

UNIVERSIDADE FEDERAL DE VIÇOSA

CARLA LARISSA FONSECA DA SILVA

**MODELAGEM DE VAZÕES MÁXIMAS E MÍNIMAS COM BASE EM
APRENDIZADO DE MÁQUINA**

**VIÇOSA - MINAS GERAIS
2022**

CARLA LARISSA FONSECA DA SILVA

**MODELAGEM DE VAZÕES MÁXIMAS E MÍNIMAS COM BASE EM
APRENDIZADO DE MÁQUINA**

Dissertação apresentada à Universidade Federal de Viçosa, como parte das exigências do Programa de Pós-Graduação em Engenharia Agrícola, para obtenção do título de *Magister Scientiae*.

Orientador: Demetrius David da Silva

Coorientadores: Michel Castro Moreira
Jackson Martins Rodrigues

**VIÇOSA - MINAS GERAIS
2022**

**Ficha catalográfica elaborada pela Biblioteca Central da Universidade
Federal de Viçosa - Campus Viçosa**

T

S586m
2022

Silva, Carla Larissa Fonseca da, 1995-

Modelagem de vazões máximas e mínimas com base em
aprendizado de máquina / Carla Larissa Fonseca da Silva. –
Viçosa, MG, 2022.

1 dissertação eletrônica (137 f.): il. (algumas color.).

Inclui apêndices.

Orientador: Demetrius David da Silva.

Dissertação (mestrado) - Universidade Federal de Viçosa,
Departamento de Engenharia Agrícola, 2022.

Inclui bibliografia.

DOI: <https://doi.org/10.47328/ufvbbt.2023.028>

Modo de acesso: World Wide Web.

1. Bacias hidrográficas - Métodos estatísticos. 2. Muriaé,
Rio (MG). 3. Desenvolvimento de recursos hídricos. 4. Extremos
climáticos. 5. Aprendizado do computador. I. Silva, Demetrius
David da, 1966-. II. Universidade Federal de Viçosa.
Departamento de Engenharia Agrícola. Programa de
Pós-Graduação em Engenharia Agrícola. III. Título.

CDD 22. ed. 333.91098151

CARLA LARISSA FONSECA DA SILVA

**MODELAGEM DE VAZÕES MÁXIMAS E MÍNIMAS COM BASE EM
APRENDIZADO DE MÁQUINA**

Dissertação apresentada à Universidade Federal de Viçosa, como parte das exigências do Programa de Pós-Graduação em Engenharia Agrícola, para obtenção do título de *Magister Scientiae*.

APROVADA: 13 de dezembro de 2022.

Assentimento:

Documento assinado digitalmente
 CARLA LARISSA FONSECA DA SILVA
Data: 27/01/2023 10:10:15-0300
Verifique em <https://verificador.itl.br>

Carla Larissa Fonseca da Silva
Autora

DEMETRIUS DAVID DA
SILVA:54293472649

Assinado de forma digital por
DEMETRIUS DAVID DA
SILVA:54293472649
Dados: 2023.01.27 09:49:23
-03'00'

Demetrius David da Silva
Orientador

À minha família.

Dedico.

AGRADECIMENTOS

A Deus por sempre iluminar os meus caminhos e não permitir eu desistir.

Aos meus pais amados, Lourdes Santana e Antônio Carlos, pelo amor e apoio incondicionais mesmo diante de muitas dificuldades encontradas ao longo das nossas vidas. Eles que sempre me incentivaram a continuar lutando e alcançar voos cada vez mais longes, embora os seus corações ficassem apertados de saudade.

À Universidade Federal de Viçosa pela oportunidade de fazer o mestrado.

À Coordenação Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Código de Financiamento 001 e à Fundação de Amparo à Pesquisa do Estado de Minas Gerais (FAPEMIG) pelo apoio à pesquisa.

Ao professor Demetrius pela orientação tão enriquecedora com os seus anos de experiência.

Aos professores Michel e Jackson pela coorientação, conversas sábias e de muito aprendizado.

Ao professor Elpídio, um exímio mestre na arte de ensinar programação.

Ao meu irmão Carlos Maurício, que caminhou comigo durante muito tempo e me ajudou a superar as dificuldades encontradas para estudar.

À minha tia Ana e à minha madrinha Socorro, pessoas de corações gentis e amorosos. Obrigada pelo carinho que sempre tiveram comigo.

Ao Marco Meireles, meu grande incentivador e companheiro, que sempre esteve ao meu lado mesmo que um oceano estivesse nos separando por longos anos.

Aos meus colegas, em especial, Uilson, Guilherme Reis, Rafael Dias, Victor Duarte, Arthur e Luís, pela ajuda no entendimento de *machine learning* e processamento dos dados. Ao Rafael Petruceli, a quem devo imensa gratidão por todas as oportunidades ao longo do mestrado para me aprofundar no estudo da hidrologia. A todas as pessoas do CRRH/UFV que estiveram ao meu lado nos últimos dois anos.

Ao estagiário Igor Steven pelo auxílio na organização dos dados.

Às minhas amigas que fiz em Viçosa, em especial Jasmine, uma pessoa que levarei para sempre no coração e Eva, um ser contagiante. Agradeço também pela amizade de Nina, Fernanda, Dreice, Ana Luísa e Rayles. Vocês tornaram essa caminhada mais feliz.

BIOGRAFIA

CARLA LARISSA FONSECA DA SILVA, filha de Lourdes Santana Fonseca da Silva e Antônio Carlos Vieira da Silva, nasceu no dia 08 de novembro de 1995, na cidade de Capanema, Estado do Pará, Brasil.

Em agosto de 2014 iniciou o curso de Engenharia Ambiental e Energias Renováveis na Universidade Federal Rural da Amazônia (UFRA) com período sanduíche no curso de Mestrado Integrado em Engenharia do Ambiente da Universidade do Porto no ano acadêmico 2018/2019. A graduação foi finalizada em novembro de 2020.

Em novembro de 2020 começou a estudar o curso de Mestrado em Engenharia Agrícola, na área de Recursos Hídricos e Ambientais na Universidade Federal de Viçosa (UFV), na qual submeteu-se à prova pública em dezembro de 2022.

RESUMO GERAL

Silva, Carla Larissa Fonseca da, M.Sc. Universidade Federal de Viçosa, dezembro de 2022. **Modelagem de vazões máximas e mínimas com base em Aprendizado de Máquina.** Orientador: Demetrius David da Silva. Coorientadores: Michel Castro Moreira e Jackson Martins Rodrigues.

O comportamento hidrológico de diversas bacias hidrográficas tem sido afetado por fatores como uso e cobertura da terra, demanda hídrica e mudanças climáticas, levando ao comprometimento da disponibilidade hídrica ou à ocorrência de eventos extremos de vazão. Esse comportamento pode ser avaliado por meio de testes não-paramétricos, que mostram a tendência de variáveis hidroclimatológicas de interesse para a gestão ambiental. Além dessa análise, em estudos hidrológicos tem-se a necessidade de aumentar os horizontes de previsão da vazão, principalmente em bacias com frequente ocorrência de eventos de alta magnitude. Assim, objetivou-se compreender o comportamento hidrológico e suas possíveis mudanças nos últimos anos na bacia hidrográfica do rio Muriaé, além de modelar a vazão com base em diferentes cenários de dias de antecedência na sua previsão. No Capítulo 1 foram aplicados testes não-paramétricos para a identificação de tendências de mudança no comportamento hidrológico em anos/meses em diferentes níveis de significância, bem como a investigação de fatores que poderiam estar associados às mudanças. Os testes não-paramétricos mostraram tendência negativa dos valores extremos de vazão, principalmente no período de estiagem, com destaque para o mês de outubro. Na análise dos fatores, a precipitação e a demanda hídrica apresentaram indicativos de influência no regime de vazões, diferentemente do uso e cobertura da terra. Os diferentes métodos estatísticos permitiram encontrar resultados satisfatórios sobre o comportamento hidrológico da bacia, enquanto as análises dos fatores climáticos e antrópicos possibilitaram conhecer as possíveis influências nas mudanças de tendência da vazão. No Capítulo 2 foi explorada a utilização de técnicas de Aprendizado de Máquina como o Modelo Linear Generalizado (GLM), as *Splines* de Regressão Adaptativa Multivariada (MARS) e o *Random Forest* (RF) na previsão da vazão considerando cenários de dados com um, dois e três dias de antecedência. A modelagem mostrou que todas as técnicas tiveram um desempenho muito bom na previsão das vazões com um dia de antecedência. Por outro lado, houve uma redução do desempenho com o aumento do horizonte de previsão, principalmente no cenário de três dias. Em todos os períodos de antecedência a técnica RF se mostrou superior às demais, mesmo que tenha apresentado subestimação das vazões máximas. A modelagem em diferentes cenários de previsão com Aprendizado de Máquina mostrou a aplicabilidade dos algoritmos na

previsão da vazão, principalmente em bacias com ocorrência elevada de eventos máximos de vazão, com possibilidade de serem utilizados nos sistemas de alerta e, assim, reduzir os impactos socioambientais causados por esses eventos.

Palavras-chave: Comportamento hidrológico. Modelagem hidrológica. Períodos de antecedência.

ABSTRACT

Silva, Carla Larissa Fonseca da, M.Sc. Universidade Federal de Viçosa, December 2022. **Modeling of maximum and minimum streamflows based on Machine Learning**. Advisor: Demetrius David da Silva. Co-advisors: Michel Castro Moreira and Jackson Martins Rodrigues.

The hydrological behavior of several River basins has been affected by factors such as land use and land cover, water demand and climate change, leading to compromised water availability or the occurrence of extreme streamflow events. This behavior can be evaluated by means of non-parametric tests, which show the tendency of hydroclimatological variables of interest for environmental management. In addition to this analysis, in hydrological studies there is a need to increase the streamflow forecast horizons, especially in watersheds with frequent occurrence of high magnitude events. Thus, the objective was to understand the hydrological behavior and its possible changes in recent years in the Muriaé River basin, in addition to modeling the streamflow based on different scenarios of days in advance in its forecast. In Chapter 1, non-parametric tests were applied to identify trends of change in hydrological behavior in years/months at different levels of significance, as well as the investigation of factors that could be associated with the changes. The non-parametric tests showed a negative trend of extreme discharge values, mainly in the dry season, with emphasis on the month of October. In the factor analysis, precipitation and water demand showed influence on the streamflow regime, differently from land use and land cover. The different statistical methods made it possible to find satisfactory results on the hydrological behavior of the watershed, while the analysis of climatic and anthropic factors made it possible to know the possible influences on changes in streamflow trends. In Chapter 2, the use of Machine Learning techniques such as the Generalized Linear Model (GLM), Multivariate Adaptive Regression Splines (MARS) and Random Forest (RF) in the streamflow forecast considering scenarios of data with one, two and three days in advance. The modeling showed that all techniques performed very well in predicting streamflows one day in advance. On the other hand, there was a reduction in performance with the increase in the forecast horizon, mainly in the three-day scenario. In all periods of advance, the RF technique was superior to the others, even if it presented an underestimation of the maximum streamflows. Modeling in different prediction scenarios with Machine Learning showed the applicability of the algorithms in streamflow prediction, especially in River basins with a high occurrence of maximum streamflow events, with the possibility of being used in warning systems and, thus, reducing socio-environmental impacts caused by these events.

Keywords: Hydrological behavior. Hydrological modeling. Advance periods.

SUMÁRIO

INTRODUÇÃO GERAL	12
REFERÊNCIAS	15
CAPÍTULO 1	17
1.1 INTRODUÇÃO.....	18
1.2 MATERIAL E MÉTODOS.....	20
1.2.1 Área de estudo	20
1.2.2 Obtenção da base de dados	22
1.2.3 Variáveis hidroclimatológicas	26
1.2.4 Tendência e identificação de meses/anos de mudança no comportamento das variáveis hidroclimatológicas.....	27
1.2.4.1 Teste de Mann-Kendall (MK) e Mann-Kendall Modificado (MMK)	27
1.2.4.2 Teste de Correlação de Spearman	30
1.2.4.3 Teste de inclinação Theil-Sen	31
1.2.4.4 Teste de Pettitt.....	32
1.2.5 Uso e cobertura da terra	33
1.2.6 Outorgas de uso da água	34
1.3 RESULTADOS E DISCUSSÃO	35
1.3.1 Análise de tendência das séries de vazão.....	35
1.3.2 Análise dos meses/anos de mudança nas séries de vazão.....	41
1.3.3 Análise dos fatores que influenciam no comportamento das vazões.....	42
1.4 CONCLUSÕES	54
1.5 REFERÊNCIAS	55
CAPÍTULO 2	62
2.1 INTRODUÇÃO.....	63
2.2 MATERIAL E MÉTODOS.....	65
2.2.1 Desenvolvimento da modelagem com Aprendizado de Máquina	65
2.2.2 Área de estudo	66
2.2.3 Base de dados para a modelagem	67
2.2.4 Análise de estacionariedade das séries temporais.....	71
2.2.5 Padronização das amostras	71
2.2.6 Algoritmos de Aprendizado de Máquina para a previsão da vazão	71
2.2.6.1 Modelo Linear Generalizado (GLM)	71

2.2.6.2 <i>Splines</i> de Regressão Adaptativa Multivariada (MARS).....	72
2.2.6.3 <i>Random Forest</i> (RF)	72
2.2.7 Seleção de covariáveis	72
2.2.8 Treinamento e teste	73
2.2.9 Critérios de avaliação da modelagem	74
2.2.9.1 Métricas estatísticas	74
2.2.9.2 Vazões de referência do SAH-Muriaé	75
2.3 RESULTADOS E DISCUSSÃO	76
2.3.1 Seleção e importância das covariáveis.....	77
2.3.2 Comparação das técnicas de Aprendizado de Máquina.....	83
2.3.3 Vazões de referência do SAH-Muriaé	90
2.4 CONCLUSÕES	95
2.5 REFERÊNCIAS	95
CONCLUSÕES GERAIS	102
APÊNDICES	104
APÊNDICE A	105
APÊNDICE B.....	107
APÊNDICE C.....	108
APÊNDICE D	130
APÊNDICE E	135

INTRODUÇÃO GERAL

A análise de extremos hidroclimatológicos permite avaliar o comportamento hidrológico de bacias hidrográficas, que tem sido expressivamente afetado em decorrência das alterações no uso e ocupação da terra, na demanda hídrica, como também pelas mudanças climáticas nas últimas décadas (SPINONI *et al.*, 2020). Essas alterações são mais críticas em variáveis como a precipitação, impactando diretamente os recursos hídricos, seja no padrão de vazão, na umidade do solo ou na disponibilidade das águas superficiais e subterrâneas (CHAULUKA; SINGH; KUMAR, 2021).

Uma das formas de analisar o comportamento de uma bacia quanto a essas mudanças é conhecer a tendência de variáveis de importância para a gestão dos recursos hídricos, como as vazões mínimas de referência e as vazões máximas (SILVA *et al.*, 2020). A análise de estacionariedade dessas variáveis identifica a existência ou não de tendência na ocorrência de eventos extremos como secas e enchentes. Diversos autores (FANG *et al.*, 2018; HU *et al.*, 2017), inclusive, já observaram aumento da intensidade e da frequência de eventos extremos de precipitação em diferentes regiões do planeta.

Os testes estatísticos comumente utilizados na análise de tendência em dados hidroclimatológicos são os não-paramétricos, como os testes de Mann-Kendall (MK) (KENDALL, 1975; MANN, 1945), correlação de Spearman (SP) (SPEARMAN, 1904), inclinação de Theil – Sen (TSS) (SEN, 1968; THEIL, 1950) e Pettitt (PT) (PETTITT, 1979). O teste de MK apresenta como limitação a necessidade de dados independentes e aleatórios, a qual foi superada pela modificação feita por Hamed e Rao (1998) no cálculo da variância dos dados. Quanto ao teste SP, ele é equivalente ao teste de Mann-Kendall, de fácil implementação e é comumente utilizado com o teste de MK para fins de comparação ou análise complementar da tendência (ASHRAF *et al.*, 2021; MOREIRA *et al.*, 2019; SÖNMEZ; KALE, 2020). A inclinação de TSS, contudo, não apresenta nenhuma suposição, sendo aplicada em conjunto com os demais testes para identificar a magnitude da tendência (MOHAMMED *et al.*, 2022; NOURANI; DANANDEH MEHR; AZAD, 2018). O teste de PT, por sua vez, apresenta uma melhor taxa de falsos positivos comparado a outros testes que identificam meses/anos de mudança nas séries hidroclimatológicas (RYBERG; HODGKINS; DUDLEY, 2020).

Além de conhecer o comportamento hidrológico da bacia hidrográfica, torna-se necessário utilizar os dados disponíveis para a previsão de eventos extremos, como por exemplo, as vazões mínimas e as máximas, de modo a fornecer subsídios para as tomadas de decisão quanto ao planejamento dos recursos hídricos (TONGAL; BOOIJ, 2018). Nessa

perspectiva, surgiram diversos modelos de base física com a finalidade de prever vazões em curto e longo prazos (CHENG *et al.*, 2017). Entretanto, as incertezas e as complexidades na previsão, advindas da variabilidade ambiental, fazem com que a modelagem baseada em Aprendizado de Máquina, mesmo não sendo um processo de base física, seja muito utilizada nesses estudos (CHENG *et al.*, 2020; ZHOU *et al.*, 2022).

Na hidrologia, diferentes algoritmos de Aprendizado de Máquina têm sido utilizados de modo a superar a não linearidade dos dados ambientais. Ke *et al.* (2020) utilizaram 14 algoritmos para classificar enchentes com base nas características da precipitação de uma bacia frequentemente atingida por esses eventos. O estudo de Adikari *et al.* (2021) mostrou que técnicas de inteligência artificial podem ser utilizadas na previsão de eventos extremos tanto para climas áridos como tropicais.

A análise de tendência de variáveis de importância para a gestão dos recursos hídricos e a modelagem de dados hidrológicos baseada em Aprendizado de Máquina tornam-se ainda mais importantes em regiões em que há maior concentração populacional e/ou demanda hídrica e já são evidenciados conflitos e impactos econômicos e sociais advindos dos extremos hidrológicos, como ocorre na bacia do rio Muriaé, no Estado de Minas Gerais.

Nesse contexto, surgem quatro questões que apoiam as hipóteses do presente trabalho: i) Nos últimos anos houve evidências de mudanças de extremos hidrológicos na bacia hidrográfica do rio Muriaé? ii) Essas possíveis alterações estão associadas às alterações no uso e cobertura da terra, precipitação e/ou demanda hídrica da bacia? iii) Diante dessas prováveis mudanças é possível prever a vazão na bacia utilizando Aprendizado de Máquina? iv) Nessa perspectiva, é possível ampliar a previsão da vazão para diferentes dias de antecedência?

A fim de pesquisar e elucidar as questões levantadas, a presente dissertação foi dividida em dois capítulos. No Capítulo 1, *Análise de tendência e identificação dos possíveis períodos de mudança na ocorrência de eventos extremos de vazão na bacia hidrográfica do rio Muriaé*, apresenta-se o estudo de tendência e a identificação dos meses/anos dos períodos de mudanças na ocorrência de eventos extremos máximos e mínimos, por meio da aplicação de testes estatísticos não-paramétricos, na bacia hidrográfica do rio Muriaé, além da indicação dos fatores responsáveis pela alteração do comportamento das vazões.

Na sequência, no Capítulo 2, *Modelagem da vazão por meio de técnicas de Aprendizado de Máquina em diferentes horizontes de previsão*, buscou-se analisar três cenários de previsão de vazão, com um, dois e três dias de antecedência, e três algoritmos de Aprendizado de Máquina na bacia hidrográfica do rio Muriaé.

Sendo assim, essa dissertação visou contribuir com o entendimento de variáveis hidroclimatológicas de importância para a gestão dos recursos hídricos, bem como subsidiar a tomada de decisão a fim de minimizar os impactos advindos das alterações do uso e ocupação da terra assim como das mudanças climáticas. Além disso, as análises constituíram diferentes abordagens em termos de complexidade de processamento de dados, desde testes não-paramétricos à implementação de técnicas de Aprendizado de Máquina, o que amplia a possibilidade de replicação dos resultados para outras áreas de interesse ambiental.

REFERÊNCIAS

- ADIKARI, K. E. *et al.* Evaluation of artificial intelligence models for flood and drought forecasting in arid and tropical regions. **Environmental Modelling & Software**, v. 144, p. 105136, 1 out. 2021.
- ASHRAF, M. S. *et al.* Streamflow Variations in Monthly, Seasonal, Annual and Extreme Values Using Mann-Kendall, Spearman's Rho and Innovative Trend Analysis. **Water Resources Management**, v. 35, n. 1, p. 243–261, 2021.
- CHAULUKA, F.; SINGH, S.; KUMAR, R. Rainfall and streamflow trends of Thuchila River, Southern Malawi. **Materials Today: Proceedings**, v. 34, p. 846–855, 1 jan. 2021.
- CHENG, M. *et al.* Long lead-time daily and monthly streamflow forecasting using machine learning methods. **Journal of Hydrology**, v. 590, 1 nov. 2020.
- CHENG, MEILING *et al.* Performance assessment of spatial interpolation of precipitation for hydrological process simulation in the Three Gorges Basin. **Water (Switzerland)**, v. 9, n. 11, 1 nov. 2017.
- FANG, J. *et al.* Observed changes in hydrological extremes and flood disaster in Yangtze River Basin: spatial-temporal variability and climate change impacts. **Natural Hazards**, v. 93, n. 1, p. 89–107, 1 ago. 2018.
- HAMED, K. H.; RAO, A. R. A modified Mann-Kendall trend test for autocorrelated data. **Journal of Hydrology**, v. 204, p. 182–196, 1998.
- HU, M. *et al.* Assessment of hydrological extremes in the Kamo River Basin, Japan. **Hydrological Sciences Journal**, v. 62, n. 8, p. 1255–1265, 2017.
- KE, Q. *et al.* Urban pluvial flooding prediction by machine learning approaches – a case study of Shenzhen city, China. **Advances in Water Resources**, v. 145, p. 103719, 2020.
- KENDALL, M. G. **Rank Correlation Methods**. 4. ed. London: Charles Griffin, 1975.
- MANN, H. B. Nonparametric Tests Against Trend. **Econometrica**, v. 13, n. 3, p. 245–259, 1945.
- MOHAMMED, J. A. *et al.* Changes in observed rainfall and temperature extremes in the Upper Blue Nile Basin of Ethiopia. **Weather and Climate Extremes**, v. 37, p. 100468, 2022.
- MOREIRA, J. G. V *et al.* Stationarity in annual daily maximum streamflow series in the upper Juruá River, western Amazon. **Revista Brasileira de Geografia Física**, v. 12, n. 2, p. 705–713, 2019.
- NOURANI, V.; DANANDEH MEHR, A.; AZAD, N. Trend analysis of hydroclimatological variables in Urmia lake basin using hybrid wavelet Mann-Kendall and Sen tests. **Environmental Earth Sciences**, v. 77, n. 207, p. 1–18, 2018.
- PETTITT, A. N. A Non-parametric Approach to the Change-point Problem. **Journal of the**

Royal Statistical Society, v. 28, n. 2, p. 126–135, 1979.

RYBERG, K. R.; HODGKINS, G. A.; DUDLEY, R. W. Change points in annual peak streamflows: Method comparisons and historical change points in the United States. **Journal of Hydrology**, v. 583, p. 124307, 1 abr. 2020.

SEN, P. K. Estimates of the Regression Coefficient Based on Kendall ' s Tau. **Journal of the American Statistical Association**, v. 63, n. 324, p. 1379–1389, 1968.

SILVA, F. B. *et al.* Pluviometric and fluviometric trends in association with future projections in areas of conflict for water use. **Journal of Environmental Management**, v. 271, 1 out. 2020.

SÖNMEZ, A. Y.; KALE, S. Climate change effects on annual streamflow of filyos river (Turkey). **Journal of Water and Climate Change**, v. 11, n. 2, p. 420–433, 2020.

SPEARMAN, C. The American Journal of Psychology. **The American Journal of Psychology**, v. 15, n. 1, p. 72–101, 1904.

SPINONI, J. *et al.* Future global meteorological drought hot spots: A study based on CORDEX data. **Journal of Climate**, v. 33, n. 9, 2020.

THEIL, H. A rank-invariant method of linear and polynomial regression analysis III. **Proceedings Nederlandse Akademie van Wetenschappen**, v. 53, p. 1397–1412, 1950.

TONGAL, H.; BOOIJ, M. J. Simulation and forecasting of streamflows using machine learning models coupled with base flow separation. **Journal of Hydrology**, v. 564, p. 266–282, 1 set. 2018.

ZHOU, Y. *et al.* Short-term flood probability density forecasting using a conceptual hydrological model with machine learning techniques. **Journal of Hydrology**, v. 604, p. 127255, 1 jan. 2022.

CAPÍTULO 1

Análise de tendência e identificação dos possíveis períodos de mudança na ocorrência de eventos extremos de vazão na bacia hidrográfica do rio Muriaé

RESUMO: O conhecimento dos padrões hidrológicos de uma bacia é importante para o gerenciamento ambiental em períodos de escassez e de enchentes. Dentre as principais variáveis que influenciam a vazão de um rio em regiões tropicais está o regime pluviométrico, o qual tem se alterado em decorrência das mudanças climáticas. Aliado aos fatores climáticos, têm-se o uso e cobertura da terra, bem como a demanda de uso da água, que podem intensificar os eventos extremos em uma bacia. O objetivo desse trabalho foi identificar tendências nas séries de vazões mínima e máxima, em escala mensal e anual, na bacia hidrográfica do rio Muriaé, a partir de técnicas não-paramétricas, e analisar os principais fatores climáticos e antrópicos que podem explicar as alterações das vazões na bacia. Foram utilizados os testes de Mann-Kendall, Mann-Kendall Modificado, Correlação de Spearman e Pettitt, em diferentes níveis de significância, e a inclinação de Thiel-Sen para identificar a magnitude das tendências encontradas. Os testes aplicados mostraram que as vazões mínimas e máximas apresentam tendência negativa no período seco da bacia, destacadamente em outubro. As mudanças de tendência nas séries mensais ocorreram, em sua maioria, no ano de 2012. A precipitação apresentou um comportamento semelhante ao observado nas vazões, com mudanças significativas fortes no período seco. O uso e cobertura da terra apresentou baixa variação no período analisado, não sendo considerado um fator determinante nas alterações de vazão. Por outro lado, a demanda hídrica na bacia exerceu uma maior pressão nos cursos d'água diante do aumento das vazões outorgadas. A aplicação de diferentes métodos estatísticos permitiu diagnosticar o comportamento hídrico da bacia, enquanto as análises dos fatores climáticos e antrópicos possibilitaram conhecer as suas possíveis influências nas mudanças de tendência da vazão.

Palavras-chave: Variáveis hidroclimatológicas. Mann-Kendall. Theil-Sen. Período seco. Uso da água.

1.1 INTRODUÇÃO

Os fenômenos naturais, como secas e enchentes, são responsáveis por diversos impactos econômicos, sociais e ambientais (PÉREZ-BLANCO; GIL-GARCÍA; SAIZ-SANTIAGO, 2021; SUGAHARA *et al.*, 2012). A avaliação da tendência das secas e enchentes, e os possíveis efeitos advindos das alterações do uso e ocupação da terra, assim como das mudanças climáticas, têm se tornado uma ferramenta no auxílio à gestão de recursos hídricos nas regiões afetadas com a recorrência desses eventos extremos (CHAULUKA; SINGH; KUMAR, 2021).

A frequência de secas severas e extremas de curto prazo tem sido observada em alguns trabalhos realizados no Brasil (AWANGE; MPELASOKA; GONCALVES, 2016; MARENGO *et al.*, 2015). Embora as alterações sejam complexas, o agravamento dos impactos da estiagem nos últimos anos foi associado à diminuição da precipitação, intensidade e magnitude dos dias secos, elevação da temperatura e, conseqüentemente, o aumento das taxas evaporativas (NAUMANN *et al.*, 2018).

Por outro lado, os eventos de enchentes ocorrem quando a vazão do rio atinge o leito maior, ou seja, acima da capacidade de drenagem, estando associados às chuvas intensas e alterações no uso e ocupação da terra em áreas urbanas e rurais, entre outros fatores (AYEDUN *et al.*, 2018; MACHADO; OLIVEIRA; LOIS-GONZÁLEZ, 2019; TUCCI, 2012). Aliado à vulnerabilidade social de algumas regiões, esse perigo natural gera um elevado custo socioeconômico e ambiental, diante dos danos causados, destacadamente em áreas urbanas (UDDIN; MATIN, 2021).

Diversos estudos têm mostrado mudanças significativas na ocorrência de chuvas de alta intensidade, com possibilidades de aumentar o risco de transbordamento de rios em áreas urbanizadas em diferentes regiões do planeta (SKANSI *et al.*, 2013; SUGAHARA *et al.*, 2012). Na região sudeste do Brasil, as chuvas deixam diversas cidades em alerta devido ao elevado risco de inundação em vista de sistemas de drenagem ineficientes (SOUZA *et al.*, 2021). Esses eventos estão acontecendo em maior intensidade de acordo com as normais climatológicas para o período de 1991 a 2020, com chuvas acima de 80 e 100 mm presentes em um maior número de dias na década de 2011 a 2020 em comparação com a década de 1991 a 2000 (INMET, 2022).

Existem diversos fatores associados à ocorrência de eventos extremos. Um dos principais é a precipitação, visto que sua ocorrência em baixa ou alta intensidade está diretamente relacionada à diminuição, ou ao aumento da vazão dos cursos d'água, respectivamente (FRAGA *et al.*, 2020; KLIMENT; MATOUŠKOVÁ, 2008; SCHLÖGL *et al.*,

2021). As chuvas impactam expressivamente o regime de variação de vazões, o conteúdo de umidade do solo e a recarga dos aquíferos (CHAULUKA; SINGH; KUMAR, 2021).

A dinâmica da vazão em uma bacia hidrográfica também é influenciada pelo uso e a cobertura da terra, com destaque aos processos de urbanização e extensa implementação de cultivos agrícolas (JANIZADEH *et al.*, 2021; SILVA; ALENCAR; ALMEIDA, 2020). A intensificação desses fatores pode modificar significativamente a frequência e magnitude de eventos extremos na bacia (LIU *et al.*, 2017; SALVIANO; PEREIRA FILHO; VEMADO, 2021; XU *et al.*, 2014).

Nos estudos hidrológicos sobre eventos de secas e enchentes, diversos autores utilizam as vazões mínima e máxima para identificar, por meio de diferentes testes estatísticos, tendências positivas ou negativas no regime de variação de vazões da bacia (SILVA *et al.*, 2017; ULIANA *et al.*, 2015). A tendência consiste num componente determinístico de uma série histórica, caracterizada como uma flutuação em escala sazonal ou de longo período da série, identificada com maior frequência por meio de testes não-paramétricos em decorrência da grande variabilidade dos dados ambientais (MALIK *et al.*, 2020).

As alterações na ocorrência dos eventos de secas e de enchentes precisam ser investigadas, a fim de subsidiar as tomadas de decisões quanto aos impactos ambientais e sociais decorrentes de suas ocorrências. A bacia hidrográfica do rio Muriaé, localizada nos Estados de Minas Gerais e Rio de Janeiro, tem sido afetada expressivamente por eventos de seca e enchente, que causaram e causam diversos impactos em ambos os Estados (SALVIANO; PEREIRA FILHO; VEMADO, 2021).

Embora essa bacia tenha sido objeto de um estudo preliminar para o controle de eventos extremos após as grandes enchentes de 1997 (PRADO *et al.*, 2005), deve-se considerar que as mudanças climáticas e as alterações no uso e ocupação da terra nos últimos 20 anos submeteram a bacia a uma nova dinâmica hidrológica. Por exemplo, o estudo de Nauditt *et al.* (2022), aplicado para o período de 2000 a 2015, identificou áreas com vulnerabilidade à seca severa em regiões próximas à foz da bacia, enquanto Gomes *et al.* (2021) identificaram um aumento de vazão num estudo de simulações de vazão entre os anos de 1990 e 2015, considerando diferentes cenários de mudanças climáticas.

Assim, diante dos impactos advindos da ocorrência dos eventos extremos e da necessidade de mais informações sobre o comportamento hidrológico da bacia integrando diferentes fatores de alteração da vazão, esse estudo teve como objetivo identificar tendências nas séries de dados de vazões mínimas e máxima em escala mensal e anual, na bacia

hidrográfica do rio Muriaé, a partir de técnicas não-paramétricas, bem como analisar os principais fatores climáticos e antrópicos que podem explicar as alterações das vazões da bacia.

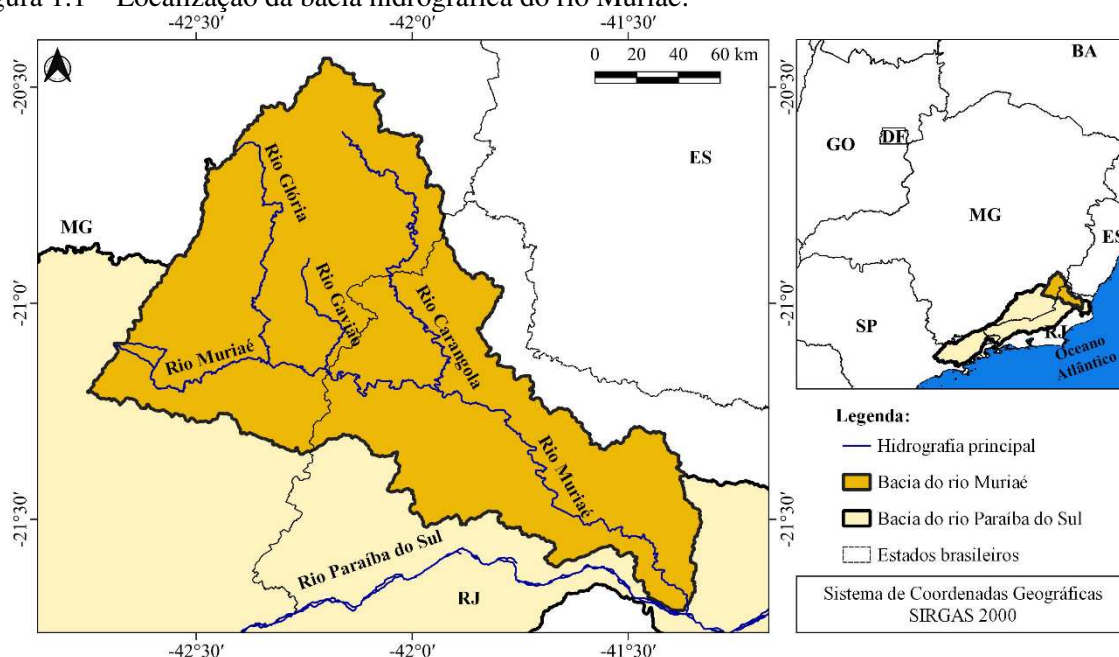
1.2 MATERIAL E MÉTODOS

1.2.1 Área de estudo

O estudo foi realizado na bacia hidrográfica do rio Muriaé (Figura 1.1), situada nos Estados de Minas Gerais e Rio de Janeiro, em razão dos recorrentes eventos extremos que a bacia tem vivenciado ao longo dos últimos anos. Nos anos hidrológicos de 2013/2014 e 2014/2015 ocorreram verões secos na região sudeste (MARENGO *et al.*, 2015), enquanto nos verões de 2016, 2017 e 2018 registraram-se vazões elevadas segundo o Sistema de Alerta Hidrológico da bacia do rio Muriaé (SAH-Muriaé) para as cinco estações fluviométricas presentes na bacia (SALVIANO, 2019).

A bacia do rio Muriaé, que pertence à bacia do rio Paraíba do Sul, possui área de drenagem de aproximadamente 8.126 km². O rio Muriaé é formado pela confluência dos ribeirões Bonsucesso e Samambaia, possuindo uma extensão de 300 km desde sua nascente até a foz, tendo os rios Glória e Carangola como os seus maiores afluentes (ÁVILA *et al.*, 2016). O bioma Mata Atlântica abrange toda a extensão territorial da bacia (IBGE, 2022).

Figura 1.1 – Localização da bacia hidrográfica do rio Muriaé.



Fonte: A autora.

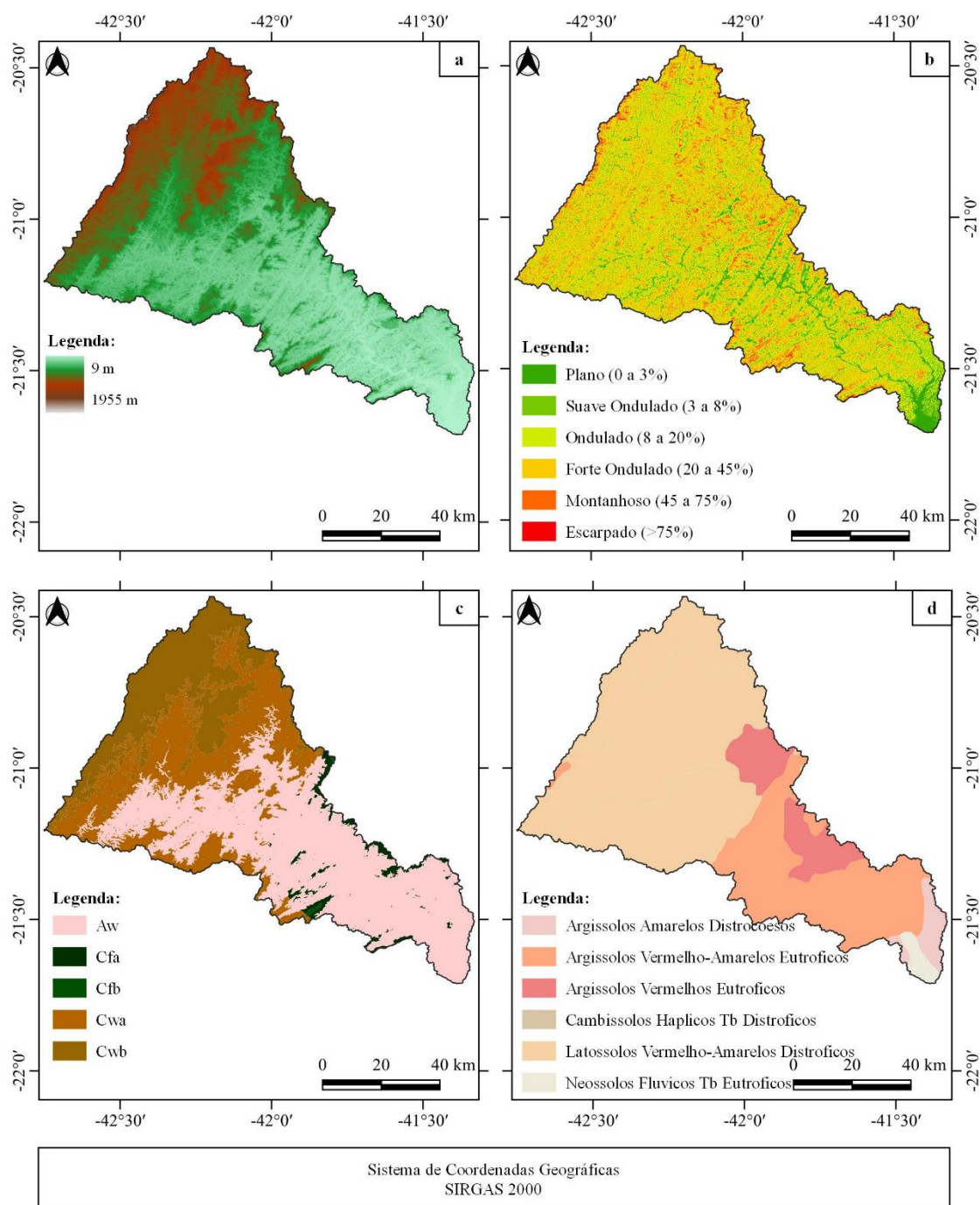
Na área de estudo existem cerca de 440 mil habitantes, situados em 26 municípios, principalmente em Muriaé/MG, Itaperuna/RJ e Carangola/MG (IBGE, 2021). Suas principais atividades econômicas se referem à atividade agropecuária como arroz, café e cana-de-açúcar (ÁVILA *et al.*, 2016).

A bacia possui altitude entre 9 e 1.955 m, com uma média de 435 m (Figura 1.2). A declividade média é cerca de 22%, caracterizando a região como de relevo fortemente ondulado (EMBRAPA, 1979).

Os tipos climáticos (Figura 1.2), segundo Alvares *et al.* (2013), pela classificação de Köppen, variam ao longo da bacia, sendo que na cabeceira predominam os tipos Cwb (Temperado úmido, com inverno seco e verão morno) e Cwa (Temperado úmido com inverno seco e verão quente), enquanto no restante da bacia predomina o clima tipo Aw (Tropical de savana com inverno seco).

Em relação ao tipo de solo (Figura 1.2), seis classes estão presentes na bacia, com maior destaque para os Latossolos Vermelho-Amarelos Distróficos e Argissolos Vermelho-Amarelos Eutróficos (EMBRAPA SOLOS, 2020).

Figura 1.2 – Bacia hidrográfica do rio Muriaé: a) Modelo Digital de Elevação (MDE); b) declividade; c) classificação climática de Köppen; e d) tipo de solo.



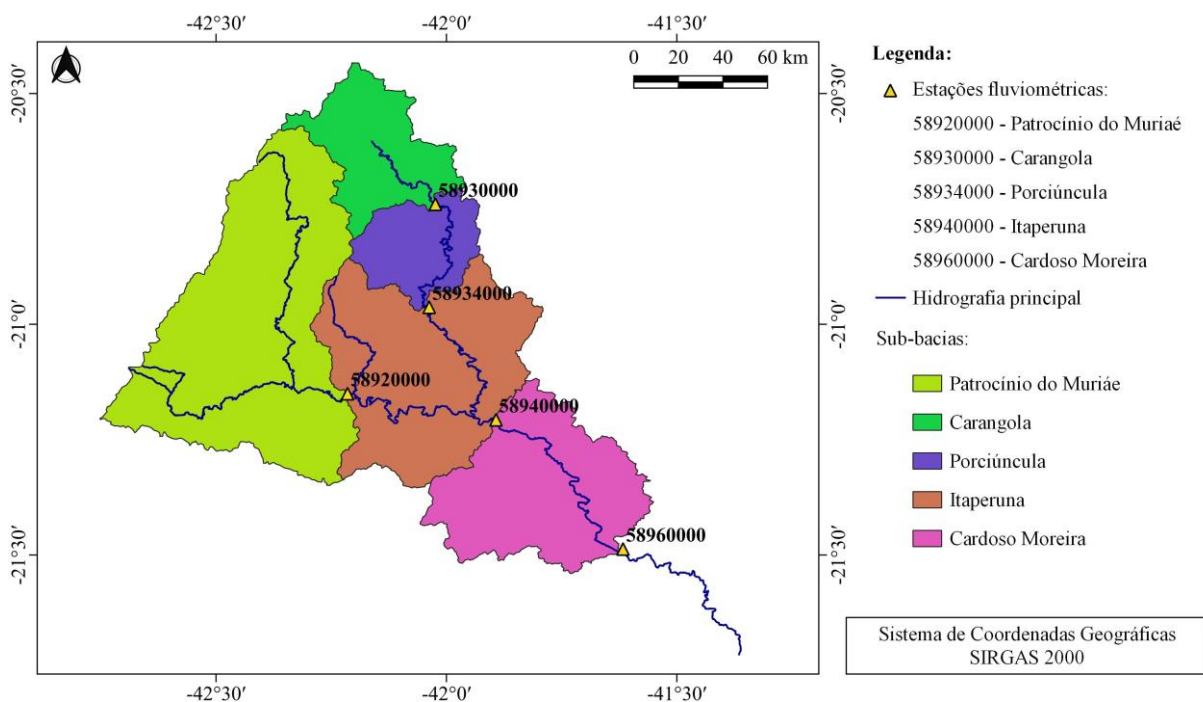
Fonte: A autora.

1.2.2 Obtenção da base de dados

As séries de vazão das estações fluviométricas localizadas na bacia hidrográfica do rio Muriaé (Figura 1.3) foram obtidas no Portal Hidroweb (<https://www.snirh.gov.br/hidroweb/serieshistoricas>), da Agência Nacional de Águas e Saneamento Básico (ANA), sendo que todas as estações pertencem ao Sistema de Alerta

Hidrológico da bacia do rio Muriáé (SAH – Muriáé) (https://www.cprm.gov.br/sace/index_bacias_monitoradas.php?getbacia=bmuriáé#). Para essas estações foi considerado um limite de falhas anuais igual ou superior a 5%.

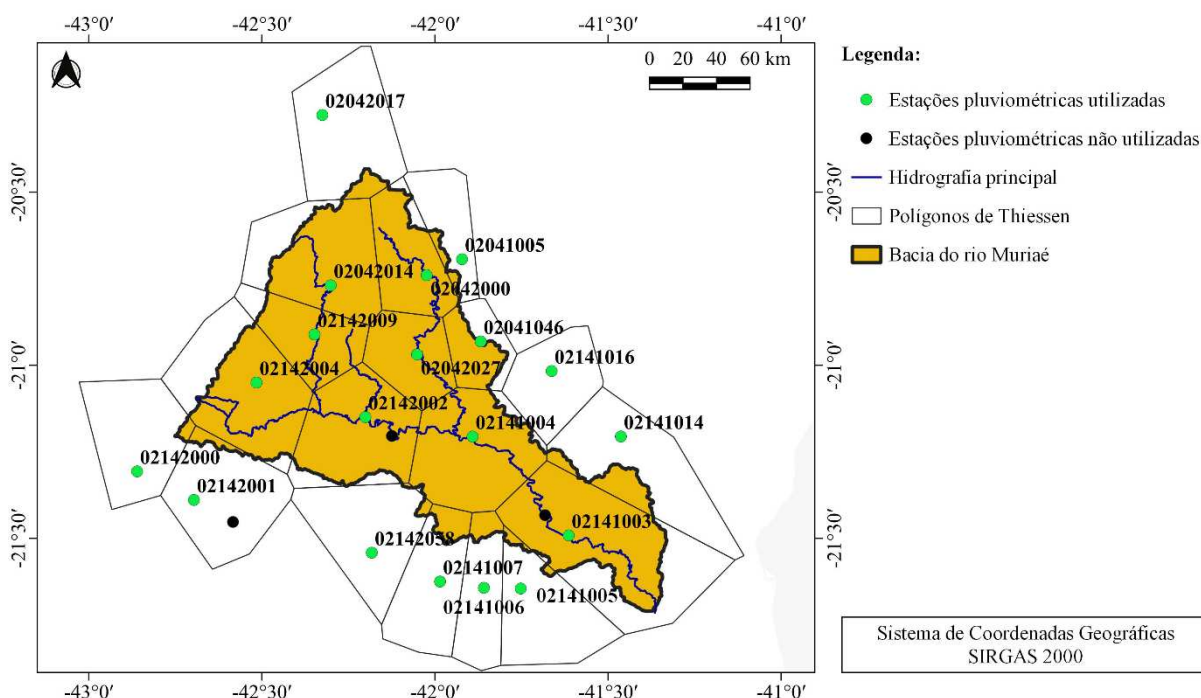
Figura 1.3 – Localização das estações fluviométricas localizadas na bacia do rio Muriáé e suas respectivas áreas de drenagem.



Fonte: A autora.

Os dados das estações pluviométricas também foram obtidos no Portal Hidroweb (<https://www.snirh.gov.br/hidroweb/serieshistoricas>). A seleção das estações utilizadas no estudo foi realizada com base na sua influência na bacia (Figura 1.4), determinada por meio dos polígonos de Thiessen (THIESSEN, 1911) e disponibilidade de dados para um período mínimo de 30 anos.

Figura 1.4 – Localização das estações pluviométricas utilizadas no estudo.



Fonte: A autora.

Diante dos critérios adotados, foram selecionadas cinco estações fluviométricas (Tabela 1.1) e 19 estações pluviométricas (Tabela 1.2) para a análise de tendência e identificação do ano/mês de mudança do comportamento hidrológico na bacia hidrográfica do rio Muriaé em caso de não estacionariedade. O período de análise dos dados foi de 49 anos, iniciando em 1971 e finalizando em 2019. O intervalo de tempo foi definido com base na disponibilidade de dados de cada estação fluviométrica e pluviométrica. A estação pluviométrica 02142009 foi a responsável por limitar o início do período-base, visto que suas medições começaram no ano de 1971, enquanto a estação fluviométrica 58940000 definiu o ano de 2019 como sendo o último ano da série.

Considerando que os dados de vazão disponibilizados pelo Hidroweb apresentam consistência até o ano de 2014 e os dados de precipitação até 2005, realizou-se uma análise prévia dos dados não consistidos para identificação de erros e/ou valores discrepantes (*outliers*). Tal análise foi fundamental para permitir a utilização de séries históricas de dados mais recentes, até o ano de 2019.

Tabela 1.1 – Estações fluviométricas selecionadas no estudo.

Código	Nome	Latitude (°)	Longitude (°)	Área de drenagem (km²)	Período de observação
58920000	Patrocínio do Muriaé	-21,1506	-42,2144	2.660	1928 – 2021
58930000	Carangola	-20,74	-42,0239	773	1935 – 2021
58934000	Porciúncula	-20,9633	-42,0372	1.340	1928 – 2020
58940000	Itaperuna	-21,2072	-41,8922	5.800	1934 – 2020
58960000	Cardoso Moreira	-21,4872	-41,6167	7.210	1928 – 2020

Fonte: A autora.

Tabela 1.2 – Estações pluviométricas selecionadas no estudo.

Código	Nome	Latitude (°)	Longitude (°)	Altitude (m)	Período de observação
02041005	Caiana	-20,6939	-41,9219	747	1939 – 2021
02041046	Varre - Sai	-20,9314	-41,8678	688	1967 – 2021
02042000	Carangola	-20,74	-42,0239	399	1935 – 2021
02042014	Bicuiba	-20,7694	-42,3011	630	1962 – 2020
02042017	Matipó	-20,2772	-42,3256	595	1967 – 2021
02042027	Porciúncula	-20,9692	-42,0517	188	1939 – 2020
02141003	Cardoso Moreira	-21,4872	-41,6136	29	1939 – 2021
02141004	Itaperuna	-21,2064	-41,8911	110	1940 – 2020
02141005	São Fidelis	-21,6453	-41,7522	10	1939 – 2021
02141006	Dois Rios	-21,6433	-41,8586	50	1939 – 2021
02141007	Três Irmãos	-21,6253	-41,9853	42	1943 – 2021
02141014	Ponte do Itabapoana	-21,2061	-41,4628	59	1937 – 2021
02141016	São José do Calçado	-21,0169	-41,6628	150	1952 – 2021
02142000	Astolfo Dutra	-21,3069	-42,8606	231	1939 – 2021
02142001	Cataguases	-21,3894	-42,6964	182	1939 – 2021
02142002	Patrocínio do Muriaé	-21,15	-42,2011	177	1936 – 2020
02142004	Fazenda Umbaúbas	-21,0503	-42,5156	490	1943 – 2021
02142009	Jussara	-20,9114	-42,3481	419	1971 – 2020
02142058	Santo Antônio de Pádua	-21,5419	-42,1825	70	1966 – 2021

Fonte: A autora.

Além dos dados de vazão e precipitação, também foram obtidos dados do uso e cobertura da terra e das concessões de outorgas da bacia. Para o uso e cobertura da terra realizou-se o *download* de mapas da Coleção 6.0 do Projeto MapBiomas (MAPBIOMAS, 2021), utilizando-se o Google Earth Engine (GEE), para os anos compreendidos entre 1985 e 2021.

Quanto aos dados de outorgas, fez-se o *download* de dados referentes ao período de 2002 a 2019, extraídos de diferentes órgãos gestores de recursos hídricos que atuam na bacia

como: Instituto Mineiro de Gestão das Águas (IGAM), Instituto Estadual do Ambiente (INEA) e Agência Nacional de Águas e Saneamento Básico (ANA).

1.2.3 Variáveis hidroclimatológicas

As vazões mínimas e máximas das estações fluviométricas localizadas na bacia hidrográfica do rio Muriaé foram definidas considerando o ano hidrológico com início em novembro e final em outubro, sendo o período chuvoso entre novembro e abril e o período seco entre maio e outubro. A definição do ano hidrológico teve como critério a ocorrência da vazão mínima de sete dias consecutivos (Q_7) para a série histórica de 1971 a 2019, em que foi observada uma mudança brusca nos meses de outubro e novembro (APÊNDICE A).

Para a análise de tendência da vazão foram definidas quatro variáveis em escalas mensal e anual, as quais estão apresentadas na Tabela 1.3. As séries em escala mensal foram compostas por cada um dos 12 meses do ano para o período de jan/1971 a dez/2019, enquanto as séries anuais compreenderam cada ano hidrológico (nov/1971 a out/2019). As séries históricas foram analisadas no *software* SisCAH 1.0 (<http://www.gprh.ufv.br/software>). Ressalta-se que não houve preenchimento de falhas para os dados de vazão, em vista de não ter sido encontrada nenhuma falha no período analisado.

Tabela 1.3 – Variáveis utilizadas na análise de tendência.

Tipo	Sigla	Variável	Unidade
Vazão	QmaxMensal	Vazão Máxima Diária Mensal	m ³ /s
	QmaxAnual	Vazão Máxima Diária Anual	m ³ /s
	Q ₇ Mensal	Vazão Mínima Mensal de 7 dias consecutivos	m ³ /s
	Q ₇ Anual	Vazão Mínima Anual de 7 dias consecutivos	m ³ /s
Precipitação	PTM	Precipitação Total Mensal	mm
	PTA	Precipitação Total Anual	mm
	PTSC	Precipitação Total do Semestre Chuvoso	mm
	PTSS	Precipitação Total do Semestre Seco	mm

Fonte: A autora.

Na análise de tendência da precipitação, primeiramente os dados pluviométricos tiveram as falhas mensais preenchidas utilizando o *software* R 4.0.4 (R CORE TEAM, 2020), a partir do método de preenchimento de regressão linear simples, desde que o coeficiente de determinação com a estação de apoio estivesse no intervalo de 0,7 a 1, indicando uma forte

relação linear (APÊNDICE B) (CALDERA; PIYATHISSE; NANDALAL, 2016; CECÍLIO *et al.*, 2021; RATNER, 2009).

A próxima etapa consistiu na determinação das séries de precipitação em escala mensal, semestral (período chuvoso e seco) e anual (Tabela 1.3). A variável mensal teve um período de 49 anos, enquanto as séries semestrais e anuais tiveram um período de 48 anos, visto que adotaram como referência o ano hidrológico encontrado na bacia hidrográfica.

1.2.4 Tendência e identificação de meses/anos de mudança no comportamento das variáveis hidroclimatológicas

As variáveis hidroclimatológicas foram submetidas aos seguintes testes estatísticos para a identificação de tendências e identificação do mês/ano da mudança de comportamento: Mann-Kendall (MK); Mann-Kendall Modificado (*Modified Mann-Kendall*, MMK) quando os princípios do método de independência e aleatoriedade não foram atendidos; Correlação de Spearman (SP); inclinação de Theil-Sen (*Theil Sen's Slope*, TSS) e teste de Pettitt (*Pettitt Test's*, PT), para identificação do mês/ano de mudança.

Nos testes de MK, MMK, PT e SP adotaram-se três níveis de significância (α), 10%, 5% e 1%, que foram considerados, respectivamente, como sinais de tendência fraca, média e forte, em consonância com trabalhos semelhantes realizados por Bhatti *et al.* (2020), Mahmood e Jia (2019) e Malik e Kumar (2020).

Os testes não paramétricos foram executados no *software* R, versão 4.0.4 (R CORE TEAM, 2020). Para os testes de Mann-Kendall e de Mann-Kendall Modificado foram utilizados, respectivamente, os pacotes *kendall* (MCLEOD, 2022) e *modifiedmk* (PATAKAMURI; O'BRIEN, 2020). O coeficiente de correlação de Spearman foi obtido por meio do pacote *stat* (BOLAR, 2019), enquanto a definição do mês/ano de mudança foi realizada pelo teste de Pettitt por meio do pacote *trend* (POHLERT, 2020).

1.2.4.1 Teste de Mann-Kendall (MK) e Mann-Kendall Modificado (MMK)

Na avaliação da estacionariedade das séries históricas com base no teste de Mann-Kendall (MK) foi considerado que os dados eram independentes e aleatórios, pois uma autocorrelação positiva poderia levar à rejeição da hipótese nula, ou então, levar a uma interpretação equivocada dos resultados (FANG *et al.*, 2018; NALLEY; ADAMOWSKI; KHALIL, 2012). Portanto, na análise da independência das séries temporais foi realizado um teste de correlação serial lag-1 (ρ_1), utilizando-se a Equação 1.1 (JÚNIOR *et al.*, 2020; PATAKAMURI; MUTHIAH; SRIDHAR, 2020):

$$\rho_1 = \frac{\sum_{t=1}^{n-1} (x_t - \bar{x})(x_{t+1} - \bar{x})}{\sum_{t=1}^n (x_t - \bar{x})^2} \quad (1.1)$$

em que:

- ρ_1 = coeficiente de correlação entre observações defasadas de um período;
- $\bar{x}$ = média da amostra;
- σ = desvio padrão da variável preditora;
- n = tamanho amostral, adimensional;
- $n - 1$ = primeiro termo; e
- $t + 1$ = último termo.

A significância da autocorrelação lag-1, para a hipótese nula em que se segue a distribuição de t de Student com $(n - 2)$ graus de liberdade, foi determinada por meio da Equação 1.2:

$$t = |\rho_1| \sqrt{\frac{n - 2}{1 - \rho_1^2}} \quad (1.2)$$

em que:

- t = distribuição t de Student com $(n - 2)$ graus de liberdade;
- ρ_1 = coeficiente de correlação entre observações defasadas em um período; e
- n = tamanho da amostra.

Na análise da aleatoriedade dos dados foi aplicado o teste de Run (THOM, 1966), a um nível de 5% de significância. Esse teste analisa a sequência de eventos em relação à mediana dos dados observados, em que cada elemento recebe um valor positivo n_+ ou negativo n_- para uma série de tamanho $n = n_+ + n_-$, com as execuções denominadas de R e estatística calculada pela Equação 1.3 (ADELOYE; MONTASERI, 2002; MACCIOTTA *et al.*, 2017).

$$z = \frac{R - \left(\frac{2n_+n_-}{n_+n_-} + 1\right)}{\sqrt{\frac{2n_+n_-(2n_+n_- - n_+ - n_-)}{(n_+n_-)^2(n_+n_- - 1)}}} \quad (1.3)$$

Para uma série temporal independente e aleatória foi utilizado o teste de MK, que pressupõe que na hipótese nula os dados $(x_1, x_2, \dots, x_n)$ constituem uma amostra de n variáveis

aleatórias com distribuição uniforme, sendo a sua matemática determinada pela Equação 1.4 e o sinal pela Equação 1.5 (ISHAK *et al.*, 2013).

$$S = \sum_{i=1}^{n-1} \sum_{j=i+1}^n \text{sgn}(x_j - x_i) \quad (1.4)$$

$$\text{sgn}(X_j - X_i) = \begin{cases} +1 & \text{se } x_j - x_i > 0 \\ 0 & \text{se } x_j - x_i = 0 \\ -1 & \text{se } x_j - x_i < 0 \end{cases} \quad (1.5)$$

em que:

- S = estatística do teste de MK;
- x = série temporal univariada;
- i e j = índices de tempo relacionados a valores individuais;
- n = tamanho da série temporal, adimensional; e
- sgn = sinal da operação $(x_j - x_i)$, adimensional.

Para um valor n superior a 10, admitiu-se que a estatística de Kendall tivesse uma distribuição normal, de modo que para a hipótese nula não existe tendência e sua variância foi calculada pela Equação 1.6 (GÜÇLÜ, 2020).

$$V(S) = \frac{n(n-1)(2n+5)}{18} \quad (1.6)$$

em que n é o tamanho da série temporal, adimensional.

Para identificar a tendência monotônica da série temporal foi calculada a estatística Z padronizada, a partir das classificações geradas na série expressa pela Equação 1.7 (GÜÇLÜ, 2020).

$$Z_{MK} = \begin{cases} \frac{S-1}{\sqrt{V(S)}} & \text{se } S > 0 \\ 0 & \text{se } S = 0 \\ \frac{S+1}{\sqrt{V(S)}} & \text{se } S < 0 \end{cases} \quad (1.7)$$

em que:

- Z_{MK} = estatística Z padronizada; e
- $V(S)$ = variância, $(\text{m}^3/\text{s})^2$ (vazão) e $(\text{mm})^2$ (precipitação).

No caso em que os dados não seguiram os requisitos básicos do teste de MK, foi aplicado o teste de Mann-Kendall Modificado (MMK), no qual há uma modificação da variância expressa pela Equação 1.8, a partir da multiplicação de fatores de correção (Equação 1.9) (HAMED; RAO, 1998).

$$V \times (S) = \frac{n(n-1)(2n+5)}{18} \times \frac{n}{n_s^*} \quad (1.8)$$

$$\frac{n}{n_s^*} = 1 + \frac{2}{n(n-1)(n-2)} \times \sum_{i=1}^{n-1} (n-i)(n-i-1)(n-i-2)\rho_k \quad (1.9)$$

em que:

- $V \times (S)$ = variância modificada, $(\text{m}^3/\text{s})^2$ (vazão) e $(\text{mm})^2$ (precipitação);
- $\frac{n}{n_s^*}$ = fator de correção, adimensional;
- n = número de observações, adimensional;
- n_s^* = número efetivo de observações para explicar a correlação nos dados, adimensional; e
- ρ_k = coeficiente de correlação entre observações defasadas de k períodos.

Após o cálculo da variância modificada $V \times (S)$, calculou-se a estatística de Z_{MMK} pela Equação 1.10.

$$Z_{MMK} = \begin{cases} \frac{S-1}{V \times (S)} & \text{se } S > 0 \\ 0 & \text{se } S = 0 \\ \frac{S+1}{V \times (S)} & \text{se } S < 0 \end{cases} \quad (1.10)$$

1.2.4.2 Teste de Correlação de Spearman

A correlação de Spearman (SP) foi aplicada para complementar a análise dos testes não-paramétricos anteriores, que em alguns casos não fornecem respostas satisfatórias nas análises de tendência (AHMAD *et al.*, 2015; SWAIN *et al.*, 2022). Este teste baseia-se na ordem de classificação, em que, para um dado conjunto de amostras $X(x_1, x_2, \dots, x_n)$, considera-se que na hipótese nula todo X_i seja independente e igualmente distribuído, enquanto para a hipótese alternativa existe tendência, ou seja, a variável X_i aumenta ou diminui com i (YUE *et al.*, 2002).

O coeficiente de correlação foi calculado pela Equação 1.11, sendo a estatística com a hipótese nula de que a correlação de Spearman apresentasse uma distribuição normal, em que

a média e a variância foram determinadas pelas Equações 1.12 e 1.13, respectivamente (SNEYERS, 1991).

$$\rho = 1 - \frac{6 \sum_{i=1}^n [R(X_i) - i]^2}{n(n^2 - 1)} \quad (1.11)$$

$$E(\rho) = 0 \quad (1.12)$$

$$V(\rho) = \frac{1}{n-1} \quad (1.13)$$

em que:

- ρ = coeficiente de correlação de Spearman;
- $R(X_i)$ = i -ésima classificação de X ;
- X_i = valores da série, m³/s (vazão) e mm (precipitação);
- n = número de valores da amostra, adimensional;
- $E(\rho)$ = média da estatística ρ ; e
- $V(\rho)$ = variância da estatística ρ .

A significância do teste foi definida por uma função de distribuição cumulativa normal, de modo que a padronização foi calculada pela Equação 1.14 (YUE *et al.*, 2002).

$$Z_{SP} = \frac{\rho}{\sqrt{V(\rho)}} \quad (1.14)$$

em que Z_{SP} é a estatística Z padronizada.

1.2.4.3 Teste de inclinação Theil-Sen

O teste de inclinação Theil-Sen consiste em um teste não-paramétrico utilizado para estimar a inclinação da tendência em uma amostra de N pares de dados, sendo calculada pela Equação 1.15 (SEN, 1968; THEIL, 1950).

$$TSS_{ij} = \text{mediana} \left(\frac{x_j - x_i}{j - i} \right) \text{ para } i = 1, \dots, n, \text{ e } j > i \quad (1.15)$$

em que:

- x_i e x_j = dados nos tempos i e j .

A magnitude da tendência, crescente, decrescente ou inexistente, foi determinada pelo valor positivo, negativo ou zero, respectivamente, de TSS_{ij} (MALIK; KUMAR, 2020). Para n

valores nas séries históricas, o número de pares estimados no teste foi calculado pelas Equações 1.16 e 1.17 (MALIK *et al.*, 2020).

$$TSS_{mediana} = TSS_{\left(\frac{n+1}{2}\right)} \text{ se } n \text{ for ímpar} \quad (1.16)$$

$$TSS_{mediana} = \frac{1}{2} \left[TSS_{\frac{n}{2}} + TSS_{\frac{(n+2)}{2}} \right] \text{ se } n \text{ for par} \quad (1.17)$$

1.2.4.4 Teste de Pettitt

O teste de Pettitt foi aplicado para detectar o mês/ano de mudança nas variáveis analisadas a um nível de significância de 10%, 5% e 1% e confirmar as análises de tendência dos testes não-paramétricos anteriores. Neste teste, que utiliza uma versão estatística de Mann-Whitney (U_t), cada série temporal $X(x_1, x_2, \dots, x_n)$ foi dividida em duas amostras $x_1, x_2, \dots, x_t$ e $x_{t+1}, x_{t+2}, \dots, x_{t+n}$, em que a estatística e o sinal foram calculados por meio das Equações 1.18 e 1.19 (ZHAO *et al.*, 2014).

$$U_{t,n} = \sum_{i=1}^t \sum_{j=1}^n \text{sgn}(x_t - x_j) \text{ se } t = 2, \dots, n \quad (1.18)$$

$$\text{sgn} = (x) \begin{cases} 1 & \text{se } x_t - x_j > 0 \\ 0 & \text{se } x_t - x_j = 0 \\ -1 & \text{se } x_t - x_j < 0 \end{cases} \quad (1.19)$$

em que:

- $U_{t,n}$ = estatística do teste de Pettitt, adimensional;
- n = tamanho da série temporal, adimensional;
- sgn = função do sinal, $x_t - x_j$, adimensional;
- t e j = posições dos valores na série temporal, adimensional; e
- x_t e x_j = valores na série, m³/s (vazão) e mm (precipitação).

O mês/ano de mudança foi definido como sendo onde o valor de $|U_{t,n}|$ alcança o seu máximo, determinado pela Equação 1.20 (ZHAO *et al.*, 2014). A hipótese alternativa (H_a), ou seja, de que a série apresenta o mês/ano de mudança, foi aceita para os níveis de significância (α) igual a 10%, 5% e 1%, com o nível de confiança estimado pela Equação 1.21 (ULIANA *et al.*, 2015).

$$K_n = \text{Max } |U_{t,n}| \text{ para } 1 \leq t \leq n \quad (1.20)$$

$$P = 2e^{\left(\frac{-6(K_n)^2}{n^3+n^2}\right)} \quad (1.21)$$

em que:

- K_n = valor crítico, adimensional;
 P = nível de confiança, adimensional; e
 n = número de anos da série histórica, adimensional.

1.2.5 Uso e cobertura da terra

Para a análise das mudanças no uso e cobertura da terra na bacia do rio Muriaé foram selecionadas imagens da Coleção 6.0 do Projeto MapBiomias (MAPBIOMAS, 2022). Os anos analisados foram determinados com base na disponibilidade das imagens, nos resultados do teste de Pettitt e no intervalo temporal da análise de tendência. Dessa forma, o primeiro ano para a análise do uso e cobertura da terra correspondeu a 1985, início da série histórica do projeto MapBiomias, seguido pelos anos de 1992 e 2012, anos em que ocorreram o maior número de mudanças pelo teste de Pettitt na análise de mês/ano de mudança nos dados de vazão e por fim, o ano de 2019, último ano da série histórica hidroclimatológica analisada.

As imagens do MapBiomias foram reclassificadas no *software* Idrisi 17.02 considerando sete classes (Tabela 1.4), com o objetivo de sintetizar determinadas classes do MapBiomias e destacar outras de relevância na dinâmica da área analisada. Após a reclassificação, as imagens foram submetidas a uma análise de transição por meio do módulo *Land Change Modeler* (LCM), disponível no Idrisi 17.02, de modo a considerar três períodos: 1985 a 1992, 1992 a 2012 e 2012 a 2019. O LCM consiste em um conjunto de ferramentas com a finalidade de explorar as alterações de uso e cobertura da terra (YIRSAW *et al.*, 2017), fornecendo resultados de transição, como perdas e ganhos das classes ocorridas em cada período analisado.

Tabela 1.4 – Classes de uso e cobertura da terra adotadas no estudo.

Legenda MapBiomias 6	Legenda adotada
Floresta	Floresta (FL)
Formação Natural não Florestal	Formação Natural não Florestal (FN)
Pastagem	Pastagem (P)
Agricultura	Agricultura (AG)
Silvicultura	Floresta Plantada (FP)
Área Urbanizada	Área não Vegetada (AnV)
Mineração	Área não Vegetada (AnV)
Outras Áreas não Vegetadas	Área não Vegetada (AnV)
Corpo D'água	Corpos Hídricos (CH)

Fonte: A autora.

1.2.6 Outorgas de uso da água

Os dados de outorgas de captação superficial, inicialmente, passaram por uma análise de consistência, no qual os dados imprecisos foram verificados por meio das informações contidas nas portarias emitidas pelos órgãos responsáveis (IGAM, INEA e a ANA), sendo descartados caso não houvesse nenhuma referência a esses dados. Posteriormente, as outorgas foram classificadas em oito classes, de modo a representar as distintas finalidades da captação de água na bacia (Tabela 1.5). Após essa etapa, calculou-se o percentual de vazões outorgadas para cada finalidade entre 2000 e 2019, considerando a disponibilidade dos dados e validade de cada concessão.

A fim de identificar a demanda hídrica na bacia e comparar o total de vazões outorgadas com as vazões passíveis de outorga em cada área de drenagem das estações fluviométricas, analisou-se a série histórica de vazões outorgadas nos rios de Minas Gerais e Rio de Janeiro (50% da vazão mínima anual de 7 dias consecutivos e período de retorno de 10 anos – $Q_{7,10}$ anual) e em rios de domínio da União (70% da vazão mínima anual associada a permanência de 95% do tempo – Q_{95} anual). A vazão $Q_{7,10}$ anual foi calculada pela função de densidade de probabilidade Lognormal 3 no *software* SisCAH, enquanto a Q_{95} anual foi calculada por meio do pacote *hidrowebr* (INOMATA, 2019), de modo a relacionar a vazão e a porcentagem do tempo em que a vazão é superada ou igualada (EDUARDO; TUCCI, 2008) (APÊNDICE A).

Tabela 1.5 – Classes das finalidades de uso da água na bacia do rio Muriaé

ID	Classe	Descrição
1	Abastecimento público	Captação de água para o abastecimento público
2	Aquicultura	Captação de água para tanques de aquicultura
3	Consumo humano	Captação de água para o consumo humano
4	Indústria	Captação de água para consumo industrial e agroindustrial
5	Irrigação	Captação de água para áreas de agricultura
6	Mineração	Captação de água para extração mineral, extração de areia/cascalho em leito de rio e contenção de rejeitos
7	Outros	Captação de água para fins não informados, lavagem de carro, paisagismo, dessedentação de animais, umectação de vias e outros usos.
8	Usos conjugados	Captação para mais de uma finalidade de uso

Fonte: A autora.

Na manipulação dos dados de outorgas foram utilizados os seguintes pacotes do *software* R (R CORE TEAM, 2020): *tidyverse* (WICKHAM, 2021), *dplyr* (WICKHAM *et al.*, 2022), *ggplot2* (WICKHAM *et al.*, 2022), *gridExtra* (AUGUIE; ANTONOV, 2017) e *ggpubr* (KASSAMBARA, 2021).

1.3 RESULTADOS E DISCUSSÃO

1.3.1 Análise de tendência das séries de vazão

A estatística das análises de tendência realizadas por meio dos testes estatísticos de MK, MMK e SP, considerando os níveis de significância de 10%, 5% e 1%, bem como a magnitude da tendência pela inclinação de TSS para as séries mensais e anuais de vazão, encontram-se relatadas nas Tabelas 1.6 e 1.7, respectivamente para vazões mínima e máxima, e representadas nas Figuras 1.5 e 1.6.

Para todas as variáveis de vazão (Q_7 e $Q_{\text{máxima}}$), as tendências significativas encontradas foram negativas. Na análise da vazão mínima Q_7 mensal (Tabela 1.6, Figuras 1.5 e 1.6), as tendências fortes ($\alpha = 0,01$) estiveram concentradas no período seco, especificamente na estação Carangola (58930000), em que o teste de MMK indicou tendência negativa forte em três meses consecutivos: agosto ($Z_{\text{MMK}} = -2,677$), setembro ($Z_{\text{MMK}} = -2,646$) e outubro ($Z_{\text{MMK}} = -2,353$). As tendências ao nível de significância de 10% e 5% para a Q_7 mensal também foram registradas, em sua maioria, para a estação seca, destacando-se os meses de julho com quatro tendências negativas fracas ($\alpha = 0,10$) e outubro com sete tendências negativas médias ($\alpha = 0,05$) considerando todos os testes de tendência. Na análise da Q_7 anual, encontrou-se tendência a um nível de 10% de significância ($Z_{\text{SP}} = -1,805$) na estação Patrocínio do Muriaé (58920000), enquanto a estação Carangola (58930000) apresentou tendência, com significâncias de 5% pelo teste de MMK ($Z_{\text{MMK}} = -2,318$) e 1% pelo teste de SP ($Z_{\text{SP}} = -2,904$).

Ainda para a vazão Q_7 , os valores da inclinação de Theil-Sen (Tabela 1.6) mostraram tendências negativas em escala mensal e anual em todas as estações que apresentaram tendências significativas nos níveis de significância de 10%, 5% e 1%. A tendência na escala mensal variou, por ano, de -0,097 a -0,525 m³/s na estação Patrocínio do Muriaé (58920000); de -0,025 a -0,102 m³/s na estação Carangola (58930000); e de -0,142 a -0,300 m³/s na estação Itaperuna (58940000). Na escala anual, encontrou-se na estação Carangola uma redução de 0,037 m³/s por ano, significativa a 5% pelo teste de MK e a 1% pelo teste de SP.

No estudo realizado na bacia do rio Itapemirim, ao norte da bacia em estudo, pertencente também ao bioma Mata Atlântica, Cecílio *et al.* (2021) verificaram redução significativa ao nível de 5% da vazão Q_7 anual em sete das doze estações fluviométricas analisadas. Isso reforça a situação preocupante quanto à disponibilidade hídrica na região sudeste, uma vez que o mesmo comportamento foi observado em duas bacias próximas, Muriaé e Itapemirim.

Quanto às séries de vazão máxima (Tabela 1.7 e Figuras 1.5 e 1.6), os testes estatísticos identificaram tendência negativa significativa entre os meses de julho e outubro, destacando-se

o mês de julho com o maior número de tendências a 5% e o mês de outubro com tendências para o nível de significância de 1%. Na escala anual, nenhum teste identificou tendência significativa, sendo que essa escala tem uma maior importância em eventos extremos máximos do que a escala mensal. Esses resultados são semelhantes aos encontrados por Ferreira *et al.* (2021), em que o teste de MK, para um intervalo de confiança de 95%, encontraram tendências decrescentes nas séries mensais de vazão máxima para a bacia do rio Santo Antônio, que apresenta características climáticas semelhantes à bacia do rio Muriaé.

Na análise pela inclinação de TSS, observou-se para as vazões máximas da bacia (Tabela 1.7) tendências negativas em todos os níveis de significância considerados, com a estação de Patrocínio do Muriaé (58920000) apresentando variação de -0,252 (julho) a -0,770 m³/s (outubro) por ano. Na Tabela 1.7 também foi possível visualizar que os maiores intervalos estiveram nas estações Itaperuna (58940000) e Cardoso Moreira (58960000), com variações de -0,573 a -1,573 m³/s e -0,564 a -1,638 m³/s, respectivamente. Na análise anual, nenhuma inclinação foi significativa para essa variável.

A redução da magnitude das vazões máximas, principalmente no período seco, reforça que a bacia está com a sua disponibilidade comprometida, de modo a apresentar períodos de escassez hídrica mais severos e impactar diversas atividades desenvolvidas na bacia, desde a agricultura ao abastecimento para o consumo humano. Essa situação se torna mais crítica devido à falta de planejamento na gestão dos recursos hídricos pelos setores que mais consomem água, agravando os conflitos existentes na bacia (SIQUEIRA *et al.*, 2021).

Tabela 1.6 – Estatística dos testes MK, MMK, TSS e SP para a análise de tendência das séries mensais e anuais de Q₇ das estações fluviométricas localizadas na bacia do rio Muriaé.

Estação	Parâmetro	Jan	Fev	Mar	Abr	Mai	Jun	Jul	Ago	Set	Out	Nov	Dez	Anual
58920000	Z _{MK}	-1,784*	-1,319	0,026	-1,526	-1,767*				-1,957*	-2,405**	-1,043	-0,646	
	Z _{MMK}						-1,382	-1,339	-1,178					-1,546
	Z _{SP}	-2,001**	-1,393	-0,136	-1,480	-1,863*	-1,583	-1,716*	-1,379	-1,893*	-2,396**	-1,169	-0,755	-1,805*
	TSS	-0,525	-0,240	0,014	-0,253	-0,172	-0,117	-0,097	-0,074	-0,122	-0,158	-0,176	-0,157	-0,074
58930000	Z _{MK}	-1,612	-1,284	0,060	-0,319	-0,595	-0,974	-1,638				-2,026	-0,595	
	Z _{MMK}								-	-	-3,543***			-2,318**
	Z _{SP}	-1,753*	-1,417	0,019	-0,499	-0,664	-1,058	-1,697*	2,677***	2,646***		-2,071**	-0,713	-
	TSS	-0,102	-0,072	0,005	-0,019	-0,015	-0,020	-0,031	-2,133**	2,638***	3,427***	-0,069	-0,047	2,904***
58934000	Z _{MK}	-1,388	-1,284	0,491	-0,353	0,060	0,112	-0,310	0,129	-0,543	-2,508**	-1,526	-0,336	
	Z _{MMK}													-0,870
	Z _{SP}	-1,587	-1,359	0,395	-0,375	-0,011	0,069	-0,395	-0,023	-0,791	-2,532**	-1,515	-0,305	-1,348
	TSS	-0,162	-0,128	0,053	-0,049	0,001	0,004	-0,006	0,001	-0,016	-0,063	-0,081	-0,040	-0,020
58940000	Z _{MK}	-1,405	-1,060	0,526						-1,405	-2,422**	-0,922		
	Z _{MMK}				-0,928	-0,893	-2,130**	-1,896*	-1,138				-0,441	-1,039
	Z _{SP}	-1,604	-1,097	0,452	-0,729	-0,925	-0,786	-1,680*	-1,088	-1,541	-2,477**	-0,960	-0,450	-0,847
	TSS	-0,797	-0,364	0,196	-0,256	-0,169	-0,142	-0,226	-0,118	-0,144	-0,300	-0,234	-0,203	-0,061
58960000	Z _{MK}	-1,181	-1,009	0,646	-0,474						-2,560**	-1,336	-0,353	
	Z _{MMK}					-0,678	-1,025	-0,648	-0,785	-1,480				-1,287
	Z _{SP}	-1,394	-1,150	0,502	-0,486	-0,846	-0,806	-0,838	-0,873	-1,464	-	-1,308	-0,385	-1,587
	TSS	-0,733	-0,338	0,258	-0,202	-0,154	-0,114	-0,112	-0,086	-0,168	2,556***	-0,396	-0,223	-0,165

MK – Mann-Kendall; MMK – Mann-Kendall Modificado; SP – correlação de Spearman; *Nível de significância de 10%; **Nível de significância de 5%;

***Nível de significância de 1%. Fonte: A autora.

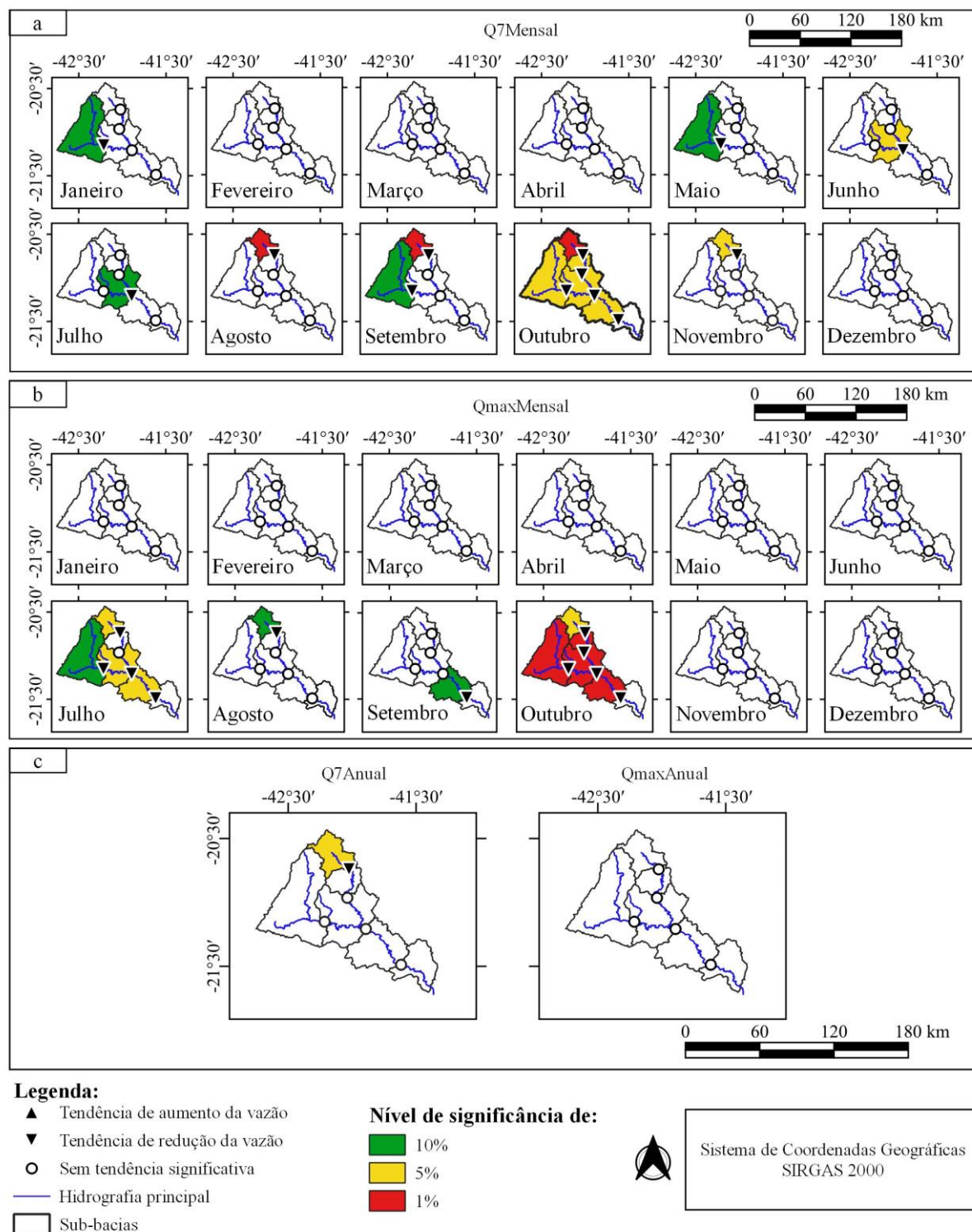
Tabela 1.7 – Estatística dos testes MK, MMK, SP e TSS para a análise de tendência das séries mensais e anuais de vazão máxima das estações fluviométricas localizadas na bacia do rio Muriaé.

Estação	Parâmetro	Jan	Fev	Mar	Abr	Mai	Jun	Jul	Ago	Set	Out	Nov	Dez	Anual
58920000	Z _{MK}	-0,612	-0,491	-0,638		-0,879	-1,362		-1,198	-1,293		-0,155	0,983	
	Z _{MMK}				-0,108			-1,840*			-2,759***			0,071
	Z _{SP}	-0,760	-0,566	-0,615	-0,141	-0,935	-1,432	-1,835*	-1,068	-1,279	-2,597***	-0,129	0,935	0,046
	TSS	-0,689	-0,427	-0,439	-0,043	-0,172	-0,175	-0,252	-0,181	-0,250	-0,770	-0,126	0,876	0,227
58930000	Z _{MK}	-0,690	-0,129	-0,759	0,069	-0,017	-1,276	-2,216**	-1,690*	-1,603	-2,276**	-1,431	-0,483	-0,498
	Z _{MMK}													
	Z _{SP}	-0,689	-0,239	-0,624	0,106	0,090	-1,294	-2,228**	-1,813*	-1,612	-2,258**	-1,334	-0,456	-0,470
	TSS	-0,178	-0,049	-0,116	0,018	0,000	-0,034	-0,069	-0,056	-0,091	-0,204	-0,362	-0,138	-0,208
58934000	Z _{MK}	-0,802	-0,517	-1,043	0,026	-0,405	-0,293	-1,639	-0,974	-0,957		-1,060	-0,095	-0,053
	Z _{MMK}										-2,965***			
	Z _{SP}	-0,887	-0,648	-1,094	0,155	-0,530	-0,269	-1,636	-0,989	-1,005	-2,709***	-1,081	-0,216	-0,071
	TSS	-0,434	-0,164	-0,303	0,002	-0,035	-0,030	-0,081	-0,056	-0,085	-0,414	-0,385	-0,077	-0,061
58940000	Z _{MK}	-0,371	-0,664	-0,328	-0,612		-1,181	-2,181**	-0,922	-1,491		-0,353	0,345	0,231
	Z _{MMK}					-0,532					-4,419***			
	Z _{SP}	-0,567	-0,794	-0,217	-0,554	-0,797	-1,205	-2,027**	-0,891	-1,470	-2,595***	-0,436	0,230	0,224
	TSS	-0,809	-0,873	-0,399	-0,557	-0,257	-0,328	-0,573	-0,174	-0,481	-1,573	-0,467	0,449	0,358
58960000	Z _{MK}	-0,767	-0,509	-0,207	-0,578	-0,345	-1,078		-0,784	-1,810*		-0,345	0,147	
	Z _{MMK}							-2,183**			-3,796***			-0,222
	Z _{SP}	-0,867	-0,578	-0,115	-0,522	-0,458	-1,106	-1,96**	-0,704	-1,771*	-2,481***	-0,301	0,098	-0,228
	TSS	-1,510	-0,800	-0,184	-0,528	-0,163	-0,298	-0,564	-0,227	-0,612	-1,638	-0,673	0,469	-0,477

MK – Mann-Kendall; MMK – Mann-Kendall Modificado; SP – correlação de Spearman; *Nível de significância de 10%; **Nível de significância de 5%;

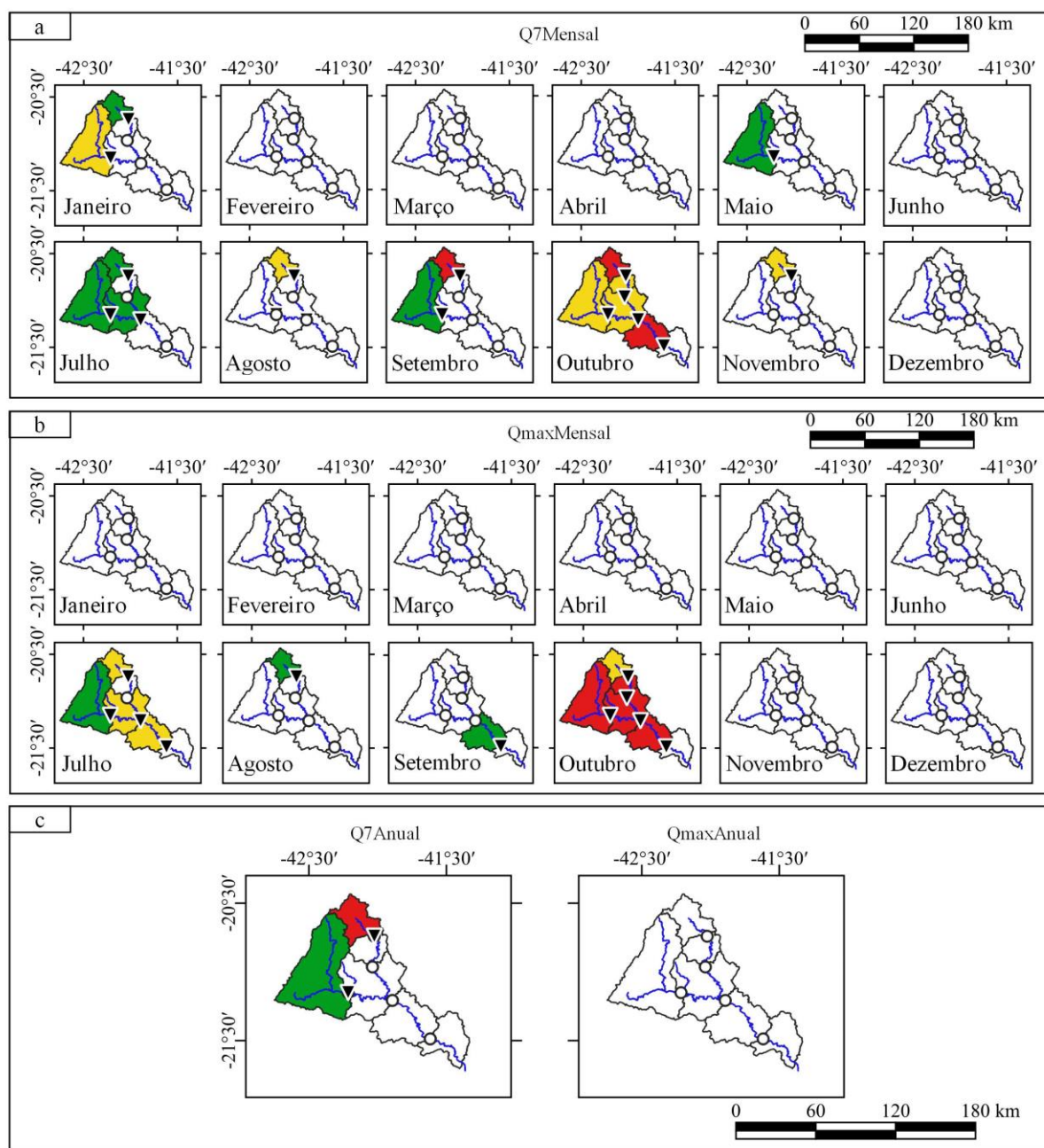
***Nível de significância de 1%. Fonte: A autora.

Figura 1.5 – Variação espacial da tendência das variáveis: a) Q_7 Mensal, b) Q_{max} Mensal e c) Q_7 Anual e Q_{max} Anual, identificadas pelos testes de MK e MMK em níveis de significância de 10%, 5% e 1%.



Fonte: A autora.

Figura 1.6 – Variação espacial da tendência das variáveis: a) Q_7 Mensal; b) Q_{max} Mensal e c) Q_7 Anual e Q_{max} Anual, identificadas pelo teste de SP em níveis de significância de 10%, 5% e 1%.



Legenda:

- ▲ Tendência de aumento da vazão
- ▼ Tendência de redução da vazão
- Sem tendência significativa
- Hidrografia principal
- Sub-bacias

Níveis de significância de:

- 10%
- 5%
- 1%



Sistema de Coordenadas Geográficas
SIRGAS 2000

Fonte: A autora.

1.3.2 Análise dos meses/anos de mudança nas séries de vazão

Os resultados dos pontos de mudança obtidos pelo teste de Pettitt considerando as análises de tendência significativas para os três níveis de confiança adotados são apresentados nas Tabelas 1.8 e 1.9 para Q_7 e $Q_{\text{máx}}$, respectivamente, apenas para os meses e anos em que foram encontrados valores significativos.

Na escala mensal, observaram-se que os meses do período seco apresentaram o maior número de pontos de mudança das variáveis analisadas, com destaque para o mês de outubro em todas as estações da bacia (Tabelas 1.8 e 1.9). Quanto às maiores mudanças da vazão mínima Q_7 (Tabela 1.8), para a significância mais forte encontrada ($\alpha = 0,05$), o teste de Pettitt indicou os meses de outubro e janeiro da estação Carangola (58930000) com 24,10% e 61,64% de diferença entre as médias das séries separadas pelos anos de 1992 e 2012, respectivamente.

Na escala anual da Q_7 , diferentemente dos testes de MK e SP, o teste de Pettitt indicou não existir estacionariedade em três estações fluviométricas, sendo encontradas mudanças significativas a 5% para a estação Carangola (58930000) e 10% para as estações Porciúncula (58934000) e Cardoso Moreira (58960000). Além disso, foi encontrada uma redução de 56,66% entre as médias de vazão para o ano hidrológico 2011/2012 na estação 58934000.

Tabela 1.8 – Meses/anos de mudança identificados pelo teste de Pettitt para as séries mensais e anuais de Q_7 das estações fluviométricas localizadas na bacia do rio Muriaé.

Estação	Parâmetro	Jan	Set	Out	Nov	Anual
58920000	Média 1 (m ³ /s)	57,49				
	Média 2 (m ³ /s)	25,85				
	PT	2012*				
58930000	Média 1 (m ³ /s)	16,48	4,21	4,08	4,21	3,63
	Média 2 (m ³ /s)	6,32	3,16	3,10	1,82	2,64
	PT	2012**	2009*	1992**	2012*	1984/1985**
58934000	Média 1 (m ³ /s)			5,78		4,13
	Média 2 (m ³ /s)			3,09		1,79
	PT			2011*		2012/2013*
58940000	Média 1 (m ³ /s)			27,40		
	Média 2 (m ³ /s)			13,54		
	PT			2011*		
58960000	Média 1 (m ³ /s)			30,74		22,92
	Média 2 (m ³ /s)			14,96		12,01
	PT			2012*		2011/2012*

*Nível de significância de 10%; **Nível de significância de 5%; ***Nível de significância de 1%.

Fonte: A autora.

Quanto às vazões máximas, a maior redução da vazão ocorreu no mês de outubro na estação Itaperuna (58940000), com 61,18% diferença entre as médias antes e depois do ano de 2012. Destacou-se, também, o mês de julho com anos de mudança na década de 1990, com reduções ao nível de 10% de significância com 35,32% na estação Carangola (58930000) e 31,48% na estação Itaperuna (58940000). Ressalta-se ainda que na escala anual, as vazões máximas apresentaram um comportamento estacionário, sem anos de mudança significativos, corroborando com os resultados obtidos nos testes de MK e SP.

No trabalho de Cecílio *et al.* (2021), realizado na bacia do rio Itapemirim, os autores encontraram a maioria dos pontos de mudança na década de 1990, com redução da vazão média nos meses do período seco da bacia e apontaram o crescimento populacional como o principal fator de alteração. Ferreira *et al.* (2021), por outro lado, identificaram mudanças no regime de vazões mínimas em uma estação fluviométrica nessa década em uma bacia próxima. Como possíveis explicações para essas alterações, os autores apontaram a influência das modificações no uso e cobertura da terra e aumento da demanda hídrica na bacia.

Tabela 1.9 – Meses/anos de mudança identificados pelo teste de Pettitt para as séries mensais e anuais de $Q_{\max}$ das estações fluviométricas localizadas na bacia do rio Muriaé.

Estação	Parâmetro	Jul	Set	Out
58920000	Média 1 (m ³ /s)			72,69
	Média 2 (m ³ /s)			39,92
	PT			2011**
58930000	Média 1 (m ³ /s)	10,22		
	Média 2 (m ³ /s)	6,61		
	PT	1991*		
58934000	Média 1 (m ³ /s)			30,72
	Média 2 (m ³ /s)			14,02
	PT			2011**
58940000	Média 1 (m ³ /s)	65,57		120,97
	Média 2 (m ³ /s)	44,93		46,96
	PT	1992*		2012*
58960000	Média 1 (m ³ /s)		78,02	138,36
	Média 2 (m ³ /s)		43,12	68,99
	PT		2009*	2011*

*Nível de significância de 10%; **Nível de significância de 5%; ***Nível de significância de 1%.

Fonte: A autora.

1.3.3 Análise dos fatores que influenciam no comportamento das vazões

Para uma melhor compreensão da análise de tendência das séries mensais e anuais de vazão, os testes estatísticos de MK, MMK e SP também foram aplicados aos dados

pluviométricos, assim como para a análise da alteração do uso e cobertura da terra e a quantificação das outorgas de direito do uso da água na bacia do rio Muriaé. As estatísticas dos testes se encontram detalhadas no Apêndice C.

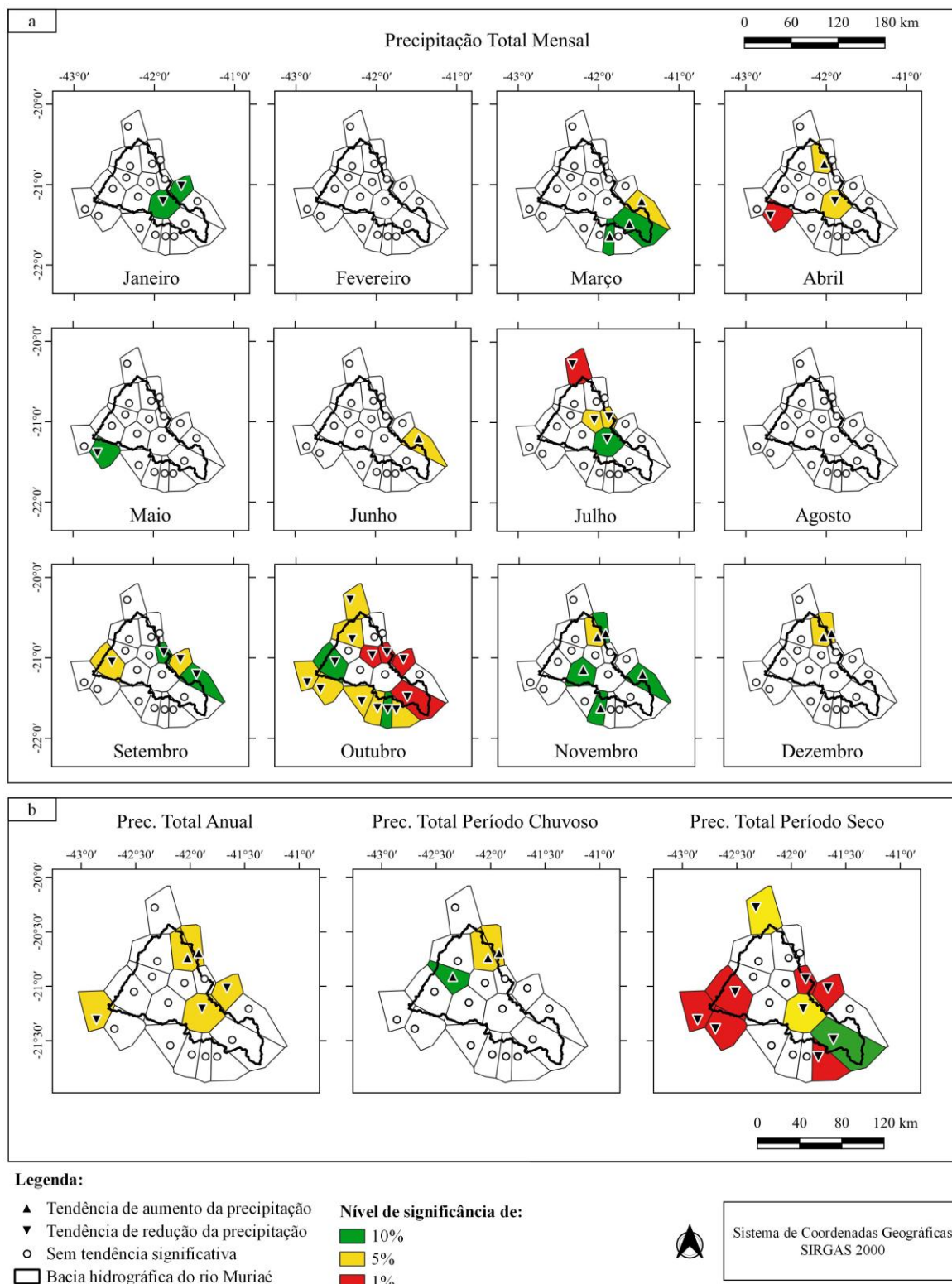
Quanto à análise das séries mensais de chuva (Figuras 1.7 e 1.8), os testes de MK, MMK e SP indicaram que os dados apresentaram uma tendência de redução para o período seco, nos meses de julho a outubro, nos diferentes níveis de significância. Nesse intervalo, observaram-se que a maioria das estações pluviométricas com as menores áreas de influências apresentou tendência negativa significativa nos níveis de 10% e 5% de significância. Por outro lado, apenas três estações pelo teste de MK e MMK e duas estações pelo teste de SP, apresentaram tendência positiva nos meses do período chuvoso, com a maioria das estações com área de influência insignificante diante da área da bacia.

Na avaliação dos resultados para os semestres chuvoso e seco (Figuras 1.7 e 1.8), os testes mostraram que o período seco apresenta o maior número de estações com tendências negativas de precipitação, com destaque para as regiões oeste e sudeste da bacia, onde se encontram as áreas de drenagem das estações fluviométricas Patrocínio do Muriaé (58920000) e Cardoso Moreira (58960000), respectivamente. O período chuvoso apresenta tendências positivas ao norte da bacia nos níveis de 5% de significância pelo teste de MK e 1% e 10% pelo teste de SP. Já para a análise dos totais anuais precipitados, verificaram-se tendências positivas e negativas ao nível de 5% de significância, destacando-se a redução das chuvas na área que abrange a estação fluviométrica Itaperuna (58940000).

As tendências negativas observadas na precipitação no período seco são coerentes com os resultados encontrados na análise das séries de vazão para o mesmo semestre. Essas condições hidroclimáticas tendem a serem agravadas ainda mais pelas mudanças climáticas, vindo a influenciar negativamente a disponibilidade hídrica nos próximos anos, como foi verificado nos estudos de Herrera-Pantoja e Hiscock (2015); Rocha *et al.* (2020); Siqueira *et al.* (2021) e Voisin *et al.* (2020).

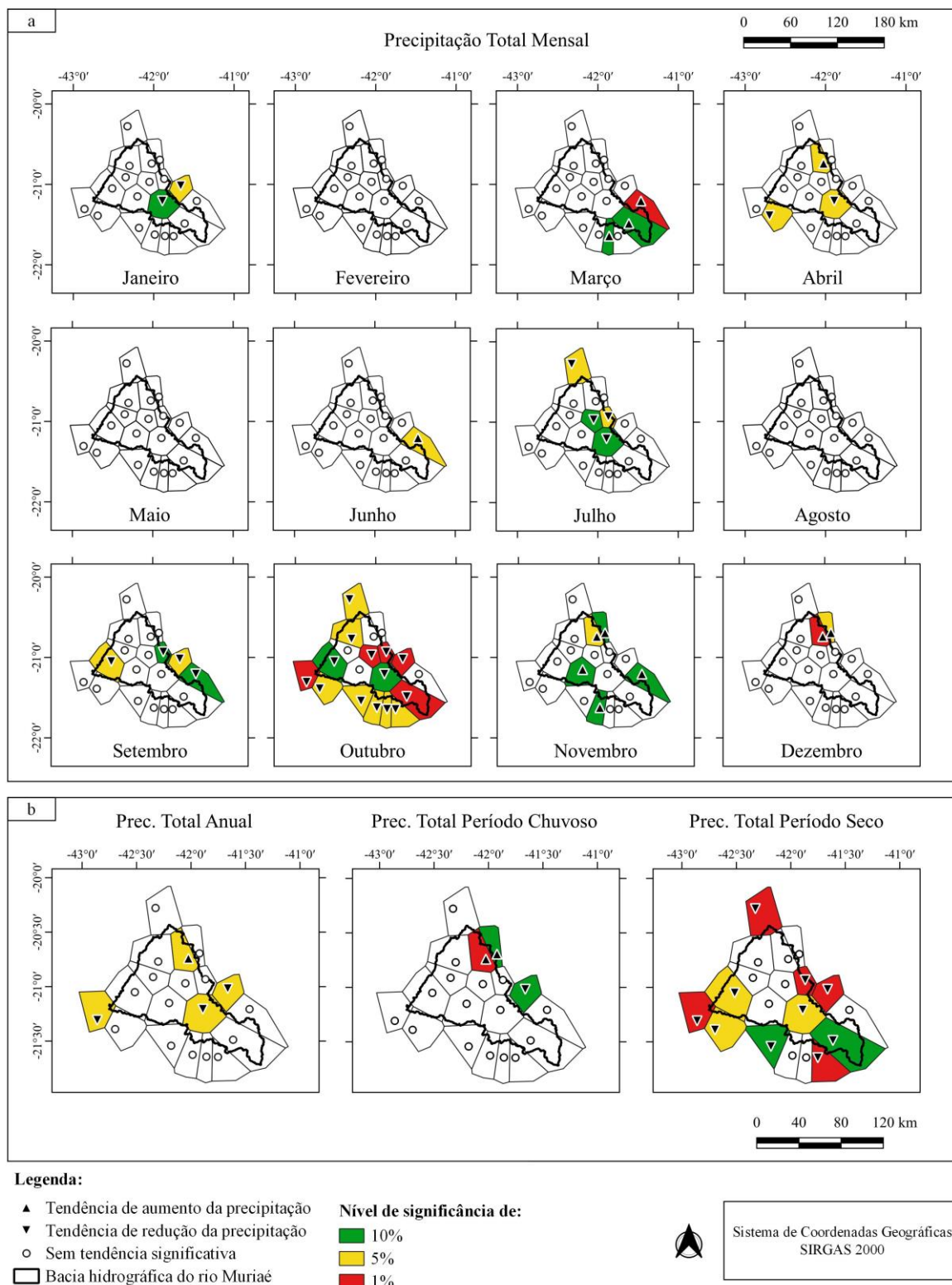
Existem, entretanto, outros fatores que regem essa disponibilidade, desde as características morfológicas da bacia, profundidade da camada de solo, evapotranspiração, interceptação da água da chuva (RÁPALO *et al.*, 2021), demanda de uso para o consumo humano e para os setores industrial e agrícola, bem como os fatores associados à governança e à socioeconomia da bacia (DELL'ANGELO; RULLI; D'ODORICO, 2018).

Figura 1.7 – Variação espacial da tendência das variáveis: a) Precipitação Total Mensal; e b) Precipitação Total Anual, Precipitação do Período Chuvoso e Período Seco, identificadas pelos testes MK e MMK em níveis de significância de 10%, 5% e 1%.



Fonte: A autora.

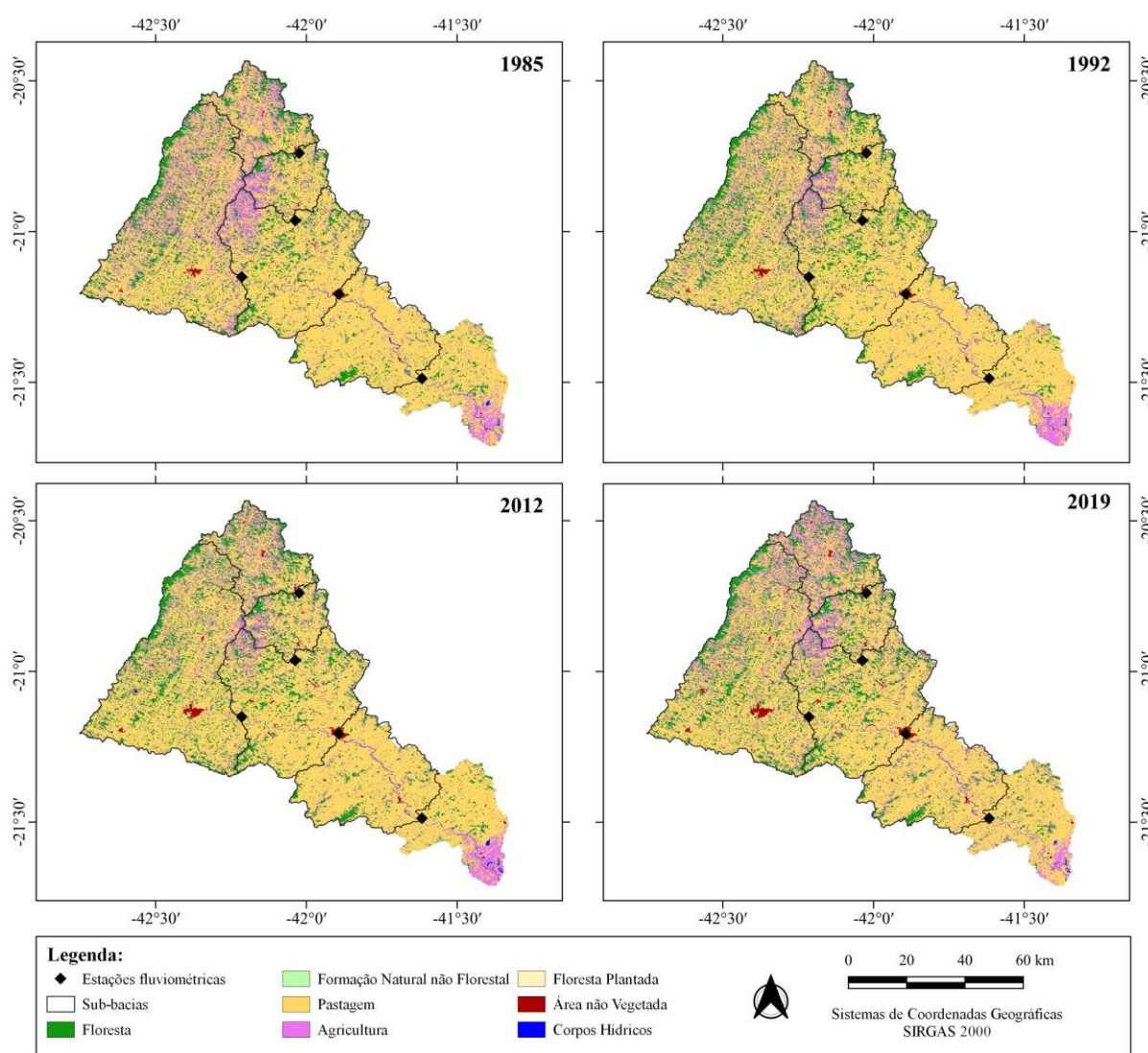
Figura 1.8 – Variação espacial da tendência das variáveis: a) Precipitação Total Mensal; e b) Precipitação Total Anual, Precipitação do Período Chuvoso e Período Seco, identificadas pelo teste de SP em níveis de significância de 10%, 5% e 1%.



Fonte: A autora.

O uso e cobertura da terra na bacia do rio Muriaé para os anos de 1985, 1992, 2012 e 2019 é apresentado na Figura 1.9. Observou-se que a bacia apresentou pouca variabilidade nas classes de uso e cobertura da terra, com pequenas alterações visualizadas em algumas classes, por exemplo, aumento das áreas sem vegetação na área de drenagem da estação Patrocínio do Muriaé (58920000). Além disso, a agropecuária apresentou um comportamento de aumento e diminuição, respectivamente, na foz da bacia. Essa classe teve um comportamento inverso nas regiões que abrangem a fronteira das áreas de drenagem Patrocínio do Muriaé (58920000), Carangola (58930000) e Porciúncula (58934000).

Figura 1.9 – Uso e cobertura da terra na bacia hidrográfica do rio Muriaé nos anos de 1985, 1992, 2012 e 2019.

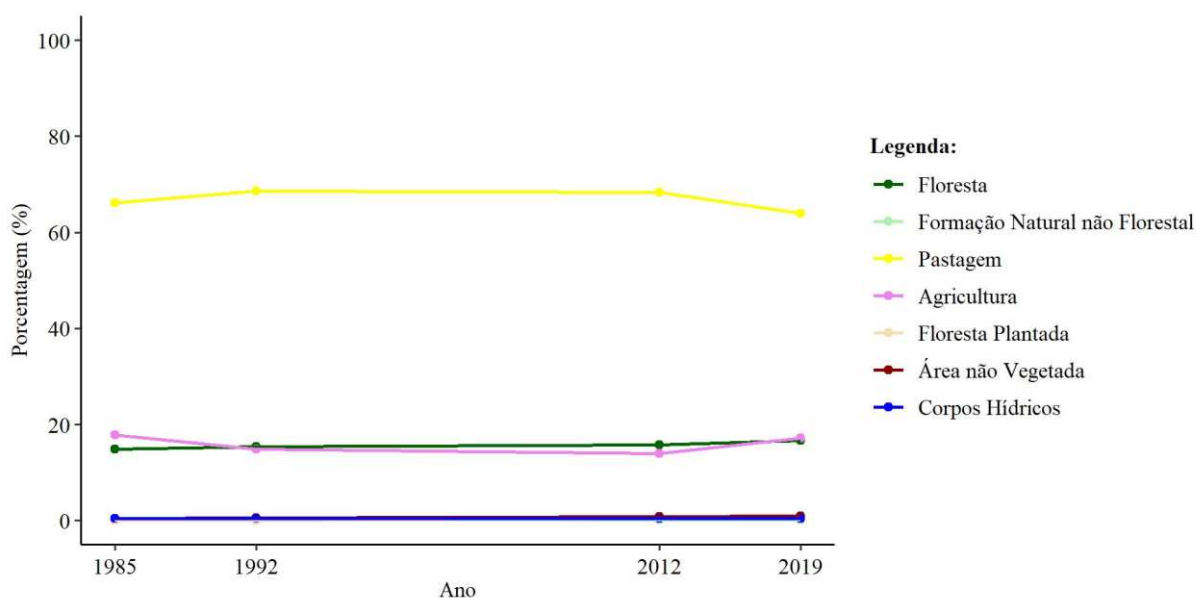


Fonte: A autora.

Segundo Rosa *et al.* (2021), o bioma Mata Atlântica, o qual a bacia do rio Muriaé pertence, apresenta estabilidade ao longo do tempo e em toda a sua extensão territorial, de modo que as principais mudanças estão associadas às transições entre as áreas ocupadas por atividades agropastoris e a vegetação nativa.

No ano de 1985, as classes de pastagem e agricultura ocupavam juntas 84,03% da área da bacia. Em 2019 essas classes continuaram a representar uma extensão territorial superior a 80% da bacia do rio Muriaé (Figura 1.10), confirmando a estabilidade do uso e cobertura da terra observada na Figura 1.9.

Figura 1.10 – Uso e cobertura da terra percentual nos anos de 1985, 1992, 2012 e 2019 na bacia hidrográfica do rio Muriaé.



Fonte: A autora.

Na bacia do rio Muriaé, observou-se que a conversão de floresta em áreas agrícolas e de pastagem ocorreu de forma mais expressiva nos anos anteriores à série de dados do projeto MapBiomias, visto que no ano de 1985, início do projeto, a bacia já apresentava uma quantidade significativa de áreas ocupadas por esses usos, que se manteve estável ao longo do tempo, mesma característica encontrada para o bioma em que se encontra (SILVA *et al.*, 2020a). No Brasil, a distribuição de perdas e ganhos de floresta, pastagem e agricultura ocorre de forma associada aos períodos de ocupação e exploração da terra, destacando-se o bioma Mata Atlântica, com um histórico de elevada degradação ambiental (CABALLERO; RUHOFF; BIGGS, 2022).

Na Tabela 1.10 tem-se as áreas de contribuições, positivas (+) e negativas (-), de cada classe para as mudanças ocorridas nas demais classes no período de 1985 a 2019.

Tabela 1.10 – Ganhos e perdas entre as classes de uso e cobertura da terra no período de 1985 a 2019 na bacia do rio Muriaé.

Classe	Área (km ²)							Total
	FL	FN	P	AG	FP	AnV	CH	
FL	0	-0,71	55,03	99,51	-7,62	-0,52	-0,36	145,33
FN	0,71	0	0,63	0,46	0	0	-0,02	1,78
P	-55,03	-0,63	0	-62,62	-25,70	-33,04	-2,51	-179,53
AG	-99,51	-0,46	62,62	0	-13,05	-6,13	4,10	-52,43
FP	7,62	0	25,7	13,05	0	0	0	46,37
AnV	0,52	0	33,04	6,13	0	0	0,38	40,07
CH	0,36	0,02	2,51	-4,10	0	-0,38	0	-1,59

FL – Floresta; FN – Formação Natural não Florestal; PA – Pastagem; AG – Agricultura; FP – Floresta Plantada; AnV – Área não Vegetada e CH – Corpos Hídricos. Fonte: A autora.

Entre 1985 e 2019, a classe floresta incorporou 55,03 km² da pastagem e 99,51 km² da agricultura. Por outro lado, a agricultura teve ganhos de 62,62 km² de áreas que antes eram destinadas à pastagem. De forma similar, a pastagem também contribuiu com as classes de floresta plantada (+25,70 km²) e áreas não vegetadas (+33,04 km²). Considerando a área da bacia, que apresenta aproximadamente 8.126,00 km², as mudanças observadas podem não exercer influência no regime de variação das vazões diante de sua baixa magnitude. Gomes *et al.* (2021) observaram que as mudanças na vazão da bacia estavam mais relacionadas à precipitação do que ao uso e cobertura da terra, afirmando ainda ser difícil mensurar a sua influência na diminuição e aumento da vazão.

Na Tabela 1.11 têm-se as áreas ocupadas pelas diferentes classes de uso e cobertura da terra nas áreas de drenagem das estações fluviométricas localizadas na bacia do rio Muriaé. As classes de pastagem e agricultura foram dominantes em todos os anos analisados (Tabela 1.11), apresentando em 2019 uma porcentagem de áreas ocupadas igual a 76,81% em Patrocínio do Muriaé (58920000), 72,68% em Carangola (58930000), 74,01% em Porciúncula (58934000), 77,50% em Itaperuna (58940000) e 79,80% em Cardoso Moreira (58960000).

Na análise da variação percentual, as estações apresentaram um comportamento similar à bacia com algumas exceções, como, por exemplo, a classe floresta, com redução de sua área na estação Carangola (58930000) nos períodos de 1985 a 1992 (-0,27%) e 2012 a 2019 (-1,07%). Por outro lado, entre 2012 e 2019 foram observados aumentos de áreas florestais nas

estações Patrocínio do Muriaé (58920000), Itaperuna (58940000) e Cardoso Moreira (58960000), correspondentes a 0,30%, 0,66% e 0,81%, respectivamente.

Tabela 1.11 – Alteração do uso e cobertura da terra entre os anos de 1985, 1992, 2012 e 2019 nas áreas de drenagem das estações fluviométricas localizadas na bacia do rio Muriaé.

Código	Classe	Porcentagem da área (%)				Variação da área (%)		
		1985	1992	2012	2019	1985-1992	1992-2012	2012-2019
58920000	FL	19,13	20,69	20,34	20,64	+1,57	-0,36	+0,30
	FN	0,20	0,19	0,22	0,21	0,00	+0,03	-0,01
	PA	57,19	60,10	63,64	59,18	2,92	+3,53	-4,45
	AG	22,71	18,06	13,77	17,62	-4,64	-4,29	+3,86
	FP	0,05	0,06	0,77	0,93	0,01	+0,71	+0,16
	AnV	0,55	0,70	1,05	1,19	0,15	+0,35	+0,14
	CH	0,18	0,19	0,22	0,23	0,01	+0,02	+0,01
58930000	FL	22,12	21,86	25,47	24,40	-0,27	+3,62	-1,07
	FN	0,53	0,51	0,61	0,58	-0,02	+0,09	-0,02
	PA	50,10	56,27	46,78	39,26	+6,17	-9,49	-7,52
	AG	26,62	20,57	25,03	33,43	-6,06	+4,46	+8,40
	FP	0,12	0,12	1,27	1,37	+0,01	+1,15	+0,10
	AnV	0,46	0,64	0,81	0,93	+0,18	+0,17	+0,12
	CH	0,04	0,04	0,03	0,03	+0,00	-0,01	+0,01
58934000	FL	20,38	20,96	23,58	23,20	+0,58	+2,62	-0,38
	FN	0,37	0,36	0,42	0,41	-0,01	+0,06	-0,01
	PA	53,25	58,70	51,07	44,65	+5,45	-7,63	-6,42
	AG	25,43	19,23	22,89	29,36	-6,20	+3,66	+6,46
	FP	0,11	0,12	1,22	1,44	+0,02	+1,10	+0,22
	AnV	0,39	0,57	0,73	0,83	+0,17	+0,16	+0,10
	CH	0,07	0,06	0,08	0,12	-0,01	+0,02	+0,03
58940000	FL	17,88	19,41	19,40	20,07	+1,52	+0,00	+0,66
	FN	0,21	0,21	0,24	0,23	+0,00	+0,03	-0,01
	PA	60,31	63,23	63,71	58,44	+2,92	+0,48	-5,27
	AG	20,79	16,20	14,76	19,06	-4,60	-1,44	+4,30
	FP	0,05	0,06	0,68	0,86	+0,01	+0,62	+0,17
	AnV	0,47	0,64	0,94	1,05	+0,16	+0,30	+0,11
	CH	0,28	0,26	0,26	0,30	-0,02	+0,00	+0,04
58960000	FL	15,78	16,77	17,11	17,92	+0,99	+0,34	+0,81
	FN	0,18	0,18	0,21	0,20	+0,00	+0,03	-0,01
	PA	65,11	67,58	67,60	62,29	+2,46	+0,02	-5,31
	AG	18,02	14,46	13,27	17,50	-3,56	-1,18	+4,23
	FP	0,04	0,05	0,55	0,69	+0,01	+0,50	+0,14
	AnV	0,47	0,62	0,91	1,01	+0,15	+0,29	+0,10
	CH	0,39	0,35	0,35	0,39	-0,04	+0,00	+0,04

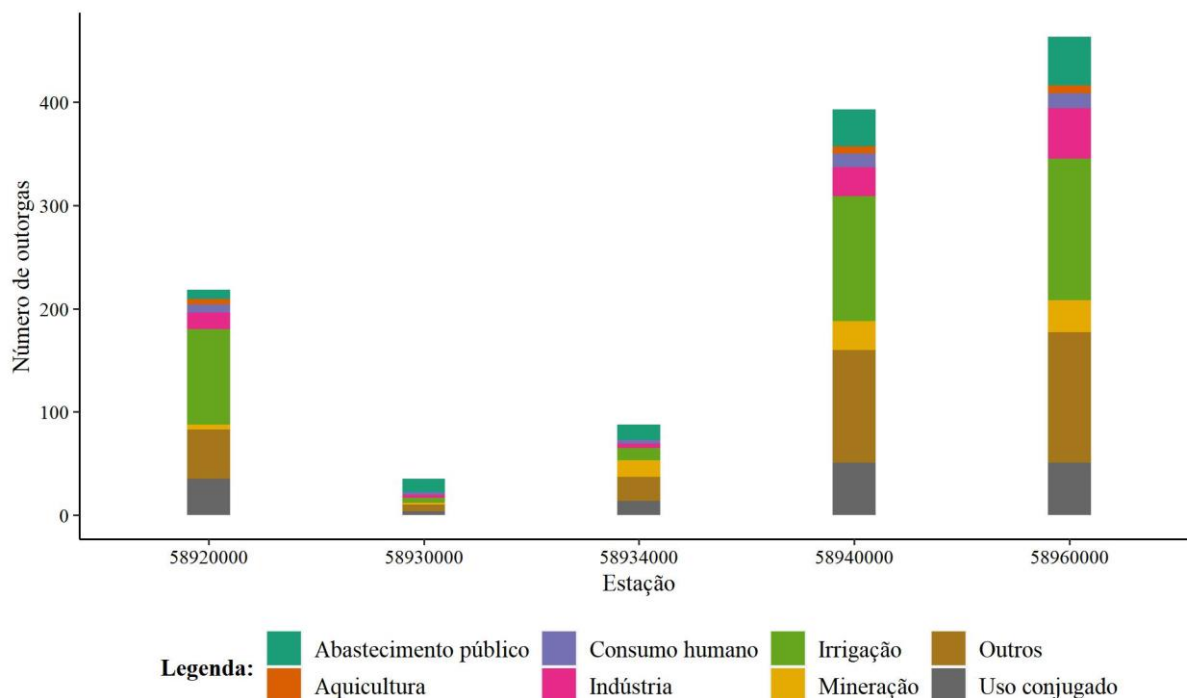
FL – Floresta; FN – Formação Natural não Florestal; PA – Pastagem; AG – Agricultura; FP – Floresta Plantada; AnV – Área não Vegetada e CH – Corpos Hídricos. Fonte: A autora.

As florestas fornecem inúmeros benefícios sob o ponto de vista hidrológico, por exemplo: redução de vazões máximas (LI *et al.*, 2017b); elevadas taxas de infiltração (ARSYAD; WAHYUNI, 2019) e recarga de águas subterrâneas (BREMER *et al.*, 2019). Entretanto, alguns autores destacam que as formações florestais podem ter uma influência negativa na disponibilidade hídrica no período seco, em razão do tamanho do dossel, proporcionando elevadas taxas evaporativas (BROWN *et al.*, 2013; CECÍLIO; PIMENTEL; ZANETTI, 2019). Todavia, considerando as variações das áreas florestais proporcionalmente às áreas de drenagem envolvidas, observou-se que as estações apresentaram mudanças pouco expressivas, reforçando a sua baixa influência nas tendências observadas nas séries de vazão.

Ao analisar o período mais recente, entre 2012 e 2019 (Tabela 1.11), a agricultura apresentou crescimento em todas as áreas de drenagem das estações fluviométricas, com a menor variação na estação Patrocínio do Muriaé (58920000) e a maior na estação Carangola (58930000), iguais a 3,86% e a 8,40%, respectivamente. Destaca-se, ainda, a pastagem com as maiores reduções em todas as estações, com perdas variando entre 4,45% (58920000 – Patrocínio do Muriaé) e 7,52% (58930000 – Carangola). Embora tenha sido observado que as áreas de pastagem apresentaram diminuição, essa classe ainda continuou sendo a mais representativa na área de drenagem das estações.

Quanto à análise da demanda hídrica na bacia do rio Muriaé, foram consideradas 465 outorgas de uso consuntivo de captação superficial para o período de 2000 a 2019 para a área de drenagem a montante das estações fluviométricas (Figura 1.11). Pela análise das outorgas, observaram-se que as classes indústria, irrigação, outros usos e usos conjugados apresentaram a maior quantidade de outorgas concedidas.

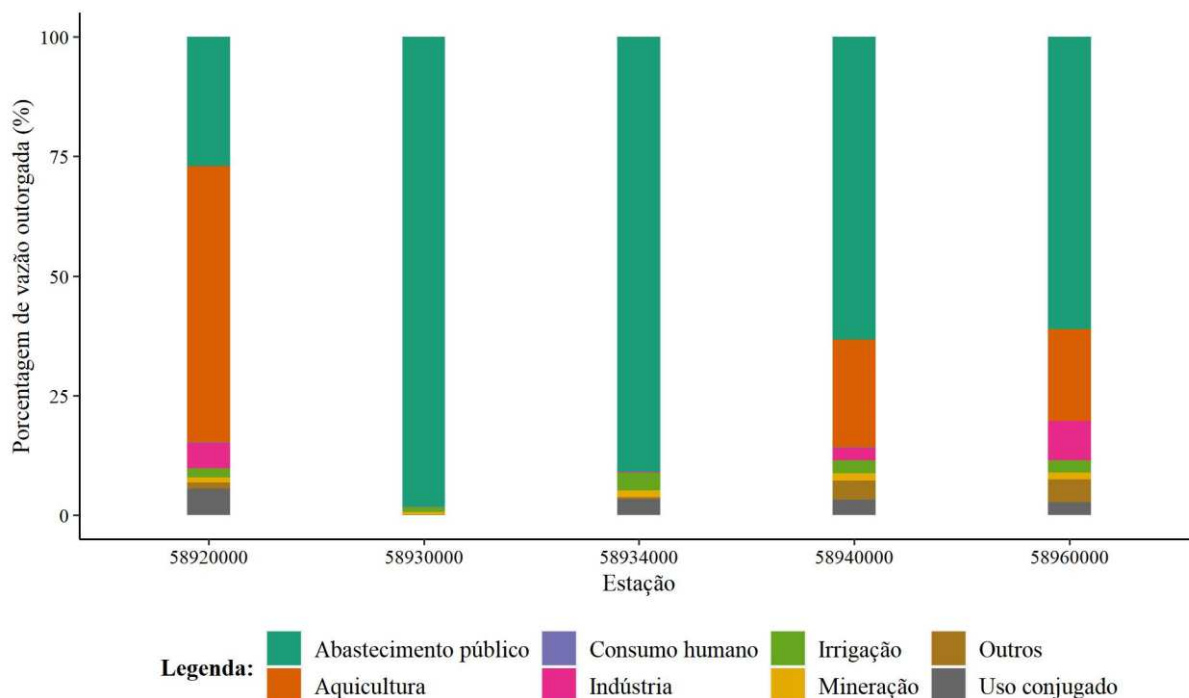
Figura 1.11 – Distribuição de outorgas de captação superficial entre 2000 e 2019, por finalidade de uso da água, na área de drenagem a montante das estações fluviométricas localizadas na bacia do rio Muriaé.



Fonte: A autora.

Ao se analisar a distribuição das porcentagens da vazão média outorgada (m^3/s) (Figura 1.12), verificou-se que as classes mais representativas nas áreas de drenagem das estações corresponderam ao abastecimento público, com um intervalo de 27,06% a 98,20%, seguido pela aquicultura, que variou entre 0% e 57,78%. Destacaram-se, também, as outorgas destinadas às atividades industriais, que tiveram a menor porcentagem na estação Carangola (58930000), com 0,06%, e a maior na estação Cardoso Moreira (58960000), com 8,21%.

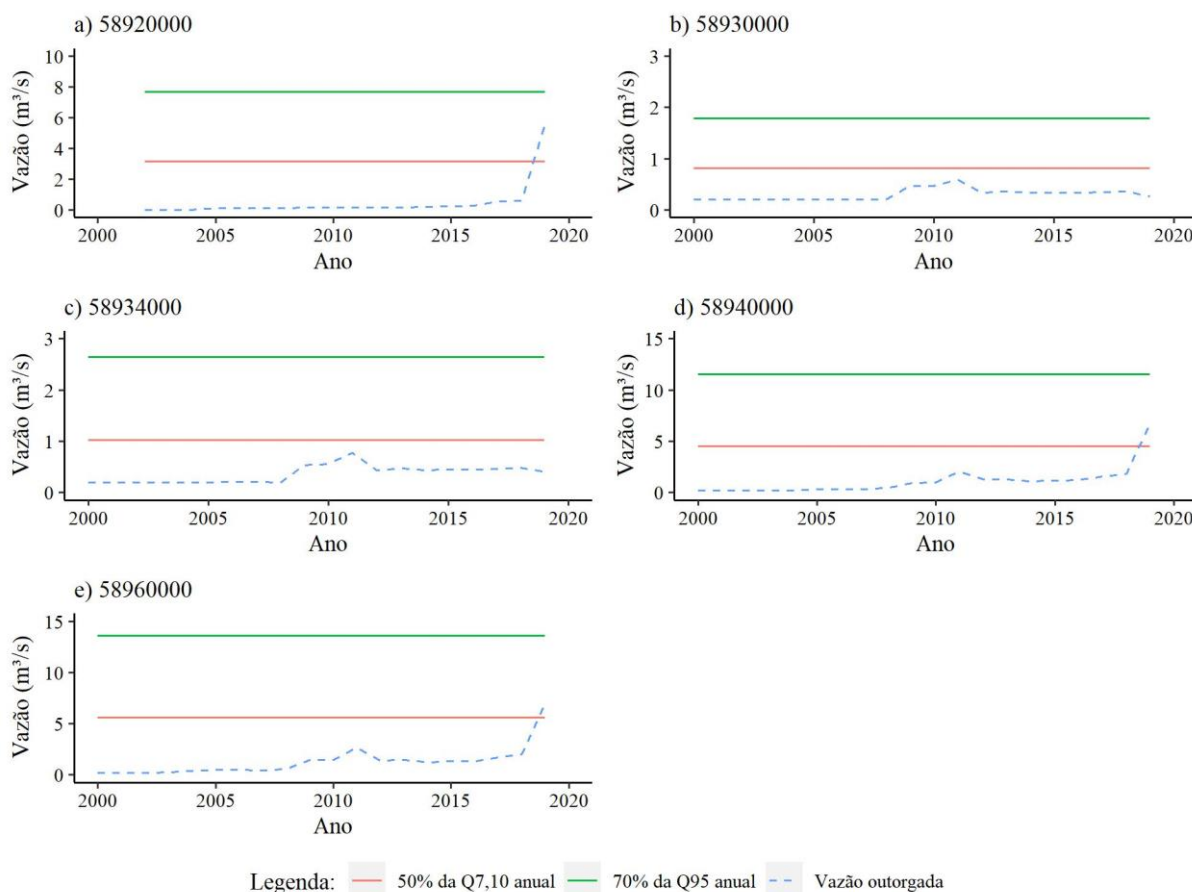
Figura 1.12 – Porcentagens de vazão média outorgada entre 2000 e 2019, por finalidade de uso da água, na área de drenagem a montante das estações fluviométricas localizadas na bacia do rio Muriaé.



Fonte: A autora.

A verificação das vazões outorgadas em consonância com os limites estabelecidos pela legislação, sendo 50% da $Q_{7,10}$ para os cursos d'água estaduais de Minas Gerais e Rio de Janeiro e 70% da Q_{95} anual para os rios de dominialidade federal, entre 2000 e 2019, está representada na Figura 1.13. Em quatro estações, Carangola (58930000), Porciúncula (58934000), Itaperuna (58940000) e Cardoso Moreira (58960000), embora a vazão outorgada não tenha ultrapassado os limites estabelecidos na legislação, houve um aumento na vazão outorgada nos anos de 2009, 2010 e 2011, o que pode ter influenciado na tendência do regime de vazões da bacia, visto que esses anos antecederam ao principal ano de mudanças nas tendências, que foi 2012, identificado pelo teste de Pettitt ao analisar a mudança na tendência das séries de vazão. O aumento na demanda da água pode ter um efeito negativo na vazão de um rio, levando a problemas futuros quanto à disponibilidade hídrica (SHI *et al.*, 2022; ZHANG *et al.*, 2015).

Figura 1.13 – Comportamento das vazões outorgadas ao longo da série histórica em comparação aos limites máximos das vazões $Q_{7,10}$ e Q_{95} anual estabelecidos para fins de outorga de uso de água nos estados de Minas Gerais/Rio de Janeiro e Brasil, respectivamente.



Fonte: A autora.

Observou-se, ainda, que na área de influência da estação de Patrocínio do Muriaé (58920000) (Figura 1.13), ocorreram vazões outorgadas superiores aos limites estabelecidos pela legislação estadual no ano de 2019, ou seja, 50% da $Q_{7,10}$. Considerando que a vazão outorgada foi calculada para cada área de drenagem, esse limite excedido impactou nas demais estações a jusante, Itaperuna (58940000) e Cardoso Moreira (58960000), em que as vazões liberadas pelos órgãos ambientais excederam a vazão máxima passível de ser outorgada (50% de $Q_{7,10}$), correspondentes a 4,52 m³/s e 5,59 m³/s, respectivamente. Segundo Miller *et al.* (2021), as alterações da vazão na bacia são influenciadas, principalmente, pelo que ocorre nas sub-bacias de cabeceira, sendo que na bacia do rio Muriaé, essas modificações estão relacionadas, principalmente, ao abastecimento público.

Ao se analisar a Figura 1.13 no que diz respeito às vazões de referência para fins de outorga nos cursos d'água de domínio federal, verificou-se que a vazão outorgada não excedeu

o limite estabelecido pela própria ANA. Entretanto, a adoção de 70% da Q₉₅ anual como limite para as captações em rios interestaduais apresenta um fator de risco quanto aos conflitos que envolvem o uso da água na bacia (MOREIRA *et al.*, 2020), uma vez que a utilização dos critérios com base na escala anual não permite a flexibilização dos valores outorgados de acordo com os períodos seco e chuvoso (SILVA; SILVA; MOREIRA, 2015).

Segundo WWAP (2018), o aumento da demanda hídrica é influenciado diretamente pelas alterações da finalidade de uso associadas ao crescimento populacional e subsequente expansão das atividades industrial e agrícola, esperando-se para os próximos anos demandas humana e industrial com valores mais expressivos. Além disso, o crescimento da população humana, principalmente nos países em desenvolvimento, ocorre de forma desigual quanto aos recursos naturais disponíveis, comprometendo a disponibilidade hídrica, agravada pelas mudanças climáticas, que tendem a ocasionar a diminuição das chuvas no período seco (BORETTI; ROSA, 2019; HUANG; YUAN; LIU, 2021).

1.4 CONCLUSÕES

Com base nos resultados obtidos e nas análises realizadas, pode-se concluir que:

- A aplicação de diferentes métodos estatísticos (Mann-Kendall, Mann-Kendall Modificado, correlação de Spearman, inclinação de Theil-Sen e teste de Pettitt) permitiu diagnosticar o comportamento hídrico temporal e espacial da bacia do rio Muriaé;
- As vazões mínimas e máximas apresentaram tendências negativas em diferentes níveis de significância, destacando-se os meses do período seco com o maior número de tendências de redução, principalmente no mês de outubro, e as mudanças mais significativas nos anos de 1992 e 2012;
- Na análise dos fatores que podem alterar a vazão, a precipitação apresentou tendência negativa e significativa nos meses de estiagem da bacia, principalmente entre julho e outubro. Diante disso, verificou-se uma correspondência dos resultados encontrados com as tendências encontradas nas séries de vazão, indicando uma possível influência das chuvas sobre o regime de vazão da bacia do rio Muriaé;
- As alterações do uso e cobertura da terra na bacia do rio Muriaé não foram expressivas ao longo do período analisado, não constituindo em fator determinante de mudança no regime de vazões da bacia;
- Quanto à demanda hídrica, observou-se que o abastecimento público foi a finalidade de uso da água mais importante no período analisado. Na análise das vazões outorgadas,

verificou-se um aumento das concessões em 2009, 2010 e 2011, que antecederam o ano de 2012, tido como principal ano em termos de significância estatística no teste de Pettitt na análise de tendência da vazão; e

- O conhecimento do comportamento hidrológico da bacia, bem como os principais fatores associados às mudanças do regime de vazões, permite o melhor gerenciamento ambiental nas bacias hidrográficas, com destaque para as que apresentam maior vulnerabilidade aos eventos extremos, de modo a minimizar as pressões humanas sobre os recursos hídricos.

1.5 REFERÊNCIAS

ADELOYE, A. J.; MONTASERI, M. Preliminary streamflow data analyses prior to water resources planning study. **Hydrological Sciences Journal**, v. 47, n. 5, p. 679–692, 2002.

AHMAD, I. *et al.* Precipitation Trends over Time Using Mann-Kendall and Spearman's rho Tests in Swat River Basin, Pakistan. **Advances in Meteorology**, p. 1–15, 2015.

ALVARES, C. A. *et al.* Köppen's climate classification map for Brazil. **Meteorologische Zeitschrift**, v. 22, n. 6, p. 711–728, 2013.

ARSYAD, U.; WAHYUNI. The characteristics of infiltration in natural forest in teaching forest of Hasanuddin University University at Maros Regency. **IOP Conference Series: Earth and Environmental Science**, v. 270, n. 1, p. 012004, 2019.

AUGUIE, B.; ANTONOV, A. **Package “gridExtra”**: Miscellaneous Functions for ‘Grid’ Graphics Type. p. 1–9, 2017. Disponível em: <https://cran.r-project.org/web/packages/gridExtra/gridExtra.pdf>. Acesso em: 10 mai 2022.

ÁVILA, M. W. DA *et al.* Gestão qualitativa dos recursos hídricos. Proposta metodológica para o planejamento de uma rede de estações para monitoramento da qualidade de águas superficiais. Estudo de caso: bacia hidrográfica do Rio Muriaé. **Revista Brasileira de Recursos Hídricos**, v. 21, n. 2, p. 401–415, 2016.

AWANGE, J. L.; MPELASOKA, F.; GONCALVES, R. M. When every drop counts: Analysis of Droughts in Brazil for the 1901-2013 period. **Science of The Total Environment**, v. 566–567, p. 1472–1488, 1 out. 2016.

AYEDUN, C. A. *et al.* The flooding effect on residential property values: A case study of shogunro residential estate, agege; lagos state nigeria. **International Journal of Civil Engineering and Technology**, v. 9, n. 6, p. 489–496, 2018.

BHATTI, A. S. *et al.* Trend in extreme precipitation indices based on long term in situ precipitation records over Pakistan. **Water (Switzerland)**, v. 12, n. 3, p. 1–19, 2020.

BOLAR, K. **Package “STAT”**: Interactive Document for Working with Basic Statistical Analysis. p. 3, 2019. Disponível em: <https://cran.r-project.org/web/packages/STAT/>. Acesso em: 20 ago. 2021.

BORETTI, A.; ROSA, L. Reassessing the projections of the World Water Development Report. **NPJ Clean Water**, v. 15, p. 1–6, 2019.

BREMER, L. L. *et al.* Contributions of native forest protection to local water supplies in East Maui. **Science of The Total Environment**, v. 688, p. 1422–1432, 20 out. 2019.

BROWN, A. E. *et al.* Impact of forest cover changes on annual streamflow and flow duration curves. **Journal of Hydrology**, v. 483, p. 39–50, 13 mar. 2013.

CALDERA, H. P. G. M.; PIYATHISSE, V. R. P. C.; NANDALAL, K. D. W. A Comparison of Methods of Estimating Missing Daily Rainfall Data. **Engineer: Journal of the Institution of Engineers**, Sri Lanka, v. 49, n. 4, p. 1–8, 2016.

CECÍLIO, R. A. *et al.* Trends in monthly and annual streamflow related to rainfall and land use at the Atlantic rainforest biome. **Journal of South American Earth Sciences**, v. 112, p. 103600, 1 dez. 2021.

CECÍLIO, R. A.; PIMENTEL, S. M.; ZANETTI, S. S. Modeling the influence of forest cover on streamflows by different approaches. **CATENA**, v. 178, p. 49–58, 1 jul. 2019.

CHAULUKA, F.; SINGH, S.; KUMAR, R. Rainfall and streamflow trends of Thuchila River, Southern Malawi. **Materials Today: Proceedings**, v. 34, p. 846–855, 1 jan. 2021.

DELL'ANGELO, J.; RULLI, M. C.; D'ODORICO, P. The Global Water Grabbing Syndrome. **Ecological Economics**, v. 143, p. 276–285, 1 jan. 2018.

EDUARDO, C.; TUCCI, M. Estimativa da Disponibilidade Hídrica Através da Curva de Permanência. **Revista Brasileira de Recursos Hídricos**, v. 13, p. 111–124, 2008.

EMBRAPA, E. B. D. P. A. **Súmula da X Reunião Técnica de Levantamento de Solos**. Rio de Janeiro: Serviço Nacional de Levantamento e Conservação de Solos, 1979.

EMBRAPA SOLOS. **Mapas de solos do Brasil**. Disponível em: http://geoinfo.cnps.embrapa.br/layers/geonode%3Aabrasil_solos_5m_20201104#more. Acesso em: 20 ago. 2021.

FANG, J. *et al.* Observed changes in hydrological extremes and flood disaster in Yangtze River Basin: spatial–temporal variability and climate change impacts. **Natural Hazards**, v. 93, n. 1, p. 89–107, 1 ago. 2018.

FERREIRA, F. L. V. *et al.* Time Series Trends of Streamflow and Rainfall in The Santo Antônio River Basin, Brazil. **Engenharia Agrícola**, v. 41, n. 1, p. 47–55, 2021.

GOMES, L. C. *et al.* Disentangling the historic and future impacts of land use changes and climate variability on the hydrology of a mountain region in Brazil. **Journal of Hydrology**, v. 594, p. 125650, 1 mar. 2021.

GÜÇLÜ, Y. S. Improved visualization for trend analysis by comparing with classical Mann-Kendall test and ITA. **Journal of Hydrology**, v. 584, p. 124674, 1 maio 2020.

HAMED, K. H.; RAO, A. R. A modified Mann-Kendall trend test for autocorrelated data. **Journal of Hydrology**, v. 204, p. 182–196, 1998.

HERRERA-PANTOJA, M.; HISCOCK, K. M. Projected impacts of climate change on water availability indicators in a semi-arid region of central Mexico. **Environmental Science & Policy**, v. 54, p. 81–89, 1 dez. 2015.

HUANG, Z.; YUAN, X.; LIU, X. The key drivers for the changes in global water scarcity: Water withdrawal versus water availability. **Journal of Hydrology**, v. 601, p. 126658, 1 out. 2021.

INMET, I. N. DE M. **Normais Climatológicas do Brasil 1961-1990**. Disponível em: <<https://portal.inmet.gov.br/normais#>>. Acesso em: 10 jan. 2022.

INOMATA, R. **Package “hidrowebr”**: uma ferramenta para análise estatística de séries hidrológicas. 2019. Disponível em: <https://rdr.io/github/renatoinomata/hidrowebr/>. Acesso em: 20 ago. 2021.

ISHAK, E. H. *et al.* Evaluating the non-stationarity of australian annual maximum flood. **Journal of Hydrology**, v. 494, p. 134–145, 8 jun. 2013.

JÚNIOR, S. F. A. X. *et al.* Precipitation trends analysis by mann-kendall test: A case study of paraíba, brazil. **Revista Brasileira de Meteorologia**, v. 35, n. 2, p. 187–196, 1 abr. 2020.

KASSAMBARA, A. **Package “Ggpubr”**: “Ggplot2” Based Publication Ready Plots. 2021. Disponível em: <https://cran.r-project.org/package=ggpubr>. Acesso em: 20 ago. 2021.

LI, Q. *et al.* Forest cover change and water yield in large forested watersheds: A global synthetic assessment. **Ecohydrology**, v. 10, p. e1838, 2017.

LIU, J. *et al.* Contribution of multiple climatic variables and human activities to streamflow changes across China. **Journal of Hydrology**, v. 545, p. 145–162, 1 fev. 2017.

MACCIOTTA, R. *et al.* Quantifying rock fall probabilities and their temporal distribution associated with weather seasonality. **Landslides**, v. 14, p. 2025–2039, 2017.

MACHADO, R. A. S.; OLIVEIRA, A. G.; LOIS-GONZÁLEZ, R. C. Urban ecological infrastructure: The importance of vegetation cover in the control of floods and landslides in Salvador / Bahia, Brazil. **Land Use Policy**, v. 89, p. 104180, 1 dez. 2019.

MAHMOOD, R.; JIA, S. Assessment of hydro-climatic trends and causes of dramatically declining stream flow to Lake Chad, Africa, using a hydrological approach. **Science of The Total Environment**, v. 675, p. 122–140, 20 jul. 2019.

MALIK, A. *et al.* Application of non-parametric approaches to identify trend in streamflow during 1976–2007 (Naula watershed). **Alexandria Engineering Journal**, v. 59, n. 3, p. 1595–1606, 1 jun. 2020.

MALIK, A.; KUMAR, A. Spatio-temporal trend analysis of rainfall using parametric and

non-parametric tests: case study in Uttarakhand, India. **Theoretical and Applied Climatology**, v. 140, n. 1–2, p. 183–207, 2020.

MAPBIOMAS. **Coleção [versão 6.0] da Série Anual de Mapas de Cobertura e Uso de Solo do Brasil**. Disponível em: <https://mapbiomas.org/>. Acesso em: 3 nov. 2021. Acesso em: 25 ago. 2021.

MARENGO, J. A. *et al.* A seca e a crise hídrica de 2014-2015 em São Paulo. **Revista USP**, n. 106, p. 31–44, 2015.

MCLEOD, A. I. **Package “Kendall”**: Kendall Rank Correlation and Mann-Kendall Trend Test. 2022. Disponível em: <https://cran.r-project.org/web/packages/Kendall/index.html>. Acesso em: 25 ago. 2021.

MILLER, O. L. *et al.* Changing climate drives future streamflow declines and challenges in meeting water demand across the southwestern United States. **Journal of Hydrology X**, v. 11, p. 100074, 1 maio 2021.

MOREIRA, H. S. *et al.* Cenários de disponibilidade hídrica para concessão de outorga: estudo de caso da Bacia Vertentes do Rio Grande, Estados de Minas Gerais e São Paulo, Brasil. **Revista Brasileira de Gestão Ambiental e Sustentabilidade**, v. 7, n. 15, p. 341–340, 2020.

NALLEY, D.; ADAMOWSKI, J.; KHALIL, B. Using discrete wavelet transforms to analyze trends in streamflow and precipitation in Quebec and Ontario (1954-2008). **Journal of Hydrology**, v. 475, p. 204–228, 19 dez. 2012.

NAUDITT, A. *et al.* Evaluating tropical drought risk by combining open access gridded vulnerability and hazard data products. **Science of The Total Environment**, v. 822, p. 153493, 20 maio 2022.

NAUMANN, G. *et al.* Global Changes in Drought Conditions Under Different Levels of Warming. **Geophysical Research Letters**, v. 45, n. 7, p. 3285–3296, 2018.

PATAKAMURI, S. K.; MUTHIAH, K.; SRIDHAR, V. Long-Term homogeneity, trend, and change-point analysis of rainfall in the arid district of ananthapuramu, Andhra Pradesh State, India. **Water (Switzerland)**, v. 12, n. 1, 1 jan. 2020.

PATAKAMURI, S. K.; O'BRIEN, N. **Package “modifiedmk”**: Modified Versions of Mann Kendall and Spearman’s Rho Trend Tests. 2020. Disponível em: <https://cran.r-project.org/web/packages/modifiedmk/index.html>. Acesso em: 25 ago. 2021.

PÉREZ-BLANCO, C. D.; GIL-GARCÍA, L.; SAIZ-SANTIAGO, P. An actionable hydroeconomic Decision Support System for the assessment of water reallocations in irrigated agriculture. A study of minimum environmental flows in the Douro River Basin, Spain. **Journal of Environmental Management**, v. 298, p. 113432, 15 nov. 2021.

POHLERT, T. **Package “trend”**: Non-Parametric Trend Tests and Change-Point Detection. 2020. Disponível em: <https://cran.r-project.org/web/packages/trend/>. Acesso em: 26 ago. 2021.

PRADO, R. B. *et al.* Diagnóstico do Meio Físico da Bacia Hidrográfica do Rio Muriaé. **EMBRAPA Solos**, v. 83, p. 75, 2005.

RÁPALO, L. M. C. *et al.* Effects of land-use and -cover changes on streamflow regime in the Brazilian Savannah. **Journal of Hydrology: Regional Studies**, v. 38, p. 100934, 1 dez. 2021.

RATNER, B. The correlation coefficient: Its values range between $+1$ / -1 , or do they? **Journal of Targeting, Measurement and Analysis for Marketing**, v. 17, p. 139–142, 2009.

ROCHA, J. *et al.* Impacts of climate change on reservoir water availability, quality and irrigation needs in a water scarce Mediterranean region (southern Portugal). **Science of The Total Environment**, v. 736, p. 139477, 20 set. 2020.

ROSA, M. R. *et al.* Hidden destruction of older forests threatens Brazil's Atlantic Forest and challenges restoration programs. **Science Advances**, v. 7, n. 4, p. 1–9, 2021.

SALVIANO, M. F. **Modelagem hidrológica da bacia do rio Muriaé com TOPMODEL, telemetria e sensoriamento remoto**. 2019. 163 f. Universidade de São Paulo, 2019.
Disponível em: <http://www.teses.usp.br/teses/disponiveis/14/14133/tde-30082019-133658/>.
Acesso em: 30 set. 2021.

SALVIANO, M. F.; PEREIRA-FILHO, A. J.; VEMADO, F. TOPMODEL Hydrometeorological Modeling with Rain Gauge Data Integrated by High-Resolution Satellite Estimates. A Case Study in Muriaé River Basin, Brazil. **Atmospheric and Climate Sciences**, v. 11, p. 486–507, 2021.

SEN, P. K. Estimates of the Regression Coefficient Based on Kendall's Tau. **Journal of the American Statistical Association**, v. 63, n. 324, p. 1379–1389, 1968.

SHI, R. *et al.* Streamflow decline threatens water security in the upper Yangtze river. **Journal of Hydrology**, v. 606, p. 127448, 1 mar. 2022.

SILVA, B. M. DA; SILVA, D. D. DA; MOREIRA, M. C. Influência da sazonalidade das vazões nos critérios de outorga de uso da água: estudo de caso da bacia do rio Paraopeba Influence of the seasonality of stream flows on water use rights criteria: a case study of the Paraopeba river basin. **Ambiente & Água - An Interdisciplinary Journal of Applied Science**, v. 10, n. 3, p. 624–634, 2015.

SILVA, L. A. DA *et al.* Vazões mínimas e de referência e rendimento específico para o estado de Minas Gerais. **Revista Brasileira de Ciências Agrárias**, v. 12, n. 4, p. 543–549, 2017.

SILVA, R. F. B. DA *et al.* Three decades of land-use and land-cover change in mountain regions of the Brazilian Atlantic Forest. **Landscape and Urban Planning**, v. 204, n. September, p. 103948, 2020. Disponível em:
<https://doi.org/10.1016/j.landurbplan.2020.103948>. Acesso em: 25 jan. 2022.

SIQUEIRA, P. P. *et al.* Effects of climate and land cover changes on water availability in a Brazilian Cerrado basin. **Journal of Hydrology: Regional Studies**, v. 37, p. 100931, 1 out. 2021.

SKANSI, M. DE LOS M. *et al.* Warming and wetting signals emerging from analysis of changes in climate extreme indices over South America. **Global and Planetary Change**, v. 100, 2013.

SOUZA, G. R. DE *et al.* Regional flood frequency analysis and uncertainties: Maximum streamflow estimates in ungauged basins in the region of Lavras, MG, Brazil. **CATENA**, v. 197, p. 104970, 1 fev. 2021.

SUGAHARA, S. *et al.* Homogeneity assessment of a station climate series (1933-2005) in the Metropolitan Area of São Paulo: Instruments change and urbanization effects. **Theoretical and Applied Climatology**, v. 107, p. 361–374, 2012.

SWAIN, S. *et al.* Exploring recent groundwater level changes using Innovative Trend Analysis (ITA) technique over three districts of Jharkhand, India. **Groundwater for Sustainable Development**, v. 18, p. 100783, 1 ago. 2022.

THEIL, H. A rank-invariant method of linear and polynomial regression analysis III. **Proceedings Nederlandse Akademie van Wetenschappen**, v. 53, p. 1397–1412, 1950.

THIESSEN, A. H. Precipitation average for large areas. **Monthly Weather Review**, v. 39, n. 7, p. 1082–1089, 1911.

THOM, H. C. S. Some methods of climatological analysis. **World Meteorological Organization- WMO. Technical note No. 81, WMO-No. 199. TP. 103**, v. 81, p. 53, 1966.

TUCCI, C. E. M. **Hidrologia: ciência e aplicação**. 4. ed. Porto Alegre: Editora da UFRGS/ABRH, 2012.

UDDIN, K.; MATIN, M. A. Potential flood hazard zonation and flood shelter suitability mapping for disaster risk mitigation in Bangladesh using geospatial technology. **Progress in Disaster Science**, v. 11, p. 100185, 2021.

ULIANA, E. M. *et al.* Análise de tendência em séries históricas de vazão e precipitação: uso de teste estatístico não paramétrico. **Ambiente e Agua - An Interdisciplinary Journal of Applied Science**, v. 10, n. 1, p. 82–88, 1 jan. 2015.

VOISIN, N. *et al.* Impact of climate change on water availability and its propagation through the Western U.S. power grid. **Applied Energy**, v. 276, p. 115467, 15 out. 2020.

WICKHAM, H.; FRANÇOIS, R.; *et al.* **Package “dplyr”**: A Grammar of Data Manipulation. 2022. Disponível em: <https://cran.r-project.org/web/packages/dplyr/>. Acesso em: 26 fev. 2022.

WICKHAM, H.; CHANG, W.; *et al.* **Package “ggplot2”**: Create Elegant Data Visualisations Using the Grammar of Graphics. 2022. Disponível em: <https://cran.r-project.org/web/packages/ggplot2/>. Acesso em: 26 fev. 2022.

WICKHAM, H. **Package “tidyverse”**: Easily Install and Load the “Tidyverse”. p. 1–6, 2021. Disponível em: <https://cran.r-project.org/web/packages/tidyverse/>. Acesso em: 26 fev. 2022.

WWAP. The United Nations world water development report 2018: **Nature-Based Solutions for Water**. [S.l.: s.n.], 2018. Disponível em: www.unwater.org/publications/%0Aworld-water-development-report-2018/. Acesso em: 28 fev. 2022.

XU, X. *et al.* Attribution analysis based on the Budyko hypothesis for detecting the dominant cause of runoff decline in Haihe basin. **Journal of Hydrology**, v. 510, p. 530–540, 14 mar. 2014.

YIRSAW, E. *et al.* Land Use/Land Cover Change Modeling and the Prediction of Subsequent Changes in Ecosystem Service Values in a Coastal Area of China, the Su-Xi-Chang Region. **Sustainability**, v. 9, p. 1204, 2017. Disponível em: www.mdpi.com/journal/sustainability. Acesso em: 20 mar. 2022.

YUE, S. *et al.* The influence of autocorrelation on the ability to detect trend in hydrological series. **Hydrological Processes**, v. 16, n. 9, p. 1807–1829, 30 jun. 2002.

ZHANG, A. *et al.* Analysis of streamflow variations in the Heihe River Basin, northwest China: Trends, abrupt changes, driving factors and ecological influences. **Journal of Hydrology: Regional Studies**, v. 3, p. 106–124, 1 mar. 2015.

ZHAO, G. *et al.* Quantifying the impact of climate variability and human activities on streamflow in the middle reaches of the Yellow River basin, China. **Journal of Hydrology**, v. 519, n. PA, p. 387–398, 7 nov. 2014.

CAPÍTULO 2

Modelagem da vazão por meio de técnicas de Aprendizado de Máquina em diferentes horizontes de previsão

RESUMO: A ocorrência de eventos extremos, principalmente as vazões máximas, tem ocasionado imensos impactos socioambientais, tanto em áreas urbanas como em áreas rurais. Nesse cenário, torna-se necessária a adoção de modelos de previsão da vazão que auxiliem nas tomadas de decisão. Modelos de base física costumam ser utilizados para esse objetivo, mas apresentam limitações no que diz respeito à quantidade de parâmetros e possibilidades de incertezas no processo preditivo. Para a superação dos obstáculos encontrados nesse tipo de modelagem, as técnicas de Aprendizado de Máquina surgem como uma importante alternativa para a previsão da vazão, funcionando como uma ferramenta importante nos sistemas de alerta de enchentes. Uma das principais vantagens do Aprendizado de Máquina está na possibilidade do aumento do horizonte de previsão, de modo que se objetivou neste trabalho analisar distintas técnicas de Aprendizado de Máquina na predição de vazões na bacia hidrográfica do rio Muriaé, adotando-se diferentes cenários de antecedência. Foram utilizadas seis categorias de dados: vazão, precipitação, sensoriamento remoto, balanço hídrico, uso e cobertura da terra e tempo. Os dados corresponderam ao período de 2000 a 2019 e inicialmente passaram por análise de estacionariedade, seguida da padronização. Foram adotados cenários de um, dois e três dias de antecedência para fins de previsão das vazões. Os algoritmos de Aprendizado de Máquina aplicados foram o Modelo Linear Generalizado (GLM), as *Splines* de Regressão Adaptativa Multivariada (MARS) e o *Random Forest* (RF). Na seleção de recursos, utilizou-se a função *findcorrelation* e o algoritmo *Recursive Feature Elimination* (RFE) ajustado a cada técnica para a eliminação de covariáveis altamente correlacionadas e pouco informativas, respectivamente. A amostragem de treinamento e teste seguiu a proporção 75:25, com amostragem aleatória dos dados e 50 repetições. O desempenho da modelagem foi verificado por meio das métricas: raiz do Erro Quadrático Médio (RMSE), Erro Médio Absoluto (MAE), Coeficiente de Determinação (R^2) e Eficiência de Nash-Sutcliffe (NSE). As técnicas de Aprendizado de Máquina apresentaram desempenho ótimo no cenário de um dia de antecedência, mas a performance diminuiu com o aumento do prazo de entrega da previsão. Em todos os cenários, o algoritmo RF apresentou melhor concordância com os dados observados, embora tenha subestimado picos de vazão. A modelagem com GLM, MARS e RF mostrou a aplicabilidade desses algoritmos na previsão da vazão em bacias hidrográficas. Destacadamente na bacia do rio Muriaé, com recorrentes eventos de elevada magnitude, a possibilidade de aumentar os dias

de antecedência na previsão da vazão permite minimizar expressivamente os impactos socioambientais decorrentes da ocorrência de eventos extremos de vazão.

Palavras-chave: Previsão de vazão. Sistema de alerta. Vazões máximas.

2.1 INTRODUÇÃO

Nos últimos anos diversas cidades brasileiras passaram por eventos de inundação, que resultaram em perdas econômicas e, mais lamentavelmente, humanas. Os eventos com vazões acima do fluxo normal do rio são conhecidos como enchentes e precedem a inundação. As enchentes e inundações constituem alguns dos principais desastres naturais que ocorrem com elevada frequência e intensidade em várias partes do planeta (KAO *et al.*, 2020; NACHAPPA *et al.*, 2020).

A intensidade dos efeitos desses desastres está relacionada às mudanças climáticas, às falhas nos sistemas de drenagem, à redução da capacidade de infiltração do solo, especialmente em áreas urbanas, e ao crescimento populacional desordenado, que resulta em construções habitacionais em áreas de risco (DARABI *et al.*, 2019).

Ainda que as enchentes possam promover alguns benefícios, como a restauração de pântanos, estes são superados por seus múltiplos impactos negativos (PHAM *et al.*, 2021). Inúmeros são os danos causados por esses eventos, sejam de ordem ambiental e/ou socioeconômica, que levam à perda de vidas humanas e propriedades (CHEN *et al.*, 2021). Este cenário requer a realização de estudos em busca de metodologias de previsão de cheias que permitam a tomada de ações para minimizar os impactos à sociedade.

Diversos modelos hidrológicos de base física foram desenvolvidos para a previsão de vazão e simulação do comportamento hidrológico sob os efeitos das atividades humanas e mudanças climáticas (e.g. CHENG, M. *et al.*, 2020; TONGAL; BOOIJ, 2018). Todavia, para a aplicação desses modelos são necessários muitos parâmetros referentes aos processos hidrológicos que ocorrem na bacia hidrográfica, além de apresentarem como fator limitante a não linearidade dos dados, o que leva a fontes de incerteza no processo preditivo (ATIKUZZAMAN; KANDASAMY, 2018).

Alternativamente, diversos autores têm utilizado técnicas de Aprendizado de Máquina (*Machine Learning*, ML), as quais utilizam algoritmos aplicados às bases de dados para a previsão da vazão, a curto e longo prazos, em razão do seu alto desempenho frente às limitações dos modelos de base física (CHENG *et al.*, 2020; FATHIAN *et al.*, 2019; MOSAVI; OZTURK; CHAU, 2018). Dentre alguns algoritmos de Aprendizado de Máquina comumente utilizados na

previsão estão o Modelo Linear Generalizado (GLM) (NELDER; WEDDERBURN, 1972), as *Splines* de Regressão Adaptativa Multivariada (MARS) (FRIEDMAN, 1991) e o *Random Forest* (RF) (BREIMAN, 2001).

Na aplicação dos algoritmos de Aprendizado de Máquina, um conjunto de dados referente às variáveis de estudo gera um subconjunto de covariáveis ambientais. Essas variáveis devem ser selecionadas com base em critérios para a otimização do modelo preditivo, por meio de algoritmos de seleção de variáveis (FERREIRA *et al.*, 2021b; MIAO; NIU, 2016). A seleção desse subconjunto de variáveis possibilita a construção de modelos mais simples e compreensíveis, além de terem um menor esforço computacional no processamento dos dados (LI *et al.*, 2017a).

Entretanto, mesmo ao superar as limitações da dimensão dos dados, não existem muitos estudos que utilizem métodos de Aprendizado de Máquina para a modelagem da vazão em diferentes horizontes de previsão em bacias tropicais. Nesse contexto, conhecer o desempenho de diferentes técnicas em prazos de entrega diferentes pode auxiliar as autoridades no aperfeiçoamento dos sistemas de previsão de cheias, que embora tenham passado por melhorias, ainda não conseguem evitar os danos causados pelas enchentes recorrentes a cada período chuvoso (SILVESTRO *et al.*, 2019), como as que ocorrem na região sudeste do Brasil (CAMPOS; PAZ, 2020).

Nessa região, uma das bacias que enfrenta de forma frequente os problemas relacionados às enchentes e inundações é a do rio Muriaé, localizada nos estados de Minas Gerais e Rio de Janeiro (AMORIM; REIS; FERREIRA, 2017). Esta bacia, além dos eventos naturais de cheia, sofreu com inundações provenientes do rompimento de uma barragem de rejeitos de mineração no ano de 2007, o que motivou a implementação do Sistema de Alerta Hidrológico da bacia do rio Muriaé (SAH-Muriaé), operado pelo Serviço Geológico do Brasil (CPRM) a partir de 2014 (CPRM, 2021).

Esse sistema prevê o nível da água em tempo real por meio de modelagem linear, com 10 horas de antecedência, estendendo-se até 24 horas, com base nos dados de vazão da estação fluviométrica localizada a montante (CPRM, 2021). Sistemas de alerta precoce são de grande utilidade na tomada de ações preventivas e mitigadoras contra as cheias (SIQUEIRA *et al.*, 2020), mas a depender da magnitude da inundação, a disponibilidade de dados de algumas estações pode ficar comprometida para a geração de alerta (CPRM, 2021).

Dessa forma, estudar métodos de Aprendizado de Máquina na predição de vazões máximas que apresentem confiabilidade torna-se crucial para um melhor planejamento da bacia hidrográfica, principalmente quanto aos eventos máximos (KAO *et al.*, 2020), de modo que o

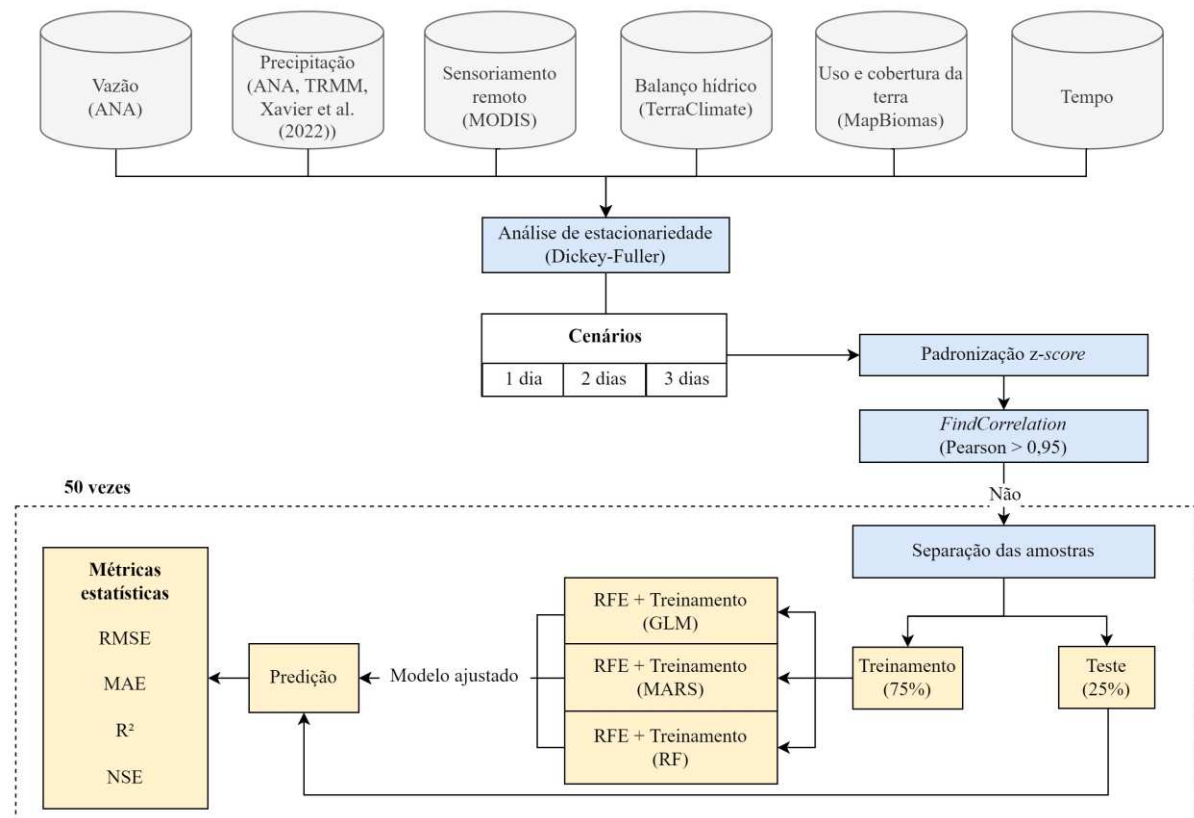
presente capítulo teve por objetivo analisar diferentes técnicas de Aprendizado de Máquina na predição de vazões nos cenários de um, dois e três dias de antecedência na bacia hidrográfica do rio Muriaé, a fim de auxiliar na superação das limitações do sistema atual, de modo a minimizar os impactos gerados por eventos de elevada magnitude.

2.2 MATERIAL E MÉTODOS

2.2.1 Desenvolvimento da modelagem com Aprendizado de Máquina

A predição da vazão foi realizada por meio de três técnicas de Aprendizado de Máquina (*Machine Learning*, ML): Modelo Linear Generalizado (GLM), *Splines* de Regressão Adaptativa Multivariada (MARS) e o *Random Forest* (RF), sendo considerados três cenários de prazos na modelagem: um, dois e três dias de antecedência (Y_{t+1} ; Y_{t+2} ; Y_{t+3}) na estação mais próxima à foz. O fluxograma adotado na modelagem é apresentado na Figura 2.1.

Figura 2.1 – Fluxograma da modelagem para predição da vazão em um, dois e três dias de antecedência utilizando técnicas de Aprendizado de Máquina.



Fonte: A autora.

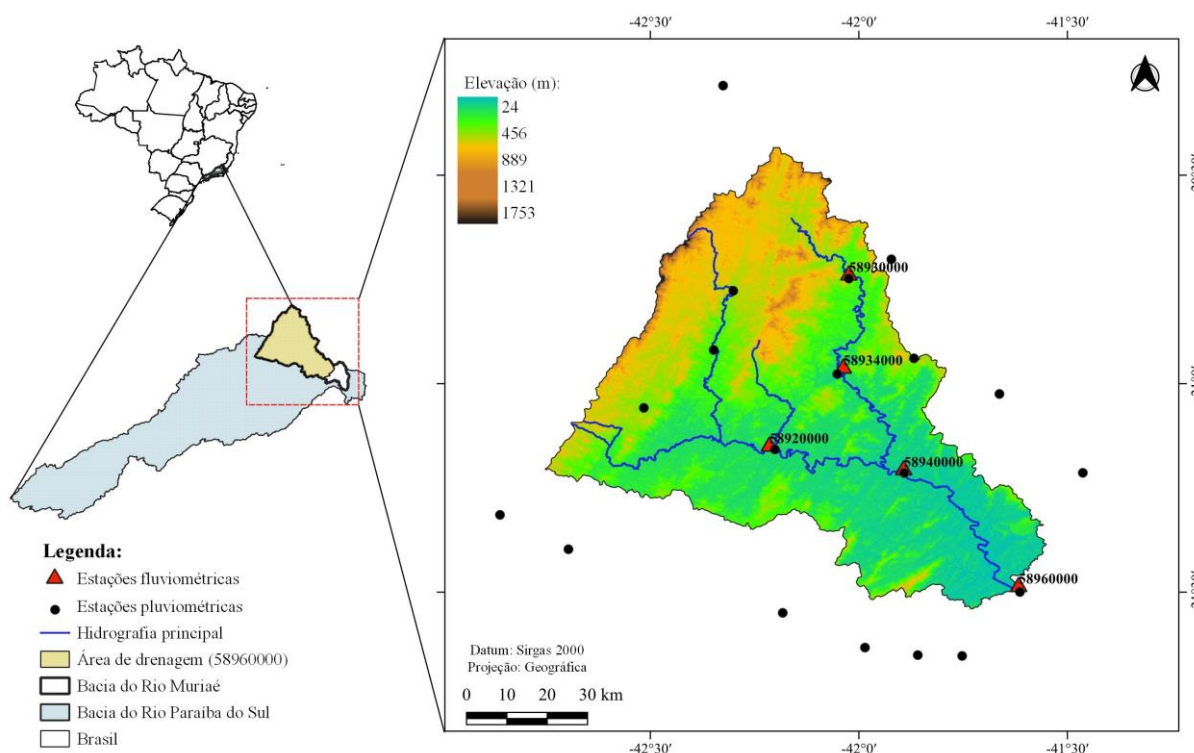
O período da modelagem foi de 2000 a 2019, sendo utilizadas seis categorias de dados, as quais passaram por análise de estacionariedade e padronização. Na seleção das covariáveis foram utilizados dois métodos de exclusão de covariáveis: eliminação de uma covariável quando a correlação de Pearson fosse acima de 0,95 por meio da função *findcorrelation* (KUHN *et al.*, 2022) e o algoritmo *Recursive Feature Elimination* (RFE) para a seleção dos melhores preditores para cada modelo (FERDINANDY *et al.*, 2020).

O conjunto de dados foi dividido em duas partes, sendo 75% para o treinamento e 25% para o teste, sendo duas validações utilizadas: interna com 10-*folds* e três repetições e a validação externa *holdout*. Posteriormente, os resultados foram avaliados por meio de quatro métricas: Raiz Quadrada do Erro Médio (RMSE), Erro Médio Absoluto (MAE), Coeficiente de Determinação (R^2) e o índice da eficiência de Nash-Sutcliffe (NSE).

2.2.2 Área de estudo

O estudo foi realizado na bacia hidrográfica do rio Muriaé, especificamente na área de drenagem da estação fluviométrica localizada mais próxima à sua foz, estação Cardoso Moreira (58960000), com área de drenagem de 2.710 km² (Figura 2.2). Essa bacia consiste em um dos afluentes do rio Paraíba do Sul, com sua nascente em Minas Gerais e foz no Rio de Janeiro (CPRM, 2021).

Figura 2.2 - Localização da área de drenagem da estação fluviométrica 58960000, próximo à foz da bacia hidrográfica do Rio Muriaé.



Fonte: A autora.

A área de drenagem considerada apresenta terreno montanhoso, com altitude entre 24 e 1.753 m. A precipitação média mensal da bacia varia entre 22,9 mm e 281,7 mm, com período chuvoso entre os meses de novembro a abril, enquanto o período seco começa em maio e termina em outubro. O clima predominante na bacia é tropical de savana, com inverno seco (Aw) (ALVARES *et al.*, 2013). Quanto aos solos, os principais são os Latossolos Vermelho-Amarelos Distróficos e Argissolos Vermelho-Amarelos Eutróficos (EMBRAPA SOLOS, 2020).

2.2.3 Base de dados para a modelagem

Nesse trabalho consideraram-se seis categorias de dados: vazão dos rios afluentes, precipitação, sensoriamento remoto, dados derivados da evapotranspiração de referência, uso e cobertura da terra e tempo. A partir dessas categorias, originaram-se covariáveis com potencial para serem preditoras da vazão na estação Cardoso Moreira (58960000), o que resultou em 59 covariáveis.

Foram utilizados dados de vazão diária dos rios afluentes correspondentes às estações fluviométricas do Sistema de Alerta Hidrológico da bacia do rio Muriaé (SAH – Muriaé)

disponíveis no Portal Hidroweb (<https://www.snirh.gov.br/hidroweb/serieshistoricas>), da Agência Nacional de Águas e Saneamento Básico (ANA).

A vazão da estação fluviométrica mais próxima à foz da bacia do rio Muriaé, Cardoso Moreira (58960000), foi definida como a variável dependente (Y), enquanto as vazões das estações a montante compuseram o conjunto de covariáveis (Tabela 2.1).

Considerando que as informações passadas de variáveis hidroclimáticas de uma bacia determinam o seu comportamento no futuro, sendo essenciais no estudo de desastres naturais como enchentes e inundações (GIRIHAGAMA *et al.*, 2022), as covariáveis de vazão foram consideradas nos tempos t , $t - 1$ e $t - 2$, em que t é o dia corrente, levando-se em consideração o tempo de concentração até a estação 58960000, o qual é de aproximadamente 50,33 horas (ANA, 2011)

Tabela 2.1 – Características das estações fluviométricas utilizadas na modelagem de vazões na bacia do rio Muriaé.

Código	Estação	Latitude (°)	Longitude (°)	Área de drenagem (km ²)	Curso d'água
58920000	Patrocínio do Muriaé (Q1)	-21,15	-42,21	2.660	Rio Muriaé
58930000	Carangola (Q2)	-20,74	-42,02	773	Rio Carangola
58934000	Porciúncula (Q3)	-20,96	-42,04	1.340	Rio Carangola
58940000	Itaperuna (Q4)	-21,21	-41,89	5.800	Rio Muriaé
58960000	Cardoso Moreira (Y)	-21,49	-41,62	7.210	Rio Muriaé

Fonte: A autora.

Os dados diários de chuva foram provenientes das estações pluviométricas disponíveis no Portal Hidroweb (<https://www.snirh.gov.br/hidroweb/serieshistoricas>), sendo selecionadas por meio de Polígonos de Thiessen (THIESSEN, 1911), de modo que estações no entorno da bacia foram consideradas, desde que suas áreas de influência sobrepusessem a área da bacia. A precipitação das estações selecionadas foi interpolada pelo método Inverso do Quadrado da Distância (IDW), utilizando o pacote *gstat* (PEBESMA; GRAELER, 2022) do *software* R 4.0.4 (R CORE TEAM, 2020), sendo posteriormente calculada a média para a área de drenagem da estação Cardoso Moreira (58960000).

Para a composição das covariáveis de precipitação também foram utilizados dados diários médios dos sensores *Tropical Rainfall Measuring Mission* (TRMM) da *National Aeronautics and Space Administration* (NASA), contidos no produto *3B42: 3-Hourly*

Precipitation Estimates com resolução de 27.830 metros (TRMM, 2011), que foram agregados em dados diários no *Google Earth Engine* (GEE).

Além disso, foi inserida a base diária de precipitação de Xavier *et al.* (2022) na resolução $0,1^\circ \times 0,1^\circ$ para o Brasil, em que foi calculada a média dos valores para a área de drenagem da estação considerada como variável dependente (58960000). Para os dados de chuva seguiu-se a mesma defasagem dos dados de vazão, em que t , $t - 1$ e $t - 2$ foram o dia corrente e os atrasos em 1 e 2 dias, respectivamente, das covariáveis de precipitação, sendo considerados também os valores acumulados em 48 horas, 72 horas, 96 horas, 120 horas e 1 semana no tempo t .

Quanto aos dados de sensoriamento remoto, consideraram-se os índices de vegetação e séries de temperatura do sensor MODIS obtidos a partir de dois produtos: MOD13A1 (DIDAN, 2021) e MOD11A2 (WAN; HOOK; HULLEY, 2021). O primeiro agrega 16 dias do sensor MODIS a bordo do satélite Terra, sendo obtidos os índices *Normalized Difference Vegetation Index* (NDVI) e o *Enhanced Vegetation Index* (EVI). O segundo produto forneceu dados sobre a média da temperatura de 8 dias. Ambos os produtos são da versão 6.1 e apresentam 500 metros de resolução.

Na base de dados também foram incluídas covariáveis mensais do balanço hídrico derivadas da evapotranspiração de referência calculada pelo método ASCE Penman-Montieth: escoamento superficial e umidade do solo. Essas variáveis foram obtidas do conjunto de dados climáticos TerraClimate (<https://www.climatologylab.org/terraclimate.html>) com resolução de $0,5^\circ \times 0,5^\circ$ por meio da plataforma GEE.

Os dados de uso e cobertura da terra foram obtidos da coleção 6.0 do MapBiomas (MAPBIOMAS, 2021), por meio do repositório do GEE. Foram consideradas as porcentagens anuais das seguintes classes: Floresta (FL), Formação Natural não Florestal (FN), Pastagem (PA), Floresta Plantada (FP), Agricultura (AG), Área não Vegetada (AnV) e Corpos Hídricos (CH). Para essas covariáveis, consideraram-se as porcentagens de cada ano da vazão a ser predita, ou seja, sem defasagem.

Considerou-se, também, a periodicidade como variável de entrada, sendo representada pelos meses do ano, sendo utilizados os valores observados no mês da vazão a ser predita.

Diante dessas covariáveis, foram selecionados três cenários com previsão de um, dois e três dias de antecedência, seguindo como critério o tempo de concentração, igual a 50,33 horas, até a estação 58960000 (ANA, 2011). Assim, a modelagem poderá ser útil para os sistemas de alerta hidrológico da bacia, sendo a previsão da vazão na estação Y nos tempos $t + 1$, $t + 2$ e $t + 3$, correspondendo aos cenários 1, 2 e 3, respectivamente.

Todas as covariáveis de entrada, para cada um dos cenários de antecedência adotados, encontram-se na Tabela 2.2 e as medidas descritivas (média, mediana, mínimo, máximo, desvio padrão e coeficiente de variação) estão apresentadas no Apêndice D.

Tabela 2.2 – Covariáveis usadas como possíveis preditores da vazão na estação Cardoso Moreira (58960000) em três diferentes cenários de antecedência na previsão.

Cenários	Vazão	Precipitação	Sensoriamento remoto	Balanço hídrico	Uso e cobertura da terra	Tempo
$Y_{t+1}; Y_{t+2}; Y_{t+3}$	$Q1_t$	p_idw_t	$temp_min_t$	$escoam_min_t$	FL	mês
	$Q1_{t-1}$	p_idw_{t-1}	$temp_med_t$	$escoam_med_t$	FN	
	$Q1_{t-2}$	p_idw_{t-2}	$temp_max_t$	$escoam_max_t$	PA	
	$Q2_t$	$p_idw_{48h_t}$	evi_min_t	$umidade_min_t$	AG	
	$Q2_{t-1}$	$p_idw_{72h_t}$	evi_med_t	$umidade_med_t$	FP	
	$Q2_{t-2}$	$p_idw_{96h_t}$	evi_max_t	$umidade_max_t$	AnV	
	$Q3_t$	$p_idw_{120h_t}$	$ndvi_min_t$		CH	
	$Q3_{t-1}$	$p_idw_{1sem_t}$	$ndvi_med_t$			
	$Q3_{t-2}$	p_xavier_t	$ndvi_max_t$			
	$Q4_t$	p_xavier_{t-1}				
	$Q4_{t-1}$	p_xavier_{t-2}				
	$Q4_{t-2}$	$p_xavier_{48h_t}$				
		$p_xavier_{72h_t}$				
		$p_xavier_{96h_t}$				
		$p_xavier_{120h_t}$				
		$p_xavier_{1sem_t}$				
		p_trmm_t				
		p_trmm_{t-1}				
		p_trmm_{t-2}				
		$p_trmm_{48h_t}$				
		$p_trmm_{72h_t}$				
		$p_trmm_{96h_t}$				
		$p_trmm_{120h_t}$				
		$p_trmm_{1sem_t}$				

Cenários: vazão a ser prevista na estação 58960000 (Y) ($t + 1$, $t + 2$ e $t + 3 = 1$, 2 e 3 dias de antecedência); **Vazões diárias dos rios afluentes:** vazão da estação 58920000 no dia corrente ($Q1_t$), vazão da estação 58930000 no dia corrente ($Q2_t$), ..., vazão na estação 58940000 no dia corrente ($Q4_t$) ($t - 1$; $t - 2 = 1$ e 2 dias de defasagem); **Precipitação:** precipitação diária na área de drenagem da estação 58960000 interpolada pelo inverso da distância ao quadrado no dia corrente (p_idw_t), calculada por Xavier *et al.* (2022) no dia corrente (p_xavier_t) e obtidas do sensor *Tropical Rainfall Measuring Mission* (TRMM) no dia corrente (p_trmm_t); ($t - 1$; $t - 2 = 1$ e 2 dias de defasagem; $48h_t$, ..., $1sem_t$ = acumulado de 48 horas no dia corrente, ..., acumulado de 1 semana no dia corrente); **Sensoriamento remoto:** temperatura ($temp$), *Normalized Difference Vegetation Index* ($ndvi$), *Enhanced Vegetation Index* (evi) (min_t , med_t , max_t = mínimo, médio e máximo no dia corrente); **Balanço hídrico:**

(cont.)

escoamento superficial (*escoam*) e umidade do solo (*umidade*) (min_t , med_t , max_t = mínimo, médio e máximo no dia corrente); **Uso e cobertura da terra:** FL – Floresta; FN – Formação Natural não Florestal; PA – Pastagem; AG – Agricultura; FP – Floresta Plantada; AnV – Área não Vegetada e CH – Corpos Hídricos; **Tempo:** periodicidade da série em meses do ano (mês). Fonte: A autora.

2.2.4 Análise de estacionariedade das séries temporais

Na modelagem de séries temporais, a não estacionariedade dos dados indica mudanças ao longo do tempo de suas métricas estatísticas como média e variância, sendo caracterizada como uma propriedade importante no desempenho dos modelos preditivos (LIVIERIS *et al.*, 2021). Dessa forma, aplicou-se o teste aumentado de Dickey-Fuller em todas as covariáveis, a um nível de significância de 5%, no qual três raízes unitárias foram testadas por meio do pacote *aTSA* (QIU, 2015).

Nesse trabalho, o teste aumentado de Dickey-Fuller indicou que todas covariáveis são estacionárias, ou seja, as séries temporais criadas a partir de processos estocásticos são constantes ao longo do período analisado, não interferindo nas inferências estatísticas dos modelos utilizados (COLLAR *et al.*, 2022; SILVEIRA *et al.*, 2022a).

2.2.5 Padronização das amostras

Todas as covariáveis citadas anteriormente passaram por uma padronização *Z –score*, na qual a média foi igualada a zero e o desvio padrão a um, a fim de equiparar as ordens de grandeza e evitar interferências nos resultados. A padronização foi calculada pela Equação 2.4 (AHANI; SHOURIAN; RAD, 2018).

$$Z = \frac{x - \bar{x}}{\sigma} \quad (2.4)$$

em que:

- Z = variável preditora padronizada;
- $\bar{x}$ = média amostral da variável preditora; e
- σ = desvio padrão da variável preditora.

2.2.6 Algoritmos de Aprendizado de Máquina para a previsão da vazão

2.2.6.1 Modelo Linear Generalizado (GLM)

O Modelo Linear Generalizado (*Generalized Linear Model*, GLM) apresenta como base a relação entre uma regressão logística dos preditores com a variável dependente (AVAND *et al.*, 2022). Nessa técnica não há hiperparâmetros a serem ajustados. Diante disso, a modelagem

com o GLM utilizou o pacote *caret* (KUHNS *et al.*, 2022) do *software* R 4.0.4 (R CORE TEAM, 2020), sendo escolhido o modelo com o melhor desempenho com base no menor Erro Médio Absoluto (MAE).

2.2.6.2 *Splines* de Regressão Adaptativa Multivariada (MARS)

As *Splines* de Regressão Adaptativa Multivariada (*Multivariate Adaptive Regression Splines*, MARS) são uma técnica não linear e não paramétrica utilizada em previsões que têm um parâmetro contínuo (FRIEDMAN, 1991). Esse parâmetro, representado pela variável dependente (estação 58960000), foi relacionado aos preditores por meio de funções básicas que melhoram o ajuste dos dados a depender de funções de segmento (*spline*) unidas por nós que dividem o conjunto de dados em partes (ADNAN *et al.*, 2020). Esse algoritmo é controlado pelos hiperparâmetros *nprune* (nós) e *degree* (interações), sendo vários modelos criados para diferentes arranjos, em que foi selecionado o modelo com o menor MAE. Nessa técnica foi utilizado o pacote *earth* (MILBORROW, 2022) do *software* R 4.0.4 (R CORE TEAM, 2020).

2.2.6.3 *Random Forest* (RF)

Desenvolvido por Breiman (2001), o *Random Forest* (RF) faz a previsão da variável dependente por meio da criação de várias árvores, em que cada uma aprende com as amostras e recursos dos preditores selecionados de forma aleatória, sendo o resultado final a média das previsões de cada árvore criada (ZAHURA; GOODALL, 2022). Para o RF, utilizou-se o pacote *caret* (KUHNS *et al.*, 2022) do *software* R 4.0.4 (R CORE TEAM, 2020). Nessa técnica, o hiperparâmetro é o número de variáveis selecionadas aleatoriamente (*mtry*) para cada árvore, de modo que foram construídos modelos para diferentes *mtry* e selecionado o que apresentou o menor MAE.

2.2.7 Seleção de covariáveis

Para selecionar o melhor conjunto de preditores em cada cenário, determinou-se a correlação de Pearson para as covariáveis. Após a determinação das correlações, o conjunto de dados foi analisado pela função *findcorrelation* do pacote *caret* (KUHNS *et al.*, 2022), disponível no *software* R 4.0.4 (R CORE TEAM, 2020), em que para cada par de covariáveis com correlação superior a 95% uma covariável foi eliminada caso apresentasse maior média absoluta (DIENGDOH *et al.*, 2022). Nessa etapa foram eliminadas 13 covariáveis, resultando em 46 covariáveis para a próxima etapa.

Após essa etapa, aplicou-se o algoritmo de seleção de variáveis *Recursive Feature Elimination* (RFE), com validação cruzada interna com 10-*folds* e uma repetição, observando a métrica de desempenho MAE, que selecionou regressivamente os melhores preditores para cada modelo a fim de reduzir o número de covariáveis e possibilitar a criação de modelos parcimoniosos (FERDINANDY *et al.*, 2020; KUHN; JOHNSON, 2013).

Considerando que as técnicas abordadas nesse trabalho foram Modelo Linear Generalizado (GLM), *Splines* de Regressão Adaptativa Multivariada (MARS) e o *Random Forest* (RF), utilizou-se na seleção dos preditores duas funções do pacote *caret* (KUHN *et al.*, 2022): *lmFuncs* para GLM e MARS e *rfFuncs* para RF, para cada cenário de previsão ($Y_{t+1}; Y_{t+2}; Y_{t+3}$) com o número de covariáveis regredindo de 46 a 1. Esse processamento foi realizado 50 vezes para capturar a variabilidade dos dados e fornecer resultados confiáveis. A importância das covariáveis foi organizada hierarquicamente pela função *VarImp*, pertencente ao pacote *caret* do *software* R (KUHN; JOHNSON, 2013). O algoritmo RFE foi aplicado de forma integrada com o treinamento e teste, que se encontram na seção 2.2.8.

2.2.8 Treinamento e teste

Após a seleção dos preditores, realizou-se, iterativamente, 50 processamentos para as fases de treinamento e teste com o melhor grupo de covariáveis para cada modelo considerando cada cenário de previsão ($Y_{t+1}; Y_{t+2}; Y_{t+3}$), de modo a serem construídos 450 modelos em cada fase, sendo 50 modelos para cada técnica em cada horizonte de previsão. Para o treinamento e teste, adotou-se a validação externa *holdout*, em que houve a seleção aleatória de amostras, sendo destinados 75% e 25% das observações, respectivamente, para essas duas fases da modelagem (DIENGDOH *et al.*, 2022; SOUZA *et al.*, 2022).

No treinamento, os hiperparâmetros de cada modelo, exceto GLM que não apresenta hiperparâmetros, foram otimizados no *software* R 4.0.4 (R CORE TEAM, 2020) com validação cruzada interna *k-folds* ($k=10$) com três repetições pelo pacote *caret* (KUHN *et al.*, 2022), mesma abordagem de Veloso, Rodrigues e Filho (2022).

Após o treinamento dos modelos, os dados do conjunto de teste foram usados para determinar as métricas de desempenho da modelagem, de modo que os dados fossem submetidos a diversos modelos ajustados que realizam as previsões para cada amostra. Os dados preditos são comparados com os dados observados e as métricas de desempenho (RMSE, MAE, R^2 e NSE) são calculadas. Essa etapa determina a performance dos modelos em dados que não foram usados no treinamento e medem a capacidade que o modelo tem para generalizar.

2.2.9 Critérios de avaliação da modelagem

2.2.9.1 Métricas estatísticas

Na avaliação das previsões obtidas pelos modelos de GLM, MARS e RF foram calculadas as seguintes métricas estatísticas: Raiz do Erro Quadrático Médio (RMSE) (Equação 2.5), Erro Médio Absoluto (MAE) (Equação 2.6), Coeficiente de Determinação (R^2) (Equação 2.7) e Eficiência de Nash-Sutcliffe (NSE) (Equação 2.8). Essas métricas também foram utilizadas por Nguyen *et al.* (2022) e podem ser expressas pelas seguintes equações:

$$RMSE = \sqrt{n^{-1} \times \left[\sum_{i=1}^n (q_{obs(i)} - q_{sim(i)})^2 \right]} \quad (2.5)$$

$$MAE = \frac{1}{n} \times \left[\sum_{i=1}^n |q_{obs(i)} - q_{sim(i)}| \right] \quad (2.6)$$

$$R^2 = 1 - \left[\frac{\frac{1}{n} \sum_{i=1}^n (q_{sim(i)} - q_{obs(i)})^2}{\frac{1}{n} \sum_{i=1}^n (q_{sim(i)} - \bar{q}_{obs})^2} \right] \quad (2.7)$$

$$NSE = 1 - \left[\frac{\sum_{i=1}^n (q_{sim(i)} - q_{obs(i)})^2}{\sum_{i=1}^n (q_{obs(i)} - \bar{q})^2} \right] \quad (2.8)$$

em que:

q_{obs} = vazão observada; e

q_{sim} = vazão simulada;

A escolha dos melhores modelos teve como referência os menores erros (RMSE e MAE) e os critérios de avaliação de desempenho definidos por Moriasi *et al.* (2015), em que os modelos com R^2 e NSE superiores a 0,85 e 0,80, respectivamente, são classificados como muito bons (Tabela 2.3).

Tabela 2.3 – Classificação das métricas de desempenho dos modelos de Aprendizado de Máquina

Métrica	Muito bom	Bom	Satisfatório	Insatisfatório
R^2	$R^2 > 0,85$	$0,75 < R^2 \leq 0,85$	$0,60 < R^2 \leq 0,75$	$R^2 \leq 0,60$
NSE	$NSE > 0,80$	$0,70 < NSE \leq 0,80$	$0,50 < NSE \leq 0,70$	$NSE \leq 0,50$

Fonte: adaptado de Moriasi *et al.* (2015)

As métricas estatísticas RMSE, MAE e R^2 foram calculadas pelo pacote *labgeo* (FERNANDES-FILHO, 2022), enquanto a NSE foi calculada pelo pacote *hydroGOF* (ZAMBRANO-BIGIARINI, 2020). Os algoritmos de Aprendizado de Máquina foram implementados por meio dos pacotes do *software* R 4.0.4 (R CORE TEAM, 2020).

2.2.9.2 Vazões de referência do SAH-Muriaé

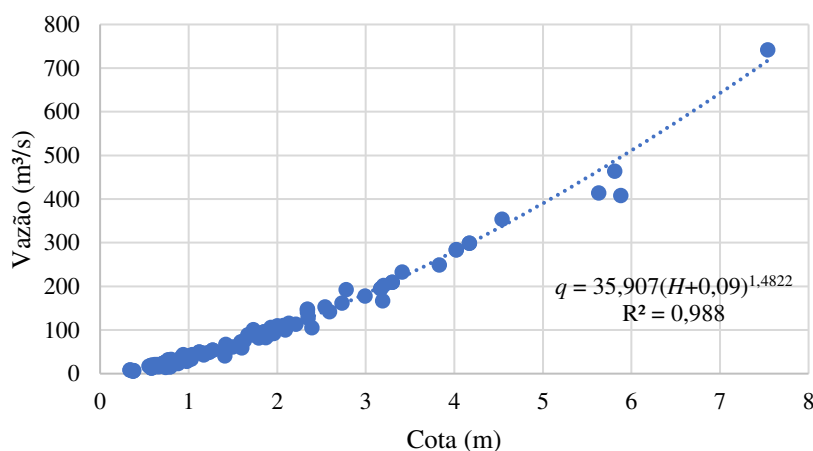
Para conhecer as vazões de alerta do SAH-Muriaé e comparar com as vazões simuladas pelos modelos de GLM, MARS e RF, fez-se a curva-chave da estação Cardoso Moreira (58960000), com os dados de cota e vazão observadas entre 2000 e 2019 disponíveis no portal Hidroweb (<https://www.snirh.gov.br/hidroweb/serieshistoricas>). Nesse processo os dados foram ajustados a uma função exponencial (JACCON; CUDO, 1989), de modo a encontrar o melhor coeficiente de determinação (R^2) e, conseqüentemente, a função da curva-chave para a estação fluviométrica (Figura 2.3), obtida com base na Equação 2.1.

$$q = \alpha \times (H - h_0)^n \quad (2.1)$$

em que:

- q = vazão calculada, em m^3/s ;
- h_0 = cota quando a vazão é igual a 0 (m);
- H = cota observada; e
- α e n = parâmetros empíricos calculados para cada estação.

Figura 2.3 – Curva-chave para a transformação de cota observada (m) em vazão (m³/s) na estação Cardoso Moreira (58960000).



Fonte: A autora.

A partir da curva chave, as vazões de referência foram determinadas de acordo com as cotas do SAH-Muriaé e estão apresentadas na Tabela 2.4.

Tabela 2.4 – Cotas e vazões de referência utilizadas na estação Cardoso Moreira do SAH-Muriaé.

Cota (m)	Vazão (m³/s)	Referência
Maior que 7,5	Maior que 724	Inundação
Entre 5 e 7,5	Entre 401 e 724	Alerta
Entre 3 e 5	Entre 191 e 401	Atenção

Fonte: Adaptado de Salviano e Matos (2021).

Diante disso, foram selecionados os dois anos hidrológicos com as maiores médias de vazão entre nov/2000 e out/2019, a fim de analisar o desempenho da modelagem tendo como referência as vazões do sistema de alertas da bacia (SAH-Muriaé).

2.3 RESULTADOS E DISCUSSÃO

Nesse estudo, as técnicas GLM, MARS e RF foram configuradas para três cenários de previsão (um, dois e três dias de antecedência) da vazão na estação mais próxima à foz da bacia do rio Muriaé, estação Y (Cardoso Moreira - 58960000), com 50 repetições. Dessa forma, foram construídos 450 modelos (3×3×50) de Aprendizado de Máquina.

2.3.1 Seleção e importância das covariáveis

Na implementação da modelagem, após a padronização das amostras, houve a seleção das covariáveis pela função *findcorrelation* e algoritmo RFE. Assim, são apresentados nas Tabelas 2.4 a 2.6 os resultados da seleção de recursos para cada cenário, enfatizando as covariáveis que estiveram presentes nas 50 repetições de cada técnica.

No cenário de um dia de antecedência, o RFE ajustado a cada técnica resultou em nove covariáveis relevantes para GLM e MARS, enquanto para RF, destacaram-se 35 covariáveis (Tabela 2.5).

Tabela 2.5 – Covariáveis selecionadas utilizando o algoritmo RFE para GLM, MARS e RF nos cenários de um, dois e três dias de antecedência.

GLM			MARS			RF		
1	2	3	1	2	3	1	2	3
$Q1_t$	$Q1_t$	$Q1_t$	$Q1_t$	$Q1_t$	$Q1_t$	$Q1_t$	$Q1_t$	$Q1_t$
$Q1_{t-1}$	$Q1_{t-1}$	$Q1_{t-1}$	$Q1_{t-1}$	$Q1_{t-1}$	$Q1_{t-1}$	$Q1_{t-1}$	$Q2_t$	$Q2_t$
$Q2_t$	$Q1_{t-2}$	$Q1_{t-2}$	$Q2_t$	$Q2_t$	$Q1_{t-2}$	$Q1_{t-2}$	$Q3_t$	$Q3_t$
$Q3_t$	$Q2_t$	$Q2_t$	$Q3_t$	$Q2_{t-1}$	$Q2_t$	$Q2_t$	$Q3_{t-1}$	$Q3_{t-1}$
$Q4_t$	$Q2_{t-1}$	$Q2_{t-1}$	$Q4_t$	$Q3_t$	$Q2_{t-1}$	$Q2_{t-1}$	$Q4_t$	$Q4_t$
$Q4_{t-1}$	$Q3_t$	$Q3_t$	$Q4_{t-1}$	$Q3_{t-1}$	$Q3_t$	$Q2_{t-2}$	$Q4_{t-1}$	$Q4_{t-1}$
p_xavier_t	$Q4_t$	$Q4_t$	p_xavier_t	$Q4_t$	$Q4_t$	$Q3_t$	$Q4_{t-2}$	$Q4_{t-2}$
PA	$Q4_{t-2}$	$Q4_{t-2}$	p_trmm_t	$Q4_{t-2}$	$Q4_{t-2}$	$Q3_{t-1}$	p_xavier_t	p_trmm_t
AG	p_xavier_t	p_xavier_t	PA	p_xavier_t	p_xavier_t	$Q3_{t-2}$	$p_xavier_{48h_t}$	$p_trmm_{48h_t}$
	p_trmm_t	p_trmm_t		p_trmm_t	p_trmm_t	$Q4_t$	$p_trmm_{48h_t}$	$ndvi_med_t$
	PA	FL		PA	FL	$Q4_{t-1}$	p_trmm_t	$escoam_min_t$
	AG	PA		AG	PA	$Q4_{t-2}$	$p_trmm_{72h_t}$	$umidade_min_t$
		AG			AG	p_xavier_t	$umidade_min_t$	$umidade_max$
		AnV			AnV	p_xavier_{t-2}	$umidade_max$	AG
						$p_xavier_{48h_t}$		AnV
						$p_xavier_{72h_t}$		
						$p_xavier_{96h_t}$		
						$p_xavier_{15em_t}$		
						p_trmm_t		
						p_trmm_{t-1}		
						$p_trmm_{48h_t}$		
						$p_trmm_{72h_t}$		
						$p_trmm_{96h_t}$		
						$p_trmm_{120h_t}$		
						evi_med_t		
						$ndvi_med_t$		
						$escoam_min_t$		
						$umidade_min_t$		
						$umidade_max_t$		
						FL		
						PA		
						AG		
						AnV		
						CH		
						mês		

Fonte: A autora.

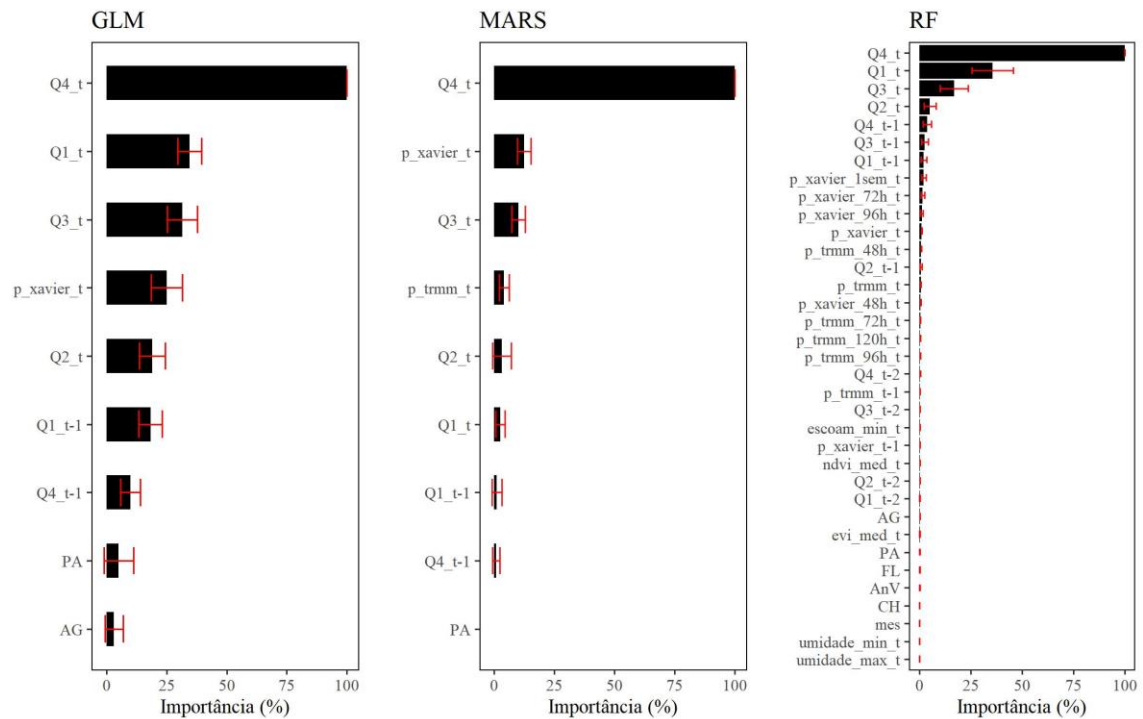
Para o prazo de entrega igual a 2 dias, a seleção de recursos mostrou que o melhor conjunto para GLM e MARS era composto por 12 covariáveis, enquanto 14 covariáveis tiveram mais relevância no RF (Tabela 2.5). Quanto ao cenário de 3 dias de antecedência, novamente GLM e MARS apresentaram o mesmo número de covariáveis relevantes, 14 covariáveis, enquanto RF apresentou 15 covariáveis (Tabela 2.5).

Nos cenários de um, dois e três dias de antecedência, a importância relativa média e o erro relativo para as 50 repetições considerando cada técnica de ML (GLM, MARS e RF) são apresentados nas Figuras 2.4 a 2.6.

No cenário em que se utilizou a vazão com a defasagem de um dia (Figura 2.4.), observou-se que a covariável Q_{4t} foi a mais importante na modelagem para os três algoritmos, além das vazões das demais estações (Q_{1t} , Q_{2t} e Q_{3t}) no tempo t (dia corrente), destacando-se também a precipitação estimada por Xavier *et al.* (2022) (p_{xavier_t}).

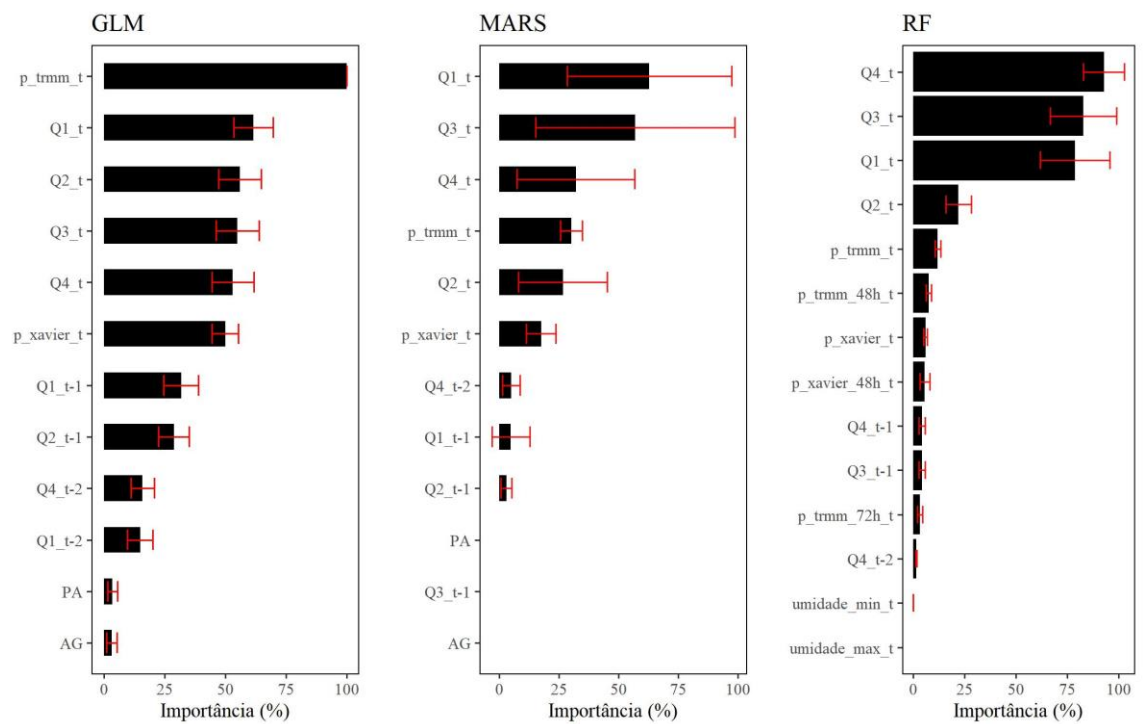
Curiosamente, embora a seleção de recursos com o RFE em associação ao RF tenha apresentado um elevado número de covariáveis no cenário de um dia de antecedência, quando a importância relativa média foi analisada, verificou-se que a maioria das covariáveis (precipitação, MODIS, TerraClimate e uso e cobertura da terra) obteve uma pequena importância, exceto as da categoria vazão. No trabalho de Yin *et al.* (2018), os autores consideraram apenas as covariáveis relacionadas à memória da vazão do rio, ou seja, dados defasados no tempo, e ignoraram covariáveis de precipitação e temperatura, por exemplo. Entretanto, a utilização de diferentes dados na modelagem possibilita maior diversidade aos preditores e pode levar ao desenvolvimento de um modelo mais adequado (AHMED *et al.*, 2021).

Figura 2.4 – Importância e erros relativos das covariáveis selecionadas para a modelagem com os algoritmos GLM, MARS e RF no cenário de um dia de antecedência.



Fonte: A autora.

Figura 2.5 – Importância e erros relativos das covariáveis selecionadas para a modelagem com os algoritmos GLM, MARS e RF no cenário de dois dias de antecedência.

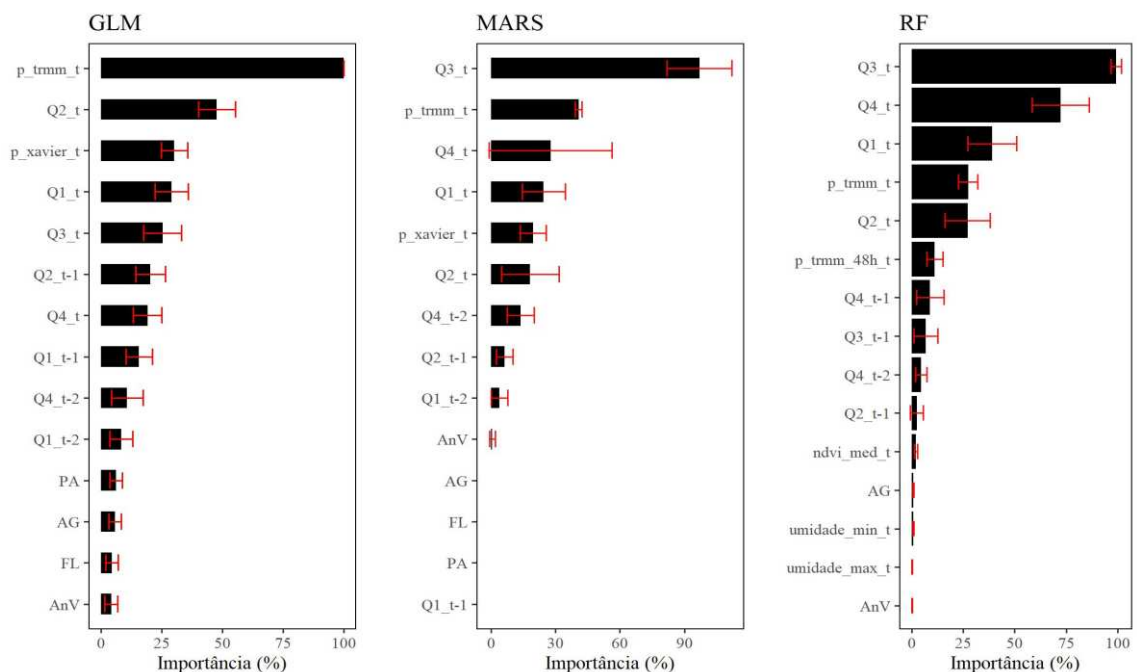


Fonte: A autora.

No cenário de dois dias de antecedência (Figura 2.5) na técnica GLM, a covariável p_trmm_t teve a maior importância, seguida das covariáveis de vazão no dia corrente (t): $Q1_t$, $Q2_t$, $Q3_t$ e $Q4_t$, que apresentaram importância relativa média semelhante. Por outro lado, nas técnicas MARS e RF, as principais covariáveis pertenceram à categoria vazão, principalmente no tempo t . As covariáveis derivadas dos dados do TerraClimate e uso e cobertura da terra apresentaram pequena e/ou nenhuma importância quando presentes na modelagem com as três técnicas de Aprendizado de Máquina.

Quanto ao cenário com três dias de antecedência, embora o RFE associado às técnicas GLM e MARS tenha fornecido o mesmo conjunto de covariáveis, a importância relativa média se apresentou diferente, com a covariável p_trmm_t sendo a mais informativa para GLM, enquanto a covariável $Q3_t$ foi para MARS. Para a técnica RF, o algoritmo RFE mostrou que as covariáveis de vazão foram as mais informativas na modelagem desse cenário, com destaque para $Q3_t$ e $Q4_t$, enquanto as covariáveis do TerraClimate e uso e cobertura da terra obtiveram pouca importância.

Figura 2.6 – Importância e erros relativos das covariáveis para a modelagem com os algoritmos GLM, MARS e RF no cenário de três dias de antecedência.

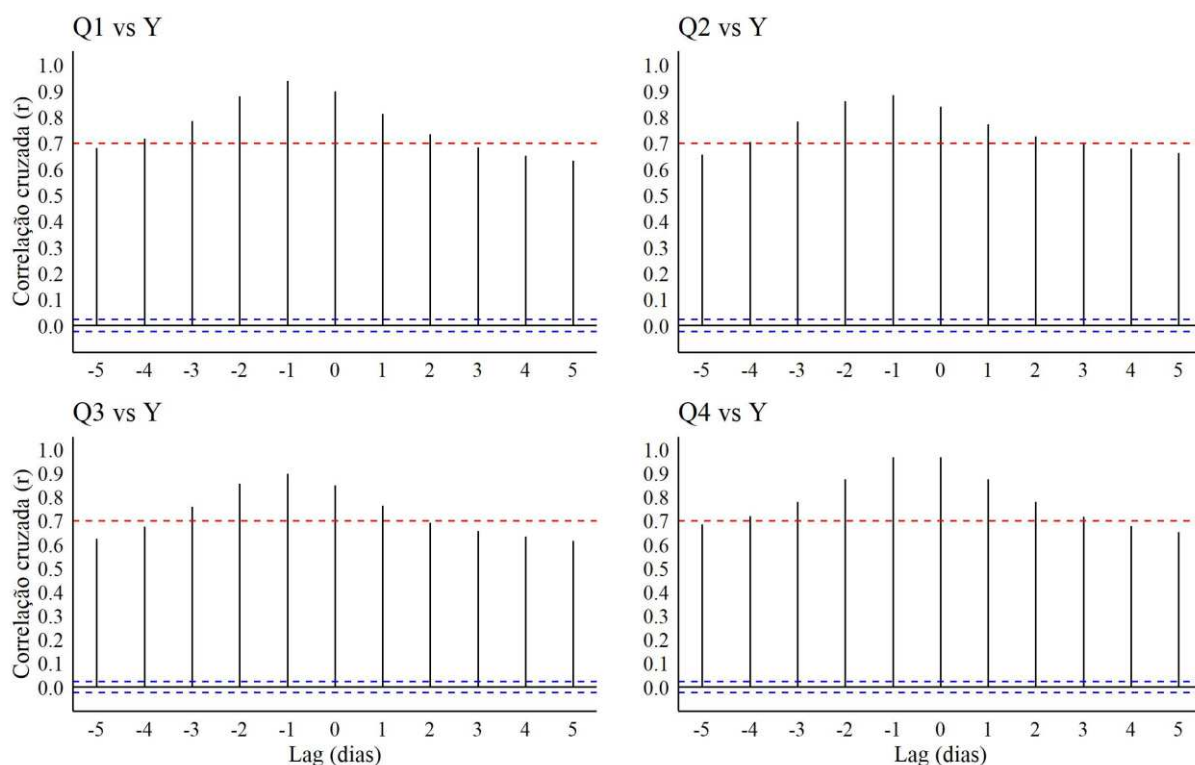


vazão teve os maiores valores de importância seguida pela precipitação. Esse resultado corrobora com diversos trabalhos que encontraram covariáveis de vazão com maior relevância na previsão da vazão com técnicas de ML, independentemente do algoritmo utilizado na seleção de recursos (DUARTE *et al.*, 2022; LIU *et al.*, 2019; REIS *et al.*, 2021).

É importante ressaltar, ainda, que a vazão e a precipitação são as variáveis que mais imprimem importância na modelagem hidrológica (SHEN *et al.*, 2022), de tal modo que suas covariáveis ficaram alternando nos primeiros lugares do *ranking* de importância nos três cenários analisados nesse estudo.

Para entender o poder informativo dessas covariáveis, fez-se a correlação cruzada destas com a vazão da estação 58960000 (Y) (Figura 2.7). Todas as correlações se apresentaram elevadas com um dia de atraso e houve um decréscimo no coeficiente de correlação (r) com o aumento da defasagem. Esse comportamento se reflete na importância das covariáveis em cada cenário, em que houve maior erro médio à medida que os dias de antecedência da previsão aumentaram, mostrando uma diminuição do poder dessas covariáveis na previsão da vazão na estação Y com o aumento de dias de antecedência. Todavia, a elevada importância encontrada na categoria vazão das covariáveis a montante revela a capacidade dos algoritmos em modelar eventos máximos na variável objetivo Y (REIS *et al.*, 2021), especialmente no cenário de um dia de antecedência.

Figura 2.7 – Correlação cruzada entre as covariáveis de vazão (Q1, Q2, Q3 e Q4) e a vazão da estação 58960000 (Y).

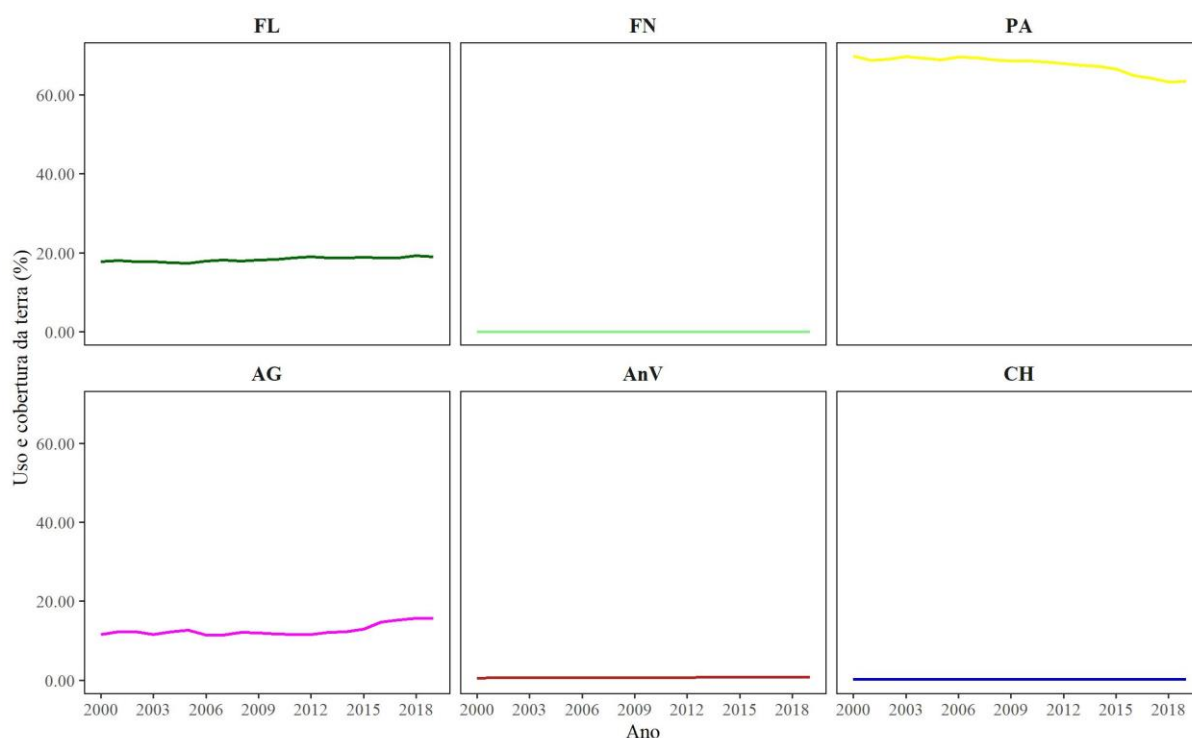


Fonte: A autora.

A seleção de recursos evidenciou, ainda, que a precipitação observada no tempo t foi a segunda covariável a ter maior contribuição na modelagem, de modo a estar presente em todos os conjuntos de covariáveis selecionados com o RFE combinado às técnicas de Aprendizado de Máquina em cada cenário. Por outro lado, a precipitação acumulada teve maior relevância na técnica RF. Para Bui, Ishidaira e Shaowei (2019), a precipitação consiste em uma das principais entradas para a modelagem da vazão e, a depender da fonte, pode influir incerteza no processo preditivo. Isso reflete a importância da seleção de variáveis, que objetiva selecionar as melhores covariáveis dentro de um conjunto amplo, a fim de reduzir as incertezas que poderiam estar associadas a determinadas covariáveis.

Para o uso e cobertura da terra, as covariáveis dessa categoria foram consideradas relevantes em todos os cenários, embora não tenham sido informativas na maioria das técnicas. Esse fato pode estar associado à baixa variabilidade das classes de uso e cobertura da terra no período analisado considerando a área de drenagem da estação mais a jusante da bacia (Figura 2.8), em que houve um crescimento de apenas 1,4% da área de floresta (FL) e redução de 6,4% da área de pastagem (PA).

Figura 2.8 – Variação temporal do uso e cobertura da terra (%) na área de drenagem da estação Cardoso Moreira (58960000).



FL – Floresta; FN – Formação Natural não Florestal; PA – Pastagem; AG – Agricultura; AnV – Área não Vegetada e CH – Corpos Hídricos. Fonte: A autora.

A bacia do rio Muriaé está localizada no bioma Mata Atlântica, que tem um histórico de elevada degradação anterior a 1985, de acordo com os dados do MAPBIOMAS (2022). Alguns autores ressaltam que a Mata Atlântica apresenta um comportamento divergente em relação aos demais biomas brasileiros, com tendência positiva da cobertura nativa e migração de áreas ocupadas por pastagem para a agricultura (SILVEIRA *et al.*, 2022b; SOUZA *et al.*, 2020).³

2.3.2 Comparação das técnicas de Aprendizado de Máquina

Na Tabela 2.6 encontram-se as vazões mínimas, máximas e médias observadas e simuladas pelas técnicas de ML (LM, MARS e RF) nos cenários de um, dois e três dias de antecedência para as 50 repetições da modelagem.

Tabela 2.6 – Vazões médias, mínimas e máximas observadas e simuladas pelas técnicas de ML (GLM, MARS e RF) na previsão da vazão nos dados de teste nos cenários de um, dois e três dias de antecedência, com base em 50 repetições.

Cenário	Modelo	Vazão média		Vazão mínima		Vazão máxima		Desvio padrão (m³/s)
		Vazão (m³/s)	Erro relativo (%)	Vazão (m³/s)	Erro relativo (%)	Vazão (m³/s)	Erro relativo (%)	
	Vazão observada	93,96	0,00	3,14	0,00	1849,99	0,00	116,89
1	GLM	94,10	0,15	-9,88	414,48	1682,39	9,06	115,55
	MARS	94,16	0,21	-0,92	129,16	2117,26	14,45	115,61
	RF	94,25	0,31	4,33	37,82	1463,62	20,88	112,65
2	GLM	94,05	0,09	-15,89	605,48	1649,49	10,84	111,81
	MARS	94,04	0,08	-10,68	439,78	2199,08	18,87	112,03
	RF	94,65	0,73	4,08	29,72	1372,07	25,83	108,17
3	GLM	94,09	0,14	-22,88	828,03	1443,63	21,97	103,79
	MARS	93,96	0,00	-12,64	502,15	1661,68	10,18	103,66
	RF	95,53	1,67	4,23	34,62	1304,54	29,48	103,38

Fonte: A autora.

Entre todos os modelos de ML (GLM, MARS e RF), para os cenários de um, dois e três dias de antecedência, houve superestimação das vazões médias com baixos erros relativos, exceto para o modelo obtido por MARS no cenário de três dias de antecedência que obteve a mesma vazão observada (Tabela 2.6).

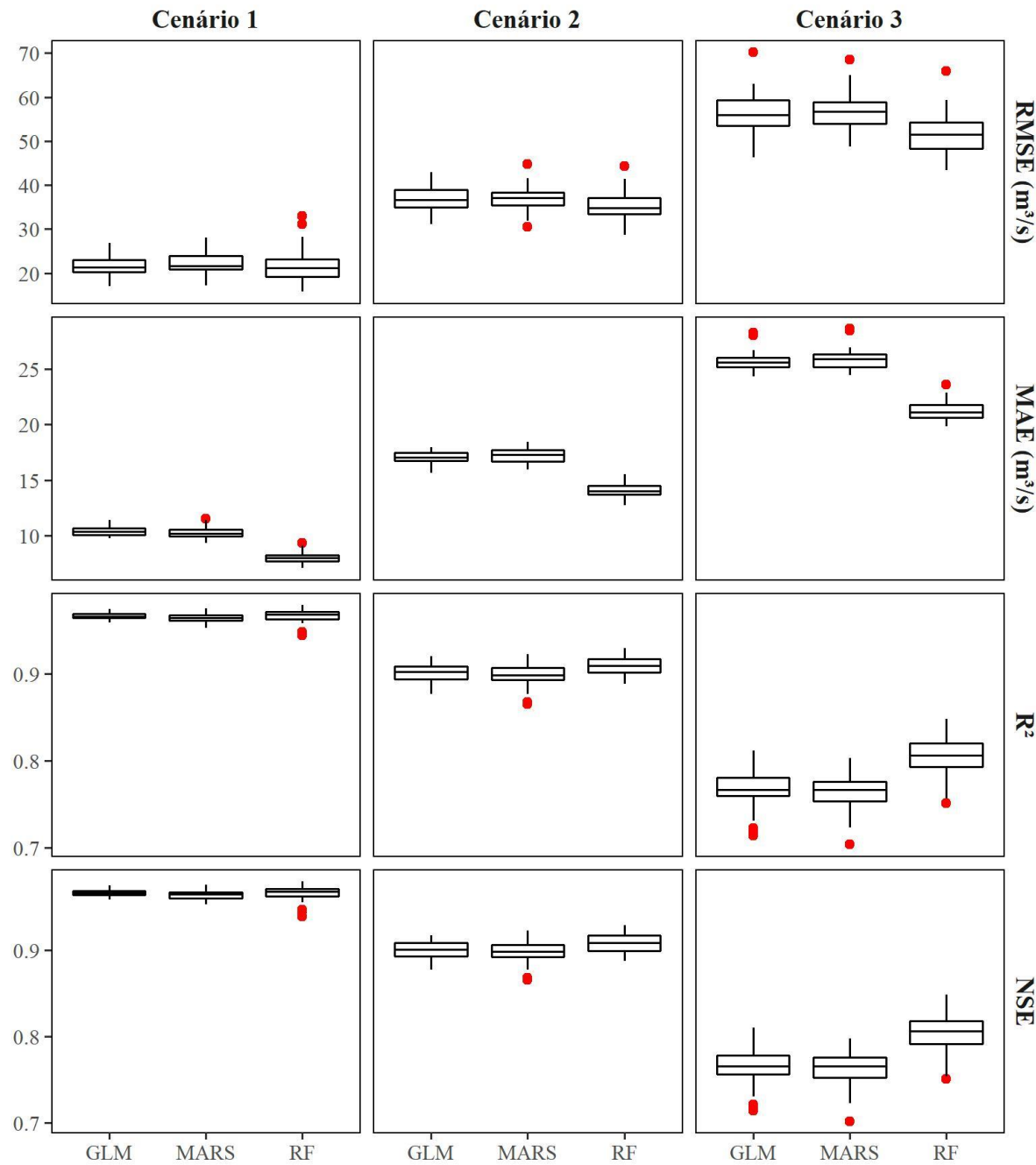
Para as vazões mínimas, apenas o RF apresentou superestimação da vazão, sendo igual a 37,82%, 29,72% e 34,62%, respectivamente, para os cenários de um, dois e três dias de antecedência. Por outro lado, as técnicas GLM e MARS apresentaram valores negativos nos três cenários de antecedência. Embora GLM e MARS constituam técnicas mais sofisticadas de regressão linear, a sua natureza matemática pode fazer os modelos predizerem valores negativos (IRVING *et al.*, 2018).

Nos valores máximos, observaram-se vazões inferiores obtidas pelas técnicas de ML (GLM, MARS e RF), em que o algoritmo MARS superestimou o valor observado nos cenários de um e dois dias de antecedência, enquanto o RF obteve as vazões mais baixas em comparação com as demais técnicas. O resultado observado para o RF é semelhante ao encontrado por Reis (2020) ao analisar a predição da vazão diária utilizando esse algoritmo. Ressalta-se ainda que as vazões máximas ocorrem com baixa frequência e apresentam-se muitas vezes como *outliers* numa série histórica de vazões, aumentando os erros na previsão (LI; SHA; WANG, 2019).

A variação das métricas RMSE, MAE, R^2 e NSE dos algoritmos de ML (GLM, MARS e RF) nos dados de teste, para os três cenários, são apresentadas na Figura 2.9. As métricas

estatísticas RMSE, MAE, R^2 e NSE são do conjunto das 50 execuções de teste para cada modelo ajustado, enquanto na Tabela 2.7 é apresentado o desempenho médio no treinamento e teste considerando as 50 repetições.

Figura 2.9 – Performance dos modelos de ML (GLM, MARS e RF) na previsão da vazão nos dados de teste nos cenários de um, dois e três dias de antecedência, com base em 50 repetições.



Fonte: A autora.

Tabela 2.7 – Média das métricas estatísticas dos modelos de ML (GLM, MARS e RF) no treinamento e teste na previsão da vazão nos cenários de um, dois e três dias de antecedência na previsão da vazão, com base em 50 repetições.

Cenário	Estatística	Treinamento			Teste		
		GLM	MARS	RF	GLM	MARS	RF
1	RMSE (m ³ /s)	21,59	22,83	21,68	21,52	22,27	21,51
	MAE (m ³ /s)	10,41	10,39	8,13	10,39	10,24	8,04
	R ²	0,97	0,96	0,97	0,97	0,96	0,97
	NSE	0,97	0,97	0,99	0,97	0,96	0,97
	Classificação	Muito bom	Muito bom	Muito bom	Muito bom	Muito bom	Muito bom
2	RMSE (m ³ /s)	36,74	37,48	35,91	36,93	37,09	35,45
	MAE (m ³ /s)	17,15	17,45	14,33	17,07	17,22	14,08
	R ²	0,90	0,90	0,91	0,90	0,90	0,91
	NSE	0,91	0,91	0,98	0,90	0,90	0,91
	Classificação	Muito bom	Muito bom	Muito bom	Muito bom	Muito bom	Muito bom
3	RMSE (m ³ /s)	56,21	57,15	52,29	56,39	56,80	51,77
	MAE (m ³ /s)	25,84	26,21	21,58	25,68	25,90	21,22
	R ²	0,78	0,77	0,81	0,77	0,76	0,80
	NSE	0,78	0,79	0,96	0,77	0,76	0,80
	Classificação	Bom	Bom	Bom	Bom	Bom	Bom

Fonte: A autora.

No cenário de um dia de antecedência, os modelos gerados pelas três técnicas de Aprendizado de Máquina apresentaram resultados semelhantes (Figura 2.9), com RF tendo um desempenho levemente acima das demais técnicas em vista de erros menores, embora tenha apresentado mais *outliers* em todas as métricas. Ressalta-se que as métricas MAE e RMSE para todos os modelos apresentaram valores distantes entre si, indicando a presença de *outliers* nos dados, uma vez que a métrica RMSE apresenta sensibilidade a eles (KARUNASINGHA, 2022). Além disso, ao analisar a média das métricas (Tabela 2.7), verificou-se que os erros do RF (RMSE = 21,51 m³/s e MAE = 8,04 m³/s) foram menores em 0,05% e 3,41% em relação ao GLM e MARS, respectivamente. Destaca-se, ainda, que os resultados obtidos confirmam o ajuste ótimo nos dados de teste pelas três técnicas de ML.

No cenário com dois dias de antecedência (Figura 2.9) houve uma redução no desempenho das técnicas, com MARS tendo maior número de *outliers* nas métricas RMSE, R² e NSE. A Tabela 2.7 mostra que os erros aumentaram em relação a modelagem com um dia de antecedência para GLM em 71,61% (RMSE = 36,93 m³/s) e 64,29% (MAE = 17,07 m³/s), no MARS em 66,55% (RMSE = 37,09 m³/s) e 68,16% (MAE = 17,22 m³/s), enquanto o RF os aumentos foram de 64,81% (RMSE = 35,45 m³/s) e 75,12% (MAE = 14,08 m³/s), respectivamente. Quanto ao R² e NSE, ambos foram iguais a 0,90 para GLM e MARS, enquanto

no RF essas métricas chegaram a 0,91. Na análise geral, o RF se sobressaiu em relação às demais técnicas por ter os menores erros, embora todas tenham recebido a mesma classificação de desempenho.

O bom desempenho do RF no cenário com dois dias de antecedência pode estar relacionado à sua capacidade de mineração de dados e habilidade em analisar a complexidade das suas iterações (FANG *et al.*, 2023). No trabalho de Papacharalampous e Tyrallis (2018), o RF manteve um desempenho superior às demais técnicas mesmo com o aumento do número de dias de antecedência na previsão da vazão.

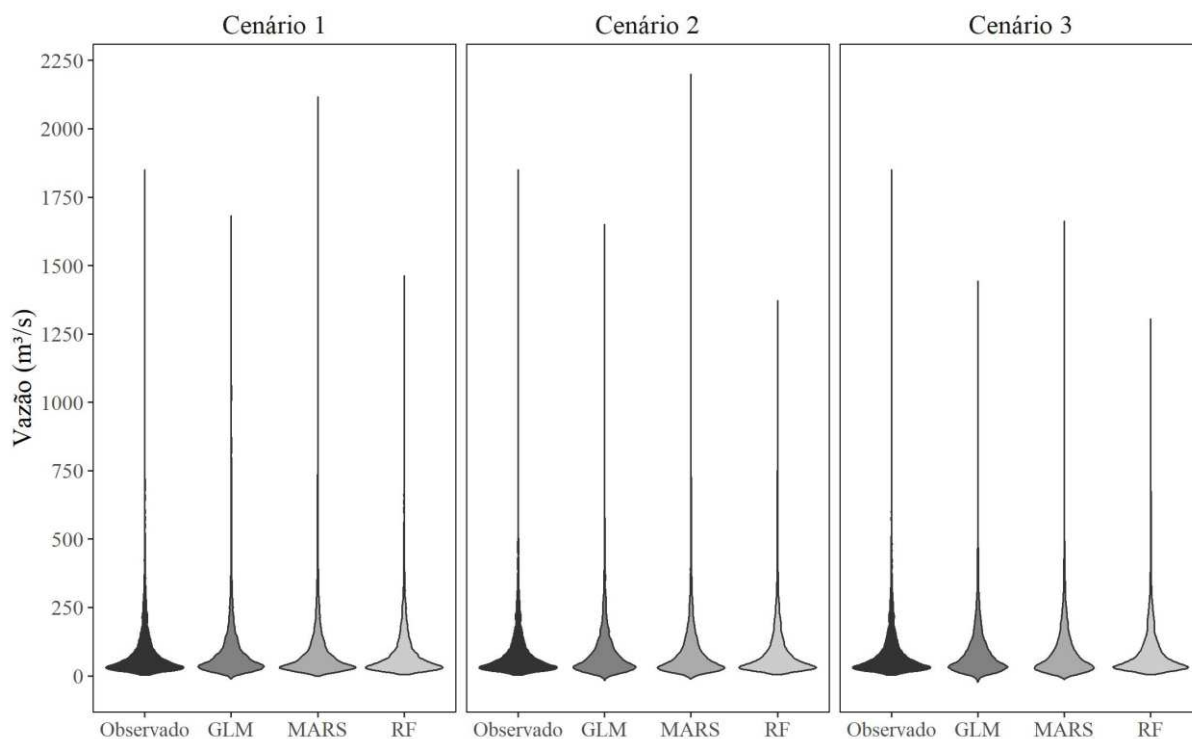
No prazo de entrega igual a três dias, o desempenho das técnicas reduziu consideravelmente (Figura 2.9), sendo que o RF manteve o melhor desempenho, em vista dos menores erros ($RMSE = 51,77 \text{ m}^3/\text{s}$ e $MAE = 21,22 \text{ m}^3/\text{s}$), com R^2 e NSE iguais a 0,80 (Tabela 2.7). Esses valores foram considerados como adequados no trabalho de Reis *et al.* (2021), em que o R^2 tem limite mínimo igual 0,80 e NSE igual a 0,75 para a previsão de vazão diária.

A redução do desempenho na medida em que o prazo de entrega da previsão aumenta também foi encontrada em diversos trabalhos com diferentes técnicas de ML (CHENG *et al.*, 2020; GIRIHAGAMA *et al.*, 2022; PAPACHARALAMPOUS; TYRALIS, 2018). Esse comportamento ocorre em decorrência da modelagem estar estruturada em dados, ou seja, as relações de dependência existentes entre as covariáveis e a variável objetivo diminuem com o aumento da defasagem, de modo a reduzir a previsibilidade da série temporal (DHANYA; NAGESH KUMAR, 2011; LIU *et al.*, 2014; TIWARI; CHATTERJEE, 2010).

Na Figura 2.10 é apresentada a forma da função densidade de probabilidade (PDF) das vazões observadas e simuladas por meio das técnicas de ML (GLM, MARS e RF) para os três prazos de entrega nos dados de teste considerando as 50 repetições.

As técnicas GLM e RF não conseguiram alcançar os valores máximos observados de vazão nos cenários com um, dois e três dias de antecedência, enquanto o MARS teve tendência a superestimar nos dois primeiros prazos de entrega da previsão. No trabalho de Singh *et al.* (2022), os autores obtiveram um outro comportamento de MARS e RF, em que RF foi superior na previsão de valores máximos. No presente trabalho, MARS teve maior eficiência em aprender os processos ocultos da modelagem associados aos picos de vazão.

Figura 2.10 – Distribuição dos dados observados e simulados pelas técnicas de ML (GLM, MARS e RF) nos dados de teste nos cenários de um, dois e três dias de antecedência, com base em 50 repetições

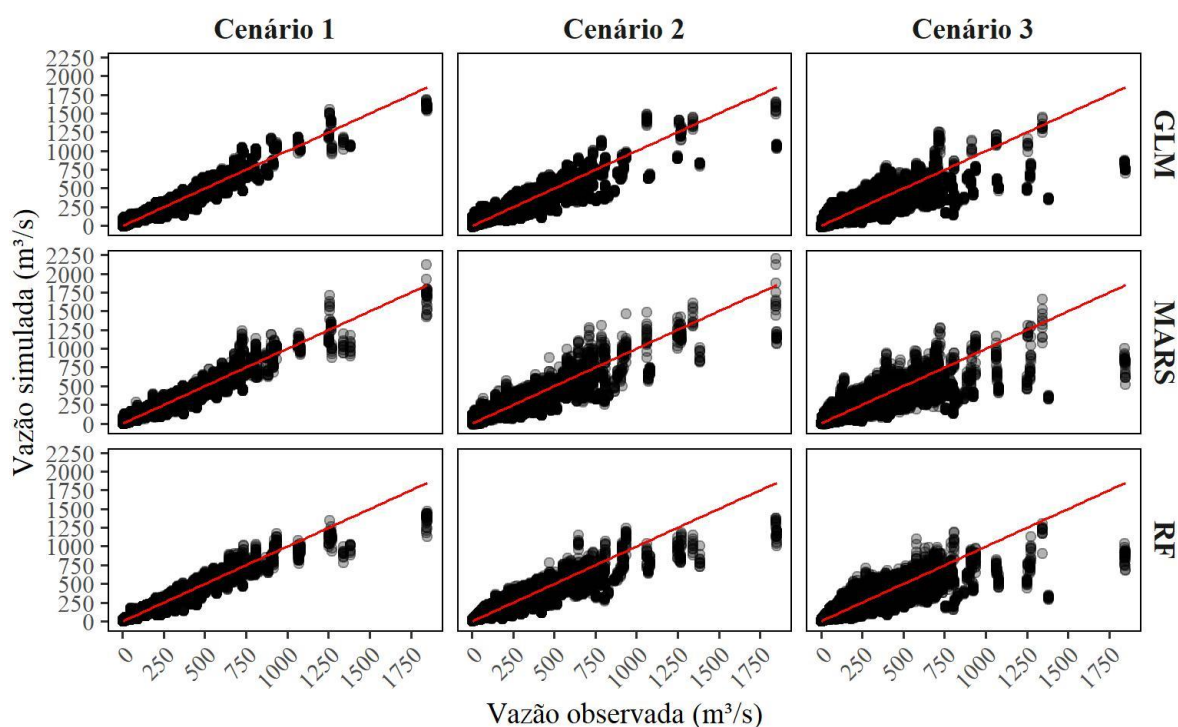


Fonte: A autora.

Na análise da previsão da vazão por meio de diferentes técnicas de ML, incluindo o MARS, Cui, Salih e Choubin (2020) encontraram um desempenho excelente dessa técnica com uma representação da função densidade de probabilidade muito próxima aos valores reais. Entre as principais razões para o alcance de vazões máximas por essa técnica, encontra-se a robustez da modelagem, que não precisa de nenhuma hipótese forte para aprender com as relações dos preditores e variável resposta que não seguem linearidade (DEO; SAMUI; KIM, 2016; YIN *et al.*, 2018).

Na Figura 2.11 é apresentada a dispersão das vazões observadas e simuladas pelas técnicas de ML (GLM, MARS e RF) nos cenários com um, dois e três dias de antecedência nas 50 repetições nos dados de teste, com a linha de ajuste 1:1 em vermelho.

Figura 2.11 – Dispersão das vazões simuladas e observadas para diferentes métodos de Aprendizado de Máquina (GLM, MARS e RF) nos três cenários nas 50 repetições nos dados de teste



Fonte: A autora.

No cenário de um dia de antecedência, a previsão de vazões acima de 1.200 m³/s apresentou dois comportamentos: maior variação entre subestimação e superestimação para as técnicas GLM e MARS e maior subestimação para RF. Para os picos de vazão, GLM e RF subestimaram os dados observados, enquanto MARS apresentou subestimação de alguns dados.

Para o prazo de entrega de dois dias, a dispersão em torno da linha de ajuste perfeito aumentou para todas as técnicas, em destaque MARS, que continuou a superestimar vazões máximas. No cenário de três dias de antecedência, GLM, MARS e RF passaram a subestimar a vazão com maior frequência a partir da vazão de 750 m³/s.

É evidente que a habilidade das técnicas em prever a vazão foi reduzindo conforme o aumento dos dias de antecedência, principalmente no prazo de entrega igual a três dias, com a previsão de valores anormais em comparação aos dados observados, como, por exemplo, no RF, que previu uma vazão de 343,32 m³/s enquanto o valor observado era igual a 1.384,50 m³/s. Esse comportamento com aumento do prazo de entrega também foi observado por Yin *et al.* (2018) ao analisar os cenários de um, dois e três dias de antecedência na previsão da vazão com diferentes técnicas de ML.

2.3.3 Vazões de referência do SAH-Muriaé

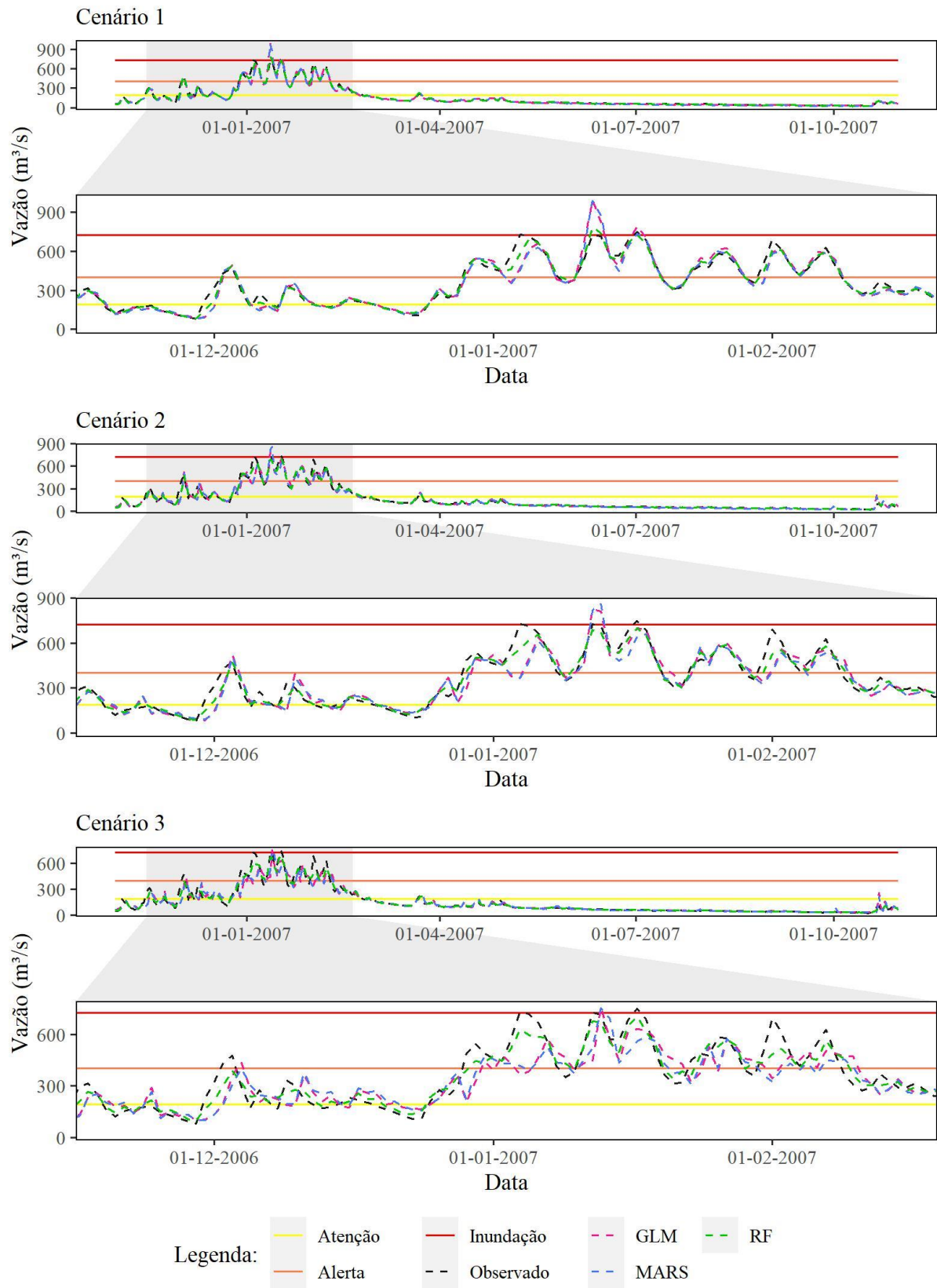
Para explorar o desempenho das técnicas de ML (GLM, MARS e RF) considerando os prazos de entrega da previsão em um dois e três dias, fizeram-se gráficos do desempenho das técnicas nos dois anos hidrológicos com as maiores médias anuais de vazão, que corresponderam aos anos 2006/2007 (Figura 2.12) e 2008/2009 (Figura 2.13). As métricas de desempenho dos anos hidrológicos se encontram no Apêndice E. Nessa análise gráfica também foram incluídos os limites de alerta do SAH-Muriaé, a fim de identificar a habilidade das técnicas levando em consideração as vazões de referência correspondentes aos níveis de: Atenção, Alerta e Inundação. Além disso, calcularam-se os eventos das vazões de referência do SAH-Muriaé e seus respectivos erros relativos para cada cenário e modelo de Aprendizado de Máquina Tabela 2.8 e 2.9.

No ano hidrológico 2006/2007, no cenário de um dia de antecedência (Figura 2.12), verificou-se que a técnica RF conseguiu se aproximar mais dos dados observados de Alerta e Inundação, enquanto GLM e MARS não registram o primeiro evento de Inundação e superestimaram a vazão no segundo evento da série analisada. Dos eventos observados e simulados (Tabela 2.8), o RF apresentou os melhores resultados, confirmando sua habilidade na previsão da vazão diária.

No cenário de dois dias de antecedência no ano 2006/2007 (Figura 2.12), novamente RF conseguiu prever valores mais próximos aos dados observados, embora tenha apresentado dificuldades de prever os picos de vazões, enquanto GLM e MARS apresentaram um atraso maior na previsão do primeiro evento de Inundação, não conseguindo atingir a vazão para a emissão do alerta pelo SAH-Muriaé. Esses resultados foram refletidos nos erros relativos (%) dos eventos das vazões de referência do SAH-Muriaé, com o RF não prevendo nenhuma vazão de Inundação (Tabela 2.8).

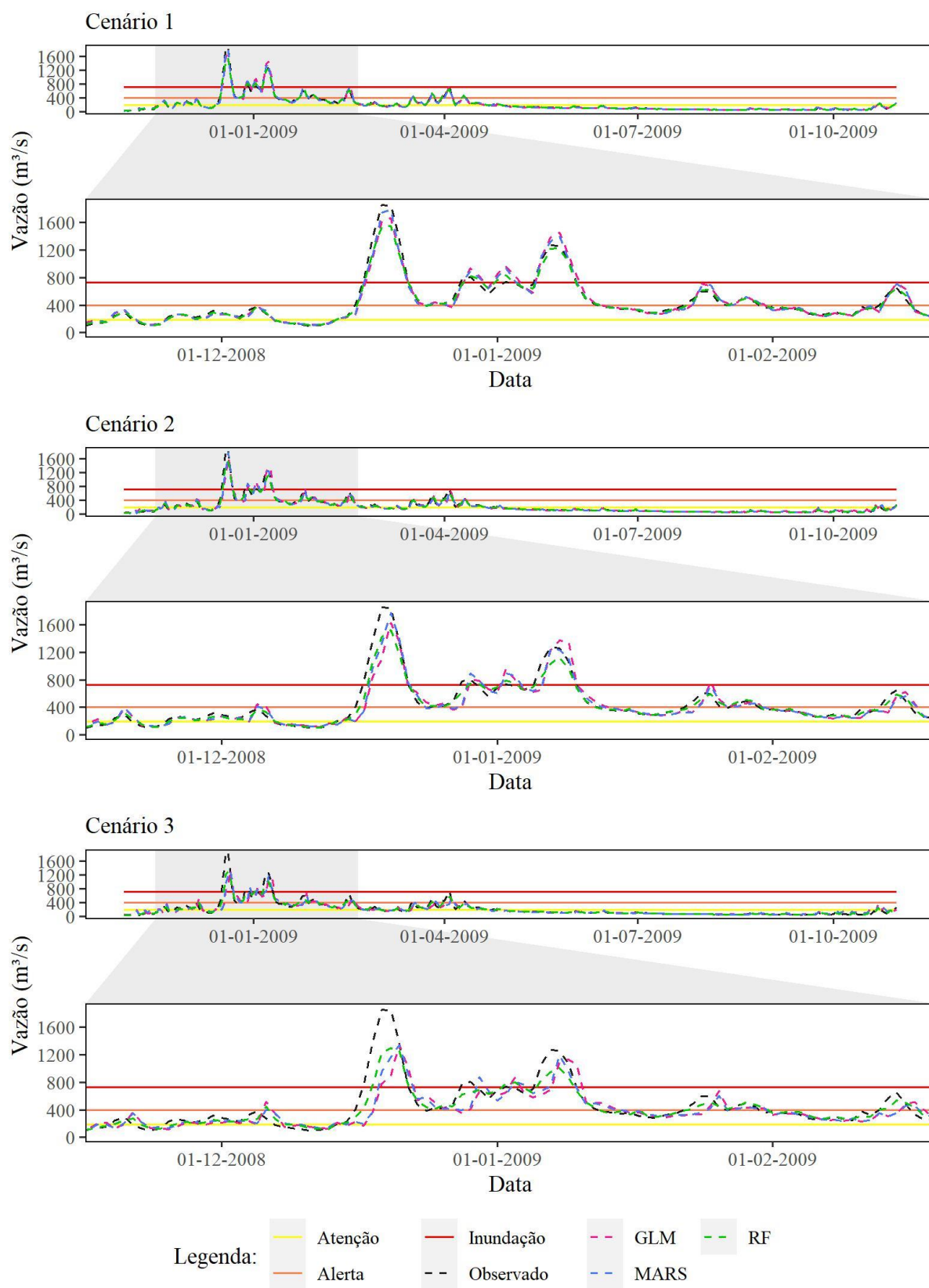
No cenário de três dias de antecedência (Figura 2.12), para o ano hidrológico 2006/2007, o desempenho das técnicas não foi satisfatório, em que os valores simulados não se aproximaram dos valores das vazões máximas observadas. Embora os erros relativos dos eventos preditos pelos modelos de MARS tenham sido os menores para esse cenário (Tabela 2.8), houve atraso na previsão como demonstrado na Figura 2.12.

Figura 2.12 – Desempenho médio das técnicas de ML (GLM, MARS e RF) na previsão da vazão com um, dois e três dias de antecedência no ano hidrológico 2006/2007.



Fonte: A autora.

Figura 2.13 – Desempenho médio das técnicas de ML (GLM, MARS e RF) na previsão da vazão com um, dois e três dias de antecedência no ano hidrológico 2008/2009.



Fonte: A autora.

Tabela 2.8 – Número de eventos das vazões de referência do SAH-Muriaé para o ano hidrológico 2006/2007 com base na média das 50 repetições.

Cenário	Vazão	Normal		Atenção		Alerta		Inundação	
		Vazão	Erro relativo (%)	Vazão	Erro relativo (%)	Vazão	Erro relativo (%)	Vazão	Erro relativo (%)
	Vazão observada	283,00	0,00	45,00	0,00	34,00	0,00	3,00	0,00
1	GLM	290,00	2,47	40,00	11,11	32,00	5,88	3,00	0,00
	MARS	289,00	2,12	42,00	6,67	30,00	11,76	3,00	0,00
	RF	286,00	1,06	42,00	6,67	34,00	0,00	3,00	0,00
2	GLM	286,00	1,06	43,00	4,44	34,00	0,00	2,00	33,33
	MARS	286,00	1,06	47,00	4,44	30,00	11,76	2,00	33,33
	RF	282,00	0,35	44,00	2,22	38,00	11,76	0,00	100,00
3	GLM	281,00	0,71	49,00	8,89	34,00	0,00	1,00	66,67
	MARS	274,00	3,18	58,00	28,89	32,00	5,88	1,00	66,67
	RF	277,00	2,12	53,00	17,78	35,00	2,94	0,00	100,00

Fonte: A autora.

Tabela 2.9 – Número de eventos das vazões de referência do SAH-Muriaé para o ano hidrológico 2008/2009 com base na média das 50 repetições.

Cenário	Modelo	Normal		Atenção		Alerta		Inundação	
		Vazão	Erro relativo (%)	Vazão	Erro relativo (%)	Vazão	Erro relativo (%)	Vazão	Erro relativo (%)
	Vazão observada	222,00	0,00	94,00	0,00	34,00	0,00	14,00	0,00
1	GLM	231,00	4,05	88,00	6,38	30,00	11,76	16,00	14,29
	MARS	229,00	3,15	91,00	3,19	30,00	11,76	14,00	0,00
	RF	224,00	0,90	95,00	1,06	33,00	2,94	13,00	7,14
2	GLM	235,00	5,86	85,00	9,57	32,00	5,88	13,00	7,14
	MARS	229,00	3,15	93,00	1,06	30,00	11,76	12,00	14,29
	RF	223,00	0,45	96,00	2,13	33,00	2,94	13,00	7,14
3	GLM	231,00	4,05	94,00	0,00	32,00	5,88	8,00	42,86
	MARS	226,00	1,80	100,00	6,38	28,00	17,65	10,00	28,57
	RF	229,00	3,15	93,00	1,06	30,00	11,76	12,00	14,29

Fonte: A autora.

No ano hidrológico 2008/2009 (Figura 2.13), no cenário com um dia de antecedência, observou-se um comportamento semelhante das técnicas GLM, MARS e RF em relação ao ano 2006/2007. Novamente RF conseguiu se aproximar dos dados observados, apresentando dificuldade em prever o maior evento de Inundação, embora tenha obtido um dos melhores desempenhos quando analisado sob a perspectiva do número de eventos das vazões de referência do SAH-Muriaé (Tabela 2.9). Além disso, os valores simulados também se

adequaram bem às vazões de Atenção e Alerta, sendo possível emitir alerta para o prazo de entrega da previsão considerado. Ekolu *et al.* (2022) também observaram resultados satisfatórios do RF na modelagem do hidrograma, de forma a representar as características básicas das vazões mínimas e máximas, por exemplo, forma e duração.

Quanto às técnicas GLM e MARS para um dia de antecedência na previsão da vazão, essas se aproximaram do maior evento de Inundação e superestimaram levemente os demais eventos. De todo modo, o desempenho das técnicas foi muito bom no ano hidrológico 2008/2009 para o cenário de um dia de antecedência, visto que forneceram vazões passíveis de alerta de acordo com os limites do SAH-Muriaé. Ressalta-se ainda que o MARS obteve o melhor desempenho quanto ao número de eventos, conseguindo prever todos no cenário de um dia de antecedência (Tabela 2.9).

Ainda sobre o ano 2008/2009, no cenário com dois dias de antecedência, as técnicas GLM e MARS tiveram maiores atrasos nos eventos de Inundação, seguidas pelo RF. GLM e MARS se aproximaram mais dos valores de pico, o que não foi verificado na técnica RF, que superou as demais técnicas por ter maior concordância com os valores reais ao longo do ano analisado. Na análise do número de eventos das vazões de referência, os modelos de RF apresentaram os menores erros relativos ao prever eventos de Alerta e Inundação, seguido pelos modelos de GLM, sendo que ambos apresentaram um erro relativo igual a 7,14% para vazões de Inundação (Tabela 2.9).

No cenário com três dias de antecedência, para o ano hidrológico 2008/2009, houve um declínio considerável do desempenho das três técnicas para as vazões de interesse nesse estudo, destacando-se GLM e MARS com os piores desempenhos. Essas técnicas previram eventos de Alerta e Inundação com mais atrasos em relação aos dados observados, apresentando erros relativos maiores nesse cenário para vazões de Inundação (Tabela 2.9).

Nos dois anos hidrológicos analisados, para fins de Atenção, Alerta e Inundação, os cenários de até dois dias de antecedência com RF se mostraram promissores. Embora para o prazo de entrega de dois dias tenha ocorrido *underfitting*, essa técnica apresentou maior concordância com a série de vazões observadas, sendo possível superar esse problema com a atualização dos dados diariamente (GRANATA; DI NUNNO; MARINIS, 2022).

Yin *et al.* (2018) ressaltam que encontrar previsões confiáveis de vazão em prazos de entrega maiores consiste em um dos grandes problemas enfrentados por gestores de bacias hidrográficas, em decorrência da intensificação da recorrência de eventos extremos, como, por exemplo, na bacia do rio Muriaé, de modo que as previsões realizadas pelas técnicas de ML,

em especial o RF com dois dias de antecedência, pode se constituir em uma ferramenta valiosa a fim de minimizar os impactos causados pelas vazões máximas.

2.4 CONCLUSÕES

Com base nos resultados obtidos e nas análises realizadas, pode-se concluir que:

- Em todos os prazos de entrega da previsão, as covariáveis mais informativas para as técnicas de Aprendizado de Máquina foram da categoria vazão, principalmente no dia corrente (t), seguida da precipitação, que alternou com a vazão nos primeiros lugares do *ranking* de importância na técnica GLM;
- Todas as técnicas de Aprendizado de Máquina apresentaram um desempenho muito bom na previsão da vazão com um dia de antecedência, com destaque para o RF;
- Houve uma redução no desempenho das técnicas com o aumento do número de dias de antecedência na previsão, embora o RF tenha obtido maior precisão nos três cenários analisados com menores erros entre as técnicas analisadas;
- Ao comparar as vazões simuladas e observadas, levando em consideração os limites para a emissão de alerta do SAH-Muriaé, comprovou-se que o RF consiste na técnica mais eficiente na previsão com até dois dias de antecedência em razão de ter apresentado maior concordância com os valores observados; e
- As técnicas de Aprendizado de Máquina GLM, MARS e RF são aplicáveis na previsão de vazões, destacadamente em bacias hidrográficas com recorrentes eventos de elevada magnitude, com possibilidade de aumentar os dias de antecedência na previsão da vazão nos sistemas de alerta e, assim, reduzir os impactos socioambientais decorrentes dos eventos extremos de vazão.

2.5 REFERÊNCIAS

- ADNAN, R. M. *et al.* Least square support vector machine and multivariate adaptive regression splines for streamflow prediction in mountainous basin using hydro-meteorological data as inputs. **Journal of Hydrology**, v. 586, p. 124371, 1 jul. 2020.
- AHANI, A.; SHOURIAN, M.; RAD, P. R. Performance Assessment of the Linear, Nonlinear and Nonparametric Data Driven Models in River Flow Forecasting. **Water Resources Management**, v. 32, n. 2, p. 383–399, 2018.
- AHMED, A. A. M. *et al.* Deep learning hybrid model with Boruta-Random forest optimiser algorithm for streamflow forecasting with climate mode indices, rainfall, and periodicity. **Journal of Hydrology**, v. 599, p. 126350, 1 ago. 2021.

ALVARES, C. A. *et al.* Köppen's climate classification map for Brazil. **Meteorologische Zeitschrift**, v. 22, n. 6, p. 711–728, 2013.

AMORIM, R. R.; REIS, C. H.; FERREIRA, C. Mapeamento dos geossistemas e dos sistemas antrópicos como subsídio ao estudo de áreas com riscos a inundações no baixo curso da bacia hidrográfica do rio Muriaé (Rio de Janeiro – Brasil). **Territorium**, n. 24, p. 89–114, 30 mar. 2017.

ANA, AGÊNCIA NACIONAL DE ÁGUAS. **Previsão de Eventos Críticos na Bacia do Rio Paraíba do Sul, R31 – Papel dos Afluentes nas Cheias e Pontos Críticos para o SIEMEC**. Brasília: ANA, SUM/SPR, 2011.

ATIQUEZZAMAN, M.; KANDASAMY, J. Robustness of Extreme Learning Machine in the prediction of hydrological flow series. **Computers and Geosciences**, v. 120, n. August, p. 105–114, 2018.

AVAND, M. *et al.* DEM resolution effects on machine learning performance for flood probability mapping. **Journal of Hydro-environment Research**, v. 40, p. 1–16, 1 jan. 2022.

BREIMAN, L. Random Forests. **Machine Learning**, v. 45, p. 5–32, 2001.

BUI, H. T.; ISHIDAIRA, H.; SHAOWEI, N. Evaluation of the use of global satellite–gauge and satellite-only precipitation products in stream flow simulations. **Applied Water Science**, v. 9, n. 53, p. 1–15, 2019.

CAMPOS, P. C. DE O.; PAZ, I. Spatial diagnosis of rain gauges' distribution and flood impacts: Case study in Itaperuna, Rio de Janeiro-Brazil. **Water (Switzerland)**, v. 12, n. 4, 1 abr. 2020.

CHEN, C. *et al.* Convolutional Neural Networks for forecasting flood process in Internet-of-Things enabled smart city. **Computer Networks**, v. 186, 26 fev. 2021.

CHENG, M. *et al.* Long lead-time daily and monthly streamflow forecasting using machine learning methods. **Journal of Hydrology**, v. 590, 1 nov. 2020.

COLLAR, N. M. *et al.* Linking fire-induced evapotranspiration shifts to streamflow magnitude and timing in the western United States. **Journal of Hydrology**, v. 612, p. 128242, 1 set. 2022.

CPRM, COMPANHIA DE PESQUISA DE RECURSOS MINERAIS. **Bacia do Rio Muriaé - Apresentação**. Disponível em: https://www.cprm.gov.br/sace/index_bacias_monitoradas.php?getbacia=bmuriae#. Acesso em: 8 set. 2021.

CUI, F.; SALIH, S. Q.; CHOUBIN, B. Newly explored machine learning model for river flow time series forecasting at Mary River, Australia. **Environ Monit Assess**, p. 192–761, 2020.

DARABI, H. *et al.* Urban flood risk mapping using the GARP and QUEST models: A comparative study of machine learning techniques. **Journal of Hydrology**, v. 569, p. 142–154, 1 fev. 2019.

DEO, R. C.; SAMUI, P.; KIM, D. Estimation of monthly evaporative loss using relevance vector machine, extreme learning machine and multivariate adaptive regression spline models. **Stochastic Environmental Research and Risk Assessment**, v. 30, n. 6, p. 1769–1784, 2016.

DHANYA, C. T.; NAGESH KUMAR, D. Predictive uncertainty of chaotic daily streamflow using ensemble wavelet networks approach. **Water Resources Research**, v. 47, n. 6, 2011.

DIDAN, K. **MODIS/Terra Vegetation Indices 16-Day L3 Global 500m SIN Grid V061 [Data set]**. [S.l.]: NASA EOSDIS Land Processes DAAC, 2021. Disponível em: <<https://doi.org/10.5067/MODIS/MOD13A1.061>>. Acesso em: 25 jul. 2022.

DIENGDOH, V. L. *et al.* Predicted impacts of climate change and extreme temperature events on the future distribution of fruit bat species in Australia. **Global Ecology and Conservation**, v. 37, p. e02181, 1 set. 2022.

DUARTE, V. B. R. *et al.* Streamflow forecasting in Tocantins river basins using machine learning. **Water Supply**, v. 22, n. 7, p. 6230–6244, 2022.

EKOLU, J. *et al.* Long-term variability in hydrological droughts and floods in sub-Saharan Africa: New perspectives from a 65-year daily streamflow dataset. **Journal of Hydrology**, v. 613, 2022.

EMBRAPA SOLOS. **Mapas de solos do Brasil**. Disponível em: <http://geoinfo.cnps.embrapa.br/layers/geonode%3Abrasil_solos_5m_20201104#more>. Acesso em: 22 jun. 2022.

FANG, W. *et al.* Study on the evolution law of performance of mid- to long-term streamflow forecasting based on data-driven models. **Sustainable Cities and Society**, v. 88, p. 104277, 2023.

FATHIAN, F. *et al.* Hybrid models to improve the monthly river flow prediction: Integrating artificial intelligence and non-linear time series models. **Journal of Hydrology**, v. 575, p. 1200–1213, 1 ago. 2019.

FERDINANDY, B. *et al.* Challenges of machine learning model validation using correlated behaviour data: Evaluation of cross-validation strategies and accuracy measures. **PLoS ONE**, v. 15, n. 7, 1 jul. 2020.

FERREIRA, R. G. *et al.* Machine learning models for streamflow regionalization in a tropical watershed. **Journal of Environmental Management**, v. 280, 15 fev. 2021.

FRIEDMAN, J. H. Multivariate Adaptive Regression Splines. **The Annals of Statistics**, v. 19, n. 1, p. 1–141, 1991.

GIRIHAGAMA, L. *et al.* Streamflow modelling and forecasting for Canadian watersheds using LSTM networks with attention mechanism. **Neural Computing and Applications**, p. 1–21, 2022.

GRANATA, F.; DI NUNNO, F.; MARINIS, G. DE. Stacked machine learning algorithms and bidirectional long short-term memory networks for multi-step ahead streamflow forecasting: A comparative study. **Journal of Hydrology**, v. 613, p. 128431, 2022.

IRVING, K. *et al.* Data descriptor: A high-resolution streamflow and hydrological metrics dataset for ecological modeling using a regression model. **Scientific Data**, v. 5, p. 1–14, 2018.

JACCON, G.; CUDO, K. J. **Hidrologia-curva-chave: Análise e Traçado**. Brasília: DNAEE, 1989.

KAO, I. F. *et al.* Exploring a Long Short-Term Memory based Encoder-Decoder framework for multi-step-ahead flood forecasting. **Journal of Hydrology**, v. 583, p. 124631, 2020.

KARUNASINGHA, D. S. K. Root mean square error or mean absolute error? Use their ratio as well. **Information Sciences**, v. 585, p. 609–629, 2022.

KUHN, M.; JOHNSON, K. **Applied Predictive Modeling**. New York: [s.n.], 2013.

KUHN, MAX *et al.* **Package “caret”**: Classification and Regression Training. 2022. Disponível em: <https://cran.r-project.org/web/packages/caret/>. Acesso em: 20 jul. 2022.

LI, J. *et al.* Feature selection: A data perspective. **ACM Computing Surveys**, v. 50, n. 6, p. 1–45, 1 dez. 2017.

LI, X.; SHA, J.; WANG, Z. L. Comparison of daily streamflow forecasts using extreme learning machines and the random forest method. **Hydrological Sciences Journal**, v. 64, n. 15, p. 1857–1866, 2019.

LIU, Y. *et al.* Middle and Long-Term Runoff Probabilistic Forecasting Based on Gaussian Mixture Regression. **Water Resources Management**, v. 33, n. 5, p. 1785–1799, 2019.

LIU, Z. *et al.* Evaluating a coupled discrete wavelet transform and support vector regression for daily and monthly streamflow forecasting. **Journal of Hydrology**, v. 519, p. 2822–2831, 2014.

LIVIERIS, I. E. *et al.* Smoothing and stationarity enforcement framework for deep learning time-series forecasting. **Neural Computing and Applications**, v. 33, p. 14021–14035, 2021.

MAPBIOMAS. **Coleção [versão 6.0] da Série Anual de Mapas de Cobertura e Uso de Solo do Brasil**. Disponível em: <https://mapbiomas.org/>. Acesso em: 3 nov. 2021.

MIAO, J.; NIU, L. A Survey on Feature Selection. **Procedia Computer Science**, v. 91, p. 919–926, 2016.

MILBORROW, S. **Package “earth”**: Multivariate Adaptive Regression Splines. **CRAN**, 2022. Disponível em: <https://cran.r-project.org/web/packages/earth/>. Acesso em: 3 jul. 2021.

MORIASI, D. N. *et al.* Hydrologic and water quality models: Performance measures and evaluation criteria. **American Society of Agricultural and Biological Engineers**, v. 58, n. 6,

p. 1763–1785, 2015.

MOSAVI, A.; OZTURK, P.; CHAU, K. W. Flood prediction using machine learning models: Literature review. **Water (Switzerland)**, v. 10, n. 11, p. 1–40, 27 out. 2018.

NACHAPPA, T. G. *et al.* Flood susceptibility mapping with machine learning, multi-criteria decision analysis and ensemble using Dempster Shafer Theory. **Journal of Hydrology**, v. 590, p. 125275, nov. 2020.

NELDER, J. A.; WEDDERBURN, R. W. M. Generalized Linear Models. **Journal of the American Statistical Association**, v. 135, n. 3, p. 370–384, 1972.

NGUYEN, D. H. *et al.* Hourly streamflow forecasting using a Bayesian additive regression tree model hybridized with a genetic algorithm. **Journal of Hydrology**, v. 606, p. 127445, 2022.

PAPACHARALAMPOUS, G. A.; TYRALIS, H. Evaluation of random forests and Prophet for daily streamflow forecasting. **Advances in Geosciences**, v. 45, p. 201–208, 2018.

PEBESMA, E.; GRAELER, B. **Package “gstat”**: Spatial and Spatio-Temporal Geostatistical Modelling, Prediction and Simulation. p. 89 p., 2022. Disponível em: <https://cran.r-project.org/web/packages/gstat/>. Acesso em: 20 jul. 2022.

PHAM, B. T. *et al.* Can deep learning algorithms outperform benchmark machine learning algorithms in flood susceptibility modeling? **Journal of Hydrology**, v. 592, 1 jan. 2021.

QIU, D. **Package “aTSA”**: Alternative Time Series Analysis. *CRAN*. [S.l.: s.n.], 2015. Disponível em: <<https://cran.r-project.org/web/packages/aTSA/>>. Acesso em: 20 jul. 2022.

REIS, G. B. **Análise do regime hidrológico em bacia hidrográfica do sudeste brasileiro e previsão de séries temporais de vazões diárias utilizando aprendizado de máquina**. 2020. 125 f. Universidade Federal de Viçosa, 2020.

REIS, G. B. *et al.* Effect of environmental covariable selection in the hydrological modeling using machine learning models to predict daily streamflow. **Journal of Environmental Management**, v. 290, p. 112625, 15 jul. 2021.

SALVIANO, M. F.; MATOS, A. J. S. **Relatório Anual do Sistema de Alerta Hidrológico da bacia do rio Muriaé - 2021**. São Paulo: CPRM, 2021. Disponível em: <<https://rigeo.cprm.gov.br/handle/doc/22571>>. Acesso em: 10 jul. 2022.

SHEN, Y. *et al.* Random forests-based error-correction of streamflow from a large-scale hydrological model: Using model state variables to estimate error terms. **Computers and Geosciences**, v. 159, n. November 2021, p. 105019, 2022. Disponível em: <<https://doi.org/10.1016/j.cageo.2021.105019>>. Acesso em: 23 ago. 2022.

SILVEIRA, A. G. *et al.* Análise do Valor p Determinado pela Estatística τ na Aplicação do Teste de Dickey-Fuller Aumentado. **Trends in Computational and Applied Mathematics**, v. 23, n. 2, p. 283–298, 2022.

SILVEIRA, J. G. DA *et al.* Land Use, Land Cover Change and Sustainable Intensification of Agriculture and Livestock in the Amazon and the Atlantic Forest in Brazil. **Sustainability (Switzerland)**, v. 14, n. 5, p. 1–23, 2022.

SILVESTRO, F. *et al.* Impact-based flash-flood forecasting system: Sensitivity to high resolution numerical weather prediction systems and soil moisture. **Journal of Hydrology**, v. 572, p. 388–402, 1 maio 2019.

SINGH, A. K. *et al.* An Integrated Statistical-Machine Learning Approach for Runoff Prediction. **Sustainability (Switzerland)**, v. 14, n. 13, p. 1–30, 2022.

SIQUEIRA, V. A. *et al.* Potential skill of continental-scale, medium-range ensemble streamflow forecasts for flood prediction in South America. **Journal of Hydrology**, v. 590, 1 nov. 2020.

SOUZA, C. M. *et al.* Reconstructing three decades of land use and land cover changes in brazilian biomes with landsat archive and earth engine. **Remote Sensing**, v. 12, n. 17, 2020.

SOUZA, C. M. P. *et al.* Geomorphic indices, machine learning and ostracod palynology chronology to assess neotectonic deformation in the continental margin – Northeastern Brazil. **Journal of South American Earth Sciences**, v. 118, n. June, 2022.

THIESSEN, A. H. Precipitation average for large areas. **Monthly Weather Review**, v. 39, n. 7, p. 1082–1089, 1911.

TIWARI, M. K.; CHATTERJEE, C. Development of an accurate and reliable hourly flood forecasting model using wavelet-bootstrap-ANN (WBANN) hybrid approach. **Journal of Hydrology**, v. 394, n. 3–4, p. 458–470, 2010.

TONGAL, H.; BOOIJ, M. J. Simulation and forecasting of streamflows using machine learning models coupled with base flow separation. **Journal of Hydrology**, v. 564, p. 266–282, 1 set. 2018.

TRMM, T. R. M. M. **TRMM (TMPA) Rainfall Estimate L3 3 hour 0.25 degree x 0.25 degree V7**. Greenbelt, MD: Goddard Earth Sciences Data and Information Services Center (GES DISC), 2011. Disponível em: <https://doi.org/10.5067/TRMM/TMPA/3H/7>. Acesso em: 10 jul. 2022.

VELOSO, M. F.; RODRIGUES, L. N.; FILHO, E. I. F. Evaluation of machine learning algorithms in the prediction of hydraulic conductivity and soil moisture at the Brazilian Savannah. **Geoderma Regional**, v. 30, p. e00569, 1 set. 2022. Disponível em: <https://www.sciencedirect.com/science/article/abs/pii/S235200942200089X>. Acesso em: 20 jul. 2022.

WAN, Z.; HOOK, S.; HULLEY, G. **MODIS/Terra Land Surface Temperature/Emissivity 8-Day L3 Global 1km SIN Grid V061 [Data set]**. [S.l.]: NASA EOSDIS Land Processes DAAC, 2021. Disponível em: <https://doi.org/10.5067/MODIS/MOD11A2.061>. Acesso em: 10 jul. 2022.

XAVIER, A. C. *et al.* New improved Brazilian daily weather gridded data (1961–2020).

International Journal of Climatology, v. 42, n. 16, p. 8390-8404, 2022.

YIN, Z. *et al.* Design and evaluation of SVR, MARS and M5Tree models for 1, 2 and 3-day lead time forecasting of river flow data in a semiarid mountainous catchment. **Stochastic Environmental Research and Risk Assessment**, v. 32, n. 9, p. 2457–2476, 2018.

ZAHURA, F. T.; GOODALL, J. L. Predicting combined tidal and pluvial flood inundation using a machine learning surrogate model. **Journal of Hydrology: Regional Studies**, v. 41, p. 101087, 2022. Disponível em: <https://www.sciencedirect.com/science/article/pii/S2214581822001008>. Acesso em: 16 set. 2022.

CONCLUSÕES GERAIS

- A aplicação de diferentes métodos estatísticos permitiu identificar a ausência ou a presença de tendência das variáveis hidroclimatológicas em escala temporal e espacial da bacia do rio Muriaé.
- Os testes não-paramétricos utilizados indicaram tendências negativas em diferentes níveis de significância para as vazões mínimas, principalmente nos meses do período seco, com maior ocorrência de redução em outubro. Na escala anual, as vazões mínimas apresentaram tendência negativa em apenas uma estação.
- As vazões máximas apresentaram resultados semelhantes aos das vazões mínimas, embora não tenha sido encontrado nenhuma tendência na escala anual. Entretanto, ambos os resultados indicam o comprometimento da disponibilidade hídrica na bacia com mudanças da tendência nos anos de 1992 e 2012.
- Na análise dos fatores que alteram a vazão, a precipitação apresentou tendência negativa no período seco, com destaque para os meses de julho a outubro. Isso mostra haver uma correspondência entre a tendência da precipitação com a da vazão, indicando uma influência das chuvas no regime de vazões da bacia. Por outro lado, o uso e cobertura da terra apresentou pouca importância no comportamento hidrológico em vista da baixa variabilidade no período analisado.
- Ainda sobre os fatores, na exploração dos dados de demanda hídrica, verificou-se que o abastecimento público foi a classe mais importante na série analisada. Nessa análise, observou-se a possível influência que as concessões tiveram na vazão.
- Na modelagem da vazão, as covariáveis mais informativas para as técnicas de ML, de acordo com a seleção de recursos, foram das categorias vazão e precipitação;
- As técnicas de Aprendizado de Máquina Modelo Linear Generalizado (GLM), *Splines* de Regressão Adaptativa Multivariada (MARS) e o *Random Forest* (RF) apresentaram diferentes desempenhos nos cenários de dias de antecedência na previsão da vazão.
- O melhor desempenho foi observado para o prazo de entrega de 1 dia, em que todos os algoritmos tiveram desempenho muito bom. Todavia, ao aumentar os dias de antecedência houve um decréscimo na performance das técnicas, embora RF tenha se apresentado superior com menores erros.

- As técnicas de Aprendizado de Máquina GLM, MARS e RF se mostraram excelentes ferramentas a serem utilizadas na previsão da vazão, principalmente em bacias que sofrem frequentemente com eventos extremos de vazão.

APÊNDICES

APÊNDICE A

Tabela 1A – Vazões mínimas (Q_7) ocorridas nos meses ao longo da série histórica das estações fluviométricas para o período de 1971 a 2019 na bacia do rio Muriaé

Estação	Jan	Fev	Mar	Abr	Mai	Jun	Jul	Ago	Set	Out	Nov	Dez	Total
58920000						1		10	19	16	3		49
58930000	1							7	18	19	4		49
58934000	1							6	18	23	1		49
58940000			1					7	21	17	3		49
58960000								7	19	20	3		49

Fonte: A autora.

Tabela 2A. – Eventos máximos de vazões das estações fluviométricas associados aos períodos de retorno de 5, 10, 20, 50 e 100 anos (T), funções de densidade de probabilidade (PDF) e número de eventos iguais ou superiores ocorridos nas séries históricas (N)

Estação	T (anos)	Evento ($m^3.s^{-1}$)	FDP	N
58920000	5	429,86	Lognormal 3	38
	10	522,90	Lognormal 3	18
	20	610,50	Lognormal 3	8
	50	779,07	Gumbel	0
	100	876,63	Gumbel	0
58930000	5	146,52	Gumbel	27
	10	174,31	Lognormal 3	12
	20	222,81	Gumbel	5
	50	271,16	Gumbel	1
	100	307,39	Gumbel	0
58934000	5	233,13	Logpearson 3	17
	10	363,58	Gumbel	8
	20	448,10	Gumbel	6
	50	557,51	Gumbel	4
	100	639,50	Gumbel	1
58940000	5	809,15	Lognormal 3	44
	10	1006,27	Lognormal 3	19
	20	1268,54	Gumbel	7
	50	1535,64	Gumbel	1
	100	1735,79	Gumbel	0
58960000	5	800,21	Logpearson 3	41
	10	992,23	Logpearson 3	15
	20	1185,13	Logpearson 3	8
	50	1505,00	Gumbel	2
	100	1693,83	Gumbel	2

Fonte: A autora.

Tabela 3A. – Eventos máximos de vazão das estações fluviométricas, períodos de retorno (T) e funções de densidade de probabilidade (FDP) correspondentes, para a bacia do rio Muriaé

Estação	Vazão máxima (m³/s)	Período de Retorno (T)	Função de Densidade de Probabilidade (FDP)
58920000	738,95	38	Gumbel
58930000	298,47	84	Gumbel
58934000	740,44	235	Gumbel
58940000	1589,32	60	Gumbel
58960000	1849,99	177	Gumbel

Fonte: A autora.

Tabela 4A. – Valores da variável $Q_{7,10}$ para as estações fluviométricas e suas respectivas funções de densidade de probabilidade (FDP) para a bacia do rio Muriaé

Estação	$Q_{7,10}$ (m³/s)	Função de Densidade de Probabilidade (FDP)
58920000	6,30	Logpearson 3
58930000	1,64	Logpearson 3
58934000	2,05	Logpearson 3
58940000	9,05	Logpearson 3
58960000	11,19	Logpearson 3

Fonte: A autora.

APÊNDICE B

Tabela 1B – Estações de apoio utilizadas no preenchimento de falhas por regressão linear e seus respectivos parâmetros e ajuste.

Estação com Falha (variável y)	Estação de Apoio (variável x)	β_0	β_1	Coefficiente de Determinação (R²)
02041046	02141016	19,847	0,894	0,771
02042000	02042014	9,487	0,843	0,780
02042014	02142009	9,962	0,995	0,835
02042017	02042014	4,781	0,819	0,696
02042027	02042014	11,058	0,829	0,777
02141004	02141016	18,651	0,719	0,732
02141005	02141003	5,875	0,936	0,740
02141006	02141007	10,392	0,825	0,781
02141007	02142058	7,546	0,848	0,848
02142000	02142001	8,813	0,892	0,844
02142001	02142000	10,462	0,946	0,844
02142002	02042014	14,866	0,996	0,686
02142004	02042014	15,141	0,933	0,784
02142009	02042014	9,447	0,839	0,835
02142058	02141007	7,290	0,999	0,848

Fonte: A autora.

APÊNDICE C

Tabela 1C – Estatística dos testes não paramétricos de Mann Kendall ou Mann Kendall Modificado, correlação de Spearman e teste de Pettitt para as séries históricas de precipitação total anual

Estação	Mann-Kendall				Pettitt			Spearman			
	p-value	Sen's Slope	tau	Var(s)	p-value	U	Ano de mudança	p-value	S	Rho	R tabelado
02041005*	0,021	3,582	0,131	4072	0,112	233	1999/2000	0,196	14930	0,190	0,285
02041046	0,282	-4,351	-0,108	12659	0,282	192	2008/2009	0,266	21438	-0,164	0,285
02042000	0,015	7,016	0,245	12659	0,003	347	1997/1998	0,013	11852	0,357	0,285
02042014*	0,581	2,626	0,067	18486	0,415	172	1976/1977	0,684	17316	0,060	0,285
02042017	0,379	-2,942	-0,089	12659	0,386	176	2012/2013	0,326	21086	-0,144	0,285
02042027	0,408	-2,048	-0,083	12659	0,408	173	2011/2012	0,467	20398	-0,107	0,285
02141003	0,244	-2,688	-0,117	12659	0,244	199	2011/2012	0,285	21322	-0,157	0,285
02141004	0,034	-5,079	-0,213	12659	0,109	234	1996/1997	0,022	24538	-0,332	0,285
02141005	0,217	-2,628	-0,124	12659	0,171	215	2009/2010	0,197	21908	-0,189	0,285
02141006	0,936	-0,262	-0,009	12659	0,254	197	2011/2012	0,812	19072	-0,035	0,285
02141007	0,866	-0,803	-0,018	12659	0,748	136	2002/2003	0,905	18750	-0,018	0,285
02141014	0,389	3,010	0,087	12659	0,408	173	1989/1990	0,479	16502	0,104	0,285
02141016	0,032	-8,454	-0,215	12659	0,037	274	2008/2009	0,028	24282	-0,318	0,285
02142000	0,038	-5,124	-0,207	12659	0,171	215	2011/2012	0,026	24368	-0,323	0,285
02142001	0,133	-4,814	-0,151	12659	0,479	164	1984/1985	0,136	22440	-0,218	0,285
02142002*	0,420	5,517	0,119	27191	0,101	237	2011/2012	0,547	16786	0,089	0,285
02142004	0,389	-3,512	-0,087	12659	0,438	169	2011/2012	0,398	20716	-0,124	0,285
02142009	0,112	5,034	0,160	12659	0,239	200	1980/1981	0,209	15030	0,184	0,285
02142058	0,810	-0,930	-0,025	12659	0,716	139	2011/2012	0,683	19534	-0,060	0,285

*Teste de Mann-Kendall Modificado. Fonte: A autora.

Tabela 2C – Estatística dos testes não paramétricos de Mann Kendall ou Mann Kendall Modificado, correlação de Spearman e teste de Pettitt para as séries históricas de precipitação total do período chuvoso

Estação	Mann-Kendall				Pettitt			Spearman			
	p-value	Sen's Slope	tau	Var(s)	p-value	U	Ano de mudança	p-value	S	Rho	R tabelado
02041005*	0,049	5,469	0,163	8609	0,055	260	2000/2001	0,095	13932	0,244	0,285
02041046	0,894	-0,467	-0,014	12659	0,860	126	2008/2009	0,894	18788	-0,020	0,285
02042000	0,014	8,046	0,247	12657	0,003	347	1997/1998	0,010	11570	0,372	0,285
02042014*	0,465	3,976	0,112	29225	0,378	177	1977/1978	0,299	15612	0,153	0,285
02042017	0,979	-0,186	-0,004	12659	0,605	150	2012/2013	0,901	18764	-0,018	0,285
02042027	0,979	0,088	0,004	12659	0,576	153	2011/2012	0,898	18076	0,019	0,285
02141003*	1,000	-0,034	-0,001	8159	0,504	161	2011/2012	0,993	18399	0,001	0,285
02141004	0,298	-2,247	-0,105	12659	0,400	174	2008/2009	0,233	21652	-0,175	0,285
02141005	0,696	-0,737	-0,040	12658	0,438	169	2009/2010	0,661	19619	-0,065	0,285
02141006*	0,816	0,665	0,029	18951	0,357	180	2010/2011	0,844	17888	0,029	0,285
02141007	0,797	0,919	0,027	12659	0,357	180	2002/2003	0,730	17484	0,051	0,285
02141014	0,158	3,987	0,142	12659	0,224	203	1991/1992	0,262	15384	0,165	0,285
02141016	0,140	-4,512	-0,148	12658	0,061	256	2010/2011	0,098	22881	-0,242	0,285
02142000	0,324	-2,374	-0,099	12659	0,288	191	2011/2012	0,307	21194	-0,150	0,285
02142001	0,472	-1,846	-0,073	12659	0,895	123	2011/2012	0,431	20560	-0,116	0,285
02142002*	0,411	5,093	0,122	27803	0,224	203	2011/2012	0,313	15688	0,149	0,285
02142004	0,922	0,493	0,011	12659	0,849	127	2011/2012	0,970	18526	-0,006	0,285
02142009	0,077	5,430	0,177	12659	0,168	216	1998/1999	0,111	14130	0,233	0,285
02142058	0,663	0,900	0,044	12659	0,634	147	2001/2002	0,743	17532	0,048	0,285

*Teste de Mann-Kendall Modificado. Fonte: A autora.

Tabela 3C – Estatística dos testes não paramétricos de Mann Kendall ou Mann Kendall Modificado, correlação de Spearman e teste de Pettitt para as séries históricas de precipitação total do período seco

Estação	Mann-Kendall				Pettitt			Spearman			
	p-value	Sen's Slope	tau	Var(s)	p-value	U	Ano de mudança	p-value	S	Rho	R tabelado
02041005*	0,255	-0,660	-0,064	3898	0,781	133	2011/2012	0,491	20296	-0,102	0,285
02041046	0,003	-2,993	-0,301	12659	0,005	337	1991/1992	0,002	26494	-0,438	0,285
02042000*	0,428	0,339	0,046	4139	1,000	88	1998/1999	0,731	17488	0,051	0,285
02042014	0,223	-1,083	-0,122	12659	0,540	157	1997/1998	0,244	21574	-0,171	0,285
02042017	0,013	-2,254	-0,250	12659	0,082	245	1997/1998	0,009	25308	-0,374	0,285
02042027	0,106	-1,797	-0,162	12658	0,184	212	2004/2005	0,114	22686	-0,231	0,285
02141003	0,075	-1,342	-0,178	12658	0,431	170	1998/1999	0,082	23103	-0,254	0,285
02141004	0,040	-1,985	-0,206	12659	0,043	269	1997/1998	0,039	23942	-0,300	0,285
02141005	0,006	-1,867	-0,276	12658	0,045	267	1995/1996	0,005	25762	-0,398	0,285
02141006	0,638	-0,555	-0,048	12659	0,549	156	2004/2005	0,732	19358	-0,051	0,285
02141007	0,419	-0,723	-0,082	12659	0,895	123	1975/1976	0,373	20842	-0,131	0,285
02141014	1,000	0,006	0,001	12658	0,895	123	1985/1986	0,949	18250	0,009	0,285
02141016	0,004	-2,613	-0,289	12659	0,013	308	1998/1999	0,004	25916	-0,407	0,285
02142000*	0,002	-2,552	-0,305	12659	0,001	377	1997/1998	0,002	26562	-0,442	0,285
02142001*	0,001	-1,934	-0,234	5902	0,184	212	1984/1985	0,028	24282	-0,318	0,285
02142002	0,612	-0,506	-0,051	12659	0,300	189	2003/2004	0,576	19946	-0,083	0,285
02142004*	0,000	-1,819	-0,246	48	0,058	258	2004/2005	0,014	24978	-0,356	0,285
02142009	0,369	-1,067	-0,090	12659	0,364	179	2005/2006	0,478	20350	-0,105	0,285
02142058	0,114	-1,440	-0,159	12658	0,549	156	1982/1983	0,085	23056	-0,251	0,285

*Teste de Mann-Kendall Modificado. Fonte: A autora.

Tabela 4C – Estatística dos testes não paramétricos de Mann Kendall ou Mann Kendall Modificado, correlação de Spearman e teste de Pettitt para as séries históricas de precipitação total da estação 02041005

Mês	Mann-Kendall				Pettitt			Spearman			
	p-value	Sen's Slope	tau	Var(s)	p-value	U	Ano de mudança	p-value	S	Rho	R tabelado
Jan	0,356	-1,255	-0,092	13459	0,212	212	2009	0,346	22290	-0,137	0,282
Fev	0,464	0,703	0,073	13459	0,751	140	1999	0,447	17426	0,111	0,282
Mar	0,138	1,437	0,147	13458	0,448	173	2003	0,184	15819	0,193	0,282
Abr	0,289	0,745	0,105	13459	0,142	230	2003	0,266	16428	0,162	0,282
Mai*	0,756	0,111	0,031	13456	1,000	110	2009	0,848	19048	0,028	0,282
Jun	0,660	0,039	0,045	13411	1,000	104	1991	0,626	18203	0,071	0,282
Jul	0,769	-0,014	-0,030	13412	1,000	86	1992	0,805	20309	-0,036	0,282
Ago	0,959	0,000	-0,006	13410	1,000	95	1978	0,969	19490	0,006	0,282
Set	0,629	-0,153	-0,048	13458	1,000	115	2010	0,664	20848	-0,064	0,282
Out	0,485	-0,368	-0,070	13459	0,611	154	2011	0,488	21584	-0,101	0,282
Nov	0,093	1,600	0,167	13459	0,052	270	2000	0,086	14744	0,248	0,282
Dez	0,013	2,686	0,247	13459	0,020	304	1999	0,011	12478	0,363	0,282

*Teste de Mann-Kendall Modificado. Fonte: A autora.

Tabela 5C – Estatística dos testes não paramétricos de Mann Kendall ou Mann Kendall Modificado, correlação de Spearman e teste de Pettitt para as séries históricas de precipitação total da estação 02041046

Mês	Mann-Kendall				Pettitt			Spearman			
	p-value	Sen's Slope	tau	Var(s)	p-value	U	Ano de mudança	p-value	S	Rho	R tabelado
Jan	0,140	-1,810	-0,146	13459	0,186	218	2009	0,143	23754	-0,212	0,282
Fev	0,938	-0,089	-0,009	13459	0,521	164	2009	0,854	20126	-0,027	0,282
Mar	0,448	0,536	0,076	13458	0,368	184	2002	0,435	17364	0,114	0,282
Abr*	0,599	0,231	0,053	13459	0,650	150	2003	0,696	18480	0,057	0,282
Mai	0,750	0,087	0,032	13457	1,000	104	1973	0,724	18585	0,052	0,282
Jun	0,945	-0,010	-0,008	13455	1,000	116	1994	0,976	19688	-0,004	0,282
Jul	0,034	-0,335	-0,211	13441	0,030	290	1992	0,039	25392	-0,296	0,282
Ago	0,541	-0,132	-0,061	13457	0,882	128	1992	0,525	21424	-0,093	0,282
Set	0,098	-0,652	-0,164	13458	0,207	213	2008	0,097	24301	-0,240	0,282
Out*	0,000	-2,019	-0,391	13459	0,006	342	1990	0,000	30244	-0,543	0,282
Nov	0,850	0,197	0,020	13458	1,000	110	2016	0,827	18972	0,032	0,282
Dez	0,897	0,098	0,014	13459	1,000	90	1975	0,824	18964	0,032	0,282

*Teste de Mann-Kendall Modificado. Fonte: A autora.

Tabela 6C – Estatística dos testes não paramétricos de Mann Kendall ou Mann Kendall Modificado, correlação de Spearman e teste de Pettitt para as séries históricas de precipitação total da estação 02042000

Mês	Mann-Kendall				Pettitt			Spearman			
	p-value	Sen's Slope	tau	Var(s)	p-value	U	Ano de mudança	p-value	S	Rho	R tabelado
Jan	0,966	-0,105	-0,005	13459	0,342	188	2013	0,796	20342	-0,038	0,282
Fev	0,231	1,038	0,119	13459	0,118	238	2000	0,171	15712	0,198	0,282
Mar*	0,333	0,978	0,118	20301	0,145	229	1998	0,132	15321	0,218	0,282
Abr	0,031	1,396	0,214	13458	0,004	350	2003	0,023	13223	0,325	0,282
Mai*	0,507	0,139	0,066	13455	0,621	153	2000	0,510	17709	0,096	0,282
Jun	0,365	0,080	0,091	13427	0,464	171	2003	0,349	16918	0,137	0,282
Jul	0,515	-0,024	-0,067	13242	1,000	105	1979	0,530	21400	-0,092	0,282
Ago	0,382	0,093	0,088	13362	0,521	164	1999	0,385	17111	0,127	0,282
Set	0,924	0,043	0,010	13450	0,974	120	1988	0,782	18804	0,041	0,282
Out	0,208	-0,670	-0,125	13458	0,565	159	2000	0,198	23268	-0,187	0,282
Nov	0,035	2,143	0,209	13459	0,032	288	1997	0,035	13664	0,303	0,282
Dez*	0,013	2,417	0,280	17247	0,011	324	1999	0,007	12087	0,383	0,282

*Teste de Mann-Kendall Modificado. Fonte: A autora.

Tabela 7C – Estatística dos testes não paramétricos de Mann Kendall ou Mann Kendall Modificado, correlação de Spearman e teste de Pettitt para as séries históricas de precipitação total da estação 02042014

Mês	Mann-Kendall				Pettitt			Spearman			
	p-value	Sen's Slope	tau	Var(s)	p-value	U	Ano de mudança	p-value	S	Rho	R tabelado
Jan	0,897	0,226	0,014	13459	0,472	170	2009	0,869	19126	0,024	0,282
Fev	0,474	0,663	0,071	13457	0,837	132	1977	0,577	17999	0,082	0,282
Mar	0,202	1,025	0,127	13458	0,335	189	1980	0,196	15913	0,188	0,282
Abr	0,134	0,666	0,149	13458	0,260	202	2003	0,121	15202	0,224	0,282
Mai	0,448	0,209	0,076	13458	0,837	132	1973	0,473	17544	0,105	0,282
Jun	0,581	0,067	0,055	13449	1,000	105	2003	0,553	17900	0,087	0,282
Jul	0,382	-0,059	-0,088	13365	0,848	131	1995	0,492	21571	-0,101	0,282
Ago	0,365	0,113	0,090	13451	0,794	136	1978	0,417	17278	0,118	0,282
Set	0,374	-0,413	-0,089	13451	0,382	182	2009	0,415	21936	-0,119	0,282
Out	0,018	-1,505	-0,235	13459	0,080	254	1998	0,020	26132	-0,333	0,282
Nov	0,121	1,732	0,154	13458	0,145	229	1994	0,123	15218	0,224	0,282
Dez*	0,287	1,447	0,128	19567	0,178	220	1999	0,255	16358	0,165	0,282

*Teste de Mann-Kendall Modificado. Fonte: A autora.

Tabela 8C – Estatística dos testes não paramétricos de Mann Kendall ou Mann Kendall Modificado, correlação de Spearman e teste de Pettitt para as séries históricas de precipitação total da estação 02042017

Mês	Mann-Kendall				Pettitt			Spearman			
	p-value	Sen's Slope	tau	Var(s)	p-value	U	Ano de mudança	p-value	S	Rho	R tabelado
Jan	0,496	1,026	0,068	13459	0,472	170	1978	0,535	17826	0,091	0,282
Fev	0,692	-0,254	-0,040	13458	0,815	134	1992	0,675	20803	-0,061	0,282
Mar	0,789	0,358	0,027	13459	1,000	112	2013	0,768	18754	0,043	0,282
Abr	0,458	0,236	0,074	13458	0,472	170	2003	0,495	17645	0,100	0,282
Mai	0,453	-0,134	-0,075	13455	0,565	159	1998	0,328	22397	-0,143	0,282
Jun	0,795	0,000	-0,027	13364	1,000	112	1997	0,742	20544	-0,048	0,282
Jul*	0,000	-0,207	-0,254	5945	0,021	302	1995	0,011	26628	-0,359	0,282
Ago	0,722	0,000	-0,037	13293	1,000	109	1993	0,770	20440	-0,043	0,282
Set	0,737	-0,140	-0,034	13454	1,000	112	2009	0,630	20984	-0,071	0,282
Out	0,028	-1,292	-0,218	13459	0,044	276	1979	0,018	26220	-0,338	0,282
Nov	0,347	-0,776	-0,094	13459	0,472	170	1981	0,361	22210	-0,133	0,282
Dez	0,413	1,031	0,082	13459	0,927	124	1999	0,454	17458	0,109	0,282

*Teste de Mann-Kendall Modificado. Fonte: A autora.

Tabela 9C – Estatística dos testes não paramétricos de Mann Kendall ou Mann Kendall Modificado, correlação de Spearman e teste de Pettitt para as séries históricas de precipitação total da estação 02042027

Mês	Mann-Kendall				Pettitt			Spearman			
	p-value	Sen's Slope	tau	Var(s)	p-value	U	Ano de mudança	p-value	S	Rho	R tabelado
Jan	0,704	-0,564	-0,038	13458	0,368	184	2013	0,638	20949	-0,069	0,282
Fev	0,221	0,968	0,122	13458	0,382	182	1987	0,224	16131	0,177	0,282
Mar	0,123	1,067	0,153	13459	0,410	178	2002	0,112	15098	0,230	0,282
Abr	0,843	0,146	0,020	13457	0,974	120	2003	0,852	19064	0,027	0,282
Mai	0,641	0,098	0,047	13447	1,000	114	2011	0,624	18195	0,072	0,282
Jun	0,842	0,000	-0,021	13365	1,000	100	1986	0,856	20121	-0,027	0,282
Jul	0,048	-0,167	-0,202	13050	0,142	230	2002	0,058	24949	-0,273	0,282
Ago	0,931	0,000	-0,009	13390	1,000	87	1986	0,852	20137	-0,027	0,282
Set	0,518	-0,288	-0,065	13450	0,997	118	2006	0,530	21399	-0,092	0,282
Out	0,006	-1,687	-0,274	13459	0,178	220	1998	0,006	27234	-0,389	0,282
Nov	0,496	0,597	0,068	13459	0,456	172	1994	0,486	17610	0,102	0,282
Dez	0,356	-0,783	-0,092	13459	0,186	218	2010	0,282	22664	-0,156	0,282

Fonte: A autora.

Tabela 10C – Estatística dos testes não paramétricos de Mann Kendall ou Mann Kendall Modificado, correlação de Spearman e teste de Pettitt para as séries históricas de precipitação total da estação 02141003

Mês	Mann-Kendall				Pettitt			Spearman			
	p-value	Sen's Slope	tau	Var(s)	p-value	U	Ano de mudança	p-value	S	Rho	R tabelado
Jan	0,617	-0,368	-0,050	13458	0,472	170	2012	0,617	21037	-0,073	0,282
Fev	0,474	0,400	0,071	13459	0,951	122	1984	0,478	17572	0,103	0,282
Mar	0,076	1,064	0,176	13458	0,240	206	1980	0,056	14214	0,275	0,282
Abr	0,850	-0,073	-0,020	13458	0,403	179	1996	0,792	20359	-0,039	0,282
Mai	0,423	0,173	0,080	13459	0,730	142	2010	0,440	17392	0,113	0,282
Jun	0,635	0,063	0,048	13450	0,882	128	1981	0,555	17906	0,086	0,282
Jul	0,580	-0,068	-0,056	13392	1,000	106	1992	0,685	20766	-0,059	0,282
Ago	0,679	-0,052	-0,042	13430	0,740	141	1986	0,571	21225	-0,083	0,282
Set	0,448	-0,341	-0,076	13458	0,505	166	2006	0,487	21591	-0,102	0,282
Out	0,009	-1,236	-0,257	13459	0,112	240	1999	0,010	26804	-0,368	0,282
Nov	0,248	0,918	0,115	13458	0,130	234	1994	0,182	15797	0,194	0,282
Dez	0,581	-0,458	-0,055	13458	0,230	208	2010	0,507	21504	-0,097	0,282

Fonte: A autora.

Tabela 11C – Estatística dos testes não paramétricos de Mann Kendall ou Mann Kendall Modificado, correlação de Spearman e teste de Pettitt para as séries históricas de precipitação total da estação 02141004

Mês	Mann-Kendall				Pettitt			Spearman			
	p-value	Sen's Slope	tau	Var(s)	p-value	U	Ano de mudança	p-value	S	Rho	R tabelado
Jan	0,066	-1,879	-0,182	13459	0,084	252	2009	0,063	24860	-0,268	0,282
Fev	0,518	0,431	0,065	13459	0,927	124	1984	0,561	17938	0,085	0,282
Mar	0,231	0,893	0,119	13459	0,317	192	2008	0,250	16326	0,167	0,282
Abr	0,032	-0,992	-0,213	13459	0,004	352	1996	0,014	26440	-0,349	0,282
Mai*	0,220	-0,223	-0,078	5507	0,539	162	1998	0,363	22198	-0,133	0,282
Jun	0,301	-0,182	-0,103	13449	0,382	182	1997	0,362	22210	-0,133	0,282
Jul	0,057	-0,390	-0,189	13454	0,102	244	1996	0,079	24559	-0,253	0,282
Ago	0,485	-0,157	-0,070	13448	0,207	213	1986	0,353	22259	-0,136	0,282
Set	0,221	-0,544	-0,122	13458	0,530	163	2003	0,259	22819	-0,164	0,282
Out	0,111	-0,918	-0,158	13459	0,382	182	1976	0,099	24268	-0,238	0,282
Nov*	0,408	-0,484	-0,048	4426	1,000	80	1981	0,615	21040	-0,073	0,282
Dez	0,966	0,069	0,005	13459	0,730	142	2010	0,938	19378	0,011	0,282

*Teste de Mann-Kendall Modificado. Fonte: A autora.

Tabela 12C – Estatística dos testes não paramétricos de Mann Kendall ou Mann Kendall Modificado, correlação de Spearman e teste de Pettitt para as séries históricas de precipitação total da estação 02141005

Mês	Mann-Kendall				Pettitt			Spearman			
	p-value	Sen's Slope	tau	Var(s)	p-value	U	Ano de mudança	p-value	S	Rho	R tabelado
Jan	0,599	-0,518	-0,053	13459	0,488	168	2013	0,591	21136	-0,078	0,282
Fev	0,730	-0,230	-0,035	13458	0,235	207	2009	0,676	20800	-0,061	0,282
Mar	0,564	0,325	0,058	13459	0,772	138	2003	0,472	17542	0,105	0,282
Abr	0,347	-0,367	-0,094	13459	0,396	180	1995	0,336	22346	-0,140	0,282
Mai	1,000	0,000	0,000	13455	0,456	172	2011	0,930	19346	0,013	0,282
Jun	0,125	-0,219	-0,153	13428	0,152	227	1994	0,137	23826	-0,216	0,282
Jul	0,214	-0,151	-0,124	13436	0,225	209	1995	0,178	23435	-0,196	0,282
Ago	0,248	-0,178	-0,116	13432	0,080	254	1986	0,141	23781	-0,213	0,282
Set	0,168	-0,523	-0,137	13456	0,152	227	2003	0,173	23473	-0,198	0,282
Out*	0,017	-1,142	-0,236	13458	0,102	244	2011	0,013	26492	-0,352	0,282
Nov	0,163	0,794	0,139	13458	0,464	171	1995	0,213	16051	0,181	0,282
Dez	0,698	0,384	0,039	13459	0,440	174	2011	0,754	18704	0,046	0,282

*Teste de Mann-Kendall Modificado. Fonte: A autora.

Tabela 13C – Estatística dos testes não paramétricos de Mann Kendall ou Mann Kendall Modificado, correlação de Spearman e teste de Pettitt para as séries históricas de precipitação total da estação 02141006

Mês	Mann-Kendall				Pettitt			Spearman			
	p-value	Sen's Slope	tau	Var(s)	p-value	U	Ano de mudança	p-value	S	Rho	R tabelado
Jan	0,464	0,786	0,073	13459	0,574	158	2012	0,614	18156	0,074	0,282
Fev	0,986	-0,033	-0,003	13458	0,985	119	2009	0,880	20034	-0,022	0,282
Mar	0,094	0,876	0,166	13458	0,178	220	2004	0,076	14587	0,256	0,282
Abr	0,524	-0,222	-0,064	13458	0,871	129	1996	0,553	21302	-0,087	0,282
Mai	0,809	-0,078	-0,025	13458	1,000	89	2011	0,864	20094	-0,025	0,282
Jun	0,836	0,025	0,021	13426	0,699	145	2003	0,795	18852	0,038	0,282
Jul	0,979	0,000	-0,003	13411	1,000	79	2001	0,961	19740	-0,007	0,282
Ago	1,000	0,000	-0,001	13441	0,859	130	1986	0,946	19796	-0,010	0,282
Set	0,654	-0,152	-0,045	13453	0,448	173	2003	0,659	20867	-0,065	0,282
Out	0,058	-1,038	-0,188	13458	0,052	270	2011	0,043	25298	-0,291	0,282
Nov	0,485	0,546	0,070	13459	0,396	180	1995	0,480	17582	0,103	0,282
Dez	0,796	0,298	0,026	13458	0,362	185	2008	0,992	19571	0,001	0,282

Fonte: A autora.

Tabela 14C – Estatística dos testes não paramétricos de Mann Kendall ou Mann Kendall Modificado, correlação de Spearman e teste de Pettitt para as séries históricas de precipitação total da estação 02141007

Mês	Mann-Kendall				Pettitt			Spearman			
	p-value	Sen's Slope	tau	Var(s)	p-value	U	Ano de mudança	p-value	S	Rho	R tabelado
Jan	0,816	-0,251	-0,024	13459	0,794	136	2013	0,835	20196	-0,030	0,282
Fev	0,803	-0,155	-0,026	13459	1,000	90	1998	0,811	20286	-0,035	0,282
Mar	0,330	0,538	0,097	13459	0,317	192	2007	0,308	16694	0,148	0,282
Abr	0,458	-0,210	-0,074	13458	0,630	152	1995	0,513	21476	-0,096	0,282
Mai*	0,709	0,116	0,038	13852	0,539	162	2010	0,751	18689	0,046	0,282
Jun	0,938	0,016	0,009	13457	0,669	148	2003	0,935	19364	0,012	0,282
Jul	0,883	0,020	0,015	13453	1,000	113	2001	0,859	19091	0,026	0,282
Ago	0,617	-0,116	-0,050	13453	0,230	208	1986	0,545	21334	-0,088	0,282
Set	0,724	-0,122	-0,036	13459	0,905	126	2003	0,815	20272	-0,034	0,282
Out	0,029	-0,998	-0,216	13459	0,118	238	2011	0,033	25588	-0,306	0,282
Nov	0,086	1,190	0,170	13459	0,124	236	1995	0,091	14812	0,244	0,282
Dez*	0,878	0,290	0,022	26441	0,440	174	2010	0,917	19898	-0,015	0,282

*Teste de Mann-Kendall Modificado. Fonte: A autora.

Tabela 15C – Estatística dos testes não paramétricos de Mann Kendall ou Mann Kendall Modificado, correlação de Spearman e teste de Pettitt para as séries históricas de precipitação total da estação 02141014

Mês	Mann-Kendall				Pettitt			Spearman			
	p-value	Sen's Slope	tau	Var(s)	p-value	U	Ano de mudança	p-value	S	Rho	R tabelado
Jan	0,704	-0,406	-0,038	13458	0,250	204	2009	0,556	21287	-0,086	0,282
Fev	0,231	0,737	0,119	13459	0,574	158	1987	0,221	16114	0,178	0,282
Mar	0,010	2,042	0,253	13459	0,027	294	2002	0,009	12304	0,372	0,282
Abr	0,202	0,547	0,127	13458	0,127	235	1985	0,193	15892	0,189	0,282
Mai	0,823	0,095	0,023	13458	1,000	100	1984	0,775	18779	0,042	0,282
Jun	0,029	0,456	0,216	13455	0,080	254	1986	0,024	13282	0,322	0,282
Jul	0,666	-0,067	-0,043	13456	0,916	125	1992	0,633	20972	-0,070	0,282
Ago	0,979	-0,003	-0,003	13457	1,000	97	1990	0,972	19498	0,005	0,282
Set	0,068	-0,671	-0,181	13458	0,216	211	2005	0,071	24707	-0,261	0,282
Out	0,322	-0,492	-0,099	13459	0,396	180	2011	0,300	22556	-0,151	0,282
Nov	0,089	1,703	0,169	13457	0,029	291	1994	0,081	14673	0,251	0,282
Dez	0,443	0,617	0,077	13459	0,282	198	1994	0,600	18100	0,077	0,282

Fonte: A autora.

Tabela 16C – Estatística dos testes não paramétricos de Mann Kendall ou Mann Kendall Modificado, correlação de Spearman e teste de Pettitt para as séries históricas de precipitação total da estação 02141016

Mês	Mann-Kendall				Pettitt			Spearman			
	p-value	Sen's Slope	tau	Var(s)	p-value	U	Ano de mudança	p-value	S	Rho	R tabelado
Jan	0,056	-2,185	-0,190	13458	0,028	292	2007	0,042	25316	-0,292	0,282
Fev	0,776	-0,203	-0,029	13459	0,505	166	2009	0,733	20576	-0,050	0,282
Mar	0,856	-0,294	-0,019	13459	1,000	78	2013	0,822	20244	-0,033	0,282
Abr	0,692	-0,120	-0,040	13458	0,794	136	1996	0,835	20199	-0,031	0,282
Mai*	0,763	-0,081	-0,026	9287	0,882	128	2011	0,770	20436	-0,043	0,282
Jun	0,581	0,104	0,055	13456	0,689	146	1987	0,593	18067	0,078	0,282
Jul	0,196	-0,236	-0,129	13439	0,133	233	1992	0,233	23002	-0,174	0,282
Ago	0,163	-0,227	-0,139	13458	0,659	149	2000	0,192	23315	-0,190	0,282
Set*	0,035	-0,784	-0,209	13459	0,072	258	2003	0,033	25590	-0,306	0,282
Out	0,001	-1,825	-0,325	13459	0,022	300	1999	0,001	28422	-0,450	0,282
Nov	0,281	-1,255	-0,107	13459	0,472	170	2006	0,241	22938	-0,170	0,282
Dez	0,274	-1,004	-0,109	13459	0,178	220	2010	0,195	23288	-0,188	0,282

*Teste de Mann-Kendall Modificado. Fonte: A autora.

Tabela 17C – Estatística dos testes não paramétricos de Mann Kendall ou Mann Kendall Modificado, correlação de Spearman e teste de Pettitt para as séries históricas de precipitação total da estação 02142000

Mês	Mann-Kendall				Pettitt			Spearman			
	p-value	Sen's Slope	tau	Var(s)	p-value	U	Ano de mudança	p-value	S	Rho	R tabelado
Jan	0,458	-1,262	-0,074	13458	0,772	138	2012	0,504	21517	-0,098	0,282
Fev	0,297	-1,191	-0,104	13459	0,611	154	2006	0,355	22238	-0,135	0,282
Mar	0,850	0,124	0,020	13458	1,000	96	1980	0,820	18946	0,033	0,282
Abr*	0,413	-0,377	-0,082	13459	0,342	188	1995	0,494	21554	-0,100	0,282
Mai	0,464	-0,257	-0,073	13457	0,859	130	1996	0,474	21654	-0,105	0,282
Jun	0,972	0,000	-0,004	13412	1,000	100	2003	0,993	19625	-0,001	0,282
Jul	0,592	-0,036	-0,054	13390	0,893	127	1992	0,683	20773	-0,060	0,282
Ago	0,198	-0,252	-0,129	13389	0,174	221	1986	0,179	23426	-0,195	0,282
Set	0,575	-0,235	-0,056	13459	0,689	146	2003	0,601	21096	-0,076	0,282
Out*	0,039	-1,552	-0,276	24652	0,012	319	1998	0,004	27570	-0,407	0,282
Nov	0,593	0,697	0,054	13458	0,804	135	1995	0,528	17789	0,092	0,282
Dez	0,552	-0,745	-0,060	13459	0,305	194	2010	0,549	21316	-0,088	0,282

*Teste de Mann-Kendall Modificado. Fonte: A autora.

Tabela 18C – Estatística dos testes não paramétricos de Mann Kendall ou Mann Kendall Modificado, correlação de Spearman e teste de Pettitt para as séries históricas de precipitação total da estação 02142001

Mês	Mann-Kendall				Pettitt			Spearman			
	p-value	Sen's Slope	tau	Var(s)	p-value	U	Ano de mudança	p-value	S	Rho	R tabelado
Jan	0,924	0,080	0,010	13459	0,669	148	1977	0,989	19558	0,002	0,282
Fev	0,535	-0,484	-0,062	13458	0,761	139	2009	0,536	21377	-0,091	0,282
Mar	0,979	-0,056	-0,003	13459	1,000	86	1985	0,941	19812	-0,011	0,282
Abr*	0,001	-1,011	-0,196	5024	0,030	290	1994	0,044	25280	-0,290	0,282
Mai*	0,055	-0,259	-0,109	4395	0,837	132	1984	0,352	22254	-0,135	0,282
Jun	0,512	-0,112	-0,066	13447	0,761	139	1989	0,496	21550	-0,099	0,282
Jul	0,724	-0,034	-0,036	13440	1,000	108	1992	0,747	20526	-0,047	0,282
Ago	0,152	-0,207	-0,142	13456	0,070	259	1986	0,145	23743	-0,211	0,282
Set	0,599	-0,259	-0,053	13459	0,505	166	2003	0,598	21110	-0,077	0,282
Out	0,030	-1,173	-0,214	13459	0,163	224	2011	0,026	25852	-0,319	0,282
Nov	0,763	0,392	0,031	13459	0,882	128	1981	0,834	18998	0,031	0,282
Dez	0,850	-0,185	-0,020	13458	0,905	126	1989	0,760	20476	-0,045	0,282

*Teste de Mann-Kendall Modificado. Fonte: A autora.

Tabela 19C – Estatística dos testes não paramétricos de Mann Kendall ou Mann Kendall Modificado, correlação de Spearman e teste de Pettitt para as séries históricas de precipitação total da estação 02142002

Mês	Mann-Kendall				Pettitt			Spearman			
	p-value	Sen's Slope	tau	Var(s)	p-value	U	Ano de mudança	p-value	S	Rho	R tabelado
Jan	0,660	0,647	0,044	13459	0,425	176	2013	0,884	19182	0,021	0,282
Fev	0,490	0,485	0,069	13458	0,418	177	1984	0,492	17630	0,101	0,282
Mar*	0,307	1,200	0,104	14027	0,521	164	1992	0,287	16559	0,155	0,282
Abr	0,783	-0,156	-0,028	13458	0,783	137	1995	0,910	19925	-0,017	0,282
Mai	0,737	-0,099	-0,034	13459	0,927	124	2004	0,769	20442	-0,043	0,282
Jun	0,816	0,050	0,024	13442	1,000	84	1999	0,950	19420	0,009	0,282
Jul	0,316	-0,159	-0,101	13408	0,521	164	2005	0,344	22305	-0,138	0,282
Ago	0,769	0,033	0,030	13451	1,000	109	1986	0,886	19188	0,021	0,282
Set	0,529	-0,271	-0,063	13459	0,410	178	2003	0,558	21278	-0,086	0,282
Out	0,297	-0,698	-0,104	13459	0,305	194	2011	0,257	22832	-0,165	0,282
Nov	0,064	2,503	0,184	13459	0,052	270	1994	0,071	14492	0,261	0,282
Dez*	0,562	1,115	0,090	32817	0,186	218	1994	0,475	17558	0,104	0,282

*Teste de Mann-Kendall Modificado. Fonte: A autora.

Tabela 20C – Estatística dos testes não paramétricos de Mann Kendall ou Mann Kendall Modificado, correlação de Spearman e teste de Pettitt para as séries históricas de precipitação total da estação 02142004

Mês	Mann-Kendall				Pettitt			Spearman			
	p-value	Sen's Slope	tau	Var(s)	p-value	U	Ano de mudança	p-value	S	Rho	R tabelado
Jan	0,743	-0,405	-0,033	13458	0,574	158	2013	0,767	20453	-0,044	0,282
Fev	0,558	-0,519	-0,059	13458	1,000	116	2009	0,722	20620	-0,052	0,282
Mar	0,289	1,026	0,105	13459	0,329	190	2003	0,285	16554	0,155	0,282
Abr*	0,149	0,286	0,077	3896	0,190	217	2002	0,407	17227	0,121	0,282
Mai	0,558	-0,213	-0,059	13458	0,521	164	1994	0,687	20757	-0,059	0,282
Jun	0,557	-0,074	-0,059	13412	0,997	118	1976	0,572	21221	-0,083	0,282
Jul	0,250	-0,118	-0,115	13391	0,583	157	2011	0,233	23000	-0,173	0,282
Ago	0,938	-0,011	-0,009	13450	0,448	173	1986	0,963	19733	-0,007	0,282
Set	0,041	-0,871	-0,203	13457	0,051	271	2003	0,044	25277	-0,290	0,282
Out	0,080	-0,961	-0,173	13459	0,072	258	2011	0,082	24514	-0,251	0,282
Nov	0,966	-0,077	-0,005	13459	0,794	136	1981	0,960	19746	-0,007	0,282
Dez	0,660	-0,478	-0,044	13459	0,574	158	1988	0,671	20816	-0,062	0,282

*Teste de Mann-Kendall Modificado. Fonte: A autora.

Tabela 21C – Estatística dos testes não paramétricos de Mann Kendall ou Mann Kendall Modificado, correlação de Spearman e teste de Pettitt para as séries históricas de precipitação total da estação 02142009

Mês	Mann-Kendall				Pettitt			Spearman			
	p-value	Sen's Slope	tau	Var(s)	p-value	U	Ano de mudança	p-value	S	Rho	R tabelado
Jan	0,507	0,603	0,066	13459	0,396	180	1978	0,566	17956	0,084	0,282
Fev	0,365	0,759	0,090	13459	0,556	160	2003	0,300	16644	0,151	0,282
Mar	0,458	0,609	0,074	13458	0,669	148	1992	0,407	17224	0,121	0,282
Abr	0,297	0,525	0,104	13459	0,317	192	2003	0,217	16088	0,179	0,282
Mai	0,356	0,202	0,092	13459	0,342	188	1982	0,383	17110	0,127	0,282
Jun	0,795	0,002	0,027	13390	1,000	99	1987	0,785	18816	0,040	0,282
Jul	0,213	-0,131	-0,125	13366	0,418	177	1992	0,169	23517	-0,200	0,282
Ago	0,301	-0,112	-0,104	13436	0,425	176	1986	0,285	22652	-0,156	0,282
Set	0,986	-0,012	-0,003	13458	0,250	204	1982	0,809	18907	0,035	0,282
Out	0,123	-1,001	-0,153	13459	0,124	236	2011	0,132	23876	-0,218	0,282
Nov	0,238	1,115	0,117	13459	0,293	196	1994	0,238	16236	0,172	0,282
Dez*	0,534	0,698	0,051	8986	0,593	156	1998	0,768	18751	0,043	0,282

*Teste de Mann-Kendall Modificado. Fonte: A autora.

Tabela 22C – Estatística dos testes não paramétricos de Mann Kendall ou Mann Kendall Modificado, correlação de Spearman e teste de Pettitt para as séries históricas de precipitação total da estação 02142058

Mês	Mann-Kendall				Pettitt			Spearman			
	p-value	Sen's Slope	tau	Var(s)	p-value	U	Ano de mudança	p-value	S	Rho	R tabelado
Jan	0,776	-0,256	-0,029	13457	0,480	169	2013	0,734	20575	-0,050	0,282
Fev	0,959	0,088	0,006	13458	1,000	114	2009	0,944	19397	0,010	0,282
Mar*	0,343	0,781	0,094	13458	0,271	200	2003	0,395	17163	0,124	0,282
Abr	0,546	-0,220	-0,060	13458	0,689	146	1997	0,563	21261	-0,085	0,282
Mai	0,623	-0,147	-0,049	13457	0,410	178	1988	0,622	21014	-0,072	0,282
Jun	0,518	-0,102	-0,065	13457	0,488	168	1989	0,530	21402	-0,092	0,282
Jul	0,763	-0,033	-0,031	13440	1,000	111	2001	0,784	20386	-0,040	0,282
Ago*	0,356	-0,177	-0,092	13457	0,194	216	1986	0,300	22562	-0,151	0,282
Set	0,289	-0,399	-0,106	13454	0,602	155	2003	0,344	22305	-0,138	0,282
Out	0,038	-1,225	-0,206	13457	0,051	271	2011	0,030	25684	-0,310	0,282
Nov	0,228	0,949	0,120	13458	0,194	216	2009	0,255	16348	0,166	0,282
Dez	0,911	-0,132	-0,012	13457	0,871	129	2011	0,828	20224	-0,032	0,282

*Teste de Mann-Kendall Modificado. Fonte: A autora.

APÊNDICE D

Tabela 1D – Estatística descritiva da variável objetivo e das covariáveis utilizadas no cenário de 1 dia de antecedência

Variável	Média	Mediana	Mínimo	Máximo	Desvio Padrão	CV (%)*
Y (m ³ /s)	94,21	54,52	3,14	1849,99	117,84	125,08
Q1 _t (m ³ /s)	45,91	27,20	1,86	649,47	54,31	118,30
Q1 _{t-1} (m ³ /s)	45,91	27,20	1,86	649,47	54,31	118,30
Q1 _{t-2} (m ³ /s)	45,91	27,20	1,86	649,47	54,31	118,30
Q2 _t (m ³ /s)	12,49	7,51	0,72	228,45	15,51	124,20
Q2 _{t-1} (m ³ /s)	12,49	7,51	0,72	228,45	15,51	124,20
Q2 _{t-2} (m ³ /s)	12,49	7,51	0,72	228,45	15,51	124,20
Q3 _t (m ³ /s)	20,75	12,22	0,90	740,44	28,75	138,54
Q3 _{t-1} (m ³ /s)	20,75	12,22	0,90	740,44	28,75	138,54
Q3 _{t-2} (m ³ /s)	20,75	12,22	0,90	740,44	28,75	138,54
Q4 _t (m ³ /s)	84,63	50,39	1,22	1589,32	110,14	130,15
Q4 _{t-1} (m ³ /s)	84,62	50,39	1,22	1589,32	110,14	130,16
Q4 _{t-2} (m ³ /s)	84,62	50,39	1,22	1589,32	110,14	130,16
p_idw _t (mm)	3,73	0,45	0,00	69,23	7,40	198,48
p_idw _{t-1} (mm)	3,73	0,45	0,00	69,23	7,40	198,49
p_idw _{t-2} (mm)	3,73	0,45	0,00	69,23	7,40	198,49
p_idw _{48h_t} (mm)	7,45	1,56	0,00	108,71	12,74	170,93
p_idw _{72h_t} (mm)	11,18	3,20	0,00	153,70	17,30	154,78
p_idw _{96h_t} (mm)	14,90	5,24	0,00	184,16	21,44	143,88
p_idw _{120h_t} (mm)	18,63	7,55	0,00	204,31	25,33	135,96
p_idw _{1sem_t} (mm)	26,08	12,21	0,00	248,20	32,81	125,80
p_xavier _t (mm)	3,73	0,40	0,00	71,64	7,52	201,70
p_xavier _{t-1} (mm)	3,73	0,40	0,00	71,64	7,52	201,71
p_xavier _{t-2} (mm)	3,73	0,40	0,00	71,64	7,52	201,70
p_xavier _{48h_t} (mm)	7,46	1,50	0,00	114,98	12,87	172,46
p_xavier _{72h_t} (mm)	11,19	3,10	0,00	162,92	17,41	155,60
p_xavier _{96h_t} (mm)	14,92	5,08	0,00	191,27	21,54	144,34
p_xavier _{120h_t} (mm)	18,65	7,42	0,00	211,18	25,41	136,21
p_xavier _{1sem_t} (mm)	26,11	12,37	0,00	253,16	32,86	125,82
p_trmm _t (mm)	3,71	0,01	0,00	99,46	8,69	234,42
p_trmm _{t-1} (mm)	3,71	0,01	0,00	99,46	8,69	234,42
p_trmm _{t-2} (mm)	3,71	0,01	0,00	99,46	8,69	234,42
p_trmm _{48h_t} (mm)	7,42	0,42	0,00	102,37	14,10	190,11
p_trmm _{72h_t} (mm)	11,12	1,62	0,00	133,85	18,62	167,34
p_trmm _{96h_t} (mm)	14,83	3,36	0,00	177,42	22,71	153,11
p_trmm _{120h_t} (mm)	18,54	5,72	0,00	179,83	26,54	143,14
p_trmm _{1sem_t} (mm)	25,96	10,71	0,00	195,19	33,95	130,81
FL (%)	18,32	18,34	17,41	19,26	0,53	2,91
FN (%)	0,06	0,06	0,06	0,07	0,00	6,43
PA (%)	67,67	68,62	63,24	69,76	2,05	3,03
AG (%)	12,67	12,20	11,39	15,72	1,41	11,11
FP (%)	0,35	0,38	0,04	0,72	0,25	70,99

(cont.)

Variável	Média	Mediana	Mínimo	Máximo	Desvio Padrão	CV (%)*
AnV (%)	0,69	0,67	0,57	0,84	0,09	12,42
CH (%)	0,24	0,24	0,22	0,24	0,01	2,59
$temp_min_t$ (°C)	18,10	18,25	-0,21	27,39	3,59	19,85
$temp_med_t$ (°C)	28,18	28,46	17,97	38,44	3,38	11,98
$temp_max_t$ (°C)	34,15	34,31	20,39	50,47	4,61	13,49
evi_min_t (adimensional)	0,07	0,07	-0,13	0,19	0,05	78,12
evi_med_t (adimensional)	0,42	0,43	0,22	0,57	0,08	19,20
evi_max_t (adimensional)	0,61	0,67	0,00	0,90	0,22	35,77
$ndvi_min_t$ (adimensional)	0,11	0,09	-0,20	0,37	0,09	90,18
$ndvi_med_t$ (adimensional)	0,66	0,68	0,41	0,78	0,08	12,58
$ndvi_max_t$ (adimensional)	0,83	0,91	0,00	0,99	0,25	30,76
$escoam_min_t$ (mm)	15,67	2,00	0,00	213,00	37,83	241,39
$escoam_med_t$ (mm)	29,09	3,40	0,00	326,11	59,80	205,61
$escoam_max_t$ (mm)	49,50	5,00	0,00	456,00	87,39	176,54
$umidade_min_t$ (mm)	52,91	39,60	9,70	107,70	35,51	67,12
$umidade_med_t$ (mm)	99,28	100,63	20,94	158,59	46,45	46,79
$umidade_max_t$ (mm)	150,59	165,30	36,70	199,30	47,34	31,43
mês	6,56	7,00	1,00	12,00	3,43	52,32

*Coeficiente de Variação. Fonte: A autora.

Tabela 2D – Estatística descritiva da variável objetivo e das covariáveis utilizadas no cenário de 2 dias de antecedência

Variável	Média	Mediana	Mínimo	Máximo	Desvio Padrão	CV (%)*
Y (m³/s)	94,21	54,52	3,14	1849,99	117,84	125,08
$Q1_t$ (m³/s)	45,91	27,20	1,86	649,47	54,31	118,30
$Q1_{t-1}$ (m³/s)	45,91	27,20	1,86	649,47	54,31	118,30
$Q1_{t-2}$ (m³/s)	45,91	27,20	1,86	649,47	54,31	118,30
$Q2_t$ (m³/s)	12,49	7,51	0,72	228,45	15,51	124,20
$Q2_{t-1}$ (m³/s)	12,49	7,51	0,72	228,45	15,51	124,20
$Q2_{t-2}$ (m³/s)	12,49	7,51	0,72	228,45	15,51	124,20
$Q3_t$ (m³/s)	20,75	12,22	0,90	740,44	28,75	138,54
$Q3_{t-1}$ (m³/s)	20,75	12,22	0,90	740,44	28,75	138,54
$Q3_{t-2}$ (m³/s)	20,75	12,22	0,90	740,44	28,75	138,54
$Q4_t$ (m³/s)	84,62	50,39	1,22	1589,32	110,14	130,16
$Q4_{t-1}$ (m³/s)	84,62	50,39	1,22	1589,32	110,14	130,16
$Q4_{t-2}$ (m³/s)	84,61	50,39	1,22	1589,32	110,14	130,17
p_idw_t (mm)	3,73	0,45	0,00	69,23	7,40	198,49
p_idw_{t-1} (mm)	3,73	0,45	0,00	69,23	7,40	198,49
p_idw_{t-2} (mm)	3,73	0,45	0,00	69,23	7,40	198,50
$p_idw_{48h_t}$ (mm)	7,45	1,56	0,00	108,71	12,74	170,94
$p_idw_{72h_t}$ (mm)	11,18	3,19	0,00	153,70	17,30	154,79
$p_idw_{96h_t}$ (mm)	14,90	5,24	0,00	184,16	21,44	143,88
$p_idw_{120h_t}$ (mm)	18,63	7,55	0,00	204,31	25,33	135,96
$p_idw_{1sem_t}$ (mm)	26,08	12,21	0,00	248,20	32,81	125,80

(cont.)

Variável	Média	Mediana	Mínimo	Máximo	Desvio Padrão	CV (%)*
p_{xavier_t} (mm)	3,73	0,40	0,00	71,64	7,52	201,71
$p_{xavier_{t-1}}$ (mm)	3,73	0,40	0,00	71,64	7,52	201,70
$p_{xavier_{t-2}}$ (mm)	3,73	0,40	0,00	71,64	7,52	201,70
$p_{xavier_{48h_t}}$ (mm)	7,46	1,50	0,00	114,98	12,87	172,46
$p_{xavier_{72h_t}}$ (mm)	11,19	3,10	0,00	162,92	17,41	155,60
$p_{xavier_{96h_t}}$ (mm)	14,92	5,08	0,00	191,27	21,54	144,34
$p_{xavier_{120h_t}}$ (mm)	18,65	7,42	0,00	211,18	25,41	136,20
$p_{xavier_{1sem_t}}$ (mm)	26,11	12,37	0,00	253,16	32,86	125,82
p_{trmm_t} (mm)	3,71	0,01	0,00	99,46	8,69	234,42
$p_{trmm_{t-1}}$ (mm)	3,71	0,01	0,00	99,46	8,69	234,42
$p_{trmm_{t-2}}$ (mm)	3,71	0,01	0,00	99,46	8,69	234,42
$p_{trmm_{48h_t}}$ (mm)	7,42	0,42	0,00	102,37	14,10	190,11
$p_{trmm_{72h_t}}$ (mm)	11,12	1,62	0,00	133,85	18,62	167,34
$p_{trmm_{96h_t}}$ (mm)	14,83	3,36	0,00	177,42	22,71	153,12
$p_{trmm_{120h_t}}$ (mm)	18,54	5,72	0,00	179,83	26,54	143,14
$p_{trmm_{1sem_t}}$ (mm)	25,96	10,72	0,00	195,19	33,95	130,80
FL (%)	18,32	18,34	17,41	19,26	0,53	2,91
FN (%)	0,06	0,06	0,06	0,07	0,00	6,43
PA (%)	67,67	68,62	63,24	69,76	2,05	3,03
AG (%)	12,67	12,20	11,39	15,72	1,41	11,11
FP (%)	0,35	0,38	0,04	0,72	0,25	70,99
AnV (%)	0,69	0,67	0,57	0,84	0,09	12,42
CH (%)	0,24	0,24	0,22	0,24	0,01	2,59
$temp_{min_t}$ (°C)	18,10	18,25	-0,21	27,39	3,59	19,85
$temp_{med_t}$ (°C)	28,18	28,46	17,97	38,44	3,38	11,98
$temp_{max_t}$ (°C)	34,15	34,31	20,39	50,47	4,61	13,49
evi_{min_t} (adimensional)	0,07	0,07	-0,13	0,19	0,05	78,12
evi_{med_t} (adimensional)	0,42	0,43	0,22	0,57	0,08	19,20
evi_{max_t} (adimensional)	0,61	0,67	0,00	0,90	0,22	35,77
$ndvi_{min_t}$ (adimensional)	0,11	0,09	-0,20	0,37	0,09	90,18
$ndvi_{med_t}$ (adimensional)	0,66	0,68	0,41	0,78	0,08	12,58
$ndvi_{max_t}$ (adimensional)	0,83	0,91	0,00	0,99	0,25	30,76
$escoam_{min_t}$ (mm)	15,67	2,00	0,00	213,00	37,83	241,39
$escoam_{med_t}$ (mm)	29,09	3,40	0,00	326,11	59,80	205,61
$escoam_{max_t}$ (mm)	49,50	5,00	0,00	456,00	87,39	176,54
$umidade_{min_t}$ (mm)	52,91	39,60	9,70	107,70	35,51	67,12
$umidade_{med_t}$ (mm)	99,28	100,63	20,94	158,59	46,45	46,79
$umidade_{max_t}$ (mm)	150,59	165,30	36,70	199,30	47,34	31,43
mês	6,56	7,00	1,00	12,00	3,43	52,32

*Coeficiente de Variação. Fonte: A autora.

Tabela 3D – Estatística descritiva da variável objetivo e das covariáveis utilizadas no cenário de 3 dias de antecedência

Variável	Média	Mediana	Mínimo	Máximo	Desvio Padrão	CV (%)*
Y (m³/s)	94,21	54,52	3,14	1849,99	117,84	125,08
Q1 _t (m³/s)	45,91	27,20	1,86	649,47	54,31	118,30
Q1 _{t-1} (m³/s)	45,91	27,20	1,86	649,47	54,31	118,30
Q1 _{t-2} (m³/s)	45,90	27,20	1,86	649,47	54,31	118,30
Q2 _t (m³/s)	12,49	7,51	0,72	228,45	15,51	124,20
Q2 _{t-1} (m³/s)	12,49	7,51	0,72	228,45	15,51	124,20
Q2 _{t-2} (m³/s)	12,49	7,51	0,72	228,45	15,51	124,19
Q3 _t (m³/s)	20,75	12,22	0,90	740,44	28,75	138,54
Q3 _{t-1} (m³/s)	20,75	12,22	0,90	740,44	28,75	138,54
Q3 _{t-2} (m³/s)	20,75	12,22	0,90	740,44	28,75	138,55
Q4 _t (m³/s)	84,62	50,39	1,22	1589,32	110,14	130,16
Q4 _{t-1} (m³/s)	84,61	50,39	1,22	1589,32	110,14	130,17
Q4 _{t-2} (m³/s)	84,61	50,39	1,22	1589,32	110,14	130,18
p_idw _t (mm)	3,73	0,45	0,00	69,23	7,40	198,49
p_idw _{t-1} (mm)	3,73	0,45	0,00	69,23	7,40	198,50
p_idw _{t-2} (mm)	3,73	0,45	0,00	69,23	7,40	198,50
p_idw _{48h_t} (mm)	7,45	1,56	0,00	108,71	12,74	170,94
p_idw _{72h_t} (mm)	11,18	3,18	0,00	153,70	17,30	154,79
p_idw _{96h_t} (mm)	14,90	5,24	0,00	184,16	21,44	143,88
p_idw _{120h_t} (mm)	18,63	7,55	0,00	204,31	25,33	135,95
p_idw _{1sem_t} (mm)	26,08	12,21	0,00	248,20	32,81	125,82
p_xavier _t (mm)	3,73	0,40	0,00	71,64	7,52	201,70
p_xavier _{t-1} (mm)	3,73	0,40	0,00	71,64	7,52	201,70
p_xavier _{t-2} (mm)	3,73	0,40	0,00	71,64	7,52	201,70
p_xavier _{48h_t} (mm)	7,46	1,50	0,00	114,98	12,87	172,46
p_xavier _{72h_t} (mm)	11,19	3,10	0,00	162,92	17,41	155,60
p_xavier _{96h_t} (mm)	14,92	5,08	0,00	191,27	21,54	144,34
p_xavier _{120h_t} (mm)	18,65	7,42	0,00	211,18	25,41	136,19
p_xavier _{1sem_t} (mm)	26,11	12,37	0,00	253,16	32,86	125,84
p_trmm _t (mm)	3,71	0,01	0,00	99,46	8,69	234,42
p_trmm _{t-1} (mm)	3,71	0,01	0,00	99,46	8,69	234,42
p_trmm _{t-2} (mm)	3,71	0,01	0,00	99,46	8,69	234,42
p_trmm _{48h_t} (mm)	7,42	0,42	0,00	102,37	14,10	190,11
p_trmm _{72h_t} (mm)	11,12	1,62	0,00	133,85	18,62	167,34
p_trmm _{96h_t} (mm)	14,83	3,36	0,00	177,42	22,71	153,12
p_trmm _{120h_t} (mm)	18,54	5,72	0,00	179,83	26,54	143,13
p_trmm _{1sem_t} (mm)	25,96	10,72	0,00	195,19	33,95	130,81
FL (%)	18,32	18,34	17,41	19,26	0,53	2,91
FN (%)	0,06	0,06	0,06	0,07	0,00	6,43
PA (%)	67,67	68,62	63,24	69,76	2,05	3,03
AG (%)	12,67	12,20	11,39	15,72	1,41	11,11
FP (%)	0,35	0,38	0,04	0,72	0,25	70,99
AnV (%)	0,69	0,67	0,57	0,84	0,09	12,42

(cont.)

Variável	Média	Mediana	Mínimo	Máximo	Desvio Padrão	CV (%)*
CH (%)	0,24	0,24	0,22	0,24	0,01	2,59
<i>temp_min_t</i> (°C)	18,10	18,25	-0,21	27,39	3,59	19,85
<i>temp_med_t</i> (°C)	28,18	28,46	17,97	38,44	3,38	11,98
<i>temp_max_t</i> (°C)	34,15	34,31	20,39	50,47	4,61	13,49
<i>evi_min_t</i> (adimensional)	0,07	0,07	-0,13	0,19	0,05	78,12
<i>evi_med_t</i> (adimensional)	0,42	0,43	0,22	0,57	0,08	19,20
<i>evi_max_t</i> (adimensional)	0,61	0,67	0,00	0,90	0,22	35,77
<i>ndvi_min_t</i> (adimensional)	0,11	0,09	-0,20	0,37	0,09	90,18
<i>ndvi_med_t</i> (adimensional)	0,66	0,68	0,41	0,78	0,08	12,58
<i>ndvi_max_t</i> (adimensional)	0,83	0,91	0,00	0,99	0,25	30,76
<i>escoam_min_t</i> (mm)	15,67	2,00	0,00	213,00	37,83	241,39
<i>escoam_med_t</i> (mm)	29,09	3,40	0,00	326,11	59,80	205,61
<i>escoam_max_t</i> (mm)	49,50	5,00	0,00	456,00	87,39	176,54
<i>umidade_min_t</i> (mm)	52,91	39,60	9,70	107,70	35,51	67,12
<i>umidade_med_t</i> (mm)	99,28	100,63	20,94	158,59	46,45	46,79
<i>umidade_max_t</i> (mm)	150,59	165,30	36,70	199,30	47,34	31,43
mês	6,56	7,00	1,00	12,00	3,43	52,32

*Coeficiente de Variação. Fonte: A autora.

APÊNDICE E

Tabela 1E – Estatística descritiva das métricas de desempenho das técnicas de ML (GLM, MARS e RF) no cenário de 1 dia de antecedência no ano hidrológico 2006/2007 considerando as 50 repetições

Métrica	Modelo	Média	Mediana	Mínimo	Máximo	Desvio Padrão	CV(%)*
RMSE (m ³ /s)	GLM	32,87	32,93	31,49	35,19	0,73	2,23
	MARS	35,09	34,55	30,52	40,74	2,38	6,79
	RF	17,90	17,20	14,06	24,47	2,79	15,56
MAE (m ³ /s)	GLM	14,36	14,27	13,40	15,53	0,50	3,49
	MARS	16,05	16,18	14,63	17,12	0,63	3,95
	RF	6,54	6,59	5,83	7,35	0,40	6,11
R ² (adimensional)	GLM	0,96	0,96	0,95	0,96	0,00	0,17
	MARS	0,95	0,96	0,94	0,97	0,01	0,64
	RF	0,99	0,99	0,98	0,99	0,00	0,41
NSE (adimensional)	GLM	0,96	0,96	0,95	0,96	0,00	0,20
	MARS	0,95	0,95	0,93	0,96	0,01	0,71
	RF	0,99	0,99	0,98	0,99	0,00	0,42

*Coeficiente de Variação. Fonte: A autora.

Tabela 2E – Estatística descritiva das métricas de desempenho das técnicas de ML (GLM, MARS e RF) no cenário de 2 dias de antecedência no ano hidrológico 2006/2007 considerando as 50 repetições

Métrica	Modelo	Média	Mediana	Mínimo	Máximo	Desvio Padrão	CV(%)*
RMSE (m ³ /s)	GLM	45,00	44,95	43,94	46,29	0,50	1,12
	MARS	46,84	46,06	43,81	54,99	2,53	5,39
	RF	25,93	26,26	20,62	33,25	2,87	11,05
MAE (m ³ /s)	GLM	21,19	21,14	20,21	23,46	0,66	3,12
	MARS	22,80	22,80	20,14	25,42	1,15	5,02
	RF	10,63	10,56	9,45	11,86	0,52	4,93
R ² (adimensional)	GLM	0,92	0,92	0,92	0,92	0,00	0,17
	MARS	0,92	0,92	0,89	0,93	0,01	1,03
	RF	0,97	0,97	0,96	0,98	0,01	0,58
NSE (adimensional)	GLM	0,92	0,92	0,92	0,92	0,00	0,20
	MARS	0,91	0,92	0,88	0,92	0,01	1,07
	RF	0,97	0,97	0,96	0,98	0,01	0,61

*Coeficiente de Variação. Fonte: A autora.

Tabela 3E – Estatística descritiva das métricas de desempenho das técnicas de ML (GLM, MARS e RF) no cenário de 3 dias de antecedência no ano hidrológico 2006/2007 considerando as 50 repetições

Métrica	Modelo	Média	Mediana	Mínimo	Máximo	Desvio Padrão	CV(%)*
RMSE (m ³ /s)	GLM	62,85	62,88	61,37	64,44	0,68	1,08
	MARS	63,60	63,43	60,11	67,46	1,81	2,84
	RF	32,88	32,87	26,83	38,92	2,93	8,92
MAE (m ³ /s)	GLM	30,56	30,53	29,92	32,10	0,40	1,29
	MARS	31,78	31,61	30,09	36,81	1,09	3,44
	RF	15,05	14,99	13,70	16,79	0,69	4,56
R ² (adimensional)	GLM	0,85	0,85	0,84	0,85	0,00	0,37
	MARS	0,85	0,85	0,83	0,86	0,01	1,10
	RF	0,96	0,96	0,94	0,97	0,01	0,77
NSE (adimensional)	GLM	0,84	0,84	0,84	0,85	0,00	0,40
	MARS	0,84	0,84	0,82	0,86	0,01	1,09
	RF	0,96	0,96	0,94	0,97	0,01	0,80

*Coeficiente de Variação. Fonte: A autora.

Tabela 4E – Estatística descritiva das métricas de desempenho das técnicas de ML (GLM, MARS e RF) no cenário de 1 dia de antecedência no ano hidrológico 2008/2009 considerando as 50 repetições

Métrica	Modelo	Média	Mediana	Mínimo	Máximo	Desvio Padrão	CV(%)*
RMSE (m ³ /s)	GLM	43,08	43,04	42,26	44,59	0,55	1,27
	MARS	37,83	37,28	31,98	49,94	3,72	9,83
	RF	35,36	34,43	24,25	63,32	7,95	22,48
MAE (m ³ /s)	GLM	21,54	21,55	20,84	22,41	0,36	1,68
	MARS	18,66	18,56	17,09	21,06	0,91	4,88
	RF	11,35	11,18	9,27	14,65	1,07	9,45
R ² (adimensional)	GLM	0,97	0,97	0,97	0,97	0,00	0,07
	MARS	0,98	0,98	0,96	0,98	0,00	0,50
	RF	0,98	0,98	0,95	0,99	0,01	0,87
NSE (adimensional)	GLM	0,97	0,97	0,97	0,97	0,00	0,09
	MARS	0,97	0,98	0,96	0,98	0,01	0,52
	RF	0,98	0,98	0,93	0,99	0,01	1,14

*Coeficiente de Variação. Fonte: A autora.

Tabela 5E – Estatística descritiva das métricas de desempenho das técnicas de ML (GLM, MARS e RF) no cenário de 2 dias de antecedência no ano hidrológico 2008/2009 considerando as 50 repetições

Métrica	Modelo	Média	Mediana	Mínimo	Máximo	Desvio Padrão	CV(%)
RMSE (m ³ /s)	GLM	80,92	80,91	78,93	84,00	1,09	1,35
	MARS	70,19	70,00	62,76	80,38	3,90	5,55
	RF	51,65	50,70	35,82	77,02	8,41	16,29
MAE (m ³ /s)	GLM	37,11	37,10	36,14	38,25	0,43	1,15
	MARS	33,35	32,94	30,87	36,60	1,20	3,58
	RF	18,82	18,73	16,16	21,91	1,37	7,29
R ² (adimensional)	GLM	0,89	0,89	0,88	0,90	0,00	0,38
	MARS	0,92	0,92	0,89	0,93	0,01	1,00
	RF	0,96	0,97	0,93	0,98	0,01	1,23
NSE (adimensional)	GLM	0,89	0,89	0,88	0,89	0,00	0,35
	MARS	0,91	0,91	0,89	0,93	0,01	1,05
	RF	0,95	0,96	0,90	0,98	0,02	1,66

*Coeficiente de Variação. Fonte: A autora.

Tabela 6E – Estatística descritiva das métricas de desempenho das técnicas de ML (GLM, MARS e RF) no cenário de 3 dias de antecedência no ano hidrológico 2008/2009 considerando as 50 repetições

Métrica	Modelo	Média	Mediana	Mínimo	Máximo	Desvio Padrão	CV(%)
RMSE (m ³ /s)	GLM	131,93	131,78	129,10	137,89	1,82	1,38
	MARS	119,43	119,24	111,45	128,77	4,16	3,48
	RF	78,70	79,47	61,51	110,77	10,51	13,35
MAE (m ³ /s)	GLM	58,07	57,91	57,15	60,33	0,70	1,20
	MARS	54,45	54,46	51,36	57,93	1,13	2,07
	RF	28,76	28,62	26,01	32,49	1,60	5,55
R ² (adimensional)	GLM	0,71	0,71	0,68	0,72	0,01	1,43
	MARS	0,77	0,77	0,71	0,80	0,02	2,22
	RF	0,91	0,91	0,81	0,95	0,03	2,98
NSE (adimensional)	GLM	0,70	0,70	0,67	0,71	0,01	1,20
	MARS	0,75	0,75	0,71	0,78	0,02	2,31
	RF	0,89	0,89	0,79	0,93	0,03	3,34

*Coeficiente de Variação. Fonte: A autora.