

Preenchimento de Falhas de Precipitação do Grupo de Bacias GI7 em Pernambuco

Raianna Galvão Dantas Silva*, Joyce Catarina Rodrigues Brainer**, Graziela Gislaine Lima Dos Santos ***
Juliana Patrícia Fernandes Guedes Barros****, Josiclêda Domiciano Galvêncio*****

* Graduanda de Geografia Bacharelado do Departamento de Ciências Geográficas da Universidade Federal de Pernambuco (UFPE), Recife/PE - Brasil; (raianna.galvão@ufpe.br)

** Graduandas de Bacharelado Geografia do Departamento de Ciências Geográficas da Universidade Federal de Pernambuco (UFPE), Recife/PE - Brasil; (graziela.gislaine@ufpe.br; joyce.catarina@ufpe.br)

*** Mestra e doutoranda, Programa de Pós-Graduação em Desenvolvimento e Meio Ambiente, Universidade Federal de Pernambuco (UFPE), Recife/PE - Brasil; (juliana.guedesbarros@ufpe.br)

**** Professora Associada do Departamento de Ciências Geográficas da Universidade Federal de Pernambuco (UFPE), Programa de Pós-Graduação em Desenvolvimento e Meio Ambiente-PRODEMA. Recife/PE - Brasil; (josicleda.galvencio@ufpe.br)

Resumo

A utilização do monitoramento pluviométrico se torna essencial para analisar o ciclo hidrológico em uma bacia hidrográfica. Este estudo analisou o manejo hidrológico do grupo de bacias hidrográficas de rios interiores - GI7 inserido na bacia hidrográfica do rio Garças no interior do estado de Pernambuco, Brasil. Foram utilizadas séries históricas pluviométricas com escala diária com 61 anos de análise de dados (1961-2021). O banco de dados disponível apresenta falhas e inconsistências, comprometendo os estudos hidroclimáticos. Nesse contexto, se fez necessário o desenvolvimento do software SUPeR-FILL para a calibração dos dados, que tem como função preencher séries históricas e temporais de dados pluviométricos. Nesse software foram combinados métodos e modelos como, Cubist e Qmap. O estudo demonstra que essas metodologias de aprendizagem de máquinas têm grande importância na era de alto poder de processamento com inteligência artificial. Conclui-se que as análises obtidas ofertam o poder de entendimento das condições climáticas das regiões em que a GI7 está inserida, propondo a partir das análises, mais segurança para as tomadas de decisões nos planejamentos hídrico, social e econômico dessas regiões.

Palavras-Chave: Monitoramento pluviométrico, Séries históricas, Calibração de dados, Bacia Hidrográfica.

Filling of Precipitation Faults of the GI7 Basin Group in Pernambuco

Abstract

The use of rainfall monitoring becomes essential for analyzing the hydrological cycle within a watershed. This study analyzed the hydrological management of the group of interior river basins - GI7, located within the Garças River Basin, in the interior of the state of Pernambuco. Daily rainfall historical series with 61 years of data analysis (1961–2021) were used. The available database presents gaps and inconsistencies, compromising hydroclimatic studies. In this context, it became necessary to develop the SUPeR-FILL software for data calibration, whose function is to fill historical and temporal rainfall data series. In this software, methods and models such as Cubist and Qmap were combined. The study demonstrates that these machine learning methodologies are of great importance in the era of high-performance computing and artificial intelligence. It is concluded that the analyses obtained provide a better understanding of the climatic conditions of the regions where GI7 is located, offering greater support and security for decision-making in the hydrological, social, and economic planning of these regions.

Keywords: Rainfall monitoring, Historical series, Data calibration, River basin.

1. Introdução

A precipitação é uma das variáveis meteorológicas mais importantes para os estudos climáticos. Determinante na caracterização e potencial

hídrico de uma região ou uma bacia hidrográfica (Nascimento *et al.*, 2013, p. 5).

As estações pluviométricas são de grande importância para análises flúvio climáticas, entretanto se

trata de equipamentos que exigem manutenção recorrente para que ocorra boa qualidade dos dados. Os pluviométricos frequentemente operam com falhas operacionais que resultam em séries temporais fragmentadas. Isso exige que pesquisadores apliquem métodos de preenchimento de falhas para um melhor manejo dos dados (Fagandini et al., 2024).

O sistema hidroclimático examina como as variáveis como: precipitação, evapotranspiração e vazão se integram como determinantes, para a disponibilidade hídrica de uma região. A intensidade da chuva e temperatura também são fatores essenciais para entender a dinâmica de água da área de estudo. Compreendendo tais características do e no espaço, se faz possível a criação de planos para o uso da água, reservatórios, estabelecendo medidas para mitigar secas ou cheias (Tercini et al. 2024).

Em séries temporais extensas as falhas de dados pluviométricos interferem diretamente na precisão das análises hidroclimáticas, afetando, conseqüentemente, a eficácia da gestão de recursos hídricos (Dias et al. 2021). Partindo dessa compreensão, séries longas fragmentadas são uma grande problemática para um exame coerente do planejamento alternativo, frente às mudanças climáticas. A ausência de dados coesos impossibilita uma dimensão pluviométrica clara dos acontecimentos sobre o espaço, dispondo que as estimativas sejam prováveis - assertivas ou improváveis - errôneas, causando transtornos e falhas na gestão de segurança hidrossocial.

Os resultados das variabilidades espaciais e temporais exercidas pela chuva, a ordenação e posição dos pluviômetros têm grande influência ao preencher dados ausentes na estação e na interpolação espacial. A garantia de dados de precipitação mais consistentes é essencial para um melhor planejamento da área (Depiné et al., 2014). O estudo do ritmo pluviométrico é uma das análises mais importantes para entender o sistema hidroclimático (Berezuk; Garcia, 2012).

A imprecisão de estimativas de precipitação alertou para a necessidade de desenvolver um software que tivesse a capacidade de preencher de maneira eficaz dados faltosos. Ruezzen, Miranda, Tech e Mauad (2021) aplicaram métodos tradicionais e de inteligência artificial para o preenchimento de falhas em dados de precipitação. O avanço da inteligência artificial fez com que criássemos a hipótese que era possível avançar na estimativa de precipitação e no preenchimento de falhas. Segundo Brubacher, Oliveira e Guasselli (2020) a

compreensão espacial da variabilidade pluviométrica, é essencial para a ação de planejamento dos recursos hídricos, a distribuição da chuva no espaço e no tempo, associadas a fatores como característica e uso do solo, são fatores ativos na resposta hidrológica da bacia.

Para o preenchimento de falhas é essencial utilizar pontos pluviométricos com alta correlação espacial. Leivas et al. (2006), propõe espaços hidrologicamente homogêneos. Propondo uma abordagem adequada de examinar as estações quando os dados são aproximados garantindo uma correlação mais consistente.

Estudos recentes demonstram que o uso de séries temporais longas é essencial para análises hidroclimáticas robustas. Pois os dados em série histórica possibilitam um melhor resultado, fornecendo uma melhor precisão nos modelos hidrológicos (Volpi et al., 2024; Portela, 2011).

Com o objetivo de estimar precipitação ou preencher falhas de regiões hidrologicamente homogêneas este estudo teve como objetivo preencher falhas de um grupo de pequenas bacias, GI7, no interior de Pernambuco.

2. Material e métodos

2.1. Caracterização da área de estudo

O grupo de bacias hidrográficas GI-7, estão situadas em uma Unidade de Planejamento Hídrico (UP26) inserida na bacia hidrográfica do rio Garças, formada por um grupo de bacias de pequenos rios interiores, na parte sudoeste do estado de Pernambuco (Figura 1), com coordenadas de 08° 44' 31" e 09° 06' 38" de latitude sul, e 39° 41' 49" e 40° 15' 02" de longitude oeste. A GI7 está nos limites, à norte com a bacia do riacho Garças, localizado na (UP12), até o sul e leste com o rio São Francisco, ficando a oeste da bacia do rio Pontal, localizado na (UP13) de acordo com plano estadual de recursos hídricos de Pernambuco, 1988 (APAC, 2017). Neste estudo foram analisadas cinco estações pluviométricas de acordo com a distribuição espacial mostrada na Figura 1. Mas, para preenchimento das falhas destas estações foram utilizadas 11, em que seis estavam no entorno da bacia.

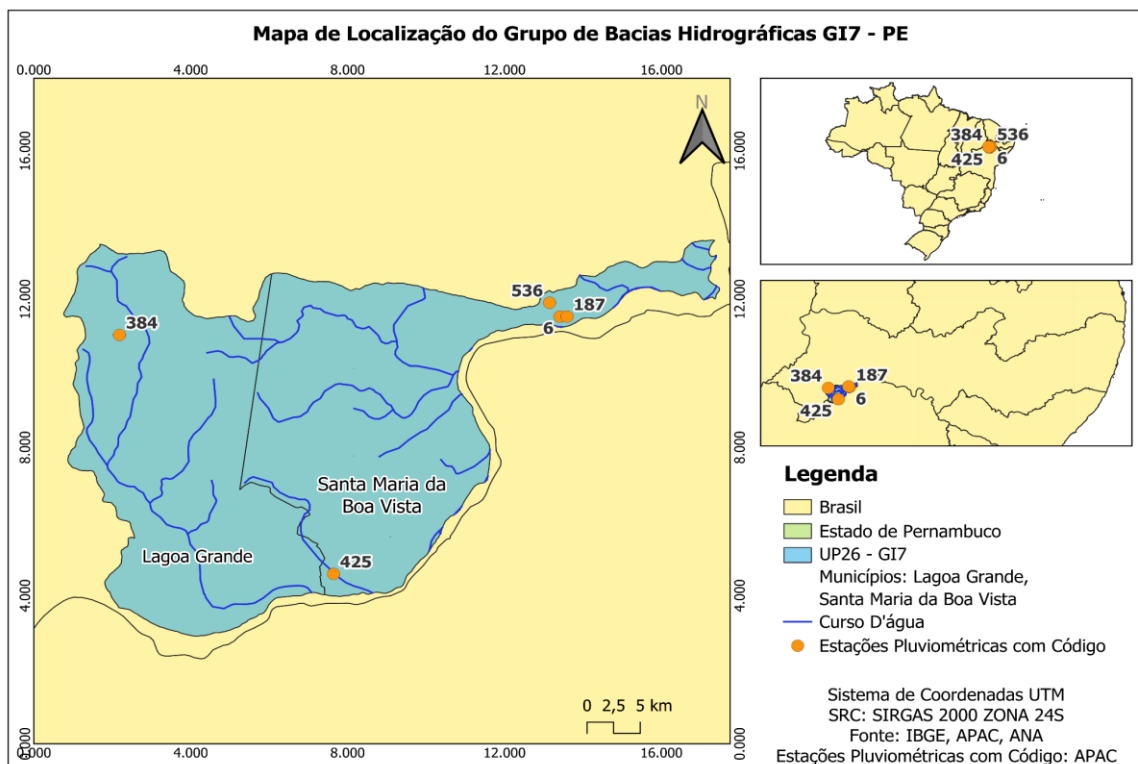


Figura 1. Localização espacial do grupo de pequenas bacias hidrográficas de rios interiores, GI7.

2.1.2 Drenagem

Ao compreender o regime hídrico de uma bacia hidrográfica, o entendimento de diferentes estudos sobre planejamentos hídricos dos diferentes usos da água é um fator indispensável para uma gestão bem alinhada dos recursos hídricos (Santos et al., 2009). O principal curso d'água da GI7, trata-se do riacho Recreio, com uma média de 59 km de extensão, o riacho Curral Novo e o

riacho Canaã também auxiliam com aproximados 14 km e 20 km de extensão, ambos deságuam a margem esquerda do rio São Francisco, irrigando os municípios de Lagoa Grande e Santa Maria da Boa Vista. O grupo de rios interiores GI-7 representa uma área de 1.205,91 km, que representa o percentual de 1,23% do estado de Pernambuco, Figura 2.

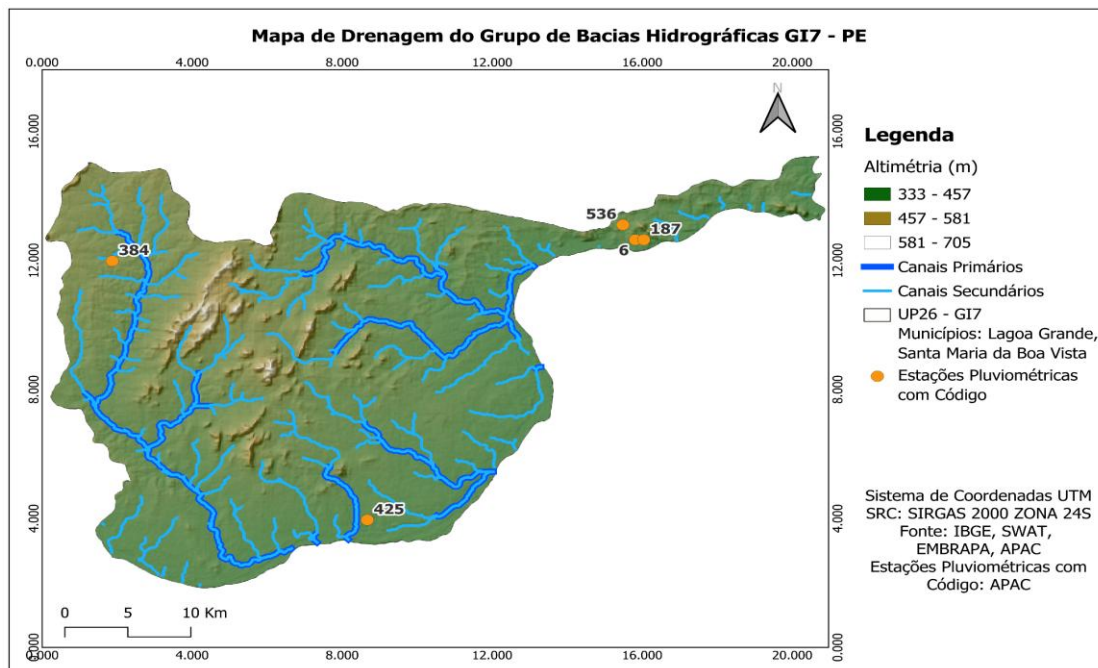


Figura 2. Drenagem do grupo de pequenas bacias hidrográficas de rios interiores, GI7.

2.1.3 Hipsometria

Ambos os municípios localizados na GI7 apresentam um relevo de terrenos planos com algumas ondulações singulares auxiliando o sistema de drenagem da região. Os solos presentes apresentam grandes porcentagens de argila que dificultam a drenagem, consequentemente o pouco de recurso hídrico que chega nessas regiões, logo sofre o processo de evapotranspiração. De acordo com Junior e Vanzela

(2016) o relevo desempenha uma influência robusta sobre a variação espacial das precipitações. O clima do grupo de pequenas bacias hidrográficas trata-se do Semiárido, pois a região está inserida em uma zona inter-tropical, clima caracterizado por altas temperaturas, precipitação sazonal, podendo chegar a escassez hídrica, Figura 3.

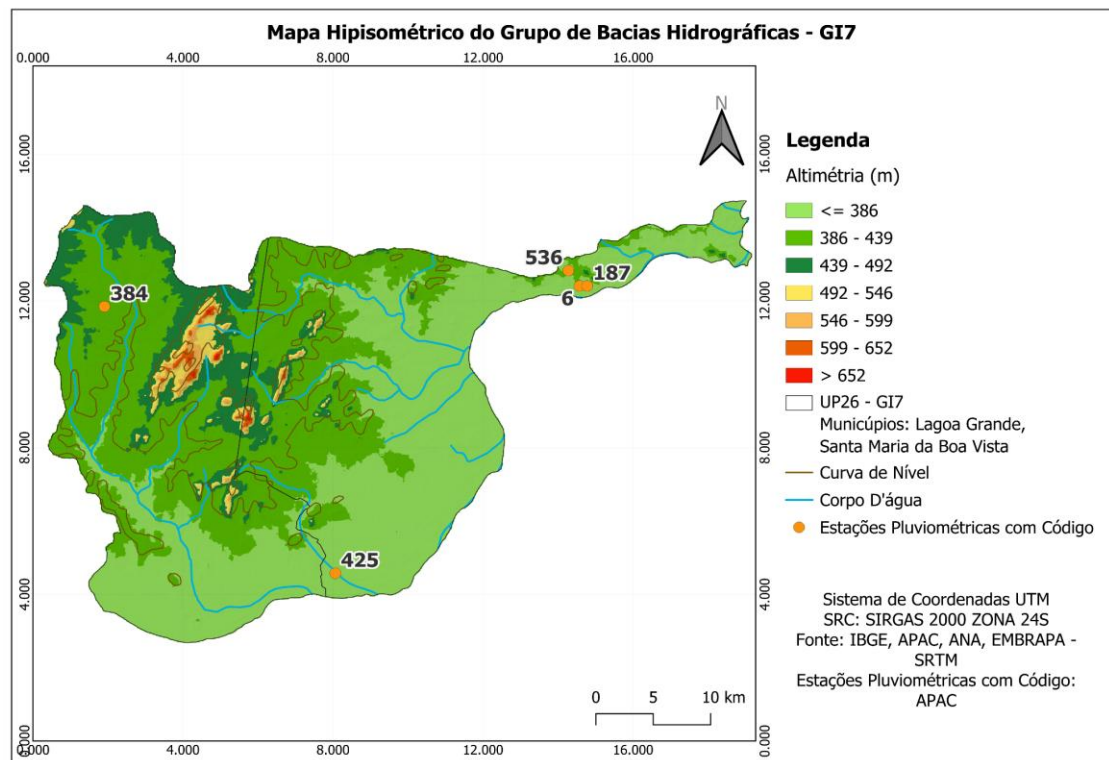


Figura 3. Hipsometria do grupo de pequenas bacias hidrográficas de rios interiores, GI7.

2.1.4 Uso do Solo

A GI7 está localizada entre dois municípios, Lagoa Grande, predominante ocupada pela fruticultura, com foco na produção de uvas e desenvolvimento no

comércio de vinhos. Nomeada como capital da uva. O município de Santa Maria da Boa Vista é abrangentemente ocupado pela agropecuária e grandes áreas de pasto. Vegetação nativa de ambas, Caatinga florestal e Caatinga, Figura 4.

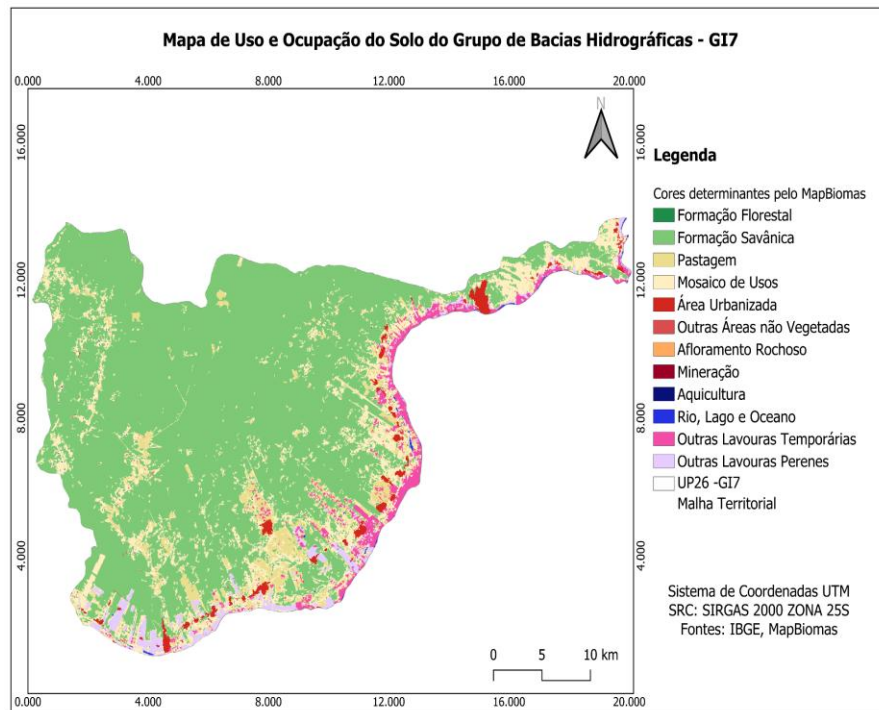


Figura 4. Uso e ocupação do solo do grupo de pequenas bacias hidrográficas de rios interiores, GI7.

2.1.5 Tipos de solos

Composto por argissolos PA - amarelo ou PVA - vermelho - amarelo, caracterizados pela grande influência erosiva, o município de Lagoa Grande enfrenta grandes desafios no manejo para adequação de

um solo frágil para um solo agrícola. Tornando necessário um estudo para analisar quais propriedades são eficazes, pois o solo argiloso exige uma preparação específica para manter tanto as plantações durante o ano todo, quanto os ecossistemas nativos da região, Figura 5.

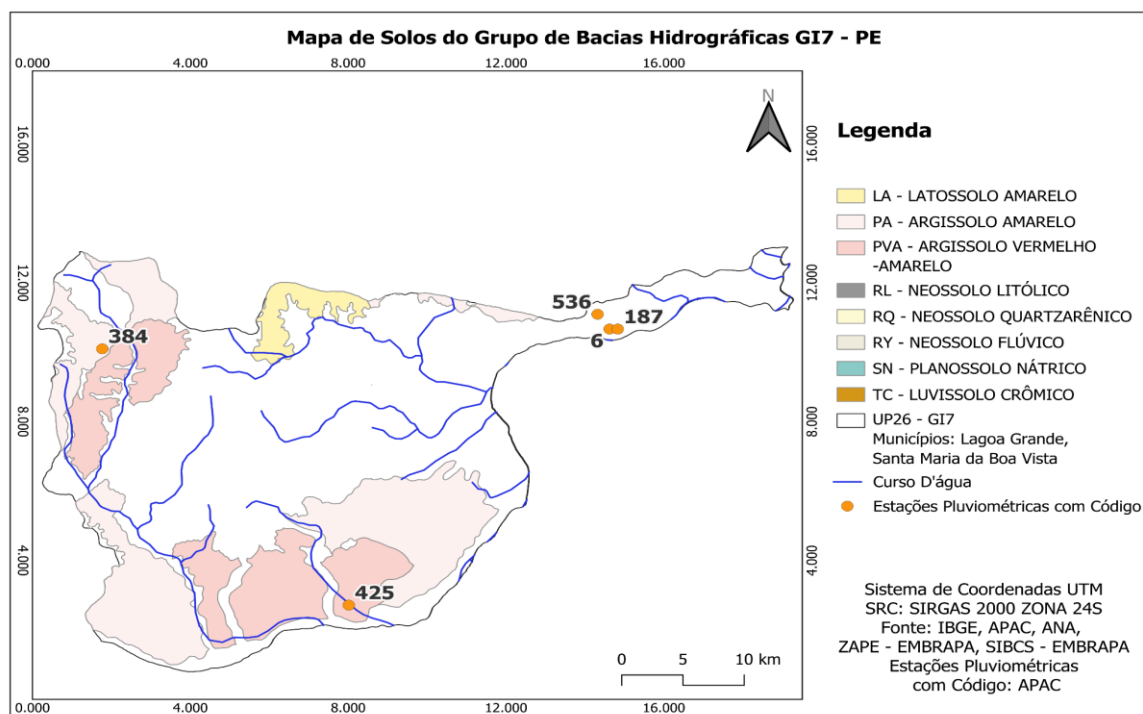


Figura 5. Solos do grupo de bacias hidrográficas de rios interiores, GI7.

3. Dados utilizados

Os dados utilizados no presente estudo foram os de precipitação, obtidos através das 11 estações pluviométricas, sendo que para análise neste estudo foram analisadas apenas as que estão no âmbito do grupo de pequenas bacias GI7, codificadas como: p06, p187, p536, p384 e p425. A análise dos mesmos está estimada em uma série histórica diária de 61 anos, entre os anos de 1961 a 2021. Esses dados foram obtidos na página da Agência Pernambucana de Águas e Clima-APAC. Essas estações estão no âmbito de dois municípios, Quadro 1.

Quadro 1. Estações pluviométricas localizadas pelos municípios da GI7.

Municípios - GI7	Estações Pluviométricas
Lagoa Grande - PE	p384
Santa Maria da Boa Vista - PE	p06, p187, p536, p425

3.1 Método CUBIST

Inicialmente foram contados os dias com falhas de dados de precipitação, utilizando o excel. Posteriormente, foi aplicado o SUPeR-Fill que tem como base o modelo CUBIST e Qmap que se utiliza

da programação em R para processamento dos dados. Vale destacar que o pacote CUBIS e QMAP faz ponderação regional dinâmica que é mais avançado do que as ponderações regionais tradicionais utilizando o IDW-Intensidade, duração e frequência e a Krigagem. No caso do pacote CUBIST a ponderação dinâmica é feita por árvore de decisão. Para cada folha se desenvolve uma equação de regressões com pesos para a estações vizinhas.

O desenvolvido do script em R contou com ajuda da Inteligência Artificial-IA e rodado no programa RStudio.

Para o preenchimento de falhas a série foi dividida em duas partes, 70% para calibração e 30% para validação. Os dados de saída do SUPeR-FILL estão organizados no formato de entrada do SWAT-Soil Water Assessment Tool. O SUPeR-Sistema de Unidades de Respostas Hidrológicas para Pernambuco possui como base metodológica o SWAT. Por isso, o programa utilizado neste estudo e desenvolvido por Brainer et al., (2025) se chamou SUPeR-FILL.

Neste estudo foi utilizado a IA generativa para auxiliar na correção do script em R. Todo o conteúdo gerado por IA foi criteriosamente revisado, adaptado e validado.

O aplicativo SUPER-FILL tem como base as seguintes metodologias: 1) usa do método de agrupamento do vizinho mais próximo, através da dissimilaridade, com uso da distância euclidiana. A vizinhança entre estações é calculada em LAT/LONG (graus) com distância Euclidiana, equação 01:

$$d((lat_i, lon_i), (lat_j, lon_j)) = \sqrt{(lat_i - lat_j)^2 + (lon_i - lon_j)^2}$$

Para cada estação-alvo são obtidos os índices dos vizinhos mais próximos. Após o agrupamento dos dados por dissimilaridade e usando o método do vizinho mais próximo é aplicado o modelo de regressão CUBIST. Ou seja, cria equações lineares por árvores de decisões e para cada árvore cria regressões para suas folhas. O grande destaque da regressão por CUBIST é que chega a esse detalhe por folha.

2) Regressão Cubist com ajuste por vizinhos de casos

O Cubist aprende regras (partições do espaço de atributos) e um modelo linear por folha. A previsão “básica” de um caso x é $\hat{y}_{rules}(x)$. Ou seja, para cada x valor eu estimo x ($y^{\wedge}$) com n rodadas (n regras). No caso, rodadas é uma condição que define um resultado. Quando usamos neighbors (vizinhos) = k no predict do Cubist, aplica-se um ajuste por casos vizinhos no espaço de atributos (isso não é a vizinhança espacial de estações, mas a vizinhança em valores similares de precipitação). É um ajuste de resíduo:

$$\hat{y}(x) = \hat{y}_{rules}(x) + \frac{1}{k} \sum_{i \in \mathcal{N}_k(x)} (y_i - \hat{y}_{rules}(x_i))$$

Em que $\mathcal{N}_k(x)$ são os k de treinamento mais próximos de x por similaridade em x , que inclui as séries dos vizinhos espaciais e o mês denominado de “mon” no aplicativo SUPER-FILL), Brainer et al., (2025). Isso leva a previsão na direção da estrutura local de resíduos. O aplicativo SUPER-FILL utiliza uma função de transformação paramétrica, que nada mais é do que um método matemático para corrigir e

ajustar os dados de precipitação. Nesse processo, aplica-se o ajuste linear pelo método dos mínimos quadrados, uma técnica estatística usada para encontrar a linha que melhor representa a relação entre os dados observados e os valores estimados. Em termos simples, o método dos mínimos quadrados procura minimizar a diferença entre os valores medidos (com possíveis falhas ou inconsistências) e os valores ajustados, reduzindo os erros ao máximo. O resultado dessa etapa é uma série de precipitação contínua, sem falhas e calibrada, pronta para ser utilizada em análises hidrológicas e modelos climáticos com maior confiabilidade.

No SUPER-FILL é efetuado três iterações. A primeira usa apenas os vizinhos espaciais e o mon (a variável *mon* funciona como um indicador do mês em que a observação ocorre. Ela ajuda o modelo a capturar a sazonalidade da precipitação (chuvas de verão, período seco, etc.), aumentando a coerência temporal. Ou seja, mon = mês do dado climático (1 = janeiro, 2 = fevereiro, ..., 12 = dezembro), usado como variável explicativa no preenchimento das falhas). Na segunda e terceira iterações, como a estação alvo ganhou valores preenchidos na primeira iteração, ela pode entrar nos preditores, enriquecendo o modelo e melhorando a coerência temporal/local. Ao final da terceira iteração grava e consolida a série temporal de precipitação preenchida.

O número de atributos utilizados no modelo varia de acordo com o número de estações, sazonalidade das estações e valores preenchidos na iteração anterior.

3.3 Critérios de avaliação do desempenho do modelo

- 1- Sensitivity (Sensibilidade ou Recall)- Mede a proporção de verdadeiros positivos identificados corretamente.
2. Specificity (Especificidade)-Mede a proporção de verdadeiros negativos identificados corretamente.
- 3- Pred Value (Valor Preditivo Positivo / Precision)- Entre todos os classificados como positivos, quantos realmente são positivos.
4. Neg Pred Value (Valor Preditivo Negativo)- Entre todos os classificados como negativos, quantos realmente são negativos.
5. Precision é o mesmo que o Valor Preditivo Positivo.

6. Recall é o mesmo que Sensibilidade. Mede a capacidade de encontrar todos os positivos.
7. F1 Score - Média harmônica entre Precisão (Precision) e Recall (Sensibilidade).
8. Prevalence – Proporção de casos positivos no conjunto de dados.
9. Detection Rate – Proporção de todos os casos corretamente detectados como positivos.
10. Detection Prevalence – Proporção de indivíduos previstos como positivos pelo modelo.
11. Balanced Accuracy – Média da Sensibilidade e Especificidade.

3.4 Critérios de avaliação da qualidade do modelo

MAE (Erro Médio Absoluto)

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - p_i|$$

Em que, n é o número de amostras, y é o valor observado para cada amostra, p é o valor previsto pelo modelo para cada amostra, | | representa o valor absoluto

3.4 Erro relativo

O erro absoluto é a diferença entre o valor previsto pelo modelo e o valor observado. O erro relativo, por outro lado, é o erro absoluto em relação ao valor observado. Ele é dado pelo seguinte cálculo:

Erro relativo = $\frac{|\text{valor observado} - \text{valor estimado}|}{\text{valor observado}}$

3.5 Coeficiente de correlação

É uma medida estatística que quantifica a relação entre duas variáveis. O coeficiente de correlação de Pearson, em particular, mede a força e a direção da relação linear entre duas variáveis contínuas. O valor do coeficiente varia entre -1 e +1.

$$r_{xy} = \frac{Cov(X, Y)}{S_x S_y} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}}$$

Interface do SUPER-FILL

Mais detalhes sobre o SUPER-FILL podem ser encontrados em Brainer et al., (2025) o download do programa/aplicativo pode ser obtido na página da Laboratório de Sensoriamento Remoto e Geoprocessamento-SERGEIO, link <https://sites.ufpe.br/sergeo/>.

4. Resultados

Critérios de avaliação do desempenho do modelo para as estações p6, p187, p384, p425, p536.

O Quadro 2 apresenta as estatísticas de desempenho do preenchimento de falhas das cinco séries temporais analisadas neste estudo. A estação p536 apresentou os menores valores de MAE e RMSE, além do maior coeficiente de determinação (R^2), indicando, em termos estatísticos, um excelente ajuste do modelo. Entretanto, esse resultado deve ser interpretado com cautela, pois a estação possui uma quantidade menor de dados disponíveis e, consequentemente, uma amostragem mais reduzida. Em situações como essa, é comum que os indicadores estatísticos apresentem desempenho aparentemente superior, sem necessariamente refletir um preenchimento mais robusto ou representativo da variabilidade climática real.

Por outro lado, a estação p384, que possuía uma série mais extensa e com menor quantidade de falhas, apresentou desempenho estatístico inferior. Isso ocorre porque séries mais longas e completas tendem a incorporar uma maior variabilidade temporal e espacial da precipitação, incluindo eventos extremos, meses secos e anos anômalos, tornando o processo de preenchimento mais complexo e reduzindo naturalmente os valores dos indicadores estatísticos.

Dessa forma, não se pode afirmar que uma estação com melhores métricas apresenta necessariamente o melhor preenchimento. É fundamental considerar, além dos resultados estatísticos, a quantidade de dados

observados, a distribuição temporal das falhas, a representatividade espacial da estação e sua capacidade de capturar os padrões sazonais e extremos de precipitação.

Esses resultados reforçam a importância de investimentos contínuos em monitoramento pluviométrico e ampliação da rede observacional, uma vez que modelos de preenchimento dependem diretamente da disponibilidade de dados consistentes para representar adequadamente a variabilidade espacial e sazonal das chuvas. A Figura 6 apresenta o

comportamento final das cinco séries temporais após o preenchimento das falhas.

Quadro 2. Análise de desempenho do modelo para as estações da GI7.

Estação	MAE Erro Médio Absoluto	RMSE Erro Médio ao Quadrado	R2 Coeficiente de Determinação
p6	0.101593072	1.225395006	0.9587355488
p187	0.07295929151	0.7692171883	0.9785142385
p384	0.2984349913	3.057714089	0.8263381907
p425	0.2087896991	2.120256197	0.9055848389
p536	0.03322663715	0.2265695499	0.9988194996

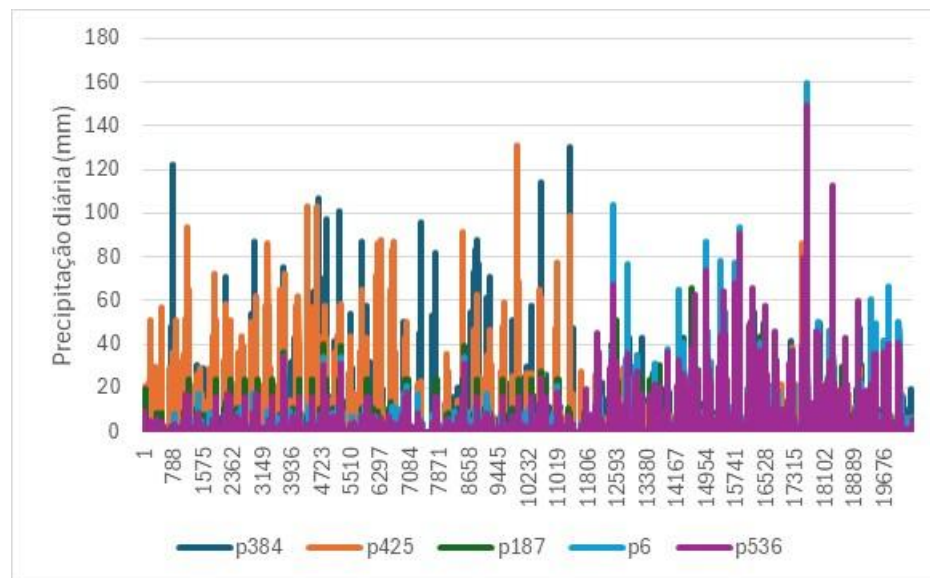


Figura 6- Resultado das estações preenchidas para o grupo de pequenas bacias GI7.

Na primeira iteração o modelo aprende com a própria estação. Por exemplo, para a estação p6 que foi observado 7261 dados existentes foi aplicado duas regras. A primeira usou 7046 casos observados, 13 atributos (11 estações, sazonalidade e preditores da primeira iteração) e obteve média de 0,26 para os dados de precipitação entre 0 a 48mm e erro de 0,09. Para os 215 casos existentes restantes dos 7261 casos foi obtida média de 28,75mm no intervalo de 10,1 a 157,5mm, com erro de 0,06. Foi desenvolvido as equações de regressão de acordo com essas regras e comparados com os valores observados. Os resultados estatísticos estão apresentados no quadro 3. As equações e regras estão listadas abaixo:

Rodada 1: [7046 casos, média 0,26, intervalo 0 a 48, est err 0,09]

Se

$$p6 \leq 10,20683$$

então

$$\text{outcome} = -0 + 0,98 p6$$

Rodada 2: [215 casos, média 28,75, intervalo 10,1 a 157,5, est err 0,06]

Se

$$p6 > 10,20683$$

então

$$\text{outcome} = 0,12 + 0,98 p6$$

Quadro 3

Average error	0.11
Relative error	0.06
Correlation coefficient	0.98

Estação p187

No modelo para a estação p187 foram utilizados 8153 casos observados e 13 atributos. Inicialmente, o modelo usou 7613 casos, com média 0,05, intervalo 0

para 19,5, est err 0,03. O quadro 4 mostra os resultados estatísticos do preenchimento dessa estação. As equações e regras estão listadas abaixo:

se
 $p_{187} \leq 1,619977$
então
 $outcome = -0 + 0,981 p_{187}$
Rodada 2: [22 casos, média 2,03, intervalo 0 a 23, est err 2,24]
se
 $p_{116} \leq 0,7787758$
 $p_{187} > 1,619977$
 $p_{187} \leq 8,302382$
 $p_{242} \leq 5,840137$
 $p_{430} \leq 0,02174362$
 $p_{536} \leq 1,811626$
então
 $outcome = -0,98 + 0,777 p_{187} - 0,66 p_{116}$
Rodada 3: [585 casos, média 3,01, r intervalo 0 a 65,3, est err 0,01]
se
 $p_{116} > 0,7787758$
então
 $outcome = -0 + 0,988 p_{187}$
Rodada 4: [27 casos, média 3,89, intervalo 0 a 13,2, est err 1,61]
se
 $p_6 \leq 6,397078$
 $p_{116} \leq 0,7787758$
 $p_{187} > 1,619977$
 $p_{187} \leq 8,302382$
 $p_{242} \leq 5,840137$
 $p_{430} > 0,02174362$
 $p_{536} \leq 1,811626$
então
 $outcome = -0,02 + 0,997 p_{187}$
Rodada 5: [393 casos, média 9,51; intervalo 0 a 96,7 e est err 0,29]
se
 $p_{242} \leq 5,840137$
 $p_{536} > 1,811626$
então
 $outcome = -0 + 0,988 p_{187}$
Rodada 6: [331 casos, média 10,13; intervalo 0 a 149,5 e est err 0,03]
se

$p_{242} > 5,840137$
então
 $outcome = -0 + 0,988 p_{187}$
Rodada 7: [243 casos, média 22,07, intervalo 1,8 a 149,5, est err 0,03]
se
 $p_6 > 6,397078$
 $p_{187} > 1,619977$
então
 $outcome = -0,01 + 0,988 p_{187}$
Rodada 8: [262 casos, média 24,03, intervalo 0 a 149,5, est err 0,58]
se
 $p_{187} > 8,302382$
então
 $outcome = -0,01 + 0,988 p_{187}$

Quadro 4.

Average error 	0.13
Relative error 	0.07
Correlation coefficient	0.98

Attribute usage:

Conds Model

86% 100% p_{187}
8% p_{242}
7% p_{116}
5% p_{536}
3% p_6

Estação p384

No modelo para a estação p384 foram utilizados 10646 casos observados e 13 atributos. Os resultados estatísticos estão apresentados no quadro 5. As equações e regras estão listadas abaixo:

Rodada 1: [10048 casos, média 0,30, intervalo 0 a 113,4, est err 0,18]
se
 $p_{384} \leq 7,642558$
então
 $outcome = -0 + 0,96 p_{384}$
Rodada 2: [10025 casos, média 1,10, intervalo 0 a 126,2, est err 0,23]
se
 $p_{31} \leq 3,892502$

então
 $\text{outcome} = -0,03 + 0,97 \text{ p384}$
 Rodada 3: [485 casos, média 10,53, intervalo 0 a 103,2, est err 0,05]
 se
 $\text{p247} > 5,926103$
 então
 $\text{outcome} = -0,03 + 0,97 \text{ p384}$
 Rodada 4: [26 casos, média 11,80, intervalo 0 a 51, est err 10,17]
 se
 $\text{p31} > 3,892502$
 $\text{p242} > 5,256171$
 $\text{p247} \leq 5,926103$
 $\text{p332} \leq 9,062278$
 $\text{p384} > 7,642558$
 então
 $\text{outcome} = -5,09 + 0,87 \text{ p384}$
 Rodada 5: [466 casos, média 13,13, intervalo 0 a 110,4, est err 0,14]
 se
 $\text{p332} > 9,062278$
 então
 $\text{outcome} = -0,03 + 0,97 \text{ p384}$
 Rodada 6: [21 casos, média 15,45, intervalo 0 a 78,6, est err 13,41]
 se
 $\text{p31} > 3,892502$
 $\text{p332} \leq 9,062278$
 $\text{p384} > 7,642558$
 $\text{p430} > 1,849119$
 então
 $\text{outcome} = -6,87 + 1 \text{ p384} - 0,11 \text{ p187} - 0,04 \text{ p31}$
 Rodada 7: [20 casos, média 16,29, intervalo 0 a 50,4, est err 1,42]
 se
 $\text{p31} > 3,892502$
 $\text{p242} \leq 5,256171$
 $\text{p247} \leq 5,926103$
 $\text{p332} \leq 9,062278$
 $\text{p384} > 7,642558$
 $\text{p430} \leq 1,849119$
 então
 $\text{outcome} = 0,02 + 1 \text{ p384} - 0,52 \text{ p242} - 0,21 \text{ p187} - 0,08 \text{ p31} + 0,06 \text{ p247}$

Quadro 5.

Average error 	0.51
Relative error 	0.18
Correlation coefficient	0.89

Attribute usage:

Conds Model

48%	100%	p384
48%		p31
3%		p332
3%		p247

Estação p425

A interação desenvolvida pelo modelo para a estação p425 foi de 9943 casos observados e 13 atributos. Os resultados estatísticos estão apresentados no quadro 6. As equações e regras estão listadas abaixo:
 Rodada 1: [9389 casos, média 0,12, intervalo 0 a 70,4, est err 0,11]

se

$\text{p425} \leq 5,452958$

então

$\text{outcome} = -0 + 0,97 \text{ p425}$

Rodada 2: [109 casos, média 2,84, intervalo 0 a 60, est err 2,06]

se

$\text{p6} > 3,941481$

$\text{p242} \leq 5,469117$

então

$\text{outcome} = -6,37 + 0,98 \text{ p425}$

Rodada 3: [290 casos, média 18,27, intervalo 0 a 89, est err 1,76]

se

$\text{p6} \leq 3,941481$

$\text{p242} \leq 5,469117$

$\text{p425} > 5,452958$

então

$\text{outcome} = 0,03 + 0,97 \text{ p425}$

Rodada 4: [234 casos, média 27,22, intervalo 0 a 127,2, est err 0,14]

se

$\text{p242} > 5,469117$

$\text{p425} > 5,452958$

então

$\text{outcome} = 0,04 + 0,97 \text{ p425}$

Quadro 6.

Average error 	0.46
Relative error 	0.19
Correlation coefficient	0.94

Attribute usage:

Conds Model

99% 100% p425

6% p242

4% p6

Estação p536

Para a estruturação do modelo estimado da estação p536 foram utilizados 4646 casos observados e 13 atributos. Os resultados estatísticos estão apresentados no quadro 7. As equações e regras estão listadas abaixo:

Rodada 1: [4646 casos, média 1,13, intervalo 0 a 147, est err 0,02]

outcome = -0 + 0,985 p536

Quadro 7.

Average error 	0.02
Relative error 	0.01
Correlation coefficient	1.0

Attribute usage:

Conds Model

100% p536

O Quadro 8 apresenta a quantidade de dados observados e os percentuais de falhas das estações pluviométricas no período de 1961 a 2021. Observa-se que a estação p536 foi a que apresentou a maior quantidade de falhas, enquanto a estação p384 registrou a menor proporção de dados ausentes.

Essas informações são fundamentais para a interpretação dos resultados estatísticos, uma vez que séries históricas com grande quantidade de lacunas tendem a comprometer a representação adequada da sazonalidade, da variabilidade interanual e da ocorrência

de eventos extremos. Em séries mais curtas ou fragmentadas, a capacidade do modelo em reproduzir padrões climáticos locais torna-se mais limitada, podendo influenciar diretamente métricas como MAE, RMSE e R².

Por outro lado, o preenchimento de falhas por meio do modelo Cubist mostrou-se altamente relevante, pois utiliza ponderação regional dinâmica, agrupamento de estações semelhantes e árvores de decisão para estimar os valores ausentes. Essa abordagem permite representar melhor a variabilidade espacial e temporal das chuvas, mesmo em séries com poucos dados disponíveis. Dessa forma, o uso do modelo contribui para tornar as séries mais consistentes e mais adequadas para análises de tendências climáticas, estudos hidrológicos e investigações relacionadas às mudanças climáticas.

Quadro 8- Quantidade de falhas nas estações analisadas.

Estação	Quant. De dados existentes	Porcentagem de dados existentes (%)	Período de dados
P6	7261	35	1961-2021
P187	8153	40	1961-2021
P384	10646	52	1961-2021
P425	9943	49	1961-2021
P536	4646	23	1961-2021

Agradecimentos

Ao CNPq através do projeto: Título: Desenvolvimento de um sistema automático, em tempo real, para assimilação, identificação e preenchimento de falhas da precipitação no SUPER e avaliação dos seus impactos na gestão de recursos hídricos de Pernambuco. Título em inglês: Real time automatic system for assimilation and gap filling of precipitation and evaluation of its impacts on the management of water resources in Pernambuco, processo número: 405853/2022-0. A FACEPE através do projeto título: Sistema de Unidades de Respostas Hidrológicas de

Pernambuco - SUPer para as pequenas bacias de Pernambuco, processo número: APQ-0044-3.07/24. Agradecemos à Universidade Federal do Agreste de Pernambuco (UFAPE), Agência Pernambucana de Águas e Clima-APAC, Agência Nacional de Águas-ANA e à Empresa Brasileira de Pesquisa Agropecuária (EMBRAPA) pelo apoio institucional ao desenvolvimento do SUPer-Fill, especialmente pela disponibilização de dados, contribuição técnica, incentivo à inovação e suporte logístico.

Conclusões

Conclui-se que o aplicativo SUPer-FILL foi eficiente no preenchimento das falhas das séries históricas de precipitação do grupo de pequenas bacias GI7, permitindo a construção de uma base de dados contínua e consistente para o período de 1961 a 2021. A combinação entre os modelos Cubist e Qmap demonstrou grande potencial para estimar valores ausentes em regiões semiáridas, onde a limitação de dados observados ainda é um dos principais entraves para análises hidroclimáticas mais robustas.

Os resultados mostraram que o desempenho estatístico do modelo varia conforme a quantidade de dados disponíveis em cada estação. Estações com menor número de observações, como a p536, apresentaram melhores indicadores estatísticos, como menor MAE, menor RMSE e maior R^2 . Entretanto, isso não significa necessariamente melhor preenchimento, pois séries com menor amostragem tendem a reduzir artificialmente os erros. Por outro lado, estações com séries mais longas e completas, como a p384, apresentaram métricas menos favoráveis, justamente por incorporarem maior variabilidade temporal, espacial e ocorrência de eventos extremos.

Dessa forma, a interpretação dos resultados não deve se limitar apenas aos indicadores estatísticos. É fundamental considerar também a extensão da série histórica, a quantidade de falhas existentes, a sazonalidade das chuvas, a representatividade espacial das estações e a capacidade do modelo em reproduzir eventos secos e chuvosos. A utilização de grupos hidrologicamente homogêneos e de variáveis sazonais mostrou-se essencial para aumentar a coerência das estimativas produzidas pelo modelo.

Além de melhorar a qualidade das séries pluviométricas, o preenchimento de falhas contribui diretamente para a confiabilidade de estudos hidrológicos, modelagem de bacias, simulações no SWAT/SUPer, avaliação de secas e cheias, planejamento agrícola e gestão dos recursos hídricos. Em áreas semiáridas como a GI7, onde a precipitação é irregular e altamente variável, a existência de bases históricas consistentes é indispensável para subsidiar decisões relacionadas ao abastecimento humano, irrigação, reservatórios e adaptação às mudanças climáticas.

Por fim, o estudo reforça a necessidade de ampliar os investimentos em monitoramento pluviométrico, manutenção das estações meteorológicas e desenvolvimento de tecnologias de inteligência artificial aplicadas às ciências ambientais. Ferramentas como o SUPer-FILL não devem substituir a coleta de dados em campo, mas sim complementar e fortalecer as redes de observação existentes, tornando os dados mais consistentes e úteis para o planejamento hidrossocial e econômico de Pernambuco.

Referências

- Amador, M. C.; Silva, C. A. da. 2020. O Preenchimento De Falhas De Séries Históricas De Precipitação. *Revista Eletrônica da Associação dos Geógrafos Brasileiros, Seção Três Lagoas*, v. 1, n. 31, p. 178-202, 1 jun. Acesso em: 15 out. 2025.
- Agência Pernambucana De Águas E Clima (APAC). Manual de padronização das publicações oficiais da Agência Pernambucana de Águas e Clima: publicações não periódicas. Recife: APAC, 2017. Disponível em: https://www.apac.pe.gov.br/images/media/1569955_307_manual_de_padronizacao_apac.pdf. Acesso em: 24 out. 2025. Acesso em: 10 set. 2025.
- Brainer, J. C. R., Silva, R. G. D., Santos, G. G. L. D., & Galvêncio, J. D. (2025). Aplicação de um Sistema Automático Inovador de preenchimento de falhas de precipitação para o SWAT/SUPer com Inteligência Artificial-SUPer-FILL para a bacia GL4. *Revista Brasileira De Geografia Física*, 18(6), 4875–4887. <https://doi.org/10.26848/rbgf.v18.6.p4875-4887>
- Brubacher, João Paulo; Oliveira, Guilherme G.; Guasselli, Laurindo A. 2020. Métodos de interpolação espacial para o mapeamento da precipitação pluvial. *Revista Brasileira de Engenharia Agrícola e Ambiental*, v. 24, n. 5, p. 321-

- 327, Disponível em: <https://www.scielo.br/j/rbeaa/a/MXGTqmPFvgFVt8VwrXpndqw/?lang=pt>. Acesso em: 23 out. 2025.
- Berezuk, A. G., & Garcia, M. R. (2011). Estudo De Tendências Do Ritmo Pluviométrico Da Bacia Hidrográfica Do Rio Pardo-MS. *Revista Brasileira De Climatologia*, 9. <https://doi.org/10.5380/abclima.v9i0.27506>. Acesso em: 18 out. 2025.
- CUBIST. *Cubist Regresion Models*. CRAN, [s. l.], 2025. Disponível em: <https://cran.r-project.org/web/packages/Cubist/vignettes/cubist.html>. Acesso em: 05 de Outubro de 2025.
- Dias, A. S.; Silva, M. R.; Oliveira, R. S.; et al. Preenchimento de falhas em séries de dados pluviométricos: uma análise para a região do semiárido brasileiro. *Revista Brasileira de Climatologia*, v. 27, p. 1–15, 2021. Disponível em: <https://www.scielo.br/j/rbclima/a/NJKZzyRfmcKGd6cyDZqxGb/>. Acesso em: 24 out. 2025.
- Depiné, H.; Castro, N. M. R.; Pinheiro, A.; Pedrollo, O. O. 2014. Preenchimento de falhas de dados horários de precipitação utilizando redes neurais artificiais. *Revista Brasileira de Recursos Hídricos*, v. 19, n. 1, p. 51-63, Acesso em: 20 out. 2025.
- Fagandini, C.; Todaro, V.; Tanda, M. G.; et al. 2024. *Missing rainfall daily data: a comparison among gap-filling approaches*. *Environmental Earth Sciences*, v. 83, n. 3, p. 1–16, DOI: 10.1007/s11004-023-10078-6. Acesso em: 23 out. 2025.
- Junior, H.N.; Vanzela, L.S. 2016. Influência do relevo sobre a distribuição espacial da precipitação no Araguatatuba - SP. In: XII Simpósio Brasileiro de Climatologia Geográfica, Goiânia, p. 2011-2024. Acesso em: 20 out. 2025.
- Leivas, J.F.; Berlato, M.A.; Fontana, D.C. 2006. Risco de deficiência hídrica decendial na metade sul do Estado do Rio Grande do Sul. *Rev. Bras. Eng. Agríc. e Ambient.*, v. 10, n. 2, p. 397-407, Acesso em: 22 out. 2025.
- Nascimento, Dênis Agüero Do; Oliveira, Mateus Ferreira De; Damasceno, Solange Batista; SILVA, Joecila Santos Da. Importância de estudo pluviométrico na gestão dos recursos hídricos: observações aplicadas em uma microbacia da Amazônia Central, https://files.abrhidro.org.br/Eventos/Trabalhos/60/PAP023240.pdf?utm_source=chatgpt.com. Acesso em: 18 out. 2025.
- Porto, Monica F. A.; Porto, Rubem La Laina. 2008. Gestão de bacias hidrográficas. *Estudos Avançados*, v. 22, n. 63, p. 43–60, Disponível em: <https://doi.org/10.1590/s0103-40142008000200004>. Acesso em: 20 out. 2025.
- Portela, Maria Manuela; Quintela, António De Carvalho; Santos, João Filipe; Vaz, Cristiana; Martins, Cláudia. 2011. *Tendências em séries temporais de variáveis hidrológicas*. *Revista Recursos Hídricos*, v. 32, n. 1, p. 43–60, Disponível em: <https://www.aprh.pt/rh/v32n1.html>. Acesso em: 24 out. 2025.
- Ruezzene, C. B., Miranda, R. B. de, Tech, A. R. B., & Mauad, F. F. (2021). Preenchimento De Falhas Em Dados De Precipitação Através De Métodos Tradicionais E Por Inteligência Artificial. *Revista Brasileira De Climatologia*, 29, 177–204. Recuperado de <https://ojs.ufgd.edu.br/rbclima/article/view/15184>. Acesso em: 15 out. 2025.
- Santos, G. G.; Figueiredo, C. C. de; Oliveira, L. F. C. de; Griebeler, N. P. 2009. Intensidade-duração-frequência de chuvas para o Estado de Mato Grosso do Sul. *Revista Brasileira de Engenharia Agrícola e Ambiental*, v.13, p.899-905, Acesso em: 22 out. 2025.
- Tercini, João Rafael Bergamaschi; Mello Júnior, Arisvaldo Vieira. 2024. Impacto das mudanças hidroclimáticas na segurança hídrica do Sistema Produtor de Água Cantareira, Brasil. *Revista Brasileira de Recursos Hídricos*, v. 29, e21, DOI: <https://doi.org/10.1590/2318-0331.292420230140>. Acesso em: 24 out. 2025.
- Volpi, Elena; Mancini, Corrado Paolo; Fiori, Aldo. *What can we learn from long hydrological time-series? The case of rainfall data at Collegio Romano, Rome, Italy. Journal of Hydrology X*, v. 230, p. 100176, 2024. DOI: <https://doi.org/10.1016/j.hydroa.2024.100176>. Acesso em: 24 out. 2025.
- Wilks, Daniel S. 2011. *Statistical Methods in the Atmospheric Sciences*. 3. ed. Amsterdam; Boston: Academic Press, 676 p. ISBN 978-0123850225. Acesso em: 20 out. 2025.