

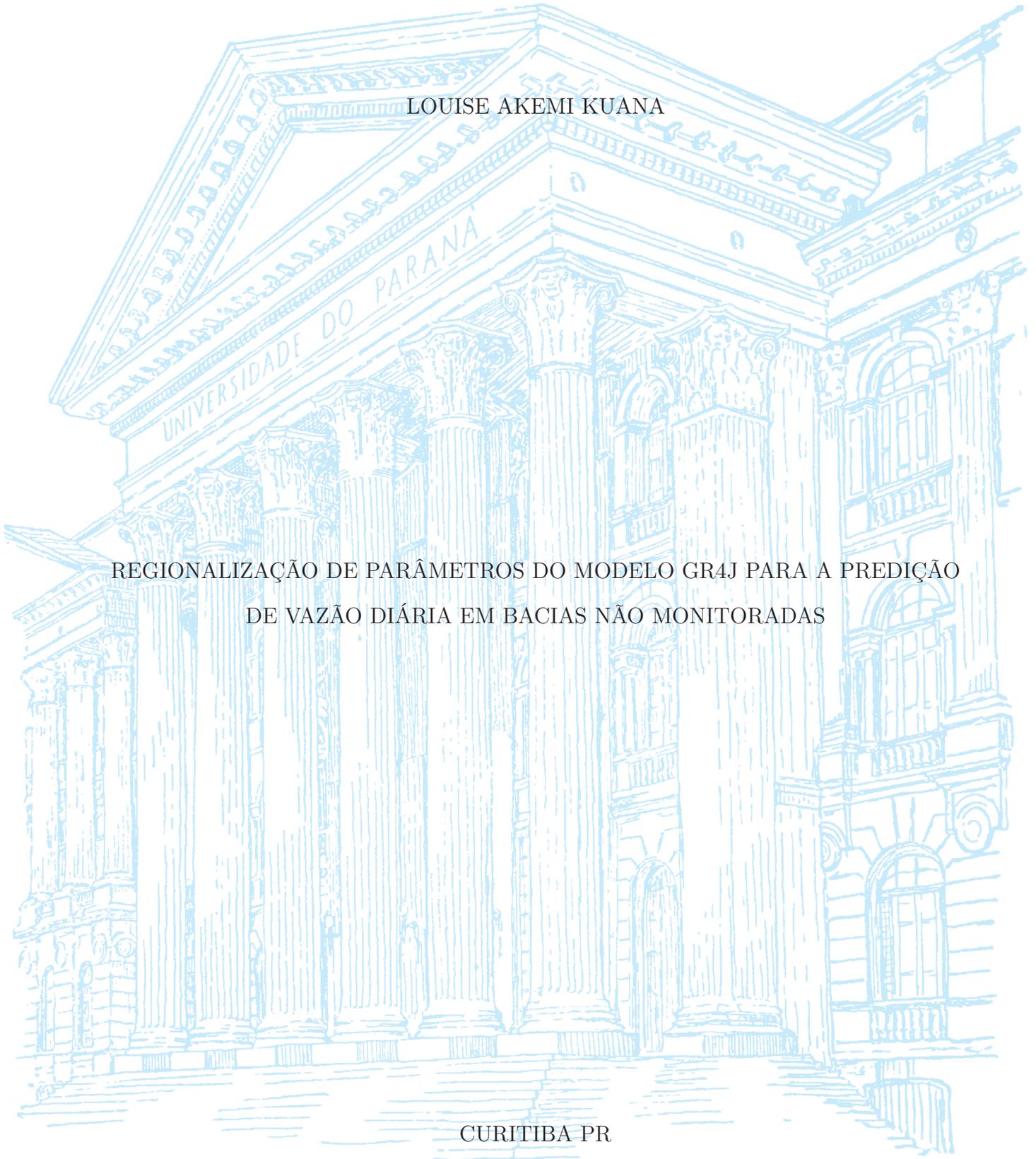
UNIVERSIDADE FEDERAL DO PARANÁ

LOUISE AKEMI KUANA

REGIONALIZAÇÃO DE PARÂMETROS DO MODELO GR4J PARA A PREDIÇÃO  
DE VAZÃO DIÁRIA EM BACIAS NÃO MONITORADAS

CURITIBA PR

2022



LOUISE AKEMI KUANA

REGIONALIZAÇÃO DE PARÂMETROS DO MODELO GR4J PARA A PREDIÇÃO  
DE VAZÃO DIÁRIA EM BACIAS NÃO MONITORADAS

Dissertação apresentada como requisito parcial à obtenção do grau de Mestre em Engenharia Ambiental no Programa de Pós-Graduação em Engenharia Ambiental, Setor de Tecnologia, da Universidade Federal do Paraná.

Área de concentração: *Engenharia Ambiental*.

Orientador: Emílio Graciliano Ferreira Mercuri.

CURITIBA PR

2022

Catálogo na Fonte: Sistema de Bibliotecas, UFPR  
Biblioteca de Ciência e Tecnologia

K95r

Kuana, Louise Akemi

Regionalização de parâmetros do modelo GR4J para a predição de vazão diária em bacias não monitoradas [recurso eletrônico] / Louise Akemi Kuana. – Curitiba, 2022.

Dissertação - Universidade Federal do Paraná, Setor de Tecnologia, Programa de Pós-Graduação em Engenharia Ambiental, 2022.

Orientador: Emílio Graciliano Ferreira Mercuri .

1. Hidrologia. 2. Bacias hidrográficas. 3. Levantamentos hidrográficos. 4. Engenharia agrônoma. I. Universidade Federal do Paraná. II. Mercuri, Emílio Graciliano Ferreira. III. Título.

CDD: 551.48

Bibliotecário: Elias Barbosa da Silva CRB-9/1894



## TERMO DE APROVAÇÃO

Os membros da Banca Examinadora designada pelo Colegiado do Programa de Pós-Graduação ENGENHARIA AMBIENTAL da Universidade Federal do Paraná foram convocados para realizar a arguição da dissertação de Mestrado de **LOUISE AKEMI KUANA** intitulada: **Regionalização de parâmetros do modelo GR4J para a predição de vazão diária em bacias não monitoradas**, sob orientação do Prof. Dr. EMILIO GRACILIANO FERREIRA MERCURI, que após terem inquirido a aluna e realizada a avaliação do trabalho, são de parecer pela sua APROVAÇÃO no rito de defesa.

A outorga do título de mestra está sujeita à homologação pelo colegiado, ao atendimento de todas as indicações e correções solicitadas pela banca e ao pleno atendimento das demandas regimentais do Programa de Pós-Graduação.

Curitiba, 14 de Fevereiro de 2022.

Assinatura Eletrônica

16/02/2022 07:09:53.0

EMILIO GRACILIANO FERREIRA MERCURI

Presidente da Banca Examinadora

Assinatura Eletrônica

16/02/2022 13:35:21.0

ELOY KAVISKI

Avaliador Externo (DEPARTAMENTO DE HIDRÁULICA E SANEAMENTO)

Assinatura Eletrônica

17/02/2022 16:55:47.0

JOEL AVRUCH GOLDENFUM

Avaliador Externo (UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL)

## AGRADECIMENTOS

Ao meu orientador, professor Emílio Graciliano Ferreira Mercuri, quem me orientou durante grande parte da minha graduação e novamente durante o mestrado. Agradeço pelo tempo que você ocupou para compartilhar seus conhecimentos, por ter confiado e acreditado no meu potencial.

À Universidade Federal do Paraná, ao Programa de Pós-Graduação em Engenharia Ambiental e aos professores, que permitiram a minha formação acadêmica e que proporcionaram um ensino de qualidade.

Ao Sistema de Tecnologia e Monitoramento Ambiental do Paraná - Simepar e à Companhia de Saneamento do Paraná (Sanepar), por terem tornado esta pesquisa realizável, por terem incentivado e apoiado este trabalho. Especialmente, ao meu supervisor Arlan Scortegagna Almeida e ao diretor-presidente Eduardo Alvim Leite, por terem me confiado esta oportunidade, pelo aperfeiçoamento profissional e pelo auxílio na execução deste trabalho. Aos colegas de trabalho da equipe de hidrologia do Simepar, Alex Bruno Kraemer e Gabriel Fiedler Bernardini.

Aos meus pais e meu irmão, pelo amor, carinho, paciência, apoio e incentivo incondicional. A presença e a educação que vocês me deram foram fundamentais, sem vocês não concluiria esta importante fase da minha vida.

Ao meu namorado, Eduardo Salmoria Fantin, pelo apoio, compreensão e carinho.

A todas as outras pessoas, que não foram poucas e nem menos importantes que ajudaram e que de alguma forma colaboraram na realização desta pesquisa.

## RESUMO

Este estudo visa avaliar a transferência de parâmetros do modelo GR4J (em francês, *modèle du Génie Rural à 4 paramètres Journalier*) calibrados em bacias hidrográficas instrumentalizadas para prever vazões diárias em bacias com pouca ou nenhuma informação hidrológica. Para isso, foi desenvolvido um conjunto de dados para o Estado do Paraná, onde foram englobadas informações atuais e históricas de diferentes órgãos, institutos e empresas que ao longo de décadas vêm investindo no monitoramento ambiental no Estado do Paraná. Além disso, foram aplicadas técnicas de controle de qualidade rigorosas para a recuperação do histórico de precipitação entre 1979 a 2020, e esforços manuais para a realização do hidrorreferenciamento das estações fluviométricas coletadas. Neste estudo foram empregados métodos para a regionalização dos parâmetros do GR4J, são eles: proximidade espacial simples, similaridade fisiográfica-climática e regressão por *Random Forest*. Adicionalmente, devido à importância da vazão Q95 para estimar a vazão máxima que pode ser outorgada dos corpos hídricos no Paraná, empregou-se a regressão direta da Q95 por meio do *Random Forest* para ser comparada com os métodos indiretos (regionalização de parâmetros do GR4J). Um conjunto de 100 bacias foram utilizadas para treinar os modelos de regionalização e outras 26 bacias foram usadas para avaliar e comparar o desempenho das regionalizações. Os resultados deste trabalho mostram que, em geral, todos os métodos de regionalização mostraram potencial e desempenhos positivos. Entretanto, de acordo com a avaliação feita para a amostra de 26 bacias, a regionalização baseada em similaridade fisiográfica-climática apontou ser o método mais robusto para a predição de vazão diária e da vazão de referência Q95 em bacias pouco ou não instrumentalizadas no Estado do Paraná.

Palavras-chave: Métodos PUBs. Regionalização. Modelo GR4J.

## ABSTRACT

This study aims to evaluate the transfer of parameters from the GR4J model (in French, *modèle du Génie Rural à 4 paramètres Journalier*) calibrated in hydrographic basins instrumented to predict daily flows in basins with little or no hydrological information. For this, a set of data for the State of Paraná was developed, which included current and historical information from different agencies, institutes and companies that for decades have been investing in environmental monitoring in the State of Paraná. In addition, rigorous quality control techniques were applied to recover the rainfall history between 1979 and 2020, and manual efforts were made to carry out the hydroreferencing of the collected fluviometric stations. In this study, methods were used to regionalize the parameters of the GR4J, which are: simple spatial proximity, physiographic-climatic similarity and regression by *Random Forest*. Additionally, due to the importance of the Q95 flow to estimate the maximum flow that can be granted from rivers in Paraná, direct regression of Q95 was used through *Random Forest* to be compared with indirect methods (regionalization of GR4J parameters). A set of 100 basins were used to train the regionalization models and another 26 basins were used to evaluate and compare the performance of regionalizations. The results of this work show that, in general, all regionalization methods showed positive potential and performance. However, according to evaluation carried out for the sample of 26 basins, regionalization based on physiographic-climatic similarity showed to be the most robust method for prediction of daily flow and reference flow Q95 in basins with little or no instrumentation in the State of Paraná.

Keywords: PUB methods. Regionalization. GR4J model.

## LISTA DE FIGURAS

3.1	Diagrama sintetizando o método. Fonte: Autor (2022).	23
4.1	Delimitação da área de estudo no estado do Paraná com base na rede hidrográfica ottocodificada do IAT. Fonte: IAT (2020).	25
4.2	Diagrama da coleta das estações.	26
4.3	Localização dos postos coletados.	28
4.4	Estações fluviométricas coletadas e bacias hidrográficas delineadas com a rede hidrográfica ottocodificada.	31
4.5	Temperaturas médias nos meses de julho e janeiro entre 2006 a 2016. Fonte: Barbieri et al. (2017).	33
4.6	Radiação solar média nos meses de inverno e de verão entre 2006 a 2016. Fonte: Barbieri et al. (2017).	34
4.7	Uso e ocupação do solo no estado do Paraná. Fonte: Souza et al. (2020).	35
4.8	Mapa de solos. Fonte: Embrapa (2020).	35
4.9	Localização das bacias hidrográficas selecionadas no estado do Paraná.	36
4.10	Disponibilidade de dados de vazão por estação. Quanto mais escura a cor verde, mais dados disponíveis há naquele ano.	37
5.1	Diagrama do modelo GR4J. Fonte: Adaptado de Perrin et al. (2003)	39
6.1	Diagrama sintetizando os métodos de regionalização. Fonte: Autor (2022).	46
7.1	Comparação entre o CABra (lado esquerdo), o conjunto do Paraná (central) e o CAMELS-BR (lado direito).	51
7.2	Desempenho do modelo calibrado no período da calibração e da validação no estado do Paraná.	52
7.3	Método <i>Elbow</i> para as bacias de treinamento com $K$ variando de 1 a 20.	53
7.4	Coefficiente de Silhueta para as bacias de treinamento com $K$ variando de 1 a 20.	53
7.5	Agrupamentos produzidos pelo método K-média para as 100 bacias de treinamento (lado esquerdo) e 26 bacias de validação (lado direito).	53
7.6	Relação dos grupos do conjunto de treinamento e os seus descritores.	54
7.7	Relação dos grupos do conjunto de treinamento e os seus parâmetros calibrados do GR4J.	54
7.8	Comparação dos hidrogramas e curvas de permanência entre bacias hidrográficas do grupo 0 pertencentes ao conjunto de dados do Paraná. Tanto os hidrogramas quanto as curvas de permanência foram produzidos com base nos registros de vazão observados no período de 1979 a 2020.	55

7.9	Comparação dos hidrogramas e curvas de permanência entre bacias hidrográficas do grupo 1 pertencentes ao conjunto de dados do Paraná. Tanto os hidrogramas quanto as curvas de permanência foram produzidos com base nos registros de vazão observados no período de 1979 a 2020.. . . . .	55
7.10	Comparação dos hidrogramas e curvas de permanência entre bacias hidrográficas do grupo 2 pertencentes ao conjunto de dados do Paraná. Tanto os hidrogramas quanto as curvas de permanência foram produzidos com base nos registros de vazão observados no período de 1979 a 2020.. . . . .	56
7.11	Comparação dos hidrogramas e curvas de permanência entre bacias hidrográficas do grupo 3 pertencentes ao conjunto de dados do Paraná. Tanto os hidrogramas quanto as curvas de permanência foram produzidos com base nos registros de vazão observados no período de 1979 a 2020.. . . . .	56
7.12	Comparação dos hidrogramas e curvas de permanência entre bacias hidrográficas do grupo 4 pertencentes ao conjunto de dados do Paraná. Tanto os hidrogramas quanto as curvas de permanência foram produzidos com base nos registros de vazão observados no período de 1979 a 2020.. . . . .	57
7.13	Comparação dos hidrogramas e curvas de permanência entre bacias hidrográficas do grupo 5 pertencentes ao conjunto de dados do Paraná. Tanto os hidrogramas quanto as curvas de permanência foram produzidos com base nos registros de vazão observados no período de 1979 a 2020.. . . . .	57
7.14	Avaliação do número de bacias doadoras com base nas medianas do logNSE durante o período de validação das simulações.. . . . .	58
7.15	Correlações dos desempenhos (logNSE) da simulação durante o período de validação entre os parâmetros calibrados e os parâmetros regionalizados pela similaridade fisiográfica-climática (superior esquerdo), pela proximidade espacial (superior direito), pelo <i>Random Forest I</i> (inferior esquerdo) e pelo <i>Random Forest II</i> (inferior direito).. . . . .	60
7.16	Gráficos chuva-vazão da estação Barbosa Ferraz (64659000) com as vazões observada e as simuladas utilizando os parâmetros calibrados (superior à esquerda), e os regionalizados pela proximidade (centro superior), pela similaridade (superior à direita), pelo <i>Random Forest I</i> (inferior à esquerda) e pelo <i>Random Forest II</i> (centro inferior).. . . . .	62
7.17	Correlações das estimativas da vazão de referência $Q_{95}$ .. . . . .	65
A.1	Disponibilidade de registros de vazão em anos e a frequência com que ocorrem dentro do conjunto de estações fluviométricas coletadas. . . . .	75
A.2	Disponibilidade de registros de vazão em anos.. . . . .	76
A.3	Disponibilidade de estações que possuem no máximo 10% de dados em cada ano.. . . . .	76
A.4	Disponibilidade de dados de vazão por estação. Quanto mais escura a cor verde mais dados disponíveis há naquele ano.. . . . .	77
A.5	Disponibilidade de dados de vazão por estação. Quanto mais escura a cor verde mais dados disponíveis há naquele ano.. . . . .	78

A.6	Disponibilidade de dados de vazão por estação. Quanto mais escura a cor verde mais dados disponíveis há naquele ano. . . . .	79
A.7	Disponibilidade de registros de precipitação em anos. . . . .	80
A.8	Disponibilidade de registros de precipitação em anos e a frequência com que ocorrem dentro do conjunto de estações fluviométricas coletadas. . . . .	80
A.9	Disponibilidade de estações que possuem no máximo 10% de dados em cada ano. . . . .	81
A.10	Coefficientes de Correlação de Pearson entre os descritores. . . . .	89

## LISTA DE TABELAS

4.1	Descrição dos índices fisiográficos das bacias hidrográficas. . . . .	33
4.2	Descrição dos índices climatológicos das bacias hidrográficas. . . . .	34
4.3	Descrição dos índices de Solo e Uso e Cobertura do Solo das bacias hidrográficas. . . . .	35
4.4	Estatísticas descritivas para o conjunto de dados do Paraná. . . . .	38
5.1	Descrições e intervalos para os parâmetros do modelo GR4J. . . . .	43
7.1	Correlação entre os parâmetros calibrados do modelo GR4J. . . . .	59
7.2	Valores medianos das estatísticas de erros calculadas para o período de validação. Em negrito estão evidenciados os melhores resultados para cada índice dentre os métodos de regionalização. . . . .	61
7.3	Correlação de Pearson (R) entre os parâmetros calibrados e os regionalizados. . . . .	61
7.4	Valores medianos das estatísticas de erros calculadas para as vazões de pico e de base. Em negrito estão evidenciados os melhores resultados para cada índice dentre os métodos de regionalização. . . . .	63
7.5	Valores medianos das estatísticas de erros calculadas para as vazões no período de recessão, o qual foi determinado pela série histórica observada. Em negrito estão evidenciados os melhores resultados para cada índice dentre os métodos de regionalização. . . . .	63
A.1	Estatísticas descritivas para as 284 bacias hidrográficas. . . . .	81

## LISTA DE ACRÔNIMOS

ANA	Agência Nacional de Águas e Saneamento Básico
AGUASPARANÁ	Instituto das Águas do Paraná
BDMEP	Banco de Dados Meteorológicos
CAMELS-BR	<i>Catchment Attributes and Meteorology for Large-sample Studies Brazil</i>
CABra	<i>Catchments Attributes for Brazil</i>
CAMELS	<i>Catchment Attributes and Meteorology for Large-sample Studies</i>
DE	<i>Differential Evolution</i>
Emater	Instituto de Assistência Técnica e Extensão Rural
EOD	<i>End of the day</i>
FIV	Fator de Inflação de Variância
GR	<i>Génie Rural</i>
GR4J	<i>Modèle du Génie Rural à 4 paramètres Journalier</i>
GRACE	<i>Gravity Recovery and Climate Experiment</i>
IAPAR	Instituto Agrônômico do Paraná
IAHS	Associação Internacional de Ciências Hidrológicas
IAT	Instituto Água e Terra
IDR-Paraná,	Instituto de Desenvolvimento Rural do Paraná - IAPAR - EMA- TER
INMET	Instituto Nacional de Meteorologia
logNSE	Logaritmo do coeficiente de Nash-Sutcliffe
MDE	Modelo Digital de Elevação
NSE	Coeficiente de Nash-Sutcliffe
PUB	<i>Predictions in Ungauged Basins</i>
Sanepar	Companhia de Saneamento do Paraná
SIH	Sistema de Informações Hídricas
Simepar	Sistema de Tecnologia e Monitoramento Ambiental do Paraná
STRM	<i>Shuttle Radar Topography Mission</i>
Suderhsa	Superintendência de Desenvolvimento de Recursos Hídricos e Saneamento Ambiental
SWAT	<i>Soil and Water Assessment Tool</i>
USGS	<i>United States Geological Survey</i>

## LISTA DE SÍMBOLOS

Variável	Unidade	Definição
$\lambda ET$	$\text{MJ m}^{-2} \text{d}^{-1}$	Densidade do fluxo de calor latente
$\Delta$	$\text{kPa}^\circ\text{C}^{-1}$	Inclinação da curva de pressão de vapor de saturação
$R_n$	$\text{MJ m}^{-2} \text{d}^{-1}$	Radiação solar
$G$	$\text{MJ m}^{-2} \text{d}^{-1}$	Densidade do fluxo de calor no solo
$\rho_a$	$\text{kg m}^{-3}$	Densidade média do ar a pressão constante
$c_p$	$\text{MJ kg}^{-1} \text{ }^\circ\text{C}^{-1}$	Calor específico do ar
$e_s$	$\text{kPa}$	Pressão de vapor de saturação média diária do ar
$e_a$	$\text{kPa}$	Pressão de vapor do ar média diária
$\gamma$	$\text{kPa }^\circ\text{C}^{-1}$	Constante psicrométrica
$r_a$	$\text{s m}^{-1}$	Resistência aerodinâmica
$r_c$	$\text{s m}^{-1}$	Resistência da superfície da cobertura vegetal
$\hat{m}(s_0)$		Valor predito da variável de interesse no ponto $s_0$ pelo método de regressão
$\hat{e}(s_0)$		Valor dos resíduos interpolados no ponto $s_0$ pelo método de geoestatística
$\lambda_i$		Pesos de interpolação
$\hat{z}(s_0)$		Valor resultante do valor predito da variável de interesse no ponto $s_0$
$e(s_i)$		Valores observados da variável de interesse no ponto $s_i$
$P$		Série de precipitação média na bacia
$P_n$		Entrada líquida de precipitação
$E$		Série de evapotranspiração potencial
$E_n$		Entrada líquida de evapotranspiração potencial
$P_s$		Parcela da precipitação líquida que infiltra no solo
$t$		Passo de tempo diário
$S$		Conteúdo de água no reservatório de produção
$E_s$		Taxa de evapotranspiração potencial no reservatório de produção
$Perc$		Percolação do reservatório de produção
$P_r$		Conteúdo de água a ser propagado
$HU_1$		Primeiro hidrograma unitário
$HU_2$		Segundo hidrograma unitário
$SH_1$		Curva-s do primeiro hidrograma unitário

$SH_2$		Curva-s do segundo hidrograma unitário
$Q_9$		Parcela de água resultante da propagada pelo primeiro hidrograma unitário
$Q_1$		Parcela de água resultante da propagada pelo segundo hidrograma unitário
$F$		Troca subterrânea de água
$R$		Conteúdo de água no reservatório de propagação
$Q_r$		Parcela de água resultante da propagada pelo reservatório de propagação
$Q_d$		Parcela de água resultante segundo hidrograma unitário sujeita às trocas subterrâneas
$Q$		Vazão resultante da simulação
$X_1$	mm	Capacidade total do reservatório de produção
$X_2$	mm/dia	Coefficiente de troca de água subterrânea
$X_3$	mm	Capacidade total do reservatório de propagação
$X_4$	dias	Atraso de resposta à precipitação
$X_{n,i}^G$		Matriz de população de tamanho N da geração G com i-énésimos parâmetros
$M_{n,i}^{G+1}$		Matriz de população com mutação de tamanho N da geração G+1 com i-énésimos parâmetros
$f$		Fator de mutação
$T_{n,i}^{G+1}$		Matriz de população teste de tamanho N da geração G+1 com i-énésimos parâmetros
$CR$		Constante de probabilidade de recombinação
$I_{rand}$		Valor escolhido de forma aleatória entre os D parâmetros
$x_n^{G+1}$		Indivíduo n da geração G+1
$dist(a,b)$		Distância Euclidiana entre os atributos da bacia a e b
$S_i$		Coefficiente de Silhueta
$a_p$		Distância Euclidiana média de um ponto p com todos os pontos pertencentes ao mesmo grupo
$b_p$		Distância média do ponto p com relação a todos os pontos pertencentes ao grupo vizinho mais próximo
$SSE$		Soma dos erros ao quadrado
$D_H$		Distância de Haversine
$r$		Raio da Terra
$:=$		Operador de atribuição

## SUMÁRIO

<b>1</b>	<b>INTRODUÇÃO</b> . . . . .	<b>15</b>
1.1	OBJETIVOS . . . . .	16
1.2	ORGANIZAÇÃO DO DOCUMENTO. . . . .	17
<b>2</b>	<b>REVISÃO BIBLIOGRÁFICA.</b> . . . . .	<b>18</b>
2.1	MÉTODOS PUBS E REGIONALIZAÇÃO. . . . .	19
2.2	MODELOS HIDROLÓGICOS . . . . .	20
2.3	TRABALHOS DESENVOLVIDOS ANTERIORMENTE NO ESTADO DO PARANÁ. . . . .	21
<b>3</b>	<b>ROTEIRO DO MÉTODO</b> . . . . .	<b>23</b>
<b>4</b>	<b>PREPARAÇÃO DOS DADOS PARA A ÁREA DE ESTUDO.</b> . . . . .	<b>25</b>
4.1	ÁREA DE ESTUDO . . . . .	25
4.2	CONJUNTO DE DADOS DO ESTADO DO PARANÁ. . . . .	25
4.2.1	Coleta de Dados Hidrometeorológicos . . . . .	26
4.2.2	Recuperação do Histórico de Precipitação (1979-2020) . . . . .	27
4.2.3	Delineamento das Bacias Hidrográficas . . . . .	30
4.2.4	Evapotranspiração Potencial . . . . .	31
4.2.5	Descritores das Bacias Hidrográficas. . . . .	32
4.2.6	Índices Fisiográficos . . . . .	32
4.2.7	Índices Climatológicos. . . . .	33
4.2.8	Índices de Solo e Uso e Cobertura do Solo . . . . .	34
4.2.9	Seleção de Bacias Hidrográficas . . . . .	36
<b>5</b>	<b>MODELO GR4J</b> . . . . .	<b>39</b>
5.1	CALIBRAÇÃO E VALIDAÇÃO . . . . .	43
<b>6</b>	<b>MÉTODOS DE REGIONALIZAÇÃO</b> . . . . .	<b>45</b>
6.1	SIMILARIDADE FISIOGRÁFICA-CLIMÁTICA . . . . .	47
6.2	PROXIMIDADE ESPACIAL SIMPLES. . . . .	49
6.3	REGRESSÃO . . . . .	49
6.4	ESTIMATIVA DA VAZÃO Q95 . . . . .	49
<b>7</b>	<b>RESULTADOS E DISCUSSÕES</b> . . . . .	<b>51</b>
7.1	CONJUNTO DE DADOS DO ESTADO DO PARANÁ. . . . .	51
7.2	DESEMPENHO DO MODELO GR4J . . . . .	51
7.3	DESEMPENHO DOS MÉTODOS DE REGIONALIZAÇÃO. . . . .	52
7.3.1	Similaridade Fisiográfica-Climática . . . . .	52

7.3.2	Proximidade Espacial Simples . . . . .	57
7.3.3	Regressão dos Parâmetros do Modelo GR4J . . . . .	58
7.3.4	Comparando os Métodos de Regionalização. . . . .	59
7.3.5	Estimação da Vazão Q95 . . . . .	63
<b>8</b>	<b>CONCLUSÃO . . . . .</b>	<b>66</b>
	<b>REFERÊNCIAS . . . . .</b>	<b>69</b>
	<b>APÊNDICE A – CONJUNTO DE DADOS DO ESTADO DO PARANÁ . . . . .</b>	<b>75</b>
A.1	DADOS FLUVIOMÉTRICOS . . . . .	75
A.2	DADOS PLUVIOMÉTRICOS . . . . .	79
A.3	DESCRITORES . . . . .	81
	<b>APÊNDICE B – FERRAMENTAS E ALGORITMOS . . . . .</b>	<b>90</b>

## 1 INTRODUÇÃO

O crescimento da população mundial, a qual ultrapassou a faixa de 7 bilhões de indivíduos na primeira década do século XXI, conduz a um dos problemas mais atuais que a humanidade terá que enfrentar que é a garantia da oferta de água para toda a população (FAO, 2021). Somando-se ao aumento da demanda pode ocorrer a ausência de chuvas, como aconteceu nos anos de 2020 e 2021 em que os baixos volumes de chuvas diminuíram a disponibilidade de água nos reservatórios e mananciais de captações do Estado do Paraná prejudicando, assim, a geração de energia proveniente de hidrelétricas, a irrigação de lavouras e o abastecimento de água para a indústria e para os municípios.

Frente a escassez hídrica a Companhia de Saneamento do Paraná (Sanepar), a qual é responsável pelos serviços de abastecimento de água e tratamento de esgoto da grande maioria dos municípios paranaenses afetados, tem adotado medidas emergenciais de rodízio de abastecimento de água, tem orientado a comunidade ao uso consciente d'água e tem investido em pesquisas que auxiliem no gerenciamento e na segurança hídrica da Companhia.

Os modelos de chuva-vazão possuem grande importância para estudos de disponibilidade hídrica, desenvolvimento de projetos de barragens, monitoramento no nível de reservatórios, prognóstico de estiagem, prevenção de inundações e enchentes, como também podem amenizar os impactos negativos ao meio ambiente. O uso de modelos capazes de estimar vazões de forma acurada pode se tornar uma ferramenta aliada para auxiliar no desenvolvimento de estratégias, no planejamento e tomadas de decisões futuras, e para amenizar os riscos e os impactos negativos para a operação da Companhia e para o meio ambiente.

O modelo hidrológico, GR4J (em Francês, *modèle du Génie Rural à 4 paramètres Journalier*), proposto por Perrin et al. (2003) vêm sendo implementado em diferentes países, como a França (Oudin et al., 2008, 2010), Austrália (Pagano et al., 2010), Brasil (Neto et al., 2021), Coreia do Sul (Shin e Kim, 2016), México (Arsenault et al., 2019), Rússia (Ayzel et al., 2019), ou seja, em regiões com características climáticas, geológicas, pedológicas e de cobertura vegetal variadas. Esse modelo parcimonioso vêm apresentando resultados promissores e destaca-se pela dependência de poucos parâmetros e o uso de duas forçantes meteorológicas de escala diária, que são: as séries de precipitação média da bacia e de evapotranspiração potencial, sendo necessárias séries históricas de vazões observadas para a realização do ajuste de seus parâmetros. Todavia, essas informações nem sempre estão disponíveis em quantidade e qualidade para a bacia hidrográfica de interesse.

Em vista da ausência recorrente de dados na área da hidrologia, a Associação Internacional de Ciências Hidrológicas (sigla em inglês, IAHS) difundiu entre 2003 a 2012 o movimento que ficou conhecido como a década PUB (em inglês, *Predictions in Ungauged Basins*), onde buscou-se iniciativas para o aprimoramento e o desenvolvimento de métodos que sejam capazes de predizerem ou preverem de forma confiável variáveis hidrológicas em bacias com pouca ou nenhuma informação hidrológica; promoverem avanços na compreensão das relações entre os processos hidrológicos e os descritores físicos, climáticos e geográficos das bacias hidrográficas, e estimularem progressos no monitoramento e no controle de qualidade dos dados medidos (Sivapalan, 2003).

Os métodos de regionalização dependentes de modelos de chuva-vazão, segundo Razavi e Coulibaly (2013), compreendem técnicas de transferência de parâmetros calibrados em bacias instrumentalizadas (bacias doadoras) para bacias não instrumentalizadas (bacias alvos). O estudo recente realizado por Arsenault et al. (2019) apresenta três técnicas baseadas em similaridade física, proximidade espacial e regressão, para se estimar os parâmetros de três modelos hidrológicos distintos, com o propósito de prever vazões contínuas em bacias hidrográficas que não possuem monitoramento. Esse estudo foi realizado no México, uma região que pode ser considerada heterogênea em múltiplos pontos de vista (Ex: climatologia e pedologia), e aponta que esses métodos de regionalização mostram ser mais promissores em regiões úmidas quando comparadas com as regiões áridas.

Embora muitos avanços já tenham sido realizados nessa área da hidrologia, mesmo após 10 anos da conclusão da década PUB ainda existem incertezas proveniente das estruturas dos modelos, dos métodos e dos procedimentos para se estimar vazões em bacias não instrumentalizadas (Guo et al., 2020). Parte disso ocorre devido à singularidade de cada região no globo, a qual diz respeito não só a unicidade de cada local como a questão da disponibilidade de informações (Ex. características descritivas das bacias e a disponibilidade de dados hidrometeorológicos), acrescentando-se há a complexidade dos sistemas hidrológicos, os quais podem ser vistos como o resultado de diferentes processos que dependem da escala temporal e espacial e das interações entre o clima, a vegetação, a topografia e o solo (Blöschl et al., 2013; Hrachowitz et al., 2013), e que, portanto, tornam a tarefa de estimar informações hidrológicas em bacias com poucas ou nenhuma informação desafiante.

A área de estudo escolhida para este trabalho é o Estado do Paraná, que está localizado na região sul do Brasil e a sua área é aproximadamente de 199.315 km<sup>2</sup>. O Paraná é banhado, principalmente, pelo rio Iguaçu, rio Paraná, rio Paranapanema, rio Tibagi, rio Ivaí e rio Piquiri, os quais proporcionam o abastecimento de água à população e às indústrias, geração de energia e irrigação de lavouras. O Estado do Paraná enfrentou uma de suas piores secas de sua história nos anos de 2020 e 2021 (Juliani et al., 2020), como citado inicialmente. Em virtude disso, destaca-se a importância de conhecermos as informações hidrometeorológicas observadas que estão disponíveis e, assim, analisarmos, compreendermos e estudarmos de forma mais aprofundada o Estado que apresenta tamanho potencial hídrico.

## 1.1 OBJETIVOS

Esta dissertação tem como objetivo principal a transferência de parâmetros do modelo GR4J (em francês, *modèle du Génie Rural à 4 paramètres Journalier*) calibrados em bacias hidrográficas instrumentalizadas para prever vazões diárias em bacias com pouca ou nenhuma informação hidrológica. A performance dos diferentes métodos aqui implementados é verificada em bacias paranaenses que possuem histórico de registros de dados hidrometeorológicos.

Os objetivos específicos são os seguintes:

- Empregar métodos de transferência dos conjuntos de parâmetros calibrados do modelo GR4J por meio de técnicas de regionalização baseadas em distância espacial, similaridade fisiográfica-climática e regressão não linear, e

- Comparar o desempenho dos métodos de regionalização de parâmetros com os métodos de regressão voltados à estimação da vazão de referência Q95.

## 1.2 ORGANIZAÇÃO DO DOCUMENTO

O trabalho está dividido em oito capítulos. O primeiro capítulo contém a introdução e os objetivos, onde estão expostos as motivações e a descrição dos propósitos e da proposta do trabalho. No segundo capítulo tem-se a revisão bibliográfica sobre os métodos PUBs e como os modelos hidrológicos podem auxiliar na predição de vazões diárias simuladas em bacias não instrumentalizadas. O terceiro capítulo inicia apresentando e descrevendo de forma sucinta o roteiro do método aqui aplicado. No quarto capítulo, é descrita a área de estudo e preparação do conjunto de dados. Depois, no quinto capítulo, tem-se a apresentação do modelo chuva-vazão e dos métodos de calibração e validação. O sexto capítulo tem-se as técnicas de regionalização que foram utilizadas. No sétimo capítulo são mostrados os resultados e discussões. Por último, no oitavo capítulo, tem-se a conclusão do trabalho e sugestões futuras de pesquisa. Adicionalmente, há dois apêndices mostrando com mais detalhes o conjunto de dados do Estado do Paraná e referências dos algoritmos utilizados para a implementação dos métodos.

## 2 REVISÃO BIBLIOGRÁFICA

A cada ano com base no consumo anual da humanidade é calculado o Dia da Sobrecarga da Terra (em inglês, *Earth Overshoot Day*). A chegada desse dia demarca o esgotamento dos serviços e dos recursos naturais que o planeta pode gerar e regenerar dentro do prazo de um ano. Em 2021 foi no dia 29 do mês de julho, ou seja, a partir desse dia até o final do ano a humanidade gerou dívidas com o planeta. De acordo com Global Footprint Network (2021), a demanda crescente por recursos naturais e a geração de poluentes faz com que a cada ano esse dia chegue mais cedo. A tendência do consumo crescente combinada com eventos extremos de escassez, sejam originados por causas naturais ou por consequência de atividades antropogênicas, podem potencializar ainda mais a ausência desses recursos.

A água é um recurso natural limitado e essencial para a vida. Por conseguinte, a outorga de direito de uso de recursos hídricos e o Plano Nacional de Segurança Hídrica (ANA, 2019) são instrumentos que auxiliam para que seja possível alcançar os objetivos da Política Nacional de Recursos Hídricos do Brasil (Brasil, 1997), em que é enaltecida a importância do consumo consciente d'água, da promoção de ações que assegurem a disponibilidade hídrica em quantidade e qualidade para a geração atual e futuras e do planejamento de estratégias de abastecimento hídrico para a garantia da segurança hídrica. Segundo Bigas (2013), a condição de segurança hídrica é alcançada quando é possível garantir: o acesso à água para as necessidades básicas e para o bem-estar da população; água adequada para a indústria, transporte, turismo e os setores de produção de energia e de alimentos; a preservação de ecossistemas e seus serviços; e a resiliência à poluição e a eventos de extremos, como de inundações e secas.

Uma gestão adequada dos recursos hídricos necessita do conhecimento da variabilidade dos regimes hidrológicos, sobretudo nos períodos de enchente ou de seca (Noumon et al., 2020). Os modelos de chuva-vazão são ferramentas que permitem compreender e simular os processos hidrológicos do passado, do presente e de cenários futuros de uma bacia hidrográfica (Tucci, 1998; Saha e Zeleke, 2015). Dessa maneira, podem ser tornar instrumentos importantes para a gestão hídrica. Eduardo et al. (2016), por exemplo, atestaram a viabilidade do uso do modelo hidrológico *Soil and Water Assessment Tool* (SWAT) na bacia do rio das Mortes (Minas Gerais) para auxiliar na gestão e no planejamento dos recursos hídricos.

Os modelos de chuva-vazão estimam a vazão utilizando a precipitação e outras forçantes climáticas ou meteorológicas, e geralmente é feita a identificação dos parâmetros que melhor reproduzem o comportamento hidrológico da bacia a partir da vazão observada (Blöschl et al., 2013). Todavia, muitas vezes as incertezas e as limitações de um modelo podem estar relacionadas com as informações que são inseridas ou utilizadas para o ajuste do mesmo. A carência de informações hidrometeorológicas é recorrente e em suma é prejudicial para a gestão dos recursos hídricos (Machado et al., 2017). Para tanto existem métodos que estão sendo desenvolvidos para amenizar os riscos e os impactos negativos da outorga de água nos mananciais de superfície e, assim, garantir a oferta de água para o abastecimento humano e as atividades produtivas.

## 2.1 MÉTODOS PUBS E REGIONALIZAÇÃO

Os métodos PUBs (em inglês, *Predictions in Ungauged Basins*) surgiram da necessidade de se prever ou prever de forma confiável variáveis hidrológicas em bacias que possuem poucas ou nenhuma informação hidrometeorológica, denominadas como não instrumentalizadas ou não monitoradas. De acordo com Hrachowitz et al. (2013), tais métodos propõem extrair o máximo possível de informações dos dados disponíveis em uma região e para os casos em que não haja dados suficientes pretendem investigar, explorar e avaliar a aplicabilidade de novos meios para adquirir essas informações. Um exemplo do avanço de tecnologias voltadas para a observação de águas subterrâneas são os resultados da missão GRACE (em inglês, *Gravity Recovery and Climate Experiment*), em que são utilizadas duas espaçonaves para detectar a variação do campo gravitacional ao longo do tempo em uma dada região, e a partir dessas mudanças os cientistas são capazes de estimar a variação do nível do lençol freático (NASA, 2021).

Os métodos de regionalização visam transferir informações hidrológicas de uma ou mais bacias hidrográficas instrumentalizadas para outras bacias não instrumentalizadas (Oudin et al., 2010; Guo et al., 2020; Blöschl et al., 2013). De acordo com Razavi e Coulibaly (2013), a predição de vazões contínuas em bacias não instrumentalizadas pode ser realizada por meio de modelos hidrológicos de base física, conceitual ou empírica. Por ora, a otimização dos modelos hidrológicos é realizada através de dados observados de vazão e, portanto, os parâmetros ou a estrutura das equações dos modelos que estimam respostas hidrológicas devem ser transferidos ou estimados para as bacias que não possuem essas medições.

As bacias hidrográficas são sistemas complexos que podem ser vistas como o resultado de diferentes processos, que dependem da escala do tempo e iterações entre o clima, a vegetação, a topografia e o solo (Blöschl et al., 2013). Intuitivamente, a escolha dos pares de bacias receptoras (não instrumentalizadas) e doadoras (instrumentalizadas) pode considerar a proximidade espacial e ou a similaridade fisiográfica-climática (Oudin et al., 2008, 2010; He et al., 2011; Razavi e Coulibaly, 2013; Guo et al., 2020). De modo que, ambas fazem parte das técnicas de regionalização baseadas em distâncias (He et al., 2011).

A similaridade fisiográfica-climática identifica as bacias hidrográficas doadoras que possuem maiores semelhanças físicas, climáticas e geográficas com as bacias receptoras. Essas semelhanças podem ser calculadas por meio de métricas de similaridade ou dissimilaridade, em geral podendo ser expressas como:

$$dist_{x,y} = \sqrt{(\mathbf{x} - \mathbf{y})^T S^{-1} (\mathbf{x} - \mathbf{y})}, \quad (2.1)$$

onde os vetores  $\mathbf{x}$  e  $\mathbf{y}$  contém dados das  $n$  características da bacia  $x$  e  $y$ , respectivamente. Por sua vez,  $S$  é uma matriz  $n \times n$  que define o tipo de distância que será adotado. Para a distância Euclidiana, distância Euclidiana padronizada e a distância de Mahalanobis as matrizes  $S$  correspondentes são: a matriz identidade, matriz diagonal da matriz de covariância e a matriz de covariância.

A aplicação do método da similaridade fisiográfica-climática implicitamente considera duas suposições. A primeira suposição é a de que, se há similaridade entre bacias, há respostas hidrológicas semelhantes. A segunda suposição é que a similaridade entre conjuntos de parâmetros calibrados do modelo hidrológico entre duas ou mais bacias hidrográficas pode refletir na semelhança de seu comportamento em relação à transformação de chuva em vazão (Oudin et al., 2008, 2010; Parajka et al., 2005; Blöschl et al., 2013).

Da mesma maneira, a proximidade espacial assume que as bacias vizinhas apresentam semelhanças climáticas, no tipo de solo, uso e cobertura do solo, na declividade, na altitude e em outras características (Arsenault et al., 2019). O estudo realizado por Kuentz et al. (2017), o qual utilizou mais de 35.000 bacias e 1.366 estações fluviométricas da Europa para explorar a correlação entre 16 índices diferentes de respostas do comportamento hidrológico (Ex. índice de vazão de base, Q5 e Q95) e 35 descritores físicos (Ex. área, declividade e índice de aridez), conclui que existem fortes conexões entre os descritores físicos e os índices de respostas do comportamento hidrológico.

De acordo com He et al. (2011), as técnicas de proximidade espacial podem ser subdivididas entre: proximidade simples e interpolação espacial. A proximidade simples identifica as bacias mais próximas, e frequentemente é calculada por meio da distância Euclidiana ou de Haversine, dependendo da projeção utilizada. Os parâmetros podem ser transferidos por completo ou estimados por meio de médias entre as bacias dentro de um certo raio (Guo et al., 2020). Dentre os métodos de interpolação espacial a krigagem é uma das técnicas mais populares conforme Razavi e Coulibaly (2013), onde os valores dos parâmetros são estimados através da média ponderada dos valores conhecidos na vizinhança.

Outra categoria de regionalização conhecida se baseia nas análises regionais de regressão (He et al., 2011). Os modelos de regressão múltipla lineares ou não-lineares buscam equacionar a relação entre a variável dependente (Ex. parâmetros do modelo hidrológico) com diferentes variáveis independentes (Ex. características descritivas). De acordo com Mohamed et al. (2019), devido à relação não linear e multidimensional entre diferentes características descritivas das bacias e os parâmetros dos modelos, vem crescendo a aplicação de métodos de regressão que utilizam técnicas de aprendizagem de máquina (Ex. *Random Forest*) para a realização da extrapolação dos parâmetros dos modelos hidrológicos.

Mais recentemente, segundo Guo et al. (2020), há uma nova classificação denominada como métodos de índices de respostas do comportamento hidrológicos (em inglês, *hydrological signature methods*). Os índices de respostas do comportamento hidrológico (Ex. curva de permanência) estimados para as bacias não instrumentalizadas são utilizados como referência para a identificação dos parâmetros do modelo que melhor reproduzem esses índices. Pinheiro e Naghettini (2013) mostraram a possibilidade de se obter performances aceitáveis para a predição de vazão contínua a partir da calibração do modelo RIO GRANDE utilizando as curvas de permanência sintéticas das bacias do rio Pará e rio Paraopeba. Por outro lado, Kim et al. (2017) ao usar as curvas de permanência regionalizadas para calibrar o modelo GR4J em um conjunto de 45 bacias Sul Coreanas observaram uma diminuição na capacidade de predição.

## 2.2 MODELOS HIDROLÓGICOS

A escolha da estrutura do modelo deve considerar a disponibilidade de dados, a restrição de recursos, a experiência do usuário e o propósito da modelagem (Blöschl et al., 2013). Os modelos distribuídos de base física são capazes de gerir um grande número de informações e criar conexões entre os parâmetros, que tentam refletir as características físicas de uma bacia, para descreverem diferentes processos físicos a serem simulados (Devia et al., 2015). Segundo Yadav et al. (2007), esses modelos são utilizados quando se busca um detalhamento espacial preciso, como a predição de áreas de inundação. Para

tal, esses modelos requerem uma quantidade de dados considerável e um conhecimento aprofundado sobre a área de estudo para a otimização dos parâmetros.

Os modelos conceituais são capazes de apresentar desempenhos semelhantes aos modelos de base física quando o enfoque do estudo é a resposta geral dos processos (Jaiswal et al., 2020). Adicionalmente, devido à facilidade de serem implementados e por demandarem um volume moderado de dados faz com que os modelos conceituais sejam largamente empregados em técnicas de regionalização (Razavi e Coulibaly, 2013; Oudin et al., 2008; Parajka et al., 2005; Zhang et al., 2015; Arsenault et al., 2019; Prieto et al., 2019; Song et al., 2019). Apesar desses modelos apresentarem capacidade para a predição de vazão contínua em bacias não monitoradas, Song et al. (2019) e Fowler et al. (2016) destacam algumas limitações ou incertezas na identificação dos parâmetros calibrados, uma vez que: estão sujeitos às condições do passado e podem não apresentar bem a situação do presente ou do futuro; e podem estar associados a função objetivo e ao método de calibração e validação escolhidos.

O uso de modelos empíricos (Ex: redes neurais artificiais, regressão múltipla linear e média móvel autoregressiva) requer uma quantidade semelhante ou menor de dados e podem apresentar estruturas mais modestas se comparados com os métodos citados anteriormente (Razavi e Coulibaly, 2013; Besaw et al., 2010). Recentemente, Waseem et al. (2015) apresentaram um novo conceito chamado de predição hidrológica de conjunto (em inglês, *ensemble hydrological prediction*), o qual agrega respostas de diferentes estimadores hidrológicos com o objetivo de diminuir a incerteza sobre as predições em bacias não monitoradas. A combinação feita por Razavi e Coulibaly (2016), a qual utilizou as respostas de dois modelos hidrológicos e dois modelos empíricos, apresentou ter uma performance superior quando comparada com o desempenho individual de cada modelo.

### 2.3 TRABALHOS DESENVOLVIDOS ANTERIORMENTE NO ESTADO DO PARANÁ

As questões apresentadas anteriormente mostram o quão desafiante, fundamental e atual é o tema de predição em bacias hidrográficas sem monitoramento (em inglês, *Predictions in Ungauged Basins*) para o desenvolvimento de atividades que estão dentro ou que se relacionam com a área da hidrologia. Tendo em vista a importância de se conhecer ou de ser capaz de estimar algumas variáveis hidrológicas para o melhor aproveitamento dos recursos hídricos, sabe-se de antemão a existência de projetos de pesquisa realizados no Estado do Paraná com o intuito de se estimar vazões contínuas e índices de resposta do comportamento hidrológico. Dentre os principais, destacam-se os trabalhos desenvolvidos por Kaviski et al. (2002) e Bazzo e Almeida (2016).

Kaviski et al. (2002) realizaram a coleta de dados fluviométricos, climáticos e fisiográficos de bacias hidrográficas presentes no Estado do Paraná, sul de São Paulo e norte de Santa Catarina. Nesse estudo fez-se o uso de 121 estações fluviométricas, as quais continham registros históricos de ao menos 10 anos e áreas de drenagens menores que 5.000 km<sup>2</sup>, para a aplicação de métodos de parametrização de diferentes índices fluviométricos, como por exemplo: vazões médias, mínimas e máximas. Contudo, constatou-se que as estimativas obtidas não foram convicidas o suficiente para determinar as estimativas seguras, isso se deve principalmente à complexidade e a carência de informações na época.

Bazzo e Almeida (2016) usaram algoritmos de Redes Neurais Artificiais (RNA) para regionalizar parâmetros de um modelo de chuva-vazão e, assim, estimar vazões contínuas em 15 bacias hidrográficas do interior do Paraná com base em séries históricas de pluviometria, temperatura e fluviometria. Concluiu-se positivamente que as técnicas empregadas por

Bazzo e Almeida (2016) alcançaram resultados satisfatórios para a estimação de vazões contínuas.

### 3 ROTEIRO DO MÉTODO

Os passos que se seguiram para a aplicação do método podem ser sintetizados em 5 etapas principais, como é apresentado no diagrama da Figura 3.1. O primeiro passo mostra a área de estudo como sendo o Estado do Paraná.

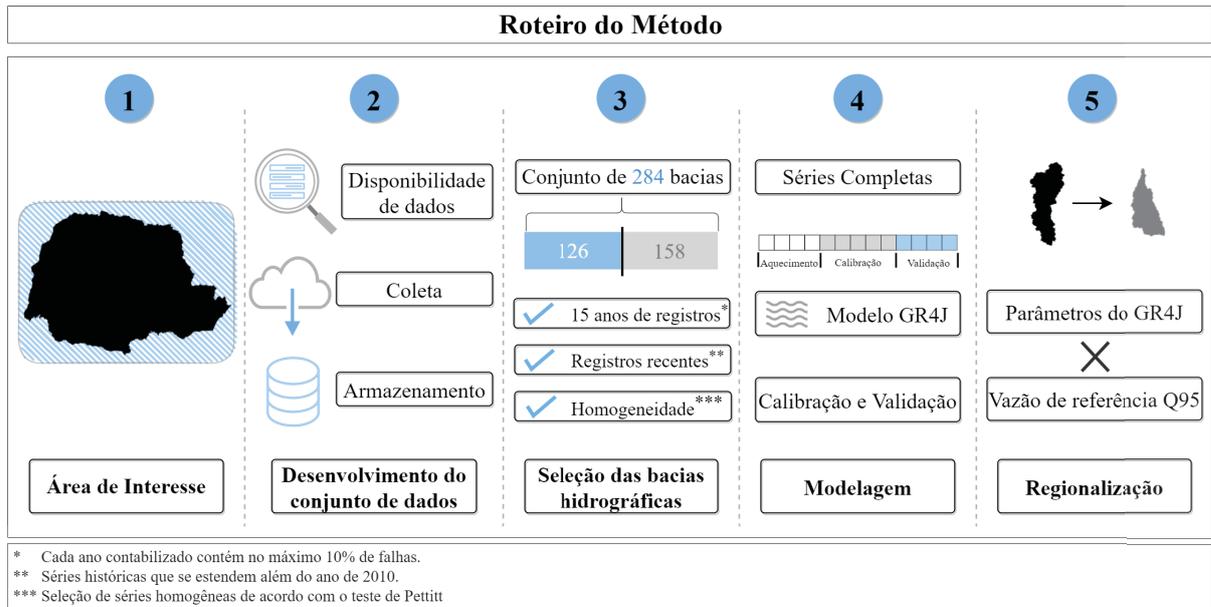


Figura 3.1: Diagrama sintetizando o método. Fonte: Autor (2022).

Continuando, no segundo passo, é explicitado como foi realizada a preparação do conjunto de dados hidrometeorológicos e os índices descritivos para a área de interesse. Nesse conjunto foram englobados dados provenientes de diferentes instituições e órgãos, que historicamente vêm contribuindo para o monitoramento ambiental no território paranaense. O conjunto de dados incluiu postos com observações fluviométricas que contivessem ao menos 5 anos de observações e desconsiderou as estações a jusante de regularizações. Ao final do segundo passo, alcançou-se um conjunto de dados com 284 bacias hidrográficas.

No terceiro passo, com o propósito de garantir a simultaneidade de dados, tanto de vazão quanto de precipitação, fez-se uma seleção de bacias hidrográficas com base no conjunto desenvolvido no passo anterior. Nessa amostra foram incluídas 126 bacias hidrográficas que possuem ao menos 15 anos de dados de vazão entre 1979 e 2020, sendo que cada ano contabilizado tem no máximo 10% de falhas, além disso preferiu-se que as séries históricas tivessem também dados mais recentes, as quais se estendiam além do ano de 2010, e limitou-se as séries históricas homogêneas que fossem aprovadas pelo teste de Pettitt (Pettitt, 1979).

A modelagem, o quarto passo, é a aplicação do modelo hidrológico GR4J, e em seguida a calibração e a validação do mesmo. No quinto passo tem-se o emprego dos métodos de regionalização dos parâmetros do modelo GR4J. Ainda no último passo, devido à importância da vazão Q95 para estimar a vazão máxima que pode ser outorgada dos corpos hídricos no Paraná, empregou-se a regressão direta da Q95 por meio do *Random Forest* para ser comparada com os métodos indiretos (regionalização de parâmetros do GR4J).

Os capítulos que se seguem apresentam mais informações e detalhes sobre os passos descritos resumidamente pelo diagrama da Figura 3.1.

## 4 PREPARAÇÃO DOS DADOS PARA A ÁREA DE ESTUDO

Neste capítulo é apresentada a área de estudo e detalha-se o processo de obtenção do conjunto de dados hidrometeorológicos com índices descritivos das bacias hidrográficas.

### 4.1 ÁREA DE ESTUDO

A área de estudo foi delimitada com base na rede hidrográfica ottocodificada do Estado do Paraná, a qual é disponibilizada via FTP pelo Instituto Água e Terra (IAT, atual Instituto das Águas do Paraná) no ano de 2020, e sobre um polígono de área retangular demarcado entre as latitudes  $22^{\circ} 15' 36''\text{S}$  e  $26^{\circ} 54' 00''\text{S}$ , e longitudes  $48^{\circ} 00' 00''\text{W}$  e  $54^{\circ} 42' 00''\text{W}$ . A área de estudo restringe-se à região, onde na Figura 4.1 há a sobreposição da rede hidrográfica ottocodificada com o polígono de área retangular. Portanto, a área de estudo se estende por todo o estado do Paraná, e avança em partes do estado de Santa Catarina e São Paulo, não abrangendo completamente as bacias do rio Paranapanema e do rio Paraná.

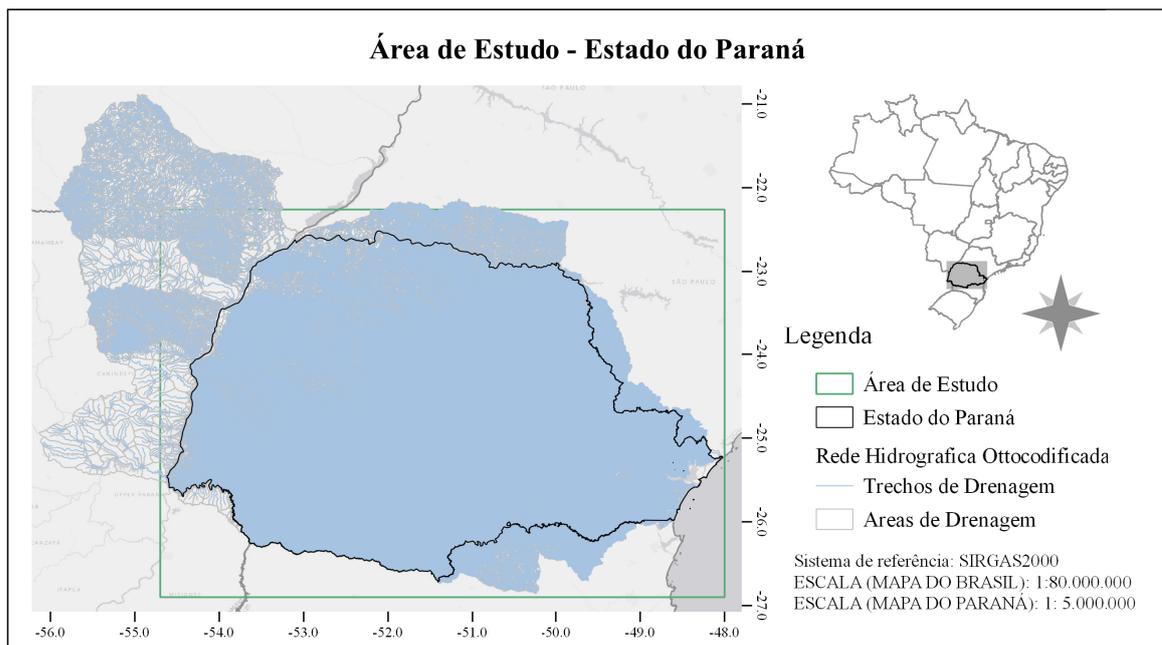


Figura 4.1: Delimitação da área de estudo no estado do Paraná com base na rede hidrográfica ottocodificada do IAT. Fonte: IAT (2020).

Como parte da elaboração deste trabalho, foi desenvolvido um conjunto de dados hidrometeorológicos com índices descritivos das bacias hidrográficas, o qual é apresentado a seguir.

### 4.2 CONJUNTO DE DADOS DO ESTADO DO PARANÁ

Os históricos de observações hidrometeorológicas no estado do Paraná são disponibilizados por instituições de nível nacional, como a Agência Nacional de Águas e

Saneamento Básico (ANA) via Portal HidroWEB e o Instituto Nacional de Meteorologia (INMET), e a nível estadual, como por exemplo: o Sistema de Tecnologia e Monitoramento Ambiental do Paraná (Simepar), o Instituto Água e Terra (IAT, atual Instituto das Águas do Paraná - AGUASPARANÁ e Superintendência de Desenvolvimento de Recursos Hídricos e Saneamento Ambiental - Suderhsa) e o Instituto de Desenvolvimento Rural do Paraná - IAPAR - EMATER (IDR-Paraná, atual Instituto de Assistência Técnica e Extensão Rural - Emater e o Instituto Agrônômico do Paraná - IAPAR). A partir desses bancos foi desenvolvido um conjunto de dados hidrometeorológicos para o estado do Paraná.

Embora haja o conhecimento de outros conjuntos de dados anteriormente confeccionados a nível nacional, como o CAMELS-BR (Chagas et al., 2020) e o CABra (Almagro et al., 2021), a vantagem do novo conjunto de dados se inicia pela priorização dos dados brutos proveniente dos órgãos responsáveis pela operação dos postos e pelo uso da rede hidrográfica ottocodificada elaborada para o estado do Paraná no ano de 2020 pelo Instituto Água e Terra (IAT). Além da rede ser constituída por uma topologia consistente, a hierarquização pela codificação de Otto Pfafstetter (Pfafstetter, 1989) possibilita a extração de informações a montante e a jusante de cada trecho (Sousa et al., 2009). O processo de obtenção do conjunto de dados desenvolvidos para o estado do Paraná é descrito a seguir e o apêndice A apresenta maiores detalhes sobre os dados coletados.

#### 4.2.1 Coleta de Dados Hidrometeorológicos

Nesta seção são descritos os esforços para a obtenção dos registros pluviométricos e fluviométricos no estado do Paraná. O diagrama da Figura 4.2 apresenta de forma esquemática os procedimentos que foram seguidos para a realização das coletas dos dados. O primeiro passo mostra a delimitação da área de estudo, a qual foi apresentada anteriormente na seção 4.1.

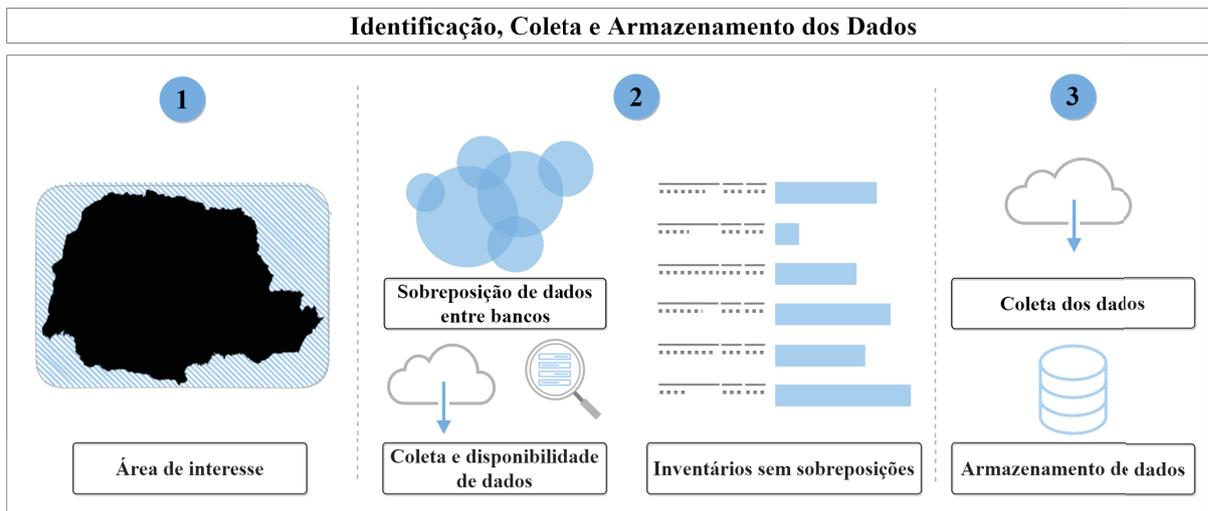


Figura 4.2: Diagrama da coleta das estações.

No passo seguinte, para cada um dos bancos de dados foram identificadas as estações que estavam localizadas dentro dos limites da área de interesse e, assim, foi produzido para cada banco uma primeira versão do inventário. Embora a busca tenha sido realizada em diferentes bancos de dados (Ex. IAT e HidroWEB), pode haver informações

duplicadas <sup>1</sup>, ou seja, os dados de uma estação podem ser obtidos em diferentes locais. Adicionalmente, há a possibilidade de estações antes convencionais tenham sido substituídas por instrumentos automáticos, e apesar de não haver a troca do código do posto pode ter sido feita a transferência da responsabilidade de operação para outro órgão <sup>2</sup>. Assim, visando que não houvesse, em um mesmo período, dados sobrepostos de uma mesma estação provenientes de diferentes bancos de dados foi adotada uma ordem de prioridade: as informações do banco do órgão responsável pela estação prevalece sobre as demais. É necessário ressaltar que durante a coleta também foram enfrentadas dificuldades, como: a inconsistência nos cadastros das estações e coordenadas geográficas imprecisas, as quais foram sanadas com a ajuda dos técnicos do órgão responsável da estação.

Ainda na segunda etapa, com base no inventário inicial foram avaliadas a duplicidade de informações entre os bancos, a disponibilidade de dados de cada estação, como também as coordenadas geográficas e o órgão responsável pela operação da estação. Ao final da análise, foi produzido um novo inventário sem sobreposição de informações para cada um dos bancos, contendo: código ANA da estação, o código da estação fornecido pelo órgão responsável, latitude, longitude e altitude.

A terceira etapa mostra a coleta dos dados, a qual utilizou o inventário produzido no passo anterior, realizada em janeiro de 2021 para que os postos coletados contivessem registros de observações de até 31 de dezembro de 2020. Os passos descritos acima foram repetidos para a coleta das estações pluviométricas e fluviométricas.

#### 4.2.2 Recuperação do Histórico de Precipitação (1979-2020)

Os registros observados de pluviometria do estado do Paraná foram adquiridos a partir de estações convencionais e automáticas. A seguir são apresentadas brevemente o número de postos adquiridos em cada banco de dados.

O IDR-Paraná contém as informações da rede de estações convencionais do antigo Instituto Agrônomo do Paraná (IAPAR), as quais foram adquiridas via e-mail pelo contato com os técnicos do Instituto. Ao todo foram fornecidas as séries históricas de 33 estações que possuem alturas de precipitações registradas uma única vez ao dia às 9h BRT.

O Banco de Dados Meteorológicos (BDMEP) do INMET é outro banco que mantém somente as medições históricas realizadas pelo próprio instituto, as quais são registradas diariamente às 9h BRT. Ao todo foram identificadas 50 estações convencionais com medições disponíveis no banco.

O Sistema de Informações Hídricas (SIH) do IAT mantém registros de estações convencionais do antigo Instituto das Águas do Paraná (AGUASPARANÁ) e da Superintendência de Desenvolvimento de Recursos Hídricos e Saneamento Ambiental (Suderhsa). Segundo o inventário fornecido pelos técnicos do Instituto, o SIH é composto por 1200 estações pluviométricas incluindo as que são operadas pelo próprio instituto, as que são operadas pelos institutos anteriormente citados e as de outros órgãos. Dessa maneira, a coleta de dados no SIH desconsiderou as estações provenientes do IDR-Paraná e do INMET, e as que não possuíam nenhum registro no banco. Logo, foram obtidas 1101 estações. Atentou-se que os técnicos do IAT realizam medições duas vezes ao dia às 7h

<sup>1</sup>A duplicidade das estações foi conferida conforme o código ANA atribuído aos postos. Para as situações em que as estações não possuíam código ANA (Ex. estações do Simepar) foi considerado que os dados dessas estações só estavam presentes no banco do órgão responsável pela operação.

<sup>2</sup>Para essas situações as séries antes convencionais foram complementadas pelos registros automáticos, os quais foram integrados ou calculadas as médias para gerar a escala diária.

BRT e às 17h BRT, sendo que a precipitação diária do dia X adotada é a soma do registro feito às 17h do dia anterior com a medição realizada às 7h do dia X.

Observa-se que dentre os institutos citados anteriormente são adotados diferentes horários de início e de término do dia, também chamado de *End of the day* (EOD). Observou-se que os registros realizados pelas estações do IAT possuem uma diferença de duas horas ao se comparar com os registros diários realizados pelo IDR-Paraná e pelo INMET. De acordo com Yatagai et al. (2020), a mistura de séries de precipitações com diferentes EOD's pode gerar incertezas sobre a natureza do evento, contudo devido à impossibilidade da agregação das alturas de precipitações para outro horário esse foi respeitado.

Por sua vez, o banco do Simepar é formado por estações telemétricas que realizam medições subdiárias a cada 15 minutos a partir de 1997. A agregação da precipitação subdiária para diária, das 205 estações selecionadas dentro da área de interesse, respeitou os mesmos horários de início e término do dia do órgão que apresentou maior número de estações convencionais coletadas. Adicionalmente, o Simepar conta com o monitoramento realizado por 4 radares e por satélites, os quais também foram utilizados para a concepção do histórico de precipitação.

O método de controle de qualidade de Llabrés-Brustenga et al. (2019) foi empregado para a avaliação dos dados diários das séries de precipitações de 1389 estações, as quais são mostradas na Figura 4.3. O método pode ser separado em quatro etapas. Primeiro, confere-se as coordenadas das estações, o período de operação e a porcentagem de dados disponíveis. No controle de qualidade básico, segunda etapa, são identificados e descartados os dados que não são fisicamente possíveis, como por exemplo: valores de precipitações negativas e eventos extremos de precipitação superiores a 300 mm.

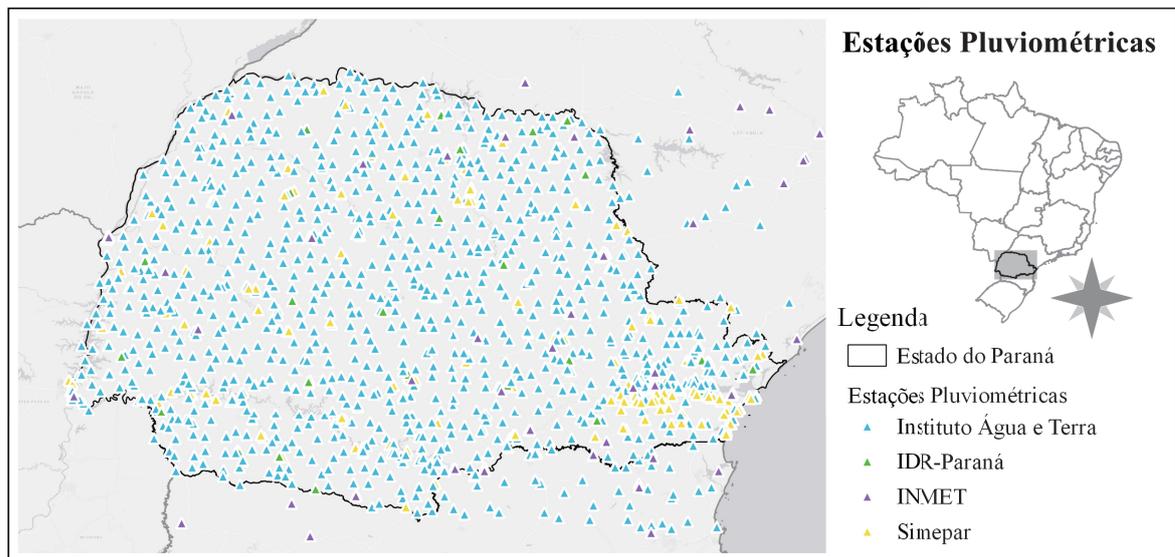


Figura 4.3: Localização dos postos coletados.

A terceira etapa, é denominada de controle de qualidade absoluto, a qual consiste na análise das séries históricas de cada estação de forma individual. Para cada ano da série é calculado um índice de qualidade (Q) que depende de 5 (cinco) fatores, são eles:

- A porcentagem de dados disponíveis em cada ano da série;

- A distribuição de falhas ao longo do ano. Para esse índice a penalização se torna maior para as séries que tiveram longos períodos contínuos de falhas;
- A probabilidade da série ser formada por possíveis falhas que foram preenchidas com zeros. Esse índice penaliza as séries que tiveram acumulados mensais iguais a zero indicando serem possíveis “falsos zeros”;
- A probabilidade do acúmulo sistemático de dois ou mais dias da semana. Esse índice penaliza a estação caso seja detectado um dia da semana com tendência de chover menos que nos outros dias da semana;
- A probabilidade da série conter *outliers*.

O índice de qualidade pode variar entre 0 (zero) a 100 (cem) por cento. Os valores iguais a 100% indicam qualidade absoluta, acima de 80% é considerada uma qualidade aceitável e abaixo de 50% a qualidade é muito baixa.

O processo de averiguação entre o que foi registrado no posto a ser analisado (posto candidato) e nas estações vizinhas (postos auxiliares) é realizado na quarta etapa, também conhecida como controle de qualidade relativo. Nessa etapa há dois índices que são relevantes para a classificação dos valores diário do posto candidato, que podem ser rotulados como: válido, “V”, duvidoso, “D”, inválido, “N”, ou informação insuficiente, “I”. Os registros identificados como informação insuficiente denotam que há menos do que duas estações auxiliares na região e no mesmo período para serem devidamente avaliadas, em consequência dessa condição observou-se que uma tendência mais significativa na quantidade de registros pluviométricos a partir de 1979. O primeiro índice, índice de representatividade, verifica os valores diários para cada dupla, posto candidato - posto auxiliar, tal índice considera a distância horizontal entre as estações <sup>3</sup>, a diferença de altitude e a correlação entre os dados medidos. O segundo índice é utilizado para analisar os acumulados mensais. A partir das estações do Simepar foram estabelecidos limites máximos de acumulados mensais, os quais foram aplicados para avaliar as séries históricas de cada posto candidato.

A etapa seguinte foi desenvolvida junto com o auxílio do Simepar. Os dados diários validados foram espacializados em uma grade de 1km x 1km por meio da krigagem espaço temporal, onde os valores de precipitação foram estimados em escala diária por meio de médias ponderadas entre os dados da vizinhança. Em seguida, um segundo método de espacialização foi aplicado para o histórico de precipitação mais recente. A partir de 2013 foi empregado o método apresentado por Calvetti et al. (2017) para se estimar a precipitação de forma espacializada dentro da área de interesse em uma grade de 1km x 1km, onde a equação de Poisson foi utilizada para combinar dados de radar e de satélite com os registros observados pelas estações telemétricas de precipitação. Por fim, foram aferidas precipitações médias por meio da média aritmética dos pontos da grade localizadas dentro da área de drenagem de cada bacia hidrográfica.

---

<sup>3</sup>Delimitou-se uma distância máxima de até 50 km entre o posto candidato e os postos auxiliares.

### 4.2.3 Delineamento das Bacias Hidrográficas

O procedimento de coleta e seleção das séries fluviométricas <sup>4</sup> iniciou com duas premissas. A primeira, em concordância com Tucci (2002), considerou aceitável utilizar postos com no mínimo 5 anos de registros fluviométricos para a aplicação dos métodos de regionalização. A segunda, as séries das estações a jusante de regularização foram desconsideradas em sua totalidade ou consideradas parcialmente nos períodos antecedentes à construção do empreendimento.

As estações convencionais foram obtidas nos bancos do IAT e do Portal HidroWEB. Apesar de ambos os bancos preservarem informações de estações de diversas operadoras, foi adotado que as informações provenientes do banco do IAT teriam prioridade sobre as do Portal HidroWEB. O inventário fornecido pelo técnico responsável do IAT informa que há 768 estações fluviométricas, dentre essas foram desconsideradas: as estações que estavam localizadas fora da área de interesse; as estações que só possuem informações de qualidade d'água, e as estações com séries com comprimento menor do que 5 anos. Obtendo-se ao final 413 estações.

A partir do catálogo de metadados da ANA, dentro do qual encontram-se os metadados de Índices e Estatísticas Hidrometeorológicas, foi obtido o inventário do HidroWEB das estações fluviométricas convencionais que possuem ao menos um ano de medições contínuas, podendo ser dados brutos ou consistidos. Majoritariamente das estações encontradas no Portal HidroWEB também estão presentes no inventário do IAT e, dessa maneira, foram identificadas apenas 15 estações diferentes com ao menos 5 anos de registros fluviométricos.

As séries telemétricas de 83 estações do IAT e de 57 estações do Simepar foram obtidas no banco do Simepar. As estações telemétricas do IAT se incluem no caso citado anteriormente, onde foi realizado o processo de modernização dos equipamentos sem que fossem alterados os códigos das estações. Para essas situações as séries convencionais foram complementadas pelos registros automáticos, os quais foram processados para a escala diária respeitando o EOD adotado pelo IAT.

Posteriormente, ao mesmo tempo em que a segunda premissa foi empregada conferiu-se as localizações dos postos por meio do procedimento manual de hidrorreferenciamento utilizando a rede hidrográfica otocodificada do IAT. O hidrorreferenciamento torna a localização geográfica conhecida para o sistema de codificação de bacias criado por Otto Pfafstetter permitindo a obtenção de informações a montante e a jusante de cada trecho e, assim, o delineamento das áreas de drenagem (Sousa et al., 2009; Pfafstetter, 1989). Por último, foi realizado um controle de qualidade manual, onde foram desconsiderados dados não consistes, como por exemplo: mudanças bruscas de régua alterando de forma evidente a vazão de base, séries com falhas grandes alternadas com período pequenos de medições, medidas de baixa precisão ou que apresentavam ter valores constantes por longos períodos. Ao final foram obtidos um total de 284 postos fluviométricos com observações que variam de 1926 a 2020, como são mostrados no mapa da Figura 4.4.

---

<sup>4</sup>A obtenção das séries fluviométricas incluiu a coleta das séries de vazões, onde as cotas observadas foram anteriormente transformadas em vazão pelo órgão responsável pela operação do posto, e séries de cotas, as quais ainda não haviam sido transformadas em vazão e, assim, quando disponível fez-se a coleta das curvas-chaves do posto e em seguida a transformação da cota em vazão.

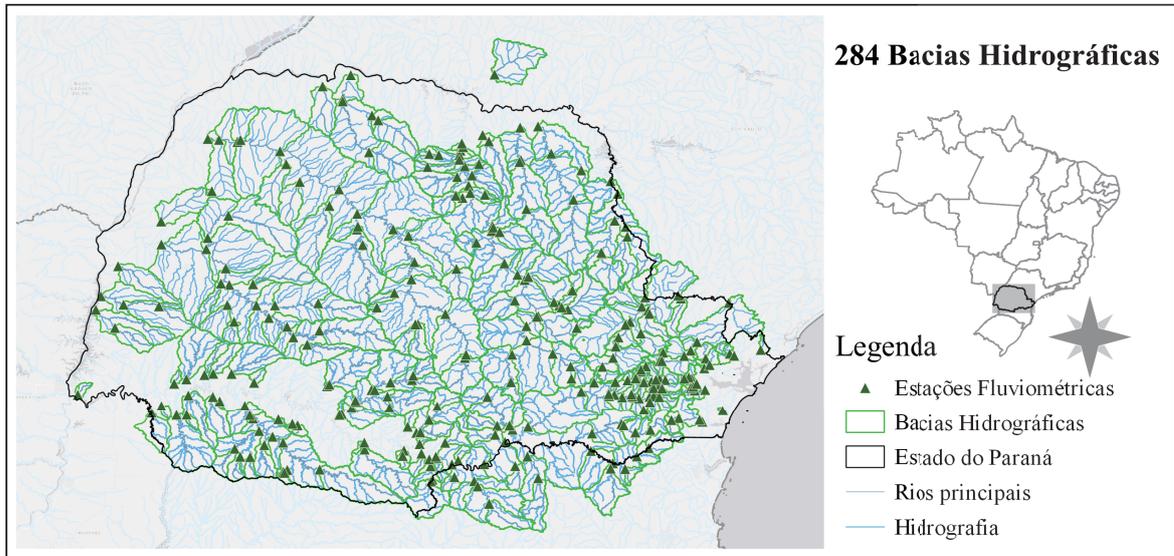


Figura 4.4: Estações fluviométricas coletadas e bacias hidrográficas delineadas com a rede hidrográfica ottocodificada.

#### 4.2.4 Evapotranspiração Potencial

A equação da FAO Penman-Monteith (Allen et al., 1998) foi utilizada para estimar a evapotranspiração potencial. Esse método requer séries de temperatura do ar, umidade relativa do ar, velocidade do vento e radiação solar, as quais foram obtidas por meio das estações telemétricas do Simepar com registros variando de 1997 a 2020. Comparando a amplitude temporal obtida na recuperação do histórico de precipitação, 1979 a 2020, e na coleta das séries fluviométricas, 1926 a 2020, as séries de evapotranspiração potencial são relativamente recentes. A equação da FAO Penman-Monteith pode ser escrita como:

$$\lambda ET = \frac{\Delta (R_n - G) + \rho_a c_p \frac{(e_s - e_a)}{r_a}}{\Delta + \gamma \left(1 + \frac{r_s}{r_a}\right)}, \quad (4.1)$$

onde  $\lambda ET$  é a densidade do fluxo de calor latente em  $\text{MJ m}^{-2} \text{d}^{-1}$ ,  $\Delta$  é a inclinação da curva de pressão de vapor de saturação em  $\text{kPa } ^\circ\text{C}^{-1}$ ,  $R_n$  é a radiação solar em  $\text{MJ m}^{-2} \text{d}^{-1}$ ,  $G$  é a densidade do fluxo de calor no solo em  $\text{MJ m}^{-2} \text{d}^{-1}$ ,  $\rho_a$  é a densidade média do ar a pressão constante em  $\text{kg m}^{-3}$ ,  $c_p$  é o calor específico do ar em  $\text{MJ kg}^{-1} \text{ } ^\circ\text{C}^{-1}$ ,  $e_s$  é a pressão de vapor de saturação média diária do ar em  $\text{kPa}$ ,  $e_a$  é a pressão de vapor do ar média diária em  $\text{kPa}$ ,  $\gamma$  é a constante psicrométrica em  $\text{kPa } ^\circ\text{C}^{-1}$ ,  $r_a$  é a resistência aerodinâmica em  $\text{s m}^{-1}$  e  $r_c$  é a resistência da superfície da cobertura vegetal em  $\text{s m}^{-1}$ . A partir da série de evapotranspiração potencial foi calculado um valor médio para cada dia juliano utilizando o período de dados disponíveis em cada posto.

Posteriormente, com o auxílio do Simepar, as informações pontuais foram espacializadas utilizando um método de regressão seguido de interpolação, também conhecido como *Regression-Kriging* ou método híbrido de interpolação. De acordo com Hengl et al. (2007), a predição de valores em um ponto  $s_0$  no espaço que não possui medições pode ser calculado pelo *Regression-Kriging* como:

$$\hat{z}(s_0) = \hat{m}(s_0) + \hat{e}(s_0), \quad (4.2)$$

onde  $\hat{z}(s_0)$  é o valor predito da variável de interesse no ponto  $s_0$ ,  $\hat{m}(s_0)$  é o valor predito pelo modelo de regressão e  $\hat{e}(s_0)$  são os resíduos interpolados.

O método de regressão *Random Forest* (Breiman, 2001) é um algoritmo de aprendizagem supervisionado em que são realizadas amostragens aleatórias com reposições provenientes de uma mesma amostra de dados de treinamento para formar subconjuntos de dados. Cada subconjunto de dados é utilizado para treinar uma árvore de decisão sem que haja nenhuma interação entre elas. As predições de todas as árvores individuais são então combinadas por meio de uma média para a realização da predição final. Fez-se o uso de 500 árvores de decisões para modelar a relação entre a variável dependente (evapotranspiração potencial) e as variáveis independentes (temperatura, precipitação e uso e cobertura do solo).

Como os resíduos da regressão geram dependência espacial esses podem ser interpolados pelo método de krigagem simples (Hengl et al., 2007). A krigagem faz parte das técnicas de geoestatística, que é capaz de estipular ou predizer valores em regiões do espaço em que não possuem medidas a partir da média ponderada de pontos observados na vizinhança (Varouchakis, 2019; Grego et al., 2014). Logo, os resíduos interpolados em um ponto  $s_0$  é escrito como:

$$\hat{e}(s_0) = \sum_{i=1}^n \lambda_i e(s_i), \quad (4.3)$$

sendo que  $\lambda_i$  são os pesos da interpolação para os resíduos ( $e$ ) observados nos pontos  $s_i$  da vizinhança.

#### 4.2.5 Descritores das Bacias Hidrográficas

Espera-se por meio de índices quantitativos descrever a paisagem, o relevo, o clima, a topologia da hidrografia, o uso e cobertura do solo e o tipo de solo. Uma vez que, tais características podem desempenhar papéis fundamentais nas respostas hidrológicas das bacias. Os descritores foram divididos em 3 categorias: fisiográficos, climatológicos e de tipo de solo e uso e cobertura do solo, como são apresentados a seguir.

#### 4.2.6 Índices Fisiográficos

Os índices fisiográficos estão relacionados com as características físicas, a localização geográfica e a topografia da rede de drenagem. A partir da rede hidrográfica otocodificada foi realizado o delineamento das bacias, sendo também obtidas as áreas e os trechos de drenagens que foram utilizados como base para o cálculo dos índices descritos na Tabela 4.1. Também foi utilizado o modelo digital de elevação (MDE) com resolução de 30 metros proveniente da missão *Shuttle Radar Topography Mission* (SRTM) da NASA para a estimativa das declividades e altitudes.

Tabela 4.1: Descrição dos índices fisiográficos das bacias hidrográficas.

Índices Fisiográficos	Unidade	Descrição
Latitude do centroide	graus	Latitude do centroide do polígono da área da bacia.
Longitude do centroide	graus	Longitude do centroide do polígono da área da bacia.
Altitude do centroide	m	Altitude do centroide do polígono da área da bacia.
Área da bacia	km <sup>2</sup>	Área do polígono.
Altitude média da bacia	m	Altitude média da área da bacia.
Declividade média da bacia	m/m	Média de todas as declividades calculadas dentro do polígono da área da bacia.
Número de Strahler	-	Número de ordem da bacia de Strahler.
Comprimento do rio principal	m	Soma do comprimento de todos os trechos de drenagem do rio principal.
Densidade de drenagem	km/km <sup>2</sup>	Razão entre o comprimento total de todos os trechos de drenagem e a área da bacia.
Declividade média do rio principal	m/m	Média da declividade de cada trecho de drenagem do rio principal ponderada pelo comprimento de cada trecho.

Observa-se na Tabela 4.1 que o número de ordem de Strahler também foi considerado como um descritor da topologia da rede de drenagem. Diferente da codificação de Otto Pfafstetter que realiza a hierarquização da rede a partir da exutória do rio principal, o número de ordem de Strahler ordena os trechos de rio no sentido da cabeceira a exutória, e a cada nova confluência o trecho localizado a jusante do afluente se eleva em uma unidade a ordem do trecho a montante do tributário.

#### 4.2.7 Índices Climatológicos

As séries de precipitações médias e as estimavas de evapotranspiração potencial, as quais foram desenvolvidas previamente para cada bacia hidrográfica, foram empregadas para o cálculo dos índices relacionados à precipitação e à evapotranspiração potencial, como mostrado na Tabela 4.2.

Ademais, Barbieri et al. (2017) disponibilizaram os atlas do estado do Paraná de médias mensais de temperaturas e médias de radiação solar para cada estação do ano, os quais são mostrados nas Figuras 4.5 e 4.6. Os atlas foram produzidos com base nas medições das estações do INMET, do Simepar e do IDR-Paraná entre o período de 2006 a 2016. Essas informações foram usadas para computar índices médios nas bacias situadas dentro do estado, e para as bacias localizadas nas fronteiras entre outros estados foram adotados os valores médios da bacia mais próxima.

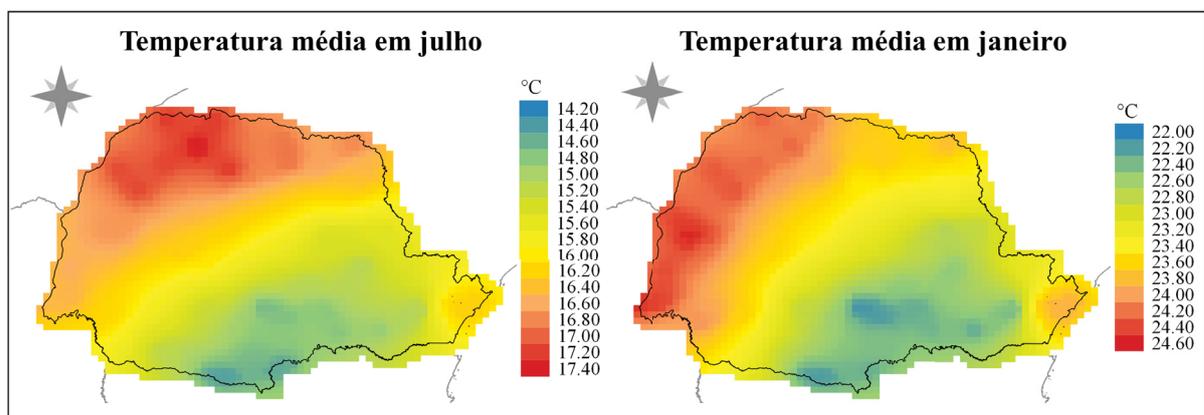


Figura 4.5: Temperaturas médias nos meses de julho e janeiro entre 2006 a 2016. Fonte: Barbieri et al. (2017).

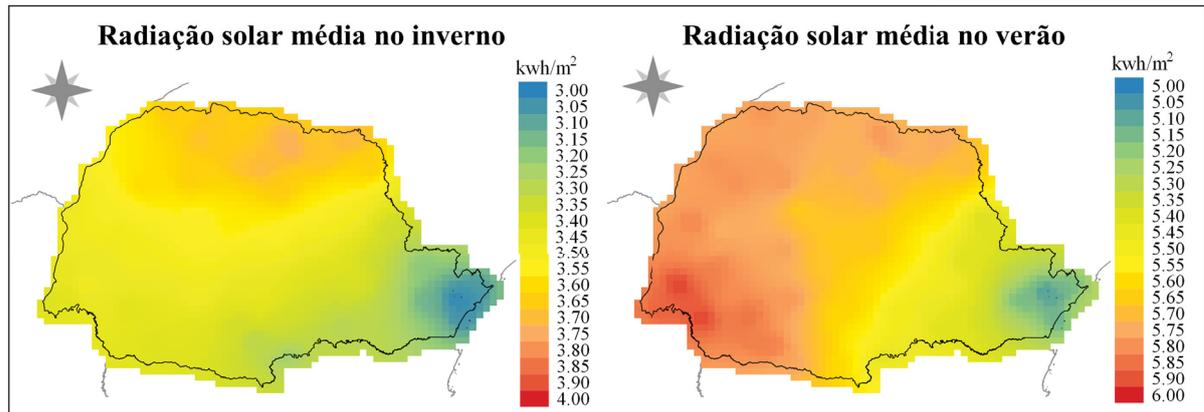


Figura 4.6: Radiação solar média nos meses de inverno e de verão entre 2006 a 2016. Fonte: Barbieri et al. (2017).

Tabela 4.2: Descrição dos índices climatológicos das bacias hidrográficas.

Índices Climatológicos	Unidade	Descrição
CV anual de precipitação	-	Coefficiente de variação anual de precipitação.
Dias de precipitação com acumulado mensal de 10mm	dias/ano	Número médio de dias anuais com precipitação (>1mm/dia)
Dias de precipitação com acumulado mensal de 50mm	dias/ano	nos meses com acumulado mensal maior que 10, 50 e 150mm.
Dias de precipitação com acumulado mensal de 150mm	dias/ano	
Precipitação média diária	mm/dia	Precipitação média diária.
Frequência de dias sem chuvas	dias/ano	Número médio de dias sem chuva (<1 mm/dia).
Duração média de dias sem chuva	dias	Duração média de dias consecutivos sem chuva (<1 mm/dia).
Precipitação	mm	Média anual de precipitação.
Evapotranspiração potencial	mm	Média anual de evapotranspiração potencial.
Índice de aridez	-	Razão entre a média anual de precipitação e média anual de evapotranspiração potencial.
Temperatura média do mês mais frio	°C	Temperatura média do mês de julho.
Temperatura média do mês mais quente	°C	Temperatura média do mês de janeiro.
Radiação solar média no mês de inverno	kWh/m <sup>2</sup>	A radiação solar média nos meses de inverno.
Radiação solar média no mês de verão	kWh/m <sup>2</sup>	A radiação solar média nos meses de verão.

#### 4.2.8 Índices de Solo e Uso e Cobertura do Solo

A paisagem pode ser descrita por meio de técnicas de sensoriamento remoto que utilizam imagens de satélites para analisar padrões espaciais sobre a superfície da Terra. O mapa de uso e cobertura do solo do ano de 2019, o qual foi utilizado, é disponibilizado pelo MapBiomias (Souza et al., 2020). As classes mostradas no mapa da Figura 4.7 foram utilizadas como base para o cálculo das frações de área que cada classe ocupa nas bacias e para determinar a classe dominante. O mapa de solos usado foi produzido e disponibilizado pela Embrapa (2020) para o estado do Paraná com escala de 1:250.000. A Figura 4.8 mostra o mapa utilizado para calcular a fração de área que cada solo ocupa e o solo dominante de cada bacia.

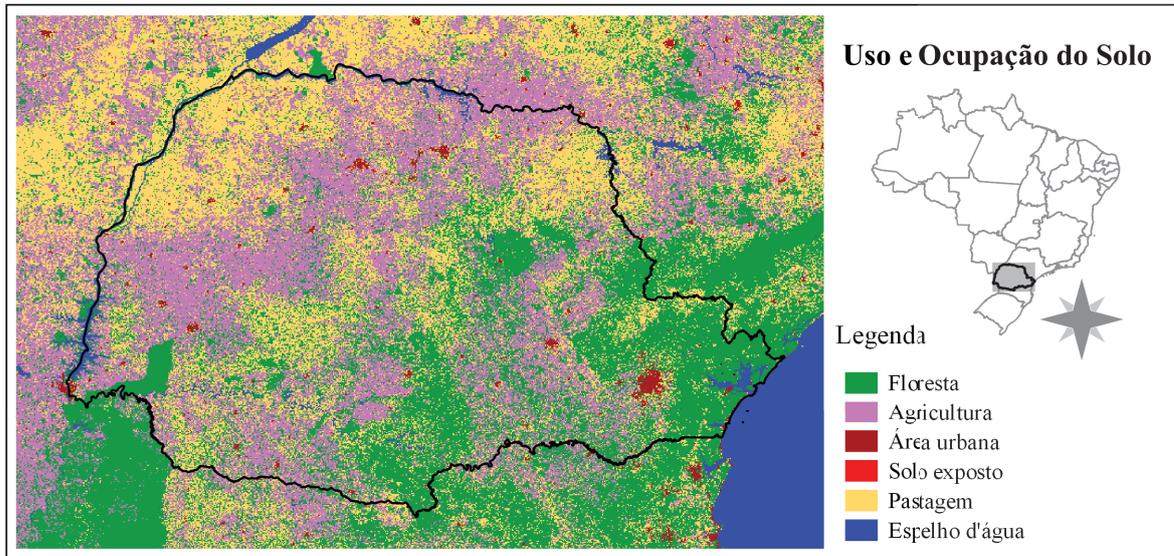


Figura 4.7: Uso e ocupação do solo no estado do Paraná. Fonte: Souza et al. (2020).

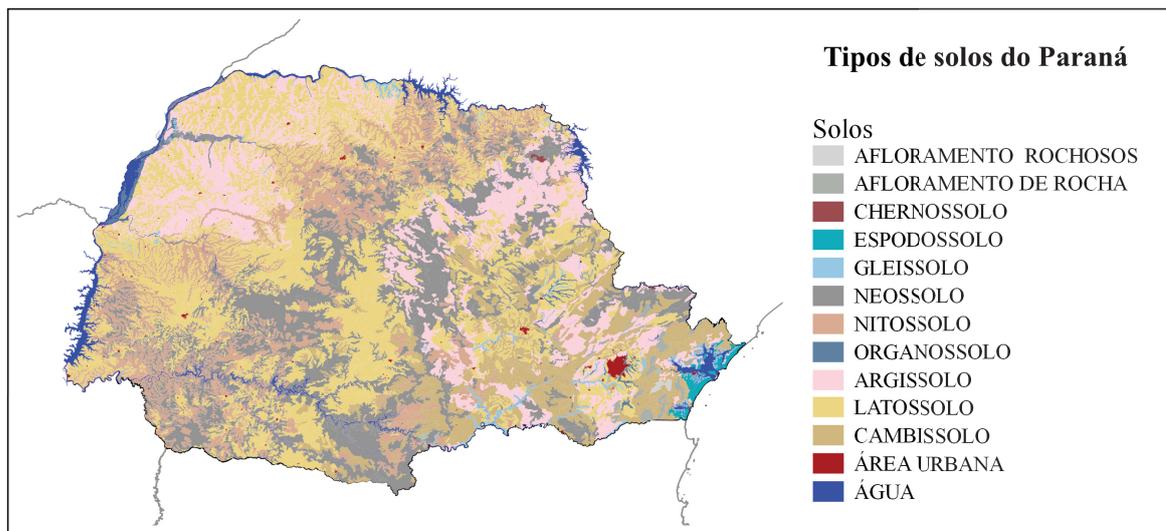


Figura 4.8: Mapa de solos. Fonte: Embrapa (2020).

Tabela 4.3: Descrição dos índices de Solo e Uso e Cobertura do Solo das bacias hidrográficas.

Tipo e uso e cobertura do solo	Unidade	Descrição
Uso e cobertura do solo	%	Frações de área da bacia com cada tipo de uso do solo
Uso e cobertura do solo dominante	-	Tipo de uso e cobertura do solo dominante da bacia
Tipo de solo	%	Frações de área da bacia com cada tipo de solo
Tipo de solo dominante	-	Tipo de solo dominante da bacia
Número curva (CN)	-	Média do número curva ponderada pela área

O método número curva (em inglês, *Curve Number* (CN)) desenvolvido por Soil Conservation Service (1972) relaciona as informações de solo e uso e cobertura do solo

para classificar a região com base no seu potencial de retenção de águas pluviais. Para a estimativa do CN nas bacias do Paraná, foram utilizados os valores calculados pela ANA e que estão disponibilizados no catálogo de metadados da ANA.

#### 4.2.9 Seleção de Bacias Hidrográficas

As Figuras A.4, A.5 e A.6 do Apêndice A mostram que os registros das estações fluviométricas coletadas, que variam de 1926 a 2020, podem estar muito dispersos no tempo e também podem não apresentar medições em períodos em que há histórico de precipitação. Desse modo, selecionou-se uma amostra do conjunto de dados do estado do Paraná para que houvesse simultaneidade de dados entre as estações fluviométricas e pluviométricas. Nessa amostra foram incluídas 126 bacias hidrográficas que possuem ao menos 15 anos de dados de vazão entre 1979 e 2020, sendo que cada ano contabilizado tem no máximo 10% de falhas, além disso preferiu-se que as séries históricas tivessem também dados mais recentes, as quais se estendiam além do ano de 2010, e limitou-se as séries históricas homogêneas que fossem aprovados pelo teste de Pettitt (Pettitt, 1979)<sup>5</sup>. As 126 bacias hidrográficas estão posicionadas em diferentes regiões do estado, como é apresentado pela Figura 4.9.



Figura 4.9: Localização das bacias hidrográficas selecionadas no estado do Paraná.

A Figura 4.10 mostra a disponibilidade de dados ao longo dos anos por estação, quanto mais escura a cor verde, maior é a quantidade de dados disponíveis naquele ano. Nessa amostra são encontradas bacias hidrográficas com características descritivas variadas, que são mostradas na Tabela 4.4. A partir da Tabela 4.4 pode-se observar que a área, a altitude e a declividade média das bacias presentes nessa amostra são dispersas. Como também, o índice de aridez mínimo igual à 0,97 indica que a região pode ser classificada como subúmido úmido (Matallo Junior, 2001). A paisagem, por sua vez, mostra ser bem diversificada podendo haver altos percentuais de áreas cobertas por floresta, agricultura, pastagem ou urbanização.

<sup>5</sup>O teste não paramétrico de Pettitt, o qual foi calculado por meio da biblioteca *Pyhomogeneity* em *Python*, é capaz de indicar o ano em que ocorreu uma mudança brusca da tendência temporal.

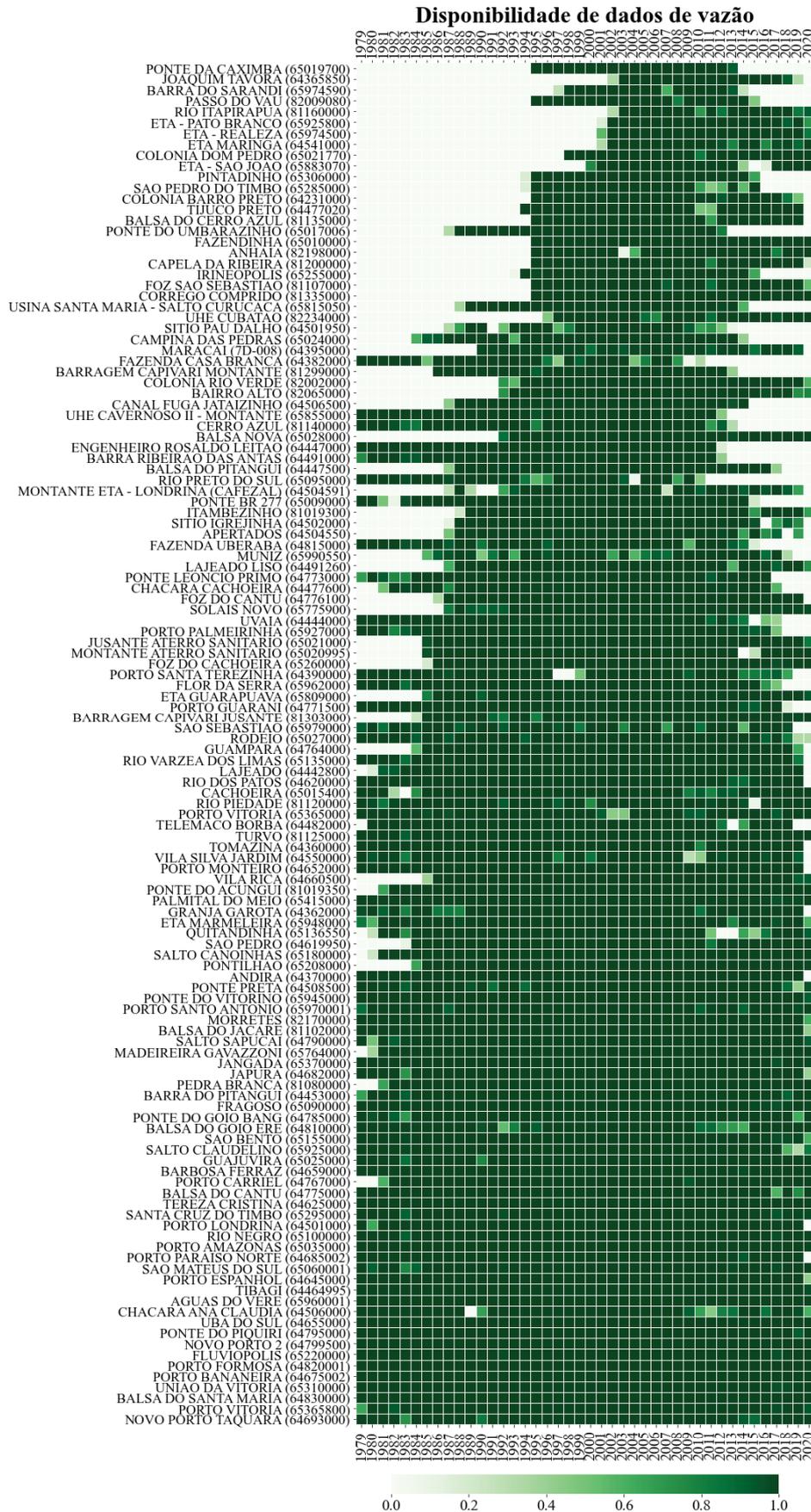


Figura 4.10: Disponibilidade de dados de vazão por estação. Quanto mais escura a cor verde, mais dados disponíveis há naquele ano.

Tabela 4.4: Estatísticas descritivas para o conjunto de dados do Paraná.

Descritores	média	desvio padrão	min	25 %	50 %	75 %	max
<b>Índices fisiográficos</b>							
Altitude média do centroide (m)	743,16	210,64	62,00	611,25	773,50	892,25	1132,00
Área (km <sup>2</sup> )	4474,15	7001,22	13,87	510,70	1523,48	4120,90	34440,18
Altitude média da bacia (m)	804,33	171,96	262,85	666,36	835,74	920,45	1150,88
Declividade média da bacia (m/m)	0,16	0,06	0,06	0,12	0,14	0,18	0,33
Número de Strahler	6,69	1,29	3,00	6,00	7,00	7,00	9,00
Comprimento do rio principal (m)	185335,43	164551,04	7336,06	66867,87	122924,13	236343,96	748033,70
Densidade de drenagem (km/km <sup>2</sup> )	2,55	1,04	0,72	1,81	2,33	3,28	5,59
Declividade do rio principal (m/m)	0,04	0,02	0,02	0,03	0,03	0,04	0,10
<b>Índices climatológicos</b>							
Coefficiente de variação da precipitação anual	0,17	0,02	0,13	0,16	0,17	0,18	0,21
Temperatura média de julho (°C)	15,55	0,58	14,56	15,16	15,42	15,98	17,02
Temperatura média de janeiro (°C)	22,96	0,46	22,26	22,56	22,89	23,31	24,07
Dias de precipitação com acumulado mensal de 10mm	152,86	17,45	112,79	142,02	151,56	162,55	213,14
Dias de precipitação com acumulado mensal de 50mm	146,07	17,55	104,38	133,40	145,44	156,40	208,45
Dias de precipitação com acumulado mensal de 150mm	82,84	17,43	51,38	70,54	81,52	93,96	159,12
Evapotranspiração potencial anual (mm)	1255,95	78,02	1139,88	1192,19	1243,22	1326,41	1423,14
Precipitação média anual (mm)	1678,61	216,08	1357,26	1511,72	1614,79	1828,01	2618,98
Radiação solar média nos meses de inverno (kwh/m <sup>2</sup> )	3,39	0,15	3,13	3,27	3,39	3,50	3,70
Radiação solar média nos meses de verão (kwh/m <sup>2</sup> )	5,53	0,19	5,16	5,37	5,54	5,71	5,86
Índice de aridez	1,34	0,18	0,97	1,24	1,30	1,45	2,02
Precipitação média diária (mm/dia)	4,60	0,59	3,72	4,14	4,42	5,00	7,17
Frequência de dias sem chuva (dias/ano)	208,28	17,94	148,24	197,65	209,87	219,21	249,17
Duração média de dias sem chuva (dias)	4,54	0,43	3,41	4,29	4,57	4,78	5,71
<b>Uso e cobertura do solo</b>							
(1) Floresta (%)	0,48	0,25	0,06	0,29	0,41	0,71	0,98
(2) Agricultura (%)	0,27	0,21	0,00	0,07	0,26	0,42	0,79
(3) Área urbana (%)	0,03	0,07	0,00	0,00	0,01	0,02	0,53
(4) Solo exposto (%)	0,00	0,00	0,00	0,00	0,00	0,00	0,01
(5) Pastagem (%)	0,23	0,13	0,01	0,14	0,19	0,29	0,80
(6) Água (%)	0,00	0,01	0,00	0,00	0,00	0,00	0,08
Número Curva	77,37	4,68	57,93	75,90	77,88	80,39	87,91
<b>Tipo de solo</b>							
(1) Latossolo (%)	0,24	0,18	0,00	0,08	0,25	0,35	0,78
(2) Neossolo (%)	0,20	0,18	0,00	0,02	0,15	0,33	0,70
(3) Argissolo (%)	0,14	0,17	0,00	0,00	0,11	0,21	0,91
(4) Nitossolo (%)	0,10	0,13	0,00	0,00	0,04	0,15	0,65
(5) Cambissolo (%)	0,20	0,23	0,00	0,02	0,11	0,31	0,95
(6) Gleissolo (%)	0,01	0,03	0,00	0,00	0,00	0,02	0,21
(7) Organossolo (%)	0,01	0,03	0,00	0,00	0,00	0,00	0,18
(8) Espodossolo (%)	0,00	0,00	0,00	0,00	0,00	0,00	0,00
(9) Afloramento rochoso (%)	0,02	0,04	0,00	0,00	0,00	0,01	0,27
(10) Área urbana (%)	0,01	0,05	0,00	0,00	0,00	0,00	0,43

## 5 MODELO GR4J

O GR4J (em francês, *modèle du Génie Rural à 4 paramètres Journalier*) proposto por Perrin et al. (2003) faz parte da família dos modelos GR (em francês, *Génie Rural*) de chuva-vazão desenvolvidos no Instituto Nacional de Pesquisa para Agricultura, Alimentação e Meio Ambiente (em francês, *L'Institut National de Recherche Pour L'Agriculture, L'Alimentation et L'Environnement*) da França. Esse modelo vem sendo amplamente aplicado em diferentes localidades como França (Oudin et al., 2008), Austrália (Pagano et al., 2010), Rússia (Ayzel et al., 2019) e Coreia do Sul (Kim et al., 2017), e apresenta ser vantajoso pela sua robustez e eficiência. Além disso, a escolha desse modelo deu-se também pela familiaridade do autor, que empregou o GR4J nas bacias hidrográficas do conjunto de dados do Estado do Paraná e do conjunto de dados dos Estados Unidos da América chamado CAMELS (em inglês, *Catchment Attributes and Meteorology for Large-sample Studies*), respectivamente apresentados nas pesquisas Kuana et al. (2021a) e Kuana et al. (2021b) no XXIV Simpósio Brasileiro de Recursos Hídricos.

A partir de forçantes meteorológicas de escala diária, precipitação média da bacia ( $P$ ) e evapotranspiração potencial ( $E$ ), o modelo é capaz de estimar a vazão, também na escala diária. Para tanto, o modelo faz uso de quatro parâmetros, dois hidrogramas unitários e dois reservatórios, como é apresentado na Figura 5.1.

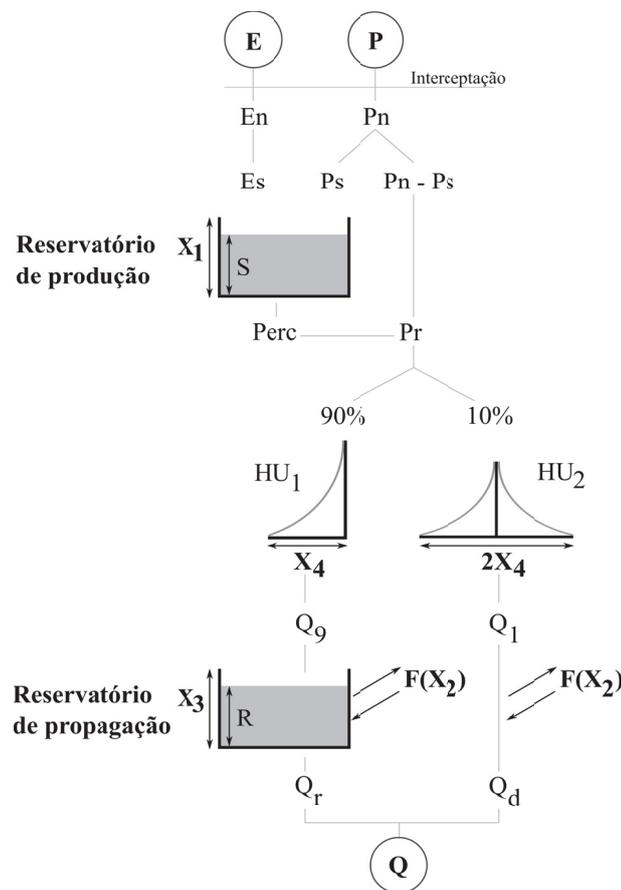


Figura 5.1: Diagrama do modelo GR4J. Fonte: Adaptado de Perrin et al. (2003)

A primeira fase do modelo é chamada de interceptação, a qual determina as entradas líquidas de precipitação ( $P_n$ ) e de evapotranspiração potencial ( $E_n$ ). Para os casos em que  $P \geq E$ , então as entradas podem ser expressas como:

$$P_n = P - E \text{ e} \quad (5.1)$$

$$E_n = 0. \quad (5.2)$$

Por outro lado, quando  $P < E$  admite-se que

$$E_n = E - P \text{ e} \quad (5.3)$$

$$P_n = 0. \quad (5.4)$$

Após a interceptação, quando há eventos de precipitação ( $P_n > 0$ ) parte dessa água pode infiltrar no solo sendo direcionada para o primeiro reservatório chamado de reservatório de produção. O reservatório de produção é responsável por controlar o armazenamento d'água, a evapotranspiração e a percolação no solo. A parcela de água,  $P_s$ , é calculada em função do conteúdo atual de água no reservatório ( $S$ ) em mm e do parâmetro  $X_1$ , que corresponde à capacidade total em mm do reservatório de produção, como é mostrado a seguir:

$$P_s = \frac{X_1 \left(1 - \left(\frac{S}{X_1}\right)^2\right) \tanh\left(\frac{P_n}{X_1}\right)}{1 + \frac{S}{X_1} \tanh\left(\frac{P_n}{X_1}\right)}. \quad (5.5)$$

Alternativamente, nos dias em que não há precipitação tem-se  $E_n > 0$ . Desse modo, a taxa de evaporação ( $E_s$ ) do reservatório de produção é

$$E_s = \frac{S \left(2 - \frac{S}{X_1}\right) \tanh\left(\frac{E_n}{X_1}\right)}{1 + \left(1 - \frac{S}{X_1}\right) \tanh\left(\frac{E_n}{X_1}\right)}. \quad (5.6)$$

Nesta etapa o valor do conteúdo de água no reservatório ( $S$ ) pode ser atualizado como:

$$S := S - E_s + P_s, \quad (5.7)$$

onde, o símbolo  $:=$  equivale ao operador de atribuição utilizado em linguagens de programação. Posteriormente, a água que percola do reservatório de produção,  $Perc$ , pode ser descrita utilizando a equação a seguir:

$$Perc = S \left\{ 1 - \left[ 1 + \left( \frac{4}{9} \frac{S}{X_1} \right)^4 \right]^{-\frac{1}{4}} \right\}. \quad (5.8)$$

O conteúdo de água no reservatório ( $S$ ) jamais pode ultrapassar o valor de  $X_1$  como também a percolação não pode ser maior que  $S$ . Conseqüentemente, o conteúdo de água do primeiro reservatório ( $S$ ) se torna:

$$S := S - Perc. \quad (5.9)$$

Quando a precipitação não infiltra mais no solo é denominada de precipitação efetiva, que é responsável pela geração do escoamento superficial. A soma da precipitação efetiva com a percolação do reservatório de produção, expressa como:

$$P_r = Perc + (P_n - P_s), \quad (5.10)$$

refere-se ao conteúdo de água a ser propagado ( $P_r$ ). Previamente,  $P_r$  é fracionada em 90% como escoamento de resposta rápida, a qual é propagada pelo primeiro hidrograma unitário e pelo segundo reservatório, denominado de reservatório de propagação, e 10% como escoamento de resposta lenta, sendo propagada pelo segundo hidrograma unitário. De acordo com Perrin et al. (2003), o uso do hidrograma linear,  $HU_1$ , e do hidrograma não linear,  $HU_2$ , permite que o modelo simule o pico de vazão com base no atraso da resposta ao evento de precipitação. O parâmetro  $X_4$  é responsável por indicar esse atraso em dias.

Os hidrogramas  $HU_1$  e  $HU_2$  na sua forma discreta possuem  $n$  e  $m$  ordenadas, as quais são definidas como os menores valores inteiros a exceder  $X_4$  e  $2X_4$ , respectivamente. Ambos os hidrogramas unitários derivam das curvas- $s$ ,  $SH_1$  e  $SH_2$ , que mostram a distribuição do escoamento de forma cumulativa ao longo do tempo. Assim,  $SH_1$  é definida ao longo do tempo  $t$  como:

$$SH_1(t) = \begin{cases} 0 & \text{para } t \leq 0 \\ \left(\frac{t}{X_4}\right)^{\frac{5}{2}} & \text{para } 0 < t < X_4 \\ 1 & \text{para } t \geq X_4. \end{cases} \quad (5.11)$$

Do mesmo modo,  $SH_2$  é definida ao longo do tempo  $t$  como:

$$SH_2(t) = \begin{cases} 0 & \text{para } t \leq 0 \\ \frac{1}{2} \left(\frac{t}{X_4}\right)^{\frac{5}{2}} & \text{para } 0 < t \leq X_4 \\ 1 - \frac{1}{2} \left(2 - \frac{t}{X_4}\right)^{\frac{5}{2}} & \text{para } X_4 < t < 2X_4 \\ 1 & \text{para } t \geq 2X_4. \end{cases} \quad (5.12)$$

Os valores de cada ordenada,  $j$ , de  $HU_1$  e  $HU_2$  são calculados por:

$$HU_1(j) = SH_1(j) - SH_1(j-1) \text{ e} \quad (5.13)$$

$$HU_2(j) = SH_2(j) - SH_2(j-1). \quad (5.14)$$

Durante o processo de propagação, seja escoamento de resposta lenta ou rápida, podem haver trocas subterrâneas de água ( $F$ ) positivas, quando há importação de água, ou negativas, quando há exportação de água, ou nulas, quando não ocorre nenhuma troca.

Esse procedimento é descrito em função do parâmetro  $X_2$  que corresponde ao coeficiente de troca de água em mm/dia, do conteúdo de água atual ( $R$ ) e da capacidade total de água ( $X_3$ ) do reservatório de propagação, como é mostrado a seguir:

$$F = X_2 \left( \frac{R}{X_3} \right)^{\frac{7}{2}}. \quad (5.15)$$

Após a propagação do escoamento pelo hidrograma unitário  $HU_1$ , a vazão resultante  $Q_9$  é direcionada para o segundo reservatório, onde ocorre uma segunda propagação com trocas subterrâneas de água. Nessa etapa o nível  $R$  do reservatório é atualizado para o maior valor entre 0 e o somatório de  $R$ ,  $Q_9$  e  $F$ , como é apresentado a seguir:

$$R = \max(0; R + Q_9 + F). \quad (5.16)$$

No início do passo de tempo  $t$  o reservatório pode chegar a receber volumes de água maiores que o deficit de saturação, ou seja, maior que  $X_3 - R$ . Contudo, quanto mais alto for o nível do reservatório maiores serão as perdas de água. Assim, a vazão  $Q_r$  que sai do reservatório de propagação pode ser calculada como:

$$Q_r = R \left\{ 1 - \left[ 1 + \left( \frac{R}{X_3} \right)^4 \right]^{-\frac{1}{4}} \right\}. \quad (5.17)$$

Ao final do passo de tempo  $t$  o nível do reservatório de produção se torna:

$$R := R - Q_r. \quad (5.18)$$

Dessa maneira, segundo Perrin et al. (2003), ao final de cada passo de tempo  $t$  o valor de  $R$  jamais ultrapassa a capacidade total do reservatório, e o reservatório de propagação é capaz de simular vazões de recessão por longos períodos.

A vazão  $Q_1$  resultante da propagação do segundo hidrograma unitário  $HU_2$  também está sujeita às trocas subterrâneas de água e, portanto, é atualizada para o maior valor entre 0 e o somatório de  $Q_1$  e  $F$ , como mostrado a seguir:

$$Q_d = \max(0; Q_1 + F). \quad (5.19)$$

Por fim, a vazão total  $Q$  é obtida calculando

$$Q = Q_r + Q_d. \quad (5.20)$$

O modelo GR4J foi empregado usando a biblioteca livre RRMPG (em inglês, *Rainfall-Runoff Modelling Playground*), a qual foi desenvolvida na linguagem de programação *Python* e disponibilizada por Kratzert et al. (2017).

## 5.1 CALIBRAÇÃO E VALIDAÇÃO

A identificação dos parâmetros do modelo GR4J é realizada por meio da calibração, onde os parâmetros são otimizados com base nas séries de vazões observadas. A Tabela 5.1 sintetiza e apresenta os intervalos mínimos e máximos utilizados.

Tabela 5.1: Descrições e intervalos para os parâmetros do modelo GR4J.

Parâmetros	Descrição	Intervalo
x1	Capacidade do reservatório de produção (mm)	0 a 6000
x2	Coefficiente de trocas subterrâneas (mm/dia)	-20 a 10
x3	Capacidade do reservatório de propagação (mm)	0 a 4000
x4	Tempo de base do hidrograma unitário (dias)	0,04 a 20

O método de otimização global *Differential Evolution* (DE), o qual foi usado para a calibração do GR4J, foi inicialmente proposto por Storn e Price (1997a) e está incluso dentro da biblioteca *Scipy* em *Python*. O DE é empregado em problemas de otimização que utilizam uma única função objetivo. A partir de uma população de tamanho  $N$  constante expressa na forma de uma matriz

$$X_{n,i}^G = [x_{n,1}^G, x_{n,2}^G, \dots, x_{n,D}^G], \quad n = 1, 2, \dots, N \quad e \quad i = 1, 2, \dots, D. \quad (5.21)$$

Onde, cada indivíduo  $x_n^G$  da geração  $G$  equivale a um conjunto de  $D$  parâmetros. A população inicial é escolhida de maneira aleatória usando os limites inferiores e superiores dos parâmetros. Em seguida, são gerados novos conjuntos de parâmetros que darão origem a uma população de indivíduos com mutações,  $M_{n,i}^{G+1}$ . Esses indivíduos com mutações são concebidos por meio da diferença ponderada entre dois indivíduos,  $x_b^G$  e  $x_c^G$ , somada a um terceiro,  $x_a^G$ , como é expresso na equação a seguir:

$$M_{n,i}^{G+1} = x_a^G + f(x_b^G - x_c^G). \quad (5.22)$$

Os indivíduos  $x_a^G$ ,  $x_b^G$  e  $x_c^G$  são escolhidos de forma aleatória. O fator de mutação,  $f$ , é um valor constante que varia entre  $[0, 2]$  que controla quão abrangente será a busca por novos conjuntos de parâmetros. Posteriormente, uma população teste ( $T_{n,i}^{G+1}$ ) é produzida a partir do cruzamento aleatório entre os indivíduos das populações  $M_{n,i}^{G+1}$  e  $X_{n,i}^G$ . Os indivíduos da população teste são formados usando a equação

$$T_{n,i}^{G+1} = \begin{cases} M_{n,i}^{G+1} & \text{se } rand(i) \leq CR \text{ ou } i = I_{rand} \\ X_{n,i}^G & \text{se } rand(i) > CR \text{ ou } i = I_{rand}. \end{cases} \quad (5.23)$$

Sendo que para o  $i$ -ésimo parâmetro é selecionado um valor entre  $[0, 1]$  a partir de um gerador de números aleatórios com distribuição uniforme,  $rand(i)$ .  $CR$  é uma constante que indica a probabilidade de recombinação, a qual pode ser escolhida pelo usuário entre os valores de  $[0, 1]$ .  $I_{rand}$  é um valor escolhido de forma aleatória entre  $[1, 2, \dots, D]$ . De acordo com Storn e Price (1997b), a equação 5.23 assegura que para cada indivíduo da população teste,  $t_n^{G+1}$ , haverá ao menos um parâmetro originado pelos indivíduos da população com mutação.

Por último, é realizado o processo de seleção. Nesse procedimento cada indivíduo de  $X_{n,i}^G$  é comparado com o da população  $T_{n,i}^{G+1}$ . O conjunto de parâmetros com menor valor de função objetivo é selecionado para a próxima geração  $X_{n,i}^{G+1}$ , como é apresentado na equação

$$x_n^{G+1} = \begin{cases} t_n^{G+1} & \text{se } f(t_n^{G+1}) < f(x_n^G) \\ x_{n,i}^G & \text{se } f(t_n^{G+1}) > f(x_n^G). \end{cases} \quad (5.24)$$

O período de simulação foi dividido em 3 partes: aquecimento, calibração e validação. Os primeiros 5 anos da simulação foram usados como aquecimento para eliminar as incertezas das condições iniciais (Daggupati et al., 2015). Os períodos de calibração e validação foram separados em 70% e 30%, respectivamente.

Segundo Krause et al. (2005) e Muleta (2012), o uso do logaritmo do coeficiente de Nash-Sutcliffe (logNSE) como função objetivo sofre maior influência das vazões baixas e, portanto, essa métrica pode ser utilizada para avaliar a performance das predição de vazões mínimas. O logNSE pode variar de  $-\infty$  (ajuste de má qualidade) a 1,0 (ajuste perfeito), e pode ser calculado como:

$$\logNSE = \frac{\sum_{i=1}^N (\ln(Q_{s,i} + 0,001) - \ln(Q_{o,i} + 0,001))^2}{\sum_{i=1}^N (\ln(Q_{o,i} + 0,001) - \bar{Q}_{o,\ln})^2}, \quad (5.25)$$

onde  $Q_s$  e  $Q_o$  correspondem à vazão simulada e observada, respectivamente.  $\bar{Q}_o$  é a vazão média das vazões observadas.

## 6 MÉTODOS DE REGIONALIZAÇÃO

Neste trabalho foram empregadas as técnicas de regionalização clássicas baseadas na similaridade fisiográfica-climática, na proximidade espacial simples e na regressão não linear.

A regionalização baseada na similaridade fisiográfica-climática inicia identificando e agrupando as bacias hidrográficas que possuem maiores semelhanças físicas, climáticas e geográficas. O propósito do agrupamento é a identificação de regiões homogêneas com base nos índices dos descritores. A regionalização baseada na proximidade espacial simples considera que a região de estudo é homogênea e, assim, as bacias próximas são semelhantes com base no clima, no relevo, na vegetação, na paisagem e no tipo de solo. Embora ambas supõem que as semelhanças físicas podem estar estreitamente correlacionadas com as respostas hidrológicas, se a região for heterogênea, a regionalização baseada na similaridade fisiográfica-climática transfere informações entre bacias que não são necessariamente vizinhas geograficamente. Uma segunda suposição a ser considerada é: a similaridade entre os conjuntos de parâmetros calibrados entre duas ou mais bacias hidrográficas pode refletir na semelhança de seu comportamento em relação a transformação de chuva em vazão (Oudin et al., 2008, 2010; Parajka et al., 2005; Blöschl et al., 2013).

Por outro lado, os métodos de regressão consideram que os parâmetros do modelo hidrológico podem estar relacionados com alguns processos físicos que ocorrem nas bacias hidrográficas e, conseqüentemente, estão associados a algumas características descritivas (Arsenault et al., 2019). Dessa maneira, é possível construir um modelo de regressão para cada parâmetro do modelo.

O diagrama da Figura 6.1 resume brevemente a aplicação dos métodos de regionalização que se seguiram. Após a calibração do modelo GR4J para cada uma das 126 bacias hidrográficas, essas foram divididas aleatoriamente em conjuntos de treinamento e de validação, sendo que 80% das bacias da amostra inicial compõem o conjunto de treinamento e 20% reúne as bacias de validação.

O conjunto de treinamento é formado pelas bacias hidrográficas consideradas como possíveis doadoras de parâmetros do modelo hidrológico GR4J, e que foram também utilizadas para treinar e construir os modelos de regionalização. O papel a ser desempenhado pelas bacias pertencentes ao conjunto de validação é o de não instrumentalizadas (indicado pela seta azul no diagrama da Figura 6.1), mesmo que se tenha o conhecimento que essas bacias possuem dados hidrometeorológicos e que já foram calibradas pelo método descrito na seção anterior. A partir dos modelos de regionalização, os quais foram treinados pelo conjunto de treinamento, são feitas estimativas dos possíveis parâmetros para as bacias do conjunto de validação. Por último, com base nos parâmetros calibrados e regionalizados para as bacias pertencentes ao agrupamento de validação, serão avaliados os desempenhos dos métodos de regionalização, e adicionalmente as vazões simuladas serão confrontadas com o método de regressão para a estimação da vazão de referência Q95.

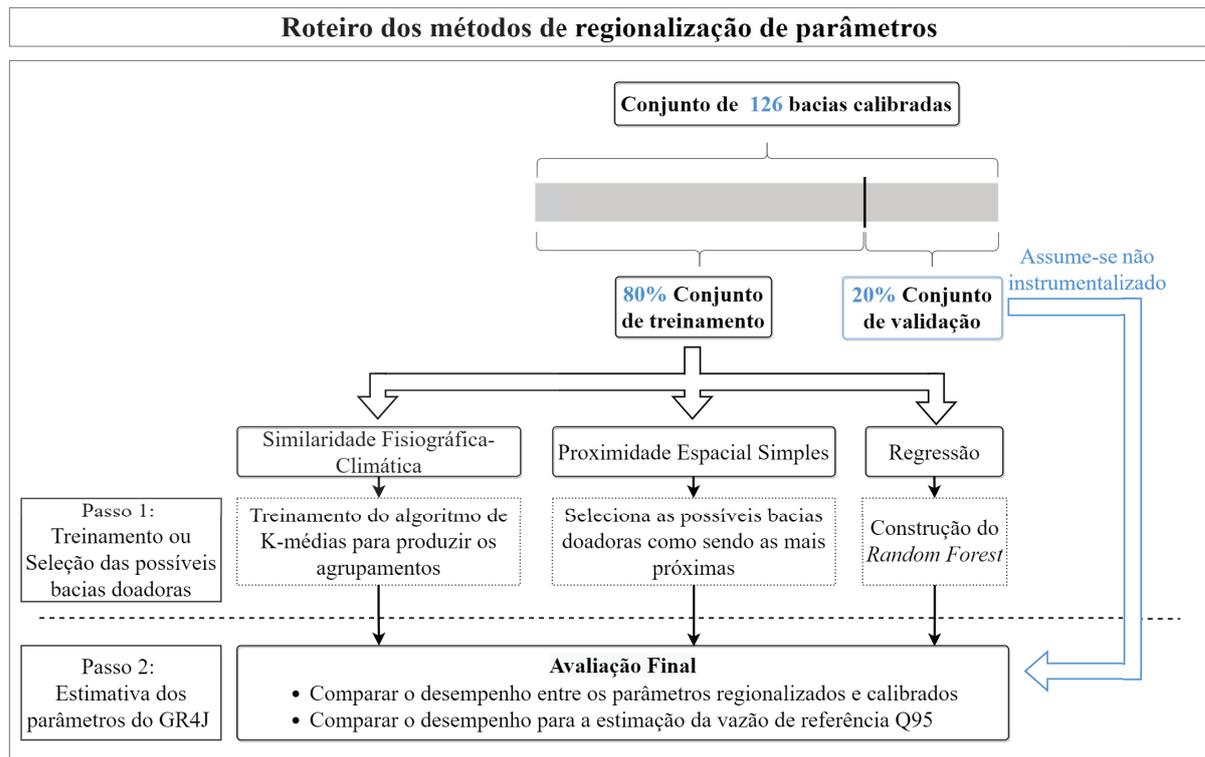


Figura 6.1: Diagrama sintetizando os métodos de regionalização. Fonte: Autor (2022).

O estudo apresentado por Oudin et al. (2008), mostrou que limitar a escolha das possíveis bacias doadoras de parâmetros com base no desempenho do modelo durante a validação, equivalentes ou superiores a 0,70 para o coeficiente de Nash-Sutcliffe (NSE), não garante altos índices de eficiência, no entanto, permite que a performance dos métodos de regionalização não sejam afetados. O impacto de escolher um limiar de desempenho superior a 0,70, durante o período de validação, pode vir a restringir o tamanho da amostra de possíveis bacias doadoras de parâmetros a um ponto em que não haja mais representatividade. Inversamente, utilizar bacias doadoras com desempenhos inferiores, durante o período de validação, pode vir a aumentar a variabilidade dentro da amostra de bacias doadoras, mas não garante fidedignidade aos parâmetros calibrados para o conjunto de possíveis bacias doadoras de parâmetros. Dessa maneira, para a construção dos modelos de regionalização foram utilizadas as bacias hidrográficas que tiveram no período de validação um  $\log NSE$  igual ou superior a 0,70.

Oudin et al. (2008) revelaram que a escolha do número de bacias doadoras também pode influenciar diretamente no desempenho e na predição de vazão das bacias alvo. Assim, para os métodos de regionalização baseados em distâncias para cada bacia alvo foram selecionadas de 1 (uma) a 10 (dez) bacias doadoras mais próximas. No caso da distância espacial foram escolhidas as bacias que estão mais perto. Já a similaridade considerou as bacias doadoras mais semelhantes e que pertencem ao mesmo agrupamento da bacia alvo. Logo, o conjunto de parâmetros de cada bacia doadora foi utilizado para estimar a vazão nas bacias consideradas não instrumentalizadas, sendo que foi feita a transferência completa dos conjuntos e foram atribuídos pesos iguais para cada bacia. Uma vez simuladas as vazões com os parâmetros das bacias doadoras, foram calculadas as médias das vazões modeladas para comparar com a vazão observada, no mesmo período em que o modelo calibrado foi validado.

## 6.1 SIMILARIDADE FISIAGRÁFICA-CLIMÁTICA

Algumas particularidades dos métodos PUBs, são: a unicidade de cada região no globo e as informações disponíveis em cada conjunto de dados. Tendo em vista a singularidade de cada local, a escolha do conjunto de descritores das bacias foi tomada com cuidado para que sintetizassem diferentes pontos de vista sobre as bacias hidrográficas e que fossem capazes de transmitir a diversidade entre as bacias dentro de uma mesma amostra. Desse modo, inicialmente foram selecionados seguintes descritores: a área, o comprimento do rio principal, a altitude e a declividade média das bacias, a latitude do centroide das bacias, as médias diárias de precipitação e evapotranspiração potencial, o índice de aridez, a duração média de dias com eventos extremos de precipitação e a fração de área coberta por floresta, agricultura e urbanização.

Inicialmente os descritores foram normalizados de maneira que a média e o desvio padrão correspondessem a zero e um, respectivamente. Esse procedimento garante que diferentes variáveis compartilhem uma mesma escala sem que haja perdas de informações significativas e, assim, permite que categorias com grandezas diferentes possam ser comparadas igualmente. Em seguida, foram selecionadas características que apresentavam variabilidades, ou seja, que descrevessem o conjunto de bacias como sendo heterogêneo.

Observando as estatísticas descritivas anteriormente apresentadas na Tabela 4.4, notou-se que as bacias paranaenses são diversificadas sob o ponto de vista de área, comprimento do rio principal e altitude média das bacias. Por outro lado, a fração de área urbana mostrou ter uma variação pouco significativa; apesar disso ocorrer foi preferível manter esse descritor, já que as infraestruturas urbanas como também outras atividades antropogênicas podem perturbar seriamente os processos do ciclo hidrológico e, assim, se torna uma fator importante para identificação de possíveis bacias afetadas pelo processo de urbanização.

As altas multicolinearidades entre os descritores podem levar os algoritmos de agrupamentos a tomar decisões erradas durante a formação dos grupos (Boutsidis et al., 2014). Por conseguinte, foram feitas duas análises para identificar as correlações entre os descritores. Primeiramente, foram calculados os coeficientes de correlação de Pearson ( $R$ ) entre cada par de descritores. Em segundo, o Fator de Inflação de Variância (FIV) foi calculado para medir o grau de multicolinearidade entre os descritores. O FIV varia entre 1 (quando não há multicolinearidade) até infinito (quando há multicolinearidade perfeita), o limiar utilizado neste trabalho foi 5.

Dentre as bacias paranaenses chamou atenção as altas correlações, com valores de  $R$  acima de 0,70, entre os pares: o índice de aridez e os dias de precipitação com acumulado mensal de 150mm, a duração média de dias sem chuva e a latitude do centroide das bacias, a declividade média da bacia e a fração de floresta, a fração de área de agricultura e a fração de floresta, a evapotranspiração potencial anual e a altitude média da bacia. Essas e outras correlações podem ser observadas na Figura A.10 do Apêndice A. Após as análises foram considerados os seguintes descritores: a área, a fração de floresta, a fração de área urbana, a duração média de eventos extremos com altas precipitações (dias de precipitação com acumulado mensal de 150mm) e a duração média de dias sem chuva.

A distância Euclidiana ( $dist$ ), a qual é expressa na equação 6.1, é uma métrica que pode expressar semelhanças (distâncias pequenas) ou diferenças (distâncias grandes) entre os atributos  $k$  de duas bacias ( $a$  e  $b$ ) em um espaço  $N$ -dimensional de atributos, podendo, também, ser aplicada em conjunto com os métodos de classificação ou de agrupamento (Viviroli et al., 2009).

$$dist(a,b) = \sqrt{\sum_{k=1}^n [atrib_k(a) - atrib_k(b)]^2} \quad (6.1)$$

Os agrupamentos foram produzidos por meio do método de K-médias, o qual foi implementado utilizando o pacote *scikit-learn* desenvolvido em *Python*. A aplicação do algoritmo K-médias envolve: primeiro, definir previamente o número de  $K$  grupos, segundo, para cada grupo inicializar um centroide aleatoriamente dentro do intervalo de cada categoria, terceiro, cada ponto é atribuído ao centroide que tiver a menor distância Euclidiana em relação a ponto, quatro, é computado uma nova localização dos  $K$  centroides com base na média de todos os pontos atribuídos a ele. O processo iterativo do terceiro ao quarto passo se repete até que não haja mais alterações nos centroides (Wilks, 2011).

O valor de  $K$  afeta diretamente como os grupos serão formados. Quanto maior o valor de  $K$  mais grupos serão formados, contudo menores esses serão (o que traz homogeneidade, mas não garante representatividade). Por outro lado quanto menor a quantidade de grupos, maiores esses serão (o que não permite identificação adequada dos diferentes grupos). Uma maneira de avaliar os agrupamentos resultantes do método de aglomeração é a partir do coeficiente de Silhueta ( $Si$ ) e do método *Elbow* ou método do cotovelo.

De acordo com Rousseeuw (1987), o coeficiente de Silhueta ( $Si$ ) consiste em calcular a distância Euclidiana média ( $a_p$ ) de um ponto  $p$  com todos os pontos pertencentes ao mesmo grupo. Em seguida, é calculada a distância média ( $b_p$ ) do ponto  $p$  com relação a todos os pontos pertencentes ao grupo vizinho mais próximo. Assim, o coeficiente pode ser calculado utilizando a equação:

$$Si = \frac{b_p - a_p}{\max(b_p, a_p)}. \quad (6.2)$$

O  $Si$  pode variar entre  $[-1, 1]$ , sendo que quanto mais próximo de 1 mais distante o ponto  $p$  está do grupo vizinho. Valores próximos de 0 indicam que o ponto  $p$  está próximo do limite que divide ambos os grupos. E medidas próximas a  $-1$  indicam que o ponto  $p$  pode ter sido associado ao grupo errado.

O método do cotovelo é outro método gráfico de se avaliar um número ótimo de agrupamentos. Essa técnica envolve calcular um coeficiente de aglomeração, neste trabalho o critério calculado foi a soma das distâncias ao quadrado de cada amostra ( $x_i$ ) com o respectivo centroide ( $\mu_j$ ) do agrupamento que a amostra faz parte, podendo ser expresso como a soma dos erros ao quadrado:

$$SSE = \sum_{i=1}^n (x_i - \mu_j)^2. \quad (6.3)$$

Conforme o número de agrupamentos cresce as distâncias entre as amostras com os respectivos centroides diminuem. Todavia, espera-se que o número de grupos e que o coeficiente de aglomeração sejam pequenos. Dessa maneira, a partir de um gráfico com o coeficiente de aglomeração do eixo  $y$  e o número de grupos no eixo  $x$  é possível identificar o ponto em que há um achatamento acentuado ou uma queda rápida desse coeficiente, sugerindo o número ótimo de agrupamentos (Ketchen Junior e Shook, 1996).

## 6.2 PROXIMIDADE ESPACIAL SIMPLES

A distância entre dois pontos, neste caso os centroides das bacias alvo e doadora, que possuem latitude e longitude conhecidas pode ser calculada usando a distância de Haversine. Assim, a equação pode ser expressa como:

$$D_H = 2r \arcsin \left[ \sqrt{\sin^2 \left( \frac{x_1 - x_2}{2} \right) + \cos(x_1) \cos(x_2) \sin^2 \left( \frac{y_1 - y_2}{2} \right)} \right], \quad (6.4)$$

onde  $D_H$  se refere a distância em km,  $r$  é o raio da Terra igual à 6371 km,  $x$  e  $y$  são respectivamente as latitudes e as longitudes dos pontos 1 e 2.

## 6.3 REGRESSÃO

Os modelos de regressão, sejam regressões múltiplas lineares ou não lineares, buscam equacionar a relação entre a variável dependente (parâmetros do modelo) com diferentes variáveis independentes (características descritivas das bacias). O método de regressão não-linear *Random Forest* (Breiman, 2001), o qual foi o método escolhido para este trabalho, é apto a apresentar bons desempenhos ao lidar com um conjunto grande de dados e é capaz de distribuir pesos às variáveis independentes de acordo com o seu grau de importância. Assim, foram construídos dois métodos de regressões diferentes.

A formulação do primeiro método chamado de *Random Forest I* deu-se, somente, a partir das informações conhecidas das bacias do conjunto de treino. Ao todo foram utilizadas 1.000 árvores de decisões e as características descritivas das bacias do conjunto de treinamento, que receberam pesos superiores à 0,01.

Para essa técnica é necessário produzir um modelo de regressão de forma independente para cada parâmetro ( $X_1$ ,  $X_2$ ,  $X_3$  e  $X_4$ ), contudo os parâmetros de um modelo hidrológico geralmente apresentam relações dependentes entre si, e por vezes não podem ser observados de maneira independente. Desse modo, o segundo método definido como *Random Forest II* incluiu os parâmetros calibrados das bacias de treino. O desenvolvimento do segundo método seguiu os seguintes passos:

- Fez-se a análise de correlação entre os parâmetros, e assim foi determinada a ordem dos parâmetros com os maiores índices de correlação, e
- Iniciou-se a regressão pelo parâmetro com menor índice de correlação, nesse caso foram empregadas apenas as características descritivas. Em seguida, fez-se a regressão para o parâmetro com segundo menor índice de correlação, aqui fez-se uso das características descritivas e o parâmetro anterior que possui o menor índice de correlação. Esse processo se seguiu até que todos os parâmetros tivessem suas regressões.

## 6.4 ESTIMATIVA DA VAZÃO Q95

O interesse sobre a estimação de vazões de referência em bacias pouco ou não instrumentalizadas, se dá pelo fato do órgão ambiental responsável pelas outorgas de recursos hídricos no Estado do Paraná, o Instituto Água e Terra (IAT), utilizar a vazão com permanência de 95% (Q95) como vazão de referência nas análises dos processos de

outorga (AGUASPARANÁ, 2010). A partir desta etapa espera-se estimar diretamente a vazão Q95 por meio de técnicas de regressão e, assim, comparar com os meios indiretos (vazões simuladas por meio da regionalização dos parâmetros do modelo GR4J).

Dando início à regressão da vazão de referência foram produzidas as curvas de permanência para as bacias presentes no grupo de treinamento. Utilizando todos os  $N$  registros disponíveis nos períodos em que o GR4J foi calibrado e validado, a construção da curva de permanência envolveu: (i) ordenar de forma crescente as vazões  $Q$ , (ii) atribuir para cada vazão ordenada  $Q_m$  a ordem de classificação  $m$  correspondente, (iii) computar a frequência ou probabilidade das vazões ordenadas  $Q_m$  de serem igualadas ou superadas  $P(Q \geq Q_m)$ , a qual pode ser calculada utilizando a posição de plotagem de Weibull mostrada na equação a seguir (Pugliese et al., 2014):

$$P(Q \geq Q_m) = 1 - \frac{m}{N + 1}. \quad (6.5)$$

Após a obtenção das vazões de referência Q95 em  $\text{m}^3/\text{s}$ , para as bacias pertencentes ao conjunto de treinamento, foi feita a transformação para a unidade de  $\text{l/s/km}^2$ , garantindo que a variável não seja dependente da área da bacia. Em seguida, o método de regressão de *Random Forest* foi treinado utilizando 1.000 árvores de decisão e os descritores das bacias do conjunto de treinamento que apresentassem pesos superiores a 0,01.

## 7 RESULTADOS E DISCUSSÕES

### 7.1 CONJUNTO DE DADOS DO ESTADO DO PARANÁ

O conjunto de dados do Estado do Paraná alcançou um total de 284 bacias hidrográficas com dados fluviométricos que variam de 1926 a 2020. Uma maneira de comparar a quantidade de informações fluviométricas aqui coletadas com outros conjuntos desenvolvidos a nível nacional (CABra e CAMELS-BR), é comparando o número de bacias hidrográficas encontradas dentro da mesma área de estudo definida neste trabalho, que possuem dados entre 1979 a 2020.

Para efeito de comparação, usando o conjunto de dados do Estado do Paraná, foram contabilizadas as estações que possuíssem ao menos 30 anos de medições de vazão entre 1979 a 2020, sendo que cada ano contabilizado contém no máximo 10% de falhas, alcançou-se a quantia de 102 estações. Nos conjuntos de dados do CABra e do CAMELS-BR, dentro da área de estudo definida neste trabalho, foram identificadas 58 estações com registros de 1980 a 2010 e 85 estações com observações de 1980 a 2018, respectivamente.

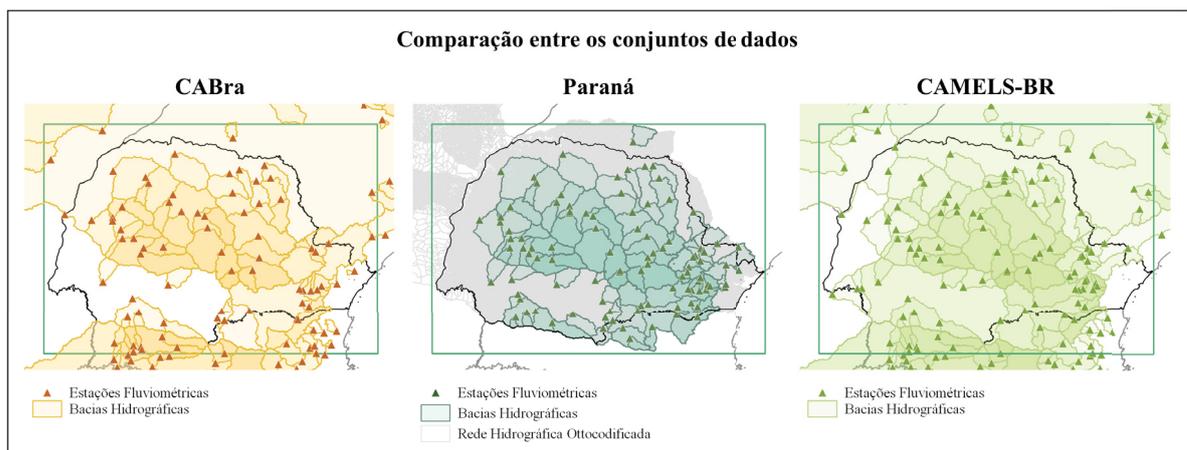


Figura 7.1: Comparação entre o CABra (lado esquerdo), o conjunto do Paraná (central) e o CAMELS-BR (lado direito).

Em sua grande maioria, desconsiderando as estações a jusante de regularizações, as estações presentes nos outros conjuntos de dados também foram incluídas no conjunto desenvolvido para o Paraná, como pode ser observado no mapa da Figura 7.1.

### 7.2 DESEMPENHO DO MODELO GR4J

O modelo GR4J apresentou desempenhos aceitáveis para a amostra de 126 bacias hidrográficas como é mostrado na Figura 7.2, sendo que cerca de 65% das bacias paranaenses apresentaram  $\log NSE$  igual ou superior à 0,70 no período da validação. Observa-se que as bacias próximas ao litoral do Paraná alcançaram uma menor eficiência quando comparadas às outras regiões. Outro ponto importante, é que algumas bacias hidrográficas apresentaram desempenhos superiores no período de validação quando comparadas com o período de calibração, contudo situações inversas também ocorrem. Supõe-se que esses fenômenos

podem estar associados como consequência da mudança ou da melhoria das técnicas e equipamentos de medição, como também podem ter sofrido influências das alterações do uso e cobertura do solo.

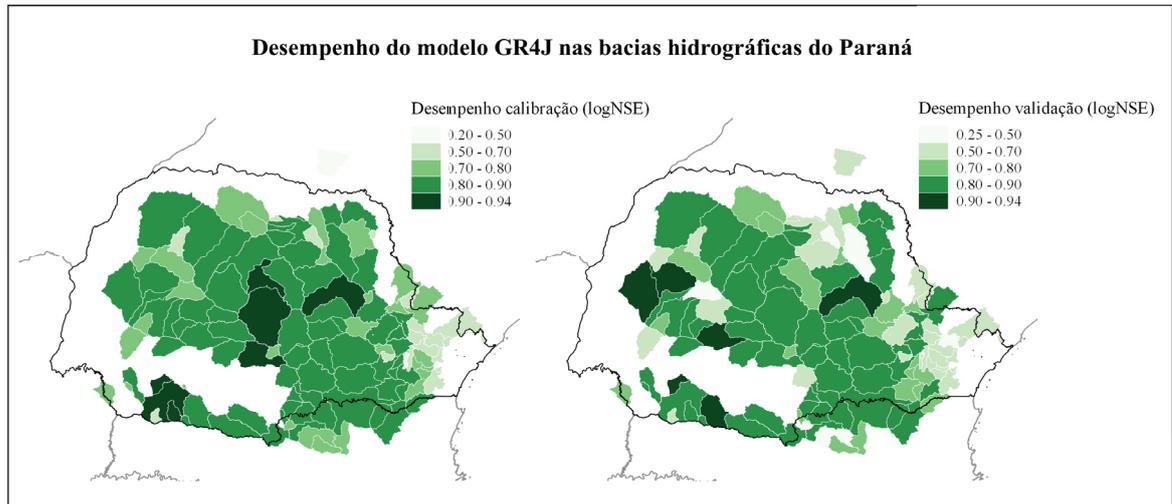


Figura 7.2: Desempenho do modelo calibrado no período da calibração e da validação no estado do Paraná.

### 7.3 DESEMPENHO DOS MÉTODOS DE REGIONALIZAÇÃO

A seguir são descritos um a um os resultados dos métodos de regionalização e por último é feita a comparação do desempenho entre eles.

#### 7.3.1 Similaridade Fisiográfica-Climática

O desenvolvimento dos agrupamentos no método de regionalização por similaridade fisiográfica-climática foi dividido em duas etapas. A primeira etapa utilizou 80% da amostra de 126 bacias hidrográficas para o treinamento do algoritmo de K-médias, o qual foi usado para agrupar as bacias semelhantes. Já os 20% restantes foram utilizados para testar e avaliar os agrupamentos formados. Assim, na fase do treinamento o número de agrupamentos sugerido pelo método *Elbow* corresponde a 6, onde ocorre o primeiro ponto de queda rápida ou achatamento da curva na Figura 7.3. Em concordância, os valores médios dos coeficientes de Silhueta ( $S_i$ ) calculados para cada valor de  $K$  foi superior quando esse corresponde à 6, como é mostrado na Figura 7.4.

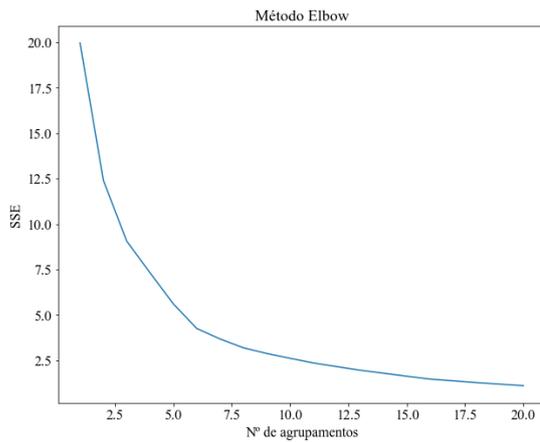


Figura 7.3: Método *Elbow* para as bacias de treinamento com  $K$  variando de 1 a 20.

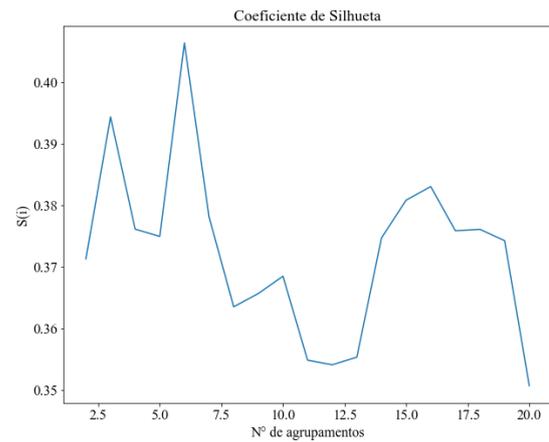


Figura 7.4: Coeficiente de Silhueta para as bacias de treinamento com  $K$  variando de 1 a 20.

O algoritmo de K-médias é um método dependente da distância Euclidiana. Portanto, observou-se que conforme o número de descritores aumentava maiores se tornavam as distâncias entre as bacias hidrográficas. Em outras palavras, quanto maior o número de descritores mais únicas as bacias hidrográficas se tornam dificultando o processo de aglomeração. A distribuição dos agrupamentos formados pelo algoritmo de K-médias pode ser observada na Figura 7.5, tanto para o conjunto de bacias de treinamento quanto para as bacias do conjunto de validação. Nota-se que a partir do algoritmo treinado alcançou-se agrupamentos com distribuições espaciais semelhantes para as bacias do conjunto de validação. Além disso, observa-se que nem sempre as bacias mais próximas geograficamente pertencem ao mesmo agrupamento formado.

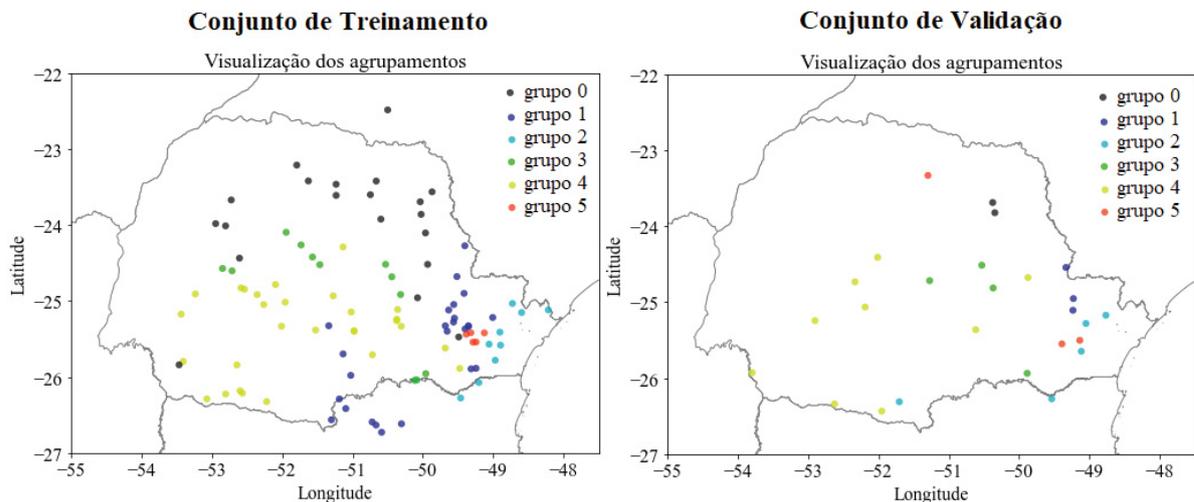


Figura 7.5: Agrupamentos produzidos pelo método K-média para as 100 bacias de treinamento (lado esquerdo) e 26 bacias de validação (lado direito).

A relação dos grupos do conjunto de treinamento e os descritores são apresentados na Figura 7.6. Diferente dos demais, o grupo 3 contém as bacias com maiores áreas de drenagem. No grupo 4 estão contidas as bacias que possuem áreas de drenagem menores, mas que acabam compartilhando características semelhantes ao do grupo 3. Já os grupos 1 e 2 possuem maior percentual de floresta, sendo que o grupo 2 possui uma tendência

maior de ter chuvas mais volumosas e períodos menores de dias secos consecutivos. Por outro lado, o grupo 0 se destaca por conter as bacias que possuem maior duração média de dias secos consecutivos. Por último, o grupo 5 se destaca das demais por conter as bacias com maiores porcentagens de área urbana e, portanto, podem sofrer maiores influências das atividades antropogênicas. Embora os descritores apontem heterogeneidades entre os grupos formados, ainda é possível ver sobreposições, principalmente, com relação aos parâmetros calibrados do modelo GR4J como é mostrada na Figura 7.7.

Os hidrogramas e as curvas de permanência, de vazões observadas, os quais são apresentados a seguir, indicam que as bacias hidrográficas classificadas como semelhantes com base nos seus atributos podem reproduzir vazões com comportamentos parecidos. Os hidrogramas das bacias do grupo 0, as quais estão localizadas mais ao norte do Estado, são apresentados na Figura 7.8. Observa-se que a Figura 7.8 indica que existem alguns períodos do ano mais chuvosos, onde percebe-se que há vazões mais altas, podendo ser seguidos por épocas mais secas em que há períodos longos de recessão. Como observado anteriormente o grupo 0 apresenta maior duração média de dias secos consecutivos e menores chances de ter chuvas mais volumosas se comparadas, por exemplo, com as bacias do litoral, que são representadas pelo grupo 2.

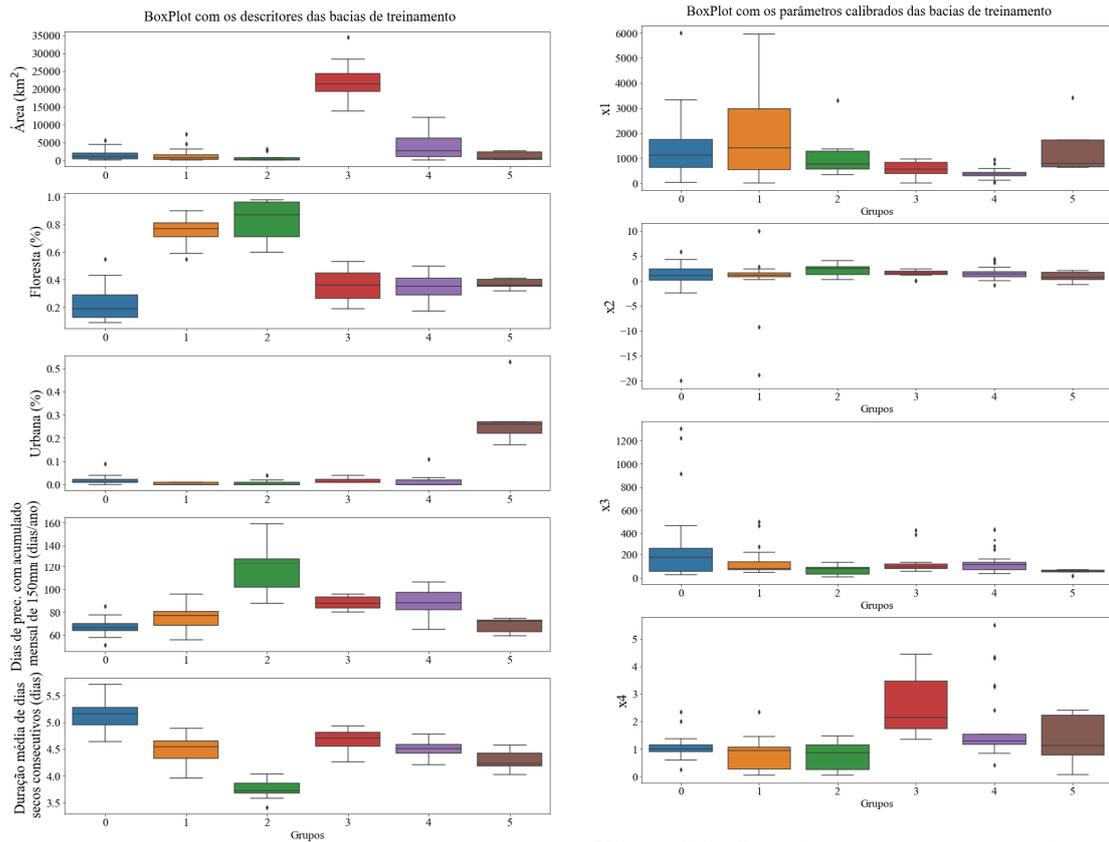


Figura 7.6: Relação dos grupos do conjunto de treinamento e os seus descritores.

Figura 7.7: Relação dos grupos do conjunto de treinamento e os seus parâmetros calibrados do GR4J.

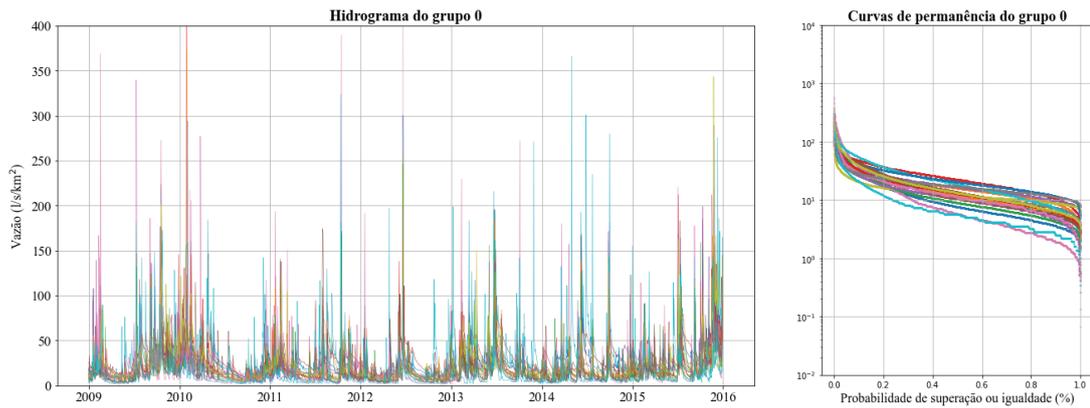


Figura 7.8: Comparação dos hidrogramas e curvas de permanência entre bacias hidrográficas do grupo 0 pertencentes ao conjunto de dados do Paraná. Tanto os hidrogramas quanto as curvas de permanência foram produzidos com base nos registros de vazão observados no período de 1979 a 2020.

As bacias do grupo 1, as quais estão localizadas tanto no primeiro e no segundo planalto, apresentam ter tamanhos semelhantes com as bacias do grupo 0, e têm como diferencial bacias com maior percentual de floresta. Como foi observado no grupo 0, as bacias do grupo 1 também apresentam ao longo do ano períodos mais chuvosos e outros mais secos. Adicionalmente, devido à tendência de ter eventos de precipitações mais volumosos nota-se vazões de base, em média, levemente superiores, e picos mais frequentes e mais intensos.

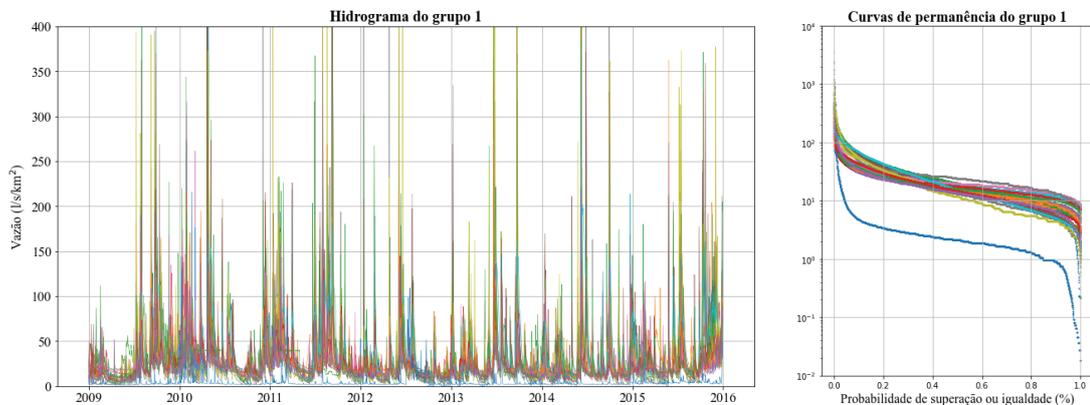


Figura 7.9: Comparação dos hidrogramas e curvas de permanência entre bacias hidrográficas do grupo 1 pertencentes ao conjunto de dados do Paraná. Tanto os hidrogramas quanto as curvas de permanência foram produzidos com base nos registros de vazão observados no período de 1979 a 2020.

As bacias do grupo 2 se encontram principalmente na região litorânea do Paraná, onde está localizada a serra do mar. Devido ao volume de precipitação ser maior nesse local, observa-se no hidrograma da Figura 7.10 picos de vazões frequentes ao longo de todo o ano, sendo que os picos mais intensos ocorrem geralmente nos meses mais quentes do ano.

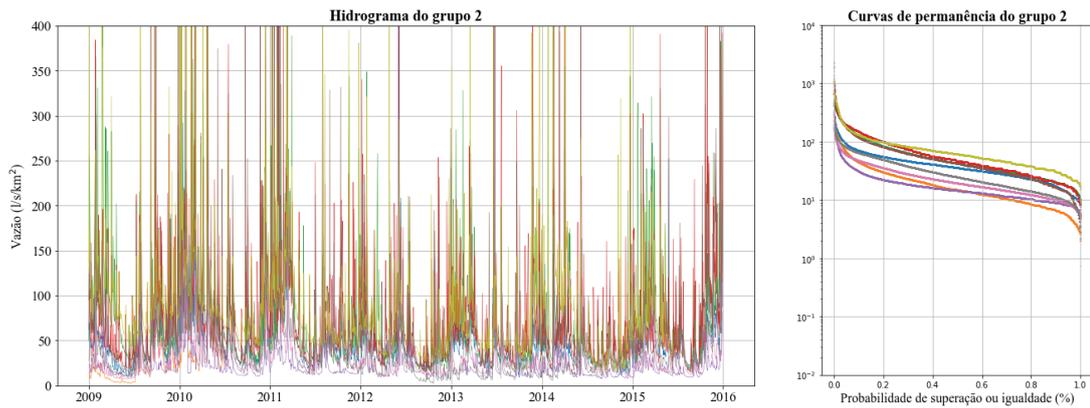


Figura 7.10: Comparação dos hidrogramas e curvas de permanência entre bacias hidrográficas do grupo 2 pertencentes ao conjunto de dados do Paraná. Tanto os hidrogramas quanto as curvas de permanência foram produzidos com base nos registros de vazão observados no período de 1979 a 2020.

O grupo 3 e 4 contém bacias hidrográficas com características descritivas parecidas, como foi apresentado na Figura 7.6, e se distinguem principalmente pela área de drenagem das bacias. O grupo 3 contém as bacias com as maiores áreas, essas se localizam nos segundo e terceiro planaltos, no centro do Estado. E o grupo 4, o qual apresenta-se em maior quantidade, contém bacias hidrográficas localizadas em todos os planaltos do Estado. Tanto o hidrograma quanto a curva de permanência apresentados na Figura 7.11 indicam que as bacias grupo 3 apresentam vazões com comportamentos semelhantes. Além disso, como visto nos grupos anteriormente citados, percebe-se nos hidrogramas das Figuras 7.11 e 7.12 que há períodos do ano em que as vazões são mais altas, indicando períodos mais chuvosos, e outras épocas mais secas, onde há recessões.

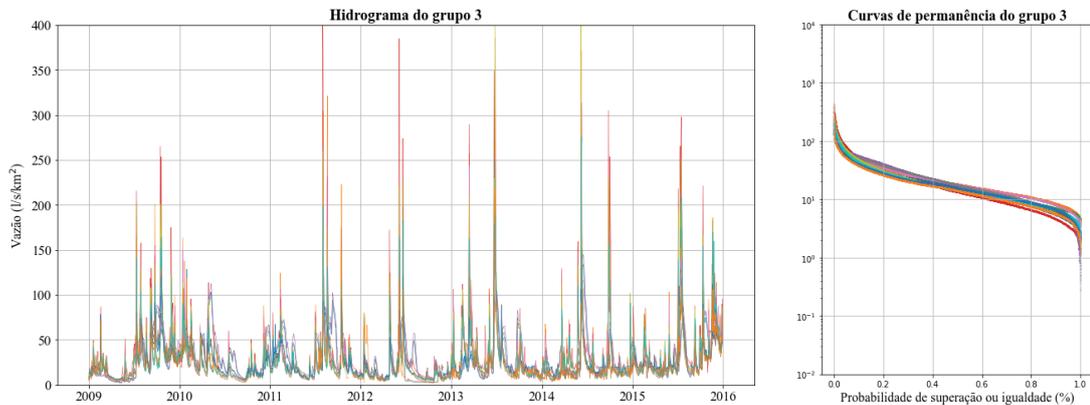


Figura 7.11: Comparação dos hidrogramas e curvas de permanência entre bacias hidrográficas do grupo 3 pertencentes ao conjunto de dados do Paraná. Tanto os hidrogramas quanto as curvas de permanência foram produzidos com base nos registros de vazão observados no período de 1979 a 2020.

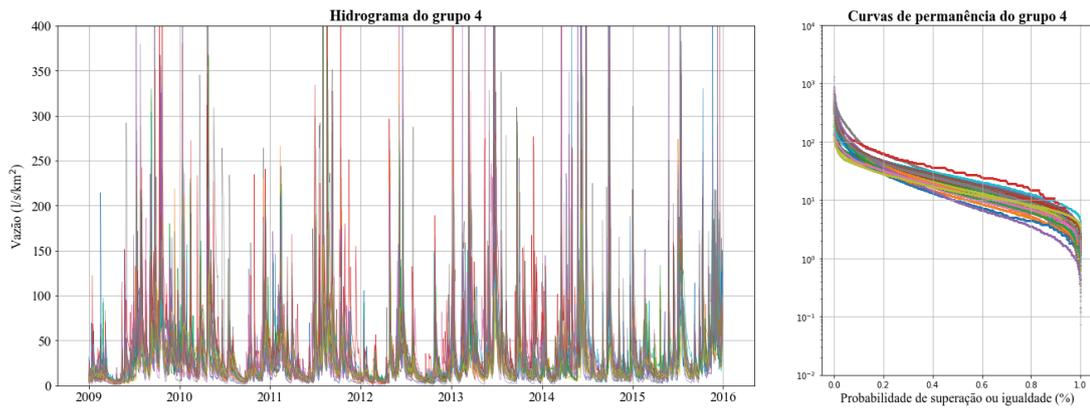


Figura 7.12: Comparação dos hidrogramas e curvas de permanência entre bacias hidrográficas do grupo 4 pertencentes ao conjunto de dados do Paraná. Tanto os hidrogramas quanto as curvas de permanência foram produzidos com base nos registros de vazão observados no período de 1979 a 2020.

Por fim, o grupo 5 inclui as bacias com maiores percentuais de áreas afetadas pela urbanização, as quais estão próximas à Curitiba e Região Metropolitana. Embora essas bacias possam ser afetadas pelas infraestruturas dos municípios e por outras ações humanas, pode se dizer, comparando os históricos de vazões apresentados pelo hidrograma e a curva de permanência da Figura 7.13, que essas bacias ainda assim apresentam comportamentos de vazões semelhantes entre si.

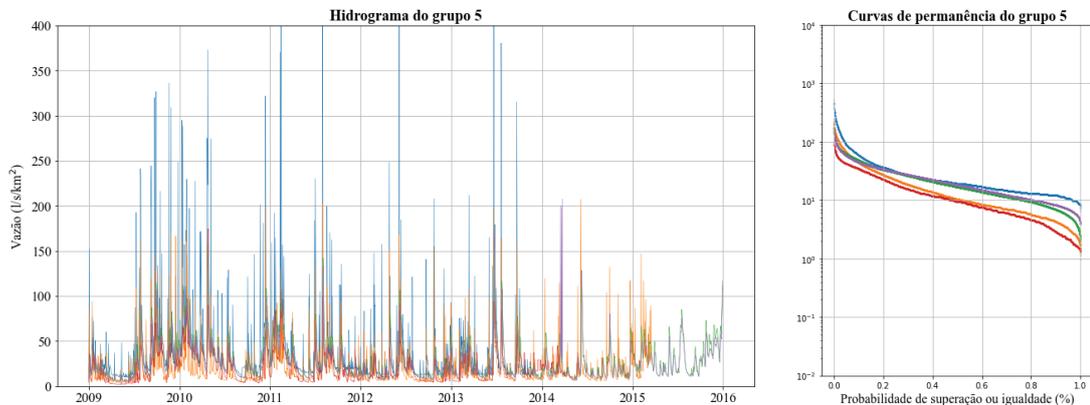


Figura 7.13: Comparação dos hidrogramas e curvas de permanência entre bacias hidrográficas do grupo 5 pertencentes ao conjunto de dados do Paraná. Tanto os hidrogramas quanto as curvas de permanência foram produzidos com base nos registros de vazão observados no período de 1979 a 2020.

### 7.3.2 Proximidade Espacial Simples

O método da proximidade espacial simples considera que a região próxima as bacias de interesse é homogênea e, portanto, somente as distâncias de Haversine entre os pares de bacias receptoras (pseudo-não monitorada) e bacias doadoras (bacias instrumentalizadas) foram consideradas como suficientes para a transferência completa dos parâmetros do modelo GR4J. Em ambos os casos, método de similaridade fisiográfica-climática e proximidade espacial simples, as bacias receptoras são todas as bacias dentro do conjunto de validação, e as possíveis bacias doadoras de parâmetros são as que fazem parte do conjunto de bacias de treinamento que alcançaram  $\log\text{NSE}$  igual ou superior a 0,70 no período da validação.

A análise para identificar o número de bacias doadoras mostrado na Figura 7.14, diz que o método de similaridade apresentou uma mediana no período de validação ótima para o total de 1 bacia, e o método de proximidade apresentou melhores resultados utilizando 7 bacias como doadoras.

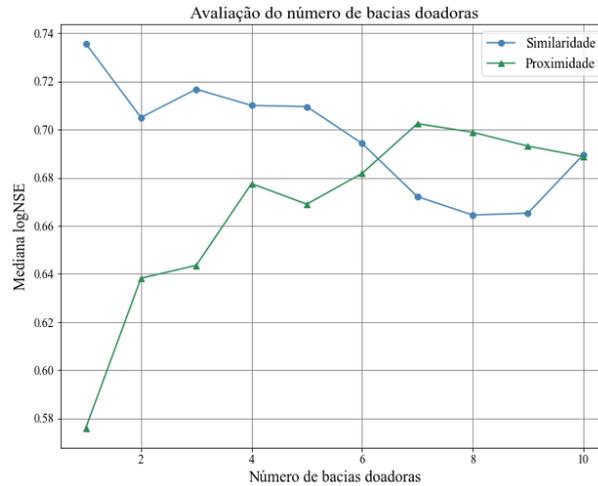


Figura 7.14: Avaliação do número de bacias doadoras com base nas medianas do logNSE durante o período de validação das simulações.

### 7.3.3 Regressão dos Parâmetros do Modelo GR4J

Os métodos de regressão dos parâmetros foram chamados de *Random Forest I* e *Random Forest II*. No método do *Random Forest I* utilizou-se, para treinar o modelo de regressão, as informações conhecidas das bacias do conjunto de treino, sendo elas: as características descritivas (variáveis independentes) e os parâmetros calibrados do modelo GR4J (variável dependente). As características descritivas que o *Random Forest I* apontou como mais relevantes foram: a declividade do rio principal e a declividade média da bacia, para o parâmetro  $X_1$ ; a altitude média e a radiação média no inverno, para o parâmetro  $X_2$ ; a radiação média no inverno, para o parâmetro  $X_3$ , e a fração de gleissolo, para o parâmetro  $X_4$ .

O *Random Forest II* também utilizou as informações conhecidas do conjunto de treino, sendo que a cada nova regressão dos quatro parâmetros foi acrescentado o parâmetro anterior como variável independente, uma vez que podem existir relações dependentes entre os parâmetros do modelo GR4J. A ordem para a construção das regressões seguiu-se do parâmetro com a correlação mais baixa para a mais alta. Segundo a tabela a seguir definiu-se a ordem como:  $X_4$ ,  $X_1$ ,  $X_3$  e  $X_2$ .

Tabela 7.1: Correlação entre os parâmetros calibrados do modelo GR4J.

	$X_1$	$X_2$	$X_3$	$X_4$
$X_1$	1,00	0,20	-0,24	-0,24
$X_2$	0,20	1,00	-0,59	0,21
$X_3$	-0,24	-0,59	1,00	0,17
$X_4$	-0,24	0,21	0,17	1,00

Assim, as características descritivas que o *Random Forest II* apontou como mais relevantes foram: a declividade do rio principal e a declividade média da bacia, para o parâmetro  $X_1$ ; os parâmetros  $X_3$  e  $X_1$  e a altitude média da bacia, para o parâmetro  $X_2$ ; a radiação média no inverno, fração de nitossolo e o parâmetro  $X_1$ , para o parâmetro  $X_3$ ; e a fração de gleissolo, para o parâmetro  $X_4$ .

#### 7.3.4 Comparando os Métodos de Regionalização

A correlação do desempenho, o qual considera o logNSE calculado para o período de validação, utilizando os parâmetros calibrados e os parâmetros regionalizados é apresentada na Figura 7.15. Observa-se que, em geral, a grande maioria das bacias que fazem parte do grupo de validação apresentou resultados com logNSE superiores a 0,50 para os diferentes métodos de regionalização de parâmetros do modelo GR4J, sendo que apenas duas bacias dentro desse grupo apresentaram ter baixo desempenho. Outro comportamento geral dentre os métodos de regionalização foi que para as bacias que apresentaram logNSE igual ou superior 0,77 utilizando os parâmetros calibrados também alcançaram desempenhos comparáveis com os parâmetros regionalizados. Adicionalmente, percebe-se, em poucos casos, que através desses métodos de regionalização é possível ter conjuntos de parâmetros diversificados, e quando combinados podem resultar em um desempenho superior se comparado com o uso dos parâmetros calibrados em um período anterior ao da validação.

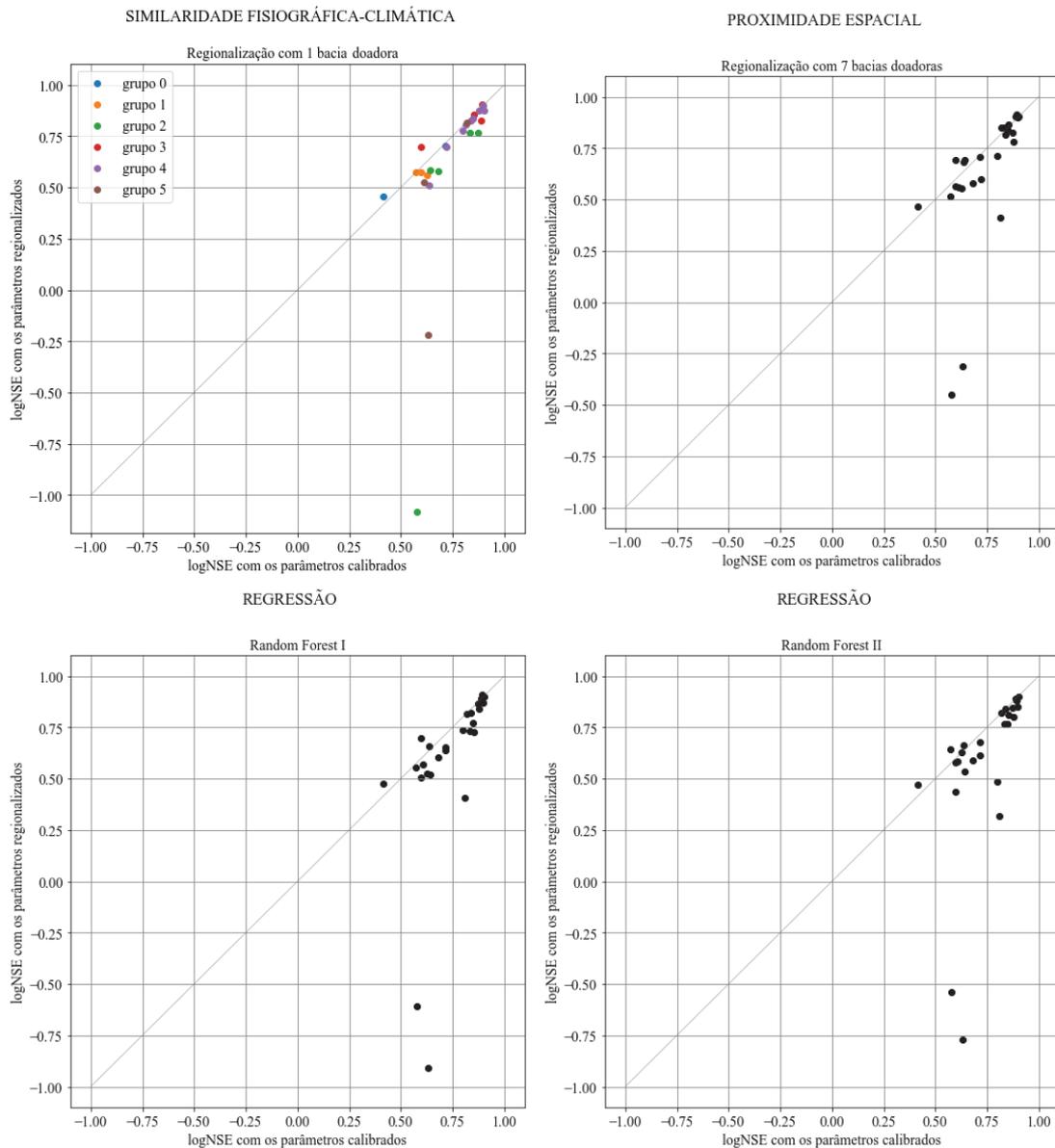


Figura 7.15: Correlações dos desempenhos ( $\log\text{NSE}$ ) da simulação durante o período de validação entre os parâmetros calibrados e os parâmetros regionalizados pela similaridade fisiográfica-climática (superior esquerdo), pela proximidade espacial (superior direito), pelo *Random Forest I* (inferior esquerdo) e pelo *Random Forest II* (inferior direito).

Os resultados foram avaliados por meio do Coeficiente de Correlação de Pearson (R), o Coeficiente de Nash-Sutcliffe (NSE) e as suas variações: a transformada da vazão pela raiz quadrada ( $\text{sqrtNSE}$ ) e pelo logaritmo ( $\log\text{NSE}$ ). O NSE dá mais ênfase na performance das vazões mais altas, já o  $\log\text{NSE}$  é mais sensível as vazões baixas e o  $\text{sqrtNSE}$  fornece um desempenho intermediário (Oudin et al., 2008). Esses índices foram calculados por meio da biblioteca *HydroErr* da linguagem *Python*.

A Tabela 7.2 apresenta os valores medianos das estatísticas de erros das vazões estimadas para todo o período de validação. Dentre os métodos de regionalização estão evidenciados em **negrito** os valores que alcançaram os melhores resultados para cada índice. Assim, a similaridade fisiográfica-climática se destaca positivamente por alcançar  $\log\text{NSE}$  e  $\text{sqrtNSE}$  iguais a 0,736 e 0,726, respectivamente. Outro ponto a ser destacado é para o método de proximidade espacial por apresentar, em geral, resultados medianos

para os três coeficientes. A Tabela 7.2 também revela que a performance do método de regionalização, em especial a proximidade e o *Ranfom Forest I*, pode alcançar valores medianos de NSE igualáveis ou superiores se comparado com os parâmetros calibrados em um período anterior ao da validação. O método de *Random Forest II* por outro lado apresentou ter baixo desempenho se comparado com os demais.

Tabela 7.2: Valores medianos das estatísticas de erros calculadas para o período de validação. Em negrito estão evidenciados os melhores resultados para cada índice dentre os métodos de regionalização.

Métricas de eficiência no período de validação					
	Calibrado	Proximidade	Similaridade	<i>Random Forest I</i>	<i>Random Forest II</i>
NSE	0,621	0,635	0,602	<b>0,643</b>	0,536
logNSE	0,758	0,702	<b>0,736</b>	0,679	0,652
sqrtNSE	0,736	0,707	<b>0,726</b>	0,713	0,693

A revisão realizada por Guo et al. (2020) englobou a análise de artigos de diferentes regiões do globo, os quais foram publicados recentemente entre os anos de 2013 e 2019, em que os pesquisadores também aplicaram técnicas semelhantes de regionalização (proximidade, similaridade e regressão). Guo et al. (2020), a partir desses estudos analisados, apresentam que existem evidências de que os métodos de regionalização baseado em distâncias (proximidade e similaridade) geralmente apresentam desempenhos superiores aos métodos de baseados em regressão.

Para fins de comparação a Tabela 7.3 apresenta as correlações (R) entre os parâmetros calibrados e os regionalizados. Para que fosse possível fazer essa análise, usou-se o conjunto de parâmetros das bacias doadoras mais próximas espacialmente para o método de proximidade e as mais similares para o método de similaridade<sup>1</sup>. As correlações (R) mostradas na Tabela 7.3 apontam que os métodos de regressão, *Random Forest I* e *Random Forest II*, apresentaram um grau significativamente alto de correlação entre os parâmetros calibrados e os que foram regionalizados, com exceção ao parâmetro  $X_2$  que apresenta ter um comportamento inversamente proporcional ao esperado. Já o método de similaridade indica maiores correlações com os parâmetros  $X_1$  e  $X_4$  e o método baseado em distância obteve maiores correlações com os parâmetros  $X_1$  e  $X_3$ . A Tabela 7.3 sugere, para as bacias do Estado do Paraná, que a proximidade/similaridade entre as bacias nem sempre é sinônimo de semelhança entre os conjuntos de parâmetros.

Tabela 7.3: Correlação de Pearson (R) entre os parâmetros calibrados e os regionalizados.

Correlação (R) entre os parâmetros				
	Proximidade	Similaridade	<i>Random Forest I</i>	<i>Random Forest II</i>
$X_1$	0,613	0,845	0,705	0,707
$X_2$	-0,320	0,357	-0,769	-0,619
$X_3$	0,896	-0,018	0,962	0,971
$X_4$	0,539	0,716	0,827	0,827

Adicionalmente, outros três critérios foram usados para avaliar a eficiência da calibração e das simulações produzidas pelas regionalizações sobre os períodos de verificação.

<sup>1</sup>As avaliações seguintes a essa, tanto para o método de proximidade quanto para o de similaridade, são referentes aos resultados obtidos usando o número ideal de bacias doadoras apontado na Figura 7.14.

Empregou-se o algoritmo desenvolvido por Hoeltgebaum (2021) e seu orientador Nelson Luís da Costa Dias, o qual foi produzido na linguagem *Python*, para identificar automaticamente os picos e os períodos das recessões dentro de uma janela móvel de 15 dias, como é exemplificado na Figura 7.16.

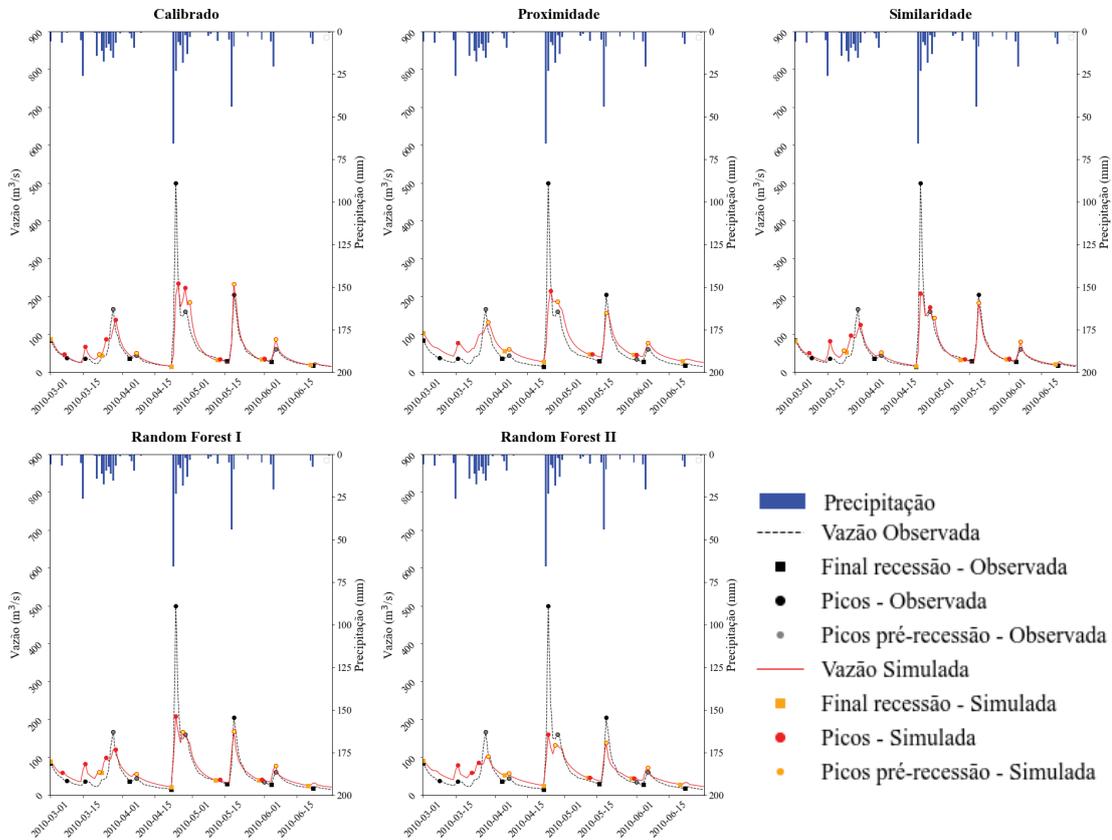


Figura 7.16: Gráficos chuva-vazão da estação Barbosa Ferraz (64659000) com as vazões observada e as simuladas utilizando os parâmetros calibrados (superior à esquerda), e os regionalizados pela proximidade (centro superior), pela similaridade (superior à direita), pelo *Random Forest I* (inferior à esquerda) e pelo *Random Forest II* (centro inferior).

Primeiramente observaram-se os picos. Para cada uma das séries históricas de vazões observadas identificou-se os dias em que ocorreram picos de vazões. Em seguida, para esses mesmos dias, a intensidade da vazão observada foi comparada com as que foram simuladas pelo modelo GR4J para os diferentes conjuntos de parâmetros regionalizados. Além disso, foram analisados o percentual de picos em fase, ou seja, após identificados os dias em que ocorreram picos de vazões na série observada fez-se a mesma análise com as séries de vazões simuladas, e por fim comparou-se os dias em que houveram aumento de vazão. A segunda análise foi feita em cima da intensidade da vazão de base. Para a estimativa de uma vazão de base variável no tempo, foi considerada a média móvel de duas vazões consecutivas identificadas como finais dos períodos de recessão (Subramanya, 2008).

Na Tabela 7.4 são mostrados os valores medianos dos percentuais de picos em fase com o observado, e os valores medianos do Coeficiente de Correlação de Pearson (R) para as intensidades das vazões de base e de pico. Embora o método de similaridade tenha alcançado, para a intensidade das vazões de pico, um valor mediano de R aproximadamente igual a 0,889, as simulações apresentaram baixa concordância com a vazão observada do ponto de vista da simultaneidade da ocorrência desses eventos.

Tabela 7.4: Valores medianos das estatísticas de erros calculadas para as vazões de pico e de base. Em negrito estão evidenciados os melhores resultados para cada índice dentre os métodos de regionalização.

Vazões de Pico e de Base					
Índices	Calibrado	Proximidade	Similaridade	<i>Random Forest I</i>	<i>Random Forest II</i>
Picos em fase (%)	35,900	30,770	29,085	34,550	<b>35,625</b>
R para a Vazão de Picos	0,855	0,870	<b>0,889</b>	0,856	0,859
R para a Vazão de Base	0,878	0,866	0,867	<b>0,887</b>	0,875

Continuando, fez-se a avaliação das simulações focando na capacidade de representar o período completo das recessões, ou seja, o período contínuo desde o pico até o fim da recessão. Na Tabela 7.5 estão indicados os valores medianos das estatísticas de erros para os períodos das recessões. Dentre os três coeficientes para determinar o desempenho das simulações o NSE, o qual tende a ser mais sensível às vazões altas, no geral apontou menor desempenho, para todos os métodos de regionalização e de calibração. Já o logNSE, o qual fornece mais ênfase às vazões baixas, alcançou valores superiores a 0,70 para o método de similaridade e de proximidade. Por sua vez, o sqrtNSE apresentou em geral um desempenho mediano para todos os métodos de regionalização e calibração.

Tabela 7.5: Valores medianos das estatísticas de erros calculadas para as vazões no período de recessão, o qual foi determinado pela série histórica observada. Em negrito estão evidenciados os melhores resultados para cada índice dentre os métodos de regionalização.

Períodos de Recessão					
	Calibrado	Proximidade	Similaridade	<i>Random Forest I</i>	<i>Random Forest II</i>
NSE	0,679	<b>0,684</b>	0,647	0,676	0,653
logNSE	0,783	0,730	<b>0,743</b>	0,717	0,687
sqrtNSE	0,778	0,731	0,713	<b>0,734</b>	0,713

### 7.3.5 Estimação da Vazão Q95

Com a finalidade de se comparar o desempenho da estimação da vazão de referência Q95 entre técnicas diretas (regressão) e indiretas (regionalização de parâmetros do modelo hidrológico). Aplicou-se o método de regressão de *Random Forest*, o qual foi nomeado como *Random Forest III*, para regionalizar diretamente a vazão Q95.

A construção do modelo de regressão de *Random Forest III* usou as informações conhecidas sobre as bacias do conjunto de treinamento, sendo elas: os descritores das bacias (variáveis independentes) e a Q95, em l/s/km<sup>2</sup>, estimada a partir das séries históricas observadas (variável dependente). As características mais relevantes apontadas pelo método de *Random Forest III* foram: dias de precipitação com acumulado mensal de 150 mm, longitude do centroide da bacia, declividade média da bacia, fração de pastagem e fração de floresta.

A estimativa da vazão Q95, tanto a partir da vazão observada quanto a partir da vazão simulada, considerou o mesmo período, sendo esse definido como a partir da data de início da calibração até a data final do período da validação. Dessa maneira, foram utilizados ao menos 15 anos de registros fluviométricos para se estimar a vazão de referência.

A Figura 7.17 mostra as correlações entre as vazões Q95 observadas e as previstas pela calibração e pelos métodos de regionalização. Observa-se que os métodos de regionalizações com as maiores performances foram o de similaridade fisiográfica-climática, que

alcançou uma correlação (R) igual a 0,973, e logo em segundo o *Random Forest III* com a correlação equivalente a 0,965.

Segundo Pugliese et al. (2014), como consequência do uso de estratégias de predição associadas à um único valor de referência, a interação contínua entre as condições climáticas e as características geomorfológicas de uma bacia hidrográfica, que fundamenta a curva de permanência, pode ser perdida, seja inteiramente ou parcialmente. Em virtude disso, durante o desenvolvimento desta pesquisa outros estudos de regionalização foram desenvolvidos usando o conjunto de dados hidrometeorológicos do Estado do Paraná, aqui desenvolvido. A pesquisa apresentada no XXIV Simpósio Brasileiro de Recursos Hídricos, Almeida et al. (2020), aplicou técnicas de geoestatística propostas por Skøien et al. (2006) e Pugliese et al. (2014), e alcançou resultados superiores aos aqui apresentados, uma vez que as curvas de permanência foram regionalizadas por completo. As curvas de permanência foram sintetizadas por meio do índice denominado como Desvio Negativo Total (DNT), o qual é obtido mediante a integração da área sobre a curva de permanência adimensional com escala log-normal. Em seguida, a regionalização da DNT é feita usando o método de krigagem-topológica, que é a interpolação e extrapolação de dados de vazão por meio da média ponderada, sendo que os pesos são distribuídos de acordo com a área das bacias e a topologia da rede hidrográfica.

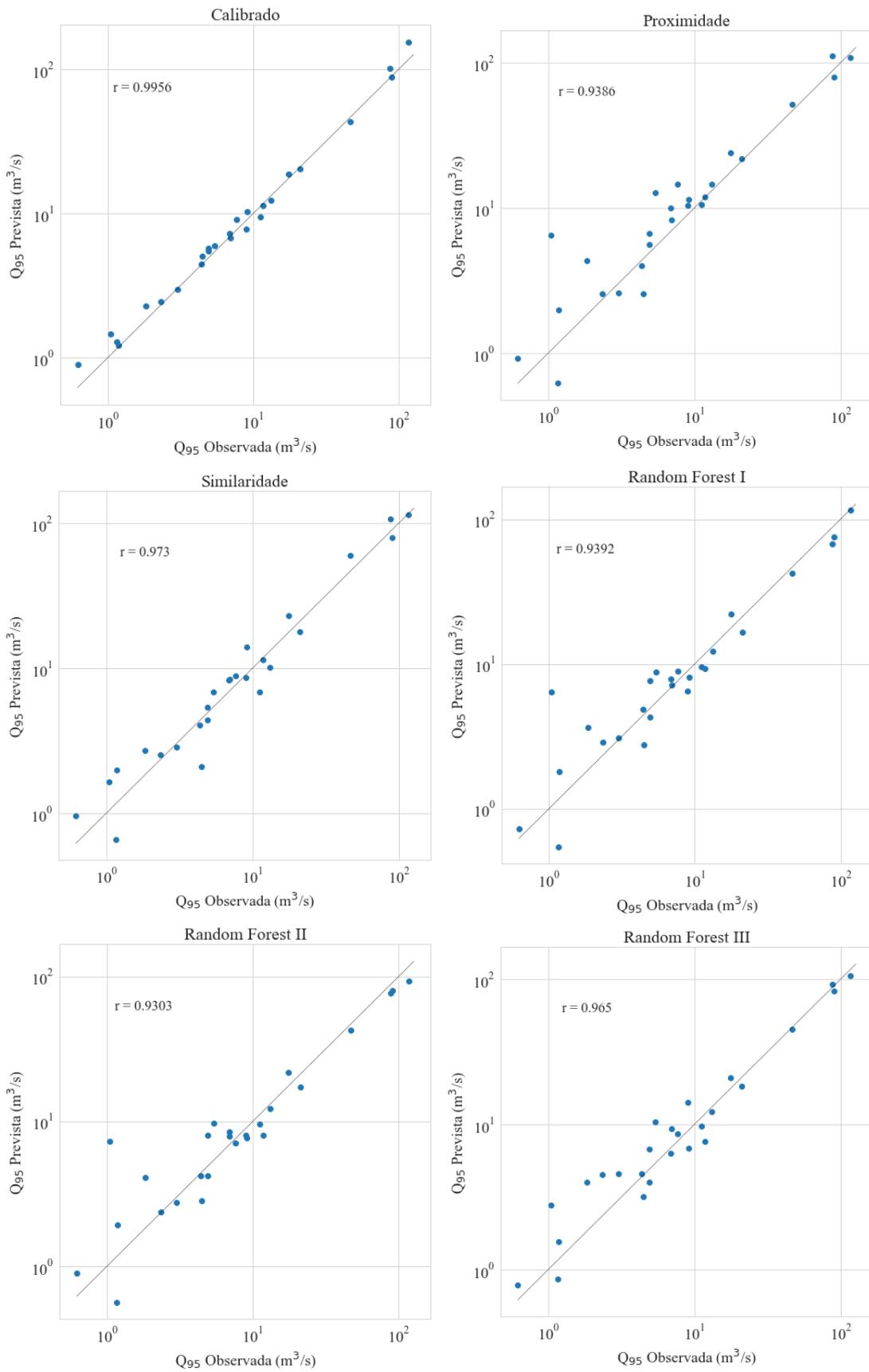


Figura 7.17: Correlações das estimativas da vazão de referência  $Q_{95}$ .

## 8 CONCLUSÃO

Neste estudo cinco métodos de regionalização foram empregados com o propósito de avaliar a predição de vazões diárias e a Q95 em bacias pouco ou não instrumentalizadas no Estado do Paraná. Dentre esses métodos quatro englobaram a estimação de parâmetros do modelo GR4J, em que foram avaliadas a capacidade de predizer vazões contínuas, de pico, de base e em períodos de recessão. O último método visou a estimação direta da vazão de referência Q95 para que fosse comparado com os métodos indiretos (regionalização de parâmetros do modelo GR4J).

Neste trabalho foram desenvolvidos e apresentados o conjunto de dados hidrometeorológico junto com os índices descritivos das bacias hidrográficas do Estado do Paraná. Os resultados das coletas desses dados, as quais foram realizadas de forma minuciosa, mostraram uma noção da quantidade de dados que estão disponíveis no Estado do Paraná e que apresenta ser maior do que é mostrado nos conjuntos de dados desenvolvidos a nível nacional, uma vez que foram obtidas uma densidade maior de estações fluviométricas dentro da mesma área de estudo. Além disso, o procedimento manual de hidrorreferenciamento utilizando a rede hidrográfica ottocodificada do IAT fornece maior confiabilidade sobre a localização das estações fluviométricas e disponibiliza informações a montante e a jusante das estações.

A simultaneidade de dados entre os registros fluviométricos e pluviométricos das estações é fundamental para esse tipo de estudo e, portanto, a parcela de bacias hidrográficas selecionada foi um pouco menor que 50% (126 de 284) do tamanho original do conjunto de bacias hidrográficas do Estado do Paraná. Além da simultaneidade de informações, atentou-se que essas bacias tivessem ao menos 15 anos de registros fluviométricos, sendo que cada ano contabilizado pudesse ter no máximo 10% de falhas, preferiu-se também que as séries históricas de vazões tivessem dados recentes, as quais se estendem além do ano de 2010, e limitou-se as séries de vazões em que o teste de Pettitt apontasse como homogênea.

O modelo chuva-vazão, GR4J, foi empregado nas 126 bacias hidrográficas e alcançou desempenhos otimistas no período de verificação ( $\log\text{NSE} \geq 0,70$ ) para aproximadamente 65% do conjunto de bacias selecionado. Excluindo a região litorânea o modelo GR4J apresentou bons resultados nas demais regiões do Estado. Com base nesses desempenhos positivos o GR4J tornou-se elegível para a aplicação de técnicas de regionalização de parâmetros entre bacias paranaenses. Em estudos de regionalização a definição do modelo hidrológico é um passo importante, uma vez que cada modelo requer uma quantidade considerável de dados e pode apresentar um certo grau de complexidade e de incertezas (Arsenault et al., 2019).

Adicionalmente, tanto o método de calibração determinística quanto a escolha da função objetivo podem também contribuir para as incertezas dos métodos de regionalização, em virtude da possibilidade de existirem múltiplos conjuntos considerados como ótimos e a alta interação entre os parâmetros (Kuczera e Mroczkowski, 1998; Fowler et al., 2016). Posta essas questões, a área de pesquisa de otimização de modelos hidrológicos é fundamental para a realização dos ajustes e, assim, para a adequação do modelo para situações específicas, sendo uma área que pode ser estudada de forma mais aprofundada futuramente para o desenvolvimento de projetos relacionados a este.

Ambos os métodos de regionalização baseados em distâncias, tanto a técnica baseada na similaridade fisiográfica-climática quanto a proximidade espacial, transferiram por completo os conjuntos de parâmetros das bacias doadoras. A regionalização por proximidade espacial considerou as regiões vizinhas às bacias como homogêneas, e elegeu as possíveis bacias doadoras de parâmetros como as mais próximas da bacia de interesse. Na análise do número ideal de bacias doadoras observou-se, para o método de proximidade, melhores resultados quando se utilizam cerca de 7 bacias. Todavia, a necessidade de se precisar um número tão alto de bacias torna o procedimento inviável. Além disso, a Tabela 7.3 sugere que mesmo a bacia doadora mais próxima não aparenta ter altas correlações com todos os parâmetros das bacias alvos.

Para o desenvolvimento da regionalização por similaridade fisiográfica-climática, primeiro foi realizado um estudo de classificação/agrupamento das bacias hidrográficas, o que permitiu avaliar se de alguma maneira as bacias dentro do mesmo grupo possuem comportamentos de vazão semelhante, e se haviam diferenças nesse comportamento entre os grupos formados. Embora as bacias consideradas como as mais similares também não apresentassem altas correlações com todos os parâmetros das bacias alvos, atingiu-se valores medianos de  $\log\text{NSE}$  e  $\sqrt{\text{NSE}}$  superiores aos demais métodos de regionalização, como foi apresentado na Tabela 7.2.

O método de regressão de *Random Forest II*, o qual esperava-se que a relação entre os descritores promovesse aperfeiçoamento da técnica de regressão para a estimação dos parâmetros, apresentou, em geral, resultados inferiores ao método do *Random Forest I* que considerou, somente, os descritores das bacias como variáveis independentes.

A vazão de referência Q95 é de grande importância no Estado do Paraná para se estimar a vazão máxima que pode ser outorgada dos corpos hídricos. Embora o método de regionalização direta tenha alcançado a segunda colocação, é propício que o método indireto de regionalização por similaridade fisiográfica-climática tenha atingido uma correlação (R) superior correspondente a 0,973.

Em estudos de regionalização é compreensível que a área de estudo, o tamanho da amostra e as bacias que compõem a amostra sejam fatores comprometedores para a avaliação do desempenho de cada método. Entretanto, mesmo com as limitações e a robustez das técnicas de regionalização empregadas, os métodos de regionalização apresentaram resultados otimistas para a predição de vazões mínimas atingindo valores medianos de  $\log\text{NSE}$ , para o período de validação, equivalentes a 0,702, 0,736, 0,679 e 0,652 para o método de proximidade, similaridade, *Random Forest I* e *Random Forest II*, respectivamente. Entretanto, de acordo com a avaliação feita para a amostra de 26 bacias, a regionalização baseada em similaridade fisiográfica-climática apontou ser o método mais robusto para a predição de vazão diária e da vazão de referência Q95 em bacias pouco ou não instrumentalizadas no Estado do Paraná.

Apesar dos baixos volumes de chuvas nos anos de 2020 e 2021 terem diminuído a disponibilidade de água no Estado do Paraná, foi o principal incentivo e encorajamento para o desenvolvimento deste trabalho. Devido à importância dos recursos hídricos para o abastecimento de água à população, a irrigação de lavouras e a geração de energia, espera-se que futuramente novas pesquisas possam ser desenvolvidas a partir do conjunto de dados e informações que foram coletadas neste trabalho.

Como recomendações de estudos futuros, considerar: o emprego de outros modelos hidrológicos, a utilização de técnicas de otimização estocástica para identificação dos conjuntos de parâmetros do modelo hidrológico e o uso de outras funções objetivos. Além disso, levar em consideração o emprego dos métodos aqui apresentados para a

regionalização de vazões médias e máximas, a estimação de intervalos de confiança para os parâmetros regionalizados e o uso de métodos de regionalização baseados nas técnicas de geoestatística, tendo em vista não só a variabilidade espacial como também a topologia da rede hidrográfica.

*O presente trabalho foi realizado com apoio do Sistema de Tecnologia e Monitoramento Ambiental do Paraná (Simepar) e Companhia de Saneamento do Paraná (Sanepar).*

## REFERÊNCIAS

- AGUASPARANÁ (2010). *Manual técnico de outorgas*. Estado do Paraná, i edition.
- Allen, R. G., Pereira, L. S., Raes, D. e Smith, M. (1998). *Crop evapotranspiration : guidelines for computing crop water requirements*. Food and Agriculture Organization of the United Nations, Rome.
- Almagro, A., Oliveira, P. T. S., Neto, A. A. M., Roy, T. e Troch, P. (2021). CABra: a novel large-sample dataset for Brazilian catchments. *Hydrology and Earth System Sciences*, 25(6):3105–3135.
- Almeida, A. S., Leite, E. A., Kuana, L. A., Mercuri, E. G. F., Gonçalves, J. E., Gobbi, E. F., Beneti, C. A. A., Marcon, R. A., Lobo, A. V. L. e Mendes, E. A. A. (2020). Regionalização de Curvas de Permanência no Estado do Paraná. Em *Anais do XXIV Simpósio Brasileiro de Recursos Hídricos*. Associação Brasileira de Recursos Hídricos, Belo Horizonte.
- ANA (Brasília, 2019.). *Plano Nacional de Segurança Hídrica*. Agência Nacional de Águas, Brasília -.
- Arsenault, R., Breton-Dufour, M., Poulin, A., Dallaire, G. e Romero-Lopez, R. (2019). Streamflow prediction in ungauged basins: analysis of regionalization methods in a hydrologically heterogeneous region of Mexico. *Hydrological Sciences Journal*, 64(11):1297–1311.
- Ayzel, G., Varentsova, N., Erina, O., Sokolov, D., Kurochkina, L. e Moreydo, V. (2019). OpenForecast: The First Open-Source Operational Runoff Forecasting System in Russia. *Water*, 11(8):1546.
- Barbieri, G. M. L., Costa, A. B. F., Olivieira, C., Jusevicius, M. e D'Ávila, V. C. (2017). ATLAS SOLARIMÉTRICO DO ESTADO DO PARANÁ. *Manuscrito não publicado*.
- Bazzo, J. P. V. e Almeida, R. C. d. (2016). Regionalização de Vazões com o Emprego de Redes Neurais Artificiais RBF. *I Simpósio de Métodos Numéricos em Engenharia*.
- Besaw, L. E., Rizzo, D. M., Bierman, P. R. e Hackett, W. R. (2010). Advances in ungauged streamflow prediction using artificial neural networks. *Journal of Hydrology*, 386(1-4):27–37.
- Bigas, H. (2013). *Water security and the global water agenda : a UN-water analytical brief*. United Nations University - Institute for Water, Environment and Health, Hamilton, Ont.
- Blöschl, G., Sivapalan, M., Wagener, T., Viglione, A. e Savenije, H. (2013). *Runoff Prediction in Ungauged Basins*. Cambridge University Press.
- Boutsidis, C., Zouzas, A., Mahoney, M. W. e Drineas, P. (2014). Randomized Dimensionality Reduction for k-Means Clustering. *IEEE Transactions on Information Theory*, 61(2):1045–1062.

- Brasil (Brasília, 9 jan. 1997.). *Lei nº 9433, 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 da Constituição Federal, e altera o art. I da Lei no 8001, de 13 de março de 1990, que modificou a Lei no 7990, de 28 de dezembro de 1989.* Diário Oficial [da]República Federativa do Brasil.
- Breiman, L. (2001). Random Forests. *Machine Learning*, 45:5–32.
- Calvetti, L., Beneti, C., Neundorf, R. L. A., Inouye, R. T., dos Santos, T. N., Gomes, A. M., Herdies, D. L. e de Gonçalves, L. G. G. (2017). Quantitative Precipitation Estimation Integrated by Poisson’s Equation Using Radar Mosaic, Satellite, and Rain Gauge Network. *Journal of Hydrologic Engineering*, 22(5).
- Chagas, V. B. P., Chaffe, P. L. B., Addor, N., Fan, F. M., Fleischmann, A. S., Paiva, R. C. D. e Siqueira, V. A. (2020). CAMELS-BR: hydrometeorological time series and landscape attributes for 897 catchments in Brazil. *Earth System Science Data*, 12(3):2075–2096.
- Daggupati, P., Pai, N., Ale, S., Douglas-Mankin, K. R., andJ. Jeong, R. W. Z., Parajuli, P. B., Saraswat, D. e Youssef, M. A. (2015). A Recommended Calibration and Validation Strategy for Hydrologic and Water Quality Models. *American Society of Agricultural and Biological Engineers*, 58(6):1705–1719.
- Devia, G. K., Ganasri, B. P. e Dwarakish, G. S. (2015). A Review on Hydrological Models. *Aquatic Procedia*, 4:1001–1007.
- Eduardo, E. N., de Mello, C. R., Viola, M. R., Owens, P. R. e Curi, N. (2016). Hydrological simulation as subside for management ofsurface water resources at the Mortes River Basin. *Ciência e Agrotecnologia*, 40(4):390–404.
- Embrapa (2020). Mapa de solos do estado do Paraná. Disponível em: [http://geoinfo.cnps.embrapa.br/layers/geonode:parana\\_solos\\_20201105](http://geoinfo.cnps.embrapa.br/layers/geonode:parana_solos_20201105). Acessado em 05 de julho de 2021.
- FAO (2021). *WASAG Strategy for 2021 to 2024 - Turning water scarcity into opportunities for sustainable agriculture, food security and nutrition.* Food and Agriculture Organization of th United Nations, Roma.
- Fowler, K. J. A., Peel, M. C., Western, A. W., Zhang, L. e Peterson, T. J. (2016). Simulating runoff under changing climatic conditions: Revisiting an apparent deficiency of conceptual rainfall-runoff models. *Water Resources Research*, 52(3):1820–1846.
- Global Footprint Network (2021). About Earth Overshoot Day. Disponível em: <https://www.overshootday.org/about-earth-overshoot-day/>. Acessado em 05 de junho de 2021.
- Grego, C. R., de Oliveira, R. P. e Vieira, S. R. (2014). *Agricultura De Precisão: Resultados De Um Novo Olhar*, capítulo: Geoestatística aplicada à Agricultura de Precisão, páginas 74–83. Embrapa.
- Guo, Y., Zhang, Y., Zhang, L. e Wang, Z. (2020). Regionalization of hydrological modeling for predicting streamflow in ungauged catchments: A comprehensive review. *Wires Water*, e1487.

- He, Y., Bardossy, A. e Zehe, E. (2011). A review of regionalisation for continuous streamflow simulation. *Hydrol. Earth Syst. Sci.*, 15:3539–3553.
- Hengl, T., Heuvelink, G. B. M. e G. Rossiter, D. (2007). About regression-kriging: From equations to case studies. *Computers & Geosciences*, 33(10):1301–1315.
- Hoeltgebaum, L. E. B. (2021). *Análise do Balanço Hídrico para Estimativas do Armazenamento de Água e Evapotranspiração na Bacia do Rio Wahoo*. Tese de doutorado, Programa de Pós-Graduação em Engenharia Ambiental, Universidade Federal do Paraná.
- Hrachowitz, M., Savenije, H. H. G., Blöschl, G., McDonnell, J. J., Sivapalan, M., Pomeroy, J. W., Arheimer, B., Blume, T., Clark, M. P., Ehret, U., Fenicia, F., Freer, J. E., Gelfan, A., Gupta, H. V., Hughes, D. A., Hut, R. W., Montanari, A., Pande, S., Tetzlaff, D., Troch, P. A., Uhlenbrook, S., Wagener, T., Winsemius, H. C., Woods, R. A., Zehe, E. e Cudennec, C. (2013). A decade of Predictions in Ungauged Basins (PUB)—a review. *Hydrological Sciences Journal*, 58(6):1–58.
- IAT (2020). Mapas e Dados Espaciais. Disponível em: <http://www.iat.pr.gov.br/Pagina/Mapas-e-Dados-Espaciais>. Acessado em 05 de julho de 2021.
- Jaiswal, R. K., Ali, S. e Bharti, B. (2020). Comparative evaluation of conceptual and physical rainfall–runoff models. *Applied Water Science*, 10(1).
- Juliani, B. H. T., de Campos, A. L., Almeida, A. S. e Leite, E. A. (2020). Estatísticas meteorológicas da seca de 2020 no estado do Paraná. Em *Anais do II END - Encontro Nacional de Desastres da ABRHidro*. ABRHidro.
- Kaviski, E., Rohn, M. d. C. e Mazer, W. (2002). Projeto HG-171: Consistência e regionalização de dados hidrológicos. *Centro de Hidráulica e Hidrologia Prof. Parigot de Souza*.
- Ketchen Junior, D. J. e Shook, C. L. (1996). The application of cluster analysis in strategic management research: an analysis and critique. *Strategic Management Journal*, 17:441–458.
- Kim, D., Jung, I. W. e Chun, J. A. (2017). A comparative assessment of rainfall–runoff modelling against regional flow duration curves for ungauged catchments. *Hydrology and Earth System Sciences*, 21(11):5647–5661.
- Kratzert, F., Čertík, O., Klotz, D., MacDonald, A., Visser, M. e Gauch, M. (2017). Rainfall-Runoff Modelling Playground (RRMPG). Disponível em: <https://github.com/kratzert/RRMPG>. Acessado em 05 de agosto de 2020.
- Krause, P., Boyle, D. P. e Bäse, F. (2005). Comparison of different efficiency criteria for hydrological model assessment. *Advances in Geosciences*, 5:89–97.
- Kuana, L. A., Almeida, A. S., Mercuri, E. G. F., Leite, E. A., de Lima Lobo, A. V., Mendes, E. A. A., Marcon, R. A., Gonçalves, J. E. e Beneti, C. A. A. (2021a). Aplicação dos Modelos GR4J e GR6J Utilizando o Conjunto de Dados Desenvolvido Para o Estado do Paraná. Em *Anais do XXIV Simpósio Brasileiro de Recursos Hídricos*. Associação Brasileira de Recursos Hídricos.

- Kuana, L. A., Almeida, A. S., Mercuri, E. G. F., Leite, E. A., de Lima Lobo, A. V., Mendes, E. A. A., Marcon, R. A., Gonçalves, J. E. e Beneti, C. A. A. (2021b). Estudo de Caso: Regionalização de Vazão Utilizando a Base de Dados Camels. Em *Anais do XXIV Simpósio Brasileiro de Recursos Hídricos*. Associação Brasileira de Recursos Hídricos.
- Kuczera, G. e Mroczkowski, M. (1998). Assessment of hydrologic parameter uncertainty and the worth of multiresponse data. *Water Resour. Res.*, 34:1481–1490.
- Kuentz, A., Arheimer, B., Hundecha, Y. e Wagener, T. (2017). Understanding hydrologic variability across Europe through catchment classification. *Hydrology and Earth System Sciences*, 21(6):2863–2879.
- Llabrés-Brustenga, A., Rius, A., Rodríguez-Sol, R., Casas-Castillo, M. C. e Redaño, A. (2019). Quality control process of the daily rainfall series available in Catalonia from 1855 to the present. *Theoretical and Applied Climatology*, 137:2715–2729.
- Machado, A. R., Junior, A. V. M. e Wendland, E. C. (2017). Avaliação do modelo J2000/JAMS para modelagem hidrológica em bacias hidrográficas brasileiras. Relatório técnico, Associação Brasileira de Engenharia Sanitária e Ambiental - ABES.
- Matallo Junior, H. (2001). *Indicadores de desertificação: histórico e perspectivas*. Edições UNESCO Brasil, Brasília, DF, Brasil.
- Mohamed, S., Ludovic, O. e Ribstein, P. (2019). Random Forest Ability in Regionalizing HourlyHydrological Model Parameters. *Water*, 11.
- Muleta, M. K. (2012). Model Performance Sensitivity to Objective Function during Automated Calibrations. *Journal of Hydrologic Engineering*, 17(6):756–767.
- NASA (2021). Measuring Earth’s Surface Mass and Water Changes. Disponível em: <https://grace.jpl.nasa.gov/>. Acessado em 05 de julho de 2021.
- Neto, W. M. P., Vieira, F. R. e Matosinhos, C. C. (2021). Avaliação da performance dos modelos GR4J, GR5J e GR6J na bacia hidrográfica do ribeirão São João, Minas Gerais. Em *Base de Conhecimentos Gerados na Engenharia Ambiental e Sanitária 3*. Atena.
- Noumon, C. J., Kodja, D. J., Kouhoundji, N., Houessou, S., Amoussou, E., Sintondji, L. O., Vissin, E. e Agbossou, E. K. (2020). Utilisation du Modèle GR4J pour la Simulation des Débitsd’Etiage du Fleuve Mono (Bénin, Togo) à l’Exutoire d’Athiémé. *International Journals of Sciences and High Technologies*, 23(1):139–144.
- Oudin, L., Andréassian, V., Perrin, C., Michel, C. e Moine, N. L. (2008). Spatial proximity, physical similarity, regression and ungaged catchments: A comparison of regionalization approaches based on 913 French catchments. *Water Resources Research*, 44(3).
- Oudin, L., Kay, A., Andréassian, V. e Perrin, C. (2010). Are seemingly physically similar catchments truly hydrologically similar? *Water Resources Research*, 46(11).
- Pagano, T., Hapuarachchi, P. e Wang, Q. J. (2010). Continuous rainfall-runoff model comparison and short-term daily streamflow forecast skill evaluation. *CSIRO Tech. Rep.*, EP103545.

- Parajka, J., Merz, R. e Blöschl, G. (2005). A comparison of regionalisation methods for catchment model parameters. *Hydrology and Earth System Sciences*, 9:157–171.
- Perrin, C., Michel, C. e Andréassian, V. (2003). Improvement of a parsimonious model for streamflow simulation. *Journal of Hydrology*, 279:275–289.
- Pettitt, A. N. (1979). A Non-Parametric Approach to the Change-Point Problem. *Applied Statistics*, 28(2):126.
- Pfaffstetter, O. (1989). Classificação de bacias hidrográficas. Manuscrito não publicado.
- Pinheiro, V. B. e Naghettini, M. (2013). Calibration of the Parameters of a Rainfall-Runoff Model in Ungauged Basins Using Synthetic Flow Duration Curves as Estimated by Regional Analysis. *Journal of Hydrologic Engineering*, 18(12):1617–1626.
- Prieto, C., Vine, N. L., Kavetski, D., García, E. e Medina, R. (2019). Flow Prediction in Ungauged Catchments Using Probabilistic Random Forests Regionalization and New Statistical Adequacy Tests. *Water Resources Research*, 55(5):4364–4392.
- Pugliese, A., Castellarin, A. e Brath, A. (2014). Geostatistical prediction of flow–duration curves in an index-flow framework. *Hydrol. Earth Syst. Sci.*, 18:3801–3816.
- Razavi, T. e Coulibaly, P. (2013). Streamflow Prediction in Ungauged Basins: Review of Regionalization Methods. *Journal of Hydrologic Engineering*, 18(8):958–975.
- Razavi, T. e Coulibaly, P. (2016). Improving streamflow estimation in ungauged basins using a multi-modelling approach. *Hydrological Sciences Journal*, 61(15):2668–2679.
- Rousseeuw, P. J. (1987). Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, 20:53–65.
- Saha, P. P. e Zeleke, K. (2015). Rainfall-Runoff Modelling for Sustainable Water Resources Management: SWAT Model Review in Australia. Em *Sustainability of Integrated Water Resources Management*, páginas 563–578. Springer International Publishing.
- Shin, M.-J. e Kim, C.-S. (2016). Assessment of the suitability of rainfall–runoff models by coupling performance statistics and sensitivity analysis. *Hydrology Research*, 48(5):1192–1213.
- Sivapalan, M. (2003). Prediction in ungauged basins: a grand challenge for theoretical hydrology. *Hydrological Processes*, 17(15):3163–3170.
- Skøien, J. O., Merz, R. e Blöschl, G. (2006). Top-kriging – geostatistics on stream networks. *Hydrol. Earth Syst. Sci.*, 10:277–287.
- Soil Conservation Service (1972). *National engineering handbook*, capítulo: Seção 4, Hydrology, página 762. Department of Agriculture, Washington.
- Song, J.-H., Her, Y., Suh, K., Kang, M.-S. e Kim, H. (2019). Regionalization of a Rainfall-Runoff Model: Limitations and Potentials. *Water*, 11(11):2257.
- Sousa, F. M. L., Neto, V. S. C., Pacheco, W. E. e Barbosa, S. A. (2009). Sistema Nacional De Informações Sobre Recursos Hídricos: Sistematização Conceitual E Modelagem Funcional. Em *Anais do XVIII Simpósio Brasileiro de Recursos Hídricos*. Associação Brasileira de Recursos Hídricos, Campo Grande.

- Souza, C. M., Shimbo, J. Z., Rosa, M. R., Parente, L. L., Alencar, A. A., Rudorff, B. F. T., Hasenack, H., Matsumoto, M., Ferreira, L. G., Souza-Filho, P. W. M., de Oliveira, S. W., Rocha, W. F., Fonseca, A. V., Marques, C. B., Diniz, C. G., Costa, D., Monteiro, D., Rosa, E. R., Vélez-Martin, E., Weber, E. J., Lenti, F. E. B., Paternost, F. F., Pareyn, F. G. C., Siqueira, J. V., Viera, J. L., Neto, L. C. F., Saraiva, M. M., Sales, M. H., Salgado, M. P. G., Vasconcelos, R., Galano, S., Mesquita, V. V. e Azevedo, T. (2020). Reconstructing Three Decades of Land Use and Land Cover Changes in Brazilian Biomes with Landsat Archive and Earth Engine. *Remote Sensing*, 12(17):2735.
- Storn, R. e Price, K. (1997a). Differential Evolution – A Simple and Efficient Heuristic for Global Optimization over Continuous Spaces. *Journal of Global Optimization*, 11(4):341–359.
- Storn, R. e Price, K. (1997b). Differential Evolution – A Simple and Efficient Heuristic for Global Optimization over Continuous Spaces. *Journal of Global Optimization*, 11:341–359.
- Subramanya, K. (2008). *Engineering hydrology*. Tata McGraw-Hill, New Delhi.
- Tucci, C. E. M. (Porto Alegre, 1998). *Modelos hidrológicos*. Universidade/UFRGS/Associação Brasileira de Recursos Hídricos.
- Varouchakis, E. A. (2019). *Spatiotemporal Analysis of Extreme Hydrological Events*. Elsevier.
- Viviroli, D., Mittelbach, H., Gurtz, J. e Weingartner, R. (2009). Continuous simulation for flood estimation in ungauged mesoscale catchments of Switzerland – Part II: Parameter regionalisation and flood estimation results. *Journal of Hydrology*, 377(1-2):208–225.
- Waseem, M., Ajmal, M. e Kim, T.-W. (2015). Ensemble hydrological prediction of streamflow percentile at ungauged basins in Pakistan. *Journal of Hydrology*, 525:130–137.
- Wilks, D. S. (2011). *Statistical Methods in the Atmospheric Sciences*. Academic Press.
- Yadav, M., Wagener, T. e Gupta, H. (2007). Regionalization of constraints on expected watershed response behavior for improved predictions in ungauged basins. *Advances in Water Resources*, 30(8):1756–1774.
- Yatagai, A., Maeda, M., Khadgarai, S., Masuda, M. e Xie, P. (2020). End of the Day (EOD) Judgment for Daily Rain-Gauge Data. *Atmosphere*, 11:15.
- Zhang, Y., Vaze, J., Chiew, F. H. S. e Li, M. (2015). Comparing flow duration curve and rainfall–runoff modelling for predicting daily runoff in ungauged catchments. *Journal of Hydrology*, 525:72–86.

## APÊNDICE A – CONJUNTO DE DADOS DO ESTADO DO PARANÁ

Nesta seção o conjunto de dados coletado para o estado do Paraná é apresentado com mais detalhes.

### A.1 DADOS FLUVIOMÉTRICOS

O conjunto de dados desenvolvido para o estado do Paraná é composto por 284 bacias hidrográficas com medições de vazões variando de 1926 a 2020. A disponibilidade de registros de vazão em anos pode variar de 5 anos e chegar até 90 anos, como é mostrado na Figura A.1. A frequência acumulada indica que mais da metade das estações coletadas possuem ao menos 15 anos de medições.

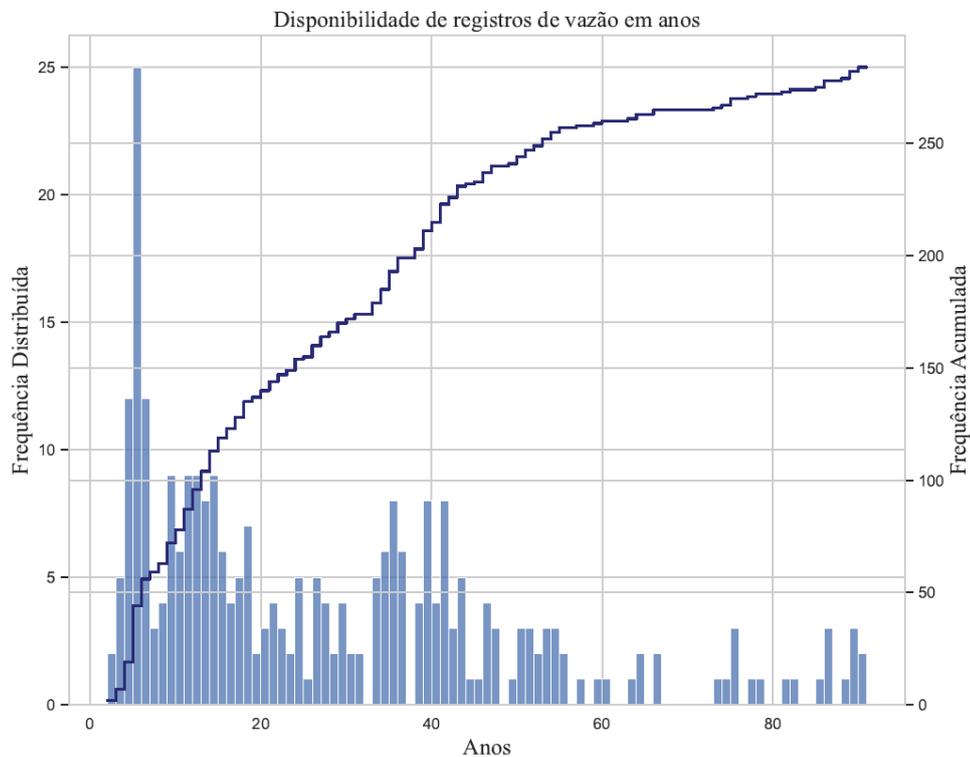


Figura A.1: Disponibilidade de registros de vazão em anos e a frequência com que ocorrem dentro do conjunto de estações fluviométricas coletadas.

A quantidade de registros em cada estação não depende da região do estado ou da área de drenagem da bacia, como é observado na Figura A.2.

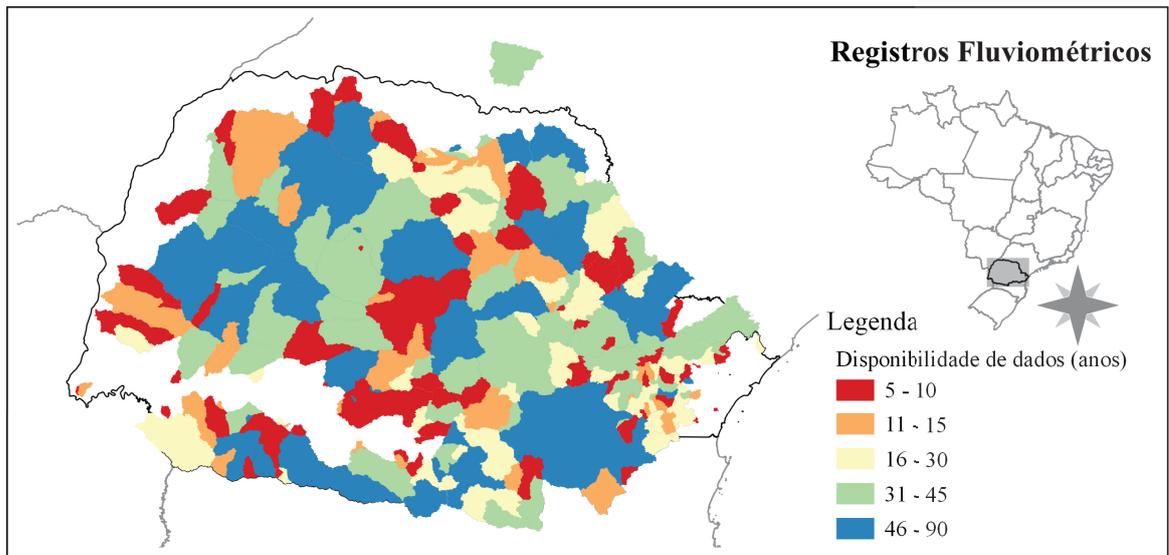


Figura A.2: Disponibilidade de registros de vazão em anos.

A Figura A.3 apresenta a disponibilidade de estações que possuem no máximo 10% de falhas em cada ano. Embora a Figura A.3 mostra uma tendência de diminuição de registros nos períodos mais recentes, é necessário ressaltar que a partir do início do século XXI o monitoramento se tornou cada vez mais privado, e além disso a Agência Nacional de Águas tem investido cada vez mais no monitoramento telemétrico, por serem estações recentes essas não foram incluídas no conjunto de dados.

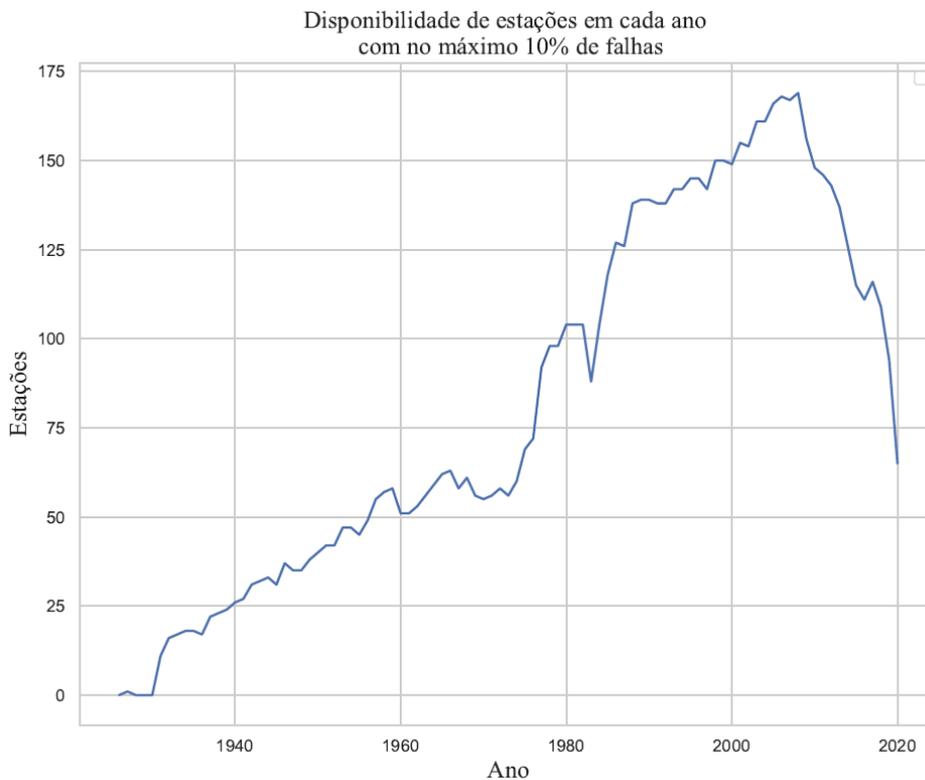


Figura A.3: Disponibilidade de estações que possuem no máximo 10% de dados em cada ano.

Os gráficos a seguir mostram a disponibilidade de dados de 1926 a 2020 em cada estação, sendo que quanto mais escura a cor verde mais dados disponíveis há naquele ano. Os códigos utilizados para referenciar as estações são majoritariamente os mesmos registrados na ANA.

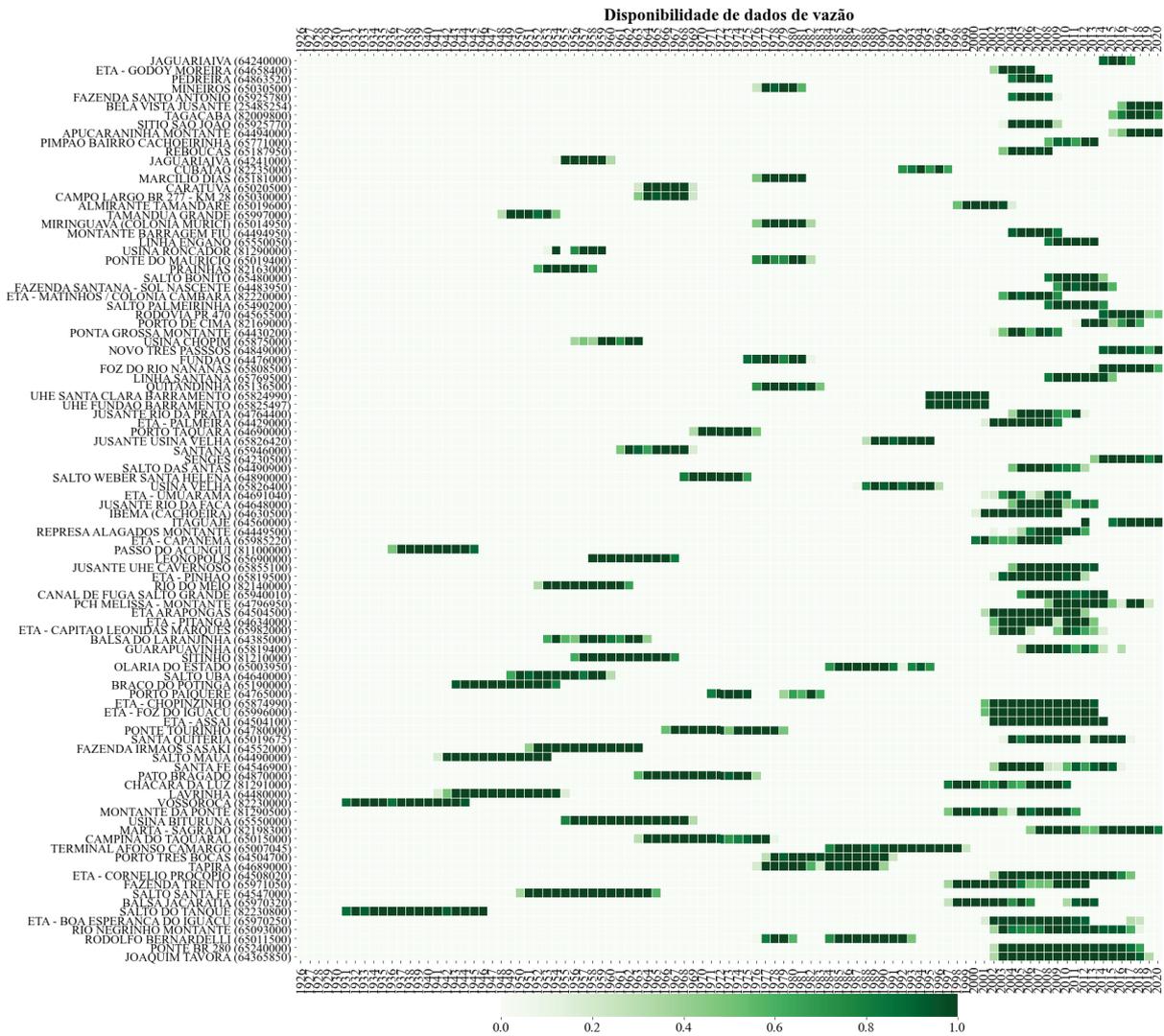


Figura A.4: Disponibilidade de dados de vazão por estação. Quanto mais escura a cor verde mais dados disponíveis há naquele ano.

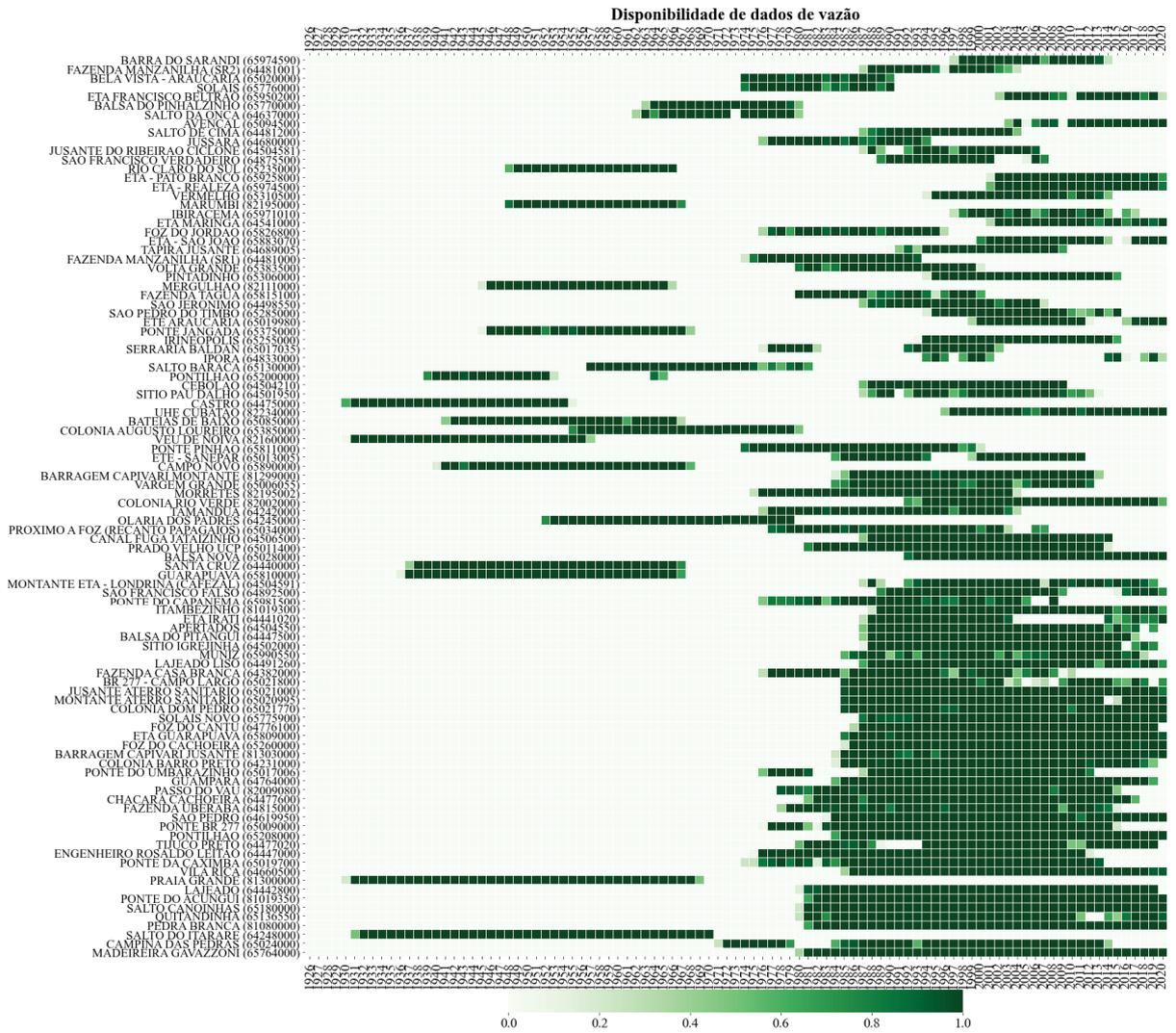


Figura A.5: Disponibilidade de dados de vazão por estação. Quanto mais escura a cor verde mais dados disponíveis há naquele ano.

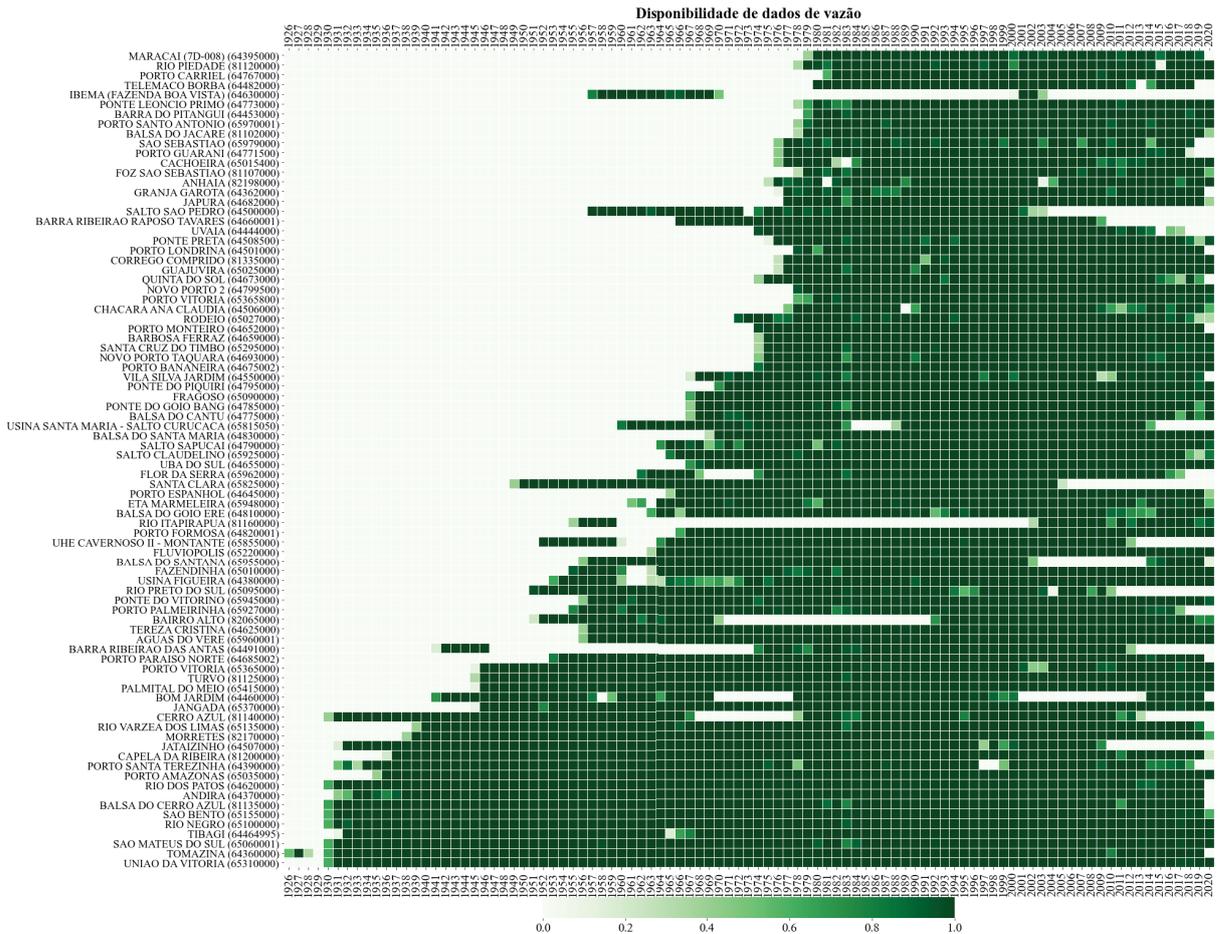


Figura A.6: Disponibilidade de dados de vazão por estação. Quanto mais escura a cor verde mais dados disponíveis há naquele ano.

## A.2 DADOS PLUVIOMÉTRICOS

A recuperação do histórico de precipitação contou com os registros de 1389 estações pluviométricas, as quais são mostradas no mapa da Figura A.7. Observa-se que ao leste do Paraná foram encontrados uma concentração maior de postos, contudo o comprimento das séries tende a ser menor. Em geral, é possível obter postos pluviométricos com registros de até 50 anos em toda a extensão do estado.

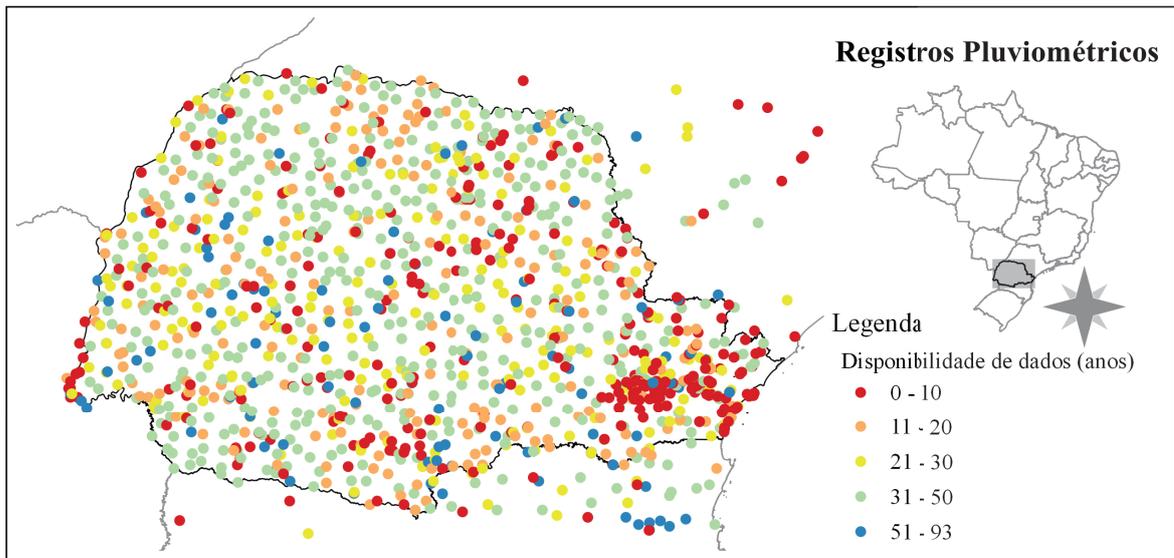


Figura A.7: Disponibilidade de registros de precipitação em anos.

A disponibilidade de registros pluviométricos nas estações coletadas pode variar a menos de 1 ano a 93 anos. A frequência acumulada, como é mostrado na Figura A.8, aponta que a maioria das estações possuem séries com comprimento de ao menos 15 anos, sendo que as estações com 43 e 44 anos são a maioria.

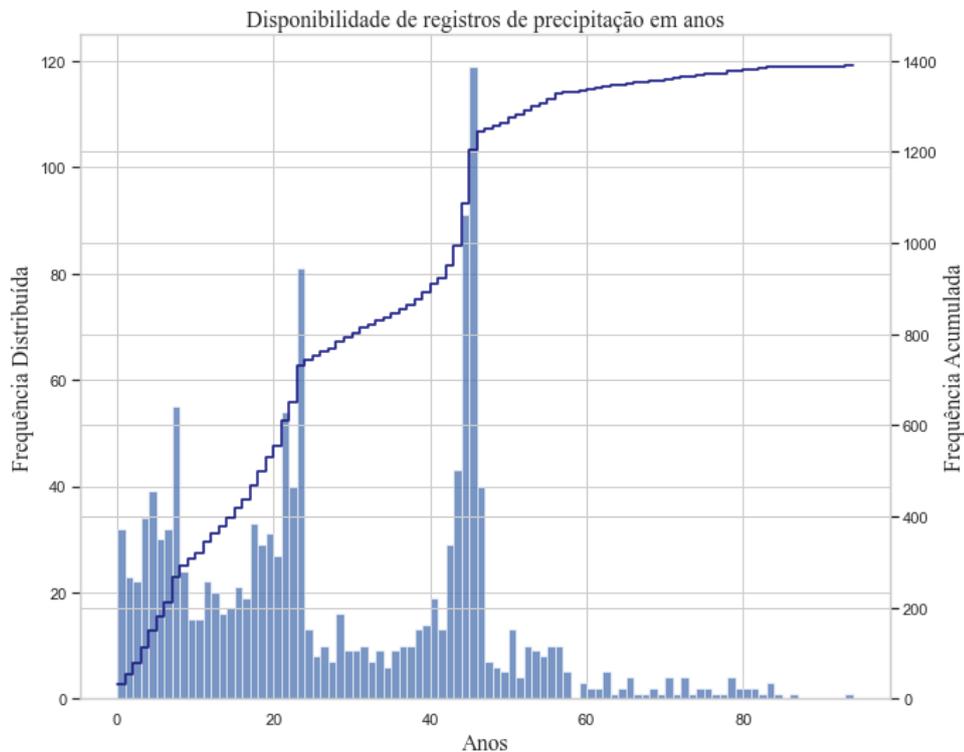


Figura A.8: Disponibilidade de registros de precipitação em anos e a frequência com que ocorrem dentro do conjunto de estações fluviométricas coletadas.

Embora a tendência de diminuição de registros seja também observada na Figura A.9, somando-se as mesmas circunstâncias pressupostas nas séries de vazões salienta-se

que não foram utilizados dados de estações telemétricas proveniente de outras operadoras além dos que são oriundos do Simepar.

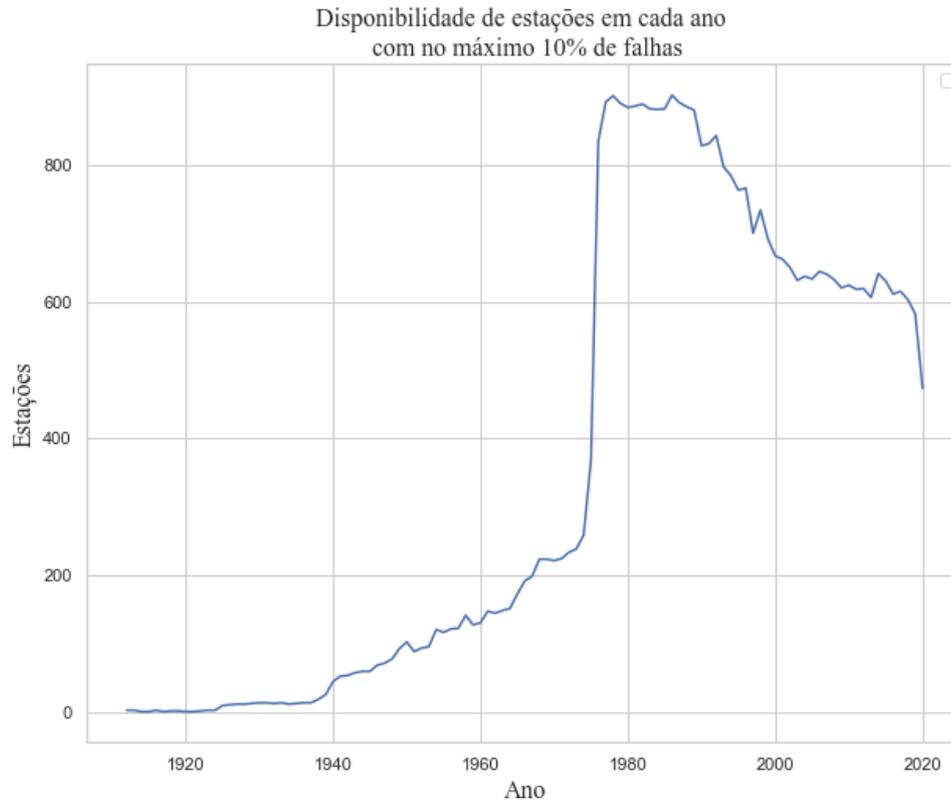


Figura A.9: Disponibilidade de estações que possuem no máximo 10% de dados em cada ano.

### A.3 DESCRITORES

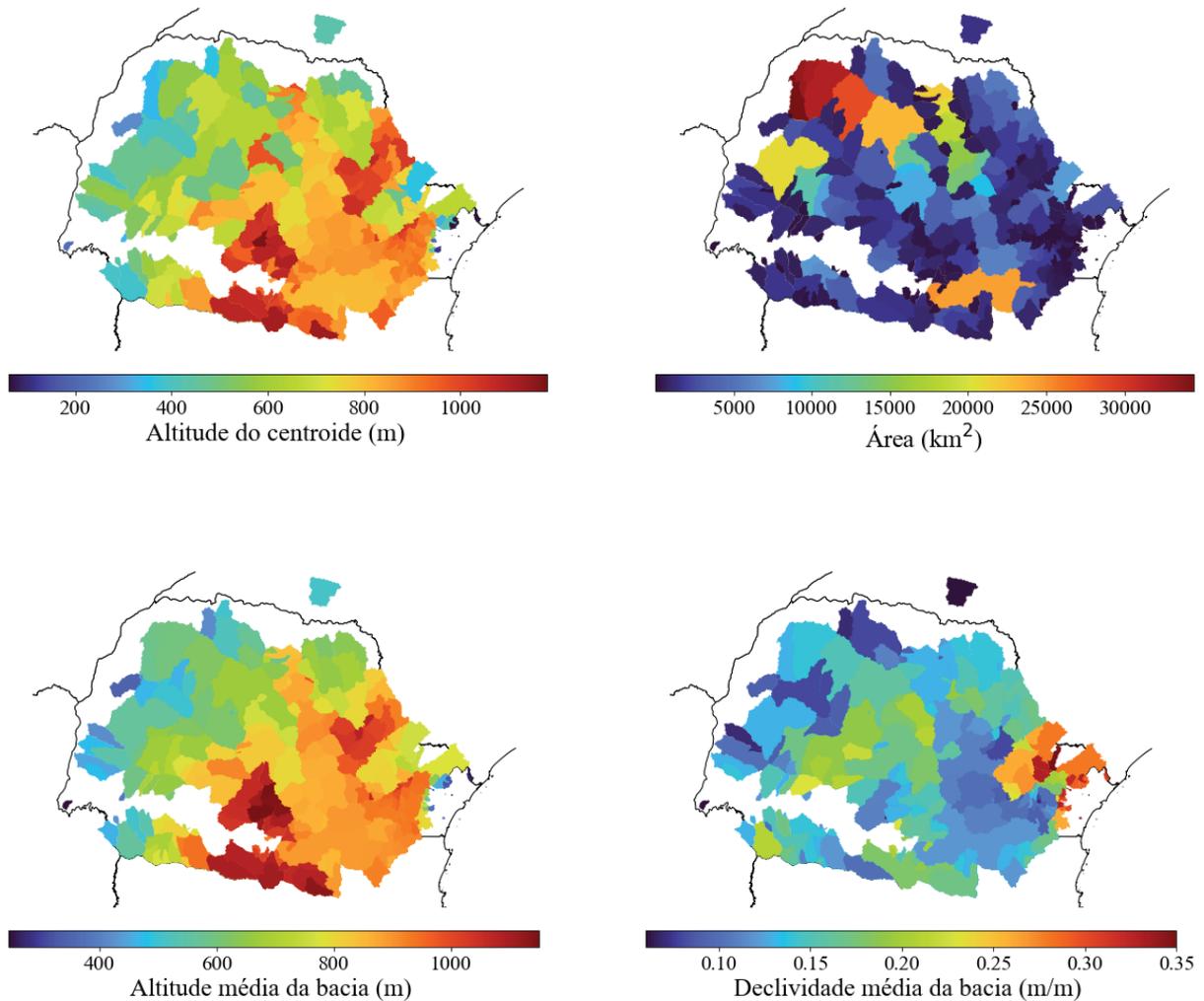
Os descritores que formam parte do conjunto de dados são apresentados na Tabela A.1 junto com algumas estatísticas descritivas.

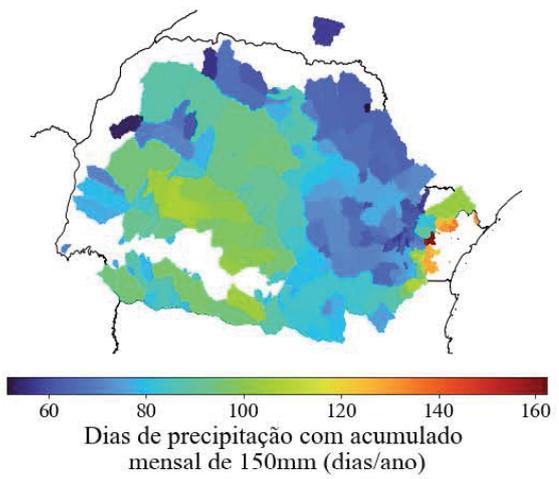
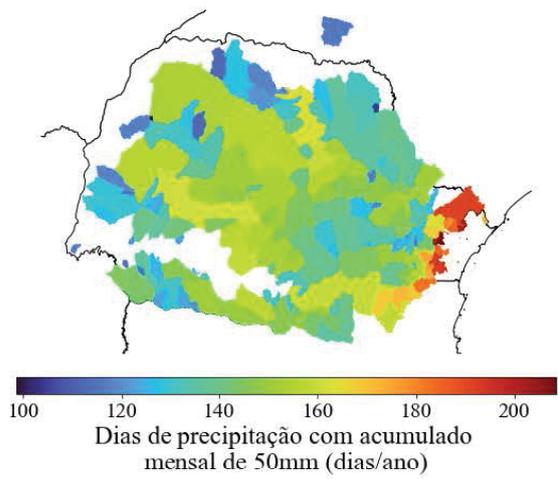
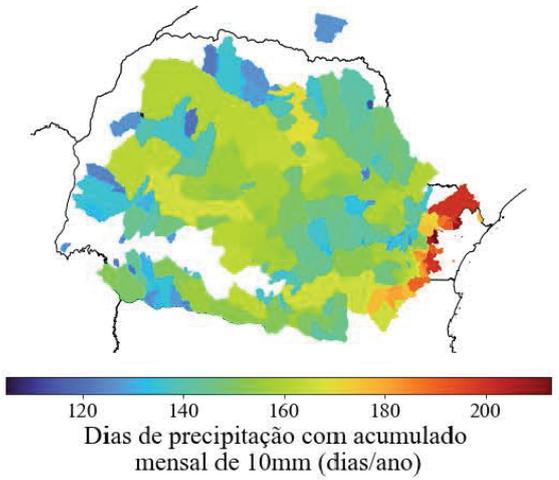
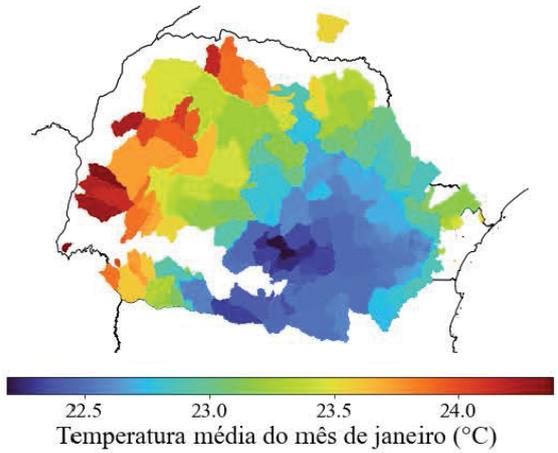
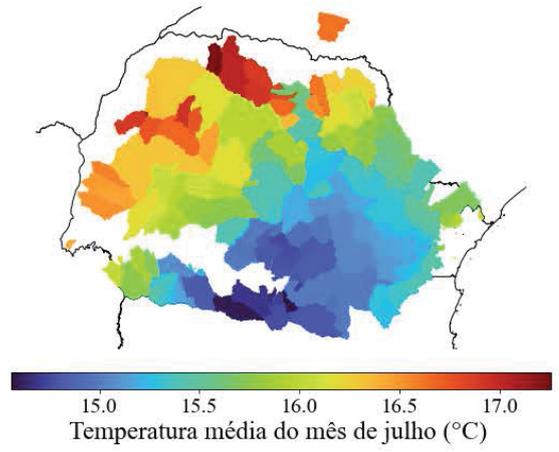
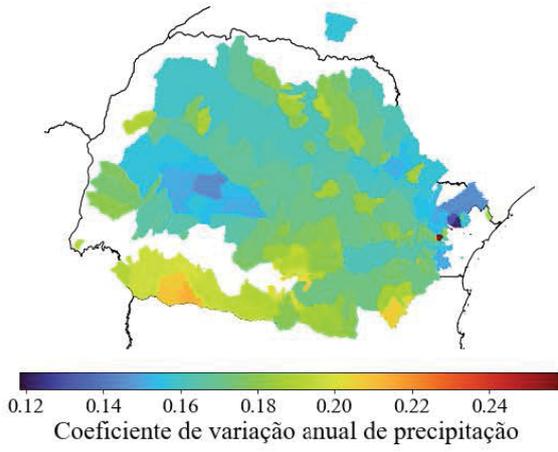
Tabela A.1: Estatísticas descritivas para as 284 bacias hidrográficas.

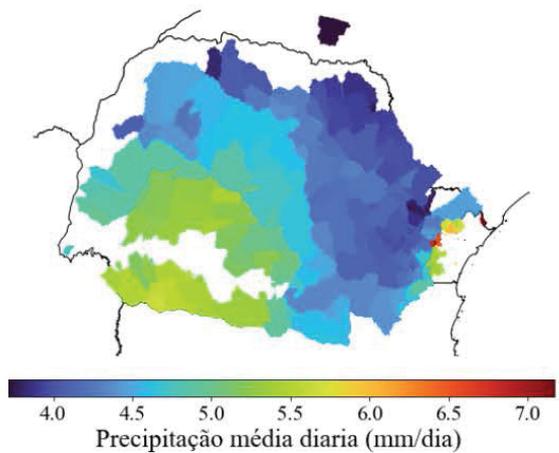
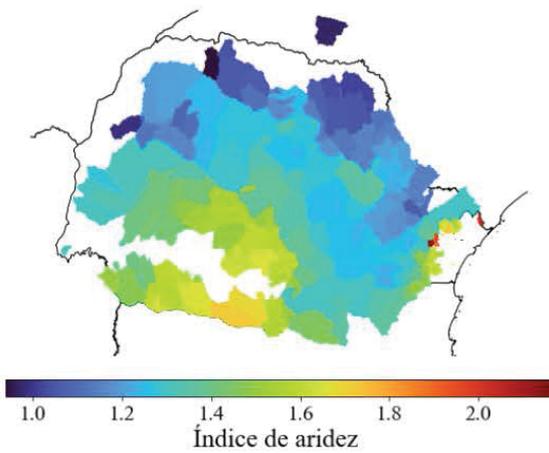
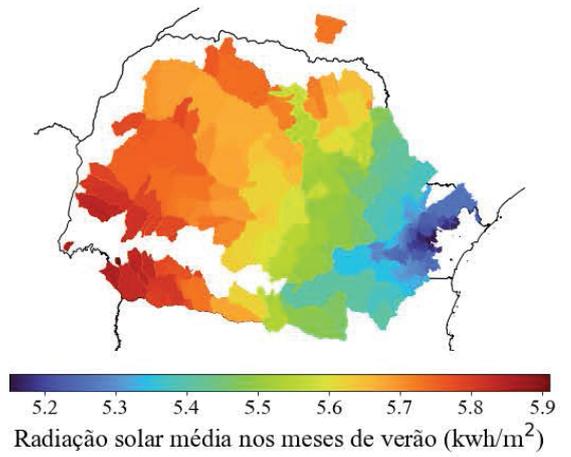
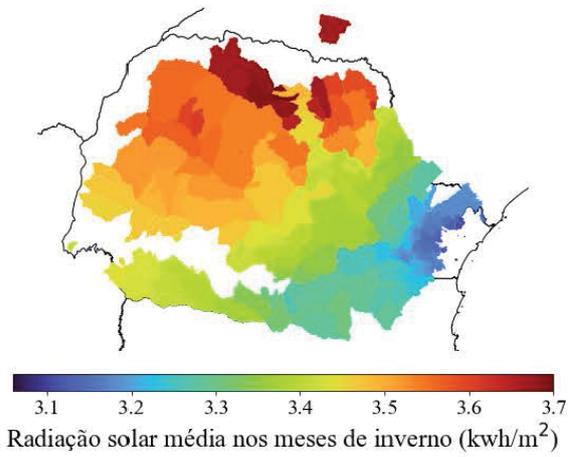
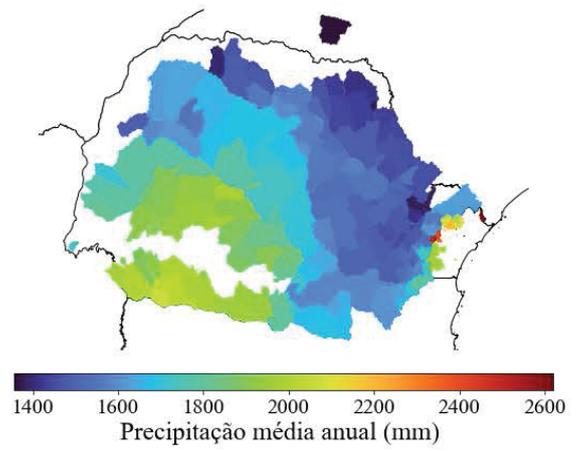
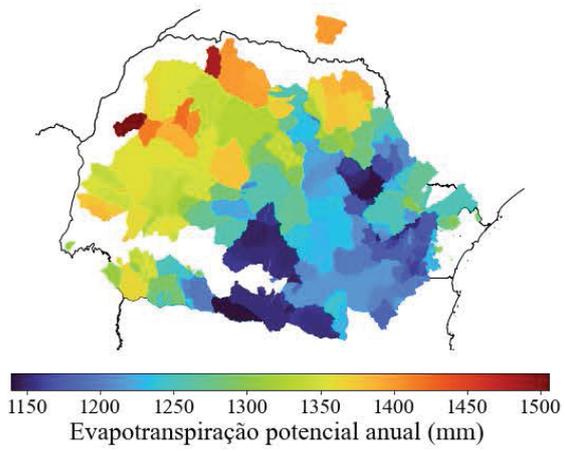
Descritores	média	desvio padrão	min	25%	50%	75%	max
<b>Índices fisiográficos</b>							
Altitude do centroide (m)	768,86	224,25	62,00	624,25	821,50	927,50	1181,00
Área (km <sup>2</sup> )	3243,51	6345,55	3,53	196,43	876,22	2610,89	34440,18
Altitude média da bacia (m)	814,35	197,82	245,03	670,09	860,62	943,33	1150,88
Declividade média da bacia (m/m)	0,15	0,06	0,06	0,11	0,14	0,17	0,35
Número de Strahler	6,30	1,40	2,00	5,00	6,00	7,00	9,00
Comprimento do rio principal (m)	143267,24	152275,70	4627,88	39842,67	89728,48	178065,81	748033,70
Densidade de drenagem (km/km <sup>2</sup> )	2,52	1,08	0,68	1,80	2,30	3,18	6,80
Declividade do rio principal (m/m)	0,04	0,02	0,02	0,03	0,03	0,04	0,12
<b>Índices climatológicos</b>							
Coefficiente de variação anual de precipitação	0,18	0,02	0,12	0,16	0,17	0,18	0,26
Temperatura média do mês de julho (°C)	15,55	0,60	14,56	15,15	15,40	15,94	17,26
Temperatura média do mês de janeiro (°C)	22,98	0,52	22,19	22,56	22,86	23,34	24,38
Dias de precipitação com acumulado mensal de 10mm (dias/ano)	150,93	19,29	104,55	136,16	148,51	161,60	213,19
Dias de precipitação com acumulado mensal de 50mm (dias/ano)	144,25	19,39	98,57	130,24	141,52	154,76	208,45
Dias de precipitação com acumulado mensal de 150mm (dias/ano)	82,97	19,01	51,38	69,39	80,35	92,92	162,21
Evapotranspiração potencial anual (mm)	1247,23	83,68	1139,11	1171,07	1230,50	1310,37	1505,70
Precipitação média anual (mm)	1701,56	222,26	1357,26	1520,40	1664,26	1866,06	2618,98
Radiação solar média nos meses de inverno (kwh/m <sup>2</sup> )	3,38	0,15	3,06	3,28	3,40	3,49	3,70
Radiação solar média nos meses de verão (kwh/m <sup>2</sup> )	5,54	0,20	5,15	5,37	5,55	5,72	5,91
Índice de aridez	1,37	0,20	0,94	1,25	1,33	1,51	2,16

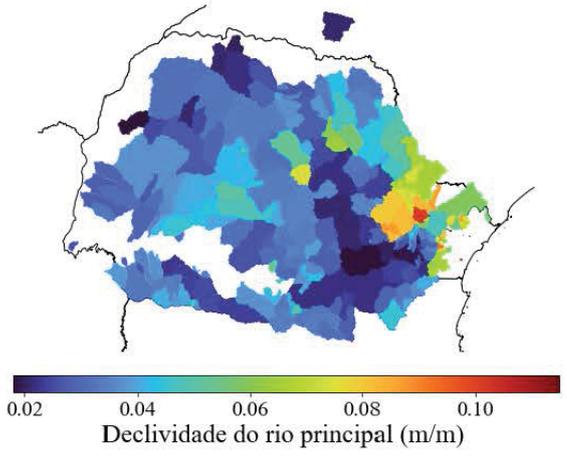
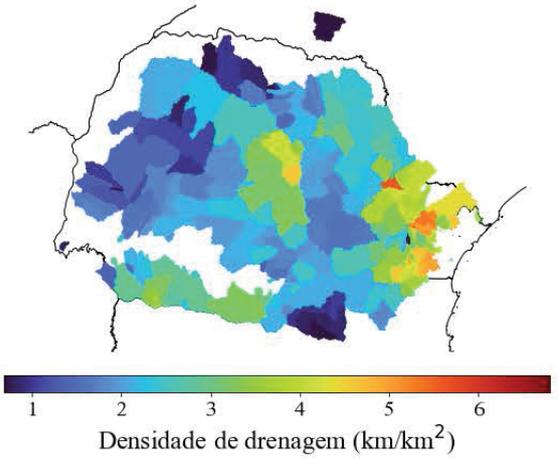
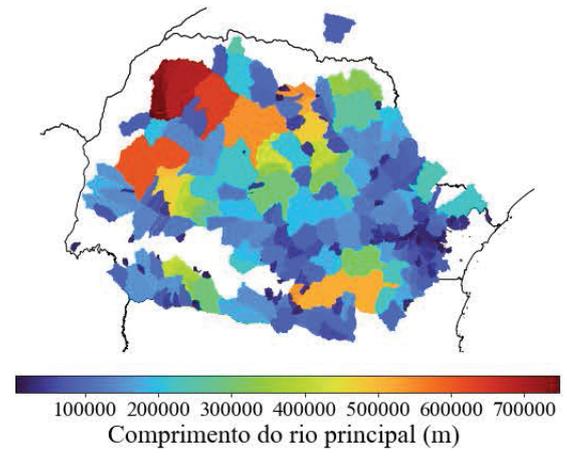
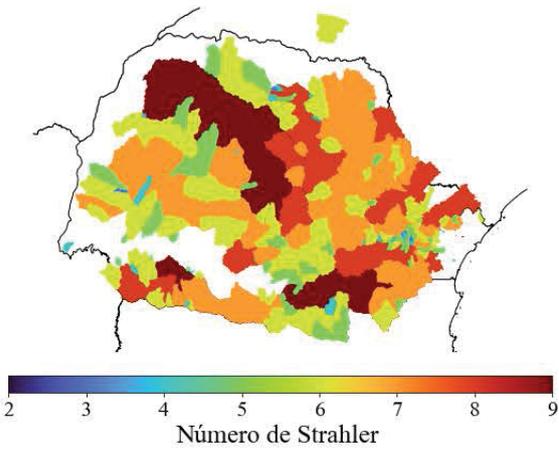
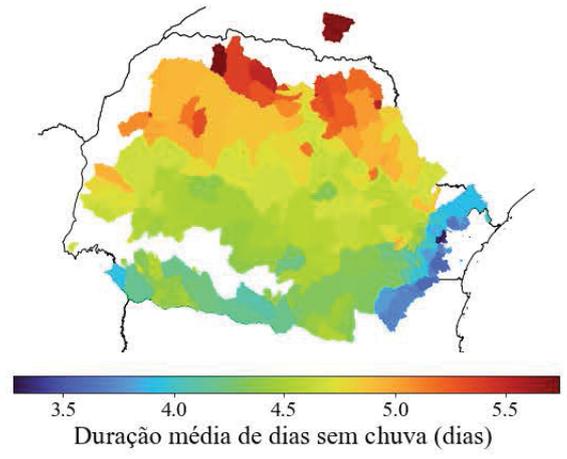
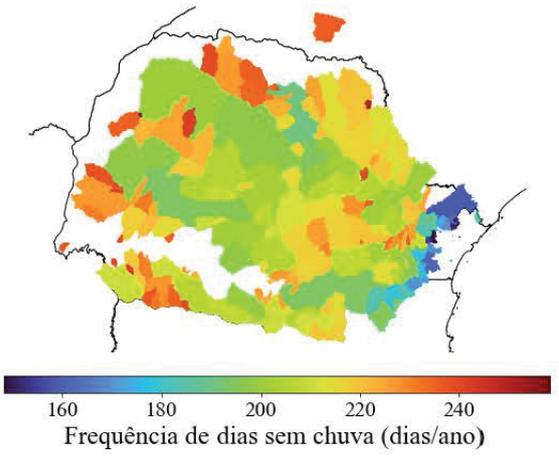
Descritores	média	desvio padrão	min	25%	50%	75%	max
Precipitação média diária (mm/dia)	4,66	0,61	3,72	4,16	4,56	5,11	7,17
Frequência de dias sem chuva (dias/ano)	210,38	19,76	148,24	199,20	212,51	225,20	258,21
Duração média de dias sem chuva (dias)	4,53	0,45	3,27	4,33	4,54	4,77	5,75
Uso e cobertura do solo							
(1) Floresta (%)	0,46	0,26	0,00	0,27	0,41	0,69	1,00
(2) Agricultura (%)	0,28	0,22	0,00	0,07	0,27	0,43	0,86
(3) Área urbana (%)	0,04	0,11	0,00	0,00	0,01	0,02	0,97
(4) Solo exposto (%)	0,00	0,00	0,00	0,00	0,00	0,00	0,02
(5) Pastagem (%)	0,21	0,13	0,00	0,14	0,19	0,27	0,80
(6) Água (%)	0,00	0,01	0,00	0,00	0,00	0,00	0,08
Tipo de solo							
(1) Latossolo (%)	0,27	0,22	0,00	0,07	0,27	0,39	1,00
(2) Neossolo (%)	0,18	0,19	0,00	0,00	0,12	0,30	0,89
(3) Argissolo (%)	0,13	0,18	0,00	0,00	0,05	0,19	0,91
(4) Nitossolo (%)	0,11	0,16	0,00	0,00	0,04	0,17	0,90
(5) Cambissolo (%)	0,20	0,23	0,00	0,00	0,11	0,30	1,00
(6) Gleissolo (%)	0,01	0,03	0,00	0,00	0,00	0,01	0,25
(7) Organossolo (%)	0,01	0,03	0,00	0,00	0,00	0,00	0,22
(8) Espodossolo (%)	0,00	0,00	0,00	0,00	0,00	0,00	0,00
(9) Afloramento rochoso (%)	0,02	0,06	0,00	0,00	0,00	0,01	0,45
(10) Área urbana (%)	0,02	0,10	0,00	0,00	0,00	0,00	0,99
Número Curva	76,91	5,89	50,60	74,58	77,89	80,49	92,90

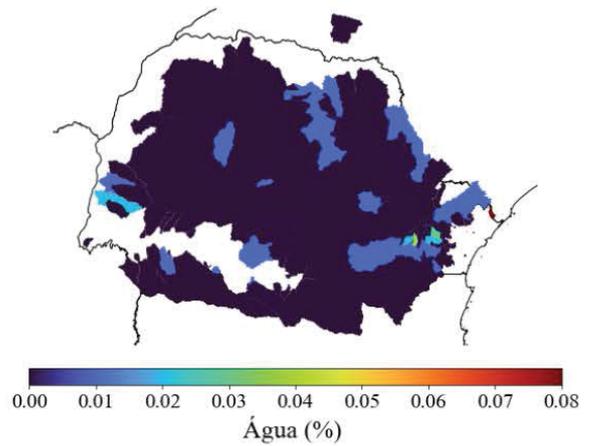
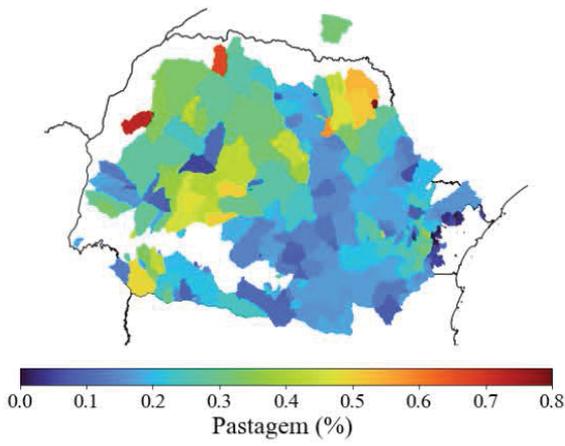
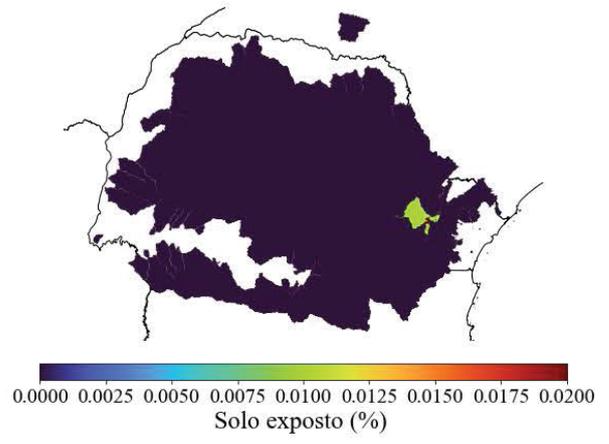
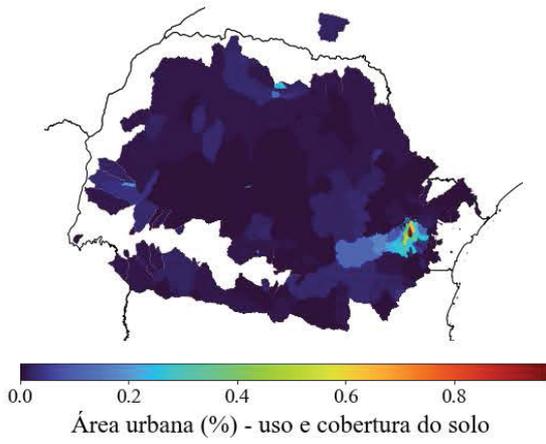
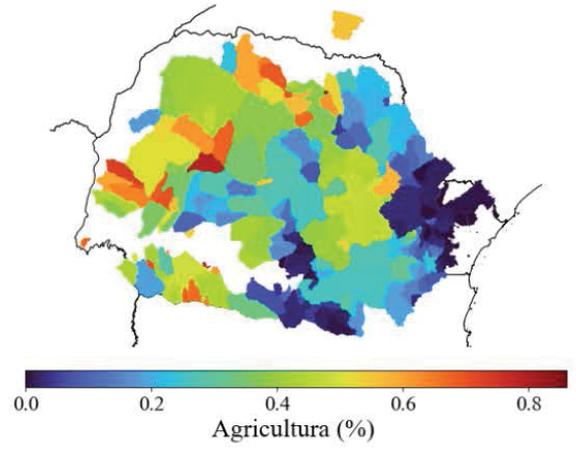
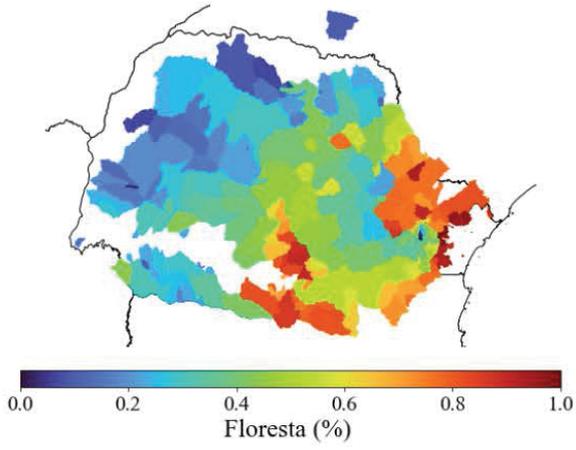
Os mapas a seguir mostram a variabilidade espacial de cada descritor nas bacias hidrográficas.

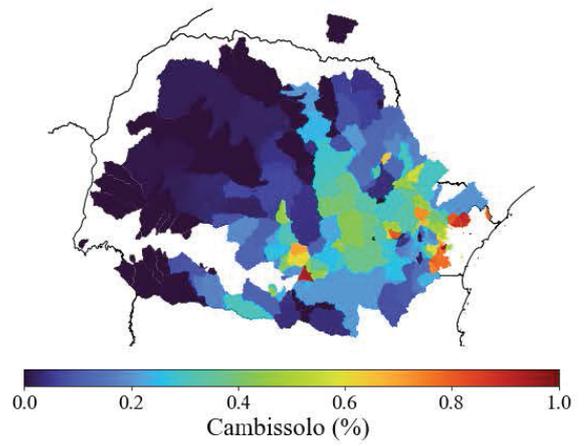
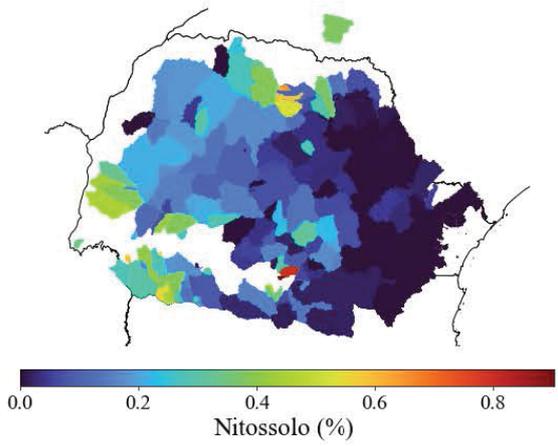
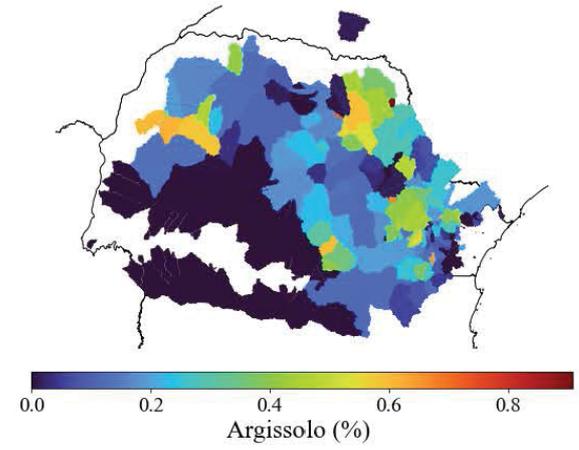
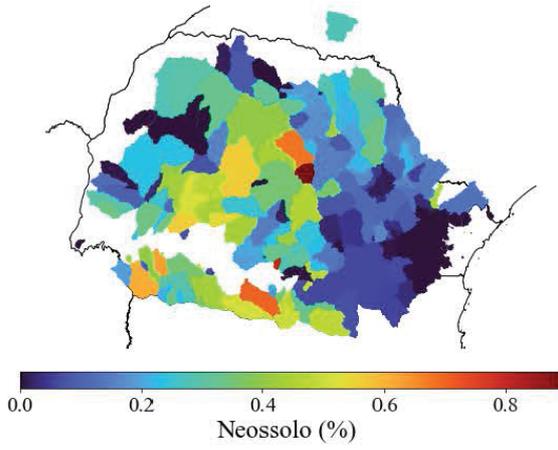
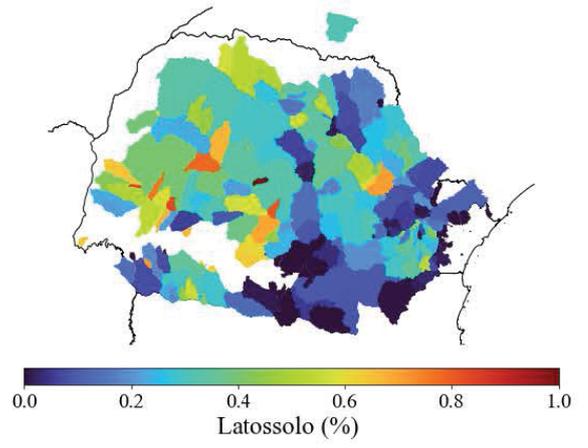
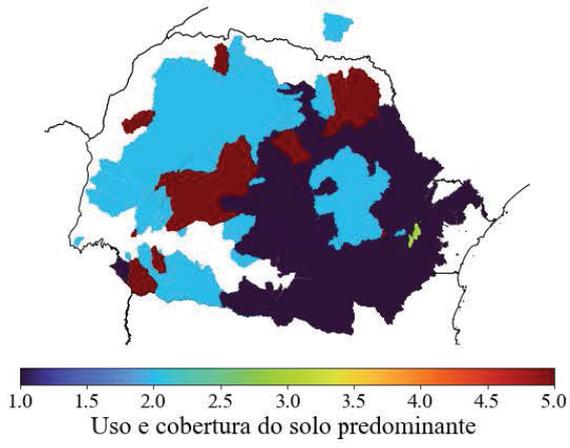


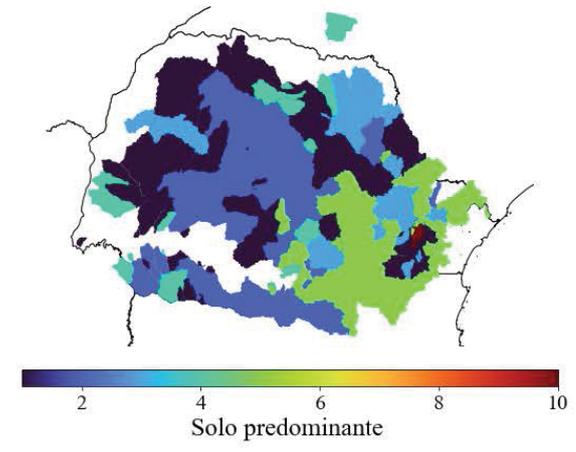
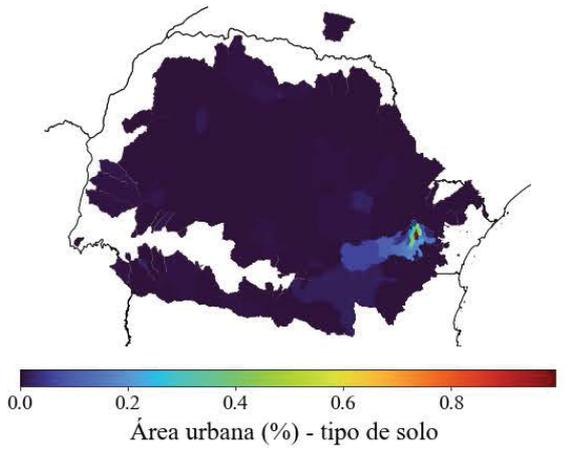
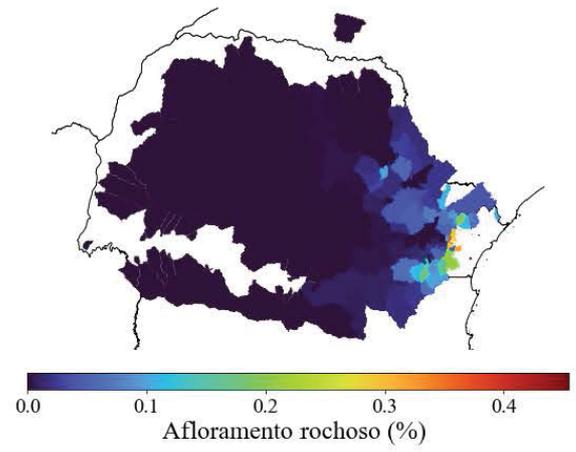
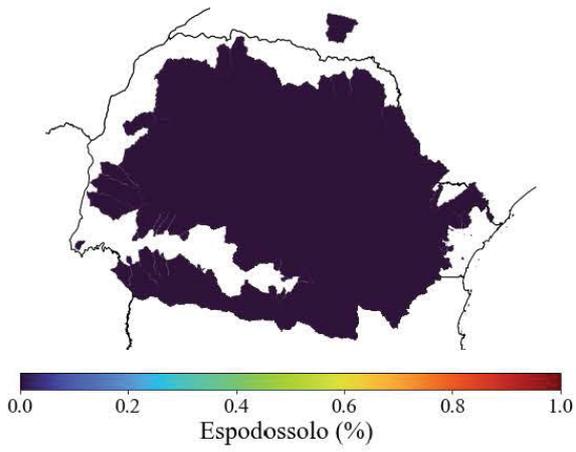
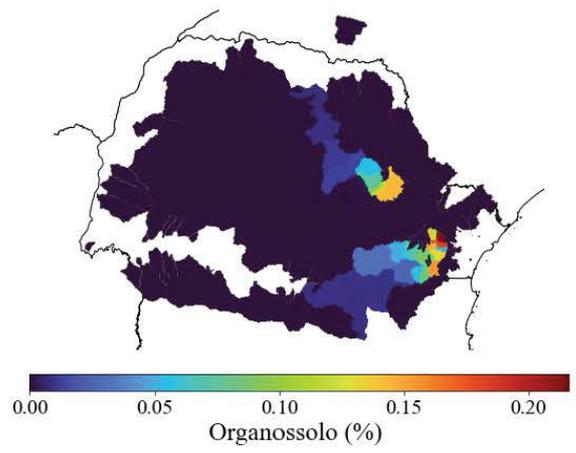
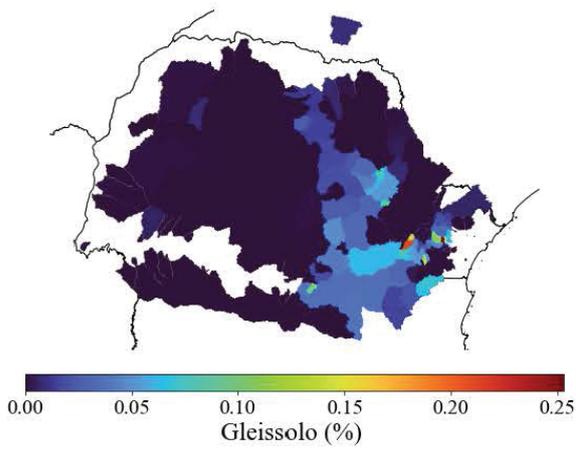


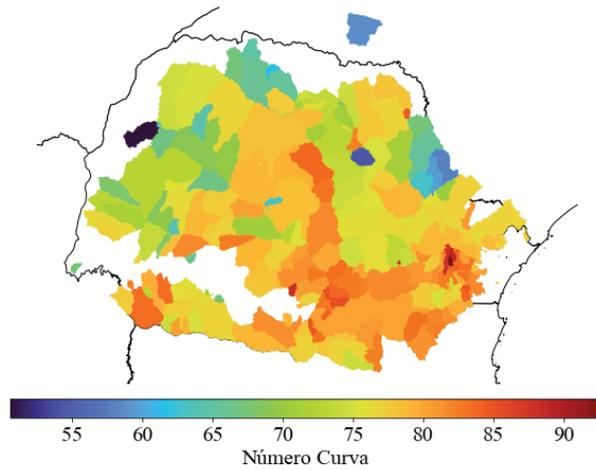












A Figura abaixo apresenta os Coeficientes de Correlação de Pearson entre os descritores.



## APÊNDICE B – FERRAMENTAS E ALGORITMOS

Nesta seção os métodos são expressos de forma resumida e estão descritas as linguagens de programação, as bibliotecas e pacotes livres que foram utilizados.

### I - Preparação dos dados

#### (1) Técnica de *Kriging* e o método de Calveti et al. (2017)

Esta etapa foi desenvolvida junto com o auxílio do Simepar. Os dados de precipitação diários, validados pelo controle de qualidade de Llabrés-Brustenga et al. (2019), foram espacializados em uma grade de 1km x 1km por meio da krigagem espaço temporal, onde os valores de precipitação foram estimados em escala diária por meio de médias ponderadas entre os dados da vizinhança. Em seguida, um segundo método de espacialização foi aplicado para o histórico de precipitação mais recente. A partir de 2013 foi empregado o método apresentado por Calveti et al. (2017) para se estimar a precipitação de forma espacializada dentro da área de interesse em uma grade de 1km x 1km, onde a equação de Poisson foi utilizada para combinar dados de radar e de satélite com os registros observados pelas estações telemétricas de precipitação. Por fim, foram aferidas precipitações médias por meio da média aritmética dos pontos da grade localizadas dentro da área de drenagem de cada bacia hidrográfica.

Uso da linguagem de programação R.

#### (2) Hidrorreferenciamento e delineamento das bacias hidrográficas

As localizações dos postos foram conferidas por meio do procedimento manual de hidrorreferenciamento utilizando a rede hidrográfica ottocodificada do IAT do ano de 2020. O hidrorreferenciamento torna a localização geográfica conhecida para o sistema de codificação de bacias criado por Otto Pfafstetter permitindo a obtenção de informações a montante e a jusante de cada trecho e, assim, o delineamento das áreas de drenagem (Sousa et al., 2009; Pfafstetter, 1989).

Uso da linguagem de programação *Python* e da biblioteca *geopandas*, a qual foi utilizada para gerar as áreas de drenagem, as hidrografias e o traçado do rio principal.

#### (3) *Regression-Kriging* ou método híbrido de interpolação

Esta etapa foi desenvolvida com o auxílio do Simepar, as informações pontuais de evapotranspiração potencial foram espacializadas utilizando um método de regressão seguido de interpolação, também conhecido como *Regression-Kriging* ou método híbrido de interpolação. O método de regressão *Random Forest* (Breiman, 2001) é um algoritmo de aprendizagem supervisionado em que são realizadas amostragens aleatórias com reposições provenientes de uma mesma amostra de dados de treinamento para formar subconjuntos de dados. Como os resíduos da regressão geram dependência espacial esses podem ser interpolados pelo método de krigagem simples (Hengl et al., 2007). A krigagem faz parte das técnicas de geoestatística, que é capaz de estipular ou predizer valores em regiões do espaço em que não possuem medidas a partir da média ponderada de pontos observados na vizinhança (Varouchakis, 2019; Grego et al., 2014).

Uso da linguagem de programação R.

#### (4) Estimativa das declividades e altitudes

O modelo digital de elevação (MDE) com resolução de 30 metros proveniente da missão *Shuttle Radar Topography Mission* (SRTM) da NASA foi usado para a estimativa das declividades e altitudes.

Uso da linguagem de programação *Python* e das bibliotecas *osgeo*, *rasterstats*, *pandas* e *numpy*.

#### (5) Cálculo dos índices climatológicos

Barbieri et al. (2017) disponibilizaram os atlas do Estado do Paraná de médias mensais de temperaturas e médias de radiação solar para cada estação do ano, os quais são mostrados nas Figuras 4.5 e 4.6. Os atlas foram produzidos com base nas medições das estações do INMET, do Simepar e do IDR-Paraná entre o período de 2006 a 2016. Essas informações foram usadas para computar índices médios nas bacias situadas dentro do estado, e para as bacias localizadas nas fronteiras entre outros estados foram adotados os valores médios da bacia mais próxima.

Uso da linguagem de programação *Python* e das bibliotecas *osgeo*, *rasterstats*, *pandas* e *numpy*.

#### (6) Cálculo dos índices de solo e uso e cobertura do solo

A paisagem pode ser descrita por meio de técnicas de sensoriamento remoto que utilizam imagens de satélites para analisar padrões espaciais sobre a superfície da Terra. O mapa de uso e cobertura do solo do ano de 2019, o qual foi utilizado, é disponibilizado pelo MapBiomias (Souza et al., 2020). As classes mostradas no mapa da Figura 4.7 foram utilizadas como base para o cálculo das frações de área que cada classe ocupa nas bacias e para determinar a classe dominante. O mapa de solos usado foi produzido e disponibilizado pela Embrapa (2020) para o estado do Paraná com escala de 1:250.000. A Figura 4.8 mostra o mapa utilizado para calcular a fração de área que cada solo ocupa e o solo dominante de cada bacia.

Uso da linguagem de programação *Python* e das bibliotecas *osgeo*, *rasterio*, *pandas*, *geopandas* e *numpy*.

## II - GR4J

### (1) Modelo GR4J

O GR4J (em Francês, *modèle du Génie Rural à 4 paramètres Journalier*) proposto por Perrin et al. (2003) faz parte da família dos modelos GR (em Francês, *Génie Rural*) de chuva-vazão desenvolvidos no Instituto Nacional de Pesquisa para Agricultura, Alimentação e Meio Ambiente (em Francês, *L'Institut National de Recherche Pour L'Agriculture, L'Alimentation et L'Environnement*) da França.

Uso da linguagem de programação *Python* e da biblioteca livre RRMFG (em inglês, *Rainfall-Runoff Modelling Playground*), a qual foi desenvolvida e disponibilizada por Kratzert et al. (2017).

## (2) Método de otimização global *Differential Evolution*(DE)

O método de otimização global *Differential Evolution* (DE), o qual foi usado para a calibração do GR4J, foi inicialmente proposto por Storn e Price (1997a). O DE é empregado em problemas de otimização que utilizam uma única função objetivo.

Uso da linguagem de programação *Python* e da biblioteca *Scipy* que possui o método *Differential Evolution* implementado.

### III - Regionalização

#### (1) Análise de agrupamentos

Os agrupamentos foram produzidos por meio do método de K-médias. A aplicação do algoritmo K-médias envolve: primeiro, definir previamente o número de  $K$  grupos, segundo, para cada grupo inicializar um centroide aleatoriamente dentro do intervalo de cada categoria, terceiro, cada ponto é atribuído ao centroide que tiver a menor distância Euclidiana em relação a ponto, quatro, é computado uma nova localização dos  $K$  centroides com base na média de todos os pontos atribuídos a ele. O processo iterativo do terceiro ao quarto passo se repete até que não haja mais alterações nos centroides (Wilks, 2011).

Uso da linguagem de programação *Python* e da biblioteca *scikit-learn* que possui o método de K-médias implementado.

#### (2) Regressão *Random Forest*

O método de regressão *Random Forest* (Breiman, 2001) é um algoritmo de aprendizagem supervisionado em que são realizadas amostragens aleatórias com reposições provenientes de uma mesma amostra de dados de treinamento para formar subconjuntos de dados. Cada subconjunto de dados é utilizado para treinar uma árvore de decisão sem que haja nenhuma interação entre elas. As predições de todas as árvores individuais são então combinadas por meio de uma média para a realização da predição final.

Uso da linguagem de programação *Python* e da biblioteca *scikit-learn* que possui o método de *Random Forest* implementado.

#### (3) Avaliação do desempenho

Os resultados foram avaliados por meio do Coeficiente de Correlação de Pearson (R), o Coeficiente de Nash-Sutcliffe (NSE) e as suas variações: a transformada da vazão pela raiz quadrada (sqrtNSE) e pelo logaritmo (logNSE). O NSE dá mais ênfase na performance das vazões mais altas, já o logNSE é mais sensível as vazões baixas e o sqrtNSE fornece um desempenho intermediário (Oudin et al., 2008).

Uso da linguagem de programação *Python* e da biblioteca *HydroErr*.

Adicionalmente, empregou-se o algoritmo desenvolvido por Hoeltgebaum (2021) e seu orientador Nelson Luís da Costa Dias, o qual foi produzido na linguagem *Python*, para identificar automaticamente os picos e os períodos das recessões dentro de uma janela móvel de 15 dias