

**Universidade de São Paulo
Escola Superior de Agricultura “Luiz de Queiroz”**

Regionalização hidrológica para o MATOPIBA, Brasil

Vinicius Costa Barreiros

Dissertação apresentada para obtenção do título de
Mestre em Ciências. Área de concentração: Engenharia
de Sistemas Agrícolas

**Piracicaba
2023**

Vinicius Costa Barreiros
Engenheiro Agrônomo

Regionalização hidrológica para o MATOPIBA, Brasil
versão revisada de acordo com a Resolução CoPGr 6018 de 2011

Orientadora:
Profa. Dra. **PATRICIA ANGÉLICA ALVES MARQUES**

Dissertação apresentada para obtenção do título de
Mestre em Ciências. Área de concentração: Engenharia
de Sistemas Agrícolas

Piracicaba
2023

Dados Internacionais de Catalogação na Publicação
DIVISÃO DE BIBLIOTECA – DIBD/ESALQ/USP

Barreiros, Vinicius Costa

Regionalização hidrológica para o MATOPIBA, Brasil / Vinicius Costa Barreiros. - - versão revisada de acordo com a Resolução CoPGr 6018 de 2011. - - Piracicaba, 2023.

134 p.

Dissertação (Mestrado) - - USP / Escola Superior de Agricultura "Luiz de Queiroz".

1. Hidrologia 2. Bacias hidrográficas 3. Modelos hidrológicos 4. Vazões de referência I. Título

DEDICATÓRIA

A minha esposa Joy Melloul
pelo apoio incondicional
pelo carinho e dedicação
pelo trabalho e criatividade
pelo companheirismo e
pelo Amor.

AGRADECIMENTOS

Agradeço, primeiramente ao Criador, que em sua infinita bondade me guia em meio aos caminhos que devo seguir.

À minha mãe Jurani, que com toda a dedicação me mostrou que a educação e a perseverança são as chaves para a mudança.

À minha esposa Joy, pela paciência, respeito, companheirismo durante estes dois anos em que me dediquei a este trabalho.

À professora Dr^a. Patrícia Angélica Alves Marques, que me abriu as portas da ESALQ, acreditando no meu potencial para a realização desta dissertação.

Ao professor Dr. Sérgio Nascimento Duarte, que não somente compartilhou seus conhecimentos, mas esteve presente e disponível para me ajudar durante minha jornada.

Aos colegas Hélio Lopes Araújo e Ângela Silviane Moura Cunha pelo compartilhamento de conhecimento e disponibilidade para melhorar o presente trabalho.

Aos senhores Davi de Oliveira Custodio e Jaime Costa Mota da Embrapa Territorial, pelo suporte e disponibilização de informações sobre a região de estudo.

À Escola Superior de Agricultura “Luiz de Queiroz”, pela oportunidade de realização deste mestrado.

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) pelo apoio financeiro concedido.

EPÍGRAFE

“Se você planeja para um ano, plante arroz. Se você planeja para dez anos, plante uma árvore. Mas se você planeja para cem anos, eduque uma criança”.

(Confúcio, 551-479 a.C.)

SUMÁRIO

RESUMO.....	9
ABSTRACT.....	10
LISTA DE FIGURAS.....	11
LISTA DE TABELAS.....	14
1. INTRODUÇÃO.....	17
2. OBJETIVOS.....	19
2.1. Objetivo geral.....	19
2.2. Objetivos específicos.....	19
2.3. Objetivos de Desenvolvimento Sustentável (ODS).....	19
3. REVISÃO DE LITERATURA.....	21
3.1. Regionalização hidrológica.....	21
3.2. Etapas da regionalização hidrológica.....	22
3.2.1. Seleção e análise dos dados de entrada.....	22
3.2.2. Identificação de regiões hidrologicamente homogêneas.....	23
3.2.3. Ajuste do modelo de regionalização hidrológica.....	24
3.3. O direito do uso da água.....	25
3.4. Vazão outorgável e vazão de referência.....	26
3.4.1. Vazão mínima com sete dias de duração consecutivos e período de retorno de dez anos ($Q_{7,10}$).....	27
3.4.2. Vazões de referência com 90% e 95% de permanência ao longo do tempo (Q_{90} e Q_{95}).....	28
3.5. Validação dos modelos de regionalização.....	29
4. MATERIAL E MÉTODOS.....	31
4.1. Localização e caracterização da área de estudo.....	31
4.2. Base de dados.....	32
4.2.1. Dados pluviométricos.....	33
4.2.1.1. Preenchimento das falhas de dados.....	33
4.2.1.2. Cálculo da precipitação média anual.....	34
4.2.1.3. Geração das isoietas para a região.....	34
4.2.1.3.1. Semivariograma.....	34
4.2.1.3.2. Modelos de semivariogramas.....	35

4.2.1.3.3. Método de ajuste dos semivariogramas.....	36
4.2.1.3.4. Seleção do modelo de semivariograma	37
4.2.1.3.5. Interpolação dos dados	38
4.2.1.3.6. Extração das isoietas.....	38
4.2.2. Dados fluviométricos.....	38
4.2.2.1. Preenchimento de falhas de dados.....	39
4.2.2.2. Cálculo das vazões de referência.....	40
4.2.2.2.1. Cálculo da vazão média de longo período (Q_{med})	40
4.2.2.2.2. Cálculo das vazões com 90% (Q_{90}) e com 95% (Q_{95}) de permanência ao longo do tempo	40
4.2.2.2.3. Cálculo da vazão mínima de 7 dias consecutivos e período de retorno de 10 anos ($Q_{7,10}$).....	42
4.2.2.2.4. Distribuições de densidade de probabilidade	42
4.2.2.2.5. Intervalo de confiança	43
4.2.2.3. Cálculo das vazões específicas	44
4.2.2.4. Delimitação das áreas de contribuição e índices morfométricos.....	45
4.2.2.4.1. Estimação da precipitação média anual por bacia	45
4.2.2.4.2. Índices morfométricos	45
4.3. Análise de agrupamento (Análise Cluster).....	47
4.4. Regressão linear multivariada	49
4.4.1. Modelos de regionalização hidrológica.....	49
4.4.2. Análise estatística dos modelos	51
4.4.3. Validação dos modelos de regionalização hidrológica.....	52
4.4.3.1. Intervalo de classes dos valores de vazão.....	52
4.4.3.2. Matriz de confusão	53
4.4.3.3. Curva ROC	54
4.4.3.4. Curva AUC	55
5. RESULTADOS E DISCUSSÃO	57
5.1. Delimitação das bacias de drenagem.....	57
5.2. Cálculo da precipitação média anual para as bacias de drenagem	58
5.3. Vazões de referência Q_{med} , Q_{90} , Q_{95} , $Q_{7,10}$ específicas obtidas através das séries históricas das estações fluviométricas	63
5.3.1. Vazões Q_{med} , Q_{90} e Q_{95} específicas.....	63
5.3.2. Vazão $Q_{7,10}$ específica.....	65

5.4. Análise de grupamento e determinação das regiões hidrologicamente homogêneas ...	67
5.5. Modelos de regionalização hidrológica	69
5.6. Avaliação dos modelos de regionalização	73
5.6.1. Vazão média específica regionalizada ($Q_{med_{espr}}$)	73
5.6.2. Vazão específica regionalizada com 90% de permanência ($Q_{90_{espr}}$)	76
5.6.3. Vazão específica regionalizada com 95% de permanência ($Q_{95_{espr}}$)	78
5.6.4. Vazão específica regionalizada mínima de 7 dias consecutivos com período de retorno de 10 anos ($Q_{7,10_{espr}}$)	80
5.7. Validação dos modelos	83
6. CONSIDERAÇÕES FINAIS.....	91
7. CONCLUSÕES	93
REFERÊNCIAS.....	95
ANEXOS	111

RESUMO

Regionalização hidrológica para a região do MATOPIBA, Brasil

Atualmente vista como a nova fronteira agrícola brasileira, a região do MATOPIBA abrange os estados do Maranhão, Tocantins, Piauí e Bahia. Segundo dados da Companhia Nacional de Abastecimento (CONAB), a região já é responsável por cerca de 14% da produção nacional de grãos de soja. Apresentando um enorme potencial de crescimento, porém, dependente do uso de irrigação, fica evidente a necessidade de dados consistentes de vazão dos cursos d'água, permitindo assim o gerenciamento adequado dos recursos hídricos, podendo desta forma atender não somente as demandas para a produção agrícola, mas também para a geração de energia hidroelétrica, o abastecimento urbano, a dessedentação de rebanhos e o uso industrial. A região apresenta uma grande quantidade de postos fluviométricos, porém, concentrados sobre os grandes cursos d'água; outro problema identificado é a presença de séries históricas incompletas, limitando desta forma a estimação da vazão. A regionalização hidrológica surge como uma alternativa para contornar este problema, consistindo na transferência de dados entre bacias hidrográficas que possuam características hidrológicas semelhantes, possibilitando assim, estimar a vazão em locais sem monitoramento ou com quantidade de dados insuficientes. Diante disto, o objetivo geral deste trabalho consistiu em desenvolver uma proposta de regionalização hidrológica de vazões de referência para o MATOPIBA, baseada em regiões hidrologicamente homogêneas e na regressão multivariada. Partindo das séries históricas de vazão disponibilizadas pela Agência Nacional de Águas e Saneamento Básico (ANA), foram calculadas as vazões: Q_{med} , Q_{90} , Q_{95} e $Q_{7,10}$ para cada estação fluviométrica. A partir dos dados das estações pluviométricas foram calculadas a precipitação anual média e posteriormente, a por meio da interpolação espacial pelo método da krigagem ordinária foram calculadas a precipitação anual média para cada bacia. Como variáveis independentes para os modelos, foram obtidas as características fisiográficas de cada bacia: densidade de drenagem, comprimento do talvegue principal, declividade média da bacia, coeficiente de compactidade (K_c) e coordenadas do centroide da bacia. De posse destes dados, foram delimitadas 7 regiões hidrologicamente homogêneas por meio da análise de agrupamento (*cluster*) com algoritmo *Ward*. Foram desenvolvidos os modelos de regionalização das vazões de referência por meio da regressão linear múltipla. A avaliação destes modelos se deu por diferentes índices de performance: coeficiente de determinação (R^2), teste de eficiência de Nash-Sutcliffe (NSE), percentual de viés (PBIAS) e o índice de concordância de Willmott (d). A validação dos modelos foi realizada de maneira agrupada, ou seja, os sete modelos para cada uma das vazões de referência foram analisados em conjuntos, tendo como base a matriz de confusão, foram gerados quatro índices de desempenho: precisão, sensibilidade, TFP (taxa de falso positivo) e acurácia, além da curva ROC (*Receiver Operator Characteristic*) e da AUC (*Area Under the Curve*). Os resultados obtidos tanto pela avaliação quanto pela validação, comprovam que os modelos, de maneira geral, apresentaram um bom desempenho para a estimação das vazões de referência para a região, apesar de algumas limitações, podendo serem utilizados como ferramenta de auxílio no processo de concessão de outorgas, para as agências de água dos estados de compõe a região.

Palavras-chave: Hidrologia, Bacia hidrográfica, Modelos hidrológicos, Vazões de referência

ABSTRACT

Hydrological regionalization for the MATOPIBA region, Brazil

Currently seen as the new Brazilian agricultural frontier, the MATOPIBA region encompasses the states of Maranhão, Tocantins, Piauí and Bahia. According to data from the Nacional Supply Company (CONAB), the region is already responsible for about 14% of the national production of soybeans. Presenting an enormous growth potential, however, dependent on the use of irrigation, it is evident the need for consistent data on watercourse flow, thus allowing the adequate management of water resources, and thus being able to meet not only the demands for agricultural production, but also for the generation of hydroelectric power, urban supply, livestock and industrial use. The region presents a large number of fluviometric stations, however, concentrated on the major waterways; another problem identified is the presence of incomplete historical series, thus limiting the estimation of the flow. Hydrological regionalization emerges as an alternative to circumvent this problem, consisting in the transfer of data between watersheds that have similar hydrological characteristics, thus making it possible to estimate the flow rate in places without monitoring or with insufficient data. In view of this, the general objective of this work was to develop a proposal for hydrological regionalization of reference water levels for MATOPIBA, based on hydrologically homogeneous regions and on multivariate regression. Based on the historical flow series made available by the National Water and Basic Sanitation Agency (ANA), the flows were calculated: Q_{med} , Q_{90} , Q_{95} and $Q_{7,10}$ for each fluviometric station. From the data of the rainfall stations the average annual precipitation was calculated and later, through spatial interpolation by the ordinary kriging method, the average annual precipitation for each basin was calculated. As independent variables for the models, the physiographic characteristics of each basin were obtained: drainage density, length of the main watershed, average basin slope, compactness coefficient (K_c) and coordinates of the basin centroid. With these data, 7 hydrologically homogeneous regions were delimited by means of cluster analysis with the Ward's algorithm. Reference flow regionalization models were developed by means of multiple linear regression. The evaluation of these models was done by different performance indexes: coefficient of determination (R^2), Nash-Sutcliffe efficiency test (NSE), bias percentage (PBIAS) and Willmott's agreement index (d). The validation of the models was carried out in a grouped manner, that is, the seven models for each of the reference flows were analyzed in sets, based on the confusion matrix, four performance indexes were generated: accuracy, sensitivity, TFP (False Positive Rate) and accuracy, besides the ROC curve (Receiver Operator Characteristic) and the AUC (Area Under the Curve). The results obtained both by evaluation and validation, prove that the models, in general, presented a good performance for the estimation of reference water levels for the region, despite some limitations, and can be used as a tool to assist the water agencies of the states that make up the region in the granting process.

Keywords: Hydrology, Watershed, Hydrological models, Reference water flow

LISTA DE FIGURAS

Figura 1 - Delimitação do MATOPIBA e relevo da região.	31
Figura 2 - Localização dos postos pluviométricos, fluviométricos e da rede hidrográfica na região do MATOPIBA.	33
Figura 3 - Distribuição dos registros diários de vazão. O eixo inferior corresponde aos anos de registro, o eixo vertical direito corresponde ao código das estações fluviométricas e a escala vertical direita representa o código de cores correspondente ao número de dias agregados por ano.	39
Figura 4 – Exemplo de curva ROC com cinco classificadores discretos, em que A é dito um classificador "conservador", B é o inverso de E, D é um classificador perfeito e C é dito aleatório.	55
Figura 5 - Postos fluviométricos na região do MATOPIBA e suas áreas de contribuição.	57
Figura 6: Mapa variográfico; (a) modelo com anisotropia; (b) modelo após a remoção da dependência espacial.	58
Figura 7: Gráfico dos semivariogramas dos modelos; (a) modelo esférico; (b) modelo exponencial e (c) modelo Matérn.	59
Figura 8: Resultado da validação cruzada para os diferentes modelos. (a) modelo esférico, (b) modelo exponencial e (c) modelo Matérn.	59
Figura 9 - Interpolação dos dados de precipitação pelo método da krigagem ordinária.	60
Figura 10: Distribuição das linhas isoietas com distância de 50 mm entre as linhas para a região do MATOPIBA.	61
Figura 11 – (a) Fluxograma das vazões médias mensais do posto 45298000; (b) Curva de permanência de vazão para o posto 45298000.	64
Figura 12 - Exemplo de curva de distribuição teórica dos dados e curvas das distribuições estatísticas para o posto 45298000.	65
Figura 13 – Delimitação dos 7 clusters pelo método hierárquico através do algoritmo de Ward.	68
Figura 14 - Regiões hidrologicamente homogêneas para a região do MATOPIBA segundo a análise de agrupamento (clusters).	68
Figura 15 - Número de postos fluviométricos por região hidrologicamente homogênea.	69
Figura 16 - Relação entre a $Q_{med_{esp}}$ calculada e a $Q_{med_{esp}}$ regionalizada para as 7 regiões hidrologicamente homogêneas no MATOPIBA.	74

Figura 17 – Relação entre a $Q90_{esp}$ calculada e a $Q90_{esp}$ regionalizada para as 7 regiões hidrologicamente homogêneas no MATOPIBA.....	76
Figura 18 - Relação entre a $Q95_{esp}$ calculada e a $Q95_{esp}$ regionalizada para as 7 regiões hidrologicamente homogêneas no MATOPIBA.....	78
Figura 19 - Relação entre a $Q7,10_{esp}$ calculada e a $Q7,10_{esp}$ regionalizada para as 7 regiões hidrologicamente homogêneas no MATOPIBA.....	81
Figura 20 - Número de observações por intervalo; (a) $Qmed_{esp}$; (b) $Q90_{esp}$; (c) $Q95_{esp}$ e (d) $Q7,10_{esp}$	84
Figura 21 - Matriz de confusão das vazões de referência; (a) $Qmed_{esp}$; (b) $Q90_{esp}$; (c) $Q95_{esp}$ e (d) $Q7,10_{esp}$	85
Figura 22 - Gráfico ROC para os quatro grupos de vazão de referência específica; (a) $Qmed_{esp}$; (b) $Q90_{esp}$; (c) $Q95_{esp}$ e (d) $Q7,10_{esp}$	88
Figura 23 - Gráfico AUC para as vazões de referência específicas; (a) $Qmed_{esp}$; (b) $Q90_{esp}$; (c) $Q95_{esp}$ e (d) $Q7,10_{esp}$	89
Figura 24 - Relação $Qmed_{esp}$ regionalizada e $Qmed_{esp}$ calculada; B) Relação $Q90_{esp}$ regionalizada e $Q90_{esp}$ calculada; C) Relação $Q95_{esp}$ regionalizada e $Q95_{esp}$ calculada; D) Relação $Q7,10_{esp}$ regionalizada e $Q7,10_{esp}$ calculada, para a região hidrologicamente homogênea 1.....	126
Figura 25 - Relação $Qmed_{esp}$ regionalizada e $Qmed_{esp}$ calculada; B) Relação $Q90_{esp}$ regionalizada e $Q90_{esp}$ calculada; C) Relação $Q95_{esp}$ regionalizada e $Q95_{esp}$ calculada; D) Relação $Q7,10_{esp}$ regionalizada e $Q7,10_{esp}$ calculada, para a região hidrologicamente homogênea 2.....	127
Figura 26 - Relação $Qmed_{esp}$ regionalizada e $Qmed_{esp}$ calculada; B) Relação $Q90_{esp}$ regionalizada e $Q90_{esp}$ calculada; C) Relação $Q95_{esp}$ regionalizada e $Q95_{esp}$ calculada; D) Relação $Q7,10_{esp}$ regionalizada e $Q7,10_{esp}$ calculada, para a região hidrologicamente homogênea 3.....	128
Figura 27 - Relação $Qmed_{esp}$ regionalizada e $Qmed_{esp}$ calculada; B) Relação $Q90_{esp}$ regionalizada e $Q90_{esp}$ calculada; C) Relação $Q95_{esp}$ regionalizada e $Q95_{esp}$ calculada; D) Relação $Q7,10_{esp}$ regionalizada e $Q7,10_{esp}$ calculada, para a região hidrologicamente homogênea 4.....	129
Figura 28 - Relação $Qmed_{esp}$ regionalizada e $Qmed_{esp}$ calculada; B) Relação $Q90_{esp}$ regionalizada e $Q90_{esp}$ calculada; C) Relação $Q95_{esp}$ regionalizada e $Q95_{esp}$ calculada; D) Relação $Q7,10_{esp}$ regionalizada e $Q7,10_{esp}$ calculada, para a região hidrologicamente homogênea 5.....	130

Figura 29 - Relação $Q_{med_{esp}}$ regionalizada e $Q_{med_{esp}}$ calculada; B) Relação $Q_{90_{esp}}$ regionalizada e $Q_{90_{esp}}$ calculada; C) Relação $Q_{95_{esp}}$ regionalizada e $Q_{95_{esp}}$ calculada; D) Relação $Q_{7,10_{esp}}$ regionalizada e $Q_{7,10_{esp}}$ calculada, para a região hidrologicamente homogênea 6..... 131

Figura 30 - Relação $Q_{med_{esp}}$ regionalizada e $Q_{med_{esp}}$ calculada; B) Relação $Q_{90_{esp}}$ regionalizada e $Q_{90_{esp}}$ calculada; C) Relação $Q_{95_{esp}}$ regionalizada e $Q_{95_{esp}}$ calculada; D) Relação $Q_{7,10_{esp}}$ regionalizada e $Q_{7,10_{esp}}$ calculada, para a região hidrologicamente homogênea 7..... 132

LISTA DE TABELAS

Tabela 1 - Critérios de avaliação qualitativa de desempenho e seus limites quantitativos para os índices estatísticos recomendados para os modelos de regionalização hidrológica.....	52
Tabela 2: Resultado do teste de validação cruzada Leave-One-Out para os três modelos.	60
Tabela 3 - Extrato dos valores estimados de precipitação média anual para a área de contribuição dos postos fluviométricos.....	61
Tabela 4 - Extrato dos dados das estações fluviométricas utilizadas para o cálculo das vazões de referência e seus respectivos valores estimados pelas séries históricas.	63
Tabela 5 - Exemplos de extrato das distribuições estatísticas que melhores se ajustaram para o cálculo da vazão mínima com duração de sete dias e período de retorno de 10 anos, valores de $Q_{7,10}$ ($m^3 s^{-1}$), intervalos de confiança e amplitude do intervalo.	66
Tabela 6 - Exemplos de extrato das vazões $Q_{7,10}$ e $Q_{7,10}$ específica de alguns postos fluviométricos na região do MATOPIBA.....	66
Tabela 7 - Parâmetros das equações de regressão para a estimação da vazão média específica regionalizada ($Q_{med_{espr}}$) para cada região hidrologicamente homogênea.....	70
Tabela 8 - Parâmetros das equações de regressão para a estimação da vazão $Q_{90_{esp}}$ regionalizada para cada região hidrologicamente homogênea.	71
Tabela 9 - Parâmetros das equações de regressão para a estimação da vazão $Q_{95_{esp}}$ regionalizada para cada região hidrologicamente homogênea.	71
Tabela 10 - Parâmetros das equações de regressão para a estimação da vazão $Q_{7,10_{esp}}$ regionalizada para cada região hidrologicamente homogênea.	72
Tabela 11 - Resultado dos critérios de avaliação de desempenho da $Q_{med_{espr}}$, segundo os índices estatísticos recomendados para modelos de regionalização hidrológica por Moriasi et al. (2015).	75
Tabela 12 - Classificação dos critérios de avaliação de desempenho da $Q_{med_{espr}}$, para os índices estatísticos recomendados para modelos de regionalização hidrológica por Moriasi et al. (2015).	75
Tabela 13 - Resultado dos critérios de avaliação de desempenho da $Q_{90_{espr}}$, segundo os índices estatísticos recomendados para modelos de regionalização hidrológica por Moriasi et al. (2015).	77
Tabela 14 - Classificação dos critérios de avaliação de desempenho da $Q_{90_{espr}}$, para os índices estatísticos recomendados para modelos de regionalização hidrológica por Moriasi et al. (2015).	77

Tabela 15 - Resultado dos critérios de avaliação de desempenho da $Q95_{\text{espr}}$, segundo os índices estatísticos recomendados para modelos de regionalização hidrológica por Moriasi et al. (2015).....	79
Tabela 16 - Classificação dos critérios de avaliação de desempenho da $Q95_{\text{espr}}$, para os índices estatísticos recomendados para modelos de regionalização hidrológica por Moriasi et al. (2015).	79
Tabela 17 - Resultado dos critérios de avaliação de desempenho da $Q7,10_{\text{espr}}$, segundo os índices estatísticos recomendados para modelos de regionalização hidrológica por Moriasi et al. (2015).....	81
Tabela 18 - Classificação dos critérios de avaliação de desempenho da $Q7,10_{\text{espr}}$, para os índices estatísticos recomendados para modelos de regionalização hidrológica por Moriasi et al. (2015).....	82
Tabela 19 - Resultado dos parâmetros para as vazões de referência, baseados na matriz de confusão.....	86
Tabela 20 - Resultado dos indicadores dos modelos para as vazões de referência específicas, baseados nas matrizes de confusão.....	86
Tabela 21 - Resultados da AUC para as diferentes classes de vazões.	89
Tabela 22 - Lista com os postos fluviométricos, rio, município e estado de localização, além da precipitação média anual para a área de contribuição.	117
Tabela 23 - Lista com os postos fluviométricos, geolocalização, área de captação e vazões de referência.	120
Tabela 24 – Lista com os postos fluviométricos, vazão $Q_{7,10}$, intervalos de confiança e tipo de distribuição.	123
Tabela 25 - Intervalos de vazão para as 8 classes segundo a Regra de Sturges, número de observações e frequência para os dados calculados e estimados para a vazão Q_{medesp}	133
Tabela 26 - Intervalos de vazão para as 8 classes segundo a Regra de Sturges, número de observações e frequência para os dados calculados e estimados para a vazão $Q90_{\text{esp}}$	133
Tabela 27 - Intervalos de vazão para as 8 classes segundo a Regra de Sturges, número de observações e frequência para os dados calculados e estimados para a vazão $Q95_{\text{esp}}$	133
Tabela 28 - Intervalos de vazão para as 8 classes segundo a Regra de Sturges, número de observações e frequência para os dados calculados e estimados para a vazão $Q7,10_{\text{esp}}$	134

1. INTRODUÇÃO

Atualmente o Brasil está entre os maiores produtores e exportadores mundiais de açúcar, café e suco de laranja. Segundo análise realizada pela OECD (Organization for Economic Co-operation and Development) em conjunto com a FAO (Food and Agriculture Organization), as projeções para a década atual (2020 – 2029), apontam que o país continuará a figurar entre os maiores produtores e exportadores de grãos, carnes, frutas e açúcar (OECD-FAO, 2020).

Para atender à demanda crescente, principalmente para grãos, a região que abrange parte dos estados do Maranhão, Tocantins, Piauí e Bahia (MATOPIBA), vem se destacando, sendo considerada como a principal fronteira para novos investimentos do agronegócio nacional (Araújo et al., 2019).

A região conta com áreas planas, condições climáticas favoráveis para o cultivo de grãos, além da proximidade de terminais portuários e acesso a modais de transporte alternativos e complementares (Belchior et al., 2017). A integração desses fatores permitiu que no ano-safra 2021/2022 a produção de grãos de soja no MATOPIBA atingisse 17,85 milhões de Mg (Megagramas), o que representa cerca de 14,1% da produção nacional (CONAB, 2022). Grande parte destes resultados são devidos ao uso da irrigação, já que a região pode apresentar períodos de veranicos durante a época chuvosa e entre 4 e 6 meses sem precipitação (Evangelista et al., 2017).

Os eventos de seca (duradouras) ou estiagens (passageiras) no Brasil estão cada vez mais frequentes, sendo que 78,5% dos municípios do Nordeste declararam situação de emergência ou estado de calamidade pública entre os anos de 2003 e 2016. Neste mesmo período, o Nordeste registrou 83% dos 5154 eventos de seca com dano humano no Brasil, o que impactou a oferta de água para o abastecimento público e para setores que dependem de água, como a geração hidroelétrica, a irrigação, a produção industrial e a navegação. Entre 1995 a 2014, as perdas relacionadas a secas ou cheias chegaram a R\$ 182,7 bilhões (Marengo et al., 2022; ANA, 2017).

A maior demanda por água associada aos eventos de seca gera cada vez mais conflitos pelo uso da água na região do semiárido. Entre os anos de 2009 e 2017, os registros passaram de 13 para 67, o que evidencia a necessidade da aplicação das políticas voltadas para uma boa gestão da água em uma região onde a oferta hídrica é limitada (Silva, 2019). A fim de garantir o melhor uso dos recursos hídricos no país, a Lei das águas (Lei 9433 / 97) instituiu a Política Nacional dos Recursos Hídricos (PNRH), tendo como principal

instrumento a outorga dos direitos de uso dos recursos hídricos. Este documento assegura o controle quantitativo e qualitativo da captação da água e do lançamento de efluentes nos corpos d'água (BRASIL, 1997).

Os dados de vazão dos cursos d'água são necessários para o gerenciamento adequado dos recursos hídricos, permitindo o planejamento e o monitoramento das atividades que necessitam deste recurso, como a geração de energia hidroelétrica, o abastecimento urbano, a irrigação e a dessedentação de rebanho no meio rural, o uso industrial, além do controle de cheias e a manutenção de ecossistemas (Araújo, 2019).

A quantificação da disponibilidade de água, que deve ser mantida no corpo hídrico nos períodos de estiagem, varia entre os estados brasileiros, já que não há um consenso entre as legislações. Porém, há três tipos de vazão comumente utilizadas para a determinação deste volume de água (vazão mínima ou de referência). Estas vazões são calculadas por modelos estatístico-hidrológicos com base em dados históricos: $Q_{7,10}$ (vazão mínima média de sete dias de duração com período de retorno de 10 anos), Q_{90} (vazão com 90% de permanência no tempo no curso d'água) e Q_{95} (vazão com 95% de permanência no tempo no curso d'água) (Santos & Cunha, 2013).

Para a obtenção de dados de precipitação e de vazão, são utilizadas estações pluviométricas e fluviométricas, respectivamente. Elas são distribuídas em todo o território nacional e para que os dados possam ser úteis na estimação da vazão, eles devem representar as variabilidades dos eventos hidrológicos da região, além de possuir um registro consistente ao longo dos anos (Charles, 2020).

A presença de séries históricas incompletas limita o cálculo da estimação da vazão; porém, uma alternativa é a utilização de técnicas de regionalização de variáveis hidrológicas. A regionalização consiste na transferência de dados entre bacias que possuam características hidrológicas semelhantes, possibilitando a estimação da vazão de referência em regiões em que não há monitoramento ou que não possuam quantidade de dados suficientes (Sivapalan et al., 2012).

2. OBJETIVOS

2.1. Objetivo geral

O objetivo geral deste trabalho consiste em propor uma metodologia de regionalização hidrológica de vazão de referência para a região do MATOPIBA, baseada na diferenciação de regiões hidrológicamente homogêneas e no uso de regressão linear multivariada.

2.2. Objetivos específicos

- a) Determinar a precipitação anual média de longo período nas bacias de drenagem das estações fluviométricas utilizadas no estudo;
- b) Determinar as regiões hidrológicamente homogêneas para a região do MATOPIBA;
- c) Desenvolver um modelo de regionalização hidrológica de vazões de referência, por meio da regressão linear múltipla entre variáveis dependentes (vazões de referência) e independentes (características fisiográficas e morfológicas das bacias hidrográficas), para cada região hidrológicamente homogênea; e
- d) Avaliar a performance dos modelos de regionalização hidrológica propostos, por meio de índices estatísticos e técnicas de machine learning.

2.3. Objetivos de Desenvolvimento Sustentável (ODS)

Os objetivos de Desenvolvimento Sustentável foram propostos pela Organização das Nações Unidas (ONU) no ano de 2015 (United Nations, 2015), tendo como principal meta a construção de uma sociedade mais igualitária e em harmonia com o meio ambiente. A principal aplicação dos resultados obtidos com o presente projeto será na concessão de outorgas de direito de uso de recursos hídricos na região do MATOPIBA. Com isso, espera-se contribuir com diferentes pontos dos ODS, já que uma boa gestão da água tem impacto direto sobre aspectos sociais, ambientais e econômicos.

3. REVISÃO DE LITERATURA

3.1. Regionalização hidrológica

Diversas bacias hidrográficas do hemisfério sul estão sofrendo um rápido processo de desenvolvimento, situação que gera grandes implicações aos setores interdependentes de água-energia-alimentação-ambiente. A variabilidade intra e interanual da precipitação e as incertezas associadas ao nível de magnitude das mudanças climáticas em curso, dificultam o planejamento do uso dos recursos hídricos, aumentando os riscos para as atividades agrícolas, de geração de energia hidrelétrica e para os indicadores ambientais (Kolusu et al., 2021).

Atualmente, um dos maiores desafios para a ciência hidrológica é a estimação dos parâmetros de entrada de modelos para a estimação das vazões para as bacias hidrográficas sem monitoramento (Guo et al., 2021), e/ou a robustez e qualidade dos dados de bacias monitoradas, sendo estes em muitos casos insuficientes para a realização de estudos visando uma melhor gestão dos recursos hídricos (Lelis et al., 2020).

Os métodos de regionalização de parâmetros hidrológicos têm sido utilizados como uma alternativa para superar estas limitações, consistindo em uma técnica que permite a transferência de parâmetros calibrados de uma estação fluviométrica operacional (doadora) para outros pontos sem monitoramento (alvo), desde que estas regiões sejam consideradas homogêneas (Beck et al., 2020).

Os métodos mais utilizados para a regionalização presentes na literatura são o da proximidade física (PF), o da similaridade física (SF) e o baseado na regressão múltipla (RM) (Yang et al., 2020). O método da RM assume que exista uma boa correlação entre os atributos da bacia hidrográfica e os parâmetros do modelo. As funções geradas a partir da regressão são transferidas para bacias não monitoradas, permitindo a estimação dos valores de vazão (Xu et al., 2020).

Como Pessoa (2015) deixou claro, os valores estimados através da regionalização não devem de forma alguma serem substitutos às observações feitas *in loco*; contudo, a tomada de decisão a partir destes modelos será baseada em uma metodologia e em análises estatísticas, evitando desta forma erros e extrapolações apoiados na experiência de campo ou sentimentos individuais, ação ainda recorrente no meio técnico.

Segundo Tucci (2002), os modelos de regionalização devem atender a quatro requisitos básicos, independentemente da informação hidrológica que se tenha como alvo, são eles: a) definir de forma clara os limites da área a ser estudada; b) determinar quais são as

variáveis dependentes e independentes do modelo; c) selecionar dados consistentes e realizar os cálculos para obtenção das variáveis de entrada; e d) definir regiões dentro da área de estudo que possuam um comportamento hidrológico muito próximo e estimar as relações entre estas regiões.

Em 1975, o atual Natural Environmental Research Council (NERC) do governo britânico, elaborou o método designado com *index-flood*, baseado na curva adimensional regional, que tem como princípio a utilização de análises de métodos estatísticos aplicados à hidrologia, seguindo três etapas para a sua realização: a) seleção e análise dos dados de entrada; b) identificação de regiões hidrológicamente homogêneas; e c) ajuste do modelo de regressão multivariada (NERC, 1975).

3.2. Etapas da regionalização hidrológica

3.2.1. Seleção e análise dos dados de entrada

A seleção das variáveis independentes é uma etapa crucial para o processo de regionalização, em que podem ser escolhidos tanto parâmetros associados à aspectos físicos da bacia hidrográfica como parâmetros associados ao clima da região. Dentre os parâmetros morfológicos mais comuns estão, a área de drenagem da bacia, o comprimento do curso d'água principal, a densidade de drenagem e o uso e cobertura do solo, sendo possível utilizar outras variáveis como o tempo de concentração e a altitude média da bacia. A precipitação média anual é a variável mais utilizada dentre os parâmetros climáticos; porém, em alguns casos a evapotranspiração é incluída nos modelos (Tasker & Stedinger, 1989).

A obtenção dos dados para o processo de regionalização passa pelo levantamento da localização geográfica, das características geomorfológicas, hidrogeológicas e climatológicas das bacias hidrográficas, além da verificação da disponibilidade de postos fluviométricos e a consistência das séries históricas de dados (Eletrobrás, 1985). As falhas de registros de vazão podem tornar os dados inapropriados para o estudo, porém, se estas representarem um pequeno período dentro da série histórica, Gustard et al. (1992) recomendam a não exclusão da estação das análises, sendo sugerida a exclusão do ano hidrológico em que ocorrer a falha da análise dos dados.

Fenicia et al. (2022) desenvolveram cinco modelos para a simulação da vazão diária de 26 áreas de captação da bacia do rio Moselle, que possui sua nascente nas montanhas do nordeste na França, passa por Luxemburgo e tem seu exutório no oeste da Alemanha.

Baseando-se em doze parâmetros de calibração, o estudo buscou ilustrar como a distribuição espacial da precipitação, a litografia e a topografia afetaram a resposta da vazão.

Em estudo realizado por Schnier & Cai (2014), visando propor uma nova metodologia para a estimação da vazão (Model Tree Ensembles - MTEs), foram utilizadas características naturais e antrópicas de 294 bacias nos Estados Unidos da América (164 no estado do Texas e 130 no estado de Illinois). Os resultados apontaram que a precipitação anual e sazonal em conjunto com a área de drenagem foram importantes variáveis independentes para a regionalização de todas as vazões (máxima, média e mínima); já a ação humana como a poluição dos corpos d'água e a captação para uso foram importantes para as vazões mínimas.

Haberlandt et al. (2001) realizaram um estudo de regionalização da vazão de base para a bacia do rio Elba localizada no nordeste da Alemanha, tendo como base os dados de 25 sub-bacias, visando criar um modelo de estimação para as 114 sub-bacias que compõe a bacia do rio Elba no território alemão. Neste estudo, como variáveis independentes foram utilizados diferentes atributos físicos e climáticos das áreas de captação, concluindo que a vazão média de base para a região está fortemente correlacionada com a topografia, as características pedológicas e hidrogeológicas, além da precipitação; já o uso e cobertura do solo foram fracamente correlacionadas com a vazão.

3.2.2. Identificação de regiões hidrologicamente homogêneas

As regiões homogêneas podem ser definidas pela similaridade hidrológica entre bacias hidrográficas, em que são considerados fatores físicos, climáticos, biológicos, geológicos, além do efeito da ação humana sobre o meio. Sendo assim, uma região homogênea ficou definida como sendo aquela em que as características morfológicas e climáticas possuem pouca variabilidade (Lelis, 2019).

Segundo Garcia-Marin et al. (2015), a determinação destas regiões é o processo mais complexo da regionalização, e a continuidade das análises depende desta etapa. Diferentes autores elencaram a importância da definição das regiões homogêneas, sejam elas climatológicas, geológicas, geomorfológicas ou a interação destes três fatores, e o seu papel na transferência de informações entre bacias hidrográficas, deixando claro a importância da definição de critérios matemáticos para este processo (Greis & Wood, 1981; Hosking et al., 1985, e Lettenmaier et al., 1987).

A delimitação destas regiões pode ser realizada por diferentes métodos como: análise de agrupamento ou análise cluster (Bonnell & Sumner, 1992), análise de componentes principais (Garcia-Marin et al., 2011), agrupamento pelo Fuzzy C-means combinado com redes neurais artificiais (Satyanarayana & Srinivas, 2011), e inspeção visual dos diagramas de L-Moments (Yürekli & Modarres, 2007), entre outros.

3.2.3. Ajuste do modelo de regionalização hidrológica

Para Riggs (1990), os modelos de regionalização hidrológica em sua maioria, relacionam a vazão de interesse com variáveis morfológicas e climatológicas da bacia em modelos empíricos de regressão; porém, vale ressaltar que há outras técnicas conhecidas no meio científico.

O planejamento e a gestão dos recursos hídricos podem ser melhorados conforme os métodos de regionalização são aperfeiçoados, já que modelos de regressão com valores de estimativa de vazão nos cursos d'água mais próximos da realidade, permitem uma maior confiabilidade e segurança ao serem utilizados (Pruski et al., 2013).

Du et al. (2020), em estudo realizado na bacia do Grande Mekong, que possui fronteiras internacionais, abrangendo os territórios do Vietnã, Laos, Camboja, Tailândia, Myanmar e parte da China, propuseram uma nova abordagem de regionalização hidrológica para as vazões de referência, em que foi combinado o método de regionalização com observações de satélites para as variáveis independentes (topografia, uso e cobertura do solo, precipitação, temperatura e evapotranspiração potencial), além dos dados de 48 estações fluviométricas ao longo do curso d'água. A proposta de regionalização baseada nos parâmetros fisiográficos e climáticos, permitiu a obtenção de um modelo com 80% de performance em relação aos dados reais.

Swain & Patra (2017) analisaram diferentes técnicas de regionalização hidrológica para as vazões como: a proximidade espacial (inverso da distância ponderada, krigagem e média global), regressão multivariada e similaridade física em conjunto com o software SWAT (Soil and Water Assessment Tool), para 32 áreas de captação localizadas no leste e no sudeste da Índia. Os resultados apontaram que os métodos baseados na proximidade espacial entre as bacias (krigagem e o IDW), foram os melhores para a estimação da vazão dos rios da região; já o método da média global apresentou os piores resultados de desempenho em relação a todos os métodos analisados.

Em estudo conduzido por Pruski et al. (2013), foram propostas duas novas variáveis exploratórias para a regionalização da vazão média de longo período (Q_{lt}) e para a vazão com 95% de permanência ao longo do tempo (Q_{95}). A primeira variável é a vazão equivalente (P_{eq}), que utiliza a área de drenagem e a precipitação anual, a segunda variável proposta é a precipitação subtraída de um valor base (P_{eq750}), que por sua vez também foi convertida em vazão. Os resultados mostraram que a substituição da área de drenagem pela vazão equivalente, melhoraram o ajuste estatístico dos modelos, sendo que os resultados foram muito melhores ao utilizar a P_{eq750} ao invés de se utilizar somente a P_{eq} .

Chaves et al. (2002) realizaram um estudo de regionalização para a bacia do rio Itapicuru, localizada no norte do estado da Bahia, em que foi proposto um método para a estimação das vazões mínimas, baseado em interpolação e extrapolação automática através de software de Sistema de Informação Geográfica (SIG). A metodologia abordada pelos autores permitiu a redução do erro relativo de 289,7% do método tradicional (curva adimensional) para 45,1%. Apesar do valor final ser significativamente elevado, devido principalmente à baixa densidade de postos fluviométricos na região, o estudo permitiu uma melhora considerável na estimação das vazões.

3.3. O direito do uso da água

Em 1997 foi assinada a Lei n° 9433/97 (Brasil, 1997), estabelecendo a Política Nacional de Recursos Hídricos (PNRH) e a criação do Sistema Nacional de Gerenciamento dos Recursos Hídricos (SINGREH), que juntos levaram as bases para a gestão nacional das águas, instituindo os fundamentos da participação e descentralização na tomada de decisões. Segundo a Lei, a água é um bem público e através da PNRH foram estabelecidos os instrumentos para a sua gestão, como a outorga dos direitos de uso da água, atribuindo o volume de água que o usuário poderá captar dos rios para atender as suas necessidades. O processo de outorga passa pela avaliação da disponibilidade hídrica no local e o volume demandado pelo usuário, sendo que a demanda é avaliada de forma dinâmica e contínua no tempo, já que a disponibilidade hídrica é variável no tempo e no espaço, sendo necessário à sua estimação por meio da avaliação do regime hídrico da bacia hidrográfica em questão. Tendo esta informação em mãos, os órgãos responsáveis são capazes de tomar as decisões, buscando o equilíbrio entre as necessidades ambiental, social, jurídica e econômica de cada região.

Segundo Silva et al. (2011), a vazão mínima de referência, corresponde ao volume de água presente no rio nas condições anuais de maior escassez, desta forma ela define o limite superior para a captação no curso d'água. A utilização deste critério para os processos de outorga, visa a proteção do meio aquático, já que garante uma vazão de base mesmo em períodos secos (Harris et al., 2000).

No Brasil, cada estado da federação tem autonomia para realizar o processo de outorga, podendo definir os critérios para a determinação das vazões mínimas de referência, mesmo que estes não sejam claros e não havendo necessidade de demonstrar as bases para a adoção dos valores (Cruz, 2001).

3.4. Vazão outorgável e vazão de referência

Segundo Cardoso da Silva & Monteiro (2004), o uso das vazões mínimas como referência para a concessão das outorgas de uso da água, permite a garantia da preservação do recurso nos mananciais. Tendo este conceito como base, é possível realizar cálculos de alocação da água, de modo a garantir as necessidades prioritárias de uso em períodos de ocorrência das vazões mínimas (Lelis et al., 2020).

O artigo 7º da Lei Federal nº 9433/1997 (Brasil, 1997) que institui a Política Nacional de Recursos Hídricos, diz que a outorga deve estar vinculada a estudos referentes ao “balanço entre disponibilidades e demandas futuras dos recursos hídricos, em quantidade e qualidade, com identificação de conflitos potenciais”. Sendo assim, os dados de vazão se tornam fundamentais para a realização de análises e na tomada de decisão dos órgãos responsáveis pelos recursos hídricos nacionais (Moreira et al., 2022).

A vazão mínima com sete dias de duração e um período de retorno de 10 anos ($Q_{7,10}$) e as vazões associadas aos níveis de permanência de 90% (Q_{90}) e 95% (Q_{95}) do tempo são utilizadas como referenciais para a concessão do uso da água (ANA, 2011). Devido à ausência ou a inconsistência destes dados em parte dos corpos d'água nacionais, faz necessário o uso de técnicas que permitam a transferência de informações entre bacias hidrográficas que possuam semelhanças morfológicas e climáticas (Sivapalan et al., 2012).

A curva de permanência é uma maneira de organizar os dados de vazão de um curso d'água seja diária, semanal, mensal ou anual, permitindo a estimativa da fração do tempo em que o fluxo é igual ou superior a um valor de interesse (Kamran et al., 2020). O formato da curva e sua interpretação dependem das características fisiográficas e hidrológicas da bacia, além do período de análise. A curva também serve como um indicador das condições

hidrológicas da bacia, sendo a inclinação da curva em um dado momento um indicativo da taxa de fluxo naquele momento (Zheng et al., 2007). A partir da curva de permanência é possível obter as vazões mínimas de referência para outorga de uso, presentes em 90% (Q_{90}) e 95% (Q_{95}) do tempo nos cursos d'água. Os valores estabelecidos como limite máximo outorgável variam de acordo com a legislação local para cada estado brasileiro. Os estados da Bahia e Pernambuco, por exemplo, adotam 80% da Q_{90} ; já a União tem como base 70% da Q_{95} (Garrido, 2003).

A ausência de padronização entre os estados no que se diz respeito aos critérios adotados para a definição das vazões de referência e os percentuais passíveis de outorga, representam desafios para os órgãos gestores, dificultando a gestão de bacias hidrográficas interestaduais (Lelis et al., 2020)

3.4.1. Vazão mínima com sete dias de duração consecutivos e período de retorno de dez anos ($Q_{7,10}$)

A vazão mínima com sete dias de duração e período de retorno de dez anos ($Q_{7,10}$) foi definida visando estipular um intervalo de tempo em que fosse possível abranger as variações de consumo hídrico ao longo dos dias, permitindo assim, que os reservatórios pudessem absorver estas variações durante a semana, minimizando desta forma os impactos a população (Mendes, 2007). Segundo Ono (2006), o período de retorno de 10 anos é considerado aceitável, já que este corresponde a 10% de probabilidade de a vazão ser inferior ao nível mínimo necessário para suprir as necessidades urbanas.

Como para o Brasil, a $Q_{7,10}$ ainda é utilizada nos Estados Unidos da América (EUA) por alguns estados e agências federais de proteção ambiental, como um indicador de padrão de qualidade para ambientes aquáticos, sendo que estudos recentes sugerem que a $Q_{7,10}$ pode representar a vazão mínima para a proteção da vida dos peixes, desde que a série histórica para a definição da vazão não sofra influência de processos antrópicos, o que permite o desenvolvimento de uma comunidade biológica permanente ao longo do curso d'água (Ubertini et al. 1996). Segundo Tucci (2002), a $Q_{7,10}$ tem como principal vantagem sofrer fraca influência de processo antrópicos e de erros operacionais, o que a torna mais precisa e mais confiável que a vazão mínima mensal.

Em trabalho realizado por Pyrce (2004), foram levantados os diferentes usos da $Q_{7,10}$ citados na literatura, entre eles: a) um dos principais índices para a definir a vazão mínima; b) referência para a proteção ou regulação da qualidade da água, visando evitar impactos

ecológicos no ambiente aquático; c) critério para avaliação dos impactos das mudanças climáticas e da irrigação sobre os corpos d'água superficiais; d) referência para projetos específicos para instalações de retenção de águas pluviais.

Kinnard et al. (2022) utilizaram a $Q_{7,10}$ para avaliaram os impactos das condições de verão e inverno sobre as vazões mínimas de doze bacias distribuídas sobre o rio St. Lawrence, uma região de planície e que sobre efeito da neve, localizada na província de Quebec no Canada, chegando à conclusão que o efeito das mudanças climáticas e a redução da quantidade de neve poderão no futuro reduzir as vazões mínimas durante o verão, aumentando os eventos de seca e impactando a integridade do ecossistema e dos serviços ecológicos.

Já Hammond & Fleming (2021) utilizaram as vazões $Q_{7,10}$ calculadas a partir de dados de 325 estações fluviométricas localizadas na bacia do rio Delaware, inserida nos estados de New York, Pennsylvania, New Jersey, Maryland e Delaware, nos Estados Unidos, visando avaliar o impacto das vazões mínimas durante o verão, sua influência sobre a competição entre humano e o uso ecológico, além de buscar antecipar e planejar ações para vazões futuras.

Cingolani et al. (2020) utilizaram os dados de 24 bacias hidrográficas localizadas da parte superior da serra de Córdoba, região central da Argentina, para avaliar o feito das queimadas sobre as vazões mínimas, chegando à conclusão que as bacias afetadas pelos incêndios de verão tiveram uma redução de cerca de 40% das vazões mínimas em comparação as bacias que não foram expostas ao fogo.

3.4.2. Vazões de referência com 90% e 95% de permanência ao longo do tempo (Q_{90} e Q_{95})

Segundo a ANA (2011), a Q_{90} representa a vazão registrada em um dado momento por um posto fluviométrico sobre um curso d'água, sendo que em 90% do tempo, as vazões observadas serão iguais ou superiores a ela. A partir disso, é possível dizer que naquela seção do curso d'água as vazões serão superiores a Q_{90} com um nível de confiança de 90%. A Q_{90} pode ser extrapolada para outras seções do curso d'água próximas, tendo como referência a área da bacia hidrográfica e a precipitação da região.

Como a Q_{90} , a Q_{95} representa uma vazão registrada em dado momento, porém, neste caso, com uma garantia de que em 95% do tempo as vazões observadas serão iguais ou superiores a Q_{95} . Neste caso, temos uma redução no valor da vazão de referência e

consequentemente uma maior garantia da permanência do fluxo d'água ao longo do tempo (Lelis et al., 2020).

3.5. Validação dos modelos de regionalização

O processo de validação de um modelo busca avaliar o quão preciso ele é para prever o comportamento do sistema de interesse. Ele deve ser realizado tanto operacionalmente, ou seja, comparando a saída do modelo com os dados observados, quando conceitualmente, determinando se o modelo atende a base teórica e as pressuposições esperadas. Os modelos podem ser validados confrontando os dados de saída com um conjunto de dados reais (observados) ou dados experimentais independentes (simulados), desde que estes se adequem ao cenário simulado. Porém, a validação através dos dados independentes pode ser impossibilitada quando o cenário simulado se estende fora do âmbito das condições observadas, por exemplo, prever respostas a futuras mudanças climáticas, ou quando são utilizadas previsões probabilísticas, ou seja, aquelas que incluem incerteza nos processos do sistema (Rykiel 1996; Aumann, 2007).

Em estudo realizado por Meira et al. (2022), foram analisados 7 anos (2014 a 2020) de dados de 3000 estações pluviométricas espalhadas pelo Brasil, visando propor um novo método automático para o controle de qualidade das séries históricas das estações, permitindo desta forma a identificação de postos com falhas. O desempenho do método foi avaliado através de uma matriz de confusão, comparando os resultados dos procedimentos com um conjunto de dados de referência. Os resultados mostraram que o procedimento proposto foi capaz de identificar corretamente 78,6% dos pluviômetros defeituosos para o período analisado.

Vidyarthi & Jain (2020) propuseram uma nova abordagem, utilizaram a técnica das redes neurais artificiais (ANNs), para a geração de um modelo de seca para a Índia, por meio de uma árvore de decisão. O resultado mostrou que a ANN foi capaz de aprender as regras definidas para a obtenção do modelo, indicando que esta técnica pode servir como uma ferramenta para a previsão de secas para o país.

Em estudo realizado por Liu et al. (2020), visando o desenvolvimento e validação de um modelo hidrológico baseado em parâmetros físicos, em escala nacional para a China, em que a área de estudo foi separada em 19406 sub-bacias, sendo as simulações dos processos hidrológicos conduzidas por 62 anos (1956 a 2017). A eficiência dos modelos foi demonstrada comparando-se as vazões simuladas com as vazões calculadas de 203 estações

fluviométricas distribuídas pelo país. A validação dos modelos foi realizada com base em dois critérios: a) a minimização do erro relativo (ER) da vazão média dos rios, que foi inferior a 10% em 95% das estações e; b) a maximização do índice de eficiência de Nash-Sutcliffe (NSE) da vazão mensal do rio, que foi superior a 0,7 em 80% das estações.

McMillan et al. (2016) conduziram um estudo buscando a validação de um modelo hidrológico nacional para a Nova Zelândia, que se baseou em 31 parâmetros relativos aos processos físicos ligados à topografia, ao uso e à cobertura do solo entre outros. A validação dos modelos se deu com base em três critérios: a) índice de eficiência de Nash-Sutcliffe, calculado para as séries diárias de vazão; b) estatísticas que descrevem as séries de vazão anuais e comparam com os dados observados e modelados (vazão máxima anual, vazão média anual, vazão mínima anual e proporção da vazão em fevereiro); e c) assinatura hidrológica calculada sobre a série de vazão diária que descreve o comportamento do regime de vazão (índice de vazão de base, densidade de crescimento da curva de vazão, ângulo da curva de permanência das vazões e ângulo da distribuição de picos de vazão). O desempenho dos modelos variou no tempo e no espaço, com melhores resultados para as bacias de porte grande e médios e para aquelas em que as variações sazonais são menores.

Em pesquisa que abordou dois estudos de caso, realizado por Kingston et al. (2005), as redes neurais artificiais (ANN) foram utilizadas para a calibração e validação dos modelos hidrológicos. No primeiro caso, foi gerado um modelo autorregressivo (AR) genérico para fins didáticos, em que se buscou demonstrar as incertezas associadas ao treinamento e validação por meio das redes neurais artificiais. No segundo caso, foi utilizado um modelo autorregressivo (AR) real para a previsão da vazão do riacho Boggy, localizado na bacia do rio Ovens no estado de Victoria na Austrália, em que 17 anos de dados de precipitação, evapotranspiração e vazão (1976 a 1992) foram disponibilizados para a calibração e validação da ANN. Os resultados mostraram que uma ANN com uma única camada e dois nós foi capaz de simular com precisão os dados de vazão através dos parâmetros de entrada fornecidos.

4. MATERIAL E MÉTODOS

4.1. Localização e caracterização da área de estudo

A região do MATOPIBA (Figura 1) se estende sobre parte dos estados do Maranhão, Tocantins, Piauí e Bahia, ocupando uma área aproximadamente 730.000 km², com altitudes variando de 0 a 1051 metros acima do nível do mar; localiza-se entre as latitudes 2°30'S – 15°15'S, e as longitudes 43°00'O – 50°00'O (Santiago et al., 2022).

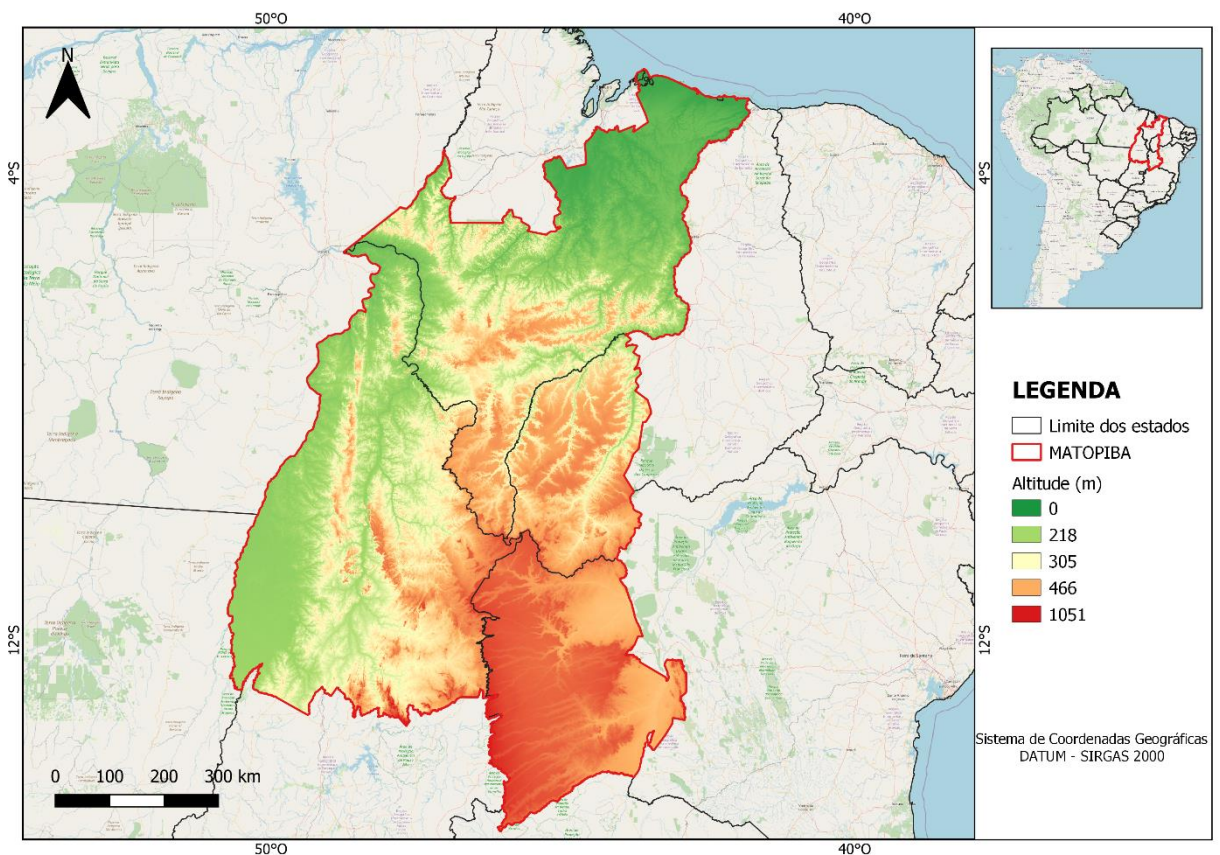


Figura 1 - Delimitação do MATOPIBA e relevo da região.

Fonte: Elaboração a partir de dados provenientes da plataforma Geoweb Matopiba (Embrapa, 2022) e IBGE (2021).

O MATOPIBA compreende 10 mesorregiões (cinco no estado do Maranhão, duas no Tocantins, uma no Piauí e duas na Bahia), 337 municípios, estando inserido em três biomas brasileiros, sendo 91% da sua área composta pelo Cerrado, 7,3% pela Amazônia e 1,7% pela Caatinga. A composição da cobertura vegetal natural é formada por 63,6% de savana, 15% de área de transição ecológica e 10,7% de floresta estacional decidual (Bolfe et al., 2017).

Conforme a classificação de Köppen-Geiger, predomina na região o clima tropical úmido com inverno seco (Aw), com temperatura média mensal entre 25°C e 27°C e

precipitação média anual variando entre 800 mm e 2000 mm, distribuída em dois períodos bem-marcados: o período seco ocorre entre junho e setembro, com valores abaixo de 600 mm e o período chuvoso que ocorre entre outubro e maio, com valores acima de 900 mm (Nascimento & Novais, 2020).

Grande parte da região é composta por latossolos e neossolos, sendo os latossolos predominantes, cobrindo cerca de 27,8 milhões de hectares (38% da região). São solos profundos e bem desenvolvidos, possuem boa permeabilidade e alta porosidade, porém baixa fertilidade natural. O Neossolo Quartzarênico é o segundo tipo de solo mais presente na região, ocupando cerca de 18 milhões de hectares (25% da região); são solos profundos e arenosos, compostos em grande parte de quartzo, possuem baixa fertilidade natural e alta porosidade e permeabilidade, com baixa capacidade de retenção de água (Magalhães & Miranda, 2014).

A região possui cerca de 324 mil estabelecimentos agrícolas, 46 unidades de conservação, 35 terras de grupos indígenas e 781 assentamentos de reforma agrária, sendo responsável por aproximadamente 14,1% da produção nacional de soja, correspondendo a cerca de 17,85 milhões de toneladas na safra 2021/2022 (CONAB, 2022).

4.2. Base de dados

Os dados de precipitação diária (mm dia^{-1}) e de vazão média diária ($\text{m}^3 \text{s}^{-1} \text{dia}^{-1}$) foram obtidos por meio das séries históricas das estações pluviométricas e fluviométricas para a região de estudo, através do acesso ao portal *Hidroweb* do Sistema Nacional de Informações sobre Recursos Hídricos - SNIRH da Agência Nacional de Águas e Saneamento Básico – ANA (ANA, 2021). Foram selecionadas as estações com no mínimo 30 anos de dados e que possuíam no máximo 20% de ausência de dados por ano, restringido o início da série ao ano de 1980. O período de 30 anos de dados é considerado o tempo mínimo para representar uma série histórica estacionária, ou seja, uma série que possua uma evolução ao longo do tempo de forma aleatória, com valores próximos a uma média constante, caracterizando uma estabilidade dos dados (Malekinezhad & Zare-Garizi, 2014).

Após a aplicação dos filtros, foram identificadas 83 estações fluviométricas e 262 estações pluviométricas, sendo 143 dentro dos limites do MATOPIBA e 119 na área externa. A Figura 2 apresenta a localização dos postos selecionados para este trabalho, além da rede hidrográfica das bacias que abastecem a região, com escala de 1:100000 (ANA, 2021).

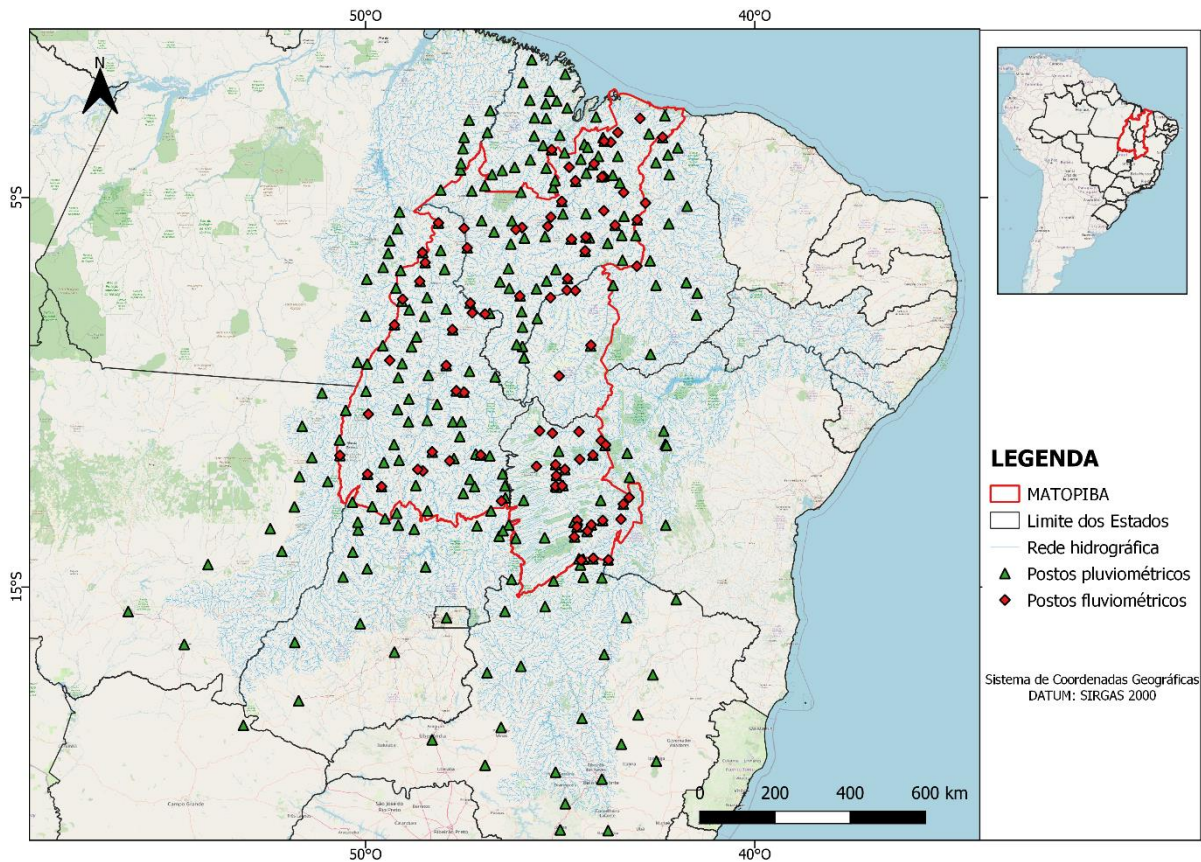


Figura 2 - Localização dos postos pluviométricos, fluviométricos e da rede hidrográfica na região do MATOPIBA.

Fonte: Elaboração a partir de dados provenientes da plataforma Hidroweb (ANA, 2021) e IBGE (2021).

4.2.1. Dados pluviométricos

4.2.1.1. Preenchimento das falhas de dados

Após os filtros iniciais foi utilizado o método da ponderação regional (PR), descrito por Bertoni & Tucci (2007) e Paulhus & Kohler (1952), visando preencher os dados faltantes de precipitação das séries. A estimação dos valores foi realizada baseando-se na Equação 1:

$$D_x = \frac{1}{n} \sum_{i=1}^n \frac{M_x}{M_i} D_i \quad (1)$$

Em que:

D_x – representa o dado mensal faltante a ser estimado para a estação teste [mm dia^{-1}];

D_i – representa o dado ocorrido na estação vizinha de ordem “i” no mês de ocorrência da falha na estação teste [mm dia^{-1}];

M_x – representa o dado médio mensal da estação teste [mm dia^{-1}]; e

M_i – representa o dado médio mensal da estação vizinha de ordem “i” e n representa o número de estações vizinhas utilizadas no cálculo [mm dia^{-1}].

4.2.1.2. Cálculo da precipitação média anual

Com os dados de precipitação devidamente organizados, por meio do software *RStudio 2022.07.1+554* (R Core Team, 2022), foram calculadas as precipitações totais anuais (Equação 2) e as precipitações médias anuais (Equação 3) dos postos pluviométricos selecionados.

$$P_a = \sum_{i=1}^n P_d \quad (2)$$

$$P_t = \frac{\sum_{i=1}^n P_a}{n} \quad (3)$$

Em que:

P_d – representa a precipitação diária [mm dia^{-1}];

P_a – representa a precipitação total anual [mm ano^{-1}];

P_t representa a precipitação média anual [mm ano^{-1}]; e

n representa o número de anos de observações [adimensional].

4.2.1.3. Geração das isoietas para a região

As etapas a seguir foram realizadas através dos softwares *RStudio 2022.07.1+554* (R Core Team, 2022) para a realização dos cálculos e seleção do modelo e do software de geoprocessamento *QGIS 3.16.11 with Grass 7.8.5* (QGIS, 2022) para a interpolação e extração das isoietas.

4.2.1.3.1. Semivariograma

Utilizada na área de geoestatística, a função de semivariância permite identificar a dependência espacial de variáveis, através da geração de semivariogramas. O gráfico do semivariograma experimental expressa a variabilidade espacial entre as amostras de interesse,

sendo uma função dependentes somente do vetor d (distância entre os pares de casos amostrados). À medida que o vetor d aumenta, o semivariograma se aproxima da variabilidade total dos dados (patamar), momento este em que há uma estabilização do semivariograma, além disso, através dele é possível realizar o cálculo dos parâmetros do modelo, como a distância da dependência espacial (range ou alcance), que representa o alcance máximo da dependência espacial da variável; o efeito pepita (nugget effect), que representa a o valor da semivariância a distância zero, ou seja, a variabilidade para escalas menores que a distância de amostragem; e altura (sill), dada pela diferença entre o patamar e o efeito pepita,; neste intervalo é possível identificar a dependência espacial dos dados (Opromolla et al., 2006). O estimador da função de semivariância para variáveis regionalizadas com distribuição normal de probabilidade é dado pela Equação 4 (Matheron, 1963).

$$y^*(d) = \frac{1}{2n(d)} \sum_{i=1}^{n(d)} [Z(S_i) - Z(S_i + d)]^2 \quad (4)$$

Em que:

$n(d)$ – Número de pares em uma mesma distância d ; e

$Z(S_i)$ e $Z(S_i + d)$ – Valores da variável na posição S_i e $S_i + d$, com $s \in S \subset R^2$.

4.2.1.3.2. Modelos de semivariogramas

Neste trabalho foram comparados três modelos de semivariograma: a) modelo esférico, b) modelo exponencial e c) modelo da família Matérn. Suas expressões são apresentadas pelas Equações 5, 6 e 7 respectivamente:

$$\rho(d) = t^2 + \sigma^2 \left[\frac{3}{2} \left(\frac{d}{\phi} \right) - \frac{1}{2} \left(\frac{d}{\phi} \right)^3 \right] \text{ se } 0 \leq d \leq \phi \quad (5)$$

$$t^2 + \sigma^2 \quad \text{se } d > 0$$

$$\rho(d) = t^2 + \sigma^2 \left[1 - \exp \left(\frac{-d}{\phi} \right) \right] \text{ se } 0 \leq d \leq \phi \quad (6)$$

$$t^2 + \sigma^2 \quad \text{se } d > \phi$$

$$\rho(d) = t^2 + \sigma^2 \left[1 - \exp\left(\frac{-h^2}{\phi^2}\right) \right] \text{ se } 0 \leq d \leq \phi \quad (7)$$

$$t^2 + \sigma^2 \quad \text{se } d > \phi$$

4.2.1.3.3. Método de ajuste dos semivariogramas

Para o ajuste dos modelos foi utilizado o método da máxima verossimilhança. Segundo Wonnacott & Wonnacott (1972), esta é uma técnica de estimação que possui propriedades assintóticas interessantes, com como eficiência e consistência. O método busca obter a partir de uma mostra, o valor “mais verossímil” dos parâmetros de um certo modelo probabilístico (Mello et al., 2005), em que é considerado um conjunto de observações (Z_1, \dots, Z_n) gerados a partir de um modelo linear gaussiano, podendo ser representado pela soma de três componentes, como dado na Equação 8:

$$Z_i = \mu(x_i) + S(x_i) + \varepsilon_i, \quad \text{com } i = 1, \dots, n \quad (8)$$

Em que:

$\mu(x_i)$ – É uma componente determinística dada por um modelo linear (Equação 9):

$$\mu(x_i) = \sum_{k=1}^p f_k(x_i)\beta_k \quad (9)$$

Em que:

(f_1, \dots, f_p) – é um conjunto de p funções que descrevem deterministicamente a variável Z a partir das coordenadas espaciais (x_1, \dots, x_n);

$S(x_i)$ – é uma componente aleatória entendida como um processo estocástico gaussiano autocorrelacionados.

A esperança de $S(x_i)$ é igual a zero e a variância constante (σ^2). Esta componente é que confere continuidade espacial aleatória à variável Z , isto é, $S(x)$ é que possui a correlação espacial. (Mello et al., 2005). O modelo definido na Equação 9 para a precipitação total anual com uma distribuição multivariada normal pode ser representado da seguinte forma: $Z \sim MVN(F\beta, \sigma^2 \Sigma + t^2 I)$. Logo, a função de verossimilhança a ser maximizada é dada pela Equação 10:

$$L(\beta, \theta) = -\frac{1}{2} \{ \log|K| + (Z - F\beta)^t [K]^{-1} (Z - F\beta) \} \quad (10)$$

4.2.1.3.4. Seleção do modelo de semivariograma

A seleção do modelo que se melhor ajustou aos dados foi feita através do método de validação cruzada Leave-One-Out (LOO). Nesta abordagem, os dados são separados em dois grupos, onde o primeiro contém uma observação $d_k = \{(y_k, x_k)\}$, $k = 1, 2, \dots, n$, (conjunto de validação), e o segundo contém as $n-1$ observações (conjunto de treinamento). O processo é repetido n vezes, até que todas as observações sejam analisadas. A estimativa do risco esperado pode ser definida pela Equação 11:

$$LOO = \frac{1}{n} \sum_{k=1}^n L(y_k, \hat{f}_{(-k)}(x_k)) \quad (11)$$

Em que:

$\hat{f}_{(-k)}(X)$ – É o preditor criado em cada interação k do processo, retirando-se a observação (y_k, x_k) da amostra treino [mm ano⁻¹].

O método LOO é um estimador quase não viesado do erro, já que o conjunto de treinamento é constituído por quase toda a base de dados, porém ele possui alta variabilidade já que cada etapa da validação possui apenas uma observação no conjunto de validação (Borra & Ciaccio, 2010).

Nesta etapa, a avaliação do desempenho de cada modelo na estimativa do semivariograma da precipitação total anual foi baseada nos valores do R^2 e do RMSE gerados a partir da validação cruzada, em que se buscou o modelo com maior R^2 e com melhor RMSE (Equações 12 e 13).

$$R^2 = \left(\frac{\sum_{i=1}^n (O_i - \bar{O})(P_i - \bar{P})}{\left(\sqrt{\sum_{i=1}^n (|O_i - \bar{O}|)^2} \right) \left(\sqrt{\sum_{i=1}^n (|P_i - \bar{P}|)^2} \right)} \right)^2 \quad (12)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (O_i - P_i)^2}{n}} \quad (13)$$

Em que:

O_i – Representa o valor do dado de precipitação observado [mm ano^{-1}];

\bar{O} – É o valor médio dos dados de precipitação observados [mm ano^{-1}];

P_i – Representa o valor do dado de precipitação predito [mm ano^{-1}]; e

\bar{P} – É o valor médio dos dados de precipitação preditos [mm ano^{-1}].

4.2.1.3.5. Interpolação dos dados

A interpolação dos dados de precipitação total anual foi realizada através do interpolador de Krigagem ordinária. Este método utiliza um modelo geoestatístico univariado baseado na autocorrelação espacial da variável de interesse expressa pelo semivariograma. Neste processo assume-se médias moveis, o que resulta na minimização da média do erro (Yamamoto & Landim, 2013). O estimador da krigagem é definido pela Equação 14 (Yamamoto, 2020):

$$Z^*_{(x_0)} = \sum_{i=1}^n \delta_i Z_{xi} \quad (14)$$

Em que:

Z^* - Representa o valor de precipitação estimado [mm ano^{-1}];

δ_i – Representa a variância [mm ano^{-1}]; e

Z – Representa o valor da amostra [mm ano^{-1}].

4.2.1.3.6. Extração das isoietas

Após o processo de krigagem ordinária foi gerado um arquivo raster, permitindo a extração das isoietas (linhas que representam pontos de igual pluviosidade) com valores de 50 mm entre as curvas.

4.2.2. Dados fluviométricos

4.2.2.1. Preenchimento de falhas de dados

As falhas encontradas nas séries históricas dos dados de vazão (Figura 3) foram corrigidas utilizando-se o módulo de preenchimento e extensão de séries do *software SisCAH 1.0* (Sistema Computacional para Análises Hidrológicas), desenvolvido pelo Grupo de Pesquisa em Recursos Hídricos - GPRH da Universidade Federal de Viçosa (GPRH, 2009). O módulo permite o preenchimento das falhas através da análise de correlação dos dados observados no período comum entre a estação a ser preenchida e uma estação utilizada como referência, determinada segundo critérios de escolha de bases para regressões.

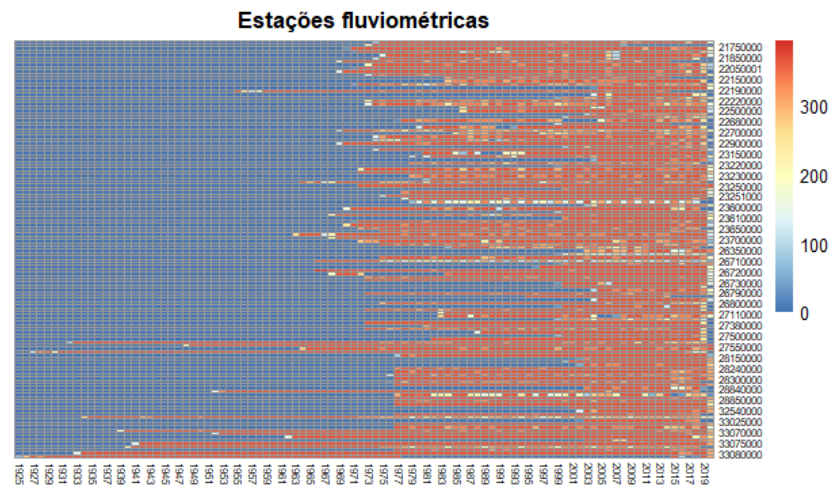


Figura 3 - Distribuição dos registros diários de vazão. O eixo inferior corresponde aos anos de registro, o eixo vertical direito corresponde ao código das estações fluviométricas e a escala vertical direita representa o código de cores correspondente ao número de dias agregados por ano.

Fonte: Elaboração a partir de dados provenientes da plataforma Hidroweb (ANA, 2021).

Em um mesmo curso d'água, inicia-se preferencialmente pela estação mais próxima da estação a ser preenchida; quando esta condição não pode ser atendida, aplica-se o princípio da semelhança hidrológica, tendo como base para a tomada de decisão os coeficientes de correlação.

Para o preenchimento de falhas, utilizou-se a regressão linear simples (Equação 15).

$$Y = \beta_0 X + \beta_1 \quad (15)$$

Em que:

Y – Vazão do posto com falhas [$\text{m}^3 \text{s}^{-1}$];

X – Vazão do posto com dados em [$\text{m}^3 \text{s}^{-1}$]; e

β_0 e β_1 – Parâmetros ajustados na regressão [adimensionais].

Adotou-se 2 (dois) critérios mínimos para a realização da regressão:

- a) Obtenção de coeficiente de determinação (R^2) superior a 0,7;
- b) Existência de pelo menos 8 (oito) pares de eventos entre as duas estações.

4.2.2.2. Cálculo das vazões de referência

Após a aplicação dos filtros e o preenchimento das falhas, os cálculos para a obtenção das vazões de referência (Q_{med} , Q_{90} , Q_{95} , $Q_{7,10}$) foram realizados com o auxílio do *software SisCAH 1.0*.

4.2.2.2.1. Cálculo da vazão média de longo período (Q_{med})

Por meio do cálculo da média aritmética das vazões das séries históricas, foi possível a determinação da vazão média de longo período (Q_{med}) para cada posto fluviométrico.

4.2.2.2.2. Cálculo das vazões com 90% (Q_{90}) e com 95% (Q_{95}) de permanência ao longo do tempo

Segundo Cruz e Tucci (2008), o método da série completa de dados permite a obtenção da curva de permanência empírica total, a qual possibilita a obtenção dos valores das vazões de referência Q_{90} e Q_{95} .

O *software SisCAH 1.0* integra a metodologia descrita por Vogel e Fennessay (1995), na qual as vazões são dispostas em ordem decrescente por classes, de acordo com a magnitude dos eventos e associadas às frequências empíricas de não superação. As curvas de permanência são obtidas através das seguintes etapas:

- a) Seleção da série de dados de vazões diárias;
- b) Definição de 50 intervalos de classes de vazões diárias;
- c) Subdivisão dos intervalos de classes com base na escala logarítmica devido à variação de magnitude das vazões;
- d) Cálculo do intervalo de classe (ΔX) (Equação 16):

$$\Delta X = \frac{[\ln(Q_{m\acute{a}x}) - \ln(Q_{m\acute{i}n})]}{50} \quad (16)$$

Em que:

ΔX – Intervalo de classe [$\text{m}^3 \text{s}^{-1}$];

$Q_{m\acute{a}x}$ – Vazão máxima da série [$\text{m}^3 \text{s}^{-1}$]; e

$Q_{m\acute{i}n}$ – Vazão mínima da série [$\text{m}^3 \text{s}^{-1}$].

- e) Cálculo dos limites dos intervalos, a partir da $Q_{m\acute{i}n}$, adicionando o intervalo calculado anteriormente, o que resulta na vazão do limite superior do intervalo i (Equação 17):

$$Q_{i+1} = \exp[\ln(Q_i) + \Delta X] \quad (17)$$

- f) Determinação do número de vazões classificadas em cada intervalo;
- g) Determinação da frequência (f_i) associada ao limite inferior de cada intervalo (Equação 18):

$$f_i = \frac{N_{qi}}{NT} \cdot 100 \quad (18)$$

Em que:

f_i – Frequência associada ao limite inferior de cada intervalo [adimensional];

N_{qi} – Número de vazões de cada intervalo [adimensional]; e

NT – Número total de vazões [adimensional].

- h) Obtenção da curva de permanência plotando-se na ordenada os limites inferiores dos intervalos de classe de vazão e na abscissa a frequência de ocorrência.

Por meio das frequências acumuladas obtidas pela curva de permanência são extraídos os valores das vazões com 90% de permanência (Q_{90}) e com 95% de permanência (Q_{95}) ao longo do tempo.

4.2.2.2.3. Cálculo da vazão mínima de 7 dias consecutivos e período de retorno de 10 anos ($Q_{7,10}$)

As funções de distribuição de densidade de probabilidade permitem estimar as vazões máximas e mínimas de séries históricas de dados para diferentes períodos de retorno. A partir dos dados de eventos mínimos é possível gerar funções para a estimativa de vazões mínimas, utilizadas para determinar a $Q_{7,10}$ (vazão média mínima de sete dias consecutivos e período de retorno de dez anos).

Em conjunto com a estimativa das vazões, os intervalos de confiança permitem estabelecer o limite de variação possível dos eventos estimados. A metodologia para a obtenção das funções de probabilidade e dos intervalos de confiança são descritas a seguir.

A estimativa da magnitude de um evento com determinado período de retorno para as distribuições de frequências utilizadas é dada pela Equação 19.

$$M = \mu + K\sigma \quad (19)$$

Em que:

M – Magnitude do evento para o período de retorno estabelecido [adimensional];

μ – Média dos eventos de vazão [$m^3 s^{-1}$];

K – Fator de frequência [adimensional]; e

σ – Desvio padrão dos eventos de vazão [$m^3 s^{-1}$].

O valor do fator de frequência (K) pode ser obtido através das funções de distribuição de frequência. O *software SisCAH 1.0* integra 6 (seis) das principais funções, atendendo assim grande parte da necessidade para a análise dos dados de vazão.

4.2.2.2.4. Distribuições de densidade de probabilidade

A distribuição de densidade de probabilidade permite descrever o comportamento aleatório de um fenômeno, ou seja, a probabilidade de cada valor de uma variável aleatória de interesse ao longo do processo (Novaes, 2022).

Por meio do *software SisCAH 1.0*, foram comparadas 6 (seis) diferentes distribuições de probabilidade (Gumbel, Log-Normal 2, Log-Normal 3, Pearson 3, Log-Pearson 3,

Weibull), visando obter aquela que melhor descrevia o comportamento do regime hídrico para cada posto fluviométrico de interesse, permitindo assim obter os valores das vazões de referência. A descrição e a equação de cada uma delas são apresentadas no Anexo A deste documento.

4.2.2.2.5. Intervalo de confiança

O intervalo de confiança utilizado pelo *software SiSCAH 1.0* é de 95%, sendo que os limites do intervalo são estimados segundo o método proposto por Kite (1988), dado pela Equação 20.

$$M - 1,96\delta < \beta < M + 1,96\delta \quad (20)$$

Em que:

M – Magnitude do evento de vazão [$\text{m}^3 \text{s}^{-1}$];

δ – Erro padrão pertinente a cada função de probabilidade [$\text{m}^3 \text{s}^{-1}$]; e

β – Intervalo de confiança em nível de probabilidade de 95% [$\text{m}^3 \text{s}^{-1}$].

O software também permite a realização do teste de aderência para a escolha da distribuição de probabilidade que melhor se ajuste aos dados. Neste caso foi utilizado o teste de Kolmogorov-Smirnov que quantifica a distância entre a função de distribuição empírica da amostra e a função de distribuição acumulada da distribuição escolhida para a comparação.

A função de distribuição empírica F_n para n observações X_i independentes e identicamente distribuídas é definida pela Equação 21.

$$F_n(X) = \frac{1}{n} \sum_{i=1}^n I_{[-\infty, x]}(X_i) \quad (21)$$

Em que:

$I_{[-\infty, x]}(X_i)$ – é a função indicadora, igual a 1 se $X_i \leq x$ e igual a 0 de outro modo.

A estatística de Kolmogorov-Smirnov para uma dada função distribuição acumulada $F(x)$ é dada pela Equação 22.

$$D_n = \sup_x |F_n(x) - F(x)| \quad (22)$$

Em que:

\sup_x – é o supremo do conjunto de distâncias [adimensional].

4.2.2.3. Cálculo das vazões específicas

A vazão específica estabelece uma relação entre a vazão de uma determinada bacia hidrográfica e sua área de captação. Devido à grande variabilidade de volume de água e tamanho das bacias dentro da área de estudo, optou-se por trabalhar com a vazão específica, o que facilita os cálculos e a compreensão dos resultados encontrados, já que os dados deste cálculo permitem “reduzir” a variabilidade entre os valores encontrados.

As vazões específicas foram utilizadas nos modelos como sendo as variáveis dependentes, sendo calculadas através das Equações 23, 24, 25 e 26.

$$Q_{med_{esp}} = \frac{Q_{med} * 1000}{A} \quad (23)$$

$$Q_{90_{esp}} = \frac{Q_{90} * 1000}{A} \quad (24)$$

$$Q_{95_{esp}} = \frac{Q_{95} * 1000}{A} \quad (25)$$

$$Q_{7,10_{esp}} = \frac{Q_{7,10} * 1000}{A} \quad (26)$$

Em que:

$Q_{med_{esp}}$ – Vazão média específica [$L s^{-1} km^{-2}$];

Q_{med} – Vazão média [$m^3 s^{-1}$];

$Q_{90_{esp}}$ – Vazão média específica com 90% de permanência [$L s^{-1} km^{-2}$]

Q_{90} – Vazão média com 90% de permanência [$m^3 s^{-1}$];

$Q_{95_{esp}}$ – Vazão média específica com 95% de permanência [$L s^{-1} km^{-2}$]

Q_{95} – Vazão média com 95% de permanência [$m^3 s^{-1}$];

$Q_{7,10_{esp}}$ – Vazão média específica da mínima de sete dias consecutivos e período de retorno de 10 anos [$L s^{-1} km^{-2}$]; e

$Q_{7,10}$ – Vazão média mínima de sete dias consecutivos e período de retorno de 10 anos [$m^3 s^{-1}$];

A – Área de drenagem [km^2].

4.2.2.4. Delimitação das áreas de contribuição e índices morfométricos

Devido a limitação computacional encontrada, dada a grande extensão da área de estudo (730.000 km^2), optou-se pela delimitação da área de contribuição de cada posto fluviométrico de forma manual. Por meio do *software* de geoprocessamento *QGIS 3.16.11 with Grass 7.8.5*, a rede hidrográfica da região foi importada e utilizada como base para definir os limites das “bacias”.

4.2.2.4.1. Estimação da precipitação média anual por bacia

A estimação da precipitação média anual por bacia foi feita pelo método das isoietas (linhas de igual precipitação). O cálculo foi através da média ponderada das alturas médias das isoietas em função da área que cada uma ocupa na área de contribuição de cada estação fluviométrica (Equação 27).

$$P_m = \frac{1}{A} * \sum A_i * \left(\frac{I_i + I_{i+1}}{2} \right) \quad (27)$$

Em que:

P_m – Precipitação média anual por bacia [$mm \text{ ano}^{-1}$];

A – Área total da bacia [km^2];

A_i – Área entre as isoietas [km^2]; e

I – Valor da isoietas [$mm \text{ ano}^{-1}$].

4.2.2.4.2. Índices morfométricos

A delimitação das áreas de contribuição, permitiu o cálculo da área e do perímetro de cada bacia; os valores obtidos foram comparados aos fornecidos pela plataforma *Hidroweb* afim de garantir a consistência dos resultados.

Tendo estabelecidas as áreas de contribuição para cada posto fluviométrico, por meio do software de geoprocessamento *QGIS 3.16.11 with Grass 7.8.5*, foi possível a obtenção do comprimento do talvegue principal, do comprimento total da rede hidrográfica e das coordenadas UTM do centroide da bacia. Os outros índices morfométricos foram obtidos através de equações, que são apresentadas a seguir.

- Densidade de drenagem (Dd), índice que representa o grau de desenvolvimento do sistema de drenagem de uma bacia hidrográfica, permitindo fazer inferências sobre a sua eficiência de drenagem (Beana, 2002). É definido como a razão entre o somatório do comprimento total dos cursos d'água (perenes, intermitentes ou efêmeros) de uma bacia, por sua área, conforme a Equação 28.

$$Dd = \frac{\sum \text{comprimento dos rios}}{A} \quad (28)$$

Em que:

Dd – Densidade de drenagem [km km⁻²]; e

A – Área da bacia [km²].

- Coeficiente de Compacidade (Kc) ou Índice de Graveis, relaciona a forma da bacia com a de um círculo, ou seja, é a razão entre o perímetro da bacia e o perímetro de um círculo de área igual à da bacia. Trata-se de um valor adimensional, variando com a forma da bacia, independentemente do seu tamanho (Villela & Mattos, 1975). Quanto maior for o resultado do coeficiente, mais irregular e menos susceptível a inundações será a bacia, é calculado conforme a Equação 29.

$$Kc = \frac{P}{2\pi \sqrt{\frac{A}{\pi}}} \quad (29)$$

Em que:

Kc – Coeficiente de compacidade [adimensional];

P – Perímetro da bacia [km]; e

A – Área da bacia [km²].

- Declividade do talvegue da bacia (I_s), consiste na razão entre a diferença de cota entre a nascente e a foz do curso d'água, e o comprimento do talvegue, sendo calculado pela Equação 30.

$$I_s = \frac{\text{cota nascente} - \text{cota foz}}{\text{comprimento do rio}} \quad (30)$$

Em que:

I_s – Declividade do talvegue da bacia [m km^{-1}];

Os valores das cotas são em metros e do comprimento do rio principal [km].

Estes parâmetros permitiram a definição dos índices morfológicos que, em conjunto com a precipitação média anual por bacia, serviram para a delimitação das regiões hidrologicamente homogêneas dentro da região do MATOPIBA, além de permitirem a definição de modelos para a estimação das vazões de referência.

4.3. Análise de agrupamento (Análise Cluster)

Após o levantamento dos índices morfométricos e da precipitação média anual das bacias, foi realizada a análise de agrupamento (*Cluster*) por meio do *software RStudio*, definindo assim as regiões homogêneas dentro da área de estudo.

A análise de agrupamento permite classificar componentes em grupos, de forma que os elementos com semelhanças pertençam a um mesmo cluster, e os elementos com diferenças pertençam a clusters distintos. A semelhança ou diferença entre os elementos é definida através da distância Manhattan (ou *City Block*) entre os elementos, na qual fornece a soma dos comprimentos da projeção da linha que combina eixos e coordenadas, sendo menos afetada por *outliers* (Equação 31).

$$d_{ij} = \sum_{k=1}^p |x_{ik} - x_{jk}|; \forall r \in R: r \geq 1 \quad (31)$$

Em que:

d – Diferença entre dois elementos [adimensional]; e

x – Representa os pontos no espaço [adimensional].

O método da clusterização hierárquica foi escolhido por permitir uma maior flexibilidade na escolha no número total de grupos, em que além das análises estatísticas, o conhecimento do pesquisador sobre a região é levado em consideração.

Neste método, os algoritmos de agrupamentos hierárquicos são formados a partir de uma matriz de similaridade ou dissimilaridade, na qual é identificado o par de parcelas que mais se assemelham. Segundo Kassambara (2017), este método segue três etapas para a formação dos grupos:

- a) Inclusão inicial de cada objeto considerando um agrupamento individual;
- b) Formação de pares de grupos com maiores similaridades entre si; e
- c) Repetição da etapa anterior até a obtenção de um único grande grupo.

O método de Ward (Ward, 1963), é baseado em um processo geral de classificação, no qual n parcelas são progressivamente reunidas dentro de grupos, por meio da minimização de uma função objetiva para cada $(n-2)$ passo de fusão.

Este algoritmo baseia-se na perda de informação resultante do agrupamento das variáveis de interesse e é medida por meio da soma dos quadrados dos desvios das parcelas individuais relativamente às médias dos grupos em que são classificadas. Para Barroso & Artes (2003), o algoritmo de Ward é interessante por se basear em uma medida com forte apelo estatístico e por gerar grupos que, assim como os do método do vizinho mais distante, apresentam alta homogeneidade interna.

Romesburg (1984) cita as quatro principais características do método:

- a) Apresenta bons resultados tanto para a distância euclidiana quanto para outras distâncias;
- b) Pode apresentar resultados insatisfatórios quando o número de parcelas em cada grupo é praticamente igual;
- c) Tem tendência a combinar grupos com poucas parcelas;
- d) É sensível à presença de *outliers*.

Nesta análise foram utilizados sete parâmetros, sendo seis morfológicos (densidade de drenagem; comprimento do talvegue principal; declividade média da bacia; coeficiente de compacidade; e coordenadas do centroide da bacia) e a precipitação média anual estimada para a bacia.

4.4. Regressão linear multivariada

4.4.1. Modelos de regionalização hidrológica

Tendo definido as regiões homogêneas, a próxima etapa foi a geração das equações para a estimação da vazão para cada um dos postos (exutório das bacias). As mesmas variáveis utilizadas para a análise de cluster, foram utilizadas como variáveis independentes dos modelos de regressão linear multivariada.

Inicialmente foi utilizado um modelo dito completo, em que considera todas as variáveis já apresentadas para a estimação da vazão específica. Nesta fase, foram gerados modelos para a estimação da $Q_{med_{esp}}$, $Q_{90_{esp}}$, $Q_{95_{esp}}$ e $Q_{7,10_{esp}}$. Abaixo é apresentado um exemplo do modelo base para os cálculos (Equação 32):

$$Q_{med_{espr}} = a.Dd + b.L + c.Is + d.Kc + e.P + f.Cx + g.Cy \quad (32)$$

Em que:

$Q_{med_{espr}}$ – Vazão média de longo período específica regionalizada [$L s^{-1} km^{-2}$];

Dd – Densidade de drenagem [$km km^{-2}$];

L – Comprimento do talvegue principal [km];

Is – Declividade média do talvegue principal [$m km^{-1}$];

Kc – Coeficiente de compacidade [adimensional];

P – Precipitação total anual da bacia [$mm ano^{-1}$];

Cx – Coordenada UTM da latitude do centroide da bacia [m];

Cy – Coordenada UTM da longitude do centroide da bacia [m]; e

a, b, c, d, e, f, g representam os coeficientes gerados pela regressão [adimensional].

A mesma equação foi utilizada como base para a definição dos modelos para a estimação da vazão específica regionalizada da Q_{90} , Q_{95} e da $Q_{7,10}$.

Além do modelo base, foi testado um segundo modelo, gerado através do critério de informação Akaike (AIC), em um processo “backward”, em que se inicia com um modelo com todas as variáveis e gradativamente elas são removidas até que se encontre um modelo com o menor valor de AIC.

Por meio do *software RStudio 2022.07.1+554*, os dois modelos gerados passaram pelos testes sobre os resíduos afim de garantir os pressupostos estatísticos da regressão:

a) Teste de Shapiro-Wilk (Shapiro & Wilk, 1965)

O teste de Shapiro-Wilk analisa a hipótese nula de que uma amostra x_1, \dots, x_n são provenientes de uma população de dados normalmente distribuídos. A estatística do teste é dada pela Equação 33.

$$W = \frac{(\sum_{i=1}^n a_i x_{(i)})^2}{\sum_{i=1}^n (x_i - x_{\text{méd}})^2} \quad (33)$$

Sendo:

Os coeficientes a_i dados pela Equação 34.

$$(a_1, \dots, a_n) = \frac{m^T V^{-1}}{C} \quad (34)$$

Sendo C é uma norma de vetor (Equação 35).

$$C = \|V^{-1}m\| = (m^T V^{-1} V^{-1} m)^{\frac{1}{2}} \quad (35)$$

E o vetor m (Equação 36).

$$m = (m_1, \dots, m_n)^T \quad (36)$$

Em que:

$x_{(i)}$ – i -ésima estatística de ordem, ou seja, o i -ésimo menor número da amostra [$L s^{-1} km^{-2}$]; e
 $x_{\text{méd}}$ – Média da amostra de vazão [$L s^{-1} km^{-2}$].

b) Teste de Durbin-Watson (Durbin & Watson, 1950)

O teste de Durbin-Watson é utilizado para detectar a presença de autocorrelação entre os resíduos de uma regressão. A estatística do teste é dada pela Equação 37.

$$d = \frac{\sum_{i=2}^N (e_i - e_{i-1})^2}{\sum_{i=1}^N e_i^2} \quad (37)$$

Em que:

N – Número total de observações [adimensional]; e

e_i – i -ésimo resíduo do modelo de regressão [$L s^{-1} km^{-2}$].

c) Teste de Breusch-Pagan (Breusch & Pagan, 1979)

O teste de Breusch-Pagan é usado para verificar a homoscedasticidade dos resíduos. Ele é baseado no teste do χ^2 , se a estatística do teste de Breusch-Pagan é superior à aquela obtida pelo teste do χ^2 , ou seja, se o p-valor for inferior a um certo limite (normalmente 5%), rejeita-se a hipótese nula de homoscedasticidade com um risco de erro do tipo I de 5%.

d) Teste de Multicolinearidade (Fox & Monette, 1992)

A análise da possível multicolinearidade entre as variáveis explicativas dos modelos de regressão linear foi avaliada através do Fator de Inflação da Variância (VIF), o que permitiu a eliminação de variáveis redundantes e a obtenção de modelos mais parcimoniosos. A estatística do teste é dada pela Equação 38.

$$VIF = \frac{1}{1-R_i^2} \quad (38)$$

Em que:

R_i^2 - Representa o coeficiente de determinação não ajustado para a regressão da i -ésima variável independente sobre as demais [adimensional].

4.4.2. Análise estatística dos modelos

A avaliação do desempenho dos modelos de regionalização hidrológica gerados deu-se pela comparação entre os valores estimados através dos modelos e os valores observados calculados para cada posto fluviométrico pelo *software SisCAH 1.0*.

As diretrizes para a avaliação dos modelos seguiram a proposta de Moriasi et al. (2015), em que são realizados 5 (cinco) testes estatísticos quantitativos: a) Índice de concordância (d); Teste de eficiência de Nash-Sutcliffe (NSE); c) Viés percentual (PBIAS); d) Razão entre o RMSE e o desvio padrão dos dados observados (RSR); e) Coeficiente de determinação (R^2). A descrição de cada teste e suas respectivas equações são apresentadas no Anexo B deste documento.

Após a realização destes testes, por meio da classificação qualitativa (muito bom, bom, satisfatório ou insatisfatório) com limites quantitativos correspondentes para cada índice estatístico foi possível definir um critério para a avaliação do desempenho dos modelos. Seguindo a recomendação de Moriasi et al. (2015), a Tabela 1 apresenta os critérios de avaliação utilizados com seus devidos limites quantitativos.

Tabela 1 - Critérios de avaliação qualitativa de desempenho e seus limites quantitativos para os índices estatísticos recomendados para os modelos de regionalização hidrológica.

Índice	Desempenho			
	Muito bom	Bom	Satisfatório	Insatisfatório
d	>0,90	0,85 < d ≤ 0,90	0,75 < d ≤ 0,85	d ≤ 0,75
NSE	>0,80	0,70 < NSE ≤ 0,80	0,50 < NSE ≤ 0,70	≤ 0,50
PBIAS (%)	<± 5	± 5 ≤ PBIAS ≤ ± 10	± 10 ≤ PBIAS ≤ ± 15	PBIAS ≥ ± 15
R²	>0,85	0,75 < R ² ≤ 0,85	0,60 < R ² ≤ 0,75	≤ 0,60

Fonte: Adaptado de Moriasi et al. (2015).

4.4.3. Validação dos modelos de regionalização hidrológica

4.4.3.1. Intervalo de classes dos valores de vazão

A primeira etapa do processo de validação, foi a separação em diferentes classes dos valores de vazão específica estimada para cada uma das vazões de referência. Realizada de maneira aleatória, geralmente são utilizadas entre 5 e 20 classes. Segundo Assis et al. (1996), a baixa ou a elevada quantidade de classes pode ocultar certas propriedades da distribuição de frequência; sendo assim, foram definidas 5 classes para as variáveis de vazões específicas calculadas e vazões específicas estimadas. Após a definição das classes, foram calculadas as frequências relativas, f_i (Equação 39), ou seja, o número de vezes que ocorrem valores dentro da amplitude da classe em relação ao número total; e as frequências acumuladas f_{ac} (Equação 40), isto é, a soma das frequências relativas ocorridas em cada classe.

$$f_i = \frac{n_i}{n} \quad (39)$$

$$f_{ac} = f_i + f_{ac-1} \quad (40)$$

Em que:

n – É o número total de observações do conjunto de dados [adimensional], neste caso, $n = 83$;

n_i – É a frequência do valor i [adimensional].

f_i – É a frequência relativa [adimensional]; e

f_{ac} – É a frequência acumulada [adimensional].

4.4.3.2. Matriz de confusão

O próximo passo foi a construção da matriz de confusão, que se resume a uma tabela que descreve o desempenho dos modelos, permitindo distingui-los e classificá-los em função dos valores verdadeiros conhecidos em um conjunto de dados de teste. Foram calculadas quatro diferentes matrizes, uma para cada vazão de referência abordada neste trabalho. Cada matriz de confusão é composta por quatro elementos (Fratini et al., 2010):

- a) Verdadeiro positivo (VP) – em que o valor estimado pelo modelo será positivo, assim como o valor real.
- b) Verdadeiro negativo (VN) – em que o valor estimado pelo modelo será negativo, assim como o valor real.
- c) Falso Positivo ou Erro de tipo 1 (FP) – em que o valor estimado pelo modelo será positivo, no entanto, o valor real é negativo. O FP também pode ser denominado como Erro de Tipo 1.
- d) Falso Negativo ou Erro do tipo 2 (FN) – em que o valor estimado pelo modelo será negativo, no entanto, o valor real é positivo. O FN também pode ser denominado como Erro de Tipo 2.

A partir destes elementos da matriz, foram calculadas quatro métricas utilizadas para a classificação dos modelos:

- 1) Sensibilidade – é a razão entre os verdadeiros positivos e os verdadeiros positivos e falsos negativos. Em outros termos, indica qual a proporção de positivos foi identificada corretamente (Equação 41).

$$\text{Sensibilidade ("recall")} = \frac{VP}{VP+FN} \quad (41)$$

- 2) Precisão – é a razão entre o verdadeiro positivo e os positivos, isto é, calcula a proporção de identificações positivas que foram realmente corretas (Equação 42).

$$\text{Precisão} = \frac{VP}{VP+FP} \quad (42)$$

- 3) Especificidade – determina a capacidade do modelo de estimar corretamente os verdadeiros negativos (Equação 43).

$$\text{Especificidade} = \frac{VN}{VN+FP} \quad (43)$$

- 4) Acurácia – determina a proporção de vazões que foram identificadas dentro de cada classe com precisão, ou seja, determina o quanto o modelo acertou das previsões possíveis (Equação 44).

$$\text{Acurácia} = \frac{VP+VN}{VP+VN+FP+FN} = \frac{\text{Predições corretas}}{\text{Todas as predições}} \quad (44)$$

4.4.3.3. Curva ROC

A análise da curva ROC (Receiver Operating Characteristic) é um método gráfico, gerado a partir da matriz de confusão, que permite a avaliação, organização e seleção de sistemas de predição e/ou diagnóstico. A curva ROC foi inicialmente utilizado para avaliar a quantidade de transmissão de um sinal em um canal com ruído (Egan, 1975), assim como, também foi utilizado para avaliar a quantidade de predição de eventos climatológicos raros (Mylne, 2002).

A Figura 4 foi obtida a partir da plotagem no eixo das ordenadas dos valores para uma determinada classe de número de classificações corretas (sensibilidade ou recall), e no eixo das abscissas a taxa de falsos positivos (TFP) (Equação 45).

$$TFP = 1 - \text{Especificidade} \quad (45)$$

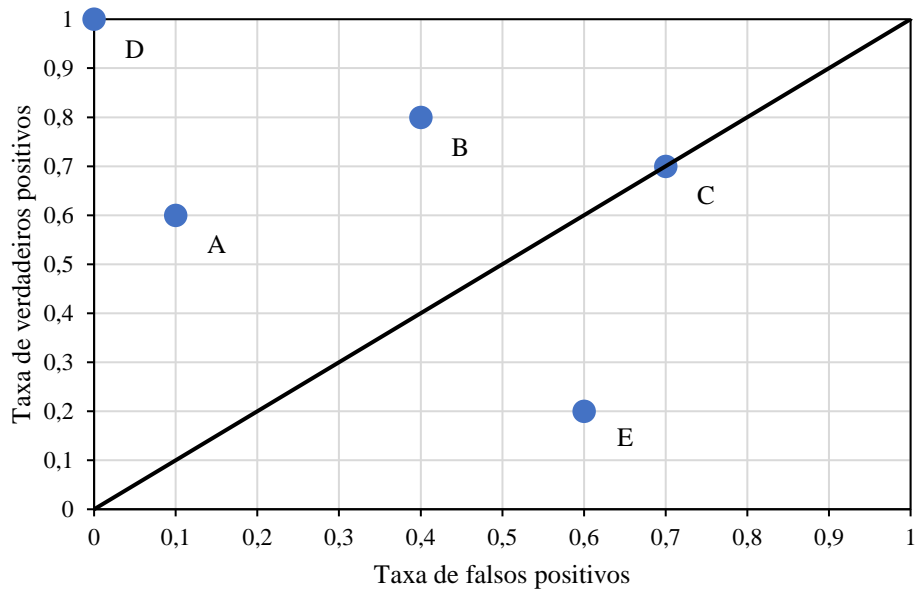


Figura 4 – Exemplo de curva ROC com cinco classificadores discretos, em que A é dito um classificador "conservador", B é o inverso de E, D é um classificador perfeito e C é dito aleatório. Fonte: Adaptado de Silva (2006).

O ponto no canto inferior esquerdo (0,0) representa um modelo que não gera classificação positiva, ou seja, não comete erro de falso positivo, mas também não classifica nenhum verdadeiro positivo. O modelo oposto, que gera somente classificação positiva (verdadeiros e falsos) é representado pelo ponto no canto superior direito (1,1).

O ponto D (0,1) representa um modelo perfeito. De maneira geral, um ponto no espaço ROC, é melhor que outro se ele está mais à noroeste. Modelos com a curva no lado esquerdo do gráfico ROC (próximos ao eixo Y) são ditos “conservadores”, já que estes fazem classificações positivas somente com uma evidência forte, sendo assim eles cometem poucos erros de falso positivo. Modelos com curvas no lado direito do gráfico, são ditos “liberais”, pois fazem classificações positivas com pouca evidência, mas cometem muito erros de falso positivo. Na Figura 4, vemos que o ponto A é mais conservador que o ponto B.

Se um modelo apresenta a curva abaixo da linha diagonal que vai do ponto (0,0) ao ponto (1,1), significa que ele tem um desempenho pior que um modelo aleatório. Sendo assim, busca-se modelos que a curva de distribuição dos valores não apareça dentro do triângulo inferior do gráfico ROC (Silva, 2006).

4.4.3.4. Curva AUC

A análise AUC (Area Under the ROC Curve) é uma simplificação da curva ROC, sendo representada por uma reta que integra todos os limiares da ROC, além de calcular a “área sob a curva”, que é um indicativo do desempenho do modelo avaliado. O AUC é uma métrica que

não possui variação em escala, já que avalia a precisão das classificações ao invés dos valores absolutos gerados pelos modelos, permitindo medir a qualidade das previsões independentemente do limiar de classificação (Hand & Anagnostopoulos, 2013).

O valor do AUC varia de 0,0 a 1,0, sendo que quanto mais próximo de 1,0 melhor é o desempenho do modelo. Valores iguais a 0,5 indicam que o modelo de interesse possui o mesmo desempenho que um modelo aleatório e valores abaixo deste limiar indicam que o modelo avaliado possui desempenho pior que de um modelo aleatório.

5. RESULTADOS E DISCUSSÃO

5.1. Delimitação das bacias de drenagem

Na Figura 5 são apresentadas as áreas de drenagem delimitadas a partir da localização das 83 estações fluviométricas definidas inicialmente. Afim de garantir a consistência das delimitações, os valores obtidos foram comparados com os disponíveis na plataforma *Hidroweb* (ANA, 2021). Os resultados indicaram que a delimitação de forma manual não apresentou diferença significativa em relação aos valores oficiais, sendo explicado pelas grandes dimensões das bacias, que “absorveram” os possíveis erros. Localizada sobre o rio Araguaia, na altura do município de Araguatins – TO, a estação 28850000 apresentou a maior área de drenagem com 382.077,13 km²; a menor área de drenagem registrada foi a da estação 34230000, com 801,65 km², estando localizada no município de Monte Alegre do Piauí - PI, sendo responsável pelo registro da vazão do riacho do Contrato.

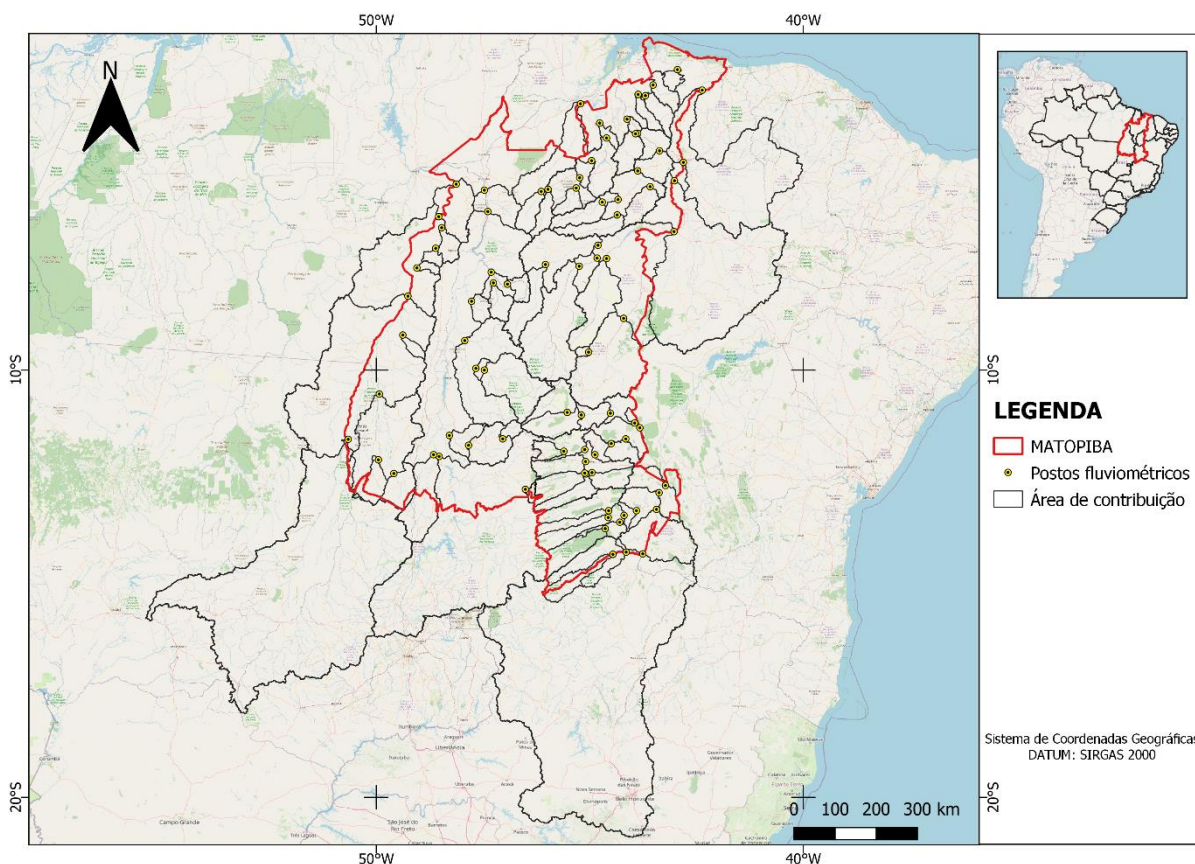


Figura 5 - Postos fluviométricos na região do MATOPIBA e suas áreas de contribuição.

Fonte: Elaboração a partir de dados provenientes da plataforma *Hidroweb* (ANA, 2021) e IBGE (2021).

5.2. Cálculo da precipitação média anual para as bacias de drenagem

Por meio do mapa variografico (Figura 6A), é possível observar que os dados de precipitação total anual possuem uma dependência espacial (anisotropia), ou seja, a sua variação é maior no sentido noroeste-sudeste. A fim de remover a tendência do modelo, foi ajustada uma nova fórmula em que os dados de latitude e longitude são considerados; o resultado é apresentado na Figura 6B.

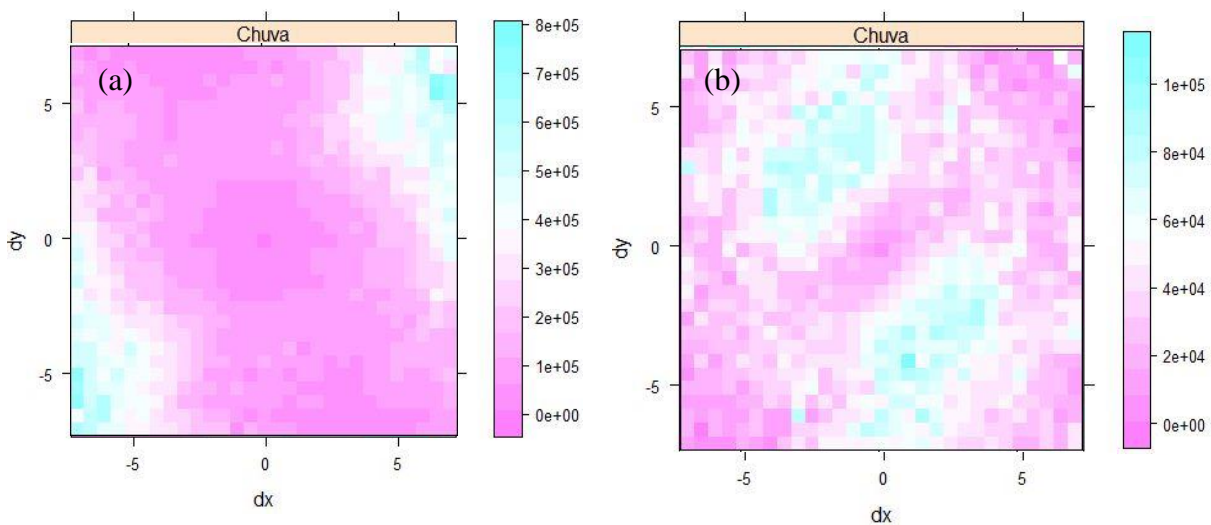


Figura 6: Mapa variografico; (a) modelo com anisotropia; (b) modelo após a remoção da dependência espacial.

Após a remoção da dependência espacial dos dados de precipitação, foi realizado o ajuste do semivariograma aos diferentes modelos (esférico, exponencial e Matérn com $k=0,4$), sendo que para cada um deles foi feito um arbítrio inicial e, após o ajuste pelo método da máxima verossimilhança, foram obtidos os valores do efeito pepita (900), da contribuição (50000) e do alcance (4). Os resultados são apresentados na Figura 7.

Para a escolha do melhor modelo ajustado foi utilizado o critério de validação cruzada (Faraco, 2008). O modelo escolhido foi aquele que apresentou o menor valor de RMSE e o maior valor de R^2 . Na figura 8 é possível visualizar o resultado da validação cruzada, em que no eixo das ordenadas são plotados os valores de precipitação estimados e no eixo das abscissas os valores calculados. Nesta mesma figura, a linha laranja representa a linha 1:1 e a linha verde representa a linha de correlação entre os dois dados.

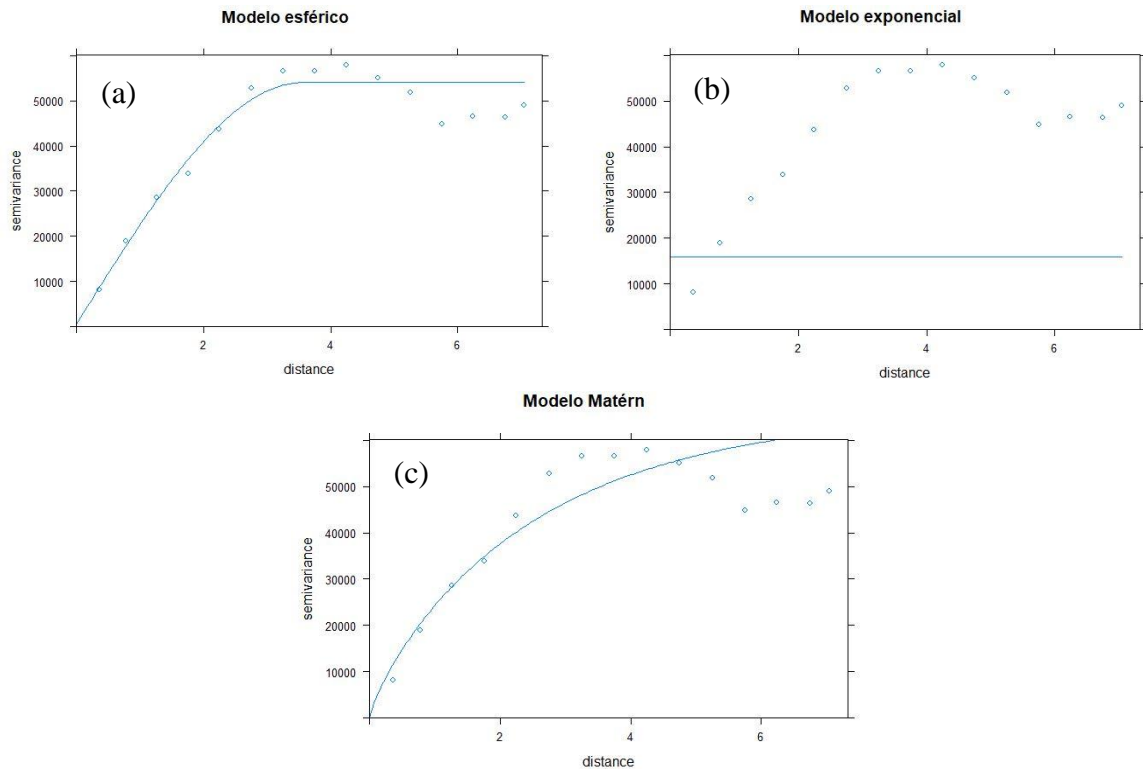


Figura 7: Gráfico dos semivariogramas dos modelos; (a) modelo esférico; (b) modelo exponencial e (c) modelo Matérn.

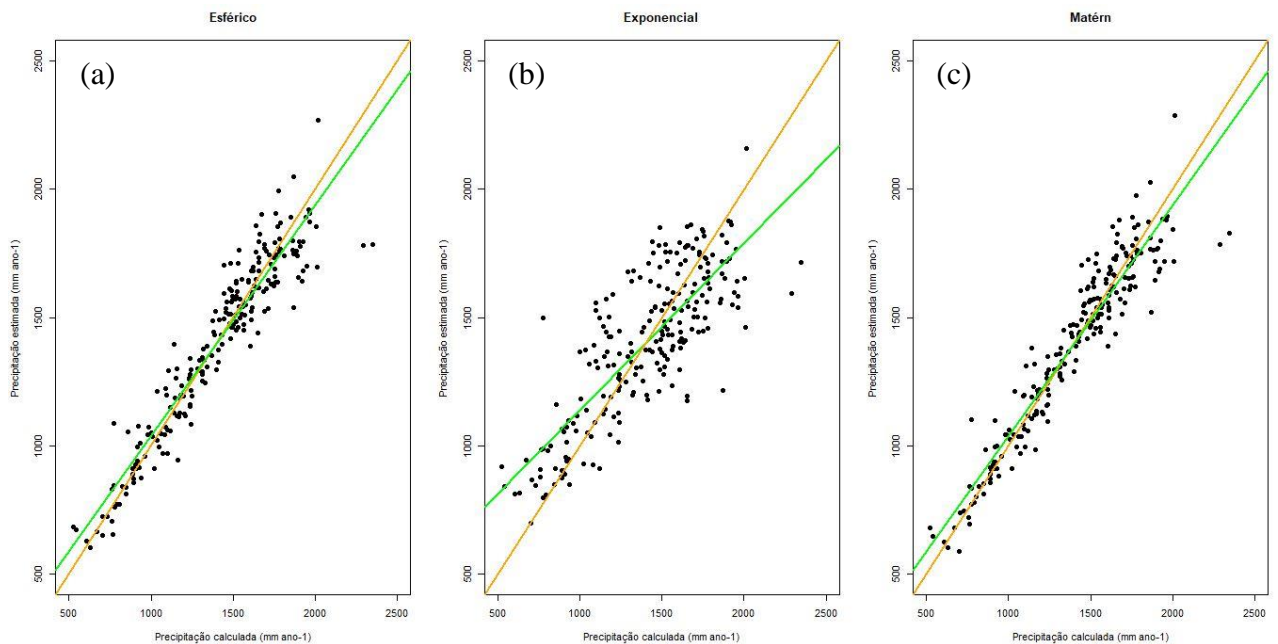


Figura 8: Resultado da validação cruzada para os diferentes modelos. (a) modelo esférico, (b) modelo exponencial e (c) modelo Matérn.

Fonte: Elaboração a partir de dados provenientes da plataforma Hidroweb (ANA, 2021).

Na Tabela 2 é possível visualizar os valores do RMSE e do R^2 para cada um dos modelos, o que permite concluir que o modelo Matérn foi o que melhor se ajustou ao semivariograma experimental pelo método da máxima verossimilhança.

Tabela 2: Resultado do teste de validação cruzada Leave-One-Out para os três modelos.

Modelo	R²	RMSE
Exponencial	0,647	213,145
Esférico	0,898	114,807
Matérn	0,902	112,259

Fonte: Elaboração própria.

A partir destes valores para os parâmetros (efeito pepita, contribuição e alcance) e do modelo a ser utilizado, foi possível realizar a interpolação dos dados de precipitação pelo método da krigagem ordinária (Figura 9). Em seguida foram extraídas as linhas de isoietas com distâncias entre elas de 50 mm (Figura 10).

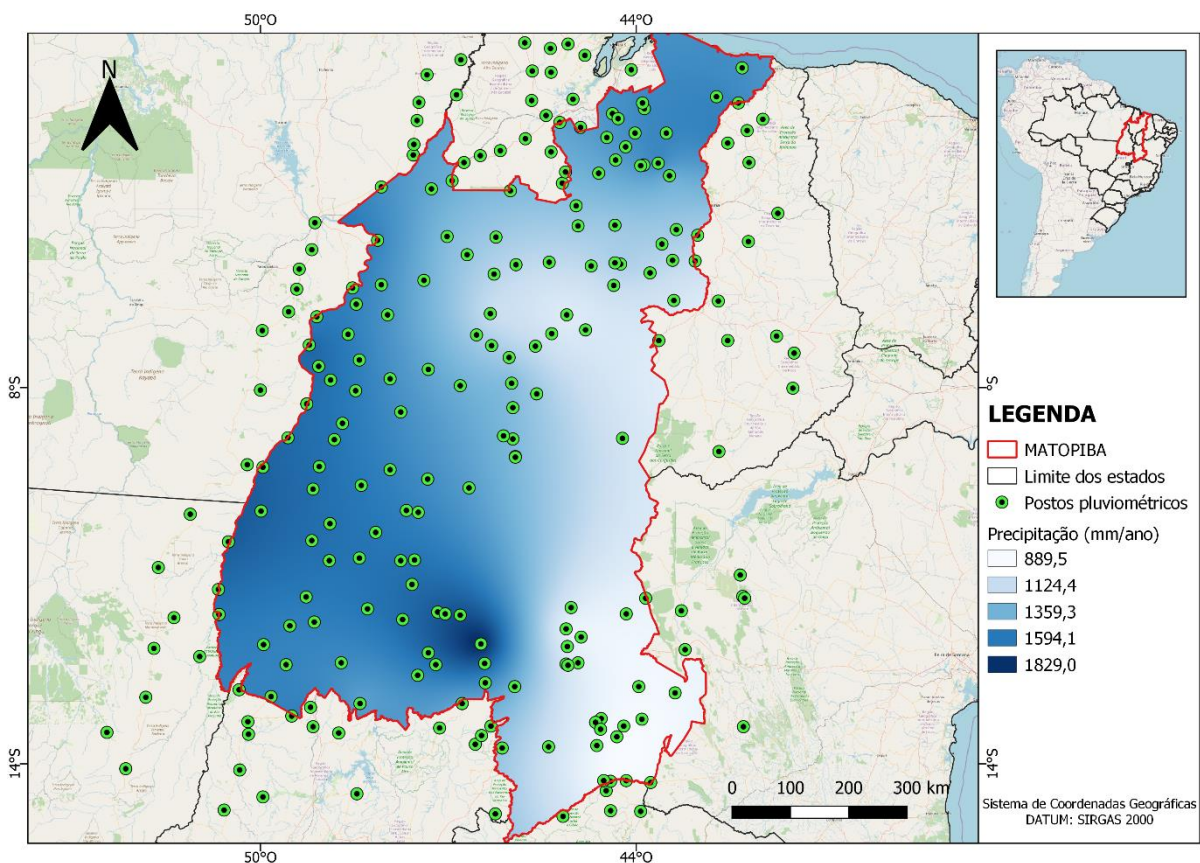


Figura 9 - Interpolação dos dados de precipitação pelo método da krigagem ordinária.

Fonte: Elaboração a partir de dados provenientes da plataforma Hidroweb (ANA, 2021) e IBGE, 2021.

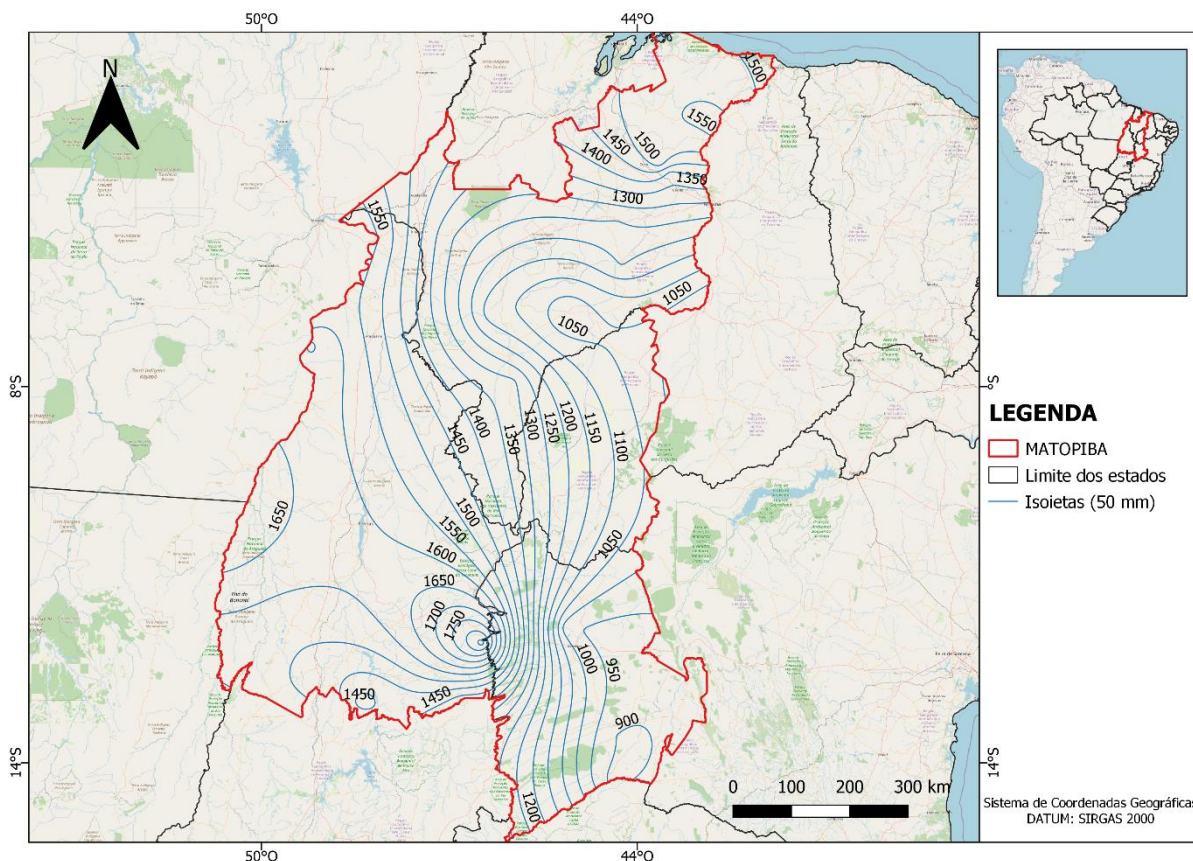


Figura 10: Distribuição das linhas isoietas com distância de 50 mm entre as linhas para a região do MATOPIBA. Fonte: Elaboração a partir de dados provenientes da plataforma Hidroweb (ANA, 2021) e IBGE, 2021.

Tendo como base os valores das isoietas geradas, foi utilizada a Equação 27 para a obtenção dos valores de precipitação total anual para cada área de contribuição, tendo como exutório a estação fluviométrica. Um extrato dos resultados é apresentado na Tabela 3. Os valores estimados para todas as áreas de contribuição são apresentados no Anexo C deste documento.

Tabela 3 - Extrato dos valores estimados de precipitação média anual para a área de contribuição dos postos fluviométricos.

Posto	Município	Estado	Rio	Precipitação (mm)
33025000	Grajau	Maranhão	Mearim	1356,84
26710000	Sandolândia	Tocantins	Javaés	1475,68
34060000	Ribeiro Gonçalves	Piauí	Parnaíba	1377,92
45131000	Montalvânia	Bahia	Carinhanha	1154,48

Fonte: Elaboração própria.

A precipitação total anual estimada para as áreas de contribuição em que os exutórios se encontram dentro do estado do Maranhão, variaram entre 1324 mm e 1476 mm. Estes

valores estão dentro do intervalo encontrado em outros estudos, como o de Mataveli et al. (2017), que analisou a precipitação em área de Cerrado do Maranhão (64% da área do estado), tendo como base dados do sensor MODIS e do satélite TRMM para o período de 2002 a 2015; os resultados indicaram uma precipitação total anual variando entre 890 mm a 1829 mm. O estudo realizado por Silva et al. (2014), para o período de 2003 a 2012, constatou que a distribuição espacial da precipitação anual para o estado do Maranhão, tendo como base a série climatológica proveniente do ISLSCP (International Land Surface Climatology Project), segue um gradiente no sentido sudeste-noroeste, com precipitações anuais próximas a 1000 mm no sudeste do estado, chegando aos 2500 mm na região noroeste.

As áreas de contribuição com exutório dentro do estado do Tocantins, apresentaram precipitação total anual estimada variando de 1349 mm a 1538 mm; os valores se mostraram coerentes com os encontrados em outros trabalhos realizados no estado. Segundo estudo realizado por Neto et al. (2021), em que foram analisados dados de 93 estações pluviométricas, para o período entre 1988 e 2019, a precipitação total anual variou de 933 a 1300 mm na porção sul do estado e entre 1760 e 1980 mm na porção centro-oeste. Valores próximos foram encontrados por Marcuzzo & Goularte (2013), em cujo trabalho foram analisados dados de 43 estações pluviométricas distribuídas no estado, com dados de 1977 a 2006. Os resultados apresentaram uma precipitação total anual variando de 1300 mm nas regiões sul e norte do estado, chegando aos 2000 mm nas porções centro-oeste e extremo sudeste.

Para as áreas de contribuição com exutório no estado do Piauí, a precipitação total anual estimada variou entre 1324 e 1378 mm. Estes valores são corroborados por estudos realizados no estado, como o de Medeiros (2017), em que foram analisados dois períodos distintos (1912 a 2011 e 1962 a 2011) para 15 municípios, tendo como resultado uma precipitação total anual variando entre 700 a 1300 mm na região sul, entre 500 e 1450 mm na região central e entre 800 e 1680 mm no norte do estado. Já no estudo de Guedes et al. (2012), foram analisados os dados de 33 estações pluviométricas para o período entre 1963 e 2000, tendo como resultado uma precipitação total anual variando de 500 mm na região sudeste a 1700 mm na região norte do estado.

Já a precipitação total anual estimada para as áreas de contribuição em que o exutório se encontram dentro do estado da Bahia, variaram entre 1158 mm e 1358 mm. Estes valores estão dentro do intervalo encontrado em outros estudos, como o de Dourado et al. (2013), que analisou os dados de 92 estações pluviométricas para o período compreendido entre 1981 e 2020, e teve como resultado uma precipitação total anual variando de 438 mm na região norte

do estado a 2033 mm na zona costeira Atlântica. Em estudo realizado por Braga et al. (1998), foram analisados os dados de 140 estações pluviométricas, com no mínimo 30 (trinta) anos de observações, tendo como resultado valores de precipitação total anual no semiárido do estado (norte) inferiores a 600 mm e na faixa litorânea ultrapassando os 1800 mm.

5.3. Vazões de referência Q_{med} , Q_{90} , Q_{95} , $Q_{7,10}$ específicas obtidas através das séries históricas das estações fluviométricas

5.3.1. Vazões Q_{med} , Q_{90} e Q_{95} específicas

As vazões de permanência Q_{90} e Q_{95} correspondem aos valores limites em que 90% ou 95% respectivamente dos registros diários de vazão são superiores a elas, ou seja, somente 10% ou 5% das vazões registradas naquele ponto serão inferiores aos da Q_{90} ou Q_{95} . Já a vazão média de longo período (Q_{med}) é definida através do cálculo da média aritmética das vazões diárias da série histórica.

A Tabela 4 apresenta um extrato dos resultados das vazões de referência $Q_{med,esp}$, $Q_{90,esp}$ e $Q_{95,esp}$ calculadas utilizando as Equações 23, 24 e 25 e os dados de vazão obtidos pelas séries históricas para cada estação fluviométrica. Os dados da Q_{med} , Q_{90} e Q_{95} para todos os postos utilizados são apresentados no Anexo D deste documento.

Tabela 4 - Extrato dos dados das estações fluviométricas utilizadas para o cálculo das vazões de referência e seus respectivos valores estimados pelas séries históricas.

Posto	Rio	Área	Lat.	Long.	$Q_{med,esp}$	$Q_{90,esp}$	$Q_{95,esp}$
		(km^2)	($^\circ$)	($^\circ$)	($L s^{-1} km^{-2}$)	($L s^{-1} km^{-2}$)	($L s^{-1} km^{-2}$)
33025000	Mearim	3728,5	-5,77	-45,98	6,48	0,87	0,61
26710000	Javaés	5586,9	-12,10	-49,59	30,83	0,44	0,19
34060000	Parnaíba	31345,6	-7,57	-45,25	6,73	4,15	3,93
45131000	Carinhanha	5995,8	-14,31	-44,46	10,08	5,44	4,81

Fonte: Elaboração própria.

Na Figura 11 são apresentados o histograma e a curva de permanência obtidos a partir dos dados observados para a estação fluviométrica 45298000, localizada sobre o rio São Francisco, na altura do município de Carinhanha no estado da Bahia.

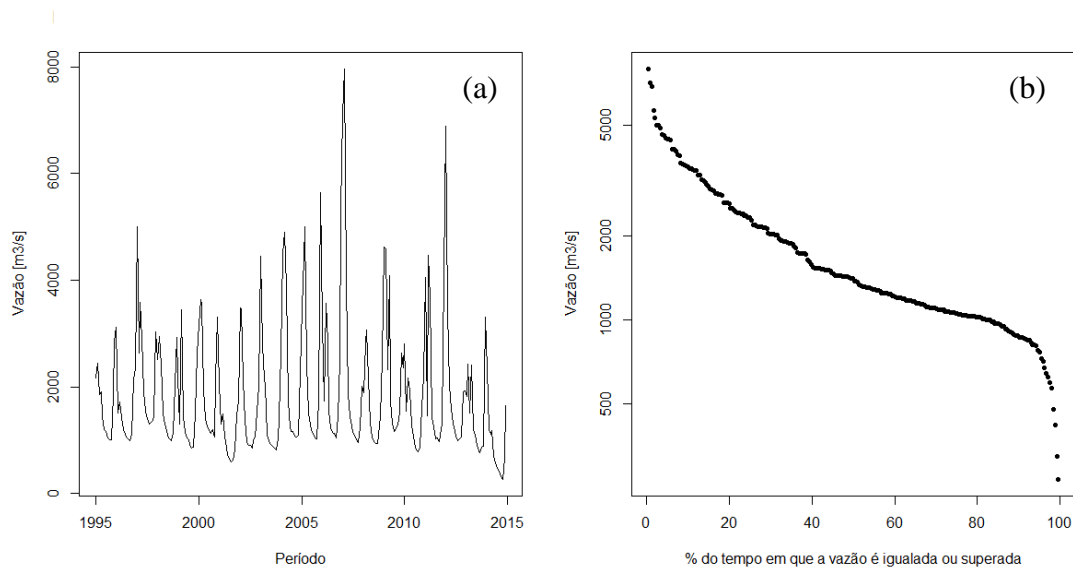


Figura 11 – (a) Fluxograma das vazões médias mensais do posto 45298000; (b) Curva de permanência de vazão para o posto 45298000.

Fonte: Elaboração a partir de dados provenientes da plataforma Hidroweb (ANA, 2021).

As maiores vazões médias foram observadas na porção norte do estado do Tocantins, como a vazão de $5891,95 \text{ m}^3 \text{ s}^{-1}$, registrada pela estação 28850000, localizada no município de Araguaia – TO sobre o rio Araguaia destaca-se também a vazão de $4208,14 \text{ m}^3 \text{ s}^{-1}$, registrada pela estação 23700000, localizada no município de Itaguatins – TO sobre o rio Tocantins. Já as maiores vazões médias específicas foram observadas no sul do estado do Tocantins, como a vazão de $30,83 \text{ L s}^{-1} \text{ km}^{-2}$, registrada pela estação 26710000, localizada no município de Sandolândia – TO sobre o rio Javaés.

As menores vazões médias foram observadas na região do semiárido brasileiro, entre os estados da Bahia e do Piauí, como a vazão de $1,88 \text{ m}^3 \text{ s}^{-1}$ registrada pela estação 34230000, localizada no município de Monte Alegre do Piauí – PI sobre o rio Contrato e a vazão de $5,00 \text{ m}^3 \text{ s}^{-1}$ registrada pela estação 46490000, localizada no município de São Desidério – BA sobre o rio São Desidério. Já as menores vazões médias específicas foram observadas no sul do estado do Piauí e no Oeste do estado da Bahia, como a vazão de $0,57 \text{ L s}^{-1} \text{ km}^{-2}$, registrada pela estação 34251000, localizada no município de Cristiano Castro – PI sobre o rio Gurguéia e a vazão de $1,11 \text{ L s}^{-1} \text{ km}^{-2}$, registrada pela estação 46490000, localizada no município de São Desidério – BA sobre o rio São Desidério.

5.3.2. Vazão $Q_{7,10}$ específica

Dentre as distribuições estatísticas analisadas, visando a estimação da vazão $Q_{7,10}$ para os 83 postos fluviométricos na área de estudo, a distribuição de Log-Pearson tipo 3, foi a que melhor se ajustou aos dados de 84,4% (70) dos postos; os outros 15,6% (13) dos postos apresentaram a distribuição de Weibull como a mais adequada para a estimativa da vazão.

Na Figura 12 é apresentado um exemplo gráfico das diversas distribuições analisadas neste trabalho, para os dados de vazão mínima da estação 45298000, que se localiza sobre o Rio São Francisco, no municíio de Carinhanha no estado da Bahia.

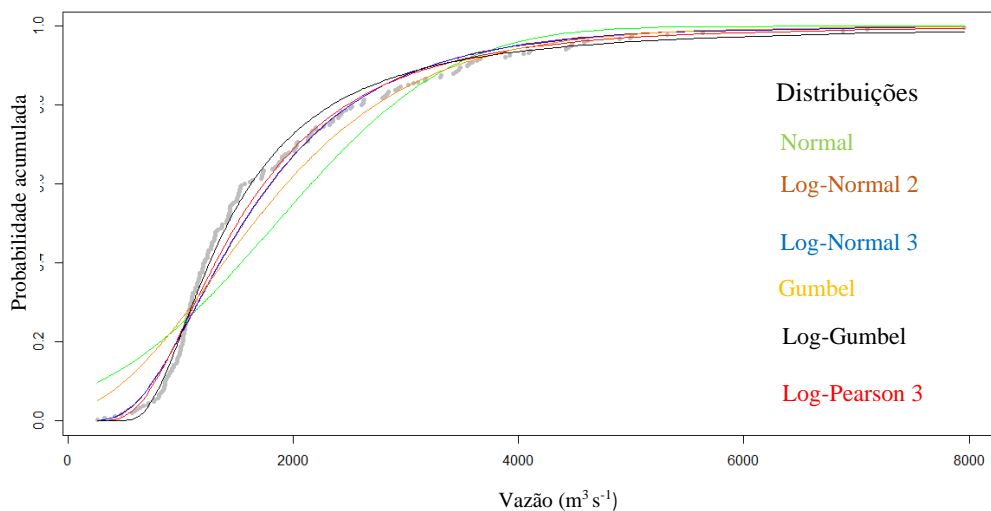


Figura 12 - Exemplo de curva de distribuição teórica dos dados e curvas das distribuições estatísticas para o posto 45298000.

Fonte: Elaboração a partir dos dados provenientes da plataforma Hidroweb (ANA, 2021).

Na Tabela 5 é apresentado um extrato dos valores obtidos para a vazão de referência $Q_{7,10}$ e os intervalos de confiança superiores e inferiores com 95% de confiança, a amplitude máxima entre esses intervalos, além da distribuição que melhor se ajustou aos dados. Os valores para todos os postos fluviométricos analisados são apresentados no Anexo E deste documento.

Tabela 5 - Exemplos de extrato das distribuições estatísticas que melhores se ajustaram para o cálculo da vazão mínima com duração de sete dias e período de retorno de 10 anos, valores de $Q_{7,10}$ ($m^3 s^{-1}$), intervalos de confiança e amplitude do intervalo.

Posto	$Q_{7,10}$ ($m^3 s^{-1}$)	Intervalos de confiança (95%)		Amplitude	Distribuição
		Superior	Inferior		
33025000	1,89	2,42	1,36	1,061	Log-Pearson 3
26710000	0,37	0,54	0,21	0,321	Log-Pearson 3
34060000	111,04	116,19	105,89	10,300	Weibull
45131000	27,50	30,05	24,95	5,098	Log-Pearson 3

Fonte: Elaboração própria.

Na Tabela 6 são apresentados os valores da $Q_{7,10}$ e da $Q_{7,10}$ específica, calculada pela Equação 23 para os mesmos postos listados na Tabela 5.

Tabela 6 - Exemplos de extrato das vazões $Q_{7,10}$ e $Q_{7,10}$ específica de alguns postos fluviométricos na região do MATOPIBA.

Posto	$Q_{7,10}$ ($m^3 s^{-1}$)	$Q_{7,10esp}$ ($L s^{-1} km^{-2}$)
33025000	1,89	0,51
26710000	0,37	0,07
34060000	111,04	3,54
45131000	27,50	4,59

Fonte: Elaboração própria.

Fikler et al. (2015) compararam 6 (seis) diferentes funções de distribuição de probabilidade para a estimação das vazões mínimas na bacia hidrográfica do Arroio Belo, localizada no município de Caxias do Sul – RS. Os resultados obtidos mostram que a distribuição de Log-Pearson III alcançou o melhor ajuste dos valores.

A pesquisa conduzida por Aiyelokun et al. (2022) testou 5 (cinco) distribuições de probabilidade para a estimação das vazões mínimas para o rio Ogun, no sudoeste da Nigéria, e concluíram que o modelo Logístico Generalizado (GLO) foi o que melhor se ajustou para a estimativa das vazões.

Silva et al. (2006) analisaram os modelos de probabilidade Log-Normal 3, Weibull e Gumbel, visando definir a melhor distribuição para a estimativa da vazão mínima de referência para outorga na região do Alto Rio Grande no estado de Minas Gerais. Os resultados obtidos apontaram a distribuição Log-Normal 3 como a que melhor representou os dados.

Em análise feita por Pinto et al. (2010), compararam-se as distribuições de probabilidade de Gumbel, Weibull e Log-Normal 3 para a estimação da vazão $Q_{7,10}$ para o rio Itabapoana no estado do Espírito Santo. Os resultados apontaram a distribuição de Weibull como a que obteve o melhor ajuste na estimativa.

Oliveira & Gonçalves (2022) analisaram as distribuições Normal, Gumbel para máximos, Gumbel para mínimos e Log-Normal para as regiões hidrográficas dos rios Paraná e São Francisco, visando determinar o tamanho ideal de séries históricas de vazões mínimas anuais para essas regiões. O estudo mostra que não é possível utilizar somente um modelo para a estimação, sendo que para cada estação fluviométrica houve uma distribuição que melhor se ajustou aos dados.

Os diferentes estudos citados anteriormente, mostram que para cada caso haverá uma distribuição de probabilidade que melhor se ajuste aos dados medidos. Com isso fica evidente a necessidade da análise prévia dos dados, já que a generalização de um modelo ou regra pode levar a um aumento dos erros associados ao cálculo da vazão mínima de referência.

5.4. Análise de agrupamento e determinação das regiões hidrologicamente homogêneas

A partir da análise de agrupamento ou análise de cluster, foi possível fazer a delimitação das regiões hidrologicamente homogêneas. Para a análise foram consideradas como variáveis independentes: as coordenadas UTM (Universal Transversa de Mercator) da latitude e longitude do centroide da bacia, a declividade média do talvegue principal, a densidade de drenagem, o comprimento do talvegue principal, o coeficiente de compacidade (K_c) e a precipitação total.

No método hierárquico o número ideal de clusters é dado pelo agrupamento com maior distância, sendo assim, a quantidade seria de 4 (quatro) grupos, porém, por se tratar de uma região extensa e que cobre 3 (três) biomas brasileiros, optou-se por “descer” um nível hierárquico, permitindo assim, agrupamentos de postos com características mais próximas. Com isso, foram definidas 7 (sete) regiões hidrologicamente homogêneas para o MATOPIBA. A Figura 13 apresenta o dendrograma dos postos e os clusters e os postos fluviométricos com os compõem, já a Figura 14 apresentam a distribuição dos clusters dentro da área de estudo.

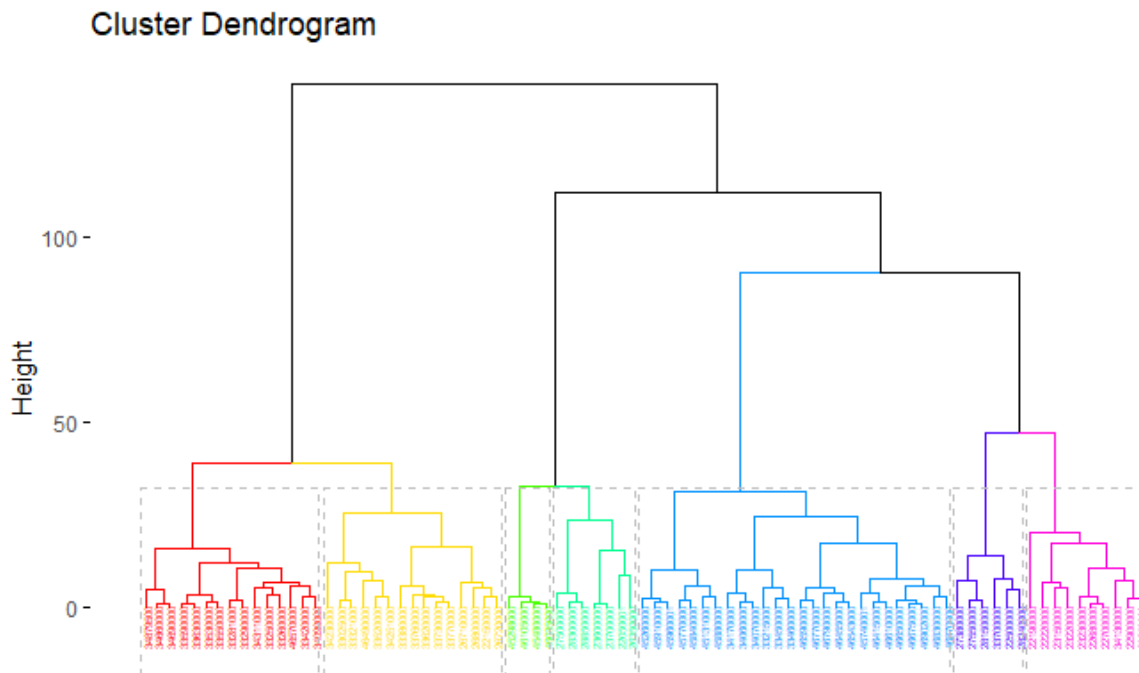


Figura 13 – Delimitação dos 7 clusters pelo método hierárquico através do algoritmo de Ward.
 Fonte: Elaboração a partir de dados provenientes da plataforma Hidroweb (ANA, 2021).

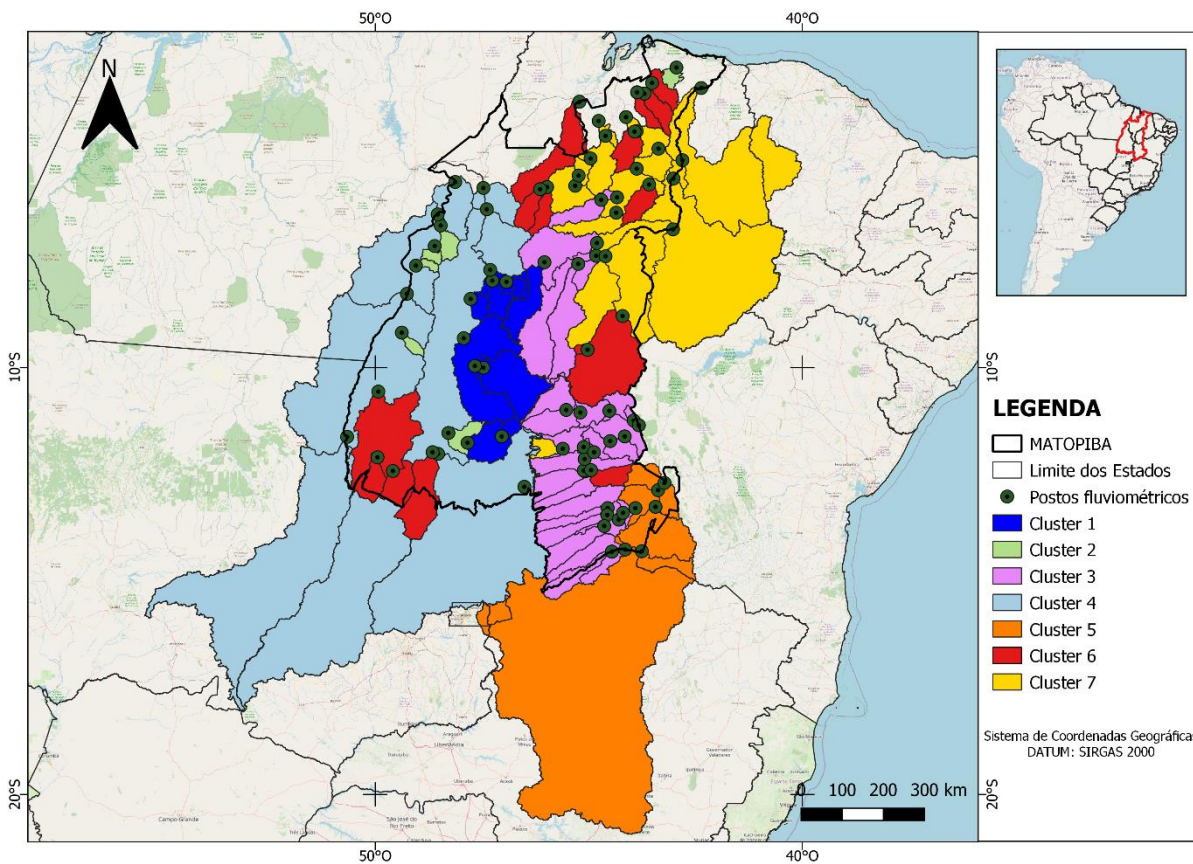


Figura 14 - Regiões hidrologicamente homogêneas para a região do MATOPIBA segundo a análise de agrupamento (clusters).
 Fonte: Elaboração a partir de dados provenientes da plataforma Hidroweb (ANA, 2021) e IBGE (2021).

A Figura 15 apresenta o número de postos fluviométricos dentro de cada região homogênea delimitada.

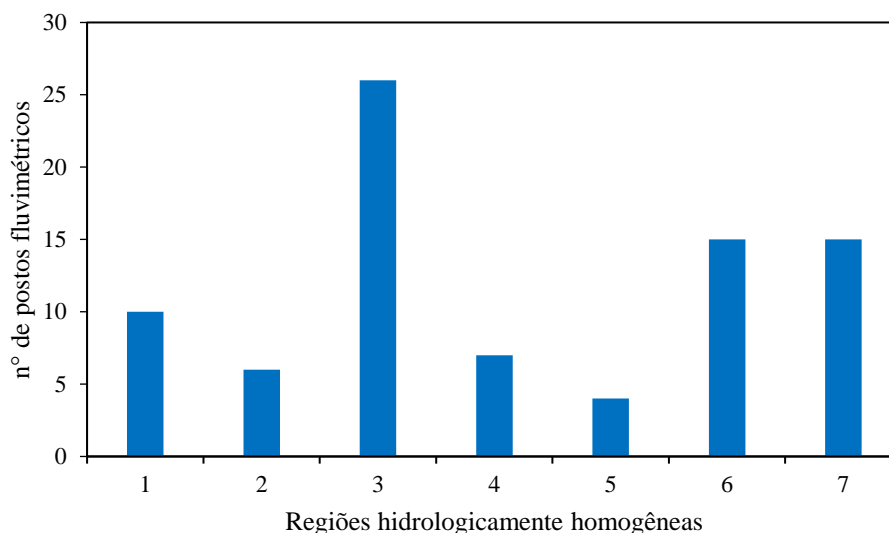


Figura 15 - Número de postos fluviométricos por região hidrológicamente homogênea.
Fonte: Elaboração própria.

As regiões hidrológicamente homogêneas delimitadas neste trabalho podem se tornar importantes ferramentas na gestão dos recursos hídricos para a região, visto que na literatura não foi encontrado nenhum trabalho que aborda a regionalização da vazão para todo o MATOPIBA.

Considerando a extensão territorial da área de estudo, o número de regiões hidrológicamente homogêneas é relativamente baixo. Segundo os dados do CPRM (2008), 20% da área do MATOPIBA é composta por formas de relevo de colinas amplas e suaves, com declividade variando entre 3 e 10° e 15% é composta por áreas de chapadas e platôs com variação de declividade entre 0 e 5°. Apesar desta baixa variabilidade, o comprimento e a declividade média do talvegue foram os parâmetros que se destacaram dentro da classificação das regiões.

5.5. Modelos de regionalização hidrológica

Por meio da regressão linear múltipla, foi possível a geração dos modelos para a estimação das vazões de referência para as regiões hidrológicamente homogêneas. O cálculo das regressões foi realizado para cada uma das sete regiões delimitadas, considerando a correlação entre diferentes combinações das variáveis independentes (Dd, L, Is, Kc, P, Cx,

Cy) e as variáveis dependentes ($Q_{med_{esp}}$, $Q_{90_{esp}}$, $Q_{95_{esp}}$ e $Q_{7,10_{esp}}$). Em seguida, a seleção dos melhores modelos foi definida através do teste t-student ($p < 0,05$). Com isso, foram geradas quatro equações de regionalização para cada uma das sete regiões hidrologicamente homogêneas, os índices das equações para cada vazão de referência são apresentados nas Tabelas 7, 8, 9 e 10.

Tabela 7 - Parâmetros das equações de regressão para a estimação da vazão média específica regionalizada ($Q_{med_{esp}}$) para cada região hidrologicamente homogênea.

Cluster	Intercepto	a	b	c	d	e	f	g
1	-106,79	-26,69	$2,35 \cdot 10^{-03}$	4,37	0,00**	0,03	$1,25 \cdot 10^{-05}$	$1,21 \cdot 10^{-05}$
2	166,60	-67,06	0,00*	-1,93	0,00*	0,00*	$-1,11 \cdot 10^{-05}$	$-3,74 \cdot 10^{-06}$
3	58,67	-4,24	$-1,84 \cdot 10^{-03}$	-1,69	-0,70	-0,02	$-3,00 \cdot 10^{-05}$	0,00**
4	8,85	0,85	$1,59 \cdot 10^{-03}$	0,00**	0,00*	0,00**	$-3,15 \cdot 10^{-07}$	0,00*
5	8,45	2,62	$-4,30 \cdot 10^{-03}$	0,00*	0,00*	0,00*	0,00*	0,00*
6	13,82	-9,48	$-2,22 \cdot 10^{-03}$	-0,59	4,17	0,03	$-1,61 \cdot 10^{-06}$	$-4,67 \cdot 10^{-06}$
7	-39,89	-0,71	$-1,83 \cdot 10^{-05}$	-0,06	1,46	0,06	0,00**	$-3,57 \cdot 10^{-06}$

a) índice da variável Dd; b) índice da variável L; c) índice da variável Is; d) índice da variável Kc; e) índice da variável P; f) índice da variável Cx e g) índice da variável Cy.

* índices excluídos através do coeficiente de correlação de Pearson

** índices excluídos através do teste de multicolinearidade VIF (Variance Inflation Factor)

Fonte: Elaboração própria.

Os modelos obtidos para a estimativa da vazão média específica regionalizada ($Q_{med_{esp}}$) apresentaram padrões diferentes entre as regiões hidrologicamente homogêneas. Dentre as variáveis iniciais, somente a densidade de drenagem (Dd) e o comprimento do talvegue (L) estiveram presentes em ao menos 86% dos modelos. A simplificação das equações se deu pela eliminação de variáveis correlacionadas através do Fator de Inflação da Variância (VIF), ou variáveis estatisticamente não significativas ($p < 0,05$).

Rodrigues et al. (2021) realizaram um estudo de regionalização hidrológica para a bacia do rio Tocantins no bioma Cerrado. Dentre as variáveis avaliadas, somente a área de drenagem das bacias foi significativa para a estimação da vazão média de longo período.

Tabela 8 - Parâmetros das equações de regressão para a estimação da vazão $Q90_{\text{esp}}$ regionalizada para cada região hidrologicamente homogênea.

Clusters	Intercepto	a	b	c	d	e	f	g
1	-75,73	-20,71	$8,79.10^{-03}$	4,27	0,00**	-0,03	$1,62.10^{-05}$	$1,55.10^{-05}$
2	-25,49	-0,57	0,00*	-0,38	0,00*	0,00*	$-2,70.10^{-05}$	$1,15.10^{-06}$
3	30,65	-2,16	$-3,91.10^{-03}$	-1,25	-1,01	-0,01	$-1,40.10^{-05}$	0,00**
4	9,64	-1,00	$-8,49.10^{-04}$	0,00*	0,00**	0,00**	$-2,89.10^{-06}$	0,00*
5	6,74	-1,53	$-2,47.10^{-04}$	0,00*	0,00*	0,00*	0,00*	0,00*
6	2,17	$-1,08.10^{-03}$	$-6,98.10^{-04}$	0,01	0,50	$-3,67.10^{-04}$	$-1,45.10^{-06}$	$-2,36.10^{-07}$
7	-10,55	0,42	$8,56.10^{-04}$	0,90	-1,57	0,02	0,00**	$-1,18.10^{-06}$

a) índice da variável Dd; b) índice da variável L; c) índice da variável Is; d) índice da variável Kc; e) índice da variável P; f) índice da variável Cx e g) índice da variável Cy.

* índices excluídos através do coeficiente de correlação de Pearson

** índices excluídos através do teste de multicolinearidade VIF (Variance Inflation Factor)

Fonte: Elaboração própria.

Seguindo o mesmo padrão dos modelos para a vazão média específica regionalizada, as equações para a estimativa das vazões específicas regionalizada com 90% de permanência ($Q90_{\text{espr}}$), apresentaram como principais variáveis a densidade de drenagem (Dd) e o comprimento do talvegue (L), estando presentes em ao menos 86% dos modelos.

Em estudo realizado por Beskow et al. (2016), em que foram empregadas técnicas de inteligência artificial para a regionalização hidrológica da vazão $Q90$ para a estado do Rio Grande do Sul, a área de drenagem das bacias foi a única variável necessária para a definição dos modelos para as diferentes regiões hidrologicamente homogêneas.

Nas equações para a estimativa das vazões específicas regionalizada com 95% de permanência ($Q95_{\text{espr}}$), somente duas variáveis independentes estiveram presentes em ao menos 86% dos modelos para a região hidrologicamente homogênea: densidade de drenagem (Dd) e comprimento do talvegue (L); as demais variáveis foram removidas por meio dos mesmos critérios utilizados para os modelos da estimativa da Q_{medespr} .

Tabela 9 - Parâmetros das equações de regressão para a estimação da vazão $Q95_{\text{esp}}$ regionalizada para cada região hidrologicamente homogênea.

Clusters	Intercepto	a	b	c	d	e	f	g
1	-57,65	-20,36	0,01	3,48	0,00**	-0,04	$1,28.10^{-05}$	$1,52.10^{-05}$
2	-18,95	0,90	0,00*	-0,46	0,00*	0,00*	$2,42.10^{-05}$	$3,58.10^{-07}$
3	25,28	-1,99	$-3,83.10^{-03}$	-1,21	-0,87	-0,01	$-1,19.10^{-05}$	0,00**
4	8,56	-0,80	$-8,26.10^{-04}$	0,00*	0,00**	0,00**	$-2,77.10^{-06}$	0,00*

5	5,79	-1,51	-1,91.10 ⁻⁰⁴	0,00*	0,00*	0,00*	0,00*	0,00*
6	2,00	-0,02	-5,24.10 ⁻⁰⁴	-0,05	0,30	-1,43.10 ⁻⁰⁴	-1,12.10 ⁻⁰⁶	-1,28.10 ⁻⁰⁷
7	-9,24	0,51	9,14.10 ⁻⁰⁴	0,83	-1,88	0,01	0,00**	-7,51.10 ⁻⁰⁷

a) índice da variável Dd; b) índice da variável L; c) índice da variável Is; d) índice da variável Kc; e) índice da variável P; f) índice da variável Cx e g) índice da variável Cy.

* índices excluídos através do coeficiente de correlação de Pearson

** índices excluídos através do teste de multicolinearidade VIF (Variance Inflation Factor)

Fonte: Elaboração própria.

Baena et al. (2004) realizaram um estudo visando a determinação de modelos para a regionalização das vazões para a bacia do Rio Paraíba do Sul, no trecho dentro do estado de São Paulo; foram utilizadas cinco variáveis morfológicas das bacias, além da precipitação total anual para a definição dos modelos. O estudo conclui que para a determinação da vazão Q_{95} , somente a área da bacia e a densidade de drenagem foram significativas.

Tabela 10 - Parâmetros das equações de regressão para a estimação da vazão $Q_{7,10_{esp}}$ regionalizada para cada região hidrologicamente homogênea.

Clusters	Intercepto	a	b	c	d	e	f	g
1	-42,95	-19,70	0,01	3,43	0,00**	-0,04	1,31.10 ⁻⁰⁵	1,40.10 ⁻⁰⁵
2	-27,29	4,33	0,00*	-0,32	0,00*	0,00*	2,15.10 ⁻⁰⁵	7,96.10 ⁻⁰⁷
3	25,52	-2,07	-4,36.10 ⁻⁰³	-1,25	-0,86	-0,01	-1,09.10 ⁻⁰⁵	0,00**
4	7,17	-0,57	-7,04.10 ⁻⁰⁴	0,00*	0,00**	0,00**	-2,78.10 ⁻⁰⁶	0,00*
5	3,44	-0,12	-6,19.10 ⁻⁰⁴	0,00*	0,00*	0,00*	0,00*	0,00*
6	2,27	0,02	-5,05.10 ⁻⁰⁴	-0,03	0,34	-6,33.10 ⁻⁰⁴	-9,40.10 ⁻⁰⁷	-1,23.10 ⁻⁰⁷
7	-10,06	0,56	9,88.10 ⁻⁰⁴	0,91	-2,06	0,01	0,00**	-3,58.10 ⁻⁰⁷

a) índice da variável Dd; b) índice da variável L; c) índice da variável Is; d) índice da variável Kc; e) índice da variável P; f) índice da variável Cx e g) índice da variável Cy.

* índices excluídos através do coeficiente de correlação de Pearson

** índices excluídos através do teste de multicolinearidade VIF (Variance Inflation Factor)

Fonte: Elaboração própria.

Para as equações para a estimativa das vazões específicas regionalizada mínima de 7 dias com período de retorno de 10 anos ($Q_{7,10_{espr}}$), o mesmo padrão encontrado anteriormente foi observado, em que as variáveis independentes: densidade de drenagem (Dd), comprimento do talvegue principal (L) estiveram presentes em ao menos 86% dos modelos para as regiões hidrologicamente homogêneas. As demais variáveis foram removidas por meio dos mesmos critérios utilizados para os modelos da estimativa da $Q_{med_{espr}}$.

Em estudo realizado por Silva et al. (2009), em que foram analisados cinco parâmetros morfológicos, além da precipitação total anual para a bacia, visando comparar três diferentes metodologias de regionalização de vazões mínimas de referência para a bacia do Rio São Francisco a montante do reservatório de Três Marias, conclui-se que as variáveis densidade de drenagem e precipitação total anual foram significativas para a determinação da vazão $Q_{7,10}$.

Observando os trabalhos citados para as diferentes vazões, fica evidente que não existe uma regra clara para a determinação das variáveis explicativas que irão compor um modelo para a estimação da vazão, devendo ser levado em consideração as características morfológicas da região de interesse, para a definição das variáveis que melhor se adaptam para a estimação.

Tendo definidos os modelos, a próxima etapa foi a de avaliação dos mesmos, na qual foram estimados os valores de vazão específica para cada posto fluviométrico dentro de cada região hidrologicamente homogênea e em seguida estes dados foram comparados com os valores de vazão específica calculados através dos dados provenientes das séries históricas de cada posto, sendo avaliados por diferentes índices de performance.

5.6. Avaliação dos modelos de regionalização

A validação dos modelos de regionalização foi realizada através de cinco índices de desempenho, para as vazões $Q_{med_{esp}}$, $Q_{90_{esp}}$, $Q_{95_{esp}}$ e $Q_{7,10_{esp}}$, em que são comparados os valores estimados de vazão específica regionalizada, com os valores de vazão específica calculados através dos dados provenientes das estações fluviométricas. Para facilitar a interpretação dos resultados, foi utilizado uma classificação qualitativa dos índices: a) coeficiente de concordância (d); b) coeficiente de eficiência de modelagem (NSE); c) erro percentual média (PBIAS) e d) coeficiente de determinação (R^2), sendo eles classificados como: Muito bom, Bom, Satisfatório ou Insatisfatório, conforme os limites quantitativos definidos na Tabela 1.

5.6.1. Vazão média específica regionalizada ($Q_{med_{espr}}$)

A Figura 16 apresenta um gráfico de dispersão dos dados de vazão média específica calculada e dos dados de vazão média específica regionalizada para as sete regiões

hidrologicamente homogêneas definidas anteriormente de forma conjunta. Os valores acima da reta 1:1 representam subestimação dos dados pelos modelos, já os valores abaixo da reta 1:1 representam superestimação da vazão pelos modelos. Nas tabelas 11 e 12 são apresentados os resultados dos critérios de avaliação e a classificação qualitativa dos mesmos respectivamente.

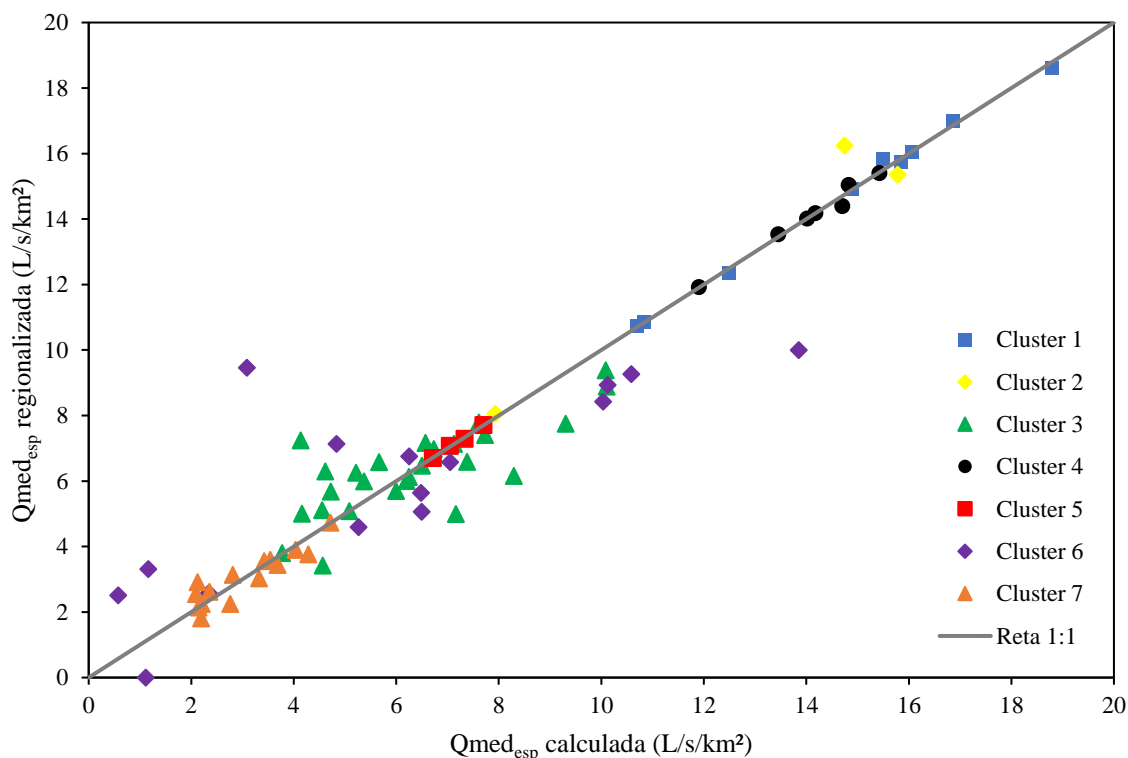


Figura 16 - Relação entre a $Q_{med_{esp}}$ calculada e a $Q_{med_{esp}}$ regionalizada para as 7 regiões hidrologicamente homogêneas no MATOPIBA.

Fonte: Elaboração própria.

Para a vazão média específica regionalizada ($Q_{med_{espr}}$), a maior parte dos modelos se ajustaram bem aos dados calculados, já que os valores estimados estão bem próximos a reta 1:1, a exceção foi o modelo para a região hidrologicamente homogênea 6, que apresentou uma maior dispersão dos dados, podendo ser explicado pela maior variância dos valores de vazão específica calculados, entre $0,57$ e $13,85 \text{ L s}^{-1} \text{ km}^{-2}$, forçando o modelo a se ajustar da melhor forma possível a esta variância.

Tabela 11 - Resultado dos critérios de avaliação de desempenho da $Q_{med_{espr}}$, segundo os índices estatísticos recomendados para modelos de regionalização hidrológica por Moriasi et al. (2015).

Região	d	NSE	PBIAS (%)	RSR	R ²
1	0,9996	0,9982	-0,118	0,0133	0,9982
2	0,9804	0,9258	0,205	0,1112	0,9259
3	0,8835	0,6548	0,029	0,1152	0,5846
4	0,9953	0,9814	-0,001	0,0516	0,9814
5	0,9984	0,9938	-0,071	0,0393	0,9940
6	0,8754	0,6486	1,103	0,1531	0,6355
7	0,9500	0,8260	-0,040	0,1077	0,8251

Fonte: Elaboração própria.

O RSR serviu como uma medida para avaliar a precisão do modelo na estimativa das vazões; quanto menor o seu valor, mais preciso foi o modelo. Todos os modelos utilizados para a estimação da $Q_{med_{espr}}$ apresentaram um bom ajuste, já que seus valores são inferiores a 0,5, ou seja, os valores de RMSE são menores que a metade do desvio padrão dos dados observados.

Tabela 12 - Classificação dos critérios de avaliação de desempenho da $Q_{med_{espr}}$, para os índices estatísticos recomendados para modelos de regionalização hidrológica por Moriasi et al. (2015).

Região	d	NSE	PBIAS (%)	R ²
1	Muito bom	Muito bom	Muito bom	Muito bom
2	Muito bom	Muito bom	Muito bom	Muito bom
3	Bom	Satisfatório	Muito bom	Insatisfatório
4	Muito bom	Muito bom	Muito bom	Muito bom
5	Muito bom	Muito bom	Muito bom	Muito bom
6	Bom	Satisfatório	Muito bom	Satisfatório
7	Muito bom	Muito bom	Muito bom	Bom

Fonte: Elaboração própria.

O índice de concordância (d) serviu como uma medida de precisão dos valores preditos em relação aos valores observados. As sete regiões apresentaram um bom desempenho, com resultados acima 0,87, ou seja, foram capazes de prever os valores calculados acima dos 87% dos casos.

O coeficiente de eficiência de modelagem (NSE) permite avaliar a capacidade preditiva dos modelos. Para a estimativa de $Q_{med_{espr}}$, o menor valor de NSE foi de 0,64 para a região hidrológicamente homogênea 6, ou seja, o modelo menos eficiente entre os sete analisados, sendo capaz de prever corretamente o valor da vazão em somente 64% dos postos fluviométricos dentro da área de estudo.

Todos os modelos apresentaram erro médio percentual (PBIAS) inferiores a 1%, sendo classificados como “Muito bom”. Entre as regiões, somente os modelos para as regiões

1, 4, 5 e 7 apresentaram um leve viés de subestimação dos valores; as demais regiões apresentaram um leve viés de superestimação das vazões.

Uma outra medida utilizada para a avaliação dos modelos foi o coeficiente de determinação (R^2). Para a $Q_{med_{espr}}$ o modelo para região hidrologicamente homogênea 3 que apresentou o menor valor ($R^2 = 0,58$) dentre os outros modelos, ou seja, este modelo foi capaz de explicar somente 58% da variabilidade do resultado final.

5.6.2. Vazão específica regionalizada com 90% de permanência ($Q90_{espr}$)

A Figura 17 apresenta um gráfico de dispersão dos dados de vazão média específica calculada e dos dados de vazão média específica regionalizada para as sete regiões hidrologicamente homogêneas definidas anteriormente. Nas Tabelas 13 e 14 são apresentados os resultados dos critérios de avaliação e a classificação qualitativa dos mesmos respectivamente.

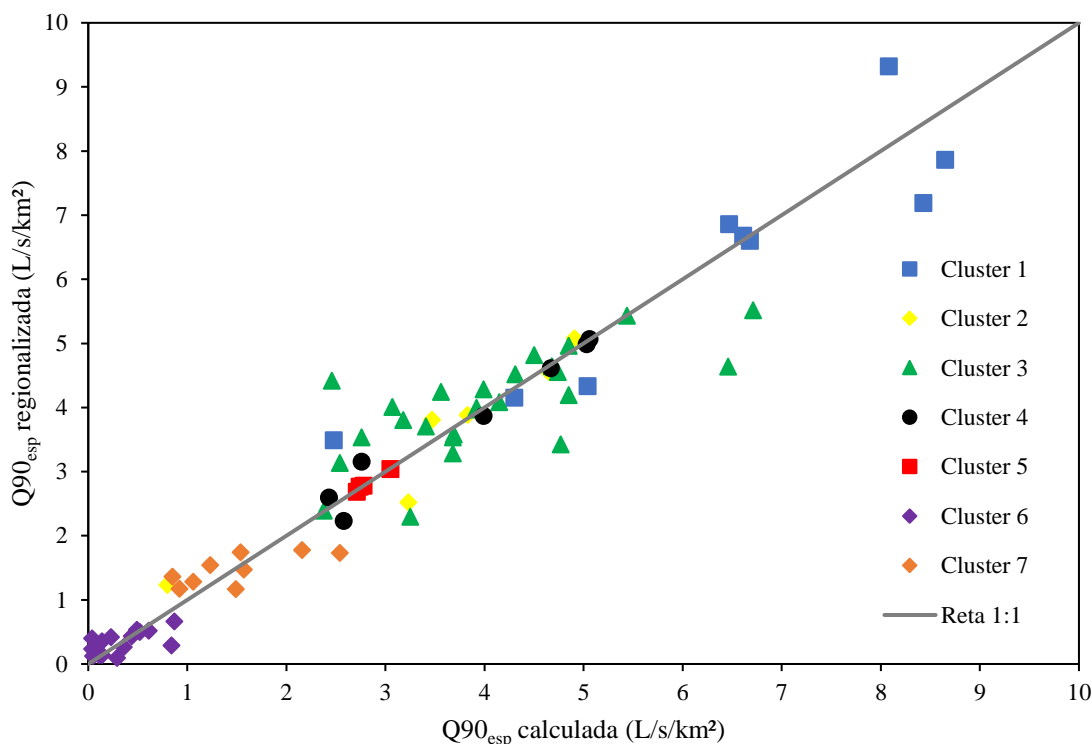


Figura 17 – Relação entre a $Q90_{espr}$ calculada e a $Q90_{espr}$ regionalizada para as 7 regiões hidrologicamente homogêneas no MATOPIBA.

Fonte: Elaboração própria.

O gráfico da vazão específica regionalizada com 90% de permanência ao longo do tempo ($Q90_{espr}$), apresentou os valores da região 3 com maior dispersão em relação a reta 1:1,

fato pode ser explicado pela área que este cluster abrange, sendo fortemente influenciado pela distribuição da precipitação e os valores da latitude, reduzindo a capacidade preditiva dos modelos.

Tabela 13 - Resultado dos critérios de avaliação de desempenho da $Q90_{espr}$, segundo os índices estatísticos recomendados para modelos de regionalização hidrológica por Moriasi et al. (2015).

Região	d	NSE	PBIAS (%)	RSR	R ²
1	0,9799	0,9241	-0,526	0,0871	0,9242
2	0,9792	0,9217	0,899	0,1142	0,9222
3	0,8470	0,5807	0,027	0,1270	0,5084
4	0,9900	0,9610	0,005	0,0746	0,9610
5	0,9955	0,9823	-0,021	0,0665	0,9823
6	0,7199	0,3906	-1,093	0,2016	0,3897
7	0,7981	0,4970	-0,217	0,1831	0,4971

Fonte: Elaboração própria.

O índice RSR para a $Q90_{espr}$ indica um bom ajuste dos modelos aos dados reais, já que seus valores são inferiores ao limite determinado na literatura (0,5); porém, nota-se um aumento do índice na região 6, com resultado acima de 0,20, podendo ser explicado pela dispersão da distribuição das bacias que compõem o cluster dentro da área de estudo.

Tabela 14 - Classificação dos critérios de avaliação de desempenho da $Q90_{espr}$, para os índices estatísticos recomendados para modelos de regionalização hidrológica por Moriasi et al. (2015).

Região	d	NSE	PBIAS (%)	R ²
1	Muito bom	Muito bom	Muito bom	Muito bom
2	Muito bom	Muito bom	Muito bom	Muito bom
3	Satisfatório	Satisfatório	Muito bom	Insatisfatório
4	Muito bom	Muito bom	Muito bom	Muito bom
5	Muito bom	Muito bom	Muito bom	Muito bom
6	Insatisfatório	Insatisfatório	Muito bom	Insatisfatório
7	Satisfatório	Insatisfatório	Muito bom	Insatisfatório

Fonte: Elaboração própria.

Os modelos para a estimativa da $Q90_{espr}$ para as sete regiões hidrologicamente homogêneas apresentaram um bom desempenho ao analisar o índice de concordância (d); porém, duas das regiões (clusters 6 e 7) apresentaram valores abaixo de 0,80, evidenciando uma redução na capacidade preditiva dos mesmos em relação aos modelos para a $Q_{med_{espr}}$.

Os coeficientes de eficiência de modelagem (NSE) para os modelos das regiões 6 e 7 ficaram abaixo de 0,50, sendo estes classificados como “Insatisfatório” pela escala qualitativa. A região 6 foi fortemente afetada pela dispersão das bacias que compõem o agrupamento, já a região 7 possui uma grande variabilidade do comprimento do talvegue (42,5 a 1556,7 km), afetando a capacidade preditiva do modelo.

Em relação ao erro médio percentual (PBIAS), as regiões 1, 5, 6 e 7 apresentaram um leve viés a subestimar os valores das vazões e as demais regiões apresentaram um viés a superestima-los. Para este índice, o desempenho de todas as regiões foi “Muito bom”, sendo que o cluster 6 apresentou o maior erro (-1,09%) em relação aos demais.

O coeficiente de determinação (R^2) manteve o mesmo comportamento do índice NSE, destacando-se as regiões 6 e 7 com valores inferiores a 0,50, ou seja, os valores previstos possuem ajuste aos dados calculados inferior a 50%.

5.6.3. Vazão específica regionalizada com 95% de permanência ($Q95_{\text{espr}}$)

A Figura 18 apresenta um gráfico de dispersão dos dados de vazão média específica calculada e dos dados de vazão média específica regionalizada para as sete regiões hidrologicamente homogêneas definidas anteriormente. Nas Tabelas 15 e 16 são apresentados os resultados dos critérios de avaliação e a classificação qualitativa dos mesmos respectivamente.

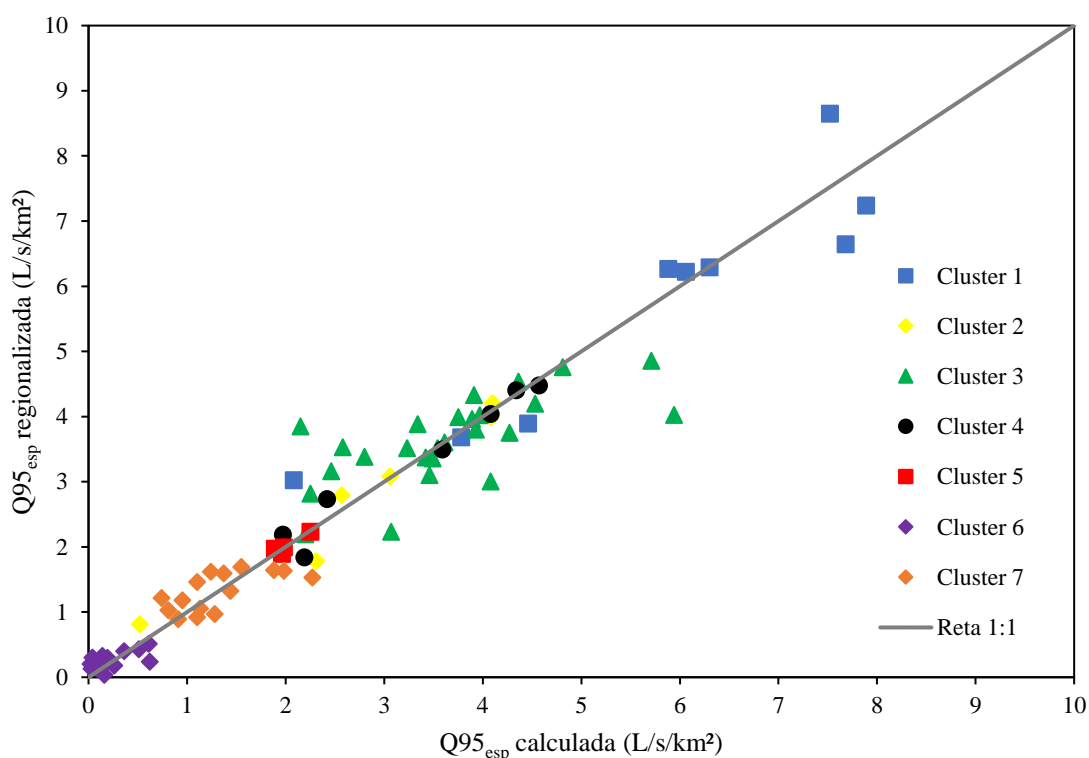


Figura 18 - Relação entre a $Q95_{\text{esp}}$ calculada e a $Q95_{\text{esp}}$ regionalizada para as 7 regiões hidrologicamente homogêneas no MATOPIBA.

Fonte: Elaboração própria.

Como para a $Q90_{espr}$, o gráfico da vazão específica regionalizada com 95% de permanência ao longo do tempo ($Q95_{espr}$), os valores do cluster 3 apresentaram uma maior dispersão dos em relação a reta 1:1, devido aos mesmos fatores ligados a precipitação e a latitude em que as bacias deste agrupamento se encontram.

Tabela 15 - Resultado dos critérios de avaliação de desempenho da $Q95_{espr}$, segundo os índices estatísticos recomendados para modelos de regionalização hidrológica por Moriasi et al. (2015).

Região	d	NSE	PBIAS (%)	RSR	R ²
1	0,9788	0,9202	0,284	0,0894	0,9202
2	0,9874	0,9513	0,031	0,0901	0,9513
3	0,8292	0,5479	-0,060	0,1319	0,4687
4	0,9895	0,9592	-0,003	0,0763	0,9592
5	0,9588	0,8534	0,030	0,1914	0,8535
6	0,7163	0,3793	10,102	0,2034	0,3916
7	0,7608	0,4428	-0,149	0,1927	0,4423

Fonte: Elaboração própria.

O índice RSR para a $Q95_{espr}$ indica que há um bom ajuste dos modelos, já que seus valores são inferiores ao limite determinado na literatura (0,5); porém, como para a $Q90_{espr}$, nota-se um aumento das regiões com valores próximos à 0,20, indicado uma redução da capacidade de precisão dos modelos.

Tabela 16 - Classificação dos critérios de avaliação de desempenho da $Q95_{espr}$, para os índices estatísticos recomendados para modelos de regionalização hidrológica por Moriasi et al. (2015).

Região	d	NSE	PBIAS (%)	R ²
1	Muito bom	Muito bom	Muito bom	Muito bom
2	Muito bom	Muito bom	Muito bom	Muito bom
3	Satisfatório	Satisfatório	Muito bom	Insatisfatório
4	Muito bom	Muito bom	Muito bom	Muito bom
5	Muito bom	Muito bom	Muito bom	Muito bom
6	Insatisfatório	Insatisfatório	Satisfatório	Insatisfatório
7	Satisfatório	Insatisfatório	Muito bom	Insatisfatório

Fonte: Elaboração própria.

Os modelos para a estimativa da $Q95_{espr}$ para as sete regiões hidrologicamente homogêneas apresentaram um bom desempenho ao analisar o índice de concordância (d); porém, é possível observar uma ligeira diminuição da capacidade preditiva dos modelos ao compara-los com os índices de concordância observados para as vazões $Qmed_{espr}$ e $Q90_{espr}$.

Seguindo a tendência vista para os modelos para a $Q90_{espr}$, os modelos para as regiões 6 e 7, apresentaram o pior desempenho para o índice NSE (coeficiente de eficiência de modelagem), tendo resultado abaixo de 0,80.

Em relação ao erro médio percentual (PBIAS), as regiões 3, 4 e 7 apresentaram um leve viés a subestimar os valores das vazões e as demais regiões apresentaram um viés a superestima-los. A região 6 se destacou, tendo como resultado um erro médio percentual de 10,1%, podendo ser explicado pela variabilidade dos valores reais (0,08 a 0,61 L s⁻¹ km⁻²), reduzindo a capacidade preditiva do modelo.

O coeficiente de determinação (R²) manteve o mesmo comportamento do índice NSE, destacando-se as regiões 3, 6 e 7 com valores inferiores a 0,50, ou seja, os valores previstos possuem ajuste aos dados calculados inferior a 50%.

5.6.4. Vazão específica regionalizada mínima de 7 dias consecutivos com período de retorno de 10 anos (Q7,10_{esp})

A Figura 19 apresenta um gráfico de dispersão dos dados de vazão média específica calculada e dos dados de vazão média específica regionalizada para as sete regiões hidrologicamente homogêneas definidas anteriormente. Nas Tabelas 17 e 18 são apresentados os resultados dos critérios de avaliação e a classificação qualitativa dos mesmos respectivamente.

Os resultados para a Q7,10_{esp} seguiram o mesmo padrão visto para as vazões anteriores, em que a região 3 se destaca por ter uma maior dispersão em relação a reta 1:1.

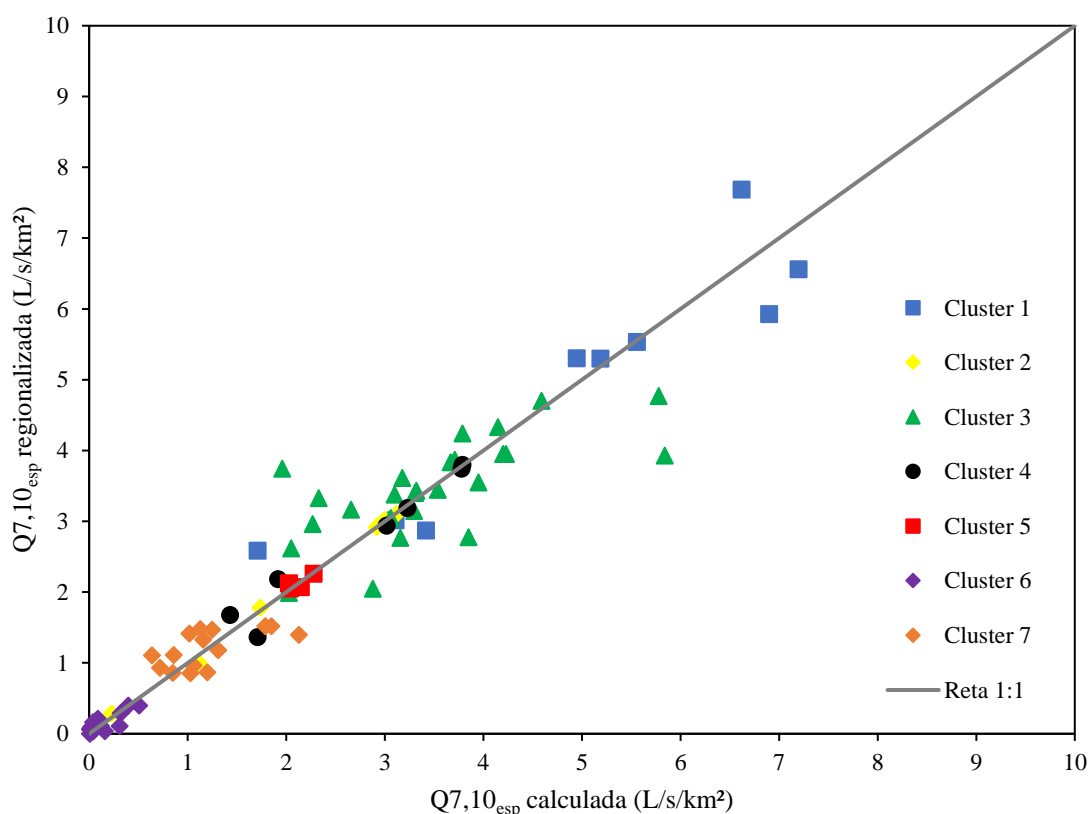


Figura 19 - Relação entre a $Q_{7,10_{esp}}$ calculada e a $Q_{7,10_{esp}}$ regionalizada para as 7 regiões hidrologicamente homogêneas no MATOPIBA.

Fonte: Elaboração própria.

Tabela 17 - Resultado dos critérios de avaliação de desempenho da $Q_{7,10_{espr}}$, segundo os índices estatísticos recomendados para modelos de regionalização hidrológica por Moriasi et al. (2015).

Região	d	NSE	PBIAS (%)	RSR	R ²
1	0,9815	0,9300	0,095	0,0837	0,9300
2	0,9994	0,9977	-0,109	0,0194	0,9977
3	0,8414	0,5705	0,060	0,4705	0,4879
4	0,9888	0,9568	-0,010	0,0786	0,9568
5	0,8700	0,6261	0,032	0,3057	0,6262
6	0,8894	0,6794	0,451	0,1462	0,6803
7	0,7102	0,3797	-0,108	0,2034	0,3764

Fonte: Elaboração própria.

Os valores dos índices RSR para as vazões $Q_{7,10_{espr}}$ indicam que há um bom ajuste dos modelos, já que seus valores são inferiores ao limite determinado na literatura (0,5); porém, o resultado para o modelo da região 3 indicou uma redução da capacidade de precisão do modelo.

Tabela 18 - Classificação dos critérios de avaliação de desempenho da $Q_{7,10_{\text{espr}}}$, para os índices estatísticos recomendados para modelos de regionalização hidrológica por Moriasi et al. (2015).

Região	d	NSE	PBIAS (%)	R ²
1	Muito bom	Muito bom	Muito bom	Muito bom
2	Muito bom	Muito bom	Muito bom	Muito bom
3	Satisfatório	Satisfatório	Muito bom	Insatisfatório
4	Muito bom	Muito bom	Muito bom	Muito bom
5	Bom	Satisfatório	Muito bom	Satisfatório
6	Bom	Satisfatório	Muito bom	Satisfatório
7	Insatisfatório	Insatisfatório	Muito bom	Insatisfatório

Fonte: Elaboração própria.

Os modelos para a estimativa da $Q_{7,10_{\text{espr}}}$ para as sete regiões hidrologicamente homogêneas apresentaram um bom desempenho ao analisar o índice de concordância (d); porém, é possível observar uma ligeira melhora da capacidade preditiva dos modelos ao compará-los com os índices de concordância observados para as vazões $Q_{95_{\text{espr}}}$.

Como para a $Q_{95_{\text{espr}}}$, o modelo para as regiões 7 apresentou coeficiente de eficiência de modelagem (NSE) inferiores a 0,50, o que mostra uma redução da capacidade de estimação em relação aos modelos para a $Q_{\text{med}_{\text{espr}}}$.

As regiões 2, 4 e 7 apresentaram um leve viés de subestimação dos valores; já as demais regiões apresentaram viés de superestimação das vazões. O resultado do PBIAS para todos os modelos foi “Muito bom”, mostrando que os valores estimados estão próximos dos reais.

O coeficiente de determinação (R^2) manteve o mesmo comportamento do índice NSE, destacando-se as regiões 3 e 7 com valores inferiores a 0,50, ou seja, os valores previstos possuem ajuste aos dados calculados inferior a 50%.

Como observado por Charles (2020), a quantidade de postos fluviométricos em conjunto com variáveis morfológicas e climáticas das regiões permitem um melhor desempenho dos modelos de regionalização hidrológica. A densidade de drenagem, variável independente que esteve presente em grande parte dos modelos aqui definidos, está diretamente ligada ao relevo da região que por sua vez tem influência sobre o regime de chuvas.

Sendo definida como um dos principais fatores do ciclo hidrológico, a precipitação total anual das áreas de contribuição dos postos fluviométricos esteve presente na maioria dos modelos de regionalização aqui definidos, visto que esta variável tem correlação direta com a vazão dos cursos d’água. A distribuição da precipitação total anual por bacia na área de estudo mostrou-se pouco variável, sendo de 1154,5 mm para um posto localizado no sudoeste do estado da Bahia e de 1570,2 mm para um posto localizado no noroeste do estado do

Tocantins. A relativa baixa variabilidade se dá graças a extensão das áreas de contribuição destes postos, que em muitos casos possuem seus divisores de águas em outros estados da federação (Pruski et al., 2004).

Em estudo realizado por Beskow et al. (2016), para o estado do Rio Grande do Sul, a sazonalidade se mostrou um importante fator sobre a disponibilidade hídrica de uma bacia hidrográfica. No presente estudo ficou evidente a importância da precipitação total anual para os modelos de regionalização; com isso, para futuros trabalhos indica-se o estudo do efeito da sazonalidade sobre as vazões para a região do MATOPIBA, já que esta possui estações bem definidas e áreas de contribuição que contemplam diferentes biomas brasileiros.

5.7. Validação dos modelos

A validação dos modelos foi realizada com base nos critérios obtidos através da matriz de confusão (precisão, sensibilidade, especificidade, acurácia e taxa de falsos positivos), da curva ROC (*Receiver Operating Characteristic*) e da AUC (*Area Under the Curve*). As Figuras 20, 21, 22 e 23, apresentam as matrizes de confusão geradas para as avaliações das vazões Q_{med} , Q_{90} , Q_{95} e $Q_{7,10}$, respectivamente.

Na Figura 20 são apresentados os cinco intervalos de valores para as diferentes vazões de referência específicas. Nela é possível notar que os valores de vazão se concentraram nas duas primeiras classes (74% para a $Q_{med_{esp}}$; 92% para a $Q_{90_{esp}}$; 92% para a $Q_{95_{esp}}$ e 96% para a $Q_{7,10_{esp}}$), fato este explicado pela grande área de captação que possuem os exutórios (postos fluviométricos), o que implica em vazões específicas menores à medida que nos deslocamos em direção à foz das bacias (Pruski et al., 2004). As tabelas detalhadas para cada conjunto de modelos são apresentadas no Anexo G deste documento.

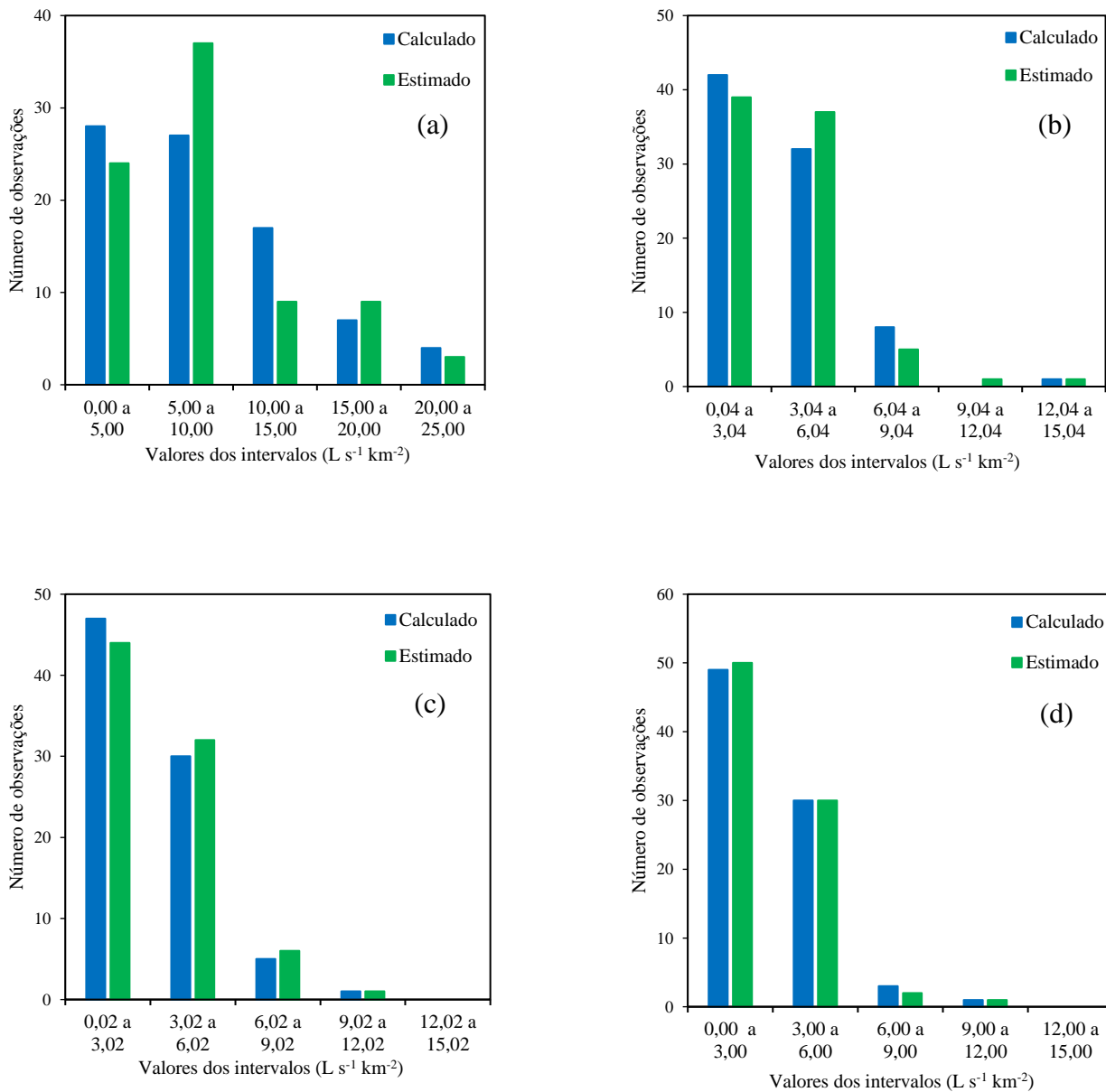


Figura 20 - Número de observações por intervalo; (a) $Q_{med_{esp}}$; (b) $Q_{90_{esp}}$; (c) $Q_{95_{esp}}$ e (d) $Q_{7,10_{esp}}$.
Fonte: Elaboração própria.

Na Figura 21 temos o resultado das matrizes de confusão para as vazões de referência específicas. A diagonal principal representa os valores que foram corretamente estimados dentro das classes. Estas matrizes serviram de base para a obtenção dos diferentes parâmetros e cálculo dos diferentes índices de avaliação dos modelos, além da curva ROC e da AUC, apresentados a seguir.

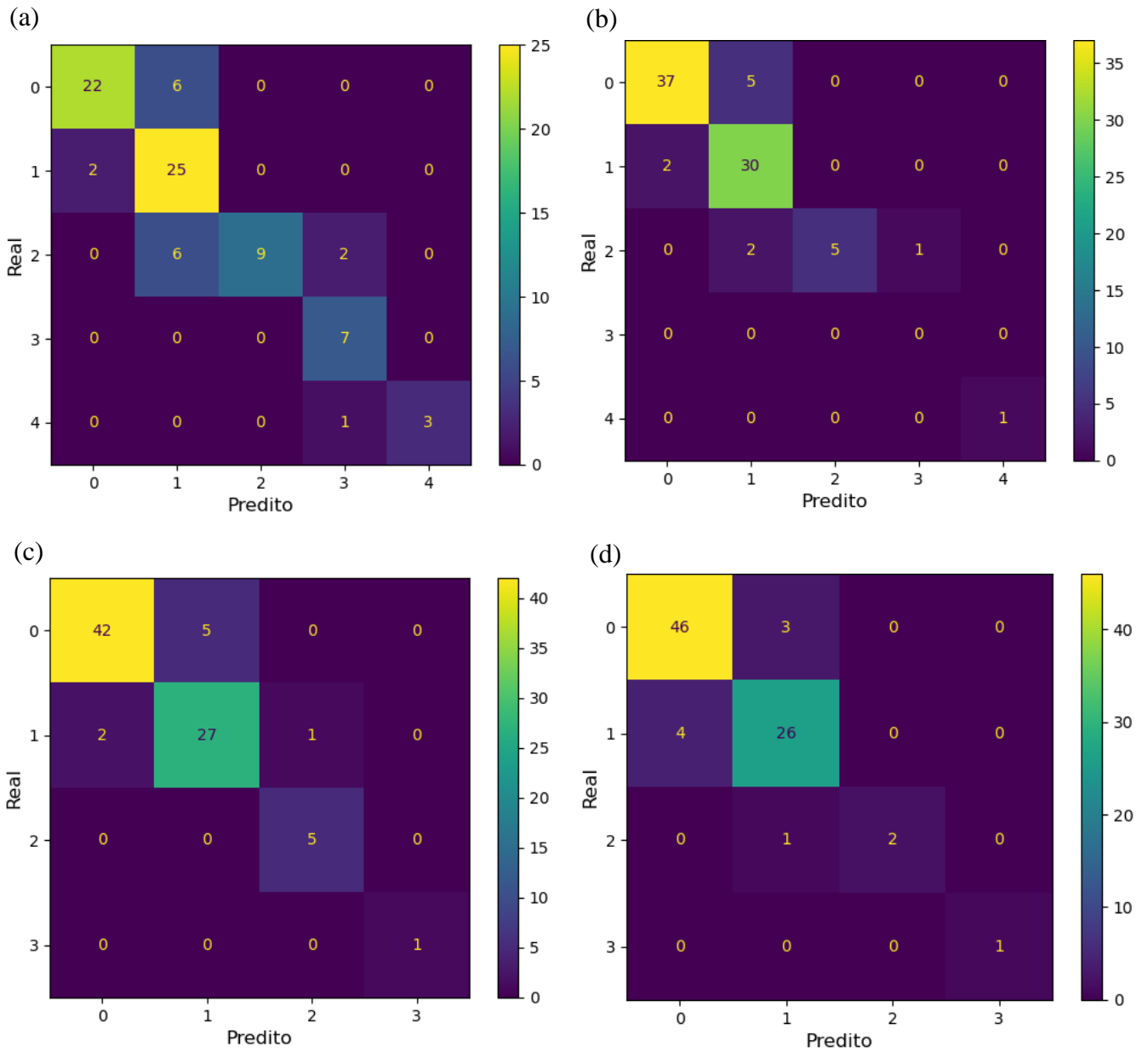


Figura 21 - Matriz de confusão das vazões de referência; (a) $Q_{med_{esp}}$; (b) $Q_{90_{esp}}$; (c) $Q_{95_{esp}}$ e (d) $Q_{7,10_{esp}}$.
Fonte: Elaboração própria.

Nas Tabelas 19 e 20 são apresentados os valores dos parâmetros e os indicadores de desempenho dos modelos, gerados a partir das matrizes de confusão.

Tabela 19 - Resultado dos parâmetros para as vazões de referência, baseados na matriz de confusão.

Qmed _{esp}						Q90 _{esp}					
Parâmetro	C0	C1	C2	C3	C4	Parâmetro	C0	C1	C2	C3	C4
VP	22	25	9	7	3	VP	37	30	5	0	1
FP	2	12	8	3	0	FP	2	7	0	1	0
FN	6	2	0	0	1	FN	5	2	3	0	0
VN	53	44	66	73	79	VN	39	44	75	82	82

Q95 _{esp}						Q7,10 _{esp}					
Parâmetro	C0	C1	C2	C3	C4	Parâmetro	C0	C1	C2	C3	C4
VP	42	27	5	1	0	VP	46	26	2	1	0
FP	2	5	1	0	0	FP	4	4	0	0	0
FN	5	3	0	0	0	FN	3	4	1	0	0
VN	34	48	77	82	83	VN	30	49	80	82	82

Fonte: Elaboração própria.

Tabela 20 - Resultado dos indicadores dos modelos para as vazões de referência específicas, baseados nas matrizes de confusão.

Vazão	Indicadores	C0	C1	C2	C3	C4
Qmed _{esp}	Precisão	0,917	0,676	1,000	0,700	1,000
	Sensibilidade	0,786	0,926	0,529	1,000	0,750
	TFP	0,036	0,214	0,108	0,039	0,000
	Acurácia	0,795				
Q90 _{esp}	Precisão	0,949	0,811	1,000	0,000	1,000
	Sensibilidade	0,881	0,938	0,625	0,000	1,000
	TFP	0,049	0,137	0,000	0,012	0,000
	Acurácia	0,880				
Q95 _{esp}	Precisão	0,955	0,844	0,833	1,000	0,000
	Sensibilidade	0,894	0,900	1,000	1,000	0,000
	TFP	0,056	0,094	0,013	0,000	0,000
	Acurácia	0,904				
Q7,10 _{esp}	Precisão	0,920	0,867	1,000	1,000	0,000
	Sensibilidade	0,939	0,867	0,667	1,000	0,000
	TFP	0,118	0,075	0,000	0,000	0,000
	Acurácia	0,904				

Fonte: Elaboração própria.

A precisão indica a proporção de valores estimados como positivos quando os mesmos fazem parte da classe dos “Positivos”. Para os valores da classe C0 das diferentes vazões, os resultados foram acima de 90%, indicando que os modelos possuem ótima capacidade de estimação para valores baixos de vazão específica. Para os valores da classe C1, houve uma queda no desempenho, evidente no grupo da $Q_{med_{esp}}$, que recebeu valores superestimados da classe C0 e subestimados da classe C2; porém para os demais grupos, a capacidade de estimação ficou acima de 80%. Os valores da classe C2 apresentaram excelente resultado, sendo que o menor valor foi de 83% para vazão $Q_{95_{esp}}$. Como para os valores da classe C2, os valores da classe C3 apresentaram excelente resultado, sendo o menor valor de 70% para a $Q_{med_{esp}}$ devido a tendência da classe C2 em superestimar alguns valores e da classe C4 de subestimar alguns valores. O resultado para a classe C4 foi excelente, sendo que para as vazões $Q_{95_{esp}}$ e $Q_{7,10_{esp}}$ não foram observados dados nesta classe.

A sensibilidade representa a capacidade de prever um valor como positivo quando este faz parte da classe dos “Positivos”. Com exceção do grupo da $Q_{95_{esp}}$, o indicador apresentou menores valores na classe C3, ficando abaixo de 70%. Estes resultados podem estar ligados ao fato de que nesta classe encontram-se os valores intermediários dos intervalos de vazão, resultando em tendência à subestimação ou superestimação dos valores desta classe pelos modelos.

A Taxa de Falsos Positivos (TFP) indica o erro de estimativa dos modelos. As classes C1 e C2 apresentaram os maiores valores para o índice, sendo explicado pela tendência de subestimação (C2, C3) ou de superestimação (C1) dos modelos para estas classes de vazão.

A acurácia representa a capacidade preditiva dos modelos. Os quatro grupos de vazão apresentaram um bom resultado, ficando próximo de 80% para a $Q_{med_{esp}}$, e próximo de 90% para as outras vazões. Este comportamento pode ser explicado pela dispersão dos dados que é para a vazão $Q_{med_{esp}}$ e vai reduzindo à medida que diminuem os valores de vazão de referência específica.

Na Figura 22 pode ser observado o resultado para a análise ROC; os modelos apresentaram um comportamento dito “conservador”, ou seja, eles precisam de uma forte evidência para classificar o dado como sendo da mesma classe que o valor calculado (positivo); sendo assim, eles cometem poucos erros de falso positivo, classificando os valores em outras classes, porém, os modelos para as quatro vazões de referência, não obtiveram um bom desempenho para estimar os valores contidos na classe C1, incluindo nesta classe valores que deveriam estar em outras. Este fato pode ser observado pela posição do ponto C1 (vermelho) em relação aos demais nos gráficos, estando estes mais deslocados para a direita.

Outro ponto a ressaltar ao analisar o gráfico ROC é que as estimativas de todas as classes ficaram acima da reta 1:1, ou seja, os modelos de estimação possuem um desempenho superior a um modelo aleatório. Os modelos para a $Q90_{esp}$, $Q95_{esp}$ e $Q7,10_{esp}$ apresentaram valores para a classe C3 (verde) e C4 (preto) na origem dos gráficos; indicando que não houveram observações nestas classes; com isso, eles não geram classificação de verdadeiro positivo e não cometem erro de falso positivo.

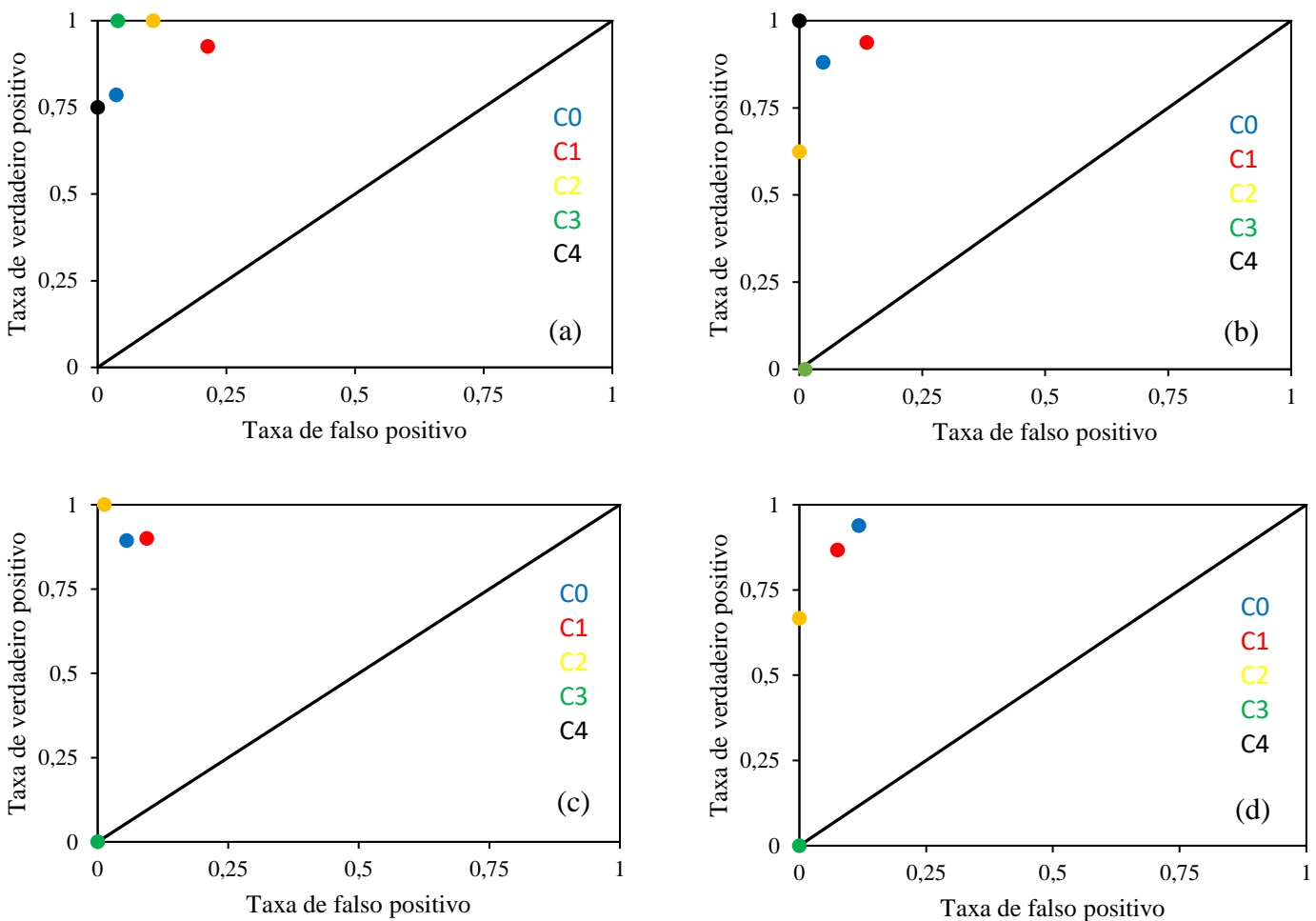


Figura 22 - Gráfico ROC para os quatro grupos de vazão de referência específica; (a) $Q_{med_{esp}}$; (b) $Q90_{esp}$; (c) $Q95_{esp}$ e (d) $Q7,10_{esp}$.
Fonte: Elaboração própria.

Na Tabela 21 e na Figura 23 são apresentados os resultados para análise da área sob a curva (AUC), em que foi adotada a estratégia um contra todos (One-vs-Rest) para medir a capacidade dos modelos de estimar os valores dentro da classe correta, integrando os valores da Taxa de Verdadeiro Positivo (Sensibilidade) e a Taxa de Falso Positivo (1 - Especificidade). Neste método de validação, são geradas curvas e quanto mais próximas ao eixo das ordenadas elas forem, maior será a área abaixo delas, e melhor será a performance

média do modelo para a estimação daquela classe de valores. Linhas sobre a reta em 45° indicam que o desempenho dos modelos para estimar os valores daquela classe igual ao de um modelo aleatório e linhas abaixo indica que o desempenho é inferior a um modelo aleatório.

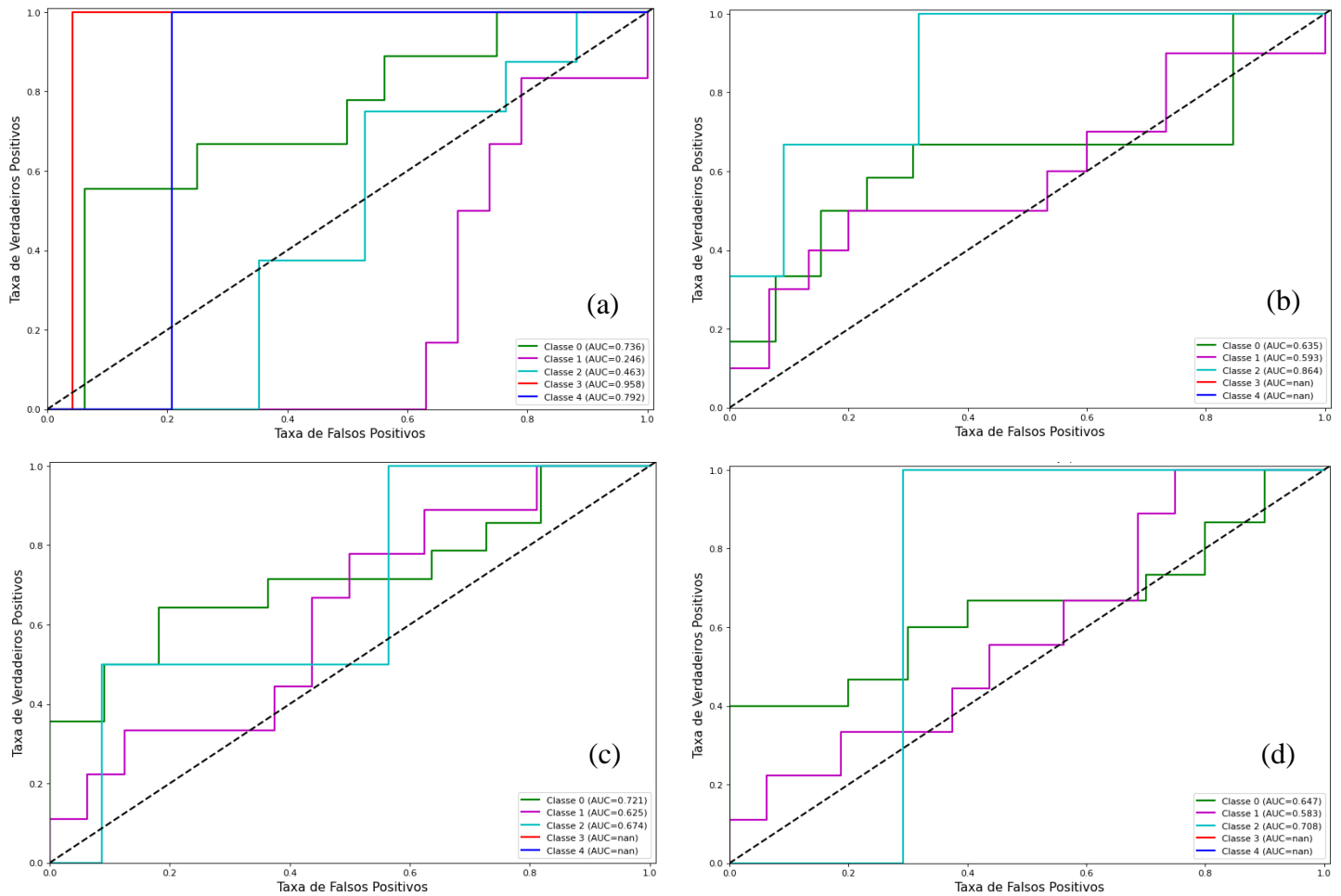


Figura 23 - Gráfico AUC para as vazões de referência específicas; (a) $Q_{med_{esp}}$; (b) $Q_{90_{esp}}$; (c) $Q_{95_{esp}}$ e (d) $Q_{7,10_{esp}}$.

Fonte: Elaboração própria.

Tabela 21 - Resultados da AUC para as diferentes classes de vazões.

Vazão	C0	C1	C2	C3	C4
$Q_{med_{esp}}$	0,736	0,246	0,463	0,958	0,792
$Q_{90_{esp}}$	0,635	0,593	0,864	-	-
$Q_{95_{esp}}$	0,721	0,625	0,674	-	-
$Q_{7,10_{esp}}$	0,647	0,583	0,708	-	-

Fonte: Elaboração própria.

Os modelos para a $Q_{med_{esp}}$ (Figura 23a), conseguiram estimar corretamente 96% dos valores de vazões presentes nas classes C3, 79% dos valores da classe C4 e 74% dos valores da classe C0, porém, para as classes C1 e C2 o resultado foi pior que o de um modelo

aleatório, estando ligado a tendência do modelo a subestimar os valores nestas classes, além da tendência a superestimar valores da classe C0.

Os modelos para as menores vazões ($Q_{90_{esp}}$, $Q_{95_{esp}}$ e $Q_{7,10_{esp}}$) não obtiveram um bom desempenho nesta análise, com exceção da estimativa dos valores presentes na classe C2 da $Q_{90_{esp}}$ que apresentou resultado de 86% de acerto. O resultado pode ser explicado pela concentração de valores nas classes menores, levando os modelos a subestimação ou superestimação das vazões, resultando em um menor desempenho dos modelos.

Após as diversas análises realizadas, ficou evidente que apesar das limitações, os modelos de regionalização para as vazões de referência específicas gerados, apresentaram uma boa capacidade de estimação; contudo, como Du et al. (2020) e Charles (2020) ressaltaram, as equações aqui apresentados devem ser destinados para fins de gerenciamento levando em consideração o erro embutido nas mesmas, que se não considerados, podem levar a erros de cálculos de outorgas por exemplo, impactando na gestão dos recursos hídricos. Contudo, os resultados aqui apresentados podem se tornar mais uma ferramenta de apoio para os órgãos locais, já que atualmente a região do MATOPIBA carece de postos de medição, principalmente para os cursos d'água de pequeno e médio porte.

6. CONSIDERAÇÕES FINAIS

Os modelos de regionalização são importantes ferramentas de suporte para a tomada de decisão dos órgãos locais, visando um planejamento sustentável dos recursos hídricos. A região do MATOPIBA, considerada atualmente como a nova fronteira agrícola brasileira, é confrontada com questões ligadas a expansão das áreas agrícolas, o desenvolvimento socioeconômico e os impactos das mudanças climáticas, necessitando de políticas públicas que visem encontrar o equilíbrio entre estes fatores.

Neste estudo foram utilizadas 6 (seis) variáveis morfométricas e uma variável climática. Foi considerado a utilização do tipo de solo e do uso e cobertura do solo, porém, para a primeira variável não foram encontrados mapas pedológicos com o detalhamento necessário para a realização do estudo. Já para a segunda variável, a capacidade de processamento de dados limitou a inclusão desta variável. Como proposta para uma continuidade do presente trabalho, indica-se a inclusão destas variáveis nas análises, sabendo que estas possuem forte influência sobre a infiltração e o escoamento superficial da precipitação, fatores diretamente ligados a vazão dos cursos d'água.

Na classificação das bacias em regiões hidrologicamente homogêneas, destacam-se os clusters 2 (dois) e 6 (seis), as quais apresentaram grupos dispersos dentro da área de estudo, não podendo ser classificadas como regiões homogêneas. As bacias que compõem o clusters 2 são as que em média, apresentaram o menor comprimento e a maior declividade de talvegue e o menor Kc. As bacias que compõem o cluster 6 são as que para algumas variáveis apresentaram valores próximos de outros grupos, porém não tão próximos para fazer parte deles, desta forma o cluster 6 é constituído pelas bacias que não se enquadraram nos outros grupos.

Os resultados dos modelos de regionalização para as vazões médias específicas de longo período ($Q_{med_{esp}}$), se mostram bem ajustados com os valores reais, tendo como resultados para a maioria dos índices de avaliação entre “Muito bom” e “Bom”. Os modelos para as regiões hidrologicamente homogêneas 3 e 6 foram as que apresentaram desempenho inferior ao das demais, com o NSE e o R^2 abaixo de 0,70. Sendo que o desempenho para a região 3 pode ser explicado pela área que este cluster abrange, sendo influenciado pela distribuição da precipitação e pela latitude, reduzindo a capacidade preditiva dos modelos. Já o desempenho do modelo para a região 6 é explicado pela distribuição das bacias que compõem o cluster, estando estas mais dispersas dentro da área de estudo.

Os modelos para as demais vazões de referência abordadas neste trabalho seguiram o mesmo padrão de desempenho dos modelos para a Qmedesp, sendo que além das regiões 3 e 6, a região 7 aparece como uma das com pior desempenho, sendo os modelos classificados entre “Satisfatório” e “Insatisfatório” para grande parte dos índices. O baixo desempenho para os modelos da região 7 pode ser explicado pela variabilidade do comprimento do talvegue (42,5 a 1556,7 km). Esta variável esteve presente em todos os modelos da região, acarretando em uma menor capacidade preditiva dos mesmos.

A validação das equações geradas através das técnicas de *machine learning* apresentaram resultados de acurácia próximos de 80% para os modelos de estimação da Qmed_{esp} e próximos a 90% para os modelos de estimação das menores vazões, indicando que apesar das limitações, principalmente em relação ao número de postos fluviométricos disponíveis, os modelos possuem um potencial para serem utilizados na região, podendo servir de ferramenta de suporte para as agências de águas da região do MATOPIBA, auxiliando no processo de concessão de outorgas e contribuindo assim, para um melhor gerenciamento dos recursos hídricos na região.

A partir dos resultados encontrados tanto nos índices de avaliação e nos índices de validação dos modelos, indica-se a realização de estudos que permitam a melhoria dos mesmos, tendo como principal algo a inclusão de mais estações fluviométricos na etapa de delimitação das regiões hidrologicamente homogêneas, o que resultaria em conjuntos de postos com características mais próximas. Outra abordagem poderia ser realizada sobre a sazonalidade das chuvas na região, permitindo a obtenção de modelos ajustados para cada época do ano. Uma terceira abordagem seria a inclusão de outras variáveis independentes que possam ter correlação com a vazão de água nos rios da região.

7. CONCLUSÕES

Com base na análise de dados e de acordo com o objetivo geral traçado inicialmente para este trabalho, conclui-se:

- a) A interpolação espacial dos dados pluviométricos, tendo como base o método da krigagem ordinária, mostrou-se uma boa opção para a obtenção da precipitação anual média das bacias para o local de estudo;
- b) A partir da análise de agrupamento (Cluster), foi possível identificar cinco regiões hidrologicamente homogêneas, para a região do MATOPIBA, sendo os clusters 2 e 6 possuindo bacias dispersas dentro da área, não constituindo assim uma região.
- c) Por intermédio do método da regressão múltipla entre as variáveis dependentes e independentes foi possível desenvolver sete modelos regionalização hidrológica para cada uma das vazões de referência abordadas neste trabalho;
- d) Os modelos desenvolvidos apresentaram um bom desempenho geral, sendo estes capazes de auxiliar no gerenciamento dos recursos hídricos para a região do MATOPIBA; e
- e) As técnicas de machine learning (matriz de confusão, curva ROC e AUC) empregadas, se mostram uma boa alternativa para a validação de modelos de regionalização hidrológica.

REFERÊNCIAS

- AGÊNCIA NACIONAL DE ÁGUAS E SANEAMENTO BÁSICO (ANA). Relatório da ANA apresenta situação das águas do Brasil no contexto de crise hídrica. p. 1-8, 2017, disponível em: http://conjuntura.ana.gov.br/static/media/Release_Conjuntura.77d10324.pdf (acesso em: 12 set. 2022).
- AGÊNCIA NACIONAL DE ÁGUAS E SANEAMENTO BÁSICO (ANA). Hidroweb: Sistemas de informações hidrológicas. Disponível em: <https://www.snirh.gov.br/hidroweb/download> (acesso em: 04 jun. 2021).
- AGÊNCIA NACIONAL DE ÁGUAS E SANEAMENTO BÁSICO (ANA). Outorga de direito de uso de recursos hídricos. **Caderno de Recursos Hídricos**. Volume 6. Brasília: ANA, 2011.
- AIYELOKUN, O.O.; ADESOGAN, S.O.; AGBEBE.O. Evaluation of probabilistic models for characterizing design low-flows of river Ogun, Southwest Nigeria. **Nigerian Journal of Technology**, vol. 41, n° 2, p. 408-418, 2022.
- ARAÚJO, F.C. Regionalização de vazões na bacia hidrográfica do rio Paranapanema. Tese (doutorado) Universidade Estadual do Oeste de Paraná, Campus Cascavel, 134p. 2019.
- ARAÚJO, M.L.S.; SANO, E.E.; BOLFE, E.L.; SANTOS, J.R.N.; SANTOS, J.L.; SILVA, F.B. Spatiotemporal dynamics of soybean crop in the Matopiba region, Brazil (1990-2015). **Land Use Policy** 80, 57-67. 2019.
- ASSIS, F.N.; ARRUDA, H.V.; PEREIRA, A.R. Aplicações de estatística à climatologia: teoria e pratica. Pelotas: Editora Universitária/UFPel, 161 p., 1996.
- AUMAMM, C.A. A methodology for developing simulation models of complex systems. **Ecol. Modell.** 202, p. 385-396, 2007.

- BAENA, L.G.N.; da SILVA, D.D.; PRUSKI, F.F.; CALIJURI, M.L. Regionalização de vazões com base em modelo digital de elevação para a bacia do Rio Paraíba do Sul. **Eng. Agric.** 24 (3), p. 612-624, 2004.
- BAENA, L.G.N. Regionalização de vazões para a bacia do Rio Paraíba do Sul, a montante de Volta Redondo, a partir de modelo digital de elevação hidrologicamente consistente. (Tese) 148 p., 2002. Disponível em: <https://www.locus.ufv.br/bitstream/123456789/9466/1/texto%20completo.pdf> (acesso em: 19 jul. 2023).
- BARROSO, L.P.; ARTES, R. Análise multivariada. In: REUNIÃO ANUAL DA RBES E SEAGRO, 48., 100., Lavras. **Curso**. Lavras: Departamento de Ciências Exatas, 155 p., 2003.
- BECK, H.E.; PAN, M.; LIN, P.; SEIBERT, J.; DIJK, A.I.J.M.; WOOD, E.F. Global fully distributed parameter regionalization based on observed streamflow from 4,229 headwater catchments. **J. Geophys. Res. Atmos.** 125(17), p. 1-16, 2020.
- BELCHIOR, E.B.; ALCANTARA, P.H.R.; BARBOSA, C.F. Perspectivas e desafios para a região do Matopiba. **Fronteira Agrícola** n° 16, p. 1-3. 2017.
- BERTONI, J.C.; TUCCI, C.E.M. Precipitação. In: Tucci, C.E.M. **Hidrologia: Ciência e aplicação**. Porto Alegre: UFRGS, p.177-241, 2007. ISBN-13: 978-8570256638;
- BESKOW, S.; MELLO, C.R.; VARGAS, M.M.; CORREA, L de L.; CALDEIRA, T.L.; DURÃES, M.F.; AGUIAR, M.S de. Artificial intelligence techniques coupled with seasonality measures for hydrological regionalization of Q90 under Brazilian conditions. **Journal of Hydrology**, v. 541, p. 1406-1419, 2016.
- BOLFE, E.L.; VICTORIA, D.C.; CONTINI, E.; BAYMA-SILVA, G.; SPINELLI-ARAÚJO, L.; GOMES, D. MATOPIBA: análise do uso da terra e a produção agrícola. **Anais do XVIII Simpósio Brasileiro de Sensoriamento Remoto – SBSR**. 2017.
- BONELL, M.; SUMNER, G. Autumn and winter daily precipitation areas in Wales, 1982-1983 to 1986-1987. **Int. J. Climatol.**, 12, p. 77-102, 1992.

- BRAGA, C.C.; MELO, M.L.D.; MELO, E.C.S. Análise de agrupamento aplicada à distribuição da precipitação no estado da Bahia. **Congresso Brasileiro de Meteorologia**, vol. 10, 1998.
- BRASIL. Lei nº 9.433, de 8 de janeiro de 1997. Institui a Política Nacional de Recursos Hídricos, cria o Sistema Nacional de Gerenciamento de Recursos Hídricos, regulamenta o inciso XIX do art. 21 de Constituição Federal, e altera o art. 1º da Lei nº 8.001, Brasília, DF, p.13. 1997.
- BREUSCH, T.; PAGAN, A. Simple test for heteroscedasticity and random coefficient variation. *Econométrica*, **The Econometric Society**, vol. 47, nº 5, 1979, p. 1287-1294.
- CARDOSO DA SILVA, L.M.; MONTEIRO, R.A. Outorga do direito de uso de recursos hídricos: uma das possíveis abordagens. *Gestão de Águas Doces/Carlos José Saldanha Machado (Organizador)*. Capítulo V, p. 135-178. Rio de Janeiro: Interciência, 2004.
- CHARLES, T.S. Regionalização hidrológica para o estado de Goiás e Distrito Federal. Dissertação (Mestrado em Ciências – concentração em Engenharia de Sistemas Agrícolas) – Escola Superior de Agricultura “Luiz de Queiroz”, Universidade de São Paulo. Piracicaba, p. 98. 2020.
- CHAVES, H.M.L.; ROSA, J.W.C.; VADAS, R.G.; OLIVEIRA, R.V.T. Regionalização de vazões mínimas em bacias através de interpolação em sistemas de informações geográficas. **Revista Brasileira de Recursos Hídricos**, v. 7, nº 3, p. 43-51, 2002.
- CINGOLANI, A.M.; POCA, M.; WHITWORTH-HULSE, J.I.; GIORGIS, M.A.; VAIERETTI, M.V.; HERRERO, L.; RAMOS, S.N.; RENISON, D. Fire reduces dry season low flows in a subtropical highland of central Argentina. **Journal of Hydrology**, volume 590, pp. 125538, 2020.
- CPRM, Serviço Geológico do Brasil. Geodiversidade do Brasil: conhecer o passado, para entender o presente e prever o futuro. Rio de Janeiro, p. 264, 2008.

- CONAB-Companhia Nacional de Abastecimento. Acompanhamento da Safra Brasileira de Grãos, Safra 2021/22 – 12º levantamento. 2022; pp. 1-88. Disponível em: <https://www.conab.gov.br/info-agro/safra/graos/boletim-da-safra-de-graos> (acesso em: 12 set. 2022).
- CRUZ, J.C.; TUCCI, C.E.M. Estimativa da disponibilidade hídrica através da curva de permanência. **Revista Brasileira de Recursos Hídricos**, v. 13, n.1, p. 111 – 124, 2008.
- CRUZ, J.C. Disponibilidade hídrica para outorga: avaliação dos aspectos técnicos e conceituais, Tese (Doutorado em Engenharia de Recursos Hídricos e Saneamento Ambiental), UFRGS, 199p., Porto Alegre, 2001.
- DOURADO, C.S.; OLIVEIRA, S.R.M.; AVIDA, A.M.H. Análise de zonas homogêneas em séries temporais de precipitação no estado da Bahia. **Agrometeorologia, Bragantia** 72 (2), p. 192-198, 2013.
- DU, T.L.T.; LEE, H.; BUI, D.D.; ARHEIMER, B.L.; LI, H.; OLSSON, J.; DARBY, S.E.; SHEFFIELD, J.; KIM, D.; HWANG, E. Streamflow prediction in “geopolitically ungauged” basins using satellite observations and regionalization at subcontinental scale. **Journal of Hydrology**, v. 588, p. 1-21, 2020.
- DURBIN, J.; WATSON, G. Testing for serial correlation in least squared regression, I. **Biometrika**, vol. 37, 1950, p. 409-428.
- EGAN, J. Signal detection theory and ROC analysis. New York, USA: Academic Press, 1975.
- ELETROBRÁS. Centrais Elétricas Brasileiras S.A. Metodologia para regionalização de vazões. Rio de Janeiro, v.1, 1985.
- EMBRAPA. Matopiba GeoWeb, [S.I] [2015?]. Disponível em: <http://mapas.cnpm.embrapa.br/matopiba2015/> (acesso em: 16 dez. 2022).

- EVANGELISTA, B.A.; SILVA, F.A.M. da; SIMON, J.; CAMPOS, L.J.M.; VALE, T.M. do. Zoneamento de risco climático para determinação de épocas de semeadura da cultura da soja na região MATOPIBA. **Embrapa Pesca e Aquicultura - Boletim de Pesquisa e Desenvolvimento** 18. 2017, disponível em: <https://www.embrapa.br/pesca-e-aquicultura/busca-de-publicacoes/-/publicacao/1078131/zoneamento-de-risco-climatico-para-determinacao-de-epocas-de-semeadura-da-cultura-da-soja-na-regiao-matopiba> (acesso em: 12 set. 2022).
- FARACO, M.A., URIBE-OPAIZA, M.A.; SILVA, E.A.A.D.; JOHANN, J.A.; BORSSOI, J.A. Seleção de modelos de variabilidade espacial para elaboração de mapas temáticos de atributos físicos do solo e produtividade de soja. **Revista Brasileira de Ciência do Solo**, v. 32, p. 463-476, 2008.
- FENICIA, F.; MEIBNER, D.; McDONNELL, J.J. Modeling streamflow variability at the regional scale: (2) Development of a bespoke distributed conceptual model. **Journal of Hydrology**, volume 605, 2022.
- FINKLER, N.R.; MENDES, L.A.; SCHNEIDER, E.H.M.; BORTOLIN, T.A.; SCHNEIDER, V.E. Comparação de funções de distribuição de probabilidade na determinação de vazão mínima anual e sazonal. **Scientia Cum Industria** (Sci. Cum Ind.), v. 3, n° 2, p. 42-49, 2015.
- FOX, J.; MONETTE, G. Generalized collinearity diagnostics. **J. Am. Stat. Assoc.** 87, 178-183, 1992.
- FRATTINI, P.; CROSTA, G.; CARRARA, A. Techniques for evaluating the performance of landslide susceptibility models. **Eng. Geol.** 111 (1-4), p. 62-72, 2010.
- GARCIA-MARIN, A.P.; ESTEVEZ, J.; MEDINA-COBO, M.T.; AYUSO-MUNÓZ, J.L. Delimiting homogeneous regions using the multifractal properties of validated rainfall data series. **Journal of Hydrology**, volume 529, p. 106-119, 2015.

- GARCIA-MARIN, A.P.; AYUSO-MUNÓZ, J.L.; TAGUAS-RUIZ, E.V.; ESTEVEZ, J. Regional analysis of the annual maximum daily rainfall in the province of Malaga (southern Spain) using the principal component analysis. **Water Environ. J.**, 25 (4), p. 522-531, 2011.
- GARRIDO, R.J.S. Subprojeto 4.2B – Avaliação dos mecanismos financeiros para o gerenciamento sustentável da bacia do rio São Francisco. In: Projeto de gerenciamento integrado das atividades desenvolvidas em terra na bacia do São Francisco. Brasília: ANA/GAF/PNUMA/OEA, 335 p. 2003.
- GREIS, N.P.; WOOD, E.F. Regional flood frequency estimation and network design. **Water Resour. Res.**, 17 (4), p. 1167-1177, 1981.
- GPRH, 2009. Sistema Computacional para Análises Hidrológicas (SISCAH 1.0). Disponível em: <http://www.gprh.ufv.br/?area=softwares> (acesso em: 12 set. 2022).
- GUEDES, R.V.S.; MACEDO, M.J.H.; SOUZA, F.A.S. Análise espacial de eventos de secas com base no índice padronizado de precipitação e análise de agrupamento. **Revista Brasileira de Ciências Ambientais** n° 23, p.55-65, 2012.
- GUO, Y.; ZHANG, Y.; ZHANG, L.; WANG, Z. Regionalization of hydrological modeling for prediction streamflow in ungauged catchments: A comprehensive review. **Wiley Interdisciplinary Reviews-Water** 8 (1), p. 1-17, 2021.
- GUPTA, H.V.; SOROOSHIAN, S.; YAPO, P.O. Status of automatic calibration for hydrologic models: Comparison with multilevel expert calibration. **J. Hydrologic Eng.** 4(2): 135-143, 1999.
- GUSTARD, A.; BULLOCK, A.; DIXON, J.M. Low flow estimation in the United Kingdom. **Institute of Hydrology Report**, n°108, 1992.
- HABERLANDT, U.; KLÖCKING, B.; KRYSANOVA, V.; BECKER, A. Regionalization of the base flow index from dynamically simulated flow components – a case study in the Elbe River Basin. **Journal of Hydrology**, volume 248, p. 35-53, 2001.

- HAMMOND, J.C.; FLEMING, B.J. Evaluating low flow patterns, drivers and trends in the Delaware River Basin. **Journal of Hydrology**, volume 598, p. 1-13, 2021.
- HAND, D.J.; ANAGNOSTOPOULOS, C. When is area under the receiver operating characteristic curve an appropriate measure of classifier performance? **Journal of Hydrology**, volume 34(5), p. 492-495, 2013.
- HARRIS, N.M.; GURNELL, A.M.; HANNAH, D.M.; PETTS, G.E. Classification of river regimes: a context for hydrogeology. **Hydrological Processes**, Hardcover, v. 14, p. 2831-2848, 2000.
- HOSKING, J.R.M.; WALLIS, J.R.; WOOD, E.F. An appraisal of the regional flood frequency procedure in the UK flood studies report. **Hydrol. Sci. J.**, 30, p. 86-109, 1985.
- IBGE – Instituto Brasileiro de Geografia e Estatística, malhas territoriais. Disponível em: <https://www.ibge.gov.br/geociencias/organizacao-do-territorio/malhas-territoriais/15774-malhas.html> (acesso em: 04 jun. 2021).
- KAMRAN, M.; FAZAL, M.R.; MUSSADAR, M. Towards empowerment of the renewable energy sector in Pakistan for sustainable energy evolution: SWOT analysis. **Renewable Energy**, 146: 543-558, 2020.
- KASSAMBARA, A. Practical Guide to Cluster Analysis in R. 1 ed. USA: **STHDA**, 189p., 2017.
- KINNARD, C.; BZEOUICH, G.; ASSANI, A. Impacts of summer and winter conditions on summer river low flow in low elevation, snow-affected catchments. **Journal of Hydrology**, volume 605, p. 1-13, 2022.
- KITE, G.W. Frequency and risk analyses in hydrology. 5 ed. Highlands Ranch, Colorado: **Water Resources Publications**, 1988, 257p.

- KOLUSU, S.R.; SIDERIUS, C.; TODD, M.C.; BHAVE, A.; CONWAY, D.; JAMES, R.; WASHINGTON, R.; GERESSU, R.; HAROU, J.J.; KASHAIGILI, J.J. Sensitivity of projected climate impacts to climate model weighting: multi-sector analysis in Eastern Africa. **Clim. Chang.**, 164, pp. 1-20. 2021.
- LEGATES, D.R.; MCCABE, G.J. Evaluating the use of “goodness-of-fit” measures in hydrologic and hydroclimatic model validation. **Water Resources Res.** 35 (1): 233-241, 1999.
- LELIS, L.C.S.; NASCIMENTO, J.G.; DUARTE, S.N.; PACHECO, A.B.; BOSQUILIA, R.W.D. Assessment of hydrological regionalization methodologies for the upper Jaguari river basin. **Journal of South American Earth Sciences** 97:102402, 2020.
- LETTENMAIER, D.P.; WALLIS, J.R.; WOOD, E.F. Effect of regional heterogeneity on flood frequency estimation. **Water Resour. Res.**, 23, p. 313-323, 1987.
- LIU, H.; JIA, Y.; NIU, C.; SU, H.; WANG, J.; DU, J.; KHAKI, M.; HU, P.; LIU, J. Development and validation of a physically-based, national-scale hydrological model in China. **Journal of Hydrology**, volume 590, p. 1-14, 2020.
- MAGALHÃES, L.A.; MIRANDA, E.E. MATOPIBA: quadro natural. Nota Técnica 5, EMBRAPA/GITE, Campinas-SP, 41p., dezembro 2014.
- MALEKINEZHAD, H.; ZARE-GARIZI, A. Regional frequency analysis of daily rainfall extremes using L-moments approach. **Atmosfera**, v. 27, n° 4, p. 411-427, 2014.
- MARCUZZO, F.F.N.; GOULARTE, E.R.P. Caracterização do ano hidrológico e mapeamento espacial das chuvas nos períodos úmido e seco do estado do Tocantins. **Revista Brasileira de Geografia Física**, v. 06, n° 01, p. 91-99, 2013.
- MARENGO, J.A.; GALDOS, M.V.; CHALLINOR, A.; CUNHA, A.P.; MARIN, F.R.; VIANNA, M.d.S.; ALVALA, R.C.S.; ALVES, L.M.; MORAES, O.L.; BENDER, F. Drought in Northeast Brazil: A review of agricultural and policy adaptation options for food security. **Climate Resilience and Sustainability**, 1, e17, 2022.

- MATAVELI, G.A.V.; SILVA, M.E.S.; PEREIRA, G.; KAWAKUBO, F.S.; BERTANI, G.; OLIVEIRA, B.S. Análise das queimadas e da precipitação em áreas de Cerrado do Maranhão a partir de dados do sensor MODIS e do satélite TRMM para o período 2002-2015. **Boletim Paulista de Geografia** v. 96, p.11-30, 2017.
- MATHERON, G. Principles of geostatistics. **Eco. Geol.**, v. 58, p. 1246-1266, 1963.
- McMILLAN, N.K.; BOOKER, D.J.; CATTOËN, C. Validation of a national hydrological model. **Journal of Hydrology**, volume 541, p. 800-815, 2016.
- MEDEIROS, R.M. Variabilidade decadal da precipitação em microrregiões do Piauí – Brasil. **Revista de Geografia (Recife)** v. 34, n° 1, 2017.
- MEIRA, M.A.; FREITAS, E.S.; COELHO, V.H.R.; TOMASELLA, J.; FOWLER, H.J.; FILHO, G.M.R.; SILVA, A.L.; ALMEIRA, C.N. Quality control procedures for sub-hourly rainfall data: An investigation in different spatio-temporal scales in Brazil. **Journal of Hydrology**, volume 613, Part A, 128358, 13p., 2022.
- MELLO, J.M.; BATISTA, J.L.F.; RIBEIRO JUNIOR, P.J.; OLIVEIRA, M.S. Ajuste e seleção de modelos espaciais de semivariograma visando à estimativa volumétrica de *Eucalyptus grandis*. **SCIENTIA FORESTALIS**, n.69, p. 25-37, 2005.
- MENDES, L.A. Análise dos critérios de outorga de direito de usos consultivos dos recursos hídricos baseados em vazões mínimas e em vazões de permanência. Dissertação (Mestrado) – Escola Politécnica, Universidade de São Paulo, São Paulo, 187 p. 2007.
- MOREIRA, G.B.; MAIA, J.L.; SILVA, J.P.O. Disponibilidade hídrica superficial da bacia hidrográfica do rio Suaçuí: análise por diferentes metodologias de definição de vazão de referência. **Research, Society and Development**, v.11, n. 3, e0211325987, 2022.
- MORIASI, D.N.; GITAU, M.W.; PAI, N.; DAGGUPATI, P. Hydrologic and water quality models: Performance measures and evaluation criteria. Trans. **ASABE**, 58(6), 1763-1785, 2015.

- MUSIEROSE, C.; KOKILAVANI, S.; DHEEBAKARAN, G.; SATHYAMOORTHY, N.K.; PANGAYAR-SELVI, R. Performance of Regional Climate Model (WRF 4.3) in Medium Range Rainfall Forecast (MRRF) for Tamil Nadu, India. **IJECC**, 12(11): 570-576, 2022.
- MYLNE, K.R. Decision-making from probability forecasts based on forecast value. **Meteorological Applications**, v.9, p. 307-315, 2002.
- NASCIMENTO, D.; NOVAIS, G. Clima do Cerrado: Dinâmica atmosférica e características, variabilidade e tipologias climáticas. **Élisée, Rev. Geo. UEG**, n.9, n.2, e922021, 2020.
- NASH, J.E.; SUTCLIFFE, J.V. River flow forecasting through conceptual models: Part 1. A discussion of principles. **Journal of Hydrology** 10(3): 282-290, 1970.
- NERC. National Environmental Research Council. Flood Studies Report, vol. 1. London: **National Environmental Research Council**, 1975.
- NETO, V.L.S.; SOUZA, L.B.; VIOLA, M.R.; MORAIS, M.A.V. Variabilidade espacial e temporal da precipitação no estado do Tocantins, Brasil. **Research, Society and Development**, v. 10, n° 3, e44310313545, 2021.
- NOVAES, Marcel. Distribuições de probabilidade. **Revista Brasileira de Ensino de Física**, vol. 44, e20210424, 2022.
- OECD-FAO. Perspectives agricoles de l'OCDE et de la FAO 2020-2029. Editions OECD, Paris/FAO, Rome. 2020. <https://doi.org/10.1787/ccc6f09c-fr>.
- OLIVEIRA, T.C.C.; GONÇALVES, F.A. Tamanho ideal de séries históricas de vazão mínimas anuais e suas aplicações em estudos hidrológicos: estudo de caso nas regiões hidrográficas Paraná e São Francisco. **Engenharia Sanitária e Ambiental**, vol. 27 (4), 2022
- ONO, S. Análise dos critérios de outorga de direito de uso de recursos hídricos considerando a disponibilidade de água superficial. Dissertação (Mestrado) – Escola Politécnica, Universidade de São Paulo, São Paulo, 118 p. 2006.

- OPROMOLLA, P.A.; DALDEN, I.; CARDIM, M. Análise geoestatística de casos de hanseníase no Estado de São Paulo, 1991-2002. **Rev. Saúde Pública**, 40(5), p. 907-913, 2006.
- PAULHUS, J.L.H.; KOHLER, M.A. Interpolation of missing precipitation records. **Monthly Weather Review**, 80, p. 129-133, 1952.
- PESSOA, F.C.L. Desenvolvimento de metodologia para regionalização de curvas de permanência de vazões na Amazônia Legal. 236 p. Tese (Doutorado em Engenharia de Recursos Naturais) – Programa de Pós-Graduação em Engenharia de Recursos Naturais da Amazônia da Universidade Federal do Pará, Belém, 2015.
- PINTO, L.C.; MACHADO, E.F.P.; MELLO, C.R.; SILVA, A.M. Análise de distribuição de probabilidade e estimativa da $Q_{7,10}$ para a região do rio Itabapoana, Espírito Santo/Rio de Janeiro. XIX congresso da pós-graduação da UFLA, 2010.
- PRATI, R.C.; BATISTA, G.E.A.P.A.; MONARD, M.C. Curvas ROC para avaliação de classificadores. **IEEE LATIN AMERICA TRANSACTIONS**, v.6, n.2, p. 215-222, 2008.
- PRUSKY, F.F.; NUNES, A.A.; PRUSKI, P.L.; RODRIGUEZ, R.G. Improved regionalization of streamflow by use of the streamflow equivalent of precipitation as an explanatory variable. **Journal of Hydrology**, volume 467, p. 52-71, 2013.
- PRUSKY, F.F.; PEREIRA, S.B.; NOVAES, L.F.; SILVA, D.D.; RAMOS, M.M. Precipitação média anual e vazão específica média de longa duração, na Bacia do São Francisco. **Revista Brasileira de Engenharia Agrícola e Ambiental**, v.8, n.2/3, p. 247-253, 2004.
- PYRCE, R. Hydrological low flow indices and their use. Peterborough, ON: **Watershed Science Centre**, Report n° 4, p. 147-186, 2004.
- QGIS Development Team, 2022. QGIS Geographic Information System. Open-Source Geospatial Foundation Project. URL: <http://qgis.osgeo.org>.

- R Core Team (2022). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <http://www.R-project.org/>.
- RIGGS, H.C. Estimating flow characteristics at ungauged sites. In: Regionalization in Hydrology, **Proceedings of the Ljubljana Symposium**, IAHS, n° 191, 1990.
- RODRIGUES, J.A.M.; VIOLA, M.R.; MELLO, C.R.; MORAIS, M.A.V. Hydrological regionalization of streamflow for the Tocantins River basin in Brazilian Cerrado biome. **Rev. Ambient. Água**, volume 16, n° 6, e2716, 2021.
- ROMESBURG, C.H. How to make classifications. In: ROMESBURG, C.H. (Ed.). **Cluster analysis for researchers**. Belmont: Lifetime Learning Publications, 203-216, 1984.
- RYKIEL Jr, E.J. Testing ecological models: the meaning of validation. **Ecol. Modell.** 90, p. 229-244, 1996.
- SANTIAGO, D.D.B.; BARBOSA, H.A.; CORREIA FILHO, W.L.F.; OLIVEIRA-JUNIOR, J.F.d. Interactions of environmental variables and water use efficiency in the Matopiba Region via multivariate analysis. **Sustainability**, 14, 8758, 2022.
- SANTOS, P.V.C.J.; CUNHA, A.C. da. Outorga de recursos hídricos e vazão ambiental no Brasil: Perspectivas metodológicas frente ao desenvolvimento do setor hidrelétrico na Amazonia. **Revista Brasileira de Recursos Hídricos**, v. 18, p. 81-95, 2013.
- SATYANARAYANA, P.; SRINIVAS, V.V. Regionalization of precipitation in data sparse areas using large scale atmospheric variables – a fuzzy clustering approach. **Journal of Hydrology**, volume 405, p. 462-473, 2011.
- SCHNIER, S.; CAI, X. Prediction of regional streamflow frequency using model tree ensembles. **Journal of Hydrology**, volume 517, p. 289-309, 2014.
- SHAPIRO, S.S.; WILK, M.B. An analysis of variance test for normality (complete samples). **Biometrika**, vol.52, n 3 e 4, 1965, p. 591-611.

- SILVA, A.M.; OLIVEIRA, P.M.; MELLO, C.R.; PIERANGELI, C. Vazões mínimas e de referência para outorga na região do Alto Rio Grande, Minas Gerais. **Revista Brasileira de Engenharia Agrícola e Ambiental**, v.10, p. 374-380, 2006.
- SILVA, C.B. da. Conflitos por água no Brasil: panoramas na região semiárida. **Boletim DATALUTA**, nº 136, abril de 2019.
- SILVA, D.D.; MARQUES, F.A.; LEMOS, A.F. Flexibilidade das vazões mínimas de referência com a adoção do período trimestral. **Revista Engenharia na Agricultura**, 19 (3), p. 244-254, 2011.
- SILVA, D.D.; MARQUES, F.A.; LEMOS, A.F. Avaliação de metodologias de regionalização de vazões mínimas de referência para bacia do Rio São Francisco. **Revista Engenharia na Agricultura**, 17 (5), p. 392-403, 2009.
- SILVA, F.B.; SANTOS, J.R.N.; ARAUJO, M.L.S.; SILVA JUNIOR, C.H.L. Análise espaço-temporal da precipitação no estado do Maranhão no período de 2003 a 2012. **XIV Safety, Health and Environment World Congress**, Cubatão, 2014.
- SILVA, F.C. Análise ROC. São José dos Campos, 6 de dezembro de 2006. Disponível em: https://www.cin.ufpe.br/~jead/bioInformatica/roc_analyzes.pdf, acesso em: 21 out. 2022.
- SIVAPALAN, M.; TAKEUCHI, K.; FRANKS, S.W.; GUPTA, V.K.; KARAMBIRI, H.; LAKSHMI, V.; LIANG, X.; MCDONNELL, J.J.; MEDIOONDO, E.M.; O'CONNEL, P.E.; OKI, T.; POMEROY, J.W.; SCHERTZER, D.; UHLENBROOK, S.; ZEHE, E. IAHS decade on predictions in ungauged basins (PUB), 2003-2012: shaping an exciting future for the hydrological sciences. **Hydrological Sciences Journal**, v.48, nº 6, p. 857-880. 2012.
- SWAIN, J.B.; PATRA, K.C. Streamflow estimation in ungauged catchments using regionalization techniques. **Journal of Hydrology**, volume 554, p. 420-433, 2017.
- TASKER, G.D.; STEDINGER, J.R. An operational GLS model for hydrologic regression. **Journal of Hydrology**, vol. 111, p. 361-375, 1989.

TUCCI, C.E.M. Regionalização de vazões. Porto Alegre, Ed. ABRH/UFRGS, 256 p. 2002.

UBERTINI, L.; CASADEI, S.; MANCIOLA, P. Evaluation of the minimal instream flow of the Tiber River Basin. **Environmental Monitoring and Assessment**, 41 (2), p. 125-136, 1996.

UNITED NATIONS. Transforming our world: the 2030 Agenda for Sustainable Development. General Assembly, 2015. Disponível em: <https://sdgs.un.org/2030agenda> (acesso em: 12 set. 2022).

VAN LIEW, M.W.; ARNOLD, J.G.; GARBRECHT, J.D. Hydrologic simulation on agricultural watersheds: Choosing between two models. Trans. **ASAE** 46(6): 1539-1551, 2003.

VIDYARTHI, V.K.; JAIN, A. Knowledge extraction from trained ANN drought classification model. **Journal of Hydrology**, volume 585, 124804, 16p., 2020.

VILLELA, S.M.; MATTOS, A. Hidrologia aplicada, São Paulo: **McGraw-Hill do Brasil**, 245p., 1975.

VOGEL, R.M.; FENNESSEY, N.M. Flow-duration curves II. A review of application in water resources planning. **Water Resource Bull**, 31 (6): 1029-1039, 1995.

WARD, J.H. Hierarchical grouping to optimize an objective function. **Journal of the American Statistical Association**, 58, 236-244, 1963.

WILLMOTT, C.J. On the validation of models. **Physical Geography** 2: 184-194, 1981.

WONNACOTT, T.H.; WONNACOTT, R.J. Introductory statistics. 2.ed. New York: **John Wiley Sons**, 510p., 1972.

XU, Z.; WU, Z.; SHAO, Q.; HE, H.; GUO, X. A two-step calibration framework for hydrological parameter regionalization based on streamflow and remote sensing evapotranspiration. **Journal of Hydrology** 613, 128320, 2022.

YAMAMOTO, J.K. Estatística, Análise e Interpolação de dados Geoespaciais. São Paulo: **Gráfica Paulos**, 308p., ISBN: 9786599072727, 2020;

YAMAMOTO, J.K.; LANDIM, P.M.B. Geoestatística, conceitos e aplicações. São Paulo: **Oficina de Textos**, 215p. ISBN: 9788579750779, 2013.

YANG, X.; MAGNUSSON, J.; HUANG, S.; BELDRING, S.; XU, C.Y. Dependence of regionalization methods on the complexity of hydrological models in multiple climatic regions. **Journal of Hydrology** 582, 124357, 2020.

YÜREKLI, K.; MODARRES, R. Regionalization of maximum daily rainfall data over Tokat province, Turkey. **Int. J. Nat. Eng. Sci.**, 1 (2), p. 1-7, 2007.

ZHENG, H.; ZHANG, L.; LIU, C.; SHAO, Q.; FUKUSHIMA, Y. Changes in stream flow regime in headwater basins of the Yellow River basin since the 1950s. **Journal of Hydrological Process**, 21(7): 886-893, 2007.

ANEXOS

ANEXO A. Distribuições de probabilidade utilizadas para a estimação da vazão $Q_{7,10}$

Distribuição de Gumbel

Também conhecida como distribuição de valores extremos do tipo I, ou distribuição do tipo I de Fisher-Tippett. Segundo Kite (1988), a magnitude do evento para séries finitas é dada pela Equação 46.

$$M = \mu + \frac{\sigma_x}{\sigma_n}(b - Y_n) \quad (46)$$

Sendo:

$$b = -\ln\left(-\ln\left(1 - \frac{1}{T}\right)\right) \quad (47)$$

O fator de frequência para séries infinitas é calculado através da Equação 48.

$$K = -\left\{0,45 + 0,7797\ln\left(-\ln\left(1 - \frac{1}{T}\right)\right)\right\} \quad (48)$$

Em que:

M – Magnitude do evento de vazão [$\text{m}^3 \text{s}^{-1}$];

μ – Média da amostra de vazão [$\text{m}^3 \text{s}^{-1}$];

σ_x – Desvio padrão da amostra de vazão [$\text{m}^3 \text{s}^{-1}$];

σ_n – Desvio padrão da variável reduzida da vazão [$\text{m}^3 \text{s}^{-1}$];

b – Variável reduzida da vazão [$\text{m}^3 \text{s}^{-1}$];

Y_n – Média da variável reduzida da vazão [$\text{m}^3 \text{s}^{-1}$]; e

T – Período de retorno [anos].

Distribuição Log-Normal tipo II

De acordo com Kite (1988), para a série gerada a partir dos logaritmos dos eventos da série de dados, o fator de frequência é calculado pela Equação 49.

$$K = \frac{\left(e^{\sqrt{\ln(z^2+1)}} \cdot T - \frac{\ln(z^2+1)}{2} \right)^{-1}}{z} \quad (49)$$

Sendo:

$$z = \frac{\sigma}{\mu} \quad (50)$$

Em que:

T – Período de retorno [anos];

σ – Desvio padrão dos eventos de vazão [$\text{m}^3 \text{s}^{-1}$]; e

μ – Média dos eventos da amostra de vazão [$\text{m}^3 \text{s}^{-1}$].

Distribuição Log-Normal tipo III

Segundo Kite (1988), para a série gerada a partir dos logaritmos dos eventos da série de dados, o fator de frequência é calculado por meio da Equação 51.

$$K = \frac{\left(e^{\sqrt{\ln(z^2+1)}} \cdot T - \frac{\ln(z^2+1)}{2} \right)^{-1}}{W} \quad (51)$$

Sendo:

$$W = \frac{1 - \omega^{\frac{2}{3}}}{\omega^{\frac{1}{3}}} \quad (52)$$

$$\omega = \frac{-y + \sqrt{y^2 + 4}}{2} \quad (53)$$

$$y = \frac{N}{(N-1)(N-2)} \sum_{i=1}^N \left(\frac{(x_i - \mu)^3}{\sigma^3} \right) \quad (54)$$

Em que:

T – Período de retorno [anos];

σ – Desvio padrão dos eventos da amostra de vazão [$\text{m}^3 \text{s}^{-1}$];

μ – Média dos eventos da amostra de vazão [$\text{m}^3 \text{s}^{-1}$];

y – Assimetria [adimensional]; e

N – Número de eventos [adimensional].

Distribuição de Pearson tipo III

Também conhecida como distribuição Gama Tipo III. Segundo Kite (1988), o fator de frequência é calculado segundo a Equação 55.

$$K = D + (D^2 - 1) \frac{y}{6} + \frac{1}{3} (D - 6D) \left(\frac{y}{6}\right)^2 - (D^2 - 1) \left(\frac{y}{6}\right)^3 + D \left(\frac{y}{6}\right)^4 + \frac{1}{3} \left(\frac{y}{6}\right)^5 \quad (55)$$

Sendo:

$$D = T - \frac{2,30753+0,27061T}{1+0,99229T+0,04481T^2} \quad (56)$$

$$y = \frac{N}{(N-1)(N-2)} \sum_{i=1}^N \left(\frac{(x_i - \mu)^3}{\sigma^3} \right) \quad (57)$$

Em que:

D – Desvio padrão padronizado da vazão [$\text{m}^3 \text{s}^{-1}$];

y – Assimetria [adimensional];

T – Período de retorno [anos];

N – Número de eventos da amostra [adimensional];

x_i – i-ésimo evento da amostra [adimensional];

μ – Média dos eventos de vazão [$\text{m}^3 \text{s}^{-1}$]; e

σ – Desvio padrão da vazão [$\text{m}^3 \text{s}^{-1}$].

Distribuição Weibull

Recomendada para série de eventos mínimos, seu fator de frequência é dado, conforme Kite (1988), segundo a Equação 58.

$$K = A_\alpha + B_\alpha \left(\left(-\ln \left(1 - \frac{1}{T} \right) \right)^{\frac{1}{\alpha}} - 1 \right) \quad (58)$$

Sendo:

$$\alpha = \frac{1}{0,2777757913+0,3132617714y+0,0575670910y^2-0,0013038566y^3-0,0081523408y^4} \quad (59)$$

Em que:

T – Período de retorno [anos]; e

y – Assimetria [adimensional].

Esta equação só é válida para valores de α entre -1,02 e 2,00.

$$y = \frac{\mu_3}{(\mu_2)^2} \quad (60)$$

$$A_\alpha = \left(1 - \Gamma\left(1 + \frac{1}{\alpha}\right)\right) B_\alpha \quad (61)$$

$$B_\alpha = \frac{1}{\sqrt{\left(\Gamma\left(1 + \frac{2}{\alpha}\right) - \Gamma^2\left(1 + \frac{1}{\alpha}\right)\right)}} \quad (62)$$

Em que:

$$\Gamma(X) = \int_0^\infty X^{\lambda-1} e^{-x} dx \quad (63)$$

Sendo $\Gamma(X)$ é a função Gama.

ANEXO B. Descrição dos testes estatísticos para a análise dos modelos de regionalização

Índice de concordância (d)

O índice de concordância (Willmott, 1981) permite a detecção de diferenças aditivas e proporcionais nos dados, além da variância observada entre os dados reais e os simulados. O índice apresenta uma sensibilidade a valores extremos devido às diferenças quadradas (Legates & McCabe, 1999), presentes na Equação 64.

$$d = 1 - \left| \frac{\sum_{i=1}^n (Est_i - Obs_i)^2}{\sum_{i=1}^n (|Est_i - Obs_i| + |Obs_i - \overline{Obs}|)^2} \right| \quad (64)$$

Em que:

Est_i – Valor de vazão específica regionalizada [$L s^{-1} km^{-2}$];

Obs_i – Valor de vazão específica calculada [$L s^{-1} km^{-2}$]; e

\overline{Obs} – Média dos valores específicos calculados [$L s^{-1} km^{-2}$].

Teste de eficiência de Nash-Sutcliffe (NSE)

O teste de eficiência de Nash-Sutcliffe (NSE) é uma estatística normalizada que descreve o grau de variação residual dos dados estimados em comparação com a variância dos dados observados (Nash & Sutcliffe, 1970). O teste é amplamente utilizado e potencialmente confiável estatisticamente para a análise do ajuste de modelos hidrológicos (Equação 65).

$$NSE = 1 - \left| \frac{\sum_{i=1}^n (Est_i - Obs_i)^2}{\sum_{i=1}^n (|Obs_i - \overline{Obs}|)^2} \right| \quad (65)$$

O teste de eficiência de Nash-Sutcliffe varia de $-\infty$ a 1, sendo que 1 corresponde ao ajuste perfeito.

Viés percentual (PBIAS)

O viés percentual (PBIAS) mede a tendência média dos valores estimados serem maiores ou menores do que os valores observados. Segundo Gupta et al. (1999), o PBIAS tem um valor ideal igual a 0 (zero), sendo que valores de baixa magnitude indicam um modelo preciso na simulação. Valores positivos indicam um viés de superestimação do modelo, enquanto que valores negativos indicam um viés de subestimação do modelo (Equação 66).

$$PBIAS = \left(\frac{\sum_{i=1}^n |Obs_i - Est_i|}{\sum_{i=1}^n Obs_i} \cdot 100 \right) \quad (66)$$

Razão entre o RMSE e o desvio padrão dos dados observados (RSR)

O RSR varia de um valor ideal 0 (zero), a um valor significativo positivo. Quanto menor for o valor do RSR, mais baixo será o RMSE (raiz do erro médio quadrático) e mais preciso será o modelo de previsão (Equação 68).

$$RSR = \frac{RMSE}{STDEV_{obs}} = \frac{\left(\frac{\sqrt{\sum_{i=1}^n (Obs_i - Est_i)^2}}{n} \right)}{\left(\sqrt{\sum_{i=1}^n (|Obs_i - Obs|)^2} \right)} \quad (68)$$

Quando o RMSE for igual a 0 (zero), os valores observados e preditos estarão perfeitamente alinhados, enquanto que um valor alto RMSE implica em um péssimo ajuste do modelo; valores menores que a metade desvio padrão dos dados observados (medidos) indicam um bom ajuste do modelo de previsão (Musierose et al., 2022).

Coefficiente de determinação (R²)

O coeficiente de determinação descreve o grau de colinearidade entre os dados simulados e medidos. Variando entre 0 (zero) e 1 (um), os valores mais altos indicam menor variação de erro. Segundo Van Liew et al. (2003), valores maiores que 0,5 são geralmente considerados aceitáveis (Equação 12).

ANEXO C. Dados de precipitação total anual estimada média para as áreas de contribuição dos 84 postos fluviométricos utilizados no estudo.

Tabela 22 - Lista com os postos fluviométricos, rio, município e estado de localização, além da precipitação média anual para a área de contribuição.

Posto	Município	Estado	Rio	Precipitação (mm)
21750000	Lavandeira	Tocantins	Mosquito	1349,41
22050001	Peixe	Tocantins	Tocantins	1383,33
22150000	Peixe	Tocantins	Santa Tereza	1434,72
22190000	Porto Alegre do Tocantins	Tocantins	Manuel Alves	1415,70
22220000	Natividade	Tocantins	Manuel Alves	1417,39
22500000	Miracema do Tocantins	Tocantins	Tocantins	1420,89
22680000	Novo Acordo	Tocantins	do Sono	1415,06
22700000	Novo Acordo	Tocantins	do Sono	1418,91
22900000	Pedro Afonso	Tocantins	do Sono	1442,67
23150000	Itacajá	Tocantins	Manuel Alves pequeno	1480,00
23220000	Campos Lindos	Tocantins	Manuel Alves grande	1421,07
23230000	Goiatins	Tocantins	Vermelho	1439,87
23250000	Goiatins	Tocantins	Manuel Alves grande	1436,27
23600000	Tocantinópolis	Tocantins	Tocantins	1425,54
23700000	Itaguatins	Tocantins	Tocantins	1425,98
26350000	São Félix do Araguaia	Mato Grosso	Araguaia	1439,61
26710000	Sandolândia	Tocantins	Javaés	1475,68
26720000	Sandolândia	Tocantins	Formoso	1473,48
26800000	Pium	Tocantins	Javaés	1485,55
27380000	Dois Irmãos do Tocantins	Tocantins	Piranhas	1570,23
27500000	Conceição do Araguaia	Para	Araguaia	1469,89
27550000	Arapoema	Tocantins	Jenipapo	1538,92
28150000	Muricilândia	Tocantins	Muricizal	1535,51
28240000	Piraquê	Tocantins	Lontra	1527,42
28300000	Xambioá	Tocantins	Araguaia	1478,02
28850000	Araguatins	Tocantins	Araguaia	1480,19
33025000	Açailândia	Maranhão	Pindaré	1356,84
33215000	Barra do Corda	Maranhão	Corda ou Capim	1365,10

33250000	Barra do Corda	Maranhão	Mearim	1366,47
33260000	Barra do Corda	Maranhão	Mearim	1369,25
33281000	Pedreiras	Maranhão	Mearim	1378,93
33290000	Bacabal	Maranhão	Mearim	1384,09
33321000	Grajau	Maranhão	Grajau	1370,55
33380000	Bela Vista do Maranhão	Maranhão	Grajau	1416,18
33420000	Mirador	Maranhão	Itapecuru	1328,41
33450000	Mirador	Maranhão	Alpercatas	1352,90
33460000	Mirador	Maranhão	Alpercatas	1348,80
33520000	Buriti Bravo	Maranhão	Correntes	1324,11
33530000	Governador Eugênio Barros	Maranhão	Itapecuru	1340,54
33550000	Caxias	Maranhão	Itapecuru	1347,39
33590000	Codo	Maranhão	Itapecuru	1357,55
33620000	Codo	Maranhão	Codozinho	1420,12
33630000	Coroatá	Maranhão	Itapecuru	1368,42
33700000	Barreirinhas	Maranhão	Preguiças	1475,00
33730000	Vargem Grande	Maranhão	Munim	1457,69
33760000	São Benedito do Rio Preto	Maranhão	Preto	1469,78
33770000	Vargem Grande	Maranhão	Iguará	1476,49
34060000	Ribeiro Gonçalves	Piauí	Paranaíba	1377,92
34070000	Uruçuí	Piauí	Paranaíba	1373,28
34090000	Uruçuí	Piauí	Uruçuí Preto	1341,75
34130000	Balsas	Maranhão	das Balsas	1401,80
34230000	Monte Alegre do Piauí	Piauí	Contrato	1369,40
34251000	Cristiano Castro	Piauí	Gurguéia	1323,98
34311000	Barão de Grajau	Maranhão	Paranaíba	1346,50
34660000	Palmeiras	Piauí	Parnaíba	1327,15
34690000	Teresina	Piauí	Parnaíba	1343,40
34879500	Luzilândia	Piauí	Paranaíba	1346,01
45131000	Montalvânia	Minas Gerais	Carinhanha	1154,48
45260000	Juvenília	Minas Gerais	Carinhanha	1227,44
45298000	Carinhanha	Bahia	São Francisco	1306,12
45480000	Bom Jesus da Lapa	Bahia	São Francisco	1303,62

45740001	Correntina	Bahia	do Meio	1192,66
45770000	Correntina	Bahia	Arrojado	1184,49
45840000	Jaborandi	Bahia	Formoso	1182,09
45880000	Coribe	Bahia	Formoso	1157,73
45910000	Santa Maria da Vitoria	Bahia	Corrente	1187,78
45960001	Santana	Bahia	Corrente	1203,13
46035000	Sítio do Mato	Bahia	São Francisco	1300,07
46105000	Paratinga	Bahia	São Francisco	1279,87
46415000	São Desidério	Bahia	Grande	1259,44
46455000	São Desidério	Bahia	das Fêmeas	1288,90
46490000	São Desidério	Bahia	São Desidério	1217,72
46543000	Barreiras	Bahia	das Ondas	1296,53
46570000	Barreiras	Bahia	de Janeiro	1344,14
46590000	Barreiras	Bahia	Branco	1305,97
46610000	Angical	Bahia	Grande	1268,84
46650000	Cotegipe	Bahia	Grande	1264,89
46675000	Cotegipe	Bahia	Grande	1260,81
46770000	Formosa do Rio Preto	Bahia	Sapão	1358,04
46790000	Formosa do Rio Preto	Bahia	Preto	1351,21
46830000	Santa Rita de Cassia	Bahia	Preto	1336,63
46870000	Mansidão	Bahia	Preto	1325,67
46902000	Mansidão	Bahia	Grande	1277,79

ANEXO D. Dados das vazões de referência Q_{med} , Q_{90} e Q_{95} calculados, da área total da bacia de captação e das coordenadas geográficas das 83 estações fluviométricas utilizadas no estudo.

Tabela 23 - Lista com os postos fluviométricos, geolocalização, área de captação e vazões de referência.

Posto	Rio	Área	Latitude	Longitude	Q_{med}	Q_{90}	Q_{95}
		(km²)	(°)	(°)	(m³ s⁻¹)	(m³ s⁻¹)	(m³ s⁻¹)
21750000	Mosquito	1030,4	-12,79	-46,51	36,13	27,00	25,90
22050001	Tocantins	128064,9	-12,02	-48,53	1523,58	598,37	522,18
22150000	Santa Tereza	13871,2	-11,98	-48,66	146,78	3,98	2,21
22190000	Manuel Alves	1785,6	-11,61	-47,04	40,18	22,81	18,93
22220000	Manuel Alves	10350,5	-11,76	-47,84	127,70	25,69	21,49
22500000	Tocantins	14422,9	-11,53	-48,29	227,65	67,02	58,81
22680000	do Sono	16933,5	-10,00	-47,47	271,81	146,49	133,62
22700000	do Sono	18036,0	-9,96	-47,67	285,73	152,04	138,62
22900000	do Sono	43939,6	-9,31	-47,93	692,03	284,18	259,27
23150000	Manuel Alves pequeno	2826,7	-8,39	-47,77	48,00	14,24	12,61
23220000	Manuel Alves grande	3000,2	-7,99	-46,93	32,52	12,91	11,35
23230000	Vermelho	4063,2	-7,96	-47,26	75,58	32,83	30,57
23250000	Manuel Alves grande	10035,3	-7,71	-47,31	149,64	66,32	60,77
23600000	Tocantins	288962,0	-6,29	-47,39	4047,31	1461,15	1321,20
23700000	Tocantins	296992,9	-5,79	-47,47	4208,14	1495,16	1289,42
26350000	Araguaia	194867,3	-11,62	-50,66	2620,53	776,82	699,59
26710000	Javaés	5586,9	-12,10	-49,59	172,23	2,47	1,09
26720000	Formoso	6611,3	-12,42	-49,53	66,90	0,91	0,54
26800000	Javaés	39129,1	-10,56	-49,38	542,03	13,96	10,11
27380000	Piranhas	1664,8	-9,18	-49,05	24,53	1,33	0,87
27500000	Araguaia	325649,6	-8,27	-49,26	4785,54	897,41	786,58
27550000	Jenipapo	1196,9	-7,61	-49,05	26,50	3,86	2,77
28150000	Muricizal	1572,8	-7,15	-48,61	31,97	5,46	4,05
28240000	Lontra	3457,6	-6,67	-48,47	76,03	16,98	14,19
28300000	Araguaia	370777,2	-6,41	-48,54	5496,09	901,85	730,63
28850000	Araguaia	382077,1	-5,65	-48,13	5891,95	984,80	825,34
33025000	Pindaré	3728,5	-5,77	-45,98	24,14	3,25	2,29
33215000	Corda ou Capim	2992,7	-5,74	-45,32	21,31	14,53	13,06

33250000	Mearim	13154,4	-5,50	-45,24	56,27	33,47	29,86
33260000	Mearim	15451,5	-5,10	-44,96	56,84	33,44	30,55
33281000	Mearim	23933,7	-4,57	-44,61	84,72	37,67	34,43
33290000	Mearim	25443,4	-4,22	-44,77	102,58	39,10	34,79
33321000	Grajau	4513,0	-5,82	-46,14	23,72	2,78	2,29
33380000	Grajau	20279,3	-3,77	-45,22	98,01	4,67	2,92
33420000	Itapecuru	6073,4	-6,37	-44,36	16,79	9,02	7,79
33450000	Alpercatas	5762,2	-6,07	-44,71	30,95	22,99	21,63
33460000	Alpercatas	6892,3	-6,01	-44,34	31,39	23,53	22,27
33520000	Correntes	5322,5	-5,71	-43,59	6,17	0,57	0,40
33530000	Itapecuru	27310,9	-5,34	-43,88	58,58	33,64	30,03
33550000	Itapecuru	32673,3	-4,87	-43,37	68,30	34,55	30,92
33590000	Itapecuru	37178,3	-4,46	-43,88	104,58	34,32	30,23
33620000	Codozinho	5499,0	-4,47	-43,93	35,68	0,77	0,60
33630000	Itapecuru	43780,2	-4,13	-44,13	149,70	37,15	32,20
33700000	Preguiças	2031,6	-2,97	-42,95	16,12	7,79	6,21
33730000	Munim	4205,3	-3,58	-43,70	29,66	0,16	0,10
33760000	Preto	3977,9	-3,33	-43,52	24,86	3,36	2,46
33770000	Iguará	2639,9	-3,55	-43,87	26,47	0,10	0,10
34060000	Paranaíba	31345,6	-7,57	-45,25	210,90	130,05	123,33
34070000	Paranaíba	35307,8	-7,38	-44,83	211,37	129,93	122,25
34090000	Uruçuí Preto	15619,0	-7,39	-44,61	33,06	21,15	19,35
34130000	das Balsas	9089,3	-7,53	-46,04	97,56	60,72	57,24
34230000	Contrato	801,6	-9,58	-45,03	1,88	0,42	0,10
34251000	Gurguéia	31023,7	-8,79	-44,21	17,70	1,56	1,01
34311000	Paranaíba	140204,8	-6,76	-43,03	465,28	273,96	264,19
34660000	Parnaíba	234891,0	-5,57	-43,02	513,25	273,87	258,39
34690000	Parnaíba	237123,2	-5,14	-42,81	557,37	284,15	269,00
34879500	Paranaíba	298236,9	-3,45	-42,37	656,14	289,93	271,96
45131000	Carinhanha	5995,8	-14,31	-44,46	60,46	32,60	28,82
45260000	Carinhanha	16274,7	-14,26	-44,15	125,73	73,16	63,65
45298000	São Francisco	253775,9	-14,30	-43,76	1954,90	774,82	571,55
45480000	São Francisco	270646,1	-13,26	-43,44	1983,60	742,73	512,32

45740001	do Meio	8981,8	-13,29	-44,56	37,08	22,08	19,31
45770000	Arrojado	5644,0	-13,45	-44,57	52,47	36,46	33,52
45840000	Formoso	7131,8	-13,71	-44,64	72,05	47,89	40,70
45880000	Formoso	9555,1	-13,56	-44,30	72,67	44,69	37,97
45910000	Corrente	29692,7	-13,40	-44,20	194,48	127,45	115,19
45960001	Corrente	31024,2	-13,29	-43,91	193,46	121,46	109,98
46035000	São Francisco	307654,6	-12,87	-43,38	2165,57	832,79	602,13
46105000	São Francisco	316165,1	-12,70	-43,23	2122,07	878,10	627,09
46415000	Grande	4967,4	-12,43	-45,09	28,11	15,24	12,80
46455000	das Fêmeas	6358,7	-12,41	-45,12	45,50	30,35	25,94
46490000	São Desidério	4487,7	-12,40	-44,95	5,00	2,22	1,61
46543000	das Ondas	5391,0	-12,14	-45,10	44,69	26,16	23,01
46570000	de Janeiro	2682,3	-11,90	-45,61	12,65	5,31	4,16
46590000	Branco	7382,4	-11,86	-45,12	45,77	28,90	26,64
46610000	Grande	33025,1	-11,98	-44,88	172,21	104,86	92,59
46650000	Grande	35919,7	-11,72	-44,50	169,57	99,01	88,19
46675000	Grande	39650,6	-11,61	-44,16	165,12	100,72	89,40
46770000	Sapão	7084,3	-10,99	-45,53	32,65	25,21	23,64
46790000	Preto	14393,4	-11,05	-45,20	93,52	68,20	65,20
46830000	Preto	18423,6	-11,01	-44,52	93,68	68,02	64,31
46870000	Preto	22190,6	-11,24	-43,95	101,09	72,10	68,11
46902000	Grande	69416,8	-11,35	-43,83	261,75	165,16	152,68

ANEXO E. Dados da vazão de referência $Q_{7,10}$ calculada, intervalo superior e inferior com 95% de confiança, amplitude e distribuição de probabilidade para as 83 estações fluviométricas utilizadas neste trabalho.

Tabela 24 – Lista com os postos fluviométricos, vazão $Q_{7,10}$, intervalos de confiança e tipo de distribuição.

Posto	$Q_{7,10}$ ($m^3 s^{-1}$)	Intervalo de confiança (95%)		Amplitude	Distribuição
		Superior	Inferior		
21750000	25,17	25,67	24,67	1,003	Weibull
22050001	413,54	453,89	373,19	80,681	Logpearson 3
22150000	0,11	0,20	0,02	0,179	Logpearson 3
22190000	18,21	19,43	16,98	2,454	Logpearson 3
22220000	18,21	19,43	16,98	2,454	Logpearson 3
22500000	450,05	489,22	410,88	78,336	Weibull
22680000	121,97	126,95	117,00	9,946	Logpearson 3
22700000	124,41	129,71	119,10	10,616	Logpearson 3
22900000	217,68	231,12	204,23	26,889	Logpearson 3
23150000	9,66	10,55	8,78	1,769	Weibull
23220000	9,32	10,14	8,50	1,642	Weibull
23230000	26,88	28,18	25,59	2,5947	Weibull
23250000	52,10	55,40	48,79	6,611	Logpearson 3
23600000	1096,50	1158,66	1034,33	124,331	Logpearson 3
23700000	1123,53	1191,69	1055,37	136,321	Logpearson 3
26350000	588,93	627,52	550,35	77,170	Weibull
26710000	0,37	0,54	0,21	0,321	Logpearson 3
26720000	0,13	0,20	0,06	0,138	Logpearson 3
26800000	6,38	7,98	4,78	3,197	Logpearson 3
27380000	0,39	0,47	0,30	0,171	Logpearson 3
27500000	626,60	684,01	569,19	114,817	Logpearson 3
27550000	1,33	1,60	1,06	0,541	Logpearson 3
28150000	2,73	3,09	2,38	0,709	Logpearson 3
28240000	10,37	11,58	9,16	2,421	Logpearson 3
28300000	528,51	583,09	473,92	109,177	Logpearson 3
28850000	653,31	707,96	598,66	109,294	Logpearson 3
33025000	1,89	2,42	1,36	1,061	Logpearson 3
33215000	12,43	13,37	11,48	1,889	Logpearson 3

33250000	28,08	30,16	26,00	4,152	Logpearson 3
33260000	28,64	30,79	26,48	4,317	Logpearson 3
33281000	31,29	34,18	28,41	5,775	Logpearson 3
33290000	31,87	34,52	29,22	5,299	Logpearson 3
33321000	1,45	1,76	1,15	0,612	Logpearson 3
33380000	1,88	2,37	1,39	0,973	Logpearson 3
33420000	7,30	7,96	6,64	1,322	Weibull
33450000	21,16	22,40	19,93	2,467	Logpearson 3
33460000	21,35	22,46	20,24	2,217	Logpearson 3
33520000	0,33	0,39	0,26	0,125	Logpearson 3
33530000	27,79	29,73	25,86	3,869	Logpearson 3
33550000	28,26	30,64	25,87	4,766	Logpearson 3
33590000	26,82	29,66	23,99	5,670	Weibull
33620000	0,21	0,29	0,13	0,168	Logpearson 3
33630000	28,08	30,95	25,22	5,735	Logpearson 3
33700000	5,94	6,74	5,13	1,608	Logpearson 3
33760000	1,22	1,62	0,81	0,807	Logpearson 3
34060000	111,04	116,19	105,89	10,300	Weibull
34070000	111,67	116,37	106,98	9,389	Logpearson 3
34090000	17,68	18,57	16,80	1,765	Logpearson 3
34130000	50,53	52,73	48,34	4,397	Weibull
34230000	0,010	0,013	0,009	0,005	Logpearson 3
34251000	0,45	0,57	0,32	0,249	Logpearson 3
34311000	250,80	261,57	240,02	21,556	Logpearson 3
34660000	241,43	249,42	233,44	15,978	Weibull
34690000	251,03	263,19	238,88	24,313	Logpearson 3
34879500	253,12	263,81	242,43	21,382	Logpearson 3
45131000	27,50	30,05	24,95	5,098	Logpearson 3
45260000	61,71	67,20	56,21	10,993	Logpearson 3
45298000	578,50	637,42	519,57	117,848	Logpearson 3
45480000	549,73	611,47	487,98	123,491	Logpearson 3
45740001	17,58	19,06	16,10	2,962	Logpearson 3
45770000	32,94	34,89	31,00	3,889	Logpearson 3

45840000	41,22	44,41	38,03	6,384	Logpearson 3
45880000	40,09	43,48	36,70	6,783	Logpearson 3
45910000	109,84	118,37	101,31	17,058	Logpearson 3
45960001	103,09	111,55	94,62	16,988	Logpearson 3
46035000	661,18	733,82	588,54	145,275	Logpearson 3
46105000	644,10	724,73	563,47	161,259	Logpearson 3
46415000	11,57	12,90	10,25	2,649	Logpearson 3
46455000	24,51	26,28	22,75	3,528	Logpearson 3
46490000	1,78	1,99	1,56	0,426	Logpearson 3
46543000	21,30	23,18	19,42	3,761	Logpearson 3
46570000	3,12	3,58	2,66	0,914	Logpearson 3
46590000	24,45	26,31	22,60	3,708	Logpearson 3
46610000	87,99	94,99	81,00	13,984	Logpearson 3
46650000	81,58	88,31	74,86	13,451	Logpearson 3
46675000	81,18	87,98	74,38	13,596	Logpearson 3
46770000	22,56	23,45	21,67	1,783	Weibull
46790000	60,95	63,22	58,69	4,529	Weibull
46830000	60,75	63,07	58,43	4,644	Logpearson 3
46870000	64,00	66,24	61,76	4,477	Logpearson 3
46902000	140,78	150,85	130,71	20,138	Logpearson 3

ANEXO F. Gráficos de dispersão dos dados calculados e regionalizados para as vazões $Q_{med_{esp}}$, $Q_{90_{esp}}$, $Q_{95_{esp}}$ e $Q_{7,10_{esp}}$ das 7 regiões hidrologicamente homogêneas.

Região hidrologicamente homogênea 1

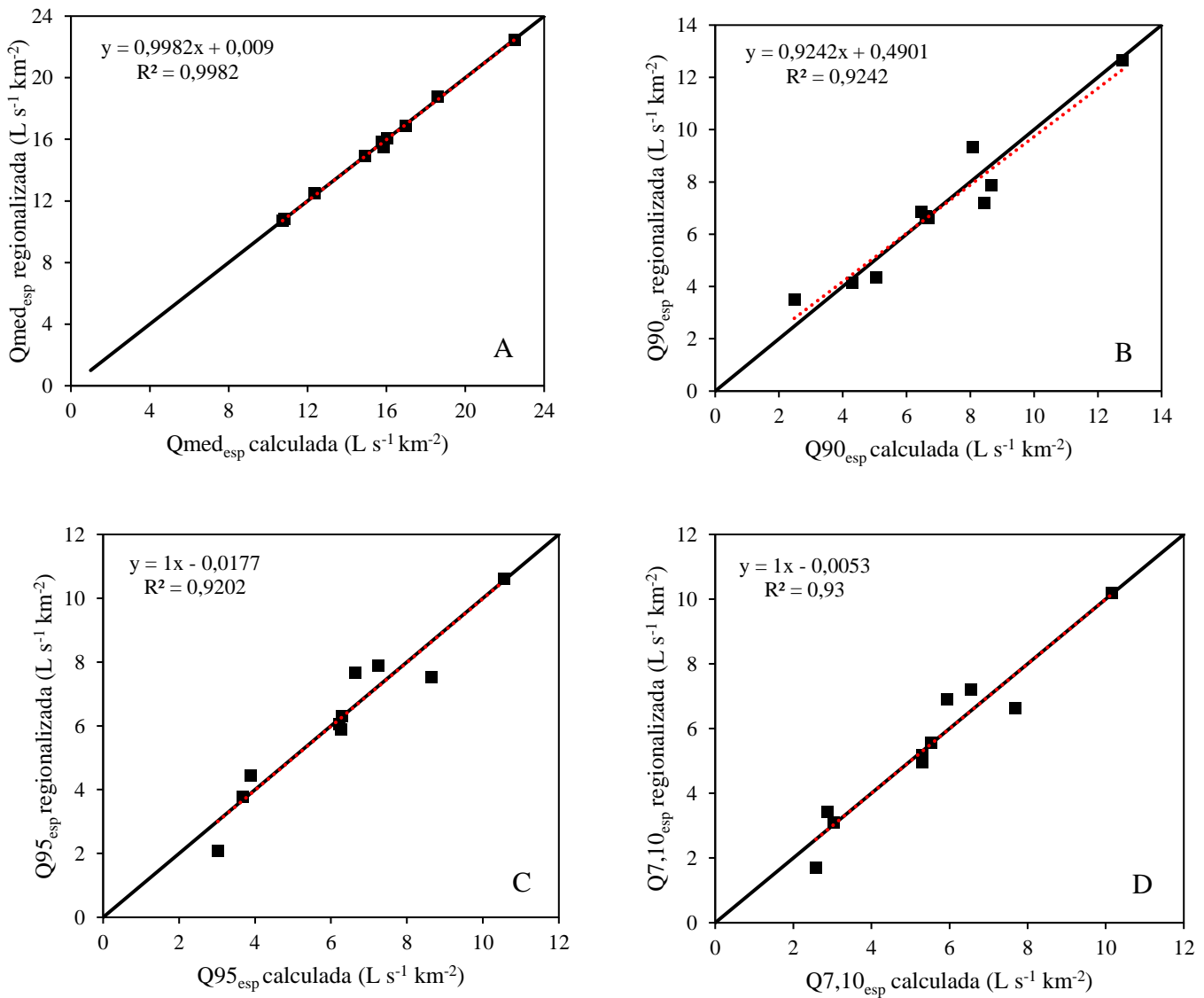


Figura 24 - Relação $Q_{med_{esp}}$ regionalizada e $Q_{med_{esp}}$ calculada; B) Relação $Q_{90_{esp}}$ regionalizada e $Q_{90_{esp}}$ calculada; C) Relação $Q_{95_{esp}}$ regionalizada e $Q_{95_{esp}}$ calculada; D) Relação $Q_{7,10_{esp}}$ regionalizada e $Q_{7,10_{esp}}$ calculada, para a região hidrologicamente homogênea 1.

Fonte: Elaboração própria.

Região hidrologicamente homogênea 2

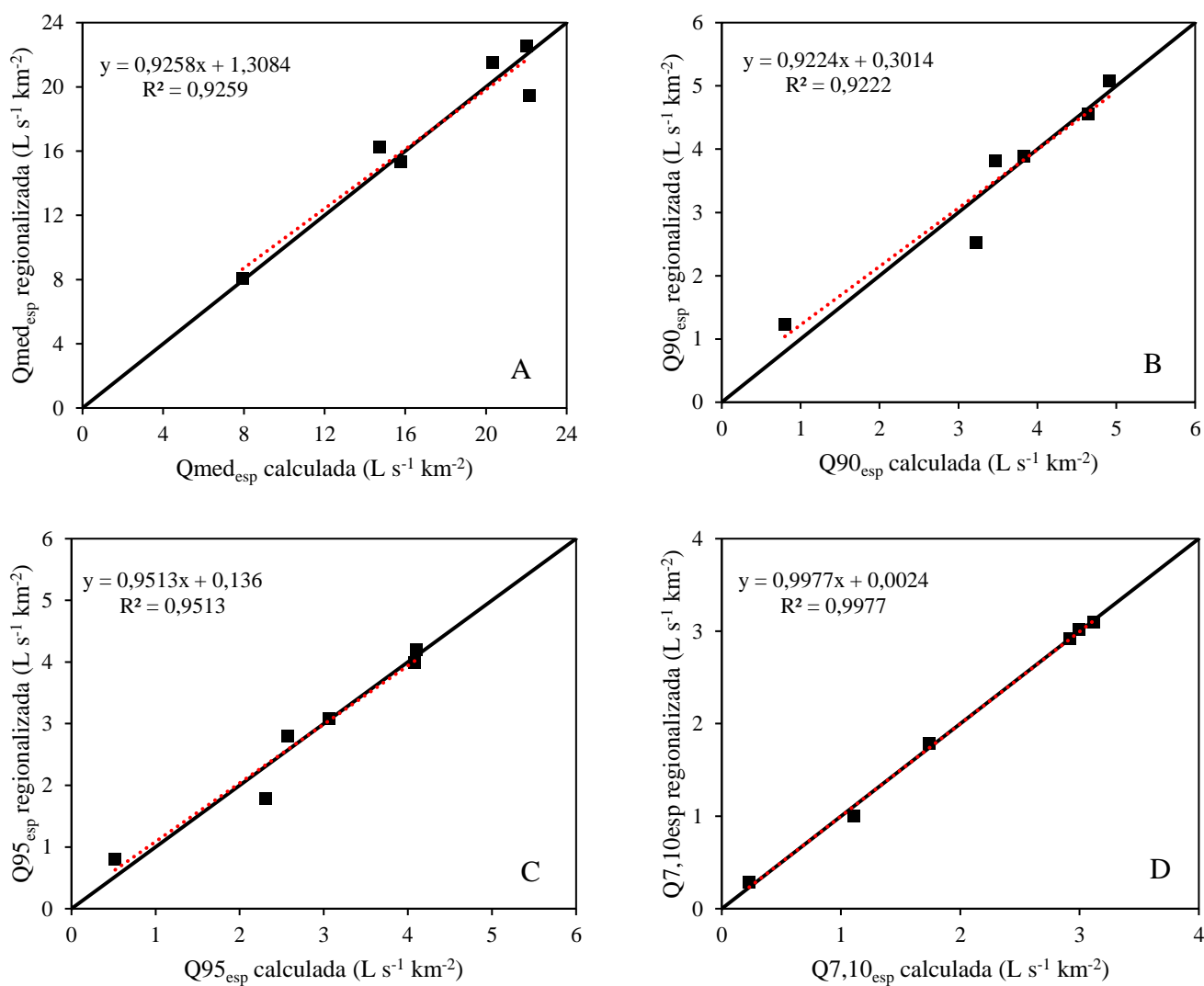


Figura 25 - Relação Q_{med_esp} regionalizada e Q_{med_esp} calculada; B) Relação Q_{90_esp} regionalizada e Q_{90_esp} calculada; C) Relação Q_{95_esp} regionalizada e Q_{95_esp} calculada; D) Relação $Q_{7,10_esp}$ regionalizada e $Q_{7,10_esp}$ calculada, para a região hidrologicamente homogênea 2.

Fonte: Elaboração própria.

Região hidrologicamente homogênea 3

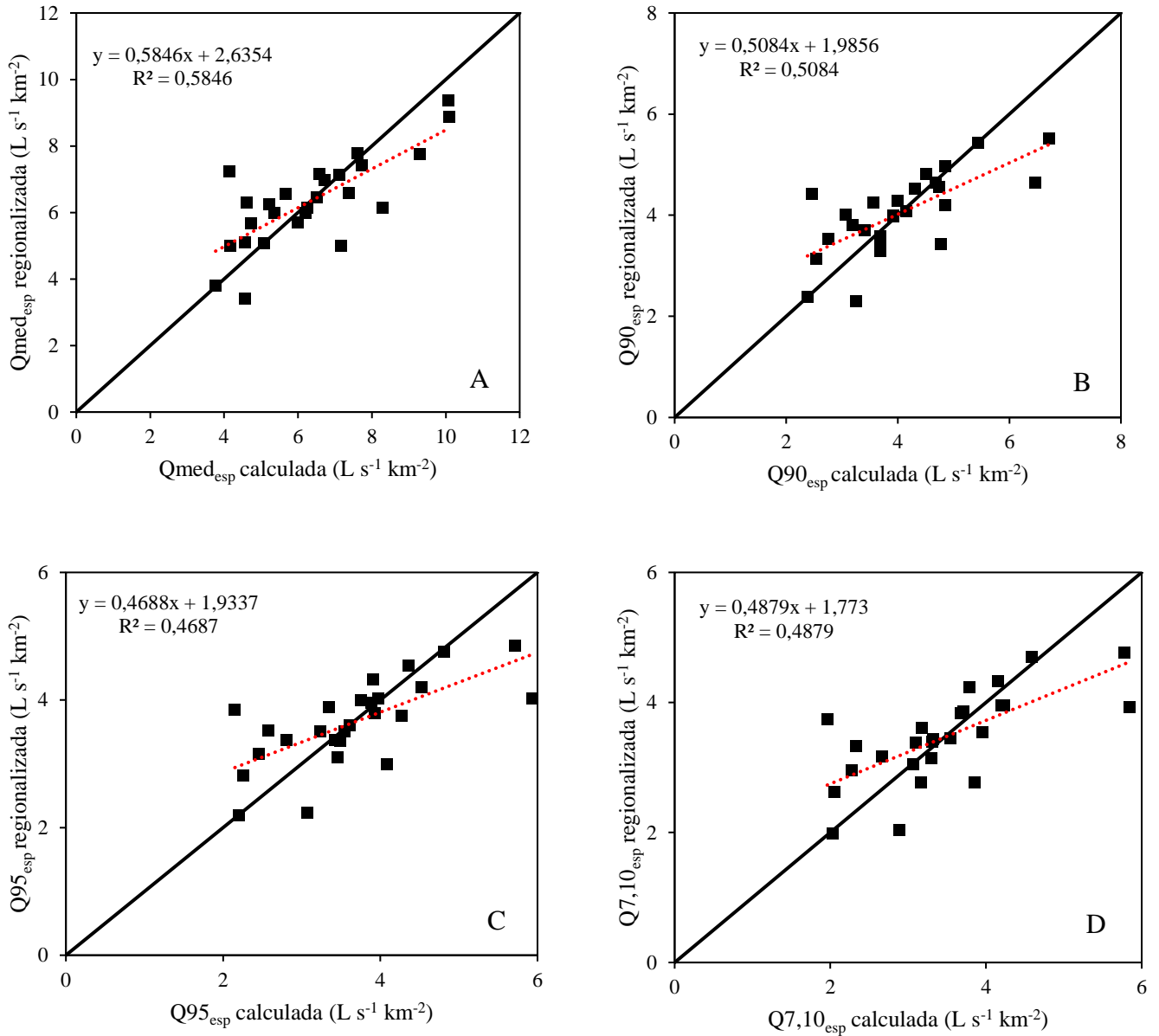


Figura 26 - Relação Q_{med_esp} regionalizada e Q_{med_esp} calculada; B) Relação Q_{90_esp} regionalizada e Q_{90_esp} calculada; C) Relação Q_{95_esp} regionalizada e Q_{95_esp} calculada; D) Relação $Q_{7,10_esp}$ regionalizada e $Q_{7,10_esp}$ calculada, para a região hidrologicamente homogênea 3.

Fonte: Elaboração própria.

Região hidrologicamente homogênea 4

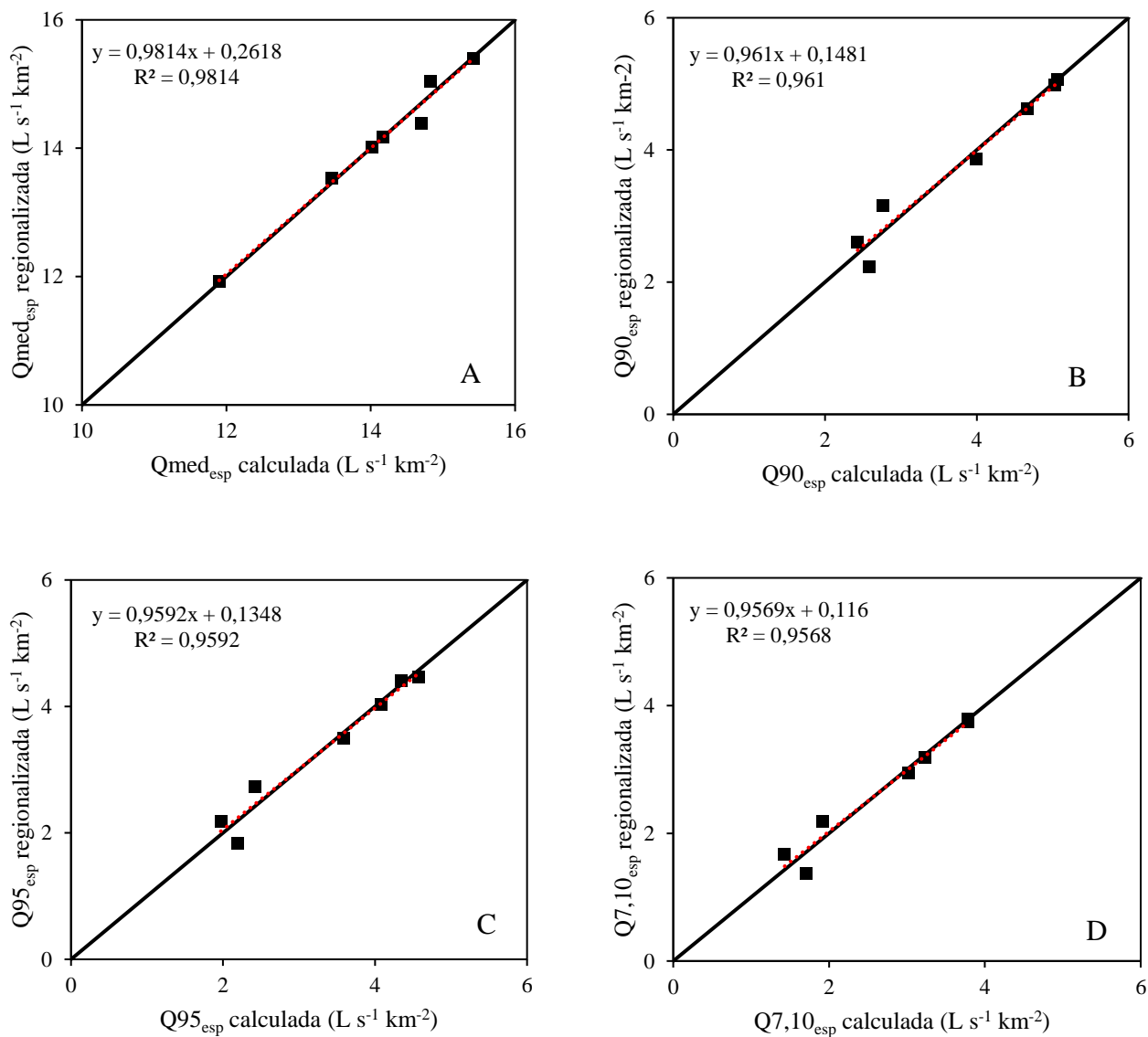


Figura 27 - Relação $Q_{med_{esp}}$ regionalizada e $Q_{med_{esp}}$ calculada; B) Relação $Q_{90_{esp}}$ regionalizada e $Q_{90_{esp}}$ calculada; C) Relação $Q_{95_{esp}}$ regionalizada e $Q_{95_{esp}}$ calculada; D) Relação $Q_{7,10_{esp}}$ regionalizada e $Q_{7,10_{esp}}$ calculada, para a região hidrologicamente homogênea 4.

Fonte: Elaboração própria.

Região hidrologicamente homogênea 5

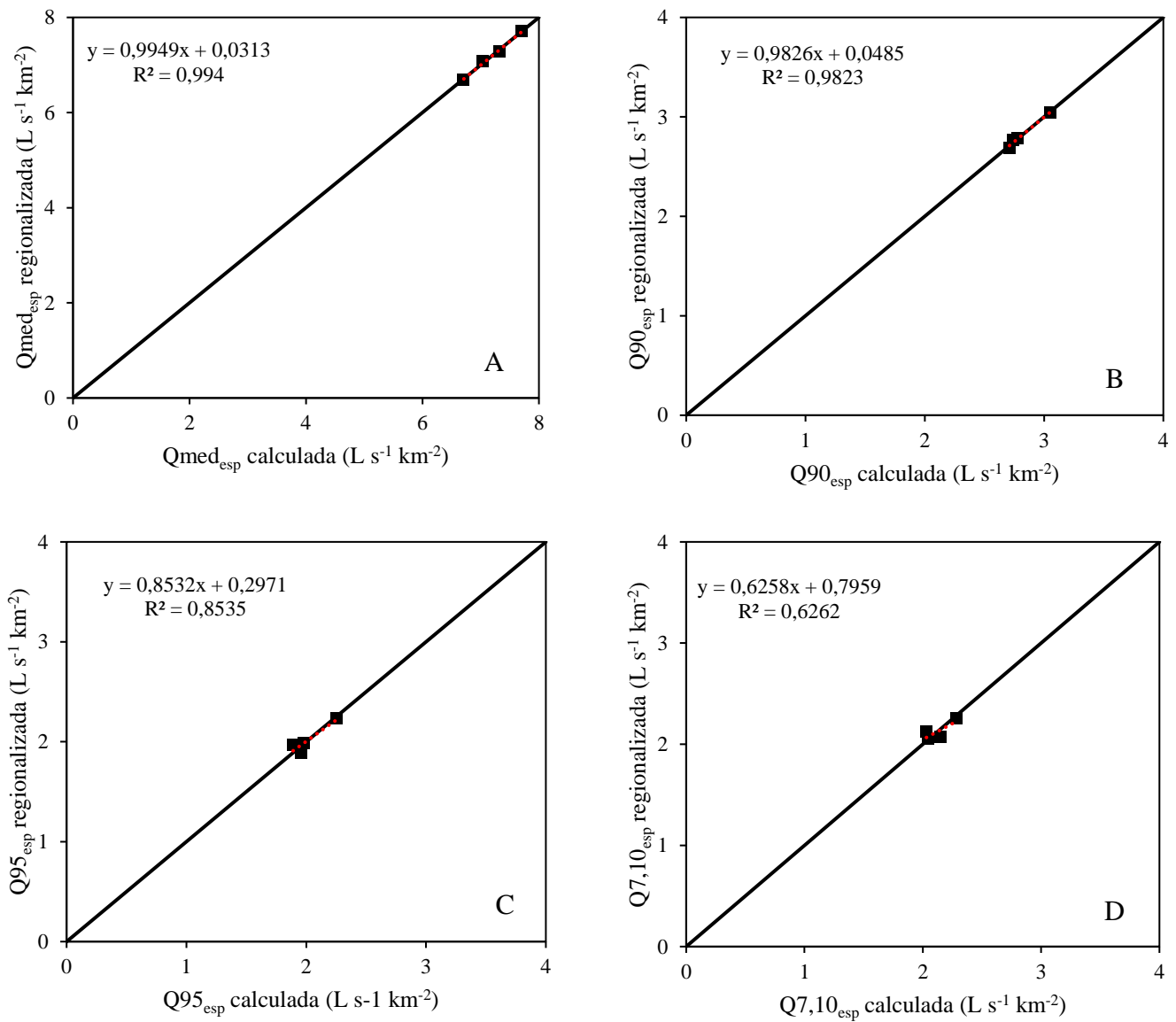


Figura 28 - Relação Q_{med_esp} regionalizada e Q_{med_esp} calculada; B) Relação Q_{90_esp} regionalizada e Q_{90_esp} calculada; C) Relação Q_{95_esp} regionalizada e Q_{95_esp} calculada; D) Relação $Q_{7,10_esp}$ regionalizada e $Q_{7,10_esp}$ calculada, para a região hidrologicamente homogênea 5.

Fonte: Elaboração própria.

Região hidrologicamente homogênea 6

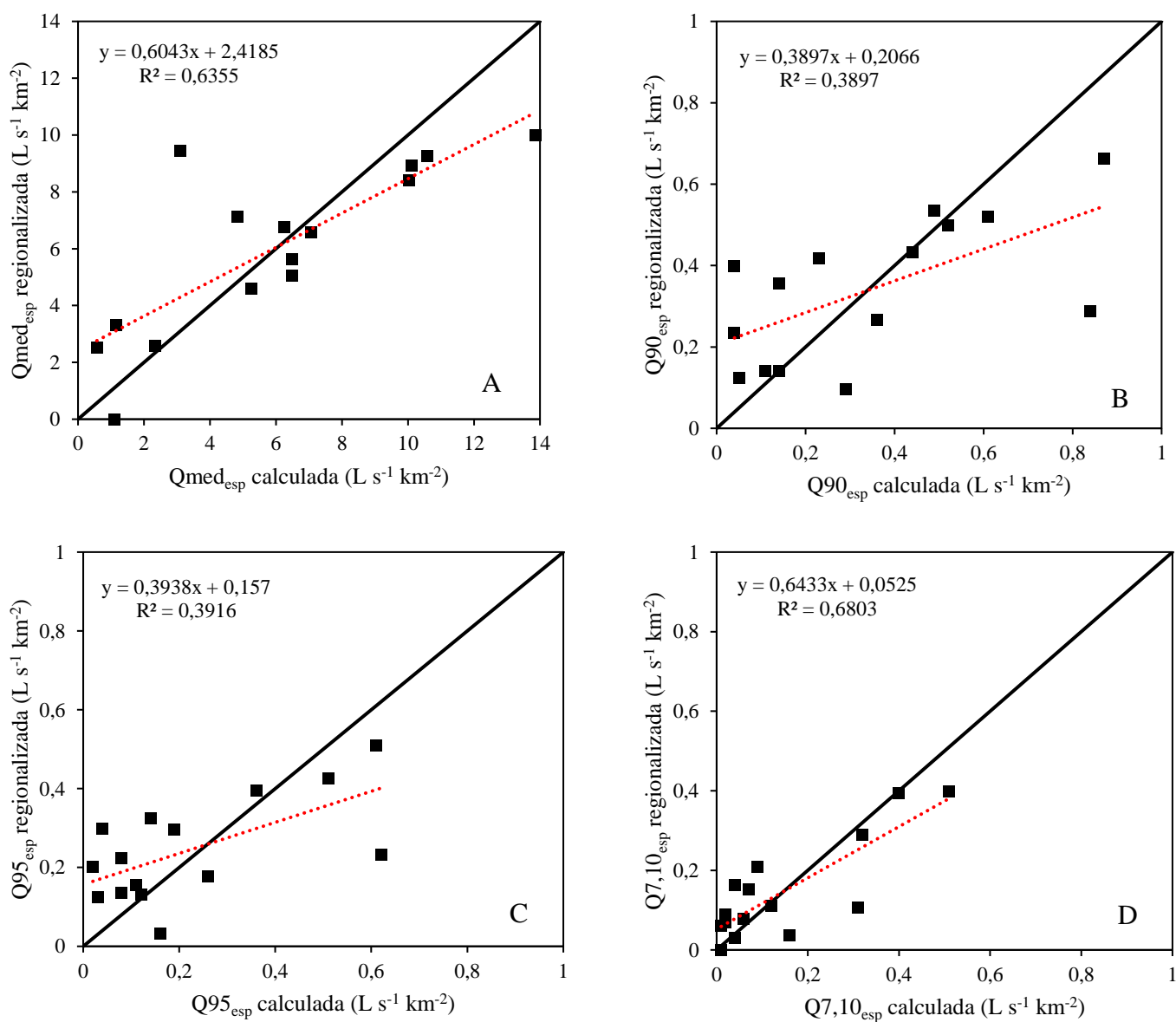


Figura 29 - Relação Q_{med_esp} regionalizada e Q_{med_esp} calculada; B) Relação Q_{90_esp} regionalizada e Q_{90_esp} calculada; C) Relação Q_{95_esp} regionalizada e Q_{95_esp} calculada; D) Relação $Q_{7,10_esp}$ regionalizada e $Q_{7,10_esp}$ calculada, para a região hidrologicamente homogênea 6.

Fonte: Elaboração própria.

Região hidrologicamente homogênea 7

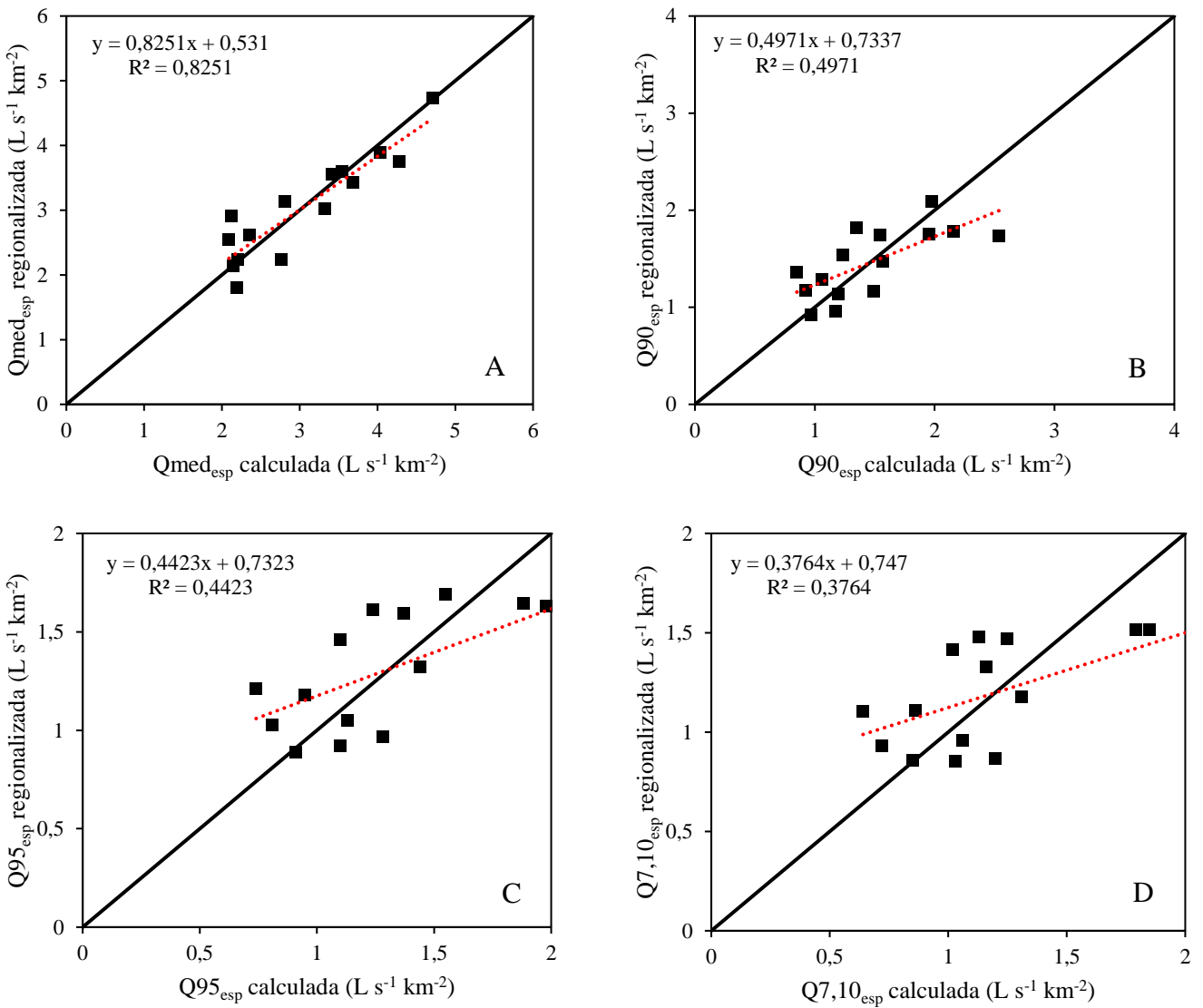


Figura 30 - Relação Q_{med_esp} regionalizada e Q_{med_esp} calculada; B) Relação Q_{90_esp} regionalizada e Q_{90_esp} calculada; C) Relação Q_{95_esp} regionalizada e Q_{95_esp} calculada; D) Relação $Q_{7,10_esp}$ regionalizada e $Q_{7,10_esp}$ calculada, para a região hidrologicamente homogênea 7.

Fonte: Elaboração própria.

ANEXO G. Intervalos de dados para a geração da matriz de confusão.Tabela 25 - Intervalos de vazão para as 8 classes segundo a Regra de Sturges, número de observações e frequência para os dados calculados e estimados para a vazão $Q_{med_{esp}}$.

Classes	Intervalo de Vazões	Vazão calculada		Vazão estimada	
		n° de Obs.	Freq. (%)	n° de Obs.	Freq. (%)
1	0,00 † 5,00	28	33,73	24	28,92
2	5,00 † 10,00	27	32,53	37	44,58
3	10,00 † 15,00	17	20,49	9	10,84
4	15,00 † 20,00	7	8,43	10	12,05
5	20,00 † 25,00	4	4,82	3	3,61

Fonte: Elaboração própria.

Tabela 26 - Intervalos de vazão para as 8 classes segundo a Regra de Sturges, número de observações e frequência para os dados calculados e estimados para a vazão $Q_{90_{esp}}$.

Classes	Intervalo de Vazões	Vazão calculada		Vazão estimada	
		n° de Obs.	Freq. (%)	n° de Obs.	Freq. (%)
1	0,00 † 2,78	42	50,60	39	46,99
2	2,78 † 5,56	32	38,55	37	44,58
3	5,56 † 8,33	8	9,65	5	6,03
4	8,33 † 11,11	0	0,00	1	1,20
5	11,11 † 13,89	1	1,20	1	1,20

Fonte: Elaboração própria.

Tabela 27 - Intervalos de vazão para as 8 classes segundo a Regra de Sturges, número de observações e frequência para os dados calculados e estimados para a vazão $Q_{95_{esp}}$.

Classes	Intervalo de Vazões	Vazão calculada		Vazão estimada	
		n° de Obs.	Freq. (%)	n° de Obs.	Freq. (%)
1	0,00 † 2,22	47	56,63	44	53,01
2	2,22 † 4,44	30	36,15	32	38,55
3	4,44 † 6,67	5	6,02	6	7,24
4	6,67 † 8,89	1	1,20	1	1,20
5	8,89 † 11,11	0	0,00	0	0,00

Fonte: Elaboração própria.

Tabela 28 - Intervalos de vazão para as 8 classes segundo a Regra de Sturges, número de observações e frequência para os dados calculados e estimados para a vazão $Q_{7,10_{esp}}$.

Classes	Intervalo de Vazões	Vazão calculada		Vazão estimada	
		n° de Obs.	Freq. (%)	n° de Obs.	Freq. (%)
1	0,00 † 2,22	49	59,04	50	60,24
2	2,22 † 4,44	30	36,15	30	36,15
3	4,44 † 6,67	3	3,61	2	2,41
4	6,67 † 8,89	1	1,20	1	1,20
5	8,89 † 11,11	0	0,00	0	0,00

Fonte: Elaboração própria.