b> | 12,0362  | 0,0038   | 0,0861   | 0,0021   |
| <b>Menor Valor</b>   | 718,0999 | 0,2536   | 12,7111  | 0,7958   |
| <b>Maior Valor</b>   | 787,4411 | 0,2739   | 13,2486  | 0,8055   |

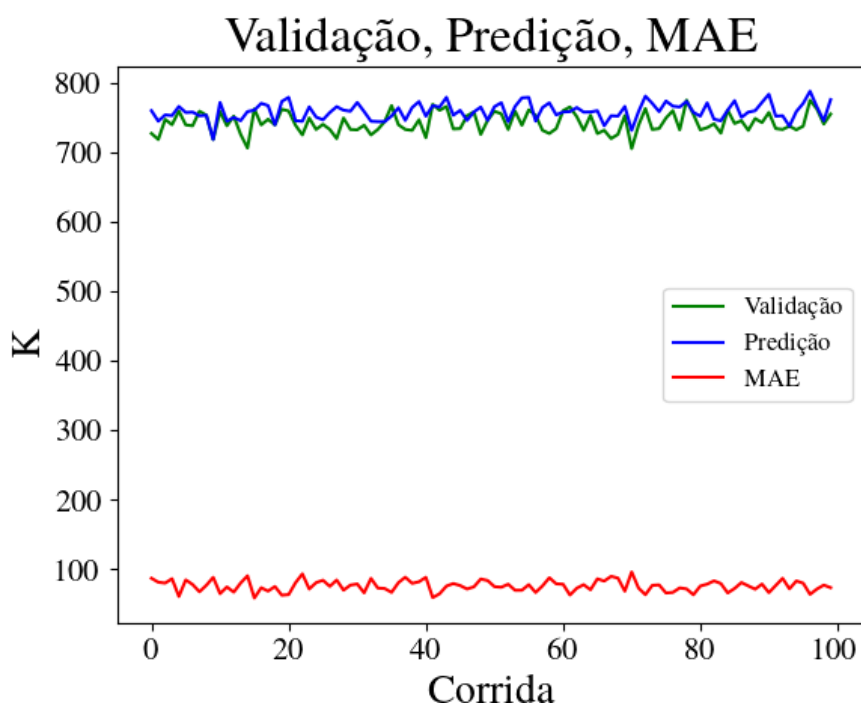
**Fonte: Elaborado pelos autores (2025).**

**Tabela 4 – Resultados obtidos na predição dos parâmetros considerando 447 estações para treino e validação**

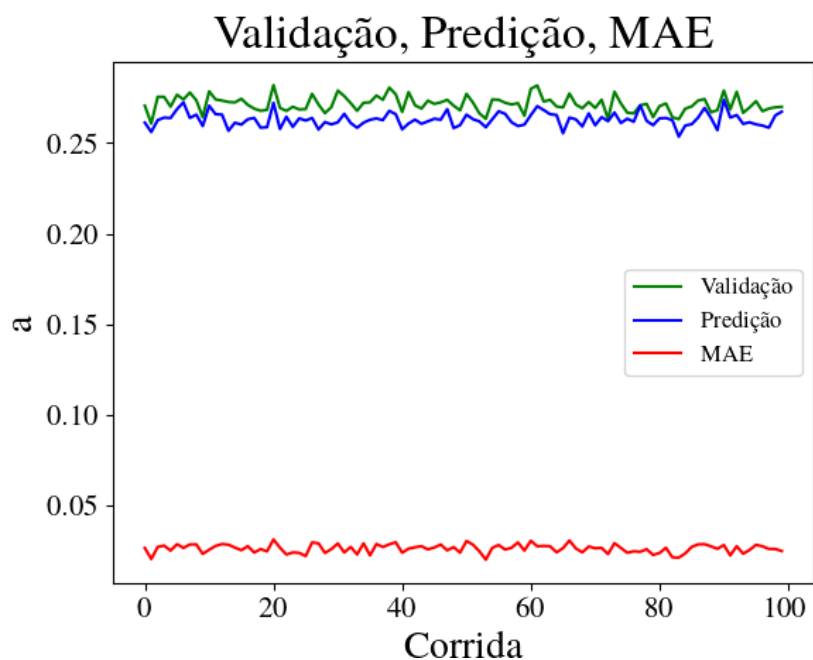
|                      | $K$      | $a$    | $b$     | $c$    |
|----------------------|----------|--------|---------|--------|
| <b>Média</b>         | 801,9276 | 0,2636 | 12,7215 | 0,7953 |
| <b>Desvio Padrão</b> | 12,1240  | 0,0079 | 0,1080  | 0,0024 |
| <b>Menor Valor</b>   | 777,6926 | 0,2452 | 12,5068 | 0,7898 |
| <b>Maior Valor</b>   | 835,2886 | 0,2805 | 12,9841 | 0,8015 |

Fonte: Elaborado pelos autores (2025).

Na Figura 5 é apresentado o gráfico dos resultados, ao longo das 100 corridas para 447 estações, do MAE, da validação e de predição do parâmetro  $K$ . O mesmo vale para as Figuras 6-8 referente a obtenção dos parâmetros  $a$ ,  $b$  e  $c$ , respectivamente. Devido a semelhança dos gráficos, neste artigo não serão exibidos os mesmos gráficos gerados com a quantidade menor estações.

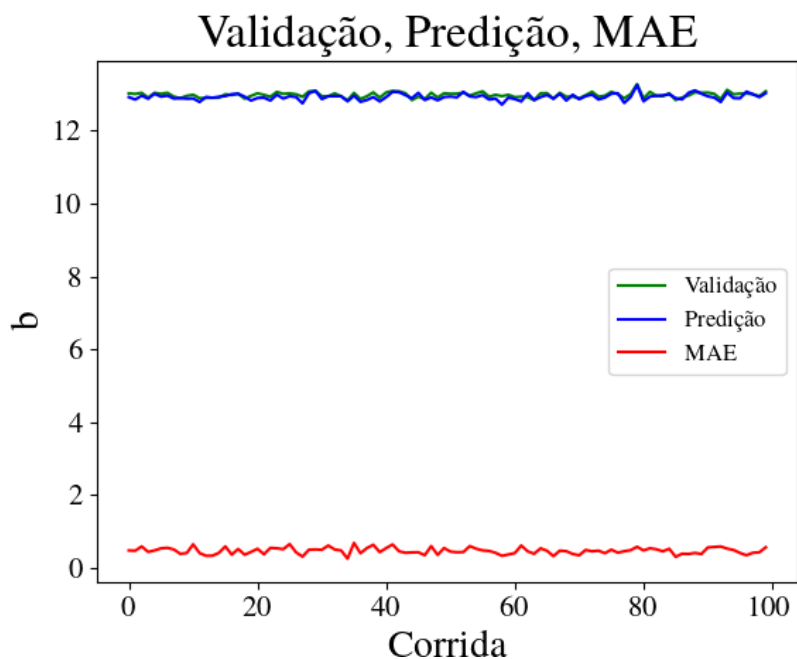
**Figura 5 – Resultado das 100 corridas na predição do parâmetro  $K$** 

**Figura 6 – Resultado das 100 corridas na previsão do parâmetro  $a$**



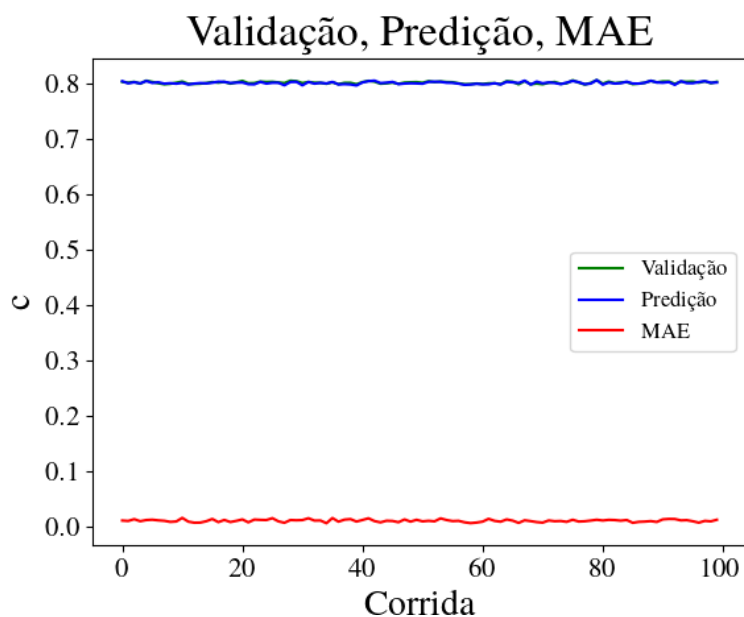
Fonte: Elaborado pelos autores (2025).

**Figura 7 – Resultado das 100 corridas na previsão do parâmetro  $b$**



Fonte: Elaborado pelos autores (2025).

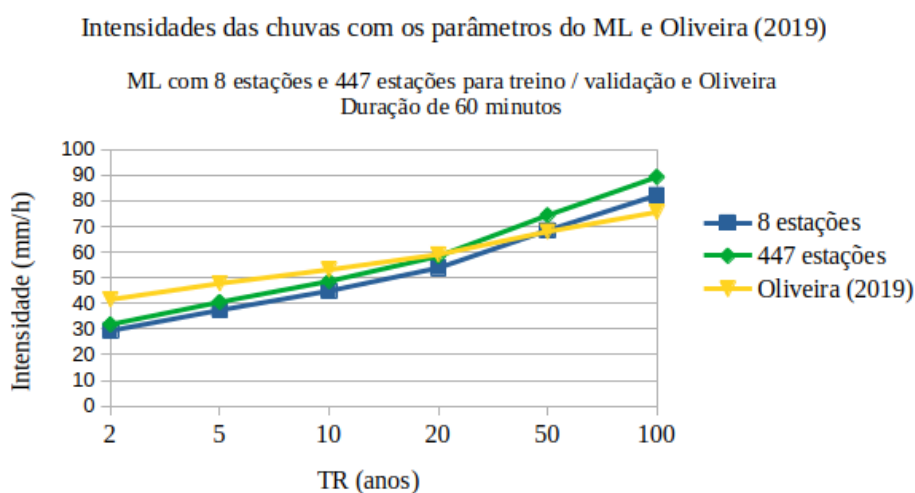
**Figura 8 – Resultado das 100 corridas na previsão do parâmetro c**



Fonte: Elaborado pelos autores (2025).

Já na Figura 9 é apresentado o perfil das chuvas com os parâmetros obtidos pelo algoritmo ML, sendo que, em um perfil, os parâmetros foram obtidos utilizando uma quantidade menor de estações para treinar e validar os resultados enquanto o outro perfil utilizou parâmetros obtidos de uma quantidade maior de estações adotadas para treino e validação. Uma comparação é feita com o perfil utilizando dados atualizados de Oliveira (2019).

**Figura 9 – Comparação da chuva de projeto do ML com oito estações, ML com 447 estações usadas para treino / validação e com os parâmetros da literatura**

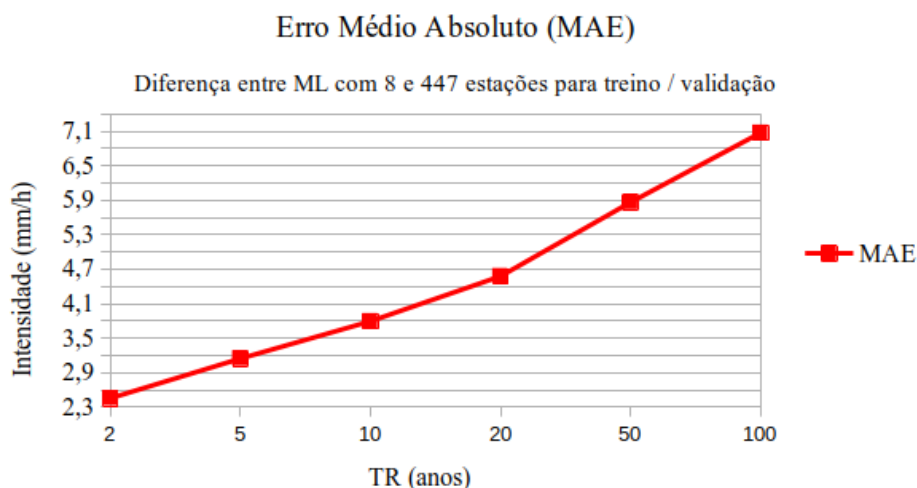


Fonte: Elaborado pelos autores (2025).

Na sequência, Figura 10, apresenta-se o MAE entre ML com menor quantidade de estações e ML com uma quantidade maior de estações utilizadas para treino / validação. Esse

erro é calculado ao utilizar ML, sendo assim, não foi realizado para os dados de Oliveira (2019), uma vez que não há treinos e validações para esses parâmetros.

**Figura 10 – Erro Médio Absoluto entre ML com oito estações e ML com 447 estações utilizadas pra treino/validação**



Fonte: Elaborado pelos autores (2025).

## 5 DISCUSSÕES

O Caelus demonstrou resultados promissores na estimativa de parâmetros coerentes com os valores encontrados na literatura atualizada, destacando-se pela aplicação de técnicas de otimização e inteligência artificial. No entanto, algumas limitações foram observadas, principalmente relacionadas a problemas nas leituras das estações em determinadas regiões, o que pode interromper o processo. Apesar disso, uma vantagem significativa do Caelus é que as regiões onde a leitura e a conversão para arquivos de precipitação foram bem-sucedidas não são perdidas, mesmo quando ocorrem erros em estações específicas. Esses erros, que surgem durante o processo de otimização pelo método Luus-Jaakola, não comprometem totalmente a funcionalidade da ferramenta.

Os resultados obtidos pelo algoritmo de ML do Caelus, tanto com oito quanto com 447 estações, apresentaram pequenas variações em relação aos valores de referência de Oliveira (2019). Essas diferenças podem ser atribuídas a diversos fatores, como a qualidade dos dados de entrada, a representatividade espacial das estações utilizadas e a própria natureza estocástica dos métodos de otimização empregados. No caso do treinamento com poucas estações, embora os resultados tenham sido próximos aos da literatura, a limitação na quantidade de dados pode ter impactado a generalização do modelo, especialmente em regiões com características hidrológicas distintas. Já o modelo treinado com muitas estações, apesar de mais abrangente, ainda apresentou discrepâncias pontuais, indicando que a simples quantidade de dados não é suficiente para garantir perfeita precisão sem um pré-processamento adequado.

Um exemplo prático da eficácia da ferramenta Caelus pode ser observado nos resultados obtidos para o município de Itaperuna, que se mostraram compatíveis com a literatura atualizada. Os parâmetros estimados pelo Caelus para essa região apresentaram alta confiabilidade, reforçando a precisão da ferramenta quando alimentada com dados adequados e utilizada corretamente. Esse desempenho positivo em Itaperuna destaca o potencial do Caelus para aplicações em outras localidades com características semelhantes.

Quanto ao algoritmo de Inteligência Artificial implementado, ele mostrou-se robusto, proporcionando estimativas de parâmetros bastante próximas às da literatura, desde que alimentado com dados corretos e adequadamente preparados. Um dos principais desafios no uso do ML para estimativa de parâmetros hidrológicos é a sensibilidade do algoritmo à qualidade e à consistência dos dados. O método de Luus-Jaakola, embora eficiente, depende de configurações iniciais (como limites de busca e número de iterações) que, se mal definidas, podem levar a soluções subótimas. Outro fator crítico é a possível superestimação ou subestimação de parâmetros em regiões com poucos dados históricos, o que pode comprometer a confiabilidade das previsões em eventos extremos, como chuvas intensas ou secas prolongadas.

A diferença de resultados ao treinar a rede com poucas e com muitas estações foi muito pequena, indicando que o algoritmo de ML não necessita de grandes volumes de dados para obter resultados satisfatórios, embora seja recomendável treiná-lo com o máximo de informações disponíveis para maior confiabilidade. Essa precisão reforça o potencial do Caelus como uma ferramenta valiosa para o apoio à tomada de decisão no gerenciamento de desastres naturais. No entanto, as diferenças observadas entre os valores de Oliveira (2019) e os obtidos pelo ML destacam a importância de validar os resultados com dados independentes e de incorporar técnicas de pós-processamento para reduzir incertezas. A dependência de dados inconsistentes ou não representativos pode levar a erros na calibração de modelos hidrológicos, afetando, por exemplo, a precisão de alertas de inundações ou a gestão de recursos hídricos.

A ferramenta ainda está em desenvolvimento e precisa aprimorar a forma como trata e comunica os erros ao usuário. Atualmente, as mensagens de erro não são suficientemente claras ou detalhadas, o que pode dificultar a identificação e a correção de problemas, especialmente para usuários menos familiarizados com a ferramenta. Para mitigar esses problemas, futuras versões do Caelus devem incluir mecanismos de detecção e correção de anomalias nos dados, além de métodos de *ensemble* que combinem múltiplas execuções do algoritmo para aumentar a robustez das estimativas. Esses avanços serão essenciais para transformar o Caelus em uma ferramenta ainda mais confiável para o apoio à tomada de decisão em cenários reais.

Outro desafio é a necessidade de conhecimento prévio para operar o Caelus de forma eficiente. Embora a ferramenta funcione bem quando utilizada seguindo as etapas corretas, usuários sem experiência podem enfrentar dificuldades para interpretar os resultados ou entender onde estão cometendo erros. A falta de mensagens mais atrativas e explicativas agrava essa situação, limitando a acessibilidade da ferramenta para um público mais amplo. Portanto, é essencial investir em melhorias na interface e na comunicação de erros, tornando o Caelus

mais intuitivo e amigável.

É importante ressaltar, também, que a utilização do Caelus não substitui a *expertise* dos profissionais responsáveis pela gestão de desastres. A ferramenta deve ser vista como um recurso complementar, que integra a análise de dados e a interpretação de resultados ao conhecimento e à experiência dos tomadores de decisão. Uma abordagem integrada, que combine tecnologia e *expertise* humana, é essencial para garantir decisões mais assertivas e eficazes no enfrentamento de eventos climáticos extremos.

## 6 CONCLUSÕES

A expansão do Caelus para novas regiões, permitindo avaliar seu desempenho em diferentes condições climáticas e geográficas, contribuirá na validação da sua confiabilidade. Simultaneamente, planejar melhorias na interface do sistema, com ênfase na clareza das mensagens de erro e na facilidade de uso, tornando a ferramenta mais acessível, inclusive para usuários sem experiência técnica é fundamental. Avançar no desenvolvimento da ferramenta permitirá consolidar o Caelus como uma solução completa, combinando rigor científico com aplicabilidade prática no gerenciamento de desastres naturais e no apoio à tomada de decisões ambientais.

## AGRADECIMENTOS

Os autores agradecem o apoio financeiro proporcionado pela CAPES, Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (Código Financeiro 001), pelo CNPq, Conselho Nacional de Desenvolvimento Científico e Tecnológico, e pela FAPERJ, Fundação Carlos Chagas Filho de Amparo à Pesquisa do Estado do Rio de Janeiro (Processo E-26/210.220/2021).

## REFERÊNCIAS

- ALLENDE-CID, Héctor. Machine learning: catalisador da ciência. **Sociedade Brasileira de Computação**, Porto Alegre, v. 1, n. 1, p. 15–18, 2019.
- BREIMAN, L. Random forests. machine learning. **R. News**, v. 2, p. 18–22, 2002.
- CARIA, G. **Rasterio**. 2021. Disponível em: <<https://github.com/rasterio/rasterio>>. Acessado em 27/07/2025.
- CARVALHO, Wallisson Moreira De. **HydroBr: A Python package to work with Brazilian hydrometeorological time series**. 2020. Disponível em: <<https://github.com/wallissoncarvalho/hydrobr>>. Acessado em 27/07/2025.
- COSTA, Julio *et al.* Validação dos dados de precipitação estimados pelo chirps para o brasil. **Revista Brasileira de Climatologia**, v. 24, 2019.
- FARIAS, José Alexandre Moreira Silva; COELHO, José Francisco Rodrigo; SILVA, Luiz da. Determinação de equação idf, utilizando regressão linear em base logarítmica. **XX Simpósio Brasileiro de Recursos Hídricos**, Bento Gonçalves, 2013.
- GOLDSCHMIDT, Ronaldo Ribeiro. **Uma introdução a inteligência computacional: fundamentos, ferramentas e aplicações**. Rio de Janeiro: IST-Rio, 2010.
- GROISMAN, Pavel Ya; KNIGHT, Richard W; KARL, Thomas R. Changes in intense precipitation over the central united states. **Journal of Hydrometeorology**, v. 13, n. 1, p. 47–66, 2012.
- LIAW, Andy; WIENER, Matthew. Classification and regression by random forest. **R. News**, v. 2, p. 18–22, 2002.
- LUUS, Rein; JAAKOLA, THI. Optimization by direct search and systematic reduction of the size of search region. **ACIhE Journal**, v. 19, n. 4, 1973.
- NETTO, Amílcar; MACIEL, Francisco. **Python para data science e machine learning descomplicado**. Rio de Janeiro: Alta Books, 2021.
- OLIVEIRA, Luiz Fernando Coutinho de. **Chuvas extremas no Brasil: modelos e aplicações**. Lavras: Editora UFLA, 2019.
- PEREIRA, Daniela Carneiro; DUARTE, Letícia Rodrigues; SARMENTO, Antover Panazzolo. Determinação da curva de intensidade, duração e frequência do município de ipameri – goiás. **REEC - Revista Eletrônica de Engenharia Civil**, Goiânia, v. 13, n. 2, 2017.
- PEREIRA, Eduardo Mendes; BEKER, João Pedro dos Santos; SCHMOELLER, Ruminiki Pavei. Aplicação de machine learning na previsão da produtividade da soja. **Revista Pleiade**, v. 18, n. 45, 2024.
- PRISTO, Marcos Vinícius de Jesus *et al.* Climatologia de chuvas intensas no município do rio de janeiro. **Revista Brasileira de Meteorologia**, v. 4, p. 33, 2018.
- SANTOS, Hellen Geremias dos *et al.* Machine learning para análises preditivas em saúde: exemplo de aplicação para predizer óbito em idosos de são paulo, brasil. **Caderno de Saúde Pública**, v. 35, n. 7, 2019.

SENNÁ, Rafael Salgado de *et al.* Metodologias para estimativa dos parâmetros da equação de chuvas intensas no estado do espírito santo. **Revista Engenharia na Agricultura**, v. 18, n. 6, p. 496–503, 2010.

SHERMAN, Charles W. Frequency and intensity of excessive rainfall at boston, massachusetts. Transaction Paper ASCE, v. 95, p. 951–960, Jan 1931.

SILVA, Alex Tavares. **Caelus: Ferramenta de Inteligência Artificial de Predição de Parâmetros da Equação de Chuvas Intensas e Apoio a Tomada de Decisão**. 2024. Tese (Doutorado) — Instituto Federal Fluminense, Campos dos Goytacazes, RJ., 2024.

SILVA, Alex Tavares; JUNIOR, Jader Lugon; TELLES, Wagner Rambaldi. Equação de chuvas intensas e vazão da cidade de itaperuna-rj. **Anais do 6º Simpósio Sobre Sistemas Sustentáveis**, v. 3, p. 104–116, 2021.

TELLES, Wagner Rambaldi. **Previsão do comportamento hidráulico de um rio com base na estimativa de coeficientes que controlam seu escoamento. Estudo de caso: Rio Bengalas, Nova Friburgo-RJ**. 2014. Tese (Doutorado) — Universidade do Estado do Rio de Janeiro, Friburgo, RJ., 2014.

TUCCI, Carlos Eduardo Morelli. **Hidrologia: ciência e aplicação**. Rio Grande do Sul: UFRGS, 2015.

WAGESHO, Negash; CLAIRE, Marie. Analysis of rainfall intensity-duration-frequency relationship for rwanda. **Journal fo Water Resource and Protection**, v. 8, n. 7, p. 706–723, 2016.