

UNIVERSIDADE FEDERAL DE SANTA MARIA
CENTRO DE TECNOLOGIA
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA AMBIENTAL

Taiara Sampaio Santos

**REGIONALIZAÇÃO DE VAZÃO MÍNIMA EM BACIAS SEM DADOS NO SEMIÁRIDO DA
BAHIA**

Santa Maria, RS
2023

Taiara Sampaio Santos

**REGIONALIZAÇÃO DE VAZÃO MÍNIMA EM BACIAS SEM DADOS NO SEMIÁRIDO DA
BAHIA**

Dissertação de mestrado apresentado ao Programa de Pós-Graduação em Engenharia Ambiental da Universidade Federal de Santa Maria (UFSM), como requisito parcial para a obtenção do título de **Mestre em Engenharia Ambiental**.

Orientador: Dr. Daniel Gustavo Allasia Piccilli

Santa Maria, RS
2023

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001

Sampaio Santos, Taiara
Regionalização de vazão mínima em bacias sem dados no
semiárido da Bahia / Taiara Sampaio Santos.- 2023.
176 p.; 30 cm

Orientador: Daniel Gustavo Allasia Piccilli
Dissertação (mestrado) - Universidade Federal de Santa
Maria, Centro de Tecnologia, Programa de Pós-Graduação em
Engenharia Ambiental, RS, 2023

1. Regionalização de vazão 2. Bacias hidrográficas sem
dados 3. Regiões homogêneas 4. Técnicas de clusterização
5. Modelo chuva-vazão GR4J I. Allasia Piccilli, Daniel
Gustavo II. Título.

Sistema de geração automática de ficha catalográfica da UFSM. Dados fornecidos pelo autor(a). Sob supervisão da Direção da Divisão de Processos Técnicos da Biblioteca Central. Bibliotecária responsável Paula Schoenfeldt Patta CRB 10/1728.

Declaro, TAIARA SAMPAIO SANTOS, para os devidos fins e sob as penas da lei, que a pesquisa constante neste trabalho de conclusão de curso (Dissertação) foi por mim elaborada e que as informações necessárias objeto de consulta em literatura e outras fontes estão devidamente referenciadas. Declaro, ainda, que este trabalho ou parte dele não foi apresentado anteriormente para obtenção de qualquer outro grau acadêmico, estando ciente de que a inveracidade da presente declaração poderá resultar na anulação da titulação pela Universidade, entre outras consequências legais.

Taiara Sampaio Santos

**REGIONALIZAÇÃO DE VAZÃO MÍNIMA EM BACIAS SEM DADOS NO
SEMIÁRIDO DA BAHIA**

Dissertação de mestrado apresentado ao Programa de Pós-Graduação em Engenharia Ambiental da Universidade Federal de Santa Maria (UFSM), como requisito parcial para a obtenção do título de **mestre em Engenharia Ambiental**.

Aprovado em 18 de dezembro de 2023:

Daniel Gustavo Allasia Picilli, Dr (UFSM)
(Presidente/Orientador)

Fernanda Helfer, Dr (Griffith University, Australia)

Raviel Eurico Basso, Dr (UFG)

Santa Maria, RS
2023

RESUMO

REGIONALIZAÇÃO DE VAZÃO MÍNIMA EM BACIAS SEM DADOS NO SEMIÁRIDO DA BAHIA

AUTOR: Taiara Sampaio Santos
ORIENTADOR: Daniel Gustavo Allasia Picilli

A inexistência ou a escassez de dados provenientes de monitoramento hidrológico é um dos grandes empecilhos para o avanço da gestão e planejamento dos recursos hídricos. Essas informações são extremamente importantes para o semiárido, uma região que sofre com desastres naturais como secas extremas, inundações e alagamentos. Nesse sentido, este estudo teve como objetivo modelar variáveis hidrológicas para bacias sem dados no semiárido da Bahia através da regionalização de vazões com o modelo *Génie Rural à 4 Paramètres Journalier (GR4J)*. Para isso, foi utilizado um conjunto de dados do *Catchment Attributes and Meteorology for Large-sample Studies (CAMELS-BR)*, onde foram englobadas informações históricas do período de 1980 a 2018 em um conjunto de 52 bacias. Neste estudo foram empregados os métodos *k-means* e *Ward* para o agrupamento das bacias em regiões com comportamento similar. Dentro de cada região homogênea, os valores dos parâmetros do modelo foram correlacionados com características físicas e hidrogeológicas das bacias, a fim de obter os valores dos parâmetros em bacias sem dados, gerando equações por meio de regressão linear para sua determinação. Os resultados mostraram que a técnica *k-means* foi satisfatória para a obtenção de regiões homogêneas e algumas das características que influenciaram na heterogeneidade entre os clusters e homogeneidade dentro dos grupos foram: a área da bacia hidrográfica, a aridez, a declividade média da bacia, a duração média dos eventos de precipitação máxima (número de dias consecutivos 5 vezes a precipitação média diária), a evapotranspiração média anual, a longitude, a porosidade subterrânea da bacia hidrográfica, a precipitação média anual e a profundidade mediana do lençol freático. De forma geral, os modelos apresentam baixa eficiência em regiões semiáridas, o que não foi diferente neste estudo. No período analisado, o GR4J obteve o Coeficiente de Nash-Sutcliffe para o logaritmo das vazões ($Nslog$) superior a 70% em 13 (25%) estações na fase de calibração e 7 (13%) na fase de validação. No processo de regionalização dos parâmetros do modelo a técnica de regressão linear não foi adequada para obtenção de equações. O modelo mostrou-se eficiente para a obtenção das vazões de referência com permanência de 90% na série histórica (Q90) e a vazão mínima referente a setes dias de duração com tempo de retorno (TR) igual a 10 anos (Q7,10), apresentando um R^2 maior que 99%. Ao relacionar a Q90 obtida por meio de dados simulados e observados com a área das bacias, apenas um dos grupos apresentou P-valor menor que 0,05, evidenciando a necessidade de testar outros métodos no intuito de subsidiar decisões relativas ao planejamento e à gestão dos recursos hídricos no semiárido da Bahia.

Palavras-chave: Bacias sem dados. Regiões homogêneas. Regionalização de vazão. Semiárido. GR4J.

ABSTRACT

REGIONALIZATION OF MINIMUM FLOW IN BASINS WITHOUT DATA IN THE SEMI-ARID OF BAHIA

AUTHOR: Taiara Sampaio Santos
ADVISOR: Daniel Gustavo Allasia Piccilli

The lack or scarcity of data from hydrological monitoring is one of the major obstacles to the advancement of water resources management and planning. This information is extremely important for the semi-arid, a region that suffers from natural disasters such as extreme droughts, floods and flooding. In this sense, this study aimed to model hydrological variables for basins without data in the semi-arid of Bahia through the regionalization of flows with the Génie Rural model to 4 Paramètres Journalier (GR4J). For this, we used a data set from the Catchment Attributes and Meteorology for Large-sample Studies (CAMELS-BR), which included historical information from 1980 to 2018 in a set of 52 basins. In this study, k-Means and Ward methods were used to group the basins in regions with similar behavior. Within each homogeneous region, the values of the parameters of the model were correlated with physical and hydrogeological characteristics of the basins, in order to obtain the values of the parameters in basins without data, generating equations through linear regression for their determination. The results showed that the k-Means technique was satisfactory for obtaining homogeneous regions and some of the characteristics that influenced the heterogeneity between the clusters and homogeneity within the groups were: the area of the watershed, aridity, the average slope of the basin, the average duration of the maximum precipitation events (number of consecutive days 5 times the average daily precipitation), the average annual evapotranspiration, the longitude, the underground porosity of the watershed, the mean annual precipitation and the median depth of the water table. In general, the models present low efficiency in semi-arid regions, which was not different in this study. In the period analyzed, the GR4J obtained the Nash-Sutcliffe coefficient for the logarithm of flows (Nslog) higher than 70% in 13 (25%) stations in the calibration phase and 7 (13%) in the validation phase. In the process of regionalization of the parameters of the model the linear regression technique was not suitable for obtaining equations. The model was efficient to obtain the reference flow rates with 90% permanence in the historical series (Q90) and the minimum flow rate for seven days of duration with return time (RT) equal to 10 years (Q7,10), presenting a R^2 greater than 99%. When relating the Q90 obtained through simulated data and observed with the area of the basins, only one of the groups presented P-value lower than 0.05, showing the need to test other methods in order to support decisions regarding the planning and management of water resources in the semi-arid region of Bahia.

Keywords: Ungauged basins. Homogeneous regions. Regionalization of flow. Semi-arid. GR4J

LISTA DE FIGURAS

Figura 1 - Procedimento a ser seguido no processo de regionalização de parâmetros.....	21
Figura 2 - Classificação das técnicas de agrupamentos.....	29
Figura 3 - Modelos do tipo chuva-vazão usados em regiões semiáridas.....	34
Figura 4 - Família GR de acordo com o espaço de tempo e seus respectivos autores.	39
Figura 5 - Esquema do modelo GR4J proposto por Perrin <i>et al.</i> (2003).....	40
Figura 6 - Modelo GR4J aplicado em regiões semiáridas.....	50
Figura 7 - Delimitação das bacias hidrográficas presente na base de dados CAMELS-BR. ...	53
Figura 8 - Procedimentos para obtenção de vazões estatisticamente consistentes usando o GR4J.	55
Figura 9 - Localização do semiárido da Bahia, divisão de acordo com as sub-regiões do Nordeste e a delimitação das bacias hidrográficas da base de dados CAMELS-BR.	56
Figura 10 - Tipologia climática segundo Thornthwaite para o estado da Bahia. Pluviometria 1943 - 1983/ Temperatura 1961 - 1990.....	60
Figura 11 - Resumo das vazões pertencentes a região hidrográfica do São Francisco e Atlântico Leste localizadas no semiárido da Bahia (1980 - 2018).....	61
Figura 12 - Formação hidrogeológica presente no semiárido da Bahia.	62
Figura 13 - Elevação média do terreno no semiárido da Bahia.....	63
Figura 14 - Caracterização do solo no semiárido da Bahia.	63
Figura 15 - Uso e ocupação das terras no semiárido da Bahia.....	64
Figura 16 - Relação entre a capacidade total de armazenamento do reservatório da bacia hidrográfica e a sua vazão anual total.....	65
Figura 17 - Reservatórios localizados no semiárido da Bahia monitorados pela ANA.	65
Figura 18 - Cobertura vegetal no semiárido da Bahia.	66
Figura 19 - Esquema para o agrupamento de bacias.	68
Figura 20 - Esquema para a entrada de dados no Pacote airGR no <i>software R</i>	73
Figura 21 - Divisão das bacias de acordo com a existência de parâmetros.....	74
Figura 22 - Espacialização dos clusters formados após as três análises pelo método de <i>Ward</i> e <i>K-means</i>	77
Figura 23 - Método <i>Elbow</i> para a determinação do número de cluster para as três análises. ..	77
Figura 24 - Espacialização do melhor conjunto de parâmetros do modelo GR4J para o semiárido da Bahia.....	88

Figura 25 - Espacialização dos coeficientes Nslog e LOGQ para os períodos de calibração e validação do modelo na área de estudo.	91
Figura 26 - Resumo comparativo dos parâmetros do modelo GR4J.	92
Figura 27 -Espacialização dos coeficientes NSE e KGE para os períodos de calibração e validação do modelo.	96
Figura 28 - Hidrograma da vazão simulada e observada e hietograma da precipitação na estação 51350000 (NSE=0,72).	97
Figura 29 - Hidrograma da vazão simulada e observada e hietograma da precipitação na estação 45298000 (NSE=0,70).	97
Figura 30 - Correlações entre as vazões de referência Q90 e Q7,10 obtidas pelas vazões observadas e as simuladas no período de calibração do modelo GR4J.	126
Figura 31 - Curvas de permanência do grupo 1 obtido pela técnica <i>k-means</i>	127
Figura 32 - Curvas de permanência do grupo 2 obtido pela técnica <i>k-means</i>	128
Figura 33 - Curvas de permanência do grupo 3 obtido pela técnica <i>k-means</i>	128
Figura 34 - Curvas de permanência dos Grupos 1 e 2.	129
Figura 35 - Curvas de permanência dos Grupos 3 e 4.	129
Figura 36 - Curvas de permanência dos Grupos 5 e 6.	129
Figura 37 - Curvas de permanência do Grupo 7.	130
Figura 38 - Espacialização dos sete clusters formados após análise das curvas de permanência	131

LISTA DE TABELA

Tabela 1 - Valores dos parâmetros do modelo GR4J com intervalo de confiança de 80%. ---	42
Tabela 2 - Limites dos parâmetros do modelo GR4J. -----	43
Tabela 3 - Parâmetros do modelo GR4J obtidos por autores que estudaram áreas semiáridas. -----	43
Tabela 4 - Código das estações da ANA com suas respectivas localizações, áreas e vazões para o período de 1980 a 2018, pertencentes ao semiárido da Bahia. -----	58
Tabela 5 - Atributos selecionados após a correlação de Pearson. -----	69
Tabela 6 - Limites dos parâmetros do modelo GR4J. -----	71
Tabela 7 - Primeira análise dos agrupamentos formados pelo método <i>Ward</i> .-----	78
Tabela 8 - Primeira análise dos agrupamentos formados pelo método <i>k-means</i> . -----	79
Tabela 9 - Estatística descritiva dos clusters formados pelos métodos <i>k-means</i> e <i>Ward</i> para a análise 1.-----	80
Tabela 10 - Segunda análise dos agrupamentos formados pelo método <i>Ward</i> . -----	81
Tabela 11 - Segunda análise dos agrupamentos formados pelo método <i>k-means</i> .-----	82
Tabela 12 - Estatística descritiva dos clusters formados pelos métodos <i>k-means</i> e <i>Ward</i> para a análise 2.-----	83
Tabela 13 - Terceira análise dos agrupamentos formados pelo método <i>Ward</i> . -----	84
Tabela 14 - Terceira análise dos agrupamentos formados pelo método <i>k-means</i> . -----	85
Tabela 15 - Estatística descritiva dos clusters formados pelos métodos <i>k-means</i> e <i>Ward</i> para a análise 3.-----	86
Tabela 16 - Melhor conjunto de parâmetros e critérios de eficiência do modelo GR4J para o período de calibração e validação para as vazões mínimas. -----	89
Tabela 17 - Melhor conjunto de parâmetros e critérios de eficiência do modelo GR4J para o período de calibração e validação para as vazões máximas.-----	95
Tabela 18 - Parâmetros do modelo GR4J com KGE igual ou superior a 0.70.-----	99
Tabela 19 - Estatística descritiva dos parâmetros do modelo para o período de calibração. -	100
Tabela 20 - Análise de correlação dos parâmetros do modelo GR4J com as características das bacias. -----	102
Tabela 21 - Estatística descritiva dos parâmetros do modelo para o período de calibração do Grupo 1.-----	103
Tabela 22 - Estatística descritiva dos parâmetros do modelo para o período de calibração do Grupo 2.-----	104

Tabela 23 - Estatística descritiva dos parâmetros do modelo para o período de calibração do Grupo 3.-----	104
Tabela 24 - Análise de correlação dos parâmetros do modelo GR4J com as características das bacias dentro de cada cluster.-----	105
Tabela 25 - Estatística descritiva das equações geradas utilizando as estações do Cluster 1, considerando apenas uma amostra de treino.-----	116
Tabela 26 - Estatística descritiva das equações geradas utilizando as estações do Cluster 1.	117
Tabela 27 - Estatística descritiva das equações geradas utilizando as estações do Cluster 2.	119
Tabela 28 - Estatística descritiva das equações geradas utilizando as estações do Cluster 3.	122
Tabela 29 - Vazões de referência para as estações localizadas no semiárido da Bahia.-----	125
Tabela 30 - Valores médio das vazões de referência para os clusters localizadas no semiárido da Bahia.-----	127
Tabela 31 - Estatística descritiva da Q90 (m ³ /s) com dados observados de vazão por cluster.-----	130
Tabela 32 - Estatística descritiva da Q90 (m ³ /s) com dados simulados de vazão por cluster.	130
Tabela 33 - Estatística descritiva das equações geradas utilizando as estações dos grupos e as áreas de correspondentes.-----	132

LISTA DE SIGLAS

ANA	Agência Nacional de Águas e Saneamento Básico
AWBM	<i>Australian Water Balance Model</i>
CAMELS	<i>Catchment Attributes and Meteorology for Large-sample Studies</i>
CAWM IV	<i>Campus Agreste Watershed Model - Version IV</i>
CEMAGREF	<i>Centre d'Étude du Machinisme Agricole et du Génie Rural des Eaux et Forêts</i>
CEREVE	<i>Centre d'Enseignement et de Recherche Eau, Ville, Environnement</i>
CONDEL	Conselho Deliberativo
CPC	Centro de Previsão Climática
ECMWF	<i>European Centre for Medium-Range Weather Forecasts</i>
FIV	Fator de Inflação de Variância
GAMLSS	<i>Generalized additive models for location, scale and shape</i>
GLEAM	<i>Global Land Evaporation Amsterdam Model</i>
GR2M	<i>Génie Rural à 2 paramètres mensuel</i>
GR3	<i>Génie Rural à trois paramètres</i>
GR3J	<i>Génie Rural à 3 paramètres Journalier</i>
GR4J	<i>Génie Rural à 4 Paramètres Journalier</i>
GR6J	<i>Génie Rural à 6 Paramètres Journalier</i>
GSIM	<i>Global Streamflow Indices and Metadata Archive</i>
HEC-HMS	<i>Hydrological Engineering Center's - Hydrologic Modeling System</i>
IAHS	<i>International Association of Hydrological Sciences</i>
IDW	Inverso da distância ponderada
KGE	Coefficiente de eficiência de Kling-Gupta
MGB-IPH	Modelo de Grandes Bacias
MSWEP	<i>Multi-Source Weighted-Ensemble Precipitation</i>
MODHAC	Modelo Hidrológico Auto-Calibrável
NSE	Coefficiente de eficiência de Nash-Sutcliffe
PUB	<i>Predictions in Ungauged Basins</i>
RMSE	Raiz do Erro Quadrático Médio
RENQ	Raiz do Erro Médio Quadrático
REMQ _{BV}	Raiz do Erro Médio Quadrático modificado para baixas vazões
SUDENE	Superintendência do Desenvolvimento do Nordeste

SMARG	<i>Soil Moisture Accounting and Routing</i>
SWAT	<i>Soil and Water Assessment Tool</i>
TOPSIMPL	TOPmodel SIMPLifiée
WCSS	<i>Within Clusters Sum of Squares</i>

SUMÁRIO

1.	INTRODUÇÃO -----	15
2.	OBJETIVOS -----	17
2.1.	OBJETIVO GERAL-----	17
2.2.	OBJETIVOS ESPECÍFICOS -----	17
3.	REVISÃO BIBLIOGRÁFICA -----	18
3.1.	REGIONALIZAÇÃO HIDROLÓGICA -----	18
3.1.1.	Método de regionalização -----	24
3.1.1.1	Regionalização baseada na proximidade espacial	25
3.1.1.2	Regionalização baseada na similaridade climáticas e/ou fisiográficas.....	25
3.1.1.3	Regionalização baseada na regressão de dados	26
3.2.	IDENTIFICAÇÃO DAS REGIÕES HOMOGÊNEAS -----	27
3.3.	MODELOS DE CHUVA-VAZÃO EM REGIÕES SEMIÁRIDAS -----	32
3.4.	MEDIDAS DE DESEMPENHO -----	36
3.5.	MODELO GR4J -----	38
3.5.1.	Contexto Histórico -----	38
3.5.2.	GR4J proposto por Perrin <i>et al.</i> (2003) -----	40
3.5.3.	Calibração do modelo GR4J -----	43
3.5.4.	Aplicabilidade do Modelo GR4J -----	45
3.5.5.	Comparação do GR4J com outros modelos -----	47
3.5.6.	Aplicação do GR4J em regiões semiáridas -----	50
3.6.	BASE DE DADOS DO CAMELS-BR-----	52
4.	MATERIAIS E MÉTODOS -----	55
4.1.	ÁREA DE ESTUDO -----	55
4.1.1.	Características pluviométricas e climáticas -----	60
4.1.2.	Características fluviométricas -----	61
4.1.3.	Hidrogeologia, relevo, pedologia, uso e ocupação do solo, índice de intervenção humana e vegetação -----	62
4.2.	IDENTIFICAÇÃO DAS REGIÕES HOMOGÊNEAS NO SEMIÁRIDO DA BAHIA	66
4.2.1.	Padronização dos dados -----	68
4.2.2.	Análise de correlação -----	68
4.2.3.	Análise Cluster -----	69

4.2.4.	Análise dos agrupamentos-----	69
4.3.	REGIONALIZAÇÃO DE PARÂMETROS DO MODELO GR4J PARA BACIAS SEM DADOS NO SEMIÁRIDO DA BAHIA -----	70
4.3.1.	Análise de dados-----	70
4.3.2.	Objetivo da regionalização -----	71
4.3.3.	Modelo GR4J-----	71
4.3.4.	Requisitos de dados no pacote airGR -----	72
4.3.5.	Análise do desempenho do modelo GR4J -----	73
4.3.6.	Regionalização de parâmetros do modelo GR4J para bacias sem dados -----	73
4.3.7.	Obtenção das vazões de referência -----	75
5.	RESULTADOS -----	77
5.1.	IDENTIFICAÇÃO DAS REGIÕES HOMOGÊNEAS -----	77
5.2.	ANÁLISE DO DESEMPENHO DO MODELO GR4J -----	87
5.3.	REGIONALIZAÇÃO DE PARÂMETROS DO MODELO GR4J PARA BACIAS SEM DADOS -----	99
5.4.	OBTENÇÃO DAS VAZÕES DE REFERÊNCIA -----	125
6.	CONSIDERAÇÕES FINAIS -----	133
	REFERÊNCIAS -----	136
	APÊNDICE A - HIDROGRAMA DA VAZÃO E HIETOGRAMA DA PRECIPITAÇÃO NAS ESTAÇÕES AVALIADAS -----	151
	APÊNDICE B - PERÍODOS COM DADOS DE VAZÃO DISPONÍVEIS EM CADA ESTAÇÃO-----	157
	APÊNDICE C - RESUMO DAS VAZÕES E PRECIPITAÇÕES DAS ESTAÇÕES AVALIADAS-----	161
	APÊNDICE D - CURVAS DE PERMANÊNCIA DAS ESTAÇÕES -----	172

1. INTRODUÇÃO

A disponibilidade de dados hidrológicos é fundamental para o adequado planejamento e gestão dos recursos hídricos (PEREIRA *et al.*, 2020; ANDRADE *et al.*, 2021). Essas informações são essenciais para estudos de previsão de cheias, para estimativa da disponibilidade hídrica, para outorga, além do cálculo de vazões afluentes anuais para o dimensionamento de reservatórios, balanço hídrico em sistemas de apoio à decisão e dimensionamento de obras hidráulicas (CHAVES *et al.*, 2002; MELLER, 2012; SILVA *et al.*, 2006; SILVA JÚNIOR *et al.*, 2003; SOUSA, 2009), dentre outras aplicações. Entretanto, a escassez de dados ainda é uma realidade em muitos cursos d'água mundialmente (HANNAH *et al.*, 2011; SIVAPALAN *et al.*, 2003), sendo particularmente crítica em regiões semiáridas (AL-QURASHI *et al.*, 2008, *apud* RODIER, 1985; SORMAN; ABDULRAZZAQ, 1993; AL-QURASHI, 1995), devido ao acometimento de desastres naturais como secas extremas, inundações e alagamentos.

Uma solução para suprir a ausência desses dados é a transferência de informações de bacias hidrográficas vizinhas em um processo denominado regionalização hidrológicas. Existem diferentes técnicas para a regionalização, como por exemplo, estatísticas (BRUSA, 2004; GASQUES *et al.*, 2018; HE; BÁRDOSSY; ZEHE, 2011; SILVA *et al.*, 2006a; TUCCI, 2017) ou ainda, recorrer a modelos do tipo chuva-vazão, desenvolvidos para fornecer dados tendo por base informações de clima, topografia, geologia, tipo de solo e uso e ocupação da terra (BECK *et al.*, 2016, VIRÃES 2018, KUANA *et al.*, 2022).

Na literatura, diversos modelos são mencionados com essa finalidade, tais como: *Campus Agreste Watershed Model Version IV* (CAWM IV) (CIRILO *et al.*, 2020; FEITOZA 2021), Modelo Hidrológico Auto-Calibrável (MODHAC) (VIRÃES 2018; VIRÃES; CIRILO 2019), Modelo de Grandes Bacias (MGB-IPH) (PESSOA, 2017), *Hydrological Engineering Center's - Hydrologic Modeling System* (HEC-HMS) (ROCHA, 2020), *Soil and Water Assessment Tool* (SWAT) (SANTOS *et al.*, 2013), *Soil Moisture Accounting Procedure* (SMAP) (ALEXANDRE *et al.*, 2005; Saraiva 2010), *Génie Rural à 4 Paramètres Journalier* (GR4J) (LEBECHEREL *et al.*, 2014; POISSANT *et al.*, 2017; ARSENAULT *et al.*, 2019; KUANA *et al.*, 2020; GOLIAN *et al.*, 2021; WANG *et al.*, 2021; KUANA, 2022), Sistema Computacional para Regionalização de Vazões (SisCoRV) (SOUSA, 2009), entre outros.

Há que se considerar, todavia, que um dos grandes problemas para a utilização de tais modelos é a determinação de seus parâmetros, uma vez que, em regiões com carência de dados

como as semiáridas, existem dificuldades para a calibração dos parâmetros dos modelos, ou ainda, dificuldades em encontrar informações comuns, que possibilitem que os dados calibrados em uma bacia possam ser utilizados em bacias não calibradas (CARNEIRO; FARIAS 2013; SILVA FILHO *et al.*, 2020). Dessa forma, trabalhar com modelos parcimoniosos, que possuam exigência de menos parâmetros, pode ser uma alternativa interessante, gerando menor incerteza na sua calibração e maior facilidade para obtenção de informações (GARRIDO 2003; WAGENER *et al.*, 2004; HRACHOWITZ *et al.*, 2013; PESSOA 2017).

No entanto, nas regiões semiáridas, assim como a maior parte do Nordeste brasileiro, que são áreas com grandes irregularidades pluviométricas, temporais e espaciais, a maioria dos modelos hidrológicos descritos na literatura, que podem representar bem o escoamento em regiões úmidas, não repetem o desempenho em territórios semiáridos (HUANG *et al.*, 2016). Al-Qurashi *et al.* (2008), *apud* RODIER, 1985; SORMAN e ABDULRAZZAQ, 1993; AL-QURASHI, 1995, atribuem essa dificuldade à escassez geral de dados de precipitação, vazão, propriedades do solo e condições iniciais de umidade, além da influência da variabilidade sazonal e interanual da vegetação; da complexidade da morfologia do curso de água; da dificuldade de quantificar as perdas por transbordamento; e da imprecisão na estimativa de evapotranspiração potencial.

O GR4J é um modelo de chuva-vazão caracterizado como parcimonioso com quatro parâmetros e funcionamento no intervalo de tempo diário (ANDRADE *et al.*, 2021; KUANA *et al.*, 2021; KUANA 2022).

É considerado um modelo robusto, com estrutura concisa, com aplicações e resultados em uma grande amostra de bacias hidrográficas (ANDRADE *et al.*, 2021). Diversos autores aplicaram o modelo em regiões semiáridas, como Araújo (2005) em bacias hidrográficas localizadas na região do Cariri e do sertão paraibano, Ruelland *et al.* (2014) e Hublart *et al.* (2015) no Chile, Bodian *et al.* (2018) no Senegal na África Ocidental, Hajhouji *et al.* (2018) na bacia do Rio Rheraya (Haut Atlas, Marrocos), Guan *et al.* (2019) em seis sub-bacias do Rio Amarelo, localizada na China, Andrade *et al.* (2021) nas bacias hidrográficas do rio Capibaribe (em Pernambuco), rio Pindaré (no Maranhão), rio Salgado (no Ceará), rio Sirinhaém (em Pernambuco) e no rio Piatinga (em Sergipe), Liu *et al.* (2021) na China, Oueldkaddour *et al.* (2021) no Marrocos, Wanzala *et al.* (2022) no Quênia, Sezen e Partal (2022) na bacia Antalya na Turquia e Cantoni *et al.* (2022) em seis bacias hidrográficas na Tunísia, no norte da África.

Frente aos pressupostos abordados, o objetivo dessa pesquisa é regionalizar as vazões mínimas para bacias sem dados em regiões semiáridas com o modelo GR4J, com o intuito de suprir a carência de dados fluviométricos de bacias localizadas no semiárido da Bahia.

2. OBJETIVOS

2.1. OBJETIVO GERAL

Regionalizar as vazões mínimas para bacias sem dados em regiões semiáridas com o modelo GR4J.

2.2. OBJETIVOS ESPECÍFICOS

- Identificar regiões homogêneas considerando as características físicas e climáticas por meio das técnicas *k-means* e *Ward*.
- Testar a aplicabilidade do modelo GR4J para a obtenção de vazões diárias em bacias hidrográficas localizadas semiárido da Bahia.
- Encontrar relações (equações) entre os valores dos parâmetros do modelo GR4J e as características físicas das bacias (área, características hidrogeológicas e o tipo de solo das bacias), dentro de cada região homogênea identificada na etapa anterior.
- Calcular as vazões de referência para cada bacia hidrográfica a partir dos dados gerados pelo modelo GR4J e regionalizar para as bacias que não possuem dados, através de equações.

3. REVISÃO BIBLIOGRÁFICA

3.1. REGIONALIZAÇÃO HIDROLÓGICA

Um dos grandes desafios atuais na hidrologia é a previsão em bacias sem dados hidrológicos (*Predictions in Ungauged Basins*, PUB). Nesse cenário, os métodos de regionalização são amplamente discutidos, como uma solução para fornecer previsões de escoamento nessas bacias, especialmente durante a década da *International Association of Hydrological Sciences* (IAHS), PUB (2003-2012) e, mais recentemente, na década científica intitulada *Panta Rhei* (2013 -2022) (SIVAPALAN *et al.*, 2003, WAGENER e WHEATER 2006, SEIBERT e BEVEN 2009, MONTANARI *et al.*, 2013, GUPTA *et al.*, 2014 e YANG *et al.*, 2020).

O processo de regionalização consiste na transferência de informações hidrológicas, trazendo dados do comportamento de uma bacia calibrada para uma bacia não calibrada, considerando que ambas apresentam características semelhantes.

A regionalização é importante na hidrologia, sendo muito utilizada para facilitar a extrapolação de dados coletados para outros locais nos quais os valores são obrigatórios, porém, indisponíveis (NATHAN e MCMAHON 1990; CHAVES *et al.*, 2002; SIVAPALAN *et al.* 2003; OUDIN *et al.*, 2010; TUCCI 2017 e GASQUES *et al.*, 2018).

Uma característica fundamental para o método de regionalização é a demarcação de regiões hidrológicamente homogêneas (HALL e MINNS, 1999). Quando muitos locais estão envolvidos em uma análise de frequência regional, a identificação de regiões homogêneas é geralmente a parte mais difícil da análise e requer uma grande quantidade de julgamento subjetivo (FARSADNIA *et al.*, 2014). Kahya *et al.* (2008) considera que a análise de agrupamento de registros de vazão em regiões homogêneas é uma ferramenta importante para a caracterização de sistemas hidrológicos. Devido a sua importância, esse assunto será abordado em um tópico específico mais adiante.

Ainda sobre o processo de regionalização, é importante considerar o objetivo final para escolher a informação a ser regionalizada. Essa informação pode ocorrer na forma de (i) uma variável, interpretada como uma expressão que identifica como se comporta um fenômeno, como a precipitação ou a vazão de um rio; (ii) uma função, a qual possui uma relação entre a variável hidrológica e uma ou mais variáveis explicativas, como a curva de permanência ou a curva de probabilidade de vazões mínimas ou (iii) um parâmetro, identificado como uma

característica do sistema hídrico como o coeficiente de rugosidade e o tempo de concentração da bacia (TUCCI, 2017).

A fim de restringir ao conteúdo do presente estudo, será abordado somente a regionalização da variável vazão e de parâmetros hidrológicos.

Os dados de vazão são essenciais para a realização de diferentes cálculos hidrológicos relacionados à simulação, à previsão e à predeterminação de várias variáveis, os quais permitem obter informações sobre os regimes, os eventos passados e também sobre o comportamento hidrológico da bacia hidrográfica (ANDRADE *et al.*, 2021; LEBECHEREL *et al.*, 2014; PEREIRA; BARBIEIRO; QUEVEDO, 2020; TUCCI, 2017).

A representatividade das informações de vazão, qualitativa e quantitativamente, é fundamental para limitar as incertezas dos cálculos hidrológicos, relacionadas à amostragem desta variável no tempo e no espaço (BLÖSCHL *et al.*, 2013; LEBECHEREL *et al.*, 2014; TUCCI, 2017). Sempre que possível, deve-se priorizar o uso de dados monitorados.

Entretanto, nem sempre é possível encontrar séries históricas de vazões, evidenciando a necessidade da busca de alternativas para esse déficit, tais como a regionalização. Baseando-se na regionalização é possível obter a vazão desejada num local onde não existem dados ou ainda, que tenha séries com um curto período (TUCCI, 2017).

Segundo Razavi *et al.* (2013) e He *et al.* (2011), a regionalização de vazões pode ser realizada por meio de modelos de chuva-vazão ou métodos independentes de modelos hidrológicos. No primeiro caso, os parâmetros do modelo são usados como instrumentos para transferir informações hidrológicas de bacias instrumentalizadas para não instrumentalizadas, enquanto o último caso transfere a vazão diretamente por meio de métodos estatísticos.

Na verdade, o modelo empregado para estimar a vazão nas bacias sem dados não utiliza informações de vazão daquela bacia, entretanto, faz uso de observações locais do clima e de características que podem ser usadas na bacia de interesse (BLÖSCHL *et al.*, 2013). De acordo com Wagener *et al.*, 2004, a aplicação da regionalização em bacias não calibradas é uma forma de gerar relações estatísticas entre os parâmetros do modelo e as características da bacia.

Yadav *et al.* (2007) realizaram um estudo apresentando uma abordagem independente de modelo, para previsão de vazão em bacias sem dados, tendo por base evidências empíricas das relações entre estrutura da bacia, clima e comportamento da resposta da bacia. Em vez de estimar diretamente os valores dos parâmetros do modelo, os autores estimaram e regionalizaram diferentes comportamentos de resposta hidrológica da bacia hidrográfica, quantificados por meio de índices de vazão independentes do modelo.

Sousa (2009) desenvolveu o Sistema Computacional para Regionalização de Vazões (SisCoRV), o qual permite a regionalização das vazões mínimas, máximas e médias e das curvas de regularização e de permanência. O software foi aplicado na bacia do rio Paracatu, sub-bacia do São Francisco, situando-se no Médio São Francisco nos Estados de Minas Gerais, Goiás e Distrito Federal. O SisCoRV apresentou interação significativa com a Internet, facilidade de delimitação das regiões homogêneas, eficiência na aplicação da metodologia de regionalização e capacidade para alimentação da base hidrográfica otocodificada da Agência Nacional de Águas e Saneamento Básico (ANA).

Morais *et al.* (2020) desenvolveram funções regionais para a vazão máxima aplicável a diferentes períodos de retorno, para a vazão média de longo prazo e para a permanência de vazão de 95%, considerando sub-bacias do alto e médio rio Araguaia no Brasil.

Como mencionado, além da regionalização da resposta hidrológica da bacia hidrográfica ou de vazões características é possível transferir diretamente os valores dos parâmetros de um modelo ajustado entre bacias hidrologicamente semelhantes. A determinação dos parâmetros deve ser realizada de forma criteriosa, para não comprometer a confiabilidade dos resultados. Os principais problemas que podem comprometer esse tipo de regionalização são: (i) amostra pouco representativa; (ii) quantidade de estações pluviométricas e fluviométricas insuficientes para representar as variabilidades espaciais da bacia; (iii) interrupções e alterações nas estações de medições; (iv) diferenças na geologia da bacia, que influenciam a vazão subterrânea e na evapotranspiração Riggs (1990 *apud* FEITOZA, 2021), etc.

Saraiva (2010) aplicou o modelo hidrológico SMAP na bacia do Rio Paraopeba, afluente do Rio São Francisco, localizada no estado de Minas Gerais. No processo de regionalização foi utilizado o método de regressão linear múltipla, relacionando os parâmetros com as características geomorfológicas, climáticas, hidrogeológicas e pedológicas da bacia. As equações obtidas foram verificadas quanto a sua adequação utilizando o método *Jack-knife*. Os índices de desempenho encontrados permitiram concluir que o método de regionalização estudado pode ser utilizado para conhecer, de forma aproximada, a disponibilidade hídrica das bacias hidrográficas sem monitoramento da região.

Wang *et al.* (2021) utilizaram dados de escoamento, clima e superfície da bacia do rio Hulan e da bacia do lago Poyang para construir um novo índice de similaridade hidrológica física com o intuito de comparar a eficiência de transferência dos parâmetros e aplicá-los em bacias não calibradas. Os autores chegaram à conclusão que o índice foi mais capaz de quantificar a similaridade hidrológica entre as bacias em termos de clima e superfície

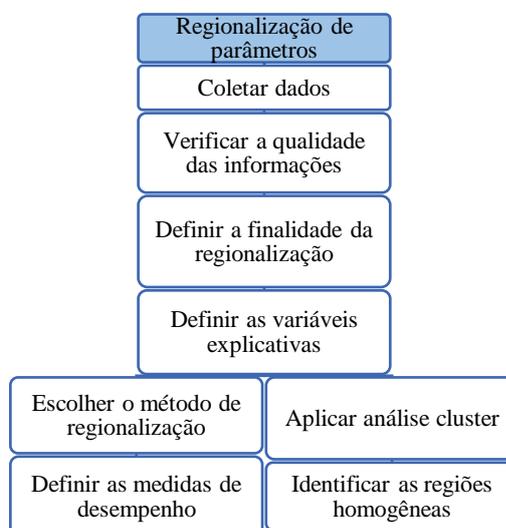
subjacente, sendo também capaz de fornecer uma referência científica para a transferência de parâmetros do modelo hidrológico em uma bacia não medida.

Lane *et al.* (2021) utilizaram a técnica de regionalização de parâmetros multiescala (MPR) para estimar incertezas de parâmetros, considerando 437 bacias hidrográficas na Grã-Bretanha (GB). A partir dessa técnica foram produzidos milhares de possíveis campos de parâmetros de modelo, restritos em uma estrutura de incerteza.

Cheng *et al.* (2021) realizaram a regionalização de parâmetros do modelo HEC-HMS em pequenas bacias hidrográficas sem dados em áreas montanhosas da província de Henan. O modelo foi aplicado a três pequenas bacias hidrográficas em Luanchuan, Gaocheng e Xiahecun e verificado na bacia hidrográfica de Zhongtang. Como resultado, os autores verificaram que o modelo HEC-HMS pode ser aplicado para simular o processo chuva-vazão de pequenas bacias hidrográficas em áreas montanhosas, e o método de regionalização de parâmetros pode efetivamente deduzir os parâmetros do modelo HEC-HMS para microbacias.

Para a regionalização de parâmetros hidrológicos são definidas algumas etapas que norteiam a aplicação da técnica. As principais são: (i) identificação das regiões homogêneas; (ii) definição do método de regionalização a ser utilizado; (iii) obtenção das características fisiográficas e meteorológicas que serão aplicadas no estudo; tais características irão variar com o método a ser empregado; (iv) ajuste de um modelo de regressão para estimativa da vazão em função das características fisiográficas e meteorológicas; e (v) avaliação do desempenho do modelo, validando os valores obtidos em testes (CARVALHO, 2017). Na Figura 1, são apresentados os procedimentos para a realização do processo de regionalização de parâmetros.

Figura 1 - Procedimento a ser seguido no processo de regionalização de parâmetros.



Fonte: Próprio autor (2023).

A qualidade dos dados é fundamental para garantir a confiabilidade dos estudos hidrológicos. Para uma melhor análise regional, Tucci (1993 *apud* Tucci, 2017) afirma que os postos devem ser classificados segundo a qualidade das informações identificadas ao analisar os dados. O processo de análise envolve a triagem preliminar e a análise das informações selecionadas (TUCCI, 2017).

Após a verificação dos dados de vazão é necessário determinar o intuito da regionalização bem como qual a informação a ser regionalizada (Quadro 1). Após este processo pode iniciar a escolha das variáveis explicativas. A partir desses dados deve-se buscar a correlação entre a variável em estudo e indicadores climáticos, de relevo e de geologia que expliquem tal variável.

Quadro 1 - Variáveis explicativas de acordo com o intuito da regionalização e a variável regionalizada.

Intuito da regionalização	Variável regionalizada	Variáveis explicativas
	Vazão diária	Área de drenagem, precipitação média anual, comprimento do rio principal, relevo da bacia, percentagem de área impermeável
Regularização da vazão do rio	Vazão média	Área da bacia e precipitação
Dimensionamento das obras hidráulicas	Vazão média de cheia	Área da bacia, precipitação, declividade e comprimento do rio
Obtenção da vazão mínima referente a setes dias, de duração com tempo de retorno (TR) igual a 10 anos (Q7,10) para estudos da qualidade da água em rios e no processo de outorga	Vazão mínima	Área da bacia e densidade de drenagem
-	Tempo de concentração	Comprimento, declividade e a área da bacia

Fonte: Alterach *et al.* (2006); Sousa (2009); Tucci (2017)

Estas relações mostram melhores resultados na medida que as variáveis independentes escolhidas são significativas e apresentam variabilidade na área de estudo (SILVA JÚNIOR *et al.*, 2003; TUCCI, 2017).

Em estudo realizado por Silva Júnior *et al.* (2003), as variáveis regionalizadas foram a vazão média de longo período, a vazão média de cheia, a curva de permanência e a vazão mínima referente a setes dias, de duração com tempo de retorno (TR) igual a 10 anos. A Q7,10 estipula a vazão que ocorre a maior parte do tempo, em um corpo hídrico, independente se for períodos de seca ou de chuva (CARVALHO, 2017; KOEFENDER, 2015; SILVA *et al.*, 2006b).

No estudo elaborado por Silva Júnior *et al.* (2003), os autores usaram apenas a área de drenagem como variável independente, não descartando a necessidade da análise do ganho estatístico da correlação entre a variável dependente e outras variáveis explicativas, tais como: a precipitação, a densidade de drenagem e a declividade.

Além de características físicas, outras variáveis podem influenciar no aumento ou na diminuição da vazão. De acordo com Tucci (2017), a urbanização produz aumento da vazão máxima em razão da impermeabilização, já a redução da cobertura de floresta e da evapotranspiração propicia o aumento da vazão média. Dessa forma, faz-se necessário que no processo de regionalização, além dos dados hidrológicos, sejam levadas em consideração informações sobre características físicas e climáticas de uma determinada região. Isso porque essas informações interferem na distribuição espacial e temporal das vazões em bacias hidrográficas (GASQUES *et al.*, 2018).

Segundo Alves *et al.* (2014), os dados das formas físicas e dos fenômenos terrestres ou fisiográficos de uma bacia hidrográfica compreende uma série de parâmetros de grande importância para o comportamento hidrológico.

As principais variáveis que explicam as características físicas na regionalização são: (i) área de drenagem; (ii) comprimento do rio principal; (iii) declividade média do rio principal; (iv) densidade de drenagem (TUCCI, 2017).

No processo de regionalização também podem ser utilizadas as variáveis hidrológicas: precipitação média anual; precipitação máxima média de 1 dia de duração; precipitação média do trimestre mais seco; evapotranspiração potencial anual; evaporação potencial do trimestre mais seco; relação entre evaporação potencial e precipitação (TUCCI, 2017).

Em pesquisa realizada por Wagener e Wheater (2006), em 10 bacias hidrográficas localizadas no sudeste da Inglaterra, os autores perceberam que a vazão de base era o descritor dominante que separava os diferentes tipos de resposta das bacias do estudo.

Kuana *et al.* (2020) realizaram a regionalização de vazão com o intuito de avaliar o desempenho das simulações para predição de disponibilidade hídrica em um conjunto de 302 bacias pertencentes aos Estados Unidos. Os autores selecionaram para análise a longitude de localização da estação fluviométrica, a área da bacia, a declividade média, o índice de aridez, a frequência e a duração de eventos de alta precipitação, a média mensal máxima do índice de área foliar, a condutividade hidráulica e a fração de areia e argila no solo.

No processo de regionalização das vazões mínimas para a bacia do Rio São Francisco, a montante do Reservatório de Três Marias, Lemos (2006) utilizou as seguintes características: área de drenagem, o comprimento do rio principal, a densidade de drenagem e a declividade média da bacia e do rio principal, extraídas a partir de modelo digital de elevação hidrologicamente consistente. As características climáticas utilizadas foram a precipitação total anual e as precipitações do semestre e trimestre mais secos. O autor também usou os modelos probabilísticos de distribuição de eventos extremos testados para as séries de vazões mínimas:

Log-Normal a dois e três parâmetros, Pearson III, Log-Pearson III e Weibull. Ressalta-se que a vazões com permanência de 90% na série histórica (Q90) é utilizada como referência para concessão de outorgas em estados como Bahia, Ceará, Paraíba, Sergipe (CARVALHO, 2017).

Dentre os métodos de escolha das variáveis, também é possível a seleção através do cálculo do coeficiente de correlação de Pearson. Segundo Figueiredo Filho e Silva Júnior (2009), o coeficiente de correlação Pearson (r) varia de -1 a 1. O sinal indica direção positiva ou negativa do relacionamento e o valor sugere a força da afinidade entre as variáveis. Uma correlação considerada perfeita (-1 ou 1) indica que o escore de uma variável pode ser determinado exatamente ao se saber o escore da outra. Uma péssima correlação possui valor zero e indica que não há relação linear entre as variáveis. Ou seja, quanto mais perto de 1, independente do sinal, maior é o grau de dependência estatística linear entre as variáveis e quanto mais próximo de zero, menor é a força dessa relação.

Antes de iniciar os procedimentos, é importante realizar uma análise das unidades de medidas das variáveis, tendo em vista que unidades distintas podem influenciar no resultado final. Para minimizar esse problema, existem técnicas como a normalização ou padronização dos dados. Ambas as técnicas têm o intuito de transformar as variáveis na mesma ordem de grandeza, permitindo que as variáveis sejam comparadas igualmente. A diferença básica é que padronizar as variáveis irá resultar em uma média igual a 0 e um desvio padrão igual a 1. Já normalizar tem como objetivo colocar as variáveis dentro do intervalo de 0 e 1, caso tenha resultado negativo -1 e 1. Se estiverem em unidades de medidas distintas, é importante optar-se pela padronização das variáveis antes de iniciar a análise de cluster. Comumente, aplica-se o z-score (FÁVERO; BELFIORE, 2017).

3.1.1. Método de regionalização

Não existe uma metodologia universal de regionalização para uma determinada região ou bacia hidrográfica. Comumente, realiza-se uma testagem de vários métodos para identificar o mais adequado para a região de interesse (SAMUEL *et al.*, 2011).

De acordo com Beck *et al.* (2016), existem seis abordagens de regionalização que vêm sendo frequentemente utilizadas: (i) calibração de bacia por bacia e posteriormente a construção de um modelo de regressão que relaciona os parâmetros do modelo calibrado com as características da bacia, (ii) construção simultânea do modelo de regressão e realização da calibração (iii) transferência do conjunto de parâmetros calibrados para regiões próximas com base apenas na proximidade geográfica, (iv) transferência do conjunto de parâmetros calibrados com base na similaridade climática e/ou fisiográfica, (v) calibração de múltiplas bacias com

características climáticas e/ou fisiográficas semelhantes para obter conjuntos de parâmetros mais generalizáveis, (vi) medidas que quantificam a forma do hidrograma, como inclinação da curva de duração da vazão e índice da vazão de base.

A nível de organização as abordagens foram divididas em 3 grupos: regionalização baseada na proximidade espacial, regionalização baseada na similaridade climática e/ou fisiográficas e regionalização baseada na regressão de dados.

3.1.1.1 Regionalização baseada na proximidade espacial

A abordagem de proximidade espacial consiste na transferência de parâmetros de bacias calibradas para as bacias vizinhas não calibradas, sendo que as bacias próximas umas das outras devem ter comportamento semelhante, uma vez que as condições climáticas e de captação devem variar uniformemente no espaço (MERZ *et al.*, 2006; OUDIN *et al.*, 2008).

Segundo Parajka *et al.* (2013) os métodos de proximidade espacial e similaridade tiveram melhor desempenho em bacias úmidas, enquanto que em bacias áridas, os métodos de similaridade e regressão de parâmetros tiveram desempenho ligeiramente melhor.

Os parâmetros do modelo hidrológico em bacias não calibradas podem ser estimados usando técnicas de interpolação, como krigagem (Vandewiele e Elias, 1995) e distância inversa ponderada (IDW). Este grupo de métodos considera as localizações geográficas entre as bacias não calibradas e calibradas e supõe-se que as bacias sem dados já estejam localizadas em regiões homogêneas e geograficamente semelhantes (SAMUEL *et al.*, 2011).

A desvantagem é que esta abordagem requer uma densa rede de estações hidrológicas (OUDIN *et al.*, 2008; LEBECHEREL *et al.*, 2014). Além disso, Johnston *et al.* (2022) afirmam que, mesmo dentro de regiões geográficas relativamente pequenas, as bacias hidrográficas podem apresentar considerável variabilidade hidrológica, tornando-se outra desvantagem para o referido método.

3.1.1.2 Regionalização baseada na similaridade climáticas e/ou fisiográficas

A regionalização baseada na similaridade física considera que as bacias com características físicas aparentemente semelhantes possuem um comportamento hidrológico semelhante (ACREMAN; SINCLAIR, 1986; OUDIN *et al.*, 2010).

Todavia, não se pode presumir que todas as bacias hidrográficas em de uma região possuam um comportamento hidrológico similar, e por isso é necessária uma abordagem de

regionalização combinando atributos físicos essenciais da bacia (KOKKONEN *et al.*, 2003; REICHL *et al.*, 2009).

No caso de parâmetros de modelos hidrológicos, essas técnicas levam em consideração duas suposições: (i) a similaridade entre os parâmetros ajustados das duas bacias deve corresponder à similaridade de comportamento de ambas na transformação chuva-vazão; (ii) a similaridade física entre ambas gera similaridade hidrológica (OUDIN *et al.*, 2010).

Autores como Kokkonen *et al.* (2003) e McIntyre *et al.* (2005), optaram por transferir conjuntos de parâmetros calibrados com base na consideração explícita de similaridade climática e/ou fisiográfica. Fernandez *et al.* (2000) e Parajka *et al.* (2007) calibraram simultaneamente várias bacias com características climáticas e/ou fisiográficas semelhantes para obter conjuntos de parâmetros mais generalistas.

Parajka *et al.* (2007), destaca que a principal motivação da calibração regional é encontrar parâmetros mais confiáveis para uso em bacias sem dados, sendo aceitável, nesse caso, uma ligeira diminuição do desempenho do modelo para bacias calibradas. Os autores salientam ainda que, para aplicações práticas, o interessante não está na eficiência do modelo mediano em muitas bacias, mas sim, na eficiência do modelo em uma bacia em particular.

Os resultados de um estudo realizado por Wang *et al.* (2013), em 196 bacias na Austrália, mostram que as previsões de escoamento diário das abordagens de subagrupamento com base nas características da bacia têm um desempenho substancialmente melhor do que aquelas de um subagrupamento aleatório.

A dificuldade em aplicar esse tipo de regionalização reside em descobrir quais parâmetros influenciam na resposta hidrológica da bacia, para determinar quais correspondem à similaridade entre elas, para a escolha adequada daqueles que serão regionalizados (OUDIN *et al.*, 2010; BECK *et al.*, 2016).

3.1.1.3 Regionalização baseada na regressão de dados

Os modelos de regressão podem ser do tipo lineares ou não lineares, possuindo o objetivo de equacionar a relação entre a variável dependente (parâmetros do modelo) com diferentes variáveis independentes (características descritivas das bacias) (KUANA, 2022).

Entre as metodologias de regionalização, a regressão de dados é a mais conhecida (OUDIN *et al.*, 2008). Sendo esses métodos os preferidos para uso em regiões de clima temperado como, por exemplo, na maioria dos países europeus (RAZAVI; COULIBARY, 2013).

Para se estabelecer as equações de regressão utilizam-se apenas aquelas bacias para as quais o desempenho do modelo foi julgado aceitável, ou seja, com critério de Nash-Sutcliffe na calibração acima de 0,70, uma vez que as bacias mal modeladas não fornecem informações relevantes suficientes. Uma limitação bem conhecida da abordagem de regressão é que os parâmetros do modelo podem não ser bem definidos produzindo equações de regressão fracas (OUDIN *et al.*, 2008).

Segundo Beck *et al.* (2016), construir simultaneamente um modelo de regressão e realizar a calibração envolve a tarefa não trivial de formular relações parâmetro-preditor a priori parcimoniosas. E calibrar bacia por bacia e posteriormente construir um modelo de regressão, que relaciona os parâmetros do modelo calibrado com as características da bacia, geralmente tem mostrado sucesso limitado. Isso porque ocorre a perda de interação dos parâmetros do modelo e também, problema de equifinalidade (BECK *et al.*, 2016). Para procurar relações não lineares, existe a família de modelos *Generalized additive models for location, scale and shape* (GAMLSS). Esses modelos possibilitam o ajuste de uma ampla família de distribuições para a variável de resposta, permitindo a modelagem direta, a partir da utilização de funções (semi) paramétricas da estrutura de regressão da variável estudada (FLORENCIO, 2010).

Os modelos de regressão podem ser avaliados de acordo com alguns critérios como (i) a significância, a qual é representada pelo P-valor. O P-valor representa a probabilidade que a variável não é relevante para o modelo, podendo ser visualizado a partir da análise de variância (ANOVA). Se o P-valor ($\Pr(>F)$) tem um valor menor que 0,05, a relação entre as duas variáveis é significativa; (ii) a linearidade, a qual indica o grau de associação entre as variáveis testadas e pode ser representada pelo “*R-squared* e *Adjusted R-squared*” do modelo. Quanto maior o coeficiente de determinação (R), maior a representatividade da variável independente para prever o comportamento da variável dependente; (iii) a homocedasticidade dos resíduos, o qual pode ser avaliado graficamente a hipótese da variância constante (homocedasticidade dos resíduos), criando um gráfico dos valores previstos e dos resíduos de cada previsão, os resíduos não devem apresentar nenhum padrão ou tendência; e (iv) normalidade dos resíduos, em que a distribuição dos resíduos deve se aproximar de uma distribuição normal (CHEIN, 2019; PIMENTA; RIBEIRO; CRUZ JÚNIOR, 2021).

3.2. IDENTIFICAÇÃO DAS REGIÕES HOMOGÊNEAS

A determinação das regiões hidrologicamente homogêneas é importante para o processo de regionalização (HALL; MINNS, 1999; KAHYA; DEMIREL; BEG, 2008). Tasker

(1982) e Nathan e McMahon (1990) listam alguns problemas na realização do agrupamento: (i) definição de variáveis adequadas para caracterizar a similaridade intragrupo e a dissimilaridade entre grupos, que pode ser resolvido através da distância euclidiana; (ii) o problema de computação, como solução é sugerido a análise discriminante; (iii) o problema de classificação; (iv) a medida de similaridade altamente dependente das escalas de medida utilizadas. (v) qualquer grupo de variáveis é capaz de gerar clusters e, portanto, é necessário selecionar variáveis de acordo com sua relevância para o problema. A solução padrão para os problemas (ii), (iii) e (iv) é dimensionar todas as variáveis de modo que elas tenham média zero e unidade variação.

Em um estudo realizado em 184 bacias no sudeste da Áustria, Nathan e McMahon (1990) descreveram um procedimento para identificar regiões de desempenho homogêneo semelhante, de forma a estimar vazões mínimas. O estudo é aplicável a qualquer outro objetivo de regionalização.

Da mesma forma, outros autores também realizaram a determinação de regiões homogêneas, são exemplos: Wiltshire (1986) na Grã-Bretanha, Burn e Goel (2000) na Índia, Gibertoni *et al.* (2004), Kuana *et al.* (2020) e Kuana (2022) no estado brasileiro do Paraná, Kahya *et al.* (2008) na Turquia, Cassalho *et al.* (2019) no sul do Brasil, dentre outros.

Há que se considerar que a determinação de regiões homogêneas compõe um problema amplo de classificação de conjuntos de dados, para o qual estão disponíveis uma variedade de ferramentas de informática moderna, como redes neurais artificiais e conjuntos fuzzy. Uma técnica que tem sido muito empregada (Nathan; McMahon, 1990; Hall; Minns, 1999; Burn; Goel, 2000; Kahya *et al.*, 2008, Farsadnia *et al.*, 2014, Cassalho *et al.*, 2019; Kuana *et al.*, 2020; Kuana, 2022) para identificar regiões hidrológicas, tendo por base características hidrológicas ou características de captação é a análise cluster (TASKER, 1982).

A análise de agrupamento (cluster) consiste na exploração e organização de dados para identificar grupos de objetos semelhantes, com base em análise estatística multivariada, associando cada posto a um vetor de dados, que contém as características e/ou estatísticas locais (TASKER, 1982, FOVELL; FOVELL, 1993; BASSO, 2019).

Para detectar os objetos que são semelhantes e agrupá-los em clusters, algumas medidas de (dis-) similaridade ou distância entre pares de objetos devem ser criadas, sendo a medida mais comumente utilizada é a distância euclidiana (ACREMAN; SINCLAIR, 1986 ; FOVELL; FOVELL, 1993).

A distância euclidiana ($D_{i,j}$) entre dois elementos i e j na matriz de dados $n \times p$ é simplesmente a diferença quadrada entre eles para cada uma das p variáveis, somadas sobre as variáveis (Equação (1) (FOVELL; FOVELL, 1993))

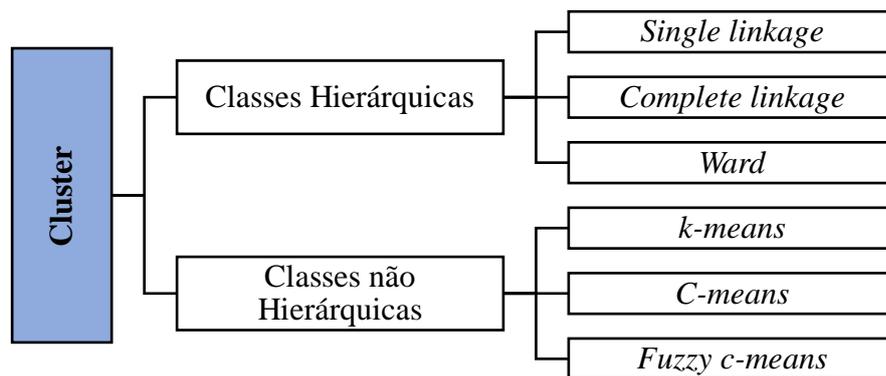
$$D_{i,j} = \left[\sum_{k=1}^p (x_{ki} - x_{kj})^2 \right]^{\frac{1}{2}} \quad (1)$$

Quanto menor $D_{i,j}$, mais semelhantes são as duas bacias i e j . Sugere-se que os “ p ” atributos sejam previamente normalizados, de modo a considerar sua variação de valores (VIVIROLI *et al.*, 2009).

Autores como Santos e Souza (2012) no Nordeste do Brasil, Kuana *et al.* (2020) na área dos Estados Unidos e Kuana (2022) no estado brasileiro do Paraná, aplicaram a distância euclidiana aliada ao método de clusterização para determinação de regiões homogêneas.

Dentre as diversas classificações das técnicas de agrupamentos, Fovell e Fovell (1993), Carvalho (2007), Han *et al.* (2012) e Tan *et al.* (2013), destacam as classes hierárquicas e não hierárquicas (Figura 2).

Figura 2 – Classificação das técnicas de agrupamentos.



Fonte: Próprio autor (2023).

As técnicas hierárquicas executam sucessivos agrupamentos, diminuindo ou aumentando o número de grupos ao longo da análise, por meio da decomposição de determinado conjunto de dados (HAN *et al.*, 2012). Um método hierárquico pode ser classificado como sendo aglomerativo ou divisivo, de acordo com a formação da decomposição hierárquica (CARVALHO 2007; HAN *et al.*, 2012; TAN *et al.*, 2013).

O método aglomerativo, também chamado de abordagem *bottom-up*, começa com cada objeto formando um grupo separado (HAN *et al.*, 2012). Inicia-se com n grupos unitários formados por cada um dos elementos e, a cada iteração, os dois clusters mais próximos se unem até que todos os indivíduos constituam um único cluster (CARVALHO, 2007). A abordagem divisiva faz o processo contrário da aglomerativa, partindo de um único grupo composto por todos os indivíduos, e a cada iteração vai se dividindo até os n clusters serem formados (CARVALHO, 2007).

A abordagem aglomerativa vem sendo amplamente utilizada devido à facilidade de programar e o método de *Ward* vem sendo bastante utilizado em hidrologia (CARVALHO, 2007).

A técnica de *Ward* (1963), produz grupos que são todos aproximadamente do mesmo tamanho, visando ligar vários elementos em clusters usando uma medida de semelhança ou distância (CANNAROZZO *et al.*, 2009; RIANNA *et al.*, 2011). Em suma, este método tenta minimizar a soma dos quadrados dentro do grupo de um cluster definido como a soma dos quadrados das distâncias entre todos os objetos do cluster e seu centroide (CANNAROZZO *et al.*, 2009; YANG *et al.*, 2010).

De acordo com Murtagh e Legendre (2014), *Ward* é o único entre os métodos de agrupamento aglomerativo que se baseia em um critério clássico de soma de quadrados, produzindo grupos que minimizam a dispersão dentro do grupo a cada fusão binária. Além disso, o método de *Ward* é interessante porque faz uso de distâncias euclidianas (KAUFMAN; ROUSSEUW, 1990; MURTAGH; LEGENDRE, 2014).

Além da medida de distância, é importante a determinação do método de encadeamento. Dentre eles, existe: *Nearest neighbor (single linkage)*, o qual privilegia menores distâncias, recomendável em casos de observações distintas; *Furthest neighbor (complete linkage)*, o qual privilegia maiores distâncias e é recomendável em casos de observações parecidas; *Between groups (average linkage)*, o qual faz a junção de grupos pela distância média entre todos os pares de observações do grupo em análise (FÁVERO; BELFIORE, 2017).

Kahya *et al.* (2008) realizaram o agrupamento usando o método de variância mínima de *Ward* para regionalização de vazão em 80 estações fluviométricas na Turquia. A fim de auxiliar na proposta de um modelo de regionalização, Ganora *et al.* (2009) aplicaram a técnica *Ward* em 95 bacias localizadas no noroeste da Itália e na Suíça para determinar as regiões homogêneas.

Melo Júnior *et al.* (2006) utilizaram o método de *Ward* para determinar regiões homogêneas quanto à distribuição de frequência de chuvas no leste do Estado de Minas Gerais no

Brasil. Santos e Souza (2012) determinaram as regiões pluviométricamente homogêneas no Nordeste do Brasil.

Para identificar regiões homogêneas, o método *Ward* também foi aplicado por Yang *et al.* (2010) na bacia do rio Pérola, na China, por Álvarez *et al.* (2011) em 17 bacias na região da Mixteca no México, Farsadnia *et al.* (2014) no norte do Irã, Pessoa (2015) na Amazônia Legal, Cupak *et al.* (2017) em 15 bacias hidrográficas localizadas no alto Vístula na Europa e Gonçalves *et al.* (2017) na bacia hidrográfica do rio Tapajós.

As técnicas não hierárquicas, também conhecidas como métodos de partição, supõem que o número de grupos é conhecido. Nesse caso, os elementos são agrupados conforme a similaridade mais próxima de maneira iterativa (CARVALHO, 2007; TAN *et al.*, 2013).

No método de partição, são escolhidos “k” elementos que constituem o centro de cada grupo. Posteriormente, os elementos são alocados ao cluster em que a distância com relação ao centro seja mínima (CARVALHO, 2007). Na segunda iteração, um novo centro de cada cluster é determinado. A cada iteração os indivíduos são deslocados para o centro mais próximo até que se chegue a uma alocação ideal, geralmente com o menor somatório dos erros entre os pontos de cada grupo e suas médias.

O *k-means*, juntamente com o método de *Ward* é uma técnica muito utilizada, sendo a mais conhecida dentre os métodos de agrupamento não hierárquicos (LIN; CHEN, 2006).

Suponha-se que existam “n” indivíduos e cada um dos elementos presentes precisam ser agrupados em k clusters U_1, \dots, U_k , com o centro do j-ésimo grupo denotado por u_j . Inicialmente, cada centro de cluster é atribuído de forma arbitrária e aleatória. Uma amostra de dados a_i ; pertence a um cluster U_j ; se o centro u_j for o mais próximo entre todos os centros de a_i , onde u_j é a média das amostras em U_j . A cada iteração, o algoritmo *k-means* redistribui os dados da amostra aos seus centros mais próximos e o novo centro do cluster u_1, \dots, u_k é gerado. A iteração termina quando há pouca ou nenhuma mudança no centro do cluster. Em que $1 \leq j \leq k$, $1 \leq i \leq n$, e $I \{c_i = j\}$ é a função índice que é igual a 1 se $c_i = j$ e 0, caso contrário (XING *et al.*, 2017).

Há que se considerar, contudo, que o *k-means* apresenta alguns problemas quando os clusters são de diferentes: tamanhos, densidade, formas não globulares e ainda, quando os dados contêm valores atípicos. Uma solução possível é remover os *outliers* antes do agrupamento (TAN *et al.*, 2013). Para a definição da quantidade necessária de clusters é utilizada algumas técnicas, dentre elas, existe o método *Elbow*, também conhecido como Cotovelo. O *Elbow* é um método visual e seu *output* é um gráfico, cujo eixo das abcissas representa o número de clusters

“ k ” e o eixo das ordenadas representa o valor da soma dos quadrados intraclusters ou wcss (*Within clusters sum of squares*) (CARVALHO, 2020).

Dikbas *et al.* (2013) utilizou o método de agrupamento *k-means* para classificar as vazões máximas anuais e identificar os grupos hidrológicamente homogêneos, considerando 117 estações operadas pela Direção Geral de Pesquisa e Desenvolvimento de Recursos de Energia Elétrica em toda a Turquia. Segundo os autores, os resultados mostram que as regiões definidas podem ser utilizadas para a análise regional da frequência de cheias, recomendando o referido método para identificar as regiões hidrológicamente homogêneas para análise de frequência regional.

Awan *et al.* (2015) empregou o *k-means* e métodos de agrupamento hierárquico para estabelecer zonas de chuva homogêneas na região de monções do leste da Ásia. A comparação de *k-means* e agrupamento hierárquico mostrou que embora ambos os métodos tenham sido capazes de definir bem as zonas homogêneas de chuva, o agrupamento *k-means* superou o agrupamento hierárquico na identificação de zonas mais distintas, que possuíam características pluviométricas diversas.

Kuana *et al.* (2020) aplicou o *k-means* em 302 bacias dos Estados Unidos e Kuana (2022) no estado brasileiro do Paraná. Ambos os estudos almejavam determinar as regiões homogêneas, para auxiliar no processo de regionalização de vazão.

É importante analisar os agrupamentos, verificando quais das variáveis selecionadas contribuíram de fato para a dissimilaridade entre grupo e homogeneidade dentro do grupo. Para isso, existe a possibilidade de aplicar o Teste F para a análise de variância. Também é possível analisar quais variáveis mais contribuíram para a formação de pelo menos um dos grupos, observando-se os maiores valores da estatística F, em conjunto com sua significância (FÁVERO; BELFIORE, 2017).

3.3. MODELOS DE CHUVA-VAZÃO EM REGIÕES SEMIÁRIDAS

Os modelos hidrológicos são utilizados para compreender melhor os fenômenos que ocorrem no ciclo hidrológico na totalidade, preencher falhas existentes nos dados, calcular e prever inundações, enchentes e secas, realizar previsões sobre determinado cenário e ainda efetuar estudos sobre o impacto das modificações do uso do solo (YADAV *et al.*, 2007, LOU 2010; BECK *et al.*, 2017).

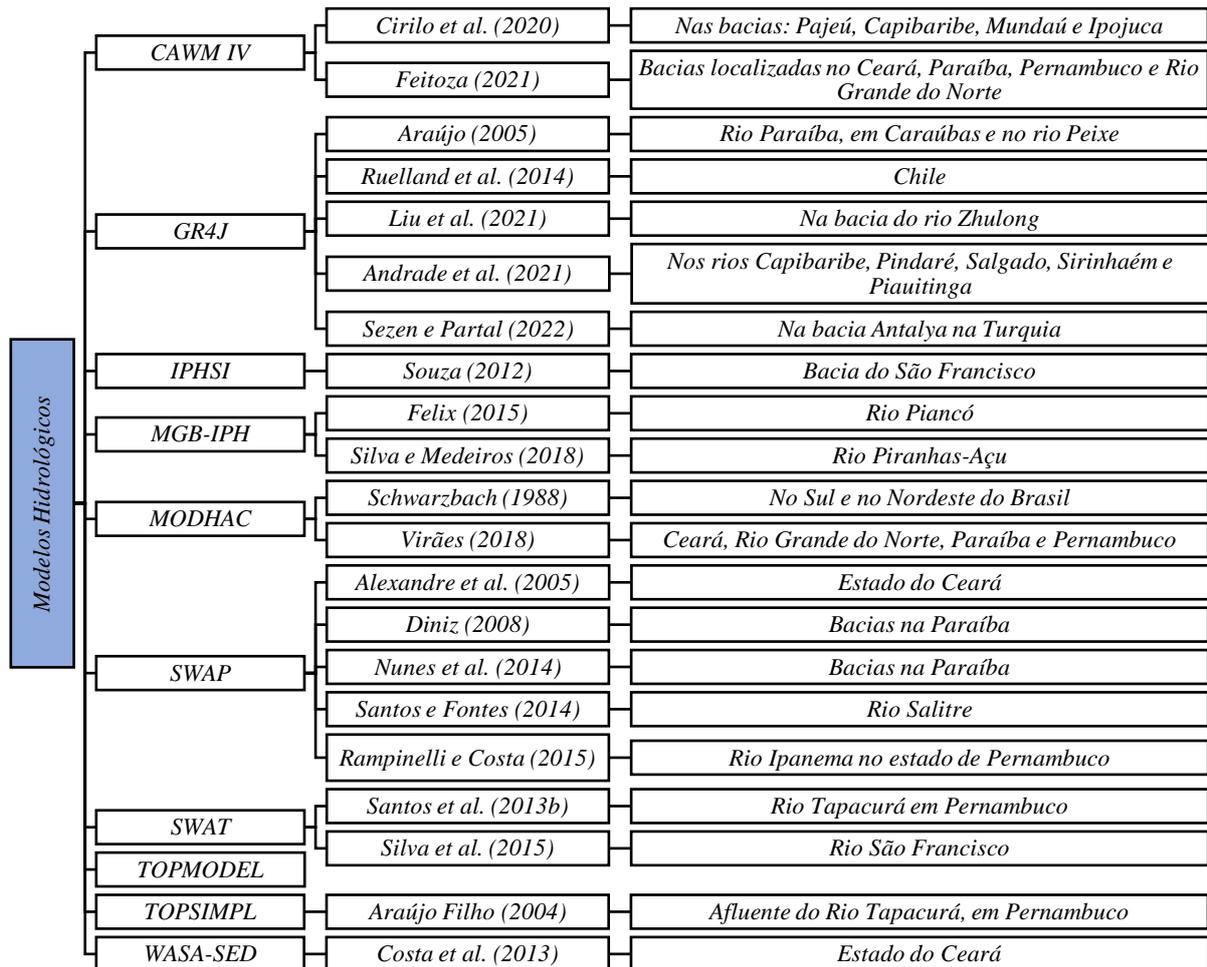
Os modelos desenvolvidos procuram responder às questões emergentes de cada época,

com destaque aos processos hidrológicos mais representativos da realidade de cada modelo desses modelos, gerando diferentes configurações (LOU, 2010). Os modelos hidrológicos apresentam particularidades e restrições que nem sempre são adequadas a determinadas características climáticas e/ou fisiográficas de uma determinada região.

Em regiões semiáridas, por exemplo, é frequentemente necessário que o sistema hídrico seja operado e monitorado de forma eficaz, em razão das altas taxas de evaporação, da má distribuição sazonal dos recursos hídricos e baixa pluviometria (CARNEIRO; FARIAS, 2013).

De acordo com Silva Filho *et al.* (2020), a escassez de dados de vazão, especialmente em ambiente semiárido, constitui um entrave para o adequado planejamento dos recursos hídricos. Nesse cenário, muitos autores recorrem à aplicação de metodologias que ajudem a definir o comportamento hidrológico de uma bacia hidrográfica com a utilização de modelos do tipo chuva-vazão (Figura 3). Modelos de chuva-vazão são ferramentas comumente usadas para extrapolar séries históricas de vazões em investigações operacionais e científicas (WAGENER; WHEATER, 2006).

Figura 3 - Modelos do tipo chuva-vazão usados em regiões semiáridas.



Fonte: Próprio autor (2023).

Araújo (2005) realizou a simulação hidrológica via modelo *Génie Rural à 4 Paramètres Journalier* (GR4J) nas bacias hidrográficas do rio Paraíba, em Caraúbas e do rio Peixe, localizadas na região do Cariri e do sertão paraibano, respectivamente. Andrade *et al.* (2021) aplicou o GR4J para a obtenção de vazões sintéticas diárias nas bacias hidrográficas do rio Capibaribe (em Pernambuco), rio Pindaré (no Maranhão), rio Salgado (no Ceará), rio Sirinhaém (em Pernambuco) e no rio Piauítinga (em Sergipe). Outros autores também utilizaram o GR4J: como Ruelland *et al.* (2014) em uma bacia hidrográfica semiárida no Chile, Liu *et al.* (2021) na bacia do rio Zhulong e Sezen e Partal (2022) em áreas semiáridas e úmidas da bacia Antalya na Turquia também utilizaram o GR4J.

Souza (2012) utilizou o modelo IPHS1 para simular o processo de chuva-vazão na bacia hidrográfica do Rio São Francisco (BHSF), a fim de verificar a resposta hidrológica sobre os reservatórios de Sobradinho e Itaparica no ano de 2007.

Felix (2015) analisou o desempenho do Modelo de Grandes Bacias (MGB-IPH) para a representação de padrões espaços-temporais dos processos hidrológicos na bacia do Rio Piancó. Silva e Medeiros (2018) aplicaram o modelo hidrológico MGB-IPH na bacia hidrográfica do Rio Piranhas-Açu.

Schwarzbach (1988) realizou simulações com o Modelo Hidrológico Autocalibrável (MODHAC) em bacias de médio porte da região Sul, de clima temperado úmido e em bacias do semiárido cristalino do Nordeste do Brasil. Os resultados obtidos para as bacias do semiárido cristalino sugeriram a realização de duas modificações no modelo original: (i) no cálculo da evapotranspiração e, (ii) nos parâmetros que definem o armazenamento subterrâneo.

Virões (2018) e Virões e Cirilo (2019) aplicaram o MODHAC para obter a regionalização dos seus parâmetros na área do semiárido nordestino brasileiro que envolve os Estados do Ceará, Rio Grande do Norte, Paraíba e Pernambuco, carente de informação de vazão.

Alexandre *et al.* (2005) aplicaram o modelo hidrológico concentrado *Soil Moisture Accounting Procedure* (SMAP) no estado do Ceará no intuito de regionalizar parâmetros de vazão. Diniz (2008) e Nunes *et al.* (2014) aplicaram o SMAP em bacias localizadas no estado da Paraíba. Saraiva (2010) aplicou o modelo na bacia do Rio Paraopeba, afluente do Rio São Francisco, localizada no estado de Minas Gerais. Santos e Fontes (2014) aplicaram o SMAP para simulação chuva-vazão na bacia do Rio Salitre, devido à sua simplicidade e êxito em aplicações em rios de região semiárida. Rampinelli; Costa (2015) calibraram o modelo chuva-vazão SMAP para avaliarem os efeitos nas estimativas das vazões médias mensais de uma estação fluviométrica no Rio Ipanema, no estado de Pernambuco.

Santos *et. al* (2013b) aplicaram o *Soil and Water Assessment Tool* (SWAT) na bacia do rio Tapacurá, estado de Pernambuco; Silva *et al.* (2015) aplicaram o SWAT para estimar a vazão na bacia hidrográfica do submédio Rio São Francisco e Sales (2019) aplicou o SWAT no semiárido com o intuito de realizar a modelagem hidrológica e avaliar a influência das atividades da agricultura irrigada na quantidade e na qualidade da água.

Araújo Filho (2004) utilizou o *TOPmodel SIMPLifié* (TOPSIMPL) na bacia do riacho Gameleira, afluente do rio Tapacurá, em Pernambuco. O modelo é uma variante do TOPMODEL, que utiliza apenas quatro parâmetros.

Costa *et al.* (2013) regionalizaram o modelo hidrológico WASA-SED para a estimativa da disponibilidade hídrica de longo-prazo em regiões sem dados, as quais eram de interesse para o planejamento dos recursos hídricos do Estado do Ceará.

3.4. MEDIDAS DE DESEMPENHO

Para avaliar o desempenho dos métodos preditivos, as vazões simuladas em bacias não calibradas precisam ser comparadas com as vazões observadas. Comumente, essa avaliação é realizada por meio de medidas de desempenho. Entender essas medidas de forma generalizada é, pois, um passo importante para reduzir a incerteza das previsões do modelo (BLÖSCHL *et al.*, 2013).

Blöschl *et al.* (2013) apresentam três grupos de medidas de desempenhos: (i) as medidas de viés, considerado um aspecto importante do desempenho do modelo, pois descreve o erro de balanço de massa do escoamento, podendo ser positivo ou negativo, considerando-se o valor zero a previsão perfeita em relação ao viés; (ii) medidas de erros aleatórios as quais são indicadores de dispersão das diferenças de vazões previstas e observadas, com valor zero indicando previsões perfeitas em relação aos erros aleatórios; e (iii) os coeficientes de correlação que denotam a força da associação entre a vazão prevista e observada, onde uma correlação de 1 implica uma associação linear perfeita do padrão observado e previsto, embora a média e a variabilidade possam ser bastante diferentes (Quadro 2).

Quadro 2 - Medidas de desempenho.

Símbolo	Nome	Equação	Significado	Valor para um desempenho perfeito	Como se relaciona com outras medidas
R ²	Coefficiente de determinação	$= 1 - \frac{\sum(Q_i - \hat{Q}_i)^2}{\sum(Q_i - \bar{Q}_i)^2}$	1 para previsão perfeita, 0 se a previsão não for melhor que a média dos dados observados, o valor negativo é ainda pior.	1	Medidas compostas de viés e erros aleatórios.
RMSNE	Raiz média quadrado normalizado	$= \sqrt{\frac{1}{n} \sum \left(\frac{\hat{Q}_i - Q_i}{Q_i} \right)^2}$	0 para previsão perfeita. Maior, para previsões mais pobres.	0	Medida composta de viés e erro aleatório

Símbolo	Nome	Equação	Significado	Valor para um desempenho perfeito	Como se relaciona com outras medidas
RRMSE	Raiz relativa ao erro quadrado médio	$= \frac{\sqrt{\frac{1}{n} \sum (\hat{Q}_i - Q_i)^2}}{\bar{Q}}$	0 para previsão perfeita. Maior, para previsões mais pobres.	0	Medida composta de viés e erro aleatório
ANE	Erro absoluto normalizado	$= \left \frac{\hat{Q}_i - Q_i}{Q_i} \right $	Diferença absoluta entre previsão e observação. 0 para uma observação perfeita.	0	
NSE	Eficiência do modelo Nash Sutcliffe	$= 1 - \frac{\sum (\hat{Q}_t - Q_t)^2}{\sum (Q_t - \bar{Q}_t)^2}$	1 para previsão perfeita, 0 se a previsão não for melhor que a média dos dados observados, negativo para previsões mais pobres.	1	Medida composta de viés e erros aleatórios.
LOGQ	Logaritmo dos erros quadrados (m ³ /s) ²	$= \sum (\log Q - \log Qt)^2$	Varia entre 0 e + ∞ Sendo que 0 são para previsões perfeitas	0	
REMQ _{BV}	Raiz do Erro Médio Quadrático (REMQ) modificado para baixas vazões	$= \sqrt{\frac{1}{b} \sum_{i=1}^n (Q - Qt)^2 * \left(\frac{Q_{max} - Q}{Q_{max}}\right)^2}$	Varia entre 0 e + ∞ Sendo que 0 são para previsões perfeitas	0	
Nslog	Coefficiente de Nash-Sutcliffe para o	$= 1 - \frac{\sum (\log Qt - \log Q)^2}{\sum (\log Q - \log \bar{Q})^2}$	Varia entre -∞ a 1 Sendo que 1 são para	1	

Símbolo	Nome	Equação	Significado	Valor para um desempenho perfeito	Como se relaciona com outras medidas
	logaritmo das vazões		previsões perfeitas		

\hat{Q}_i : vazão estimada no local i ; \hat{Q}_t : vazão estimada no tempo t ; Q : vazão observada; \bar{Q} : vazão média observada no tempo ou no espaço; n : número de dias, $Q_{\text{máx}}$: vazão máxima observada; b : número de informações.

Fonte: Adaptado de Blöschl *et al.*, (2013); Fenicia *et al.*, (2007); Krause; Boyle; Bäse, (2005); Parajka; Blöschl; Merz, (2007); Romanowicz; Osuch; Grabowiecka, (2013).

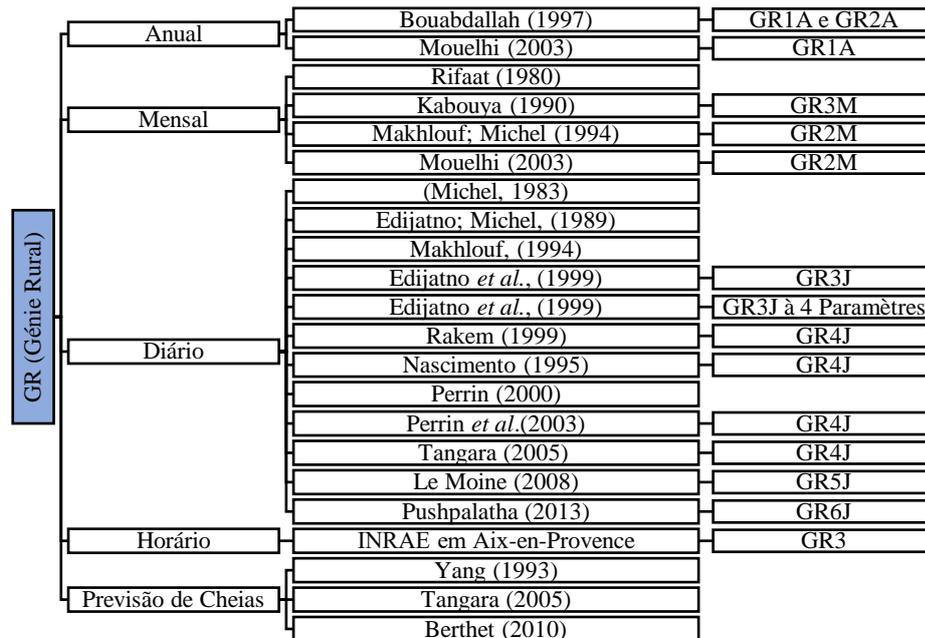
3.5. MODELO GR4J

3.5.1. Contexto Histórico

Os modelos hidrológicos da família *Génie Rural* (GR) foram inicialmente elaborados a partir de pesquisas promovidas pelo *Institut national de recherche en sciences et technologies pour l'environnement et l'agriculture* (IRSTEA) do governo Francês e que, recentemente, se fundiu ao *Institut national de la recherche agronomique* (INRA) formando o *Institut national de recherche pour l'agriculture, l'alimentation et l'environnement* (INRAE) (RAMPINELLI, 2020).

Os primeiros indícios dos modelos de chuva-vazão da família GR foram descritos em 1980 por Samer Rifaat (INRAE, 2018). Iniciou-se trabalhando com modelos no espaço de tempo mensal e, posteriormente, outros autores começaram a estudar com dados anuais, diários, horários e instantâneos (Figura 4).

Figura 4 - Família GR de acordo com o espaço de tempo e seus respectivos autores.



Fonte: Próprio autor (2023).

Tratando-se dos modelos da família GR com entrada de dados diários, uma estrutura simples de dois reservatórios foi proposta e aplicada à bacia do Orgeval por Claude Michel em 1983. O autor partiu de três modelos: CREC à nove parâmetros, criado por Galea (1972 *apud* MICHEL, 1983) e Rifaat (1980 *apud* MICHEL, 1983), CEQUEAU e DEJOREG de Tiercelin e Leviandier (1977 *apud* MICHEL, 1983).

Utilizado por Loumagne (1988 *apud* INRAE, 2018) e Devred (1989 *apud* INRAE, 2018), este modelo foi retomado por Edijatno (1987, 1991 *apud* INRAE, 2018) que, por meio de uma abordagem empírica, adaptou a sua estrutura modificando a função de produção e acrescentando um terceiro parâmetro, uma constante de tempo de um hidrograma unitário (INRAE, 2018). Um novo modelo diário, o *Génie Rural à trois paramètres* (GR3), foi testado por Edijatno e Michel (1989) em uma amostra de 110 bacias hidrográficas francesas. O modelo apresentou resultados satisfatórios, com facilidade de utilização, pequenas incertezas na inicialização e facilidade de regulação, quando comparado a outros modelos. Esse modelo foi adaptado em 1991 por Nilo de Oliveira Nascimento (INRAE, 2018) para aplicação em bacias intermitentes. Já Makhlouf (1994), complementou o modelo de chuva-vazão GR3J, buscando estimar seus parâmetros nas bacias do Mosela e da Bretanha.

Nascimento (1995) analisou as razões da falha do *Génie Rural à trois paramètres Journalier* (GR3J) em simular os processos hidrológicos em sete bacias de córregos

intermitentes, e iniciando uma reflexão sobre as origens desse fenômeno. Para isso, introduziu no modelo um quarto parâmetro, passando a ser conhecido como GR4J. Ele usou uma amostra de 120 bacias hidrográficas, das quais 113 são cursos d'água perenes.

Em 1999 foi proposta uma versão do GR3J por Edijatno *et al.* (1999). Esta versão é capaz de trabalhar continuamente em uma ampla gama de bacias. Apesar de sua grande simplicidade, este modelo mostrou bons resultados em mais de 140 bacias francesas. No mesmo ano, Rakem (1999), propôs uma reformulação matemática do modelo no intuito de tornar sua extensão em tempo contínuo.

Perrin (2000) propôs uma nova versão do modelo GR4J a partir da adaptação do GR3J existente. Essa versão é considerada mais satisfatória do que as versões anteriores, principalmente na simulação das vazões mínimas. Posteriormente, Perrin *et al.* (2003) publicaram a formulação matemática de uma nova versão de quatro parâmetros do GR4J, são eles: (i) X1 - a capacidade do reservatório de produção (mm); (ii) X2 - o coeficiente de troca subterrânea (mm/dia); (iii) X3 - a capacidade máxima do reservatório de propagação (mm); e (iv) X4 - tempo de base do hidrograma unitário (dias) (KUANA *et al.*, 2021). O desempenho do modelo foi avaliado em uma grande amostra de bacias hidrográficas, com desempenho considerado entre os melhores, quando comparado com versões anteriores.

Mathevet (2005) realçou o interesse em ligar os melhores modelos resultantes do estudo da transformação chuva-vazão a diferentes passos de tempo. Esta análise realizada com a estrutura de quatro parâmetros permitiu desenvolver uma variante comum para o modelo, também eficaz em intervalos de tempo diário e horário.

3.5.2. GR4J proposto por Perrin *et al.* (2003)

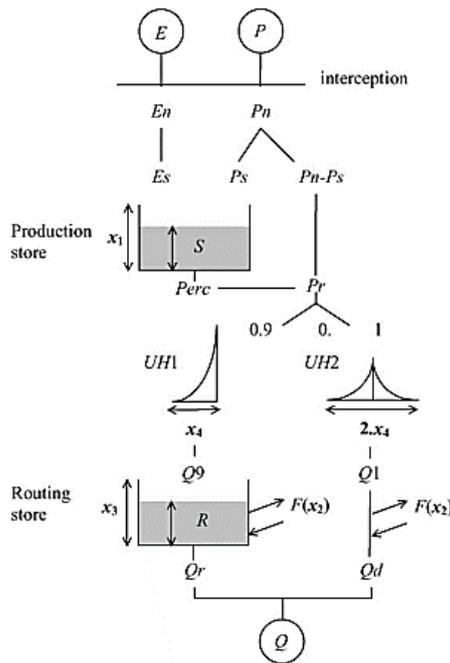
O GR4J proposto por Perrin *et al.* (2003) é um modelo robusto, com estrutura concisa, desenvolvido para obter os melhores resultados em uma grande amostra de bacias hidrográficas (PERRIN *et al.*, 2003) (Figura 5).

Essa versão do GR4J foi uma modificação do modelo GR3J originalmente proposto por Edward Michel (1989), sucessivamente melhorado por Nascimento (1995) e Edijatno *et al.* (1999). As principais equações do modelo, até a obtenção da vazão sintética total (13), estão listadas na Figura 5.

Figura 5 - Esquema do modelo GR4J proposto por Perrin *et al.* (2003).

$$\text{Se } P > E, \text{ então } P_n = P - E \text{ e } E_n = 0 \quad (2)$$

Se $P < E$, então $P_n = 0$ e $E_n = P - E$ (3)



$$P_s = \frac{X_1(1 - (\frac{S}{X_1})^2) \tanh(\frac{P_n}{X_1})}{1 + \frac{S}{X_1} \tanh(\frac{P_n}{X_1})} \quad (4)$$

$$E_s = \frac{S(2 - \frac{S}{X_1}) \tanh(\frac{E_n}{X_1})}{1 + (1 - \frac{S}{X_1}) \tanh(\frac{E_n}{X_1})} \quad (5)$$

$$S = S - E_s + P_s \quad (6)$$

$$Perc = S \{1 - [1 + (\frac{4S}{9X_1})^4]^{-\frac{1}{4}}\} \quad (7)$$

$$S = S - Perc \quad (8)$$

$$P_r = Perc + (P_n - P_s) \quad (9)$$

$$F = X_2 (\frac{R}{X_3})^{\frac{7}{3}} \quad (10)$$

$$Q_r = R \{1 - [1 + (\frac{R}{X_3})^4]^{-\frac{1}{4}}\} \quad (11)$$

$$R = R - Q_r \quad (12)$$

$$Q = Q_r + Q_d, \text{ com } Q_d = \max(0; Q_1 + F) \quad (13)$$

Em que: E_n é a evapotranspiração total; P_n é a chuva total; P_s é a parte da chuva total que vai para o estoque de produção; S é o nível do estoque de produção; E_s é uma parte da evaporação removida do reservatório de produção; $Perc$ é a percolação vindo reservatório de produção; P_r é a quantidade de água que finalmente atinge o próximo reservatório hipotético; F é um termo de troca de água subterrânea (para perda ou ganho); R é o nível de armazenamento; Q_r é a componente de vazão de saída do reservatório de nível R ; Q_d é uma componente de vazão; Q_1 representa uma parcela de saída do hidrograma 2; e Q é a vazão final.

Fonte: Adaptado de Perrin *et al.* (2003); Andrade *et al.* (2021).

Para determinar a chuva total (P_n) subtrai-se P de E segundo a equação (2). Deve-se observar que a operação é feita considerando-se a capacidade de armazenamento por interceptação igual a zero e se a precipitação for maior que a evapotranspiração, o valor de E_n é igual a zero. Nos casos em que a precipitação for menor que a evapotranspiração, P_n é igual a zero e a capacidade de evapotranspiração total (E_n) é calculada conforme a equação (3). Quando a chuva total for diferente de zero, parte dela vai para o reservatório de produção, também conhecida como P_s e deverá ser calculada através da equação (4). P_s é determinada em função do nível S do reservatório de produção e X_1 é a sua capacidade máxima. O reservatório de produção representa o movimento e o armazenamento de água através do perfil do solo, bem como de múltiplas camadas de água subterrânea.

Por outro lado, se a evapotranspiração total não for nula, a taxa de evaporação é determinada como função do nível do reservatório de produção para cálculo da quantidade de água (E_s), que evapora desse reservatório e é representada pela equação (5). O conteúdo total do reservatório de produção será dado pela equação (6).

Outro fator que influencia é a percolação. Ela ocorre quando a capacidade máxima do reservatório alcança $9/4$ de $X1$ e está descrita pela equação (7). A equação (8) representa a quantidade de água que atinge o próximo reservatório hipotético (Pr).

Pr, representada pela equação (9), é então dividida em duas componentes de escoamento, em uma delas 90% de Pr é roteada para o hidrograma unitário 1 (UH1) e os outros 10% para a hidrograma unitário 2 (UH2). Os hidrogramas dependem do parâmetro de tempo ($X4$) e suas ordenadas são usadas para distribuição da chuva efetiva sobre os sucessivos intervalos de tempo. Com os hidrogramas, o modelo pode simular o retardo entre o evento de chuva e o pico de vazão resultante. O termo de troca de água, que atua nas componentes de vazão Q_r e Q_d , é calculado pela equação (10), nesta equação é utilizada o coeficiente de troca subterrânea em milímetros por dia ($X2$) e a capacidade máxima do reservatório de propagação em milímetro ($X3$). A partir do cálculo do nível do reservatório R, é possível calcular a componente de escoamento de saída do reservatório a partir da equação (11). A partir da vazão de saída Q_r depois da entrada de uma parcela de água que saiu do hidrograma (Q_9) é possível calcular o novo nível do reservatório, por meio da equação (12). Por fim, a vazão total é obtida por meio da equação (13).

Segundo Perrin *et al.* (2003), essa versão é mais eficiente quando comparada com a versão anterior, especialmente no que diz respeito à simulação das vazões mínimas. Para sua elaboração, os autores utilizaram uma amostra de ensaio de 429 bacias hidrográficas situadas em condições climáticas diferentes, de semiárido a temperado e úmido tropical.

Perrin *et al.* (2003) afirma que a maioria dos algoritmos de otimização usados para calibrar os parâmetros do modelo requer um conjunto inicial de parâmetros, devido a esse fato, apresentou valores para cada parâmetro em um intervalo de confiança de 80% após testagem em um grupo de 429 bacias (Tabela 1).

Tabela 1 - Valores dos parâmetros do modelo GR4J com intervalo de confiança de 80%.

Parâmetros	Valor médio	Intervalo de 80% de confiança
X1 - capacidade do reservatório de produção (mm)	350	100 a 1200
X2 - coeficiente de troca subterrânea (mm/dia)	0	-5 a 3
X3 - capacidade máxima do reservatório de propagação (mm)	90	20 a 300
X4 - tempo de base do hidrograma unitário (dias)	1,7	1,1 a 2,9

Fonte: Perrin *et al.* (2003)

O limite dos parâmetros sugeridos por Perrin *et al.* (2003) é muito restrito, e o uso de uma gama de parâmetros muito ampla pode requerer um número muito grande de iterações para

encontrar um conjunto de parâmetros aceitáveis, sugerindo outro conjunto de parâmetros (Tabela 2) (ZAMBRANO-BIGIARINI, 2020). Ressalta-se que, o valor de X2 igual a zero indica que não há troca de água na bacia.

Tabela 2 - Limites dos parâmetros do modelo GR4J.

X1	X2	X3	X4
0	-100	0	0
1200	100	5000	5

Fonte: Zambrano-Bigiarini (2020).

Realizando uma busca dos autores que aplicaram o modelo GR4J em regiões semiáridas, pode ser visualizado outros conjuntos de parâmetros (Tabela 3), diferentes daquele originalmente proposto.

Tabela 3 - Parâmetros do modelo GR4J obtidos por autores que estudaram áreas semiáridas.

Autores	Local	*X1	*X2	*X3	*X4
(OUELDKADDOUR <i>et al.</i> , 2021)	Marrocos	169,017	-14,893	74,969	2,035
	Rivadavia (Chile)	11	2,700	473	10
	Varillar (Chile)	1005	-1,5	16	12
(RUELLAND; DEZETTER; HUBLART, 2014)	Riva. & Var (Chile)	517	-2,1	63	3
	Capibaribe em Pernambuco	5,439	0,381	2,090	-0,444
(ANDRADE <i>et al.</i> , 2021)	Pindaré no Maranhão	8,654	1,522	3,569	-0,750
	Salgado no Ceará	4,901	-4,990	6,129	-0,043
	Sirinhaém em Pernambuco	8,711	-0,275	4,051	-0,415
	Piauítinga em Sergipe	7,501	0,941	2,817	0,207
	MÍNIMO		4,901	-14,893	2,090
MÁXIMO		1005	2,700	437	12

* Segundo Perrin *et al.* (2003) todos os parâmetros possuem valores reais. Os parâmetros x 1 e x3 são positivos, x4 deve ser maior que 0,5 e x2 pode ser negativo, positivo ou zero.

Fonte: Próprio autor (2023).

Os parâmetros do modelo GR4J são utilizados para representar as características hidrológicas das bacias. Esses parâmetros possuem relação com as características físicas da bacia, como a área de drenagem, a capacidade de armazenamento de água no solo, a taxa de escoamento superficial, entre outros. A escolha adequada dos parâmetros do modelo é fundamental para garantir uma representação precisa do comportamento hidrológico da bacia.

3.5.3. Calibração do modelo GR4J

Um problema comum nos modelos chuva-vazão é que todos eles exigem algum grau de calibração de parâmetros de modo a obter previsões de vazões confiáveis. Na calibração, os

parâmetros do modelo são ajustados, de forma manual ou automática, até que o valor da vazão simulada corresponda o mais próximo possível da vazão observada (YADAV *et al.*, 2007).

No GR4J, existe a possibilidade de utilizar métodos de calibração como a otimização global Differential Evolution (DE), empregada por Kuana *et al.* (2020) e Kuana (2022), que pode ser aplicada de forma automática, utilizando o algoritmo *Dynamically Dimensioned Search* (DDS) usado por Kuana *et al.* (2021) ou ainda, empregar a técnica *Gauss -Newton* aplicada por Boumenni *et al.* (2017).

Outra alternativa é o uso da otimização global *Particle Swarm Optimization* (PSO), desenvolvida por Eberhart e Kennedy (1995). Segundo Zambrano-Bigiarini (2020) a PSO é uma técnica de otimização estocástica de base populacional usada para explorar um espaço de pesquisa delimitado com um enxame de partículas, de modo a encontrar o melhor conjunto de parâmetros necessários para maximizar ou minimizar uma função objetivo definida pelo usuário.

Aufar *et al.* (2020), obtiveram sucesso ao aplicar a otimização PSO no modelo GR4J nas bacias hidrográficas do Progo e Wuryantoro, ambas localizadas na Indonésia. Os autores utilizaram o NSE e o Raiz do Erro Quadrático Médio (RMSE) para análise de desempenho, observando que o NSE aumentou e o RMSE diminuiu em cada bacia hidrográfica após o uso da otimização PSO.

A otimização PSO também foi utilizada por Zhou *et al.* (2015) para otimização do GR4J e dos modelos AWRA-L em três bacias hidrográficas do sudeste da Austrália e por Althoff *et al.* (2021) em bacias correspondentes a três estações de medição dentro do bioma Cerrado brasileiro.

Além do tipo de otimização, existem outros fatores que precisam ser levados em consideração, como, por exemplo, se a calibração será local ou regional. Segundo Wang *et al.* (2013), a calibração regional permite encontrar um conjunto de valores de parâmetros para representar toda a região hidrologicamente semelhante, enquanto que a calibração local considera cada bacia de forma independente. Na calibração regional, os parâmetros do modelo são otimizados para produzir uma melhor simulação geral para todas as bacias calibradas dentro da região (WANG *et al.*, 2013).

Wang *et al.* (2013) utilizou o GR4J em 196 bacias no sudeste da Austrália. Os autores mostraram que a abordagem clássica de calibração e validação local apresenta melhor desempenho quando comparada à calibração regional. Os resultados também mostram que as previsões de escoamento diário das abordagens de subgrupos, tendo por base as características da bacia, possuem desempenho substancialmente melhor do que a de clusters aleatórios.

3.5.4. Aplicabilidade do Modelo GR4J

O modelo GR4J já foi aplicado na operação de um reservatório dedicado ao aumento das vazões mínimas e ao controle de enchentes (Yang *et al.*, 1991), na previsão de cheias a curto prazo (Yang e Michel 2000), na previsão de secas a longo prazo (PERRIN *et al.*, 2001b *apud* PERRIN *et al.*, 2003), na avaliação do impacto de estimativas aprimoradas de evapotranspiração potencial (PE) (Andréassian *et al.*, 2004), em previsões de vazão (Anshuman *et al.*, 2021) e em investigações das mudanças nos processos hidrológicos em diferentes períodos (Liu *et al.*, 2021).

O modelo também tem sido utilizado para avaliar as mudanças climáticas em bacias hidrográficas. Tian *et al.* (2013), por exemplo, calibraram o modelo GR4J para simular a vazão diária na bacia do rio Jinhua na China. Já Zhou *et al.* (2015) quantificaram os impactos de incêndios florestais e variabilidade climática em três bacias hidrográficas do sudeste da Austrália. Li *et al.* (2013) simulou o escoamento mensal e anual ao longo da bacia do rio Yarlung Tsangpo sob condições climáticas históricas e futuras. Guan *et al.* (2019) avaliaram a capacidade e o desempenho do modelo em simular a vazão sob um ambiente em mudança climática.

Traore (2014) aplicou o GR4J para avaliar a disponibilidade de água na bacia do rio Koulountou, um afluente do rio Gâmbia e Auffer *et al.* (2020) utilizaram o modelo GR4J para decidir a vazão de água para o plantio de arroz nas bacias hidrográficas do Progo e Wuryantoro, na Indonésia.

Autores como Wanzala *et al.* (2022) usaram o modelo GR4J para simular a vazão usando dados de reanálise (ERA5, ERA-Interim, CFSR e JRA55) e Cantoni *et al.* (2022) que usaram o ERA5 para reproduzir escoamento e inundações em 6 bacias hidrográficas na Tunísia, no norte da África.

Grouillet *et al.* (2016) analisaram a sensibilidade de um modelo hidrológico a diferentes métodos para reduzir estatisticamente a precipitação e a temperatura do clima em quatro bacias do Mediterrâneo ocidental, ilustrando diferentes situações hidrometeorológicas. O modelo GR4J foi usado para simular a vazão. Através da validação cruzada, os autores verificaram que o modelo foi capaz de reproduzir corretamente o escoamento em anos secos e úmidos quando as forçantes climáticas observadas de alta resolução foram usadas como entradas (GROUILLET *et al.*, 2016). Segundo os autores, essas simulações podem, portanto, ser usadas

como referência para testar a capacidade de diferentes conjuntos de dados estatisticamente reduzidos de reproduzir vários aspectos do hidrograma.

Lebecherel *et al.* (2014) utilizaram o modelo GR4J para avaliar o impacto da distância da rede de medição de vazão em cálculos hidrológicos em um contexto de bacias sem dados. A regionalização aplicada teve por base a proximidade espacial das bacias hidrográficas. Segundo os autores a avaliação mostrou que para todos os cálculos de vazão considerados, a eficiência do processo de regionalização diminuiu quando a densidade da rede de medição de vazão foi reduzida, ou seja, quando as bacias mais próximas foram removidas. Os resultados também mostraram que a queda na eficiência impactou de forma semelhante todas as variáveis hidrológicas analisadas.

Poissant *et al.* (2017) implementaram as técnicas de proximidade espacial, similaridade física e regressão linear múltipla em 266 bacias hidrográficas no Québec, Canadá. O modelo hidrológico utilizado foi o GR4J acoplado ao modelo de neve CemaNeige. Os resultados mostram que a similaridade física tem um desempenho melhor do que a proximidade espacial e ambos os métodos superam a regressão linear múltipla. Dentre os 12 descritores de captação, o percentual de água e as coordenadas geográficas foram os mais relevantes para esta região.

Arsenault *et al.* (2019) testaram a regressão linear múltipla, proximidade espacial e similaridade física a fim de prever a vazão em bacias não calibradas no México. Para isso, os autores utilizaram os modelos hidrológicos GR4J, HMETS e MOHYSE em 30 bacias hidrográficas. O estudo mostrou que a regionalização em uma área hidrologicamente heterogênea, como o local estudado, apresentou problemas para abordagens de regionalização que dependem de descritores de captação física, como regressão linear múltipla e proximidade espacial. A transferência de conjuntos completos de parâmetros de uma bacia vizinha forneceu o método mais robusto para estimar a vazão em bacias semiáridas e úmidas sem dados e o desempenho do modelo para bacias áridas foi pior no contexto de regionalização, sendo o GR4J mais robusto que os demais modelos, em razão da simplicidade de sua estrutura (ARSENAULT *et al.*, 2019).

Kuana *et al.* (2020) avaliaram a transferência de parâmetros do modelo GR4J em bacias calibradas para bacias não calibradas com características físicas, climáticas e geográficas similares. Para as 181 bacias analisadas, localizadas nos Estados Unidos, o GR4J calibrado obteve um logaritmo do coeficiente de NSE igual ou superior à 0,70 no período da validação.

Golian *et al.* (2021) regionalizaram os modelos hidrológicos GR4J e GR6J para simulação de vazão em bacias sem dados, utilizando 44 bacias irlandesas. Ao aplicarem regressão linear múltipla, regressão não linear e *random forests* (método baseado em *Machine*

Learning com alto custo computacional), descobriram que o método *random forests* superou as técnicas tradicionais para vincular parâmetros de modelo com descritores de bacias físicas. Ao comparar o desempenho com outras abordagens para regionalizar vazões mínimas, médias e máximas (K vizinhos mais próximos, regressão em descritores físicos), os autores concluíram que os modelos regionalizados se comparam favoravelmente em termos de erro relativo.

Wang *et al.* (2021) usaram dados de vazão, clima e a superfície subjacente da Bacia do Rio Hulan e da Bacia do Lago Poyang para construir um novo índice de similaridade hidrológica física (HSI). O índice foi utilizado para comparar a eficiência de transferência dos parâmetros de métodos de regionalização comumente usados e, posteriormente, aplicar parâmetros a bacias não calibradas. Alguns dos resultados apresentados pelos autores foram: (i) a precipitação é o principal fator climático que regula a magnitude do escoamento na Bacia do Lago Poyang; (ii) os modelos GR4J e CemaNeigeGR4J alcançaram simulações razoáveis de escoamento da Bacia do Lago Poyang e da Bacia do Rio Hulan; (iii) houve uma correlação significativa entre o clima, a superfície subjacente e os parâmetros do modelo hidrológico; (iv) existe uma correlação positiva significativa entre o HSI e a eficiência de transferência de parâmetros; (v) o HSI foi mais capaz de quantificar a similaridade hidrológica entre bacias em termos de clima e superfície subjacente e pode fornecer uma referência científica para a transferência de parâmetros de modelos hidrológicos em uma bacia não calibrada (WANG *et al.*, 2021).

Kuana (2022) avaliou a regionalização de parâmetros do modelo GR4J em bacias hidrográficas instrumentalizadas, buscando prever vazões diárias em bacias com pouca ou nenhuma informação hidrológica no estado do Paraná. A autora empregou os métodos de regionalização por proximidade espacial simples, por similaridade fisiográfica-climática e por regressão por *Random Forest*. Os resultados mostraram que todos os métodos de regionalização apresentaram potencial e desempenhos positivos. Entretanto, a regionalização baseada em similaridade fisiográfica-climática demonstrou ser o método mais robusto para a previsão de vazão diária e da vazão de referência Q95, em bacias com poucos dados ou não instrumentalizadas.

3.5.5. Comparação do GR4J com outros modelos

Um mesmo modelo pode resultar em desempenhos diferentes a depender da área a ser aplicada, do período de dados, da função de calibração, dentre outros fatores. Por tais motivos,

alguns autores optaram por testar dois ou mais modelos e decidir qual atende melhor aos objetivos pretendidos.

Andréassian *et al.* (2004) usaram o modelo GR4J e uma versão modificada de oito parâmetros do TOPMODEL para avaliar o impacto de estimativas aprimoradas de evapotranspiração potencial (EP). Para isso, os autores utilizaram uma rede de 42 estações para uma amostra de 62 bacias hidrográficas, buscando testar a sensibilidade dos modelos chuva-vazão às bacias hidrográficas. A EP foi estimada pela equação de Penman Monteith. Em ambos os modelos verificou-se a mesma média de entrada da EP para todas as bacias, produzindo os mesmos resultados que entradas mais precisas obtidas da regionalização. A avaliação detalhada do modelo GR4J calibrado com diferentes cenários de entrada de EP, mostrou que o modelo é claramente sensível à entrada da evapotranspiração, utilizando seus dois parâmetros de produção para se adaptar aos vários cenários da EP.

Li *et al.* (2013) usaram dois modelos de chuva-vazão (SIMHYD e GR4J) para simular o escoamento mensal e anual ao longo da bacia do rio Yarlung Tsangpo (YTR) no Planalto Tibetano (TP), considerando condições climáticas históricas (1962 -2002) e futuras (até aproximadamente 2030). Os dois modelos de chuva-vazão simularam com sucesso o escoamento histórico para as oito bacias hidrográficas nas bacias do YTR.

Traore (2014) aplicou o GR4J e o modelo *Génie Rural à 2 Paramètres Mensuel* (GR2M) para avaliar a disponibilidade de água na bacia do rio Koulountou, um afluente do rio Gâmbia. Os resultados mostraram que ambos os modelos apresentaram bom desempenho, resultando em NSE superior a 0,8.

Benkaci e Dechemi (2018) testaram os modelos de chuva-vazão GR4J, IHACRES, Mordor e *Soil Moisture Accounting and Routing* (SMARG), em escala de tempo diária em quatro bacias situadas no norte da Argélia, caracterizadas por um clima árido e regime hidrológico diferente. O objetivo do estudo foi escolher o melhor modelo de chuva-vazão para uso em simulação e previsão de escoamento. Os resultados da simulação confirmam que o modelo GR4J apresentou os melhores desempenhos, sendo considerado o mais robusto para regiões semiáridas e de regime hidrológico flutuante (BENKACI; DECHEMI, 2018).

Kunnath-Poovakka e Eldho (2019) aplicaram os modelos GR4J, Australian Water Balance Model (AWBM) e o Sacramento para simulação de escoamento superficial no rio Godavari, localizado na Índia. Os autores perceberam que o GR4J resultou em desempenho mais adequado, em termos de previsão e eficiência computacional, em comparação com os outros dois modelos.

Guan *et al.* (2019) utilizaram seis sub-bacias da bacia do Rio Amarelo, localizada na China e situada em uma zona de clima semiárido, a fim de avaliar a capacidade e o desempenho dos modelos hidrológicos *Xin'anjiang* (XAJ), o GR4J, o SIMHYD e o modelo RCCC-WBM (Modelo de Balanço Hídrico desenvolvido pelo Centro de Pesquisa para Mudanças Climáticas) em simular a descarga sob um ambiente em mudança climática. Os resultados mostram que a temperatura anual da área de todas as sub-bacias apresentou uma tendência significativa de aumento, e a precipitação anual apresentou uma tendência de declínio insignificante; e como resultado das mudanças climáticas e atividades humanas intensivas, a série anual de escoamento apresentou uma tendência de declínio com mudanças abruptas ocorrendo principalmente na década de 1980. Com exceção de uma das estações avaliadas, os quatro modelos hidrológicos mostraram bom desempenho para simulação de escoamento para todas as sub-bacias, sob o período natural.

Kuana *et al.* (2021) aplicaram os modelos GR4J e GR6J para a simulação de vazões mínimas e para a estimativa da vazão Q95 em 136 bacias no estado do Paraná, no Brasil. Ambos os modelos mostraram correlações acima de 0,98 entre as vazões Q95 observadas e simuladas para os períodos de validação. Entretanto, o GR4J alcançou valores medianos do coeficiente de Nash-Sutcliffe (NSE) e do logaritmo do coeficiente de Nash-Sutcliffe (LNSE) significativamente mais altos nos períodos de validação, quando comparado ao GR6J.

Anshuman *et al.* (2021) avaliaram e compararam o desempenho do GR4J e AWBM, trabalhando na plataforma de modelagem Source. Tendo como área de estudo duas bacias hidrográficas em Godavari (Maharashtra, Índia). Os autores verificaram que o modelo GR4J fornece melhores previsões de vazão para as bacias de estudo em comparação com o modelo AWBM.

Sezen e Partal (2022), utilizaram modelos conceituais agrupados GR4J, GR6J e modelos baseados em wavelets [Wavelet-Genetic algorithm-Artificial neural network (WGANN)], regressão do vetor de suporte baseado em Wavelet (WSVR)] para modelagem hidrológica em áreas semiáridas e úmidas da bacia Antalya na Turquia. Os autores perceberam que o GR4J e o GR6J, resultaram em desempenhos semelhantes.

Cantoni *et al.* (2022) empregaram os modelos LISFLOOD e o GR4J para reproduzir escoamento e inundações, em 6 bacias hidrográficas na Tunísia, no norte da África, sob um clima semiárido: O desempenho dos dois modelos hidrológicos testados em bacias não calibradas melhorou devido ao processo de calibração, mostrando que ambos os modelos têm capacidade semelhante em determinar se um dado evento de inundação ocorrerá. Os resultados também mostraram que as diferenças dos quantis dos dados observados e com os simulados

com os modelos orientados por ERA5 são muito semelhantes. O ERA5 é dado de reanálise da quinta geração do *European Centre for Medium-Range Weather Forecasts* (ECMWF), disponíveis a partir de 1940 para o clima e condições meteorológicas globais nas últimas 8 décadas (HERSBACH *et al.*, 2023).

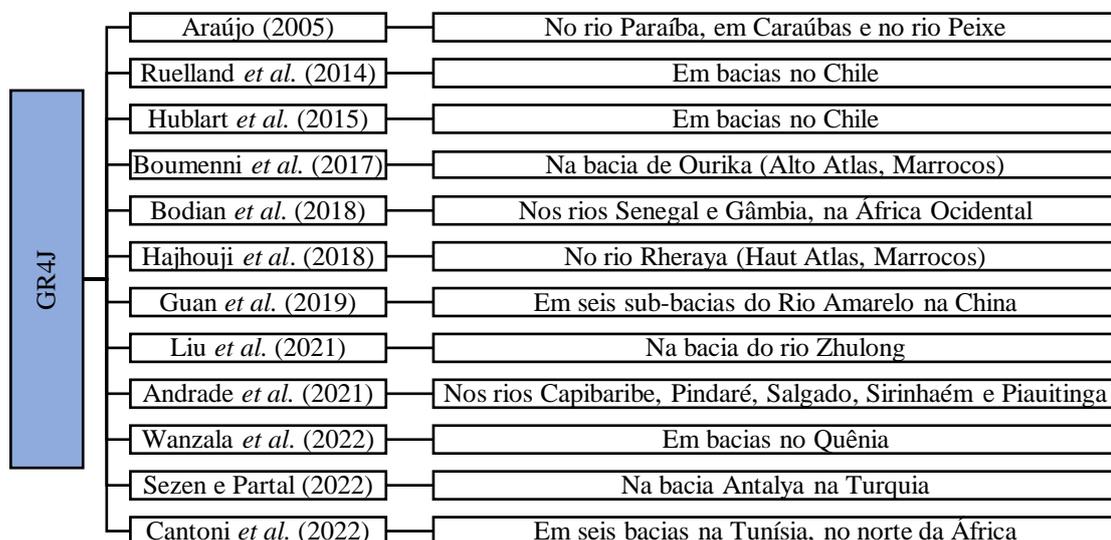
Utilizando a precipitação observada, Cantoni *et al.* (2022) obtiveram um KGE médio de 0,65 e de 0,59 com ERA5 utilizando o GR4J e um KGE médio de 0,59 para o modelo LISFLOOD utilizando a precipitação observada e 0,48 com ERA5, indicando o seu valor como uma alternativa às observações em áreas com escassez de dados, como o semiárido da Tunísia.

Meresa *et al.* (2022) utilizaram os modelos SMART e GR4J para avaliar as mudanças nas vazões médias sazonais e vazões mínimas anuais (Q95) para 37 bacias hidrográficas localizadas na República da Irlanda para as décadas de 2020, 2050 e 2080. Os resultados mostram amplas faixas de mudanças plausíveis na média sazonal e nas vazões mínimas, com aumentos nas vazões médias de inverno e grandes reduções nas vazões médias de verão e nas vazões mínimas anuais (Q95) tanto para SMART quanto para GR4J.

3.5.6. Aplicação do GR4J em regiões semiáridas

As bacias semiáridas caracterizam-se por uma forte variabilidade de precipitação e escoamento e também, pela falta de dados observados de vazão, um grande desafio para o adequado planejamento dos recursos hídricos (SILVA FILHO *et al.*, 2020; CANTONI *et al.*, 2022). Nesse sentido, muitos autores recorrem a modelos do tipo chuva - vazão como o GR4J (Figura 6).

Figura 6 - Modelo GR4J aplicado em regiões semiáridas.



Fonte: Próprio autor (2023).

Araújo (2005) realizou a simulação hidrológica com o modelo GR4J nas bacias hidrográficas do rio Paraíba, em Caraúbas e do rio Peixe, localizadas na região do Cariri e do Sertão Paraibano, respectivamente. Os resultados indicaram que o modelo foi capaz de produzir séries sintéticas de vazão com desempenho acima de 97% para a bacia do rio Paraíba, porém inferior a 70% na bacia do rio Peixe, ambas avaliadas pelo índice estatístico NSE.

Boumenni *et al.* (2017) testaram o desempenho do modelo GR4J na estimativa do escoamento pluvial em bacias de ambiente semiárido no Marrocos. Os autores também destacaram a contribuição de novas técnicas de otimização na determinação dos parâmetros ótimos do modelo. No período de calibração os autores obtiveram um coeficiente Nash variando entre 0,5156 a 0,8492, RMSE entre 0,0122 e 0,0893 e o Bilan variou de 0,6441 até 1,0368. Para o período de validação, o coeficiente Nash obtido variou entre 0,7082 à 0,8869, o RMSE variou entre 0,0413 à 0,3129 e o Bilan variou de 0,8178 à 1,0138, para os diferentes períodos analisados.

Pensando em quantificar a contribuição do derretimento da neve e das chuvas para o escoamento, Hajhouji *et al.* (2018) modelaram o escoamento da bacia do Rio Rheraya (Haut Atlas, Marrocos) levando em consideração sua componente de neve. Para tanto, foi aplicado o modelo conceitual e agrupado GR4J no período 1989-2009, acoplado ao módulo CemaNeige para dinâmica de neve semi-distribuída.

Andrade *et al.* (2021) utilizou o modelo GR4J para a obtenção de vazões simuladas bacias hidrográficas do rio Capibaribe (em Pernambuco), rio Pindaré (no Maranhão), rio Salgado (no Ceará), rio Sirinhaém (em Pernambuco) e no rio Piatinga (em Sergipe). Os resultados indicaram que o GR4J foi capaz de produzir séries sintéticas de vazão com bom desempenho, quando comparado com as vazões observadas, sendo representados pelos índices estatísticos de desempenho NSE, R^2 , RMSE, o viés relativo e o quociente da raiz do erro médio quadrático pela raiz do desvio padrão dos dados observados, amplamente utilizados na literatura para modelos hidrológicos. Na fase de calibração os valores de NSE variaram entre 0,51 e 0,82, o NSE raiz variou entre 0,63 à 0,81, o Pbias variou entre -0,80% à 17,30%, o RMSE variou entre 9,26 à 63,10, RSR o variou entre 0,42 à 0,67 e o R^2 entre 0,51 a 0,82. Na fase de validação os valores de NSE variaram entre 0,42 e 0,78, o NSE raiz variou entre 0,42 à 0,76, o Pbias variou entre -21,40% à 21,30%, o RMSE variou entre 5,08 à 74,29, RSR o variou entre 0,47 à 0,76 e o R^2 entre 0,43 a 0,85.

Liu *et al.* (2021) investigaram as mudanças no processo hidrológico em diferentes períodos e testaram o desempenho do GR4J para aplicação do modelo às bacias hidrográficas

em regiões semiáridas. O modelo GR4J teve um bom desempenho para simulações de escoamento na bacia do rio Zhulong durante o período natural. No entanto, o desempenho do modelo não foi satisfatório em períodos de modificações antrópicas.

Wanzala *et al.* (2022) avaliaram a capacidade de quatro conjuntos de dados de reanálise (ERA5, ERA-Interim, CFSR e JRA55) no Quênia para o período 1981-2016 em escalas de tempo diárias, mensais e sazonais, para simular o escoamento usando o modelo GR4J. Os autores concluíram que a reanálise pode ser uma alternativa viável para observações em bacias não calibradas, mas as incertezas associadas precisam ser tratadas de forma cuidadosa para uma escolha criteriosa de ferramentas de modelagem hidrológica.

Outros autores que também aplicaram o GR4J em regiões semiáridas foram Hublart *et al.* (2015) no Chile, Guan *et al.* (2019) em seis sub-bacias do Rio Amarelo, localizada na China, Sezen e Partal (2022) na bacia Antalya na Turquia e Cantoni *et al.* (2022) em seis bacias hidrográficas na Tunísia, no norte da África.

3.6. BASE DE DADOS DO CAMELS-BR

Segundo Hrachowitz *et al.* (2013) um aumento nos sistemas de observação via satélite deu origem a uma ampla variedade de dados ambientais agora disponíveis de forma fácil e frequentemente gratuita. Esses dados permitem a continuidade das pesquisas, anteriormente prejudicadas devido à ausência de informações.

O *Catchment Attributes and Meteorology for Large-sample Studies - Brazil* (CAMELS-BR) possui séries temporais hidrometeorológicas diárias que incluem vazão observada, precipitação, evapotranspiração real e potencial e temperatura, além de 65 descritores da bacia que abarcam propriedades como topografia, clima, hidrologia, cobertura do solo, geologia, solo e intervenção humana (CHAGAS *et al.*, 2020). O CAMELS-BR segue o mesmo padrão de dados dos Estados Unidos, Chile e Grã-Bretanha (CHAGAS *et al.*, 2020), possibilitando a comparação de estudos.

A base de dados do CAMELS-BR foi utilizada para avaliar como o clima e a gestão da terra aceleram o ciclo da água (CHAGAS *et al.*, 2020), e também, para analisar a viabilidade de relacionar estruturas modelo adequadas às características de clima e vazão em 508 bacias hidrográficas brasileiras (DAVID *et al.*, 2022).

A área de drenagem das bacias da base de dados do CAMELS - BR utiliza os limites das bacias fornecidas por Do *et al.* (2018) (Figura 7), que utilizou o modelo de elevação digital *HydroSHEDS 15 arcsec* de resolução (DEM) para delimitar as bacias hidrográficas,

empregando um procedimento análogo a Lehner (2012) para mais de 3.000 estações no Brasil (CHAGAS *et al.*, 2020).

Figura 7 - Delimitação das bacias hidrográficas presente na base de dados CAMELS-BR.



Fonte: Dados obtidos de Chagas *et al.* (2020).

Para cada estação fluviométrica, Do *et al.* (2018) posicionou a saída no centro de todas as células da grade DEM em um raio de 5 km das coordenadas do medidor indicadas pelos metadados. Eles então selecionaram a célula da grade e os limites de captação associados, levando à área de captação mais próxima da indicada pela ANA (2019a) (CHAGAS *et al.*, 2020). A área das bacias variou de 10,8 km² na região hidrográfica do alto São Francisco a 4,7 milhões de quilômetros quadrados na bacia amazônica, sendo que aproximadamente 30% das bacias analisadas são menores que 1.000 km², 43% estão entre 1.000 e 10.000 km² e 27% são maiores que 10.000 km² (CHAGAS *et al.*, 2020).

Os dados de evapotranspiração são extraídos da base do satélite *Global Land Evaporation Amsterdam Model (GLEAM) v3.3a* (Miralles *et al.*, 2011; Martens *et al.*, 2017 *apud* CHAGAS *et al.*, 2020). A evapotranspiração potencial é calculada a partir de radiação líquida de superfície e dados de temperatura do ar próximo à superfície, utilizando a equação de Priestley & Taylor. Borges Júnior *et al.* (2012) observaram que a equação de Priestley & Taylor apresentou melhor

desempenho em comparação aos métodos FAO-24 Radiação, Hargreaves Samani, Camargo-71 e Blaney-Criddle, para estimativa da evapotranspiração de referência em base diária na microrregião de Garanhuns - PE, por meio de comparações com o método padrão FAO Penman Monteith. Martins (MARTINS, 2019) comparou a evapotranspiração de Priestley & Taylor com a evapotranspiração de Penman Monteith e observou que pode ser uma alternativa eficaz na estimativa da evapotranspiração em regiões áridas e semiáridas.

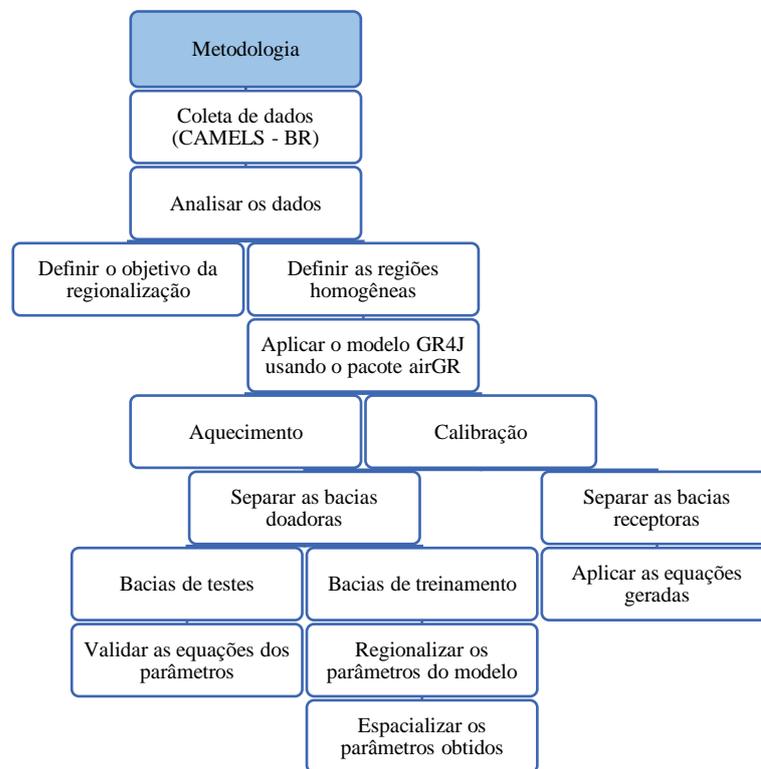
A base de dados do CAMELS-BR também possui um diretório com séries temporais diárias de precipitação extraídas dos satélites CPC NOAA (2019a *apud* CHAGAS *et al.*, 2020), CHIRPS v2.0 FUNK *et al.* (2015 *apud* CHAGAS *et al.*, 2020) e MSWEP v2.2 (Beck *et al.* (2019 *apud* CHAGAS *et al.*, 2020). Os dados indicam as médias de precipitação da bacia em milímetros por dia.

A base de dados CAMELS-BR possui três diretórios de vazão: o primeiro contém arquivos com séries temporais de vazão diária obtidas no site da Agência Nacional de Águas e Saneamento Básico (ANA). As vazões diárias observadas estão em metros cúbicos por segundo. O segundo diretório contém arquivos com séries temporais de vazão diária em milímetros por dia. E o terceiro diretório contém arquivos com séries temporais de vazão diária em metros cúbicos por segundo, simulada usando o modelo MGB-SA de larga escala (Modelo de Grandes Bacias; Siqueira *et al.*, 2018 *apud* CHAGAS *et al.*, 2020).

4. MATERIAIS E MÉTODOS

A metodologia geral (i) identificou regiões onde o comportamento das bacias tenha sido similar utilizando as metodologias das K-médias e o método de *Ward*, a partir da análise das características físicas e climáticas da área de estudo; (ii) aplicou o modelo GR4J para a obtenção de vazões diárias em um conjunto de bacias hidrográficas no semiárido da Bahia de características variadas; (iii) relacionou os valores dos parâmetros do modelo GR4J e características físicas das bacias, como, por exemplo, área, perímetro, características hidrogeológicas e o tipo de solo das bacias, dentro de cada região homogênea verificada na etapa anterior; (iv) calculou as vazões de referências a partir dos dados gerados pelo modelo GR4J (Figura 8).

Figura 8 - Procedimentos para obtenção de vazões estatisticamente consistentes usando o GR4J.



Fonte: Próprio autor (2023).

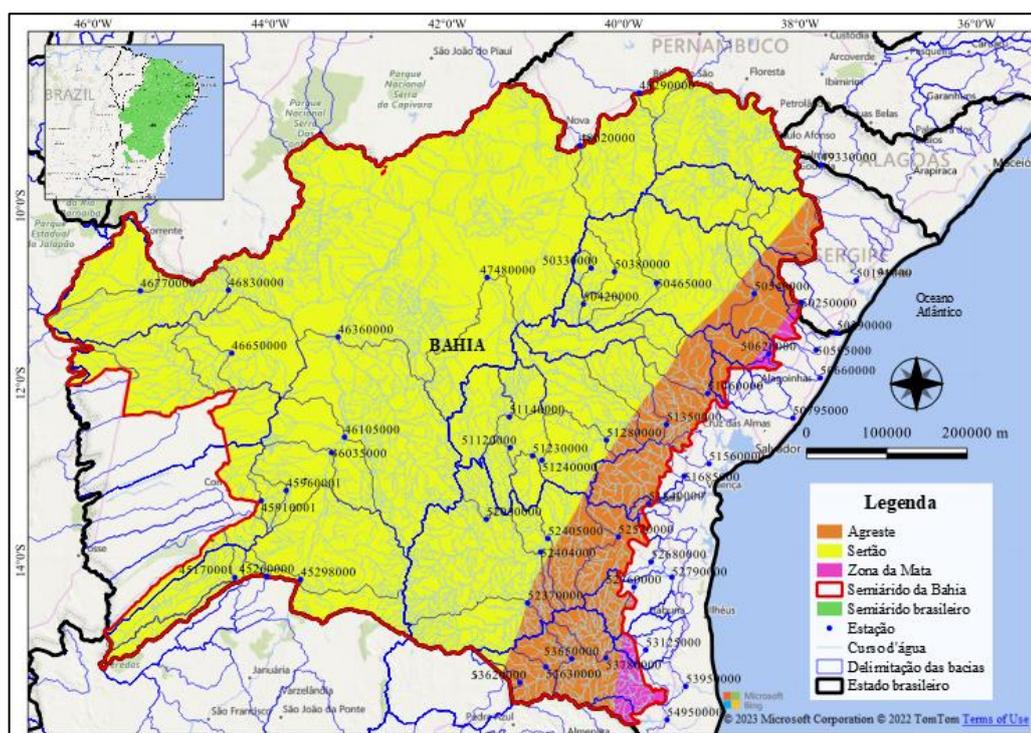
4.1. ÁREA DE ESTUDO

O semiárido da Bahia é definido conforme critérios estabelecidos pelo Conselho Deliberativo (CONDEL) da Superintendência do Desenvolvimento do Nordeste (SUDENE) sendo: (i) precipitação pluviométrica média anual igual ou inferior a 800 mm; (ii) índice de

Aridez de Thornthwaite igual ou inferior a 0,50 e; (iii) percentual diário de déficit hídrico igual ou superior a 60%, considerando todos os dias do ano (BRASIL, 2017).

O semiárido baiano é composto por 278 municípios (BRASIL, 2017), abrangendo uma área de aproximadamente 445.948 km², o que, de acordo com Ribeiro e Oliveira (2019), corresponde a mais de 66% do território do estado da Bahia. A área de estudo está situada em três sub-regiões do Nordeste: Agreste, Sertão e Zona da Mata (Figura 9).

Figura 9 - Localização do semiárido da Bahia, divisão de acordo com as sub-regiões do Nordeste e a delimitação das bacias hidrográficas da base de dados CAMELS-BR.



Fonte: Próprio autor (2023).

O Sertão ocupa a maior parte do semiárido da Bahia. Essa sub-região apresenta como vegetação típica a Caatinga, solos rasos e pedregosos, com chuvas escassas e mal distribuídas, o que dificulta o desenvolvimento de atividades agrícolas (EMBRAPA, 2022). A sub-região apresenta totais pluviométricos em torno 400 a 600 mm por ano e evapotranspiração anual em torno de 2.000 mm por ano, caracterizando o polígono das secas (EMBRAPA, 2021).

O Agreste está situado entre o Sertão e a Zona da Mata. Nessa sub-região, o solo é essencialmente pedregoso, a vegetação rala e pequena (EMBRAPA, 2022). Nas áreas altas, há maior umidade, sendo os terrenos mais férteis ocupados por minifúndios, com predomínio das culturas de subsistência e da pecuária leiteira (EMBRAPA, 2022). Na referida sub-região a precipitação pluviométrica média anual oscila entre 600 a 800 mm (EMBRAPA, 2021).

A Zona da Mata, localizada no sul da Bahia, possui um clima tropical úmido, com chuvas mais frequentes no outono e no inverno. O solo é fértil e a vegetação natural é a Mata Atlântica (EMBRAPA, 2022). Também denominada de Zona Úmida Costeira, é considerada a área mais úmida da região Nordeste, com precipitação pluviométrica média anual que varia de 1.200 a 2.000 mm (EMBRAPA, 2021).

Na Tabela 4 são apresentados os códigos das estações pertencentes ao semiárido da Bahia, com suas respectivas localizações, áreas e vazões, para o período de 1980 a 2018. 6% (3) das estações estão localizadas em bacias com áreas de até 1000 km², 38% (20) se encontram em áreas maiores que 1000 km² e menores que 10000 km² e 56% (29) estão situadas em bacias com áreas superiores a 10000 km². No período analisado, as estações não tiveram as suas respectivas vazões interrompidas completamente, indicando um comportamento perene, porém apresentaram uma redução significativa entre os meses de maio e agosto (outono e inverno), como pode ser visualizado no APÊNDICE A.

Na base de dados do CAMELS-BR, cada bacia apresenta o código da identificação da região hidrográfica, a localização e dois dados referentes a área: um da Agência Nacional de Águas e Saneamento Básico (ANA) e outro do *Global Streamflow Indices and Metadata Archive* (GSIM).

Tabela 4 - Código das estações da ANA com suas respectivas localizações, áreas e vazões para o período de 1980 a 2018, pertencentes ao semiárido da Bahia.

Código	Nome da estação	Região hidrográfica	Latitude	Longitude	Área - ANA (km ²)	Área - GSIM (km ²)	Vazão em m ³ /dia			Registro com dados
							Mínima	Média	Máxima	
45170001	Fazenda Porto Alegre	São Francisco	-14,2642	-44,5225	5850,00	5823,41	0,00	56,60	204,87	14187
45260000	Juvenília	São Francisco	-14,2600	-44,1522	16300,00	16284,56	49,01	127,80	698,75	14187
45298000	Carinhanha	São Francisco	-14,3044	-43,7633	254000,00	256995,20	319,48	1975,05	18332,72	14187
45910001	Santa Maria da Vitória	São Francisco	-13,4006	-44,1989	29600,00	29484,01	90,99	193,10	823,83	14187
45960001	Porto Novo	São Francisco	-13,2914	-43,9089	31000,00	30996,37	95,48	198,19	816,22	14187
46035000	Gameleira	São Francisco	-12,8686	-43,3797	309000,00	312707,10	358,20	2200,01	11571,47	14187
46105000	Paratinga	São Francisco	-12,6967	-43,2264	314000,00	321332,30	0,00	2166,89	12847,72	14187
46360000	Morpará	São Francisco	-11,5583	-43,2825	345000,00	351975,40	296,45	2165,70	15284,00	14187
46650000	Tanguá	São Francisco	-11,7208	-44,5022	36000,00	35821,92	0,00	177,27	540,30	14187
46770000	Fazenda Bom Jardim	São Francisco	-10,9936	-45,5278	6930,00	6926,20	18,52	33,47	62,89	14187
46830000	Ibipetuba	São Francisco	-11,0058	-44,5242	17900,00	18301,11	59,79	95,36	339,57	14187
47480000	Jaguaraci	São Francisco	-10,9017	-41,5703	-	9270,73	0,00	0,69	44,26	11081
48020000	Juazeiro	São Francisco	-9,4064	-40,5036	516000,00	525840,70	922,13	2196,82	10806,07	14187
48290000	Santa Maria da Boa Vista	São Francisco	-8,8097	-39,8239	535000,00	545413,30	554,38	2069,86	10819,03	14095
49330000	Piranhas	São Francisco	-9,6261	-37,7561	610000,00	619185,70	457,975	2061,49	12601,9	14187
50191000	Fazenda Belém	Atlântico Leste	-10,9431	-37,3469	15900,00	15805,63	457,98	2061,49	12601,90	14187
50250000	Fazenda Tourão	Atlântico Leste	-11,1931	-37,9750	2940,00	2928,95	0,00	13,05	3183,58	14187
50290000	Itanhy	Atlântico Leste	-11,5397	-37,5681	4480,00	4462,40	0,00	2,44	159,18	14187
50330000	Pindobaçu	Atlântico Leste	-10,8017	-40,3800	-	665,01	0,00	10,58	565,71	14187
50380000	Ponto Novo	Atlântico Leste	-10,8450	-40,1128	2850,00	2855,01	0,00	2,40	98,17	11143
50420000	Jacobina	Atlântico Leste	-11,2114	-40,4703	1290,00	1304,52	0,00	4,91	147,02	14187
50465000	Queimadas	Atlântico Leste	-10,9733	-39,6328	11800,00	11834,96	0,00	1,76	423,19	14187
50540000	Cipó	Atlântico Leste	-11,0975	-38,5125	28000,00	27982,18	0,00	9,76	837,66	14187
50595000	Usina Altamira	Atlântico Leste	-11,7347	-37,8033	35200,00	35133,49	0,00	13,44	945,39	14187
50620000	Inhambupe	Atlântico Leste	-11,7822	-38,3467	2890,00	2896,08	1,46	21,21	747,59	14187
50660000	Corte Grande	Atlântico Leste	-12,0536	-37,7550	4380,00	4393,69	0,00	2,01	264,78	14187

Código	Nome da estação	Região hidrográfica	Latitude	Longitude	Área - ANA (km ²)	Área - GSIM (km ²)	Vazão em m ³ /dia			Registro com dados
							Mínima	Média	Máxima	
50795000	Tiririca	Atlântico Leste	-12,5119	-38,0667	4700,00	4703,06	0,33	8,82	228,79	14187
51120000	Andaraí	Atlântico Leste	-12,8406	-41,3222	2350,00	2344,33	2,09	23,88	337,57	14187
51140000	Porto	Atlântico Leste	-12,4911	-41,3289	6180,00	6181,84	0,00	15,30	630,94	14187
51230000	Fazenda Iguaçú	Atlântico Leste	-12,9361	-41,0644	2130,00	2102,95	0,79	9,71	751,52	14187
51240000	Itaeté	Atlântico Leste	-12,9861	-40,9581	16500,00	16417,18	1,06	49,81	1184,45	14187
51280000	Iaçú	Atlântico Leste	-12,7622	-40,2136	22800,00	22760,99	0,45	53,63	1848,54	14187
51350000	Argoim	Atlântico Leste	-12,5856	-39,5222	37900,00	37863,95	0,00	13,44	500,95	14187
51460000	Ponte Rio Branco	Atlântico Leste	-12,2331	-39,0461	11600,00	11566,03	1,06	49,81	1184,45	14187
51560000	Nazaré	Atlântico Leste	-13,0339	-39,0314	1350,00	1348,28	0,45	53,63	1848,54	14187
51685000	Jiquiriçá	Atlântico Leste	-13,1736	-39,3183	6520,00	6475,22	0,00	59,69	3287,03	14187
51840000	Tesouras	Atlântico Leste	-13,4872	-39,7294	345,00	355,44	0,00	7,38	6575,00	14187
52050000	Fazenda Canabrava	Atlântico Leste	-13,6528	-41,6061	2260,00	2266,14	0,00	6,59	408,68	14187
52370000	Anagé	Atlântico Leste	-14,6097	-41,1372	7930,00	7934,55	0,40	9,92	296,73	14187
52404000	Areião	Atlântico Leste	-14,0314	-40,9825	29600,00	29678,15	0,00	1,06	19,72	14187
52405000	Lagoa do Tamburi	Atlântico Leste	-13,8778	-40,8964	32500,00	32611,95	0,00	5,05	305,70	11841
52570000	Jequié	Atlântico Leste	-13,8647	-40,0808	40400,00	40498,39	0,00	4,64	1291,13	14187
52680000	Ipiaú	Atlântico Leste	-14,1503	-39,7036	46500,00	46547,88	0,00	19,07	2396,63	14187
52760000	Dário Meira	Atlântico Leste	-14,4367	-39,9036	2730,00	2726,18	0,00	22,19	1800,66	12454
52790000	Pedrinhas	Atlântico Leste	-14,3253	-39,4636	5270,00	5246,77	0,00	24,13	2585,37	14187
53125000	Fazenda Manaus	Atlântico Leste	-15,1531	-39,7692	-	831,63	1,46	38,99	2200,49	14187
53620000	Cândido Sales	Atlântico Leste	-15,5133	-41,2367	13000,00	12923,24	0,78	14,60	657,79	14187
53630000	Inhobim	Atlântico Leste	-15,3397	-40,9328	16400,00	16339,08	0,00	31,35	877,95	14187
53650000	Itambé	Atlântico Leste	-15,2483	-40,6311	18500,00	18414,00	0,00	4,70	641,07	10959
53780000	Itapetinga	Atlântico Leste	-15,2414	-40,2333	2870,00	2831,02	0,00	13,77	1425,77	14187
53950000	Mascote	Atlântico Leste	-15,5686	-39,3044	30700,00	30473,57	0,00	12,02	538,22	14187
54950000	Itapebi	Atlântico Leste	-15,9483	-39,5236	68900,00	68283,32	0,00	15,12	595,26	14187

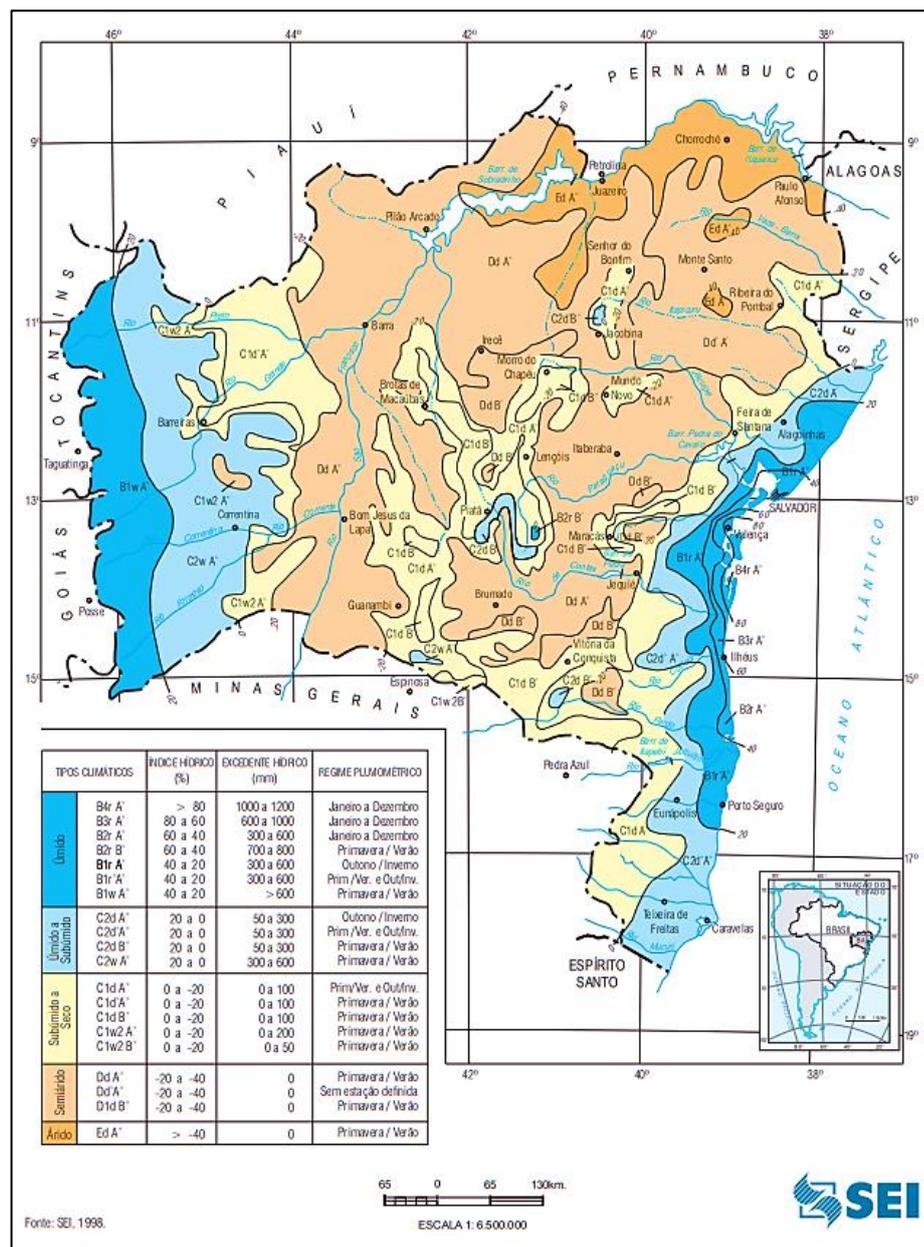
Fonte: Dados obtidos de Chagas *et al.* (2020).

Quinze das bacias (29%) pertencem à região hidrográfica brasileira do São Francisco e as outras 37 (71%) à região do Atlântico Leste.

4.1.1. Características pluviométricas e climáticas

De acordo com Thornthwaite, a tipologia climática da área de estudo varia do úmido até o árido, com predominância do semiárido (Figura 10).

Figura 10 – Tipologia climática segundo Thornthwaite para o estado da Bahia. Pluviometria 1943 - 1983/ Temperatura 1961 - 1990.



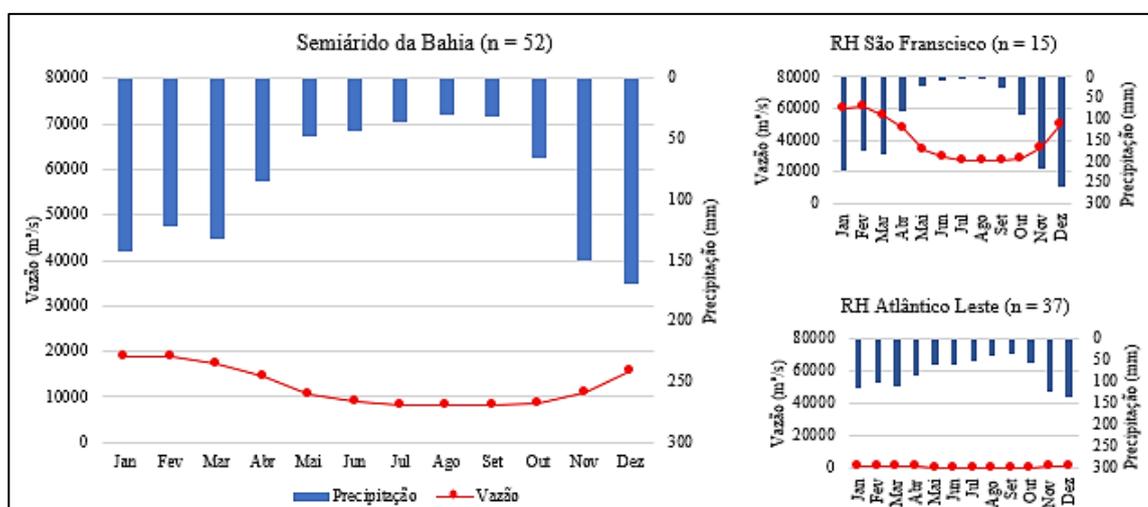
Ao analisar a quantidade de informações das séries históricas disponíveis, observou-se que 97% dos dados estavam disponíveis no CHIRPS v2.0, com falha apenas no ano de 1980, 92% dos dados estavam disponíveis no *Multi-Source Weighted - Ensemble Precipitation* (MSWEP v2.2), com falha no ano de 2017 e 2018 e 100% para os dados Centro de Previsão Climática (CPC). Em razão disso, optou-se pelo uso dos dados pluviométricos do CPC.

A base de dados do CAMELS apresenta uma série histórica da evapotranspiração potencial obtida do GLEAM v3.3^a para o período de 1980 a 2018.

4.1.2. Características fluviométricas

Ao avaliar os dados de vazões diárias do período de 1980 a 2018, observou-se que todas as bacias deste estudo possuem acima de 77% dos dados disponíveis para o período, exceto a estação 50330000 que possui apenas 67%. No APÊNDICE B é possível visualizar as estações e os períodos com dados de vazão disponíveis. Além disso, percebe-se que as vazões mais elevadas pertencem à região hidrográfica do São Francisco quando comparada à região hidrográfica do Atlântico Leste. Os meses de novembro, dezembro e janeiro apresentam vazões mais elevadas, em razão das altas precipitações (Figura 11).

Figura 11 - Resumo das vazões pertencentes a região hidrográfica do São Francisco e Atlântico Leste localizadas no semiárido da Bahia (1980 – 2018).



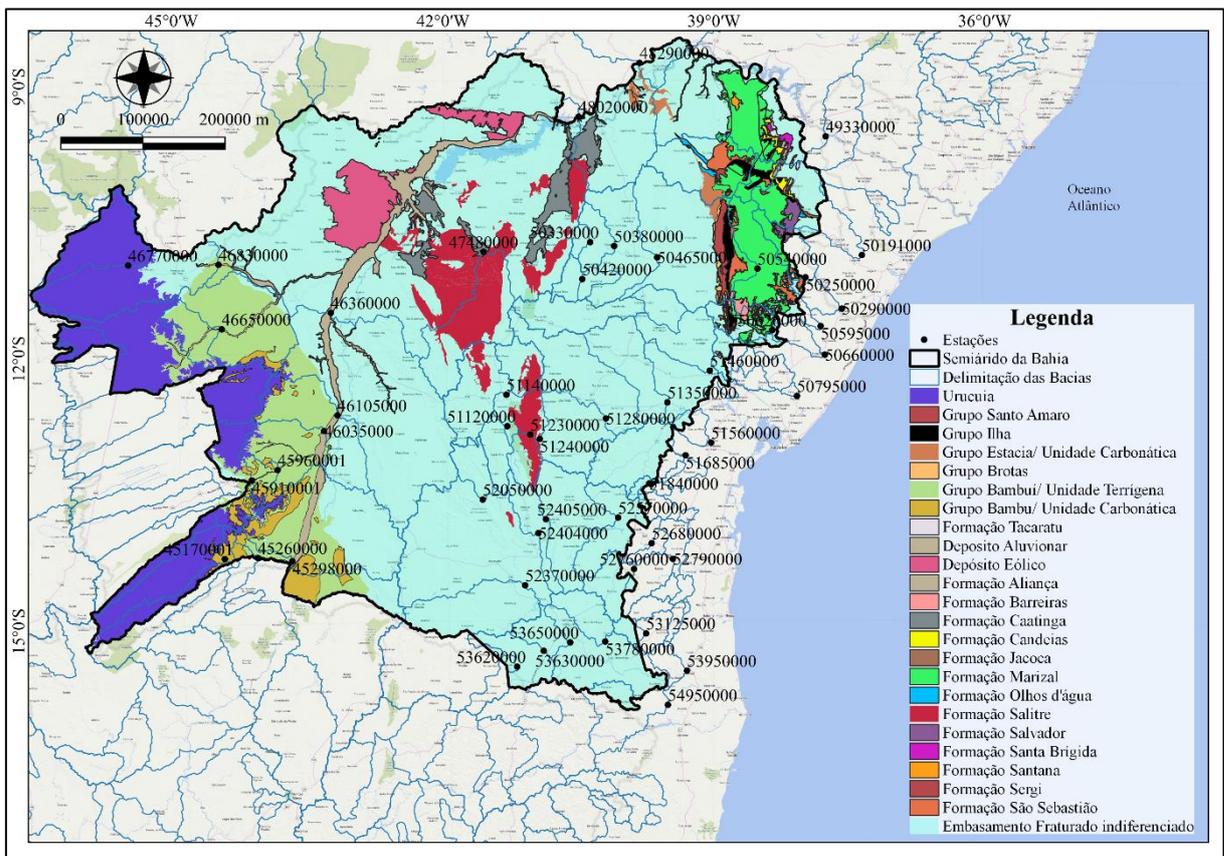
Fonte: Dados obtidos de Chagas *et al.* (2020).

É interessante observar que apesar da redução da precipitação nos meses de maio a setembro na região hidrográfica do São Francisco, as bacias apresentam vazões significativamente altas, podendo ser justificada pela influência do Aquífero Urucuia.

4.1.3. Hidrogeologia, relevo, pedologia, uso e ocupação do solo, índice de intervenção humana e vegetação

O semiárido da Bahia apresenta 24 formações hidrogeológicas diferentes. O Embasamento Fraturado Indiferenciado, Grupo Urucua, Grupo Bambuí Unidade Terrígena, Formação Salitre, Formação Marizal e Depósito Eólico ocupam áreas significativas (Figura 12).

Figura 12 - Formação hidrogeológica presente no semiárido da Bahia.

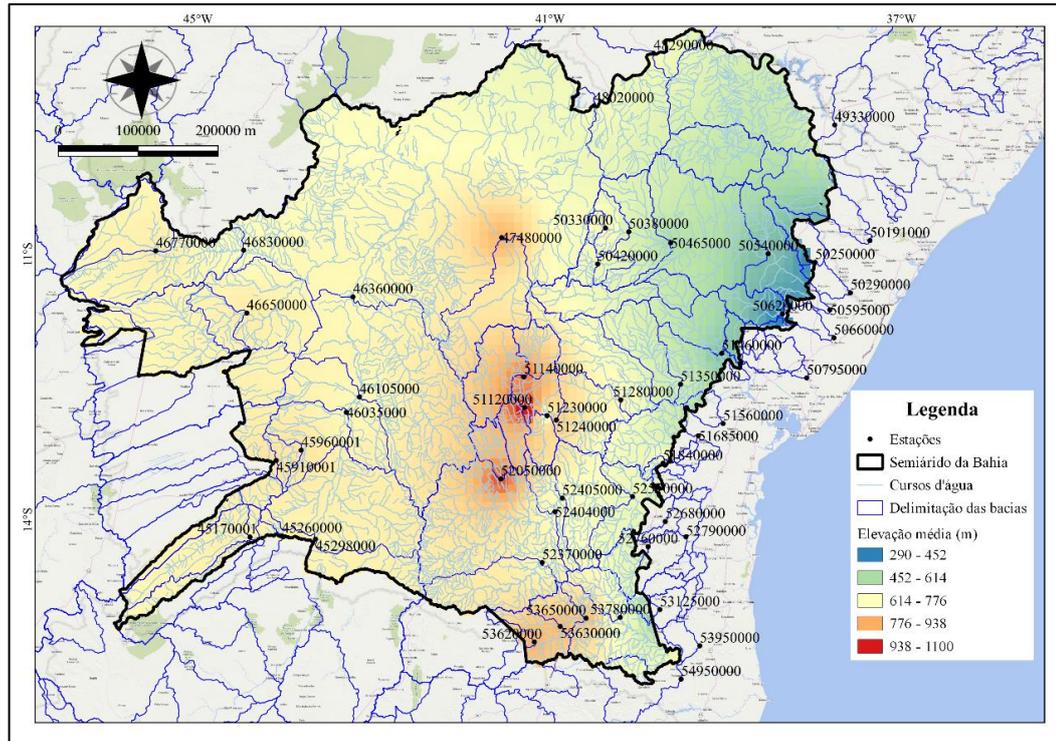


Fonte de dados: Mapa hidrogeológico do Brasil (SGB - CPRM, 2007), Curso d'água (IBGE, 2017), semiárido (SUDENE, 2017), Bacias Hidrográficas (Do et al 2018 and Gudmundsson *et al.*, 2018 apud Chagas 2020).

No que diz respeito à elevação do terreno, o semiárido da Bahia apresenta valores médios de cota entre 290 a 1100 m (Figura 13).

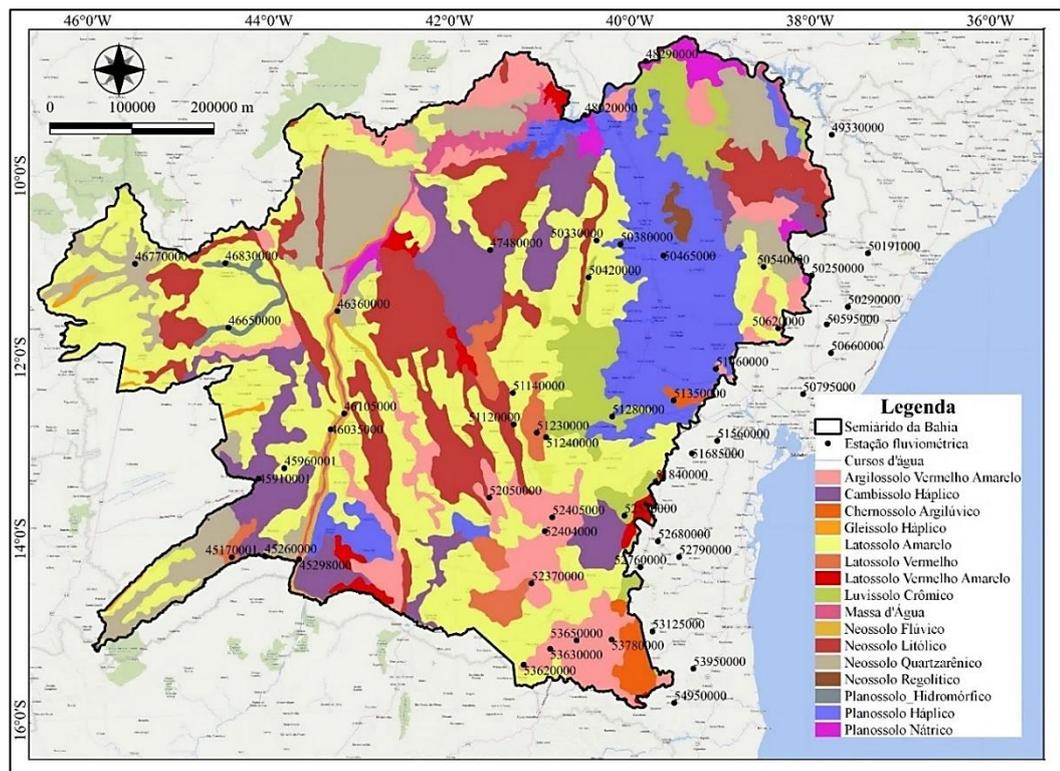
O solo do semiárido da Bahia apresenta em média 50,63% de areia, 33,07% de argila e 16,30% de silte. A profundidade mediana do lençol freático é de 30,61 m e a profundidade do leito rochoso atinge uma média de 28,98 m. Conforme a classificação da EMBRAPA, foram identificados 16 tipos diferentes de solos, tendo o Latossolo Amarelo como solo predominante (Figura 14).

Figura 13 - Elevação média do terreno no semiárido da Bahia.



Fonte: Próprio autor (2023).

Figura 14 - Caracterização do solo no semiárido da Bahia.

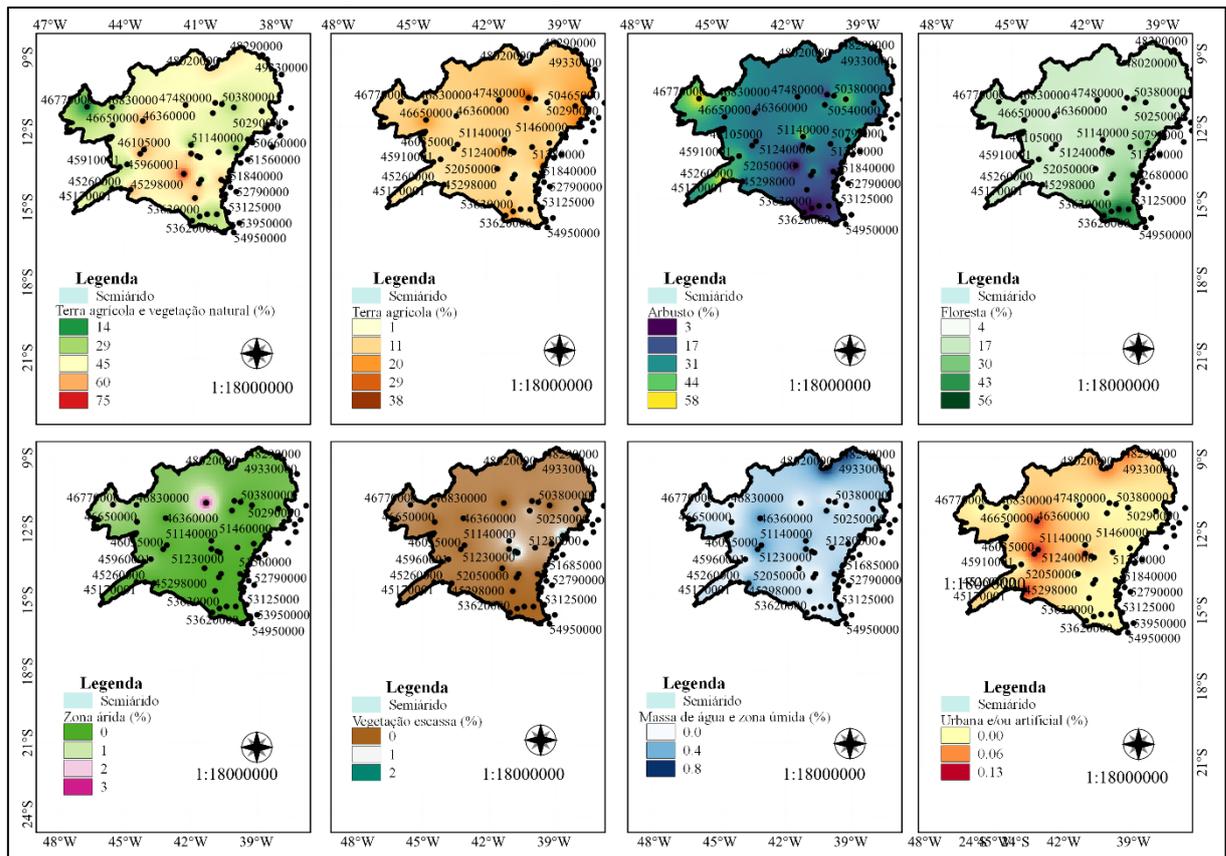


Fonte de dados: Mapa de solos do Brasil (IBGE, 2001), Cursos d'água (IBGE, 2017), semiárido da Bahia (SUDENE, 2017), Bacias Hidrográficas (Do *et al.*, 2018 e Gudmundsson *et al.*, 2018 apud Chagas *et al.*, 2020).

Em 31 bacias (60%) o uso do solo predominante é terras agrícolas e vegetação natural, em 12 bacias (23%) a predominância é a cobertura por arbustos e nas outras 9 bacias (17%) predomina florestas folhosas ou de folhas caducas, perenes.

Também foram avaliados a cobertura por terras agrícolas, prados ou zonas com vegetação escassa (<15%), cobertura por zonas áridas, cobertura por superfícies artificiais ou zonas urbanas, cobertura por massas de água ou zonas úmidas (Figura 15).

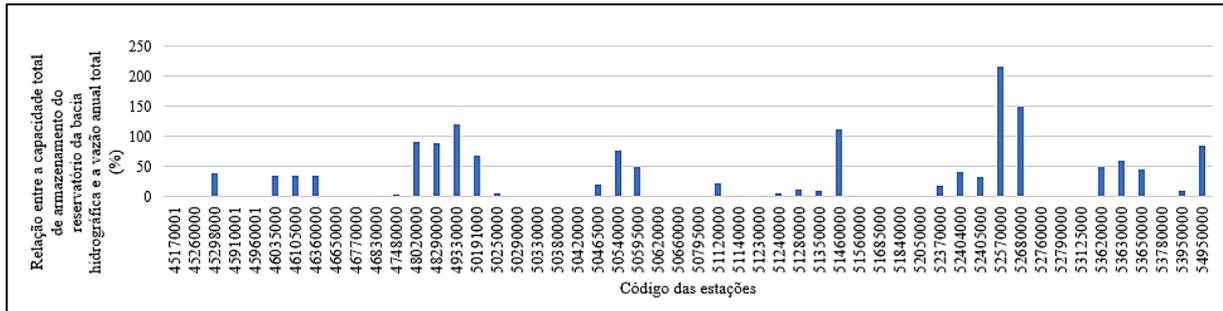
Figura 15 - Uso e ocupação das terras no semiárido da Bahia.



Fonte: Dados obtidos de ESA GlobCover (2009) *apud* Chagas *et al.* (2020).

Para o índice de intervenção humana foi avaliado a relação entre a capacidade total de armazenamento do reservatório da bacia hidrográfica e a sua vazão anual total (Figura 16).

Figura 16 - Relação entre a capacidade total de armazenamento do reservatório da bacia hidrográfica e a sua vazão anual total.

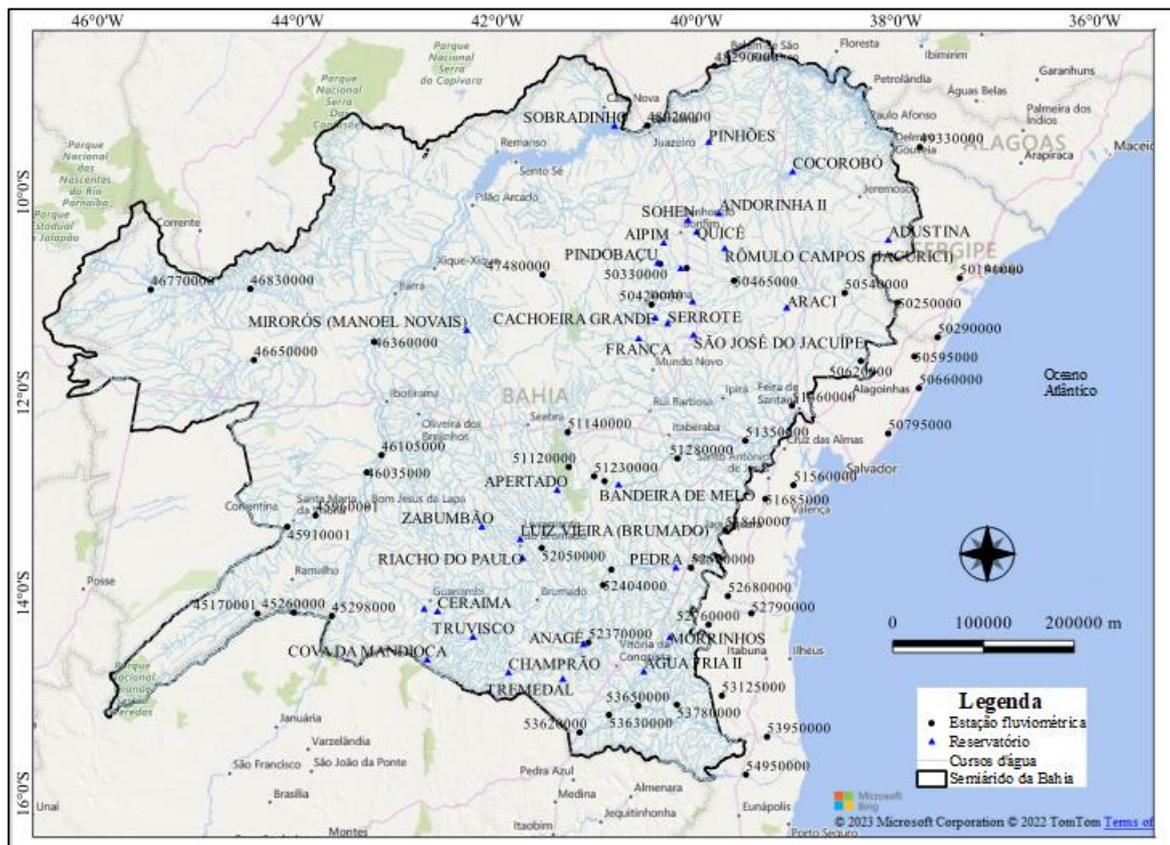


Fonte: Dados obtidos de GRanD v1.3, ONS e ANA (2018) *apud* Chagas *et al.* (2020).

Segundo Chagas *et al.* (2020) 17,8% das bacias hidrográficas presentes na base de dados do CAMELS - BR possuem usos consuntivos anuais de água superiores a 5% da vazão média anual. Essas bacias estão localizadas nas regiões mais secas do país, ou seja, nas regiões hidrográficas do São Francisco, Atlântico Leste, Atlântico Nordeste oriental e alto Paraná.

Na área de estudo foram identificados um total de 33 reservatórios monitorados pela Agência Nacional de Águas e Saneamento Básico (Figura 17).

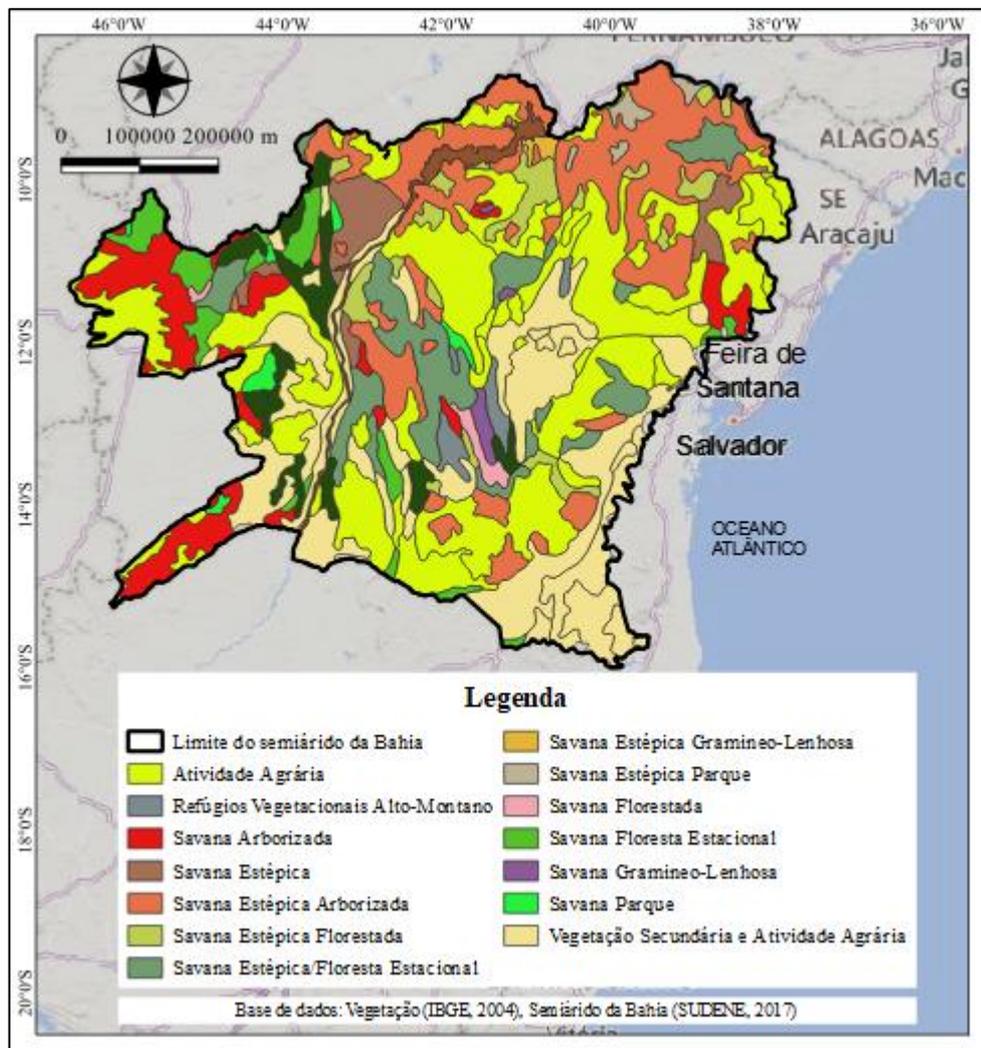
Figura 17 – Reservatórios localizados no semiárido da Bahia monitorados pela ANA.



Fonte de dados: Reservatório (ANA, 2017), semiárido da Bahia (SUDENE, 2017), Cursos d'água (IBGE, 2017), Estação Fluviométrica (Do *et al.*, 2018 e Gudmundsson *et al.*, 2018 *apud* Chagas *et al.*, 2020).

Em relação à vegetação, foram identificados 14 tipos diferentes de cobertura vegetal no semiárido da Bahia. Sendo que a maior parcela é composta por atividades agrárias, outra vegetação que se destaca é a savana (Figura 18).

Figura 18 - Cobertura vegetal no semiárido da Bahia.



Fonte de dados: Semiárido da Bahia (SUDENE, 2017), Vegetação (IBGE, 2004)

4.2. IDENTIFICAÇÃO DAS REGIÕES HOMOGÊNEAS NO SEMIÁRIDO DA BAHIA

A identificação das regiões homogêneas considerou as características fisiográficas e climáticas das bacias pertencentes ao semiárido da Bahia. Para a descrição mais aprofundada do semiárido da Bahia, foram avaliados vinte e seis atributos presentes na base de dados do CAMELS-BR. Esses atributos foram separados em sete grupos: topografia, clima, geologia,

hidrologia, uso e ocupação da terra, solo e intervenção humana (Quadro 3). Considerando apenas as características que tinham valores quantitativos, descartando as variáveis qualitativas.

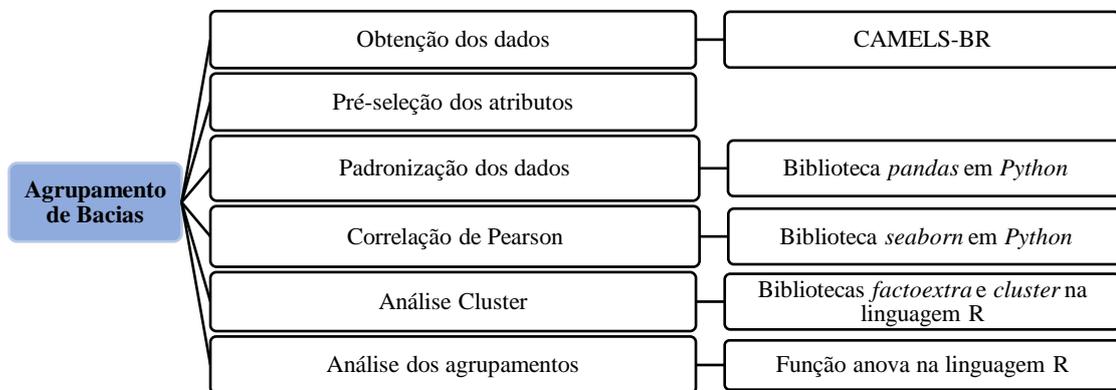
Quadro 3 - Atributos usados para descrever o semiárido da Bahia usando a base de dados do CAMELS-BR.

Índice	Atributo	Unidade	Fonte
Topografia	Elevação da estação	m.a.s.l.	HydroSHEDS 15 arc-sec DEM
	Elevação média da bacia	m.a.s.l.	HydroSHEDS 15 arc-sec DEM
	Declive médio da bacia hidrográfica	m km ⁻¹	HydroSHEDS 15 arc-sec DEM
	Latitude	-	-
	Longitude	-	-
	área da Bacia hidrográfica de acordo com GSIM	km ²	Do <i>et al.</i> (2018)
Climático	Precipitação média diária	mm day ⁻¹	CHIRPS v2.0
	Evapotranspiração potencial diária média (PET)	mm day ⁻¹	GLEAM v3.3a
	Aridez, calculada como a razão entre a PET média e a precipitação média	-	GLEAM v3.3a and CHIRPS v2.0
	Frequência dos dias de precipitação máxima (5 vezes a precipitação média diária)	days yr ⁻¹	CHIRPS v2.0
	Duração média dos eventos de precipitação máxima (número de dias consecutivos 5 vezes a precipitação média diária)	days	CHIRPS v2.0
	Frequência dos dias secos (< 1 mm dia ⁻¹)	days yr ⁻¹	CHIRPS v2.0
Uso e ocupação da terra	Porcentagem coberta por um mosaico de terras agrícolas e vegetação natural	%	ESA GlobCover2009
	Porcentagem coberta por florestas folhosas ou de folhas caducas, perenes ou caducas	%	ESA GlobCover2009
	Porcentagem coberta por arbustos	%	ESA GlobCover2009
	Porcentagem coberta por superfícies artificiais ou zonas urbanas	%	ESA GlobCover2009
	Porcentagem coberta por massas de água ou zonas úmidas	%	ESA GlobCover2009
Geológico	Porosidade subterrânea da bacia hidrográfica	-	GLHYMPS v2.0
	Permeabilidade da superfície (escala log10) da bacia hidrográfica, extrair para cada bacia hidrográfica utilizando a média geométrica	m ²	GLHYMPS v2.0
Solo	Porcentagem do teor de areia do material do solo inferior a 2 mm a uma profundidade de 30 cm	%	SoilGrids250m
	Porcentagem de silte do material do solo inferior a 2 mm a uma profundidade de 30 cm	%	SoilGrids250m
	Porcentagem do teor de argila do material do solo inferior a 2 mm a uma profundidade de 30 cm	%	SoilGrids250m
	Profundidade do leito rochoso	cm	SoilGrids250m
	Profundidade mediana do lençol freático	cm	Fan <i>et al.</i> (2013)
Intervenção humana	Capacidade máxima total de armazenamento dos reservatórios na bacia hidrográfica	10 ⁶ m ³	GRanD v1.3, ONS, and ANA (2018)
	Razão entre a capacidade total de armazenamento do reservatório da bacia hidrográfica e a vazão anual total	%	GRanD v1.3, ONS, and ANA (2018)

Fonte: Dados obtidos de Chagas *et al.* (2020).

Tendo em vista a particularidade de cada bacia hidrográfica, a escolha do conjunto de atributos foi realizada de forma criteriosa, para que sintetizassem diferentes pontos de vista sobre as bacias e transmitem a heterogeneidade entre elas, em uma mesma amostra e, posteriormente, no agrupamento utilizando a clusterização (Figura 19).

Figura 19 – Esquema para o agrupamento de bacias.



Fonte: Próprio autor (2023).

4.2.1. Padronização dos dados

A transformação dos dados é uma etapa importante do processo de análise. Esse processo consistiu em transformar todas as variáveis em distribuição com uma média de 0 e um desvio típico de 1. Cada valor padronizado foi calculado subtraindo a média do recurso correspondente e dividindo pelo desvio padrão. Para isso, foi utilizada a padronização *z-score* com auxílio da biblioteca *Pandas* (The Pandas Development Team, 2020) no *Python* (Python Software Foundation, 2021) nos atributos das 52 bacias.

4.2.2. Análise de correlação

A análise de correlação foi realizada pela aplicação da matriz de Pearson, descrito no item 3.1, buscando identificar se existe dependência entre os atributos e qual seria esse grau, ou seja, o quanto uma característica da bacia interfere em outra. A intenção foi determinar quais descritores físicos e climáticos podem melhor dividir as bacias hidrográficas nos agrupamentos hidrológicos, sempre comparando em relação à precipitação média anual de cada estação.

Nesse processo foi utilizada a biblioteca *seaborn* (Waskom, 2021) disponível na linguagem *Python*. Foram selecionados apenas os atributos que tiveram correlação igual ou superior a $\pm 0,70$ em relação à precipitação média anual (Tabela 5).

Tabela 5 - Atributos selecionados após a correlação de Pearson.

Atributos	Correlação de Pearson
Precipitação média anual	1,0 (atributo de referência)
Frequência dos dias de precipitação máxima (5 vezes a precipitação média diária)	-0,96
Duração média dos eventos de precipitação máxima (número de dias consecutivos 5 vezes a precipitação média diária)	-0,82
Frequência dos dias secos ($< 1 \text{ mm dia}^{-1}$)	-0,90
Aridez	-0,96
Profundidade mediana do lençol freático	-0,83
Superfícies artificiais ou zonas urbanas	0,80
Longitude	-0,76
Porosidade subterrânea da bacia hidrográfica	0,71

Fonte: Próprio autor (2023).

Além das características acima, também foram adicionadas a declividade média do rio principal, elevação média da bacia, precipitação e evapotranspiração potencial anual, fração de areia no solo, área de drenagem, declividade média da bacia, tendo em vista que essas características foram citadas no processo de regionalização de vazão por autores como Lemos (2006), Tucci (2017), Kuana *et al.* (2020), Kuana (2022).

4.2.3. Análise Cluster

Após a seleção dos descritores, foi realizado o agrupamento das bacias pertencentes ao semiárido da Bahia usando o método de clusterização *k-means* e a técnica *Ward*, já discutidos no item 3.2. Em ambas as técnicas foram aplicadas a distância euclidiana.

Para a aplicação dos métodos *k-means* e *Ward* foram utilizadas as bibliotecas *Extract and Visualize the Results of Multivariate Data Analyses (factoextra)* (Kassambara; Mundt, 2020) e *cluster* (MAEHLER *et al.*, 2023) na linguagem R. A quantidade de grupos necessários foi definida utilizando o método *Elbow*. Para a aplicação do método *Ward* foi utilizado o método de encadeamento “*single linkage*”.

4.2.4. Análise dos agrupamentos

A análise da estrutura dos agrupamentos foi realizada utilizando a função *aov* (CHAMBERS; FREENY; HEIBERGER, 1992) na linguagem R para o cálculo da Anova. A princípio foi determinada a hipótese H_0 , em que a variável selecionada não auxiliou na separação entre grupos e H_1 , em que a variável selecionada auxiliou na separação entre grupos. Ao executar a função *anova*, foi mostrado como resultado a média de dissimilaridade entre

grupos e dentro do grupo, Teste F e o P-valor (ver item 3.2) para todas as treze variáveis selecionadas anteriormente. Nos casos em que o P-valor foi maior que o nível de significância de 0,05, rejeitou-se H_1 , ficando com a hipótese de que a variável selecionada não auxiliou na separação entre grupos. Neste caso, retirou-se a variável que obteve o nível de significância superior à 0,05. Nas situações em que mais de um atributo não foi significativo, retirou-se aquele que obteve o menor valor no Teste F. A aplicação da função anova foi repetida até que todas as características obtiveram um P-valor menor ou igual a 0,05.

4.3. REGIONALIZAÇÃO DE PARÂMETROS DO MODELO GR4J PARA BACIAS SEM DADOS NO SEMIÁRIDO DA BAHIA

Da regionalização dos parâmetros do modelo GR4J até o processo de obtenção de vazões mínimas em bacias sem dados foi necessária uma série de procedimentos para garantir a confiabilidade do resultado final. Para isso, foi preciso coletar informações de vazão, precipitação e evapotranspiração das estações pertencentes à área de estudo, aplicar o modelo GR4J para calibração das bacias, realizar a validação, comparar estatisticamente os valores encontrados e por fim, realizar a regionalização dos parâmetros do modelo.

4.3.1. Análise de dados

No artigo elaborado por Chagas *et al.* (2022) apresenta de forma detalhada como foi realizada a análise dos dados do CAMELS BR. Resumidamente, foi selecionada apenas estações fluviométricas que apresentaram menos de 5% de falhas entre os anos de 1990 (começando em 1º de setembro de 1989) e 2009 (terminando em 31 de agosto de 2009). Os autores escolheram os anos de 1990 a 2009 porque (i) é o período com o maior número de estações com dados disponíveis e (ii) coincide com o período de análise de outros conjuntos de dados do CAMELS, permitindo comparações diretas com eles. Ressaltando que as séries temporais de algumas estações fluviométricas incluem dados de 1980 a 2018, quando disponíveis, para permitir análises complementares por outros usuários. A base de dados do CAMELS foca nos postos com melhores informações, isso reduz o número de postos disponíveis, dessa forma, informações podem estar sendo desprezadas. Por outro lado, ao focar nos melhores postos, e, que foram consistidos pelo grupo da CAMELS se evita trazer postos com informações duvidosas.

Em relação aos dados da precipitação CPC e evapotranspiração potencial de GLEAM v3.3^a, os valores diários representam a média de todas as células com seus centroides

interceptados pela captação, das quais todas as células contribuem para a média igualmente, quer a bacia as cubra totalmente ou não (CHAGAS *et al.*, 2020).

4.3.2. Objetivo da regionalização

Considerando que a área de estudo se localiza no semiárido da Bahia, o presente trabalho optou pela regionalização das vazões mínimas. As vazões mínimas asseguram a qualidade e quantidade de água, no tempo e no espaço, fundamentais para manter os componentes, as funções e os processos dos ecossistemas aquáticos. A regionalização de vazões mínimas vem sendo aplicada em diversos estudos e indicam a disponibilidade hídrica de uma bacia hidrográfica, sendo assim, conhecê-las é importante para projetos de barragens e usinas hidrelétricas, avaliação de disponibilidade hídrica para irrigação, dentre outros projetos hidrológicos (GASQUES *et al.*, 2018) e também para a outorga de uso da água.

4.3.3. Modelo GR4J

A versão utilizada no presente estudo foi proposta por Perrin *et al.* (2003). Foi utilizado o conjunto de parâmetros sugerido por Zambrano-Bigiarini, (2020) (Tabela 6), a fim de limitar o intervalo desses valores.

Tabela 6 - Limites dos parâmetros do modelo GR4J.

X1	X2	X3	X4
0	-100	0	0
1200	100	5000	5

Fonte: Zambrano-Bigiarini (2020).

O limite dos parâmetros sugeridos por Perrin *et al.* (2003) são muito restritos. O autor afirma que o uso de uma gama de medidas muito ampla pode levar a exigir um número muito grande de iterações para encontrar um conjunto de valores aceitáveis.

Outro ponto fundamental para a utilização do GR4J é a separação dos dados das bacias para aquecimento, calibração e validação do modelo. Os dados foram separados da seguinte forma: 5 anos de dados para o aquecimento (KUANA, 2022). Considerando apenas os dados restantes, sem as informações utilizadas na fase de aquecimento, 70% foi para a calibração do modelo e 30% para validação (GOLIAN; MURPHY; MERESA, 2021; KUANA, 2022).

Para a aplicação do GR4J foi utilizado o pacote *airGR* versão 1.7.0 (CORON *et al.*, 2022) no *software R* (R Development Core Team, 2019), seguindo o código estabelecido por Zambrano-Bigiarini (2020), o qual utiliza a Otimização por Enxame de Partículas (*Particle*

Swarm Optimization - PSO) a partir da biblioteca *HydroPSO* (ZAMBRANO-BIGIARINI; ROJAS, 2018). A PSO é uma otimização global proposta por Eberhart e Kennedy (1995), a qual faz a atualização da velocidade e da posição das partículas no espaço de busca até encontrar uma solução suficientemente aceitável. A PSO já foi utilizada por autores como Aouf *et al.* (2020) e Zhou *et al.* (2015) para a otimização do modelo GR4J. Na calibração do modelo o autor utilizou o Kling-Gupta modificado. O KGE modificado também foi utilizado por Kuana *et al.* (2021) como função objetivo para a calibração dos modelos GR4J e GR6J, os autores alcançaram valores medianos aceitáveis ($\geq 0,70$).

4.3.4. Requisitos de dados no pacote airGR

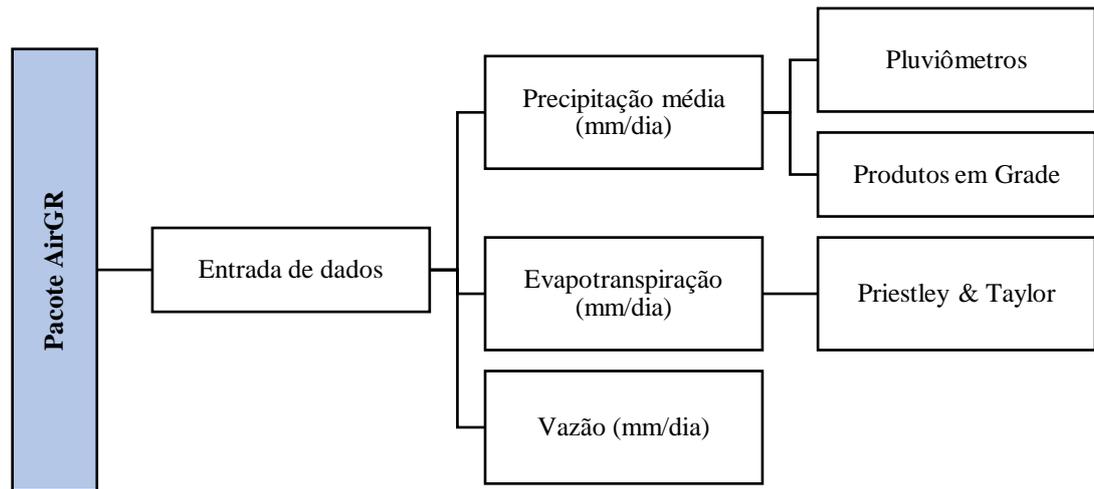
O pacote airGR é uma ferramenta útil para modelagem hidrológica e traz para o R várias das características hidrológicas dos modelos da família GR, desenvolvidos pelo Irstea na França.

O airGR serve para propósitos educacionais e investigação científica, com base numa análise de uma única bacia ou experimento com uma grande quantidade de amostras (CORON *et al.*, 2017). O pacote em questão também foi usado por Golian *et al.* (2021) na versão 1.4.3.6.0 e Althoff e Rodrigues (2021).

Os dados inseridos no modelo seguiram alguns requisitos como estabelecidos por Coron *et al.* (2017):

- Séries históricas diárias e contínuas de precipitação média e estimativas de evapotranspiração.
- A precipitação calculada com base em produtos em grade, processados fora do modelo.
- Foram utilizados observações de vazão para calibração de parâmetros e avaliação de desempenho do modelo.
- As séries temporais de vazão não necessariamente foram contínuas, pois a calibração e as funções de avaliação lidam com a falta de dados.
- Devido à indisponibilidade de dados, o presente trabalho utilizou dados de evapotranspiração calculados por Priestley & Taylor. Ressalta-se que, a equação não leva em consideração todos os fatores que explicam a evapotranspiração. Conforme Allan *et al.* (1998), o comportamento da evapotranspiração pode ser explicado por fatores como a radiação solar, a temperatura e a umidade relativa do ar, a velocidade dos ventos à 2m de altura, a umidade relativa do ar e a pressão atmosférica.

Figura 20 - Esquema para a entrada de dados no Pacote airGR no *software R*.



Fonte: Próprio autor (2023).

4.3.5. Análise do desempenho do modelo GR4J

Após a geração dos parâmetros do modelo e das vazões simuladas, foi comparado estatisticamente os valores encontrados. O desempenho do modelo hidrológico foi verificado por meio de índices estatísticos como o Coeficiente de Nash-Sutcliffe para o logaritmo das vazões (Nslog), a Raiz do Erro Médio Quadrático modificado para baixas vazões (REMQ_{BV}) e o Logaritmo dos erros quadrados (LOGQ) (detalhados no item 3.1.2), a fim de verificar as vazões mínimas. O coeficiente Nslog varia entre $-\infty$ a 1, sendo que 1 são para previsões perfeitas, já os coeficientes LOGQ e REMQ_{BV} variam entre 0 e $+\infty$, sendo que 0 são para previsões perfeitas. No intuito de verificar as vazões máximas e comparar com as vazões mínimas, foram calculados o coeficiente de determinação (R^2), viés e a Raiz do Erro Quadrático Médio (RMSE), o coeficiente de eficiência de Nash-Sutcliffe (NSE) e o coeficiente de eficiência de Kling-Gupta (KGE). O índice de eficiência dos critérios NSE e KGE apresenta um valor numérico entre $-\infty$ e 1 e é comumente adotado em modelagem hidrológica para expressar aderência entre as séries de vazão histórica e sintética. Para o cálculo dos critérios foi utilizado o *software R* e o *Excel* (Microsoft, 2016).

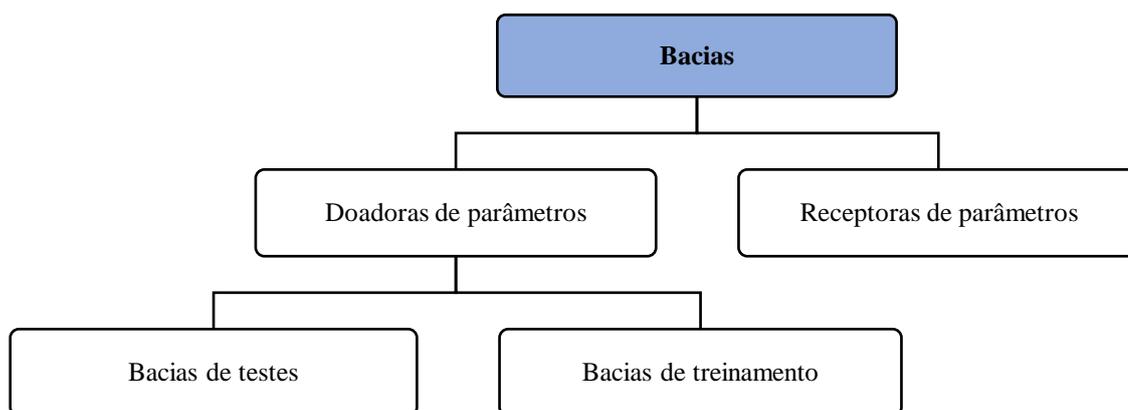
4.3.6. Regionalização de parâmetros do modelo GR4J para bacias sem dados

A regionalização de parâmetros hidrológicos é a ação de obter dados de uma bacia para simular a disponibilidade hídrica de outra bacia com características semelhantes, na qual não há séries hidrológicas disponíveis (VIRÃES; CIRILO, 2019).

A métrica de similaridade entre as bacias foi realizada por clusterização. E supõe-se que, se há similaridade entre bacias, há respostas hidrológicas semelhantes. Posteriormente foram separadas as bacias doadoras e as receptoras dentro de cada grupo para a transferência de parâmetros. As bacias consideradas doadoras são aquelas em que o desempenho do modelo foi julgado aceitável, ou seja, com o coeficiente de eficiência de Kling-Gupta (KGE) no período de calibração acima de 0,70. Foi escolhido o coeficiente KGE tendo em vista que esta é a função de calibração do modelo GR4J. Oudin *et al.* (2008) afirma que bacias mal modeladas não fornecem informações relevantes suficientes.

Após a separação das bacias doadoras de parâmetros, foi realizado um novo agrupamento: 70% das estações para treino e 30% das estações para teste (Figura 21). Para isso foi utilizada a função *sample.split* da biblioteca *caTools* (Tuszynski, (TUSZYNSKI, 2021)) disponível na linguagem R. A intenção foi aproveitar os dados coletados para poder realizar as previsões de dados futuros, já que, se for utilizado todos os dados de treino não teria dados para as verificações.

Figura 21 - Divisão das bacias de acordo com a existência de parâmetros.



Fonte: Próprio autor (2023).

Posteriormente foram geradas as equações de cada parâmetro, as quais foram validadas nas bacias de testes.

Foi utilizada a técnica de regressão linear para gerar as equações. Utilizou-se como variáveis dependentes os parâmetros calibrados do modelo GR4J e como variáveis independentes, as características descritivas. Ressalta-se que, os parâmetros do GR4J são gerados ao aplicar o modelo. A partir dos dados de vazão, evapotranspiração e precipitação, além das vazões simuladas, é gerado os valores para X1, X2, X3 e X4 para cada bacia.

A nível de padronização, foram escolhidas as características descritivas que apresentaram o coeficiente de correlação superior a 60% em pelo menos um dos grupos formados (Quadro 4).

Quadro 4 - Variáveis dependentes e independentes utilizadas para gerar as equações.

Variável dependente	Variável independente
X1	A frequência dos dias de precipitação elevada
	Razão entre a capacidade total de armazenamento do reservatório da bacia hidrográfica e a sua vazão anual total
	Vazão diária média
	Taxa de escoamento
	Declive médio da bacia hidrográfica
	Duração média dos eventos de vazão máxima
X2	Razão entre a capacidade total de armazenamento do reservatório da bacia hidrográfica e a sua vazão anual total
X3	Classe geológica mais comum na bacia hidrográfica (rochas metamórficas)
	Classe geológica mais comum na bacia hidrográfica (rochas metamórficas)
X4	Porosidade subterrânea da bacia hidrográfica
	Índice de vazão de base
	Frequência dos dias de vazão máxima
	Frequência dos dias de vazão mínima
	Duração média dos eventos de vazão mínima
	Porcentagem coberta por zonas áridas
	longitude
	Elevação da estação
	Vazão diária média
	Taxa de escoamento
	Q5 (vazão mínima)
	Duração média dos eventos de vazão máxima
	Declive médio da bacia hidrográfica

Fonte: Próprio autor (2023).

Foi utilizada a função nativa do R para criar o modelo de regressão. O objetivo foi verificar a possibilidade de prever os parâmetros do modelo GR4J utilizando as características físicas e climáticas das bacias para criar um modelo preditivo com regressão linear.

Após aplicar a técnica de regressão linear, foi avaliado o modelo a fim de verificar se dados utilizados foram adequados, se o modelo conseguiu representar os dados e principalmente se o modelo foi capaz de realizar previsões para novos dados. O modelo foi avaliado conforme os seguintes critérios: i) significância; ii) linearidade, iii) homocedasticidade dos resíduos (variância constante dos erros experimentais); iv) normalidade dos resíduos.

4.3.7. Obtenção das vazões de referência

Foram elaboradas as curvas de permanência, e estimadas as vazões com permanência de 90% na série histórica (Q90) a partir de dados diários para cada estação e a vazão de 7 dias de duração e 10 anos de tempo de retorno (Q7,10). As vazões Q90 e Q7,10 foram escolhidas

pela sua importância na outorga de vazões e gestão dos recursos hídricos. A construção da curva de permanência para todos os “n” registros disponíveis nos períodos em que o GR4J foi calibrado (1986 - 2005), envolveu: (i) determinar os percentis, que variaram de 0,05 a 0,95, (ii) calcular a vazão para cada percentil estabelecido através do Excel, (iii) realizar a plotagem das curvas utilizando o percentual e a vazão relacionada a cada percentil. As curvas também auxiliaram no processo de agrupamento das regiões homogêneas, como mostrado por Koefender (2015) no rio Ibicuí, localizado na porção oeste do estado rio Grande do Sul.

Utilizando todos os “n” registros disponíveis nos períodos em que o GR4J foi calibrado (1986 - 2005), o cálculo da $Q_{7,10}$ envolveu: (i) cálculo das médias móveis de 7 dias, (ii) cálculo da menor vazão média no ano, (iii) cálculo das vazões médias dos anos, (iv) cálculo do desvio padrão das vazões médias dos anos, (v) aplicação do método de Gumbel para valores mínimos, (vi) obtenção da $Q_{7,10}$. Ressalta-se que as vazões mínimas foram calculadas utilizando os dados diários observados e simulados através do modelo GR4J com intuito de comparar os resultados.

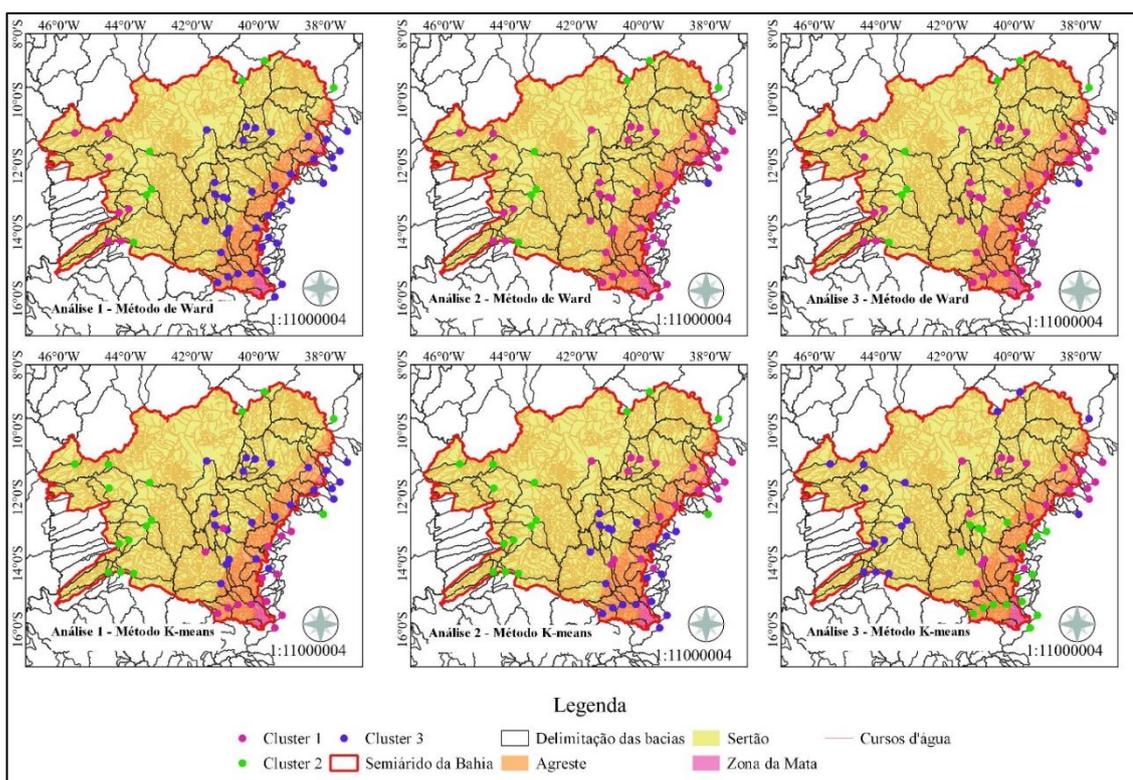
Posteriormente foram geradas as equações de cada parâmetro, as quais foram validadas nas bacias de testes. Foi utilizada a técnica de regressão linear para gerar as equações. Utilizou-se como variáveis dependentes as vazões simuladas e observadas e como variável independente, a área das bacias.

5. RESULTADOS

5.1. IDENTIFICAÇÃO DAS REGIÕES HOMOGÊNEAS

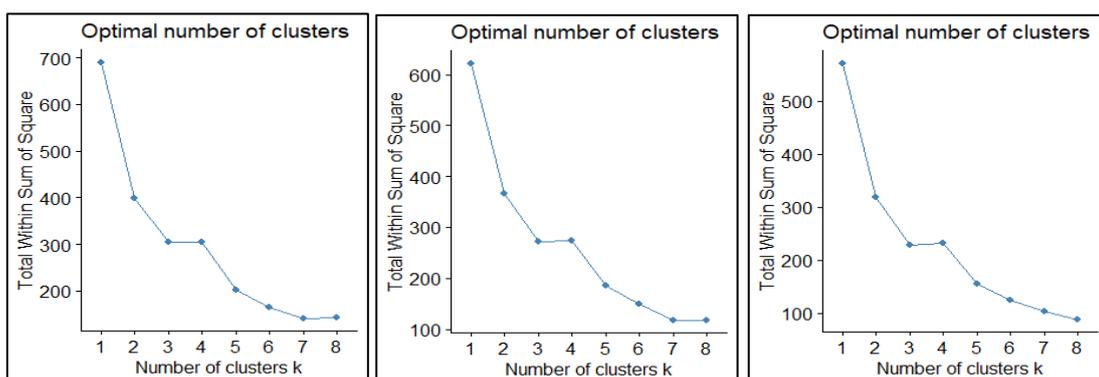
Para a identificação das regiões homogêneas utilizando as técnicas *Ward* e *k-means*, observou-se a formação de três clusters (Figura 22). A quantidade de clusters foi definida ao aplicar o método *Elbow* (Figura 23).

Figura 22 - Espacialização dos clusters formados após as três análises pelo método de *Ward* e *K-means*.



Fonte: Próprio autor (2023).

Figura 23 - Método *Elbow* para a determinação do número de cluster para as três análises.



Fonte: Próprio autor (2023).

Inicialmente, para a aplicação das técnicas *k-means* e *Ward*, foram utilizadas a precipitação média anual, a frequência dos dias de precipitação máxima (5 vezes a precipitação média diária), a duração média dos eventos de precipitação máxima (número de dias consecutivos 5 vezes a precipitação média diária), a frequência dos dias secos (< 1 mm dia-1), a aridez, a profundidade mediana do lençol freático, a superfícies artificiais ou zonas urbanas, a longitude, a porosidade subterrânea da bacia hidrográfica, a declividade média do rio principal, a elevação média da bacia, a precipitação e a evapotranspiração potencial anual, a fração de areia no solo, a área de drenagem e a declividade média da bacia. Foi observado que algumas das variáveis selecionadas não influenciaram na formação de pelo menos um dos clusters. Essa observação foi realizada após a primeira execução dos métodos de *Ward* e *k-means* para o conjunto de parâmetros selecionados e da análise dos agrupamentos. Logo, houve a necessidade de retirar esses atributos da amostra. Considerando a primeira análise, o processo foi executado três vezes, como apresentado a seguir.

5.1.1. Análise 1

O objetivo desta análise foi verificar quais as variáveis significativas para a formação dos grupos. A partir da avaliação dos agrupamentos formados pelo método *Ward*, a elevação média da bacia foi a única variável que apresentou um P-valor maior que o nível de significância de 0,05, ou seja, aceitou-se H_0 . Ficando com a hipótese de que a variável citada não auxiliou na separação entre grupos de pelo menos um cluster. Ao observar os maiores valores da estatística F, em conjunto com sua significância, foi possível analisar quais variáveis mais contribuíram para a formação de pelo menos um dos grupos. De acordo com a estatística F, as quatro variáveis que mais contribuíram para a formação de pelo menos um cluster foram: área da bacia hidrográfica (187,20), frequência dos dias secos (116,50), frequência dos dias de precipitação máxima (89,33) e superfícies artificiais ou zonas urbanas (65,55) (Tabela 7).

Tabela 7 - Primeira análise dos agrupamentos formados pelo método *Ward*.

Atributos	Variabilidade		Teste F	P valor
	Entre grupo	Dentro do grupo		
Area da bacia hidrográfica	22,55	0,12	187,20	$<2e^{-16}$
Aridez	9,72	0,64	15,08	$7,88e^{-06}$
Declividade média da bacia	8,52	0,69	12,29	$4,71e^{-05}$
Duração Média dos eventos de precipitação máxima	4,19	0,87	4,81	0,01
Elevação média da bacia	1,32	0,99	1,34	0,27*
Evapotranspiração média anual	12,02	0,55	21,85	$1,65e^{-7}$
Frequência dos dias de precipitação máxima	20,01	0,22	89,33	$<2e^{-16}$
Frequências dos dias secos	21,07	0,18	116,50	$<2e^{-16}$
Longitude	15,46	0,41	37,75	$1,2e^{-10}$

Porosidade subterrânea da bacia hidrográfica	5,64	0,81	6,97	0,02
Precipitação média anual	14,96	0,43	34,77	3,98e ⁽⁻¹⁰⁾
Profundidade mediana do lençol freático	15,44	0,41	37,63	1,26e ⁽⁻¹⁰⁾
Superfícies artificiais ou zonas urbanas	18,56	0,28	65,55	1,41e ⁽⁻¹⁴⁾
Teor de areia	6,39	0,78	8,20	0,00

* P valor maior que o nível de significância de 0,05, ou seja, a variável não foi significativa.

Fonte: Próprio autor (2023).

A partir da avaliação dos agrupamentos formados pelo método *k-means*, as variáveis duração média dos eventos de precipitação máxima e a elevação média da bacia apresentaram um P valor maior que o nível de significância de 0,05, ou seja, aceita-se H₀. Ficando com a hipótese de que duas variáveis citadas não auxiliaram na separação entre grupos de pelo menos um cluster. Conforme a estatística F, as quatro variáveis que mais contribuíram para a formação de pelo menos um cluster foram: precipitação média anual (104,80), aridez (73,51), frequência dos dias secos (70,14) e profundidade mediana do lençol freático (46,01) (Tabela 8).

Tabela 8 - Primeira análise dos agrupamentos formados pelo método *k-means*.

Atributos	Variabilidade		Teste F	P valor
	Entre grupo	Dentro do grupo		
Área da bacia hidrográfica	8,87	0,68	13,07	2,82e ⁽⁻⁰⁵⁾
Aridez	19,13	0,26	73,51	1,77e ⁽⁻¹⁵⁾
Declividade média da bacia	11,04	0,59	18,70	9,25e ⁽⁻⁰⁷⁾
Duração Média dos eventos de precipitação máxima	2,28	0,95	2,41	0,10*
Elevação média da bacia	0,84	1,01	0,84	0,44*
Evapotranspiração média anual	13,62	0,49	28,07	7,53e ⁽⁻⁹⁾
Frequência dos dias de precipitação máxima	14,80	0,44	33,87	6,53e ⁽⁻¹⁵⁾
Frequências dos dias secos	18,90	0,27	70,14	4,18e ⁽⁻¹⁵⁾
Longitude	10,27	0,62	16,52	3,28e ⁽⁻⁰⁶⁾
Porosidade subterrânea da bacia hidrográfica	6,77	0,76	8,86	0,00
Precipitação média anual	20,67	0,20	104,80	<2e ⁽⁻¹⁶⁾
Profundidade mediana do lençol freático	16,64	0,36	46,01	5,66e ⁽⁻¹²⁾
Superfícies artificiais ou zonas urbanas	10,10	0,63	16,08	4,29e ⁽⁻⁰⁶⁾
Teor de areia	12,61	0,53	23,95	5,55e ⁽⁻⁰⁸⁾

* P valor maior que o nível de significância de 0,05, ou seja, a variável não foi significativa.

Fonte: Próprio autor (2023).

Tabela 9 - Estatística descritiva dos clusters formados pelos métodos *k-means* e *Ward* para a análise 1.

	Grupo	Estatística	Elevação Média da Bacia	Declividade média da bacia	Área da bacia hidrográfica	Aridiz	Superfícies artificiais ou zonas urbanas	Porosidade da bacia hidrográfica	Profundidade mediana do lençol freático	Teor de areia	Frequência dos dias secos	Frequência dos dias de precipitação máxima	Longitude	Precipitação média anual	Evapotranspiração média anual	Duração Média dos eventos de precipitação máxima
			m.a.s.l.	m km ⁻¹	Km ²	-	%	-	cm	%	dia/ano	dia/ano	-	mm/ano	mm/ano	dia
<i>k-means</i>	1	Desvio	225,05	19,48	18359,88	0,11	0,01	0,05	1031,00	2,46	12,37	1,56	0,83	94,20	44,57	0,13
		Média	665,27	57,93	12186,92	1,26	0,00	0,09	4105,05	46,76	268,05	22,06	-40,12	868,89	1129,98	1,51
		Máximo	1027,39	108,35	68283,32	1,43	0,05	0,21	5845,68	51,19	242,00	18,35	-39,03	1014,42	1221,93	1,66
		Mínimo	193,94	26,25	355,44	1,08	0,00	0,01	1844,28	43,10	242,00	18,35	-41,61	774,17	1060,73	1,25
	2	Desvio	140,57	6,53	226621,73	0,08	0,05	0,02	416,69	3,18	16,40	3,07	2,46	77,77	36,56	0,12
		Média	679,37	24,85	205452,69	1,10	0,05	0,17	1274,77	51,84	225,72	17,10	-42,74	1095,23	1228,67	1,45
		Máximo	764,77	34,28	619185,70	1,28	0,13	0,19	1915,54	57,28	207,70	12,50	-37,76	1188,47	1307,55	1,68
		Mínimo	184,36	16,02	4703,06	0,96	0,00	0,14	289,41	47,12	207,70	12,50	-45,53	938,98	1178,25	1,28
	3	Desvio	235,37	15,02	14605,40	0,20	0,00	0,06	972,11	1,80	9,14	1,23	1,37	71,87	28,16	0,17
		Média	597,63	43,37	16258,17	1,67	0,00	0,12	3589,99	52,17	272,28	22,20	-39,68	710,80	1174,93	1,56
		Máximo	1139,56	70,18	46547,88	2,04	0,00	0,22	4944,23	55,30	250,15	19,80	-37,35	883,93	1255,39	1,79
		Mínimo	239,81	24,02	665,01	1,30	0,00	0,04	1902,50	49,36	250,15	19,80	-41,57	563,66	1129,95	1,30
<i>Ward</i>	1	Desvio	22,39	2,79	11887,67	0,03	0,00	0,01	519,24	1,42	7,82	1,45	0,52	57,11	29,91	0,04
		Média	736,39	18,68	20519,65	1,10	0,00	0,19	1088,67	54,72	239,27	19,87	-44,48	1109,58	1256,33	1,36
		Máximo	764,77	22,38	35821,92	1,14	0,00	0,19	1764,57	57,28	228,85	17,65	-43,91	1188,47	1307,55	1,41
		Mínimo	696,42	16,02	5823,41	1,07	0,00	0,17	289,41	53,15	228,85	17,65	-45,53	1048,16	1234,31	1,28
	2	Desvio	27,44	2,33	140857,23	0,12	0,02	0,00	127,55	1,51	0,99	1,00	2,32	91,40	22,62	0,07
		Média	693,06	30,81	419064,24	1,10	0,09	0,15	1369,32	48,96	209,67	14,21	-41,68	1067,87	1206,52	1,56
		Máximo	723,80	34,28	619185,70	1,28	0,13	0,16	1572,70	50,95	207,70	12,50	-37,76	1179,69	1230,67	1,68
		Mínimo	648,06	28,05	256995,20	0,96	0,07	0,14	1214,95	47,12	207,70	12,50	-43,76	938,98	1178,25	1,49
	3	Desvio	238,50	18,19	15868,55	0,27	0,02	0,06	1043,89	3,32	11,30	1,49	1,22	128,90	40,89	0,16
		Média	611,67	48,28	14454,15	1,50	0,00	0,11	3735,68	50,17	269,96	22,03	-39,80	781,56	1158,77	1,53
		Máximo	1139,56	108,35	68283,32	2,04	0,11	0,22	5845,68	55,30	242,00	17,95	-37,35	1186,33	1255,39	1,79
		Mínimo	184,36	24,02	355,44	1,06	0,00	0,01	1844,28	43,10	242,00	17,95	-41,61	563,66	1060,73	1,25

Fonte: Próprio autor (2023).

5.1.2. Análise 2

Para essa análise foi desconsiderada a elevação média da bacia, tendo em vista que foi o único atributo que não contribuiu para a formação de pelo menos um dos clusters em ambos os métodos da análise 1.

Para o novo conjunto de atributos, a declividade média da bacia, a duração média dos eventos de precipitação máxima, a longitude, a porosidade subterrânea da bacia hidrográfica e o teor de areia apresentaram um P valor maior que o nível de significância de 0,05, ou seja, aceita-se H_0 . Ficando com a hipótese de que as cinco variáveis citadas não auxiliaram na separação entre grupos de pelo menos um cluster. Conforme a estatística F, as quatro variáveis que mais contribuíram para a formação de pelo menos um cluster foram: superfícies artificiais ou zonas urbanas (235,80), área da bacia hidrográfica (187,00), frequência dos dias de precipitação máxima (76,07) e frequência dos dias secos (46,20) (Tabela 10).

Tabela 10 - Segunda análise dos agrupamentos formados pelo método *Ward*.

Atributos	Variabilidade		Teste F	P valor
	Entre grupo	Dentro do grupo		
Área da bacia hidrográfica	22,55	0,12	187,00	$<2e^{-16}$
Aridez	4,93	0,84	5,88	0,01
Declividade média da bacia	1,81	0,97	1,87	0,17*
Duração Média dos eventos de precipitação máxima	1,45	0,98	1,47	0,24*
Evapotranspiração média anual	1,28	0,99	1,30	0,28*
Frequência dos dias de precipitação máxima	19,29	0,25	76,07	$9,41e^{-16}$
Frequências dos dias secos	16,66	0,36	46,20	$5,3e^{-12}$
Longitude	1,60	0,98	1,64	0,21*
Porosidade subterrânea da bacia hidrográfica	1,22	0,99	1,23	0,30*
Precipitação média anual	7,14	0,75	9,52	0,00
Profundidade mediana do lençol freático	6,02	0,80	7,57	0,00
Superfícies artificiais ou zonas urbanas	23,10	0,10	235,80	$<2e^{-16}$
Teor de areia	1,02	1,00	1,02	0,37*

* P-valor maior que o nível de significância de 0,05, ou seja, a variável não foi significativa.

Fonte: Próprio autor (2023).

A partir da avaliação dos agrupamentos formados pelo método *k-means*, das variáveis selecionadas após a Análise 1, a duração média dos eventos de precipitação máxima continuou apresentando um P valor maior que o nível de significância de 0,05, ou seja, aceita-se H_0 . Ficando com a hipótese de que a variável citada não auxiliou na separação entre grupos de pelo menos um cluster. Segundo a estatística F, as quatro variáveis que mais contribuíram para a formação de pelo menos um cluster foram as mesmas apresentadas na Análise 1: precipitação média anual (102,20), aridez (71,86), frequência dos dias secos (69,36) e profundidade mediana do lençol freático (57,26) (Tabela 11).

Tabela 11 - Segunda análise dos agrupamentos formados pelo método *k-means*.

Atributos	Variabilidade		Teste F	P valor
	Entre grupo	Dentro do grupo		
Área da bacia hidrográfica	8,875	0,679	13,08	2,81e ⁽⁻⁰⁵⁾
Aridez	19,02	0,27	71,86	2,69e ⁽⁻¹⁵⁾
Declividade média da bacia	11,64	0,57	20,57	3,27e ⁽⁻⁰⁷⁾
Duração Média dos eventos de precipitação máxima	2,34	0,95	2,48	0,09*
Evapotranspiração média anual	12,81	0,52	24,73	3,75e ⁽⁻⁰⁸⁾
Frequência dos dias de precipitação máxima	14,79	0,44	33,85	5,84e ⁽⁻¹⁰⁾
Frequências dos dias secos	18,84	0,27	69,36	5,11e ⁽⁻¹⁵⁾
Longitude	11,20	0,58	19,18	7,06e ⁽⁻⁰⁷⁾
Porosidade subterrânea da bacia hidrográfica	5,71	0,81	7,06	0,00
Precipitação média anual	20,57	0,20	102,20	<2e ⁻¹⁶
Profundidade mediana do lençol freático	17,86	0,31	57,26	1,51e ⁽⁻¹³⁾
Superfícies artificiais ou zonas urbanas	10,08	0,63	16,02	4,43e ⁽⁻⁰⁶⁾
Teor de areia	12,18	0,54	22,39	1,24e ⁽⁻⁰⁷⁾

* P-valor maior que o nível de significância de 0,05, ou seja, a variável não foi significativa.

Fonte: Próprio autor (2023).

Tabela 12 - Estatística descritiva dos clusters formados pelos métodos *k-means* e *Ward* para a análise 2.

	Grupo	Estatística	Declividade média da bacia	Área da bacia hidrográfica	Aridez	Superfícies artificiais ou zonas urbanas	Porosidade subterrânea da bacia hidrográfica	Profundidade mediana do lençol freático	Teor de areia	Frequência dos dias secos	Frequência dos dias de precipitação máxima	Longitude	Precipitação média anual	Evapotranspiração média anual	Duração Média dos eventos de precipitação máxima
			m km ⁻¹	Km ²	-	%	-	cm	%	dia/ano	dia/ano	-	mm/ano	mm/ano	dia
<i>k-means</i>	1	Desvio	15,75	15810,85	0,19	0,00	0,06	929,42	1,55	9,57	1,27	1,32	68,07	29,33	0,17
		Média	40,81	17683,28	1,71	0,00	0,11	3325,26	52,76	272,31	22,09	-39,32	693,06	1177,47	1,52
		Máximo	70,18	46547,88	2,04	0,00	0,21	4729,63	55,30	250,15	19,80	-37,35	883,93	1255,39	1,76
		Mínimo	24,02	665,01	1,30	0,00	0,04	1902,50	49,36	250,15	19,80	-41,57	563,66	1129,95	1,30
	2	Desvio	6,53	226621,73	0,08	0,05	0,02	416,69	3,18	16,40	3,07	2,46	77,77	36,56	0,12
		Média	24,85	205452,69	1,10	0,05	0,17	1274,77	51,84	225,72	17,10	-42,74	1095,23	1228,67	1,45
		Máximo	34,28	619185,70	1,28	0,13	0,19	1915,54	57,28	207,70	12,50	-37,76	1188,47	1307,55	1,68
		Mínimo	16,02	4703,06	0,96	0,00	0,14	289,41	47,12	207,70	12,50	-45,53	938,98	1178,25	1,28
	3	Desvio	17,07	16092,36	0,15	0,01	0,06	907,59	2,57	11,38	1,44	0,83	93,35	42,72	0,14
		Média	56,52	11908,20	1,32	0,00	0,11	4220,31	47,62	269,14	22,19	-40,35	844,10	1139,41	1,56
		Máximo	108,35	68283,32	1,62	0,05	0,22	5845,68	51,19	242,00	18,35	-39,03	1014,42	1221,93	1,79
		Mínimo	26,25	355,44	1,08	0,00	0,01	1844,28	43,10	242,00	18,35	-41,61	698,69	1060,73	1,25
<i>Ward</i>	1	Desvio	19,98	15450,89	0,28	0,01	0,06	1375,24	3,55	15,37	1,59	2,04	162,99	53,53	0,16
		Média	44,07	15640,73	1,45	0,00	0,12	3355,94	50,85	265,68	21,78	-40,59	824,55	1173,58	1,51
		Máximo	108,35	68283,32	2,04	0,05	0,22	5845,68	57,28	228,85	17,65	-37,35	1188,47	1307,55	1,79
		Mínimo	16,02	355,44	1,07	0,00	0,01	289,41	43,10	228,85	17,65	-45,53	563,66	1060,73	1,25
	2	Desvio	2,33	140857,23	0,12	0,02	0,00	127,55	1,51	0,99	1,00	2,32	91,40	22,62	0,07
		Média	30,81	419064,24	1,10	0,09	0,15	1369,32	48,96	209,67	14,21	-41,68	1067,87	1206,52	1,56
		Máximo	34,28	619185,70	1,28	0,13	0,16	1572,70	50,95	207,70	12,50	-37,76	1179,69	1230,67	1,68
		Mínimo	28,05	256995,20	0,96	0,07	0,14	1214,95	47,12	207,70	12,50	-43,76	938,98	1178,25	1,49
	3	Desvio	-	-	-	-	-	-	-	-	-	-	-	-	-
		Média	26,26	4703,06	1,06	0,11	0,17	1915,54	51,85	243,15	17,95	-38,07	1186,33	1190,10	1,29
		Máximo	26,26	4703,06	1,06	0,11	0,17	1915,54	51,85	243,15	17,95	-38,07	1186,33	1190,10	1,29
		Mínimo	26,26	4703,06	1,06	0,11	0,17	1915,54	51,85	243,15	17,95	-38,07	1186,33	1190,10	1,29

Fonte: Próprio autor (2023).

5.1.3. Análise 3

A partir da avaliação dos agrupamentos formados pelo método *Ward* e *k-means* a duração média dos eventos de precipitação máxima foi a única que apresentou um P valor maior que o nível de significância de 0,05 em ambos os métodos. Por esse motivo, a variável foi retirada da Análise 3.

Ao utilizar o novo conjunto de atributos no método de *Ward*, foi observado que além das variáveis (declividade média da bacia, longitude, porosidade subterrânea da bacia hidrográfica e teor de areia) que não contribuíram para a formação de pelo menos o cluster na análise 2, a evapotranspiração média anual também apresentou um P valor maior que o nível de significância de 0,05, ou seja, aceita-se H_0 . Ficando com a hipótese de que as cinco variáveis citadas não auxiliaram na separação entre grupos de pelo menos um cluster. Segundo a estatística F, as quatro variáveis que mais contribuíram para a formação de pelo menos um cluster foram: superfícies artificiais ou zonas urbanas (235,80), área da bacia hidrográfica (187,00), frequência dos dias de precipitação máxima (76,07) e frequência dos dias secos (46,20) (Tabela 13).

Tabela 13 - Terceira análise dos agrupamentos formados pelo método *Ward*.

Atributos	Variabilidade		Teste F	P valor
	Entre grupo	Dentro do grupo		
Área da bacia hidrográfica	22,55	0,12	187,00	$<2e^{-16}$
Aridez	4,93	0,84	5,88	0,01
Declividade média da bacia	1,81	0,97	1,87	0,17*
Evapotranspiração média anual	1,28	0,99	1,30	0,28*
Frequência dos dias de precipitação máxima	19,29	0,25	76,07	$9,41e^{-16}$
Frequências dos dias secos	16,66	0,36	46,20	$5,3e^{-12}$
Longitude	1,60	0,98	1,64	0,21*
Porosidade subterrânea da bacia hidrográfica	1,22	0,99	1,23	0,30*
Precipitação média anual	7,14	0,75	9,52	0,00
Profundidade mediana do lençol freático	6,02	0,80	7,57	0,00
Superfícies artificiais ou zonas urbanas	23,10	0,10	235,80	$<2e^{-16}$
Teor de areia	1,02	1,00	1,02	0,37*

* P-valor maior que o nível de significância de 0,05, ou seja, a variável não foi significativa.

Fonte: Próprio autor (2023).

A partir da avaliação dos agrupamentos formados pelo método *k-means*, das variáveis selecionadas na análise anterior, todas apresentaram P valor maior que o nível de significância de 0,05, ou seja, aceita-se H_1 . Ficando com a hipótese de que todas as variáveis citadas auxiliaram na separação entre grupos de pelo menos um cluster. Conforme a estatística F, as quatro variáveis que mais contribuíram para a formação de pelo menos um cluster foram:

precipitação média anual (109,90), aridez (84,58), frequência dos dias secos (70,61) e profundidade mediana do lençol freático (51,48) (Tabela 14).

Tabela 14 - Terceira análise dos agrupamentos formados pelo método *k-means*.

Atributos	Variabilidade		Teste F	P valor
	Entre grupo	Dentro do grupo		
Área da bacia hidrográfica	8,872	0,679	13,07	$2,82e^{-05}$
Aridez	19,77	0,23	84,58	$2e^{-16}$
Declividade média da bacia	11,78	0,56	21,03	$2,54e^{-07}$
Evapotranspiração média anual	13,13	0,51	26,00	$2,01e^{-08}$
Frequência dos dias de precipitação máxima	14,79	0,44	33,85	$5,84e^{-10}$
Frequências dos dias secos	18,93	0,27	70,61	$3,7e^{-15}$
Longitude	10,65	0,61	17,56	$1,78e^{-06}$
Porosidade subterrânea da bacia hidrográfica	5,70	0,81	7,05	0,00
Precipitação média anual	20,85	0,19	109,90	$<2e^{-16}$
Profundidade mediana do lençol freático	17,28	0,34	51,48	$9,06e^{-13}$
Superfícies artificiais ou zonas urbanas	10,09	0,63	16,04	$4,38e^{-06}$
Teor de areia	12,11	0,55	22,15	$1,4e^{-07}$

* P-valor maior que o nível de significância de 0,05, ou seja, a variável não foi significativa.

Fonte: Próprio autor (2023).

Foi observado que não houve diferença nos valores apresentados na análise 2 e 3 para o método de *Ward*.

Tabela 15 - Estatística descritiva dos clusters formados pelos métodos *k-means* e *Ward* para a análise 3.

	Grupo	Estatística	Declividade médio da bacia	Área da bacia hidrográfica	Aridez	Superfícies artificiais ou zonas urbanas	Porosidade subterrânea da bacia hidrográfica	Profundidade mediana do lençol freático	Teor de areia	Frequência dos dias secos	Frequência dos dias de precipitação máxima	Longitude	Precipitação média anual	Evapotranspiração média anual
			mk ⁻¹	Km ²	-	%	-	cm	%	dia/ano	dia/ano	-	mm/dia	mm/dia
<i>k-means</i>	1	Desvio	15,15	15311,60	0,18	0,00	0,06	956,79	1,71	9,40	1,29	1,38	67,73	28,85
		Média	41,47	16620,77	1,70	0,00	0,11	3446,75	52,48	272,79	22,20	-39,51	698,04	1176,52
		Máximo	70,18	46547,88	2,04	0,00	0,21	4729,63	55,30	250,15	19,80	-37,35	883,93	1255,39
		Mínimo	24,02	665,01	1,30	0,00	0,04	1902,50	49,36	250,15	19,80	-41,57	563,66	1129,95
	2	Desvio	17,67	16969,16	0,12	0,01	0,06	953,50	2,57	11,48	1,44	0,82	90,15	43,29
		Média	57,59	12478,79	1,29	0,00	0,11	4182,68	47,34	268,20	22,09	-40,25	856,01	1136,05
		Máximo	108,35	68283,32	1,49	0,05	0,22	5845,68	51,19	242,00	18,35	-39,03	1014,42	1221,93
		Mínimo	26,25	355,44	1,08	0,00	0,01	1844,28	43,10	242,00	18,35	-41,61	774,17	1060,73
	3	Desvio	6,53	226621,73	0,08	0,05	0,02	416,69	3,18	16,40	3,07	2,46	77,77	36,56
		Média	24,85	205452,69	1,10	0,05	0,17	1274,77	51,84	225,72	17,10	-42,74	1095,23	1228,67
		Máximo	34,28	619185,70	1,28	0,13	0,19	1915,54	57,28	207,70	12,50	-37,76	1188,47	1307,55
		Mínimo	16,02	4703,06	0,96	0,00	0,14	289,41	47,12	207,70	12,50	-45,53	938,98	1178,25
<i>Ward</i>	1	Desvio	19,98	15450,89	0,28	0,01	0,06	1375,24	3,55	15,37	1,59	2,04	162,99	53,53
		Média	44,07	15640,73	1,45	0,00	0,12	3355,94	50,85	265,68	21,78	-40,59	824,55	1173,58
		Máximo	108,35	68283,32	2,04	0,05	0,22	5845,68	57,28	228,85	17,65	-37,35	1188,47	1307,55
		Mínimo	16,02	355,44	1,07	0,00	0,01	289,41	43,10	228,85	17,65	-45,53	563,66	1060,73
	2	Desvio	2,33	140857,23	0,12	0,02	0,00	127,55	1,51	0,99	1,00	2,32	91,40	22,62
		Média	30,81	419064,24	1,10	0,09	0,15	1369,32	48,96	209,67	14,21	-41,68	1067,87	1206,52
		Máximo	34,28	619185,70	1,28	0,13	0,16	1572,70	50,95	207,70	12,50	-37,76	1179,69	1230,67
		Mínimo	28,05	256995,20	0,96	0,07	0,14	1214,95	47,12	207,70	12,50	-43,76	938,98	1178,25
	3	Desvio	-	-	-	-	-	-	-	-	-	-	-	-
		Média	26,26	4703,06	1,06	0,11	0,17	1915,54	51,85	243,15	17,95	-38,07	1186,33	1190,10
		Máximo	26,26	4703,06	1,06	0,11	0,17	1915,54	51,85	243,15	17,95	-38,07	1186,33	1190,10
		Mínimo	26,26	4703,06	1,06	0,11	0,17	1915,54	51,85	243,15	17,95	-38,07	1186,33	1190,10

Fonte: Próprio autor (2023).

O método *k-means* apresentou melhor distribuição entre os grupos (Figura 22). A técnica de *Ward* obteve pouca ou nenhuma variação nas análises realizadas. Nas duas últimas análises o grupo 3 foi formado por apenas uma estação, podendo indicar que essa estação possui características diferentes das demais, exigindo uma análise mais aprofundada. Porém, esta configuração dificulta o processo de regionalização, tendo em vista que seria necessária mais de uma estação por grupo para realizar a transferência de dados no processo de regionalização.

Sendo assim, o método *k-means* foi o mais adequado para a separação dos clusters. Os atributos que auxiliaram foram a área da bacia hidrográfica, a aridez, a declividade média da bacia, a duração média dos eventos de precipitação máxima, a evapotranspiração média anual, a frequência dos dias de precipitação máxima, a frequências dos dias secos, a longitude, a porosidade subterrânea da bacia hidrográfica, a precipitação média anual, a profundidade mediana do lençol freático, as superfícies artificiais ou zonas urbanas e o teor de areia.

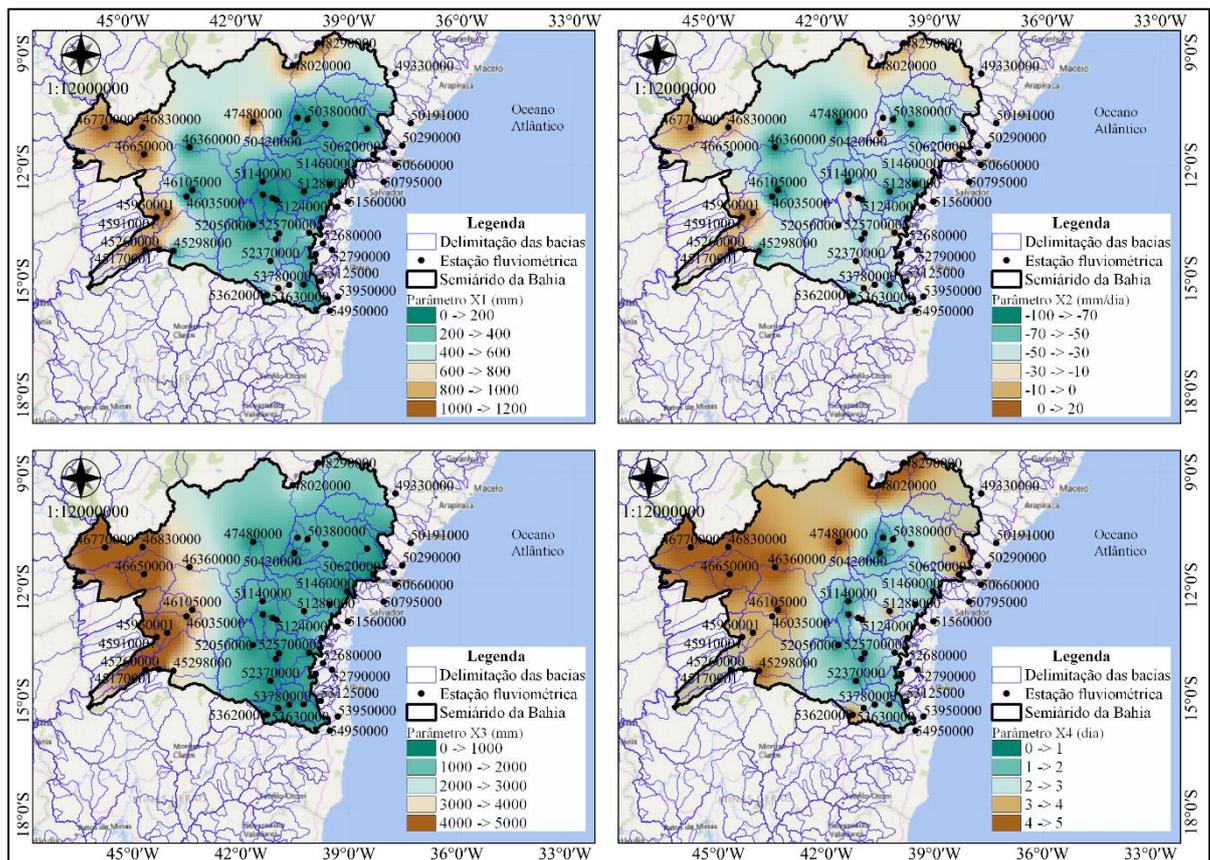
As características citadas, também foram determinadas como explicativas para o processo de regionalização de vazão, como citadas por Silva Júnior *et al.* (2003), em que os autores não descartaram a possibilidade de utilizar a precipitação, a densidade de drenagem e a declividade. Tucci (2017), o qual citou a urbanização como um fator de contribuição para o aumento da vazão máxima em razão da impermeabilização e a redução da cobertura de floresta e da evapotranspiração favoráveis ao aumento da vazão média. Já Kuana *et al.* (2020) selecionaram para análise a longitude de localização da estação fluviométrica, a área da bacia, a declividade média, o índice de aridez, a frequência e a duração de eventos de alta precipitação, a média mensal máxima do índice de área foliar, a condutividade hidráulica e a fração de areia e argila no solo. Tratando-se da regionalização das vazões mínimas para a bacia do Rio São Francisco, Lemos (2006) utilizou a área de drenagem, o comprimento do rio principal, a densidade de drenagem e a declividade média da bacia e do rio principal, extraídas a partir de modelo digital de elevação hidrologicamente consistente, a precipitação total anual e as precipitações do semestre e trimestre mais secos.

5.2. ANÁLISE DO DESEMPENHO DO MODELO GR4J

Após a aplicação do GR4J foram obtidos o melhor conjunto dos parâmetros do modelo. A especialização dos parâmetros é possível de ser visualizada na Figura 24. Nota-se que os valores mais altos de X1, X2, X3 e X4 se encontram no Oeste do semiárido baiano. O local faz parte do Sertão. Esta sub-região possui solos rasos e pedregosos, com chuvas escassas e mal distribuídas (em torno 400 a 600 mm por ano) e evapotranspiração anual em torno de 2.000 mm

por ano, caracterizando o polígono das secas (EMBRAPA, 2021; EMBRAPA, 2022). As principais formações hidrogeológicas são o Sistema Aquífero Urucuia (SAU) e o Grupo Bambuí (Unidade Terrígena). O aquífero Urucuia é um dos responsáveis para a manutenção da vazão do Rio São Francisco (BARBOSA, 2017; CUNHA, 2017; GASPAR, 2006; GASPAR; CAMPOS, 2007; GONÇALVES; ENGELBRECHT; CHANG, 2017; SGB-CPRM, 2022). A elevação do terreno varia entre 614 à 776 metros e o tipo de solo predominante é o latossolo amarelo (ver Figura 12, Figura 13 e Figura 14).

Figura 24 - Espacialização do melhor conjunto de parâmetros do modelo GR4J para o semiárido da Bahia.



Fonte: Próprio autor (2023).

Ressalta-se que, o melhor conjunto de parâmetros são os valores mais altos das medidas de desempenho que o modelo conseguiu gerar para cada estação. Não necessariamente significa dizer que as estações geraram valores significativos. Portanto, após a geração dos parâmetros do modelo e das vazões simuladas, foi obtido e avaliado o resumo dos parâmetros e as métricas estatísticas para as vazões mínimas LOGQ, REMQ_{BV} e Nslog (Tabela 16).

Tabela 16 - Melhor conjunto de parâmetros e critérios de eficiência do modelo GR4J para o período de calibração e validação para as vazões mínimas.

Código da estação	Melhor conjunto de parâmetro				Calibração (1986 - 2005)			Validação (2011 - 2018)		
	X1	X2	X3	X4	LOGQ	REMQ _{BV}	Nslog	LOGQ	REMQ _{BV}	Nslog
45170001	1200,00	17,91	5000,00	2,58	0,06	0,24	0,50	0,02	0,21	0,83
45260000	1200,00	-5,96	5000,00	2,81	0,21	1,22	0,27	0,05	0,95	0,70
45910001	1200,00	20,63	5000,00	2,92	1,43	43,49	-5,10	0,72	32,30	-3,21
45960001	1198,90	15,33	5000,00	3,53	0,01	0,80	0,78	0,00	0,38	0,96
46650000	1200,00	-48,56	5000,00	4,85	0,00	0,46	0,81	0,11	1,48	0,70
46770000	1200,00	14,23	5000,00	4,45	3,66	33,49	-21,28	3,39	30,00	-30,09
46830000	1200,00	-3,67	5000,00	5,00	1,78	35,92	-9,92	2,23	29,64	-17,32
51120000	0,00	-14,48	213,33	0,83	4,36	40,25	-20,00	3,33	25,72	-103,02
47480000	1200,00	-97,22	314,53	4,57	5,23	1,07	-37,93	7,10	1,70	-158,97
50191000	13,89	-30,33	130,47	1,35	0,38	0,08	-7,21	0,38	0,17	-12,95
50250000	329,39	-0,93	4,72	4,54	8,86	0,38	-48,32	8,70	1,48	-363,82
50290000	113,14	-0,33	61,60	2,33	0,05	0,02	0,91			
50330000	24,89	-31,25	737,86	0,80		21,27			25,29	
50380000	46,32	-85,25	859,74	2,00	5,16	22,54	-1053,28	5,00	15,25	-383,70
50420000	110,17	-18,27	231,77	0,68	6,75	26,24	-479,09	5,52	12,46	-123,07
50465000	99,81	-90,90	840,14	2,05	0,94	0,52	-18,67	0,61	0,32	-3,71
50540000	60,88	-90,14	594,23	3,12	5,26	0,64	-404,25	8,78	0,58	-2,52
50595000	40,66	-85,79	760,92	4,52	3,00	0,95	-9,45	2,19	0,82	-6,30
50620000	691,19	-1,74	8,22	3,53	0,02	0,00	0,99			
50660000	48,44	-56,00	706,90	2,05	0,04	0,08	-0,06	0,03	0,07	-1,58
51140000	37,79	-94,37	852,39	1,00	1,64	0,72	-2,32	1,05	0,24	-1226,06
51230000	39,97	-1,87	191,00	2,56	0,02	0,01	0,99	0,25	0,18	-115,84
51240000	0,04	-98,08	793,01	3,10	0,48	1,92	-1,72	1,95	1,89	-224,15
51280000	10,75	-93,22	890,29	3,15	0,42	1,80	-1,26	0,46	1,02	-12,24
51350000	14,45	-99,75	725,18	2,33	4,04	2,65	-2933,55	3,01	1,02	-75,44
51460000	146,71	-22,82	51,98	1,41	0,00	0,02	0,17	0,02	0,01	0,96
51560000	0,00	-38,51	306,84	0,97	0,70	0,31	-37,63	0,04	0,04	0,65
51685000	91,58	-68,75	1472,80	1,93	0,05	0,26	0,99	0,01	0,04	0,83
51840000	65,17	-77,82	1843,36	2,96	0,63	0,33	-1,01	1,73	0,17	-82,64
52050000	34,98	-37,76	196,62	0,96	0,01	0,04	0,90	0,19	0,05	-0,53
52370000	268,74	-35,27	123,82	0,98	0,11	0,75	0,85	0,44	0,45	-720,35
52404000	241,78	-60,82	216,08	1,27	0,00	1,01	0,60	0,70	0,59	-136,79
52405000	195,80	-66,80	256,37	1,18	0,00	0,17	0,89	0,54	0,75	-28,06
52570000	496,61	-9,72	115,91	2,67	5,04	5,34	-200,63	2,75	0,36	-0,69
52680000	106,21	-97,89	798,14	3,03	3,88	3,88	-101,86	1,40	0,50	-3,38
52760000	20,52	-29,71	621,30	1,03	0,17	0,47	-0,14	0,72	0,37	-11,33
52790000	0,00	-36,08	394,06	1,40	1,78	0,62	-4,88	4,97	1,23	-546,81
53125000	129,98	-0,27	10,71	0,80		1,78		0,01	0,01	-511,47

Código da estação	Melhor conjunto de parâmetro				Calibração (1986 - 2005)			Validação (2011 - 2018)		
	X1	X2	X3	X4	LOGQ	REMQ _{Bv}	Nslog	LOGQ	REMQ _{Bv}	Nslog
53620000	226,57	-83,57	411,63	3,96	0,11	1,28	0,99	1,91	0,40	-5,34
53630000	913,22	-21,50	166,93	2,59	0,09	1,57	0,87	0,28	0,23	0,85
53650000	619,99	-31,35	251,56	2,19	0,03	0,85	0,93	0,16	0,46	-0,04
53780000	0,00	-86,32	674,60	1,92	1,31	4,71	0,30	0,71	0,56	-53,37
53950000	28,95	-92,64	825,82	2,00	0,03	1,65	0,99	1,37	2,38	-3,56
54950000	9,82	-64,13	818,50	2,10	4,31	26,11	-99,92	4,23	19,45	-1416,72
45298000	25,55	-96,59	2741,11	4,40	3,62	41,86	-11,93	8,21	92,43	-271,41
46035000	75,85	-92,93	3866,60	2,71	5,69	41,48	-0,65			
46105000	46,16	-99,09	3437,23	4,82	1,68	40,44	0,52	6,63	81,35	-40,38
46360000	53,86	-95,48	3284,63	4,84	2,37	35,92	0,09	6,67	77,85	-30,70
48020000	1155,11	-1,65	2010,50	4,97	1,86	26,98	0,40	3,81	36,10	-32,42
48290000	1198,05	1,45	1771,88	4,95	4,47	47,70	-3,79	5,94	36,56	-1167,39
49330000	835,51	3,37	1916,53	2,92	1,32	35,12	-0,03	1,55	32,26	-5,93
50795000	4,48	-71,99	656,71	2,30	2,55	14,70	-5,16		2,25	

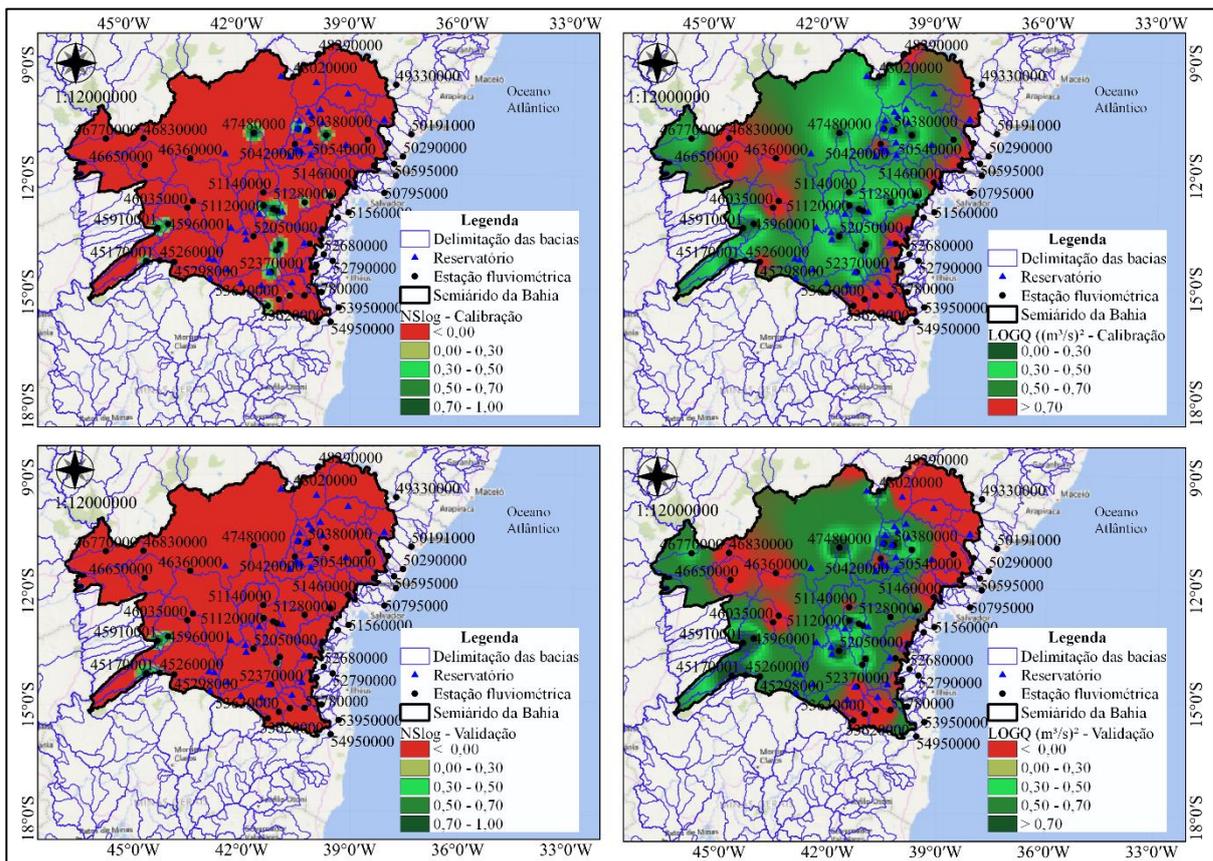
Fonte: Próprio autor (2023).

A partir do cálculo do coeficiente Nslog foi possível verificar que, na fase de calibração, o desempenho do modelo GR4J variou entre -2933,5 e 0,99 e 13 (25%) estações tiveram um desempenho acima de 70%. Ressalta-se que todas as estações que tiveram desempenho superior a 70% considerando o Nslog, também tiveram desempenho positivo para o coeficiente LOGQ, variando entre 0,00 e 0,11, sendo 0 o valor ótimo. A espacialização dos resultados pode ser visualizada na Figura 25. As estações com as melhores simulações na fase de calibração foram: 50620000 (0,99), 51230000 (0,99), 51685000 (0,99), 53950000 (0,99), 53620000 (0,99), 53650000 (0,93), 50290000 (0,91), 52050000 (0,90), 52405000 (0,89), 53630000 (0,87), 52370000 (0,85), 46650000 (0,81), 45960001 (0,78). Os tipos de solo identificado nas áreas foram Latossolo (Vermelho, Amarelo, Vermelho Amarelo), Planossolo Háplico, Argissolo Vermelho Amarelo, Cambissolo Háplico e as formações hidrogeológicas foram Embasamento Fraturado Indiferenciado, Depósito Aluvionar e Formação Caatinga (ver Figura 12 e Figura 14).

Na fase de validação o Nslog variou entre -1416,72 e 0,96 e 7 (13%) na fase de validação tiveram um desempenho acima de 70%. Comparativamente, Kuana (2022) obteve que 65% das bacias paranaenses no período da validação obtiveram um Nslog superior a 0,70. A autora ressalta que algumas bacias hidrográficas apresentaram desempenhos superiores no período de validação quando comparadas com o período de calibração, contudo situações inversas também ocorrem, esse comportamento também foi observado nas bacias do semiárido da Bahia. Kuana

(2022) explica que esses fenômenos podem estar relacionados a mudança ou a melhoria das técnicas e equipamentos de medição, como também podem ter sofrido influências das alterações do uso e cobertura do solo. Já na fase de validação as estações com melhores desempenho foram: 45960001 (0,96), 51460000 (0,96), 51685000 (0,83), 45170001 (0,83), 53630000 (0,85), 45260000 (0,70), 46650000 (0,70). Os tipos de solo identificado nas áreas foram Latossolo Amarelo, Argissolo vermelho amarelo, Cambissolo Háplico e as formações hidrogeológicas foram Embasamento Fraturado Indiferenciado, Grupo Bambuí Unidade Carbonática e Depósito Aluvionar ver Figura 12 e Figura 14).

Figura 25 - Espacialização dos coeficientes Nslog e LOGQ para os períodos de calibração e validação do modelo na área de estudo.

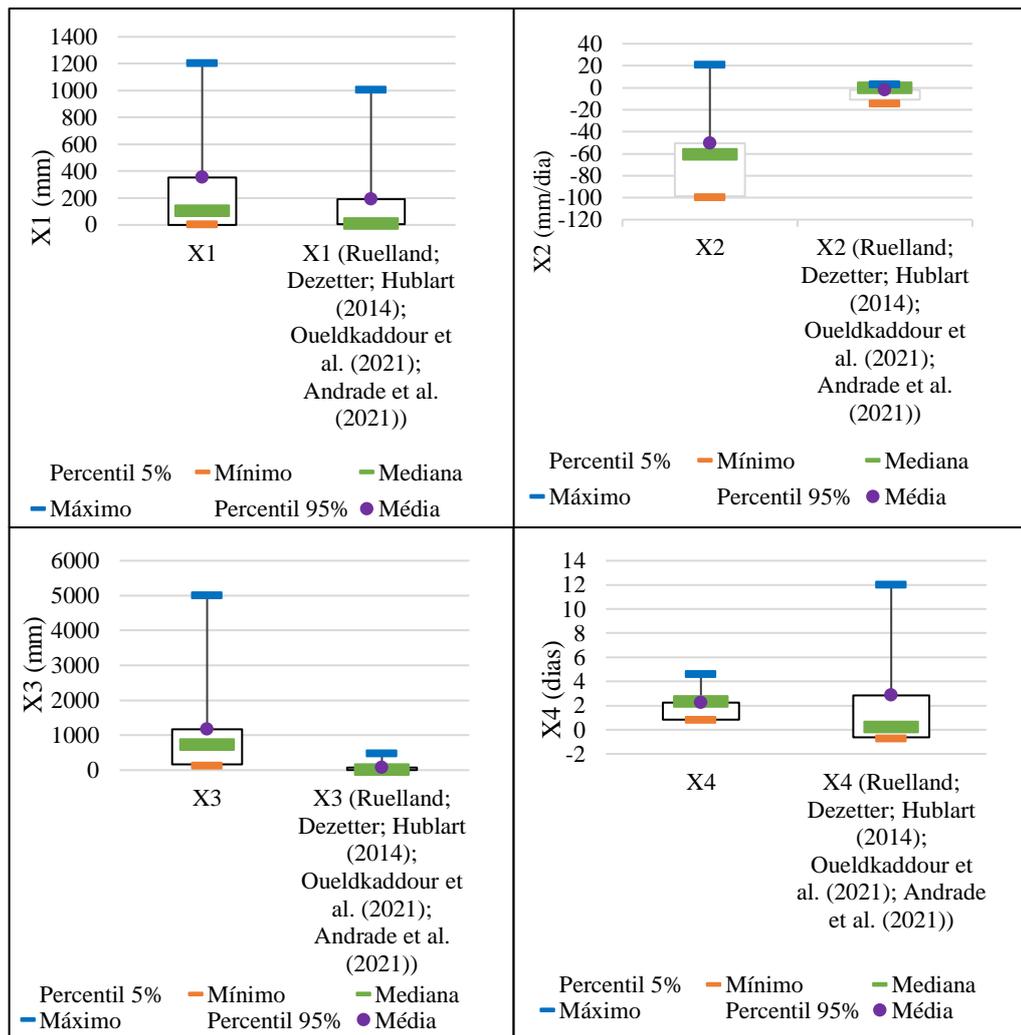


Fonte: Próprio autor (2023).

Um fator que pode ter contribuído para o desempenho do modelo é a presença de 33 reservatórios monitorados pela Agência Nacional de Águas e Saneamento Básico (ANA) distribuídos na área de estudo (ver Figura 17 e Figura 25). Os reservatórios auxiliam no processo de regularização de vazão, podendo acarretar mudanças hidrológicas a jusante do curso d'água. A modificação da vazão pela presença dos reservatórios, pode não ter sido captada pelo modelo matemático GR4J.

Para contextualizar os resultados, estes foram comparados com os obtidos por Ruelland; Dezetter; Hublart (2014), Oueldkaddour *et al.* (2021) e Andrade *et al.* (2021) em bacias semiáridas localizadas no Marrocos, Chile (Rivadavia, Varillar, Riva. & Var) e Brasil (Capibaribe em Pernambuco, Pindaré no Maranhão, Salgado no Ceará, Sirinhaém em Pernambuco e Piauitinga em Sergipe) (Figura 26). Ressalta-se que foram utilizados dados das estações que obtiveram o valor de N_{slog} igual ou maior que 0,70.

Figura 26 - Resumo comparativo dos parâmetros do modelo GR4J.



Fonte: Próprio autor (2023).

Tratando-se do parâmetro X1, o valor mínimo encontrado na bibliografia foi de 4,901 mm para a bacia do Rio Salgado no Ceará, descrito por Andrade *et al.* (2021) e o valor máximo foi de 1005 mm, encontrado por Ruelland, Dezetter e Hublart (2014) em Varillar (Chile). Dentre as estações que tiveram o coeficiente N_{slog} superior a 70%, X1 apresentou valor mínimo de 0 mm e o máximo 1200 mm, indicando uma maior capacidade do reservatório de

produção. As estações que apresentaram os maiores valores de X1 foram 45910001 (1200 mm), 47480000 (1200 mm), 45960001 (1198,9 mm), elas possuem uma área de 29484,01 km², 9270,73 km², 30996,37 km², respectivamente. Não foi identificada correlação entre o parâmetro X1 e a área. A estação 45910001 encontra-se sobre o Grupo Bambuí/Unidade Terrígena e o tipo de solo predominante é o Latossolo Amarelo, a estação 45960001 se encontra sobre o Depósito Aluvionar e o tipo de solo é o Planossolo Hidromórfico e a estação 47480000 está sobre a Formação Caatinga e o tipo de solo é o Cambissolo Háplico. As estações que tiveram menores valores de X1 foram: 51120000 (0 mm) e 51240000 (0,04 mm) as quais possuem área de 2344,33 km² e 16417,18 km², respectivamente. O tipo de solo predominante é o Latossolo Amarelo e estão sobre o Embasamento Fraturado Indiferenciado.

Para o parâmetro X2, o valor mínimo encontrado na bibliografia foi -14,893 mm/dia, em uma bacia localizada no Marrocos, como descrito por Ouelkaddour *et al.* (2021) e o valor máximo foi de 2,700 mm/dia, encontrado por Ruelland, Dezetter e Hublart (2014) em Rivadavia (Chile). Dentre as estações que tiveram o coeficiente Nslog superior a 70%, o valor mínimo para X2 foi -99,75 mm e o valor máximo foi de 20,65 mm/dia, indicando que há uma variação em relação à troca de água, com algumas bacias apresentando infiltração de água no solo superior à contribuição subterrânea (troca negativa) e em outras a troca subterrânea é superior (troca positiva). Como esperado, as estações que apresentaram troca positiva estão localizadas na região hidrográfica do Rio São Francisco. As estações que apresentaram maiores valores de X2 foram 45910001 (20,65 mm/dia), a qual o tipo de solo predominante é o Cambissolo Háplico, 45960001 (15,33 mm/dia), a qual possui o tipo de solo predominante é o Latossolo Amarelo. Ressalta-se que ambas as estações estão sobre o Depósito Aluvionar, próximo ao Sistema Aquífero Urucuia (SAU). O SAU corresponde a um dos principais aquíferos contribuintes para a vazão do Rio São Francisco (BARBOSA, 2017; CUNHA, 2017; GASPAR, 2006; GASPAR; CAMPOS, 2007; GONÇALVES; ENGELBRECHT; CHANG, 2017; SGB-CPRM, 2022). A estação 47480000 pertence à região hidrográfica do São Francisco, porém, está sobre a Formação Caatinga. As demais estações estão sobre o Embasamento Fraturado Diferenciado e pertencem à região hidrográfica do Atlântico Leste.

Em relação ao parâmetro X3, o valor mínimo encontrado entre os autores citados acima foi 2,090 mm para a bacia do Rio Capibaribe em Pernambuco, como relatado por Andrade *et al.* (2021) e o valor máximo foi de 473 mm, encontrado por Ruelland, Dezetter e Hublart (2014) em Rivadavia (Chile). Neste estudo, X3 apresentou valor mínimo de 123,83 mm e o máximo 5000 mm, indicando que alta capacidade máxima de armazenamento de água no aquífero subterrâneo das bacias avaliadas. As estações que apresentaram altos valores de X3 foram

45910001 e 45960001, o tipo de solo predominante é Cambissolo Háplico e o Latossolo Amarelo, respectivamente. Ambas as estações estão sobre o Depósito Aluvionar, próximo ao aquífero Urucuia e estão situadas em terrenos com elevação média entre 617 e 776 metros.

Por fim, o parâmetro X4 obteve valor mínimo encontrado entre os autores citados acima de -0,444 dia para a bacia do Rio Salgado no Ceará, descrito por Andrade *et al.* (2021) e o valor máximo foi encontrado por Ruelland, Dezetter e Hublart (2014) em Varillar (Chile) (12 dias). Dentre as estações que tiveram o coeficiente Nslog superior a 70%, X4 apresentou valor mínimo de 0,80 dias e o máximo 4,57 dias. As estações que apresentaram os maiores valores de X4 foram 47480000 (4,57 dias) e 45960001 (3,53 dias). Não foi identificadas relações entre a área, tipo de solo e hidrogeologia das estações. Ressalta-se que as bacias modeladas por Andrade apresentaram valores de X4 inferiores a 0,5 dias. Conforme Perrin *et al.* (2003), todos os parâmetros possuem valores reais. Os parâmetros X1 e X3 são positivos, X4 deve ser maior que 0,5 e X2 pode ser negativo, positivo ou zero.

Bodian *et al.* (2018) investigaram o efeito das mudanças climáticas nas bacias hidrográficas dos rios Senegal e Gâmbia, localizadas no Senegal, na África Ocidental. Para isso, os autores usaram dados projetados de precipitação em escala reduzida e evapotranspiração potencial com base na temperatura projetada de seis Modelos de Circulação Geral e dois cenários para forçar o modelo GR4J. A eficiência do modelo foi avaliada usando uma função multicritério que agrega os critérios de NSE, erro de volume cumulativo e erro de volume médio. Os resultados mostraram que o GR4J apresentou resultados ruins para simular as vazões mínimas. Segundo os autores, o modelo apresenta algumas dificuldades para simular a vazão de base porque não leva em conta totalmente as relações entre as águas subterrâneas superficiais. Os autores também pontuaram sobre a dificuldade em aplicar o modelo nas bacias hidrográficas do alto Senegal e da Gâmbia devido à complexidade do contexto hidrogeológico, bem como à escassez de dados hidrogeológicos. Além disso, o GR4J não leva em consideração a dinâmica do uso do solo na simulação da vazão e qualquer mudança no uso do solo no futuro pode impactar no escoamento e o fato do modelo não considerar o ciclo integrado atmosfera-água-vegetação teria introduzido uma incerteza significativa nos resultados da modelagem, sugerindo que, para pesquisas futuras, seria interessante utilizar modelos hidrológicos que integrassem a dinâmica do uso da terra, como o SWAT (BODIAN *et al.*, 2018).

Também foram avaliados os parâmetros do modelo e as métricas estatísticas para as vazões máximas RMSE, NSE, R² e KGE (Tabela 17).

Tabela 17 - Melhor conjunto de parâmetros e critérios de eficiência do modelo GR4J para o período de calibração e validação para as vazões máximas.

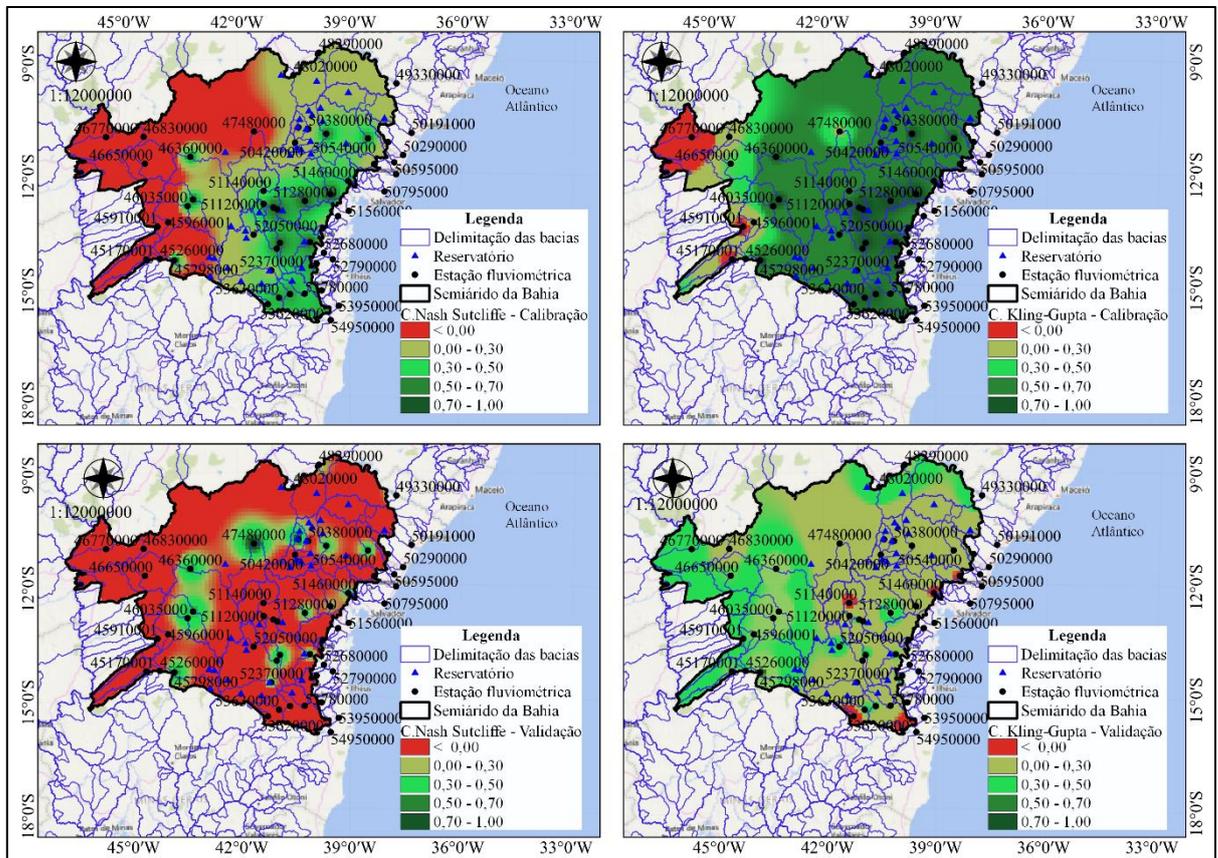
Código da estação	Melhor conjunto de parâmetro				Calibração (1986 - 2005)					Validação (2011 - 2018)				
	X1	X2	X3	X4	RMSE	PBIAS	NSE	R2	KGE	RMSE	PBIAS	NSE	R2	KGE
45170001	1200,00	17,91	5000,00	2,58	0,48	29,90	-3,51	0,33	-0,28	0,21	-20,50	-0,62	0,37	0,51
45260000	1200,00	-5,96	5000,00	2,81	0,34	9,10	-0,42	0,35	0,39	0,23	-33,80	-0,31	0,36	0,42
45910001	1200,00	20,63	5000,00	2,92	0,36	28,00	-2,95	0,31	-0,24	0,17	13,40	-0,40	0,33	0,46
45960001	1198,90	15,33	5000,00	3,53	0,30	17,80	-1,26	0,33	0,14	0,16	1,90	-0,05	0,28	0,52
46650000	1200,00	-48,56	5000,00	4,85	0,19	13,50	-0,71	0,23	0,36	0,16	-34,80	-0,56	0,28	0,31
46770000	1200,00	14,23	5000,00	4,45	0,36	54,70	-16,13	0,05	-1,46	0,11	-15,00	-0,71	0,15	0,37
46830000	1200,00	-3,67	5000,00	5,00	0,18	11,10	-0,91	0,21	0,29	0,24	-50,30	-2,45	0,51	0,15
51120000	0,00	-14,48	213,33	0,83	0,97	0,90	0,44	0,52	0,72	1,40	45,80	-0,05	0,23	0,30
47480000	1200,00	-97,22	314,53	4,57	0,02	19,70	-0,70	0,28	0,28	-	-	-	-	-
50191000	13,89	-30,33	130,47	1,35	0,36	-3,80	-0,21	0,12	0,33	0,40	37,70	-4,56	0,01	-0,57
50250000	329,39	-0,93	4,72	4,54	0,32	1,00	-0,08	0,24	0,49	0,16	-0,70	-0,10	0,15	0,38
50290000	113,14	-0,33	61,60	2,33	0,59	-0,50	0,29	0,41	0,64	0,31	79,50	-0,01	0,39	0,10
50330000	24,89	-31,25	737,86	0,80	0,54	0,30	-0,05	0,25	0,50	-	-	-	-	-
50380000	46,32	-85,25	859,74	2,00	0,25	2,60	0,24	0,43	0,64	0,23	58,90	-0,28	0,09	0,09
50420000	110,17	-18,27	231,77	0,68	0,60	0,00	0,01	0,27	0,52	0,20	37,70	-0,58	0,27	0,25
50465000	99,81	-90,90	840,14	2,05	0,16	1,60	0,57	0,63	0,79	0,15	82,60	0,29	0,31	0,00
50540000	60,88	-90,14	594,23	3,12	0,09	1,30	0,54	0,61	0,78	0,09	74,20	0,42	0,43	0,13
50595000	40,66	-85,79	760,92	4,52	0,09	1,40	0,54	0,61	0,77	0,07	9,60	0,20	0,30	0,52
50620000	691,19	-1,74	8,22	3,53	0,24	-0,50	0,03	0,26	0,51	0,06	-68,60	-1,44	0,01	-0,15
50660000	48,44	-56,00	706,90	2,05	0,22	2,10	0,32	0,47	0,68	0,12	18,30	0,27	0,33	0,50
51140000	37,79	-94,37	852,39	1,00	0,25	4,50	0,32	0,49	0,67	0,29	68,50	-1,66	0,37	-0,26
51230000	39,97	-1,87	191,00	2,56	0,91	1,30	0,68	0,71	0,84	1,02	29,30	0,04	0,41	0,48
51240000	0,04	-98,08	793,01	3,10	0,28	3,60	0,63	0,70	0,81	0,35	90,10	-0,33	0,53	-0,07
51280000	10,75	-93,22	890,29	3,15	0,25	-0,60	0,66	0,70	0,83	0,28	26,90	0,31	0,44	0,57
51350000	14,45	-99,75	725,18	2,33	0,20	2,70	0,72	0,76	0,85	0,20	63,20	0,11	0,47	0,25
51460000	146,71	-22,82	51,98	1,41	0,33	3,20	0,42	0,50	0,71	1,01	-75,00	0,25	0,64	-0,13
51560000	0,00	-38,51	306,84	0,97	0,73	3,00	0,12	0,35	0,59	0,49	42,10	0,35	0,44	0,45
51685000	91,58	-68,75	1472,80	1,93	0,18	-0,10	0,29	0,42	0,65	0,12	23,40	-0,49	0,09	0,26
51840000	65,17	-77,82	1843,36	2,96	0,22	1,10	0,19	0,36	0,60	0,15	59,50	-0,76	0,07	0,06
52050000	34,98	-37,76	196,62	0,96	0,48	2,70	0,22	0,40	0,63	0,45	-17,00	-0,33	0,23	0,41
52370000	268,74	-35,27	123,82	0,98	0,37	-2,70	0,19	0,30	0,54	0,17	-40,70	0,28	0,33	0,13
52404000	241,78	-60,82	216,08	1,27	0,16	-1,90	0,53	0,57	0,74	0,17	-60,90	0,34	0,47	0,06
52405000	195,80	-66,80	256,37	1,18	0,14	-0,30	0,62	0,65	0,80	0,14	-9,70	0,51	0,58	0,43
52570000	496,61	-9,72	115,91	2,67	0,16	-1,00	0,63	0,66	0,81	0,08	7,70	0,35	0,38	0,33
52680000	106,21	-97,89	798,14	3,03	0,12	1,90	0,50	0,60	0,75	0,11	-27,00	-0,41	0,17	0,33
52760000	20,52	-29,71	621,30	1,03	0,54	0,40	0,52	0,60	0,77	0,34	-15,50	-0,05	0,22	0,44
52790000	0,00	-36,08	394,06	1,40	0,61	-1,00	0,62	0,65	0,81	0,48	63,70	-0,73	0,28	0,09
53125000	129,98	-0,27	10,71	0,80	2,06	-9,00	0,48	0,49	0,61	-	-	-	-	-
53620000	226,57	-83,57	411,63	3,96	0,19	-0,80	0,45	0,54	0,73	0,11	51,40	-4,32	0,05	-0,59
53630000	913,22	-21,50	166,93	2,59	0,10	-1,60	0,47	0,57	0,75	0,03	10,70	-0,07	0,24	0,48
53650000	619,99	-31,35	251,56	2,19	0,12	-0,70	0,53	0,60	0,77	0,04	-14,10	0,10	0,19	0,36
53780000	0,00	-86,32	674,60	1,92	0,41	1,10	0,40	0,53	0,71	0,40	27,30	-1,56	0,10	0,06
53950000	28,95	-92,64	825,82	2,00	0,28	2,40	0,41	0,55	0,72	0,21	56,70	-1,73	0,16	-0,10
54950000	9,82	-64,13	818,50	2,10	0,33	-0,50	0,67	0,70	0,84	0,31	227,20	-10,81	0,10	-1,49
45298000	25,55	-96,59	2741,11	4,40	0,28	1,70	0,70	0,73	0,85	0,22	4,80	0,69	0,70	0,80

Código da estação	Melhor conjunto de parâmetro				Calibração (1986 - 2005)					Validação (2011 - 2018)				
	X1	X2	X3	X4	RMSE	PBIAS	NSE	R2	KGE	RMSE	PBIAS	NSE	R2	KGE
46035000	75,85	-92,93	3866,60	2,71	0,28	2,80	0,50	0,59	0,76	0,24	-5,00	0,48	0,50	0,65
46105000	46,16	-99,09	3437,23	4,82	0,25	3,70	0,60	0,67	0,80	0,20	4,00	0,60	0,60	0,71
46360000	53,86	-95,48	3284,63	4,84	0,24	5,00	0,55	0,65	0,77	0,18	-5,10	0,58	0,59	0,72
48020000	1155,11	-1,65	2010,50	4,97	0,14	-0,60	0,24	0,41	0,63	0,12	-36,90	-0,78	0,48	0,51
48290000	1198,05	1,45	1771,88	4,95	0,14	1,90	0,23	0,41	0,63	0,08	-32,10	0,12	0,64	0,57
49330000	835,51	3,37	1916,53	2,92	0,12	0,50	0,22	0,39	0,62	0,06	-27,00	0,31	0,64	0,60
50795000	4,48	-71,99	656,71	2,30	0,41	3,30	0,58	0,64	0,79	0,20	27,50	0,43	0,55	0,62

Fonte: Próprio autor (2023).

Para uma melhor visualização, foram espacializadas as métricas NSE e KGE na área de estudo (Figura 27).

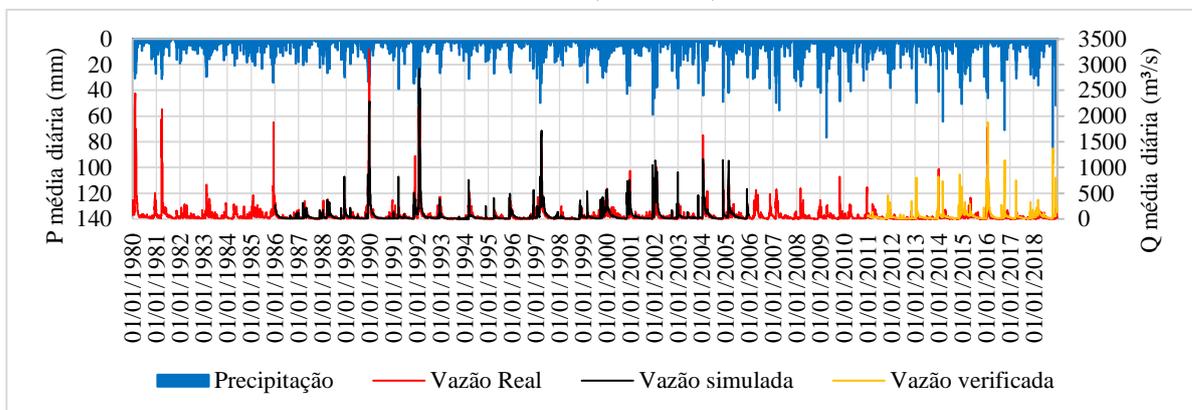
Figura 27 -Espacialização dos coeficientes NSE e KGE para os períodos de calibração e validação do modelo.



Fonte: Próprio autor (2023).

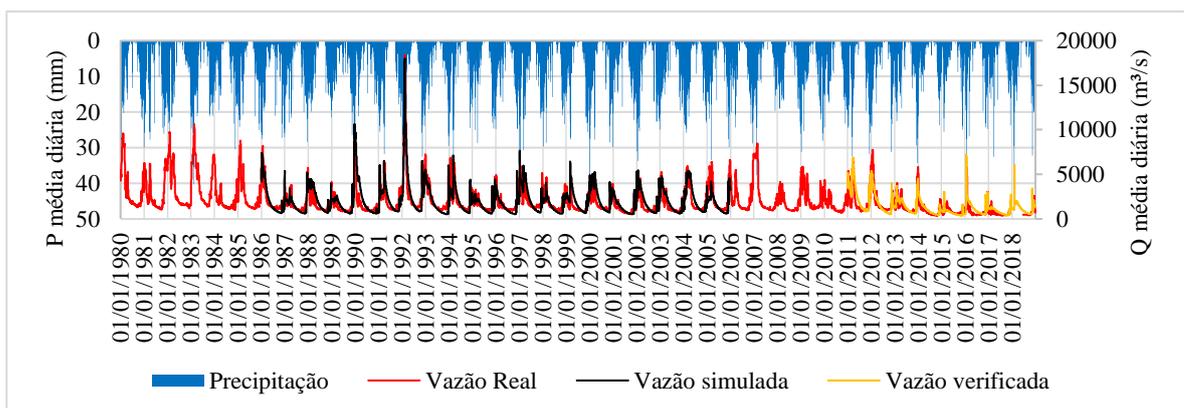
As métricas devem ser analisadas em conjunto, tendo em vista que estão intimamente relacionadas. O *NSE* dá uma ideia se o modelo simulado é melhor ou não quando comparado ao modelo com dados observados com (MURPHY, 1988). Nesse sentido, foi observado que o *NSE* médio das estações foi de -0,17 para a fase de calibração e de -0,59 na fase de validação, indicando que a soma dos quadrados dos resíduos do modelo previsto foi maior do que a do modelo de referência, o que significa que o modelo é pior do que o modelo de referência. Nas Figura 28 e Figura 29 constam os hidrogramas das vazões simuladas e observadas e hietogramas da precipitação nas estações 51350000 ($NSE=0,72$) e 55298000 ($NSE=0,70$), respectivamente. No primeiro caso é possível visualizar que o modelo conseguiu simular as vazões mínimas, porém apresentou valores de picos divergentes das vazões observadas para os anos de 1986, 1990, 1992 e 2004. No segundo caso é possível visualizar que o modelo conseguiu simular as vazões mínimas, porém na estação 51350000 os resultados foram mais satisfatórios, os demais hidrogramas podem ser avaliados no APÊNDICE C.

Figura 28 - Hidrograma da vazão simulada e observada e hietograma da precipitação na estação 51350000 ($NSE=0,72$).



Fonte: Próprio autor (2023).

Figura 29 - Hidrograma da vazão simulada e observada e hietograma da precipitação na estação 45298000 ($NSE=0,70$).



Próprio autor (2023).

Araújo (2005) obteve o valor do coeficiente de Nash-Sutcliffe acima 97%, quando o modelo foi aplicado a bacia do Rio Paraíba, em Caraubas para o período de 1996-1999, porém quando o modelo foi ajustado para a bacia do Rio do Peixe os valores de NSE foram inferiores a 70%. Evidenciando que o modelo pode apresentar resultados satisfatórios em bacias localizadas no semiárido.

O R^2 médio foi de 0,48 na fase de calibração e 0,34 na fase de validação, proveniente de análise de regressão linear, teve como objetivo indicar o grau de relação linear entre duas variáveis, nesse caso as vazões simuladas e observadas. Já o KGE foi de 0,58 na fase de calibração e de 0,25 na fase de validação, foram superiores quando comparados com o NSE.

Comparando com os valores obtidos por Kuana *et al.* (2021), o modelo GR4J alcançou valores medianos de KGE modificado para o período de calibração igual a 0,77 e 0,74 para a primeira e a segunda fase, respectivamente. Já no período de validação atingiu valores medianos de NSE na primeira fase igual a 0,67, e na segunda fase obteve valores iguais a 0,63. Evidenciando que o modelo obteve um melhor ajuste quando aplicado ao estado do Paraná.

Em um estudo realizado por Neto, Vieira e Matosinhos (2021), o modelo GR4J teve seu desempenho prejudicado na representação da simulação dos períodos de pico de chuva. Os autores explicaram que foi devido aos períodos de alta variação climática, com elevada pluviosidade na área de estudo, que contribuem aos processos hidrológicos da bacia e devido ao comportamento dos modelos em situações de altas vazões. Após a análise dos hidrogramas e dos valores de NSE e KGE gerados, os autores observaram também que o modelo GR4J se mostrou eficaz na modelagem de bacias de baixo escoamento, como a bacia do Ribeirão São João em Minas Gerais. A mudança e variabilidade climática podem interferir no sistema hidrológico da bacia de estudo. Os autores sugeriram, novas simulações considerando os períodos de calibração e validação, diferentes dos avaliados, com o intuito de obter uma maior compreensão do comportamento da bacia estudada.

Ao avaliar o Pbias, foi observado um valor médio de 4,16 na fase de calibração e de 15,38 na fase de validação nas estações localizadas no semiárido da Bahia. Na fase de calibração do modelo GR4J, Andrade *et al.* (2021) obteve um viés de 17,30 para bacia do rio Capibaribe, concluindo que houve uma leve tendência da série sintética em superestimar os dados observados, com base no baixo valor positivo quantificado. Em contrapartida, verificou-se uma leve subestimação na etapa de validação das vazões sintéticas da bacia do rio Salgado, pelo baixo valor obtido (-21,40). Em comparação aos resultados obtidos por Andrade *et al.* (2021), os valores dos Pbias obtidos no presente estudo foram melhores, tendo em vista que quanto

mais próximo de 0, melhor. Em relação à raiz do erro médio quadrático (RMSE), foram observados os valores médios de 0,34 na fase de calibração e 0,26 na fase de validação. Ao aplicar o modelo nas bacias do rio Capibaribe e do rio Piautinga, Andrade *et al.* (2021) obtiveram os valores de 9,94 e 5,08 na validação, respectivamente.

De forma geral, o modelo GR4J não apresentou resultados satisfatórios, como é possível visualizar nos hidrogramas presentes no APÊNDICE C. Conforme com Araújo (2005), o solo da região semiárida geralmente é pouco profundo, evidenciando que este possa ter sido o motivo para que houvesse resultados contrastantes quanto a aplicação do modelo na bacia do Rio Paraíba em Caraúbas, região do Cariri. A autora afirma que, do ponto de vista hidrológico, o solo pode ser entendido como um reservatório, cujo volume de água armazenado pode ser bastante variável no tempo e no espaço, que ao calcular todas as entradas e saídas do sistema resolve-se o balanço hídrico no solo. Os solos rasos foram identificados na sub-região do Sertão, porém apenas 11 (21,15%) estações estão localizadas no Sertão.

Uma vez que os resultados mostraram que, embora, de forma geral, o modelo não tenha apresentado um bom ajuste nas bacias analisadas, ele conseguiu representar as vazões mínimas.

5.3. REGIONALIZAÇÃO DE PARÂMETROS DO MODELO GR4J PARA BACIAS SEM DADOS

Como mencionado, inicialmente a regionalização focou nos parâmetros do modelo. Em itens posteriores, será explorada a regionalização de vazões típicas de mínimas (Q90 e Q7,10). Ao aplicar o modelo GR4J nas 52 estações foi observado que 50% apresentaram KGE maior ou igual a 0,70 no processo de calibração (Tabela 18). O coeficiente KGE foi escolhido porque é a função de calibração do modelo.

Tabela 18 - Parâmetros do modelo GR4J com KGE igual ou superior a 0.70.

Estação	Cluster	KGE	Área*	Melhor conjunto de parâmetros			
				X1*	X2*	X3*	X4*
50465000	1	0,79	11834,96	99,81	-90,90	840,14	2,05
50540000		0,78	27982,18	60,88	-90,14	594,23	3,12
50595000		0,77	35133,49	40,66	-85,79	760,92	4,52
51350000		0,85	37863,95	14,45	-99,75	725,18	2,33
51460000		0,71	11566,03	146,71	-22,82	51,98	1,41
52404000		0,74	29678,15	241,78	-60,82	216,08	1,27
52405000		0,80	32611,95	195,80	-66,80	256,37	1,18
52570000		0,81	40498,39	496,61	-9,72	115,91	2,67
52680000		0,75	46547,88	106,21	-97,89	798,14	3,03
51120000		2	0,72	2344,33	0,00	-14,48	213,33
51230000	0,84		2102,95	39,97	-1,87	191,00	2,56

Estação	Cluster	KGE	Área*	Melhor conjunto de parâmetros				
				X1*	X2*	X3*	X4*	
51240000		0,81	16417,18	0,04	-98,08	793,01	3,10	
51280000		0,83	22760,99	10,75	-93,22	890,29	3,15	
52760000		0,77	2726,18	20,52	-29,71	621,30	1,03	
52790000		0,81	5246,77	0,00	-36,08	394,06	1,40	
53620000		0,73	12923,24	226,57	-83,57	411,63	3,96	
53630000		0,75	16339,08	913,22	-21,50	166,93	2,59	
53650000		0,77	18414,00	619,99	-31,35	251,56	2,19	
53780000		0,71	2831,02	0,00	-86,32	674,60	1,92	
53950000		0,72	30473,57	28,95	-92,64	825,82	2,00	
54950000		0,84	68283,32	9,82	-64,13	818,50	2,10	
45298000		3	0,85	256995,20	25,55	-96,59	2741,11	4,40
46035000			0,76	312707,10	75,85	-92,93	3866,60	2,71
46105000			0,80	321332,30	46,16	-99,09	3437,23	4,82
46360000			0,77	351975,40	53,86	-95,48	3284,63	4,84
50795000			0,79	4703,06	4,48	-71,99	656,71	2,30

*X1 (mm); X2 (mm/dia); X3 (mm); X4 (dias); Área (Km²).

Fonte: Próprio autor (2023).

Em um estudo realizado por Kuana *et al.* (2021), foram considerados dois períodos com as séries de vazões, P1 e P2. Na primeira fase, o período P1 foi calibrado e P2 foi validado, e na segunda fase, o período P2 foi calibrado e P1 foi validado. Neste estudo, o modelo GR4J alcançou valores medianos de KGE modificado para o período de calibração igual a 0,77 e 0,74 para a primeira e a segunda fase, respectivamente. Valores parecidos com os obtidos na tabela acima.

Foi obtida a estatística descritiva dos parâmetros do modelo GR4J para o período de calibração. No cálculo foi considerado os dados de 26 (50%) estações, as quais obtiveram o índice KGE igual ou superior a 70%. Para melhor visualização, os dados foram sintetizados abaixo (Tabela 19).

Tabela 19 - Estatística descritiva dos parâmetros do modelo para o período de calibração.

Estatística Descritiva	X1	X2	X3	X4
Percentil 5%	0,00	-98,84	128,67	1,07
Mínimo	0,00	-99,75	51,98	0,83
Mediana	43,41	-84,68	665,66	2,45
Máximo	913,22	-1,87	3866,60	4,84
Percentil 95%	589,15	-10,91	3399,08	4,75
Média	133,79	-66,68	946,05	2,60
Quantidade de dados	26	26	26	26

*X1 (mm); X2 (mm/dia); X3 (mm); X4 (dias).

Fonte: Próprio autor (2023).

Para contextualizar os resultados, estes foram comparados com os obtidos por Ruelland; Dezetter; Hublart (2014), Oueldkaddour *et al.* (2021) e Andrade *et al.* (2021) em bacias semiáridas localizadas no Marrocos, no Chile (Rivadavia, Varillar, Riva. & Var) e no Brasil (Capibaribe em Pernambuco, Pindaré no Maranhão, Salgado no Ceará, Sirinhaém em Pernambuco e Piauitinga em Sergipe).

Tratando-se do parâmetro X1, o valor mínimo encontrado entre os autores citados acima foi de 4,901 mm para a bacia do Rio Salgado no Ceará, descrito por Andrade *et al.* (2021) e o valor máximo foi de 1005 mm, encontrado por Ruelland, Dezetter e Hublart (2014) em Varillar (Chile). Neste estudo, X1 apresentou valor mínimo de 0 mm e o máximo 913,22 mm, indicando uma maior capacidade do reservatório de produção. As estações que apresentaram alto valor de X1 foram 53630000 (913,22 mm), 53650000 (619,99 mm) e 52570000 (496,61 mm), elas possuem uma área de 16339,08 km², 18414,00 km² e 40498,39 km², respectivamente. Apesar das três estações terem apresentado um comportamento inversamente proporcional entre X1 e a área, não foi identificado este comportamento quando avaliado o conjunto de dados apresentado na Tabela 18. As três estações se encontram sobre a Embasamento Fraturado indiferenciado e o tipo de solo é o Argissolo vermelho para as estações 53630000 e 53650000 e Luviossolo crômico para 52570000 (ver Figura 12 e Figura 14). Para o parâmetro X2, o valor mínimo encontrado entre os autores foi -14,893 mm/dia, em uma bacia localizada no Marrocos, como descrito por Oueldkaddour *et al.* (2021) e o valor máximo foi de 2,700 mm/dia, encontrado por Ruelland, Dezetter e Hublart (2014) em Rivadavia (Chile). Neste estudo, o valor mínimo para X2 foi -99,75 mm e o valor máximo foi de -1,87 mm/dia, indicando que, de forma geral, a troca de água nas bacias é negativa, tendo infiltração de água no solo superior à contribuição subterrânea. As estações que apresentaram maiores valores de X2 foram 51230000 (-1,87 mm/dia), a qual está sobre a Formação Salitre e o tipo de solo predominante é o Latossolo vermelho, 52570000 (-9,72 mm/dia), localizada sobre o Embasamento Fraturado Indiferenciado e o tipo de solo é o Luvisolo Crônico e a estação 51120000 (-14,48 mm/dia), que está sobre Grupo Bambuí Terrigena e o solo é o Neossolo Litólico.

Em relação ao parâmetro X3, o valor mínimo encontrado entre os autores citados acima foi 2,090 mm para a bacia do Rio Capibaribe em Pernambuco, como relatado por Andrade *et al.* (2021) e o valor máximo foi de 473 mm, encontrado por Ruelland, Dezetter e Hublart (2014) em Rivadavia (Chile). Neste estudo, X3 apresentou valor mínimo de 51,98 mm e o máximo 3866,60 mm, indicando que alta capacidade máxima de armazenamento de água no aquífero subterrâneo das bacias avaliadas. As estações que apresentaram altos valores de X3 foram 46360000, 46105000, 46035000, 45298000. Foi observado que todas as estações estão situadas

em terrenos com elevação média entre 614 e 776 metros, o tipo de solo é o latossolo vermelho amarelo e estão sobre o Depósito Aluvionar, próximo ao aquífero Urucuia.

Por fim, o parâmetro X4 obteve valor mínimo encontrado entre os autores citados acima de -0,444 dia para a bacia do Rio Salgado no Ceará, descrito por Andrade *et al.* (2021) e o valor máximo foi encontrado por Ruelland, Dezetter e Hublart (2014) em Varillar (Chile) (12 dias). Neste estudo X4 apresentou valor mínimo de 0,83 dias e o máximo 4,84 dias. As estações que apresentaram altos valores de X4 foram 46360000, 46105000, 452980000, ambas estão situadas em terrenos onde o tipo de solo predominante é o Latossolo vermelho amarelo e estão sobre o Depósito Aluvionar, próximo ao aquífero Urucuia.

Observando a variabilidade entre os valores dos parâmetros, houve a necessidade de realizar uma análise mais aprofundada. Foi realizada uma análise de correlação entre as características utilizadas para a determinação das regiões homogêneas e os parâmetros do modelo (Tabela 20).

Tabela 20 - Análise de correlação dos parâmetros do modelo GR4J com as características das bacias.

Característica da bacia	Correlação			
	X1	X2	X3	X4
Declividade média da bacia	-0,07	0,39	-0,42	-0,52
Área da bacia hidrográfica	-0,15	-0,42	0,96	0,62
Aridez	0,15	0,07	-0,57	-0,17
Duração média dos eventos de precipitação máxima	0,20	0,08	-0,26	-0,09
Superfícies artificiais ou zonas urbanas	-0,23	-0,40	0,82	0,49
Porosidade subterrânea da bacia hidrográfica	-0,09	0,04	0,29	0,27
Profundidade mediana do lençol freático	0,28	0,44	-0,78	-0,51
Teor de areia	-0,04	-0,11	-0,21	0,05
Longitude	-0,07	0,13	-0,71	-0,37
precipitação média anual	-0,26	-0,22	0,74	0,31
Evapotranspiração	-0,06	-0,29	0,47	0,40
Frequência dos dias secos	0,25	0,46	-0,92	-0,57
Frequência dos dias de precipitação máxima	0,32	0,43	-0,91	-0,53

Fonte: Próprio autor (2023).

Nessa primeira análise foi observado que os parâmetros X1 e X2 não apresentaram correlações significativas com as características utilizadas. O parâmetro X3 apresentou uma correlação de 0,96 com a área da bacia, de -0,92 com a frequência dos dias secos, de -0,91 com frequência dos dias de precipitação máxima, de 0,82 com as superfícies artificiais ou zonas urbanas, de 0,74 com a precipitação média anual e de -0,71 com a longitude. O parâmetro X4 apresentou uma correlação de 0,62 com a área da bacia, -0,57 com a frequência dos dias secos,

-0,53 com frequência dos dias de precipitação máxima, -0,52 com a declividade média da bacia e -0,51 com a profundidade mediana do lençol freático.

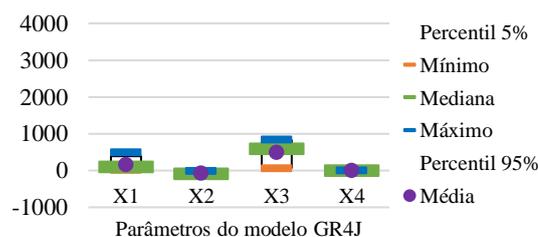
A baixa correlação entre os parâmetros e os atributos das estações pode estar correlacionada aos *outliers* presentes. O conjunto de estações apresenta uma grande variabilidade de dados quando analisadas como um único grupo.

Foi observada uma discrepância entre os valores mínimos e máximos de todos os parâmetros observados. Por tais motivos, houve a necessidade de avaliar os valores dos parâmetros por grupo, a fim de verificar se dentro de cada grupo as estações obtiveram valores de parâmetros próximos.

O Grupo 1 foi o segundo grupo com o maior número de estações. A capacidade do reservatório de produção variou entre 14,45 e 496,61 mm, o coeficiente de troca subterrânea variou entre -99,75 e -9,72 mm/dia; a capacidade máxima do reservatório de propagação ficou entre 51,98 e 840,14 mm e o tempo de base do hidrograma unitário variou entre 1,18 e 4,52 dias (Tabela 21).

Tabela 21 - Estatística descritiva dos parâmetros do modelo para o período de calibração do Grupo 1.

Estatística	X1	X2	X3	X4
Percentil 5%	24,93	-99,01	77,55	1,22
Mínimo	14,45	-99,75	51,98	1,18
Mediana	106,21	-85,79	594,23	2,33
Máximo	496,61	-9,72	840,14	4,52
Percentil 95%	394,68	-14,96	823,34	3,96
Média	155,88	-69,40	484,33	2,40
Desvio padrão	147,08	32,99	319,81	1,08
N° de estações	9	9	9	9



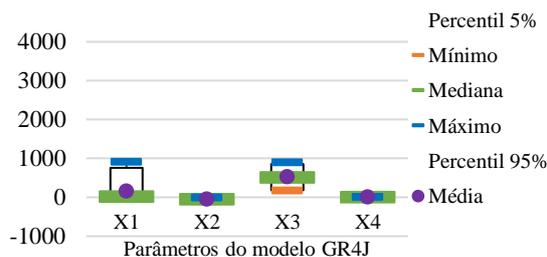
*X1 (mm); X2 (mm/dia); X3 (mm); X4 (dias).

Fonte: Próprio autor (2023).

O Grupo 2 foi o grupo com o maior número de estações, também foi o cluster que apresentou maiores variações de parâmetros. A capacidade do reservatório de produção variou entre 0 e 913,22 mm, o coeficiente de troca subterrânea variou entre -98,08 e -8,81 mm/dia; a capacidade máxima do reservatório de propagação ficou entre 166,93 e 890,29 mm e o tempo de base do hidrograma unitário variou entre 0,83 e 3,96 dias (Tabela 22).

Tabela 22 - Estatística descritiva dos parâmetros do modelo para o período de calibração do Grupo 2.

Estatística	X1	X2	X3	X4
Percentil 5%	0,00	-95,41	180,17	0,94
Mínimo	0,00	-98,08	166,93	0,83
Mediana	15,64	-50,11	516,47	2,15
Máximo	913,22	-1,87	890,29	3,96
Percentil 95%	751,94	-8,81	854,83	3,51
Média	155,82	-54,41	521,00	2,24
Desvio padrão	298,70	35,36	278,80	0,91
N° de estações	12	12	12	12



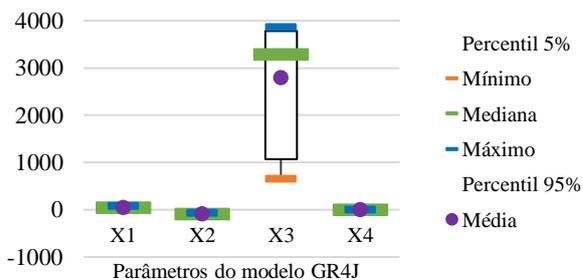
*X1 (mm); X2 (mm/dia); X3 (mm); X4 (dias).

Fonte: Próprio autor (2023).

O Grupo 3 apresentou o melhor ajuste quando comparado aos demais. Foi observado que a capacidade máxima do reservatório de propagação (X3), foi o parâmetro com maior variabilidade dentro do grupo, com valores variando entre 656,71 a 3866,60 mm (Tabela 23).

Tabela 23 - Estatística descritiva dos parâmetros do modelo para o período de calibração do Grupo 3.

Estatística	X1	X2	X3	X4
Percentil 5%	8,69	-98,59	1073,59	2,38
Mínimo	4,48	-99,09	656,71	2,30
Mediana	46,16	-95,48	3284,63	4,40
Máximo	75,85	-71,99	3866,60	4,84
Percentil 95%	71,45	-76,18	3780,73	4,84
Média	41,18	-91,22	2797,26	3,81
Desvio padrão	27,29	10,97	1262,51	1,22
N° de estações	5	5	5	5



*X1 (mm); X2 (mm/dia); X3 (mm); X4 (dias).

Fonte: Próprio autor (2023).

Os parâmetros do modelo GR4J são utilizados para representar as características hidrológicas das bacias. Esses parâmetros estão relacionados às características físicas da bacia, como a área de drenagem, a capacidade de armazenamento de água no solo, a taxa de escoamento superficial, entre outros. A escolha adequada dos parâmetros do modelo é fundamental para garantir uma representação precisa do comportamento hidrológico da bacia. Por esse motivo, foi realizada uma análise de correlação entre os parâmetros do modelo e as características das estações dentro de cada cluster. Nessa nova avaliação também foram considerados mais atributos (Tabela 24).

Tabela 24 - Análise de correlação dos parâmetros do modelo GR4J com as características das bacias dentro de cada cluster.

Atributos	Grupo 1				Grupo 2				Grupo 3			
	X1	X2	X3	X4	X1	X2	X3	X4	X1	X2	X3	X4
Precipitação média diária	0,33	0,02	-0,13	-0,52	-0,25	0,25	-0,03	-0,48	0,24	-0,66	0,51	0,34
Evapotranspiração potencial diária média (PET)	0,13	-0,07	0,10	0,02	0,10	0,06	-0,27	0,19	0,07	0,40	-0,20	-0,17
Evapotranspiração média diária real	-0,06	-0,38	0,36	-0,31	-0,07	0,20	-0,20	-0,31	-0,81	0,96	-0,96	-0,71
Aridez, calculada como a razão entre a PET média e a precipitação média	-0,31	-0,06	0,15	0,51	0,22	-0,22	0,00	0,48	-0,22	0,64	-0,49	-0,32
Sazonalidade e calendário da precipitação	0,64	0,26	-0,41	-0,52	0,60	-0,01	-0,23	0,53	0,75	-0,98	0,95	0,69
Assíncronidade entre os ciclos anuais de precipitação e PET, em que valores elevados representam grandes diferenças de magnitude e fase	-0,42	-0,34	0,42	0,59	0,35	-0,08	-0,26	0,57	-0,61	0,95	-0,86	-0,67
Frequência dos dias de precipitação elevada (5 vezes a precipitação média diária)	0,71	0,33	-0,45	-0,26	0,35	0,09	-0,44	0,36	-0,83	0,95	-0,97	-0,68
Duração média dos eventos de precipitação elevada (número de dias consecutivos 5 vezes a precipitação média diária)	0,51	0,18	-0,32	-0,30	-0,05	-0,26	0,25	0,47	0,75	-0,97	0,94	0,77
Frequência dos dias secos (< 1 mm dia-1)	0,49	0,43	-0,51	-0,53	0,15	0,39	-0,56	0,01	-0,77	0,98	-0,96	-0,69
Classe geológica mais comum na bacia hidrográfica*	0,59	0,49	-0,55	-0,53	0,22	0,58	-0,63	-0,55	-0,66	0,97	-0,89	-0,75
Porcentagem da bacia hidrográfica abrangida pela segunda classe geológica mais comum**	-0,56	-0,22	0,18	0,03	0,27	0,13	-0,25	-0,04	-0,21	-0,50	0,14	0,61
Porcentagem da bacia hidrográfica coberta por rochas sedimentares carbonatadas	-0,36	-0,34	0,28	-0,02	-0,34	-0,34	0,38	0,43	0,69	-0,97	0,91	0,66
Porosidade subterrânea da bacia hidrográfica	-0,44	-0,53	0,69	0,40	0,48	0,18	-0,50	0,41	0,85	-0,95	0,99	0,62
Consumo total de água em 2017, normalizado por bacia hidrográfica	0,37	0,02	-0,21	-0,42	-0,23	0,33	-0,32	-0,37	0,73	-0,98	0,94	0,66
Consumo total de água em 2017, normalizado pelo caudal médio anual	0,66	0,29	-0,47	-0,42	0,34	0,23	-0,45	0,03	0,81	-0,95	0,96	0,72
Capacidade máxima total de armazenamento dos reservatórios na bacia hidrográfica	0,56	0,28	-0,08	0,27	-0,14	-0,10	0,34	-0,03	0,75	-0,98	0,95	0,70
Razão entre a capacidade total de armazenamento do reservatório da bacia hidrográfica e a sua vazão anual total	0,68	0,62	-0,39	0,18	0,53	0,04	-0,13	0,27	0,70	-0,98	0,92	0,71
Vazão diária média	-0,33	-0,40	0,38	-0,13	-0,64	0,47	-0,11	-0,62	0,43	-0,79	0,69	0,43
Taxa de escoamento, calculada como a razão entre a descarga média diária e a precipitação média diária	-0,43	-0,45	0,45	-0,06	-0,66	0,46	-0,11	-0,59	0,50	-0,83	0,74	0,45
Elasticidade da precipitação (isto é, a sensibilidade da vazão às mudanças na precipitação na escala de tempo anual, usando a descarga diária média como referência). Ver equação 7 em Sankarasubramanian <i>et al.</i> (2001), sendo o último elemento P/Q não Q/P	-0,20	0,09	-0,06	-0,11	0,32	0,32	-0,44	-0,01	0,54	-0,21	0,39	0,27

Atributos	Grupo 1				Grupo 2				Grupo 3			
	X1	X2	X3	X4	X1	X2	X3	X4	X1	X2	X3	X4
Declive da curva de duração do caudal entre os percentis das correntes 33ª e 66ª	-0,03	0,35	-0,57	-0,77	-0,40	0,26	-0,12	0,10	-0,78	0,96	-0,95	-0,73
Índice de vazão de base, calculado como a razão entre a vazão média diária de base e a vazão média diária, sendo a separação por hidrografia efetuada com o filtro digital de Ladson <i>et al.</i> (2013)	-0,27	-0,64	0,81	0,76	0,19	-0,32	0,36	-0,17	0,83	-0,96	0,98	0,66
Data média de semi-fluxo (ou seja, a data em que a vazão cumulativa desde 1 de setembro atinge metade da vazão anual)	-0,12	-0,35	0,68	0,60	-0,32	-0,24	0,40	-0,12	-0,72	0,98	-0,93	-0,69
Q5 (vazão mínima)	-0,18	-0,38	0,50	0,58	-0,54	-0,04	0,38	-0,67	0,73	-0,97	0,93	0,65
Q95 (vazão máxima)	-0,28	-0,44	0,34	-0,21	-0,54	0,54	-0,27	-0,44	-0,21	-0,26	0,05	0,06
Frequência dos dias de vazão máxima (> 9 vezes a mediana da vazão diária)	0,20	0,44	-0,69	-0,85	0,10	0,43	-0,54	0,27	-0,81	0,97	-0,98	-0,64
Duração média dos eventos de vazão máxima (número de dias consecutivos > 9 vezes a mediana da vazão diária)	0,74	0,91	-0,74	-0,13	0,61	-0,15	-0,25	0,83	-0,32	-0,18	-0,11	0,43
Frequência dos dias de vazão mínima (< 0,2 vezes a vazão média diária)	0,21	0,53	-0,74	-0,84	-0,25	0,40	-0,31	0,06	-0,75	0,98	-0,95	-0,70
Duração média dos eventos de vazão mínima (número de dias consecutivos < 9 vezes a mediana da vazão diária)	-0,13	0,29	-0,53	-0,73	-0,21	-0,25	0,29	0,39	-0,75	0,98	-0,95	-0,70
Percentagem coberta por terras agrícolas	-0,51	-0,04	0,18	0,35	-0,49	-0,01	0,26	-0,54	-0,74	0,98	-0,94	-0,71
Percentagem coberta por um mosaico de terras agrícolas e vegetação natural	0,57	0,38	-0,58	-0,16	0,21	-0,54	0,37	0,55	-0,86	0,94	-0,99	-0,62
Percentagem coberta por florestas folhosas ou de folhas caducas, perenes ou caducas	0,60	0,20	-0,29	-0,39	0,29	0,07	-0,05	-0,17	0,87	-0,90	0,96	0,65
Percentagem coberta por arbustos	-0,56	-0,53	0,60	0,14	-0,11	0,13	-0,19	0,16	0,77	-0,98	0,96	0,68
Percentagem coberta por prados ou zonas com vegetação escassa (<15%)	-0,59	-0,49	0,51	0,14	-0,31	-0,30	0,29	0,20	0,81	-0,97	0,98	0,66
Percentagem coberta por zonas áridas	-0,45	-0,36	0,38	0,71	-0,19	0,19	-0,18	-0,33	0,55	0,03	0,29	-0,15
Percentagem coberta por superfícies artificiais ou zonas urbanas	0,60	0,31	-0,08	0,23	-0,13	-0,34	0,34	-0,08	-0,46	0,03	-0,25	-0,11
Percentagem coberta por massas de água ou zonas úmidas	0,46	0,65	-0,42	-0,09	-0,03	0,21	-0,26	0,05	0,39	-0,84	0,68	0,58
Cobertura de terras dominante	0,57	0,38	-0,58	-0,16	0,00	0,14	-0,07	-0,60	-0,86	0,94	-0,99	-0,62
Latitude	-0,57	-0,32	0,45	0,34	-0,39	0,17	-0,07	-0,04	0,25	0,18	0,01	0,03
Longitude	-0,54	-0,25	0,45	0,78	-0,33	-0,28	0,59	-0,37	-0,71	0,97	-0,92	-0,69
área da Bacia hidrográfica de acordo com GSIM	0,18	-0,19	0,18	0,42	0,00	-0,35	0,50	0,18	0,83	-0,95	0,96	0,70
Percentagem do teor de areia do material do solo inferior a 2 mm a uma profundidade de 30 cm	-0,42	-0,04	0,17	0,32	-0,04	-0,04	-0,08	0,36	-0,63	0,96	-0,88	-0,68
Percentagem de silte do material do solo inferior a 2 mm a uma profundidade de 30 cm	-0,38	0,03	-0,01	0,32	-0,49	-0,17	0,37	-0,23	0,73	-0,97	0,92	0,77

Atributos	Grupo 1				Grupo 2				Grupo 3			
	X1	X2	X3	X4	X1	X2	X3	X4	X1	X2	X3	X4
Percentagem do teor de argila do material do solo inferior a 2 mm a uma profundidade de 30 cm	0,42	0,03	-0,13	-0,33	0,22	0,11	-0,05	-0,30	0,58	-0,94	0,84	0,63
Teor de carbono orgânico do solo a uma profundidade de 30 cm	-0,03	-0,16	0,09	-0,17	-0,53	0,29	0,09	-0,78	0,67	-0,96	0,90	0,64
Profundidade do leito rochoso	0,40	-0,04	-0,08	-0,51	-0,07	0,02	-0,17	-0,07	-0,60	0,95	-0,85	-0,64
Profundidade mediana do lençol freático	0,54	0,13	-0,26	-0,16	0,10	0,33	-0,48	0,01	-0,86	0,93	-0,98	-0,65
Elevação do medidor	0,42	0,17	-0,31	-0,79	0,55	0,17	-0,58	0,52	0,73	-0,98	0,93	0,71
Elevação média da bacia	0,53	0,14	-0,30	-0,53	0,31	0,13	-0,35	0,08	0,74	-0,98	0,94	0,69
Declive médio da bacia hidrográfica	0,51	0,04	-0,12	-0,26	-0,64	0,28	0,20	-0,62	0,50	-0,90	0,76	0,73
Área de captação	0,18	-0,19	0,18	0,42	0,00	-0,35	0,50	0,18	0,83	-0,95	0,96	0,70

*A classe geológica mais comum no Grupo 1 é a metamórfica (100%), no grupo 2 são as rochas metamórficas (50%), os sedimentos não consolidados (33%) e as rochas sedimentares siliciclásticas (17%). No Grupo 3 são as rochas sedimentares siliciclásticas (100%).

**A segunda classe geológica mais comum no Grupo 1 são os sedimentos não consolidados (67%), as rochas plutônicas ácidas (22%) e as rochas sedimentares mistas (11%). No Grupo 2 são os Sedimentos não consolidados (33%), as metamórficas (25%), as Rochas sedimentares mistas (17%), Rochas sedimentares siliciclásticas (17%) e as rochas plutônicas ácidas (8%). No Grupo 3 são os sedimentos não consolidados (80%) e as rochas sedimentares mistas (20%).

Fonte: Próprio autor (2023).

Ao realizar a análise por cluster, foi observado que os parâmetros do modelo GR4J apresentaram correlação significativa com algumas características físicas da área em que estão localizadas as estações do Grupo 1. Os parâmetros apresentaram as seguintes correlações: X1 apresentou correção de 0,71 com a frequência dos dias de precipitação elevada (5 vezes a precipitação média diária) é de 0,68 com a razão entre a capacidade total de armazenamento do reservatório da bacia hidrográfica e a sua vazão anual total; X2 apresentou correlação de 0,62 com a razão entre a capacidade total de armazenamento do reservatório da bacia hidrográfica e a sua vazão anual total; X3 apresentou correlação de 0,69 com a porosidade subterrânea da bacia hidrográfica; X4 apresentou correlação de 0,74 com o índice de vazão de base, calculado como a razão entre a vazão média diária de base e a vazão média diária, de -0,85 com a frequência dos dias de vazão máxima (> 9 vezes a mediana da vazão diária), de -0,84 com a frequência dos dias de vazão mínima (< 0,2 vezes a vazão média diária), de -0,73 com a duração média dos eventos de vazão mínima (número de dias consecutivos < 9 vezes a mediana da vazão diária), de 0,71 com a percentagem coberta por zonas áridas, de 0,78 com a longitude e de 0,79 com a elevação do medidor.

Ao realizar a análise do cluster 2, foi observado que os parâmetros do modelo GR4J apresentaram correlação significativa com algumas características físicas da área que estão localizadas as estações que compõem este grupo. Os parâmetros apresentaram as seguintes correlações: X1 apresentou correlação de 0,60 com a sazonalidade e calendário da precipitação, de -0,64 com a vazão diária média, de 0,66 com a taxa de escoamento, de -0,64 com o declive médio da bacia hidrográfica e de 0,61 com a duração média dos eventos de vazão máxima (número de dias consecutivos > 9 vezes a mediana da vazão diária) calculada como a razão entre a descarga média diária e a precipitação média diária; X2 apresentou a maior correlação de 0,58 que foi com a classe geológica mais comum na bacia hidrográfica (rochas metamórficas); X3 também apresentou correlação de -0,63 com a classe geológica mais comum na bacia hidrográfica. Esse parâmetro também apresentou correlação de 0,59 com a longitude e de -0,58 com a elevação do medidor; X4 apresentou correlação de -0,62 com a vazão diária média, de -0,59 com a taxa de escoamento, calculada como a razão entre a descarga média diária e a precipitação média diária, de -0,67 com a Q5 (vazão mínima), de 0,83 com a duração média dos eventos de vazão máxima (número de dias consecutivos > 9 vezes a mediana da vazão diária), de -0,78 com o teor de carbono orgânico do solo a uma profundidade de 30 cm e de -0,62 com o declive médio da bacia hidrográfica.

O grupo 3 foi o cluster que apresentou maior número de características correlacionadas com os parâmetros do modelo GR4J. Os quatro parâmetros do modelo apresentaram correlação

igual ou superior a 0,60 com: a evapotranspiração média diária real; com a sazonalidade e calendário da precipitação; a assíncronidade entre os ciclos anuais de precipitação e PET; a frequência dos dias de precipitação elevada (5 vezes a precipitação média diária); a duração média dos eventos de precipitação elevada (número de dias consecutivos 5 vezes a precipitação média diária); a frequência dos dias secos (< 1 mm dia⁻¹); a classe geológica mais comum na bacia hidrográfica (rochas sedimentares siliciclásticas (100%)); a porosidade subterrânea da bacia hidrográfica; a razão entre a capacidade total de armazenamento do reservatório da bacia hidrográfica e a sua vazão anual total; o índice de vazão de base, calculado como a razão entre a vazão média diária de base e a vazão média diária; a Q5 (vazão mínima); a frequência dos dias de vazão máxima (> 9 vezes a mediana da vazão diária); a frequência dos dias de vazão mínima ($< 0,2$ vezes a vazão média diária); a percentagem coberta por terras agrícolas; a longitude; a área da bacia hidrográfica; a Percentagem do teor de área, silte e argila do material do solo inferior a 2 mm a uma profundidade de 30 cm; a profundidade do leito rochoso; a profundidade mediana do lençol freático; a elevação do medidor; a elevação média da bacia; o declive médio da bacia hidrográfica e a área de captação.

Kuana (2022), a partir da técnica *Random Forest I*, encontrou que a declividade do rio principal e a declividade média da bacia se correlacionou com o parâmetro X1; a altitude média e a radiação média no inverno, com o parâmetro X2; a radiação média no inverno, com o parâmetro X3, e a fração de gleissolo, com parâmetro X4. Conforme a descrição de cada parâmetro, o esperado seria que a capacidade do reservatório de produção (X1) apresentasse correlação significativa com o tamanho e as características físicas da bacia; o coeficiente de troca subterrânea (X2), com as características físicas do solo e da vegetação e intensidade das chuvas; a capacidade máxima do reservatório de propagação (X3), com as características geológicas da bacia; e o tempo de base do hidrograma unitário (X4) com a permeabilidade do solo e do subsolo e da estrutura do aquífero.

Não foi identificada uma relação direta dos grupos formados com as três sub-regiões do Nordeste, tendo em vista que, no primeiro grupo, 5 (56%) estações estão situadas na sub-região do Agreste, 3 (33%) na sub-região do sertão e 1 (11%) na Zona da Mata. No segundo grupo, 5 (42%) estações estão situadas na sub-região do Agreste, 4 (33%) na sub-região do Sertão e 3 (25%) na Zona da Mata. Já no terceiro grupo, 4 (80%) estações estão situadas na sub-região do Sertão e 1 (20%) na Zona da Mata.

Após realizar a correlação dos parâmetros com os atributos, foi verificada se a técnica de regressão linear se ajustou aos parâmetros e as características físicas das estações do Grupo 1 (Tabela 25).

Foi observado que apesar da frequência dos dias de precipitação elevada (5 vezes a precipitação média diária) manter uma correlação acima de 70% com o parâmetro X1, o modelo de previsão criado a partir da variável não é indicado para prever o parâmetro. Ao avaliar a significância, o P-valor foi superior a 0,05, deixando evidências de que a relação entre as duas variáveis não é significativa. A linearidade da amostra verificada através dos coeficientes de determinação, sendo observado que 54,32% da variação do comportamento de X1 é devido ao comportamento da variação da Frequência dos dias de precipitação elevada (5 vezes a precipitação média diária), porém os outros 45,68% é devido ao comportamento de outras variáveis omitidas no modelo. O erro absoluto médio foi de 43,083% para a amostra de treino e aumento para 69,97% para a amostra de teste.

A Razão entre a capacidade total de armazenamento do reservatório da bacia hidrográfica e a sua vazão anual total também apresentou correlação acima de 70% com o parâmetro X1, porém o modelo de previsão criado também não é indicado para prever o parâmetro. Ao avaliar a significância, o P-valor foi superior a 0,05, indicando que a relação entre as duas variáveis não é significativa. A linearidade da amostra, avaliada através dos coeficientes de determinação, apontou que 63% da variação do comportamento de X1 é devido ao comportamento da variação da Razão entre a capacidade total de armazenamento do reservatório da bacia hidrográfica e a sua vazão anual, os outros 37% é devido ao comportamento de outras variáveis omitidas no modelo. O modelo apresentou um erro de 83,64% para a amostra de treino e 42,32% na amostra de teste.

O atributo nomeado como razão entre a capacidade total de armazenamento do reservatório da bacia hidrográfica e a sua vazão anual apresentou correlação acima de 80% com o parâmetro X2, com um P-valor foi inferior a 0,05, indicando que a relação entre as duas variáveis é significativa. O coeficiente de determinação indicou que 70% da variação do comportamento de X2 é devido ao comportamento da variação da razão entre a capacidade total de armazenamento do reservatório da bacia hidrográfica e a sua vazão anual, os outros 30% é devido ao comportamento de outras variáveis omitidas no modelo. O modelo apresentou um erro de 6,62% para a amostra de treino e 31,31% na amostra de teste.

A Porosidade subterrânea da bacia hidrográfica apresentou correlação acima de 70% com o parâmetro X3, com um P-valor superior a 0,05, indicando que a relação entre as duas variáveis não é significativa. O coeficiente de determinação indicou que 51,90% da variação do comportamento de X3 é devido ao comportamento da variação de da porosidade subterrânea, os outros 48,10% é devido ao comportamento de outras variáveis omitidas no modelo. O modelo apresentou um erro de 47,80% para a amostra de treino. Já para amostra de teste, a

amostra escolhida não foi significativa, tendo em vista que todos os valores extrapolaram os valores da amostra de treino. Geralmente modelos lineares não respondem bem para dados extrapolados. O recomendado seria gerar uma amostra de treino/teste para cada modelo.

O índice de vazão de base apresentou correlação de aproximadamente 77% com o parâmetro X4, com um P-valor superior a 0,05, indicando que a relação entre as duas variáveis não é significativa. O coeficiente de determinação indicou que 60,40% da variação do comportamento de X4 é devido ao comportamento da variação do Índice de vazão de base. O modelo apresentou um erro de 6,41% para a amostra de treino e 24,17% para a amostra de teste.

A frequência dos dias de vazão máxima (> 9 vezes a mediana da vazão diária) apresentou correlação de aproximadamente 96% com o parâmetro X4, com um P-valor inferior a 0,05, indicando que a relação entre as duas variáveis é significativa. O coeficiente de determinação indicou que aproximadamente 92% da variação do comportamento de X4 é devido ao comportamento da variação da frequência dos dias de vazão máxima. O modelo apresentou um erro de 1,1% para a amostra de treino e 17,43% para a amostra de teste.

Frequência dos dias de vazão mínima ($< 0,2$ vezes a vazão média diária) apresentou correlação de aproximadamente 87% com o parâmetro X4, com um P-valor inferior a 0,05, indicando que a relação entre as duas variáveis é significativa. O coeficiente de determinação indicou que aproximadamente 75% da variação do comportamento de X4 é devido ao comportamento da variação da frequência dos dias de vazão mínima. O modelo apresentou um erro de 3,83% para a amostra de treino e 21,25% para a amostra de teste.

As demais variáveis apresentaram correlação abaixo de 70%, um P-valor superior a 0,05, indicando que a relação entre as variáveis e o parâmetro X4 não foram significativas. E o coeficiente de determinação foi menor que 50%. O erro gerado entre as equações foi menor que 10% na fase de treino e menor que 27% na fase de teste.

A partir destas informações é possível afirmar que, para a amostra de treino e teste utilizada, as variáveis simuladas não foram significativas para prever os parâmetros X1 e X3. Assim como, uma correlação alta não é suficiente para afirmar que a variável é significativa para prever o comportamento dos parâmetros estudados. Foi observado também que, variáveis que apresentaram uma correlação inferior a 70% não foram significativas para gerar equações significativas.

Outro fator que pode ter comprometido a qualidade das equações foi a utilização de uma única amostra para treinar e testar o modelo. Uma vez que os parâmetros do GR4J não obedecem a um comportamento similar. O parâmetro X4, por exemplo, ficou com dados extrapolados na amostra de teste, sendo necessário a extração de uma estação. O parâmetro X3

apresentou todos os dados extrapolados na amostra de teste em relação ao limite máximo da amostra de treino. Sendo assim, houve a necessidade de gerar novos dados, com amostras de treino e teste para cada equação (Tabela 26). Com essa configuração foi possível considerar os limites mínimos e máximos dentro da amostra de treino. No grupo 1 o parâmetro X1 variou entre 14,45 e 496,61 mm, o parâmetro X2 variou entre -99,75 e -9,72 mm/dia, o parâmetro X3 variou entre 51,98 e 840,14 mm e o parâmetro X4 variou entre 1,18 e 4,52 dias.

A frequência dos dias de precipitação elevada (5 vezes a precipitação média diária) apresentou correlação de aproximadamente 88% com o parâmetro X1, com um P-valor inferior a 0,05, indicando que a relação entre as duas variáveis é significativa. O coeficiente de determinação indicou que aproximadamente 77% da variação do comportamento de X1 é devido ao comportamento da variação da frequência dos dias de precipitação elevada. O modelo apresentou um erro médio de -28,49% para a amostra de treino e -134,655% para a amostra de teste, em ambos os casos os valores previstos foram maiores que os valores obtidos através do GR4J.

A duração média dos eventos de vazão máxima apresentou correlação de aproximadamente 89% com o parâmetro X1. O P-valor foi inferior a 0,05, indicando que a relação entre as duas variáveis é significativa. O coeficiente de determinação indicou que aproximadamente 80% da variação do comportamento de X1 é devido ao comportamento da variação da duração média dos eventos de vazão máxima. A equação apresentou um erro médio de -77,88% para a amostra de treino, indicando que os valores previstos para a amostra de treino foram menores que os obtidos através do modelo GR4J. Para a amostra de teste o erro foi de 29,94% evidenciando que os valores obtidos através do GR4J foram menores que os gerados através da equação.

As equações geradas para os parâmetros X2 e X3 não obtiveram resultados significativos. O X4 é o parâmetro que obteve os melhores resultados.

A frequência dos dias de vazão máxima apresentou correlação de aproximadamente 82% com o parâmetro X4, com um P-valor inferior a 0,05, indicando que a relação entre as duas variáveis é significativa. O coeficiente de determinação indicou que aproximadamente 67% da variação do comportamento de X4 é devido ao comportamento da variação da frequência dos dias de vazão máxima. O modelo apresentou um erro de -4,294% para a amostra de treino, o sinal negativo indica que os valores previstos para a amostra de treino foram menores que os obtidos através do modelo GR4J e 2,136% para a amostra de teste, evidenciando que os valores obtidos através do GR4J foram menores que os gerados através da equação.

A frequência dos dias de vazão mínima apresentou correlação de aproximadamente 96% com o parâmetro X4, com um P-valor inferior a 0,05, indicando que a relação entre as duas variáveis é significativa. O coeficiente de determinação indicou que aproximadamente 93% da variação do comportamento de X4 é devido ao comportamento da variação da frequência dos dias de vazão mínima. O modelo apresentou um erro de -0,068% para a amostra de treino, o sinal negativo indica que os valores previstos para a amostra de treino foram menores que os obtidos através do modelo GR4J e 10,98% para a amostra de teste, evidenciando que os valores obtidos através do GR4J foram menores que os gerados através da equação.

A longitude dos eventos de vazão mínima apresentou correlação de aproximadamente 83% com o parâmetro X4, com um P-valor inferior a 0,05, indicando que a relação entre as duas variáveis é significativa. O coeficiente de determinação indicou que aproximadamente 91% da variação do comportamento de X4 é devido ao comportamento da variação da longitude. O modelo apresentou um erro de -3,263% para a amostra de treino, o sinal negativo indica que os valores previstos para a amostra de treino foram menores que os obtidos através do modelo GR4J e -53,676% para a amostra de teste, evidenciando que os valores obtidos através do GR4J foram maiores que os gerados através da equação.

A elevação do medidor apresentou correlação de aproximadamente 82% com o parâmetro X4, com um P-valor inferior a 0,05, indicando que a relação entre as duas variáveis é significativa. O coeficiente de determinação indicou que aproximadamente 91% da variação do comportamento de X4 é devido ao comportamento da variação da elevação do medidor. O modelo apresentou um erro de -4,056% para a amostra de treino e -43,704% para a amostra de teste, o sinal negativo indica que os valores previstos para a amostra de treino e teste foram menores que os obtidos através do modelo GR4J.

A Q5 (vazão mínima) apresentou correlação de aproximadamente 95% com o parâmetro X4, com um P-valor inferior a 0,05, indicando que a relação entre as duas variáveis é significativa. O coeficiente de determinação indicou que aproximadamente 91% da variação do comportamento de X4 é devido ao comportamento da variação da Q5. O modelo apresentou um erro de -3,665% para a amostra de treino e -65,703% para a amostra de teste, o sinal negativo indica que os valores previstos para a amostra de treino e teste foram menores que os obtidos através do modelo GR4J.

Após realizar a correlação dos parâmetros com os atributos, foi verificada se a técnica de regressão linear se ajustou aos parâmetros e as características físicas das estações pertencentes ao Grupo 2 (Tabela 27). Os limites mínimos e máximos dentro da amostra de treino para o parâmetro X1 variou entre 0 e 913,22 mm, o parâmetro X2 variou entre -98,08 e

-1,87 mm/dia, o parâmetro X3 variou entre 166,90 e 890,30 mm e o parâmetro X4 variou entre 0,83 e 3,96 dias.

A vazão diária média apresentou correlação de aproximadamente 73% com o parâmetro X1, com um P-valor inferior a 0,05, indicando que a relação entre as duas variáveis é significativa. O coeficiente de determinação indicou que aproximadamente 54% da variação do comportamento de X1 é devido ao comportamento da variação da vazão diária. Porém, o modelo apresentou um erro médio muito acima de 100% para as amostras de treino e teste, em ambos os casos os valores previstos foram muito maiores que os valores obtidos através do GR4J.

A taxa de escoamento apresentou correlação de aproximadamente 88% com o parâmetro X1. O P-valor foi inferior a 0,05, indicando que a relação entre as duas variáveis é significativa. O coeficiente de determinação indicou que aproximadamente 54% da variação do comportamento de X1 é devido ao comportamento da variação da taxa de escoamento. A equação apresentou um erro médio superior a 100% para a amostra de treino, indicando que os valores previstos para a amostra de treino foram menores que os obtidos através do modelo GR4J. Para a amostra de teste o erro foi de 83% evidenciando que os valores obtidos através do GR4J foram menores que os gerados através da equação. As equações geradas para os parâmetros X2 e X3 não obtiveram resultados significativos.

A taxa de escoamento apresentou correlação de aproximadamente 81% com o parâmetro X4. O P-valor foi inferior a 0,05, indicando que a relação entre as duas variáveis é significativa. O coeficiente de determinação indicou que aproximadamente 59% da variação do comportamento de X4 é devido ao comportamento da variação da taxa de escoamento. A equação apresentou um erro médio -4,27% para a amostra de treino, indicando que os valores previstos para a amostra de treino foram maiores que os obtidos através do modelo GR4J. Para a amostra de teste o erro foi de 16,92% evidenciando que os valores obtidos através do GR4J foram maiores que os gerados através da equação.

Após realizar a correlação dos parâmetros com os atributos, foi verificada se a técnica de regressão linear se ajustou aos parâmetros e as características físicas das estações pertencentes ao Grupo 3 (Tabela 28). Os limites mínimos e máximos dentro da amostra de treino para o parâmetro X1 variou entre 4,48 e 75,85 mm, o parâmetro X2 variou entre -99,09 e -71,99 mm/dia, o parâmetro X3 variou entre 656,70 e 3866,60 mm e o parâmetro X4 variou entre 2,30 e 4,84 dias. As equações geradas para os parâmetros X1, X2 e X3 não obtiveram resultados significativos.

A duração média dos eventos de vazão mínima e a longitude apresentaram correlação de aproximadamente 100% com o parâmetro X4. O P-valor de ambas as variáveis foram inferiores a 0,05, indicando que a relação entre elas e o parâmetro X4 é significativa. O coeficiente de determinação indicou que aproximadamente 100% da variação do comportamento de X4 pode ser explicado pelo comportamento da variação da duração média dos eventos de vazão mínima e da longitude. As equações apresentaram um erro médio 0% para as amostras de treino, indicando que os valores previstos foram praticamente iguais aos obtidos através do modelo GR4J. Já para as amostras de teste o erro foi de -44% para a equação utilizando a duração média dos eventos de vazão mínima e -47,95% utilizando a longitude. O sinal negativo evidenciou que os valores previstos utilizando a equação linear foram maiores que os obtidos através do GR4J.

Tabela 25 - Estatística descritiva das equações geradas utilizando as estações do Cluster 1, considerando apenas uma amostra de treino.

Variável dependente	Variável independente	Est.	2,50%	97,50%	t val.	p	R ²	Adj. R ²	Erro Absoluto Médio do grupo de treino (%)	Correlação	Erro Absoluto Médio do grupo de teste (%)
X1	interseção	-2115,478	-5063,793	832,838	-1,992	0,117					
	Frequência dos dias de precipitação elevada (5 vezes a precipitação média diária)	103,661	-28,585	235,907	2,176	0,095	0,542	0,428	-43,083	0,736	-69,971
	interseção	44,189	-161,499	249,878	0,597	0,583					
	Razão entre a capacidade total de armazenamento do reservatório da bacia hidrográfica e a sua vazão anual total	1,818	-0,116	3,751	2,610	0,059	0,630	0,538	-83,637	0,794	-42,320
X2	interseção	-91,251	-129,936	-52,567	-6,549	0,003					
	Razão entre a capacidade total de armazenamento do reservatório da bacia hidrográfica e a sua vazão anual total	0,403	0,039	0,767	3,075	0,037	0,703	0,629	-6,620	0,838	31,309
X3	interseção	-333,522	-1247,710	580,666	-1,013	0,368					
	Porosidade subterrânea da bacia hidrográfica	9560,414	-3221,352	22342,179	2,077	0,106	0,519	0,399	-47,800	0,720	-
X4	interseção	-0,081	-2,509	2,347	-0,093	0,931					
	Índice de vazão de base	6,097	-0,763	12,956	2,468	0,069	0,604	0,504	-6,405	0,777	-24,116
	interseção	3,882	3,049	4,715	12,935	0,000					
	Frequência dos dias de vazão máxima (> 9 vezes a mediana da vazão diária)	-0,052	-0,074	-0,031	-6,721	0,003	0,919	0,898	-1,097	-0,958	-17,432
	interseção	3,814	2,289	5,340	6,942	0,002					
	Frequência dos dias de vazão mínima (< 0,2 vezes a vazão média diária)	-0,010	-0,018	-0,002	-3,514	0,025	0,755	0,694	-3,832	-0,869	-21,248
	interseção	3,057	1,306	4,809	4,846	0,008					
Duração média dos eventos de vazão mínima (número de dias)	-0,040	-0,099	0,019	-1,866	0,136	0,465	0,332	-8,406	-0,682	-4,021	

Variável dependente	Variável independente	Est.	2,50%	97,50%	t val.	p	R ²	Adj. R ²	Erro Absoluto Médio do grupo de treino (%)	Correlação	Erro Absoluto Médio do grupo de teste (%)
	consecutivos < 9 vezes a mediana da vazão diária)										
	interseção	1,729	0,876	2,581	5,629	0,005					
	Percentagem coberta por zonas áridas	1,294	-0,608	3,195	1,889	0,132	0,471	0,339	-9,228	0,687	25,374
	interseção	23,325	-11,231	57,882	1,874	0,134					
	Longitude	0,535	-0,332	1,403	1,714	0,162	0,424	0,279	-8,638	0,651	14,440
	interseção	3,496	0,782	6,210	3,577	0,023					
	Elevação do medidor	-0,008	-0,022	0,006	-1,607	0,183	0,392	0,241	-8,767	-0,626	26,979

Fonte: Próprio autor (2023).

Tabela 26 - Estatística descritiva das equações geradas utilizando as estações do Cluster 1.

Variável dependente	Variável independente	Est.	2,50%	97,50%	t val.	p	R ²	Adj. R ²	Erro Absoluto Médio do grupo de treino (%)	Correlação	Erro Absoluto Médio do grupo de teste (%)
	Interseção	-3273,252	-5873,079	-673,424	-3,496	0,025					
	Frequência dos dias de precipitação elevada (5 vezes a precipitação média diária)	158,352	37,956	278,747	3,652	0,022	0,769	0,712	-28,490	0,877	-134,655
	Interseção	-16,326	-191,304	158,651	-0,259	0,808					
	Razão entre a capacidade total de armazenamento do reservatório da bacia hidrográfica e a sua vazão anual total	2,115	0,479	3,752	3,589	0,023	0,763	0,704	-27,874	0,874	-13,179
	Interseção	394,931	-56,098	845,960	2,431	0,072					
	vazão diária média	-2617,639	-7616,388	2381,111	-1,454	0,220	0,346	0,182	4,909	-0,588	-299,815
	Interseção	328,986	-169,140	827,113	1,834	0,141					
X1	Taxa de escoamento	-3883,047	-15295,830	7529,735	-0,945	0,398	0,182	-0,022	0,000	-0,427	-47,902

Variável dependente	Variável independente	Est.	2,50%	97,50%	t val.	p	R²	Adj. R²	Erro Absoluto Médio do grupo de treino (%)	Correlação	Erro Absoluto Médio do grupo de teste (%)
	Interseção	-209,160	-961,811	543,491	-0,772	0,483					
	Declive médio da bacia hidrográfica	8,099	-7,874	24,072	1,408	0,232	0,331	0,164	0,000	0,576	32,702
	Interseção	-841,878	-1563,914	-119,843	-3,237	0,032					
	Duração média dos eventos de vazão máxima	119,461	35,525	203,396	3,952	0,017	0,796	0,745	-77,880	0,892	29,942
	Interseção	-93,030	-148,094	-37,967	-4,691	0,009					
	Razão entre a capacidade total de armazenamento do reservatório da bacia hidrográfica e a sua vazão anual total	0,252	-0,231	0,735	1,447	0,222	0,344	0,180	-48,208	0,586	-58,250
	Interseção	-143,476	-461,365	174,412	-1,253	0,278					
X2	Classe geológica mais comum na bacia hidrográfica (rochas metamórficas)	1,207	-4,234	6,647	0,616	0,571	0,087	-0,142	-89,310	0,328	-65,947
	Interseção	2477,435	248,673	4706,196	3,086	0,037					
	*Classe geológica mais comum na bacia hidrográfica	-35,547	-73,959	2,865	-2,569	0,062	0,623	0,528	-69,890	-0,789	18,596
	Interseção	169,191	-1232,726	1571,108	0,335	0,754					
X3	Porosidade subterrânea da bacia hidrográfica	5529,271	-14687,014	25745,556	0,759	0,490	0,126	-0,093	-132,975	0,355	-152,391
	Interseção	-0,698	-3,843	2,448	-0,616	0,571					
	Índice de vazão de base	8,033	-0,310	16,375	2,673	0,056	0,641	0,551	-7,834	0,801	-15,331
	Interseção	4,009	2,336	5,683	6,652	0,003					
	Frequência dos dias de vazão máxima	-0,055	-0,109	-0,002	-2,891	0,045	0,676	0,595	-4,294	-0,822	2,136
	Interseção	4,961	3,894	6,028	12,904	0,000					
	Frequência dos dias de vazão mínima	-0,018	-0,025	-0,011	-7,229	0,002	0,929	0,911	-0,068	-0,964	10,997
	Interseção	3,819	1,908	5,730	5,549	0,005					
X4	Duração média dos eventos de vazão mínima	-0,064	-0,131	0,004	-2,615	0,059	0,631	0,539	-9,061	0,794	-11,951

Variável dependente	Variável independente	Est.	2,50%	97,50%	t val.	p	R ²	Adj. R ²	Erro Absoluto Médio do grupo de treino (%)	Correlação	Erro Absoluto Médio do grupo de teste (%)
	Interseção	1,850	0,486	3,213	3,766	0,020					
	Percentagem coberta por zonas áridas	1,872	-0,434	4,177	2,254	0,087	0,560	0,449	-11,986	0,748	2,326
	Interseção	38,296	16,259	60,332	4,825	0,009					
	Longitude	0,900	0,341	1,459	4,473	0,011	0,833	0,792	-3,263	0,913	-53,676
	Interseção	4,639	3,221	6,057	9,084	0,001					
	Elevação do medidor	-0,012	-0,019	-0,004	-4,333	0,012	0,824	0,781	-4,056	-0,908	-43,704
	Interseção	4,594	-4,965	14,153	1,334	0,253					
	Vazão diária média	-37,038	-197,636	123,560	-0,640	0,557	0,093	-0,134	-27,110	-0,305	64,599
	Interseção	3,451	-0,316	7,218	2,543	0,064					
	Taxa de escoamento	-16,982	-96,774	62,811	-0,591	0,586	0,080	-0,150	-18,603	-0,283	-81,439
	Interseção	1,296	0,607	1,984	5,224	0,006					
	Q5 (vazão mínima)	353,344	198,630	508,058	6,341	0,003	0,910	0,887	-3,665	0,954	-65,703
	Interseção	5,668	-11,372	22,708	0,924	0,408					
	Duração média dos eventos de vazão máxima	-0,393	-2,457	1,671	-0,529	0,625	0,065	-0,168	-19,955	-0,256	-16,915
	Interseção	3,332	-2,070	8,733	1,713	0,162					
	Declive médio da bacia hidrográfica	-0,024	-0,150	0,103	-0,519	0,631	0,063	-0,171	-19,832	-0,251	-3,187

*A classe geológica mais comum no Grupo 1 é a metamórfica (100%).

Fonte: Próprio autor (2023).

Tabela 27 - Estatística descritiva das equações geradas utilizando as estações do Cluster 2.

Variável dependente	Variável independente	Est.	2,50%	97,50%	t val.	p	R ²	Adj. R ²	Erro Absoluto Médio do grupo de treino (%)	Correlação	Erro Absoluto Médio do grupo de teste (%)
X1	Interseção	-2148,967	-9168,829	4870,895	-0,749	0,482					

Variável dependente	Variável independente	Est.	2,50%	97,50%	t val.	p	R ²	Adj. R ²	Erro Absoluto Médio do grupo de treino (%)	Correlação	Erro Absoluto Médio do grupo de teste (%)
	Frequência dos dias de precipitação elevada (5 vezes a precipitação média diária)	101,953	-208,756	412,661	0,803	0,453	0,097	-0,054	-44690,582	-0,301	30,397
	Interseção	42,033	-350,282	434,347	0,262	0,802					
	Razão entre a capacidade total de armazenamento do reservatório da bacia hidrográfica e a sua vazão anual total	5,736	-4,000	15,472	1,442	0,200	0,257	0,134	-33706,923	0,507	-51,528
	Interseção	593,784	170,596	1016,972	3,433	0,014					
	Vazão diária média	-1108,360	-2127,042	-89,678	-2,662	0,037	0,542	0,465	-98761,405	-0,736	-1631,873
	Interseção	586,679	183,345	990,013	3,559	0,012					
	Taxa de escoamento	-3411,077	-6574,981	-247,173	-2,638	0,039	0,537	0,460	-270,297	0,880	82,975
	Interseção	1385,428	34,612	2736,243	2,510	0,046					
	Declive médio da bacia hidrográfica	-21,576	-46,241	3,089	-2,140	0,076	0,433	0,339	-106982,640	0,916	466,481
	Interseção	-421,179	-934,107	91,748	-2,009	0,091					
	Duração média dos eventos de vazão máxima	104,294	13,452	195,136	2,809	0,031	0,568	0,496	-84222,218	-0,662	-157,298
	Interseção	-65,801	-111,640	-19,962	-3,513	0,013					
	Razão entre a capacidade total de armazenamento do reservatório da bacia hidrográfica e a sua vazão anual total	-0,043	-1,318	1,232	-0,082	0,937	0,001	-0,165	-521,015	-0,034	-134,471
	Interseção	-158,698	-352,823	35,428	-2,000	0,092					
X2	*Classe geológica mais comum na bacia hidrográfica	1,937	-1,587	5,461	1,345	0,227	0,232	0,104	-427,991	0,094	-3,980
	Interseção	1257,067	220,803	2293,331	2,968	0,025					
	*Classe geológica mais comum na bacia hidrográfica	-14,292	-33,378	4,793	-1,832	0,117	0,359	0,252	-30,418	-0,599	3,151
	Interseção	997,666	272,172	1723,160	3,365	0,015					
X3	Porosidade subterrânea da bacia hidrográfica	-3631,864	-8458,558	1194,831	-1,841	0,115	0,361	0,255	-36,390	-0,601	-15,710
	Interseção	1,503	-6,492	9,499	0,483	0,649					
X4	Índice de vazão de base	0,842	-13,693	15,376	0,149	0,888	0,004	-0,195	-26,715	-0,122	30,172

Variável dependente	Variável independente	Est.	2,50%	97,50%	t val.	p	R²	Adj. R²	Erro Absoluto Médio do grupo de treino (%)	Correlação	Erro Absoluto Médio do grupo de teste (%)
	Interseção	1,942	-0,801	4,685	1,732	0,134					
	Frequência dos dias de vazão máxima	0,024	-0,121	0,168	0,396	0,706	0,026	-0,137	-3,583	0,160	-20,376
	Interseção	2,513	0,264	4,763	2,734	0,034					
	Frequência dos dias de vazão mínima	-0,004	-0,034	0,027	-0,285	0,785	0,013	-0,151	-14,185	-0,115	-8,407
	Interseção	-0,466	-5,635	4,704	-0,220	0,833					
	Duração média dos eventos de vazão mínima	0,151	-0,130	0,432	1,313	0,237	0,223	0,094	-5,849	0,472	17,174
	Interseção	2,435	1,449	3,420	6,047	0,001					
	Porcentagem coberta por zonas áridas	-7,570	-22,453	7,312	-1,245	0,260	0,205	0,073	-16,123	-0,453	-9,833
	Interseção	-10,246	-61,078	40,587	-0,493	0,639					
	Longitude	-0,309	-1,567	0,948	-0,602	0,569	0,057	-0,100	2,546	-0,239	-17,943
	Interseção	1,573	-0,032	3,179	2,398	0,053					
	Elevação do medidor	0,003	-0,002	0,008	1,477	0,190	0,267	0,144	5,319	0,516	-30,793
	Interseção	3,450	2,084	4,815	6,182	0,001					
	Vazão diária média	-2,833	-6,024	0,358	-2,172	0,073	0,440	0,347	0,134	-0,663	-48,168
	Interseção	3,522	2,395	4,648	7,647	0,000					
	Taxa de escoamento	-9,982	-17,280	-2,683	-3,346	0,016	0,651	0,593	-4,268	-0,807	16,925
	Interseção	3,141	1,996	4,286	6,710	0,001					
	Q5 (vazão mínima)	-20,729	-50,673	9,215	-1,694	0,141	0,324	0,211	6,646	-0,569	-30,572
	Interseção	0,745	-0,887	2,376	1,117	0,307					
	Duração média dos eventos de vazão máxima	0,258	0,040	0,476	2,895	0,028	0,583	0,513	3,048	0,763	-8,710
	Interseção	5,048	0,617	9,479	2,788	0,032					
	Declive médio da bacia hidrográfica	-0,048	-0,127	0,031	-1,493	0,186	0,271	0,149	3,177	-0,520	-20,536

*A classe geológica mais comum no Grupo 1 é a metamórfica (100%), no grupo 2 são as rochas metamórficas (50%), os sedimentos não consolidados (33%) e as rochas sedimentares siliciclásticas (17%). No Grupo 3 são as rochas sedimentares siliciclásticas (100%).

Fonte: Próprio autor (2023).

Tabela 28 - Estatística descritiva das equações geradas utilizando as estações do Cluster 3.

Variável dependente	Variável independente	Est.	2,50%	97,50%	t val.	p	R ²	Adj. R ²	Erro Absoluto Médio do grupo de treino (%)	Correlação	Erro Absoluto Médio do grupo de teste (%)
	Interseção	336,253	-1307,224	1979,729	2,600	0,234					
	Frequência dos dias de precipitação elevada (5 vezes a precipitação média diária)	-18,496	-121,433	84,440	-2,283	0,263	0,839	0,678	-2,461	-0,916	-63,024
	Interseção	4,699	-269,327	278,726	0,218	0,863					
	Razão entre a capacidade total de armazenamento do reservatório da bacia hidrográfica e a sua vazão anual total	1,609	-8,001	11,219	2,128	0,280	0,819	0,638	-5,816	0,905	-87,165
	Interseção	-59,345	-2020,621	1901,931	-0,385	0,766					
	Vazão diária média	161,774	-3145,012	3468,560	0,622	0,646	0,279	-0,443	-86,506	0,528	34,127
	Interseção	-229,981	-758,581	298,619	-5,528	0,114					
	Taxa de escoamento	1571,841	-1435,402	4579,085	6,641	0,095	0,978	0,956	-18,209	0,989	-167,552
	Interseção	-89,651	-3241,500	3062,199	-0,361	0,779					
	Declive médio da bacia hidrográfica	4,072	-98,046	106,189	0,507	0,702	0,204	-0,592	-110,775	0,452	18,041
	Interseção	49,128	-505,999	604,255	1,125	0,463					
X1	Duração média dos eventos de vazão máxima	-0,751	-70,525	69,023	-0,137	0,914	0,018	-0,963	-285,949	-0,135	-22,906
	Interseção	-72,012	-107,046	-36,979	-26,118	0,024					
	Razão entre a capacidade total de armazenamento do reservatório da bacia hidrográfica e a sua vazão anual total	-0,714	-1,927	0,499	-7,481	0,085	0,982	0,965	-0,027	-0,991	-3,483
	Interseção	-131,187	-242,900	-19,474	-14,921	0,043					
X2	*Classe geológica mais comum na bacia hidrográfica	1,161	-1,791	4,113	4,998	0,126	0,962	0,923	-0,066	0,981	-2,973
X3	Interseção	6468,382	-23146,144	36082,909	2,775	0,220					

Variável dependente	Variável independente	Est.	2,50%	97,50%	t val.	p	R ²	Adj. R ²	Erro Absoluto Médio do grupo de treino (%)	Correlação	Erro Absoluto Médio do grupo de teste (%)
	*Classe geológica mais comum na bacia hidrográfica	-111,742	-897,614	674,131	-1,807	0,322	0,766	0,531	-8,352	-0,875	4,854
	Interseção	28283,646	-7626,600	64193,892	10,008	0,063					
	Porosidade subterrânea da bacia hidrográfica	-159988,346	-383806,710	63830,018	-9,083	0,070	0,988	0,976	-0,078	-0,994	-38,299
X4	Interseção	-2,581	-101,206	96,045	-0,333	0,796					
	Índice de vazão de base	7,941	-124,903	140,785	0,760	0,587	0,366	-0,268	-5,955	0,605	20,785
	Interseção	3,784	-11,470	19,037	3,152	0,196					
	Frequência dos dias de vazão máxima	-0,212	-4,302	3,878	-0,659	0,629	0,303	-0,395	-6,730	-0,550	19,926
	Interseção	4,620	1,825	7,415	21,000	0,030					
	Frequência dos dias de vazão mínima	-0,119	-0,366	0,129	-6,088	0,104	0,974	0,948	-0,152	-0,987	-33,165
	Interseção	4,830	4,703	4,957	483,000	0,001					
	Duração média dos eventos de vazão mínima	-0,136	-0,148	-0,124	-146,070	0,004	1,000	1,000	0,000	-1,000	-44,001
	Interseção	-8,105	-70,184	53,973	-1,659	0,345					
	Porcentagem coberta por zonas áridas	216,186	-889,058	1321,430	2,485	0,244	0,861	0,721	-1,768	0,928	49,278
	Interseção	-16,264	-16,913	-15,615	-318,288	0,002					
	Longitude	-0,488	-0,503	-0,472	-396,994	0,002	1,000	1,000	0,000	-1,000	-47,947
	Interseção	2,152	-18,101	22,406	1,350	0,406					
	Elevação do medidor	0,004	-0,054	0,062	0,840	0,555	0,414	-0,173	-5,509	0,643	17,176
	Interseção	-5,735	-60,373	48,904	-1,334	0,410					
	Vazão diária média	18,394	-84,343	121,130	2,275	0,263	0,838	0,676	-2,057	0,915	-82,470
	Interseção	-2,231	-75,343	70,880	-0,388	0,765					
	Taxa de escoamento	33,984	-371,236	439,204	1,066	0,480	0,532	0,064	-4,870	0,729	-26,106
	Interseção	1,257	-9,031	11,546	1,553	0,364					
	Q5 (vazão mínima)	12,211	-32,664	57,087	3,458	0,179	0,923	0,846	-0,633	0,961	-34,211
Interseção	2,223	-7,887	12,334	2,794	0,219						

Variável dependente	Variável independente	Est.	2,50%	97,50%	t val.	p	R²	Adj. R²	Erro Absoluto Médio do grupo de treino (%)	Correlação	Erro Absoluto Médio do grupo de teste (%)
	Duração média dos eventos de vazão máxima	0,181	-1,090	1,452	1,810	0,321	0,766	0,532	-3,075	0,875	14,653
	Interseção	-8,436	-44,917	28,045	-2,938	0,209					
	Declive médio da bacia hidrográfica	0,412	-0,793	1,617	4,346	0,144	0,950	0,899	-0,654	0,975	-48,623

*A classe geológica mais comum no Grupo 1 é a metamórfica (100%), no grupo 2 são as rochas metamórficas (50%), os sedimentos não consolidados (33%) e as rochas sedimentares siliciclásticas (17%). No Grupo 3 são as rochas sedimentares siliciclásticas (100%).

Fonte: Próprio autor (2023).

5.4. OBTENÇÃO DAS VAZÕES DE REFERÊNCIA

Após a obtenção das vazões de referência, foi observado que o erro absoluto de algumas estações foi relativamente alto (Tabela 29). O erro absoluto médio da Q90 foi de -16,60%, o sinal negativo indica que a Q90 média das vazões simuladas apresentou valores superiores que a observada. Já para a Q7,10, o erro médio foi de 103,80%, o sinal positivo indica que a Q7,10 média das vazões observadas foram superiores. Ressalta-se que foram obtidos valores negativos para a Q7,10 de algumas estações. A justificativa é que as vazões são calculadas de forma estatísticas.

Tabela 29 - Vazões de referência para as estações localizadas no semiárido da Bahia.

Estação	Cluster	Q7,10 (m³/s)			Q90 (m³/s)		
		Observada	Simulada	Erro (%)	Observada	Simulada	Erro (%)
45170001	Grupo 1	39,96	33,64	15,81	44,25	43,10	2,59
45260000		78,15	56,20	28,08	90,50	78,79	12,94
45910001		127,73	112,85	11,65	144,59	143,32	0,88
45960001		131,28	108,29	17,51	147,57	137,78	6,63
46650000		115,64*	100,81	12,82	124,78*	125,29	-0,41
46770000		24,05	22,54	6,25	26,55	30,76	-15,86
46830000		66,00	50,61	23,31	72,95	69,49	4,75
51120000		0,33	0,47	-40,71	1,40	1,24	11,34
47480000		Grupo 2	-0,06*	0,04	165,36	0,13*	0,08
50191000	-0,14*		0,09	168,12	1,11*	0,49	55,83
50250000	-0,04*		-0,01	58,71	0,00*	0,01	-145,07
50290000	0,31*		0,19	41,13	1,24*	0,78	37,25
50330000	-0,02*		-	-	0,20*	-	-
50380000	-0,06*		0,43	869,75	0,37*	0,65	-72,66
50420000	-0,03*		0,06	275,09	0,05*	0,11	-124,15
50465000	-0,24		0,97	508,75	0,69	1,30	-89,00
50540000	-0,36*		0,71	296,02	1,36*	1,12	17,33
50595000	2,06		1,81	11,89	4,33	2,87	33,77
50620000	0,46*		-0,03	107,22	0,81*	0,05	94,34
50660000	1,05*		0,69	34,41	2,55*	1,84	27,93
51140000	1,23		0,57	53,85	1,98	1,19	40,09
51230000	-0,15*		0,54	451,09	0,35*	1,20	-242,97
51240000	2,82		4,07	-44,36	7,08	8,34	-17,71
51280000	1,32		4,28	-225,35	5,72	7,96	-39,08
51350000	-0,26*		3,31	1356,13	3,55*	6,89	-93,98
51460000	-0,07*		0,00	106,48	0,01*	0,04	-673,34
51560000	0,11		0,46	-319,01	1,34	1,31	2,70
51685000	0,83*		1,14	-36,45	2,56*	2,31	9,56
51840000	-0,12*		0,14	218,09	0,18*	0,29	-63,80
52050000	-0,08*		0,01	113,50	0,06*	0,09	-63,58
52404000	-0,26*		0,18	170,94	0,11*	0,34	-226,40
52405000	-0,50		0,27	152,93	0,46	0,49	-7,16
52570000	-1,06*		0,75	170,61	2,68*	1,55	42,14
52680000	4,44		2,45	44,79	11,51	4,20	63,50
52760000	0,18		0,77	-323,22	3,59*	2,17	39,64
52790000	-0,79		2,15	373,28	5,55*	6,21	-11,82
53125000	-0,07		0,03	140,89	0,21*	0,10	50,46
53620000	0,23		0,30	-31,69	1,07	0,52	50,83
53630000	0,46	0,37	18,89	1,27*	0,83	34,52	
53650000	0,42	0,59	-40,05	1,49*	1,02	31,54	
53780000	0,87	0,65	25,75	2,92*	1,50	48,61	
53950000	6,48	4,31	33,41	13,34	7,29	45,35	

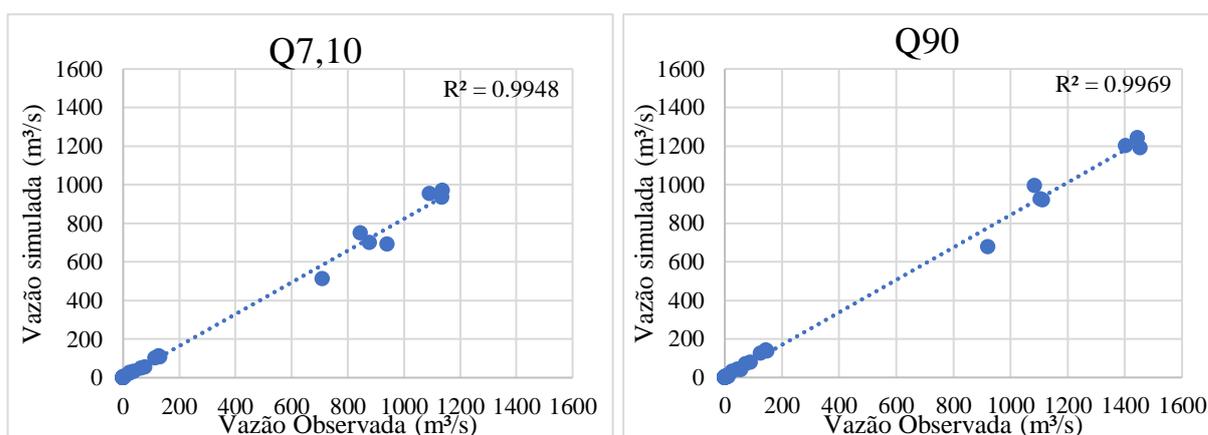
Estação	Cluster	Q7,10 (m ³ /s)			Q90 (m ³ /s)		
		Observada	Simulada	Erro (%)	Observada	Simulada	Erro (%)
54950000	Grupo 3	25,39	26,68	-5,10	56,57	39,46	30,25
45298000		709,35	511,99	27,82	920,21	676,92	26,44
46035000		845,95	751,26	11,19	1084,32	995,76	8,17
46105000		877,95	701,46	20,10	1105,43	927,03	16,14
46360000		940,88*	691,05	26,55	1112,93*	921,57	17,19
48020000		1135,58	936,02	17,57	1453,33	1190,76	18,07
48290000		1137,39	971,26	14,61	1444,66	1243,14	13,95
49330000		1092,31	955,14	12,56	1403,40	1202,96	14,28
50795000		2,65	1,52	42,48	5,87	4,72	19,65
52370000		-0,70*	0,03	103,90	0,32*	0,07	76,74

*estações com falhas.

Fonte: Próprio autor (2023).

A grande diferença em relação ao erro absoluto médio pode estar relacionada a algumas estações que o modelo não conseguiu simular de forma eficiente. As falhas contidas nos dados observados também impossibilitaram uma comparação mais eficiente. Neste caso, o preenchimento das falhas por regressão, não é indicado, porque, podem ser introduzidos dados tendenciosos (TUCCI, 2017). Tucci (2017) afirma ainda que, caso as falhas ocorram em períodos chuvosos, com vazões mais altas, as informações podem ser utilizadas, porque o intervalo de vazões altas geralmente não é utilizado associado a curva de permanência. Esta também pode ser uma justificativa para explicar o R^2 médio superior a 0,99 entre as vazões simuladas e observadas da Q90 e Q7,10 (Figura 30).

Figura 30 - Correlações entre as vazões de referência Q90 e Q7,10 obtidas pelas vazões observadas e as simuladas no período de calibração do modelo GR4J.



Fonte: Próprio autor (2023).

O R^2 médio alto indica uma forte relação linear entre duas variáveis, nesse caso as vazões simuladas e observadas tanto para a Q90 quanto para a Q7,10. No estudo realizado por Kuana *et al.* (2021) em 136 bacias do estado do Paraná, as correlações entre as vazões Q95 obtidas pelas vazões observadas e as simuladas, apresentam correlações superiores a 0,98,

indicando que o modelo GR4J foi capaz de reproduzir índices de vazões mínimas semelhantes aos observados. Os autores obtiveram mediana dos erros relativos para o modelo GR4J na primeira e na segunda fase de 14,09% e 12,92%, respectivamente.

Ao analisar as médias da Q90 e Q7,10 dentro de cada cluster, foi observado que no Grupo 2 apresentou erro relativamente alto para a Q7,10 (145%) e Q90 (-31,76%) (Tabela 30).

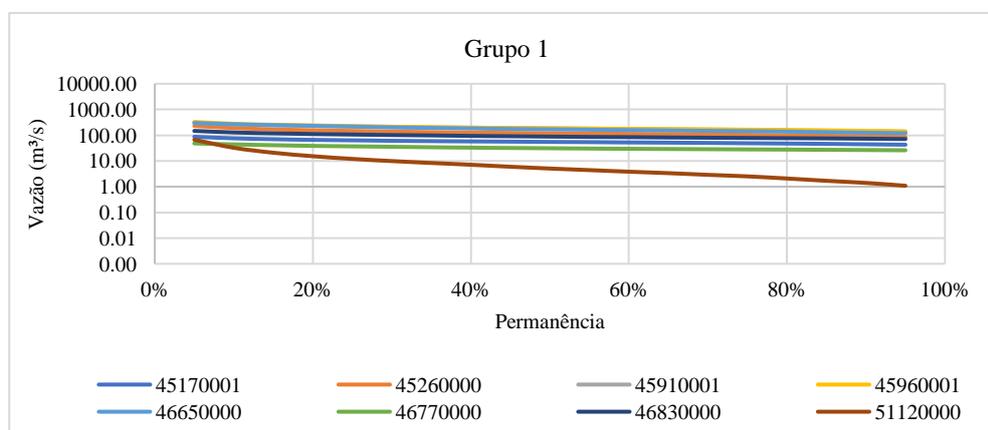
Tabela 30 - Valores médio das vazões de referência para os clusters localizadas no semiárido da Bahia.

Grupo	Quantidade de estações	Q7,10 (m ³ /s)			Q90 (m ³ /s)		
		Observada	Simulada	Erro (%)	Observada	Simulada	Erro (%)
1	8	66,79	60,68	9,34	75,40	78,72	2,86
2	35	2,61	1,73	145,35	9,46	3,08	-31,76
3	9	828,74	613,30	30,75	1059,60	795,88	23,40

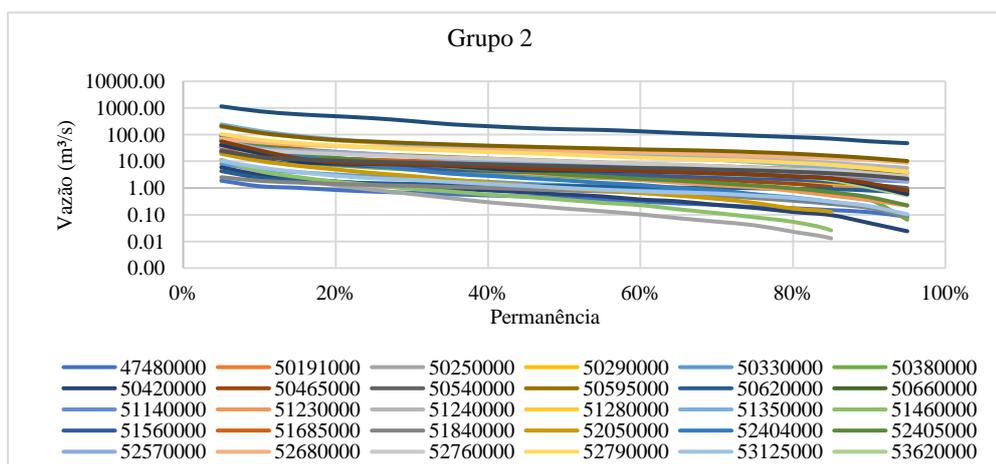
Fonte: Próprio autor (2023).

A fim de verificar o comportamento das curvas de permanência dentro de cada grupo, foi realizado os agrupamentos expostos na Figura 31, Figura 32 e Figura 33. Ressalta-se que no APÊNDICE D é possível a visualização das curvas de permanência de cada estação, considerando os dados observados, obtidos utilizando o CAMELS e os dados simulados, através do modelo GR4J.

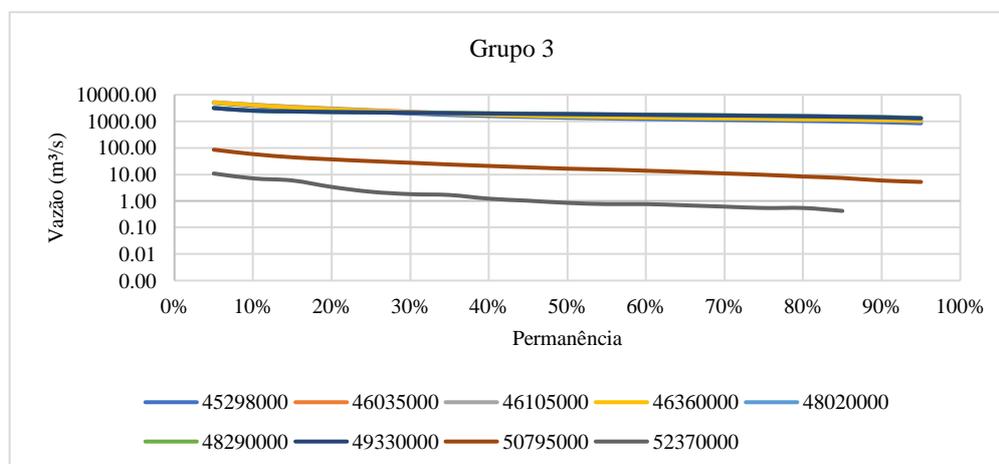
Figura 31 - Curvas de permanência do grupo 1 obtido pela técnica *k-means*.



Fonte: Próprio autor (2023).

Figura 32 - Curvas de permanência do grupo 2 obtido pela técnica *k-means*.

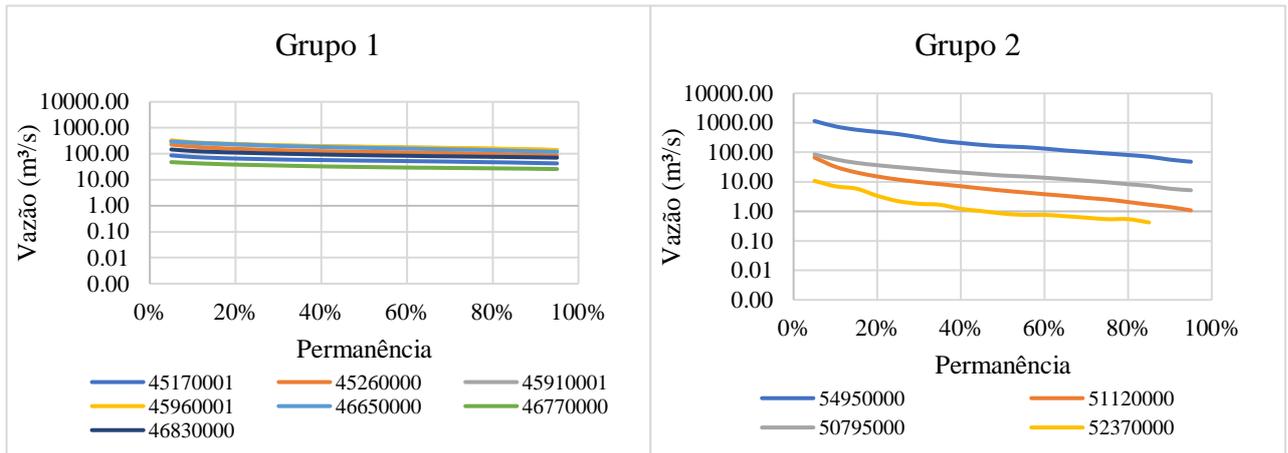
Fonte: Próprio autor (2023).

Figura 33 - Curvas de permanência do grupo 3 obtido pela técnica *k-means*.

Fonte: Próprio autor (2023).

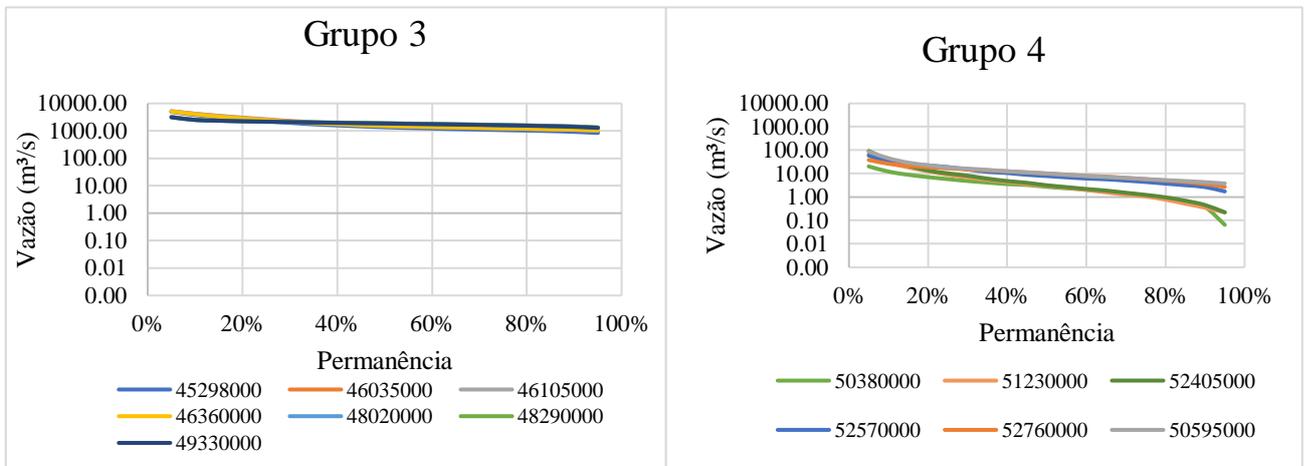
Por se tratar do processo de regionalização, houve a necessidade de reagrupar as estações de forma manual, de modo que o comportamento das curvas de permanências fosse o mais similar possível (Figura 34, Figura 35, Figura 36, e Figura 37).

Figura 34 - Curvas de permanência dos Grupos 1 e 2.



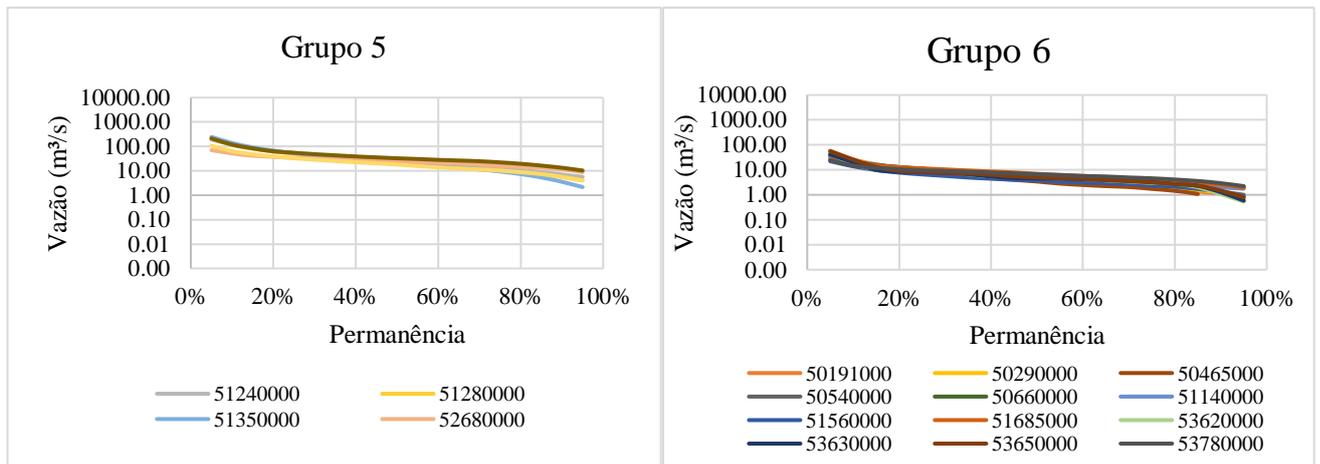
Fonte: Próprio autor (2023).

Figura 35 - Curvas de permanência dos Grupos 3 e 4.



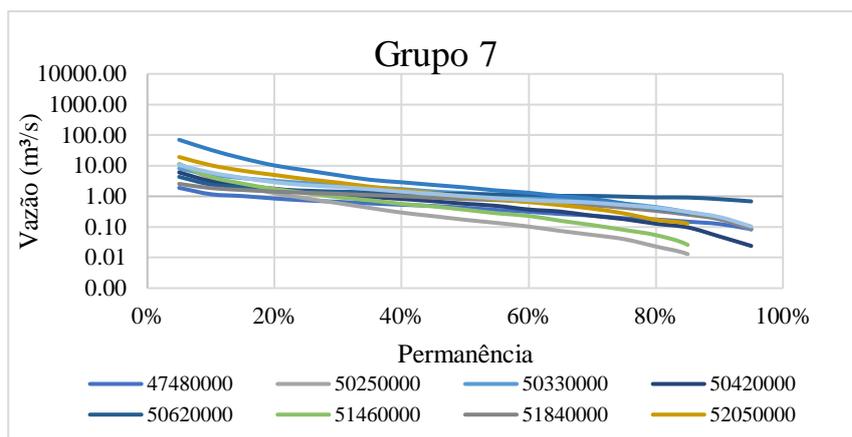
Fonte: Próprio autor (2023).

Figura 36 - Curvas de permanência dos Grupos 5 e 6.



Fonte: Próprio autor (2023).

Figura 37 - Curvas de permanência do Grupo 7.



Fonte: Próprio autor (2023).

Com o novo agrupamento das estações, observa-se uma menor discrepância em relação aos valores da Q90 tanto observada quanto simulada (Tabela 31 e Tabela 32). Ressalta-se que o Grupo 2 é formado pelas estações que não apresentaram similaridade com os demais grupos.

Tabela 31 - Estatística descritiva da Q90 (m³/s) com dados observados de vazão por cluster.

Estatística Descritiva	Grupo 1	Grupo 2	Grupo 3	Grupo 4	Grupo 5	Grupo 6	Grupo 7
Mínimo	26,55	0,32	920,21	0,35	3,55	0,69	0,00
Percentil 5%	31,86	0,48	969,44	0,36	4,05	0,90	0,00
Percentil 95%	146,68	48,97	1450,73	4,15	12,88	2,72	0,54
Média	93,03	16,04	1217,75	1,96	7,79	1,63	0,18
Mediana	90,50	3,64	1112,93	1,57	6,40	1,35	0,12
Máximo	147,57	56,57	1453,33	4,33	13,34	2,92	0,81

Fonte: Próprio autor (2023).

Tabela 32 - Estatística descritiva da Q90 (m³/s) com dados simulados de vazão por cluster.

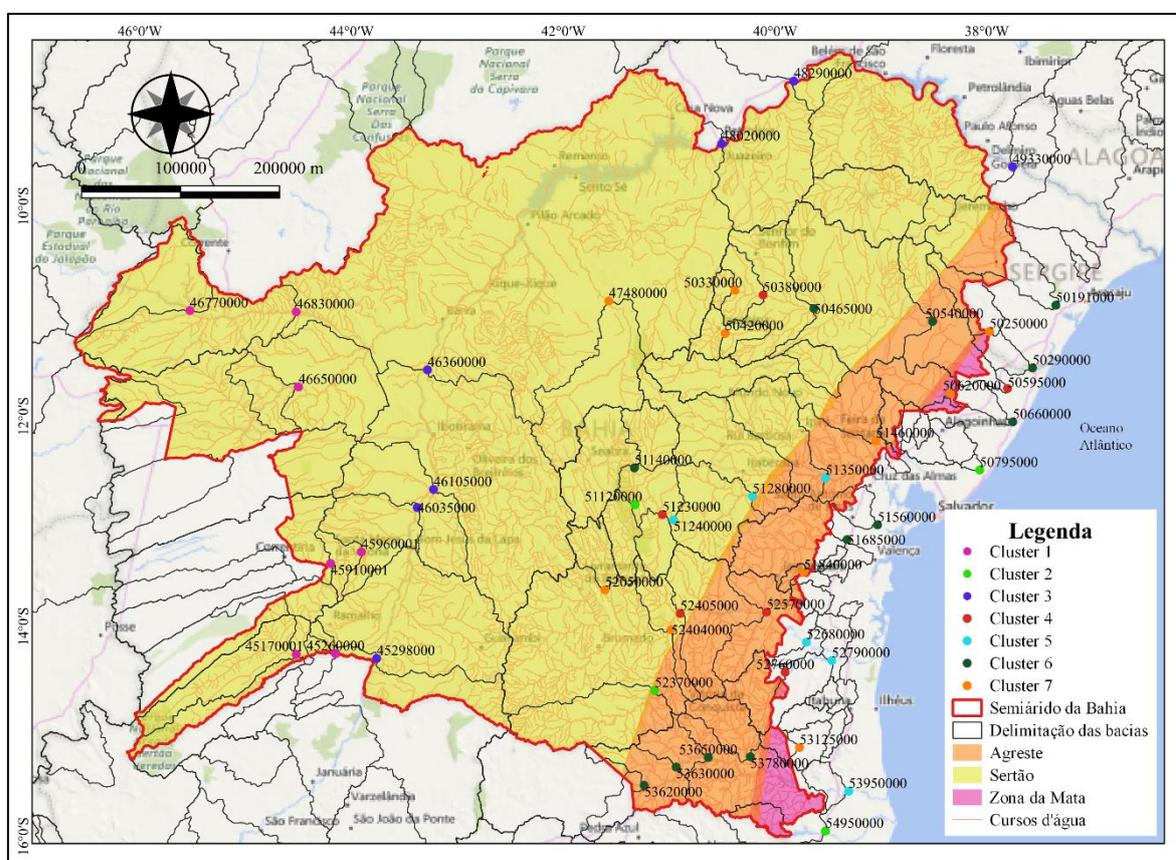
Estatística Descritiva	Grupo 1	Grupo 2	Grupo 3	Grupo 4	Grupo 5	Grupo 6	Grupo 7
Mínimo	30,76	0,07	676,92	0,49	4,20	0,49	0,01
Percentil 5%	34,46	0,05	750,32	0,53	4,70	0,51	0,02
Percentil 95%	141,66	34,25	1231,09	2,70	8,25	2,05	0,32
Média	89,79	11,37	1022,59	1,49	6,82	1,18	0,12
Mediana	78,79	2,98	995,76	1,38	7,09	1,16	0,09
Máximo	143,32	39,46	1243,14	2,87	8,34	2,31	0,34

Fonte: Próprio autor (2023).

Ao regionalizar a Q90 a partir da área foi obtido resultado satisfatório apenas para o grupo 3 (Tabela 33). A área apresentou correlação de 99% com a Q90. O P-valor foi inferior a 0,05, indicando que a relação entre a área e a Q90 é significativa. O coeficiente de determinação indicou que aproximadamente 99% da variação do comportamento da Q90 neste grupo pode ser explicada pelo comportamento da variação da área. As equações apresentaram um erro médio -0,091% para as amostras de treino, indicando que os valores previstos foram praticamente iguais aos obtidos através das vazões observadas. Já para as amostras de teste o erro foi de -3,746%. O sinal negativo evidenciou que os valores previstos utilizando a equação linear foram maiores que os obtidos através dos dados do CAMELS.

Ao especializar os grupos, foi observado que as estações do cluster 1 estão localizadas ao longo da bacia do São Francisco e as estações do cluster 3 estão geograficamente próximas (Figura 38). Não foram encontradas relações geográficas para os demais clusters. Estes apresentaram estações com grande variabilidade em relação a localização.

Figura 38 - Espacialização dos sete clusters formados após análise das curvas de permanência.



Fonte: Próprio autor (2023).

Tabela 33 - Estatística descritiva das equações geradas utilizando as estações dos grupos e as áreas de correspondentes.

Grupo	Variável dependente	Variável independente	Est.	2,50%	97,50%	t val.	p	R ²	Adj. R ²	Erro Absoluto Médio do grupo de treino (%)	Correlação	Erro Absoluto Médio do grupo de teste (%)
Grupo 1	Vazão Simulada	Interseção	9,848	-112,070	131,767	0,348	0,761					
		Área	0,004	-0,001	0,008	3,823	0,062	0,880	0,819	-4,194	0,938	5,767
	Vazão Observada	Interseção	4,694	-137,294	146,682	0,142	0,900					
		Área	0,004	-0,001	0,009	3,505	0,073	0,860	0,790	-5,907	0,927	15,459
Grupo 2	Vazão Simulada	Interseção	-2,377	-34,063	29,308	-0,953	0,515					
		Área	0,001	0,000	0,001	9,710	0,065	0,990	0,979	-1079,399	0,995	89,613
	Vazão Observada	Interseção	-3,510	-44,873	37,853	-1,078	0,476					
		Área	0,001	0,000	0,002	10,686	0,059	0,991	0,983	-256,919	0,996	89,608
Grupo 3	Vazão Simulada	Interseção	361,181	-483,320	1205,682	1,840	0,207					
		Área	0,002	0,000	0,004	3,550	0,071	0,863	0,795	-1,086	0,929	-3,284
	Vazão Observada	Interseção	486,069	203,525	768,613	7,402	0,018					
		Área	0,002	0,001	0,003	11,791	0,007	0,986	0,979	-0,091	0,993	-3,746
Grupo 4	Vazão Simulada	Interseção	1,625	-2,585	5,835	1,661	0,239					
		Área	0,000	0,000	0,000	0,078	0,945	0,003	-0,496	-56,119	0,055	-82,256
	Vazão Observada	Interseção	1,831	-6,387	10,048	0,959	0,439					
		Área	0,000	0,000	0,000	0,244	0,830	0,029	-0,457	-56,119	0,170	-203,670
Grupo 5	Vazão Simulada	Interseção	8,205	1,115	15,295	4,979	0,038					
		Área	0,000	0,000	0,000	-1,113	0,382	0,382	0,073	-4,102	-0,618	16,671
	Vazão Observada	Interseção	4,171	-1,380	9,722	3,233	0,084					
		Área	0,000	0,000	0,000	3,065	0,092	0,825	0,737	-1,677	0,908	-68,161
Grupo 6	Vazão Simulada	Interseção	1,438	0,563	2,314	4,019	0,007					
		Área	0,000	0,000	0,000	-0,950	0,379	0,131	-0,014	-18,151	-0,361	-22,294
	Vazão Observada	Interseção	2,319	1,412	3,227	6,255	0,001					
		Área	0,000	0,000	0,000	-1,872	0,110	0,369	0,263	-7,760	-0,607	-66,719
Grupo 7	Vazão Simulada	Interseção	0,108	-0,043	0,259	1,841	0,125					
		Área	0,000	0,000	0,000	1,151	0,302	0,209	0,051	-216,121	0,458	-46,355
	Vazão Observada	Interseção	0,263	-0,084	0,609	1,951	0,109					
		Área	0,000	0,000	0,000	-0,590	0,581	0,065	-0,122	0,022	-0,255	-259,247

Fonte: Próprio autor (2023).

6. CONSIDERAÇÕES FINAIS

A ausência de dados hidrológicos é um fator limitante para a gestão e planejamento dos recursos hídricos no semiárido da Bahia. É compreensível que a disponibilidade e a qualidade das informações, a área de estudo e o tamanho da amostra também sejam fatores limitantes para a avaliação do desempenho de cada técnica empregada neste estudo. Entretanto, mesmo com as limitações, a técnica K-médias apresentou resultados satisfatórios para identificar regiões homogêneas no semiárido da Bahia. As características que contribuíram para a heterogeneidade entre os grupos e homogeneidade dentro dos clusters foram a área da bacia hidrográfica, a aridez, a declividade média da bacia, a duração média dos eventos de precipitação máxima, a evapotranspiração média anual, a frequência dos dias de precipitação máxima, a frequências dos dias secos, a longitude, a porosidade subterrânea da bacia hidrográfica, a precipitação média anual, a profundidade mediana do lençol freático e as superfícies artificiais ou zonas urbanas. Um dos fatores observado é que as características físicas e climáticas separadas a partir da análise realizada já foram validadas por autores como Lemos (2006), Kuana *et al.* (2020) e Kuana (2022) na determinação de regiões homogêneas e/ou no processo de regionalização das vazões mínimas. Em contrapartida, o método de *Ward* não foi significativo. Em todas as análises realizadas, a técnica apresentou pelo menos umas das características com o P-valor superior a 0,05. À medida que ia retirando a característica que não apresentou significância, outros atributos começaram a apresentar um P-valor maior que 0,05.

Analisando individualmente o GR4J, o modelo pode auxiliar no processo de obtenção de vazões diárias de algumas estações. Avaliando de forma geral, os resultados não foram satisfatórios. As simulações apresentaram valores de RMSE entre 0,02 e 2,06 na fase de calibração e 0,03 até 1,4 na fase de validação; os intervalos de PBIAS foram -9,00 a 54,70 na fase de calibração e -75 a 227,20 na fase de validação; o NSE variou entre -16,13 e 0,72 na fase de calibração, sendo que menos de 1% das estações apresentaram NSE maior ou igual a 0,70. Na fase de validação os valores de NSE variaram entre -10,81 a 0,69; o R² variou entre 0,05 e 0,76 na fase de calibração, sendo que menos de 1% das estações apresentaram R² maior ou igual a 0,70. Na fase de validação o R² variou entre 0,01 a 0,70; o KGE variou de -1,46 a 0,84 na fase de calibração, sendo que 50% das estações apresentaram valores superiores a 0,70 e de -1,46 a 0,80 na fase de validação. Menos de 1% das estações apresentaram KGE maior ou igual a 0,70. Avaliando as vazões mínimas, o resultado foi melhor, porém os valores de Nslog variaram entre -2933,5 e 0,99 e entre -1416,72 a 0,96. Segundo Oudin *et al.* (2008) bacias mal modeladas não fornecem informações suficientemente relevantes.

Os testes não foram suficientes para afirmar que o modelo não é indicado para ser usado em outras bacias semiáridas, tendo em vista que outros autores, como Araújo (2005) e Andrade *et al.* (2021) obtiveram resultados positivos em bacias localizadas em regiões como esta. A sugestão é que sejam testados outros intervalos para os períodos de calibração e validação do modelo, considerando o ano hidrológico de cada bacia. Ressalta-se que, utilizar um período fixo para todas as estações pode ter interferido no resultado. Principalmente, porque, apesar de aceitar falhas nos dados de vazão, os períodos muito fragmentados podem ter interferido na resposta do modelo.

Também é sugerido que seja utilizada outra função de calibração do modelo GR4J, tendo em vista que a KGE não é a função mais adequada para as vazões mínimas, além de aumentar o tamanho da amostra com o intuito de obter melhores resultados, Perrin *et al.* (2003) utilizou 429 bacias e Kuana (2022) utilizou 126 bacias.

Devido à disponibilidade de dados, o presente trabalho utilizou dados de evapotranspiração calculados por Priestley & Taylor, obtida do produto GLEAM e presente na base de dados do CAMELS - BR. Em um estudo realizado por Kolling Neto (2021) na bacia do rio Paraguaçu na Bahia, também utilizou este produto. O autor atribuiu que o desempenho obtido foi devido a equação que o produto GLEAM é calculado. Também considerou que a Evapotranspiração real do GLEAM tende a acompanhar a ocorrência das chuvas. Porém, conforme Allan *et al.* (1998), o comportamento da evapotranspiração pode ser explicado por fatores como a radiação solar, a temperatura e a umidade relativa do ar, a velocidade dos ventos à 2m de altura, a umidade relativa do ar e a pressão atmosférica. Recomenda-se estimar a evapotranspiração usando fórmulas comuns como Penman-Monteith (Monteith, 1965), ou pela fórmula proposta por Oudin *et al.* (2005).

Tratando-se das relações (equações) entre os valores dos parâmetros do modelo GR4J e as características físicas das bacias, dentro de cada região homogênea, o método de regressão de linear, o qual esperava-se que as equações fossem capazes de prever os parâmetros do modelo, apresentou, em geral, resultados não conclusivos. O esperado era que tamanho e características físicas da bacia fossem capazes de auxiliar na regionalização da capacidade máxima de armazenamento de água na bacia (X1). As características físicas do solo e da vegetação da bacia e a intensidade das chuvas fossem capazes de auxiliar na regionalização do coeficiente de troca subterrânea (X2). As características geológicas da bacia fossem capazes de prever a capacidade máxima de armazenamento de água no aquífero subterrâneo da bacia (X3) e a permeabilidade do solo e do subsolo, e a estrutura do aquífero fossem capazes de auxiliar

na previsão do tempo de base do hidrograma unitário (X4). O resultado pouco satisfatório pode estar relacionado a pequena quantidade de bacias utilizadas para gerar as equações.

Sendo assim, é indicado testar outros tipos de equações como a polinomial (grau dois e três), exponencial e/ou obter as características das bacias de outra fonte de dados, a fim de comparar com os resultados obtidos neste estudo e utilizar apenas as bacias que obtiveram valores de N_{slog} e LOGQ igual ou maior que 70% de eficiência.

O modelo GR4J obteve resultados satisfatórios para a obtenção da vazão de referência Q90, vazão de grande importância no Estado da Bahia para a outorga do uso da água. Ao verificar o desempenho das vazões obtidas, calculando as vazões de referências com dados do modelo GR4J, 40% das estações apresentaram um erro absoluto médio de igual ou inferior a 20% obtido entre as vazões com dados simulados e observados. Já no processo de regionalização das vazões de referência (Q90), é indicado testar outro tipo de equação como, por exemplo, a equação de potência para auxiliar no processo de regionalização e determinar as regiões homogêneas utilizando características como a área de drenagem, densidade de drenagem, o comprimento do rio principal e a declividade média da bacia e do rio principal. Ressalta-se que, mesmo que para algumas bacias os valores das vazões de referências simuladas foram próximos das vazões de referência com os dados observados, é recomendado optar sempre pelos dados medidos. A transferência de dados é indicada somente para os casos de bacias que não possuem dados ou as informações não são confiáveis.

De forma geral, sugere-se a melhoria da qualidade de medição nos dados de evapotranspiração, a fim de que se possa ajustar melhor os modelos hidrológicos e proporcionar estimativas confiáveis de vazões. Essas estimativas são de grande valia para regiões semiáridas. Também seria interessante aplicar a metodologia utilizando outro modelo do tipo chuva-vazão como: IPHS1, Modelo de Grandes Bacias (MGB-IPH), Modelo Hidrológico Autocalibrável (MODHAC), *Soil Moisture Accounting Procedure* (SMAP), *Soil and Water Assessment Tool* (SWAT), *TOPmodel SIMPLifié* (TOPSIMPL), WASA-SED, tendo em vista que estes modelos também foram testados em regiões semiáridas brasileiras, para posteriormente comparar com o modelo GR4J.

REFERÊNCIAS

- ACREMAN, M. C.; SINCLAIR, C. D. Classification of drainage basins according to their physical characteristics; an application for flood frequency analysis in Scotland. **Journal of Hydrology**, v. 84, n. 3-4, p. 365-380, maio 1986.
- ALEXANDRE, A. M. B. *et al.* Regionalização de parâmetros de modelos hidrológicos . 2005.
- ALLAN, R.; PEREIRA, L.; SMITH, M. Crop evapotranspiration-Guidelines for computing crop water requirements-FAO Irrigation and drainage paper 56. . 1998.
- AL-QURASHI, A. *et al.* Application of the Kineros2 rainfall-runoff model to an arid catchment in Oman. **Journal of Hydrology**, v. 355, n. 1-4, p. 91-105, jun. 2008.
- ALTHOFF, D.; BAZAME, H. C.; NASCIMENTO, J. G. Untangling hybrid hydrological models with explainable artificial intelligence. **H2Open Journal**, v. 4, n. 1, p. 13-28, 1 jan. 2021.
- ALTHOFF, D.; RODRIGUES, L. N. Goodness-of-fit criteria for hydrological models: Model calibration and performance assessment. **Journal of Hydrology**, v. 600, p. 126674, set. 2021.
- ÁLVAREZ, O.; HOTAIT, S.; SUSTAITA, R. Identificación de regiones hidrológicas homogéneas mediante análisis multivariado. **Ingeniería Investigación y Tecnología**, v. 12, n. n, p. 277-284, 2011.
- ALVES, T. L. B.; AZEVEDO, P. V. DE; SILVA, M. T. Análise morfométrica da microbacia hidrográfica Riacho Namorado, São João do Cariri-PB: uma ferramenta ao diagnóstico físico. **Revista de Geografia (UFPE)**, v. 31, 2014.
- ANA. **Reservatórios**. Disponível em: <<https://metadados.snirh.gov.br/geonetwork/srv/api/records/41f4f780-dcda-43d5-80e3-5c9042eb718a>>. Acesso em: 12 mar. 2023.
- ANDRADE, J. *et al.* Oficina de modelagem hidrológica da UFPE - Aplicação do GR4J para geração de séries sintéticas de vazões diárias em cinco bacias hidrográficas nordestinas. **XXIV Simpósio Brasileiro de Recursos Hídricos**, nov. 2021.
- ANDRÉASSIAN, V.; PERRIN, C.; MICHEL, C. Impact of imperfect potential evapotranspiration knowledge on the efficiency and parameters of watershed models. **Journal of Hydrology**, v. 286, n. 1-4, p. 19-35, jan. 2004.
- ANSHUMAN, A.; KUNNATH-POOVAKKA, A.; ELDHO, T. I. Performance evaluation of conceptual rainfall-runoff models GR4J and AWBM. **ISH Journal of Hydraulic Engineering**, v. 27, n. 4, p. 365-374, 2 out. 2021.
- ARAÚJO FILHO, P. F. DE. **Modelagem hidrológica, com implantação de um limite máximo no déficit hídrico do modelo TOPMODEL, e aplicação na região de**

transição entre a Zona da Mata e o Agreste pernambucano. Recife: Universidade Federal de Pernambuco, maio 2004.

ARAÚJO, P. C. **Aplicação do modelo GR4J à bacia hidrográfica do rio Paraíba em Caraúbas.** Campina Grande-PB: Universidade Federal de Campina Grande, 2005.

ARSENAULT, R. *et al.* Streamflow prediction in ungauged basins: analysis of regionalization methods in a hydrologically heterogeneous region of Mexico. **Hydrological Sciences Journal**, v. 64, n. 11, p. 1297-1311, 18 ago. 2019.

AUFAR, Y.; SITANGGANG, I. S.; ANNISA, -. Parameter Optimization of Rainfall-runoff Model GR4J using Particle Swarm Optimization on Planting Calendar. **International Journal on Advanced Science, Engineering and Information Technology**, v. 10, n. 6, p. 2575, 27 dez. 2020.

AWAN, J. A.; BAE, D.-H.; KIM, K.-J. Identification and trend analysis of homogeneous rainfall zones over the East Asia monsoon region. **International Journal of Climatology**, v. 35, n. 7, p. 1422-1433, 15 jun. 2015.

BARBOSA, N. DA S. **Hidrogeologia do Sistema Aquífero Urucuaia, Bahia.** [s.l.] Universidade Federal da Bahia, 13 maio 2017.

BASSO, R. E. **Estimativa de chuvas intensas em locais sem dados no Rio Grande do Sul.** Santa Maria/RS: Universidade Federal de Santa Maria, 2019.

BECK, H. E. *et al.* Global-scale regionalization of hydrologic model parameters. **Water Resources Research**, v. 52, n. 5, p. 3599-3622, maio 2016.

BECK, H. E. *et al.* Global evaluation of runoff from 10 state-of-the-art hydrological models. **Hydrology and Earth System Sciences**, v. 21, n. 6, p. 2881-2903, 12 jun. 2017.

BENKACI, T.; DECHEMI, N. Comparative performances analysis of four daily rainfall-runoff models, applied on semi-arid watersheds (Algeria). **Larhyss Journal**, p. 21-39, 2018.

BLÖSCHL, G. *et al.* **Runoff prediction in ungauged basins: synthesis across processes, places and scales.** 2013. ed. [s.l.: s.n.].

BODIAN, A. *et al.* Future Climate Change Impacts on Streamflows of Two Main West Africa River Basins: Senegal and Gambia. **Hydrology**, v. 5, n. 1, p. 21, 16 mar. 2018.

BORGES JÚNIOR, J. C. F. *et al.* Métodos de estimativa da evapotranspiração de referência diária para a microrregião de Garanhuns, PE. **Revista Brasileira de Engenharia Agrícola e Ambiental**, 2012.

BOUMENNI, H.; BACHNOU, A.; ALAA, N. E. The rainfall-runoff model GR4J optimization of parameter by genetic algorithms and Gauss-Newton method: application for the watershed Ourika (High Atlas, Morocco). **Arabian Journal of Geosciences**, v. 10, n. 15, p. 343, 8 ago. 2017.

BRASIL. **Portaria n°102, de 28 de novembro de 2017.** , 2017.

- BRUSA, L. C. **Aprimoramento estatístico da regionalização de vazões máximas e médias. Aplicação a bacias hidrográficas do Rio Grande do Sul e Santa Catarina.** Porto Alegre - RS: Universidade Federal do Rio Grande do Sul, dez. 2004.
- BURN, D. H.; GOEL, N. K. The formation of groups for regional flood frequency analysis. **Hydrological Sciences Journal**, v. 45, n. 1, p. 97-112, fev. 2000.
- CANNAROZZO, M. *et al.* Annual runoff regional frequency analysis in Sicily. **Physics and Chemistry of the Earth, Parts A/B/C**, v. 34, n. 10-12, p. 679-687, 2009.
- CANTONI, E. *et al.* Hydrological performance of the ERA5 reanalysis for flood modeling in Tunisia with the LISFLOOD and GR4J models. **Journal of Hydrology: Regional Studies**, v. 42, p. 101169, ago. 2022.
- CARNEIRO, T.; FARIAS, C. Otimização estocástica implícita e redes neurais artificiais para auxílio na operação mensal dos reservatórios Coremas - Mãe d'Água. **Revista Brasileira de Recursos Hídricos**, v. 18, n. 4, p. 115-124, 2013.
- CARVALHO, G. A. DE. **Algoritmos de clusterização e o modelo RFV (Recência, Frequência e Valor) aplicados em dados de e-commerce.** Curitiba: Universidade Federal do Paraná, 2020.
- CARVALHO, S. R. DE. **Regionalização de vazões para a parte baiana da bacia hidrográfica do Rio Pardo.** Itapetinga: Universidade Estadual do Sudoeste da Bahia, 2017.
- CARVALHO, T. L. L. DE. **Análise regional de frequências aplicadas à precipitação pluvial.** . Dissertação de Mestrado— Porto Alegre - RS: Universidade Federal do Rio Grande do Sul, 2007.
- CASSALHO, F. *et al.* Artificial intelligence for identifying hydrologically homogeneous regions: A state-of-the-art regional flood frequency analysis. **Hydrological Processes**, v. 33, n. 7, p. 1101-1116, 30 mar. 2019.
- CHAGAS, V. B. P. *et al.* CAMELS-BR: hydrometeorological time series and landscape attributes for 897 catchments in Brazil. **Earth System Science Data**, v. 12, n. 3, p. 2075-2096, 8 set. 2020.
- CHAGAS, V. B. P.; CHAFFE, P. L. B.; BLÖSCHL, G. Climate and land management accelerate the Brazilian water cycle. **Nature Communications**, v. 13, n. 1, p. 5136, 1 set. 2022.
- CHAMBERS, J. M.; FREENY, A.; HEIBERGER, R. M. **Analysis of variance; designed experiments.** 1. ed. [s.l: s.n.].
- CHAVES, H. M. L. *et al.* Regionalização de Vazões Mínimas em Bacias Através de Interpolação em Sistemas de Informação Geográfica . **Revista Brasileira de Recursos Hídricos - RBRH** , v. 7, p. 43-51, set. 2002.
- CHEIN, F. Introdução aos modelos de regressão linear: um passo inicial para compreensão da econometria como uma ferramenta de avaliação de políticas públicas. p. 77, 2019.

CHENG, X. *et al.* Application of HEC-HMS Parameter Regionalization in Small Watershed of Hilly Area. **Water Resources Management**, v. 35, n. 6, p. 1961-1976, 10 abr. 2021.

CIRILO, J. A. *et al.* Development and application of a rainfall-runoff model for semi-arid regions. **RBRH**, v. 25, 2020.

CORON, L. *et al.* The suite of lumped GR hydrological models in an R package. **Environmental Modelling & Software**, v. 94, p. 166-171, ago. 2017.

CORON, L. *et al.* **airGR: Suite of GR Hydrological Models for Precipitation-Runoff Modelling (v. 1.7.0)**. , 2022.

COSTA, A. C. *et al.* Regionalização de modelagem hidrológica semidistribuída para todo o estado Ceará. **XX Simpósio Brasileiro de Recursos Hídricos**, 2013.

CUNHA, V. C. V. DA. **Avaliação da interação entre as águas subterrâneas e superficiais na bacia do Rio das Fêmas, sistema aquífero Urucuia, Bahia**. Belo Horizonte: Comissão Nacional de Energia Nuclear, jun. 2017.

CUPAK, A.; WAŁĘGA, A.; MICHAŁEC, B. Cluster analysis in determination of hydrologically homogeneous regions with low flow. **Acta Scientiarum Polonorum Formatio Circumiectus**, v. 1, p. 53-63, 2017.

DAVID, P. C. *et al.* Correspondence Between Model Structures and Hydrological Signatures: A Large-Sample Case Study Using 508 Brazilian Catchments. **Water Resources Research**, v. 58, n. 3, mar. 2022.

DIKBAS, F. *et al.* Defining Homogeneous Regions for Streamflow Processes in Turkey Using a K-Means Clustering Method. **Arabian Journal for Science and Engineering**, v. 38, n. 6, p. 1313-1319, 15 jun. 2013.

DINIZ, L. DA S. **Regionalização de parâmetros de modelo chuva-vazão usando redes neurais**. Porto Alegre: Universidade Federal do Rio Grande do Sul, 2008.

EBERHART, R.; KENNEDY, J. **A new optimizer using particle swarm theory**. MHS'95. Proceedings of the Sixth International Symposium on Micro Machine and Human Science. **Anais...IEEE**, 1995.

EDIJATNO *et al.* GR3J: a daily watershed model with three free parameters. **Hydrological Sciences Journal**, v. 44, n. 2, p. 263-277, abr. 1999.

EDIJATNO; MICHEL, C. Un modèle pluie-débit journalier à trois paramètres. **La Houille Blanche**, v. 75, n. 2, p. 113-122, 1 abr. 1989.

EMBRAPA. Potencialidades da Zona da Mata e do Agreste Nordestinos para a implantação de sistemas de Integração lavoura-pecuária-floresta - Capítulo 4. Em: [s.l: s.n.].

EMBRAPA, E. B. DE P. A. **Região Nordeste**. Disponível em: <<https://www.embrapa.br/contando-ciencia/regiao->

GOLIAN, S.; MURPHY, C.; MERESA, H. Regionalization of hydrological models for flow estimation in ungauged catchments in Ireland. **Journal of Hydrology: Regional Studies**, v. 36, p. 100859, ago. 2021.

GONÇALVES, E. D. *et al.* Identificação de regiões homogêneas e análise de regressão múltipla para regionalização de vazão na bacia hidrográfica do Rio Tapajós. **Revista Brasileira de Cartografia**, 2017.

GONÇALVES, R. D.; ENGELBRECHT, B. Z.; CHANG, H. K. Evolução da contribuição do Sistema Aquífero Urucua para o Rio São Francisco, Brasil. **Águas Subterrâneas**, v. 32, n. 1, p. 1-10, 22 dez. 2017.

GROUILLET, B. *et al.* Sensitivity analysis of runoff modeling to statistical downscaling models in the western Mediterranean. **Hydrology and Earth System Sciences**, v. 20, n. 3, p. 1031-1047, 8 mar. 2016.

GUAN, X. *et al.* The Capacity of the Hydrological Modeling for Water Resource Assessment under the Changing Environment in Semi-Arid River Basins in China. **Water**, v. 11, n. 7, p. 1328, 27 jun. 2019.

GUPTA, H. V. *et al.* Large-sample hydrology: a need to balance depth with breadth. **Hydrology and Earth System Sciences**, v. 18, n. 2, p. 463-477, 6 fev. 2014.

HAJHOUI, Y. *et al.* Modélisation pluie-débit et analyse du régime d'un bassin versant semi-aride sous influence nivale. Cas du bassin versant du Rheraya (Haut Atlas, Maroc). **La Houille Blanche**, v. 104, n. 3, p. 49-62, 1 jun. 2018.

HALL, M. J.; MINNS, A. W. The classification of hydrologically homogeneous regions. **Hydrological Sciences Journal**, v. 44, n. 5, p. 693-704, out. 1999.

HAN, J.; KAMBER, M.; PEI, J. **Data Mining: Concepts and Techniques**. 3^a ed. [s.l.: s.n.].

HANNAH, D. M. *et al.* Large-scale river flow archives: importance, current status and future needs. **Hydrological Processes**, v. 25, n. 7, p. 1191-1200, 30 mar. 2011.

HE, Y.; BÁRDOSSY, A.; ZEHE, E. A review of regionalisation for continuous streamflow simulation. **Hydrology and Earth System Sciences**, v. 15, n. 11, p. 3539-3553, 23 nov. 2011.

HERSBACH, H. *et al.* ERA5 hourly data on single levels from 1940 to present. Copernicus Climate Change Service (C3S) Climate Data Store (CDS). 2023.

HRACHOWITZ, M. *et al.* A decade of Predictions in Ungauged Basins (PUB)—a review. **Hydrological Sciences Journal**, v. 58, n. 6, p. 1198-1255, 14 ago. 2013.

HUANG, P. *et al.* Event-based hydrological modeling for detecting dominant hydrological process and suitable model strategy for semi-arid catchments. **Journal of Hydrology**, v. 542, p. 292-303, nov. 2016.

HUBLART, P. *et al.* Reliability of a conceptual hydrological model in a semi-arid Andean catchment facing water-use changes. **Proceedings of the International Association of Hydrological Sciences**, v. 371, p. 203-209, 12 jun. 2015.

IBGE. **Vegetação Brasileira 1:5.000.000 | 2004.**

INRAE. INSTITUT NATIONAL DE RECHERCHE POUR L'AGRICULTURE, L'ALIMENTATION ET L'ENVIRONNEMENT. **Bref historique des travaux ayant conduit au développement des modèles GR.**

JOHNSTON, L. H. *et al.* Identifying hydrologic regimes and drivers in Nova Scotia, Canada: catchment classification efforts for a data-limited region. **Journal of Hydrologic Engineering**, v. 27, n. 11, 2022.

JULIANA MENEZES GARRIDO. **Aplicação de modelo matemático de simulação com utilização de SIG à bacia do rio Jiquiriçá - Bahia.** Brasília - DF: Universidade de Brasília, 30 set. 2003.

KAHYA, E.; DEMIREL, M. C.; BEG, O. A. Hydrologic homogeneous regions using monthly streamflow in Turkey. **Earth Sciences Research Journal**, v. 12, p. 181-193, 2008.

KASSAMBARA, A.; MUNDT, F. Factoextra: Extraia e visualize os resultados de análises multivariadas de dados. 2020.

KAUFMAN, L.; ROUSSEEUW, P. J. **Finding Groups in Data: An Introduction to Cluster Analysis.** Hoboken, NJ, USA: John Wiley & Sons, Inc., 1990.

KOEFENDER, A. **REGIONALIZAÇÃO DE VAZÕES Q7,10, Q90% E Q50% NA BACIA HIDROGRÁFICA DO RIO IBICUI (SUB-BACIA 76).** Porto Alegre: Universidade Federal do Rio Grande do Sul, 2015.

KOKKONEN, T. S. *et al.* Predicting daily flows in ungauged catchments: model regionalization from catchment descriptors at the Coweeta Hydrologic Laboratory, North Carolina. **Hydrological Processes**, v. 17, n. 11, p. 2219-2238, 15 ago. 2003.

KOLLING NETO, A. **Aprimoramento da estimativa da disponibilidade hídrica por meio de dados espaciais. Estudo de caso para a bacia do Paraguaçu.** Viçosa - MG: Universidade Federal de Viçosa, 2021.

KRAUSE, P.; BOYLE, D. P.; BÄSE, F. Comparison of different efficiency criteria for hydrological model assessment. **Advances in Geosciences**, v. 5, p. 89-97, 16 dez. 2005.

KUANA, L. A. *et al.* Estudo de caso: Regionalização de vazão utilizando a base de dados CAMELS. **Associação Brasileira de Recursos Hídricos**, 2020.

KUANA, L. A. *et al.* Aplicação dos modelos GR4J e GR6J utilizando o conjunto de dados desenvolvido para o Estado do Paraná. **ABRhidro**, 2021.

KUANA, L. A. **Regionalização de parâmetros do modelo GR4J para a predição de vazão diária em bacias não monitoradas.** Curitiba - PR: Universidade Federal do Paraná, 2022.

- KUNNATH-POOVAKKA, A.; ELDHO, T. I. A comparative study of conceptual rainfall-runoff models GR4J, AWBM and Sacramento at catchments in the upper Godavari river basin, India. **Journal of Earth System Science**, v. 128, n. 2, p. 33, 25 mar. 2019.
- LANE, R. A. *et al.* Incorporating Uncertainty Into Multiscale Parameter Regionalization to Evaluate the Performance of Nationally Consistent Parameter Fields for a Hydrological Model. **Water Resources Research**, v. 57, n. 10, 28 out. 2021.
- LEBECHEREL, L. *et al.* Analyse de la sensibilité des calculs hydrologiques à la densité spatiale des réseaux hydrométriques. **La Houille Blanche**, v. 100, n. 1, p. 39-44, 1 fev. 2014.
- LEMOS, A. F. **Avaliação de metodologias de regionalização de vazões mínimas de referência para bacia hidrográfica do Rio São Francisco, a montante do Reservatório de Três Marias**. Viçosa: Universidade Federal de Viçosa, 2006.
- LI, F. *et al.* The impact of climate change on runoff in the southeastern Tibetan Plateau. **Journal of Hydrology**, v. 505, p. 188-201, nov. 2013.
- LIN, G.-F.; CHEN, L.-H. Identification of homogeneous regions for regional frequency analysis using the self-organizing map. **Journal of Hydrology**, v. 324, n. 1-4, p. 1-9, jun. 2006.
- LIU, Y. *et al.* Changes in and Modelling of Hydrological Process for a Semi-Arid Catchment in the Context of Human Disturbance. **Frontiers in Earth Science**, v. 9, 29 set. 2021.
- LOU, R. F. **Modelagem hidrológica chuva-vazão e hidrodinâmica aplicada na bacia experimental do Rio Piabanha/RJ**. Rio de Janeiro: Universidade Federal do Rio de Janeiro, 2010.
- MAECHLER, M. *et al.* cluster: Cluster Analysis Basics and Extensions. R package version 2.1.5. 2023.
- MAKHLOUF, Z. **Compléments sur le modèle pluie-débit GR4J et essai d'estimation de ses paramètres**. Paris: Université Paris XI Orsay, 1994.
- MARTINS, C. M. R. **Evapotranspiração de canais em diferentes fases fenológicas por sensoriamento remoto no vale do submédio do São Francisco**. Recife - PE: Universidade Federal Rural de Pernambuco, 2019.
- MATHEVET, T. **Quels modèles pluie-débit globaux au pas de temps horaire ? Développement empiriques et comparaison de modèles sur un large échantillon de bassins versants**. Paris: Ecole Nationale du Génie Rural, des Eaux et Forêts, 4 nov. 2005.
- MCINTYRE, N. *et al.* Ensemble predictions of runoff in ungauged catchments. **Water Resources Research**, v. 41, n. 12, dez. 2005.
- MELLER, A. **Previsão de cheias por conjunto em curto prazo**. Porto Alegre - RS: Universidade Federal do Rio Grande do Sul, 18 dez. 2012.

- MELO JÚNIOR, J. C. F. DE *et al.* Determinação de regiões homogêneas quanto à distribuição de frequência de chuvas no leste do Estado de Minas Gerais. **Revista Brasileira de Engenharia Agrícola e Ambiental**, v. 10, n. 2, p. 408-416, jun. 2006.
- MERESA, H. *et al.* Simulated Changes in Seasonal and Low Flows with Climate Change for Irish Catchments. **Water**, v. 14, n. 10, p. 1556, 12 maio 2022.
- MERZ, R.; BLÖSCHL, G.; PARAJKA, J. Regionalization methods in rainfall-runoff modelling using large catchment samples. **IAHS-AISH**, p. 117-125, 2006.
- MICHEL, C. Que peut-on faire en hydrologie avec modèle conceptuel à un seul paramètre ? **La Houille Blanche**, v. 69, n. 1, p. 39-44, 1 jan. 1983.
- MICROSOFT. **Atualização do Microsoft Excel 2016 (KB3178719) Edição de 64 Bits**, 2016.
- MONTANARI, A. *et al.* “Panta Rhei—Everything Flows”: Change in hydrology and society—The IAHS Scientific Decade 2013-2022. **Hydrological Sciences Journal**, v. 58, n. 6, p. 1256-1275, 5 ago. 2013.
- MORAIS, M. A. V. *et al.* Regionalization of reference streamflows for the Araguaia River basin in Brazil. **Semina: Ciências Agrárias**, v. 41, n. 3, p. 829, 7 abr. 2020.
- MURPHY, A. H. Skill Scores based on the Mean Square Error and their Relationships to the Correlation Coefficient. **Monthly Weather Review**, p. 2417-2424, 1988.
- MURTAGH, F.; LEGENDRE, P. Ward’s Hierarchical Agglomerative Clustering Method: Which Algorithms Implement Ward’s Criterion? **Journal of Classification**, v. 31, n. 3, p. 274-295, 18 out. 2014.
- NASCIMENTO, N. DE O. **Appréciation l’aide d’un modèle empirique des effets d’actions anthropiques sur la relation pluie-débit l’échelle d’un bassin versant**. [s.l.] Ecole nationale des Ponts et Chaussées, 3 jul. 1995.
- NATHAN, R. J.; MCMAHON, T. A. Identification of homogeneous regions for the purposes of regionalisation. **Journal of Hydrology**, v. 121, n. 1-4, p. 217-238, dez. 1990.
- NETO, W. M. P.; VIEIRA, F. R.; MATOSINHOS, C. C. 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**. [s.l.] Atena Editora, 2021. p. 274-288.
- NUNES, F. M. S. *et al.* Modelagem hidrológica via SMAP para estimativa de vazões mensais na bacia do rio Piancó. **Revista Verde de Agroecologia e Desenvolvimento Sustentável**, v. 9, p. 289-295, 2014.
- OLIVEIRA, L. T. *et al.* Comportamento hidrogeológico do Grupo Urucua: Subbacias dos rios Formoso e Arrojado. **Águas Subterrâneas**, 2008.

- LOUDIN, L. *et al.* Spatial proximity, physical similarity, regression and ungauged catchments: A comparison of regionalization approaches based on 913 French catchments. **Water Resources Research**, v. 44, n. 3, mar. 2008.
- LOUDIN, L. *et al.* Are seemingly physically similar catchments truly hydrologically similar? **Water Resources Research**, v. 46, n. 11, nov. 2010.
- OUELDKADDOUR, F.-Z. E. *et al.* Hydrological modeling of rainfall-runoff of the semi-arid Aguiabat Ezziar watershed through the GR4J model. **Limnological Review**, v. 21, n. 3, p. 119-126, 1 set. 2021.
- PARAJKA, J. *et al.* Comparative assessment of predictions in ungauged basins - Part 1: Runoff-hydrograph studies. **Hydrology and Earth System Sciences**, v. 17, n. 5, p. 1783-1795, 7 maio 2013.
- PARAJKA, J.; BLÖSCHL, G.; MERZ, R. Regional calibration of catchment models: Potential for ungauged catchments. **Water Resources Research**, v. 43, n. 6, jun. 2007.
- PEREIRA, M. A. F.; BARBIEIRO, B. L.; QUEVEDO, D. M. DE. Importância do monitoramento e disponibilização de dados hidrológicos para a gestão integrada dos recursos hídricos. **Sociedade & Natureza**, v. 32, p. 308-320, 6 maio 2020.
- PERRIN, C. **Vers une amélioration d'un modèle global pluie-débit au travers d'une approche comparative**. Grenoble: Institut National Polytechnique de Grenoble, 20 out. 2000.
- PERRIN, C.; MICHEL, C.; ANDRÉASSIAN, V. Improvement of a parsimonious model for streamflow simulation. **Journal of Hydrology**, v. 279, n. 1-4, p. 275-289, ago. 2003a.
- PERRIN, C.; MICHEL, C.; ANDRÉASSIAN, V. Improvement of a parsimonious model for streamflow simulation. **Journal of Hydrology**, v. 279, n. 1-4, p. 275-289, 25 ago. 2003b.
- PESSOA, F. C. L. **Regionalização de curvas de permanência de vazões na Amazônia legal**. Pará: Universidade Federal do Pará, 2015.
- PESSOA, M. M. E. P. **Obtenção de vazões em locais sem monitoramento através de um modelo hidrológico com assimilação de dados**. Porto Alegre - RS : Universidade Federal do Rio Grande do Sul, out. 2017.
- PIMENTA, F. DE S.; RIBEIRO, F. V.; CRUZ JÚNIOR, D. C. Modelagem de Regressão Espacial para Estimativa de Valores em Massa a partir de Cartografia Cadastral. **Revista Brasileira de Cartografia**, v. 73, n. 1, p. 36-52, 19 fev. 2021.
- POISSANT, D.; ARSENAULT, R.; BRISSETTE, F. Impact of parameter set dimensionality and calibration procedures on streamflow prediction at ungauged catchments. **Journal of Hydrology: Regional Studies**, v. 12, p. 220-237, ago. 2017.
- PYTHON SOFTWARE FOUNDATION. **Python 3.9.8**. , 2021.
- R DEVELOPMENT CORE TEAM. **The R Project for Statistical Computing**. , 2019.

- RAKEM, Y. **Analyse critique et reformulation mathématique d'un modèle pluie-débit (GR4)**. Marne-la-vallée: Ecole Nationale des Ponts et Chaussées , 1999.
- RAMPINELLI, C. **Modelo Hidrológico GR4J. Um modelo francês com uma pitada de DNA brasileiro**. Disponível em: <<https://rpubs.com/cassiorampinelli/600688>>. Acesso em: 24 abr. 2022.
- RAMPINELLI, C.; COSTA, M. G. DA. Calibração do modelo chuva-vazão SMAP em uma estação do semiárido nordestino com os algoritmos PSO, AG e Nelder Mead e seus efeitos na definição das séries de vazões médias mensais . **XXI Simpósio Brasileiro de Recursos Hídricos**, 2015.
- RAZAVI, T.; COULIBALY, P. Streamflow prediction in ungauged basins: Review of regionalization methods. **Journal of Hydrologic Engineering** , v. 18, n. 8, 2013.
- REICHL, J. P. C. *et al.* Optimization of a similarity measure for estimating ungauged streamflow. **Water Resources Research**, v. 45, n. 10, out. 2009.
- RIANNA, M.; RUSSO, F.; NAPOLITANO, F. Stochastic index model for intermittent regimes: from preliminary analysis to regionalisation. **Natural Hazards and Earth System Sciences**, v. 11, n. 4, p. 1189-1203, 27 abr. 2011.
- RIBEIRO, C. S.; OLIVEIRA, G. G. DE. A questão hídrica no semiárido baiano: conflitos pelo uso da água e as tecnologias sociais de aproveitamento de água de chuva. **Revista del CESLA**, p. 355-382, 2019.
- ROCHA, J. S. **Regionalização hidrológica para o estado do Ceará: uma abordagem estatística das cheias**. Fortaleza - CE: Universidade Federal do Ceará, 2020.
- ROMANOWICZ, R. J.; OSUCH, M.; GRABOWIECKA, M. On the choice of calibration periods and objective functions: A practical guide to model parameter identification. **Acta Geophysica**, 2013.
- RUELLAND, D.; DEZETTER, A.; HUBLART, & P. **Sensitivity analysis of hydrological modelling to climate forcing in a semi-arid mountainous catchment**. [s.l.] IAHS Publ, 2014.
- SALES, J. M. DE J. **Monitoramento e modelagem hidrológica em uma bacia hidrográfica no semiárido do Brasil**. São Cristóvão/SE: Universidade Federal de Sergipe, 2019.
- SAMUEL, J.; COULIBALY, P.; METCALFE, R. A. Estimation of Continuous Streamflow in Ontario Ungauged Basins: Comparison of Regionalization Methods. **Journal of Hydrologic Engineering**, v. 16, n. 5, p. 447-459, maio 2011.
- SANTOS, B. V. C. D. DOS; FONTES, A. S. Modelagem hidrológica de sistema hídrico complexo do semiárido baiano. **Revista Eletrônica de Gestão e Tecnologias Ambientais**, v. 2, n. 1, p. 152, 4 set. 2014.
- SANTOS, J. Y. G. DOS *et al.* Aplicação do modelo SWAT para a simulação chuva-vazão na bacia do rio Tapacurá, estado de Pernambuco. **XX Simpósio Brasileiro de Recursos Hídricos**, p. 8, 2013a.

- SANTOS, J. Y. G. DOS *et al.* Aplicação do modelo SWAT para a simulação chuva-vazão na bacia do Rio Tapacurá, estado de Pernambuco. **XX Simpósio Brasileiro de Recursos Hídricos**, 2013b.
- SANTOS, W. A.; SOUZA, F. D. A. S. DE. Identificação de regiões pluviometricamente homogêneas no nordeste do Brasil usando análise multivariada. **Revista Brasileira de Climatologia**, v. 10, n. 1, 30 jun. 2012.
- SARAIVA, I. S. **Regionalização de parâmetros de modelos de simulação hidrológica para emprego em bacias hidrográficas sem monitoramento fluviométrico**. Belo Horizonte: Universidade Federal de Minas Gerais, 2010.
- SCHWARZBACH, M. S. R. **Estudo comparativo do comportamento hidrológico de bacias situadas na região sul e no cristalino semi-árido do Brasil, através do desenvolvimento do modelo MOHDAC: modelo hidrológico auto-calibrável**. Porto Alegre: Universidade Federal do Rio Grande do Sul, 1988.
- SEI. **Tipologia climática segundo Thornthwaite. Pluviometria 1943 - 1983/ Temperatura 1961 - 1990**. , 1998.
- SEIBERT, J.; BEVEN, K. J. Gauging the ungauged basin: how many discharge measurements are needed? **Hydrology and Earth System Sciences**, v. 13, n. 6, p. 883-892, 22 jun. 2009.
- SEZEN, C.; PARTAL, T. The utilisation of conceptual and data-driven models for hydrological modelling in semi-arid and humid areas of the Antalya basin in Turkey. **Acta Geophysica**, v. 70, n. 2, p. 897-915, 7 abr. 2022.
- SGB - CPRM. **Mapa Hidrogeológico do Brasil ao Milionésimo**.
- SGB-CPRM. **Estudos Hidrológicos e Hidrogeológicos Integrados na Região do Aquífero Urucuia - SAU**. Disponível em: <<http://www.cprm.gov.br/publique/Hidrologia/Pesquisa-e-Inovacao/Estudos-Hidrologicos-e-Hidrogeologicos-Integrados-na-Regiao-do-Aquifero-Urucuia---SAU-5208.html>>. Acesso em: 14 maio. 2022.
- SILVA, A. M. DA *et al.* Vazões mínimas e de referência para outorga na região do Alto Rio Grande, Minas Gerais. **Revista Brasileira de Engenharia Agrícola e Ambiental**, v. 10, p. 374-380, 2006a.
- SILVA, A. M. DA *et al.* Vazões mínimas e de referência para outorga na região do Alto Rio Grande, Minas Gerais. **Revista Brasileira de Engenharia Agrícola e Ambiental**, v. 10, n. 2, p. 374-380, jun. 2006b.
- SILVA, D. S. DA; MEDEIROS, J. D. F. DE. Aplicação do modelo hidrológico MGB-IPH na bacia hidrográfica do Rio Piranhas-Açu. **XIV Simpósio de Recursos Hídricos do Nordeste**, 2018.
- SILVA FILHO, J. A. DA; RODRIGUES JUNIOR, J. C.; ARAÚJO, S. C. DE. Aplicação de um modelo chuva-vazão na bacia hidrográfica do rio Brígida. **XV SRHNE - Simpósio de Recursos Hídricos do Nordeste**, 2020.

- SILVA JÚNIOR, O. B. DA *et al.* Extrapolação Espacial na Regionalização da Vazão. **Revista Brasileira de Recursos Hídricos - RBRH**, v. 8, p. 21-37, 2003.
- SILVA, M. T. *et al.* SWAT model application to estimate the flow in the basin of lower-middle São Francisco River. **Revista Brasileira de Geografia Física**, v. 8, n. 6, p. 1615-1627, 2015.
- SIVAPALAN, M. *et al.* IAHS Decade on Predictions in Ungauged Basins (PUB), 2003-2012: Shaping an exciting future for the hydrological sciences. **Hydrological Sciences Journal**, v. 48, n. 6, p. 857-880, dez. 2003.
- SOUSA, H. T. DE. **Sistema computacional para Regionalização de Vazões**. Viçosa/MG: Universidade Federal de Viçosa, 2009.
- SOUZA, L. R. DE. **Simulação hidrológica e aplicação de uma análise multivariada no estudo de chuva-vazão na bacia hidrográfica do rio São Francisco**. Paraíba: Universidade Federal de Campina Grande, 2012.
- SUDENE. **Delimitação do Semiárido**. Disponível em: <<http://antigo.sudene.gov.br/delimitacao-do-semiarido>>. Acesso em: 12 mar. 2023.
- TAN, P.-N.; STEINBACH, M.; KUMAR, V. **Data mining cluster analysis: basic concepts and algorithms**. [s.l.: s.n.].
- TASKER, G. D. Comparing methods of hydrologic regionalization. **Journal of the American Water Resources Association**, v. 18, n. 6, p. 965-970, dez. 1982.
- THE PANDAS DEVELOPMENT TEAM. **pandas-dev/pandas: Pandas**. , 2020.
- TIAN, Y.; XU, Y.-P.; ZHANG, X.-J. Assessment of Climate Change Impacts on River High Flows through Comparative Use of GR4J, HBV and Xinanjiang Models. **Water Resources Management**, v. 27, n. 8, p. 2871-2888, 15 jun. 2013.
- TRAORE, V. B. Calibrating the Rainfall-Runoff Model GR4J and GR2M on the Koulountou River Basin, a Tributary of the Gambia River. **American Journal of Environmental Protection**, v. 3, n. 1, p. 36, 2014.
- TUCCI, C. E. M. **Regionalização de Vazões**. 1º reimpressão da 1ª ed. Porto Alegre / RS: [s.n.].
- TUSZYNSKI, J. **caTools: Tools: Moving Window Statistics, GIF, Base64, ROC AUC, etc.** , 2021.
- VANDEWIELE, G. L.; ELIAS, A. Monthly water balance of ungauged catchments obtained by geographical regionalization. **Journal of Hydrology**, v. 170, n. 1-4, p. 277-291, ago. 1995.
- VIRÃES, M. V. **Regionalização de parâmetros de modelo hidrológico para a região semiárida do Nordeste brasileiro**. Recife - PE: Universidade Federal de Pernambuco, 2018.

- VIRÃES, M. V.; CIRILO, J. A. Regionalization of hydrological model parameters for the semi-arid region of the northeast Brazil. **RBRH**, v. 24, 2019.
- VIVIROLI, D. *et al.* Continuous simulation for flood estimation in ungauged mesoscale catchments of Switzerland - Part II: Parameter regionalisation and flood estimation results. **Journal of Hydrology**, v. 377, n. 1-2, p. 208-225, out. 2009.
- WAGENER, T.; WHEATER, H.; GUPTA, H. V. **Rainfall-runoff Modelling in Gauged and Ungauged Catchments**. [s.l: s.n.].
- WAGENER, T.; WHEATER, H. S. Parameter estimation and regionalization for continuous rainfall-runoff models including uncertainty. **Journal of Hydrology**, v. 320, n. 1-2, p. 132-154, mar. 2006.
- WANG, B. *et al.* **Catchment grouping and regional calibration for predictions in ungauged basins**. Modelling and Simulation Society of Australia and New Zealand (MSSANZ), Inc., 1 dez. 2013.
- WANG, H.; CAO, L.; FENG, R. Hydrological Similarity-Based Parameter Regionalization under Different Climate and Underlying Surfaces in Ungauged Basins. **Water**, v. 13, n. 18, p. 2508, 13 set. 2021a.
- WANG, H.; CAO, L.; FENG, R. Hydrological Similarity-Based Parameter Regionalization under Different Climate and Underlying Surfaces in Ungauged Basins. **Water**, v. 13, n. 18, p. 2508, 13 set. 2021b.
- WANZALA, M. A. *et al.* Assessment of global reanalysis precipitation for hydrological modelling in data-scarce regions: A case study of Kenya. **Journal of Hydrology: Regional Studies**, v. 41, p. 101105, jun. 2022.
- WARD, J. J. H. Hierarchical grouping to optimize an objective function. **Journal of the American statistical association**, p. 236-244, 1963.
- WASKOM, M. seaborn: statistical data visualization. **Journal of Open Source Software**, v. 6, n. 60, p. 3021, 6 abr. 2021.
- WILTSHIRE, S. E. Identification of homogeneous regions for flood frequency analysis. **Journal of Hydrology**, v. 84, n. 3-4, p. 287-302, maio 1986.
- XING, K. *et al.* Mutual privacy preserving k -means clustering in social participatory sensing. **IEEE Transactions on Industrial Informatics**, v. 13, n. 4, p. 2066-2076, ago. 2017.
- YADAV, M.; WAGENER, T.; GUPTA, H. Regionalization of constraints on expected watershed response behavior for improved predictions in ungauged basins. **Advances in Water Resources**, v. 30, n. 8, p. 1756-1774, ago. 2007.
- YANG, T. *et al.* Regional frequency analysis and spatio-temporal pattern characterization of rainfall extremes in the Pearl River Basin, China. **Journal of Hydrology**, v. 380, n. 3-4, p. 386-405, jan. 2010.

YANG, X. *et al.* Gestion d'un réservoir pour la régularisation des débits. **La Houille Blanche**, v. 77, n. 6, p. 433-440, 1 out. 1991.

YANG, X. *et al.* Dependence of regionalization methods on the complexity of hydrological models in multiple climatic regions. **Journal of Hydrology**, v. 582, p. 124357, mar. 2020.

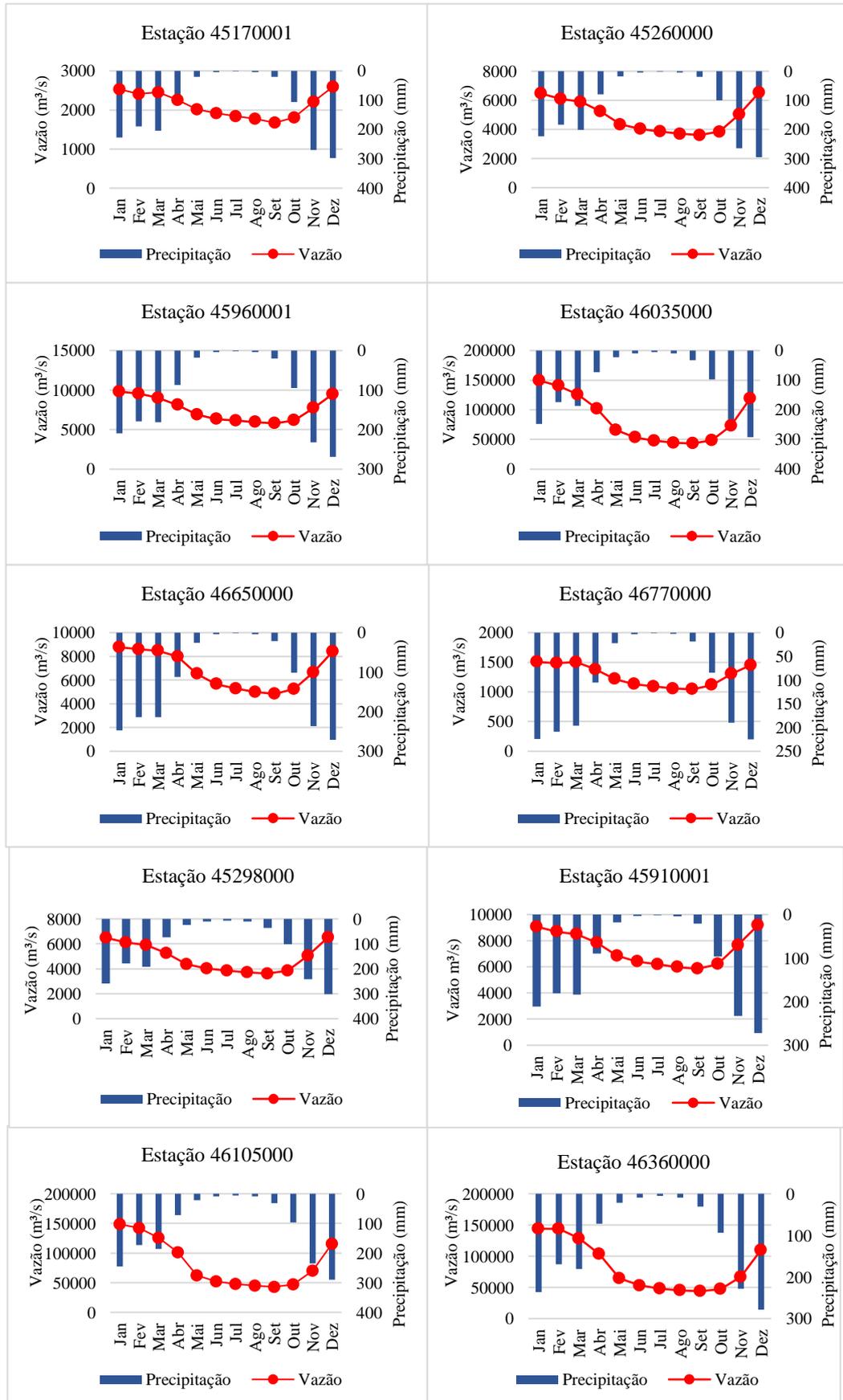
YANG, X.; MICHEL, C. Flood forecasting with a watershed model: a new method of parameter updating. **Hydrological Sciences Journal**, v. 45, n. 4, p. 537-546, ago. 2000.

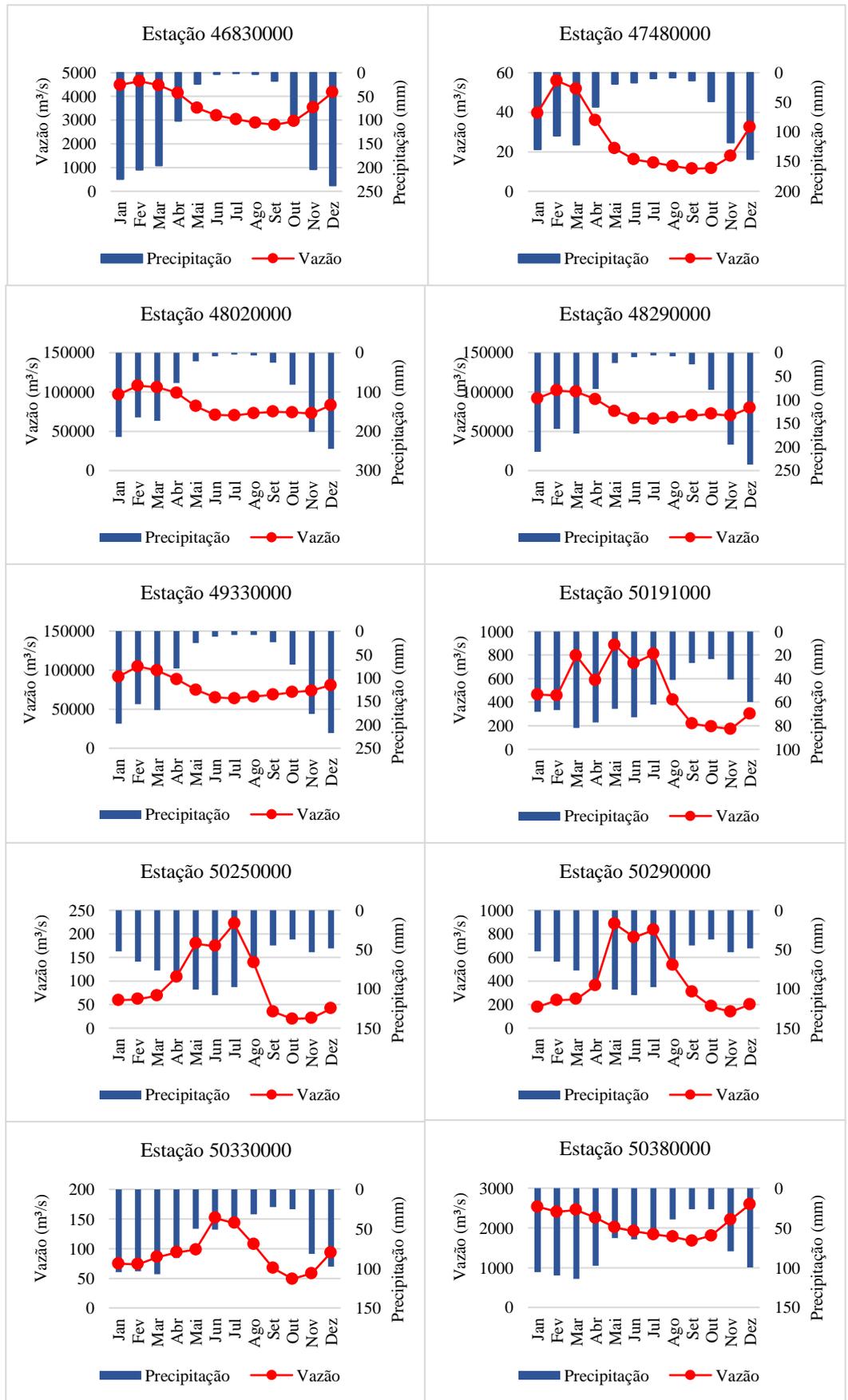
ZAMBRANO-BIGIARINI, M. Tutorial for using hydroPSO to calibrate the GR4J hydrological model (version 0.3). 2020.

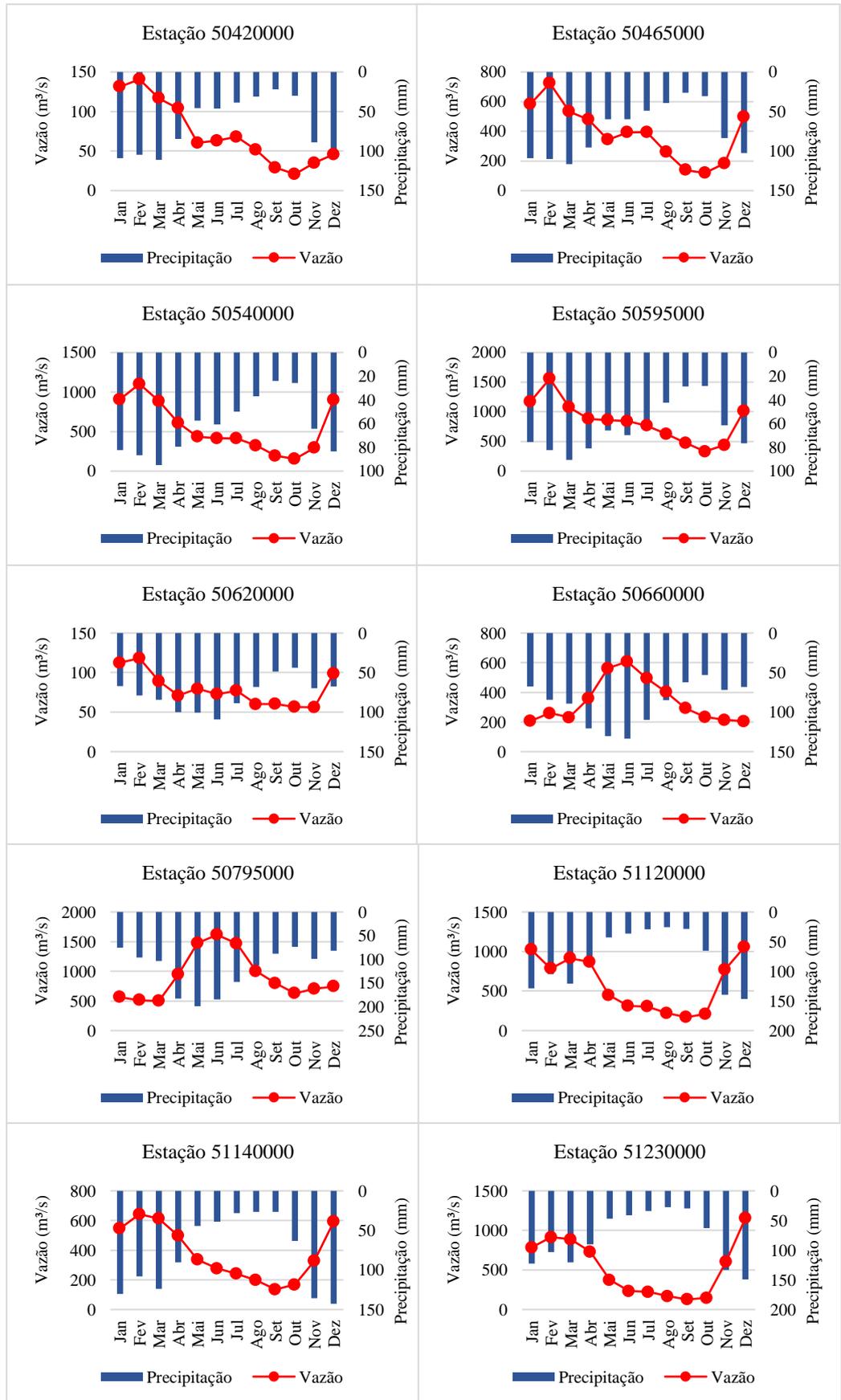
ZAMBRANO-BIGIARINI, M.; ROJAS, R. **hydroPSO: Particle Swarm Optimisation, with Focus on Environmental Models.** , 2018.

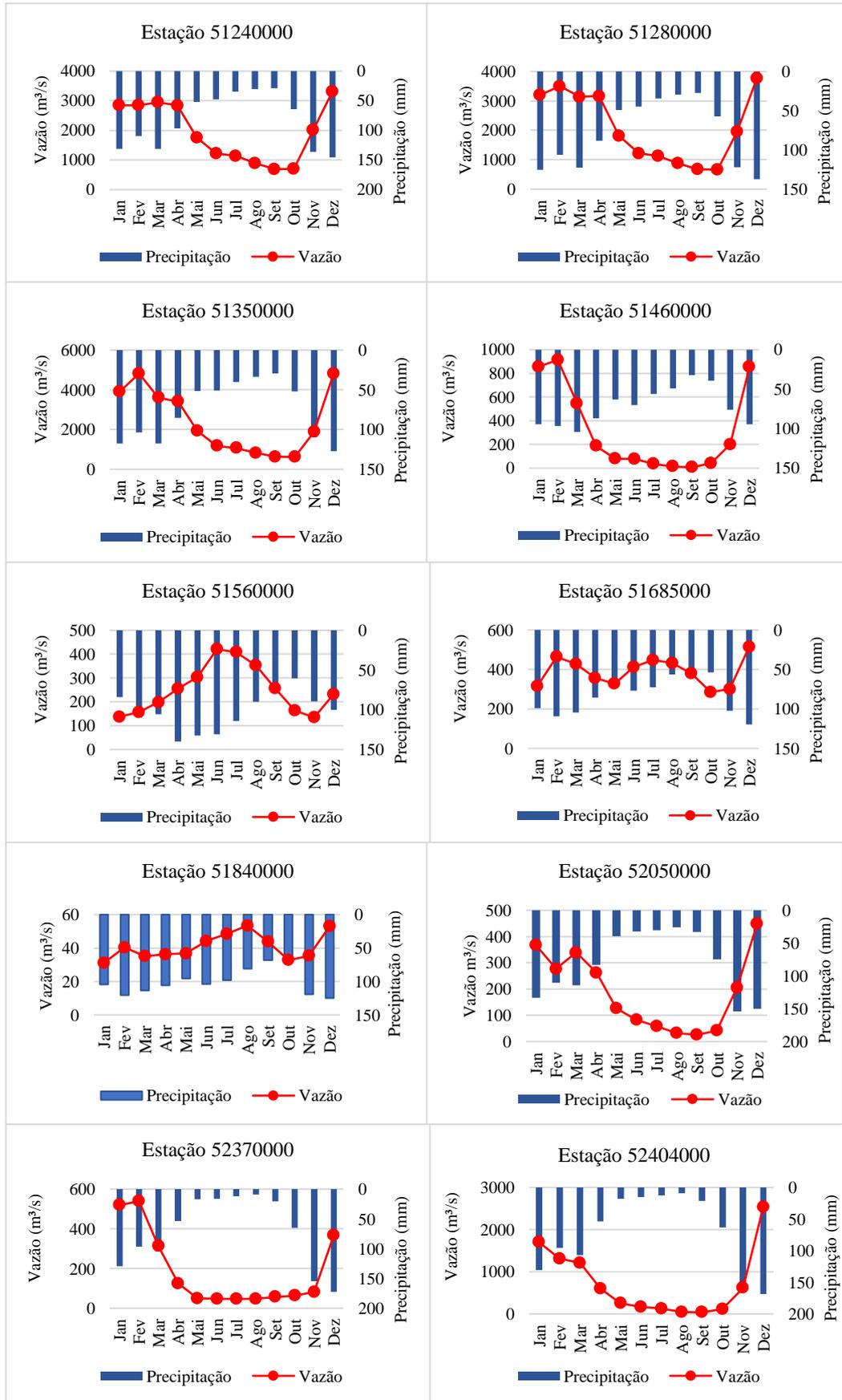
ZHOU, Y. *et al.* Impact of bushfire and climate variability on streamflow from forested catchments in southeast Australia. **Hydrological Sciences Journal**, v. 60, n. 7-8, p. 1340-1360, 3 ago. 2015.

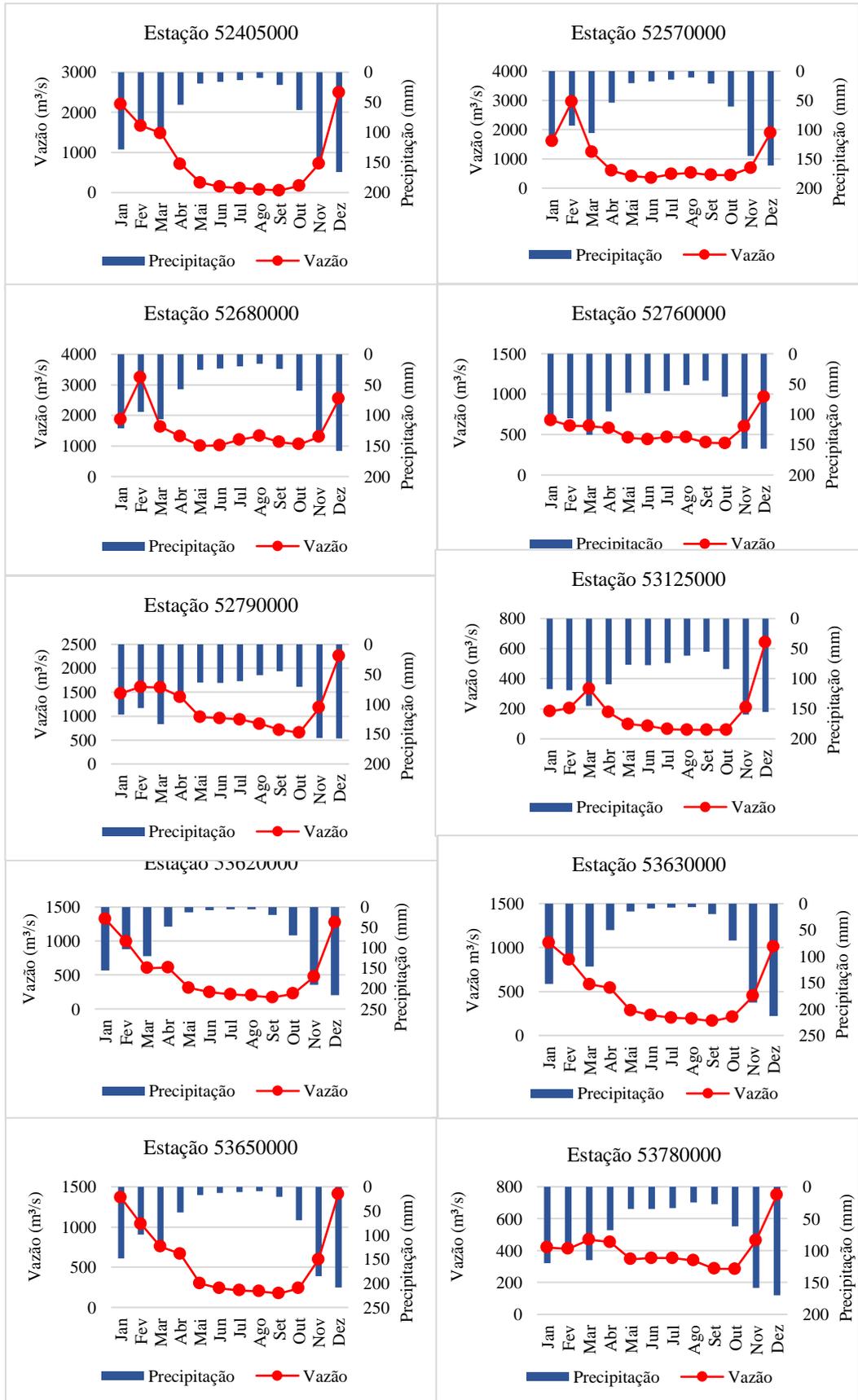
7. APÊNDICE A - HIDROGRAMA DA VAZÃO E HIETOGRAMA DA PRECIPITAÇÃO NAS ESTAÇÕES AVALIADAS

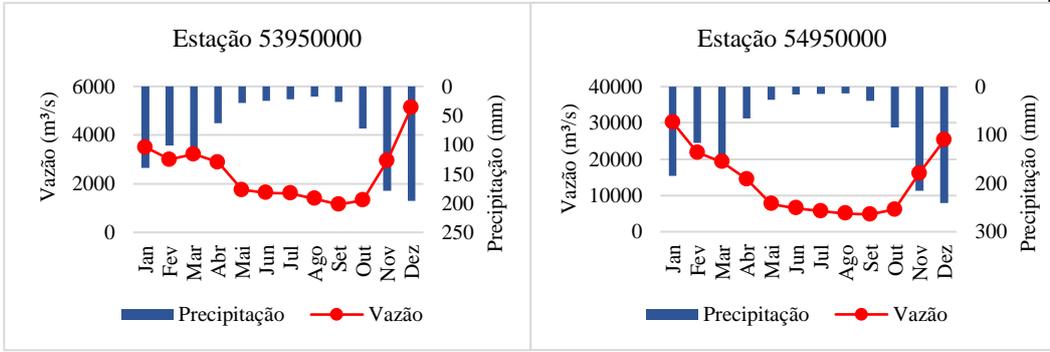












8. APÊNDICE B - PERÍODOS COM DADOS DE VAZÃO DISPONÍVEIS EM CADA ESTAÇÃO

Estação	Início	Fim	Estação	Início	Fim									
46650000	01/01/1980	31/12/1991	50290000	01/01/1980	17/03/1992	54950000	01/01/1980	23/07/1986	53125000	01/01/1980	30/09/1985	50620000	01/01/1980	29/10/1995
	01/02/1992	14/02/1992		07/04/1992	10/04/1992		01/08/1986	11/08/1986		23/10/1985	24/10/1987		01/01/1996	30/11/2014
	16/02/1992	29/02/1992		20/06/1992	31/10/1992		21/08/1986	11/08/1987		02/11/1987	18/02/1988		01/01/2015	07/05/2017
	16/05/1992	18/05/1998		30/11/1992	01/12/1992		17/08/1987	31/05/1989		21/02/1988	29/02/1988		01/06/2017	31/12/2018
	20/06/1998	31/03/2001		16/12/1992	17/12/1992		01/01/1990	28/02/1991		02/03/1988	29/09/1988	01/01/1980	16/02/1992	
	01/05/2001	30/04/2002		01/01/1993	31/01/1993		29/05/1991	31/05/1992		13/10/1988	01/11/1988	47480000	25/02/1992	30/09/2008
	31/05/2002	30/11/2003		01/05/1993	30/06/2006		15/08/1992	04/04/1993		07/11/1988	18/12/1988		01/11/2008	30/06/2010
	01/01/2004	30/04/2004		01/08/2006	30/11/2014		22/05/1993	31/12/2009		20/12/1988	23/01/1989	45960001	01/01/1980	31/01/2015
	01/06/2004	04/01/2008		01/01/2015	10/04/2015		03/02/2015	04/02/2015		28/01/1989	05/02/1989		01/03/2015	15/10/2015
	10/01/2008	19/07/2009		12/04/2015	28/05/2015		10/02/2015	10/02/2015		10/02/1989	12/02/1989		29/10/2015	01/07/2016
	01/08/2009	29/09/2010		01/06/2015	06/06/2015		12/02/2015	12/02/2015		14/02/1989	14/02/1989		01/09/2016	31/10/2017
	01/10/2010	05/10/2011		11/06/2015	17/06/2015		14/02/2015	15/02/2015		25/02/1989	25/02/1989	01/12/2017	31/12/2018	
	07/10/2011	05/12/2011		19/06/2015	22/06/2015		19/02/2015	19/02/2015		30/03/1989	30/03/1989	50330000	01/03/1984	02/05/1984
	17/12/2011	31/01/2012		24/06/2015	12/08/2015		21/02/2015	22/02/2015		01/04/1989	06/04/1989		01/06/1984	30/04/1985
	07/02/2012	10/02/2012		14/08/2015	17/12/2015		01/03/2015	01/03/2015		10/04/1989	19/04/1989		01/06/1985	03/07/1987
	12/02/2012	04/06/2012		19/12/2015	19/12/2015		07/03/2015	08/03/2015		04/05/1989	04/05/1989	01/08/1987	30/06/2006	
	12/06/2012	07/11/2012		21/12/2015	19/01/2016		11/03/2015	11/03/2015		08/05/1989	15/05/1989	01/08/2006	02/08/2010	
	11/11/2012	26/11/2012		22/01/2016	25/06/2016		14/03/2015	15/03/2015		23/05/1989	14/06/1989	01/01/1980	30/06/1984	
	29/11/2012	20/12/2012		28/06/2016	24/08/2016		19/03/2015	19/03/2015		17/06/1989	27/06/1989	20/07/1984	18/06/1999	
	01/01/2013	06/02/2013		26/08/2016	19/12/2016		21/03/2015	22/03/2015		29/06/1989	27/07/1989	50191000	24/07/1999	04/04/2017
	12/02/2013	04/04/2013		23/12/2016	05/03/2017		21/03/2015	22/03/2015		05/08/1989	10/08/1989		08/04/2017	31/12/2017
	10/04/2013	12/06/2013		07/03/2017	06/05/2017		25/03/2015	25/03/2015		17/08/1989	16/09/1989		01/02/2018	04/12/2018
	16/06/2013	20/10/2013		08/05/2017	12/05/2017		27/03/2015	28/03/2015		20/09/1989	08/12/1989	12/12/2018	31/12/2018	
	24/10/2013	15/01/2014		14/05/2017	16/06/2017		30/03/2015	03/04/2015		13/12/1989	13/12/1989	51120000	01/01/1980	14/05/1982
	19/01/2014	30/06/2014		18/06/2017	25/08/2017		10/04/2015	12/04/2015		16/12/1989	21/12/1989		16/05/1982	18/12/2014
	01/08/2014	30/09/2014		27/08/2017	23/09/2017		14/04/2015	17/04/2015		23/12/1989	29/12/1989		01/01/2015	31/12/2015
	01/01/2015	10/01/2015		25/09/2017	31/12/2018		24/04/2015	24/04/2015		01/01/1990	04/01/1990		04/01/2016	30/03/2018

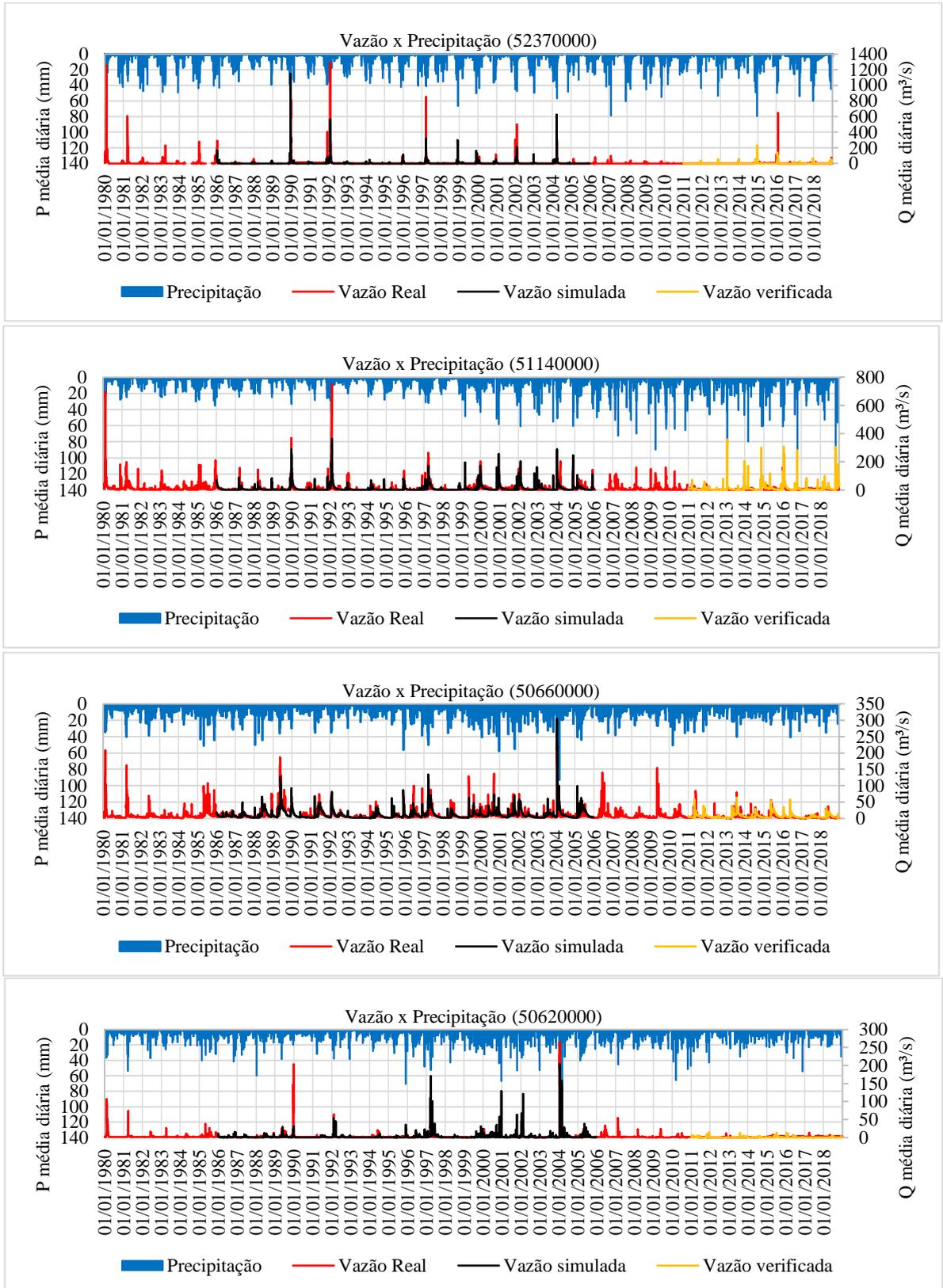
	15/01/2015	10/11/2015		01/01/1980	01/03/1981		30/04/2015	30/04/2015		08/01/1990	14/04/1990		01/04/2018	31/12/2018
	18/11/2015	06/05/2016		11/03/1981	31/08/1982		07/05/2015	07/05/2015		17/04/1990	14/11/1990	45260000	01/01/1980	12/05/2007
	01/06/2016	31/05/2017		01/10/1982	31/08/1983		18/10/1981	30/09/1982		30/11/1990	08/12/1990		01/06/2007	31/08/2009
	01/07/2017	31/12/2018		01/10/1983	31/01/1984		01/05/1984	31/08/1987		11/12/1990	12/11/1991		01/10/2009	28/03/2010
53650000	01/01/1980	25/08/1989	52370000	01/03/1984	30/04/1984	52404000	01/10/1987	30/06/2006	50250000	14/11/1991	07/01/1992		01/04/2010	30/03/2011
	19/10/1989	30/06/2006		01/09/1984	31/05/1985		01/08/2006	05/11/2008		10/01/1992	10/01/1992	01/04/2011	30/03/2015	
	01/08/2006	30/11/2014		01/10/1985	31/05/1986		13/11/2008	29/01/2012		13/01/1992	17/01/1992	01/04/2015	31/12/2018	
	01/01/2015	30/09/2015		01/07/1986	31/07/1986		01/11/2012	28/02/2013		19/01/1992	26/01/1992	01/01/1980	08/09/2001	
	27/11/2015	22/12/2015		01/10/1986	30/04/1987		01/04/2013	30/06/2013		29/01/1992	29/01/1992	10/09/2001	14/08/2009	
	09/01/2016	24/01/2016		01/11/1987	31/03/1988		01/08/2013	30/11/2014		05/02/1992	05/02/1992	17/08/2009	30/09/2012	
	04/02/2016	17/02/2016		01/12/1988	31/01/1989		01/01/2015	09/01/2016		17/02/1992	25/01/1995	01/11/2012	26/06/2014	
	22/02/2016	19/07/2016		01/04/1989	30/04/1989		19/01/2016	31/01/2016		20/03/1995	03/01/1998	02/07/2014	31/12/2018	
	26/07/2016	31/07/2016		01/06/1989	30/06/1989		02/02/2016	02/02/2016		22/01/1998	01/02/1998	01/01/1980	28/02/2011	
	01/09/2016	19/03/2017		01/11/1989	30/06/2006		01/03/2016	31/03/2016		01/03/1998	29/12/2003	31/03/2011	30/05/2014	
	30/03/2017	31/10/2017		01/08/2006	30/11/2014		01/05/2016	31/03/2017		01/01/2004	31/01/2007	01/06/2014	30/06/2015	
	01/12/2017	31/12/2018		01/01/2015	06/08/2016		01/05/2017	31/01/2018		01/03/2007	17/03/2008	28/08/2015	31/05/2016	
	50465000	01/01/1980		30/11/2014	12/08/2016		25/11/2016	04/02/2018		31/12/2018	20/03/2008	28/02/2010	06/06/2016	25/04/2017
		01/01/2015		31/12/2015	30/11/2016		29/11/2018	01/01/1980		30/04/2016	01/01/1980	30/11/2014	29/04/2017	31/12/2018
07/01/2016		31/12/2018	01/12/2018	31/12/2018	01/06/2016	31/12/2018	01/01/2015	31/12/2018						
52405000	26/11/1984	30/06/2006	46770000	01/01/1980	29/07/2011	50420000	01/01/1980	13/11/1989	50250000	01/01/1980	31/12/1993			
	01/08/2006	31/05/2012		01/08/2011	22/12/2011		01/06/1990	06/10/1990		01/02/1994	30/06/2006			
	01/10/2012	31/03/2014		26/12/2011	28/02/2012		20/11/1990	30/11/2014		01/08/2006	09/11/2007			
	01/05/2014	30/11/2014		01/03/2012	29/04/2014		01/01/2015	26/01/2016		28/11/2007	30/06/2014			
	01/01/2015	28/12/2015		01/05/2014	29/09/2014		08/03/2016	31/03/2016		01/08/2014	30/11/2014			
	01/01/2016	05/03/2017		01/10/2014	31/03/2018		01/05/2016	31/07/2017		01/01/2015	22/05/2015			
	30/03/2017	18/11/2017		22/08/2018	05/12/2018		01/09/2017	31/12/2018		27/05/2015	17/06/2015			
	20/11/2017	31/01/2018		07/12/2018	31/12/2018		01/01/1980	01/02/2005		01/07/2015	29/09/2015			
	05/02/2018	30/11/2018	48020000	01/01/1980	09/08/2010	27/07/2005	16/02/2006	01/10/2015		29/11/2015				

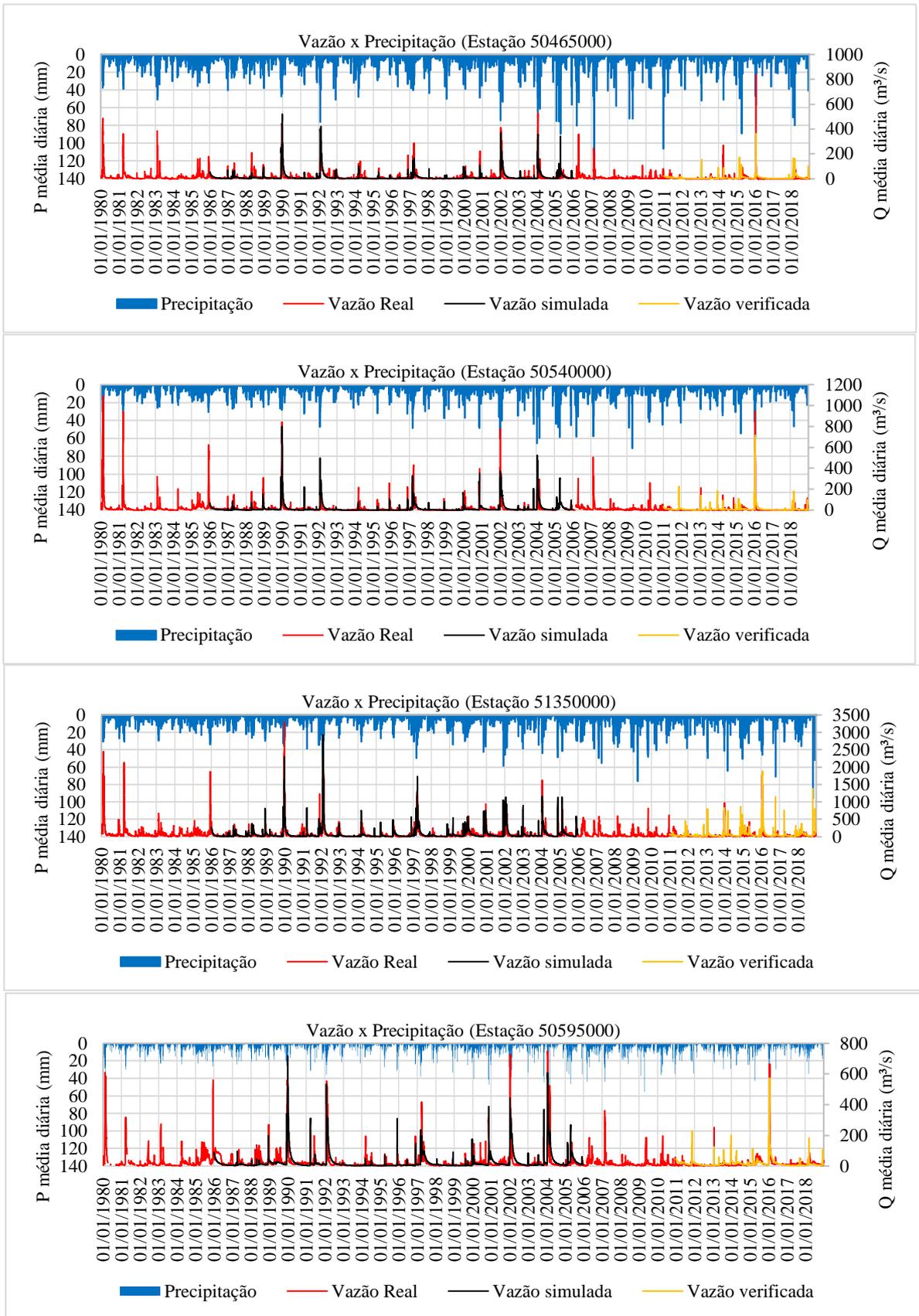
50660000	01/01/1980	04/09/1989		15/10/2010	23/03/2013		16/03/2006	18/03/2006		01/12/2015	13/04/2017			
	22/10/1989	30/10/1989		12/05/2013	19/06/2013		05/04/2006	30/11/2014		26/04/2017	04/09/2017			
	22/06/1990	30/11/2014		25/06/2013	19/07/2013		01/01/2015	05/11/2015		11/09/2017	11/09/2017			
	01/01/2015	30/06/2015		21/07/2013	21/07/2013		07/11/2015	09/11/2015		14/09/2017	15/09/2017			
	01/08/2015	31/12/2018		19/09/2014	30/04/2015		12/11/2015	14/11/2015		22/09/2017	31/12/2018			
50795000	01/01/1980	30/06/2006		01/06/2015	31/03/2016		08/12/2015	08/12/2015	52790000	01/01/1980	05/09/1993			
	01/08/2006	30/11/2014		01/10/2018	31/12/2018		29/12/2015	01/01/2016		01/10/1993	03/08/1994			
	01/01/2015	30/04/2015		01/01/1980	31/08/2009		04/01/2016	06/12/2018		01/10/1994	30/06/2006			
	01/06/2015	31/12/2018		01/10/2009	25/01/2016		08/12/2018	31/12/2018		01/08/2006	30/11/2014			
51230000	01/01/1980	14/12/1993	45170001	25/02/2016	05/07/2017	51140000	01/01/1980	30/11/2003	52680000	01/01/1980	30/06/2006			
	23/12/1993	24/12/1993		02/11/2017	04/11/2017		01/01/2004	31/01/2006		01/08/2006	30/11/2014			
	16/01/1994	20/01/1994		07/11/2017	28/11/2017		01/08/2006	31/10/2014		01/01/2015	31/12/2018			
	17/02/1994	30/11/2014		02/12/2017	26/12/2017		01/01/2015	31/03/2017		02/04/2017	08/10/2017			
	01/01/2015	29/11/2018		04/01/2018	13/01/2018		18/10/2017	30/01/2018		01/01/1980	30/11/2014			
	01/12/2018	31/12/2018		29/01/2018	03/09/2018		01/02/2018	31/12/2018		01/01/2015	31/05/2018			
51240000	01/01/1980	01/11/2012		23/10/2018	24/10/2018		01/02/2018	31/12/2018	45910001	22/10/2018	31/12/2018			
	03/11/2012	08/01/2013		29/10/2018	31/12/2018		01/08/1986	30/06/1996		01/01/1980	31/05/1991			
	13/01/2013	16/12/2014		01/01/1980	30/06/2006		01/10/1996	30/04/1998		01/10/1991	31/12/1993			
51460000	01/01/2015	31/12/2018	53950000	01/08/2006	30/11/2014	52050000	01/11/1998	30/06/2006	51840000	01/03/1994	30/06/2006			
	01/01/1980	29/02/1984		01/01/2015	18/10/2015		01/08/2006	31/03/2012		01/08/2006	30/11/2014			
	01/04/1984	30/09/1984		01/11/2015	01/11/2015		01/11/2012	31/12/2012		01/01/2015	30/01/2015			
	01/12/1984	17/04/1993		29/11/2015	29/11/2015		23/01/2013	06/02/2013		01/02/2015	29/03/2015			
	01/05/1993	31/10/1993		03/12/2015	24/12/2015		09/04/2013	30/11/2014		05/04/2015	17/04/2015			
	01/12/1993	01/11/1994		16/01/2016	16/03/2016		01/01/2015	13/03/2017		25/04/2015	15/11/2015			
	21/11/1994	31/10/2002		23/06/2016	31/07/2016		02/04/2017	26/04/2017		01/12/2015	08/12/2015			
	01/01/2003	29/02/2012		01/09/2016	31/12/2018		01/05/2017	31/01/2018		08/01/2016	07/02/2016			
	01/11/2012	30/11/2014		01/01/1980	31/01/1990		05/02/2018	31/12/2018		19/02/2016	27/02/2016			
	01/01/2015	31/12/2018		15/02/1990	30/06/2006		53620000	01/01/1980		30/06/2006	14/04/2016	26/04/2016		

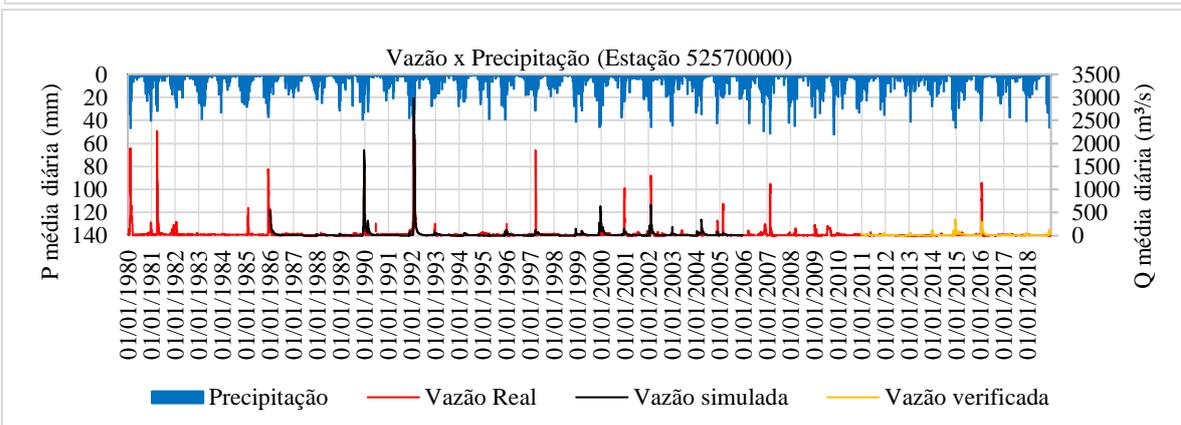
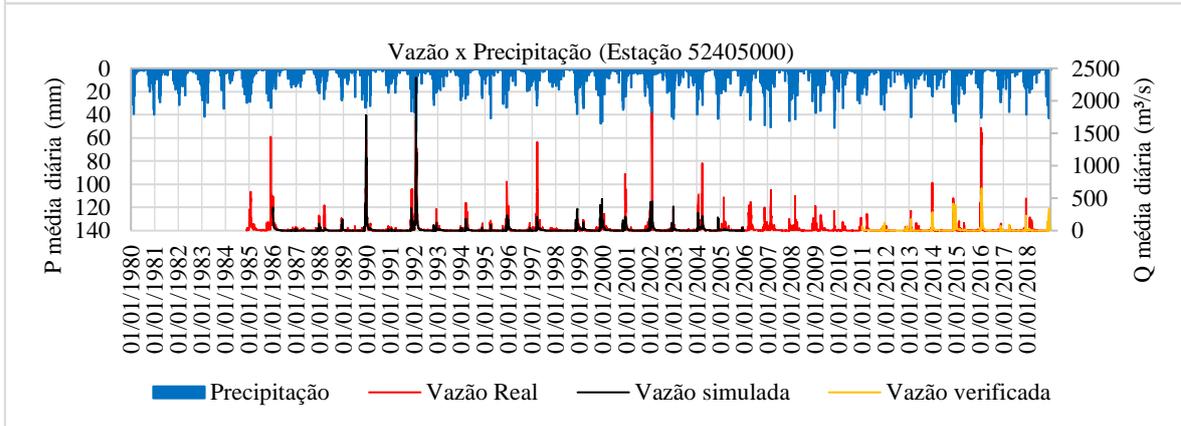
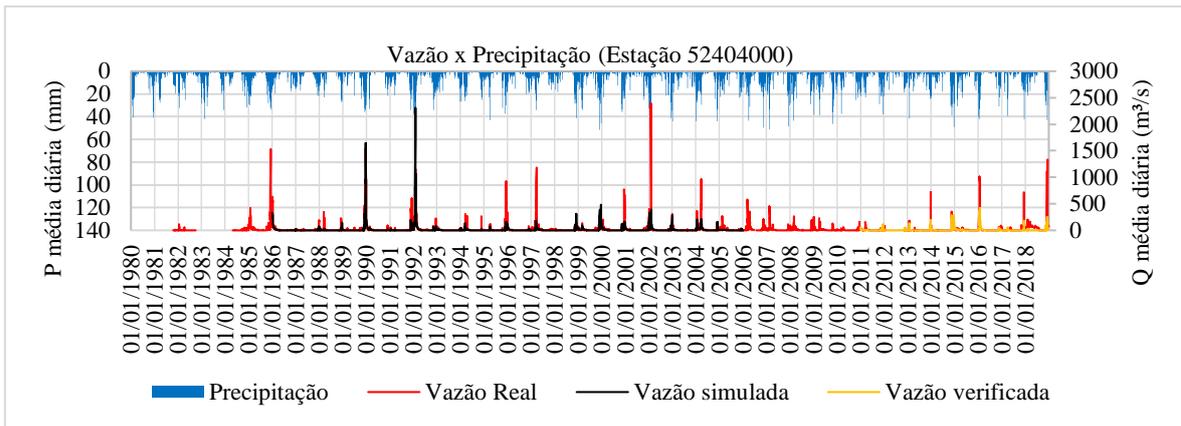
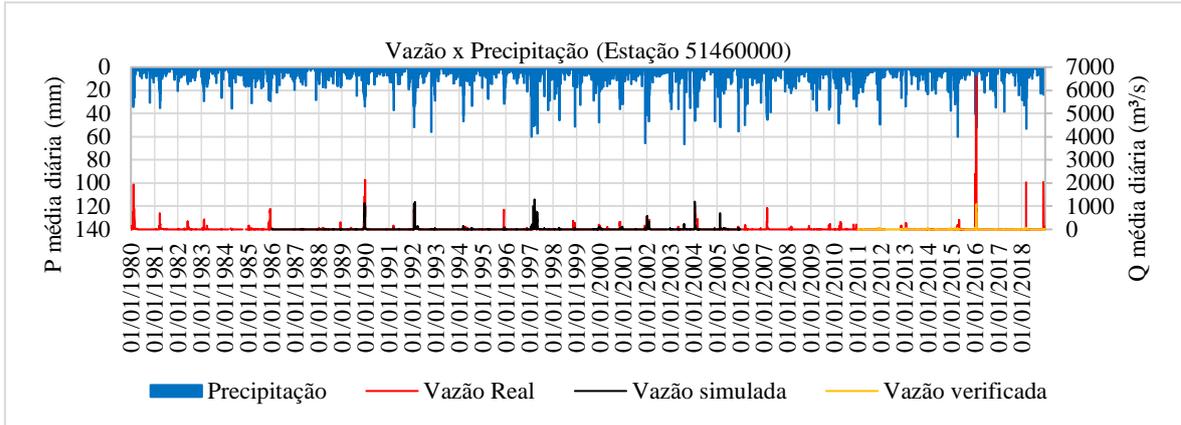
51685000	01/01/1980	31/12/1989		01/08/2006	30/11/2014		01/08/2006	30/11/2014		19/06/2016	28/11/2016			
	21/02/1990	30/06/2006		01/01/2015	31/07/2016		01/01/2015	28/06/2015		01/04/2017	31/12/2018			
	01/08/2006	31/07/2014		01/09/2016	31/12/2018		01/07/2015	31/10/2017		01/01/1980	04/04/1980			
	01/09/2014	30/11/2014		01/01/1980	30/06/2006		01/12/2017	31/12/2018		21/05/1980	04/06/1980			
	01/01/2015	31/12/2018		01/08/2006	30/11/2014		01/01/1980	18/12/2014		16/06/1980	16/06/1980			
52760000	01/01/1984	16/03/1993	53950000	01/01/2015	18/10/2015	51280000	01/01/2015	23/01/2016	50595000	11/09/1980	18/04/1981			
	20/04/1993	30/06/2006		01/11/2015	01/11/2015	26/01/2016	31/12/2018	17/05/1981		19/09/1981				
	01/08/2006	30/11/2014		29/11/2015	29/11/2015	01/01/1980	30/11/2003	21/10/1981		31/07/1982				
	01/01/2015	14/11/2017		03/12/2015	24/12/2015	01/01/2004	19/12/2014	01/09/1982		30/06/2006				
	16/11/2017	31/12/2018		16/01/2016	16/03/2016	01/01/2015	25/01/2016	01/08/2006		30/11/2014				
51560000	01/01/1980	30/06/2006		23/06/2016	31/07/2016	51350000	28/01/2016	21/02/2016		01/01/2015	28/07/2016			
	01/08/2006	30/11/2014		01/09/2016	31/12/2018		03/03/2016	31/12/2018		21/09/2016	09/11/2016			
	01/01/2015	31/12/2018		01/01/1980	07/06/1990		01/01/1980	01/02/2005		08/12/2016	31/12/2018			
53630000	01/01/1980	06/01/1990	52570000	21/06/1990	22/06/1990	50540000	27/07/2005	16/02/2006	45298000	01/01/1980	18/10/2017			
	13/02/1990	30/06/2006		01/07/1990	31/07/1990		16/03/2006	18/03/2006		01/11/2017	30/11/2017			
	01/08/2006	30/11/2014		01/09/1990	27/02/1992		05/04/2006	30/11/2014		25/06/2018	31/10/2018			
	01/01/2015	31/01/2015		01/08/1992	30/04/1993		01/01/2015	05/11/2015	01/12/2018	31/12/2018				
	01/03/2015	31/07/2016		20/07/1993	30/06/2006		07/11/2015	09/11/2015	01/01/1984	27/01/1984				
	01/09/2016	31/10/2017		01/08/2006	31/12/2013		12/11/2015	14/11/2015	18/03/1984	31/03/1999				
	01/12/2017	31/12/2018		01/02/2014	31/03/2014		08/12/2015	08/12/2015	01/08/1999	30/06/2006				
49330000	01/01/1980	30/12/2012	48290000	01/05/2014	30/11/2014		29/12/2015	01/01/2016	50380000	01/08/2006	30/06/2014			
	01/01/2013	31/12/2015		01/01/2015	31/12/2018		04/01/2016	06/12/2018		01/08/2014	30/11/2014			
	02/01/2016	31/12/2018		01/01/1980	30/09/2018		08/12/2018	31/12/2018		01/01/2015	31/12/2018			

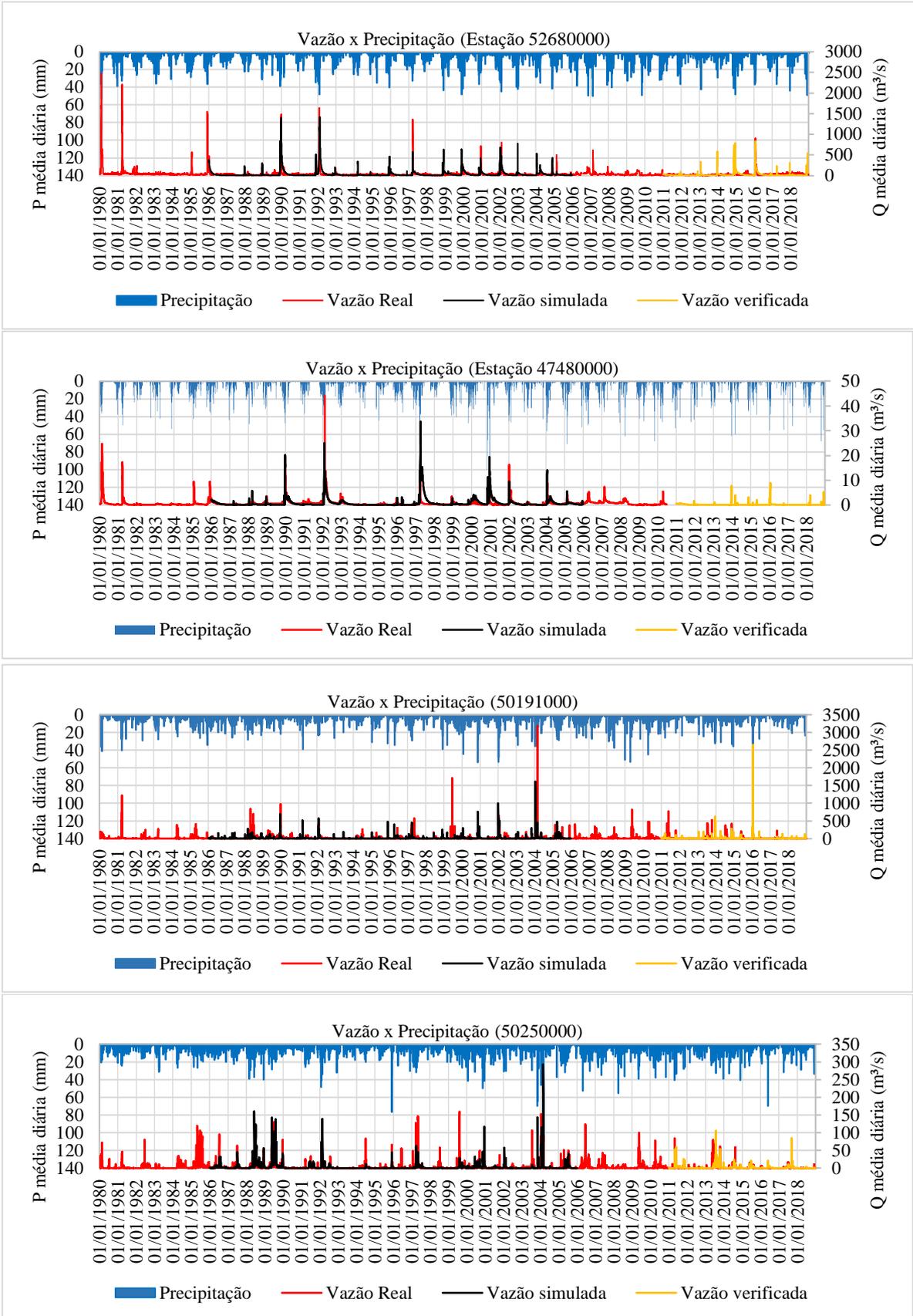
9. APÊNDICE C - RESUMO DAS VAZÕES E PRECIPITAÇÕES DAS ESTAÇÕES AVALIADAS

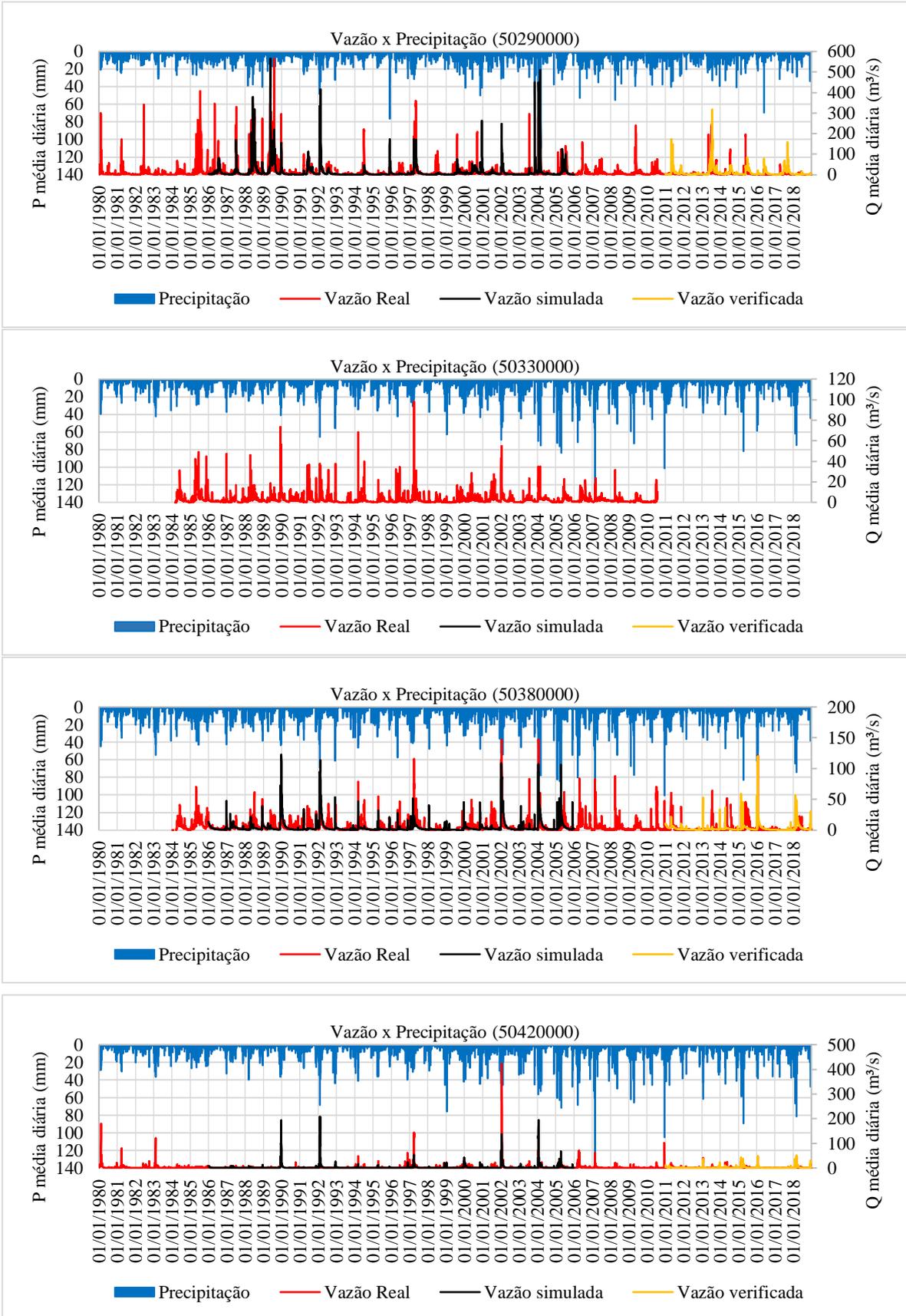
9.1. GRUPO 1



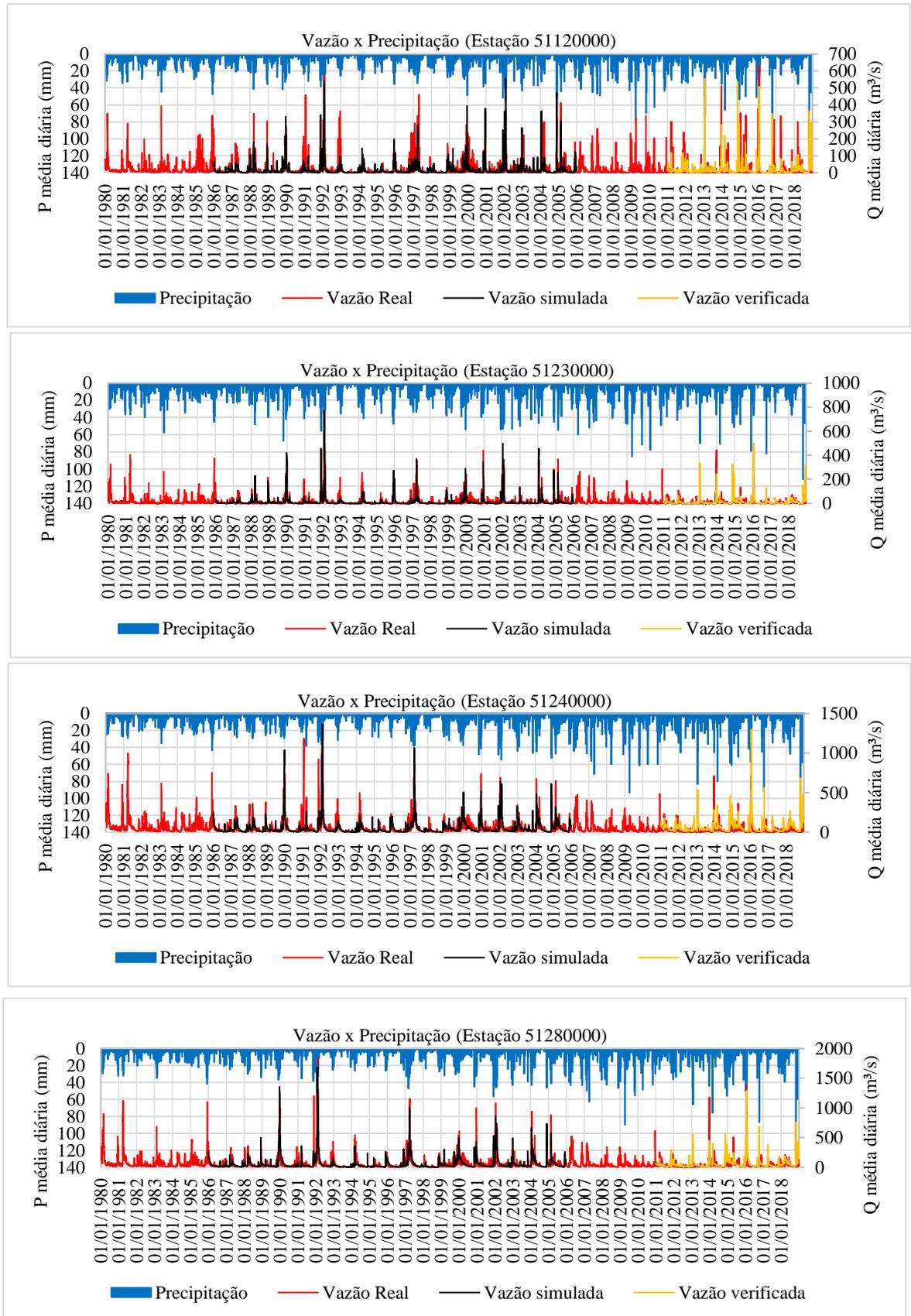


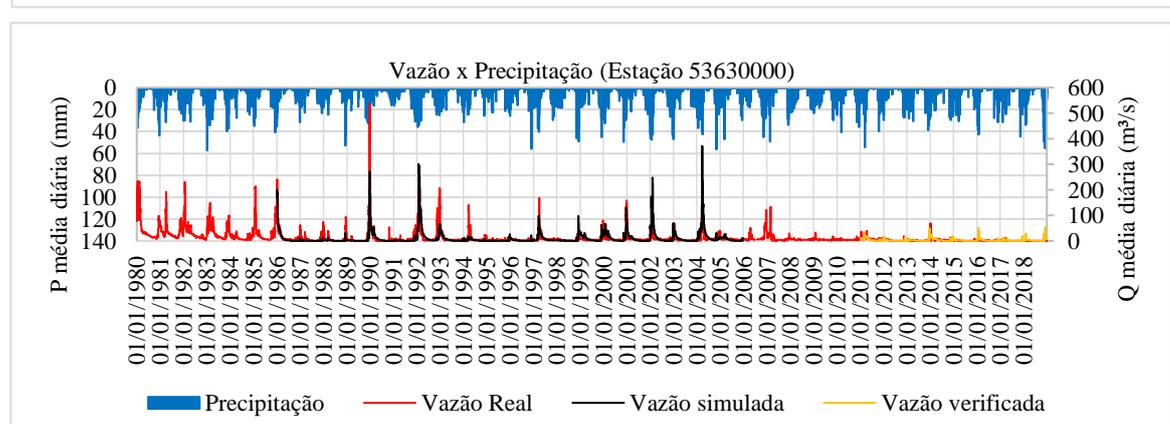
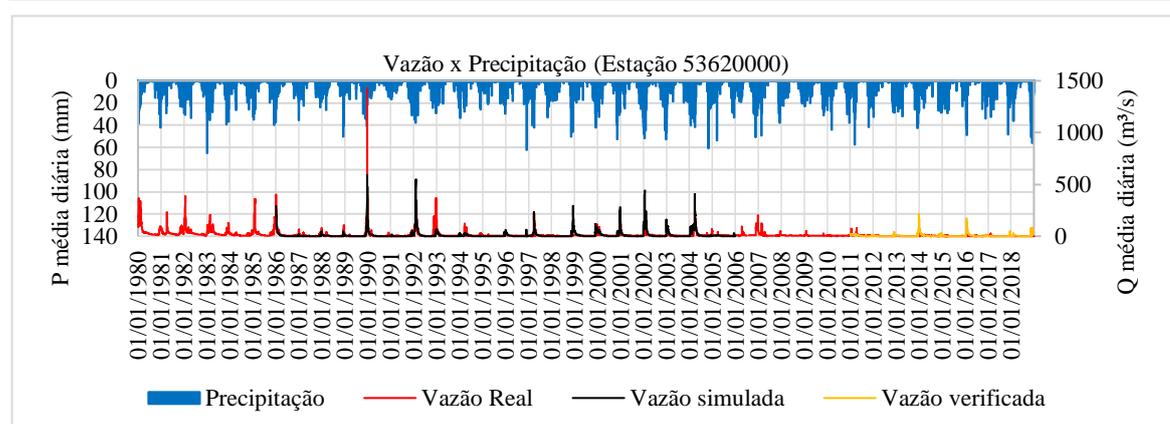
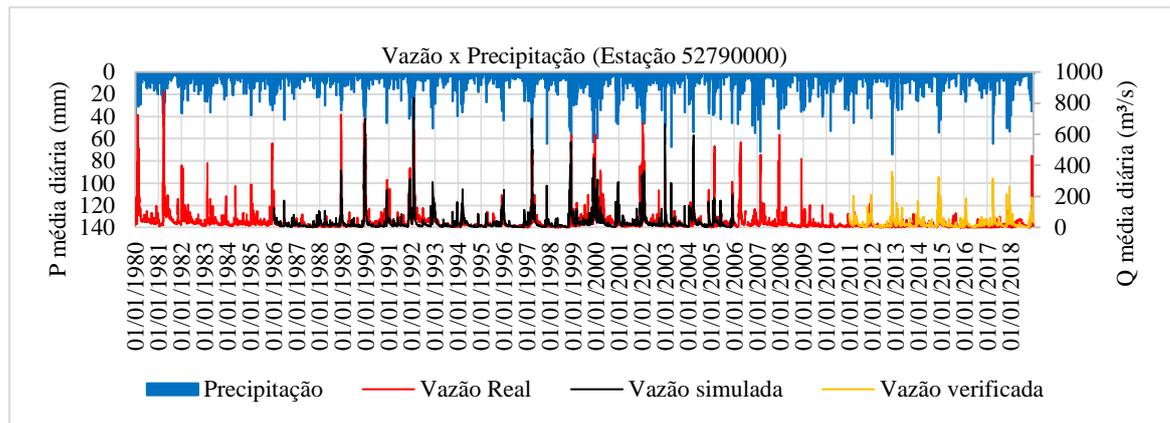
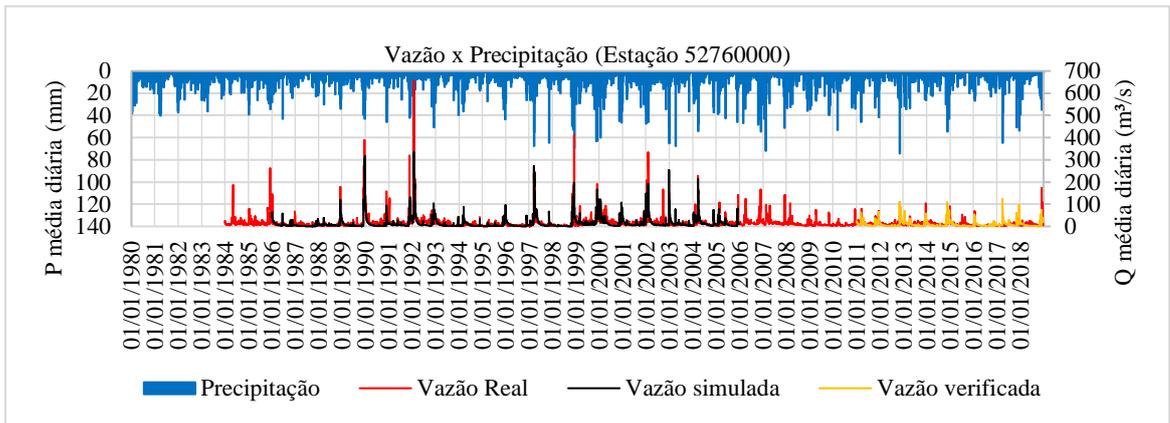


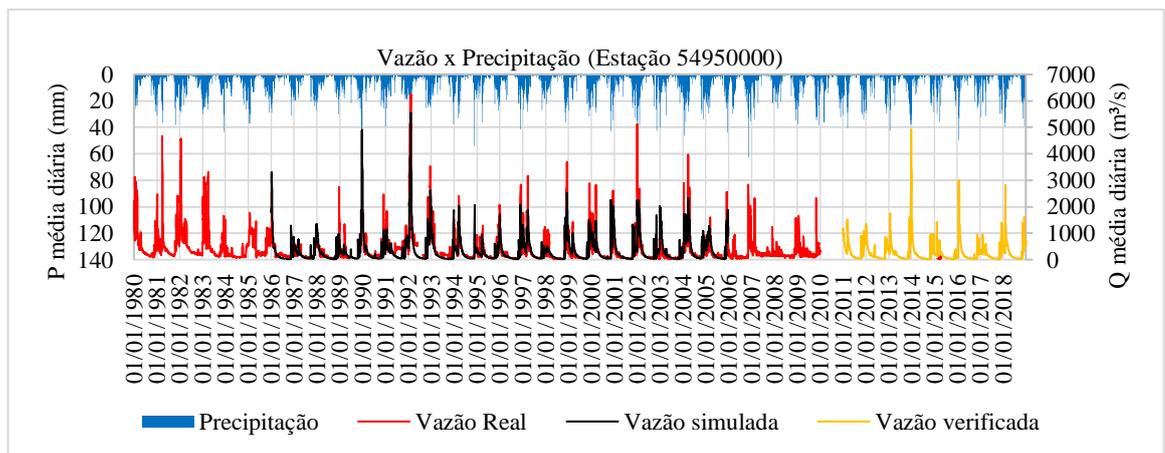
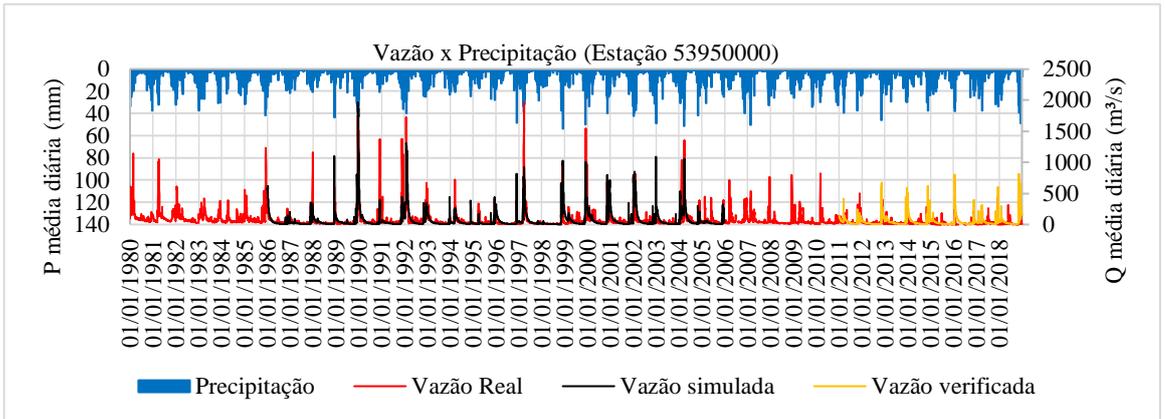
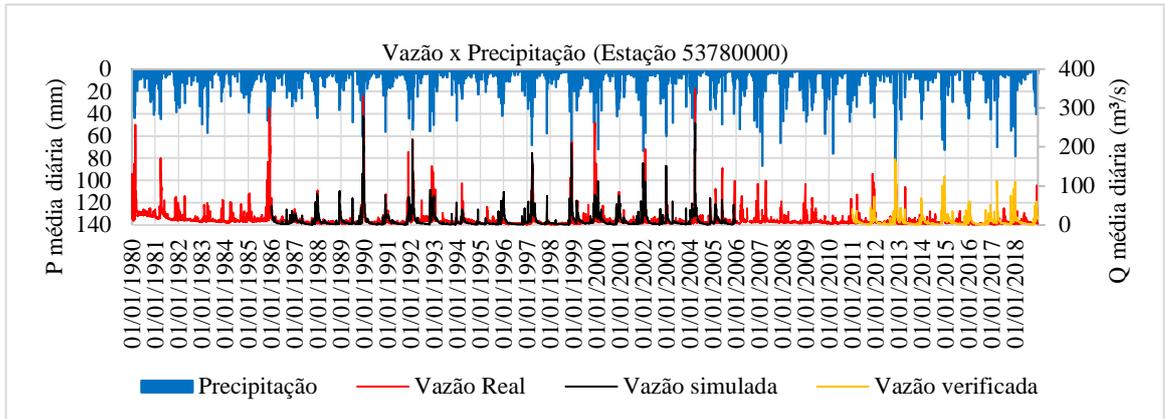
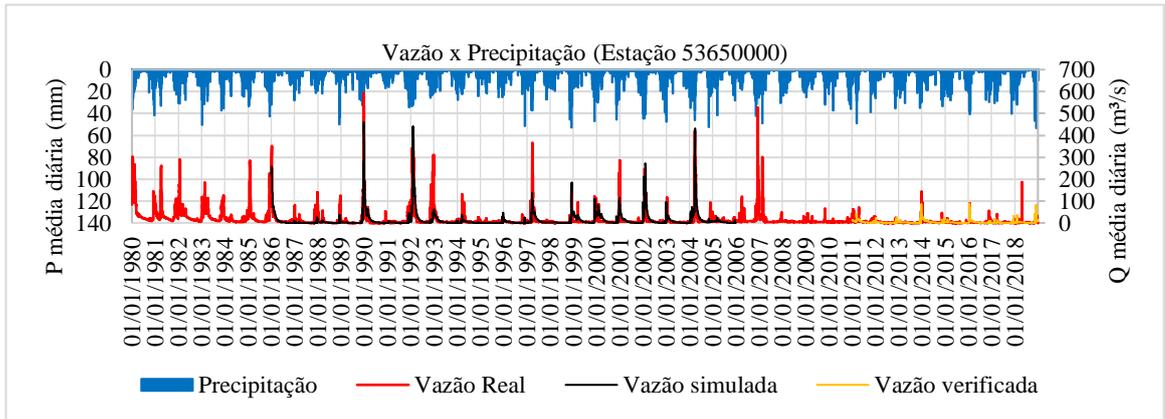


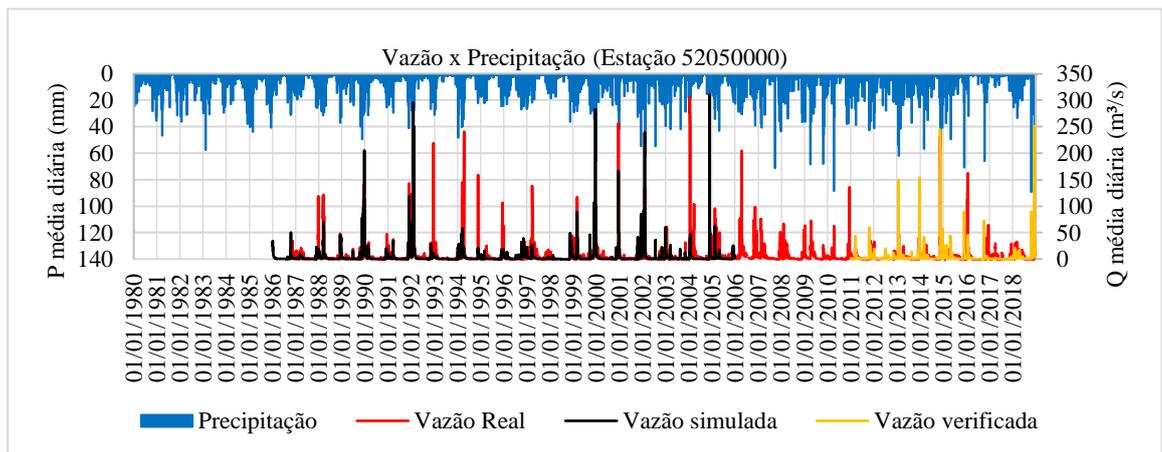
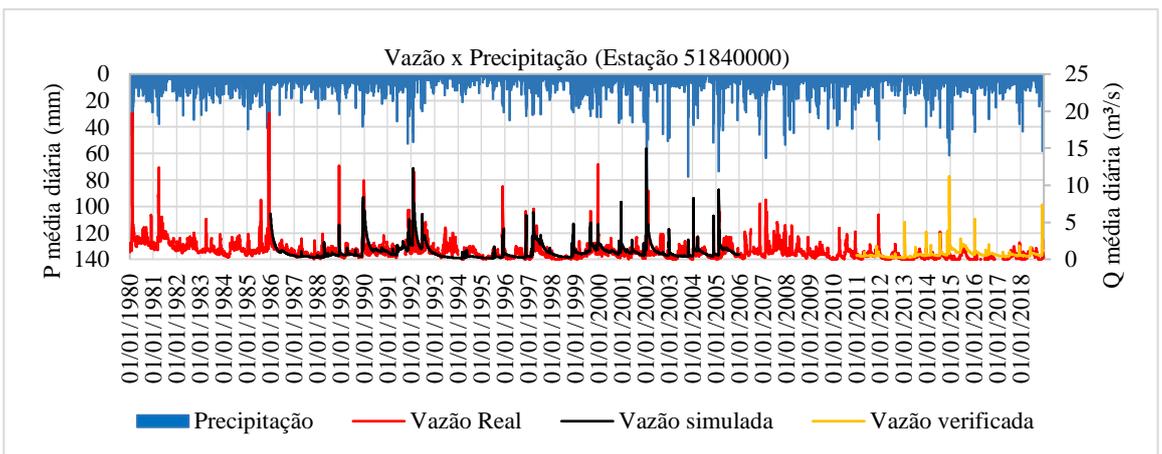
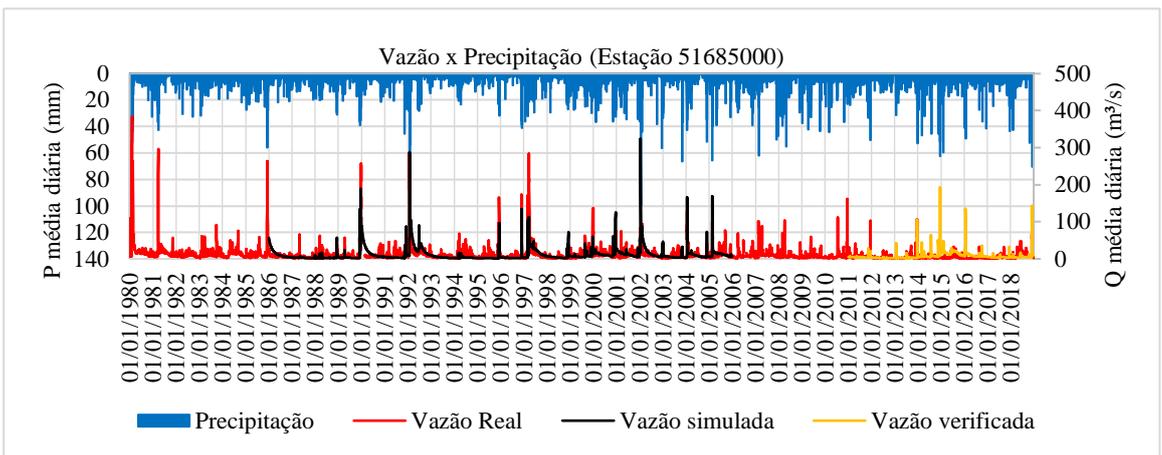
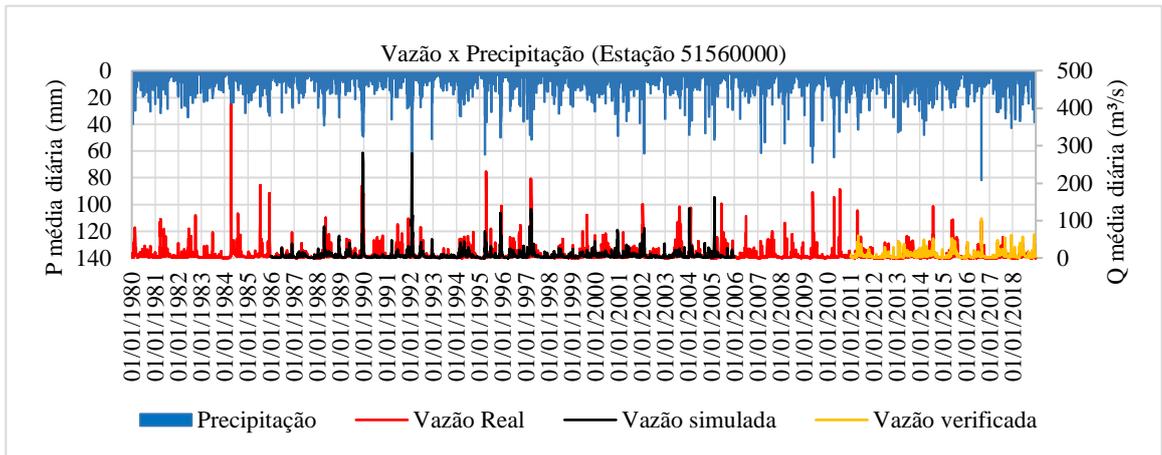


9.2. GRUPO 2

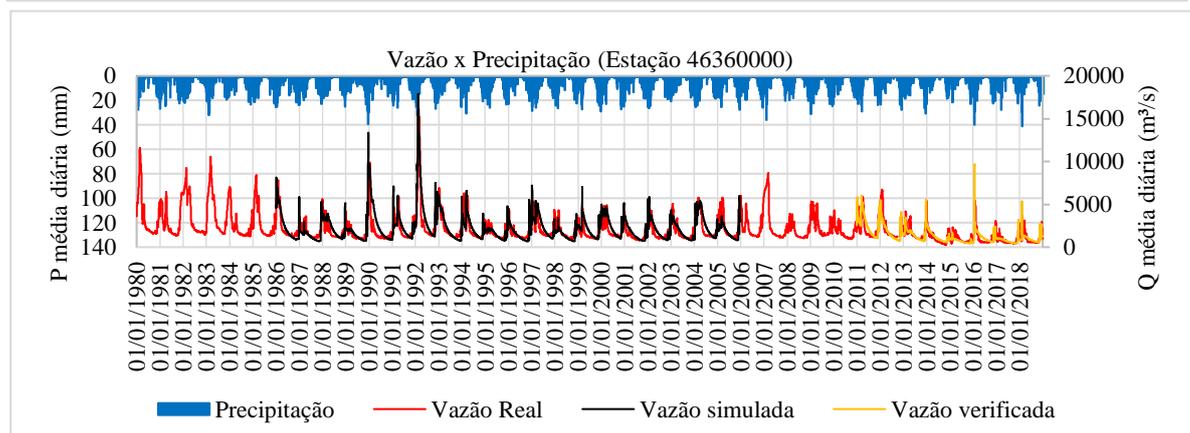
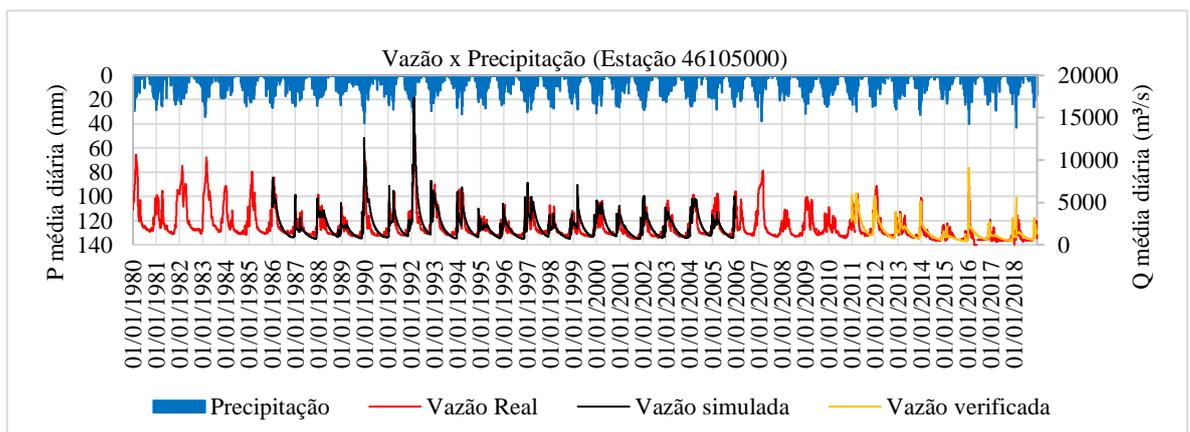
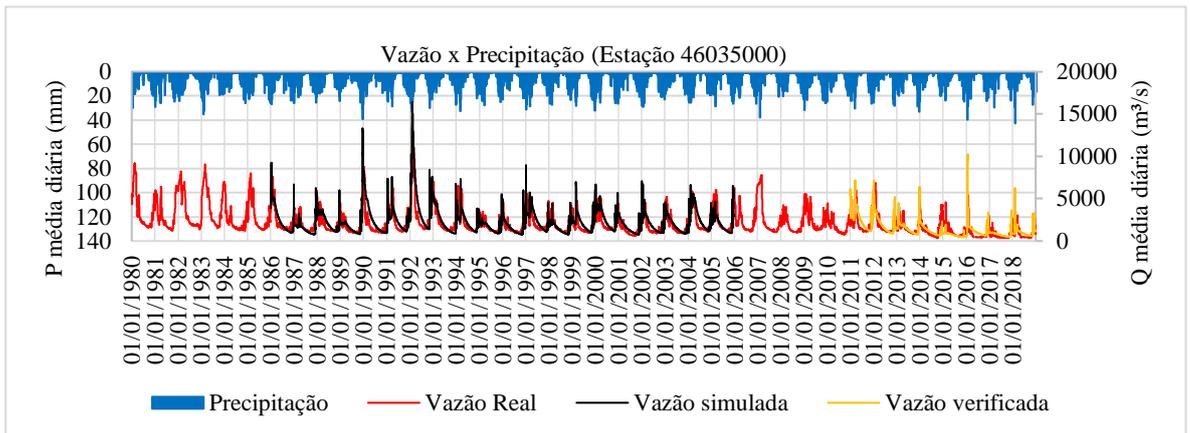
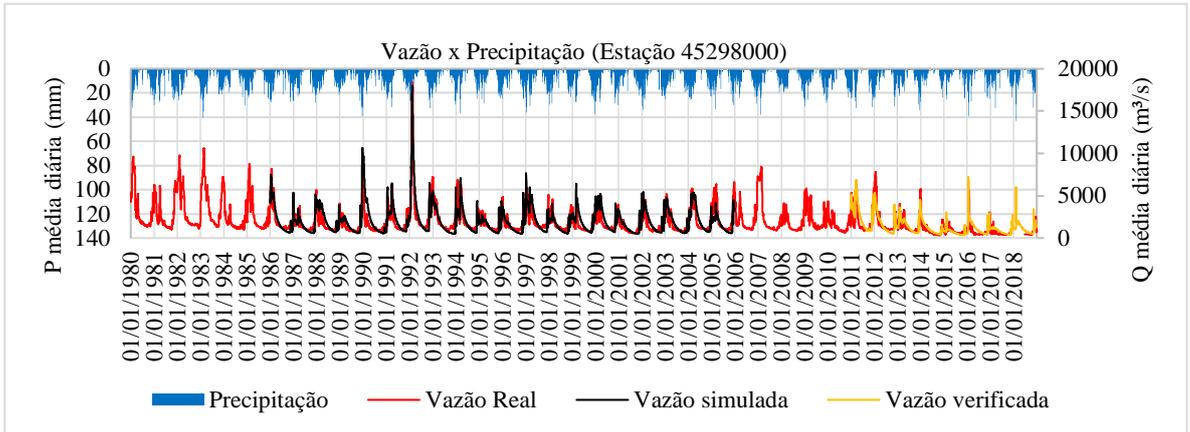


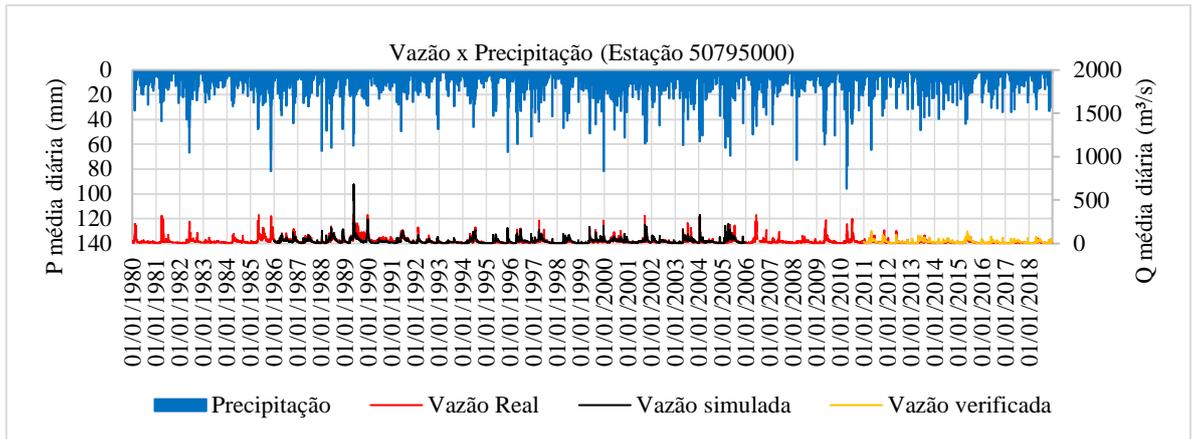






9.3. GRUPO 3





10. APÊNDICE D - CURVAS DE PERMANÊNCIA DAS ESTAÇÕES

