

UNIVERSIDADE FEDERAL DE PELOTAS
Centro de Desenvolvimento Tecnológico
Programa de Pós-Graduação em Recursos Hídricos



Dissertação de Mestrado

Regionalização de vazões mínimas para o estado do Rio Grande do Sul

Carina Krüger Bork

Pelotas, 2018.

CARINA KRÜGER BORK

REGIONALIZAÇÃO DE VAZÕES MÍNIMAS PARA O ESTADO DO RIO GRANDE
DO SUL

Dissertação apresentada ao Programa de Pós-Graduação em Recursos Hídricos do Centro de Desenvolvimento Tecnológico da Universidade Federal de Pelotas, como requisito parcial à obtenção do título de Mestre em Recursos Hídricos.

Orientador: Prof. Dr. Hugo Alexandre Soares Guedes
Co-orientador: Prof. Dr. Samuel Beskow

Pelotas, 2018.

Carina Krüger Bork

Regionalização de vazões mínimas para o estado do Rio Grande do Sul

Dissertação de Mestrado aprovada, como requisito parcial, para obtenção do grau de Mestre em Recursos Hídricos, Programa de Pós-Graduação em Recursos Hídricos do Centro de Desenvolvimento Tecnológico, da Universidade Federal de Pelotas.

Data da Defesa: 06/03/2018

Banca examinadora:

Prof. Dr. Hugo Alexandre Soares Guedes (Orientador)
Doutor em Engenharia Agrícola pela Universidade Federal de Viçosa

Prof. Dr. Bruno Müller Vieira
Doutor em Ciência e Engenharia de Materiais pela Universidade Federal de Pelotas

Profa. Dra. Claudia Fernanda Almeida Teixeira-Gandra
Doutora em Agronomia pela Universidade de São Paulo

Prof. Dr. Fabrício da Silva Terra
Doutor em Agronomia pela Universidade de São Paulo

Aos meus pais
Erno Bork e Serlei Krüger Bork
Aos meus avós
Udo Bork (*in memorian*) e Célia Bork (*in memorian*)
Albino Krüger (*in memorian*) e Irena Krüger
DEDICO.

Agradecimentos

À minha família por todo o suporte nesta caminhada, em especial aos meus pais Erno e Serlei, pela educação que me proporcionaram, e pela confiança a mim depositada, ao meu irmão André por todo apoio, meu respeito e amor.

Ao meu orientador, Professor Dr. Hugo Alexandre Soares Guedes, e co-orientador, Professor Dr. Samuel Beskow, pela orientação e, sobretudo, pela competência técnico-científica, meu respeito e admiração.

Aos demais professores do Programa de Pós-Graduação em Recursos Hídricos, por terem contribuído de forma significativa para o meu crescimento profissional e pessoal, muito obrigada.

Aos demais colegas do Programa de Pós-Graduação em Recursos Hídricos, por terem compartilhado esta etapa, muito obrigada.

À minha bolsista de iniciação científica Mylena Feitosa Tormam, por todo apoio e confiança a mim depositada.

Ao Centro de Engenharias, em especial ao Curso de Graduação em Engenharia Ambiental e Sanitária, por terem proporcionado a minha formação acadêmica.

À Universidade Federal de Pelotas, mais especificamente ao Programa de Pós-Graduação em Recursos Hídricos pela oportunidade de realização do mestrado.

"Feliz aquele que transfere o que sabe e aprende o que ensina."

Cora Coralina

Resumo

BORK, Carina Krüger. **Regionalização de vazões mínimas para o estado do Rio Grande do Sul**. 2018. 156f. Dissertação (Mestrado em Recursos Hídricos) – Programa de Pós-Graduação em Recursos Hídricos, Centro de Desenvolvimento Tecnológico, Universidade Federal de Pelotas, Pelotas, 2018.

Os conflitos pelo uso da água decorrem da falta de planejamento e gestão dos recursos hídricos, os quais estão ligados à inexistência de informações que associem as vazões já outorgadas com a atual disponibilidade hídrica. A determinação das vazões mínimas Q_{90} e Q_{95} , utilizadas como referência para os processos de outorga no estado do Rio Grande do Sul, é realizada a partir da coleta de dados de estações fluviométricas. Entretanto estas não contemplam toda a hidrografia, sendo necessário o uso da técnica de regionalização de vazões, com o objetivo de suprir a carência de dados em locais com pouca ou nenhuma informação. A análise estatística multivariada de agrupamento consiste numa ferramenta exploratória dos dados com intuito de classificar grupos homogêneos. Desta forma, utilizou-se o método hierárquico por meio das distâncias de dissimilaridade Euclidiana, Manhattan e Mahalanobis combinadas com os algoritmos de agrupamento de ligação simples, completa e de Ward, e o método não hierárquico k-means para obter relações entre as características pré-estabelecidas das bacias que contemplam a região em análise. O presente estudo teve como objetivo regionalizar as vazões mínimas Q_{90} e a Q_{95} no estado do Rio Grande do Sul utilizando a análise de agrupamento para a demarcação das regiões homogêneas, além de avaliar quais parâmetros influenciam em maior magnitude na determinação das vazões regionalizadas. Os resultados apontam que no geral as distâncias Euclidiana e Manhattan concordam mais do que a distância de Mahalanobis, entre os algoritmos de agrupamento destaca-se o algoritmo de Ward com o maior número de possibilidades para o corte no dendrograma. O método não hierárquico k-means também se mostrou favorável para determinação das regiões. Depois de testadas e analisadas todas as alternativas de cortes dos dendrogramas, juntamente com os grupos pelo método k-means, observou-se que a divisão de cinco regiões para o estado é a que proporciona os melhores resultados e ajustes para as equações de regionalização das vazões Q_{90} e Q_{95} . A medida de heterogeneidade também mostrou-se importante para avaliar as regiões delimitadas, havendo a necessidade de exclusão de cinco estações de uma região para que fosse considerada homogênea. Entre as variáveis que melhor explicam as vazões encontra-se a área ou o perímetro nas quatro primeiras regiões, sendo que apenas na região V as variáveis foram o perímetro e a precipitação total anual média. Desta forma, a demarcação das regiões homogêneas, juntamente com as equações encontradas podem ser utilizadas para um melhor planejamento dos recursos hídricos pelos órgãos gestores no estado do Rio Grande do Sul.

Palavras-chaves: Q_{90} e Q_{95} ; análise de agrupamento; gestão de recursos hídricos.

Abstract

BORK, Carina Krüger. **Regionalization of minimum flows for the of Rio Grande do Sul state**. 2018. 156p. Dissertation (Master's Degree in Water Resources) – Graduate Program in Water Resources, Center of Technological Development, Federal University of Pelotas, Pelotas, 2018.

Conflicts over water use arise from the lack of planning and management of water resources, which are linked to the inexistence of information that associates flows already granted with current water availability. The determination of the minimum flows Q_{90} and Q_{95} , used as reference for the granting processes in the state of Rio Grande do Sul, is performed from the data collection of fluviometric stations. However, these do not contemplate all the hydrography, being necessary the use of the regionalization technique of flow, with the purpose of supplying the lack of data in places with little or no information. The multivariate cluster analysis consists of an exploratory data tool to classify homogeneous groups. In this way, the use of the hierarchical method for the Euclidian, Manhattan and Mahalanobis dissimilarity grouping combined with the simple link clustering algorithms, complete and Ward, and the non-hierarchical method k-means for the relations between the banks of the basins that contemplate the region under analysis. The present study aimed at regionalizing the Q_{90} and Q_{95} minimum flows in the state of Rio Grande do Sul using cluster analysis for the homogeneous regions demarcation, as well as evaluating which parameters have a greater influence on the determination of regionalized flows. The results show that, in general, the Euclidean and Manhattan distances agree more than the distance of Mahalanobis, among the grouping algorithms, the Ward algorithm is the most likely to cut the dendrogram. The non-hierarchical k-means method was also favorable in determining the regions. After testing and analyzing all dendrogram cutting alternatives, together with the groups by the k-means method, it was observed that the division of five regions into the state is the one that provides the best results and adjustments for the regionalization equations of the dendrograms flow rates Q_{90} and Q_{95} . The measure of heterogeneity was also important to evaluate the delimited regions, and it was necessary to exclude five stations from one region to be considered homogeneous. Among the variables that best explain the flows are the area or the perimeter in the first four regions, and only in the V region were the perimeter and mean annual total precipitation. In this way, the demarcation of the homogeneous regions together with the equations can be used for a better planning of the water resources by the management organs in the state of Rio Grande do Sul.

Key-words: Q_{90} e Q_{95} ; cluster analysis; management of water resources.

Lista de Figuras

Figura 1 - Localização da área de estudo e regiões hidrográficas.	57
Figura 2 - Fluxograma utilizado para delimitação das bacias hidrográficas, a partir do banco de dados TOPODATA.	63
Figura 3 - Localização das estações pluviométricas e fluviométricas situadas no estado do Rio Grande do Sul e que atenderam aos critérios pré-estabelecidos.....	71
Figura 4 - Modelo digital de elevação do Rio Grande do Sul com destaque para as bacias hidrográficas em estudo.....	72
Figura 5 - Variabilidade espacial da precipitação total anual média no Rio Grande do Sul, de acordo com a metodologia dos polígonos de Thiessen.	73
Figura 6 - Dendrogramas obtidos a partir da combinação do algoritmo de ligação simples de agrupamento com as variáveis independentes para a Q_{90} com as distâncias: (a) Euclidiana, (b) Mahalanobis e (c) Manhattan.....	78
Figura 7 - Dendrogramas obtidos a partir da combinação do algoritmo de ligação completa de agrupamento com as variáveis independentes para a Q_{90} com as distâncias: (a) Euclidiana, (b) Mahalanobis e (c) Manhattan.....	79
Figura 8 - Dendrogramas obtidos a partir da combinação do algoritmo de Ward de agrupamento com as variáveis independentes para a Q_{90} com as distâncias: (a) Euclidiana, (b) Mahalanobis e (c) Manhattan.	80
Figura 9 - Dendrogramas obtidos a partir da combinação do algoritmo de ligação simples de agrupamento com as variáveis independentes para a Q_{95} com as distâncias: (a) Euclidiana, (b) Mahalanobis e (c) Manhattan.....	82
Figura 10 - Dendrogramas obtidos a partir da combinação do algoritmo de ligação completa de agrupamento com as variáveis independentes para a Q_{95} com as distâncias: (a) Euclidiana, (b) Mahalanobis e (c) Manhattan.....	83
Figura 11 - Dendrogramas obtidos a partir da combinação do algoritmo de Ward de agrupamento com as variáveis independentes para a Q_{95} com as distâncias: (a) Euclidiana, (b) Mahalanobis e (c) Manhattan.	84
Figura 12 - Regiões para o estado do Rio Grande do Sul determinadas usando o algoritmo de ligação completa de agrupamento combinado com a distância Euclidiana e Manhattan.	86

Figura 13 - Regiões para o estado do Rio Grande do Sul determinadas usando o algoritmo de Ward de agrupamento combinado com a distância Mahalanobis.	86
Figura 14 - Regiões para o estado do Rio Grande do Sul determinadas usando o algoritmo de Ward de agrupamento combinado com a distância Euclidiana e Manhattan.	87
Figura 15 - Regiões para o estado do Rio Grande do Sul determinadas usando o método k-means com k=2.	87
Figura 16 - Relação entre os valores da Q_{90} observados e estimados (m^3s^{-1}) para as regiões homogêneas delimitadas pela distância Euclidiana e Manhattan, com o algoritmo de ligação completa de agrupamento.	89
Figura 17 - Relação dos valores da Q_{90} observados e estimados (m^3s^{-1}) para as regiões homogêneas pela distância de Mahalanobis, com o algoritmo de Ward de agrupamento.	90
Figura 18 - Relação dos valores da Q_{90} observados e estimados (m^3s^{-1}) para as regiões homogêneas pela distância Euclidiana e Manhattan, com o algoritmo de Ward de agrupamento.	92
Figura 19 - Relação dos valores da Q_{90} observados e estimados (m^3s^{-1}) para as regiões homogêneas pelo método k-means com k=2.	93
Figura 20 - Regiões para o estado do Rio Grande do Sul determinadas usando o algoritmo de Ward de agrupamento combinado com a distância Euclidiana e Manhattan.	94
Figura 21 - Regiões determinadas para o estado do Rio Grande do Sul pelo método k-means com k=3.	95
Figura 22 - Relação dos valores da Q_{90} observados e estimados (m^3s^{-1}) para as regiões homogêneas determinadas pela distância Euclidiana e Manhattan, com o algoritmo de Ward de agrupamento.	96
Figura 23 - Relação dos valores da Q_{90} observados e estimados (m^3s^{-1}) para as regiões homogêneas pelo método k-means com k=3.	98
Figura 24 - Regiões para o estado do Rio Grande do Sul determinadas usando o algoritmo de Ward de agrupamento combinado com a distância Euclidiana e Manhattan.	99
Figura 25 - Regiões determinadas para o estado do Rio Grande do Sul pelo método k-means, com k=4.	99

Figura 26 - Relação dos valores da Q_{90} observados e estimados (m^3s^{-1}) para as regiões homogêneas determinadas pela distância Euclidiana e Manhattan, com o algoritmo de Ward de agrupamento.	100
Figura 27 - Relação dos valores da Q_{90} observados e estimados (m^3s^{-1}) para as regiões homogêneas pelo método k-means com $k=4$	102
Figura 28 - Regiões para o estado do Rio Grande do Sul determinadas usando o algoritmo de Ward de agrupamento combinado com a distância Euclidiana.	103
Figura 29 - Regiões para o estado do Rio Grande do Sul determinadas usando o algoritmo de Ward de agrupamento combinado com a distância Manhattan.	104
Figura 30 - Regiões determinadas para o estado do Rio Grande do Sul pelo método k-means, com $k=5$	104
Figura 31 - Relação dos valores da Q_{90} observados e estimados (m^3s^{-1}) para as regiões delimitadas pela distância Euclidiana, com o algoritmo de Ward de agrupamento.	106
Figura 32 - Relação dos valores da Q_{90} observados e estimados (m^3s^{-1}) para as regiões pela distância Manhattan, com o algoritmo de Ward de agrupamento.	108
Figura 33 - Relação dos valores da Q_{90} observados e estimados (m^3s^{-1}) para as regiões delimitadas pelo método k-means com $k=5$	110
Figura 34 - Regiões determinadas para o estado do Rio Grande do Sul pelo método k-means com $k=6$	111
Figura 35 - Relação dos valores da Q_{90} observados e estimados (m^3s^{-1}) para as regiões determinadas pelo método k-means com $k=6$	113
Figura 36 - Regiões homogêneas para o estado do Rio Grande do Sul determinadas usando o algoritmo de Ward de agrupamento combinado com a distância Euclidiana e Manhattan.	114
Figura 37 - Relação dos valores da Q_{90} observados e estimados (m^3s^{-1}) para as regiões homogêneas determinadas pela distância Euclidiana e Manhattan, com o algoritmo de Ward de agrupamento.	116
Figura 38 - Relação dos valores da Q_{95} observados e estimados (m^3s^{-1}) para as regiões homogêneas determinadas pela distância Euclidiana e Manhattan, com o algoritmo de ligação completa de agrupamento.	118
Figura 39 - Regiões para o estado do Rio Grande do Sul determinadas usando o algoritmo de Ward combinado com a distância Mahalanobis, com o algoritmo de Ward de agrupamento.	119

Figura 40 - Relação dos valores da Q_{95} observados e estimados (m^3s^{-1}) para as regiões homogêneas determinadas pela distância Mahalanobis, com o algoritmo de Ward de agrupamento.....	120
Figura 41 - Relação dos valores da Q_{95} observados e estimados (m^3s^{-1}) para as regiões homogêneas determinadas pela distância Euclidiana e Manhattan, com o algoritmo de Ward de agrupamento.	121
Figura 42 - Relação dos valores da Q_{95} observados e estimados (m^3s^{-1}) para as regiões homogêneas determinadas pelo método k-means com $k=2$	122
Figura 43 - Relação dos valores da Q_{95} observados e estimados (m^3s^{-1}) para as regiões homogêneas determinadas pela distância Euclidiana e Manhattan, com o algoritmo de Ward de agrupamento.	123
Figura 44 - Relação dos valores da Q_{95} observados e estimados (m^3s^{-1}) para as regiões homogêneas determinadas pelo método k-means com $k=3$	125
Figura 45 - Relação dos valores da Q_{95} observados e estimados (m^3s^{-1}) para as regiões homogêneas determinadas pela distância Euclidiana e Manhattan, com o algoritmo de Ward de agrupamento.	126
Figura 46 - Relação dos valores da Q_{95} observados e estimados (m^3s^{-1}) para as regiões homogêneas determinadas pelo método k-means com $k=4$	128
Figura 47 - Relação dos valores da Q_{95} observados e estimados (m^3s^{-1}) para as regiões determinadas pela distância Euclidiana e Manhattan, com o algoritmo de Ward de agrupamento.	130
Figura 48 - Regiões determinadas para o estado do Rio Grande do Sul pelo método k-means com $k=5$	131
Figura 49 - Regiões determinadas para o estado do Rio Grande do Sul pelo método k-means com $k=6$	132
Figura 50 - Relação dos valores da Q_{95} observados e estimados (m^3s^{-1}) para as regiões determinadas pelo método k-means com $k=6$	133
Figura 51 - Relação dos valores da Q_{95} observados e estimados (m^3s^{-1}) para as regiões homogêneas determinadas pela distância Euclidiana e Manhattan, com o algoritmo de Ward de agrupamento modificado.	137

Lista de Tabelas

Tabela 1 - Comitês das bacias hidrográficas do estado do Rio Grande do Sul	26
Tabela 2 - Critério de interpretação do índice “c”	55
Tabela 3 - Cenas referentes ao estado do Rio Grande do Sul no banco de dados TOPODATA.....	62
Tabela 4 - Classificação da declividade segundo a EMBRAPA (1979).....	64
Tabela 5 - Variáveis utilizadas no estudo para cada uma das 100 bacias hidrográficas.....	74
Tabela 6 - Matriz de correlação de Spearman entre as variáveis independentes analisadas	76
Tabela 7 - Características estatísticas das regiões determinadas usando o algoritmo de ligação completa de agrupamento combinado com a distância Euclidiana e Manhattan	88
Tabela 8 - Características estatísticas das regiões determinadas usando o algoritmo de Ward de agrupamento combinado com a distância Mahalanobis	90
Tabela 9 - Características estatísticas das regiões determinadas usando o algoritmo de Ward de agrupamento combinado com a distância Euclidiana e Manhattan	91
Tabela 10 - Características estatísticas das regiões determinadas pelo método k-means com k=2	93
Tabela 11 - Características estatísticas das regiões determinadas usando o algoritmo de Ward de agrupamento combinado com a distância Euclidiana e Manhattan	95
Tabela 12 - Características estatísticas das regiões determinadas pelo método k-means com k=3	97
Tabela 13 - Características estatísticas das regiões determinadas usando o algoritmo de Ward de agrupamento combinado com a distância Euclidiana e Manhattan	100
Tabela 14 - Características estatísticas das regiões determinadas pelo método k-means com k=4	101
Tabela 15 - Características estatísticas das regiões determinadas usando o algoritmo de Ward de agrupamento combinado com a distância Euclidiana	105
Tabela 16 - Características estatísticas das regiões determinadas usando o algoritmo de Ward de agrupamento combinado com a distância Manhattan.....	107

Tabela 17 - Características estatísticas das regiões determinadas pelo método k-means com k=5	109
Tabela 18 - Características estatísticas das regiões determinadas pelo método k-means com k=6	112
Tabela 19 - Características estatísticas das regiões determinadas usando o algoritmo de Ward de agrupamento combinado com a distância Euclidiana e Manhattan	115
Tabela 20 - Características estatísticas das regiões determinadas usando o algoritmo de ligação completa de agrupamento combinado com a distância Euclidiana e Manhattan	117
Tabela 21 - Características estatísticas das regiões determinadas usando o algoritmo de Ward de agrupamento combinado com a distância Mahalanobis	119
Tabela 22 - Características estatísticas das regiões determinadas usando o algoritmo de Ward de agrupamento combinado com a distância Euclidiana e Manhattan	120
Tabela 23 - Características estatísticas das regiões determinadas pelo método k-means com k=2	121
Tabela 24 - Características estatísticas das regiões determinadas usando o algoritmo de Ward de agrupamento combinado com a distância Euclidiana e Manhattan	122
Tabela 25 - Características estatísticas das regiões determinadas pelo método k-means com k=3	124
Tabela 26 - Características estatísticas das regiões determinadas usando o algoritmo de Ward de agrupamento combinado com a distância Euclidiana e Manhattan	126
Tabela 27 - Características estatísticas das regiões determinadas pelo método k-means com k=4	127
Tabela 28 - Características estatísticas das regiões determinadas usando o algoritmo Ward de agrupamento combinado com a distância Euclidiana e Manhattan	129
Tabela 29 - Características estatísticas das regiões determinadas pelo método k-means com k=6	132
Tabela 30 - Características estatísticas das regiões homogêneas determinadas usando o algoritmo Ward de agrupamento combinado com a distância Euclidiana e Manhattan	135
Tabela 31 - Agrupamento das estações fluviométricas em regiões hidrologicamente homogêneas para as vazões Q_{90} e Q_{95}	138

Lista de Abreviaturas e Siglas

ANA	Agência Nacional de Águas
ASI	Agenzia Spaziale Italiana
ASTER	Advanced Spaceborne Emission and Reflection Radiometer
CNRH	Conselho Nacional dos Recursos Hídricos
DIOUT	Divisão de Outorga e Fiscalização
DLR	Deutschen Zentrum für Luft - und Raumfahrt
DRH	Departamento de Recursos Hídricos
EMBRAPA	Empresa Brasileira de Pesquisa Agropecuária
EP	Erro Padrão de Estimativa
ESRI	Environmental Systems Research Institute
FEPAM	Fundação de Proteção Ambiental Henrique Luis Hoesler
H	Medida de Heterogeneidade
IBGE	Instituto Brasileiro de Geografia e Estatística
INPE	Instituto Nacional de Pesquisas Espaciais
InSAR	Interferometria de radar de abertura sintética
IVP	Infravermelho próximo
MAE	Erro Médio Absoluto
MDE	Modelo Digital de Elevação
MDEHC	Modelo Digital de Elevação Hidrograficamente Consistente
NASA	National Aeronautics and Space Administration
PBH	Plano de Bacia Hidrográfica
PNRH	Política Nacional dos Recursos Hídricos
RMSE	Raiz Quadrada do Erro Médio
SEMA	Secretaria Estadual do Ambiente e Desenvolvimento Sustentável
SHYDA	System of Hydrological Data Acquisition and Analysis
SIG	Sistema de Informação Geográfica
SRTM	Shuttle Radar Topographic Mission
UTM	Universal Transversa de Mercator
VIS	Imagens no Espectro Visível

Lista de Símbolos

ΣL	Comprimento total dos cursos d'água
A	Área de drenagem da bacia
Cfa	Clima subtropical de verão quente
Cfb	Clima subtropical de verão úmido
D	Declividade média da bacia
H ₀	Hipótese nula
L	Maior distância percorrida pela água
L _p	Comprimento do rio principal
P	Perímetro
p	Precipitação total anual
P _{eq}	Vazão equivalente ao volume precipitado
P _{sc}	Precipitação total do semestre chuvoso
P _{ss}	Precipitação total do semestre seco
Q ₅₀	Vazão de permanência em 50% do tempo
Q _{7, 10}	Vazão mínima de 7 dias de duração com 10 anos de recorrência
Q ₇₅	Vazão de permanência em 75% do tempo
Q ₉₀	Vazão de permanência em 90% do tempo
Q ₉₅	Vazão de permanência em 95% do tempo
R ² _a	Coeficiente de determinação ajustado
R ²	Coeficiente de determinação

Sumário

1	INTRODUÇÃO	19
1.1.	Objetivo geral	21
1.2.	Objetivos específicos.....	21
1.3.	Hipóteses.....	21
2	REVISÃO BIBLIOGRÁFICA	23
2.1.	Vazão mínima de referência.....	23
2.2.	Outorga de uso de recursos hídricos.....	25
2.3.	Rede de monitoramento	29
2.4.	Regionalização de vazões.....	30
2.5.	Identificação de regiões homogêneas	32
2.6.	Modelos digitais de elevação.....	35
2.7.	Estatística aplicada ao estudo de regionalização de vazões.....	39
2.8.	Análise multivariada	40
2.8.1	Métodos de agrupamento.....	42
2.8.2	Índices de validação de agrupamentos	48
2.9	Regressão múltipla.....	49
2.9.1	Modelos de regressão	50
2.9.2	Verificação dos modelos de regionalização.....	52
3	METODOLOGIA.....	57
3.1.	Área de estudo	57
3.2.	Seleção e processamento de dados	58
3.3.	Modelo digital de elevação e extração de parâmetros	62
3.4.	Análise estatística multivariada	65
3.5.	Medida de heterogeneidade	67
3.6.	Equações de regionalização.....	69
4	RESULTADOS E DISCUSSÃO	71
4.1.	Análise preliminar dos dados levantados	71
4.2.	Análise de agrupamento.....	77
4.3.	Regiões homogêneas para a Q_{90}	85
4.4.	Regiões homogêneas para a Q_{95}	117
5	CONCLUSÃO	139

REFERÊNCIAS BIBLIOGRÁFICAS140

1 INTRODUÇÃO

A contínua demanda de água para atender os diversos usos tem contribuído para o aumento de seu consumo, instalando-se, assim, o conflito entre usuários em diferentes regiões do Brasil. A maioria dos conflitos pelo uso da água decorre da falta de planejamento e gestão de recursos hídricos, os quais estão intimamente ligados à inexistência de informações que associem as vazões já outorgadas com a atual disponibilidade hídrica.

A Lei nº 9.443 de 1997, que instituiu a Política Nacional dos Recursos Hídricos (PNRH) é considerada a etapa inicial para o gerenciamento de tais recursos no Brasil (BRASIL, 1997), garantindo a proteção das águas contra as diversas formas de poluição e de uso impróprio. A PNRH, em seu art. 5º, apresenta os seis instrumentos de gestão, dentre eles, a outorga dos direitos de uso dos recursos hídricos, a qual aborda as questões relacionadas à gestão e aos múltiplos usos das águas.

Para dar suporte a esse instrumento de gestão, o conhecimento da disponibilidade hídrica e da respectiva demanda dos usuários torna-se uma prioridade ao planejamento. A disponibilidade hídrica em uma determinada bacia hidrográfica pode ser representada pelas vazões médias e mínimas, sendo elas responsáveis pela preservação das condições naturais do sistema. As vazões mínimas necessitam de especial atenção, pois é nos períodos de sua ocorrência que a disponibilidade de água é considerada crítica para atender todas as suas demandas.

As vazões mínimas Q_{90} e Q_{95} representam a vazão de permanência em 90% e 95% do tempo, respectivamente, as quais servem de referência para o processo de outorga de direito de uso dos recursos hídricos no estado do Rio Grande do Sul (SEMA, 2016). A quantificação dessas vazões mínimas é possível por meio de séries históricas de vazões, originárias de estações fluviométricas. Contudo, tais estações contemplam apenas algumas seções de controle específicas ao longo da hidrografia, dificultando, dessa forma, a tomada de decisão em seções não monitoradas.

No Rio Grande do Sul, a atual distribuição espacial da rede hidrometeorológica, a disponibilidade limitada de séries de dados consistentes, bem como a grande variabilidade das características físicas e hidrológicas das bacias hidrográficas, constituem em um grande desafio na estimativa das vazões de referência, para os diferentes corpos hídricos, de forma a subsidiar à gestão dos recursos hídricos.

Com a finalidade de contornar esta dificuldade, a regionalização de vazões tem sido uma técnica amplamente empregada nacional e internacionalmente. A regionalização de vazões é conceituada como um processo de transferência de informações hidrológicas, dentro de uma determinada região considerada hidrologicamente homogênea, com o objetivo de suprir a carência de dados em locais com pouca ou nenhuma informação (OUDIN *et al.*, 2010). Contudo, a determinação das regiões homogêneas é apontada por Gonçalves *et al.* (2016) como a etapa mais subjetiva e trabalhosa em virtude das várias metodologias existentes na literatura para delimitá-las.

De acordo com Lemos; Magalhães (2015), a regionalização é uma metodologia relevante para a gestão de recursos hídricos, uma vez que busca reproduzir e prever os comportamentos hidrológicos de uma realidade complexa; a técnica serve de referencial para a tomada de decisão dos órgãos de administração e planejamento, mas nunca poderá reproduzir a realidade como um todo. A técnica aplicada é sempre uma simplificação da realidade, a qual é muito mais complexa e dinâmica (LEMOS; MAGALHÃES, 2015).

Entretanto, a possibilidade do emprego de diferentes metodologias pode ocasionar estimativas desiguais, resultando em diferentes configurações e equações de regionalização. Consequentemente, produz variações nas estimativas das vazões, fazendo com que o estudo do método de identificação de regiões homogêneas e dos parâmetros de caracterização da região de interesse e a sua decorrente influência, sejam de fundamental importância para uma representação adequada da variação, estatística e física, das vazões ao longo da hidrografia na bacia.

Após a compilação das características da região de interesse faz-se necessário a análise e a interpretação, de forma adequada, de tais dados para a extração das informações, principalmente para dar suporte às decisões de forma a possibilitar um gerenciamento mais eficiente. A estatística multivariada tem o propósito de simplificar ou facilitar a interpretação dos fenômenos em estudo, sendo considerada uma ferramenta que permite que diversas variáveis sejam testadas simultaneamente, mesmo quando não se conhece o modelo teórico das relações entre as mesmas (BAKKE; LEITE; SILVA, 2008).

A análise estatística multivariada de agrupamento (cluster) consiste numa ferramenta exploratória dos dados com intuito de classificar grupos homogêneos (WILKS, 2006), tornando-se uma ferramenta aplicável para a determinação das

regiões homogêneas, uma vez que obtém relações entre as características pré-estabelecidas das bacias que contemplam a região em análise.

1.1. Objetivo geral

Avaliar técnicas de agrupamento com o propósito de fornecer uma ferramenta robusta para a regionalização das vazões mínimas Q_{90} e Q_{95} , de referência para os processos de outorga, no estado do Rio Grande do Sul.

1.2. Objetivos específicos

- Avaliar a aplicabilidade da análise de agrupamento, métodos hierárquicos e não hierárquicos, na delimitação de regiões hidrologicamente homogêneas para regionalização de vazões mínimas no Estado.
- Avaliar as regiões homogêneas pela medida de heterogeneidade.
- Identificar as variáveis fisiográficas e climáticas das bacias hidrográficas que exercem mais influência nas regiões homogêneas e, consequentemente, na estimativa das vazões mínimas no Estado.
- Determinar por meio de regressões múltiplas, uma equação de regionalização para cada região considerada hidrologicamente homogênea.

1.3. Hipóteses

- O estado do Rio Grande do Sul tem que ser dividido em diferentes regiões homogêneas na regionalização de vazões, para fins de gestão relacionada a processos de outorga de direito de uso da água.
- As regiões hidrologicamente homogêneas, delimitadas por meio da análise de agrupamento hierárquico e não hierárquico, diferem estatisticamente entre si, o que demandaria a obtenção de equações de regionalização de vazões para cada região no estado do Rio Grande do Sul.

- As características fisiográficas e climáticas são variáveis preditoras adequadas para a modelagem de vazões mínimas no estado do Rio Grande do Sul.
- Entre as distâncias de dissimilaridade a que proporciona os melhores resultados é a distância Euclidiana.
- Entre os algoritmo de agrupamentos o que proporciona a melhor separação entre os grupos é o algoritmo de Ward.

2 REVISÃO BIBLIOGRÁFICA

2.1. Vazão mínima de referência

O conhecimento das características hidrológicas de uma bacia hidrográfica é a premissa para a conservação dos recursos naturais e para a aplicação de técnicas adequadas para o desenvolvimento sustentável (PEREIRA *et al.*, 2016). Por serem limitados, vulneráveis e essenciais à vida e às atividades humanas, os recursos hídricos são alvo de conflitos e disputas para a garantia de seu uso (GAO *et al.*, 2017; LEMOS; MAGALHÃES, 2015).

O termo vazão significa o volume de água ou afluente que passa, por unidade de tempo, em uma determinada seção transversal de corpo hídrico, sendo uma informação necessária para a realização de estudos voltados à gestão de recursos hídricos (CUPAK, 2017). Neste contexto, é de suma importância quantificar as vazões que servem de referência para os processos de outorga. Segundo Wolff; Duarte; Mingoti (2014), isto é possível com a coleta de dados de estações fluviométricas, localizadas em seções específicas ao longo da hidrografia.

A vazão outorgável é a vazão disponível após estabelecida a vazão de referência, conforme a Resolução CONAMA nº 357/05 (BRASIL, 2005), tendo em vista o uso múltiplo das águas. As vazões de referência utilizadas, segundo Santos; Cunha (2013), são as vazões mínimas, de forma a caracterizar uma condição de garantia de água no manancial. A partir dessa condição, são realizados os cálculos de alocação da água, de modo que, quando essas vazões mínimas ocorram, os usuários e/ou os usos prioritários mantenham, de certa forma, seus consumos (ANA, 2013).

Segundo Almeida; Curi (2016), a vazão de referência é considerada a vazão máxima outorgável sazonalmente. Os autores ainda destacam que a vantagem de usar o critério da vazão de referência é que obtêm-se maiores garantias de que não ocorrerão falhas, pois a vazão de referência mais utilizada é uma vazão mínima que caracteriza uma condição de escassez hídrica no manancial. Santos; Cunha (2013) destacam que essa vazão de referência, baseada apenas em dados hidrológicos, é suficientemente adequada para a proteção dos rios, já que as alocações de água são geralmente realizadas a partir de uma vazão de base de pequeno risco.

Para Santos; Cunha (2013), uma desvantagem é que cada Estado regulamenta a vazão de referência individualmente, utilizando sempre os mesmos critérios hidrológicos e estatísticos, desprezando variáveis referentes às características sociais, econômicas e ecossistêmicas da utilização da água. Contudo, Silva B.; Silva D.; Moreira (2015) relatam que a análise quantitativa sazonal da disponibilidade hídrica permite elevar a possibilidade de exploração da água. No entanto, a fim de garantir os usos múltiplos da água de forma racional, é necessário contemplar a análise qualitativa da água, pois certos efluentes também possuem características sazonais.

As vazões mínimas aplicadas como referência são vazões de elevada permanência no tempo, calculadas de forma estatística a partir de dados históricos de estações fluviométricas, sendo que as mais empregadas são as vazões Q_{90} , Q_{95} e $Q_{7,10}$ (ALMEIDA; CURI, 2016).

A determinação das vazões de referência para um curso d'água faz-se importante já que possuem como objetivo oferecer uma base técnica para garantir os usos múltiplos e proteger os corpos hídricos, impedindo que os volumes outorgados venham a comprometer as condições necessárias à manutenção dos ecossistemas terrestres e aquáticos (GRANZIERA, 2013; LEMOS; MAGALHÃES, 2015). A definição da vazão de referência a ser aplicada depende da garantia de atendimento que se deseja considerar para os usos a serem instalados em determinada bacia. Se os usos exigirem maior garantia, deve-se optar por vazões mais conservadoras, como a Q_{95} e a $Q_{7,10}$ (PEREIRA *et al.*, 2016).

Após a definição das vazões de referência, deve ser determinado o percentual máximo a ser alocado para a divisão entre os diversos usos da bacia. A determinação desse percentual deve ser realizada em função da possibilidade de atendimento, aos diversos usos na bacia, e das vazões mínimas remanescentes que se deseja manter nos cursos d'água para sustentar a vida no ambiente (ANA, 2013). Dessa forma, quando o poder público analisa uma solicitação de outorga de um dado usuário, considerando uma captação a fio d'água em cursos d'água superficiais, ele deve considerar a vazão solicitada para o empreendimento frente ao percentual definido como outorgável, em relação à vazão de referência adotada.

A relação entre a magnitude e a frequência de vazões diárias, semanais, mensais, anuais (ou de qualquer outro intervalo de tempo) de uma determinada área de drenagem, é representada pela curva de permanência, a qual fornece a

porcentagem de tempo em que um valor é igualado ou excedido para um determinado período (VOGEL; FENESSEY, 1994).

A Q_{90} é a vazão determinada a partir das observações em um posto fluviométrico, em certo período de tempo, em que 90% daquele período, as vazões foram iguais ou superiores a ela. A Q_{95} tem o mesmo significado que a Q_{90} , entretanto a garantia corresponde a 95% do tempo de observação. A $Q_{7,10}$ é a menor vazão média consecutiva de sete dias de duração que ocorreria com um período de retorno de 10 anos. O cálculo da $Q_{7,10}$ é probabilístico, enquanto os da Q_{90} e da Q_{95} decorrem de uma análise de frequência.

A curva de permanência é uma metodologia importante para o estabelecimento da vazão de referência, visando manter o equilíbrio do ecossistema aquático e estabelecendo os possíveis volumes outorgáveis para os demais usos. As vazões de referência adotadas para a outorga de uso da água por parte dos órgãos gestores de recursos hídricos, na maioria dos estados brasileiros, são valores geralmente oriundos de vazões representativas da curva de permanência de vazões, tais como Q_{50} , Q_{90} e Q_{95} (MARÇAL; SILVA, 2017).

Para Detzel; Fernandes; Mine (2016), a curva de permanência é uma ferramenta amplamente difundida entre hidrólogos, principalmente por concentrar um grande número de informações sobre determinado rio em apenas um gráfico. A curva de permanência de um curso d'água é uma função hidrológica, que possui ampla variedade de utilizações na engenharia de recursos hídricos, dentre as quais podem ser citados estudos de disponibilidade hídrica para abastecimento público, irrigação, navegabilidade, análise de qualidade de águas, estudos hidrelétricos, análise de frequência de vazões ecológicas, entre outros.

2.2. Outorga de uso de recursos hídricos

O conhecimento da disponibilidade de água é a informação básica para a tomada de decisão referente ao planejamento e a gestão de recursos hídricos (MOREIRA; SILVA, 2014). Segundo Lisboa *et al.* (2014), a probabilidade de conflitos pelo uso da água é maior à medida que a demanda pelos recursos hídricos aumenta, em relação à disponibilidade hídrica.

Nessa perspectiva, a PNRH tem o objetivo de ajustar as demandas econômicas, sociais e ambientais por água em níveis sustentáveis de modo a permitir,

sem conflitos, a convivência de seus usos atuais e futuros (CUPAK, 2017). Tal finalidade também é assegurada em nível estadual pela Lei nº 10.350 de 1994, a qual instituiu o Sistema Estadual de Recursos Hídricos no estado do Rio Grande do Sul (RIO GRANDE DO SUL, 1994).

De acordo com a Lei Estadual nº10.350 de 1994 foi determinada a existência de três regiões hidrográficas, cujas bacias foram agrupadas para fins de gerenciamento. As regiões hidrográficas correspondem à região do rio Uruguai, incluindo as áreas de drenagem do rio Uruguai e do rio Negro; a região do Guaíba, compreendendo as áreas de drenagem do Guaíba; e a região do Litoral, abrangendo as áreas de drenagem dos corpos de água não incluídos nas demais regiões. A lei mencionada também estabeleceu que para cada bacia hidrográfica fosse instituído um Comitê de Gerenciamento, ou seja, o Comitê de Bacia, o qual determinaria o plano de bacia hidrográfica, com a referida vazão de referência e suas porcentagens outorgáveis, como observado na Tabela 1 (RIO GRANDE DO SUL, 1994).

Tabela 1 - Comitês das bacias hidrográficas do estado do Rio Grande do Sul

Região Hidrográfica	Comitê de Bacia Hidrográfica	Situação do plano	Vazão de referência
Litoral	Litoral médio	não foi elaborado	-
	rio Camaquã	aprovado	Q ₉₀
	Lagoa Mirim e do Canal São Gonçalo	não foi elaborado	-
	rio Mampituba	não foi elaborado	-
Guaíba	rio Tramandaí	em elaboração	-
	rios Vacacaí e Vacacaí-Mirim	não foi elaborado	-
	rio Baixo Jacuí	em elaboração	Q ₉₀
	rio Alto Jacuí	aprovado	Q ₉₀
	rio Pardo	em elaboração	Q ₉₀
	rio Caí	aprovado	Q ₉₀
	rio dos Sinos	aprovado	Q ₉₀
	Lago Guaíba	em elaboração	Q ₉₀
	rio Gravataí	aprovado	Q ₈₅
Uruguai	rios Taquari-Antas	aprovado	Q ₉₅
	Passo Fundo	aprovado	Q ₉₀
	rios Turvo-Santa Rosa- Santo Cristo	aprovado	Q ₉₀
	rio Ijuí	aprovado	Q ₉₀
	Santa Maria	aprovado	Q ₉₀
	rio Ibicuí	aprovado	Q ₉₀
	rio Apuaê-Inhandava	em elaboração	-
	Butuí-Icamaquã	não foi elaborado	-
	Piratinim	não foi elaborado	-
	Várzea	não foi elaborado	-
	rio Negro	não foi elaborado	-
	rio Quaraí	em elaboração	-

Fonte: SEMA (2017).

Na Tabela 1 estão apresentados os comitês das bacias do Estado, juntamente com a situação dos planos. Observa-se que apenas alguns foram aprovados pelo CNRH e outros ainda aguardam avaliação, ou estão em processo de elaboração. Independentemente da aprovação, alguns comitês já estão com os planos bem encaminhados e já definiram as suas vazões de referência.

Desta forma, a PNRH, por meio de seus instrumentos, quando efetivamente implementados, assegura a gestão integrada dos aspectos quantitativos e qualitativos da água, de forma a garantir o atendimento às demandas com a preservação do meio ambiente e promoção do desenvolvimento socioeconômico de uma região (ALMEIDA; CURTI, 2016). Dentre os instrumentos de gestão, está a outorga de direito de uso da água, em que o poder público dispõe para autorizar, conceder ou permitir aos usuários a utilização desse bem público, assegurando o controle e o efetivo exercício dos direitos de acesso aos recursos hídricos (BRASIL, 1997).

De acordo com Pruski *et al.* (2015), a outorga é um ato administrativo no qual o poder público promove a harmonização entre os múltiplos usos, garantindo a todos os usuários o acesso aos recursos hídricos, conforme a disponibilidade em cada bacia hidrográfica. Na emissão da licença para o outorgado (requerente), que assegura o aproveitamento deste recurso, alguns critérios como a quantidade de água, o prazo e o tipo de uso são definidos previamente. Esses critérios têm como finalidade ajustar demandas e disponibilidades para garantir o atendimento aos usos atuais e futuros da água (OLIVEIRA; PRUSKI; NUNES, 2013). Nessa perspectiva, o uso de ferramentas para apoiar a tomada de decisões, que forneçam informações tão detalhadas quanto possível sobre a disponibilidade de água, torna-se essencial para o desenvolvimento de estratégias de planejamento e gerenciamento de recursos hídricos (PRUSKI *et al.*, 2015).

A Lei Federal nº 9.984 de 2000 designa a Agência Nacional de Águas (ANA) como responsável pela emissão de outorgas de direito de uso de recursos hídricos em corpos hídricos de domínio da União (BRASIL, 2000). Em relação aos recursos hídricos de domínio dos Estados, existem órgãos específicos com competência legal para emitir as outorgas.

No estado do Rio Grande do Sul, o órgão de administração que tem competência de emissão de outorga é o Departamento de Recursos Hídricos (DRH), vinculado à Secretaria Estadual do Ambiente e Desenvolvimento Sustentável (SEMA), para os usos que alteram as condições quantitativas das águas. Para os usos que

alteram as condições qualitativas das águas, o órgão responsável pela emissão da outorga é a Fundação de Proteção Ambiental Henrique Luis Hoesler (FEPAM), determinado pelo decreto nº 37.033/1996 (RIO GRANDE DO SUL, 1996). Atualmente, a coordenação do procedimento de emissão da outorga de direito de uso da água, que é realizada pelo DRH, criou a Divisão de Outorga e Fiscalização (DIOUT) para melhor gerenciamento do processo.

O Art. 7º do decreto nº 37.033/1996 estabelece que os valores referentes aos parâmetros técnicos necessários para nortear os processos de outorgas deverão ser estabelecidos pelos respectivos Planos de Bacias Hidrográficas (PBH), cuja responsabilidade pela sua elaboração é dos comitês de bacias, os quais posteriormente devem ser aprovados pelo Conselho Nacional de Recursos Hídricos (CNRH). Entretanto, quando não estiver definido o plano de uma determinada bacia, cabe ao DRH o estabelecimento da vazão de referência para os processos de outorga.

O modelo de gestão brasileiro e de seus estados federados, baseado na lei 9.433/1997, define a bacia hidrográfica como unidade territorial de gestão, e os comitês de gerenciamento de bacia como instância responsável pela deliberação com a sociedade e aplicação dos instrumentos propostos na implementação da PNRH. Lemos; Magalhães (2015) discutem que mesmo contemplados na PNRH, a aplicação dos instrumentos de gestão apresentam vários desafios. Segundo os autores, para a aproximação da realidade e do contexto local de cada bacia hidrográfica, o cálculo de valores possíveis de serem autorizados para os diferentes usos tem regulação específica a ser definida por cada Estado da Federação, de acordo com as políticas estaduais de recursos hídricos e, mais especificamente, pelos planos diretores de recursos hídricos das bacias.

Os PBHs de cada unidade de planejamento no Rio Grande do Sul empregam diferentes porcentagens da vazão de referência para a concessão das outorgas, alternando entre a Q_{90} e a Q_{95} . Além do Rio Grande do Sul, os Estados da região nordeste também empregam as vazões Q_{90} e a Q_{95} como vazão de referência nos processos de outorga (WOLFF; DUARTE; MINGOTI, 2014). É importante destacar que, para os cursos d'água federais, a ANA impõe um limite para outorga de 70% da Q_{95} . Já os estados de São Paulo, Espírito Santo e Minas Gerais utilizam a $Q_{7,10}$ como vazão de referência (ANA, 2007). De certa forma, o Brasil apresenta uma tendência regional para adoção das vazões de referência.

Internacionalmente também é possível notar a utilização dessas vazões de referência. De acordo com Pushpalatha *et al.* (2012), a França utiliza a Q_{90} na avaliação dos critérios de eficiência para vazões mínimas. Na Cracóvia se utiliza a vazão mínima de referência Q_{95} (GARCIA; FOLTON; OUDIN, 2017). Cupak (2017) destaca que a Q_{95} é comumente empregada na Europa como vazão de referência para a gestão dos recursos hídricos.

2.3. Rede de monitoramento

As séries temporais são fontes valiosas de informações que podem ser consultadas para caracterização de variáveis nas mais diversas áreas do conhecimento. Em hidrologia, tais séries são de fundamental importância para o entendimento do ciclo hidrológico, sendo empregadas na gestão de sistemas hídricos (DETZEL; FERNANDES; MINE, 2016).

Os processos de monitoramento de águas fluviais e pluviais, em quantidade ou qualidade, exigem um aparato de gestão e uma infraestrutura que geralmente não estão disponíveis (MOLINA *et al.*, 2014). A rede hidrometeorológica nacional, administrada pela ANA, é a principal fornecedora de dados hidrológicos no país. Mesmo extensa e com ampliações nos últimos anos, a rede de monitoramento hídrico no Brasil, principalmente no que se refere às vazões fluviais, está defasada quanto à quantidade de estações e quanto à eficiência do monitoramento (BRESSIANI *et al.*, 2015). Estações com medições automáticas de vazão são encontradas em menor número e poucos cursos de água possuem medições sistemáticas (LEMOS; MAGALHÃES, 2015).

No Brasil, os cursos de água monitorados são aqueles com extensas áreas de drenagem e maiores vazões, já que são muito demandados para geração de energia hidroelétrica, irrigação ou outros usos mais tradicionais e estratégicos no país, não fornecendo informações hidrológicas capazes de subsidiar uma adequada gestão dos recursos hídricos como previsto na Lei nº 9.433 de 1997. Entretanto, para Pruski *et al.* (2012) os cursos de água de menores dimensões e áreas de drenagem são justamente os mais susceptíveis a impactos a partir das atividades humanas, e não costumam apresentar estações de monitoramento, concordando com Beskow *et al.* (2014) que destacam a carência de dados de vazão, especialmente em pequenas bacias hidrográficas.

Segundo Aissia; Chebana; Ouarda (2017); Costa *et al.* (2012) e Mishra; Coulibaly (2009), a rede de monitoramento não é somente caracterizada em função da sua densidade de estações, mas principalmente pela qualidade e confiabilidade dos dados. Segundo Tucci (2002), a rede de monitoramento hidrológico dificilmente cobrirá completamente a hidrografia. Além disso, podem existir estações de medições recentemente implantadas, o que indica séries temporais curtas, impossibilitando a caracterização do regime hidrológico de um determinado local (GARCÍA-MARÍN *et al.*, 2015; RAZAVI; COULIBALY, 2016). Tucci (2002) ainda destaca a importância da implementação de métodos que permitam a otimização das informações existentes.

As séries limitadas de dados fluviométricos disponíveis e a necessidade de se conhecer a vazão ao longo da rede hidrográfica dificultam ou, muitas vezes impedem, a realização de uma adequada gestão de recursos hídricos (MOREIRA; SILVA, 2014).

Visando minimizar o problema da carência de monitoramento fluviométrico no Brasil e em vários outros países, algumas metodologias têm sido desenvolvidas no intuito de oportunizar a estimativa de variáveis hidrológicas, como a vazão de estiagem associada a uma permanência, em especial em bacias hidrográficas não monitoradas (BESKOW *et al.*, 2014).

Com as deficiências nos processos de monitoramento hidrológico, gestores e pesquisadores adotam, geralmente, modelos hidrológicos para a estimativa da disponibilidade de água, a partir da técnica de regionalização, a qual permite estimar valores para áreas sem dados (LEMO; MAGALHÃES, 2015).

2.4. Regionalização de vazões

A regionalização é uma técnica que permite transferir informações entre bacias hidrográficas semelhantes, a fim de estimar, em sítios que não dispõem de dados, as variáveis fluviométricas de interesse (WOLFF; DUARTE; MINGOTI, 2014). Segundo Pruski *et al.* (2012), essa técnica relaciona os processos hidrológicos com características físicas e climáticas de uma bacia. De acordo com Costa *et al.* (2012), a regionalização consiste em explorar ao máximo as informações existentes, permitindo a estimativa das variáveis hidrológicas em locais sem dados ou com dados insuficientes.

O conhecimento das características de uma bacia hidrográfica é de extrema importância para descrever os processos e compreender o seu funcionamento

(PEREIRA *et al.*, 2016). Particularidades físicas, químicas, climáticas, morfométricas e hidrológicas são utilizadas na tentativa de identificar regiões semelhantes/homogêneas. Segundo Singh (2016), a similaridade entre duas bacias hidrográficas é um problema multivariável complexo. Melati (2016) pressupõe que a semelhança espacial das regiões permite que a vazão de regiões sem informações possa ser considerada através da relação direta com variáveis fisiográficas ou climáticas da bacia. O levantamento de tais características demanda esforços e recursos que são necessários para o planejamento de recursos hídricos em uma bacia. Quando essa realidade não é satisfeita, utiliza-se variáveis que possam descrever e explicar ao máximo os fenômenos que ocorrem.

A área de drenagem é a variável fisiográfica mais utilizada para representar as características de vazões mínimas em diferentes estudos de regionalização (BAZZO *et al.*, 2017; BESKOW *et al.*, 2016; CHAVES *et al.*, 2002; NOVAES *et al.*, 2009; LISBOA *et al.*, 2014; MOREIRA; SILVA, 2014), principalmente por ser facilmente obtida, possibilitando o emprego nas equações geradas.

Pruski *et al.* (2012), em estudo realizado para a bacia do rio Pará, sub-bacia do rio São Francisco, inseriram a variável climática precipitação média anual no estudo de regionalização de vazões mínimas. Para os autores, a precipitação que ocorre na área de drenagem de uma bacia hidrográfica é um fator que interfere diretamente no comportamento das vazões, motivo pelo qual a inclusão da precipitação como variável explicativa pode representar uma melhoria expressiva do modelo de regionalização das vazões, concordando com Melati *et al.* (2016). Além disso, Pruski *et al.* (2011) e Pruski *et al.* (2013), em estudo realizado para a bacia do rio Paracatu, sub-bacia do rio São Francisco, testaram uma única variável para representar as características físicas e climáticas da bacia (área de drenagem e precipitação, respectivamente), definida como P_{eq} vazão equivalente ao volume precipitado em m^3s^{-1} . Os autores visavam contemplar a parte da precipitação que não é convertida em escoamento ao longo da hidrografia, em decorrência de outros processos, como a evapotranspiração.

Entretanto, segundo Oliveira; Pruski; Nunes (2013), por mais que a precipitação média anual seja uma variável explicativa do processo de formação das vazões mínimas e médias, considera-se que essa não reflita efetivamente a contribuição para a formação dessas vazões, pois para que haja a ocorrência do escoamento no leito do rio, advindo da contribuição subterrânea é necessário que, primeiramente, a precipitação venha suprir o déficit de água existente na zona de aeração do solo, que,

por sua vez, depende das características do solo, da cobertura vegetal e do processo de evapotranspiração. Como também observado por Mohamoud (2008), a influência da precipitação pode ser baixa, uma vez que as vazões mínimas provavelmente provêm de liberações de armazenamento de solo, águas subterrâneas ou canais. Contudo, o levantamento de tais características e a sua consequente atuação na formação de vazões é dificultado nos padrões atuais de monitoramento das bacias.

Além da difícil questão de determinar quais características de captação pré-determinam o comportamento hidrológico, existe o problema de simplesmente definir a similaridade hidrológica (OUDIN *et al.*, 2010). Na tentativa de contornar a falta de variáveis explicativas adequadas para os estudos de vazões mínimas, Uliana *et al.* (2017) inseriram características fisiográficas das bacias hidrográficas sob análise, tais como perímetro, comprimento axial e comprimento total dos cursos d'água. Já Elesbon *et al.* (2015) introduziram também a declividade média da bacia, a precipitação do semestre mais seco e do semestre mais chuvoso, na tentativa de observar alguma relação, corroborando com Srinivas (2009), que comprovem que a similaridade entre características puramente fisiográficas não implica necessariamente em semelhança na resposta hidrológica entre bacias hidrográficas.

Swain; Patra (2017) incluíram o centroide da bacia, latitude e longitude, para estimar regiões homogêneas com bacias hidrográficas próximas, e a evapotranspiração para descrever melhor as vazões mínimas. Singh *et al.* (2016) constataram que a proximidade espacial é um bom indicador para similaridade entre os possíveis grupos formados, apresentando assim um maior grau de conectividade.

2.5. Identificação de regiões homogêneas

A técnica de regionalização de vazões destina-se a delinear regiões homogêneas, onde as bacias hidrográficas são semelhantes em relação a vários atributos, tais como características físicas, climáticas e hidrológicas, a fim de transferir informações ou modelos das bacias hidrográficas monitoradas para as não monitoradas (CUPAK, 2017).

O termo região homogênea está associado a regiões que possuem similaridade hidrológica. Essa similaridade inclui fatores físicos, climáticos, biológicos, geológicos e efeitos antrópicos. Como há grande complexidade na consideração de todos esses fatores, Elesbon *et al.* (2015) e Gonçalves *et al.* (2016) conceituaram a região

homogênea como uma região na qual diferentes características climáticas e fisiográficas teriam variabilidade mínima. A similaridade, nesse caso, seria observada com respeito aos fenômenos de maior interesse no processo hidrológico em estudo. Entretanto, muitos autores consideram a identificação de regiões homogêneas como a etapa da regionalização que possui maior grau de dificuldade, por requerer muitas vezes, julgamento subjetivo (HOSKING; WALLIS, 1997; ILORME; GRIFFIS, 2013). De fato, Gonçalves *et al.* (2016) reconheceram que a delimitação de regiões homogêneas é construída com base em premissas difíceis de serem tratadas com rigor matemático.

Regiões geograficamente contíguas com base em fronteiras geográficas, políticas, administrativas ou fisiográficas foram usadas por um longo período em hidrologia para análise de frequência regional. No entanto, esta prática não garante a homogeneidade hidrológica (RAO; SRINIVAS, 2006). Consequentemente, foram desenvolvidos para a regionalização diversos métodos que consideram a semelhança entre locais, em um espaço multidimensional dos atributos relacionados às bacias hidrográficas, tais como características fisiográficas, localização geográfica e estatísticas locais da variável hidrológica (RAO; SRINIVAS, 2006).

Uma das formas de se obter resultados, quando se deseja dividir uma área de estudo em regiões homogêneas, é por meio da análise de agrupamento cuja finalidade primária é agregar objetos com base em uma medida de similaridade e nas características que eles possuem.

As técnicas de regionalização de vazões buscam uma melhor estimativa das vazões em seções que não possuem medições fluviométricas, não sendo recomendada a utilização dessa técnica em seções que possuem medições, pois não substituem as informações reais (SILVA JR. *et al.*, 2002).

Segundo Samuel; Coulibaly; Metcalfe (2011), não existe um método de regionalização adequado, ou seja, universal. A abordagem mais comum continua a de testar vários métodos de regionalização para identificar o mais adequado para a região de interesse. Arsenault; Brissette (2016) salientam que existem lacunas importantes para serem preenchidas em regiões sem monitoramento, sendo a regionalização dependente fundamentalmente da qualidade dos dados informados.

De certa forma, o conceito de regionalização está intimamente ligado com o de regiões homogêneas, isto é, aquelas que possuem comportamento hidrológico similar. Elesbon *et al.* (2015) enfatizam que a regionalização dos índices fluviométricos é baseada na premissa de que bacias com clima, geologia, topografia, vegetação e

solos similares terão respostas hidrológicas semelhantes, contudo bacias hidrográficas com grandes áreas de drenagem podem apresentar comportamentos hidrológicos distintos, ao longo da hidrografia.

Na literatura é possível encontrar inúmeros métodos para definir e delimitar uma região homogênea, tais como: regressão simples ou múltipla (LAAHA; BLÖSCHL, 2006; MELATI, 2016; PRUSKI *et al.*, 2012), interpolador geoestatístico krigagem (WOLFF; DUARTE; MINGOTI, 2014), análise multivariada (ELESBON *et al.*, 2015), entropia (RIDOLFI *et al.*, 2016; RODRIGUES *et al.*, 2016;).

Ilorme; Griffis (2013) ressaltam que a delimitação de regiões homogêneas é altamente subjetiva e dependente da similaridade de medidas e técnicas de classificação. A não concordância de uma técnica objetiva incentiva o aperfeiçoamento e o desenvolvimento de novas (NAGHETTINI; PINTO 2007).

Rajsekhar; Mishra; Singh (2013), na tentativa de diminuir a subjetividade na determinação das regiões homogêneas, aplicaram a medida de heterogeneidade, utilizando a técnica dos momentos-L, proposta inicialmente por Hosking; Wallis (1997).

Hosking; Wallis (1997) defendem a opinião de que se as séries são homogêneas e representativas da variável em análise, não é necessário o uso de períodos comuns. Samuel; Coulibaly; Metcalfe (2011) admitiram até 20% de falhas nas séries de dados diários das estações selecionadas para garantir uma melhor distribuição espacial. Vezza *et al.* (2010), em estudos relacionados com vazões mínimas, recomendam a utilização de no mínimo 10 anos de dados.

Pruski *et al.* (2012), consideram que no atual planejamento e na gestão dos recursos hídricos, o aprimoramento das técnicas de regionalização adquirem um papel importante, em virtude de melhor estimar as vazões nos cursos d'água e garantir maior confiabilidade e segurança na utilização dos modelos de regressão, principalmente nas zonas de extrapolação das equações de regionalização. Os autores ainda evidenciaram melhor desempenho estatístico, quando utilizaram um número maior de variáveis explicativas no processo de formação de vazões. Segundo os autores, a utilização de mais variáveis tende a melhorar a regionalização e diminuir o risco associado à utilização dos modelos de regressão, para além dos limites dos dados amostrais.

Samuel; Coulibaly; Metcalfe (2011) avaliaram o desempenho de seis métodos para regionalização de vazões em 93 bacias localizadas na província de Ontário,

Canadá, e concluíram que a metodologia que alia o emprego de técnicas de interpolação e extrapolação às semelhanças físicas entre bacias foi a mais apropriada para a região.

Mamum *et al.* (2010), com dados de 82 estações fluviométricas localizadas na Península da Malásia, desenvolveram e validaram um modelo que utiliza regressão multivariada na regionalização de vazões mínimas.

Algumas técnicas convencionais de regionalização são inerentemente envolvidas com a classificação da bacia hidrográfica. Por exemplo, em semelhanças físicas ou abordagens de proximidade espacial, as respostas hidrológicas são transferidas das bacias hidrográficas calibradas para não calibradas em agrupamentos de atributos físicos similares ou local similar. Outros tipos de abordagens de regionalização, como regressão linear ou redes neurais artificiais, podem ser aplicadas em grupos homogêneos de bacias hidrográficas ou as que ainda não foram classificadas (RAZAVI; COULIBALY, 2016).

Neste contexto, as análises estatísticas multivariadas podem auxiliar na realização dos estudos de regionalização hidrológica, reduzindo o tempo de processamento de dados e aumentando a confiabilidade dos resultados obtidos (ELESBON *et al.*, 2015). Em nível internacional pode-se evidenciar esta afirmação pelo desenvolvimento de inúmeros trabalhos visando à regionalização hidrológica, com base em análises estatísticas multivariadas (ASSANI *et al.*, 2011; ENGELAND; HISDAL, 2009; KAHYA; KALAYCI; PIECHOTA, 2008; MWALE *et al.*, 2011; RAZIEI, 2017; SAMUEL; COULIBALY; METCALFE, 2011).

2.6. Modelos digitais de elevação

O levantamento das variáveis explicativas a campo muitas vezes é impossibilitado por questões financeiras e práticas. Uma forma, encontrada para contornar essa realidade, tem sido a utilização de Modelos Digitais de Elevação (MDE). De acordo com Silva *et al.* (2015) esses modelos agilizam o processo, viabilizando o maior número de estudos.

Segundo Baena *et al.* (2004) e Silva *et al.* (2015), o MDE gerado deve ser fiel na representação do relevo e deve assegurar a convergência do escoamento superficial para e ao longo da drenagem mapeada, garantindo, assim, a sua consistência hidrológica. Os autores consideram que o Modelo Digital de Elevação

Hidrologicamente Consistente (MDEHC) pode ser definido como a melhor representação digital do relevo capaz de reproduzir, com a máxima exatidão, o caminho preferencial de escoamento da água superficial observado em situações reais. Dessa forma, com a aplicação desta ferramenta torna-se possível integrar e manipular base de dados, o que viabiliza trabalhos de campo e promove maior eficiência do monitoramento e fiscalização ambiental (COSTA; REIS, 2017).

Os MDEs são representações de dados topográficos passíveis de tratamento em Sistemas de Informações Geográficas (SIG) (SILVA *et al.*, 2015). Segundo Alves Sobrinho *et al.* (2010), os MDEs são representados por uma estrutura de dados, correspondente à distribuição espacial da altitude, que pode ser obtido por meio da interpolação de curvas de nível extraídas de cartas topográficas ou imagens de sensores remotos.

Por muito tempo os estudos no Brasil utilizaram as curvas de nível e pontos de altitude conhecidos para a geração de MDEs, juntamente com a rede hidrográfica, que também são obtidas das cartas topográficas. Tais informações geralmente são elaboradas pelo Instituto Brasileiro de Geografia e Estatística (IBGE). Moura *et al.* (2017) destacam que as imagens de sensores remotos estão sendo cada vez mais utilizadas para a geração de MDEs, devido à confiabilidade e facilidade de aquisição de dados, destacando-se principalmente as imagens da missão Shuttle Radar Topographic Mission (SRTM), do sensor Advanced Spaceborne Emission and Reflection Radiometer (ASTER) e do banco de dados TOPODATA.

A missão espacial SRTM foi liderada pela agência espacial americana - National Aeronautics and Space Administration (NASA) e envolveu as agências espaciais da Alemanha (Deutschen Zentrum für Luft - und Raumfahrt - DLR) e da Itália (Agenzia Spaziale Italiana - ASI), mapeando aproximadamente 80% da superfície do planeta. Essa missão utilizou a técnica de interferometria de radar de abertura sintética (InSAR), a partir da qual a altitude é determinada pela defasagem entre reflexões de um mesmo sinal de radar captado por dois receptores (RABUS *et al.*, 2003). Para tanto é necessária a utilização de dois pares de antenas, operadas nas bandas X e C, coletando dados com uma resolução espacial de 30m e 90m, respectivamente. No território brasileiro, os dados foram processados pela Empresa Brasileira de Pesquisa Agropecuária - Embrapa e estão disponibilizados segundo as articulações das folhas topográficas em escala 1:250.000, com resolução de 90m (MIRANDA, 2005).

Nesse contexto, a tomada de dados em cobertura global pela missão SRTM representa uma oferta de MDE, por áreas relativamente desprovidas de mapeamentos topográficos. Contudo, a iniciativa TOPODATA propôs que derivações básicas da informação topográfica (declividade, orientação de vertentes, curvaturas etc.), a partir de dados SRTM, estivessem prontamente à disposição da comunidade para seu uso (VALERIANO; ROSSITTI, 2010).

O banco de dados TOPODATA foi desenvolvido para fornecer planos de informação das variáveis geomorfométricas locais básicas, para uso diretamente em ambiente de SIG. Esses dados foram produzidos com padronização de processamento (VALERIANO; ALBUQUERQUE, 2010) e de especificações técnicas (VALERIANO, 2008), a partir de um refinamento unificado (VALERIANO; ROSSETTI, 2012) dos dados SRTM (RABUS *et al.*, 2003), sendo disponibilizados para a América do Sul.

O Instituto Nacional de Pesquisas Espaciais (INPE) fornece, de forma gratuita, os dados que foram reamostrados por krigagem de 90m para 30m por meio do banco TOPODATA iniciado no ano de 2008 (VALERIANO; ROSSETTI, 2010). No final de 2015, a NASA passou então a fornecer os dados com resolução espacial de 30m de forma gratuita. Desta forma, o Brasil é um dos países com mapeamento digital completo de elevação do terreno.

Já o sensor ASTER é composto por um imageador multiespectral a bordo do satélite TERRA, lançado em dezembro 1999, com cobertura de 99% da superfície do globo terrestre e resolução espacial de 30m (FARHAN; SHAIKH, 2017). Para a obtenção dos dados, o ASTER conta com dois telescópios, fazendo aquisições de imagens no espectro visível (VIS) e infravermelho próximo (IVP) (ASTER GDEM, 2011). Além de uma maior cobertura do globo terrestre, no que concerne ao MDE, o ASTER possui algumas vantagens em relação ao SRTM, visto que a construção do produto é realizada de forma direta por processo estereoscópico, sem a necessidade de interpolação de dados, e não apresenta ausência de informações. Contudo, como as imagens são óticas e não de radar, tal como ocorre no SRTM, há a desvantagem de poder haver algumas falhas de correlação, causadas por coberturas de nuvens nas imagens de entrada (ASTER GDEM, 2011).

Desta forma, é possível encontrar na literatura diversos estudos que utilizam diferentes configurações dos MDEs para a aquisição de variáveis explicativas como Beskow *et al.* (2016) e Pruski *et al.* (2012). Os autores obtiveram a área de drenagem

à montante de cada seção da hidrografia utilizando o MDE gerado a partir de cartas planialtimétricas do IBGE, e de hidrografia mapeada na escala 1:50.000. Melati (2016) utilizou o MDE SRTM com 30m de resolução espacial para calcular a variável área de drenagem. Samuel; Coulibaly; Metcalfe, (2011) utilizaram o MDE SRTM com 30m para delinear áreas de contribuição, calcular sua morfologia e determinar a localização do centro de gravidade. Farhan; Shaikh (2017) fizeram uso do MDE a partir dos dados ASTER, com 30m de resolução espacial, para demarcar 76 bacias hidrográficas no sul da Jordânia e o levantamento de 22 variáveis morfométricas.

Moura *et al.* (2017) analisaram a influência de diferentes fontes e escalas de informação do relevo na caracterização de bacias hidrográficas. Concluíram que, em geral, dados de relevo provenientes do banco de dados TOPODATA apresentam menores erros, comparativamente aos MDEs SRTM e ASTER, corroborando com Miceli *et al.* (2011), que também encontraram melhores comportamentos nos dados provenientes do banco TOPODATA.

A utilização dos MDEs possibilita a obtenção automática das características físicas das bacias hidrográficas (FERNANDES; COSTA; STUDART, 2017). As análises hidrológicas e ambientais geralmente são realizadas tomando-se como base a bacia hidrográfica, cuja delimitação é um dos primeiros e mais comuns procedimentos executados (HASS *et al.*, 2014).

Segundo Freire; Lage; Christófar (2013), a delimitação de bacias é um processo de grande importância para o melhor entendimento da dinâmica dos recursos hídricos, uma vez que as características físicas e bióticas de uma bacia exercem relevante papel nos processos do ciclo hidrológico, influenciando diretamente na quantidade de água produzida.

A área de drenagem da bacia hidrográfica consiste na área plana, inclusa entre os divisores topográficos que a limitam. É um dado fundamental para definir a potencialidade hídrica de uma bacia hidrográfica, uma vez que a multiplicação dessa área pela altura da lâmina d'água precipitada define o volume recebido pela bacia (TUCCI, 1998). Assim como a área de drenagem, o perímetro fornece subsídios para a avaliação do nível de irregularidade que a forma da bacia apresenta e constituem os elementos básicos para o cálculo das demais características físicas (ZHANG *et al.*, 2017).

Os *softwares* de SIGs também podem auxiliar no cálculo da precipitação média, em uma superfície qualquer fazendo-se necessário utilizar as observações de postos

pluviométricos dentro da respectiva superfície e/ou nas suas proximidades. Entre os vários métodos para o cálculo da precipitação média pode-se citar o método dos polígonos de Thiessen (THIESSEN, 1911).

Segundo Cho *et al.* (2017) e Olawoyin; Acheampong (2017), embora existam outras técnicas usadas para estimar a lâmina média precipitada, o método dos polígonos de Thiessen está entre os melhores, mais populares e mais direto para estimar a precipitação. O método é indicado quando não existe uma distribuição uniforme de pluviômetros na área em análise (MACÊDO *et al.*, 2013; SOUZA; SOUZA; CARDOSO, 2017).

O método dos polígonos de Thiessen tem sua origem nos diagramas de Voronoi (AURENHAMMER, 1991), assumindo que em qualquer ponto da bacia o valor da precipitação é igual à média ponderada dos postos pluviométricos mais próximos, ou seja, é possível traçar as áreas de influência das estações para caracterizar espacialmente a precipitação (SOUZA; SOUZA; CARDOSO, 2017).

Vários autores utilizaram o método dos polígonos de Thiessen para aplicações em recursos hídricos e previsões climáticas, podendo-se citar Abreu; Tonello (2017), Cabral *et al.* (2016), Cabral; Sakuragi; Silveira (2017), Caldeira *et al.* (2015), Hasson; Lucarini; Pascale (2013), Sales *et al.* (2017) e Zema *et al.* (2016).

2.7. Estatística aplicada ao estudo de regionalização de vazões

A análise de frequência visa estimar com que periodicidade média determinado evento ocorre (MODARRES, 2010). Por sua vez, a análise regional de frequência objetiva melhorar a confiabilidade da análise por meio da utilização de informações de uma região considerada homogênea, reduzindo as incertezas na estimativa de parâmetros regionais (FARSADNIA *et al.*, 2014). Os autores, destacam ainda que, a análise regional permite a regionalização de variáveis hidrológicas, ou seja, a estimativa de quantis característicos dessas variáveis, onde não existe monitoramento.

Segundo Naghettini; Pinto (2007), os principais métodos utilizados na análise regional de frequência são os que regionalizam: eventos com determinado risco; parâmetros de distribuição de probabilidades; e quantis adimensionalizados de probabilidades. Independentemente do método utilizado, além de verificar se a não estacionariedade de uma série hidrológica está relacionada com variações do clima

ou alterações antrópicas, a realização da análise de tendência é importante na maioria das aplicações de hidrologia estatística, visto que as séries hidrológicas devem ter como pré-requisito os atributos de estacionariedade e homogeneidade (NAGHETTINI; PINTO, 2007). Rajsekhar; Mishra; Singh (2013); Yang *et al.* (2010) e Zhang *et al.* (2015) ressaltam que na análise de frequência regional é importante verificar as hipóteses de estacionariedade e independência das séries históricas utilizadas.

Atualmente, o interesse em estudos de tendência dos componentes do ciclo hidrológico tem aumentado devido às controvérsias sobre as mudanças climáticas (UILIANA *et al.*, 2015). De acordo com Joseph; Falcon; Sharif (2013), a identificação de tendências sazonais nas séries de precipitação e vazão contribui para a compreensão da variabilidade climática global e é essencial para o desenvolvimento de modelos hidrológicos, previsão hidrológica e gestão dos recursos hídricos.

Desta forma, Ishak *et al.* (2013), Joseph; Falcon; Sharif (2013); Wang *et al.* (2015); Yang *et al.* (2010) e Zhang *et al.* (2014; 2015) recomendam o teste de Mann-Kendall (KENDALL, 1975; MANN, 1945) para a análise de tendência em séries de precipitação e vazão.

2.8. Análise multivariada

As análises estatísticas multivariadas podem auxiliar na realização dos estudos de regionalização hidrológica, reduzindo o tempo de processamento da base de dados e aumentando a confiabilidade dos resultados obtidos (ELESBON *et al.*, 2015).

Entre as técnicas de estatística multivariada pode-se citar a análise de agrupamento, a análise de componentes principais, a análise de correlação canônica e a análise discriminante múltipla, que são comumente empregadas como suporte para delimitar regiões homogêneas (ILORME; GRIFFIS, 2013).

A análise estatística multivariada de agrupamento consiste na aglomeração sequencial de indivíduos a grupos cada vez maiores, de acordo com algum critério, de distância ou medida de dissimilaridade (NAGHETTINI; PINTO, 2007). A análise de componentes principais tem como objetivo reduzir a dimensionalidade de um conjunto de dados, transformando-o em funções lineares não correlacionadas das variáveis originais (MODARRES, 2010). Segundo Ilorme; Griffis (2013), a correlação canônica revela associações lineares entre dois conjuntos de variáveis e a análise de

discriminante múltipla identifica as variáveis que são capazes de explicar as diferenças entre um conjunto de grupos predefinidos.

Elesbon *et al.* (2015) estudaram a regionalização de vazões mínimas na bacia hidrográfica do rio Doce, no estado de Minas Gerais. Os autores propuseram uma metodologia baseada na análise estatística multivariada de componentes principais e de agrupamento com a finalidade de identificar as variáveis mais representativas na obtenção das regiões hidrologicamente homogêneas. No estudo foram utilizadas 61 estações fluviométricas da bacia, contendo variáveis explicativas referentes às características climáticas e morfométricas das áreas de contribuição, sendo as vazões mínimas de referência $Q_{7,10}$, Q_{90} e Q_{95} as variáveis dependentes no estudo.

Os autores supracitados utilizaram inicialmente as variáveis: precipitação total anual – p ; precipitação total do semestre seco – P_{ss} e chuvoso – P_{sc} ; área de drenagem da bacia – A ; comprimento do rio principal – L_p e comprimento total dos cursos d'água da bacia – L ; e declividade média da bacia – D . A partir da matriz de correlação entre as variáveis excluiu-se a variável D , por apresentar menor representatividade. O uso da análise de componentes principais permitiu determinar que as variáveis A e P_{sc} foram as mais representativas. As quatro regiões hidrologicamente homogêneas foram obtidas pela utilização da matriz de similaridade de Mahalanobis e do método de agrupamento do vizinho mais distante.

Modarres (2010), na determinação das regiões homogêneas no Irã, aplicou a metodologia proposta inicialmente por Hosking; Wallis (1997), a dos momentos-L. O autor obteve apenas uma região homogênea. Contudo, aplicou a técnica de análise multivariada baseada em componentes principais e agrupamento para encontrar menores regiões homogêneas, a fim de melhorar a gestão dos recursos hídricos. Lujano *et al.* (2017) determinaram 3 regiões homogêneas para a Q_{75} (vazão de permanência em 75% do tempo), na bacia do rio Titicaca no Peru, usando o método de Ward e a distância o coeficiente de correlação, sendo validadas posteriormente pela metodologia dos momentos-L.

Para o desenvolvimento das técnicas de estatística multivariada é fundamental o conhecimento das variáveis em análise, se são fortemente ou fracamente correlacionadas. De acordo com Farhan; Shaikh (2017), a análise de correlação quando positiva entre as variáveis implica que essas são interdependentes.

Desta forma, primeiramente é necessário descobrir se as variáveis seguem uma distribuição normal, caracterizando a correlação de Pearson, ou quando não

atendem os critérios, indicando correlação de Spearman (HAIR JR. *et al.*, 2009). Segundo Mello; Silva (2013), em análises hidrológicas, os testes de aderência mais comuns são o Qui-Quadrado (χ^2), Kolmogorov-Smirnov (KS) e Anderson-Darling (AD).

2.8.1 Métodos de agrupamento

A análise de agrupamento ou “*cluster analysis*” é um conjunto de técnicas que tem por objetivo identificar padrões ao formar grupos homogêneos (HSU; HUANG, 2017), a partir de n observações ou características existentes (SINGH *et al.*, 2016). A construção dos grupos é realizada de modo que as observações de um mesmo grupo assemelham-se mais entre si, do que com as observações existentes nos outros grupos formados. Singh *et al.* (2016) destacam que o agrupamento com base em respostas hidrológicas ou características de drenagem é o primeiro passo para a regionalização das vazões.

A técnica vem sendo utilizada em importantes áreas científicas para a identificação de padrões de comportamento nos dados analisados (SINGH *et al.*, 2016), auxiliando no processo de busca de relações, pois a divisão em grupos ou classes facilita a compreensão das observações e o desenvolvimento subsequente de teorias científicas. A técnica de agrupamento segundo Kumar; Singh; Mishra (2013), busca as relações entre um conjunto de variáveis, uma vez que não faz distinção entre variáveis dependentes e independentes.

Em análise de agrupamento, para construir um simples grupo, a partir de um conjunto de elementos, é necessário utilizar algum critério de proximidade ou tipo de medida que possibilite a comparação entre os componentes desse conjunto, tornando possível verificar se um dado elemento X é mais parecido com Y, do que com W (MURTAGH; CONTRERAS, 2012). A utilização dessa medida ou critério fornece a distância dimensional entre os objetos, permitindo que se quantifique o quanto eles são parecidos. Essa medida é conhecida por coeficiente de parecença ou coeficiente de similaridade (ou dissimilaridade) (HAIR JR. *et al.*, 2009), sendo essa última definição a adotada na presente dissertação.

Quando o maior valor calculado, relativo a um coeficiente, for utilizado para indicar que dois objetos são parecidos, utiliza-se um coeficiente de similaridade, em contrapartida, se o menor valor para um coeficiente indicar que dois objetos são

parecidos, trabalha-se com um coeficiente de dissimilaridade (MURTAGH; CONTRERAS, 2012).

A distância euclidiana é um dos coeficientes de dissimilaridade mais conhecidos e utilizados, enquanto que o coeficiente de correlação é um exemplo de coeficiente de similaridade (HAIR JR. *et al.*, 2009). Segundo os autores, necessita-se determinar os coeficientes que melhor se adaptam a determinados tipos de variáveis e situações, sendo, geralmente, os coeficientes de dissimilaridade mais adequados para as variáveis quantitativas, e os de similaridade para as variáveis qualitativas. Farsadnia *et al.* (2014) concluíram que a análise de agrupamento foi bem sucedida quando empregada para identificar regiões homogêneas na regionalização hidrológica, listando os métodos hierárquicos e não hierárquicos.

I) Métodos Hierárquicos

Na técnica de análise de agrupamento existem duas categorias que são utilizadas para a formação dos grupos, cada uma delas incluindo diferentes tipos de algoritmos: os métodos hierárquicos são subdivididos em métodos aglomerativos, indicando a fusão de grupos, e divisivos, indicando a separação dos grupos em cada etapa; já os não hierárquicos caracterizam os métodos por partição (CONRADT; GORNOTT; WECHSUNG, 2016; RAO; SRINIVAS, 2006).

Esses métodos se distinguem, principalmente, pelo procedimento utilizado na construção dos grupos. A escolha de um determinado método dessa técnica exige o conhecimento das propriedades desse particular algoritmo, aliado aos objetos da pesquisa (SINGH *et al.*, 2016). Os métodos hierárquicos consistem em uma série de sucessivos agrupamentos ou contínuas divisões de elementos, onde os elementos são agregados ou desagregados. Segundo Conradt; Gornott; Wechsung (2016), os métodos de hierarquização conduzem a uma hierarquia de partições.

Desta forma, cada um dos objetos que participarão do processo de análise de agrupamento são considerados como um grupo, que se juntarão a um outro elemento ou a um outro grupo, de acordo com os valores existentes na matriz de dissimilaridade ou distâncias (que é atualizada a cada agrupamento que surge). Ao fim desse processo, todos os elementos estarão reunidos em um único grupo.

A representação dos métodos hierárquicos geralmente é dada por um diagrama bidimensional, comumente chamado de dendrograma ou árvore de ligação (FARHAN;

SHAIKH, 2017), que registra as sequências de fusões ou divisões dos clusters (SINGH *et al.*, 2016). No dendrograma, cada ramo representa um elemento, enquanto a raiz representa o agrupamento de todos os elementos. Para Singh *et al.* (2016), quanto maior o ramo vertical de uma fusão, menos comuns são os dois agrupamentos.

a) Medidas de Distância

Anterior ao cálculo das distâncias é necessário realizar a padronização dos dados/variáveis, visto que, as variáveis apresentam discrepância entre as unidades de medida, o que pode acarretar em uma falsa correlação entre elas (HAIR JR. *et al.*, 2009). Essa metodologia de padronização de variáveis é conhecida por z-scores conforme Mingoti (2013) e Wilks (2006), ou seja, subtrai-se do valor original a média e divide-se pelo desvio padrão associados a cada uma das variáveis.

Segundo Abdi; Hassanzadeh; Ouarda (2017) e Chiang; Tsay; Nix (2002), todas as variáveis são igualmente importantes para a determinação das distâncias. Na literatura é possível encontrar estudos que utilizam a distância Euclidiana (CHIANG; TSAY; NIX, 2002; OUDIN *et al.*, 2010; SANTOS; LUCIO; SILVA, 2014; SINGH *et al.*, 2016; VEZZA *et al.*, 2010; YANG *et al.*, 2010) a distância de Manhattan (JOTHEESWARAN; KUMARASWAMY, 2013; YACOB; SAKIM; ISA, 2012) e a distância de Mahalanobis (CORDUAS, 2011; CUNDERLIK; BURN, 2006; OUARDA *et al.*, 2006), para calcular as matrizes de dissimilaridade em análises hidrológicas.

b) Algoritmos de Agrupamento

- **Ligação Simples**

O método também é conhecido como método do vizinho mais próximo ou da distância mínima (HAIR JR. *et al.*, 2009). Quando aplicado para fornecer a distância entre um grupo e outro, seleciona-se a distância correspondente a maior semelhança entre os elementos de grupos distintos. Se o coeficiente utilizado na construção da matriz de distâncias for um coeficiente de similaridade, essa maior semelhança será representada pela maior distância existente entre esses elementos; caso o coeficiente seja de dissimilaridade, essa maior semelhança será a menor distância existente (HAIR JR. *et al.*, 2009; MANLY, 1994).

No algoritmo de ligação simples, a distância entre dois clusters é a distância entre o par mais próximo de vetores característicos, ou seja, a distância mínima (BI; LUO; SUN, 2017). Este algoritmo tende a formar um pequeno número de grandes clusters, com restrições de pequenos agrupamentos periféricos nas margens do espaço das características do local. Segundo Hosking; Wallis (1997) e Rao; Srinivas (2006) este algoritmo não produz boas regiões homogêneas na análise regional.

- **Ligação Completa**

O método da ligação completa, ou método do vizinho mais distante, forma grupos mais homogêneos comparativamente ao método da ligação simples, gerando resultados opostos (HAIR JR. *et al.*, 2009). A aplicação desse método busca encontrar a distância entre os elementos de um grupo e de outro, que irá representar a menor semelhança entre esses elementos. Se o coeficiente utilizado na construção da matriz de distâncias for de similaridade, a distância escolhida será o menor valor dentre as existentes nos dois grupos distintos; caso o coeficiente for de dissimilaridade, a distância escolhida será a maior existente (HAIR JR. *et al.*, 2009; MANLY, 1994).

No algoritmo de ligação completa, a distância entre dois clusters é definida como a maior distância entre um par de vetores característicos, ou seja, distância máxima (LANCE; WILLIAMS, 1967; MURTAGH; CONTRERAS, 2012). Este algoritmo tende a formar pequenos aglomerados bem ligados e geralmente não é adequado para aplicação em grandes conjuntos de dados (HOSKING; WALLIS, 1997; RAO; SRINIVAS, 2006).

- **Ward**

O algoritmo de Ward (WARD, JR. 1963) é uma técnica frequentemente utilizada para estudos de regionalização em hidrologia e climatologia (FARHAN; SHAIKH, 2017; FARSADNIA *et al.*, 2014; HOSKING; WALLIS, 1997; MALEKINEZHAD *et al.*, 2011; YANG *et al.*, 2010). Baseia-se no pressuposto de que, se dois núcleos forem fundidos, a perda de informação resultante ou a mudança no valor da função objetiva dependerão apenas da relação entre os dois clusters combinados, e não nas relações com outros clusters (FARSADNIA *et al.*, 2014; RAO; SRINIVAS, 2008).

O método emprega a análise de variância para determinar as distâncias entre clusters e, a cada nova interação, reunir de forma a minimizar a soma dos quadrados de quaisquer pares de clusters hipotéticos. De acordo com YANG *et al.* (2010), é considerado como eficiente e geralmente tende a produzir clusters pouco extensos e de igual número de elementos.

Segundo Ilorme; Griffis (2013), o método calcula a distância entre dois agrupamentos e a soma dos quadrados entre os dois agrupamentos realizada sobre todas as variáveis (Equação 1). Nesse método, a formação dos grupos se dá pela maximização da homogeneidade dentro dos grupos.

$$W = \sum_{i=1}^n x_i^2 - \frac{1}{n} \left(\sum x_i \right)^2 \quad (1)$$

Em que:

W = homogeneidade intragrupos e heterogeneidade intergrupos através da soma do quadrado dos desvios;

n = número de valores analisados; e

x_i = i-ésimo elemento do agrupamento.

RAO; SRINIVAS (2006) usaram três algoritmos híbridos nos quais aplicaram um processo de agrupamento particional para refinar os resultados obtidos das técnicas de agrupamento hierárquicas. Foram usados os algoritmos hierárquicos de ligação simples, ligação completa e Ward, enquanto o algoritmo particional utilizado foi o k-means. O método híbrido entre k-means e Ward apresentou melhores resultados e foi recomendado por oferecer perspectivas de melhoria em estudos de regionalização em bacias hidrográficas em Indiana, EUA.

Farsadnia *et al.* (2014) testaram os métodos fuzzy c-mean, k-means, e Ward com a distância euclidiana, sendo esse último o melhor resultado com 4 regiões homogêneas. Nazari *et al.* (2017); Raziei (2017) observaram que o algoritmo de Ward em seus estudos apresentaram uma classificação mais realística em relação as divisões das regiões. Estes resultados também confirmam a afirmação de Gonçalves *et al.* (2016) de que o método de Ward supera outras abordagens de agrupamento na classificação de variáveis climatológicas e produz grupos, onde cada um deles,

segundo Hair Jr. *et al.* (2009), possui características semelhantes e independência relativamente alta para cada grupo.

II) Métodos não Hierárquicos

Os métodos não hierárquicos, também chamados de método de partição, têm como objetivo encontrar diretamente uma partição de n elementos em k grupos, de modo que a partição satisfaça dois requisitos básicos: “coesão” interna (ou “semelhança” interna) e isolamento (ou separação) dos grupos formados (MINGOTI, 2013).

Segundo Naldi *et al.* (2011), nem todos os valores de k apresentam grupos satisfatórios, sendo assim, aplica-se o método várias vezes para diferentes valores de k , escolhendo os resultados que apresentem melhor interpretação dos grupos. Diferentemente dos métodos hierárquicos, os procedimentos não hierárquicos não envolvem o processo de construção em árvore (HAIR JR. *et al.*, 2009). Entre os métodos não hierárquicos, o algoritmo k-means é o mais conhecido e utilizado (HANCER; KARABOGA, 2017).

A principal desvantagem do método não hierárquico é a necessidade de uma suposição inicial do número de agrupamentos (k) e centros de fragmentação (CUPAK, 2017), o que poderá sugerir interpretações erradas sobre a estrutura dos dados, caso o número de agrupamentos não seja o ideal, visto que o algoritmo é em geral sensível às condições iniciais, podendo gerar resultados diferentes a cada interação (GOYAL; GUPTA, 2014). Por mais que haja subjetividade nessa decisão, deve ser baseada em conhecimento teórico que possa sugerir um número natural de agrupamentos de melhor qualidade.

Segundo Hair Jr. *et al.* (2009), infelizmente, não existe qualquer procedimento de seleção padrão ou objetivo. Inicialmente, são escolhidos arbitrariamente k elementos, os quais serão os centros iniciais (provisórios) dos grupos. A partir dessa escolha, os k primeiros elementos já foram designados. Cada um dos demais elementos amostrais será, então, colocado no grupo mais próximo, iterativamente, baseando-se na distância euclidiana (HANCER; KARABOGA, 2017; NALDI *et al.*, 2011). Assim, serão formados os k grupos iniciais, nos quais cada elemento possui menor distância em relação ao centro do grupo ao qual foi alocado. O algoritmo

procura, dentro do possível, a partição em que os padrões de cada agrupamento estão mais próximos entre si e mais distante dos padrões dos outros agrupamentos.

Cupak (2017), empregando o algoritmo k-means juntamente com características fisiográficas e meteorológicas, obteve duas regiões homogêneas para a Q₉₅ na Cracóvia. Kulkarni (2016), utilizando o mesmo algoritmo na Índia, observaram que quanto maior a suposição inicial (k), mais fragmentadas ficaram as regiões homogêneas.

Goyal; Gupta (2014) identificaram regiões homogêneas de precipitação no nordeste da Índia empregando o algoritmo k-means. Hao; Zhang; Yao (2015) classificaram em seis regiões heterogêneas, o sudoeste da China pelo k-means com base em diferentes variáveis explicativas, onde cada sub-região pode ser vista como uma área homogênea, em relação as propriedades de seca.

Na literatura também é comumente encontrada uma abordagem de agrupamento em duas etapas, a qual obtém primeiramente o número de clusters por um método hierárquico e depois reagrupam-se, introduzindo um método não hierárquico usando os centroides do cluster, a partir dos resultados hierárquicos como as suposições iniciais (HSU; HUANG, 2017), como no estudo de Rao; Srinivas (2006).

2.8.2 Índices de validação de agrupamentos

Em estudos que envolvem análise de agrupamento é comum empregar medidas de qualidade, como os índices de validação nos grupos formados, para identificar, quanto a coesão e separação dos mesmos, assim como a qualidade dos resultados gerados pelos algoritmos de agrupamento.

Vendramin; Campello; Hruschka (2010), indicam os índices de Calinski-Harabasz, Dunn e Silhouette para a validação dos agrupamentos. Os autores ainda ressaltam a dificuldade de decidir o melhor número de grupos, contudo, revelam que a alternativa seja a comparação dos desempenhos dos índices.

Dos diferentes critérios de validação é preferível os de maximização, visto que critérios de minimização como Davies-Bouldin e C-índice devem ser convertidos em maximização (VENDRAMIN; CAMPELLO; HRUSCHKA, 2010). Um cluster deverá ser composto de pelo menos dois objetos, caso contrário, caracterizam-se como um *outlier* ao invés de um grupo de objetos (BIVAND; WILK; KOSSOWSKI, 2017).

Na regionalização, o uso desses índices de validação ainda é pouco comum, pois refere-se a uma medida estatística, preferindo-se o emprego de testes como a medida de heterogeneidade (H) de Hosking; Wallis (1993;1997) para avaliar os agrupamentos gerados por diferentes algoritmos, como os trabalhos de Goyal; Gupta (2014); Hao; Zhang; Yao (2015); Ilorme; Griffis (2013); Malekinezhad *et al.* (2011); Modarres (2010; Modarres; Sarhadi (2011); She *et al.* (2014; 2016); Zhang *et al.* (2012); e Yang *et al.* (2010).

2.9 Regressão Múltipla

A regressão múltipla é uma ferramenta de grande importância nos estudos de regionalização hidrológica (LAAHA; BLÖSCHL, 2006). É a partir dela que as variáveis dependentes, como parâmetros de distribuição teórica de probabilidades, *quantis* de curvas de permanência de vazões mínimas e máximas, dentre outros, são relacionadas a características físicas, climáticas e morfológicas das bacias hidrográficas.

A análise de regressão múltipla é uma técnica estatística que pode ser usada para verificar a relação entre uma única variável dependente e várias independentes. O objetivo da análise é usar as variáveis independentes, cujos valores são conhecidos, para prever os valores da variável dependente de estudo (HAIR JR. *et al.*, 2009).

A regressão regional é realizada em relação às variáveis dependentes Q_{90} e Q_{95} , por exemplo, com as variáveis independentes área, precipitação total anual, comprimento do rio e densidade de drenagem. Tais variáveis relacionadas devem pertencer a uma região hidrologicamente homogênea, prosseguindo o ajuste das equações de regionalização, por meio de regressões.

Segundo Naghettini; Pinto, (2007), os modelos de regressão mais frequentes são o potencial, o exponencial e o logarítmico. Para os autores, independentemente do tipo de função empregada, o modelo ideal é aquele com o menor número de variáveis explicativas e que apresente pequeno erro padrão de estimativa e alto coeficiente de determinação.

A utilização dos modelos de regressão para além dos limites dos dados amostrais não é recomendada por dois motivos (NAGHETTINI; PINTO, 2007): (a) o intervalo de confiança sobre a linha de regressão expande à medida que a variável

independente se afasta da média; e (b) a relação da variável dependente e independente pode ser diferente da encontrada para o limite utilizado na construção do modelo. A recomendação dos autores, de certa forma, causa a limitação da utilização das equações de regionalização em áreas não monitoradas, além de evitar valores de vazões de referência não condizentes com a realidade.

2.9.1 Modelos de regressão

Os modelos são resultantes da aplicação das regressões múltiplas às regiões hidrologicamente homogêneas fornecidas pela análise de agrupamento. A relação entre a variável dependente e as demais variáveis independentes pode, segundo Hair Jr. *et al.* (2009), ser formulada de acordo com modelos dos tipos linear (Equação 2), potencial (Equação 3), exponencial (Equação 4), logarítmico (Equação 5), recíproco (Equação 6):

$$Q = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n + \varepsilon \quad (2)$$

$$Q = \beta_0 \cdot X_1^{\beta_1} \cdot X_2^{\beta_2} \dots X_n^{\beta_n} \cdot \varepsilon \quad (3)$$

$$Q = \beta_0 \cdot e^{\beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n + \varepsilon} \quad (4)$$

$$\ln Q = \beta_0 + \beta_1 \ln X_1 + \beta_2 \ln X_2 + \dots + \beta_n \ln X_n + \varepsilon \quad (5)$$

$$Q = (\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n + \varepsilon)^{-1} \quad (6)$$

Em que:

Q = representa a vazão característica de interesse em m^3s^{-1} ;

β_i = coeficientes das n variáveis independentes adimensionais;

X_i = variáveis independentes envolvidas no estudo (variáveis climáticas e morfométricas); e

ε = erro de estimativa.

Para verificação do melhor ajuste dos modelos de regressão adotam-se critérios. Elesbon *et al.* (2015) utilizaram o coeficiente de determinação ajustado (R^2_a

$\geq 0,70$), o erro-padrão da estimativa menor que 0,5 ($EP < 0,5$) e nível de significância de 5% pelo teste F, concluindo no estudo que o modelo potencial apresentou o melhor ajuste para as vazões mínimas estudadas.

Costa *et al.* (2012), avaliando as curvas de permanência de vazões, utilizaram 5 modelos matemáticos de regressão (exponencial, logarítmico, potencial, quadrático e cúbico). Os melhores resultados foram obtidos pelos modelos cúbico e exponencial, pois apresentaram maiores coeficientes de Nash-Sutcliffe e R^2_a , menores erros quadráticos relativos médios percentuais e melhores ajustes das curvas.

Baena *et al.* (2004), Lisboa *et al.* (2008), Lujano *et al.* (2017) e Silva; Marques; Lemos (2011) constataram que o melhor ajuste na realização das regressões em estudos de regionalização de vazões para diferentes níveis de vazão de permanência foi obtido com a equação potencial.

Diferentes formas de modelagem hidrológica utilizando regressão foram analisadas e discutidas por Melati (2016), onde o objetivo foi estudar e discutir a conceituação de regressão matemática, logaritimização (linearização) de modelos de regressão com uma ou duas variáveis explicativas, as possíveis diferenças entre modelos logaritimizadas e “naturais”, e a aplicação e avaliação de modelos, destacando o modelo potencial.

Segundo Naghettini; Pinto (2007), algumas funções podem ser linearizadas, mediante o uso de transformações adequadas, permitindo a aplicação da regressão linear simples e múltipla, sendo a transformação dos eixos x e y na escala logarítmica de base 10 muito utilizada para regressões onde a função é do tipo potencial, o qual é comumente utilizado em estudos que empregam vazões mínimas.

a) Coeficiente de determinação

O coeficiente de determinação (R^2) é sempre positivo, e expressa a proporção da variância total da variável dependente, que é explicada pelo modelo de regressão (NAGHETTINI; PINTO, 2007), o qual pode ser estimada pela Equação 7.

$$R^2 = b^2 \cdot \frac{S_x^2}{S_y^2} \quad (7)$$

Em que:

b = coeficiente angular da reta de regressão;

S_X^2 = variância amostral de X; e

S_Y^2 = variância amostral de Y.

b) Coeficiente de determinação ajustado

O valor do coeficiente de determinação ajustado (R_a^2), ou não tendencioso, é calculado considerando o número de variáveis independentes da equação de regressão expresso pela Equação 8 (NAGHETTINI; PINTO, 2007). Costa *et al.* (2012), Elesbon *et al.* (2015) e Vezza *et al.* (2010) utilizaram o R_a^2 para avaliar modelos de regionalização. Segundo Vezza *et al.* (2010), esse coeficiente é útil para escolher o melhor modelo entre os pertencentes a uma determinada classe, mas não pode ser usado para comparar modelos de natureza diferente.

$$R_a^2 = 1 - \frac{(n-1)}{(n-p-1)} \cdot (1-R^2) \quad (8)$$

Em que:

n = número de valores observados;

p = número de variáveis independentes; e

R^2 = coeficiente de determinação.

2.9.2 Verificação dos modelos de regionalização

Segundo Garcia; Folton; Oudin (2017) há uma dificuldade potencial em encontrar um conjunto robusto de parâmetros para avaliar os modelos, quando as variáveis de destino são índices de vazões mínimas. A maioria das funções objetivas existentes favorecem a qualidade de ajuste de modelos que trabalham com vazões elevadas.

A raiz quadrada e os logaritmos estão entre as transformações mais amplamente utilizadas em valores de vazões mínimas, segundo Pushpalatha *et al.* (2012). Jain; Sudheer (2008) e Oudin *et al.* (2006) utilizaram a transformação do critério de eficiência de Nash e Sutcliffe aplicado à vazão logarítmica.

a) Coeficiente de Nash-Sutcliffe (C_{NS})

O coeficiente de Nash-Sutcliffe (NASH; SUTCLIFFE, 1970), segundo Baltokoski *et al.* (2010) e Razavi; Coulibaly (2013), é um dos mais importantes critérios estatísticos para avaliar o ajuste de modelos hidrológicos (Equação 9).

$$(C_{NS}) = 1 - \frac{\sum_{i=1}^n ((Q_{obs_i}) - (Q_{est_i}))^2}{\sum_{i=1}^n ((Q_{obs_i}) - (\overline{Q_{obs}}))^2} \quad (9)$$

Conhecido também como coeficiente de eficiência de Nash-Sutcliffe, traduz a eficiência de realizar previsões mais acertadas nas cheias, ou seja, quando o aproveitamento encontra-se com vazões bastante elevadas. Quanto maior a eficiência, ou seja, maior o valor do índice, melhores serão as previsões. Este índice é limitado em 1 (GUILHON; ROCHA; MOREIRA, 2007). Já na sua forma logarítmica, há o amortecimento da amplitude, o que possibilita ao coeficiente buscar melhores ajustes aos valores mínimos. Pushpalatha *et al.* (2012) ressaltam que os logaritmos estão entre as transformações mais amplamente utilizados em valores de vazões mínimas.

O coeficiente de eficiência de Nash-Sutcliffe aplicado aos logaritmos ($\log(C_{NS})$) (Equação 10) valoriza mais as vazões baixas/mínimas, tornando-se menos sensível a cheias (RAZAVI; COULIBALY, 2017). Quanto maior a eficiência, ou seja, maior o valor do índice, melhores serão as previsões. Este índice também é limitado em 1 (GUILHON; ROCHA; MOREIRA, 2007).

$$\log(C_{NS}) = 1 - \frac{\sum_{i=1}^n (\log(Q_{obs_i}) - \log(Q_{est_i}))^2}{\sum_{i=1}^n (\log(Q_{obs_i}) - \log(\overline{Q_{obs}}))^2} \quad (10)$$

Em que:

n = número de postos fluviométricos;

C_{NS} = coeficiente de eficiência de Nash-Sutcliffe;

$\log(C_{NS})$ = coeficiente de eficiência de Nash-Sutcliffe do logaritmo das vazões;

Q_{obsi} = vazão observada em (m^3s^{-1}) dos i-ésimos postos;

$\log(Q_{obsi})$ = logaritmo da vazão observada em (m^3s^{-1}) dos i-ésimos postos;

Q_{esti} = vazão estimada em (m^3s^{-1}) dos i-ésimos postos;

$\log(Q_{esti})$ = logaritmo da vazão estimada em (m^3s^{-1}) dos i-ésimos postos;

$\overline{Q_{obs}}$ = média das vazões observadas em (m^3s^{-1}) e;

$\log(\overline{Q_{obs}})$ = logaritmo da média das vazões observadas em (m^3s^{-1}).

Beldring *et al.* (2003); Moriasi *et al.* (2007); Viola *et al.* (2009) e Price *et al.* (2012) utilizaram a classificação de Motovilov *et al.* (1999), para avaliar a precisão do modelo com base nesses coeficientes, sendo que: $C_{NS} > 0,75$ o modelo é considerado adequado e bom; $0,36 < C_{NS} \leq 0,75$ o modelo é considerado aceitável e $C_{NS} \leq 0,36$ o modelo é considerado insatisfatório.

b) Índice de desempenho RSME e EMA

A raiz quadrada da média dos erros quadráticos (Root Mean Squared Error - RMSE), dada pela Equação 11, é utilizada para estimar a qualidade de um modelo. O uso do RMSE tem grande vantagem em relação ao coeficiente de determinação (R^2), contudo não estabelece um percentual ou tipo de erro como critério de comparação. Já o erro médio absoluto - EMA (Mean Absolute Error – MAE) dado pela Equação 12, é uma medida que indica a média do afastamento entre todos os valores estimados e os observados. Quanto menores os valores encontrados nesses índices, melhor é a qualidade de ajuste, logo, menores erros.

$$RSME = \frac{\sum_{i=1}^n (Q_{esti} - Q_{obsi})^2}{n} \quad (11)$$

$$EMA = \frac{\sum_{i=1}^n (Q_{esti} - Q_{obsi})}{n} \quad (12)$$

Em que:

n = números de postos fluviométricos;

Q_{est_i} = valor estimado dos i-ésimos postos fluviométricos; e

Q_{obs_i} = valor observado dos i-ésimos postos fluviométricos.

O EMA traduz o quanto as previsões de vazões se desviam (em m^3s^{-1}), em relação aos valores observados, ou seja, em relação ao seu objetivo, que é acertar a previsão. A tendência é que esse valor seja mais elevado para locais com maiores vazões e seja menor para locais com valores menores de vazões. Quanto menor o seu valor, melhores são as previsões (GUILHON; ROCHA; MOREIRA, 2007). Para Portals *et al.* (2010), a diferença entre eles é que o RMSE coloca muito peso em altos erros, enquanto o EMA é menos sensível aos valores extremos.

c) Índice de desempenho “c” e de Willmott “d”

O índice de desempenho “c” dado pela Equação 13, proposto por Camargo; Sentelhas (1997), utilizado por Alves; Biudes; Vecchia (2012), Beskow *et al.* (2016) e Wolff; Duarte; Mingoti (2014), visa quantificar a performance dos modelos, conforme critérios da Tabela 1.

$$c = r \cdot d \quad (13)$$

O índice avalia estatisticamente o desempenho dos modelos, de forma a correlacionar os valores estimados com os observados (CAMARGO; SENTELHAS, 1997).

Tabela 2 - Critério de interpretação do índice “c”

Valor “c”	Desempenho
> 0,85	Ótimo
0,76 a 0,85	Muito Bom
0,66 a 0,75	Bom
0,61 a 0,65	Mediano
0,51 a 0,60	Sufrível
0,41 a 0,50	Mau
≤ 0,40	Péssimo

A precisão é dada pelo coeficiente de correlação “r”, que indica o grau de dispersão dos dados observados em relação à média, ou seja, o erro aleatório. A exatidão está relacionada ao afastamento dos valores estimados em relação aos

observados. Esta aproximação é dada pelo índice de Willmott “d” (Equação 14) (WILLMOTT *et al.*, 1985).

$$d = 1 - \left[\frac{\sum (Q_{\text{est}i} - Q_{\text{obs}i})^2}{\sum (|Q_{\text{est}i} - \overline{Q_{\text{obs}}}| + |Q_{\text{obs}i} - \overline{Q_{\text{obs}}}|)^2} \right] \quad (14)$$

Em que:

$Q_{\text{est}i}$ = valor estimado dos i-ésimos postos fluviométricos;

$Q_{\text{obs}i}$ = valor observado dos i-ésimos postos fluviométricos; e

$\overline{Q_{\text{obs}}}$ = média dos valores observados.

3 METODOLOGIA

3.1. Área de estudo

O estudo abrange as bacias hidrográficas que contemplam as suas áreas inteiramente dentro nos limites do estado do Rio Grande do Sul, aproximadamente entre as coordenadas geodésicas 27° e 34°S latitude sul e 49° e 58°W longitude oeste (Figura 1). O estado ocupa uma área de aproximadamente 282.000km², com uma população de 10.693.929 habitantes, distribuída em 497 municípios, de modo que 85,1% residem em áreas urbanas e 14,9% em áreas rurais (IBGE, 2010).

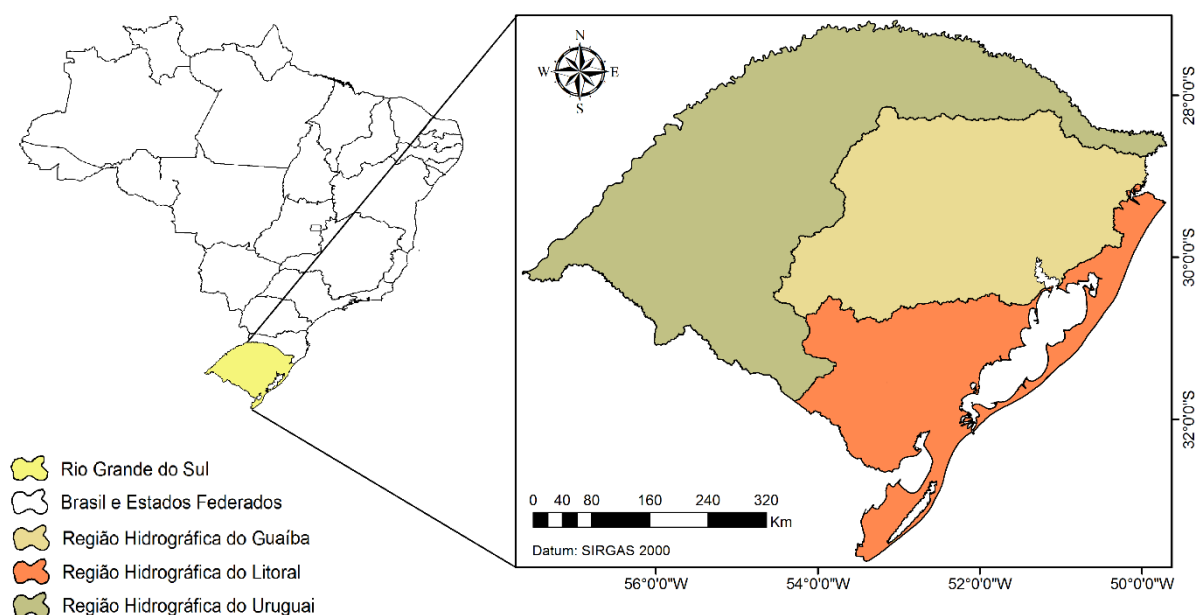


Figura 1 - Localização da área de estudo e regiões hidrográficas.

A região hidrográfica do Guaíba situa-se na porção nordeste, abrangendo uma área que corresponde a 30% da área total do Estado; a região hidrográfica do Litoral está localizada na parte leste e sul do território rio-grandense e ocupa uma superfície a 20,11 % da área do Estado; e a região hidrográfica do Uruguai abrange a metade norte, noroeste e oeste do território, com equivalente a 47,88% da área do Estado (FEPAM, 2018).

De acordo com Alvares (2013), o estado do Rio Grande do Sul caracteriza-se pelo clima subtropical de verão quente e úmido, do tipo Cfa e Cfb, em consonância

com a classificação de Köppen. O clima subtropical úmido com verões amenos (Cfb) ocorre na Serra do Sudeste e na Serra do Nordeste, onde as temperaturas médias dos meses de verão ficam abaixo dos 22°C, e o tipo Cfa nas demais regiões, onde a temperatura média do mês mais quente ultrapassa os 22°C. Para Mello *et al.* (2013), a precipitação se encontra bem distribuída em todas as regiões, sem concentração de precipitação em uma determinada estação do ano.

3.2. Seleção e processamento de dados

Os dados das estações fluviométricas e pluviométricas foram obtidos junto ao banco de dados hidrometeorológicos da ANA, disponibilizado pelo portal HidroWeb – Sistema de Informações Hidrológicas (<http://hidroweb.ana.gov.br>).

O critério de seleção das estações fluviométricas foi baseado na extensão das séries históricas, adotando-se a indicação de Cupak (2017), Mamum *et al.* (2010), Molina *et al.* (2014) e Vezza *et al.* (2010), os quais destacam a utilização de um período mínimo de 10 anos de dados observados de vazões diárias. Garcia; Folton; Oudin (2017) enfatizam que para essas séries não seja ultrapassado o limite de 10% de dados faltantes, permitindo-se no presente estudo até 31 dias de falhas ao ano. Para a seleção das estações pluviométricas empregou-se o critério indicado por Caldeira *et al.* (2015), adotando-se também estações com no mínimo 10 anos de registros sem falhas.

Salviano; Groppo; Pelledrino (2016) e Uliana *et al.* (2015) recomendam que sejam aplicados testes não paramétricos nas séries hidrometeorológicas selecionadas, com o objetivo de eliminar as estações que possuem tendência temporal em suas séries de dados. Assim como observado para Santos *et al.* (2016), o teste de tendência para séries anuais fornece uma visão global da mudança nos processos hidrológicos. No presente estudo foi utilizado o teste de Mann-Kendall (KENDALL, 1975; MANN, 1945), considerando um nível de 5% de significância ($p < 0,05$), e o teste de homogeneidade de Mann-Whitney (MANN; WHITNEY, 1947) também com um nível de 5% de significância ($p < 0,05$), para identificar séries que possuíam dados atípicos.

As vazões mínimas de referência foram estimadas a partir das séries de vazões médias diárias empregando-se o quantil de vazão Q_{90} e Q_{95} , que corresponde à vazão que é igualada ou excedida em 90% e 95% do tempo de monitoramento,

respectivamente. A manipulação e a elaboração de ambas as séries hidrológicas com as propriedades estabelecidas foram realizadas com o auxílio do *software* SHYDA - System of Hydrological Data Acquisition and Analysis (BESKOW *et al.*, 2013).

- **Teste de Mann-Kendall**

O teste de Mann-Kendall (KENDALL, 1975; MANN, 1945) é aplicado para evidenciar tendências em séries temporais somente se as mesmas forem independentes. Desta forma, as observações da série $X_1, X_2, \dots, X_n$ são testadas quanto a sua independência e distribuição. A hipótese nula H_0 é que as observações da série são independentes e identicamente distribuídas, logo não possuem tendência (MELLO; SILVA, 2013). A estatística do teste é obtida por:

$$S = \sum_{k=1}^{n-1} \sum_{j=k+1}^n \text{sign}(x_j - x_k) \quad (15)$$

Sendo,

$$\text{sign}(x) = \begin{cases} 1, & \text{se } x > 0 \\ 0, & \text{se } x = 0 \\ -1, & \text{se } x < 0 \end{cases}$$

É possível mostrar que S é normalmente distribuída, ou seja, $S \sim N(\mu, \sigma^2)$, sendo μ a média e σ^2 a variância com:

$$\mu = 0 \quad (16)$$

$$\sigma^2 = \frac{n(n-1)(2n+5) - \sum_{j=1}^P t_j(t_j-1)(2t_j+5)}{18} \quad (17)$$

Em que:

n = número de observações;

P = número de grupos com observações iguais; e

t_j = número de observações iguais no grupo j .

No caso em que o número de observações seja superior a 30, a estatística do teste é calculada por:

$$Z = \begin{cases} \frac{S-1}{\sigma}, & \text{se } S > 0 \\ 0, & \text{se } S = 0 \\ \frac{S+1}{\sigma}, & \text{se } S < 0 \end{cases}$$

A decisão de rejeitar a hipótese nula H_0 é dada por:

$$|Z| > Z_{(1-\frac{\alpha}{2})} \quad (18)$$

α = nível de significância de Z .

- **Teste de Mann-Whitney**

O teste de Mann-Whitney (MANN; WHITNEY, 1947) implica que todos elementos de uma certa amostra provêm de uma única e idêntica população. De acordo com Mello; Silva (2013), o teste ainda sugere que uma dada série histórica de vazões esteja associada a chuvas, cuja frequência de ocorrência seja normal (esperada). Vazões de pico originárias de eventos de chuvas históricas, oriundas de forma atípica, causam heterogeneidade à série, caracterizando uma observação na forma de um “*outlier*”.

A hipótese nula, H_0 , do teste considera que a série é homogênea. Para sua aplicação é recomendada que a série histórica seja dividida em duas partes com aproximadamente o mesmo número de dados, constituindo duas subamostras (MELLO; SILVA, 2013). A estatística do teste é obtida por:

$$T = \frac{V - E(V)}{\sqrt{\text{var}(V)}} \quad (19)$$

Primeiramente, calcula-se o valor de V para a primeira subamostra:

$$V_1 = N_1 \cdot N_2 \frac{N_1 \cdot (N_1 + 1)}{2} - H_1 \quad (20)$$

Na sequência, calcula-se o valor de V para a segunda subamostra.

$$V_2 = N_1 \cdot N_2 - V_1 \quad (21)$$

O valor de V no cálculo de T é o menor entre V_1 e V_2 .

$$E(V) = \frac{N_1 \cdot N_2}{2} \quad (22)$$

$$\text{var}(V) = \frac{N_1 \cdot N_2 \cdot (N_1 + N_2 + 1)}{12} \quad (23)$$

Em que:

N_1 = número de dados da primeira subamostra;

N_2 = número de dados da segunda subamostra;

H_1 = soma das ordens de classificação da série histórica da primeira subamostra (soma de F_i do teste de estacionariedade); e

$\text{var}(V)$ = variância do coeficiente.

A decisão de rejeitar a hipótese nula, H_0 , é dada por:

$$|T| > Z_{\left(1 - \frac{\alpha}{2}\right)} \quad (24)$$

α = nível de significância de Z.

3.3. Modelo digital de elevação e extração de parâmetros

Para a extração dos parâmetros usados como variáveis explicativas foi utilizado o banco de dados TOPODATA, conforme recomendações de Moura *et al.* (2017); Miceli *et al.* (2011) e Polidori; Hage; Valeriano (2014). A etapa inicial consistiu da seleção dos arquivos resultantes do projeto TOPODATA (Tabela 3), para a cobertura da área do estado do Rio Grande do Sul.

De posse das cenas, as etapas seguintes consistiram na definição da referência espacial, utilizando a ferramenta Define Projection, presente na ArcToolbox do SIG ArcGIS, definindo o sistema de coordenadas projetadas SIRGAS 2000, fusos 21 Sul e 22 Sul.

Tabela 2 - Cenas referentes ao estado do Rio Grande do Sul no banco de dados TOPODATA

n	Cena	n	Cena	n	Cena
1	31S51	11	27S525	21	29S555
2	30S51	12	33S54	22	28S555
3	29S51	13	32S54	23	27S555
4	28S51	14	31S54	24	31S57
5	27S51	15	30S54	25	30S57
6	32S525	16	29S54	26	29S57
7	31S525	17	28S54	27	28S57
8	30S525	18	27S54	28	30S585
9	29S525	19	31S555	29	29S585
10	28S525	20	30S555		

As informações geográficas foram georreferenciados utilizando o sistema de projeção cartográfica Universal Transversa de Mercator - UTM e datum SIRGAS 2000, conforme normativa do decreto nº 5.334 de 6 de janeiro de 2005 (BRASIL, 2005). Ainda para melhor visualização dos resultados foi utilizado a rede hidrográfica na escala 1:2.500.000 disponibilizada pela ANA, na construção das imagens.

Para a manipulação das cenas e delimitação das bacias hidrográficas, a partir do MDE TOPODATA, foram empregadas as funções matemáticas presentes no *software* ArcGIS® v.10.3, desenvolvido pelo “Environmental Systems Research Institute – ESRI”, de acordo com a metodologia descrita por Alves Sobrinho *et al.* (2010); Dias *et al.* (2004), a qual é subdividida em quatro etapas (Figura 2).

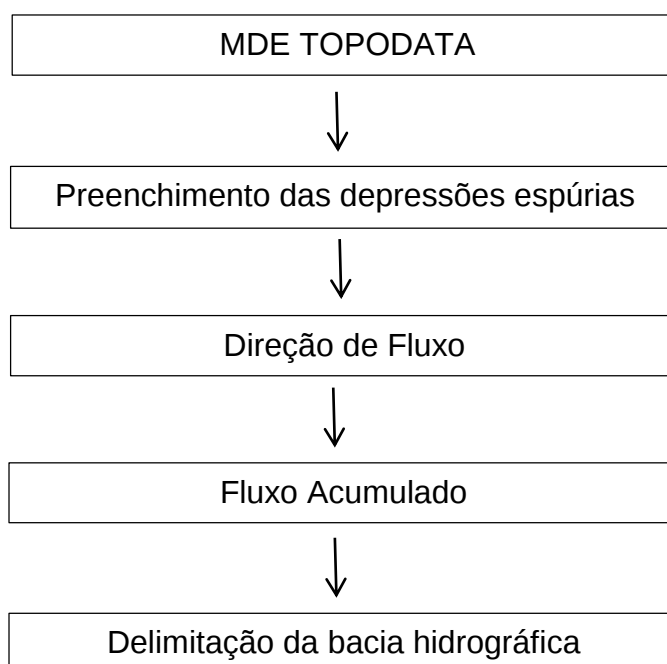


Figura 2 - Fluxograma utilizado para delimitação das bacias hidrográficas, a partir do banco de dados TOPODATA.

As depressões espúrias são falhas advindas do MDE bruto, sendo células cercadas por outras com maiores valores de elevação. O preenchimento das depressões espúrias foi executado na função “fill sink” (JIMME; BASHIR; ADEBAYO, 2016), e consiste na correção dos valores conforme os encontrados na vizinhança (MENDES; CIRILO, 2001).

A direção de fluxo ou direção de escoamento superficial foi calculada pelo método D8 (O’CALLAGHAN; MARK, 1984), obtido pela função “flow direction”, que considerou apenas uma das oito direções possíveis de escoamento para cada uma das células do MDE – norte, nordeste, leste, sudeste, sul, sudoeste, oeste ou noroeste. A direção do escoamento descreve e espacializa o sentido do fluxo entre cada célula do MDE e de suas oito células vizinhas (ZHANG *et al.*, 2017), em direção à célula vizinha com maior declive (SHIN; PAIK, 2017).

A partir da direção de escoamento do MDE, calculou-se o número de células que drenam para a célula considerada, podendo variar de zero, para os divisores de água, até valores mais elevados ao longo do curso d’água, aumentando em direção à foz (JIMME; BASHIR; ADEBAYO, 2016). O escoamento acumulado, obtido pela função “flow accumulation” indica o grau de confluência do escoamento, e, de acordo com Mendes; Cirilo (2001), representa a rede hidrográfica da bacia.

A delimitação das bacias hidrográficas pela função “watershed” foi realizada processando o mapa de direção de fluxo das etapas anteriores, sendo alocado o ponto do exutório da respectiva estação fluviométrica. O valor da área de cada bacia

hidrográfica gerada corresponde à quantidade de células que drenam até a célula que contém o exutório que representa a estação fluviométrica de interesse. Como cada célula do MDE obtido do TOPODATA possuía “pixels” de 30m, a área de cada “pixel” corresponde a 900m²; obtendo-se assim diretamente a área de drenagem de um ponto qualquer definido na superfície do MDE.

Posteriormente, calculadas as áreas à montante de cada exutório (estação fluviométrica), prosseguiu-se para o cálculo do perímetro (P em km) e centroide (X e Y em km) (BAENA *et al.*, 2004; ELESBON *et al.*, 2015; SAMUEL; COULIBALY; METCALFE, 2011). Quando a bacia de interesse ultrapassava o limite da cena, construía-se um mosaico com as mesmas, sendo necessário utilizar os recursos da janela Image Analysis do ArcGIS.

A declividade da bacia foi estimada pelo cálculo das declividades de todas as células situadas no interior de cada área de drenagem. Essa etapa foi automatizada em ambiente ArcGIS por meio do comando Slope, da ferramenta Spatial Analyst, utilizando como dados de entrada os limites de cada área de drenagem e as declividades geradas a partir do MDE, sem depressões espúrias. Posteriormente, a declividade foi classificada (Tabela 4), adaptando-se a divisão proposta pela Empresa Brasileira de Pesquisa Agropecuária (EMBRAPA, 1979), e seguidamente verificado o valor da declividade média.

Tabela 3 - Classificação da declividade segundo a EMBRAPA (1979)

Declividade (%)	Discriminação
0 - 3	Relevo plano
3 - 8	Relevo suavemente ondulado
8 - 20	Relevo ondulado
20 - 45	Relevo fortemente ondulado
45 - 75	Relevo montanhoso
> 75	Relevo fortemente montanhoso

Para o cálculo da precipitação total anual média (p em mm), foi utilizado o método dos polígonos de Thiessen (Equação 25), conforme descrito por Barbosa *et al.* (2005) e Cabral *et al.* (2016). A obtenção dessa informação também foi feita por meio do *software* ArcGIS® v.10.3, utilizando o comando Create Thiessen Polygons, formando inicialmente uma imagem com todas as estações pluviométricas e, posteriormente, por meio da ferramenta “clip”, extraídas as áreas dos polígonos referentes a cada bacia.

$$\bar{P} = \frac{\sum_{i=1}^n A_i \cdot P_i}{\sum_{i=1}^n A_i} \quad (25)$$

Em que:

$\bar{P}$ = precipitação média da bacia (mm);

n = número de postos pluviométricos;

A_i = área de influência do i-ésimo posto (km²); e

P_i = precipitação do i-ésimo posto de observação (mm).

3.4. Análise estatística multivariada

De posse de todas as variáveis explicativas foi montada uma matriz com todos os dados: variáveis independentes – área (A em km²), perímetro (P em km), centroide (X e Y em km), declividade média (D em %) e precipitação total anual média (p em mm); e variáveis dependentes – vazões de referência Q₉₀ e Q₉₅ (Q₉₀ e Q₉₅ em m³s⁻¹). Em seguida foi realizada a padronização dos dados originais das variáveis de acordo com a Equação 26, de modo a eliminar os efeitos de dependência das unidades e escalas em que as mesmas foram medidas.

$$Z = \frac{X - \bar{X}}{\sigma} \quad (26)$$

Em que,

Z = variável padronizada;

X = variável original;

$\bar{X}$ = média da variável original; e

σ = desvio padrão da variável.

Com a nova matriz de dados, foi verificado o ajuste a distribuição normal das variáveis independentes a partir do teste de Kolmogorov-Smirnov (KOLMOGOROV, 1933; SMIRNOV, 1948), e em seguida a matriz de correlação entre as variáveis independentes padronizadas, eliminando-se as que são menos correlacionadas/representativas (ELESBON *et al.*, 2015). Conforme Fávero *et al.* (2009), ao trabalhar com variáveis correlacionada indica-se a regressão método

Stepwise, retirando o efeito da multicolinearidade, ou ainda a realização da análise de componentes principais.

Em seguida, foram construídas as matrizes de dissimilaridade empregando as distâncias Euclidiana (HARRIS, 1955), Manhattan (SOKAL; MICHENER, 1957) e Mahalanobis (MAHALANOBIS, 1936), conforme as Equações (27, 28 e 29), juntamente com os diferentes algoritmos de agrupamento: ligação simples (FLOREK *et al.*, 1951), de ligação completa (LANCE; WILLIAMS, 1967) e Ward (WARD, 1963).

- Distância Euclidiana

$$\sqrt{\sum_{k=1}^n (X_{ik} - X_{jk})^2} \quad (27)$$

- Distância Manhattan

$$\sum_{k=1}^n |X_{ik} - X_{jk}| \quad (28)$$

- Distância Mahalanobis

$$\sqrt{(X_i - X_j)^T S^{-1} (X_i - X_j)} \quad (29)$$

Em que:

n = número de atributos;

X_{ik} = atributo k do vetor de características X_i ;

X_{jk} = atributo k do vetor de características X_j ;

T = matriz transposta; e

S^{-1} = matriz de covariância.

Para a aplicação do método não hierárquico, que não gera dendrogramas, foi utilizado os números de grupos dos métodos hierárquicos gerados pelos cortes dos dendrogramas, como suposições iniciais (valores de k), no algoritmo k-means, conforme sugerido por HSU; HUANG (2017).

A metodologia do k-means consistiu em dividir os vetores característicos em k grupos, minimizando a função objetiva da Equação 30 (HARTIGAN; WONG, 1979).

$$F = \sum_{i=1}^n \sum_{j=1}^K d(y_i; v_j) \quad (30)$$

A $d(y_i; v_j)$ representou a distância Euclidiana, ou seja, entre um ponto de dados y_i e o centro do cluster v_j , utilizando como indicador da distância dos n pontos dos dados e de seus respectivos centros. A rotina de trabalho relativo à análise estatística multivariada de agrupamento pelo método hierárquico e não hierárquico foi implementada no programa R, versão 3.3.3, com o auxílio do pacote “Biotools”, para o cálculo da distância de Mahalanobis (SILVA, 2017).

Para a definição das regiões homogêneas pelo método hierárquico, partiu-se do critério de inércia entre saltos, em que a primeira descontinuidade visível do gráfico defini o ponto de corte, ou seja, observou-se a melhor separação dos grupos no dendrograma (ELESBON *et al.*, 2015; MELO JR. *et al.*, 2006; RENCHER, 2002; WILKS, 2006). Posteriormente foi empregado a medida de heterogeneidade (H) de Hosking; Wallis (1997) para avaliar os grupos formados, e assim verificar se os mesmos são heterogêneos (SHE *et al.*, 2016).

3.5. Medida de heterogeneidade

O fundamento desta estatística, denominada H, é verificar o grau de heterogeneidade de uma região, de forma que uma dada região contenha n postos de observação, cada um deles indexado por i , com amostra de tamanho n_i e quocientes de momentos-L amostrais, representados por t^1 , t^3 e t^4 . Foi considerado também que t^R , t^{R_3} e t^{R_4} denotam, respectivamente, as médias regionais dos quocientes coeficiente de variação-L (CV-L), assimetria-L e curtose-L, conforme Equação 31.

$$t^R = \frac{\sum_{i=1}^N n_i t^i}{\sum_{i=1}^N n_i} \quad (31)$$

Hosking; Wallis (1997) recomendaram que essa estatística seja baseada, preferencialmente, no cálculo da dispersão de t , ou seja, o CV-L para as regiões propostas e simuladas. Inicialmente foi calculado o desvio padrão ponderado V , dos CV-L's das amostras observadas pela Equação 32.

$$V = \left\{ \frac{\sum_{i=1}^N n_i (t^i - t^R)^2}{\sum_{i=1}^N n_i} \right\}^{1/2} \quad (32)$$

Em seguida, para a simulação da região homogênea, Hosking; Wallis (1997) sugerem a utilização da distribuição Kappa. Os parâmetros da distribuição são estimados de modo a reproduzir os quocientes de momentos-L regionais (t^R , t^{R_3} , t^{R_4}). Com os parâmetros populacionais, foram simulados N_{SIM} regiões homogêneas, sem correlação cruzada e/ou serial, contendo N amostras individuais, sendo cada n_i valores da variável normalizada. As estatísticas V_i ($i=1, 2, \dots, N_{SIM}$) foram calculadas pela Equação (32), até que o número de simulação seja igual a 500 (HOSKING; WALLIS, 1997).

A média aritmética das estatísticas V_i , calculadas para cada simulação, forneceu a dispersão média esperada para a região homogênea (Equação 33).

$$\mu_V = \frac{\sum_{i=1}^{N_{SIM}} V_i}{N_{SIM}} \quad (33)$$

A medida de heterogeneidade H foi obtida comparando-se a dispersão observada e a dispersão simulada (Equação 34).

$$H = \frac{(V - \mu_V)}{\sigma_V} \quad (34)$$

Desta forma, V representou a estatística calculada pela Equação (32) utilizando os dados observados na região supostamente homogênea; μ_V é a média aritmética das estatísticas V_i calculada para cada simulação; e σ_V é o desvio padrão entre os N_{SIM} valores da medida de dispersão V_i (Equação 35).

$$\sigma_V = \sqrt{\frac{\sum_{i=1}^{N_{SIM}} (V_i - \mu_V)^2}{N_{SIM} - 1}} \quad (35)$$

De acordo com o proposto por Hosking; Wallis (1997), se $H < 1$, a região é definida como “aceitavelmente homogênea”, se $1 \leq H < 2$, a região é “possivelmente heterogênea” e, finalmente, se $H \geq 2$, a região deve ser classificada como “definitivamente heterogênea”. A rotina de trabalho relativo à medida de heterogeneidade H foi implementada no programa R, versão 3.3.3, com o auxílio dos pacotes Lmom (HOSKING, 2017a), e LmomRFA (HOSKING, 2017b).

3.6. Equações de regionalização

Entre as variáveis independentes (entrada) podem haver variáveis que pouco influenciam o conjunto de variáveis dependentes (saída). Desta forma, o método Stepwise (AISSIA; CHEBANA; OUARDA, 2017; BOOKER; SNELDER, 2012; MOHAMOUD, 2008) é frequentemente utilizado para selecionar quais variáveis são mais influentes no conjunto de saída, podendo, assim, diminuir o número de variáveis na equação de regressão. O método caracteriza-se por ser iterativo, adicionando (passo forward) e removendo variáveis (passo backward), a partir de um critério de seleção, como, por exemplo, o teste F, coeficiente de correlação linear múltipla ou o erro quadrático total. Neste trabalho optou-se por utilizar o método Stepwise backward com significância de 5% ($p < 0,05$) com o critério de seleção de variáveis sendo o teste F. A rotina de trabalho relativo à formulação das equações de regressão foram implementadas no programa R, versão 3.3.3.

O modelo que apresentou os melhores resultados em equações de regionalização foi o potencial, selecionando-se o mesmo para a presente dissertação, segundo Lopes *et al.* (2016); Melati *et al.* (2017); Moreira; Silva (2017); Pruski *et al.* (2011; 2012; 2013 e 2015); Samuel *et al.* (2011) e Uliana *et al.* (2017). Desta forma, paralelamente à geração das equações potenciais de regionalização pelo método Stepwise, ocorre a linearização das equações potenciais, transformação mais utilizada em regionalização de vazões mínimas. A Equação 36 foi transformada para a base logarítmica, obtendo assim uma equação linear (Equações 37), conforme indicação de Lujano *et al.* (2017).

$$y = ax^b \quad (36)$$

$$\ln y = \ln a + b \ln x \quad (37)$$

Alterando as variáveis de forma que $z = \ln y$, $k = \ln a$ e $v = \ln x$, obtém-se a Equação 38 da reta.

$$z = k + bv \quad (38)$$

Considerando as variáveis linearizadas, foi possível estimar os parâmetros k e v . Posteriormente, calculando o antilogaritmo de k , foi estimado o parâmetro a da Equação 36, sendo b o expoente de x .

Para avaliar os modelos, foi calculado o coeficiente de determinação R^2 (Equação 7), pois refere-se a uma quantidade que dá a qualidade de um ajuste de mínimos quadrados aos dados originais, e, portanto, pode ser incluído como um primeiro cálculo de confiabilidade do modelo, segundo Portals *et al.* (2010). Em virtude de se trabalhar com mais variáveis, também foi calculado o coeficiente de determinação ajustado (R^2_a) (Equação 8) (ELESBON *et al.*, 2015). Os modelos ainda foram analisados pelo logaritmo do coeficiente de eficiência de Nash-Sutcliffe ($\log C_{NS}$) (Equação 10), (He *et al.*, 2014; Tsakiris; Nalbantis; Cavadias 2011), seguindo a classificação de Motovilov *et al.* (1999), e o erro médio absoluto (EMA) (Equação 12), também foi calculado por ser menos sensível aos valores extremos (PORTALS *et al.*, 2010). Seguidamente, foi utilizado o índice de desempenho “ c ”, (Equação 13) proposto por Camargo; Sentelhas (1997), visando quantificar a performance do modelo para as diferentes regiões homogêneas (Alves; Biudes; Vecchia, 2012; Beskow *et al.*, 2016; Wolff; Duarte; Mingoti, 2014).

4 RESULTADOS E DISCUSSÃO

4.1. Análise preliminar dos dados levantados

Na Figura 3 estão apresentadas a localização das estações pluviométricas e fluviométricas que atenderam aos critérios estabelecidos, visto que em consulta ao HidroWeb foram encontradas 1109 estações pluviométricas, mas apenas 365 continham série histórica, com extensão igual ou superior a 10 anos, sem falhas.

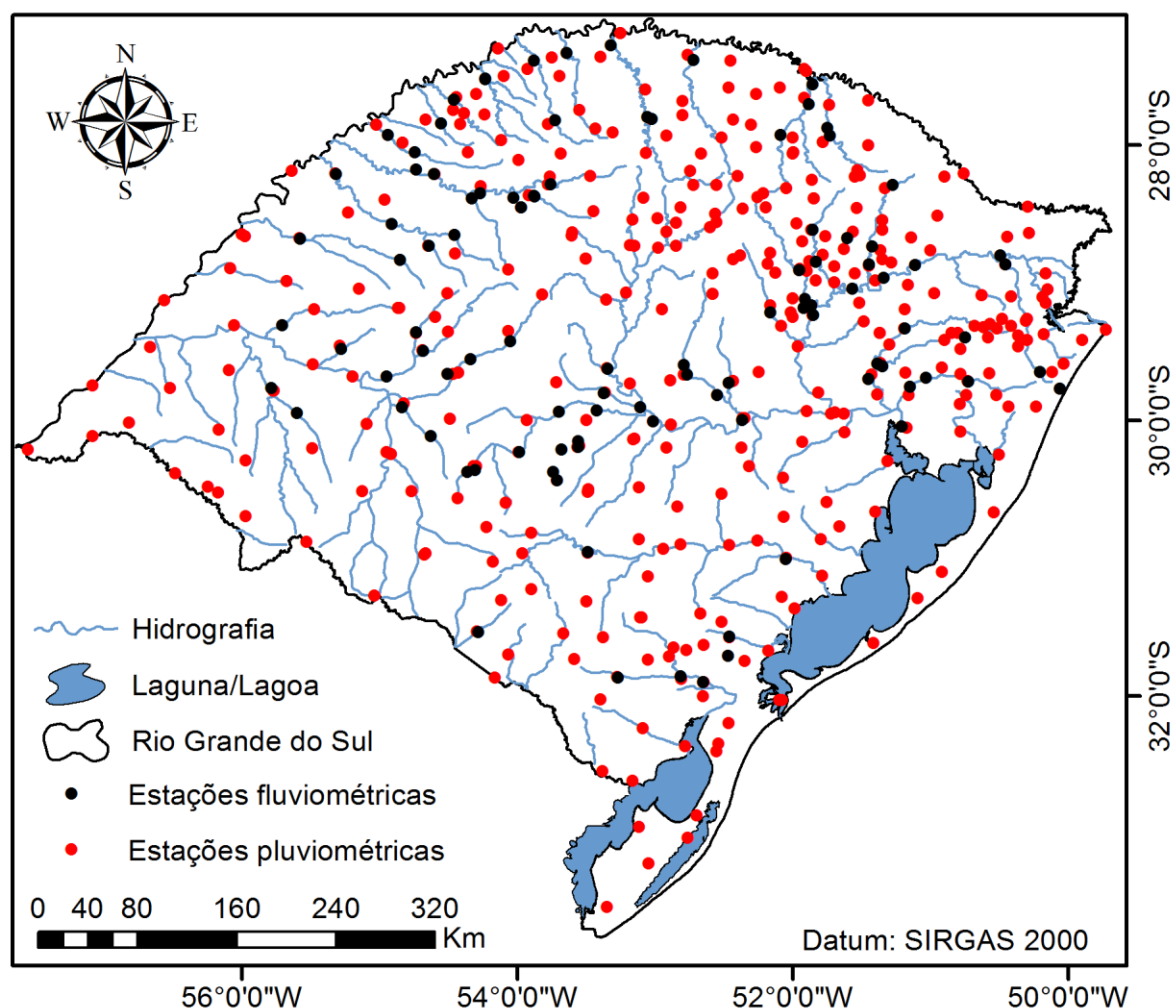


Figura 3 – Localização das estações pluviométricas e fluviométricas situadas no estado do Rio Grande do Sul e que atenderam aos critérios pré-estabelecidos.

As respectivas séries foram aceitas de acordo com o teste de Mann-Kendall e Mann-Whitney, com nível de 5% de significância. Já para as 305 estações fluviométricas encontradas no banco de dados do Hidroweb/ANA, apenas 170

possuem dados e somente 120, com série histórica igual ou superior a 10 anos e com no máximo 31 dias de falhas. Aplicando o teste de Mann-Kendall houve a necessidade de excluir 3 estações fluviométricas (72530000, 85830000 e 87865000), enquanto as demais também foram aceitas no teste Mann-Whitney. Para Moreira; Naghettini (2016), a estacionariedade é uma das premissas fundamentais, isto é, os registros amostrais representam eventos de uma população, cujas características distributivas são invariantes no tempo. Desta forma, identificando as séries com não estacionariedade garantir-se-á maior acurácia nas análises subsequentes.

O MDE gerado foi usado para a delimitação das áreas de drenagem das 117 estações fluviométricas apresentadas na Figura 4.

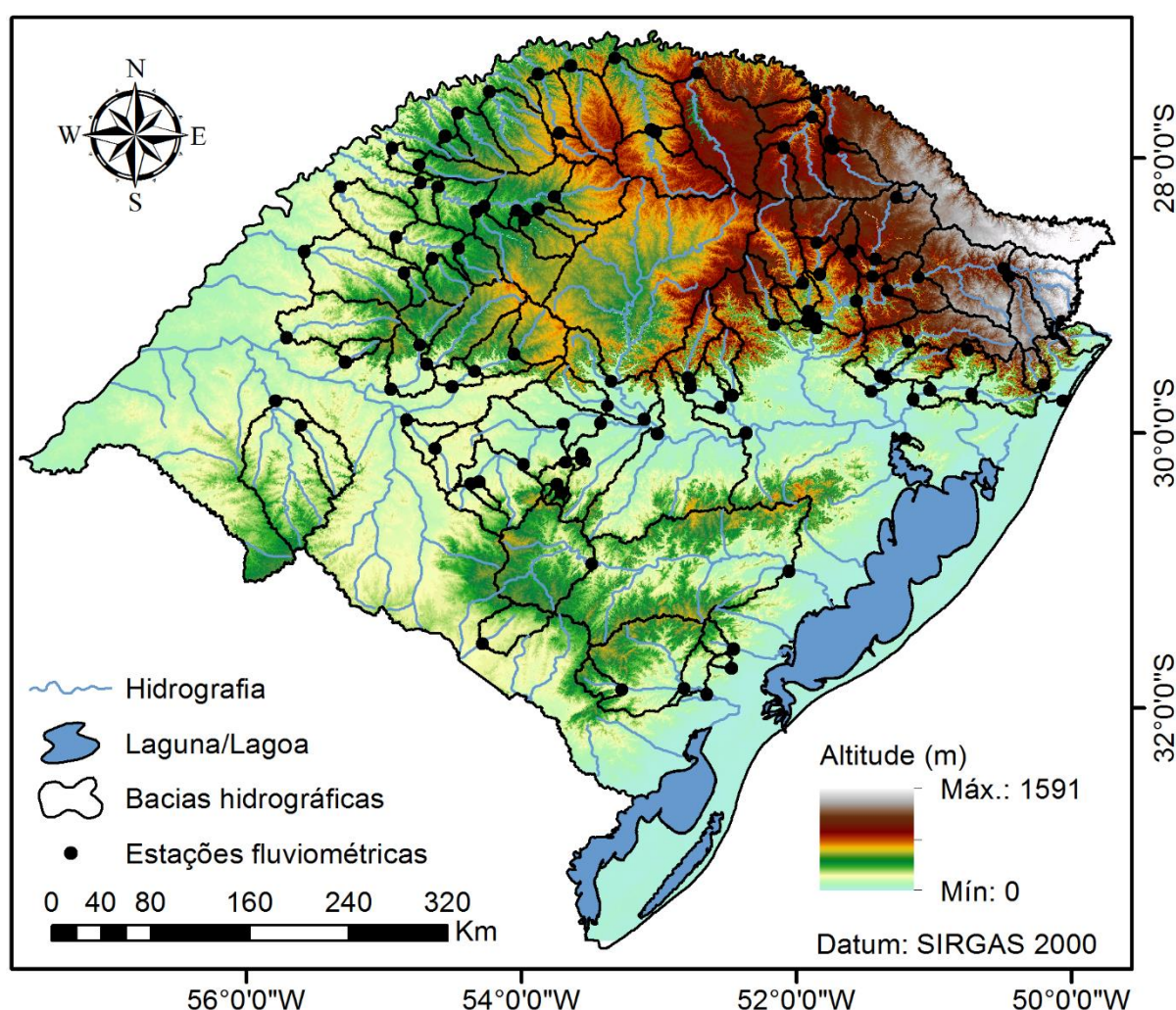


Figura 4 - Modelo digital de elevação do Rio Grande do Sul com destaque para as bacias hidrográficas em estudo.

Contudo, foi necessário eliminar mais 17 estações, cujas áreas de contribuição ultrapassaram os limites do Estado, restando 100 bacias hidrográficas no

levantamento das demais características, para a análise de agrupamento. Observa-se na Figura 4 que existe uma maior defasagem no número de estações fluviométricas na porção sudoeste e sudeste (litoral) do Estado em relação às demais regiões. Beskow *et al.* (2016), em estudo de regionalização da vazão Q_{90} no estado do Rio Grande do Sul utilizaram 78 estações, ressaltando-se assim a importância da atualização das séries, e também a aplicabilidade de testes de estacionariedade. Para o levantamento das características das bacias, primeiramente foi elaborado o mapa dos polígonos de Thiessen (Figura 5), com todas as 365 estações pluviométricas, visando o cálculo da precipitação total anual média (p em mm).

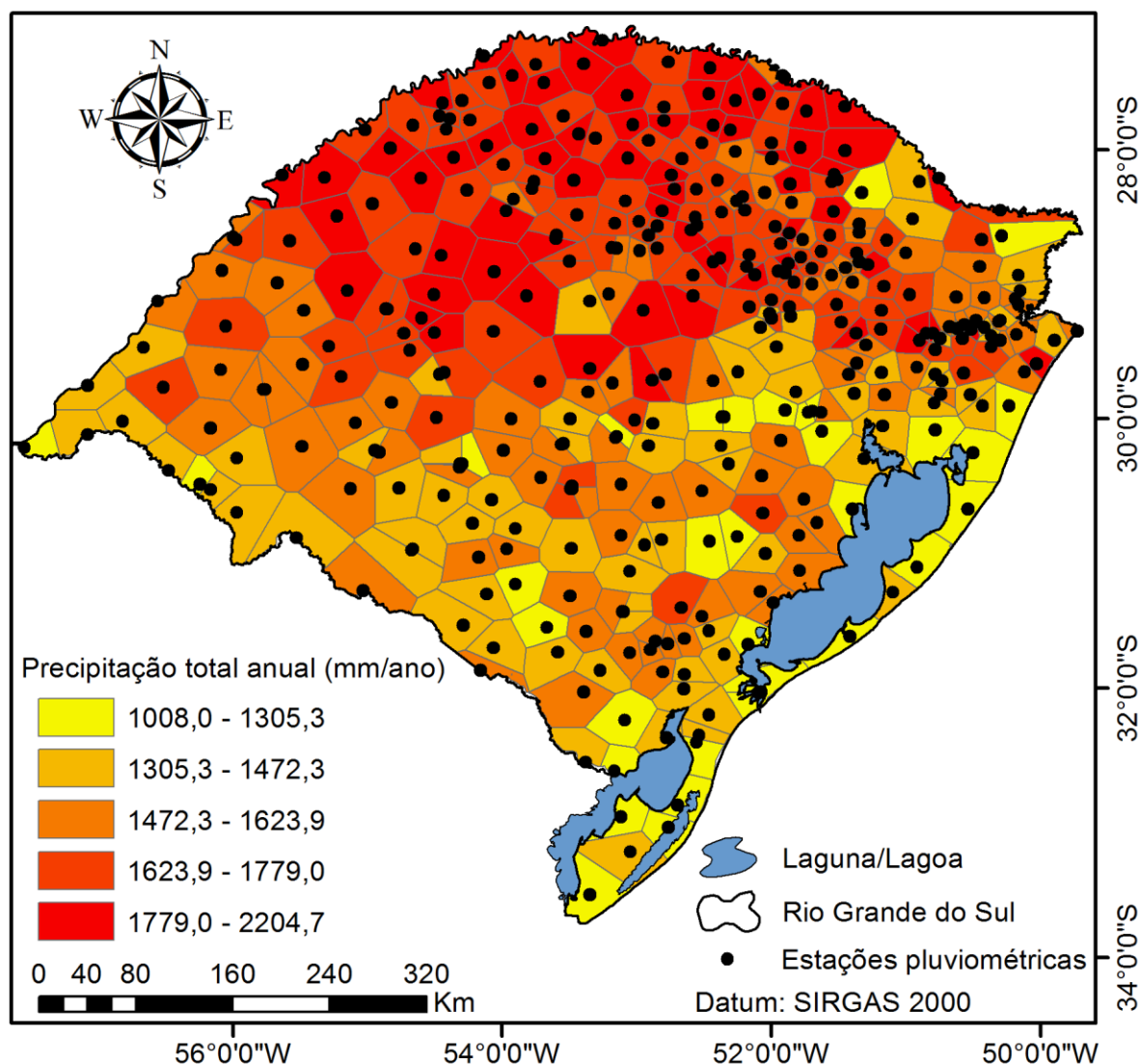


Figura 5 – Variabilidade espacial da precipitação total anual média no Rio Grande do Sul, de acordo com a metodologia dos polígonos de Thiessen.

Na Tabela 5 estão compiladas as informações de todas as variáveis levantadas das 100 bacias, juntamente com os códigos das estações fluviométricas que as caracterizam. Na sequência, os valores correspondentes a cada bacia foram extraídos, por conseguinte, a que representou a menor precipitação média anual (86340000), com 1172,1mm, localiza-se a nordeste do Estado, e a maior (75295000), com 1914,7mm, localiza-se na porção noroeste do estado. Posteriormente, foram determinados os seguintes atributos para cada uma das 100 bacias hidrográficas analisadas: área de drenagem (A em km²), perímetro (P em km), centroide (X, latitude e Y, longitude em km) e declividade (D em %). As bacias 85610000 e 85900000 apresentaram a menor (68,6km²) e a maior (38096,2km²) área de drenagem, respectivamente, ambas localizadas na parte central do estado.

Na sequência, foram elaboradas a curvas de permanência e extraídos os valores dos quantis 90 e 95 que caracterizam as vazões Q₉₀ e Q₉₅, em m³s⁻¹, respectivamente. As bacias 85615000 e 85642000 caracterizam-se por apresentar as menores e as maiores vazões de permanência, respectivamente, estando localizadas também na parte central do estado.

Tabela 4 - (continua) Variáveis utilizadas no estudo para cada uma das 100 bacias hidrográficas

n	Código	A (km ²)	P (km)	Centroide (km)		D (%)	p (mm)	Q (m ³ s ⁻¹)	
				X	Y			Q ₉₀	Q ₉₅
1	72400000	1573,6	206,6	442,2	6888,5	9,5	1814,7	3,32	2,31
2	72430000	1737,8	218,2	441,5	6890,7	9,9	1822,2	7,60	5,38
3	72580000	1159,7	170,3	387,9	6889,3	9,7	1604,3	6,24	5,15
4	72630000	2993,3	298,8	398,8	6898,6	13,2	1720,7	13,00	9,64
5	72680000	3907,7	354,6	398,3	6907,9	14,2	1744,9	18,60	14,00
6	73480000	3965,4	370,8	347,7	6922,3	10,3	1764,5	23,50	17,33
7	74205000	466,2	106,6	314,5	6915,4	11,6	1727,6	0,49	0,16
8	74210000	2746,3	327,9	309,8	6892,5	9,0	1814,5	17,00	13,86
9	74270000	5708,8	468,8	301,2	6918,6	12,2	1841,1	29,26	21,65
10	74370000	2153,9	257,9	255,1	6937,0	10,4	1774,9	21,50	15,00
11	74460000	536,0	126,4	241,2	6906,9	7,1	1874,4	4,58	3,74
12	74470000	1639,2	260,0	232,4	6930,9	9,6	1830,2	9,97	7,87
13	74600000	2407,7	245,6	793,9	6920,9	9,4	1779,9	8,89	6,95
14	74610000	2407,6	245,3	793,9	6920,9	9,4	1779,9	18,96	17,59
15	74700000	1216,9	196,7	767,8	6911,9	7,3	1807,7	7,46	5,62
16	74750000	349,7	96,4	749,0	6907,2	7,3	1830,9	1,73	1,30
17	74880000	878,2	181,9	749,1	6890,4	6,9	1897,5	3,82	2,74
18	74900000	1346,7	253,0	736,1	6893,5	7,2	1891,4	4,80	3,61
19	75155000	2080,6	242,6	257,4	6872,7	7,9	1845,2	15,38	12,52
20	75185000	655,2	156,6	238,7	6846,2	6,9	1707,8	4,80	4,27
21	75186000	655,9	156,7	238,7	6846,2	6,9	1707,7	4,80	4,27
22	75200000	867,0	156,7	225,2	6833,9	6,6	1859,7	5,24	4,13
23	75205000	1039,2	184,8	221,6	6836,4	6,5	1859,3	7,13	5,50
24	75230000	5833,6	430,4	232,4	6863,1	7,3	1792,2	38,00	27,90
25	75295000	2321,6	312,8	791,9	6817,9	5,9	1914,7	12,17	8,92
26	75300000	9573,0	589,5	803,5	6851,5	7,0	1822,7	42,00	33,20
27	75320000	10153,8	633,7	799,1	6852,3	6,9	1822,8	50,00	37,50

Tabela 5 – *continuação*

n	Código	A (km²)	P (km)	Centroide (km)		D (%)	p (mm)	Q (m³s⁻¹)	
				X	Y			Q ₉₀	Q ₉₅
28	75400000	1003,5	155,3	762,5	6816,3	6,1	1886,8	3,20	2,45
29	75430000	957,6	155,8	739,9	6799,3	5,7	1824,8	2,96	2,34
30	75450000	3474,6	317,1	741,0	6817,7	5,7	1805,6	10,33	7,28
31	75500000	5626,0	478,4	720,8	6832,2	5,4	1795,8	17,12	12,96
32	75600000	406,0	109,1	719,0	6796,7	5,8	1765,4	0,72	0,47
33	75700000	3995,9	363,1	680,8	6809,0	4,7	1807,4	4,14	2,44
34	76085000	1668,0	212,1	803,3	6755,8	6,6	1763,0	2,99	1,64
35	76100000	3028,5	299,1	790,1	6750,4	9,4	1739,5	5,76	2,99
36	76120000	3590,6	375,8	783,1	6746,4	9,3	1723,0	4,77	3,48
37	76360001	1260,5	235,8	742,4	6660,0	4,9	1554,4	0,97	0,59
38	76380000	1986,7	301,1	735,3	6665,7	5,0	1564,9	1,32	0,81
39	76440000	2514,1	298,8	758,2	6767,7	7,7	1878,4	6,20	3,40
40	76460000	996,3	166,5	725,1	6770,6	7,6	1816,7	2,60	1,61
41	76490000	4965,2	447,9	737,9	6758,5	7,9	1824,6	8,87	5,98
42	76550000	408,2	113,3	675,4	6747,4	7,8	1651,7	5,15	4,55
43	76650000	2774,9	323,2	663,2	6770,1	6,6	1711,8	21,85	19,88
44	76742000	1273,4	221,7	653,7	6660,7	7,8	1531,4	1,19	0,40
45	76750000	6503,8	457,8	629,8	6650,3	6,4	1482,8	5,94	4,22
46	79400000	1367,0	206,7	778,5	6521,3	4,0	1361,4	0,17	0,07
47	85400000	15109,2	767,7	299,2	6808,2	9,3	1693,6	149,00	116,6
48	85438000	994,2	175,4	249,2	6712,1	10,9	1699,6	0,98	0,41
49	85440000	18418,1	881,6	293,1	6792,2	9,7	1689,2	92,93	83,45
50	85460000	816,0	191,2	182,3	6614,7	5,3	1376,9	0,15	0,08
51	85470000	1058,1	201,2	183,0	6617,5	5,2	1408,0	1,71	1,25
52	85480000	3234,3	343,0	193,9	6631,0	5,7	1426,9	3,69	2,23
53	85580000	5644,7	497,0	201,3	6649,3	5,1	1472,9	3,98	1,93
54	85600000	7891,9	579,2	209,1	6660,0	5,1	1502,1	6,49	1,08
55	85610000	68,6	40,5	238,3	6623,8	9,1	1606,7	0,07	0,05
56	85615000	74,8	47,6	230,2	6631,5	8,1	1606,9	0,05	0,04
57	85620000	653,8	135,7	235,7	6636,7	7,2	1599,8	0,56	0,43
58	85623000	756,1	157,5	237,3	6639,0	7,4	1596,9	1,86	1,54
59	85630000	804,6	175,8	238,4	6640,1	7,4	1589,7	0,62	0,22
60	85642000	29600,6	1349,3	268,4	6741,8	8,1	1625,3	226,42	168,69
61	85730000	1329,0	252,1	332,2	6755,5	22,8	1813,0	1,67	1,14
62	85740000	1475,7	264,4	331,3	6752,0	22,3	1807,0	4,07	2,40
63	85780000	2248,4	347,7	334,1	6739,4	18,4	1767,4	2,27	1,52
64	85850000	903,0	193,1	352,5	6740,6	25,7	1685,9	1,44	1,06
65	85900000	38096,2	1533,7	278,7	6721,3	8,0	1599,4	172,2	106,35
66	86100000	1931,2	273,3	580,2	6802,7	16,2	1424,1	10,04	7,84
67	86160000	1202,5	228,0	564,6	6775,8	9,5	1509,9	3,57	2,20
68	86250000	6996,8	551,1	546,9	6801,5	15,3	1526,5	10,10	6,50
69	86300000	8292,3	614,9	537,5	6801,2	16,1	1550,8	13,00	8,50
70	86340000	173,1	71,2	480,6	6874,8	6,3	1172,1	1,05	0,82
71	86410000	2989,2	302,4	471,8	6850,9	9,9	1485,4	9,04	5,94
72	86420000	337,1	97,8	439,5	6840,6	9,1	1695,7	1,35	0,87
73	86440000	3867,8	344,2	466,7	6845,7	11,1	1525,8	13,72	9,28
74	86470000	13472,3	822,3	509,8	6812,4	15,4	1550,8	49,66	37,2
75	86480000	1375,2	213,5	418,8	6854,4	15,2	1660,4	2,86	2,05
76	86500000	1954,8	259,4	418,2	6845,7	16,2	1680,5	13,72	9,28
77	86510000	17291,3	928,7	491,1	6813,9	16,2	1576,9	53,77	36,82
78	86560000	2185,9	284,4	388,6	6834,0	13,9	1731,4	7,34	5,36
79	86580000	2643,9	336,3	392,1	6826,8	18,2	1725,5	7,34	5,36
80	86700000	471,8	120,9	395,8	6782,6	26,0	1708,4	0,81	0,54
81	86720000	20617,2	1027,0	475,4	6814,4	16,8	1598,9	61,3	45,40
82	86745000	852,7	197,5	377,1	6798,0	22,3	1753,7	1,50	0,75
83	87160000	2192,3	333,6	516,5	6762,1	17,8	1709,0	5,07	3,27

Tabela 5 - *continuação*

n	Código	A (km ²)	P (km)	Centroide (km)		D (%)	p (mm)	Q (m ³ s ⁻¹)	
				X	Y			Q ₉₀	Q ₉₅
84	87170000	3263,3	438,7	500,1	6757,7	19,8	1702,7	8,57	6,18
85	87250000	942,6	168,6	491,2	6735,5	21,1	1596,6	2,53	1,34
86	87270000	4731,3	498,2	493,0	6750,0	19,3	1662,1	16,01	11,88
87	87317030	478,6	126,7	570,6	6731,4	31,9	1623,1	1,22	0,90
88	87317060	1866,8	280,6	584,5	6732,8	22,4	1614,8	21,30	16,90
89	87366000	111,0	58,6	527,1	6751,0	26,4	1583,3	0,19	0,14
90	87374000	1581,1	265,6	551,8	6718,6	22,7	1522,5	5,35	4,34
91	87380000	3053,3	398,9	536,5	6723,9	21,2	1548,8	17,05	13,46
92	87382000	3397,9	433,6	532,1	6722,6	20,1	1540,6	20,33	15,97
93	87450100	70,7	46,8	486,0	6673,1	10,4	1402,0	0,45	0,36
94	87590000	4490,1	371,5	228,7	6576,4	7,9	1432,1	3,43	2,15
95	87905000	17144,2	921,7	301,4	6572,5	10,0	1455,2	37,6	27,45
96	88575000	594,6	133,6	267,8	6478,2	9,2	1436,2	1,64	1,29
97	88641000	5208,0	381,9	298,0	6492,7	9,3	1505,7	12,10	10,51
98	88680000	5960,9	421,7	302,7	6491,9	9,0	1503,6	8,57	6,60
99	88750000	143,9	56,2	353,7	6495,8	9,0	1466,4	0,13	0,05
100	88850000	409,5	109,1	348,0	6513,6	10,9	1567,2	0,88	0,54

Pelo teste de Kolmogorov-Smirnov, as variáveis área, perímetro e declividade não seguem a distribuição normal, desta forma prosseguiu-se calculando a matriz de correlação de Spearman (Tabela 6) entre as variáveis independentes. De acordo com Elesbon *et al.* (2015), desta forma é possível avaliar a importância de cada uma das variáveis, promovendo a eliminação daquelas que contribuem pouco, em termos de variação, no grupo de indivíduos avaliados na análise de regionalização de vazões.

Tabela 5 - Matriz de correlação de Spearman entre as variáveis independentes analisadas

Variável	A	P	X	Y	D	p
A	1,00	0,98**	0,90	0,08	0,05	-0,03
P	-	1,00	0,07	0,02	0,10	-0,07
X	-	-	1,00	0,20*	-0,05	0,25*
Y	-	-	-	1,00	0,06	0,66**
D	-	-	-	-	1,00	-0,10
p	-	-	-	-	-	1,00

** significância a 0,01 e * significância a 0,05

Verifica-se na Tabela 6 que as variáveis morfométricas área e perímetro são fortemente correlacionadas entre si. Já a variável precipitação total anual média é fortemente correlacionada com a longitude e moderada relação com a latitude, enquanto a longitude também possui moderada correlação com a latitude. A variável morfométrica declividade não possui correlação com as demais variáveis analisadas, indicando a possibilidade de exclusão da mesma do estudo. Os resultados da matriz de correlação corroboram com os obtidos por Elesbon *et al.* (2015), que também observaram a não correlação da variável declividade com as demais variáveis

independentes, partindo para a exclusão da mesma no estudo de regionalização de vazões mínimas.

4.2. Análise de agrupamento

Eliminando a variável declividade, reformulou-se a matriz de dados padronizada com as variáveis independentes restantes (A, P, X, Y e p) e dependentes (Q_{90} e Q_{95}). Contudo, vale lembrar que na análise de agrupamento não há distinção entre as variáveis independentes e dependentes, segundo Kumar; Singh; Mishra (2013), desta forma, primeiramente busca-se a relação das variáveis (A, P, X, Y e p) com a Q_{90} e seguidamente com a Q_{95} . Realizando-se assim, a análise de agrupamento pelo método hierárquico com os algoritmos de ligação simples, ligação completa e Ward, e as distâncias Euclidiana, Mahalanobis e Manhattan para as duas vazões de referência.

Patidar; Verma (2017) destacam que cada um desses algoritmos de ligação produz resultados diferentes quando usado no mesmo conjunto de dados, ou seja, cada um tem suas propriedades específicas. Os métodos hierárquicos têm como característica apresentar os resultados na forma de dendrogramas. Melo *et al.* (2006) salientam que a determinação do número de grupos para os dendrogramas pode ser feita de forma visual, e diferentes alternativas de cortes podem ser avaliadas.

Os dendrogramas apresentados na sequência foram derivados da aplicação dos algoritmos de ligação simples (Figura 6), ligação completa (Figura 7) e Ward (Figura 8) de agrupamento com as variáveis independentes (A, P, X, Y e p) para a vazão Q_{90} , combinados com as distâncias Euclidiana, Mahalanobis e Manhattan.

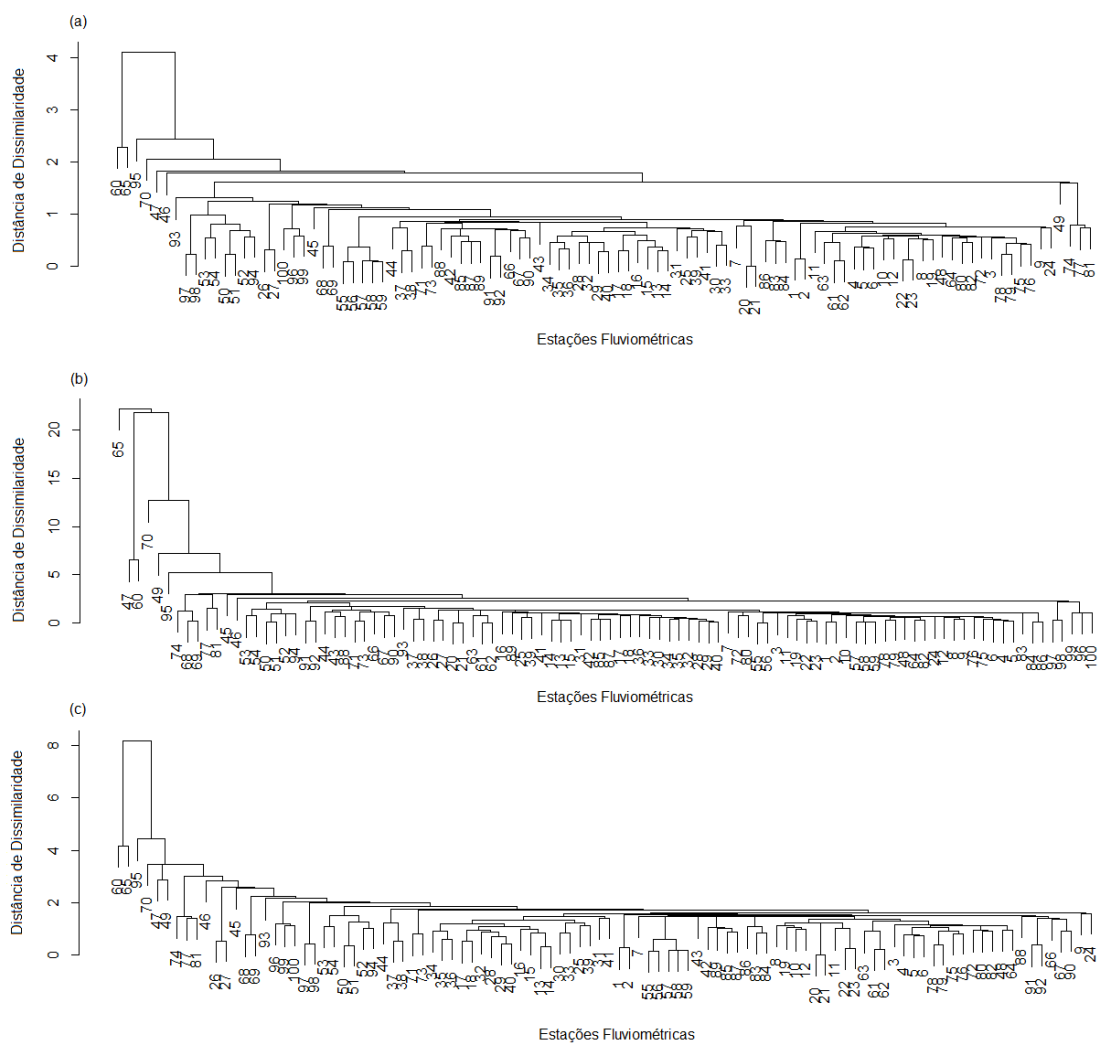


Figura 6 – Dendrogramas obtidos a partir da combinação do algoritmo de ligação simples de agrupamento com as variáveis independentes para a Q_{90} com as distâncias: (a) Euclidiana, (b) Mahalanobis e (c) Manhattan.

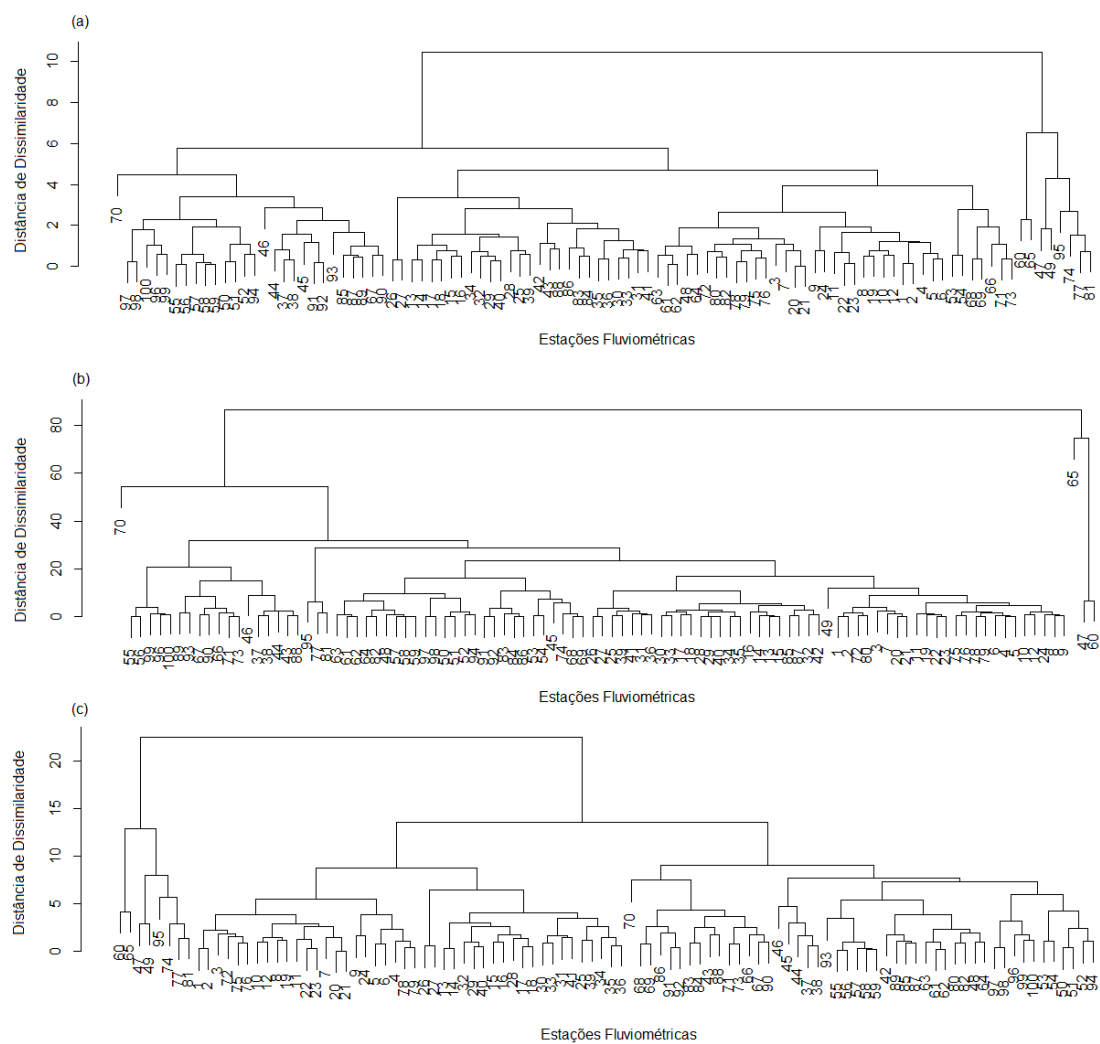


Figura 7 - Dendrogramas obtidos a partir da combinação do algoritmo de ligação completa de agrupamento com as variáveis independentes para a Q_{90} com as distâncias: (a) Euclidiana, (b) Mahalanobis e (c) Manhattan.

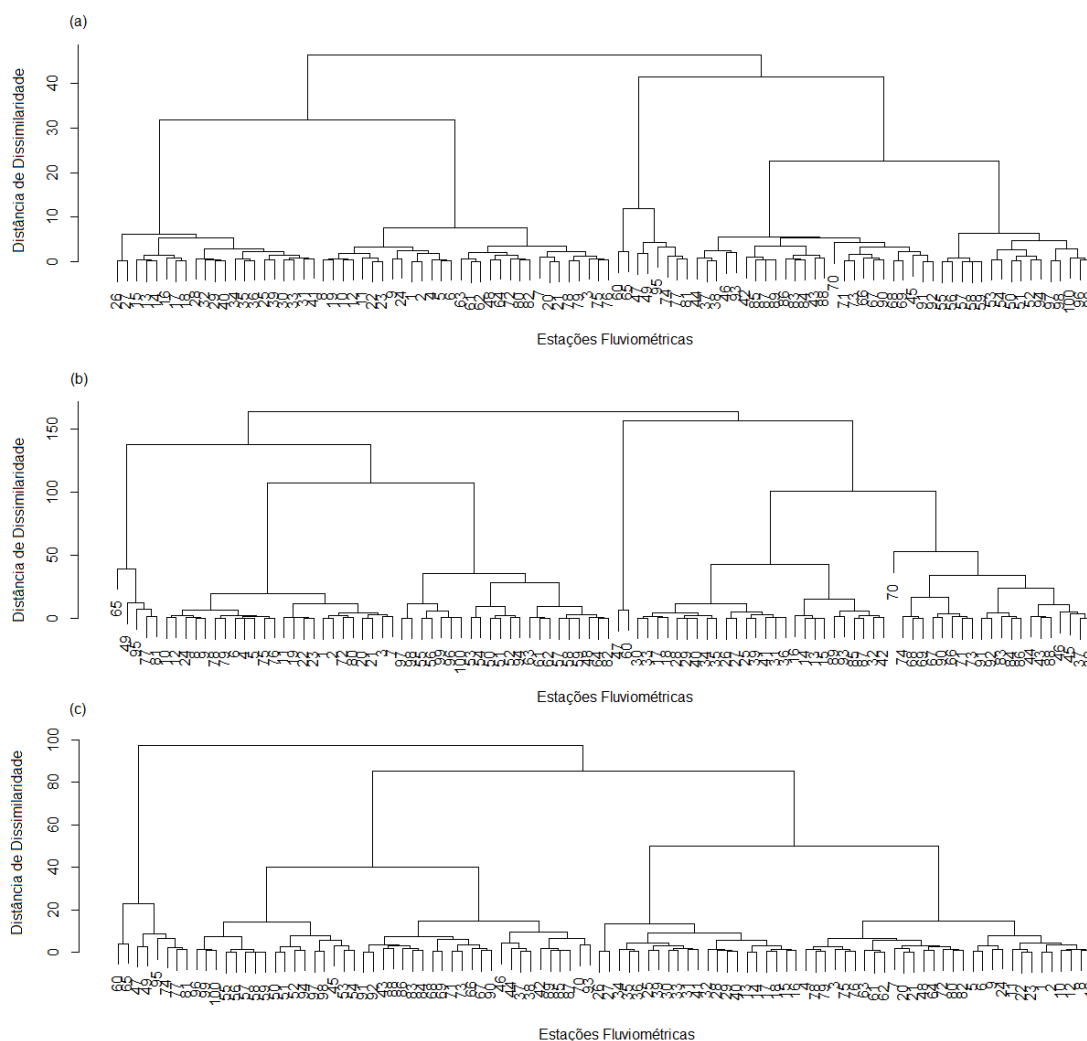


Figura 8 – Dendrogramas obtidos a partir da combinação do algoritmo de Ward de agrupamento com as variáveis independentes para a Q_{90} com as distâncias: (a) Euclidiana, (b) Mahalanobis e (c) Manhattan.

Como observado na Figura 6, as distâncias Euclidiana e Manhattan não proporcionaram a formação de grupos pelo algoritmo de ligação simples, sugerindo apenas que fosse formado um grupo com as estações 60 (85642000) e 65 (85900000), as quais caracterizam-se por apresentarem as maiores áreas e vazões mínimas, e o restante das estações formando um grande grupo. Já a distância de Mahalanobis com este algoritmo sugere a formação ainda de grupos com apenas uma estação. Para os autores Bivand; Wilk; Kossowski, (2017) e Melo, *et al.* (2017) o ideal é que os grupos formados apresentem números mais homogêneos de estações, ou seja, um agrupamento com uma ou duas estações acaba caracterizando um *outlier*. Desta forma, os cortes nos dendrogramas foram realizados com o objetivo de evitar este fato.

Em estudo de regionalização realizado por Rao; Srinivas (2006), os grupos obtidos pelo algoritmo de ligação simples consistiram em um grande grupo e vários pequenos ramos, indicando que o algoritmo não foi adequado para regionalização de vazões. Melo *et al.* (2006) também observaram que este método apresentou agrupamentos irregulares e foi descartado pelos autores no estudo de chuvas intensas no estado de Minas Gerais.

Observa-se na Figura 7 que o algoritmo de ligação completa representa melhor os grupos comparativamente àqueles obtidos com auxílio do algoritmo de ligação simples. Na aplicação da distância Euclidiana, houve a formação de 2 agrupamentos; a distância de Mahalanobis não respondeu adequadamente; e, no caso de uso da distância de Manhattan, houve possibilidade de formação de 2 e 3 agrupamentos, conforme o ajuste do corte do dendrograma.

No estudo de Elesbon *et al.* (2015), o algoritmo de ligação completa foi o que apresentou os melhores resultados, juntamente com a distância de Mahalanobis que pode ser ajustada à distância Euclidiana. Segundo os autores, quando a matriz de covariância for análoga à matriz identidade, pode ser realizada a referida simplificação. No presente estudo não foi possível realizar esta simplificação uma vez que as matrizes continham resultados diferentes. De forma geral, pode-se observar que as distâncias Euclidiana e Manhattan produzem maior semelhança quanto as possibilidades de cortes dos dendrogramas. Na literatura ainda é possível observar o maior emprego da distância Euclidiana, comparativamente às demais distâncias métricas (PEREIRA; KOBIYAMA; CASTRO, 2014).

Segundo Patidar; Verma (2017), o método de agrupamento de ligação completa geralmente produz grupos mais compactos e hierarquias mais úteis do que o método de agrupamento de ligação simples. Os autores, comparando os algoritmos de ligação simples e completa, observaram que à medida que cresce o número de objetos, aumenta a precisão dos algoritmos e após atingir 100 objetos diminui novamente, entretanto o algoritmo de ligação completa supera o de ligação simples.

É possível observar que o método de Ward na Figura 8 proporcionou um maior número de possibilidades para o corte do dendrograma. Pelas distâncias Euclidiana e Manhattan, o dendrograma pode ser cortado em diferentes alturas gerando 2, 3, 4 e 5 possibilidades de grupos, já pela distância de Mahalanobis somente 2 grupos puderam ser identificados. Rao; Srinivas (2006) observaram em seu estudo que o algoritmo de Ward apresentou os grupos de forma mais distinta. De acordo com YANG

et al. (2010), o algoritmo de Ward geralmente tende a produzir grupos pouco extensos e de igual número de elementos, o que de certa forma, concorda com os resultados obtidos, visto que foi o algoritmo que possibilitou o maior número de grupos, dentre os algoritmos analisados. Hassan; Ping (2012) e Melo *et al.* (2006) também constataram que o algoritmo de Ward forneceu os melhores resultados comparando com outros algoritmos de classificação. Nas Figuras 9, 10 e 11 estão apresentados os dendrogramas com os algoritmos de agrupamento, ligação simples, ligação completa e Ward respectivamente, utilizando as variáveis independentes (A, P, X, Y e p) para a vazão Q_{95} , combinados com as distâncias Euclidiana, Mahalanobis e Manhattan.

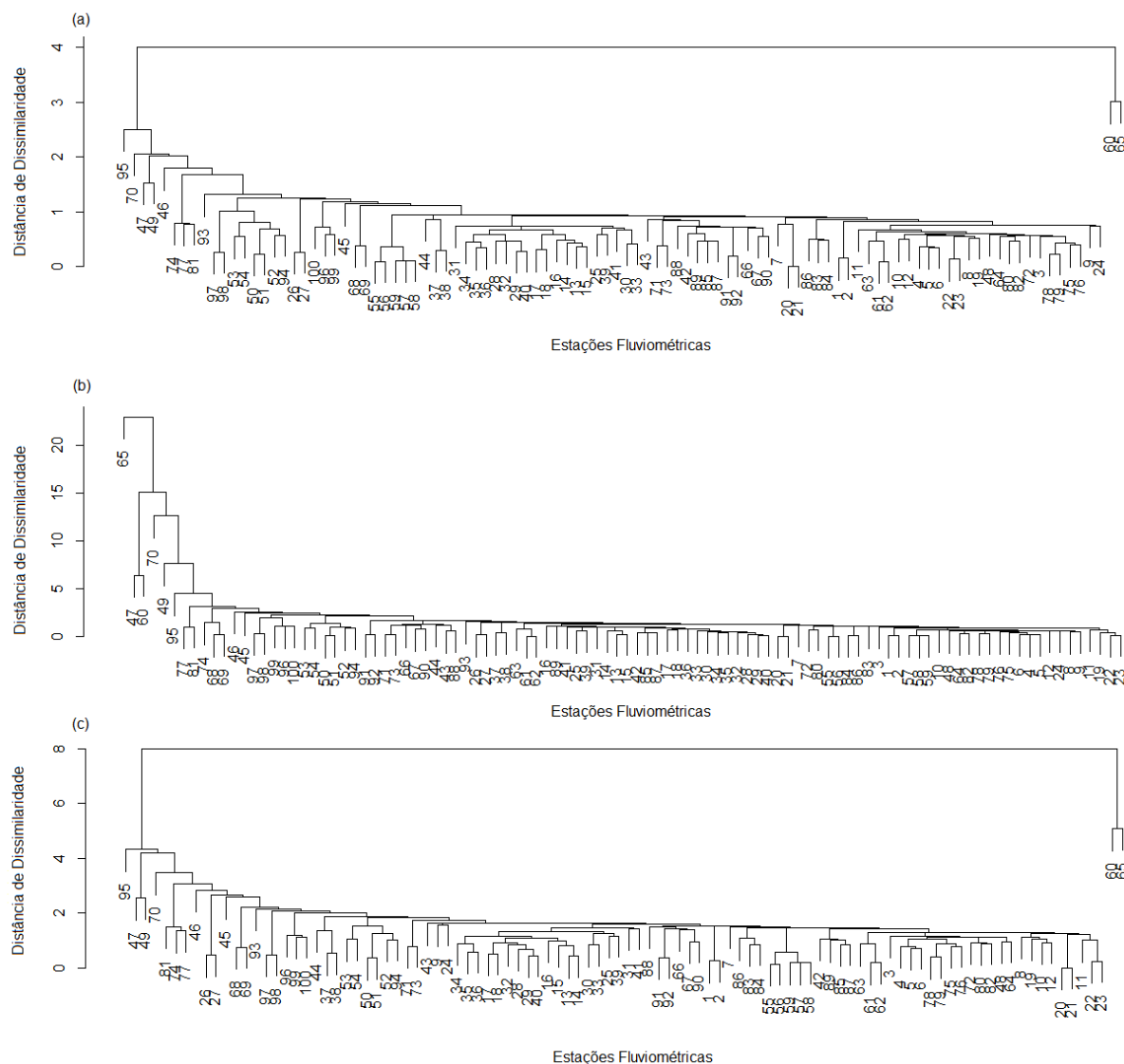


Figura 9 – Dendrogramas obtidos a partir da combinação do algoritmo de ligação simples de agrupamento com as variáveis independentes para a Q_{95} com as distâncias: (a) Euclidiana, (b) Mahalanobis e (c) Manhattan.

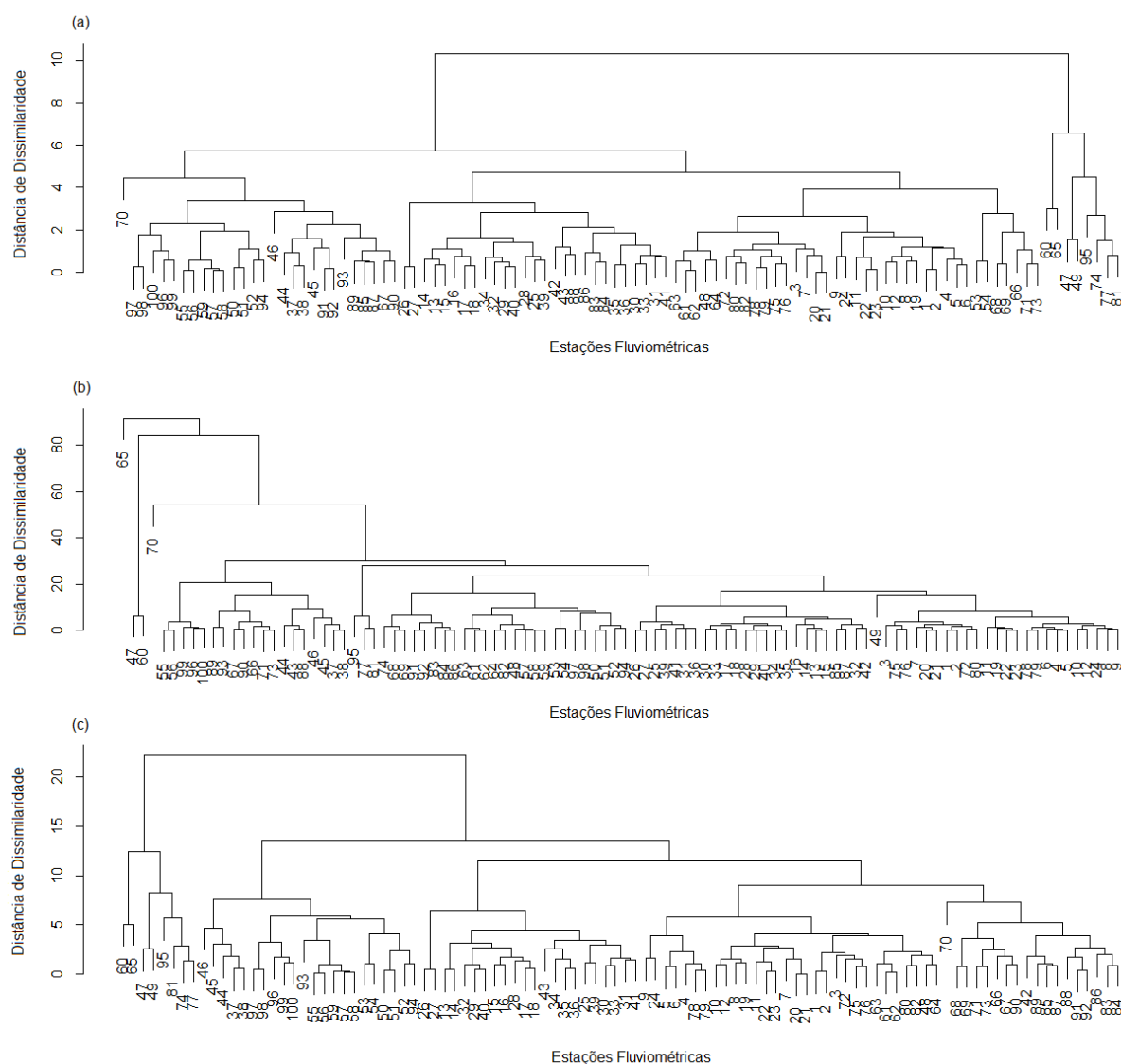


Figura 10 - Dendrogramas obtidos a partir da combinação do algoritmo de ligação completa de agrupamento com as variáveis independentes para a Q_{95} com as distâncias: (a) Euclidiana, (b) Mahalanobis e (c) Manhattan.

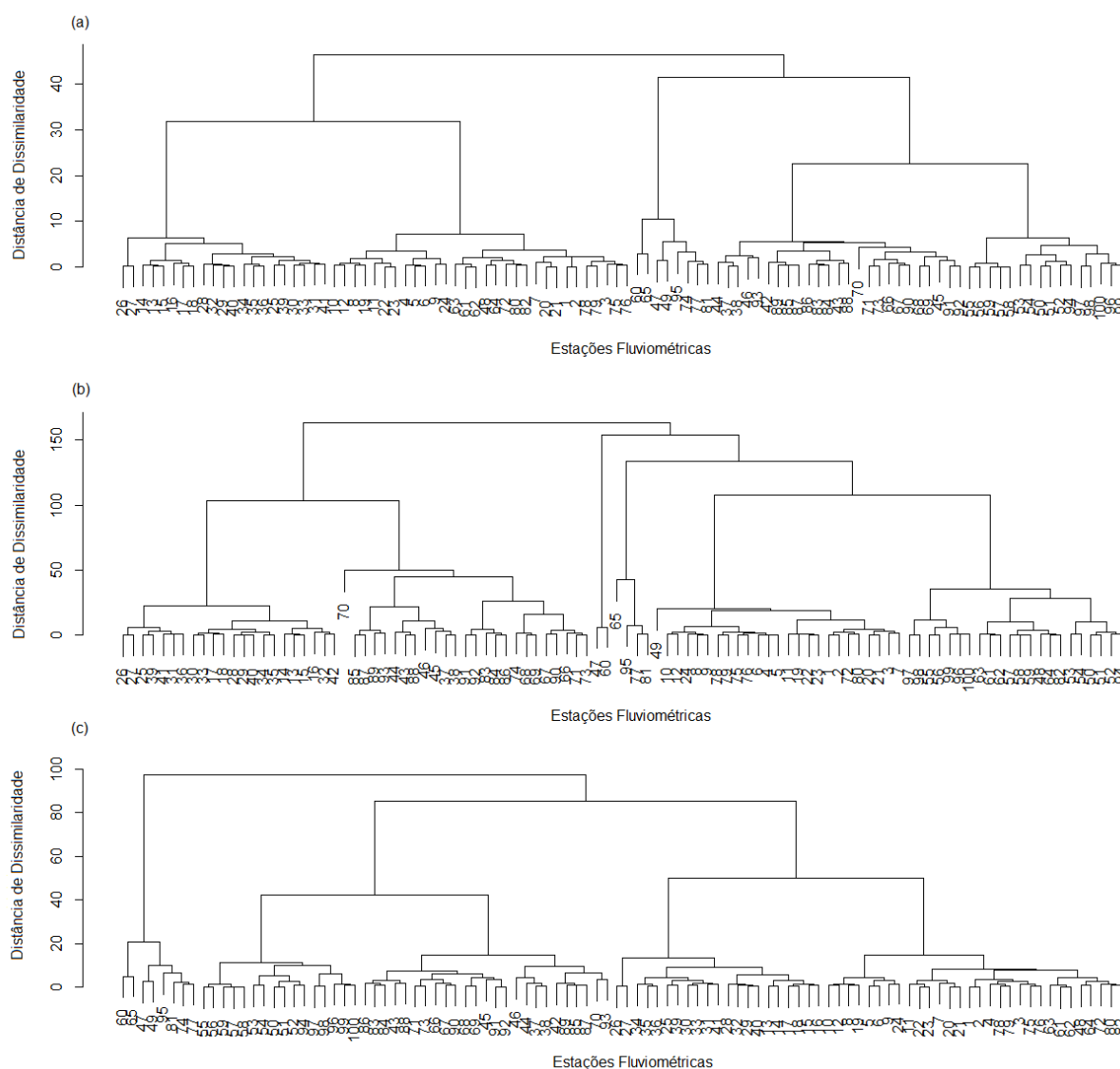


Figura 11 – Dendrogramas obtidos a partir da combinação do algoritmo de Ward de agrupamento com as variáveis independentes para a Q_{95} com as distâncias: (a) Euclidiana, (b) Mahalanobis e (c) Manhattan.

Os dendrogramas para a vazão Q_{95} pelo método de ligação simples apresentados na Figura 9 foram semelhantes, quanto à estrutura dos grupos gerados, à vazão Q_{90} , diferindo apenas quanto ao local do posicionamento das estações.

O algoritmo de ligação completa para a Q_{95} (Figura 10) possibilitou a formação de pelo menos 2 grupos pela distância Euclidiana e 2 e 3 grupos pela distância Manhattan. A distância de Mahalanobis também não se mostrou adequada com este algoritmo, dificultando a separação em grupos.

Como observado para a vazão Q_{90} , o algoritmo de Ward também apresentou maiores possibilidades para a Q_{95} , na formação dos grupos (Figura 11). Pelas distâncias Euclidiana e Manhattan, o corte do dendrograma possibilitou a formação

de 2, 3, 4 e 5 grupos; já no caso do dendrograma obtido segundo a distância de Mahalanobis, o corte permitiu apenas 2 grupos.

O algoritmo de Ward com a distância Euclidiana mostrou-se satisfatório na definição das sete regiões hidrologicamente homogêneas no nordeste brasileiro (RODRIGUEZ *et al.*, 2016). Os autores também inseriram as variáveis latitude e longitude, a fim de obter regiões homogêneas contínuas, e observaram maior influência da latitude para a identificação das mesmas.

4.3. Regiões homogêneas para a Q_{90}

De posse dos dendrogramas, optou-se em testar todas as possíveis configurações de grupos disponíveis pelos cortes dos mesmos. Desta forma, na sequência são apresentadas as regiões formadas para a vazão Q_{90} no estado do Rio Grande do Sul, além das equações de regressão originárias do método Stepwise para a Q_{90} , com as variáveis independentes (A, P e p) e as respectivas estatísticas.

Nesta dissertação optou-se por utilizar somente o modelo matemático potencial para todas as alternativas analisadas, que combinam diferentes agrupamentos para as vazões de referência, conforme conduzido por Beskow *et al.* (2016). Em todas as equações de regressão obtidas as variáveis que melhor se ajustaram devem ser inseridas com as suas unidades de medidas padrão, área em km^2 , perímetro em km, precipitação em mm e as vazões em m^3s^{-1} .

Pruski *et al.* (2015) lembram que a extrapolação das equações de regressão para além dos limites dos dados da amostra utilizados para estimar os parâmetros do modelo de regressão não é recomendado. Sendo assim, para todas as equações geradas, serão apresentados os seus respectivos limites de aplicação.

Cada região foi avaliada pela medida de heterogeneidade (H), visto que essa metodologia se fundamenta na suposição de que todas as séries tenham os mesmos momentos-L populacionais, embora seus momentos-L amostrais sejam diferentes devido à variabilidade natural (HOSKING; WALLIS, 1997). Com a aplicação da medida H é estabelecido se os grupos/regiões definidos pelo corte dos dendrogramas, compreendem uma região homogênea ou não. Primeiramente, foram identificadas todas as alternativas de cortes do dendrograma com a formação de 2 regiões (Figuras 12, 13, 14 e 15).

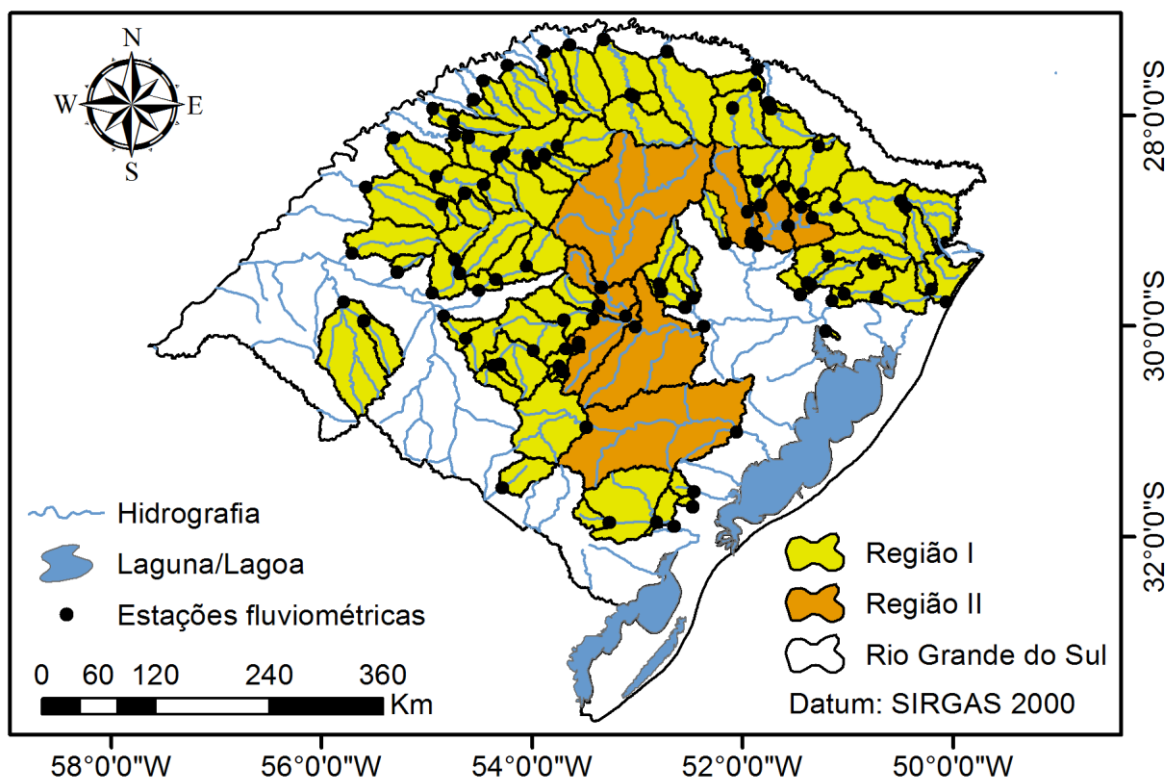


Figura 12 - Regiões para o estado do Rio Grande do Sul determinadas usando o algoritmo de ligação completa de agrupamento combinado com a distância Euclidiana e Manhattan.

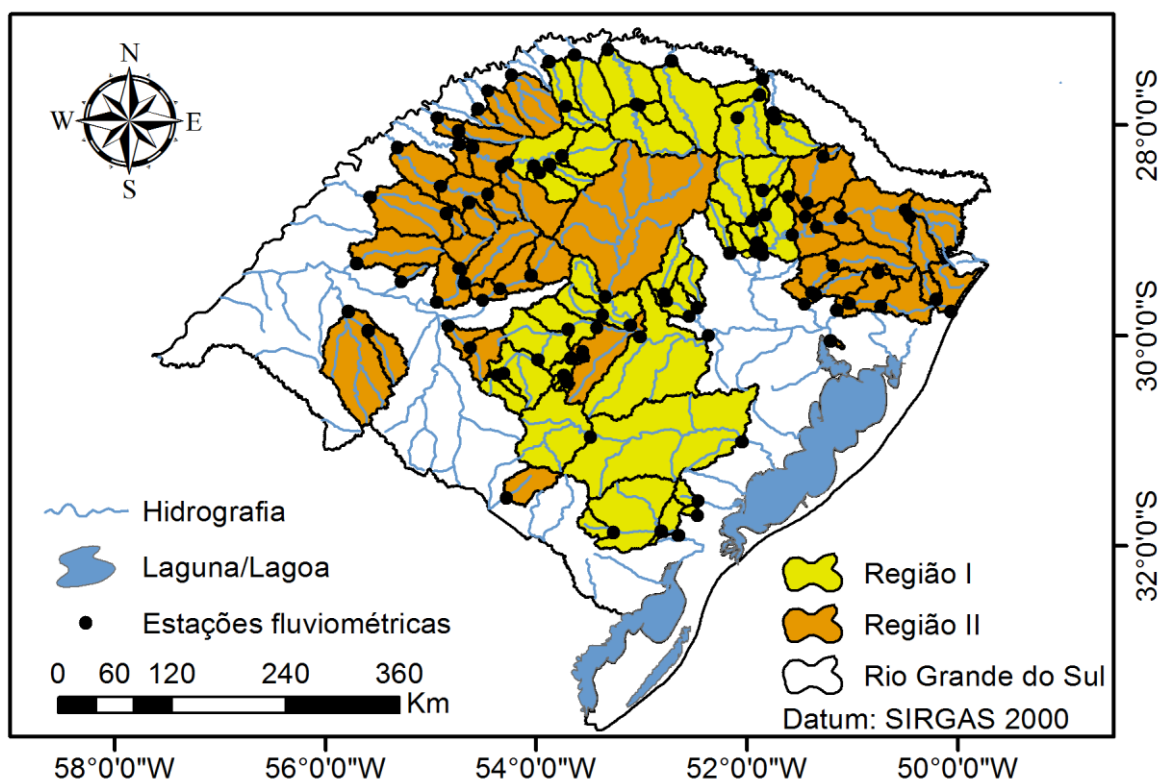


Figura 13 – Regiões para o estado do Rio Grande do Sul determinadas usando o algoritmo de Ward de agrupamento combinado com a distância Mahalanobis.

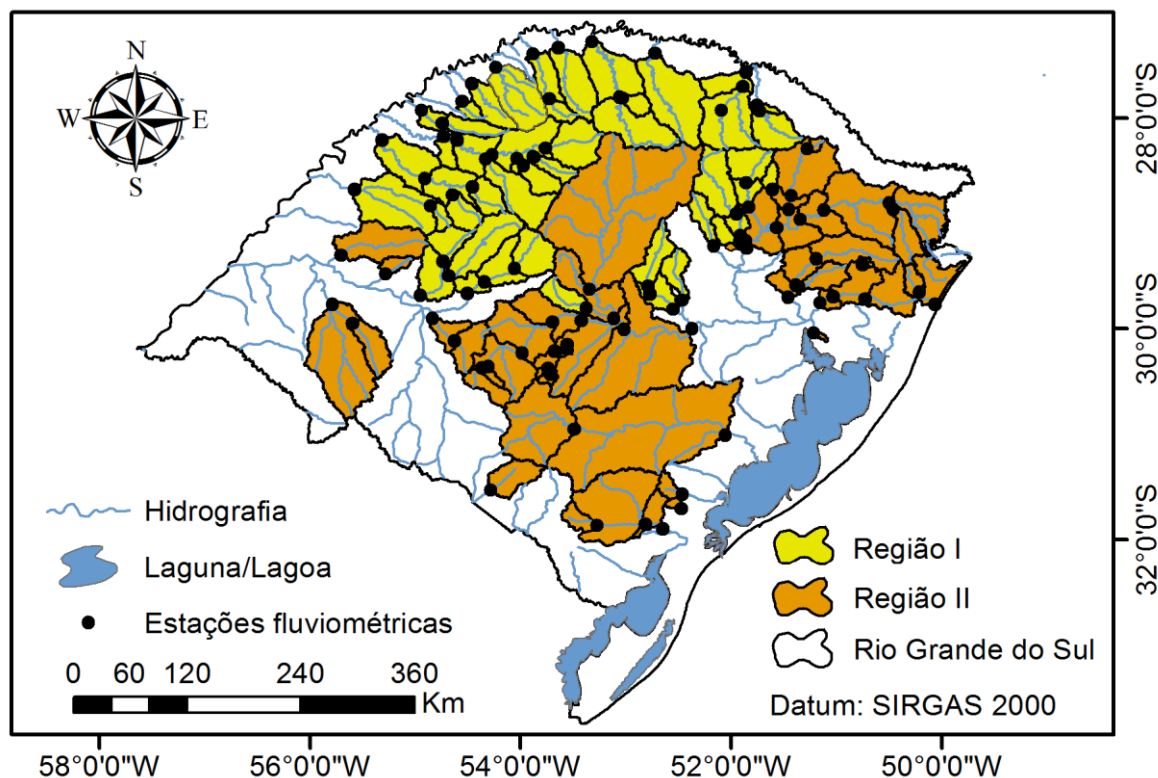


Figura 14 - Regiões para o estado do Rio Grande do Sul determinadas usando o algoritmo de Ward de agrupamento combinado com a distância Euclidiana e Manhattan.

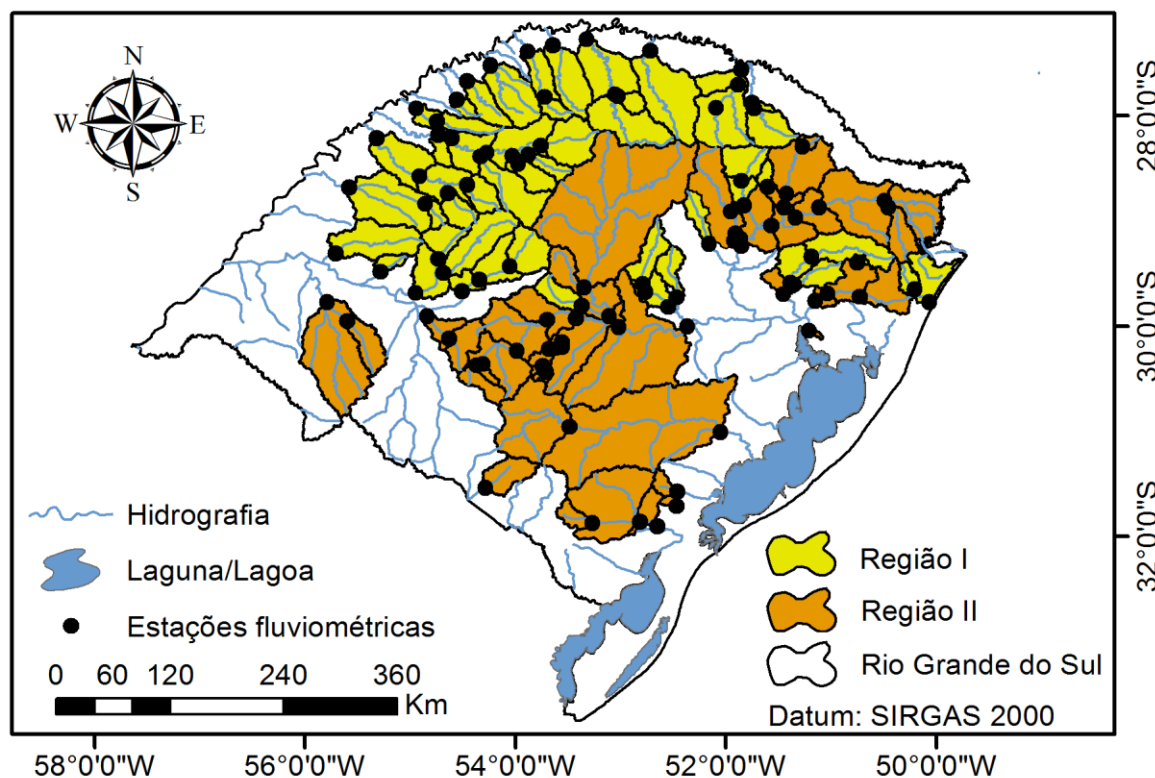


Figura 15 - Regiões para o estado do Rio Grande do Sul determinadas usando o método k-means com $k=2$.

Analisando as bacias que contemplavam cada grupo nos dendrogramas gerados pelo algoritmo de ligação completa combinados com as distâncias Euclidiana e Manhattan, observa-se que os mesmos apresentam características iguais quanto as bacias que possuem, fato que também ocorre com o algoritmo de Ward para essas distâncias. Desta forma, na Figura 12, as bacias foram agrupadas pela distância Euclidiana e Manhattan com o algoritmo de ligação completa. Essa configuração não proporcionou uma separação homogênea quanto ao número de bacias (n) que contemplam cada região. A região II caracteriza-se por reunir as bacias de maior porte (13472,33 a 38096,24 km²), e a região I as demais bacias. Na Tabela 7 são apresentadas as análises estatísticas das duas regiões por esta configuração.

Tabela 6 - Características estatísticas das regiões determinadas usando o algoritmo de ligação completa de agrupamento combinado com a distância Euclidiana e Manhattan

Região	n	Equação Q ₉₀	H	R ² _a	R ²	c	log C _{NS}	EMA (m ³ s ⁻¹)
I	92	$2,79 \cdot 10^{-14} \cdot (A^{0,988} \cdot p^{3,421})$	-0,19	0,67	0,68	0,71	0,69	3,74
II	8	$6,16 \cdot 10^{-34} \cdot (P^{1,6} \cdot p^{9,474})$	-0,18	0,73	0,81	0,82	0,81	24,14

Observa-se na Tabela 7 que a região I apresentou coeficiente de desempenho de Camargo e Sentelhas “c” considerado bom e de eficiência de Nash-Sutcliffe, aplicado ao logaritmos “log C_{NS}”, como aceitável. Com relação à região II, o “c” foi muito bom e log C_{NS} adequado e bom, conforme classificação de Camargo; Sentelhas (1997) e Motovilov *et al.* (1999), respectivamente. Na região I, o R²_a e o R² apresentam valores inferiores a 0,7, não seguindo as indicações de Elesbon *et al.* (2015), que ressaltam que tal valor deve ser superior ou igual a 0,7.

Nas duas regiões o melhor ajuste ocorreu com a inserção de duas variáveis. Em virtude da região II reunir as bacias de maior dimensão e, conseqüentemente, maior vazão de permanência (Q₉₀), o erro médio absoluto (EMA) encontrado para a Q₉₀, em m³s⁻¹, foi superior ao da região I. Entretanto, quando as regiões foram avaliadas pela medida de heterogeneidade (H), confirmou-se a sua homogeneidade, não sendo necessário excluir nenhuma bacia.

Os limites de aplicação para as equações são: para a região homogênea I, as áreas de drenagem das bacias variam de 68,61 a 10153,78km² e a precipitação média anual de 1172,11 a 1914,73mm; para a região homogênea II, o perímetro varia de 767,68 a 1533,71km e a precipitação média anual de 1455,16 a 1695,71mm. A Figura

16 ilustra a relação entre a Q_{90} observada e estimada para ambas as regiões homogêneas.

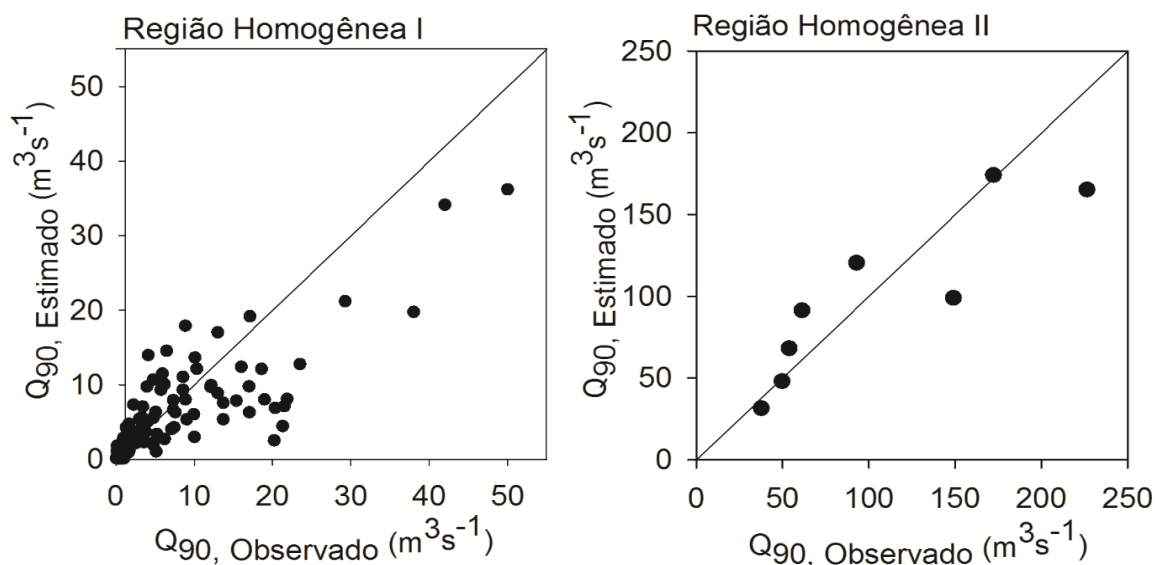


Figura 16 - Relação entre os valores da Q_{90} observados e estimados (m^3s^{-1}) para as regiões homogêneas delimitadas pela distância Euclidiana e Manhattan, com o algoritmo de ligação completa de agrupamento.

Pode-se verificar que na região homogênea I na (Figura 16), a maioria dos pontos encontra-se na parte inferior da reta 1:1, o que caracteriza-se por gerar uma subestimativa nos valores. Desta forma, o modelo acaba trabalhando em benefício da conservação dos recursos hídricos, entretanto, em alguns casos, essa subestimativa pode ser elevada, o que pode levar os usuários a conflitos com o órgão gestor de recursos hídricos. Já para a região homogênea II o modelo apresentou um melhor ajuste.

Dentro desta perspectiva, segundo Wolff *et al.* (2014) é plausível que se estabeleça critérios de outorga menos conservadores, bem como a adoção de vazões mínimas de referência sazonais, que correspondam às condições trimestrais de maior escassez hídrica, em virtude de que nos períodos chuvosos, uma maior quantidade do recurso estaria passível de ser outorgado. Pruski *et al.* (2011) salientam que as vazões mínimas mostram ser bem mais suscetíveis às alterações antrópicas do que as vazões médias, por exemplo. Queiroz *et al.* (2010) lembram que essas vazões servem para a manutenção dos ecossistemas e também podem ser utilizadas para os cálculos de diluição de efluentes, onde os critérios devem ser mais rigorosos.

Na Figura 13, as bacias foram agrupadas pela distância de Mahalanobis com o algoritmo de Ward. Observa-se que essa configuração produziu grupos mais

homogêneos no que se refere aos números de bacias que abrangem cada região. Na Tabela 8 são indicadas as características das duas regiões para esta configuração.

Tabela 7 - Características estatísticas das regiões determinadas usando o algoritmo de Ward de agrupamento combinado com a distância Mahalanobis

Região	n	Equação Q_{90}	H	R^2_a	R^2	c	$\log C_{NS}$	EMA (m^3s^{-1})
I	51	$3,17 \cdot 10^{-25} \cdot (A^{1,185} \cdot p^{6,615})$	-0,32	0,88	0,97	0,97	0,87	3,73
II	49	$3,23 \cdot 10^{-4} \cdot (P^{1,762})$	-0,31	0,65	0,66	0,65	0,67	9,2

Em comparação com a configuração anterior (Tabela 7), os resultados exibidos para a região I na Tabela 8, por meio da distância de Mahalanobis e algoritmo de Ward, apresentaram melhores estatísticas, mas o EMA praticamente igual; já a região II resultou em coeficientes estatísticos de menor qualidade do que a configuração anterior, e ainda aceitáveis. Observa-se também que para a região II, apenas a variável perímetro foi necessária para se obter o melhor ajuste. Aplicando-se a medida H confirmou-se a homogeneidade das duas regiões.

Os limites de aplicação para as equações são: para a região homogênea II, o perímetro varia de 46,82 a 1349,31km; e para a região homogênea I, as áreas de drenagem das bacias variam de 68,61 a 38096,24km² e a precipitação média anual de 1376,92 a 1874,44mm. A Figura 17 mostra a relação entre a Q_{90} observada e estimada, para ambas as regiões homogêneas.

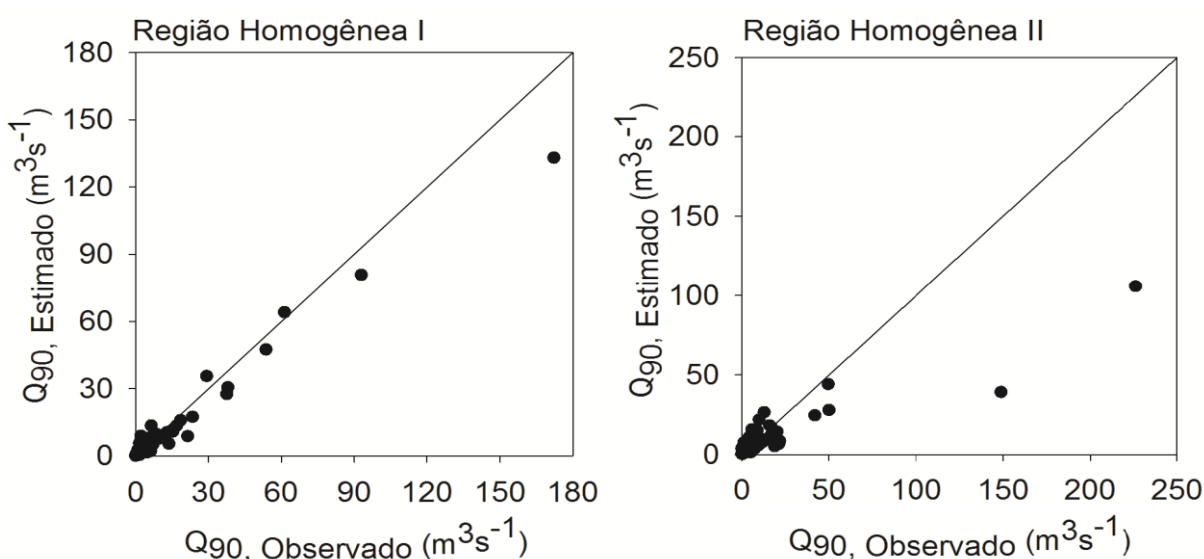


Figura 17 - Relação dos valores da Q_{90} observados e estimados (m^3s^{-1}) para as regiões homogêneas pela distância de Mahalanobis, com o algoritmo de Ward de agrupamento.

Nota-se na Figura 17 que os ajustes das vazões foi melhor para a região homogênea I o que é confirmado pelo R^2 na Tabela 8, o qual explica 97% da variabilidade dos dados no modelo, já a região homogênea II o desempenho foi inferior. Contudo, busca-se uma divisão do Estado em regiões que o desempenho seja favorável para todas as regiões.

Na Figura 14, as bacias foram agrupadas pela distância Euclidiana e Manhattan, com o algoritmo de Ward. Observa-se, em comparação com a distância de Mahalanobis (Figura 13), que as duas distâncias produziram formações iguais, apresentando regiões com o mesmo número de bacias, mas com localizações geográficas diferentes. Na Tabela 9 estão apresentadas as características dessas regiões.

Tabela 8 - Características estatísticas das regiões determinadas usando o algoritmo de Ward de agrupamento combinado com a distância Euclidiana e Manhattan

Região	n	Equação Q_{90}	H	R^2_a	R^2	c	$\log C_{NS}$	EMA (m^3s^{-1})
I	51	$2,24 \cdot 10^{-3} \cdot (A^{1,056})$	-0,12	0,70	0,70	0,73	0,65	4,04
II	49	$5,7 \cdot 10^{-5} \cdot (P^{2,005})$	-0,21	0,76	0,77	0,79	0,90	10,43

Na Tabela 9, observa-se que os coeficientes estatísticos da região II são melhores que os apresentados pela configuração anterior (Tabela 8), com mesmo número de bacias, mas o valor do EMA possui pequena diferença. O mesmo não acontece para a região I, contudo os coeficientes ainda são aceitáveis.

Nota-se que, para as duas equações, os ajustes foram apenas com uma variável independente (área ou perímetro). A opção de usar a área de drenagem como variável explicativa também foi adotada por Beskow *et al.* (2016) para regionalizar a Q_{90} no estado do Rio Grande do Sul. Os autores também concluíram que a função potencial representou adequadamente a relação entre Q_{90} e a área de drenagem.

Os limites de aplicação para as equações são: para a região homogênea I, as áreas de drenagem variam de 337,09 a 10153,78km²; e para a região homogênea II, o perímetro varia de 40,46 a 1533,71km. A Figura 18 ilustra a relação entre a Q_{90} observada e estimada para ambas as regiões homogêneas.

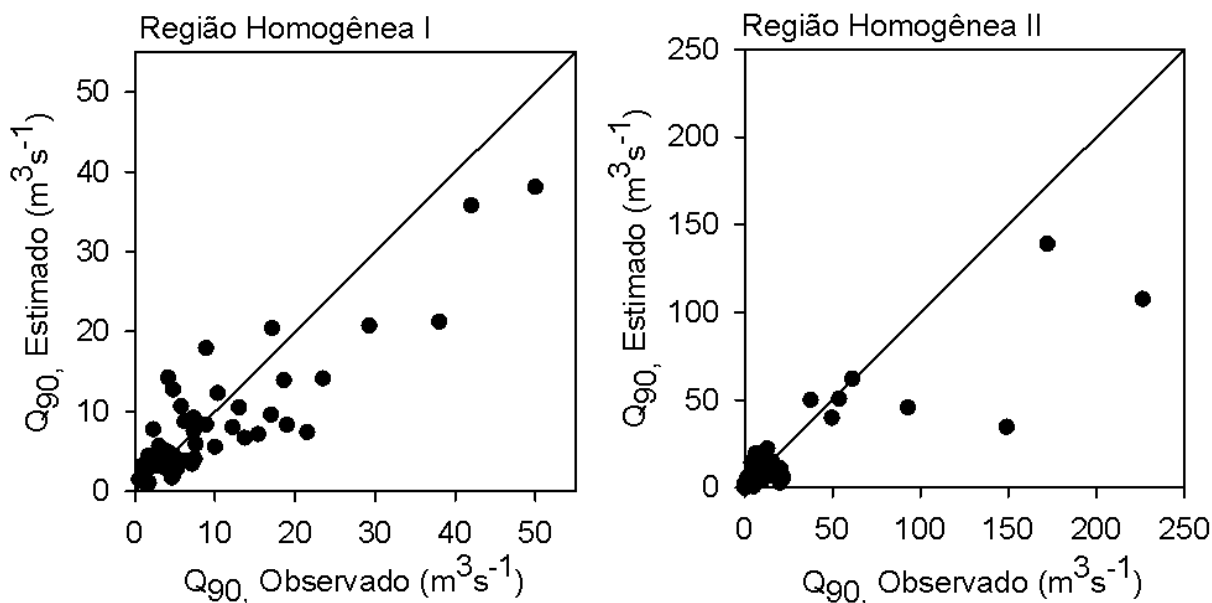


Figura 18 - Relação dos valores da Q_{90} observados e estimados (m^3s^{-1}) para as regiões homogêneas pela distância Euclidiana e Manhattan, com o algoritmo de Ward de agrupamento.

Entre os algoritmos de agrupamento, o que mais se destaca é o de Ward por formar grupos mais homogêneos, Farhan; Shaikh (2017) utilizaram o referido algoritmo juntamente com a correlação de Pearson para o cálculo da matriz de dissimilaridade e encontraram 5 regiões homogêneas para o sul da Jordânia. Malekinezhad; Nachtnebel; Klik (2011) empregando o algoritmo de Ward e a distância Euclidiana descobriram 3 regiões para a bacia Namak-Lake no Irã, sendo as regiões aceitas como homogêneas quando testadas pela medida H. Ilorme; Griffis (2013), localizaram 7 regiões no sudeste dos Estados Unidos, em que duas regiões foram consideradas heterogêneas pela medida H, necessitando realizar ajustes e também a exclusão de uma região.

Na Figura 15 estão apresentadas as bacias que contemplam as regiões pelo método não hierárquico k-means para $k=2$. O número de bacias não diferiu muito dos métodos hierárquicos, exceto pela distância Euclidiana e Manhattan, combinado com o algoritmo de ligação completa. Contudo, observa-se a maior semelhança com a configuração das distâncias Euclidiana e Manhattan com o algoritmo Ward (Figura 14). Desta forma, se não for possível obter outras regiões para fins de planejamento, recomenda-se utilizar as equações fornecidas pela regressão dos grupos gerados pela configuração da distância Euclidiana e Manhattan pelo algoritmo de Ward ou as originárias do método não hierárquico com $k=2$. Na Tabela 10 estão apresentadas as propriedades estatísticas de cada região pelo método k-means com $k=2$.

Tabela 9 - Características estatísticas das regiões determinadas pelo método k-means com k=2

Região	n	Equação Q_{90}	H	R^2_a	R^2	c	$\log C_{NS}$	EMA (m^3s^{-1})
I	59	$2,22 \cdot 10^{-3} \cdot (A^{1,062})$	-0,24	0,67	0,68	0,74	0,68	4,07
II	41	$3,38 \cdot 10^{-5} \cdot (P^{2,075})$	-0,07	0,79	0,79	0,77	0,81	11,17

A Tabela 10 é possível observar que apenas uma variável independente foi necessária para obter o melhor ajuste. Contudo, a região I apresenta valores de R^2_a e R^2 inferiores a 0,7, mas ainda quando avaliada pela medida H, as duas regiões são consideradas como homogêneas. Os limites de aplicação para as equações são: para a região homogênea II, o perímetro varia de 40,46 a 1533,71km; e para a região homogênea I, as áreas de drenagem das bacias variam de 111,00 a 10153,78km². A Figura 19 mostra a relação entre a Q_{90} observada e estimada para ambas as regiões homogêneas.

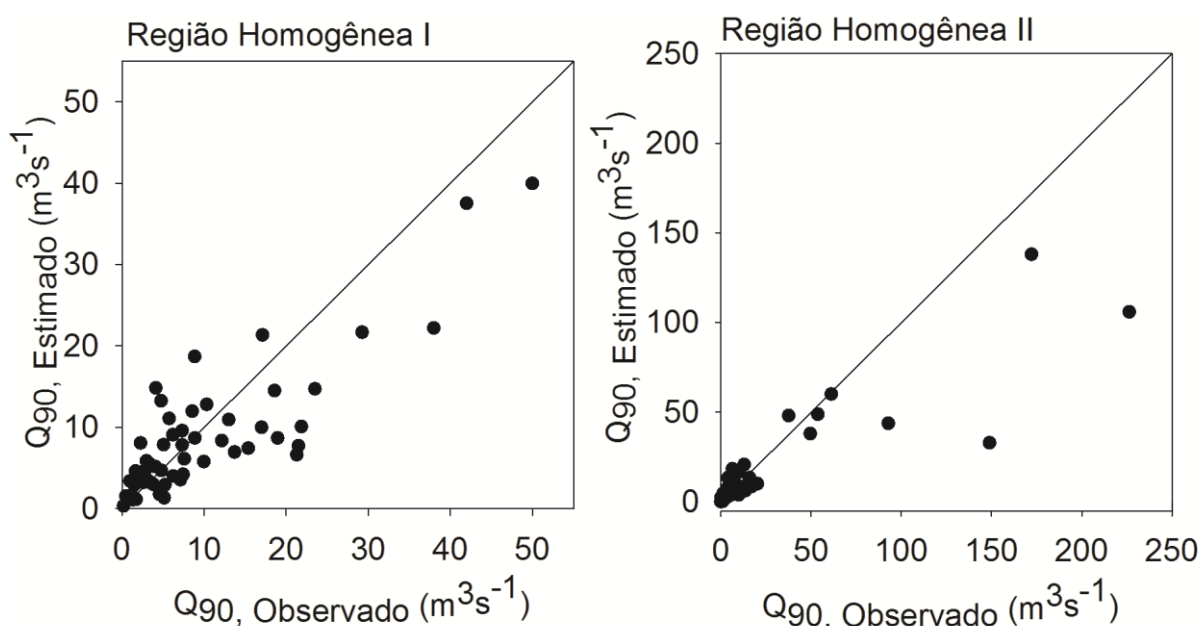


Figura 19 – Relação dos valores da Q_{90} observados e estimados (m^3s^{-1}) para as regiões homogêneas pelo método k-means com k=2.

Modarres (2010) já salientou que para fim de planejamento é recomendado um maior número de grupos, para que se consiga interpretar as características das regiões e utilizá-las com maior confiança na aplicabilidade das equações. Vezza *et al.* (2010) afirmaram que a possível consideração de uma única região homogênea é a maneira mais simples de executar a regionalização. Porém, esses pesquisadores apontaram que a hipótese de aplicar um modelo geral, generaliza todos os diferentes processos importantes para análise de vazões mínimas, o que pode comprometer, de

certa forma, a gestão dos recursos hídricos. Já Hosking; Wallis (1997) recomendam que na regionalização, uma região homogênea tenha até 20 estações.

Sendo assim comparando a divisão de 2 regiões homogêneas para o Estado, observa-se que o método da distância Euclidiana e Manhattan, com o algoritmo de ligação completa, reuniu as bacias que apresentavam os maiores valores da vazão Q_{90} na região II, o que contempla 8 bacias, cujas vazões variam de 37,60 a 226,42 m^3s^{-1} , enquanto que na região I, a Q_{90} varia de 0,05 a 50,00 m^3s^{-1} . Para fins de planejamento, uma região com 8 estações seria adequada, contudo ainda é necessário obter mais grupos, a fim de dividir o restante das estações.

Desta forma, pelas contribuições dos autores supracitados e os resultados encontrados até o momento justifica-se a busca por mais regiões no estado do Rio Grande do Sul, uma vez que, existem 100 bacias para caracterizá-lo, e a generalização em apenas uma ou duas regiões não permitiria que as particularidades fossem inseridas nas equações, para a estimativa das vazões de permanência. Consequentemente, foram identificadas todas as alternativas com a separação de 3 regiões para o Estado (Figuras 20 e 21).

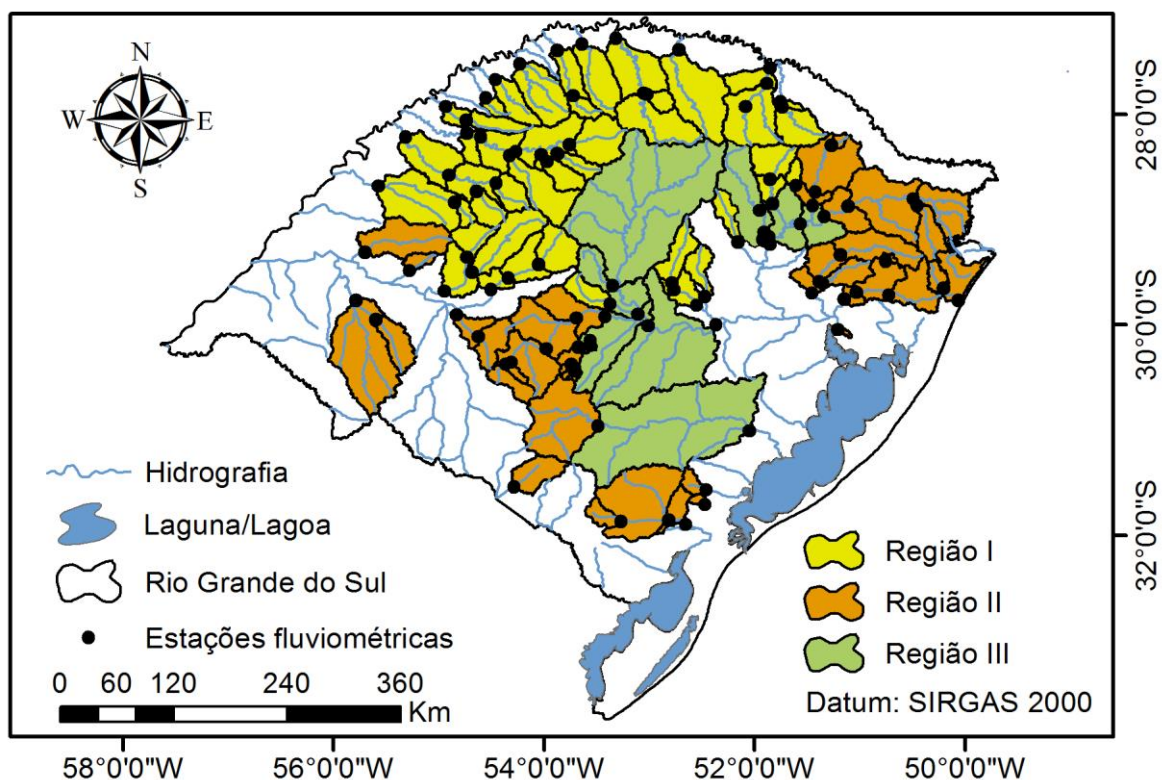


Figura 20 – Regiões para o estado do Rio Grande do Sul determinadas usando o algoritmo de Ward de agrupamento combinado com a distância Euclidiana e Manhattan.

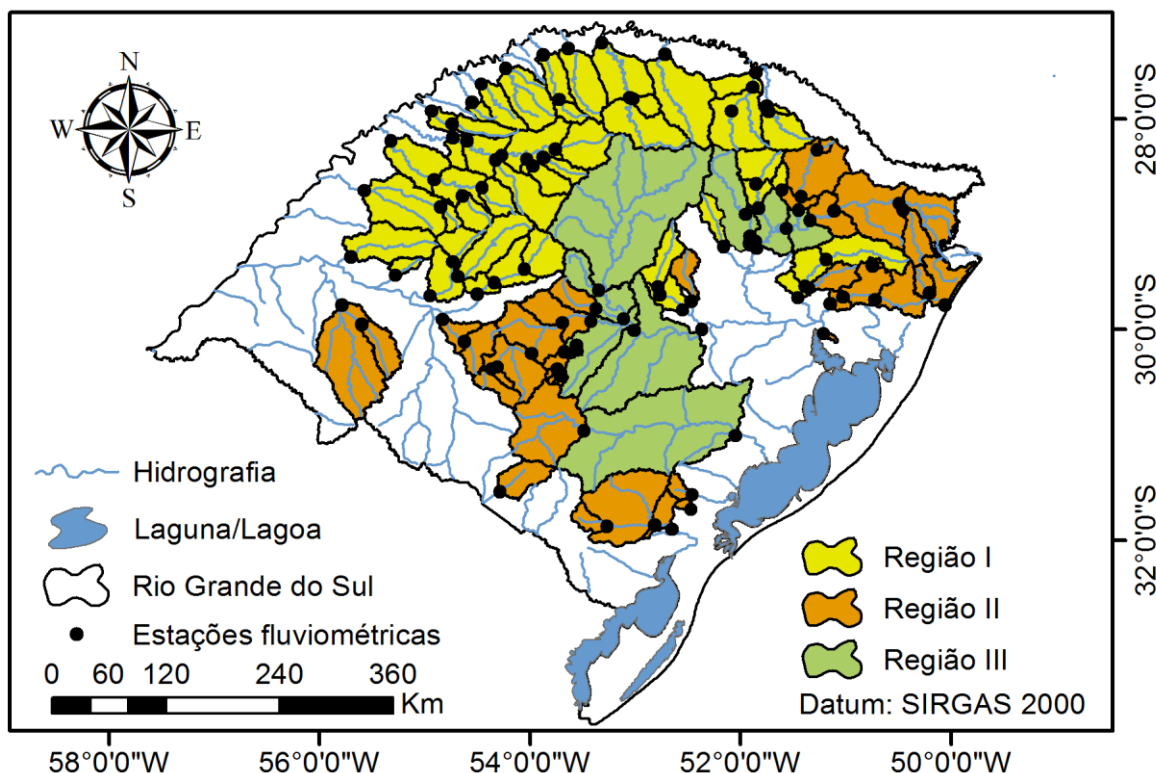


Figura 21 - Regiões determinadas para o estado do Rio Grande do Sul pelo método k-means com k=3.

Na Figura 20, as bacias foram agrupadas pela distância Euclidiana e Manhattan, combinadas com o algoritmo de Ward, respectivamente. Conforme foi mencionado anteriormente, essas distâncias já haviam gerado agrupamentos idênticos, no entanto analisando os cortes dos dendrogramas para a divisão de 3 regiões no Estado, pode-se observar que o algoritmo de ligação completa, combinado com a distância de Manhattan também produziu a mesma configuração. Sendo assim, pela primeira vez além das distâncias produzirem configurações iguais os algoritmos também proporcionam tal semelhança. Na Tabela 11 estão apresentados os coeficientes estatísticos das equações para as 3 regiões geradas por todas as configurações que apresentaram analogamente divisão e na Figura 22, a relação entre a Q_{90} observada e estimada para as regiões.

Tabela 10 - Características estatísticas das regiões determinadas usando o algoritmo de Ward de agrupamento combinado com a distância Euclidiana e Manhattan

Região	n	Equação Q_{90}	H	R^2_a	R^2	c	$\log C_{NS}$	EMA (m^3s^{-1})
I	51	$1,03 \cdot 10^{-14} \cdot (A^{1,031} \cdot p^{3,512})$	-0,12	0,65	0,67	0,74	0,67	3,95
II	41	$1,57 \cdot 10^{-4} \cdot (P^{1,802})$	-0,16	0,62	0,63	0,54	0,65	3,41
III	8	$6,16 \cdot 10^{-34} \cdot (P^{1,6} \cdot p^{9,474})$	-0,18	0,73	0,81	0,82	0,81	24,14

Observa-se que apenas para a região II o coeficiente de Camargo; Sentelhas (1997) “c” revelou que a equação obtida tem caráter de sofrível, não sendo adequada para uso, entretanto para o coeficiente de Nash-Sutcliffe se mostrou aceitável e quanto a medida H também foi aprovada como uma região homogênea. Em relação às regiões I e III é possível observar que o melhor ajuste ocorreu com a inserção da variável precipitação total anual, assim como observado por Pruski *et al.* (2012) na bacia do rio Pará, sub-bacia do rio São Francisco, e por Melati *et al.* (2016) na bacia hidrográfica do Taquari-Antas.

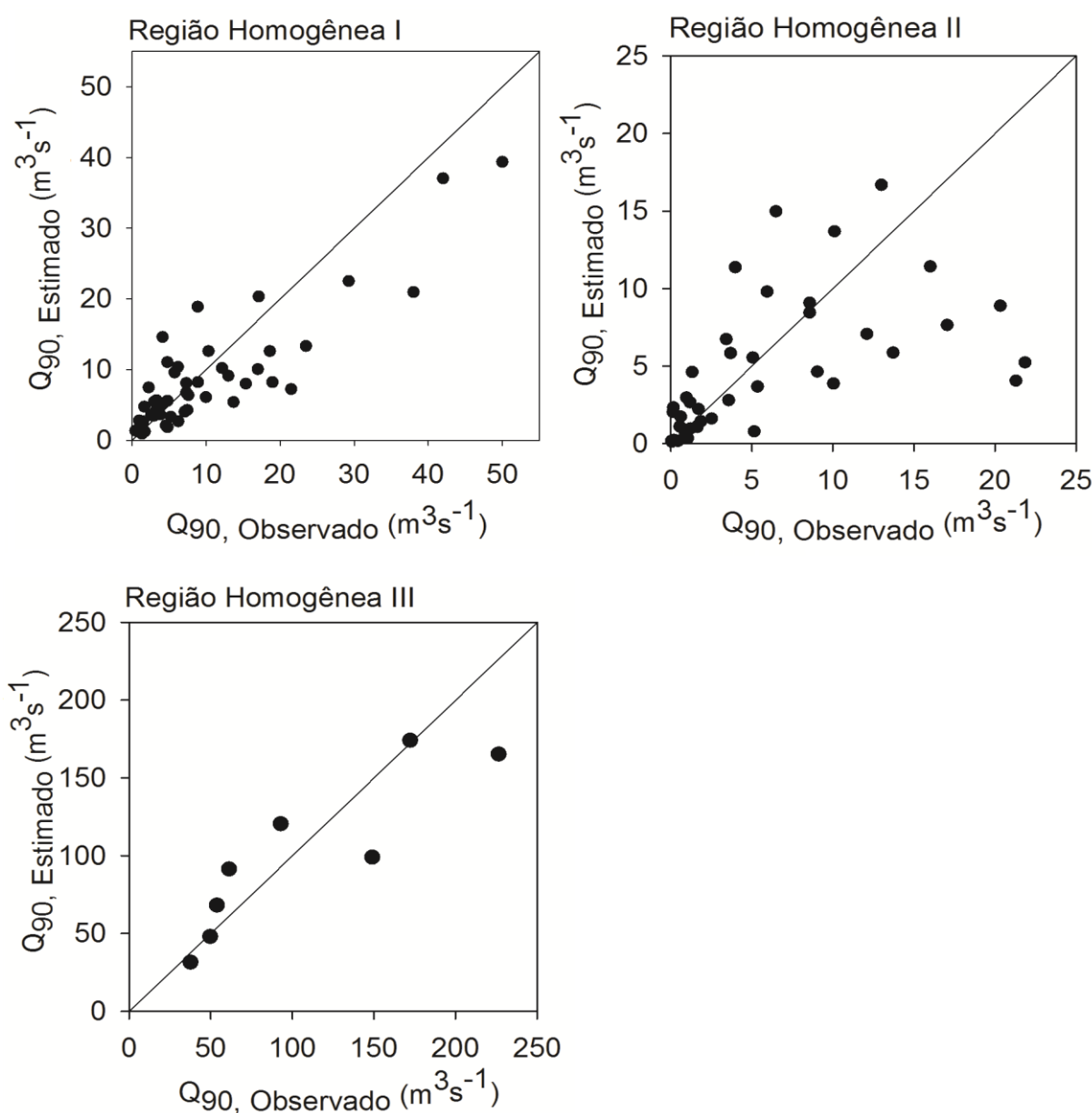


Figura 22 – Relação dos valores da Q_{90} observados e estimados (m^3s^{-1}) para as regiões homogêneas determinadas pela distância Euclidiana e Manhattan, com o algoritmo de Ward de agrupamento.

Os limites de aplicação para as equações são: para a região homogênea III, o perímetro varia de 767,68 a 1533,71km, e a precipitação de 1455,16 a 1693,62mm; para a região homogênea II, o perímetro varia de 40,46 a 614,93km; e para a região homogênea I, a área varia de 337,09 a 10153,78km² e a precipitação de 1604,34 a 1914,73mm.

Ainda na Figura 22 é possível notar que a relação para as bacias da região II mostrou-se com a maior dispersão dos pontos, quando comparada com as regiões I e II, justificando os coeficientes estatísticos de ajuste inferior.

Na Figura 21 estão apresentadas as bacias que contemplam as regiões pelo método não hierárquico k-means para k=3, cujo número de bacias não diferiu muito do método hierárquico (Figura 20), observando-se que a região III permanece com a mesma disposição. Na Tabela 12 estão presentes as características estatísticas para as regiões geradas pelo método k-means, com k=3.

Tabela 11 - Características estatísticas das regiões determinadas pelo método k-means com k=3

Região	n	Equação Q_{90}	H	R^2_a	R^2	c	$\log C_{NS}$	EMA (m ³ s ⁻¹)
I	54	$4,5 \cdot 10^{-3} \cdot (A^{0,971})$	0,18	0,62	0,63	0,70	0,63	4,26
II	38	$1,27 \cdot 10^{-4} \cdot (A^{1,82})$	-0,38	0,64	0,65	0,56	0,67	2,92
III	8	$6,16 \cdot 10^{-34} \cdot (P^{1,6} \cdot p^{9,474})$	-0,18	0,73	0,81	0,82	0,81	24,14

Observa-se que novamente na Tabela 12 a região II apresenta o coeficiente de Camargo; Sentelhas (1997) “c” como sofrível, sendo que as demais regiões exibem coeficientes adequados. Para as regiões I e II, o melhor ajuste nas equações foi proporcionado utilizando apenas a variável independente área de drenagem, o que corrobora com Bazzo *et al.* (2017); Beskow *et al.* (2016); Chaves *et al.* (2002); Novaes *et al.* (2009); Lisboa *et al.* (2014); Moreira; Silva (2014), que utilizaram somente essa variável para a formação das equações, uma vez que é de simples obtenção.

Os limites de aplicação para as equações são: para as regiões homogêneas I e II, possuem a área de drenagem que variam de 337,09 a 10153,78 km², 68,61 a 8292,29 km², respectivamente; e a região homogênea III, o perímetro varia de 767,68 a 1533,71km, e a precipitação de 1455,16 a 1693,62mm. A Figura 23 mostra a relação entre a Q_{90} observada e estimada para as regiões homogêneas geradas.

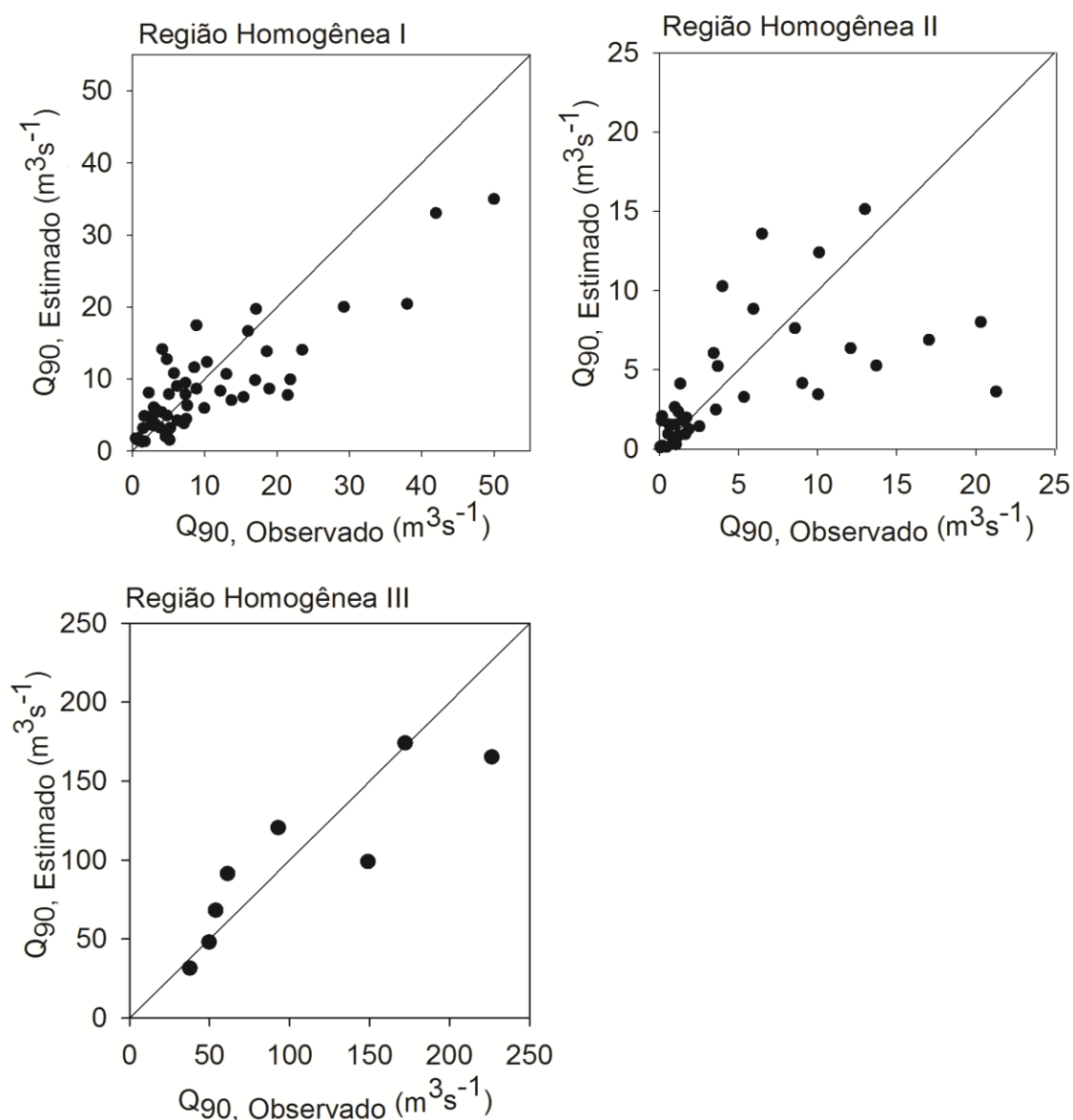


Figura 23 - Relação dos valores da Q_{90} observados e estimados (m^3s^{-1}) para as regiões homogêneas pelo método k-means com $k=3$.

Observa-se que a região II precisa ser melhorada, em virtude dos seus coeficientes estatísticos, além disso é a única que apresenta bacias distantes geograficamente, contudo para Singh *et al.* (2016), os quais constataram que a proximidade espacial é um bom indicador para similaridade entre os possíveis grupos formados, apresentando assim um maior grau de conectividade. Entretanto, Corduas (2011) acredita que a proximidade geográfica dos locais agrupados representa uma restrição que muitas vezes deve ser considerada. Desta forma, a análise sobre a região II precisa ser cuidadosa, uma vez que esta separação também pode indicar a necessidade de adicionar mais uma região no Estado, para fins de planejamento da

vazão Q_{90} . Desta forma, foram identificadas todas as alternativas com a formação de 4 regiões (Figuras 24 e 25).

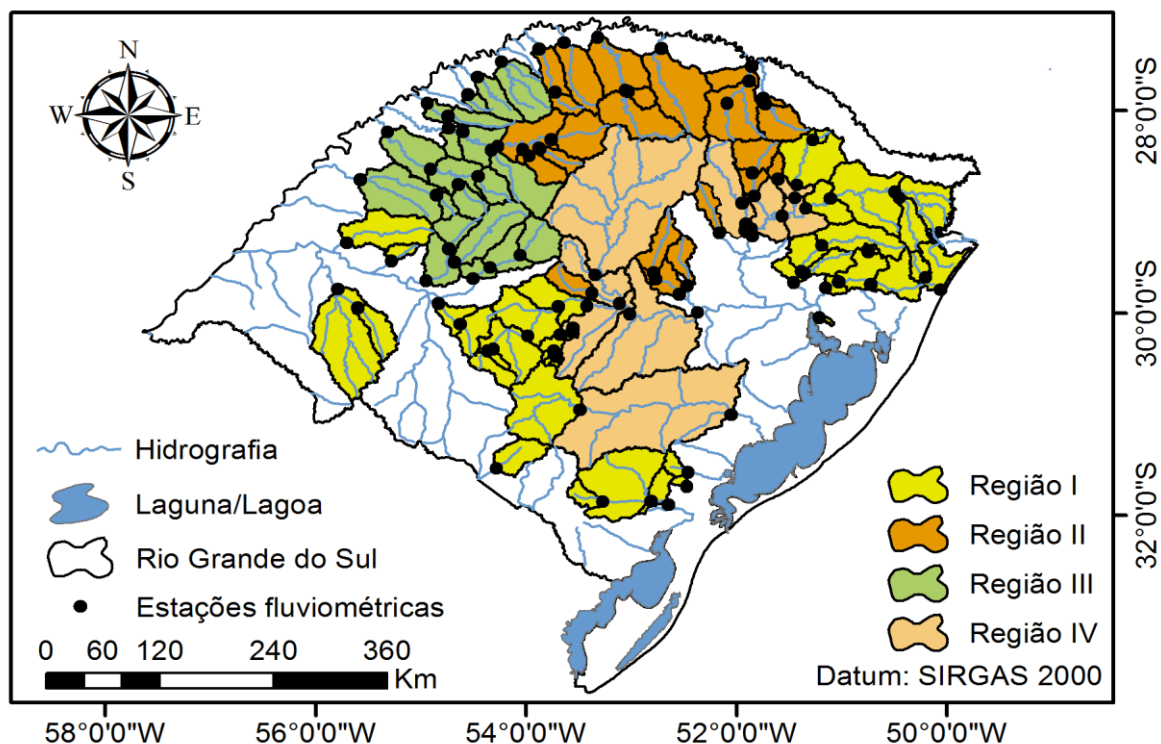


Figura 24 – Regiões para o estado do Rio Grande do Sul determinadas usando o algoritmo de Ward de agrupamento combinado com a distância Euclidiana e Manhattan.

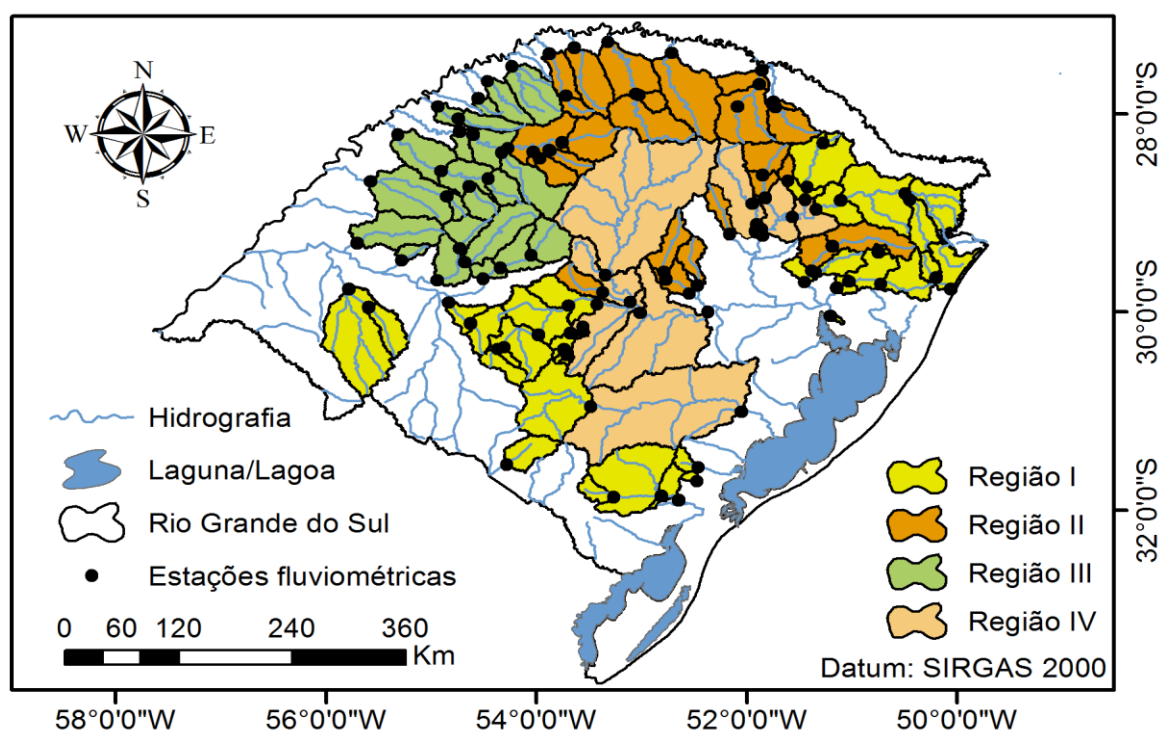


Figura 25 - Regiões determinadas para o estado do Rio Grande do Sul pelo método k-means, com $k=4$.

Na Figura 24, as bacias foram agrupadas pela distância Euclidiana e Manhattan com o algoritmo de Ward, na Tabela 13 são apresentados os coeficientes estatísticos das regiões, e na Figura 26 a relação entre a Q_{90} observada e estimada.

Tabela 12 - Características estatísticas das regiões determinadas usando o algoritmo de Ward de agrupamento combinado com a distância Euclidiana e Manhattan

Região	n	Equação Q_{90}	H	R^2_a	R^2	c	$\log C_{NS}$	EMA (m^3s^{-1})
I	41	$2,6 \cdot 10^{-4} \cdot (P^{0,983})$	-0,16	0,62	0,60	0,54	0,65	3,41
II	30	$7,0 \cdot 10^{-4} \cdot (A^{1,228})$	-0,09	0,63	0,64	0,74	0,64	3,20
III	21	$4,5 \cdot 10^{-3} \cdot (A^{0,949})$	-0,32	0,72	0,73	0,75	0,73	4,23
IV	8	$6,16 \cdot 10^{-34} \cdot (P^{1,6} \cdot p^{9,474})$	-0,18	0,73	0,81	0,82	0,81	24,14

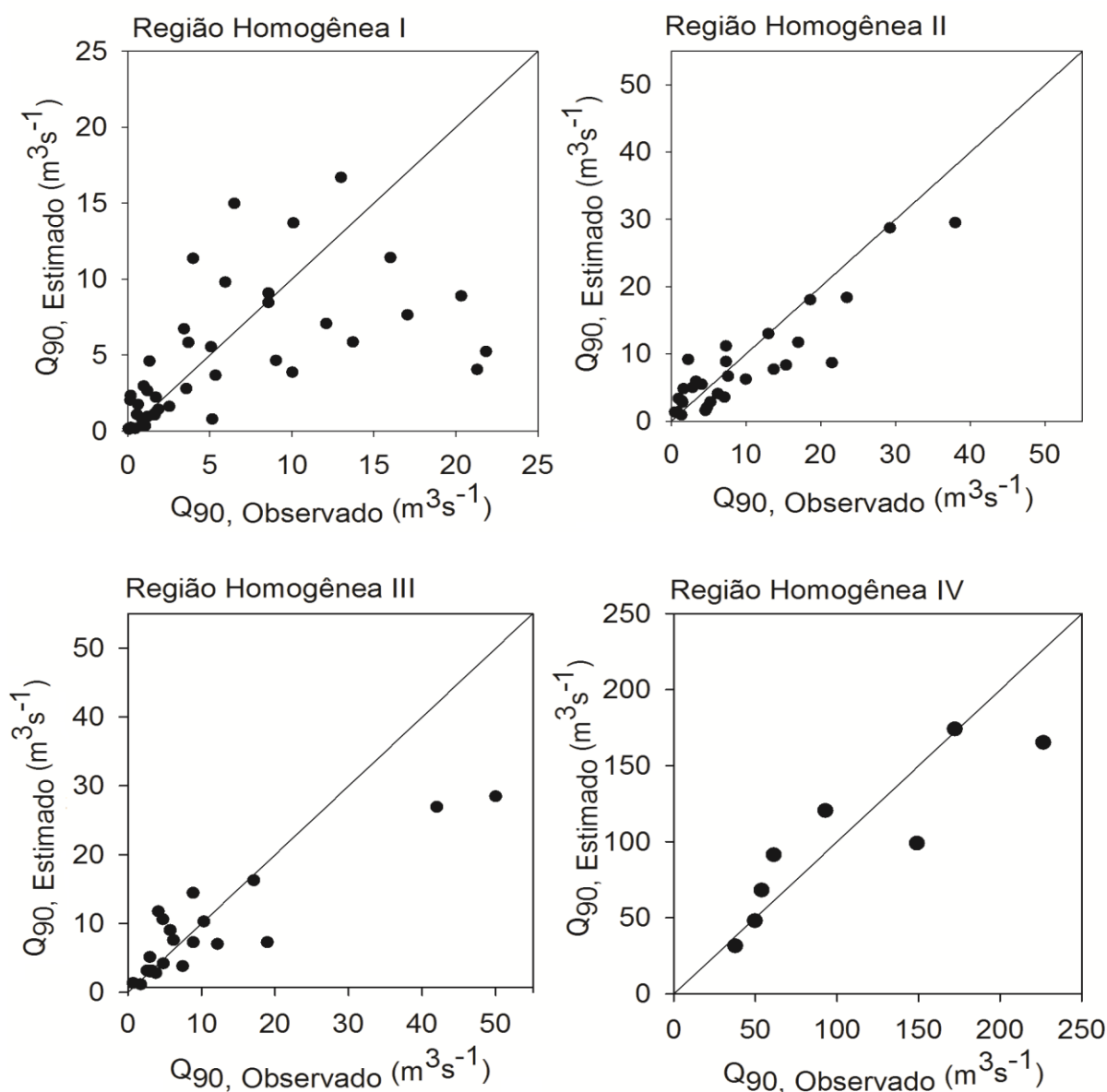


Figura 26 – Relação dos valores da Q_{90} observados e estimados (m^3s^{-1}) para as regiões homogêneas determinadas pela distância Euclidiana e Manhattan, com o algoritmo de Ward de agrupamento.

As duas distâncias Euclidiana e Manhattan produziram formações idênticas, como já observado nas divisões anteriores analisadas. Na Tabela 13 observa-se que a região I apresenta o coeficiente de Camargo; Sentelhas (1997) “c” como sofrível, sendo que o $\log C_{NS}$ é considerado aceitável. As demais regiões apresentam o coeficiente de desempenho “c” de bom a muito bom, e todas as regiões quando avaliadas pela medida H foram aceitas como regiões homogêneas.

Os limites de aplicação para as equações são: para a região homogênea IV, o perímetro varia de 767,68 a 1533,71km, e a precipitação de 1455,16 a 1693,62mm; para a região homogênea III possui área de 349,72 a 10153,78km²; para a região homogênea II, a área varia de 337,09 a 5833,55km²; e para a região homogênea I, os perímetros variam de 40,46 a 614,93km.

A Figura 26 ilustra a relação entre a Q_{90} observada e estimada para as regiões homogêneas, observa-se que agora a maior dispersão dos pontos encontra-se na região homogênea I, que por consequência é a região com coeficientes estatísticos de menor desempenho.

Na Figura 25 estão apresentadas as bacias que contemplam as regiões pelo método não hierárquico k-means, para k=4. O número de bacias não diferiu muito dos métodos hierárquicos (Figura 24), fato que também ocorreu com as divisões anteriores. Na Tabela 14 estão apresentadas as características das quatro regiões formadas e na Figura 27 é mostrada a relação entre a vazão Q_{90} observada e estimada.

Tabela 13 - Características estatísticas das regiões determinadas pelo método k-means com k=4

Região	n	Equação Q_{90}	H	R^2_a	R^2	c	$\log C_{NS}$	EMA (m ³ s ⁻¹)
I	37	$1,17 \cdot 10^{-4} \cdot (P^{1,826})$	-0,09	0,68	0,69	0,58	0,69	3,10
II	32	$8,85 \cdot 10^{-4} \cdot (A^{1,191})$	-0,07	0,61	0,63	0,72	0,63	3,52
III	23	$1,32 \cdot 10^{-3} \cdot (P^{1,54})$	-0,20	0,62	0,64	0,66	0,64	4,87
IV	8	$6,16 \cdot 10^{-34} \cdot (P^{1,6} \cdot p^{9,474})$	-0,18	0,73	0,81	0,82	0,81	24,14

Na Tabela 14 observa-se que novamente a região I apresenta o coeficiente de Camargo; Sentelhas (1997) “c” como sofrível, sendo que o valor do $\log C_{NS}$ é considerado aceitável, segundo a classificação de Motovilov *et al.* (1999), a região é considerada homogênea, assim como as demais. Como notado anteriormente esta indicação talvez seja um indício de que a região precise sofrer alterações para que

possa ser utilizada para fins de planejamento dos recursos hídricos, Farsadnia *et al.* (2014) sugere que seja realizado o realocamento de bacias.

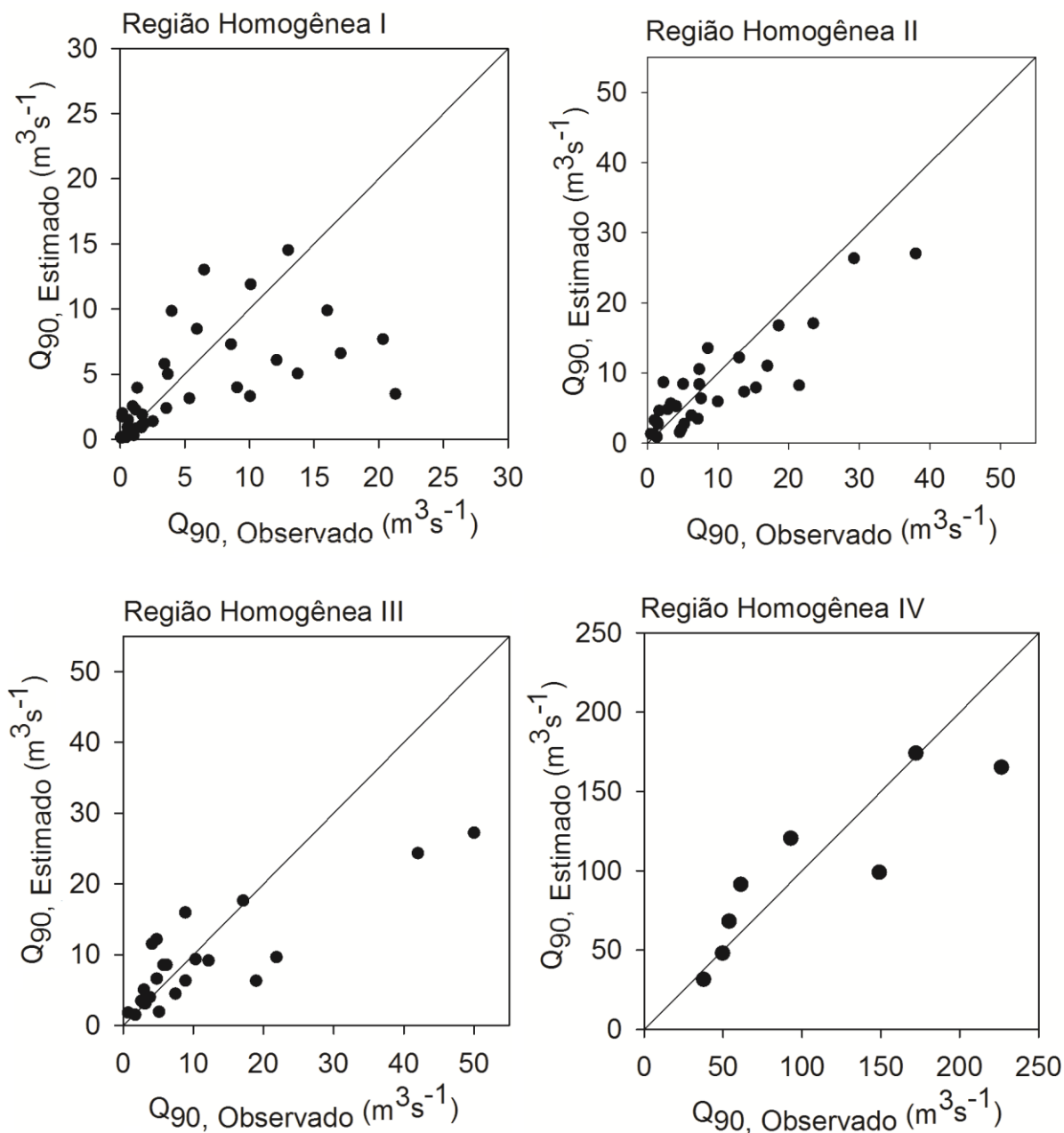


Figura 27 – Relação dos valores da Q_{90} observados e estimados (m^3s^{-1}) para as regiões homogêneas pelo método k-means com $k=4$.

Entretanto, os limites de aplicação para as equações da Tabela 13 são: para as regiões homogêneas I o perímetro varia de 40,46 a 614,93km; região homogênea II a área de drenagem varia de 337,09 a 5833, 55km²; região homogênea III, o perímetro varia de 113,26 a 447,94km; e a região homogênea IV conforme a variação anterior.

Em comparação com as regiões formadas pelos métodos hierárquicos e não hierárquicos é possível notar que a região I precisa ser aperfeiçoada para apresentar coeficientes estatísticos de melhor ajuste. Também foi observado quando a divisão do estado era de 3 regiões, lembrando que as regiões são ordenadas em ordem crescente de bacias, desta forma, a região I (divisão de 4 regiões) e a região II (divisão de 3 regiões) apresentam características semelhantes, indicando que há necessidade de buscar nova divisão, sendo assim foram identificadas todas as alternativas com a formação de 5 regiões para o Estado (Figuras 28, 29 e 30).

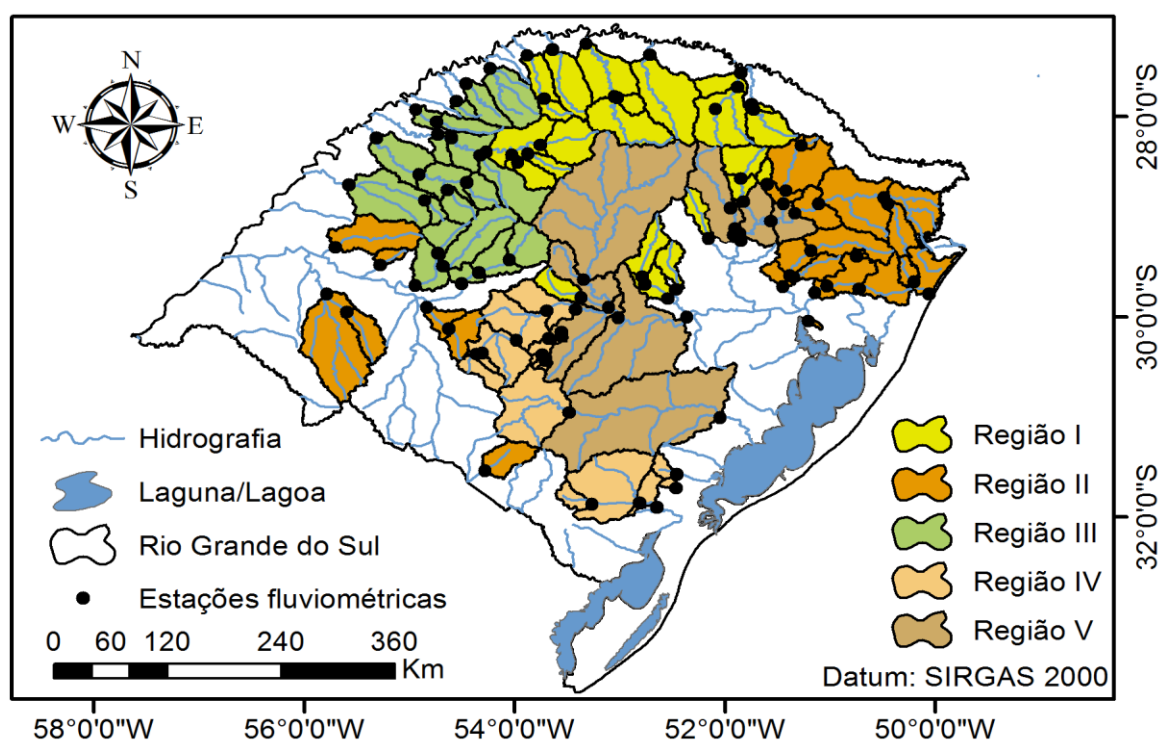


Figura 28 – Regiões para o estado do Rio Grande do Sul determinadas usando o algoritmo de Ward de agrupamento combinado com a distância Euclidiana.

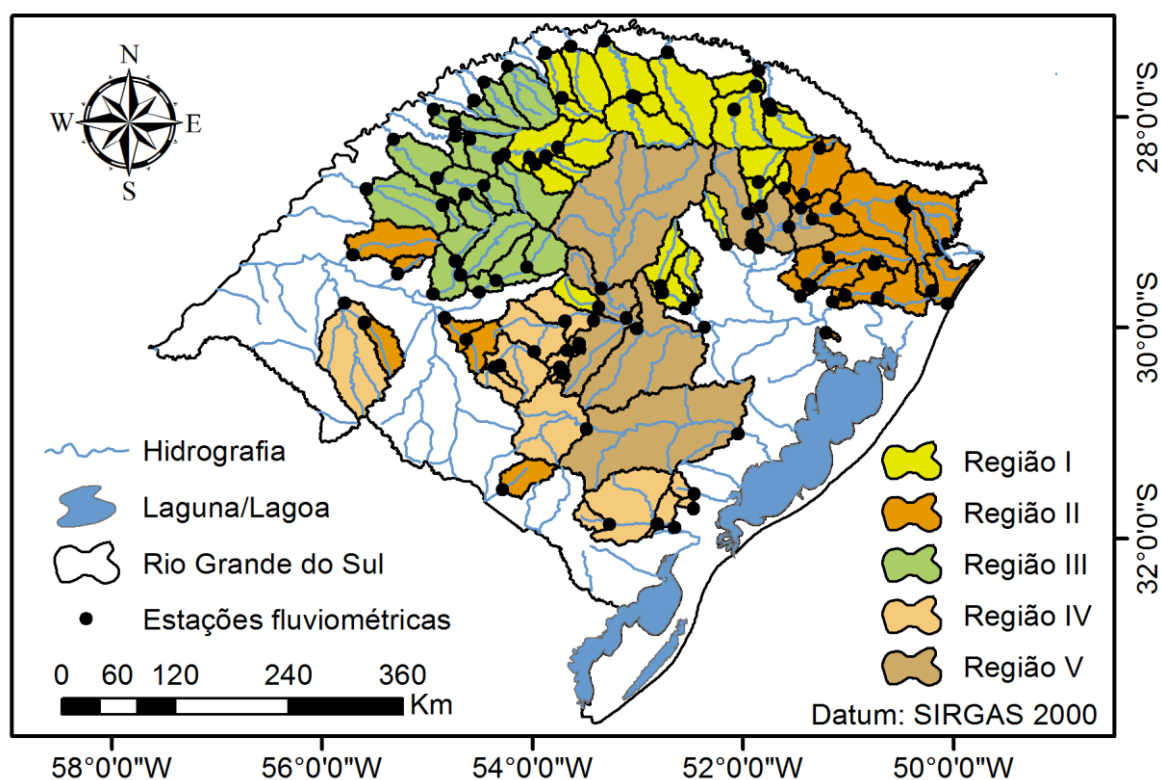


Figura 29 - Regiões para o estado do Rio Grande do Sul determinadas usando o algoritmo de Ward de agrupamento combinado com a distância Manhattan.

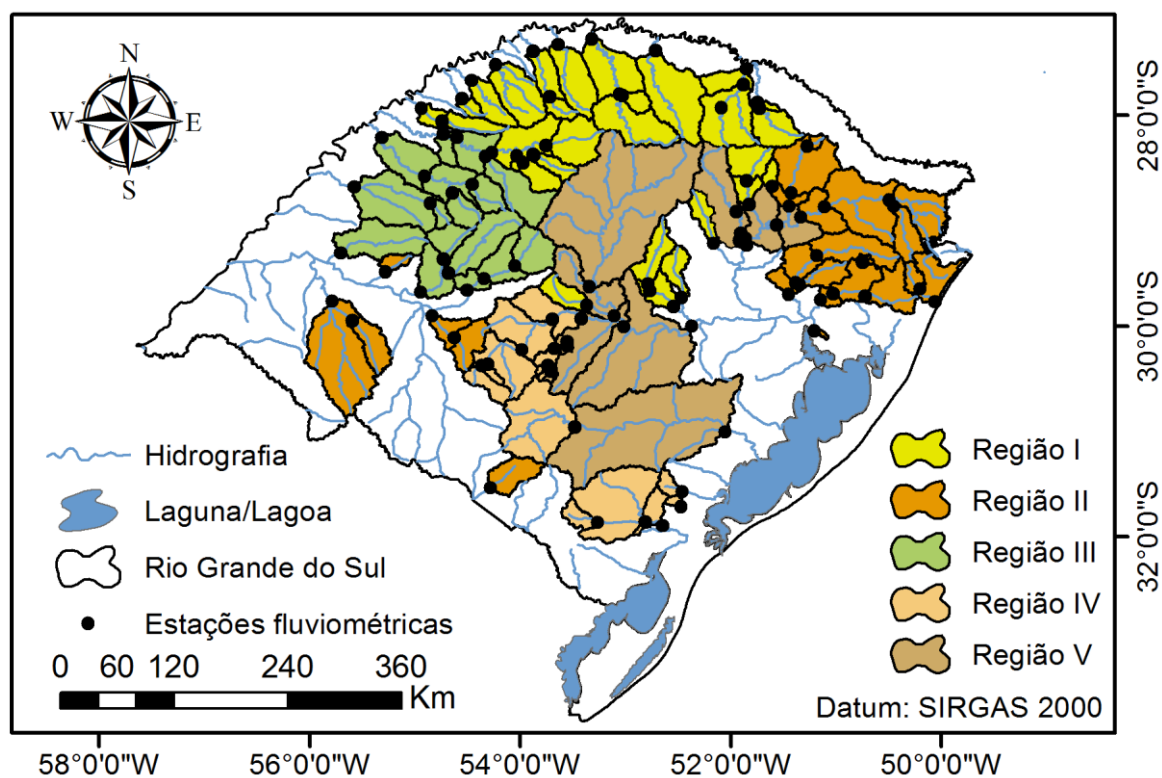


Figura 30 - Regiões determinadas para o estado do Rio Grande do Sul pelo método k-means, com k=5.

Na Figura 28, as bacias foram agrupadas pela distância Euclidiana com o algoritmo de Ward e na Tabela 15 estão apresentadas as características das 5 regiões formadas por esta configuração.

Tabela 14 - Características estatísticas das regiões determinadas usando o algoritmo de Ward de agrupamento combinado com a distância Euclidiana

Região	n	Equação Q_{90}	H	R^2_a	R^2	c	log C_{NS}	EMA (m^3s^{-1})
I	30	$7,0 \cdot 10^{-4} \cdot (A^{1,228})$	-0,09	0,63	0,64	0,74	0,64	3,20
II	25	$9,62 \cdot 10^{-3} \cdot (A^{0,837})$	-0,21	0,48	0,50	0,46	0,52	4,27
III	21	$4,49 \cdot 10^{-3} \cdot (A^{0,949})$	-0,32	0,72	0,73	0,75	0,73	4,23
IV	16	$8,86 \cdot 10^{-4} \cdot (A^{1,027})$	36,49	0,84	0,85	0,82	0,85	1,18
V	8	$6,16 \cdot 10^{-34} \cdot (P^{1,6} \cdot p^{9,474})$	-0,18	0,73	0,81	0,82	0,81	24,14

Observa-se na Tabela 15 um número mais homogêneo de bacias que compõem cada região, contudo a região II permanece com coeficiente não adequado de Camargo; Sentelhas (1997) “c” como mau e a região IV não foi aceita como homogênea pela medida H. Segundo Arsenault; Brissette (2016) e Elesbon *et al.* (2015), diversos métodos de regionalização apresentam dificuldades em definir regiões homogêneas com características hidrológicas similares, sendo particularmente desafiador para a análise de vazões, associadas a cheias e a estiagens. Os autores atribuem esse fato às influências do tamanho das bacias, topografia, solos e geometria da rede de drenagem, havendo a necessidade de identificar tais bacias para que a homogeneidade seja garantida na região. Segundo Azam *et al.* (2017) e Hosking; Wallis (1997) os ajustes para certificar uma região homogênea incluem: a exclusão de bacias ou região; o realocamento de bacias; a subdivisão de uma região e a união de regiões.

Os limites de aplicação para as equações são: para a região homogênea V o perímetro varia de 767,68 a 1533,71km, e a precipitação de 1455,16 a 1693,62mm; e para a região IV, a área varia de 68,61 a 7891,90km²; já para as regiões homogêneas II e III a área varia de 70,68 a 8292,29km² e 349,72 a 10153,78km², respectivamente; e para a região homogênea I a área varia de 337,09 a 5833,55km² (Tabela 15).

Na Figura 31 está apresentada a relação entre a Q_{90} observada e estimada para as regiões, observa-se que agora a maior dispersão dos pontos encontra-se na região homogênea II, ou seja é a região com os coeficientes estatísticos que apresentam os menores valores.

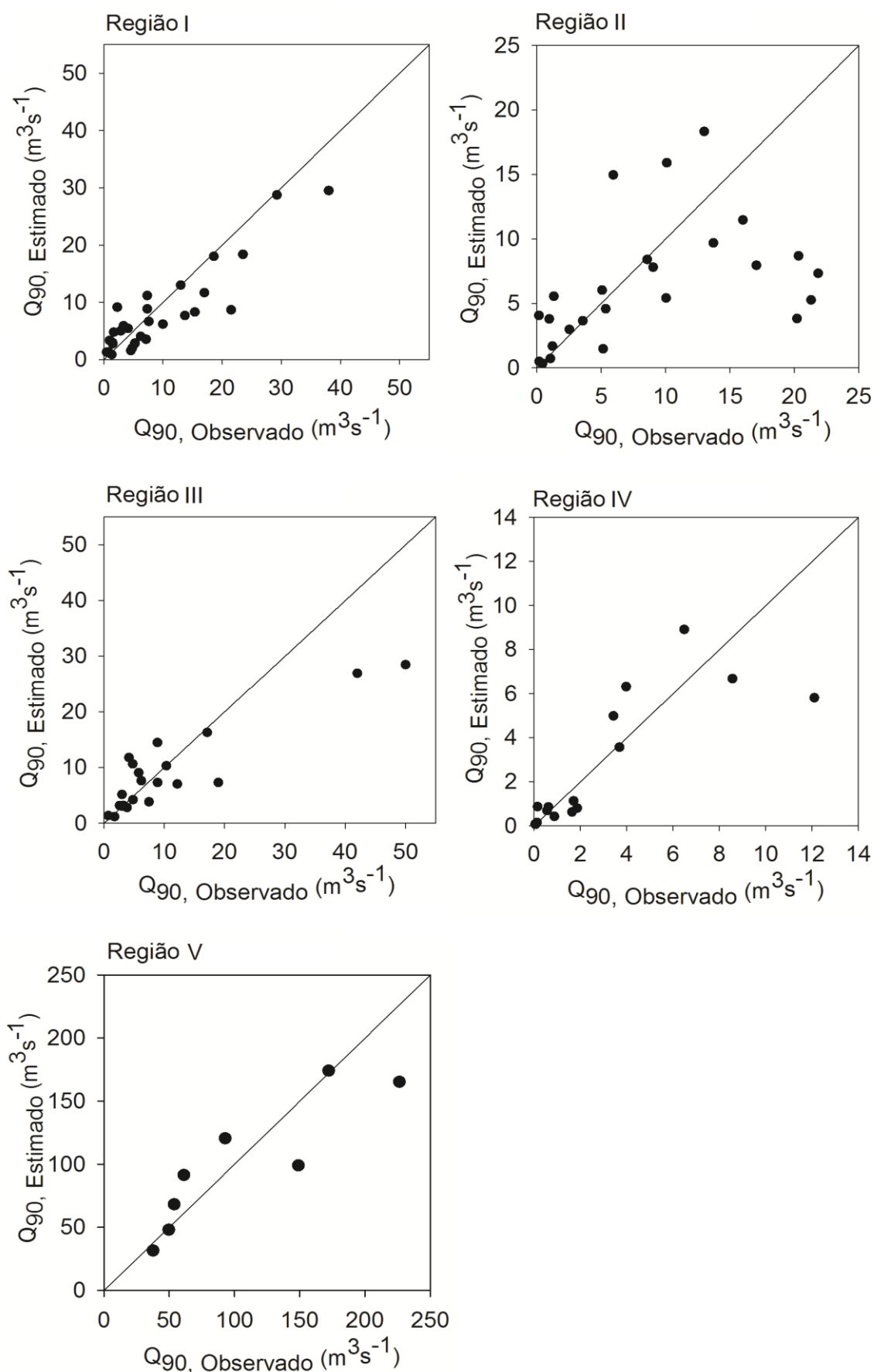


Figura 31 – Relação dos valores da Q_{90} observados e estimados (m^3s^{-1}) para as regiões delimitadas pela distância Euclidiana, com o algoritmo de Ward de agrupamento.

Na Figura 29, as bacias foram agrupadas pela distância Manhattan com o algoritmo de Ward. Até então as distâncias Euclidiana e Manhattan estavam apresentando a mesma configuração de estações, contudo ao separar o Estado em 5 grupos, uma estação (76750000) migrou da região II para a IV, cujos coeficientes estatísticos das 5 regiões são apresentados na Tabela 16.

Tabela 15 - Características estatísticas das regiões determinadas usando o algoritmo de Ward de agrupamento combinado com a distância Manhattan

Região	n	Equação Q_{90}	H	R^2_a	R^2	c	$\log C_{NS}$	EMA (m^3s^{-1})
I	30	$7,0 \cdot 10^{-4} \cdot (A^{1,228})$	-0,09	0,63	0,64	0,74	0,64	3,20
II	24	$7,4 \cdot 10^{-3} \cdot (A^{0,878})$	-0,08	0,50	0,52	0,50	0,53	4,10
III	21	$4,49 \cdot 10^{-3} \cdot (A^{0,949})$	-0,32	0,72	0,73	0,75	0,73	4,23
IV	17	$9,34 \cdot 10^{-4} \cdot (A^{1,018})$	31,67	0,85	0,86	0,83	0,86	1,17
V	8	$6,16 \cdot 10^{-34} \cdot (P^{1,6} \cdot p^{9,474})$	-0,18	0,73	0,81	0,82	0,81	24,14

As regiões I, III e V são iguais pela formação da distância Euclidiana (Figura 28), contudo a mudança de uma estação não contribuiu para a melhora dos coeficientes estatísticos das equações das regiões II e IV. A região IV não foi aceita novamente como homogênea pela medida H. Os valores limitantes das variáveis explicativas nas equações são iguais ao método anterior (Tabela 15), modificando apenas o número de bacias que contemplam as regiões II e IV.

Na Figura 32 observa-se novamente que a maior dispersão dos pontos encontra-se na região homogênea II, conseqüentemente, é a região com coeficientes estatísticos de menor valor, indicando mais uma vez, que esta região precisa ser aperfeiçoada para a gestão dos recursos hídricos. Rajsekhar *et al.* (2013) lembram que para cada uma dessas regiões, um plano de mitigação específico deverá ser elaborado, caso a demanda supere a disponibilidade hídrica nos próximos anos ou até mesmo quando contabilizado os processos de outorgas. Então as regiões homogêneas devem apresentar os melhores coeficientes possíveis uma vez que, ajudará os órgãos gestores a contornar tal situação. Entretanto, o que se observa até o momento nos planos de bacias que já foram aprovados, é que se, tal situação venha acontecer, os mesmos especificam que a vazão de referência se torne ainda mais restritiva. Desta forma, este estudo busca resultados coerentes e que possam ser utilizados pelos comitês, que são responsáveis pelos planos, assim como para os órgãos responsáveis pela sua avaliação e aprovação.

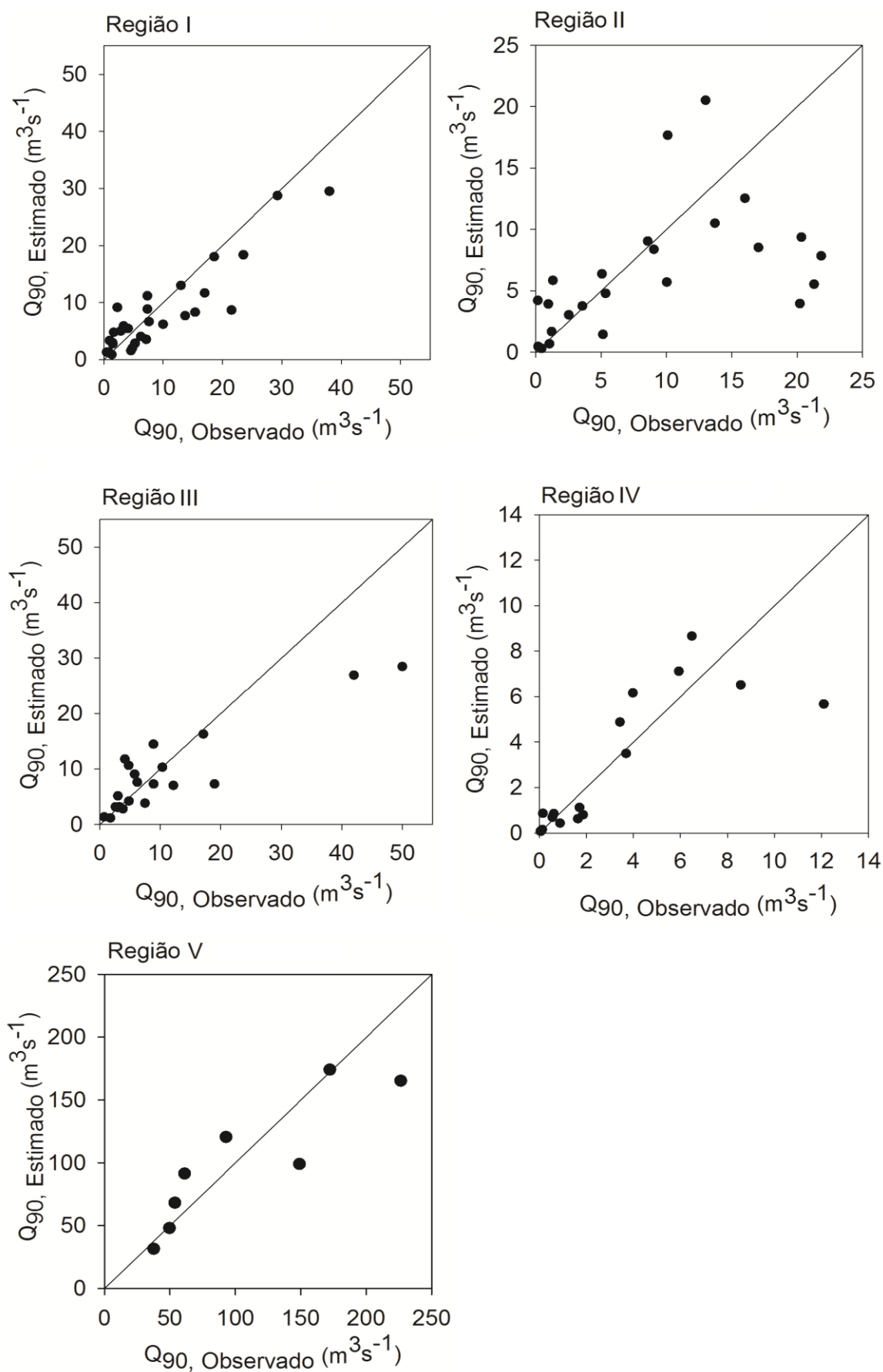


Figura 32 - Relação dos valores da Q_{90} observados e estimados (m^3s^{-1}) para as regiões pela distância Manhattan, com o algoritmo de Ward de agrupamento.

Na Figura 30 estão apresentadas as estações que contemplam as regiões pelo método não hierárquico k-means, para $k=5$. Observa-se que a região IV permanece com a mesma configuração do método hierárquico com a distância Euclidiana, e a região II com a distância Manhattan, já a região V é igual em todos os métodos. As características das regiões estão apresentadas na Tabela 17.

Tabela 16 - Características estatísticas das regiões determinadas pelo método k-means com $k=5$

Região	n	Equação Q_{90}	H	R^2_a	R^2	c	log C_{NS}	EMA (m^3s^{-1})
I	36	$1,01 \cdot 10^{-3} \cdot (A^{1,186})$	-0,16	0,63	0,64	0,74	0,64	3,06
II	24	$1,09 \cdot 10^{-3} \cdot (P^{1,503})$	-0,13	0,49	0,52	0,52	0,54	3,69
III	16	$1,23 \cdot 10^{-4} \cdot (P^{1,928})$	-0,18	0,76	0,78	0,77	0,78	5,05
IV	16	$8,86 \cdot 10^{-4} \cdot (A^{1,027})$	36,49	0,84	0,85	0,82	0,85	1,18
V	8	$6,16 \cdot 10^{-34} \cdot (P^{1,6} \cdot p^{9,474})$	-0,18	0,73	0,81	0,82	0,81	24,14

Na Tabela 17, as estatísticas se assemelham com as configurações anteriores (Tabelas 15 e 16). Diferentemente das equações anteriores onde a variável área apresentava os melhores ajustes, nessa configuração a variável perímetro caracterizou melhor as regiões II e III, tal variável já havia sido inserida no trabalho de Uliana *et al.* (2017) para delimitação de regiões homogêneas, mas não alcançou os melhores ajustes nas equações.

Os limites de aplicação para as equações são: para a região homogênea I a área varia de 337,09 a 5833,55km²; regiões homogêneas II e III o perímetro varia de 46,82 a 614,93km e 109,13 a 633,71km, respectivamente; região IV, a área varia de 68,61 a 7891,90km²; e a região homogênea V igual às demais.

A região IV ainda permanece como heterogênea nessa configuração, situação que se manteve em todos os métodos que proporcionaram a separação de 5 regiões no Estado. Desta forma, seguindo indicações de Azam *et al.* (2017) e Hosking; Wallis (1997) primeiramente, tentou-se dividir a região II em duas, contudo observou-se heterogeneidade na parte da região II localizada a sudoeste, com $H=6,87$, indicando que a separação da região não foi apropriada. Seguidamente, uniu-se parte da região II com a região IV e obteve-se $H=40,18$, sendo assim até o momento verifica-se que as tentativas de eliminar a heterogeneidade não foram alcançadas. A Figura 33 mostra a relação entre a Q_{90} observada e estimada para as regiões pelo método k-means, com $k=5$.

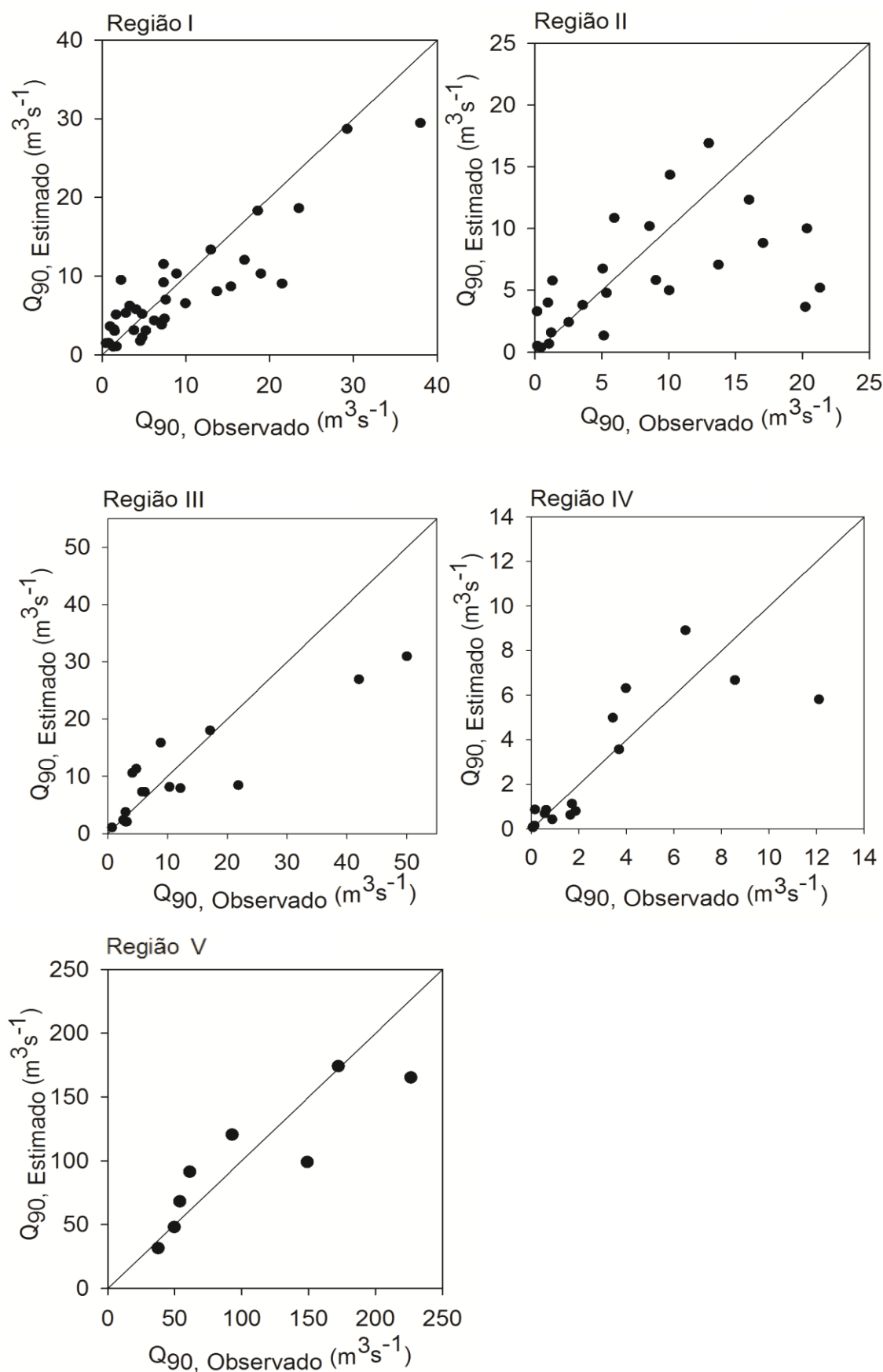


Figura 33 – Relação dos valores da Q_{90} observados e estimados (m^3s^{-1}) para as regiões delimitadas pelo método k-means com $k=5$.

Anteriormente à tentativa de identificar quais estações fluviométricas estavam causando a heterogeneidade na região IV, aplicou-se método não hierárquico k-means com $k=6$ (Figura 34), uma vez que pelos cortes dos dendrogramas não foi possível obter esse número de grupos.

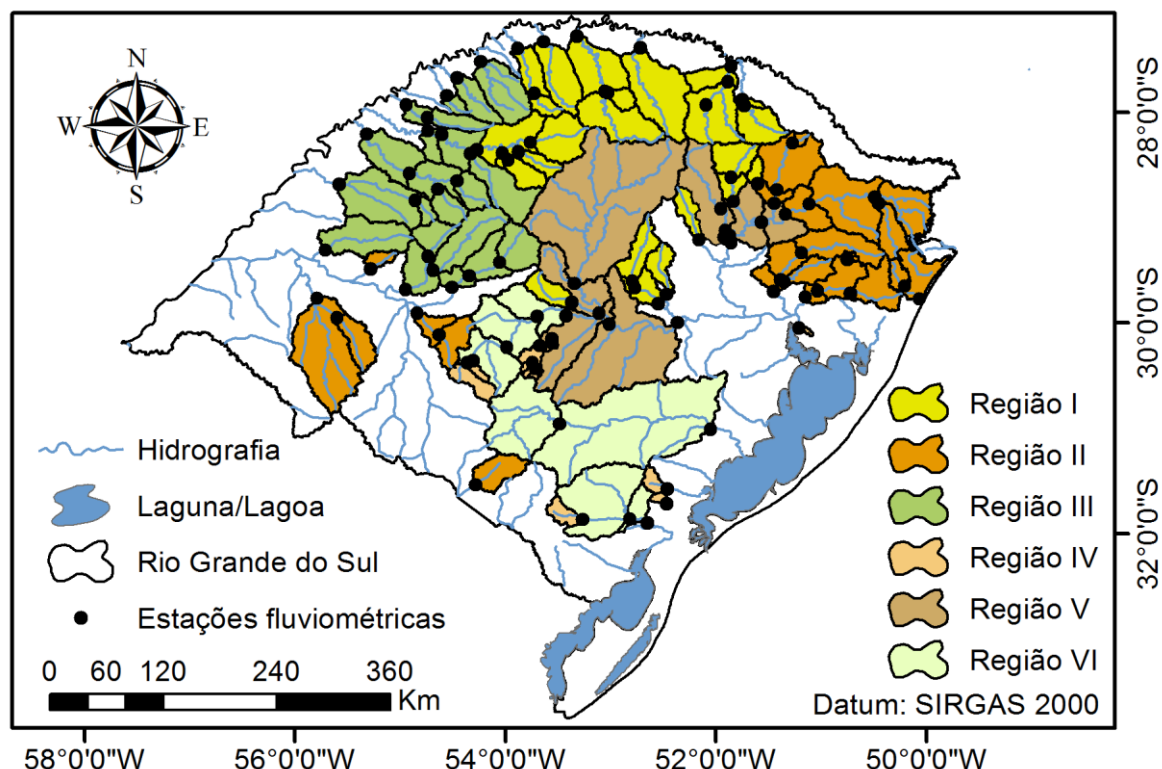


Figura 34 - Regiões determinadas para o estado do Rio Grande do Sul pelo método k-means com $k=6$.

Com a formação de 6 regiões para o estado do Rio Grande do Sul, observou-se que a heterogeneidade da região IV permanece (Tabela 18), indicando que as estações que causam a heterogeneidade precisam ser identificadas e retiradas. Com a atual configuração notou-se que a região IV ficou dividida com a região VI, sugerindo que essas duas regiões formem uma única região homogênea.

Beskow *et al.* (2016), em estudo da Q_{90} no estado do Rio Grande do Sul, também estabeleceram 6 regiões utilizando técnicas de inteligência artificial, contudo notaram a presença de uma região heterogênea, quando aplicada a medida de heterogeneidade H .

Tabela 17 - Características estatísticas das regiões determinadas pelo método k-means com k=6

Região	n	Equação Q_{90}	H	R^2_a	R^2	c	$\log C_{NS}$	EMA (m^3s^{-1})
I	30	$7,00 \cdot 10^{-4} \cdot (A^{1,228})$	-0,09	0,63	0,64	0,74	0,64	3,20
II	23	$5,38 \cdot 10^{-4} \cdot (P^{1,605})$	-0,10	0,48	0,50	0,48	0,50	3,84
III	22	$4,15 \cdot 10^{-3} \cdot (A^{0,965})$	-0,23	0,70	0,71	0,74	0,71	4,54
IV	11	$3,16 \cdot 10^{-3} \cdot (A^{0,839})$	26,44	0,45	0,51	0,47	0,51	0,43
V	7	$4,43 \cdot 10^{-5} \cdot (A^{1,378})$	-0,20	0,66	0,72	0,81	0,72	3,36
VI	7	$1,62 \cdot 10^{-3} \cdot (A^{1,11})$	0,05	0,87	0,46	0,53	0,46	37,77

Observa-se na Tabela 18 que apenas para a região II a variável perímetro obteve o melhor ajuste, enquanto que para as demais a variável área de drenagem foi selecionada. As regiões II, IV e VI apresentaram coeficiente de desempenho “c” como mau e sofrível segundo Camargo; Sentelhas (1997), o que não favorece a divisão de 6 regiões para o Estado.

Os limites de aplicação para as equações são: para as regiões homogêneas I, III, IV, V e VI a área de drenagem varia de 337,09 a 5833,55km², 349,72 a 10153,78km², 68,61 a 1058,13km², 3234,32 a 17144,18km e 13472,33 a 38096,24 km², respectivamente; e a região homogênea II, o perímetro varia de 58,58 a 498,21km. A Figura 35 ilustra a relação entre a Q_{90} observada e estimada para as regiões geradas pelo k-means, com k=6.

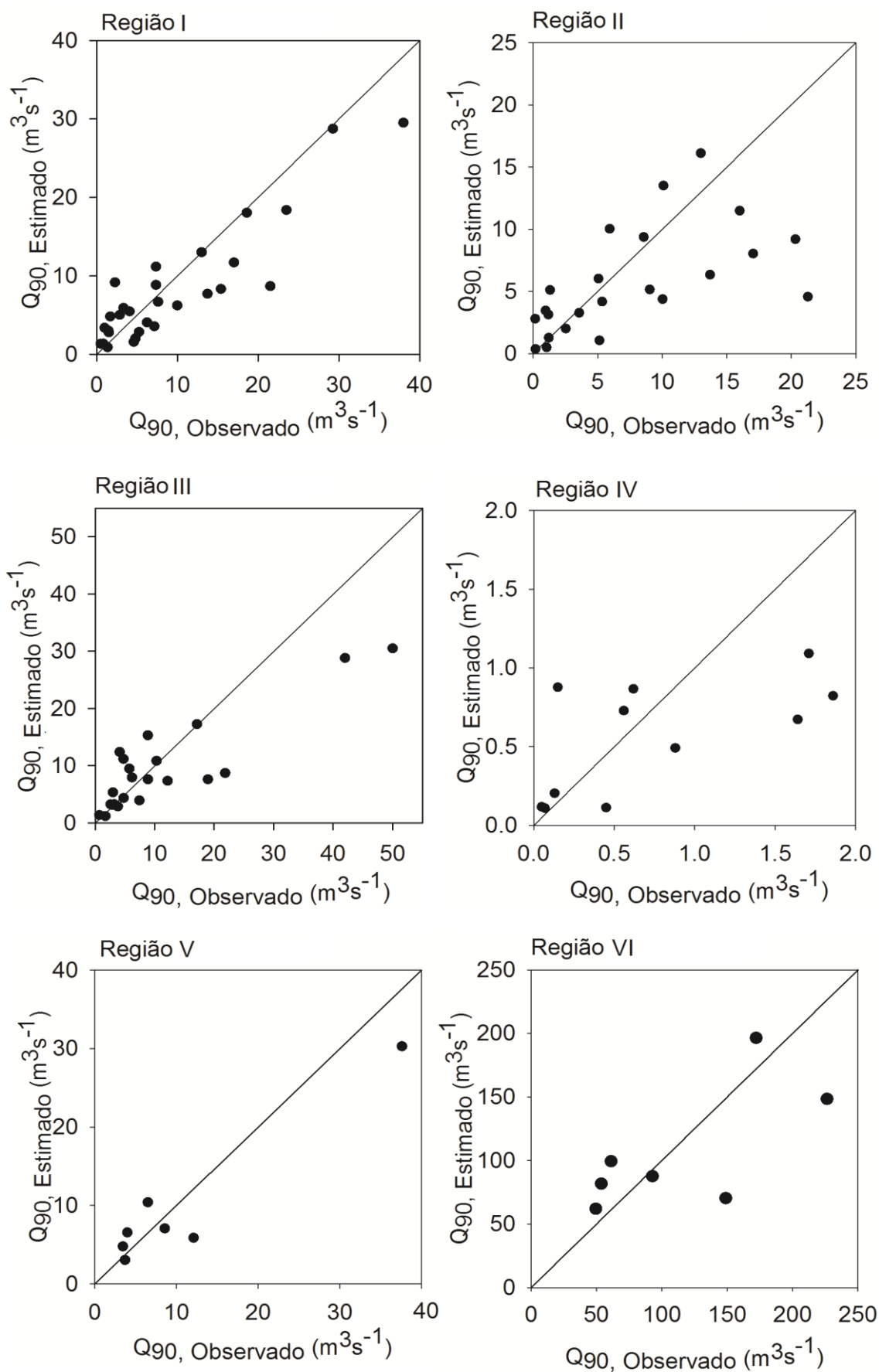


Figura 35 – Relação dos valores da Q_{90} observados e estimados (m^3s^{-1}) para as regiões determinadas pelo método k-means com $k=6$.

Com a tentativa de reunir as estações em 6 regiões homogêneas, observou-se que os resultados não se mostraram adequados, e a divisão de 3 e 4 regiões indicavam a necessidade de mais divisões, então retornou-se para a divisão de 5 regiões no Estado, com o objetivo de identificar as estações que causavam a heterogeneidade. Na tentativa de obter as regiões homogêneas também foi realizado o realocamento de estações como indicado por Azam *et al.* (2018) e Farsadnia *et al.* (2014). Sendo assim, as estações da região II, que não apresentaram coeficientes estatísticos adequados, foram realocadas até que fosse atingido melhores coeficientes para as equações geradas. Desta forma, a Figura 36 representa a melhor configuração para as 5 regiões homogêneas no Estado utilizando o algoritmo de Ward de agrupamento combinado com a distância Euclidiana e Manhattan, respectivamente.

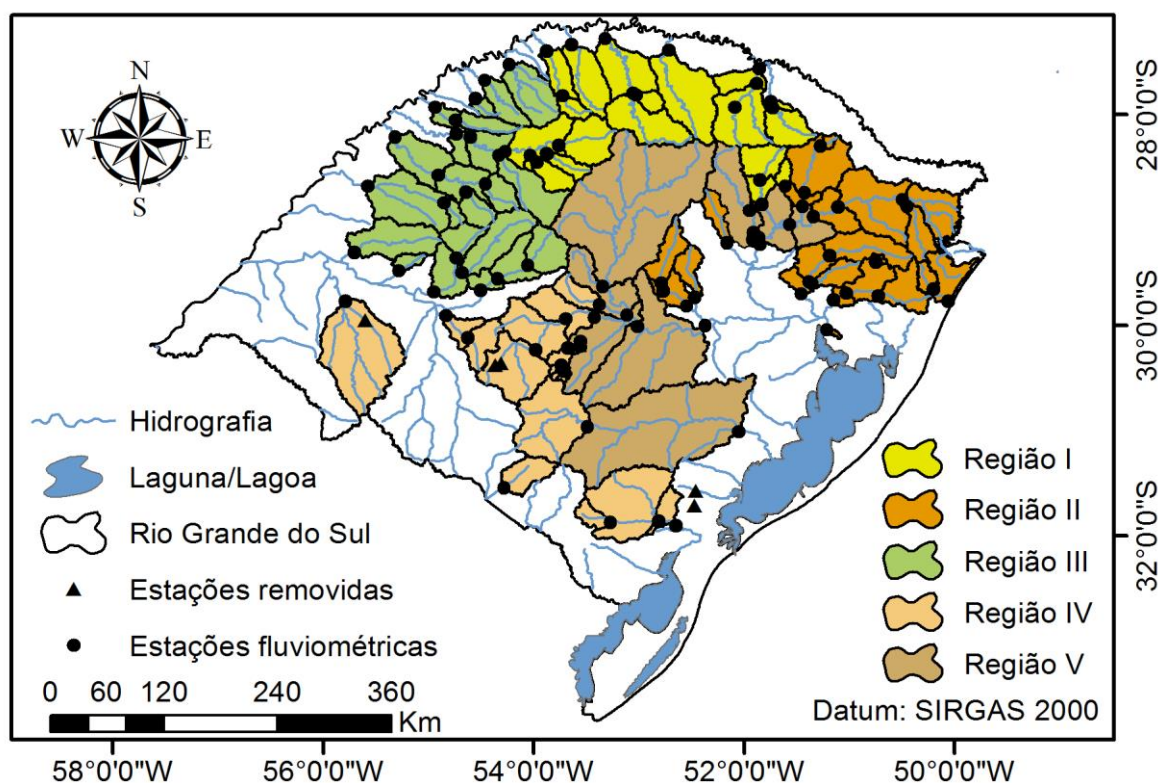


Figura 36 - Regiões homogêneas para o estado do Rio Grande do Sul determinadas usando o algoritmo de Ward de agrupamento combinado com a distância Euclidiana e Manhattan.

Primeiramente tentou-se apenas realizar o realocamento das estações da região II para a região IV, como testado anteriormente (k-means, com k=5), permanecendo contudo, a heterogeneidade da região. Desta maneira, foram analisadas as estações uma a uma e assim identificadas e removidas as estações

85460000, 85470000, 88750000, 88850000 e 76742000, para que a região formada fosse considerada homogênea, ou seja, com $H < 1$. As características das regiões homogêneas estão apresentadas na Tabela 19.

Este mesmo procedimento foi realizado nos estudos de Beskow *et al.* (2016); Farsadnia *et al.* (2014) e Rao; Srinivas (2006), que também observaram que, mesmo após a definição de grupos de bacias hidrográficas usando técnicas de agrupamento, os grupos resultantes podem ainda não atender aos requisitos de homogeneidade. Portanto, nos estudos dos autores mencionados, foi enfatizado que a revisão dos grupos formados é de extrema importância.

Tabela 18 - Características estatísticas das regiões determinadas usando o algoritmo de Ward de agrupamento combinado com a distância Euclidiana e Manhattan

Região	n	Equação Q_{90}	H	R^2_a	R^2	c	log C_{NS}	EMA (m^3s^{-1})
I	24	$5,78 \cdot 10^{-5} \cdot (P^{2,169})$	-0,11	0,75	0,76	0,81	0,76	3,22
II	23	$1,32 \cdot 10^{-3} \cdot (P^{1,54})$	-0,20	0,72	0,74	0,66	0,64	4,87
III	23	$5,10 \cdot 10^{-3} \cdot (A^{0,923})$	-0,09	0,77	0,78	0,66	0,78	3,32
IV	17	$8,52 \cdot 10^{-4} \cdot (A^{1,012})$	0,64	0,80	0,82	0,78	0,82	1,09
V	8	$6,16 \cdot 10^{-34} \cdot (P^{1,6} \cdot p^{9,474})$	-0,18	0,73	0,81	0,82	0,81	24,14

Na Tabela 19 estão apresentadas as características das 5 regiões, sendo que todas apresentam coeficiente de desempenho “c” de bom a muito bom, conforme Camargo; Sentelhas (1997), o log C_{NS} , segundo a classificação de Motovilov *et al.* (1999), encontra-se como “aceitável” e “adequado e bom”. Apenas a equação da região V tem duas variáveis, a qual permaneceu constante desde a divisão do Estado para duas regiões. As demais equações empregaram uma única variável, alternando entre a área de drenagem e o perímetro. Segundo Elesbon *et al.* (2015), os quais recomendam que o R^2_a seja superior a 0,7, fato que ainda não tinha sido alcançado em todas as equações para as divisões das regiões anteriores.

Os limites de aplicação das equações das 5 regiões homogêneas para Q_{90} no estado do Rio Grande do Sul são: região homogênea I e II o perímetro varia de 97,78 a 468,82km, 96,40 a 633,71km, respectivamente; região homogênea III e IV a área varia de 70,68 a 8292,29km², 68,61 a 7891,90km², respectivamente; e região homogênea V o perímetro varia de 767,68 a 1533,71km, e a precipitação de 1455,16 a 1693,62mm. A Figura 37 ilustra a relação entre a Q_{90} observada e estimada para todas as regiões homogêneas.

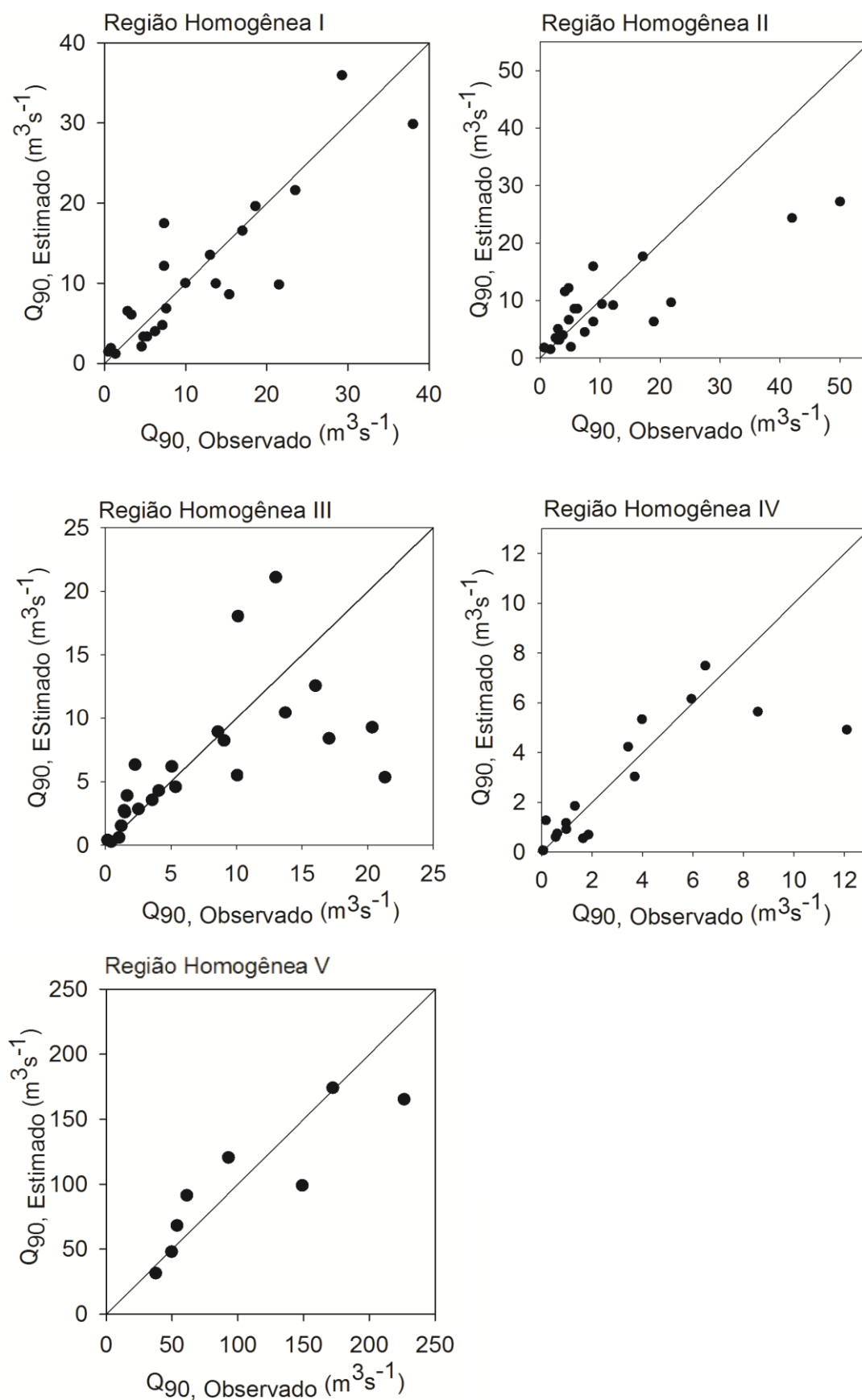


Figura 37 - Relação dos valores da Q_{90} observados e estimados (m^3s^{-1}) para as regiões homogêneas determinadas pela distância Euclidiana e Manhattan, com o algoritmo de Ward de agrupamento.

4.4. Regiões homogêneas para a Q_{95}

De posse dos dendrogramas para a Q_{95} (Figuras 9, 10 e 11), optou-se em testar todas as possíveis configurações de grupos disponíveis pelos cortes dos dendrogramas para esta vazão de permanência. Na sequência são apresentadas as regiões para a Q_{95} no estado do Rio Grande do Sul e as equações de regressão originárias do método Stepwise com as variáveis independentes (A, P e p), além das análises estatísticas.

Primeiramente, foram identificadas todas as alternativas com a formação de 2 regiões. Observou-se que a separação das regiões da vazão Q_{95} foram iguais a vazão Q_{90} para o método hierárquico, com as distâncias Euclidiana e Manhattan, com os algoritmos de ligação completa (Figura 12) e Ward (Figura 14), e o método não hierárquico k-means (Figura 15), diferindo apenas do método hierárquico com a distância de Mahalanobis e algoritmo de Ward (Figura 39).

Como mencionado, a Figura 12 apresenta também a disposição das bacias para a vazão Q_{95} pela distância Euclidiana e Manhattan com o algoritmo de ligação completa. Observou-se que as duas distâncias produziram formações idênticas para a vazão Q_{90} e Q_{95} . Na Tabela 20 estão apresentadas as características para as equações das regiões da Q_{95} .

Tabela 19 - Características estatísticas das regiões determinadas usando o algoritmo de ligação completa de agrupamento combinado com a distância Euclidiana e Manhattan

Região	n	Equação Q_{95}	H	R^2_a	R^2	c	log C_{NS}	EMA ($m^3 s^{-1}$)
I	92	$4,64 \cdot 10^{-16} \cdot (A^{0,973} \cdot p^{3,936})$	-0,19	0,59	0,60	0,63	0,61	3,11
II	8	$3,19 \cdot 10^{-36} \cdot (P^{1,337} \cdot p^{10,394})$	-0,18	0,72	0,80	0,80	0,80	19,12

Para a região I, o coeficiente de desempenho de Camargo; Sentelhas (1997) “c” foi considerado mediano e o log C_{NS} como aceitável, conforme a classificação de Motovilov *et al.* (1999). Apesar da região não apresentar o melhor desempenho, foi aceita como homogênea pela medida H, assim como a região II.

A região II constitui-se das mesmas bacias da Q_{90} , contudo em função da Q_{95} apresentar valores ainda menores que a Q_{90} , o erro consequentemente ainda permaneceu maior, em relação às demais regiões. Observa-se também que para as duas regiões a variável precipitação foi incluída para o melhor desempenho.

Os limites de aplicação das equações são: para a região homogênea II o perímetro varia de 767,68 a 1533,71km e a precipitação média de 1455,16 a 1693,62mm; e para a região homogênea I as áreas de drenagem variam de 68,61 a 10153,78km² e a precipitação média 1172,11 a 1914,73mm. A Figura 38 mostra a relação entre a Q₉₅ observada e estimada para ambas as regiões homogêneas.

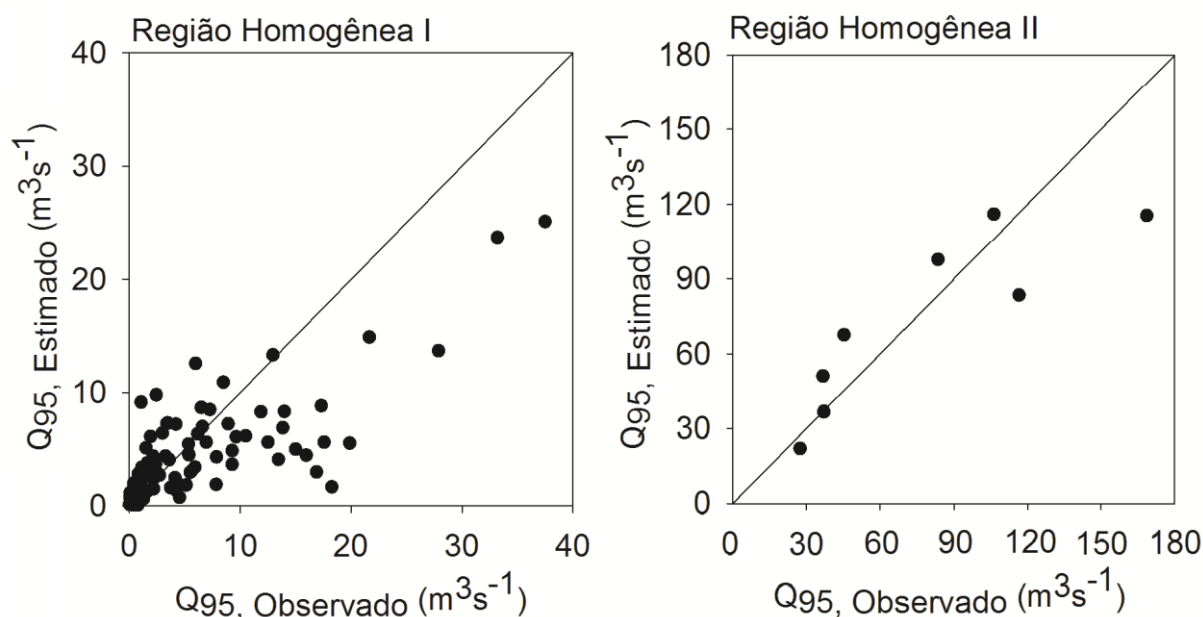


Figura 38 - Relação dos valores da Q₉₅ observados e estimados (m³s⁻¹) para as regiões homogêneas determinadas pela distância Euclidiana e Manhattan, com o algoritmo de ligação completa de agrupamento.

Diferentemente da distância Euclidiana e Manhattan pelos algoritmos de agrupamento aplicados, a distância de Mahalanobis com o algoritmo de Ward para a Q₉₅ (Figura 39) demonstrou comportamento distinto ao da vazão Q₉₀. Tal comportamento refere-se a diferença quanto a disposição e no número das bacias hidrográficas nas respectivas regiões homogêneas encontradas para a vazão.

Na Tabela 21 estão apresentados os coeficientes estatísticos das equações para a Q₉₅ nessa configuração e na Figura 40, a relação entre a Q₉₅ observada e estimada para ambas as regiões.

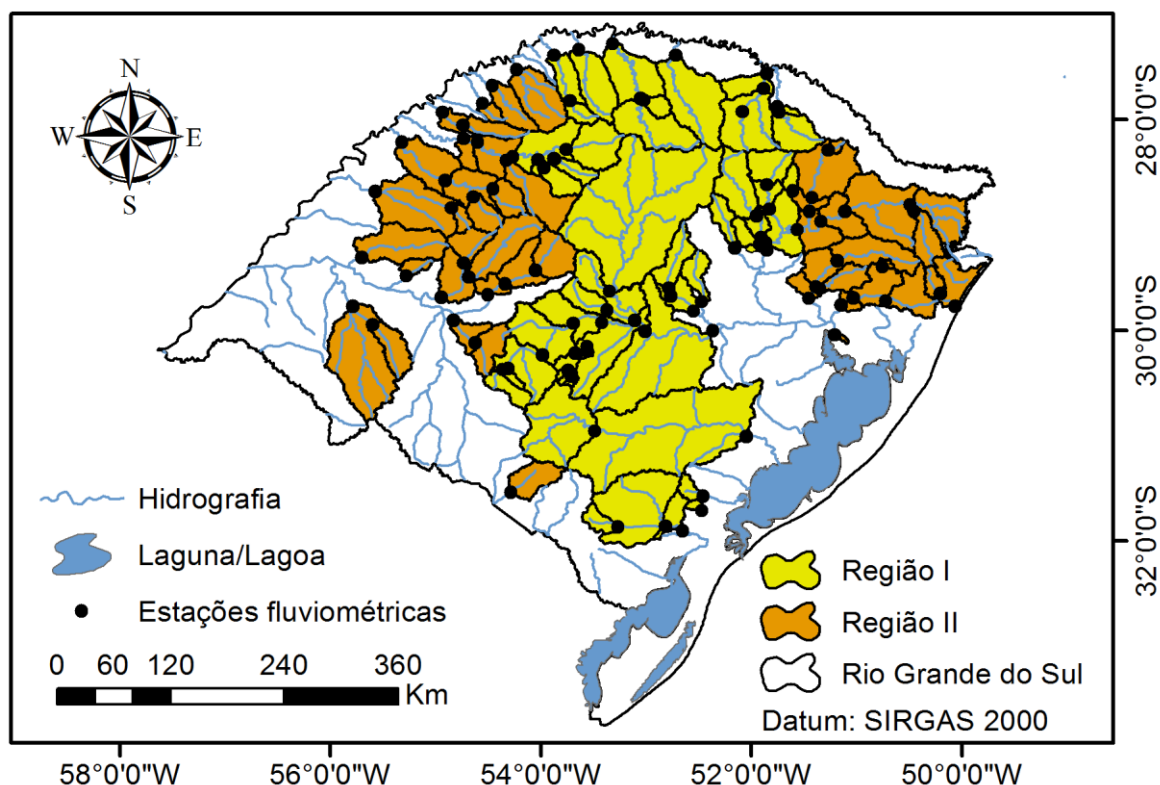


Figura 39 - Regiões para o estado do Rio Grande do Sul determinadas usando o algoritmo de Ward combinado com a distância Mahalanobis, com o algoritmo de Ward de agrupamento.

Tabela 20 - Características estatísticas das regiões determinadas usando o algoritmo de Ward de agrupamento combinado com a distância Mahalanobis

Região	n	Equação Q_{95}	H	R^2_a	R^2	c	$\log C_{NS}$	EMA ($m^3 s^{-1}$)
I	53	$7,55 \cdot 10^{-29} \cdot (A^{1,239} \cdot p^{7,635})$	-0,16	0,82	0,83	0,84	0,83	5,34
II	47	$5,1 \cdot 10^{-3} \cdot (A^{0,881})$	-0,17	0,51	0,52	0,55	0,53	4,04

Observa-se na Tabela 21 que na região II o coeficiente de desempenho “c” de Camargo; Sentelhas (1997) foi classificado como sofrível, o R^2_a e o R^2 também estão abaixo das recomendações, mas segundo a classificação do $\log C_{NS}$ de Motovilov *et al.* (1999) a equação é considerada aceitável. Nessa configuração, apenas a região I apresentou o melhor ajuste, incluindo a variável precipitação, ambas regiões foram consideradas homogêneas, quando analisadas pela medida H.

Os limites de aplicação das equações são: para a região homogênea I as áreas de drenagem variam de 68,61 a 38096,24km² e a precipitação média 1376,92 a 1877,77mm; e para a região homogênea II a área varia de 70,68 a 13472,33km².

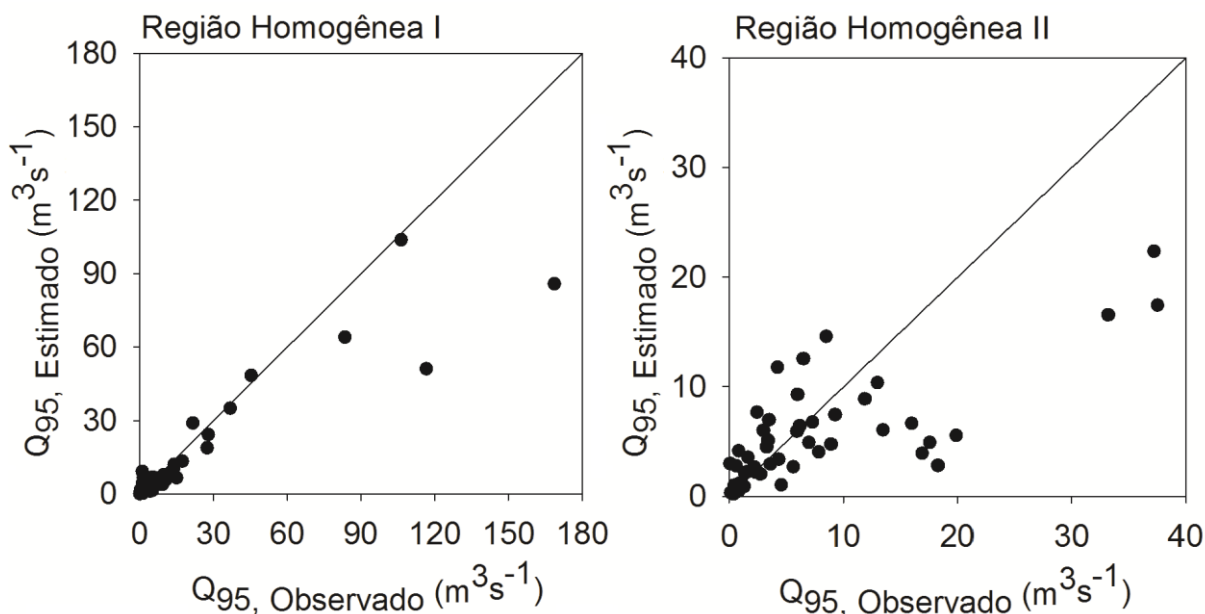


Figura 40 - Relação dos valores da Q_{95} observados e estimados (m^3s^{-1}) para as regiões homogêneas determinadas pela distância Mahalanobis, com o algoritmo de Ward de agrupamento.

Para as distâncias Euclidiana e Manhattan pelo algoritmo Ward, a disposição das bacias nas respectivas regiões são idênticas para as vazões Q_{90} e Q_{95} . Desta forma, a Figura 14 também representa a configuração das bacias para a vazão Q_{95} , contudo as características das equações dessas regiões estão indicadas na Tabela 22.

Tabela 21 - Características estatísticas das regiões determinadas usando o algoritmo de Ward de agrupamento combinado com a distância Euclidiana e Manhattan

Região	n	Equação Q_{95}	H	R^2_a	R^2	c	$\log C_{NS}$	EMA (m^3s^{-1})
I	51	$1,18 \cdot 10^{-3} \cdot (A^{1,094})$	-0,12	0,57	0,58	0,68	0,59	3,32
II	49	$3,93 \cdot 10^{-5} \cdot (P^{1,999})$	-0,21	0,70	0,70	0,71	0,85	8,67

Observa-se que o desempenho da equação para a região II foi superior ao encontrado no método anterior (Tabela 21). Já para os coeficientes da equação da região I foram inferiores, porém ainda aceitáveis conforme Camargo; Sentelhas (1997) e Motovilov *et al.* (1999). O melhor ajuste ocorreu apenas incluindo uma variável explicativa, em cada caso.

Os limites de aplicação para as equações são: para a região homogênea I as áreas de drenagem variam de 337,09 a 10153,78km², e para a região homogênea II o perímetro varia de 40,46 a 1533,71km. A Figura 41 mostra a relação entre a Q_{95} observada e estimada para ambas as regiões homogêneas.

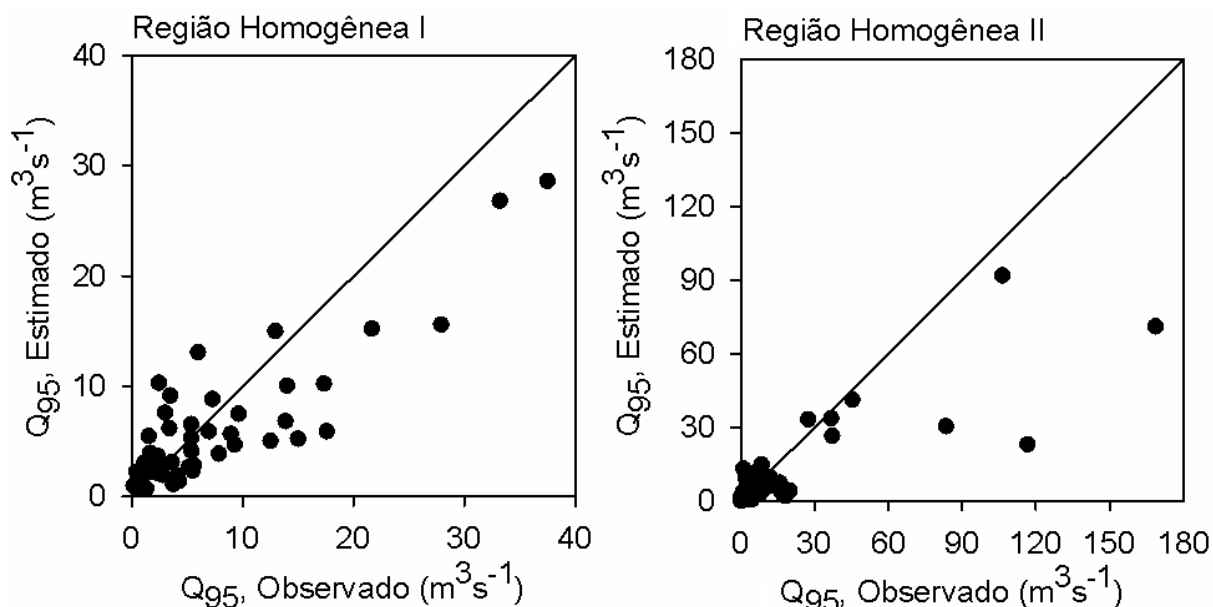


Figura 41 - Relação dos valores da Q_{95} observados e estimados (m^3s^{-1}) para as regiões homogêneas determinadas pela distância Euclidiana e Manhattan, com o algoritmo de Ward de agrupamento.

No método não hierárquico k-means para $k=2$, objetivando encontrar as regiões homogêneas para a Q_{95} observou-se que a disposição das bacias também foi idêntica quando aplicado para a Q_{90} . Desta forma, a Figura 15 também representa a configuração das bacias para a vazão Q_{95} , contudo as características das equações dessas regiões estão indicadas na Tabela 23.

Tabela 22 - Características estatísticas das regiões determinadas pelo método k-means com $k=2$

Região	n	Equação Q_{95}	H	R^2_a	R^2	c	$\log C_{NS}$	EMA ($m^3 s^{-1}$)
I	59	$1,32 \cdot 10^{-3} \cdot (A^{1,085})$	-0,24	0,59	0,60	0,67	0,60	3,42
II	41	$2,18 \cdot 10^{-5} \cdot (P^{2,079})$	-0,07	0,72	0,73	0,69	0,75	9,21

Observa-se que o desempenho das duas equações é considerado como “bom” para Camargo; Sentelhas (1997) e também foram aceitas como regiões homogêneas, pela medida H. Nota-se que para todas as divisões do Estado com o algoritmo Ward e também com o método k-means, as regiões possuem um número mais uniforme de bacias que contemplam cada região, diferindo do algoritmo de ligação completa que não proporcionou uma divisão homogênea.

Sendo que os limites de aplicação das equações são: para a região homogênea II o perímetro varia de 40,46 a 1533,71km; e para a região homogênea I as áreas de drenagem variam de 111,00 a 10153,78km². A Figura 42 ilustra a relação entre a Q_{95} observada e estimada para ambas as regiões homogêneas.

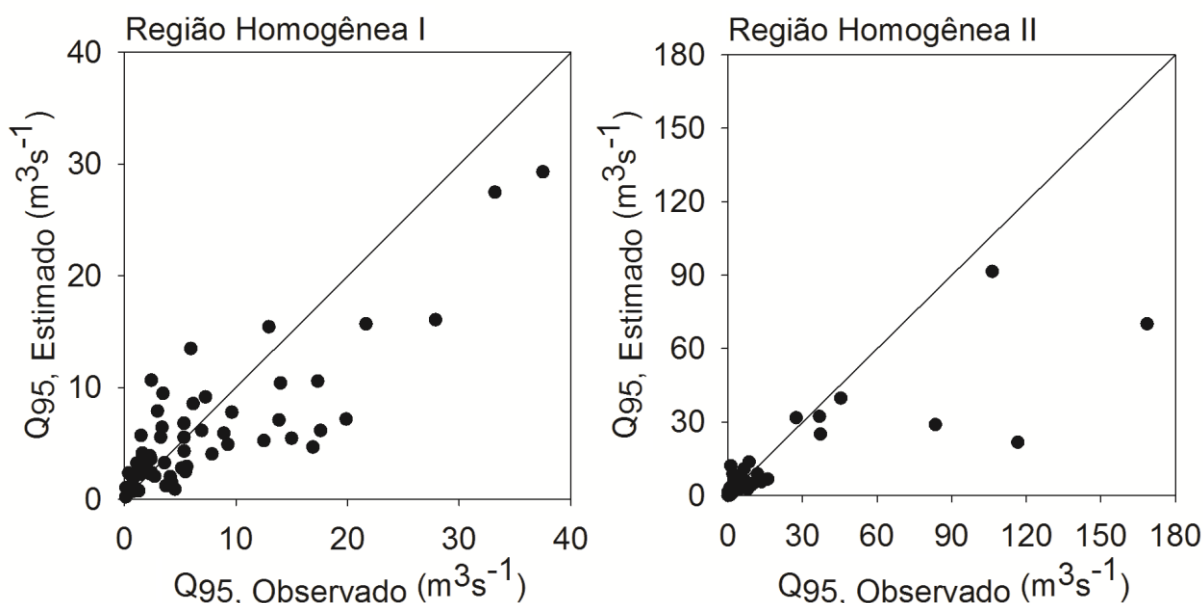


Figura 42 - Relação dos valores da Q_{95} observados e estimados (m^3s^{-1}) para as regiões homogêneas determinadas pelo método k-means com $k=2$.

Como mencionado para a separação de duas regiões no Estado, quando testado três regiões, os resultados também foram iguais quanto à disposição das bacias (Figuras 20 e 21). Observou-se que para a distância Euclidiana e Manhattan pelo algoritmo Ward, respectivamente a disposição das bacias nas regiões são idênticas para as vazões Q_{90} e Q_{95} , o que era esperado, uma vez que se utilizou as mesmas características, e as vazões Q_{90} e Q_{95} compreendem valores semelhantes.

Desta forma, a Figura 20 também representa a configuração das bacias para a vazão Q_{95} . Analisando-se o dendrograma observou-se que o algoritmo de ligação completa de agrupamento combinado com a distância de Manhattan, também produziu esta configuração, sendo assim as características das equações dessas regiões estão indicadas na Tabela 24.

Tabela 23 - Características estatísticas das regiões determinadas usando o algoritmo de Ward de agrupamento combinado com a distância Euclidiana e Manhattan

Região	n	Equação Q_{95}	H	R^2_a	R^2	c	log C_{NS}	EMA ($m^3 s^{-1}$)
I	51	$1,18 \cdot 10^{-3} \cdot (A^{1,094})$	-0,12	0,57	0,58	0,68	0,59	3,32
II	41	$1,41 \cdot 10^{-4} \cdot (P^{1,744})$	-0,16	0,52	0,53	0,38	0,54	2,86
III	8	$3,19 \cdot 10^{-36} \cdot (P^{1,337} \cdot p^{10,394})$	-0,18	0,72	0,80	0,80	0,80	19,12

Observa-se que, por meio dos índices estatísticos, que o desempenho da equação da região II foi considerado péssimo, conforme Camargo; Sentelhas (1997),

mas ainda aceitável para Motovilov *et al.* (1999), as outras duas regiões obtiveram coeficientes melhores.

Os limites de aplicação das equações são: para a região homogênea I a área varia de 337,09 a 10153,78km²; a região homogênea II o perímetro varia de 40,46 a 614,93km; e para a região homogênea III o perímetro varia de 767,68 a 1533,71km e a precipitação de 1455,16 a 1693,62mm. A Figura 43 mostra a relação entre a Q₉₅ observada e estimada para as regiões homogêneas.

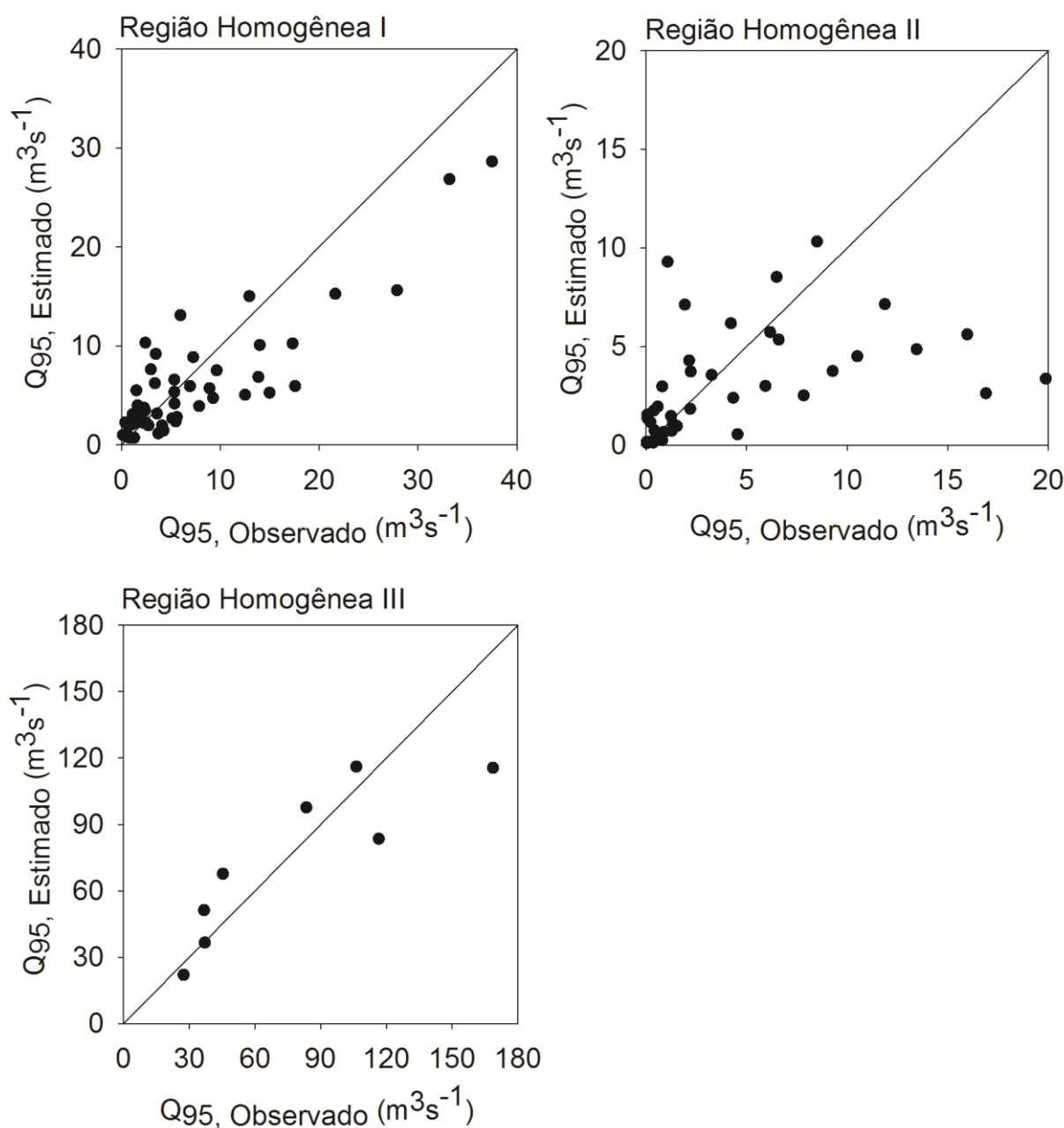


Figura 43 - Relação dos valores da Q₉₅ observados e estimados (m³s⁻¹) para as regiões homogêneas determinadas pela distância Euclidiana e Manhattan, com o algoritmo de Ward de agrupamento.

No método não hierárquico k-means para k=3 objetivando encontrar as regiões homogêneas para a Q_{95} observou-se que a disposição das bacias também foi idêntica quando aplicado para a Q_{90} . Desta forma, a Figura 21 também representa a configuração das bacias para a vazão Q_{95} , contudo os coeficientes estatísticos das equações dessas regiões estão indicados na Tabela 25.

Tabela 24 - Características estatísticas das regiões determinadas pelo método k-means com k=3

Região	n	Equação Q_{95}	H	R^2_a	R^2	c	log C_{NS}	EMA ($m^3 s^{-1}$)
I	54	$2,69 \cdot 10^{-3} \cdot (A^{0,994})$	-0,18	0,54	0,55	0,63	0,55	3,55
II	38	$1,16 \cdot 10^{-4} \cdot (P^{1,753})$	-0,38	0,53	0,55	0,41	0,56	2,35
III	8	$3,19 \cdot 10^{-36} \cdot (P^{1,337} \cdot p^{10,394})$	-0,18	0,72	0,80	0,80	0,80	19,12

Observa-se por meio dos índices estatísticos, que o desempenho da equação da região II por este método também não mostrou resultados satisfatórios conforme Camargo; Sentelhas (1997), mas ainda aceitável para Motovilov *et al.* (1999), as outras duas regiões obtiveram coeficientes melhores, contudo os valores de R^2_a e R^2 das regiões I e II também estão abaixo do recomendado (ELESBON *et al.*, 2015). Nota-se ainda que apenas na região III o melhor ajuste ocorreu com duas variáveis explicativas

Desta forma, demonstra-se que para o estado do Rio Grande do Sul a configuração de três regiões homogêneas não caracteriza os melhores resultados para a aplicação no planejamento e gestão dos recursos hídricos em relação a Q_{95} , tal resultado já estava presente nas equações da Q_{90} , contudo os valores encontrados nesta etapa são ainda menores.

Os limites de aplicação para as equações são os mesmos valores limitantes do método da distância Euclidiana e Manhattan, com o algorítmico de Ward de agrupamento (Figura 20), diferenciando apenas nos números das bacias que compreendem cada uma das regiões homogêneas. A Figura 44 ilustra a relação entre a Q_{95} observada e estimada para as regiões homogêneas.

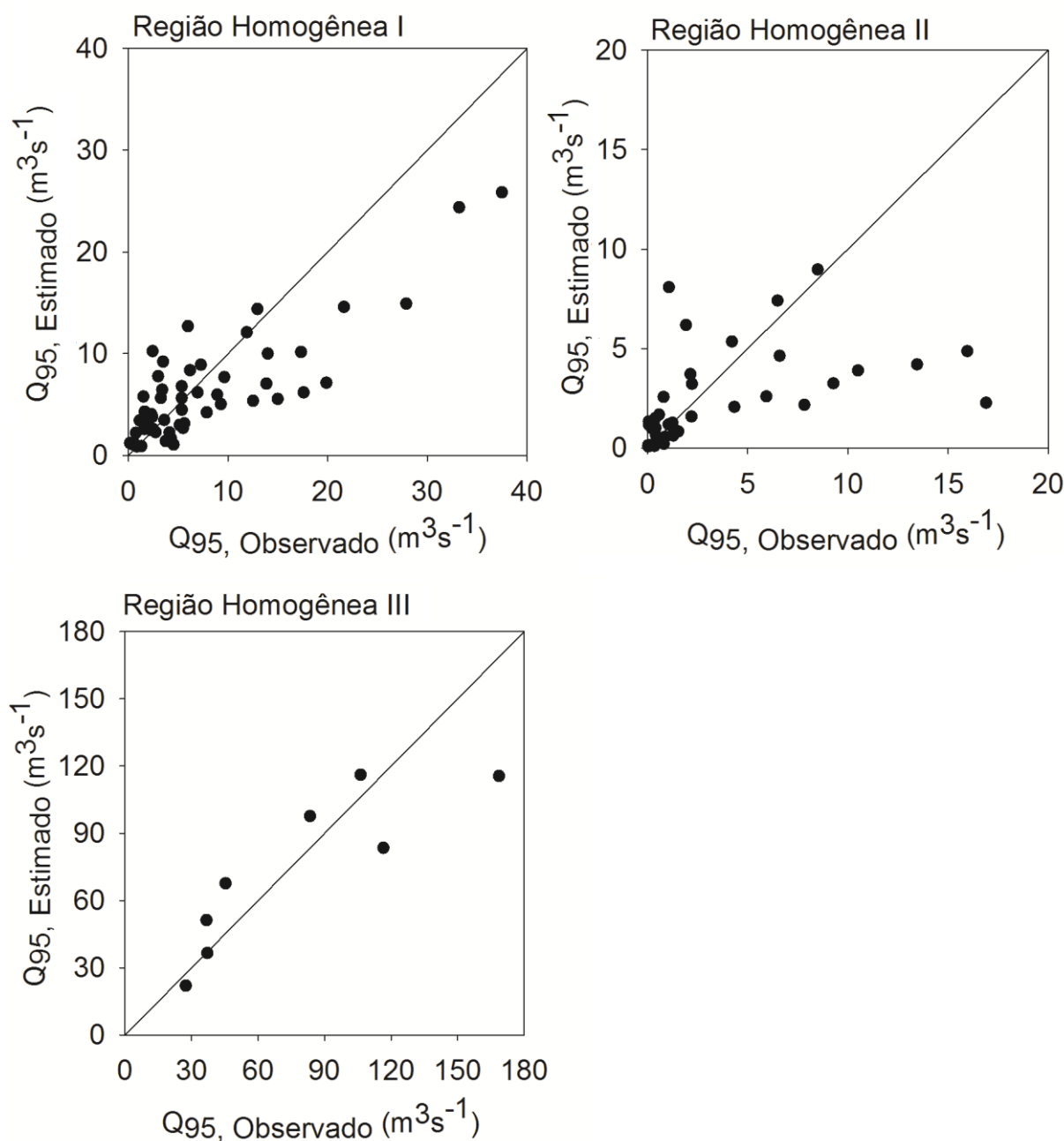


Figura 44 - Relação dos valores da Q_{95} observados e estimados (m^3s^{-1}) para as regiões homogêneas determinadas pelo método k-means com $k=3$.

Como mencionado para a separação de duas e três regiões no Estado, quando testado quatro regiões os resultados também foram iguais quanto à disposição das bacias (Figuras 24 e 25). Observou-se que para a distância Euclidiana e Manhattan pelo algoritmo Ward, a disposição das bacias nas respectivas regiões são idênticas para as vazões Q_{90} e Q_{95} . Desta forma, a Figura 24 também representa a configuração das bacias para a vazão Q_{95} , contudo as características das equações dessas regiões estão indicadas na Tabela 26 e a relação entre a Q_{95} observada e estimada é apresentada na Figura 45.

Tabela 25 - Características estatísticas das regiões determinadas usando o algoritmo de Ward de agrupamento combinado com a distância Euclidiana e Manhattan

Região	n	Equação Q_{95}	H	R^2_a	R^2	c	$\log C_{NS}$	EMA (m^3s^{-1})
I	41	$1,41 \cdot 10^{-4} \cdot (P^{1,744})$	-0,16	0,52	0,53	0,40	0,54	2,86
II	30	$2,80 \cdot 10^{-4} \cdot (A^{1,304})$	-0,09	0,56	0,57	0,71	0,57	2,52
III	21	$3,09 \cdot 10^{-3} \cdot (A^{0,951})$	-0,32	0,64	0,65	0,67	0,65	3,78
IV	8	$3,19 \cdot 10^{-36} \cdot (P^{1,337} \cdot p^{10,394})$	-0,18	0,72	0,80	0,80	0,80	19,12

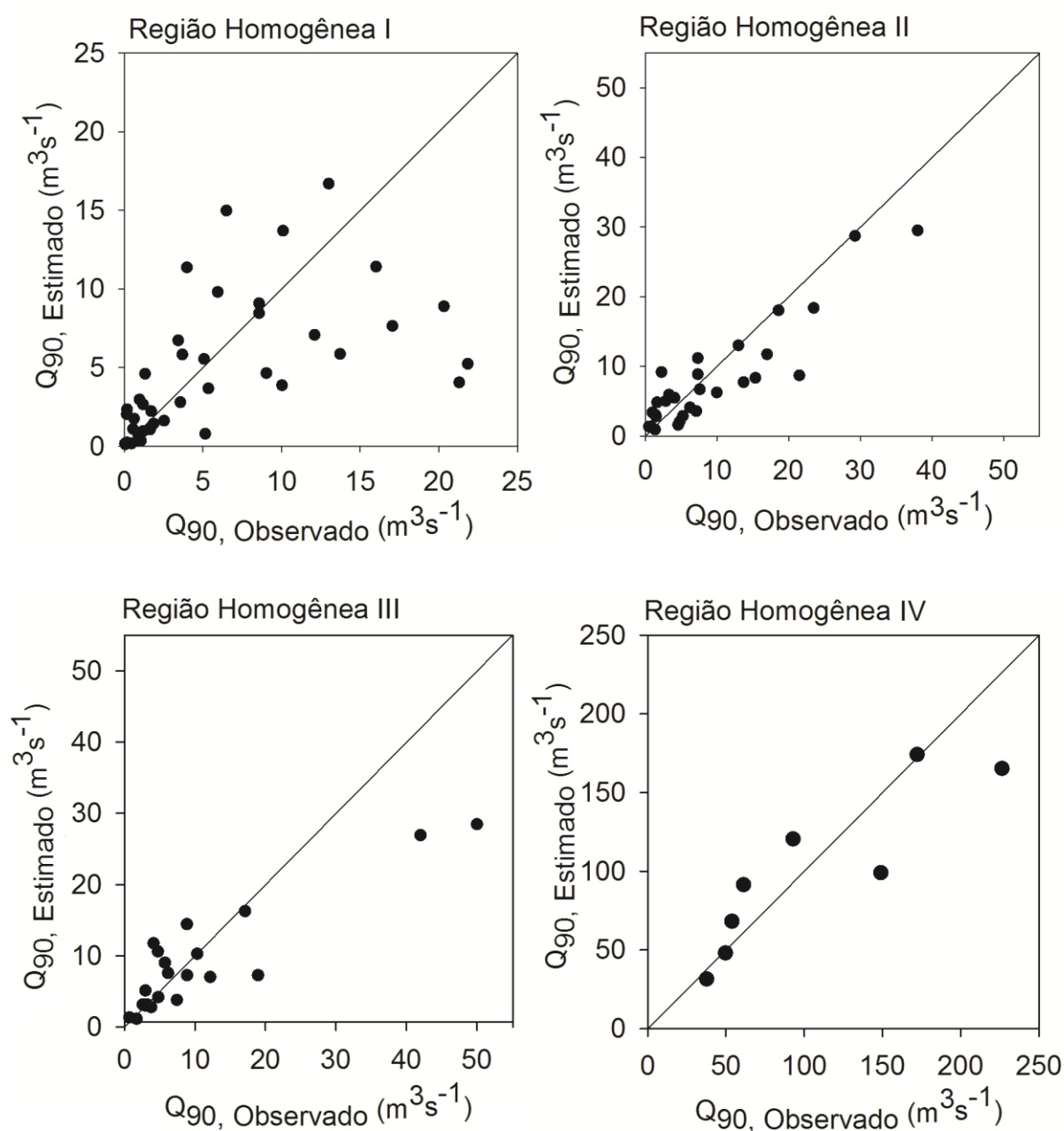


Figura 45 - Relação dos valores da Q_{95} observados e estimados (m^3s^{-1}) para as regiões homogêneas determinadas pela distância Euclidiana e Manhattan, com o algoritmo de Ward de agrupamento.

Observa-se que, por meio dos índices estatísticos, que o desempenho da equação da região I foi considerado péssimo conforme Camargo; Sentelhas (1997), mas ainda aceitável para Motovilov *et al.* (1999), as outras três regiões obtiveram coeficientes melhores para este índice. Ainda que os outros parâmetros estatísticos não foram os mais adequados, todas as regiões foram consideradas homogêneas pela medida H.

Os limites de aplicação para as equações são: para a região homogênea I os perímetros variam de 40,46 a 614,93km; região homogênea II a área varia de 337,09 a 5833,55km²; região homogênea III possui área de 349,72 a 10153,78km²; e a região homogênea IV o perímetro varia de 767,68 a 1533,71km e a precipitação de 1455,16 a 1693,62mm.

No método não hierárquico k-means para k=4 objetivando encontrar as regiões homogêneas para a Q₉₅ observou-se que a disposição das bacias também foi idêntica quando aplicado para a Q₉₀. Desta forma, a Figura 25 também representa a configuração das bacias para a vazão Q₉₅, contudo as características das equações dessas regiões estão indicadas na Tabela 27.

Tabela 26 - Características estatísticas das regiões determinadas pelo método k-means com k=4

Região	n	Equação Q ₉₅	H	R ² _a	R ²	c	log C _{NS}	EMA (m ³ s ⁻¹)
I	37	$1,04 \cdot 10^{-4} \cdot (P^{1,763})$	-0,09	0,57	0,58	0,42	0,58	2,53
II	32	$3,62 \cdot 10^{-4} \cdot (A^{1,264})$	-0,07	0,55	0,56	0,68	0,56	2,74
III	23	$1,04 \cdot 10^{-3} \cdot (P^{1,522})$	-0,20	0,52	0,55	0,56	0,55	4,32
IV	8	$3,19 \cdot 10^{-36} \cdot (P^{1,337} \cdot p^{10,394})$	-0,18	0,72	0,80	0,80	0,80	19,12

Observa-se que, por meio dos índices estatísticos, que o desempenho das equações das regiões I e III segundo a classificação de Camargo; Sentelhas (1997) foram “mau” e “sofrível”, respectivamente, as outras duas regiões obtiveram coeficientes mais adequados. Desta forma, demonstra-se que para o estado do Rio Grande do Sul a configuração de quatro regiões homogêneas não caracteriza os melhores resultados para a aplicação no planejamento e gestão dos recursos hídricos.

Os limites de aplicação das equações são: para a região IV o perímetro varia de 767,68 a 1533,71km e a precipitação de 1455,16 a 1693,62mm; a região homogênea III o perímetro varia de 96,40 a 633,71km; para a região homogênea II a área varia de 337,09 a 5833,55km²; e para a região homogênea I o perímetro varia de

40,46 a 614,93km. A Figura 46 ilustra a relação entre a Q_{95} observada e estimada para as regiões homogêneas.

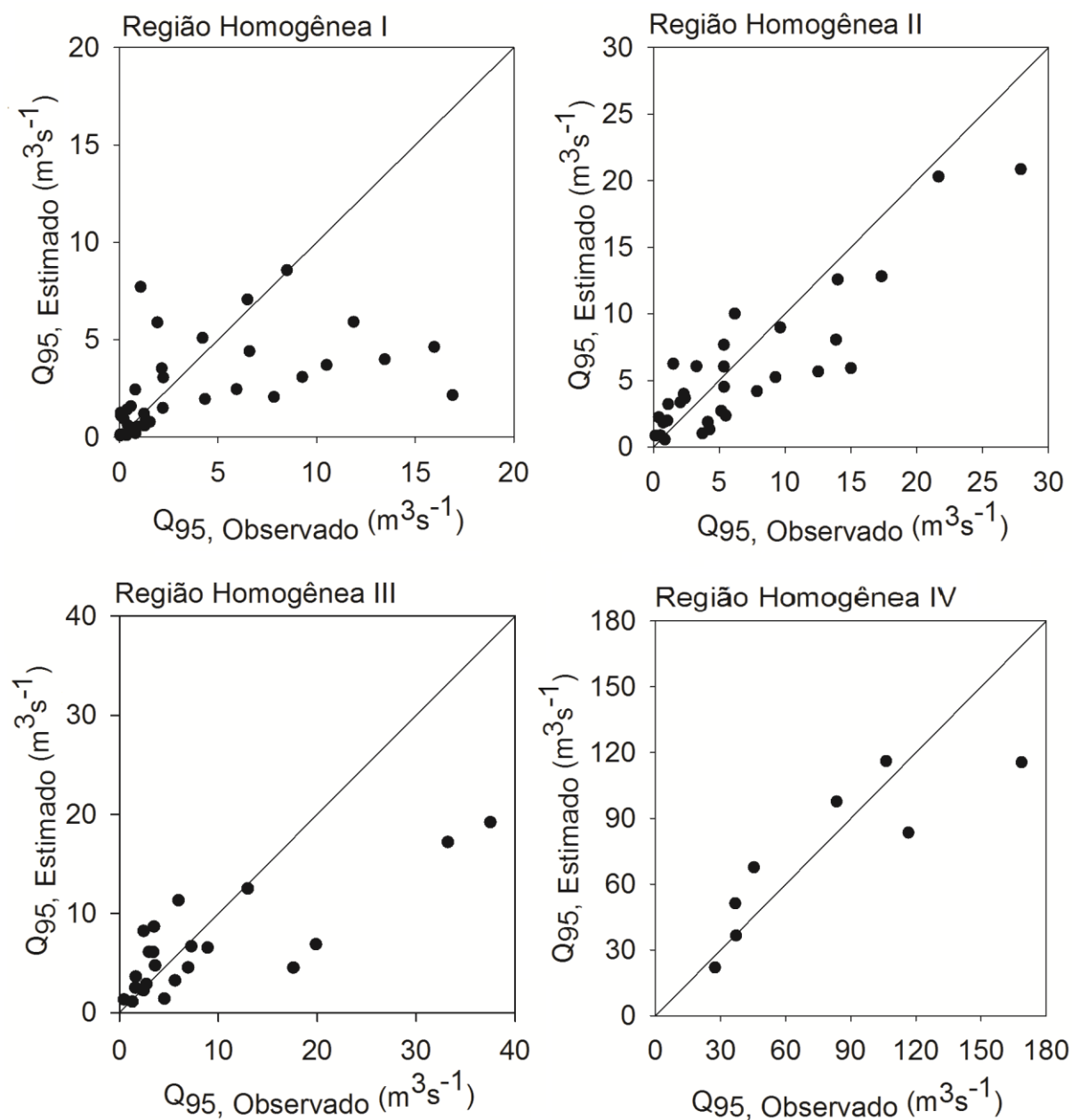


Figura 46 - Relação dos valores da Q_{95} observados e estimados (m^3s^{-1}) para as regiões homogêneas determinadas pelo método k-means com $k=4$.

Como mencionado para as separações anteriores das regiões no Estado, quando testado cinco regiões os resultados também foram iguais quanto à disposição das bacias para a vazão Q_{95} (Figura 28).

Na Figura 28 foi apresentada a conformação para a distância Euclidiana pelo algoritmo Ward da Q_{90} , sendo igual a divisão proposta para a distância Euclidiana e

Manhattan pelo algoritmo Ward para a Q_{95} , cujas características das equações para essas regiões indicadas na Tabela 28.

Tabela 27 - Características estatísticas das regiões determinadas usando o algoritmo Ward de agrupamento combinado com a distância Euclidiana e Manhattan

Região	n	Equação Q_{95}	H	R^2_a	R^2	c	$\log C_{NS}$	EMA (m^3s^{-1})
I	30	$2,80 \cdot 10^{-4} \cdot (A^{1,304})$	-0,09	0,56	0,57	0,71	0,57	2,52
II	25	$8,03 \cdot 10^{-3} \cdot (A^{0,814})$	-0,21	0,39	0,42	0,35	0,42	3,58
III	21	$3,09 \cdot 10^{-3} \cdot (A^{0,951})$	-0,32	0,64	0,65	0,67	0,65	3,78
IV	16	$8,30 \cdot 10^{-4} \cdot (A^{0,959})$	36,49	0,70	0,72	0,54	0,72	1,24
V	8	$3,19 \cdot 10^{-36} \cdot (P^{1,337} \cdot p^{10,394})$	-0,18	0,72	0,80	0,80	0,80	19,12

Observa-se que, por meio dos índices estatísticos, que o desempenho das equações das regiões II e IV segundo a classificação de Camargo; Sentelhas (1997) foram “péssimo” e “sofrível”, respectivamente, as outras três regiões obtiveram coeficientes mais favoráveis.

Entretanto, nas divisões anteriores não havia a presença de uma região heterogênea, visto que a região IV apresentou um H superior a 1, tentou-se primeiramente dividir a região II e unir parte com a região IV, similar ao que foi testado para a Q_{90} com o método k-means com $k=5$, mas a heterogeneidade permaneceu.

Apesar disso, os limites de aplicação das equações da Tabela 28 são: para a região homogênea I a área limita-se de 337,09 a 5833,55km²; regiões homogêneas II e III a área varia de 70,68 a 8292,29km² e 349,72 a 10153,78km², respectivamente; região homogênea IV, a área varia de 68,61 a 7891,90km²; e a região homogênea V o perímetro varia de 767,68 a 1533,71km e a precipitação de 1455,16 a 1693,62mm. A Figura 47 mostra a relação entre a Q_{95} observada e estimada para as regiões.

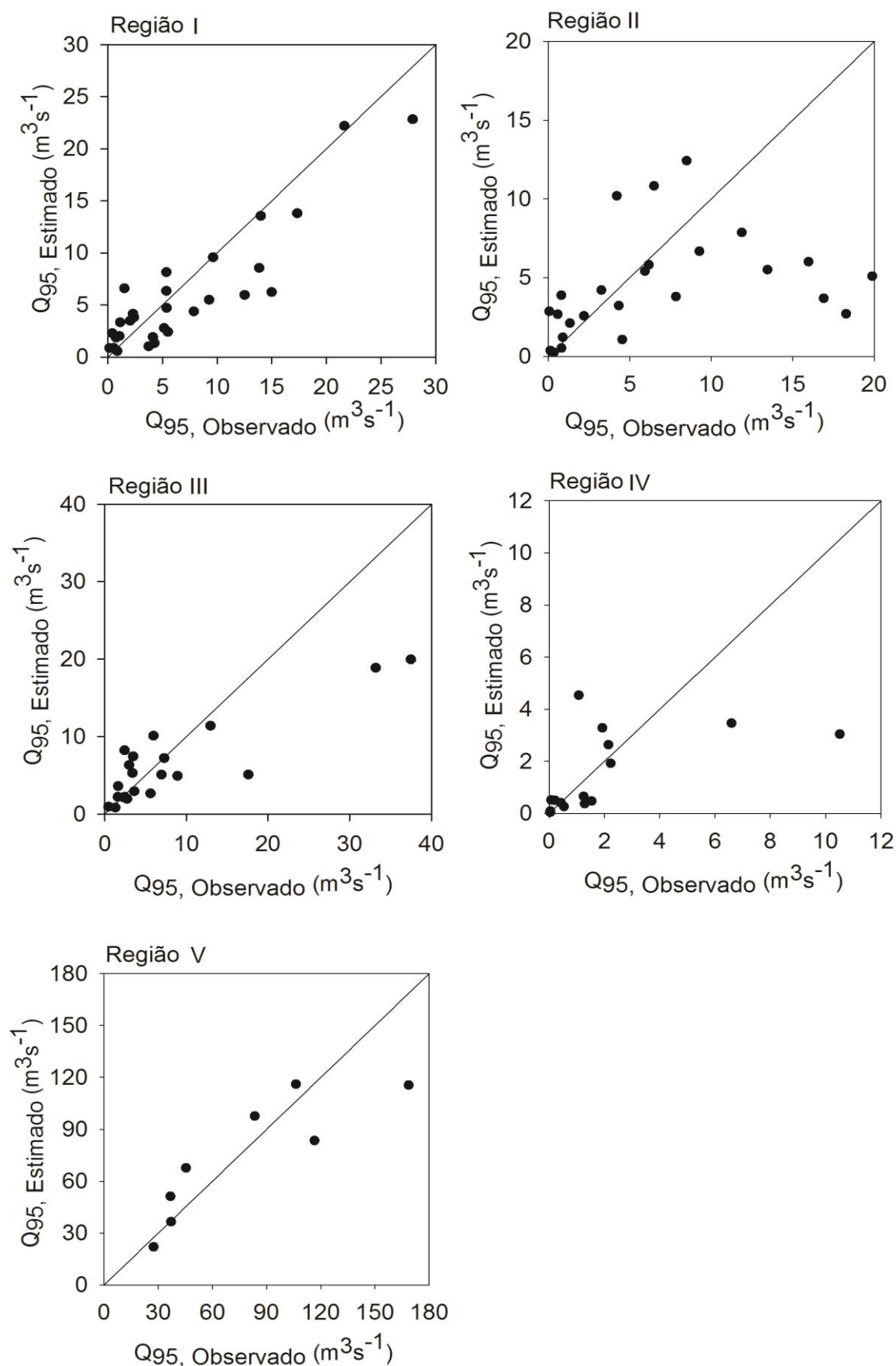


Figura 47 - Relação dos valores da Q₉₅ observados e estimados (m^3s^{-1}) para as regiões determinadas pela distância Euclidiana e Manhattan, com o algoritmo de Ward de agrupamento.

Aplicando-se o método hierárquico k-means com $k=5$ a configuração das bacias no estado do Rio Grande do Sul foram dispostas na Figura 48.

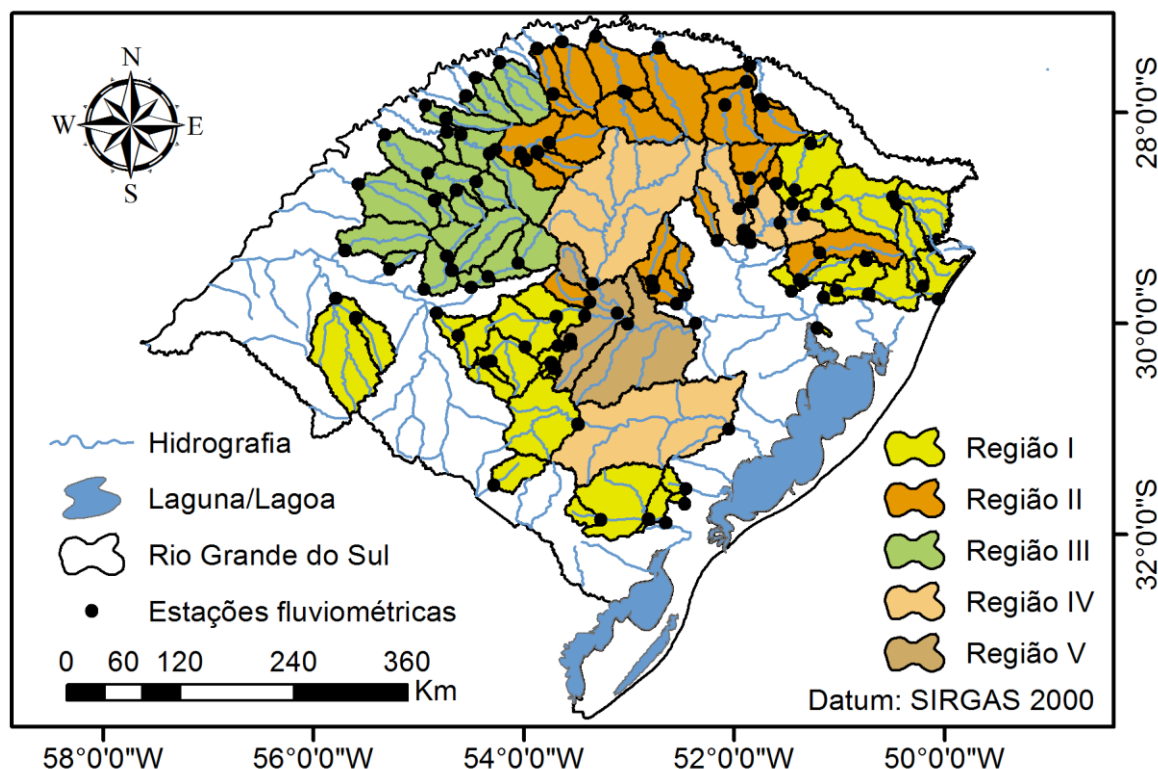


Figura 48 - Regiões determinadas para o estado do Rio Grande do Sul pelo método k-means com $k=5$.

Observa-se que a região V é representada apenas por duas estações. Silva et al. (2009) enfatizam que um número mínimo de estações em cada região precisa ser respeitado, visto que a utilização de um número maior de variáveis explicativas no modelo de regressão múltipla testado torna-se desfavorável. Deste modo, procurou-se não individualizar regiões com menos do que cinco estações fluviométricas.

Para isso uniu-se a região V com a região IV com essa nova formação em que, o método k-means com $k=5$ fica análogo ao k-means com $k=4$ (Figura 25, Tabela 27 e Figura 46).

Igualmente como na obtenção das regiões homogêneas para a vazão Q_{90} , também foi testada a divisão do Estado para $k=6$ pelo método não hierárquico k-means, obtendo-se a disposição das bacias conforme a Figura 49 e na Tabela 29 são apresentadas as características dessas regiões. Já na Figura 50 ilustra-se a relação entre a Q_{95} observada e estimada para as regiões. Diferentemente do observado para as equações da vazão Q_{90} quando aplicado o método k-means com $k=6$, indicam a continuidade da região VI com oito bacias para a Q_{95} .

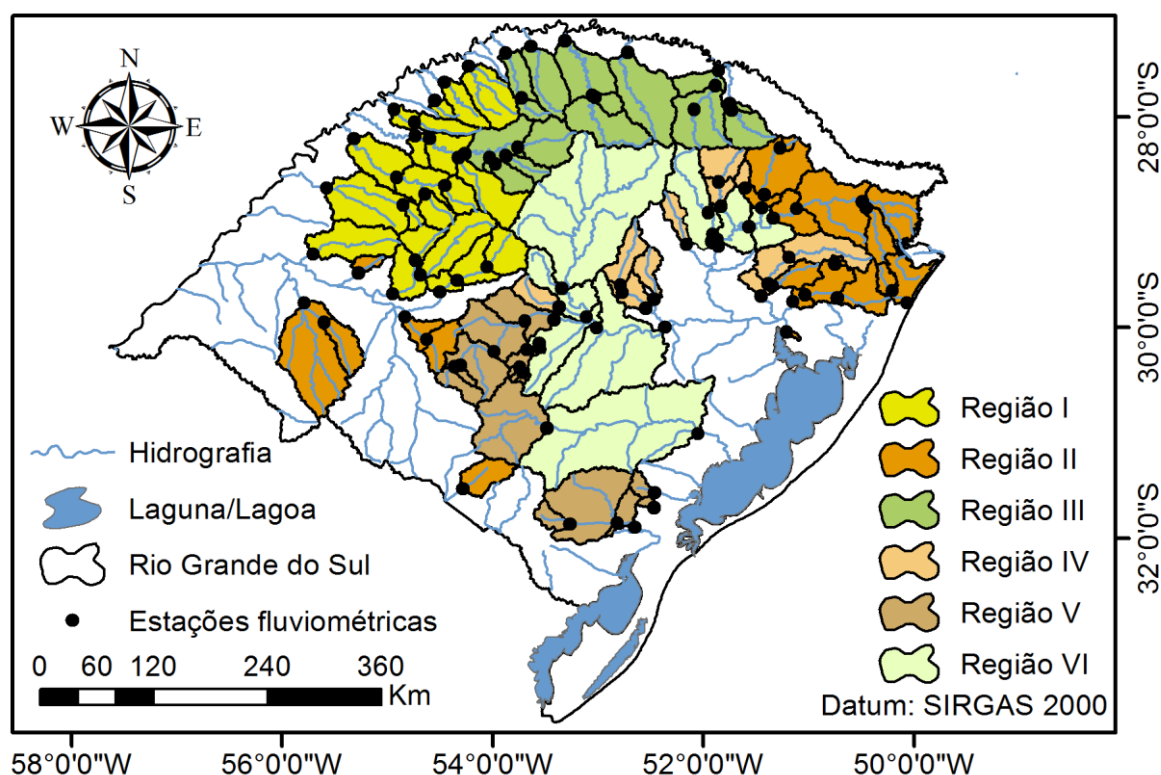


Figura 49 - Regiões determinadas para o estado do Rio Grande do Sul pelo método k-means com k=6.

Tabela 28 - Características estatísticas das regiões determinadas pelo método k-means com k=6

Região	n	Equação Q_{95}	H	R^2_a	R^2	c	$\log C_{NS}$	EMA (m^3s^{-1})
I	22	$2,79 \cdot 10^{-3} \cdot (A^{0,971})$	-0,23	0,61	0,63	0,65	0,63	4,12
II	21	$8,09 \cdot 10^{-4} \cdot (P^{1,46})$	-0,21	0,39	0,42	0,36	0,42	3,05
III	17	$2,17 \cdot 10^{-5} \cdot (P^{2,315})$	-0,12	0,69	0,71	0,79	0,71	2,66
IV	16	$3,81 \cdot 10^{-4} \cdot (A^{1,194})$	-0,07	0,69	0,70	0,73	0,62	1,29
V	16	$8,30 \cdot 10^{-4} \cdot (A^{0,959})$	36,49	0,70	0,72	0,54	0,72	1,24
VI	8	$3,19 \cdot 10^{-36} \cdot (P^{1,337} \cdot P^{10,394})$	-0,18	0,72	0,80	0,80	0,80	19,12

Na Tabela 29 observa-se a permanência da heterogeneidade da região V, indicando que a divisão de 6 regiões para o Estado também não satisfaz todos os critérios necessários para que as equações possam ser empregadas em projetos de gestão e planejamentos dos recursos hídricos, visando as vazões de permanência.

Nessa disposição, as regiões II e V permanecem com valores do coeficiente de desempenho “c” não adequados, conforme Camargo; Sentelhas (1997); na região II, o valor do R^2 obtido foi de 0,42, indicando que o modelo explica apenas de 42% da variabilidade dos dados. Esse resultado indica mais uma vez que a configuração com k=5 representa a melhor divisão para o Estado, depois da identificação das estações que causam a heterogeneidade.

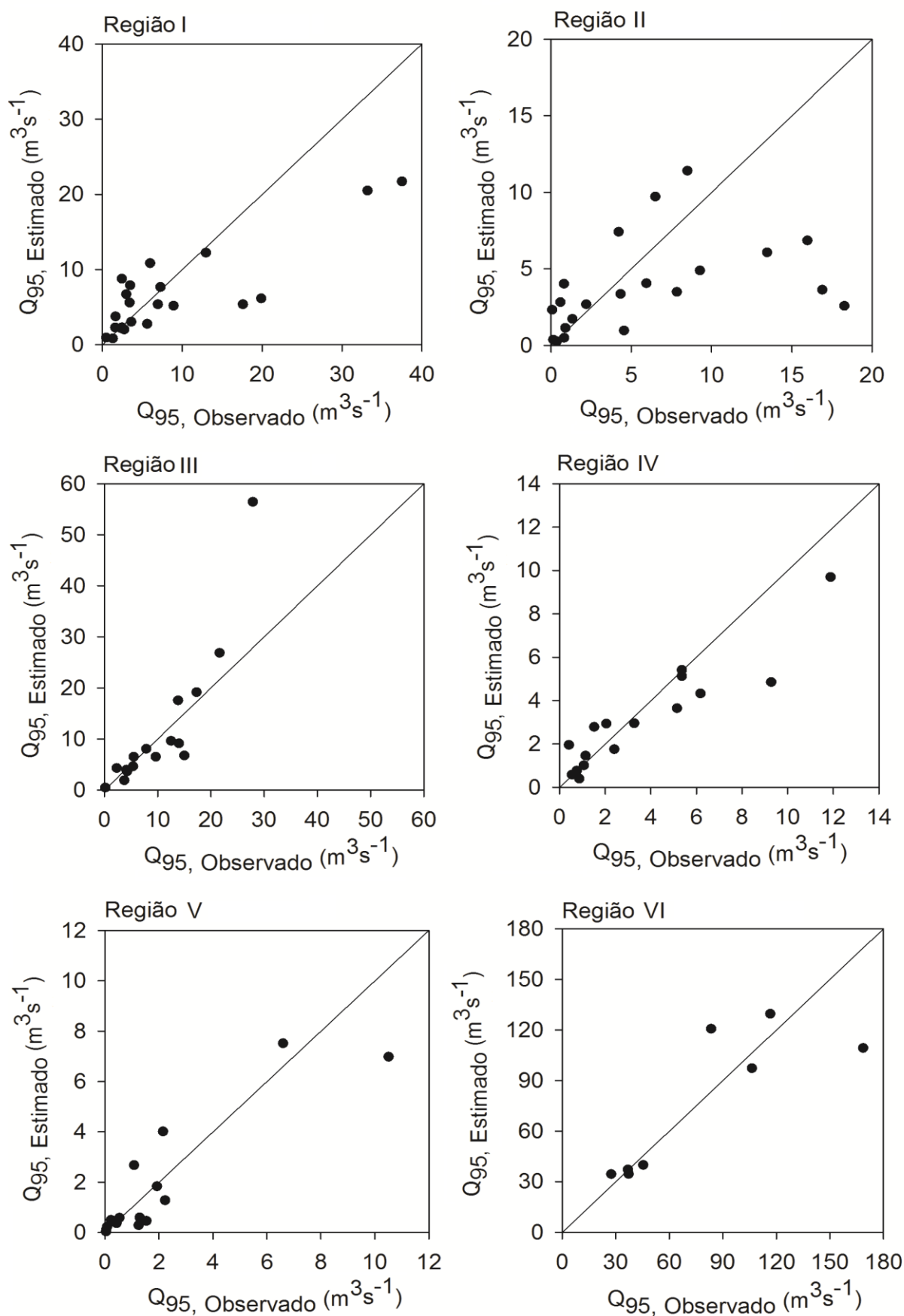


Figura 50 - Relação dos valores da Q_{95} observados e estimados (m^3s^{-1}) para as regiões determinadas pelo método k-means com $k=6$.

Os limites de aplicação para as equações são: para a região homogênea I a área limita-se de 349,72 a 10153,78km²; região homogênea II o perímetro limita-se de 46,82 a 614,93km; região homogênea III o perímetro limita-se de 96,40 a 633,71km; região homogênea IV a área limita-se de 337,09 a 4731,26km²; região V a área varia de 68,61 a 7891,90km²; e a região homogênea IV o perímetro varia de 767,68 a 1533,71km e a precipitação de 1455,16 a 1693,62mm.

Embora todo o esforço gerado na delimitação de regiões homogêneas, salienta-se que a informação de tais regiões no estado do Rio Grande do Sul para as vazões de referência precisa estar acompanhada do levantamento das outorgas já concedidas para que, dessa forma, a disponibilidade hídrica das regiões seja respeitada. Oliveira *et al.* (2013) destacam que com conhecimento dos pontos de outorgas é possível observar quais seções da bacia estão sendo mais requeridas, visto que regiões de cabeceira estão associadas a baixa disponibilidade de água e merecem maior cuidado para a preservação desse recurso. Para os autores, a totalização das outorgas nas confluências do corpo hídrico é um reflexo da situação da região, dando uma indicação da utilização da água na bacia e permitindo a identificação da existência ou não de sazonalidade das demandas ao longo do ano, bem como da sua amplitude de variação.

Para Moreira *et al.* (2012), a utilização dos recursos hídricos sem o seu respectivo documento legal inviabiliza o cálculo da real demanda de água, fato que somente pode ser contornado com a realização de um amplo cadastro de usuários de água da bacia.

O DRH do Estado começou, por meio da DIOUT, a garantir mais segurança e transparência de seus processos de outorgas aos empreendedores. Ao mesmo tempo, pretende controlar a adoção de padrões que tenham o menor impacto possível sobre a natureza. A reunião dessas informações não é uma tarefa fácil e rápida, porém é de extrema importância para uma melhor gestão e planejamento dos recursos hídricos. Moreira *et al.* (2012) observaram em seu estudo de vazões mínimas no rio do Sono em Minas Gerais que a maioria dos processos de outorgas são relacionados com a irrigação, situação similar ao Rio Grande do Sul, visto que é um Estado onde a agricultura representa umas das principais atividades econômicas.

Neste sentido, o decreto estadual nº 52.931, de 07 de março de 2016, busca implementar alterações na operacionalização dos órgãos ambientais nas autorizações

e licenças necessárias aos empreendimentos produtivos que fazem uso de sistemas de irrigação (RIO GRANDE DO SUL, 2016).

O decreto ainda dispõe dos procedimentos para a regularização ambiental relativo à outorga e alvarás dos reservatórios de água construídos no passado sem a observação dos devidos procedimentos atualmente necessários. Já a resolução do Consema nº323 de 2016, dispõe exclusivamente sobre os procedimentos de licenciamento ambiental dos empreendimentos de irrigação (RIO GRANDE DO SUL, 2016). Desta forma, o poder público estadual procura por meio desses arsenais jurídicos, sustentar e trazer maior objetividade no que se refere aos documentos, procedimentos e estudos necessários para obtenção das autorizações a serem requeridas.

Para identificar a melhor divisão no Estado em relação a Q_{95} retornou-se ao método que utiliza o algoritmo Ward de agrupamento combinado com a distância Euclidiana e Manhattan, identificando as estações (85460000, 85470000, 88750000, 88850000 e 76742000), as quais eram responsáveis pela heterogeneidade na região IV, sendo assim removidas e recalculadas as novas equações para as regiões, consequentemente as estações eram as mesmas da região IV para a Q_{90} . Paralelamente, realizou-se o realocamento das estações da região II para a região IV.

Tal configuração representou as estações de forma análoga à encontrada anteriormente para Q_{90} (Figura 36), ou seja, tanto para a vazão Q_{90} ou Q_{95} a disposição das bacias quanto às regiões homogêneas foram iguais. Na Tabela 30 são apresentadas as características das 5 regiões homogêneas.

Tabela 29 - Características estatísticas das regiões homogêneas determinadas usando o algoritmo Ward de agrupamento combinado com a distância Euclidiana e Manhattan

Região	n	Equação Q_{95}	H	R^2_a	R^2	c	log C_{NS}	EMA (m^3s^{-1})
I	24	$2,10 \cdot 10^{-5} \cdot (P^{2,298})$	-0,11	0,70	0,71	0,78	0,70	2,59
II	23	$2,79 \cdot 10^{-3} \cdot (A^{0,971})$	-0,20	0,61	0,63	0,65	0,63	4,12
III	23	$3,99 \cdot 10^{-3} \cdot (A^{0,906})$	-0,09	0,72	0,73	0,61	0,73	2,73
IV	17	$2,28 \cdot 10^{-3} \cdot (A^{0,824})$	0,64	0,62	0,67	0,72	0,67	1,39
V	8	$3,19 \cdot 10^{-36} \cdot (P^{1,337} \cdot p^{10,394})$	-0,18	0,72	0,80	0,80	0,80	19,12

Sendo que todas apresentam coeficiente de desempenho “c” de mediano a muito bom, conforme Camargo; Sentelhas (1997). O log C_{NS} segundo a classificação de Motovilov *et al.* (1999) encontra-se como “aceitável” e “adequado e bom”. Apenas

na equação da região II o melhor ajuste foi com a variável área de drenagem diferindo das demais que permaneceram com as mesmas variáveis explicativas encontradas nas equações das regiões homogêneas da Q_{90} .

Baena *et al.* (2004) e Silva; Marques; Lemos (2009) já haviam observado que para as vazões mínimas associadas às curvas de permanência (Q_{90} e Q_{95}), a área de drenagem foi o parâmetro físico que melhor se ajustou aos modelos de regressão corroborando com o presente estudo. Elesbon *et al.* (2015) recomendam que o R^2_a seja superior a 0,7, fato que não foi atingido por todas as equações; contudo, Galli; Fraga; Santos. (2016) e Ganora *et al.* (2009) evidenciam que valores menores ainda podem ser utilizados sem grandes prejuízos aos modelos.

Os limites de aplicação das equações da Tabela 30 são: para a região homogênea I o perímetro limita-se de 97,78 a 468,82km; região homogênea II a área varia de 349,72 a 10153,78km²; região homogênea III a área varia de 70,68 a 8292,29km²; região homogênea IV, a área varia de 68,61 a 7891,90km; e a região homogênea V o perímetro varia de 767,68 a 1533,71km e a precipitação de 1455,16 a 1693,62mm. A Figura 51 ilustra a relação entre a Q_{95} observada e estimada para as regiões homogêneas.

Na Tabela 31 estão apresentados os agrupamento das estações fluviométricas em regiões hidrologicamente homogêneas, identificadas com base no algoritmo de Ward de agrupamento combinado com a distância Euclidiana e Manhattan, respectivamente. A tabela refere-se tanto à vazão Q_{90} como a Q_{95} , visto que o agrupamento mais favorável foi o mesmo para as duas vazões.

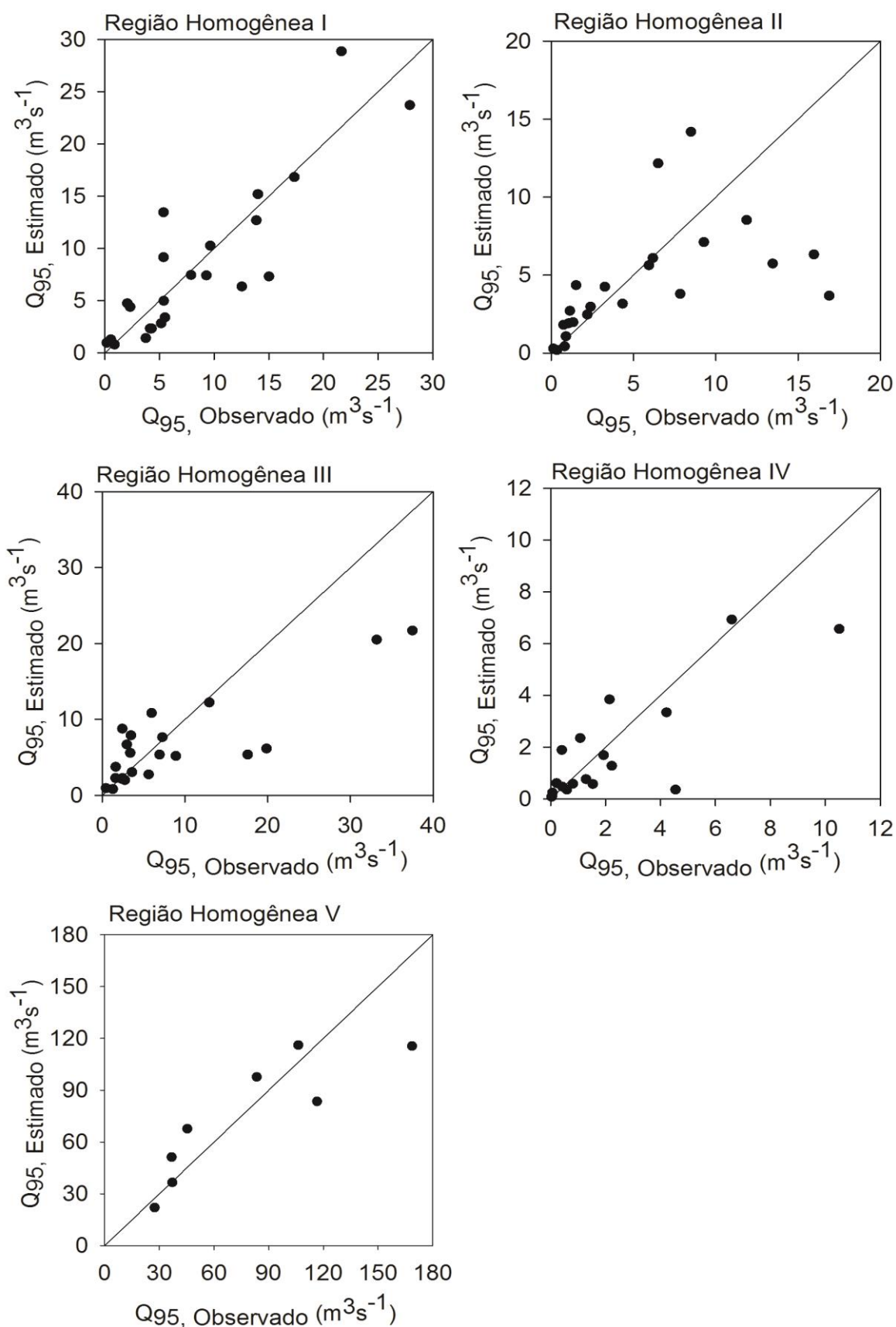


Figura 51 - Relação dos valores da Q_{95} observados e estimados (m^3s^{-1}) para as regiões homogêneas determinadas pela distância Euclidiana e Manhattan, com o algoritmo de Ward de agrupamento modificado.

Tabela 30 - Agrupamento das estações fluviométricas em regiões hidrologicamente homogêneas para as vazões Q_{90} e Q_{95}

Código	n	Região	n	Código	Região	n	Código	Região
72400000	1	I	33	75700000	II	65	86250000	III
72430000	2	I	34	76085000	II	66	86300000	III
72580000	3	I	35	76100000	II	67	86340000	III
72630000	4	I	36	76120000	II	68	86410000	III
72680000	5	I	37	76360001	IV	69	86420000	I
73480000	6	I	38	76380000	IV	70	86440000	III
74205000	7	I	39	76440000	II	71	86470000	V
74210000	8	I	40	76460000	II	72	86480000	I
74270000	9	I	41	76490000	II	73	86500000	I
74370000	10	I	42	76550000	II	74	86510000	V
74460000	11	I	43	76650000	II	75	86560000	I
74470000	12	I	44	76750000	IV	76	86580000	I
74600000	13	II	45	79400000	IV	77	86700000	I
74610000	14	II	46	85400000	V	78	86720000	V
74700000	15	II	47	85438000	IV	79	86745000	III
74750000	16	II	48	85440000	V	80	87160000	III
74880000	17	II	49	85480000	IV	81	87170000	III
74900000	18	II	50	85580000	IV	82	87250000	III
75155000	19	I	51	85600000	IV	83	87270000	III
75185000	20	I	52	85610000	IV	84	87317030	III
75186000	21	I	53	85615000	IV	85	87317060	III
75200000	22	I	54	85620000	IV	86	87366000	III
75205000	23	I	55	85623000	IV	87	87374000	III
75230000	24	I	56	85630000	IV	88	87380000	III
75295000	25	II	57	85642000	V	89	87382000	III
75300000	26	II	58	85730000	III	90	87450100	III
75320000	27	II	59	85740000	III	91	87590000	IV
75400000	28	II	60	85780000	III	92	87905000	V
75430000	29	II	61	85850000	III	93	88575000	IV
75450000	30	II	62	85900000	V	94	88641000	IV
75500000	31	II	63	86100000	III	95	88680000	IV
75600000	32	II	64	86160000	III			

5 CONCLUSÃO

Os testes aplicados para identificar a tendência nas séries das estações pluviométricas e fluviométricas mostraram-se apropriados e adequados para a formação das respectivas séries históricas, visando o levantamento das variáveis e identificação das bacias a serem empregadas para a regionalização. O modelo digital de elevação obtido pelo banco de dados TOPODATA viabilizou a determinação automática das características físicas das bacias hidrográficas.

A análise de agrupamento possibilitou a determinação de regiões homogêneas no Estado, as diferentes configurações mostram-se apropriadas e no geral, os métodos hierárquicos e não hierárquicos não diferiam muito no número de bacias que contemplam cada região, conforme as análises realizadas de 2 a 6 grupos. No método hierárquico destaca-se principalmente, que as distâncias Euclidiana e Manhattan não diferiram muito nos grupos formados, a distância de Mahalanobis foi a que possibilitou o número menor de alternativas de grupos, e em relação aos algoritmos de agrupamento, o Ward foi superior aos demais. Também ressalta-se a importância da utilização do centroide da bacia para que as regiões demarcadas contemplassem as estações com certa continuidade geográfica, o que pode ser interessante quando for avaliado projetos de outorga.

Por meio da medida de H foi possível identificar as estações que causavam a heterogeneidade na região, confirmando que tal teste é de grande importância para as análises hidrológicas. Desta forma, confirmou-se que a melhor divisão para o Estado a nível de gestão das vazões de permanência é de 5 regiões homogêneas.

Identificou-se também pela regressão método Stepwise, modo backward que as variáveis área de drenagem e perímetro foram as que mais estiveram presentes nas equações de vazões mínimas, sendo que a variável precipitação total anual média inseriu-se apenas em uma região tanto para a Q_{90} como para a Q_{95} .

As equações ajustadas apresentaram desempenho aceitável para estimar Q_{90} e Q_{95} , variando a variável explicativa de ajuste entre área de drenagem, perímetro e precipitação total anual média. Na maioria dos casos foi utilizado apenas uma variável, o que torna bastante fácil a obtenção e a aplicação, possibilitando o uso no estado do Rio Grande do Sul, para estimar as vazões de referência.

REFERÊNCIAS BIBLIOGRÁFICAS

- ABDI, A.; HASSANZADEH, Y.; OUARDA, T. B. M.J. Regional frequency analysis using Growing Neural Gas network. **Journal of Hydrology**, v. 550, p. 92-102, 2017. doi: 10.1016/j.jhydrol.2017.04.047
- ABREU, M. C.; TONELLO, K. C. Avaliação dos Parâmetros Hidrometeorológicos na Bacia do Rio Sorocaba/SP. **Revista Brasileira de Meteorologia**, v. 32, n. 1, p. 99-109, 2017. doi:10.1590/0102-778632120150164
- AGÊNCIA NACIONAL DE ÁGUAS (ANA). **GEO Brasil recursos hídricos**: componente da série de relatórios sobre o estado e perspectiva do meio ambiente no Brasil. Brasília: ANA, 2007. 264 p.
- AGÊNCIA NACIONAL DE ÁGUAS (ANA). **Manual de procedimentos técnicos e administrativos de outorga de direito de uso de recursos hídricos** 2013. Brasília: ANA, 2013. 252p.
- AISSIA, M-A. B.; CHEBANA, F.; OUARDA, T. B. M. J. Multivariate missing data in hydrology – Review and applications. **Advances in Water Resources**, v. 110, p. 299-309, 2017. doi: 10.1016/j.advwatres.2017.10.002
- ALMEIDA M. A.; CURI W. F. Gestão do uso de água na bacia do Rio Paraíba, PB, Brasil com base em modelos de outorga e cobrança. **Revista Ambiente e Água**, Tabuaté, v. 11, n. 4, p. 989-1005, 2016.
- ALVARES, C. A.; STAPE, J. L.; SENTELHAS, P. C.; DE MORAES GONÇALVES, J.L.; SPAROVEK, G. Köppen's climate classification map for Brazil. **Meteorologische Zeitschrift**, Berlin, v. 22, n.6, p.711-728, 2013. doi: 10.1127/0941-2948/2013/0507
- ALVES SOBRINHO, T.; OLIVEIRA, P.T.S.; RODRIGUES, D.B.B.; AYRES, F.M. Delimitação automática de bacias hidrográficas utilizando dados SRTM. **Engenharia Agrícola**, v. 30, n. 1, p.46-57, 2010.
- ALVES, E. D. L.; BIUDES, M. S.; VECCHIA, F.A.S. Interpolação especial na climatologia: Análise dos critérios que precede sua aplicação. **Revista Geonorte**, v.1, n.5, p. 606-618, 2012.
- ARSENAULT, R.; BRISSETTE, F. Analysis of continuous streamflow regionalization methods within a virtual setting. **Hydrological Sciences Journal**, p.1-14, 2016. doi: 10.1080/02626667.2016.1154557.
- ASSANI, A. A.; CHALIFOUR, A. LÉGARÉ, G.; MANOUANE, C-S.; LEROUX, D. Temporal regionalization of 7-day low flows in the St. Laurence watershed in Quebec (Canada). **Water Resources Management**, v. 25, n. 14, p. 3559-3574, 2011. doi:10.1007/s11269-011-9870-6
- ASTER GDEM. Aster Global Digital Elevation Model: Release of ASTER GDEM Version, 2011. Disponível em: <<http://www.jspacesystems.or.jp/ersdac/GDEM/E/4.html>>. Acesso em 24 de ago. de 2017.

AURENHAMMER, FRANZ. A survey of fundamental geometric data structure. **ACM Computing Surveys**, v. 23, n. 3, sept., p. 345-405, 1991.

AZAM, M.; PARK, H. K.; MAENG, S. J.; KIM, H. S. Regionalization of Drought across South Korea Using Multivariate Methods. **Water**, v. 10, n. 1, 24p., 2017. doi:10.3390/w10010024

BAENA, L. G. N.; SILVA, D. D. DA; PRUSKI, F. F.; CALIJURI, M.L. Regionalização de vazões com base em modelo digital de elevação para a bacia do Rio Paraíba do Sul. **Engenharia Agrícola, Jaboticabal**, v. 24, n. 3, p. 612-624, 2004.

BAKKE, H. A.; LEITE, A. S. M.; SILVA, L. B. Multivariate statistics: factorial analysis application. **Revista Gestão Industrial**, v. 04, n. 04, p. 01-14, 2008. doi: 10.3895/S1808-04482008000400001

BALTOKOSKI, V.; TAVARES, M. H. F.; MACHADO, R. E.; OLIVEIRA, M. P. Calibração de modelo para a simulação de vazão e de fósforo total nas sub-bacias dos rios Conrado e Pinheiro – Pato Branco (PR). **Revista Brasileira de Ciência do Solo**, v. 34, p. 253-261, 2010.

BARBOSA, S. E. DA S.; BARBOSA, A. R. J.; SILVA, G. Q.; CAMPOS, E. N. B.; RODRIGUES, V. DE C. Geração de modelos de regionalização de vazões máximas, médias de longo período e mínimas de sete dias para a bacia do Rio do Carmo, Minas Gerais. **Revista Brasileira de Engenharia Sanitária e Ambiental**, v. 10, n. 1, p. 64-71. 2005.

BAZZO K. R.; GUEDES H. A. S.; CASTRO A. S.; SIQUEIRA T. M.; TEIXEIRA-GANDRA C. F. A. Regionalização da vazão Q_{95} : comparação de métodos para a bacia hidrográfica do Rio Taquari-Antas, RS. **Revista Ambiente e Água**, Taubaté, v. 12, n. 5, p. 855-870, 2017. doi:10.4136/ambi-agua.2032

BELDRING, S.; ENGELAND, K.; ROALD, L. A.; SÆLTHUN, N. R.; VOKSO, A. Estimation of parameters in a distributed precipitation-runoff model for Norway. **Hydrology and Earth Sciences**, v. 7, n. 3, p. 304-316, 2003.

BESKOW, S.; CORREA, L. L.; MAHL, M.; SIMÕES, M. C.; CALDEIRA, T. L.; NUNES, G. S.; HUND, E. L.; FARIA, L. C.; MELLO, C. R. DE. Desenvolvimento de um sistema computacional de aquisição e análise de dados hidrológicos. In: Simpósio Brasileiro de Recursos Hídricos, 20, 2013, Bento Gonçalves. **Anais...** Porto Alegre: ABRH, 2013. CD Rom.

BESKOW, S.; MELLO, C. R. de; VARGAS, M. M.; CORRÊA, L. de L.; CALDEIRA, T. L.; DURÃES, M. S. F. et al. Artificial intelligence techniques coupled with seasonality measures for hydrological regionalization of Q_{90} under Brazilian conditions. **Journal of Hydrology**, v. 541, p. 1406-1419, 2016. doi: 10.1016/j.jhydrol.2016.08.046

BESKOW, S.; MELLO, C. R.; FARIA, L. C.; SIMÕES, M. C.; CALDEIRA, T. L.; N., G. S. 2014. Índices de sazonalidade para regionalização hidrológica de vazões de estiagem no Rio Grande do Sul. **Revista Brasileira de Engenharia Agrícola e Ambiental**, v. 18, n. 7, p. 748-754, 2014.

BI, X-A.; LUO, X.; SUN, Q. Branch tire packet classification algorithm based on single-linkage clustering. **Mathematics and Computers in Simulation**, 2017. doi:10.1016/j.matcom.2017.11.003

BIVAND, R.; WILK, J.; KOSSOWSKI, T. Spatial association of population pyramids across Europe: The application of symbolic data, cluster analysis and join-count tests. **Spatial Statistics**, v. 21, p. 339-361, 2017. doi:10.1016/j.spasta.2017.03.003

BOOKER, D. J.; SNELDER, T. H. Comparing methods for estimating flow duration curves at ungauged sites. **Journal of Hydrology**, v. 434-435, p. 78-94, 2012. doi:10.1016/j.jhydrol.2012.02.031

BRASIL. Decreto no 5.334, de 06 de janeiro de 2005, a nova redação ao art. 21 e revoga o art. 22 do Decreto no 89.817, de 20 de junho de 1984, que estabelece as Instruções Reguladoras das Normas Técnicas da Cartografia Nacional. **Diário Oficial da República Federativa do Brasil**, em 07 de janeiro de 2005. Disponível em: < http://www.planalto.gov.br/ccivil_03/_ato2004-2006/2005/decreto/d5334.htm > Acesso em: nov. 2016.

BRASIL. Lei nº 9.433, de 8 de janeiro de 1997. **Diário Oficial da República Federativa do Brasil**. Disponível em < <http://www2.planalto.gov.br/> >. Acesso em: set. 2016.

BRASIL. Lei nº 9.984 de 17 de julho de 2000. **Diário Oficial da República Federativa do Brasil**. Disponível em < <http://www2.planalto.gov.br/> >. Acesso em: set. 2016.

BRASIL. Resolução nº 357, de 17 de março de 2005. **Diário Oficial da República Federativa do Brasil**. Disponível em <<http://www.mma.gov.br/port/conama/res/res05/res35705.pdf>>. Acesso em: set. 2016.

BRESSIANI, D. A.; SRINIVASAN, R.; JONES, C. A.; MENDIONDO, E. M. Effects of spatial and temporal weather data resolutions on streamflow modeling of a semi-arid basin, Northeast Brazil, **International Journal of Agricultural and Biological Engineering**, v. 8, n. 3, p. 125-139, 2015.

CABRAL, S. L.; CAMPOS, J. N. B.; SILVEIRA, C. S. DA; PEREIRA, J. M. R. O intervalo de tempo para uma máxima previsibilidade da precipitação sobre o semiárido brasileiro. **Revista Brasileira de Meteorologia**, v. 31, n.2, p 105-113, 2016.

CABRAL, S. L.; SAKURAGI, J.; SILVEIRA, C. S. Incertezas e erros na estimativa de vazões usando modelagem hidrológica e precipitação por RADAR, **Revista Ambiente e Água**, v. 12, n. 1, Taubaté, 2017. doi:10.4136/ambi-agua.1924

CALDEIRA, T. C.; BESKOW, S.; MELLO, C. R. DE; FARIA, L. C.; SOUZA, M. R. DE; GUEDES, H. A. S. Modelagem probabilística de eventos de precipitação extrema no estado do Rio Grande do Sul. **Revista Brasileira de Engenharia Agrícola e Ambiental**, v.19, n.3, p. 197-203, 2015.

CAMARGO, A. P.; SENTELHAS, P. C. Avaliação do desempenho de diferentes métodos de estimativas da evapotranspiração potencial no Estado de São Paulo, Brasil. **Revista Brasileira de Agrometeorologia**, v.5, n.1, p. 89 – 97, 1997.

CHAVES, H. M. L.; ROSA, J. W. C.; VADAS, R. G.; OLIVEIRA, R. V. T. Regionalização de vazões mínimas em bacias através de interpolação em Sistemas de Informação Geográfica. **Revista Brasileira de Recursos Hídricos**, v. 7, n. 3, p. 43–51, 2002. doi:10.21168/rbrh.v7n3.p43-51

CHIANG, S-M.; TSAY, T-K.; NIX, S. J. Hydrologic Regionalization of Watersheds. I: Methodology Development. **Journal of water resources planning and management**, v. 128, p. 3-11, 2002. doi: 10.1061/(ASCE)0733-9496(2002)128:1(3)

CHO, W.; LEE, J.; PARK, J. KIM, D. Radar polygon method: an areal rainfall estimation based on radar rainfall imageries. **Stoch Environ Res Risk Asses**, n. 31, p. 275-289, 2017. doi: 10.1007/s00477-016-1348-x

CONDUAS, M. Clustering streamflow time series for regional classification. **Journal of Hydrology**, v.407, p. 73–80, 2011. doi:10.1016/j.jhydrol.2011.07.008

CONRADT, T. WECHSUNG, C. G.; WECHSUNG, F. Extending and improving regionalized winter wheat and silage maize yield regression models for Germany: Enhancing the predictive skill by panel definition through cluster analysis. **Agricultural and Forest Meteorology**, v. 216, p. 68-81, 2016. doi:10.1016/j.agrformet.2015.10.003

COSTA, A. M.; REIS, R. S. Análise de conflitos no uso do solo em APPs: O uso de sensoriamento remoto no entorno de um sistema estuarino lagunar. **Revista em Agronegócio e Meio Ambiente**, Maringá- PR, v.10, n.3, p. 775-788, 2017. doi: 10.17765/2176-9168.2017v10n3p775-788

COSTA, A. S.; CARIELLO, B. L.; BLANCO, C. J. C.; PESSOA, F. C. Regionalização de curvas de permanência de vazão de regiões hidrográficas do estado do Pará. **Revista Brasileira de Meteorologia**, v. 27, n. 4, p. 413-422, 2012.

CUNDERLIK, J. M.; BURN, D. H. Switching the pooling similarity distances: Mahalanobis for Euclidean. **WATER RESOURCES RESEARCH**, v. 42, 2006, doi:10.1029/2005WR004245

CUPAK A. Initial Results Of Nonhierarchical Cluster Methods Use For Low Flow Grouping. **Journal of Ecological Engineering**, Poland, v. 18, n. 2, p. 44–50, 2017.

DETZEL D. H. M.; FERNANDES C. V. S.; MINE M. R. M. Não Estacionariedade na Construção de Curvas de Permanência com Vistas à Outorga de Recursos Hídricos. **Revista Brasileira de Recursos Hídricos**, Porto Alegre, v. 21, n.1, p. 80 – 87, 2016.

DIAS, L. S. O.; ROCHA, G. A.; BARROS, E. U. A.; MAIA, P. H. P. Utilização do radar interferométrico para delimitação automática de bacias hidrográficas. **Bahia Análise & Dados**, Salvador, v. 14, n. 2, p. 265-271, set. 2004.

ELESBON, A. A. A.; SILVA, D. D.; SEDIYAMA, G. C.; GUEDES, H. A. S.; RIBEIRO, C. A. A. S.; RIBEIRO, C. B. M. Multivariate statistical analysis to support the

minimum streamflow regionalization. **Engenharia Agrícola, Jaboticabal**, v.35, n.5, p.838-851, 2015.

EMBRAPA - Empresa Brasileira de Pesquisa Agropecuária, Serviço Nacional de Levantamento e Conservação de Solos (Rio de Janeiro, RJ). Súmula da 10.

Reunião Técnica de Levantamento de Solos. Rio de Janeiro, 839, 1979.

ENGELAND, K.; HISDAL, H. A comparison of low flow estimates in ungauged catchments using regional regression and the HBV-Model. **Water Resources Management**, v. 23, p. 2567-2586, 2009.

FARHAN, Y.; AL-SHAikh, N. Quantitative Regionalization of W. Mujib-Wala Sub-Watersheds (Southern Jordan) Using GIS and Multivariate Statistical Techniques. **Open Journal of Modern Hydrology**, v. 7, p.165-199, 2017. doi: 10.4236/ojmh.2017.72010

FARSADNIA, F.; KAMROOD, M. R.; NIA, A. M.; MODARRES, R. BRAY, M. T. HAN, D.; SADATINEJAD, J. Identification of homogeneous regions for regionalization of watersheds by two-level self-organizing feature maps, **Journal of Hydrology**, v. 509, p. 387-397, 2014. doi:10.1016/j.jhydrol.2013.11.050

FÁVERO, L. P. L.; BELFIORE, P.; SILVA, F. L.; CHAN, B. L. **Análise de dados: modelagem multivariada para tomada de decisões**. 1. Ed. Campus: Elsevier 2009. 646p.

FEPAM. Fundação Estadual de Proteção Ambiental Henrique Luiz Roessler – RS. Disponível em:< http://www.fepam.rs.gov.br/qualidade/regioes_hidro.asp> Acesso em: mar. de 2018.

FERNANDES, R. O.; COSTA, C. T. F.; STUDART, T. M. C. Sensitivity analysis of hydrograph peak flows obtained by the SCS method in an urban watershed. **Águas Subterrâneas**, v. 31, n. 3, p.243-254, 2017. doi: 10.14295/ras.v31i3.28812

FLOREK, K.; LUKASZEWICS, J.; PERKAL, J.; STEINHAUS, H.; ZUBRZYCKI, S. Sur la liaison et la division des points d'un ensemble fini. **Colloquium Mathematicum**, v.2, p.282–285, 1951.

Freire, A.J.; Lage, G.B.; Christófar, C. Comparação entre parâmetros morfométricos de bacias hidrográficas gerados por dados SRTM e ASTER GDEM: estudo de caso para bacias do Vale do Jequitinhonha-MG. In: Simpósio Brasileiro de Sensoriamento Remoto (SBSR), 16., 2013, Foz do Iguaçu. **Anais...** São José dos Campos: INPE, 2013. Artigos, p. 5443-5450.

GALLI, P. FRAGA, C. SANTOS, M. P. S. Gravitational force exerted by Brazilian tourist destinations on foreign air travelers. **Journal of Air Transport Management**, v. 55, p.76-83, 2016. doi:10.1016/j.jairtraman.2016.04.01

GANORA, D.; CLAPS, P. LAIO, F.; VIGLIONE, A. An approach to estimate nonparametric flow duration curves in ungauged basins. **Water Resources Research**, v. 45, p.10, 2009. doi:10.1029/2008wr007472

GAO, S.; LIU, P.; PAN, Z.; MING, B.; GUO, S.; XIONG, L. Derivation of low flow frequency distributions under human activities and its implications, **Journal of Hydrology**, v.549, p. 294-300, 2017. doi:10.1016/j.jhydrol.2017.03.071

GARCIA F.; FOLTON N.; OUDIN L. Which objective function to calibrate rainfall–runoff models for low-flow index simulations? **Hydrological Sciences Journal**, p. 1149-1166, 2017.

GARCÍA-MARÍN A.P.; ESTÉVEZ J.; MEDINA-COBO M.T.; AYUSO-MUÑOZ J.L. Delimiting homogeneous regions using the multifractal properties of validated rainfall data series. **Journal of Hydrology**, v. 529, p. 106–119, 2015. doi:10.1080/02626667.2017.1308511

GONÇALVES, M. F.; BLANCO, C. J. C.; SANTOS, V. C.; OLIVEIRA, L. L. S.; PESSOA, F. C. L. Identification of Rainfall Homogenous Regions taking into account El Niño and La Niña and Rainfall Decrease in the state of Pará, Brazilian Amazon. **Acta Scientiarum Technology**, Maringá, v. 38, n. 2, p. 209-216, 2016. doi: 10.4025/actascitechnol.v28i2.26534

GOYAL, M. K.; GUPTA, V. Identification of Homogeneous Rainfall Regimes in Northeast Region of India using Fuzzy Cluster Analysis. **Water Resour Manage**, v.28, p.4491–4511, 2014. doi 10.1007/s11269-014-0699-7

GRANZIERA, M. L. M. A fixação de vazões de referência. **Revista de Direito Ambiental**, v. 18, n. 70, p. 127–148, 2013.

GUILHON, L. G. F.; ROCHA, V. F.; MOREIRA, J. C. Comparação de Métodos de Previsão de Vazões Naturais Afluentes a Aproveitamentos Hidroelétricos. **Revista Brasileira de Recursos Hídricos**, v. 12 n.3 jul/set, p.13-20, 2007.

HAIR JR., J. F.; BLACK, W. C.; BABIN, B. J.; ANDERSON, R. E.; TATHAM, R. L. **Análise multivariada de dados**. 6. Ed. Porto Alegre: Bookman, 2009. 688p.

HANCER, E.; KARABOGA, D. A comprehensive survey of traditional, merge-split and evolutionary approaches proposed for determination of cluster number. **Swarmand Evolutionary Computation**, v.32, p. 49–67, 2017. doi:10.1016/j.swevo.2016.06.004

HAO, C.; ZHANG, J.; YAO, F. Multivariate drought frequency estimation using copula method in Southwest China. **Theor Appl Climatol**, 2015. doi:10.1007/s00704-015-1678-5

HARRIS, C. W. Characteristics of two measures of profile similarity. **Psychometrika**, v.20, p. 289–297, 1955.

HARTIGAN, J. A.; WONG, M. A. A K-means clustering algorithm. **Journal of the Royal Statistical Society - Applied Statistics**. V.28, n. 1, p. 100-108, 1979.

HASS, A.; CONCEIÇÃO, S. R.; HENRICHSEN, J. L.; CHINAZZO, I. B.; RITTER, L. G.; RENZ, V. D. Análise e caracterização da bacia hidrográfica do córrego Fátima localizado no município de Frederico Westphalen-RS. **Revista do Centro do Ciências Naturais e Exatas** - UFSM, Santa Maria, v. 18 n. 1 abr, p. 187-195, 2014. doi: 10.5902/2236117011212

HASSAN, B. G. H.; PING, F. Formation of homogenous regions for Luanhe basin – by using I-moments and cluster techniques. **International Journal of Environmental Science and Development**, v. 3, n. 2, p. 205-210, 2012.

HASSON, S.; LUCARINI, V.; PASCALE, S. Hydrological cycle over South and Southeast Asian river basins as simulated by PCMDI/CMIP3 experiments. **Earth System Dynamics**, v. 4, p. 199-217, 2013.

HE, Z.; WEN, X.; LIU, H.; DU, J. A comparative study of artificial neural network, adaptive neuro fuzzy inference system and support vector machine for forecasting river flow in the semiarid mountain region. **Journal of Hydrology**, v. 509, p. 379-386, 2014.

HOSKING, J. R. M.; WALLIS, J. R. **Regional Frequency Analysis – An Approach Based on L-Moments**. Cambridge: Ed. Cambridge University Press, 1997. 224 p.

HOSKING, J. R. M. Package lmom - L-Moments. 2017. Disponível em :<<https://CRAN.R-project.org/package=lmom> >. Acesso em: 11 de nov. de 2017a.

HOSKING, J. R. M. Package lmomRFA - Regional Frequency Analysis using L-Moments. 2017. Disponível em:<<https://CRAN.R-project.org/package=lmomRFA>>. Acesso em: 11 de nov. de 2017b.

HOSKING, J. R. M.; WALLIS, J. R. Some statistics useful in regional frequency analysis. **Water Resources Research**, v. 29, n.2, p. 271 – 281, 1993.

HSU, N-S.; HUANG, C-J. Estimation of Flow Duration Curve at Ungauged Locations in Taiwan. **Journal of Hydrologic Engineering**, 2017.doi: 10.1061/(ASCE)HE.1943-5584.0001511.

IBGE – Instituto Brasileiro de Geografia e Estatística, 2010. Censo demográfico 2010. Disponível em: < <http://censo2010.ibge.gov.br/estadosat>>. Acesso em 14.10.17.

ILORME, F., GRIFFIS, V.W., A novel procedure for delineation of hydrologically homogeneous regions and the classification of ungauged sites for design flood estimation, **Journal of Hydrology**, v.492, p. 151-162, 2013.

ISHAK, E. H.; RAHMAN, A.; WESTRA, S.; SHARMA, A.; KUCZERA, G. Evaluating the non-stationarity of Australian annual maximum flood. **Journal of Hydrology**, v. 494, p. 134-145, 2013. doi:10.1016/j.jhydrol.2013.04.021

JAIN, S.K.; SUDHEER, K.P. Fitting of hydrologic models: a close look at the Nash–Sutcliffe index. **Journal of Hydrologic Engineering**, v.13, n.10, p.981–986, 2008.

JIMME, M. A.; BASHIR, A.; ADEBAYO, A. A. Spatial distribution pattern and terrain analysis of urban flash floods and inundated areas in Maiduguri metropolis, Borno State, Northeast, Nigeria. **Journal of Geographic Information System**, v. 8, p. 108-120, 2016. doi: 10.4236/jgis.2016.81011

JOSEPH, J. F.; FALCON, E.; SHARIF, H. O. Hydrologic Trends and Correlations in South Texas River Basins: 1950–2009. **Journal of Hydrologic Engineering**, v. 18, p. 1653- 1662, 2013. doi: 10.1061/ (ASCE)HE.1943-5584.0000709.

JOTHEESWARAN, J.; KUMARASWAMY, Y. S. Opinion mining using decision tree based feature selection through Manhattan hierarchical cluster measure. **Journal of Theoretical and Applied Information Technology**, v. 58, n. 1 p. 72-80, 2013.

KAHYA, E.; KALAYCI, S.; PIECHOTA, T.C. Streamflow regionalization: Case study of Turkey. **Journal of Hydrologic Engineering**, v. 13, n. 4, p. 205-214, 2008.

KENDALL, M. G. **Rank Correlation Methods**. 4 ed. London: Charles Griffin, 1975.

KOLMOGOROV, A.N. Grundbegriffe der wahrscheinlichkeitrechnung. Ergebnisse der Mathematik, 1933. English translation: KOLMOGOROV, A.N. Foundations of the theory of probability. **Chelsea Publishing Company**, Second English Edition, New York, 1956.

KULKARNI, A. Homogeneous clusters over India using probability density function of daily rainfall. **Theor Appl Climatol**, 2016. doi :10.1007/s00704-016-1808-8

KUMAR, S.; SINGH, S.K.; MISHRA, P. Multivariate Analysis: An Overview. **Journal of Dentofacial Sciences**, v. 2, n.3, p. 19-26, 2013.

LAAHA, G.; BLÖSCHL, G. Seasonality indices for regionalizing low flows. **Hydrological Processes**, v.20, p.3851-3878, 2006.

LANCE, G. N.; WILLIAMS, W. T. A general theory of classificatory sorting strategies: 1. hierarchical systems. **The Computer Journal**, v.9, n. 4, p. 373–380, 1967
doi:10.1093/comjnl/9.4.373

LEMOS, R. S.; MAGALHÃES, A. P. JR. Reflexões sobre os critérios de cálculo de vazões outorgáveis em áreas de conflito do estado de Minas Gerais: o caso da Bacia do Ribeirão Ribeiro Bonito. **Revista Espinhaço**, p. 4-12, 2015.

LISBOA, L.; CORRÊDO, L. P.; SILVA, D. D.; ULIANA, E. M.; MOREIRA, M. C. Estimativa e regionalização das vazões mínimas anuais e quadrimestrais na bacia do Piracicaba-MG. **Enciclopédia Biosfera**, Centro Científico Conhecer – Goiânia, v. 10, n. 18, p.1580 – 1592, 2014.

LISBOA, L.; MOREIRA, M. C.; DA SILVA, D. D.; PRUSKI, F. F. Estimativa e Regionalização das vazões mínimas e média na bacia do Rio Paracatu. **Engenharia na Agricultura: REVENG**, Viçosa, v. 4, n. 16, p. 471-479, out. 2008.

LOPES, T. R.; PRADO, G.; ZOLIN, C. A.; PAULINO, J.; ANTONIEL, L. S. Regionalização de vazões máximas e mínimas para a bacia do Rio IVAÍ – PR. **Irriga, Botucatu**, v. 21, n. 1, p. 188-201, janeiro-março, 2016.

LUJANO, A. L.; QUISPE, J. P. A.; LUJANO, E. L.; TAPIA, B. A. Regionalizacion de Caudales Mensuales en la Región Hidrográfica del Titicaca, Perú. **Revista de Investigaciones Altoandinas**, v. 19, n. 2, p.219 – 230, 2017. doi: org/10.18271/ria.2017.281

MACÊDO, M. N. C.; DIAS, H. C. T.; COELHO, F. M. G.; ARAÚJO, E. A.; SOUZA, M. L. H.; SILVA, E. Precipitação pluviométrica e vazão da bacia hidrográfica do Riozinho do Rôla, Amazônia Ocidental, **Ambi-Agua, Taubaté**, v. 8, n. 1, p. 206-221, 2013. doi: 10.4136/ambi-agua.809

MAHALANOBIS, P. C. On the Generalized Distance in Statistics. **Proceedings of National Institute of Sciences (India)**, v. 2, n. 1, p. 49-55, 1936.

- MALEKINEZHAD, H.; NACHTNEBEL, H. P.; KLIK, A. Comparing the index-flood and multiple-regression methods using L-moments. **Physics and Chemistry of the Earth**, v.36, p. 54-60, 2011.
- MAMUM, A.A., HASHIM, A., DAOUD, J.I. Regionalization of low flow frequency curves for the Peninsular Malaysia. **Journal of Hydrology**, v.381, p.174–180, 2010.
- MANLY, B. F. J. **Multivariate statistical methods**. 3. Ed. London: Chapman and Hall, 2004. 220p.
- MANN, H. B. Non-parametric test against trend. **Econometrica**, v.13, p.245-259, 1945.
- MARÇAL D. A.; SILVA C. E. Avaliação do impacto do efluente da estação de tratamento de esgoto ETE-Pirajá sobre o Rio Parnaíba, Teresina (PI). **Revista Brasileira de Engenharia Sanitária e Ambiental**, v.22 n.4, p. 761-772, 2017.
- MELATI, M. D. Regressões simples e robusta na regionalização da vazão Q_{95} na Bacia Hidrográfica do Taquari-Antas. **Ciência e Natura**, v.38, n.2, p. 722-739, 2016.
- MELLO, C. R. D.; SILVA, A. M. D. **Hidrologia: princípios e aplicações em sistemas agrícolas**. 1. ed. Lavras: UFLA, 2013. 455 p.
- MELLO, C.R.; VIOLA, M.R.; BESKOW, S.; NORTON, L.D. Multivariate models for annual rainfall erosivity in Brazil. **Geoderma**, v. 202–203, p. 88–102, 2013.
- MELO JÚNIOR, J. C. F.; SEDIYAMA, G. C.; FERREIRA, P. A.; LEAL, B. G. 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, 2006. doi: 10.1590/S1415-43662006000200023.
- MELO, C.C. A.; OLIBEIRA, K. S.; ANGÉLICA, R. S.; PAZ, S. P. A. Análise de agrupamentos associada a difratometria de raios-x: Uma classificação mineralógica prática de bauxitas e seus produtos de digestão bayer. **Holos**, ano 33, v. 6, 2017. doi: 10.15628/holos.2017.6387
- MENDES, C. A. B.; CIRILO, J. A. **Geoprocessamento em recursos hídricos: princípios, integração e aplicação**. Porto Alegre. ABRH, 2001. 536 p.
- MICELI, B. S.; DIAS, F. M.; SEABRA, F. M.; SANTOS, P. R. A.; Fernandes, M. C. Avaliação vertical de modelos digitais de elevação (MDEs) em diferentes configurações topográficas para médias e pequenas escalas. **Revista Brasileira de Cartografia**, n. 63, p. 191-201, 2011.
- MINGOTI, S. A. **Análise de dados através de métodos de estatística multivariada: uma abordagem aplicada**. 2. ed. Minas Gerais: UFMG, 2013, 297p.
- MIRANDA, E.E. Brasil em Relevo. Campinas: Embrapa Monitoramento por Satélite, 2005. Disponível em: <https://www.cnpm.embrapa.br/projetos/relevobr/download/>. Acesso em 21 de nov. de 2016.
- MISHRA, A. K.; COULIBALY, P. Development in hydrometric networks design: A review. **Reviews of Geophysics**, v.47, n.2, p.1-24. 2009.

MODARRES, R. Regional dry spells frequency analysis by L-Moment and multivariate analysis. **Water Resource Management**, v.24, p. 2365-2380, 2010.

MODARRES, R.; SARHADI, A. Statistically-based regionalization of rainfall climates of Iran. **Global and Planetary Change**, v.75, p. 67–75, 2011. doi:10.1016/j.gloplacha.2010.10.009

MOHAMOUD, Y., M. Prediction of daily flow duration curves and streamflow for ungauged catchments using regional flow duration curves. **Hydrological Sciences Journal**, v. 53, n. 4, p. 706-724, 2008.

MOLINA S. K. M.; ALMEIDA I. K.; STEFFEN J. L.; ALVES SOBRINHO T. Caracterização De Disponibilidade Hídrica E Comparação De Metodologias De Regionalização De Vazões. **Geociências**, São Paulo, UNESP, v. 33, n. 3, p.506-515, 2014.

MOREIRA, J. G. V.; NAGHETTINI, M. Detecção de tendências monotônicas temporais e relação com erros dos tipos I e II: Estudo de caso em séries de precipitações diárias máximas anuais do estado do Acre. **Revista Brasileira de Meteorologia**, v. 31, n. 4, 394-402, 2016. doi: 10.1590/0102-778631231420140155

MOREIRA, M. C.; SILVA, D. D. da. Análise de Métodos para Estimativa das Vazões da Bacia do Rio Paraopeba. **Revista Brasileira de Recursos Hídricos**, v. 19, n. 2, p. 313–324, 2014.

MOREIRA, M. C.; SILVA, D. D.; PRUSKI, F. F. LARA, M. S. Índices para identificação de conflitos pelo uso da água: Proposição metodológica e estudo de caso. **Revista Brasileira de Recursos Hídricos**, v. 17 n.3, p. 7-15, 2012.

MORIASI, D. N.; ARNOLD, J. G.; VAN LIEW, M. W.; BINGNER, R. L.; HARMEL, R. D.; VEITH, T. L. Model evaluation guidelines for systematic quantification of accuracy in watershed simulations. **Transactions of the ASABE**, v. 50, n. 3, p. 885–900, 2007. doi:10.13031/2013.23153

MOTOVILOV, Y. G.; GOTTSCHALK, L. ENGELAND, K. RODHE, A. Validation of a distributed hydrological model against spatial observations. **Agricultural and Forest Meteorology**, v.98, n.99, p. 257-277, 1999.

MOURA, M. M.; CUNHA, Z. A.; BESKOW, S.; TERRA, F. S. CASSALHO, F. Influência de diferentes fontes e escalas de informação do relevo na caracterização de bacias hidrográficas. In: Simpósio Brasileiro de Sensoriamento Remoto - SBSR, 18., 2017, Santos. **Anais...Santos: INPE**, 2017. p.5575-5582.

MURTAGH, F.; CONTRERAS, P. Algorithms for hierarchial clustering: an overview. **WIREs Data Mining Knowl Discov**, v. 2, p. 86–97, 2012. doi: 10.1002/widm.53

MWALE, D.; GAN, T. Y.; DEVITO, K.J.; SILINS, U.; MENDOZA, C.; PEDRONE, R. Regionalization of Runoff Variability of Alberta, Canada, by Wavelet, Independent Component, Empirical Orthogonal Function, and Geographical Information System Analysis. **Journal of Hydrologic Engineering**, v. 16, p. 93-107. 2011. doi: 1943-5584.0000284

NAGHETTINI, M.; PINTO, É. J. de A. **Hidrologia estatística**. Belo Horizonte, CPRM, 2007. 552p.

NALDI, M. C.; CAMPELLO, R. J. G. B.; HRUSCHKA, E. R.; CARVALHO, A. C. P. L. F. Efficiency issues of evolutionary k-means. **Applied Soft Computing**, v. 11, n. 2, p. 1938-1952, 2011. doi:10.1016/j.asoc.2010.06.010

NASH, J.E.; SUTCLIFFE, J.V.; 1970. River flow forecasting through conceptual models: Part I – a discussion of principles. **Journal of Hydrologic**, v.10, p.282–290, 1970.

NAZARI, M.; RAZZAGHI, F.; KHALILI, D.; KAMGAR-HAGHIGHI, A. A. ZARANDI, S. M. T. Regionalization of dryland farming potential as influenced by droughts in western Iran. **International Journal of Plant Production**, v. 11, n. 2, p. 315-332, 2017.

NOVAES, L. F. de et al. Gestão de recursos hídricos: uma nova metodologia para a estimativa das vazões mínimas. **Engenharia na Agricultura**, v. 17, n. 1, p. 62–74, 2009.

O'CALLAGHAN, J.F.; MARK, D.M. The extraction of drainage networks from digital elevation data. **Computer vision, graphics, and image processing**, n. 28, p 323-344, 1984.

OLAWOYIN, R.; ACHEAMPONG, P. K. Objective assessment of the Thiessen polygon method for estimating areal rainfall depths in the River Volta catchment in Ghana. **Journal of Geography**, v. 9, n. 2, p. 151-174, 2017.

OLIVEIRA, J.R.; PRUSKI, F.F.; NUNES, A.A. Otimização do aproveitamento da disponibilidade de águas superficiais na bacia do ribeirão Entre Ribeiros. **Revista Brasileira de Recursos Hídricos**, Porto Alegre, v. 18, n. 4, p. 157-172, 2013.

OUARDA, T.B.M.J.; CUNDERLIK, J.M.; ST-HILAIRE, A.; BARBET, M.; BRUNEAU, P.; BOBÉE, B. Data-based comparison of seasonality-based regional flood frequency methods. **Journal of Hydrology**, v. 330, p. 329 – 339, 2006. doi:10.1016/j.jhydrol.2006.03.023

LOUDIN, L.; ANDREASSIAN, V.; MATHEVET, T.; PERRIN, C.; MICHEL, C. Dynamic averaging of rainfall-runoff model simulations from complementary model parameterizations. **Water Resources Research**, v. 42, n. 7, p. 10, 2006. doi:10.1029/2005WR004636

LOUDIN, L.; KAY A.; ANDRÉASSIAN V.; PERRIN C. Are seemingly physically similar catchments truly hydrologically similar?. **Water Resources Research**, v. 46, n. 11, p. 15, 2010. doi:10.1029/2009WR008887

PATIDAR, D.; VERMA, V. K. Identify best similarity matrix to find accurate cluster using dendrogram distance. **International Journal of Scientific Engineering and Science**. v. 1, n. 10, p. 11-14, 2017.

PEREIRA M. A. F.; CAMPO G. F. N.; KOBAYAMA M.; CASTRO N. M. R. Regionalização com geometria hidráulica e fractal: estudo de caso com hidrograma unitário instantâneo geomorfológico. **Revista Brasileira de Recursos Hídricos**, Porto Alegre, v. 21, n.2, p. 347 – 359, 2016.

PEREIRA, M. A. F.; KOBAYAMA, M.; CASTRO, N. M. R. Análise de variâncias pluviométricas na bacia hidrográfica do rio Taboão – RS. *Revista Brasileira de Meteorologia*, v.29, n.3, p. 409 - 419, 2014. doi:10.1590/0102-778620130089

POLIDORI, L.; HAGE, M. EL; VALERIANO, M. DE M. Validação de modelos digitais de elevação sem dados de referência: Aplicação ao MDT Topodata no Brasil. *Boletim de Ciências Geodésicas*, v.20, n.2, p. 467-479, 2014.

PORTALÉS, C.; BORONAT, N.; PARDO-PASCUAL, J. E.; BALAGUER-BESER, A. Seasonal precipitation interpolation at the Valencia region with multivariate methods using geographic and topographic information. *International Journal of Climatology*, v. 30, p.1547–1563, 2010. doi: 10.1002/joc.1988

PRICE, K.; PURUCKER, T.; KRAEMER, S. R.; BABENDREIER, J. E. Tradeoffs among watershed model calibration targets for parameter estimation. *Water Resources Research*, v. 48, 2012. doi:10.1029/2012WR012005

PRUSKI F. F.; RODRIGUEZ R. G.; NUNES A. A.; PRUSKI P. L.; SINGH V. P. Low-flow estimates in regions of extrapolation of the regionalization equations: A new concept. *Journal of the Brazilian Association of Agricultural Engineering*, Jaboticabal, v.35, n.5, p.808-816, 2015.

PRUSKI, F.F.; NUNES, A.A.; PRUSKI, P.L.; RODRIGUEZ, R.G. Improved regionalization of streamflow by use of the streamflow equivalent of precipitation as an explanatory variable. *Journal of Hydrology*, Amsterdam, v. 476, p. 52-71, 2013.

PRUSKI, F.F.; NUNES, A.A.; REGO, F. S.; SOUZA, M. F. Extrapolação de equações de regionalização de vazões mínimas: Alternativas para atenuar os riscos. *Water Resources and Irrigation Management*, v. 1, p. 51-59, 2012.

PRUSKI, F.F.; RODRIGUEZ, R. del G.; SOUZA, J.F.; DA SILVA, B.M.B.; SARAIVA, I.S. Conhecimento da disponibilidade hídrica natural para a gestão dos recursos hídricos. *Engenharia Agrícola, Jaboticabal*, v. 31, n.1, p. 67-77, 2011.

PUSHPALATHA R.; PERRIN C.; MOINE N. L.; ANDRÉASSIAN V. A review of efficiency criteria suitable for evaluating low-flow simulations. *Journal of Hydrology*, v. 420–421, p. 171–182, 2012.

QUEIROZ, M. M. F.; SAMPAIO, S. C.; GOMES, B. M.; LOST, C. Estudo de vazões mínimas $Q_{1,10}$ e $Q_{7,10}$ de rios do Paraná segundo distribuição generalizada. *Revista Verde*, v.5, n.3, p. 32 – 46, 2010.

R Core Team (2013). **R: A language and environment for statistical computing**. R Foundation for Statistical Computing, Vienna, Austria. Disponível em: <<http://www.R-project.org/>>. Acesso em: 8 mar. 2017.

RABUS, B.; EINEDER, M.; ROTH, R.; BAMLER, R. The shuttle radar topography mission—a new class of digital elevation models acquired by spaceborne radar. *ISPRS Journal of Photogrammetry & Remote Sensing*, v.57, p.241-262. 2003.

RAJSEKHAR D., MISHRA, A.K.; SINGH, V.P. Regionalization of drought characteristics using an entropy approach. *Journal of Hydrologic Engineering*, v.18, p.870-887 .2013.

RAO, A. R.; SRINIVAS, V. V. **Regionalization of Watersheds An Approach Based on Cluster Analysis**. Water Science and Technology Library, v. 58, 248p. 2008.

RAO, A. RAMACHANDRA; SRINIVAS, V.V. Regionalization of watersheds by hybrid-cluster analysis. **Journal of Hydrology**, v. 318, p. 37-56, 2006.

RAZARI, T.; COULIBALY, P. An evaluation of regionalization and watershed classification schemes for continuous daily streamflow prediction in ungauged watersheds. **Canadian Water Resources Journal**, v. 42, n. 1, p. 2-20, 2017. doi:10.1080/07011784.2016.1184590

RAZAVI, T.; COULIBALY, P. Streamflow prediction in ungauged basins: review of regionalization methods. **Journal of Hydrologic Engineering**, v.18, n. 8, p. 958–975, 2013.

RAZIEI, T. A precipitation regionalization and regime for Iran based on multivariate analysis. **Theor Appl Climatol**, 2017. doi: 10.1007/s00704-017-2065-1

RENCHE, A. C. **Methods os Multivariate Analysis**. Ed. 2. Wiley-Interscience. 708 p. Canada, 2002.

RIDOLFI, E.; RIANNA M.; TRANI, G.; ALFONSO, L.; BALDASSARRE, G.; NAPOLITANO, F.; RUSSO, F. A new methodology to define homogeneous regions through an entropy based clustering method. **Advances in Water Resources**, v. 96, p. 237-250, 2016.

RIO GRANDE DO SUL. Decreto no 37.033, de 21 de novembro de 1996, regulamenta a outorga do direito de uso da água no Estado do Rio Grande do Sul, prevista nos arts. 29, 30 e 31 da Lei nº 10.350, de 30 de dezembro de 1994. **Diário Oficial da República Federativa do Brasil**, em 22 de novembro de 1996. Disponível em: < <http://www.legislacao.sefaz.rs.gov.br/>>. Acesso em: set. 2016.

RIO GRANDE DO SUL. Decreto no 52.931, de 07 de março de 2016, Dispõe sobre os procedimentos para a Outorga do Direito de Uso da Água e obtenção de Alvará de Obra de Reservatórios em empreendimentos de irrigação, bem como sobre procedimentos para acompanhamento da Segurança de Barragens. **Diário Oficial da República Federativa do Brasil**, em 08 de março de 2016. Disponível em: < goo.gl/YZjCwL>. Acesso em: dez. 2017.

RIO GRANDE DO SUL. LEI Nº 10.350, DE 30 DE DEZEMBRO DE 1994. Institui o Sistema Estadual de Recursos Hídricos, regulamentando o artigo 171 da Constituição do Estado do Rio Grande do Sul. **Diário Oficial da República Federativa do Brasil**, em 01 de janeiro de 1995. Disponível em: < <http://www.al.rs.gov.br/>>. Acesso em: set. 2016.

RIO GRANDE DO SUL. Resolução Consema nº 323, de 08 de Setembro de 2016, Dispõe sobre os procedimentos de Licenciamento Ambiental dos empreendimentos de irrigação. **Diário Oficial da República Federativa do Brasil**. Disponível em < goo.gl/jFQgCH >. Acesso em: dez. 2017.

RODRIGUEZ, R. D. G.; SINGH, V. P.; PRUSKI, F. F.; CALEGARIO, A. T. Using entropy theory to improve the definition of homogeneous regions in the semi-arid region of Brazil. **Hydrological Sciences Journal**, v. 61, n 11, p.2096-2109, 2016.

SALES, M. A. L.; ROMÁN, R. M. S.; SINOBAS, L. R.; MONTEIRO, R. N. F.; SOUZA, J. V. R. S. Avaliação da disponibilidade hídrica na sub-bacia do boi branco através do balanço hídrico climatológico e de cultivo, **Irriga, Botucatu**, v. 22, n. 1, p. 1-17, janeiro-março, 2017. doi: 10.15809/irriga.2017v22n1p1-17

SALVIANO, M. F.; GROppo, J. D.; PELLEGRINO, G. Q. Análise de tendência em dados de precipitação e temperatura no Brasil. **Revista Brasileira de Meteorologia**, v. 31, n. 1, 64-73, 2016.

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, 2011.

SANTOS P. V. C. J.; CUNHA A. C.; Outorga de recursos hídricos e vazão ambiental no Brasil: Perspectivas metodológicas frente ao desenvolvimento do setor hidrelétrico na Amazônia. **Revista Brasileira de Recursos Hídricos**, v. 18, n.3, p. 81-95, 2013.

SANTOS, C. A.; LIMA, A. M.M.; FARIAS, M. H. C. S.; AIRES, U. R. V.; SERRÃO, E. A. O. Análise estatística da não estacionariedade de séries temporais de vazão máxima anual diária na bacia hidrográfica do rio Pardo. **Holos**, v. 7, ano 32, p. 179-193, 2016. doi: 10.15628/holos.2016.4892

SANTOS, E. B.; LUCIO, P. S.; SILVA, C. M. S. Precipitation regionalization of the Brazilian Amazon, **Atmospheric Science Letters**, 2014. doi: 10.1002/asl2.535

SEMA. Secretaria do ambiente e desenvolvimento sustentável. Bacias hidrográficas. Disponível em: < <http://www.sema.rs.gov.br/bacias-hidrograficas> >. Acesso em: set. de 2017.

SEMA. Secretaria do ambiente e desenvolvimento sustentável. Planos de bacias. Disponível em: <<http://www.sema.rs.gov.br/plano-estadual-de-recursos-hidricos>>. Acesso em: ago. de 2016.

SHE, D.; XIA, J.; ZHANG, Y.; SHAN, L. Regional frequency analysis of extreme dry spells during rainy season in the wei river basin, China. **Hindawi Publishing Corporation Advances in Meteorology**, 2016. doi: 10.1155/2016/6427568

SHE, D-X.; XIA, J.; ZHANG, D.; YE, A-Z.; SOOD, A. Regional extreme-dry-spell frequency analysis using the L-moments method in the middle reaches of the Yellow River Basin, China, **Hydrological Processes**, v.28, p.4694–4707, 2014. doi: 10.1002/hyp.9930

SHINA, S.; PAIKA, K. An improved method for single flow direction calculation in grid digital elevation models. **Hydrological Processes**, v. 31, n. 8, p. 1650–1661, 2017. doi: 10.1002/hyp.11135

SILVA B. M. A.; SILVA D. D.; MOREIRA M. C. Influência da sazonalidade das vazões nos critérios de outorga de uso da água: estudo de caso da bacia do rio Paraopeba. **Revista Ambiente e Água**, v. 10, n. 3, p. 123-134, 2015.

SILVA JÚNIOR, O. B., BUENO, E. O., TUCCI, C. E. M., CASTRO, N. M. R. Extrapolação espacial na regionalização da vazão. **Revista Brasileira de Recursos Hídricos**, v.8, n.1, p.21-37, 2002.

SILVA, A. R. Biotools: Tools for biometry and applied statistics in agricultural science. 2017. R package version 3.1. Disponível em: < <http://CRAN.R-project.org/package=biotools> >. Acesso em: 16 de jun. de 2017.

SILVA, C. R. P.; DEMARQUI, E. N.; ALMEIDA, F. T.; MINGOTI, R.; SOUZA, A. P. Diferentes modelos digitais de elevação na caracterização física da bacia hidrográfica do Rio Nandico, MT, Brasil. **Scientia Plena**, v. 11, n. 5, 11p. 2015.

SILVA, D. D.; MARQUES, F. A.; LEMOS, A. F. Avaliação de metodologias de regionalização de vazões mínimas de referência para bacia do rio São Francisco, **Engenharia na Agricultura**: REVENG, Viçosa, v. 17, n. 5, p.392-403, setembro/outubro, 2009.

SILVA, D. D.; MARQUES, F. A.; LEMOS, A. F. Flexibilidade das vazões mínimas de referência com a adoção do período trimestral. **Engenharia na Agricultura**: REVENG, Viçosa, v. 3, n. 19, p.244-254, maio 2011.

SINGH S. K.; MCMILLAN H.; BÁRDOSSY A.; CHEBANA F. Non-parametric catchment clustering using the data depth function, **Hydrological Sciences Journal**, v. 61, n. 15, p. 2649-2667, 2016.

SMIRNOV, N. Table for estimating the goodness-of-fit of empirical distributions. **The Annals of Mathematical Statistics**, v. 19 p. 279–281, 1948.

SOKAL, R. R.; MICHENER, C. D. The effects of different numerical techniques on the phenetic classification of bees of the Hoplitis complex (Megachilidae). Proceedings of the Linnean Society of London. London, v.178, p.59-74, 1957. doi: 10.1111/j.1095-8312.1967.tb00963.x

SOUZA, N. S.; SOUZA, W. J.; CARDOSO, J. M. S. Caracterização hidrológica e influência da cobertura do solo nos parâmetros de vazão do Rio das Fêmeas. **Revista Brasileira de Engenharia Sanitária e Ambiental**, v. 22, n. 3, p. 453-462, 2017. doi: 10.1590/S1413-41522017155279

SRINIVAS V. V. Regionalization of watersheds using soft computing techniques. **Journal of Hydraulic Engineering**, v. 15, n. 1, p. 170-193, 2009.

SWAIN B. J; PATRA K. C. Streamflow Estimation in Ungauged Catchments Using Regional Flow Duration Curve: Comparative Study. **Journal of Hydrologic Engineering**, v. 22, n. 7, 2017.

THIESSEN, A. H. Precipitation averages for large areas. **Monthly Weather Review**, v.39, p. 1082-1089, 1911.

TSAKIRIS, G.; NALBANTIS, I.; CAVADIAS, G. Regionalization of low flows based on canonical correlation Analysis. **Advances in Water Resources**, v. 34, p. 865-872, 2011.

TUCCI, C.E.M. **Modelos hidrológicos**. Porto Alegre, Ed. Universidade UFRGS, 668 p. 1998.

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

ULIANA E. M.; SOUZA L. G. S.; SILVA D. D.; ALMEIDA F. T.; ARAUJO H. B. Regionalização de vazões para o médio e alto Rio Teles Pires – MT. **Revista de Ciências Agrárias**, v. 59, n. 4, p. 333-338, 2017.

ULIANA, E. M.; SILVA, D. D.; ULIANA, E. M.; RODRIGUES, B. S.; CORRÊDO, L. P. Análise de tendência em séries históricas de vazão e precipitação: uso de teste estatístico não paramétrico. **Revista Ambiente e Água**, v. 10, n. 1, p. 82-88, Taubaté – Jan. / Mar. 2015. doi.org/10.4136/ambi-agua.1427

VALERIANO, M. M. **Topodata**: guia para utilização de dados geomorfológicos locais. São José dos Campos, SP: INPE: Coordenação de Ensino, Documentação e Programas Especiais (INPE-15318-RPE/818). 72p., 2008.

VALERIANO, M. M.; ALBUQUERQUE, P. C. G. **Topodata**: processamento dos dados SRTM. São José dos Campos, SP: INPE: Coordenação de Ensino, Documentação e Programas Especiais (INPE-16702-RPQ/854). 79p., 2010.

VALERIANO, M. M.; ROSSETTI, D. F. Topodata: Brazilian full coverage refinement of SRTM data. **Applied Geography**, v. 32, p.300-309, 2012.

VENDRAMIN, L.; CAMPELLO, R. J. G. B.; HRUSHKA, E. R. Relative clustering validity criteria: a comparative overview. **Statistical Analysis and Data Mining**, v. 3, p.209-235, 2010. doi:10.1002/sam

VEZZA, P.; COMOGLIO, C.; ROSSO, M.; VIGLIONE, A. Low flows regionalization in North-Western Italy. **Water Resources Management**, v.24, p.4049-4074, 2010.

VIOLA, M. R.; MELLO, C. R.; ACERBI JR., F. W.; SILVA, A. M. Modelagem hidrológica na bacia hidrográfica do Rio Aiuruoca, MG. **Revista Brasileira de Engenharia Agrícola e Ambiental**, v.13, n.5, p.581–590, 2009.

VOGEL, R. M.; FENNESSEY, N. M. Flow duration curves I: new interpretation and confidence intervals. **Water Resources Planning Management**, v. 120, n. 4, p. 485-504, 1994. doi: 10.1061/(ASCE)0733-9496(1994)120:4(485)

WANG, Y.; LI, J.; FENG, P. HU, R. A Time-Dependent Drought Index for Non-Stationary Precipitation Series. **Water Resources Management**, 2015. doi: 10.1007/s11269-015-1138-0

WARD, JÚNIOR H. Hierarchical grouping to optimize an objective function. **Journal of the American Statistical Association**, Alexandria, v.58, n.301, p.236- 244, 1963.

WILKS, D. S. **Statistical Methods in the Atmospheric Sciences**. Academic Press, 630 p. USA, 2006.

WILLMOTT. C. J.; CKELESON. S. G.; DAVIS. R. E.; FEDDEMA. J. J.; KLINK. K. M.; LEGATES. D.R.; O'DONNELL. J. ROWE. C. M. Statistics for the evaluation and comparison of models. **Journal of Geophysical Research**, Ottawa, v. 90, n. 5, p. 8995-9005, 1985.

WOLFF, W.; DUARTE, S.N.; MINGOTI, R. Nova metodologia de regionalização de vazões, estudo de caso para o Estado de São Paulo. **Revista Brasileira de Recursos Hídricos**, v. 19, n. 4, p.21-33, 2014.

YACOB, Y. M.; SAKIM, H. A. M.; ISA, N. A. M. Decision tree-based feature ranking using Manhattan hierarchical cluster criterion. **World Academy of Science, Engineering and Technology**, v. 62, p. 765-771, 2012.

YANG, T.; SHAO, Q.; HAO, Z-C.; CHEN, X. ZHANG, Z.; XU, C-Y.; SUN, L. Regional frequency analysis and spatiotemporal pattern characterization of rainfall extremes in the Pearl River Basin, China. **Journal of Hydrology**, v.380, p.386–405, 2010. doi:10.1016/j.jhydrol.2009.11.013

ZEMA, D. A.; LABATE, A.; MARTINO, D.; ZIMBONE, S. M. Comparing different infiltration methods of the hec-hms model: the case study of the mésima torrent (Southern Italy), **Land Degradation & Development**, p. 15, 2016. doi: 10.1002/ldr.2591

ZHANG, H.; YAO, Z.; YANG, Q.; LI, S.; BAARTMAN, J. E. M.; GAI, L.; YAO, M.; YANG, X.; RITSEMA, C. J.; GEISSEN, V. An integrated algorithm to evaluate flow direction and flow accumulation in flat regions of hydrologically corrected DEMs. **Catena**, v. 151, p. 174–181, 2017. doi: 10.1016/j.catena.2016.12.009

ZHANG, Q.; GU, X.; SINGH, V. P.; XIAO, M.; XU, C-Y. Stationarity of annual flood peaks during 1951-2010 in the Pearl River basin, China. **Journal of Hydrology**, v. 519, p. 3263-3274, 2014.

ZHANG, Q.; QI, T.; SINGH, V. P.; CHEN, Y. D.; XIAO, M. Regional Frequency Analysis of Droughts in China: A Multivariate Perspective. **Water Resources Management**, 2015. doi: 10.1007/s11269-014-0910-x

ZHANG, Q.; XIAO, M.; SINGH, V. P.; LI, J. Regionalization and spatial changing properties of droughts across the Pearl River basin, China. **Journal of Hydrology**, v. 472-473, p. 355–366, 2012.