

ARTHUR TELLES CALEGARIO

**IDENTIFICAÇÃO DE REGIÕES HIDROLOGICAMENTE HOMOGÊNEAS E
COMPORTAMENTO DAS VAZÕES MÍNIMA E MÉDIA REGIONALIZADAS**

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

VIÇOSA
MINAS GERAIS – BRASIL
2014

**Ficha catalográfica preparada pela Biblioteca Central da Universidade
Federal de Viçosa - Câmpus Viçosa**

T

C148i
2014
Calegario, Arthur Telles, 1987-
Identificação de regiões hidrologicamente homogêneas e
comportamento das vazões mínima e média regionalizadas /
Arthur Telles Calegario. – Viçosa, MG, 2014.
xiv, 80f. : il. (algumas color.) ; 29 cm.

Inclui apêndices.

Orientador: Fernando Falco Pruski.

Dissertação (mestrado) - Universidade Federal de Viçosa.

Referências bibliográficas: f.67-71.

1. Recursos hídricos - Gestão. 2. Medidores de fluxos.
3. Regionalização de vazões. I. Universidade Federal de Viçosa.
Departamento de Engenharia Agrícola. Programa de
Pós-graduação em Engenharia Agrícola. II. Título.

CDD 22. ed. 551.48

ESPAÇO RESERVADO PARA A FICHA CATALOGRÁFICA

ARTHUR TELLES CALEGARIO

**IDENTIFICAÇÃO DE REGIÕES HIDROLOGICAMENTE HOMOGÊNEAS E
COMPORTAMENTO DAS VAZÕES MÉDIA E MÍNIMA REGIONALIZADAS**

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

APROVADA: 15 de abril de 2014.

Eber José de Andrade Pinto

Sílvio Bueno Pereira

Fernando Falco Pruski
(Orientador)

AGRADECIMENTOS

Agradeço a todos os reinos conhecidos, desconhecidos e, principalmente, aos inimagináveis e infinitos.

Agradeço em especial aos meus ídolos e pais queridos, Seu Joaquim e Dona Tânia, e aos meus irmãos amigos, Talita, Tagore e Lharissa.

Aos meus avós, em especial à Dona Mirtes Rola Telles, por rezar e guardar nossa 'grande família'. Você é especial vózinha.

Agradeço a meus tios e tias, em especial aos mais próximos. Vocês são como pais e mães só que um pouco mais distantes.

Aos meus primos, em especial Brunão, Camilo, Carol, Hernonzin, Hugo, Livia, Marcela, Gabriela, os quais compartilho imenso afeto. Obrigado pela infância e juventude. Espero conservar a amizade e carinho para continuar compartilhando.

Ao professor Fernando Falco Pruski, o qual tinha medo no início, mas com o passar do tempo se tornou um amigo, parceiro, às vezes até Pai, o qual tenho muito respeito e admiração. Obrigado por não ter puxado minha orelha, literalmente.

Aos amigos do DEA, em especial ao Rego, Tonin, Aline, Renan, Nívia, Morgana, Raphael, Chicão, Eduardo, Almir, Ciclone, Gláááucio, Manuééél, Wartin, Evandro, Délio, Galinari, Rosária, Claudeniiiiisu, Zé Mauro, Evaristo...

Aos amigos do grupo de almoço e a Dona Rosilda por compartilhar alimentos e momentos saborosos. Abstraíam das brincadeiras sem contexto.

Aos demais amigos que cativei e me cativaram. A galera do Paraíso, dos Cristais, do '30', do 'Cambojah', do Romão do Reis, da Violeira, da Capoeira, do Pólo Aquático, do CEP/Sauípe, do Entre-Folhas, do Leão e Teddys, da Engenharia Agrícola, do DCE, do Eurico Salles etc. Valeu pelos momentos e doideras.

À banda Pink Floyd.

À UFV, principalmente a Engenharia Ambiental e ao Departamento de Engenharia Agrícola, por todas as oportunidades que me foram proporcionadas. Agradeço também a todos os Professores que contribuíram para a minha formação.

À Bela Normando Mascarenhas e Clara Rosa Mascarenhas Calegario pelo presente.

BIOGRAFIA

Arthur Telles Calegario, filho de Joaquim Calegario Filho e Tânia Maria Telles Calegario, nasceu em Vitória – ES, no dia 16 de maio de 1987.

Em maio de 2006 iniciou o curso de graduação em Engenharia Agrícola e Ambiental pela Universidade Federal de Viçosa, o qual concluiu em janeiro de 2012.

Em março de 2012 iniciou o curso de pós-graduação em Engenharia Agrícola, na área de Recursos Hídricos e Ambientais, pela Universidade Federal de Viçosa, submetendo-se a defesa em março de 2014.

SUMÁRIO

LISTA DE TABELAS.....	vii
LISTA DE FIGURAS.....	viii
LISTA DE SÍMBOLOS.....	ix
RESUMO.....	xi
ABSTRACT.....	xiii
1 INTRODUÇÃO.....	1
2 REVISÃO DE LITERATURA.....	3
2.1 Regionalização de vazões e identificação de regiões hidrologicamente homogêneas.....	3
3 MATERIAL E MÉTODOS.....	11
3.1 Região de estudo e dados utilizados.....	11
3.2 Métodos de identificação de regiões hidrologicamente homogêneas.....	14
3.2.1 Conveniência geográfica.....	14
3.2.2 Análise de cluster.....	15
3.2.3 Entropia.....	21
3.3 Avaliação do comportamento estatístico e físico das vazões regionalizadas.....	25
3.3.1 Regionalização das vazões.....	25
3.3.1.1 Variáveis dependentes.....	25
3.3.1.2 Variável independente.....	25
3.3.2 Análise do comportamento estatístico da regionalização.....	26
3.3.3 Análise de comportamento das vazões regionalizadas.....	29
3.3.3.1 Distribuição espacial do CE e da $q_{7,10}$	30
3.3.3.2 Análise de segurança.....	30
4 RESULTADOS E DISCUSSÃO.....	33
4.1 Identificação de regiões hidrologicamente homogêneas.....	33
4.1.1 Conveniência geográfica.....	33

4.1.2	Análise de cluster.....	35
4.1.3	Entropia.....	38
4.1.3.1	Identificação de regiões hidrologicamente homogêneas para regionalização da Q_{mld}	38
4.1.3.2	Identificação de regiões hidrologicamente homogêneas para regionalização da $Q_{7,10}$	39
4.2	Análise do comportamento estatístico e físico das Q_{mld} regionalizadas.....	41
4.2.1	Regionalização da Q_{mld}	41
4.2.2	Análise do comportamento estatístico das Q_{mld} regionalizadas.....	42
4.2.3	Análise do comportamento físico das Q_{mld} regionalizadas.....	44
4.3	Análise do comportamento estatístico e físico das $Q_{7,10}$ regionalizadas.....	54
4.3.1	Regionalização da $Q_{7,10}$	54
4.3.2	Análise do comportamento estatístico das $Q_{7,10}$ regionalizadas.....	55
4.3.3	Análise do comportamento físico das $Q_{7,10}$ regionalizadas.....	56
5	CONCLUSÕES.....	66
6	REFERÊNCIAS.....	67
	APÊNDICE A.....	72
	APÊNDICE B.....	77
	APÊNDICE C.....	79
	APÊNDICE D.....	80

LISTA DE TABELAS

Tabela 1 -.....	22
Tabela 2 -.....	29
Tabela 3 -.....	30
Tabela 4 -.....	33
Tabela 5 -.....	34
Tabela 6 -.....	39
Tabela 7 -.....	40
Tabela 8 -.....	45
Tabela 9 -.....	46
Tabela 10 -.....	47
Tabela 11 -.....	49
Tabela 12 -.....	50
Tabela 13 -.....	56
Tabela 14 -.....	57
Tabela 15 -.....	58

LISTA DE FIGURAS

Figura 1	10
Figura 2	11
Figura 3	12
Figura 4	14
Figura 5	16
Figura 6	32
Figura 7 -	35
Figura 8 -	37
Figura 9 -	38
Figura 10 -	41
Figura 11 -	42
Figura 12 -	43
Figura 13 -	44
Figura 14 -	45
Figura 15 -	47
Figura 16 -	48
Figura 17 -	53
Figura 18 -	54
Figura 19 -	55
Figura 20 -	56
Figura 21 -	57
Figura 22 -	59
Figura 23 -	61

LISTA DE SÍMBOLOS

AE	Amplitude do erro relativo (%)
CK	Calinski Harabasz, índice de validação de agrupamento
Cluster 1	Agrupamento obtido a partir da análise de cluster utilizando as variáveis precipitação e evapotranspiração
Cluster 2	Agrupamento obtido a partir da análise de cluster utilizando a variável precipitação
Cluster 3	Agrupamento obtido a partir da análise de cluster utilizando as variáveis latitude, longitude, precipitação e evapotranspiração
D	Dunn, índice de validação de agrupamento Dunn
E1	Índice de Nash-Sutcliffe utilizando dados, observados e preditos, logaritmados
E2	Índice de Nash-Sutcliffe a partir dos desvios relativos
E3	Índice de Nash-Sutcliffe modificado
Entropia	Agrupamento obtido a partir da análise de entropia
ER_i	Erro relativo na estação 'i' (%)
ER_m	Erro relativo médio (%)
ETR	Evapotranspiração real (mm)
Geo 1	Agrupamento obtido a partir da análise de conveniência geográfica
Geo 2	Agrupamento obtido a partir da análise de conveniência geográfica
Geo 3	Agrupamento obtido a partir da análise de conveniência geográfica
$H(Z)$	Entropia de uma variável aleatória observada na estação Z
$H(Z/W)$	Informação não compartilhada entre estações Z e W
IMD_{wz}	Informação mútua direcional, fração de informação da estação W inferida pela estação Z
IMD_{zw}	Informação mútua direcional, fração de informação da estação Z inferida pela estação W
MDEHC	Modelo digital de elevação hidrograficamente condicionado
Peq	Precipitação equivalente (m^3/s)
Peq_{750}	Precipitação equivalente ao volume precipitado menos 750 mm de abstração (m^3/s)
R^2	Coefficiente de determinação do modelo de regressão

R_{cv}^2	Coeficiente de determinação baseado na validação cruzada
RHH	Região hidrologicamente homogênea
SW	Silhouette Width, índice de validação de agrupamento
T(Z,W)	Informação compartilhada entre estações Z e W
XB	Xie-Beni, índice de validação de agrupamento

RESUMO

CALEGARIO, Arthur Telles, M.Sc., Universidade Federal de Viçosa, abril de 2014. **Identificação de regiões hidrologicamente homogêneas e comportamento das vazões média e mínima regionalizadas**. Orientador: Fernando Falco Pruski.

A regionalização de vazões é uma técnica amplamente utilizada para estimar a disponibilidade hídrica em seções desprovidas de medições. Um dos primeiros passos em um estudo de regionalização de vazões é a identificação de regiões hidrologicamente homogêneas (RHH), entretanto esta é considerada a etapa mais subjetiva e trabalhosa, pois existem diversas metodologias para este fim, sendo que não há um consenso de qual é a mais apropriada. Este trabalho teve como objetivos: identificar, para a parte mineira da bacia do rio Doce, as RHH com base em três metodologias, denominadas análise de cluster, entropia e conveniência geográfica; e avaliar o comportamento estatístico e físico das vazões mínima e média regionalizadas. As variáveis dependentes utilizadas no estudo foram a vazão média de longa duração (Q_{mld}) e a vazão mínima com 7 dias de duração associada a um período de retorno de 10 anos ($Q_{7,10}$), enquanto que a variável independente foi a vazão equivalente ao volume precipitado menos 750 mm (Peq_{750}). Pela análise de cluster obteve-se três agrupamentos de RHH, denominados 'Cluster 1', 'Cluster 2' e 'Cluster 3', enquanto que pela análise a partir da entropia obteve-se um agrupamento, denominado 'Entropia', e pela análise considerando a conveniência geográfica obteve-se três agrupamentos, denominados 'Geo 1', 'Geo 2' e 'Geo 3'. Pela análise estatística se evidenciou que os comportamentos estatísticos dos agrupamentos 'Geo 2' e 'Geo 3' foram superiores aos demais, enquanto que para os resultados para os agrupamentos 'Entropia' e 'Geo 1' foram os piores. Na análise de comportamento físico, os agrupamentos 'Cluster 1', 'Cluster 2', 'Entropia' e 'Geo 1' apresentaram comportamentos inconsistentes e foram considerados não satisfatórios para a

regionalização da Q_{mld} e da $Q_{7,10}$, enquanto que, dentre os agrupamentos considerados satisfatórios, 'Cluster 3' apresentou o comportamento físico mais consistente.

ABSTRACT

CALEGARIO, Arthur Telles, M.Sc., Universidade Federal de Viçosa, April, 2014. **Identification of hydrologically homogeneous regions and behavior of long-term average streamflow and minimum streamflow regionalized.** Adviser: Fernando Falco Pruski.

Streamflow regionalization is a widely used technique to estimate water availability in sections devoid of measurements. One of the first steps in a study of Streamflow regionalization is the identification of hydrologically homogeneous regions (RHH), however this is considered the most subjective and laborious step, because there are several methods for this purpose and there is no consensus on which one is more appropriate. This study aimed to: identify, for the Minas Gerais State part of the Doce river basin, the RHH based on three methodologies, called cluster analysis, entropy and geographical convenience and to evaluate the statistical and physical behavior of regionalized minimum and average streamflow. The dependent variables used in the study were the long-term average streamflow (Q_{mld}) and the minimum streamflow of seven days duration associated with a ten years return period ($Q_{7,10}$) duration, while the independent variable was the streamflow equivalent to the rainfall volume minus 750 mm (Peq_{750}). By cluster analysis yielded three groupings of RHH, called 'Cluster 1', 'Cluster 2' and 'Cluster 3', while the analysis based on the entropy obtained a grouping called 'Entropy', and the analysis considering the geographical convenience obtained three groupings, termed 'Geo 1', 'Geo 2' and 'Geo 3'. By statistical analysis disclosed that the statistical behavior of the groupings 'Geo 2' and 'Geo 3' were superior to the others, while for the results to 'Entropy' and 'Geo 1' were the worst. In the analysis of physical behavior, the 'Cluster 1', 'Cluster 2', 'Entropy' and 'Geo 1' groupings showed inconsistent behavior and were considered insatisfactory for the regionalization of Q_{mld} and $Q_{7,10}$,

whereas among groups considered satisfactory, 'Cluster 3' showed the most consistent physical behavior.

1 INTRODUÇÃO

A crescente demanda pelos recursos hídricos, decorrente do crescimento populacional e das atividades industrial e agropecuária, tem aumentado a necessidade do controle sobre a utilização dos recursos hídricos de maneira a assegurar à atual e às futuras gerações a necessária disponibilidade de água.

O conhecimento da disponibilidade hídrica ao longo da hidrografia é de fundamental importância pois com base nela é possível gerir os recursos hídricos de maneira a atender as demandas antrópicas e ambientais.

A disponibilidade hídrica em uma seção da hidrografia é estimada a partir de séries históricas obtidas em estações fluviométricas, contudo essas não contemplam toda a hidrografia. A regionalização de vazões é uma técnica utilizada para estimar a disponibilidade hídrica em seções da hidrografia desprovidas de estações fluviométricas.

Uma das primeiras etapas no estudo de regionalização das vazões é, em geral, a identificação das regiões hidrologicamente homogêneas (RHH), que é considerada a etapa mais subjetiva e trabalhosa na regionalização de vazões, uma vez que existem diversas metodologias com finalidade de identificar tais regiões e não há consenso de qual é a mais apropriada.

As diferentes metodologias utilizadas na identificação de RHH geram diferentes resultados, propiciando diferentes equações de regionalização, o que, conseqüentemente, produz diferentes estimativas das vazões, fazendo com que o estudo dos métodos de identificação de RHH seja de fundamental importância na correta representatividade da variação, estatística e física, das vazões regionalizadas ao longo da hidrografia.

No trabalho foram utilizados três metodologias a fim de identificar as RHH, denominadas análise de cluster, entropia e conveniência geográfica. A análise de cluster é um método mais comumente utilizado na identificação de RHH em estudos de regionalização. A entropia é uma técnica que envolve maior complexidade estatística e sua utilização é mais recente e menos evidenciada, quando comparada à análise de cluster, enquanto que a identificação de RHH a partir da conveniência geográfica apresenta como principal vantagem a facilidade prática, contudo esta apresenta grande subjetividade na delimitação das RHH.

Objetivou-se nesse trabalho: identificar, para a parte mineira da bacia do Rio Doce, regiões hidrologicamente homogêneas com base em três metodologias, denominadas análise de cluster, entropia e conveniência geográfica; e avaliar o comportamento estatístico e físico das vazões mínima e média regionalizadas.

2 REVISÃO DE LITERATURA

2.1 Regionalização de vazões e identificação de regiões hidrologicamente homogêneas

O conhecimento da vazão que escoar ao longo da hidrografia é imprescindível para o planejamento e a gestão dos recursos hídricos, sendo que o desconhecimento da disponibilidade hídrica dificulta planejar, de maneira racional, a utilização dos recursos hídricos em uma bacia (MAMUM et al., 2010).

A vazão média de longa duração (Q_{mld}) representa a disponibilidade potencial dos recursos hídricos (PRUSKI e PRUSKI, 2011), sendo esta a máxima vazão possível de ser regularizada, abstraindo-se as perdas por evaporação e infiltração.

A vazão mínima é caracterizada pela magnitude, duração e frequência, e representa a disponibilidade natural ao longo da hidrografia (SMAKHTIN, 2001). Quanto ao aspecto qualitativo da água, a vazão mínima é utilizada para determinar, por exemplo, a permissibilidade da descarga de efluentes (BARI e SADEK, 2002). Em termos quantitativos, a vazão mínima é utilizada em projetos de navegação e de concessão do uso d'água, como é o caso de projetos de irrigação, abastecimento urbano e transposição (TSAKIRIS et al., 2011). São representadas, geralmente, pela vazão mínima com sete dias de duração associado a um período de retorno de 10 anos ($Q_{7,10}$) ou pelas vazões associadas a um certo nível de permanência no tempo, como é o caso das vazões associadas a permanência de 90% (Q_{90}) e 95% (Q_{95}) do tempo.

A disponibilidade de informações destas variáveis ao longo da hidrografia é, em geral, deficiente ou, até mesmo, inexistente (EUCLYDES et al., 1999; PANDEY e NGUYEN, 1999; SIVAPALAN et al., 2003). Nesse caso, é necessário utilizar técnicas

que possibilitem a transferência de informações de outros locais cujo comportamento hidrológico seja semelhante ao do local de interesse.

A regionalização hidrológica, ou análise regional, consiste em um conjunto de metodologias estatísticas que visam estimar vazão de interesse (máxima, média ou mínima) em locais cujos dados são escassos ou inexistentes, a partir de dados obtidos em estações fluviométricas localizadas em outras seções (NATHAN e MCMAHON, 1990; EUCLYDES et al., 1999; SIVAPALAN et al., 2003; RAO e SRINIVAS, 2006; LAAHA e BLÖSCHL, 2006; ISIK e SINGH, 2008; VEZZA et al., 2010). Dessa maneira, a regionalização de vazões pode ser utilizada para aumentar a confiabilidade dos quantis estimados em uma seção já monitorada e estimar quantis em áreas desprovidas de monitoramento.

Diversos métodos de regionalização são citados na literatura, sendo que os mais comumente utilizados envolvem a construção de modelos empíricos de regressão que relacionam a vazão de interesse e características fisiográficas e/ou climatológicas da bacia (GAN et al., 1990; RIGGS, 1990). Outros métodos são baseados, por exemplo, no uso de procedimentos de interpolação e calibração hídrica (VOGEL, 2005).

Naghattini e Pinto (2007) citam três tipos de procedimentos de análise regional: métodos que regionalizam os quantis associados a um risco específico, métodos que regionalizam os parâmetros das distribuições de probabilidades e métodos que regionalizam uma curva de quantis adimensionais, geralmente denominados de métodos Index-Flood ou métodos da cheia-índice.

Para os métodos que regionalizam os quantis associados a um risco específico, num primeiro momento, é realizada a análise de frequência local para a série histórica de cada estação, associando uma distribuição de probabilidades que seja mais bem adequada. Posteriormente, determinado o período de retorno T , procura-se encontrar uma relação, baseada na regressão, entre o quantil estimado nas estações e suas respectivas características fisiográficas e/ou climatológicas. Nesse método, o ajuste de uma função de probabilidades comum às amostras provenientes das diferentes estações de coleta de dados não é necessário. Portanto, segundo esse método, é possível estimar, a partir da relação encontrada, o quantil escolhido em qualquer seção da hidrografia. Exemplos de utilização desse método são encontrados em Nathan e McMahon (1990), Laaha e Blöschl (2006), Pruski e Pruski (2011), Pruski et al. (2012) e Pruski et al. (2013).

Os métodos que regionalizam os parâmetros das distribuições de probabilidades assumem, como pressuposto, que todas as estações de coleta de dados possuem mesma distribuição de probabilidades, ou seja, localizadas dentro de uma região considerada estatisticamente homogênea do ponto de vista da variável a ser regionalizada. Após a determinação da distribuição de probabilidades associada à região estatisticamente homogênea em estudo, os parâmetros da função de probabilidades são estimados para cada estação fluviométrica pertencentes a esta região. Posteriormente, é realizada uma análise de regressão entre os parâmetros que definem a distribuição de probabilidades e as características fisiográficas e/ou climatológicas locais e, a partir do modelo obtido, é possível determinar os parâmetros da distribuição de probabilidades em qualquer seção da hidrografia e, conseqüentemente, os quantis associados a diferentes períodos de retorno. Exemplo de utilização desse método é encontrado em Naghettini e Pinto (2007).

O terceiro método é o proposto por Dalrymple (1960) e que consiste em metodologia que visa adimensionalizar quaisquer dados obtidos em pontos de observação, dentro de uma mesma região homogênea, com o intuito de utilizá-los como um conjunto amostral único. A hipótese primordial do método é que os postos de observação formam uma região estatisticamente homogênea e as distribuições de probabilidade dos N postos são idênticas, exceto por um fator de escala local denominado de índice de cheia ou fator de adimensionalização. Determinada a distribuição de probabilidades para a região, pode-se relacionar o índice de cheia com as características fisiográficas e/ou climáticas dos pontos de coleta e utilizar tal relação para a região homogênea considerada, estimando a vazão de estudo na seção de interesse. Exemplos de utilização do método são encontrados em Hosking e Wallis (1997), Bari e Sadek (2002), Chen et al. (2006) e Mamum et al. (2010).

Dos três métodos de regionalização comentados apenas o primeiro não carece da determinação de regiões homogêneas do ponto de vista estatístico, ou seja, mesma distribuição de probabilidades para cada região hidrológica (NAGHETTINI e PINTO, 2007).

A utilização dos modelos de regressão para além dos limites dos dados amostrais não é recomendada por dois motivos (NAGHETTINI e PINTO, 2007): (a) o intervalo de confiança sobre a linha de regressão expande à medida que a variável independente se afasta da média; (b) a relação da variável dependente e independente pode ser

diferente da encontrada para o limite utilizado na construção do modelo. Na prática, essa recomendação limita a utilização da equação de regionalização e, conseqüentemente, o planejamento e a gestão dos recursos hídricos, a uma parte inexpressiva da hidrografia, visto que as estações fluviométricas brasileiras estão associadas a grandes áreas de drenagem (SILVA JÚNIOR et al., 2003; TUCCI, 2002). Contudo, salienta-se que a utilização da equação de regressão para além dos limites dos dados amostrais pode levar a valores de vazões de referências “fictícias”.

O aperfeiçoamento das técnicas de regionalização adquire caráter fundamental para o planejamento e a gestão dos recursos hídricos pois contribuí para melhorar as estimativas das vazões nos cursos d'água e garantir maior confiabilidade e segurança na utilização dos modelos de regressão, principalmente nas zonas de extrapolação da equação de regionalização. Nesse sentido, o uso de variáveis mais explicativas do processo de formação de vazões tende a melhorar a regionalização e diminuir o risco associado à utilização dos modelos de regressão para além dos limites dos dados amostrais (PRUSKI et al., 2012).

Dentre as variáveis explicativas utilizadas nos modelos de regressão, a área de drenagem é a característica física mais utilizada. Segundo Dinpashoh et al. (2004), o uso da variável climática precipitação média de longa duração pode representar uma importante melhoria no modelo de regionalização, uma vez que a variação da precipitação reflete diretamente no comportamento das vazões específicas.

Pruski et al. (2013) propuseram duas novas variáveis explicativas, a vazão equivalente ao volume precipitado (Peq) e a vazão equivalente ao volume precipitado menos um fator de abstração da precipitação correspondente a 750 mm (Peq_{750}), que representa a porção da precipitação que não é convertida em vazão devido a outros processos, em especial à evapotranspiração, o qual foi estimado por Novaes (2005) para estações fluviométricas da bacia do Paracatu. Essas variáveis representam as características físicas e climáticas da bacia em uma só variável permitindo uma representação bidimensional da relação entre a variável dependente e independente. Pruski et al. (2013) e Pruski et al. (2012) evidenciaram melhor desempenho estatístico das vazões regionalizadas e maior segurança em regiões de extrapolação dos limites dos dados amostrais quando considerada a Peq_{750} como variável explicativa do modelo de regionalização.

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 (ELESBON, 2012).

A divisão da área em estudo em regiões com comportamento hidrológico similar é um dos primeiros passos no estudo de regionalização de vazões (MALEKINEZHAD et al., 2011) e visa garantir maior confiabilidade preditiva e segurança nestes (NATHAN e MCMAHON, 1990; SMAKHTIN, 2001), entretanto a identificação das RHH é considerada a parte mais subjetiva e difícil no estudo de regionalização, pois não existe concordância de uma técnica objetiva para determinação das mesmas (NAGHETTINI e PINTO, 2007).

Na literatura são citados diversos métodos de agrupamento a fim de dividir uma área de drenagem em RHH, tais como: análise de cluster, ou conglomerados, (NATHAN e MCMAHON, 1992; VEZZA et al., 2010), padrão residual (LAAHA e BLÖSH, 2006), índice de sazonalidade (LAAHA e BLÖSH, 2006), árvore de classificação e regressão (VEZZA et al., 2010; LAAHA e BLÖSH, 2006), análise de correlação canônica (TSAKIRIS et al., 2011) e entropia (YANG e BURN, 1994).

A análise de cluster é uma das técnicas mais utilizadas a fim de identificar RHH e visa agrupar um conjunto de objetos (estações fluviométricas), baseado em suas variáveis descritivas, de maneira que os objetos dentro de um grupo sejam mais similares entre si quando comparados a objetos contidos em outros grupos. Exemplos de trabalhos que utilizam a análise de cluster como ferramenta para identificação de RHH são Laaha e Blösh (2006), Rao e Srinivas (2006), Isik e Singh (2008) e Elesbon (2012).

Na análise de cluster pode-se utilizar métodos hierárquicos e não hierárquicos. Os primeiros constituem de um processo sucessivo de união de pequenos grupos formando grupos maiores, ou divisão de grandes grupos em pequenos, de maneira a encontrar, em cada passo, a união ou divisão mais eficiente, e assim dividir o conjunto amostral em grupos mais homogêneos em relação as variáveis utilizadas. A classificação dos indivíduos pode ser representada, ao final, por um diagrama bidimensional – dendrograma ou diagrama de árvore – que ilustra a divisão realizada em cada etapa (EVERITT et al., 2001). Dentre os mais utilizados pode-se citar o método da variância mínima de Ward que visa, em cada interação, formar aglomerados de maneira a

minimizar a soma de quadrados dos grupos formados. Exemplos de trabalhos que utilizam o método Ward são Nathan e McMahon (1990), Hosking e Wallis (1997); Rao e Srinivas (2006) e Elesbon (2012).

Os métodos não hierárquicos procuram encontrar os grupos presentes no conjunto de dados através de uma partição única. Dentre os métodos mais utilizados, evidencia-se o K-médias. O método é iniciado com a divisão dos indivíduos em K grupos iniciais e, posteriormente, estes são movidos iterativamente de um grupo para outro de maneira a minimizar a variabilidade dentro dos grupos e maximizar a variabilidade entre os grupos (RENCHER, 2002). Exemplos de trabalhos que utilizam o método k-médias são Isik e Singh (2008) e Dikbas et al. (2013).

As variáveis utilizadas na análise de cluster podem ser divididas em estatísticas locais e características locais (NAGHETTINI e PINTO, 2007). As estatísticas locais são variáveis estimadas a partir dos dados objetos (séries observadas nas estações) como, por exemplo, os coeficientes de variação e de assimetria. As características locais são, a priori, variáveis obtidas a partir de informações pontuais como, por exemplo, latitude, longitude e altitude. Outras características espaciais podem ser incluídas indiretamente como, por exemplo, a precipitação média anual e a declividade média.

Alguns problemas relacionados à análise de cluster são: (a) a escolha das variáveis utilizadas para determinar a similaridade das regiões hidrológicas, pois diferentes tipos de combinações destas dão origem à RHH distintas (NATHAN e MCMAHON, 1990); (b) os diferentes métodos de agrupamentos e medidas de similaridades disponíveis, quando utilizados na mesma base de dados, dão origem a diferentes resultados (NATHAN e MCMAHON, 1990); (c) o fato de que o número de RHH não é conhecido a priori, sendo determinada por índices de validação de agrupamento.

Outro método utilizado na identificação de RHH é baseado no conceito de entropia, proposto por Shannon (1948), no campo da comunicação, o qual mensura as incertezas associadas a um evento aleatório com base em série de observações do evento em si. A entropia pode ser interpretada como uma medida, quantidade, de informação contida no evento e quanto maior a certeza associada a um evento, menores são as possibilidades e, portanto, menos informações esta contida no mesmo (YANG e BURN, 1994). Outra medida dispersão, comumente interpretada como incerteza

associada a um evento aleatório, é a variância, contudo a entropia é uma alternativa à variância e mais indicada quando o tamanho da amostra é pequeno (SING, 2013).

A informação contida nas séries hidrológicas obtidas em uma estação pode ser, em parte, compartilhada com outras estações e, com base nisso, a informação contida em uma estação pode ser inferida por outras estações, sendo que essa característica provê o suporte para a utilização da entropia em estudos hidrológicos (YANG e BURN, 1994).

A identificação de RHH considerando o conceito de entropia possui relativa vantagem sobre outras técnicas de agrupamento, pois prove medidas de informação das estações, informação transferida e perdida entre estações, assim como descrição da relação entre estações de acordo com as características da informação transmitida (YANG e BURN, 1994). Além disso, o método depende somente da série histórica de dados das estações em estudo, o que diminui a subjetividade na escolha das variáveis utilizadas. Exemplos de utilização dessa metodologia são encontrados em Yang e Burn (1994) e Rajsekhar et. al. (2011).

As regiões hidrológicas delimitadas considerando a continuidade geográfica/hidrológica associada ao comportamento de fatores climáticos e fisiográficos, além da combinação de estações que proporcionam melhor ajuste estatístico, foi utilizado por Pruski et. al. (2013) na delimitação de RHH em uma sub-bacia do rio São Francisco. Procedimento semelhante foi utilizado por GPRH e IGAM (2012) a fim de delimitar RHH para a parte mineira da bacia do rio Doce.

As regiões hidrológicas obtidas por esse método podem não ser suficientemente homogêneas, mas tal método pode ser adequado em situações em que poucos dados são disponíveis (SMAKHTIN, 2001). Além disso, segundo o mesmo autor, a utilização de técnicas estatísticas sofisticadas não garante melhores resultados práticos e com significado hidrológico dos que os obtidos pela análise combinada do julgamento científico e da conveniência geográfica.

Os métodos agrupamentos especificados anteriormente são utilizados quando o método de regionalização é o associado a um risco específico, o qual foi utilizado no presente estudo. Quando o trabalho visa regionalizar a distribuição de probabilidades ou os parâmetros da curva de probabilidade, é necessário um teste adicional que visa aferir se as RHH identificadas podem ser consideradas estatisticamente homogêneas, ou seja,

possuírem mesma distribuição probabilística. As estatísticas H , proposta por Hosking e Wallis (1997), são amplamente utilizadas para este fim.

3 MATERIAL E MÉTODOS

3.1 Região de estudo e dados utilizados

A região de estudo é a parte mineira da bacia do rio Doce (Figura 1), a qual abrange uma área de 71.724 km², que corresponde a 86% da área total da bacia.

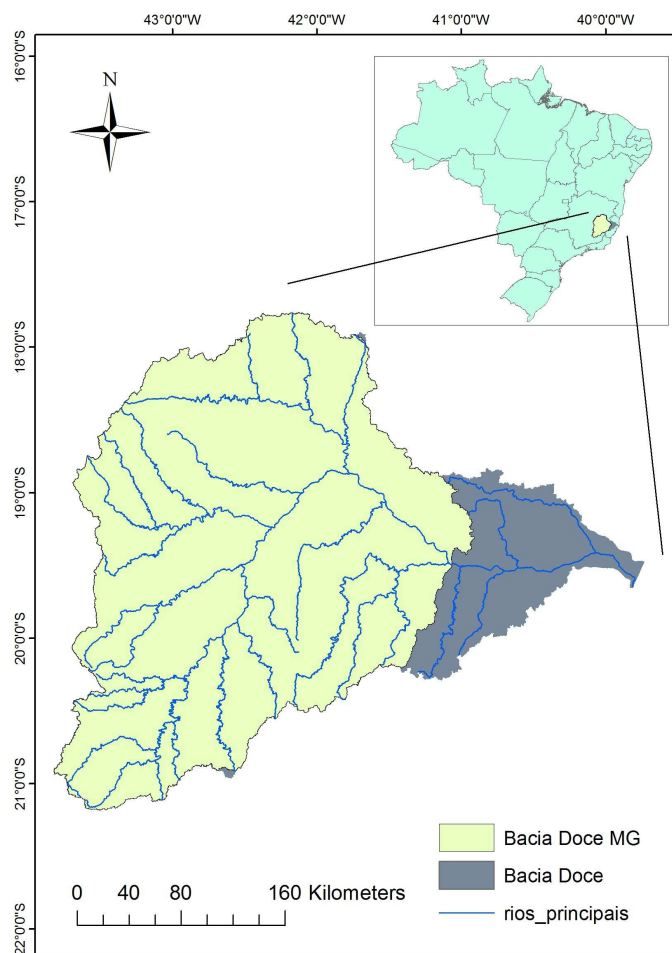


Figura 1 - Localização da bacia do Doce no Brasil e da área em estudo.

O trabalho foi desenvolvido utilizando 38 estações fluviométricas, 80 estações pluviométricas e 14 estações climatológicas (Figura 2). Os dados fluviométricos e pluviométricos pertencem à rede hidrometeorológicas do Sistema de Informações Hidrológicas (Hidroweb) da Agência Nacional de Águas (ANA) e as estações climatológicas ao Instituto Nacional de Meteorologia (INMET). O período base utilizado para estações fluviométricas e pluviométricas corresponde aos anos de 1975 a 2005, cujos dados foram consistidos, e, para as estações climatológicas, de 1961 a 1990, correspondente à normal climatológica. As estações utilizadas no estudo são apresentadas no Apêndice A.

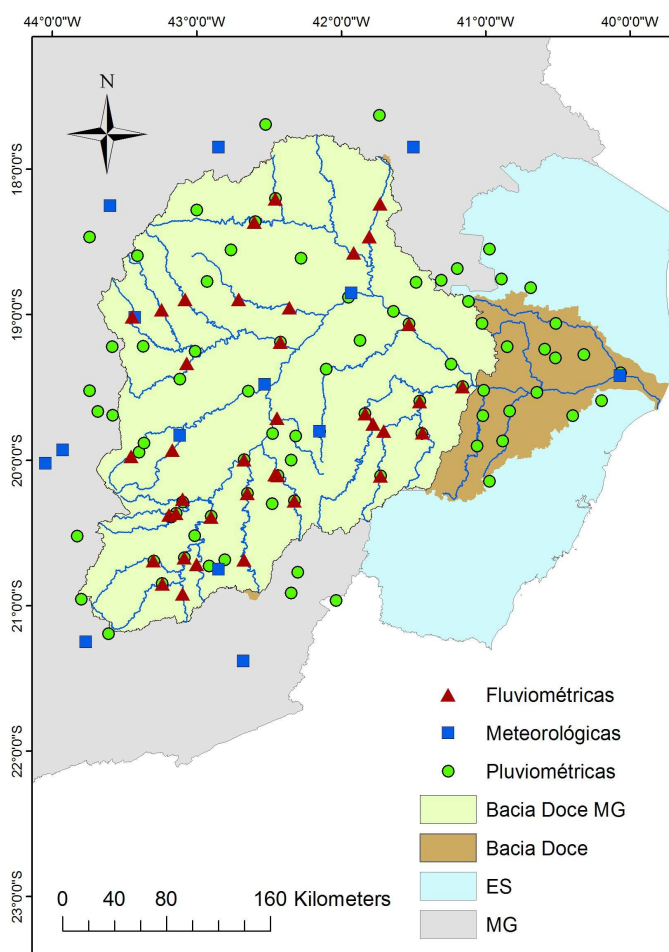


Figura 2 - Localização das estações fluviométricas, pluviométricas e climatológicas utilizadas no estudo.

O modelo digital de elevação hidrograficamente condicionado (Figura 3), disponibilizado pelo Centro de Referência em Recursos Hídricos (CRRH) da Universidade Federal de Viçosa, foi gerado a partir das cartas planialtimétricas e da

hidrografia mapeada, pelo IBGE, na escala de 1:100.000 ou 1:50.000, dependendo da disponibilidade. As altitudes mínima e máxima da área em estudo são 71 m e 2.860 m.

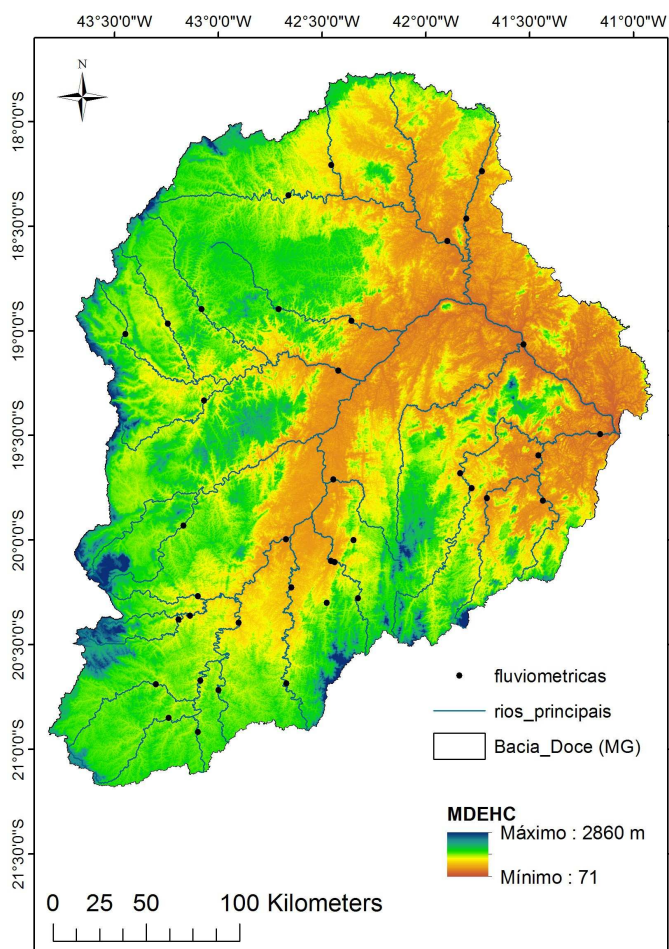


Figura 3 - Modelo digital de elevação hidrograficamente condicionado.
Fonte: CRRH, UFV, 2012.

O Mapa de Domínio Hidrogeológico (Figura 4), definido como “entidade resultante do agrupamento de unidades geológicas com afinidades hidrogeológicas, tendo como base principalmente as características litológicas das rochas” (CPRM, 2007), é disponibilizado pela COMPANHIA DE PESQUISA E RECURSOS MINERAIS (CPRM).

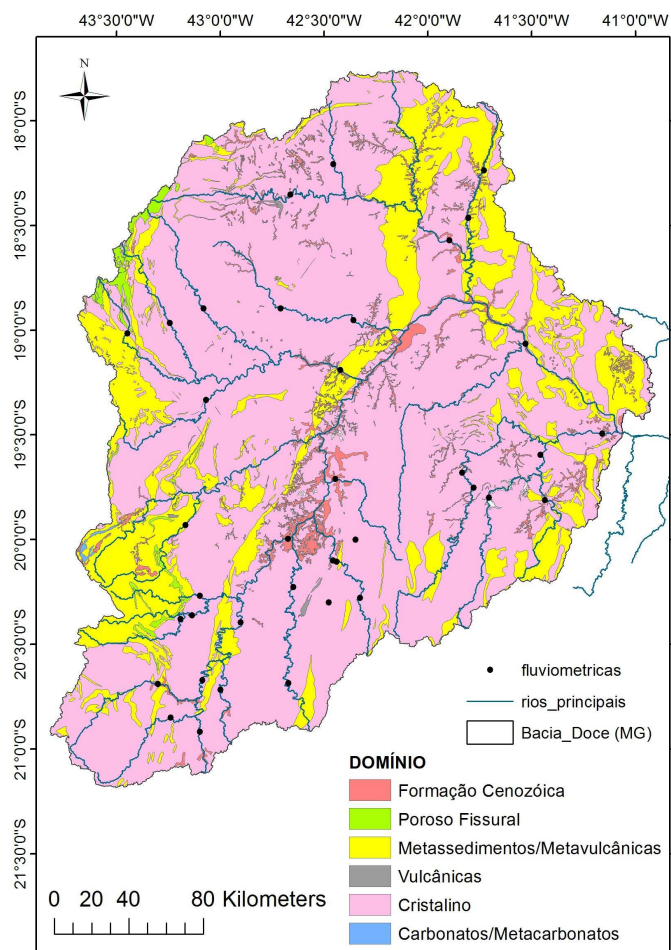


Figura 4: Mapa do domínio hidrogeológico da área em estudo.

3.2 Métodos de identificação de regiões hidrologicamente homogêneas

A identificação das RHH foi realizada utilizando três métodos: conveniência geográfica, análise de cluster e entropia, os quais são descritos na sequência.

3.2.1 Conveniência geográfica

O agrupamento consiste na delimitação de limites geograficamente contínuos e posterior análise estatística dos modelos de regressão, além da análise do comportamento das vazões regionalizadas. É um processo subjetivo de agrupamento que visa identificar regiões em que os erros associados aos modelos de regressão são minimizados. Na delimitação também podem ser consideradas características hidrológicas, climáticas, geológicas dentre outras.

3.2.2 Análise de cluster

Na análise de cluster os objetos são dispostos em uma matriz em que as linhas constituem os indivíduos (N) e as colunas as variáveis descritivas (n) dos indivíduos (equação 1). Cada objeto (estação) é representado por um vetor no espaço dimensional 'n', também denominado vetor característico.

$$A = \begin{bmatrix} X_{11} & X_{12} & \cdots & X_{1n} \\ X_{21} & X_{22} & \cdots & X_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ X_{N1} & X_{N2} & \cdots & X_{Nn} \end{bmatrix}, \quad (1)$$

As variáveis utilizadas no presente trabalho para fim de agrupamento das estações, foram latitude, longitude, precipitação média de longa duração e evapotranspiração real (Apêndice B). A latitude e longitude foram utilizadas a fim de obter regiões geograficamente contínuas (RAO; SRINIVAS, 2006). A precipitação média de longa de duração e a evapotranspiração real foram utilizadas com o intuito de representar a entrada e saída, respectivamente, de água na bacia hidrográfica, pois são, geralmente, os principais componentes do balanço hídrico em uma bacia (HANN et al., 1994).

O mapa de precipitação média de longa duração (Figura 5) usado neste trabalho foi o obtido por GPRH e IGAM (2012), o qual foi utilizado a krigagem como método interpolador.

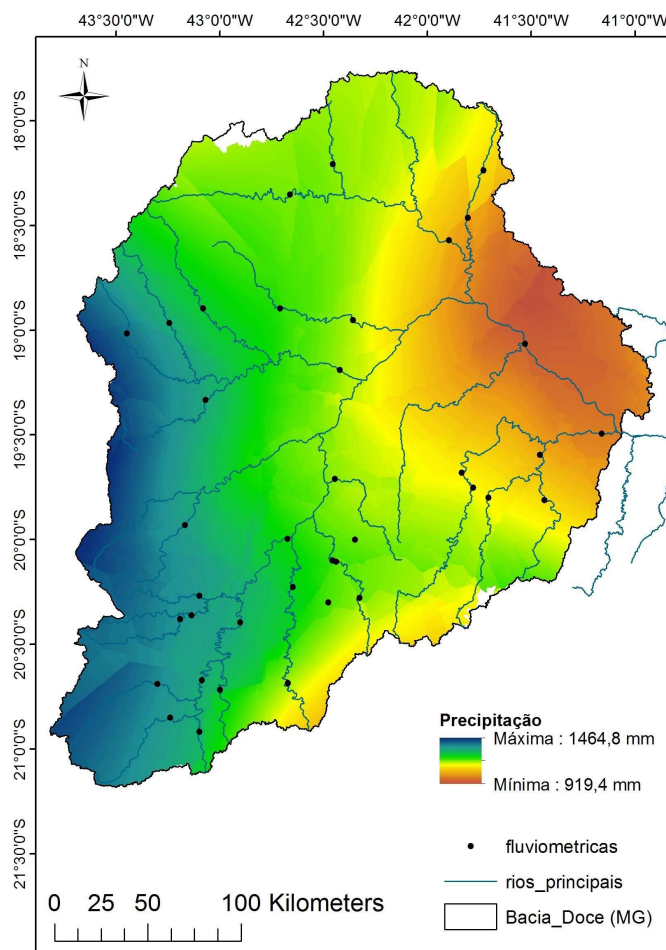


Figura 5 - Mapa de precipitação média de longa duração da área em estudo.
 Fonte: GPRH e IGAM (2012).

A evapotranspiração real (ETR) foi obtida em duas etapas. A primeira referente ao cálculo da evapotranspiração de referência (ET_0), a qual foi estimada pelo método Penman-Monteith FAO 56 (equação 2), descrito por Allen et al. (1998), tido como padrão e considerado mais preciso por considerar maior número de parâmetros envolvidos (SYPERRECK et al., 2008).

$$ET_0 = \frac{0,408 \Delta (R_n - G) + \lambda \frac{900}{T+273} u_2 (e_s - e_a)}{\Delta + \lambda (1 + 0,34 u_2)} \quad (2)$$

em que: ET_0 = evapotranspiração de referência (mm d^{-1}); R_n = saldo de radiação ($\text{MJ m}^{-2} \text{d}^{-1}$); G = fluxo de calor no solo ($\text{MJ m}^{-2} \text{d}^{-1}$); T = temperatura média do ar a 2 m de

altura ($^{\circ}\text{C}$); u_2 = velocidade do vento a 2 m de altura (m s^{-1}); e_s é a pressão do vapor de saturação (kPa); e_a = pressão de vapor real (kPa); $(e_s - e_a)$ = déficit de pressão de vapor (kPa); Δ = declividade da curva de pressão de vapor ($\text{kPa } ^{\circ}\text{C}^{-1}$); e λ = coeficiente psicrométrico ($\text{kPa } ^{\circ}\text{C}^{-1}$).

O cálculo da evapotranspiração de referência foi realizado com auxílio do programa Reference Evapotranspiration Calculator – Ref Et (ALLEN, 2000), obtendo-se a evapotranspiração de referência (ET_o) mensal nas estações climatológicas. A ET_o mensal foi interpolada pelo método do inverso da distância ponderada (IDP), devido à falta de dados amostrais para utilização da krigagem como método interpolador, obtendo-se os mapas mensais de evapotranspiração de referência para a área em estudo.

A segunda etapa constituiu na estimativa da evapotranspiração real pelo balanço hídrico climatológico (BHC) de Thornthwaite e Mather (1955). O balanço hídrico anual foi realizado nas estações pluviométricas (Figura 1), usando a capacidade de água disponível de 100mm, a evapotranspiração de referência mensal e a precipitação média mensal. Pelo somatório dos valores mensais foi obtida a ETR média anual nos locais correspondentes às estações pluviométricas e, por krigagem, obtido o mapa de ETR média anual (Figura 6).

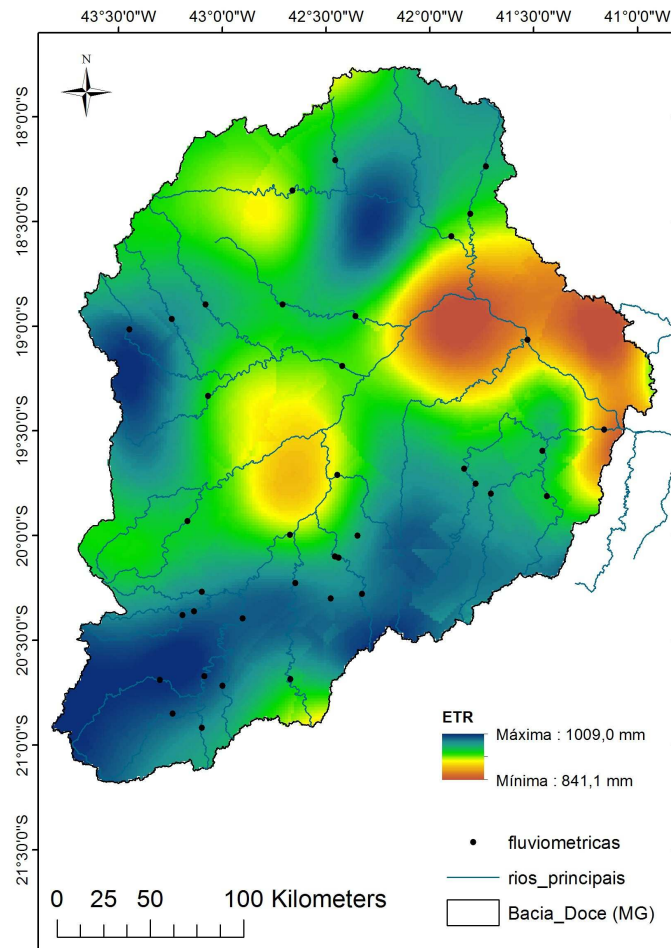


Figura 6 - Mapa de evapotranspiração real média anual.

A análise de cluster é dependente das unidades e escalas em que as variáveis descritivas utilizadas foram medidas (NATHAN & MCMAHON, 1990), assim foi necessário padronizar as variáveis utilizadas de maneira a evitar tal problema (Apêndice B). Para isso, utilizou-se a equação 3.

$$y_{ik} = \frac{x_{ik} - \bar{x}}{\sigma_k}, \quad \text{para} \quad 1 \leq i \leq n, 1 \leq k \leq N \quad (3)$$

em que: n = número de indivíduos; N = número total de variáveis utilizadas; y_{ik} = variável x_{ik} padronizada; σ_k = desvio padrão da variável 'k'; e \bar{x} = média da variável 'k'.

A análise de cluster utilizada no estudo foi a k-médias e consiste em dividir os vetores característicos em 'C' grupos minimizando a função objetiva da equação 4 (HARTIGAN; WONG, 1979).

$$F = \sum_{i=1}^N \sum_{k=1}^C d(y_i; v_k), \quad (4)$$

em que $d(y_i; v_k)$ é a distância euclidiana do vetor característico Y_i ao centróide do cluster 'K'.

A utilização da metodologia K-médias necessita, a priori, do número de grupos a serem formados. Como esse valor não é conhecido de antemão, utilizou-se quatro índices de validação de agrupamento com intuito de estimar o número ótimo de grupos. Os índices utilizados foram Calinski Harabasz (CALINSKI; HARABASZ, 1974), Dunn (DUNN, 1973), Silhouette width (ROUSSEEUW, 1987) e Xie-Beni (XIE; BENI, 1991).

O índice de validação Calinski Harabasz (CK) é estimado pela equação 5 e o número de clusters que maximiza o índice é escolhido como número ótimo de grupos.

$$CK = \frac{N-K}{K-1} \frac{\sum_{k=1}^K d(v_k; v)}{\sum_{k=1}^K \sum_{i=1}^{N_k} d(x_i^{(k)}; v^{(k)})}, \quad (5)$$

em que: N = número de vetores característicos; K = número de clusters; N_k = número de vetores no cluster 'K'; $d(v_k; v)$ = a distância entre o centroide do cluster 'K' e o centroide de todos os vetores característicos pertencentes ao conjunto de dados; e $d(x_i^{(k)}; v^{(k)})$ = a distância do vetor 'i', pertencente ao cluster 'K', e o centroide do cluster 'K'.

O índice de validação Dunn (D) é estimado pela equação 6 e o número de clusters que minimiza o índice é escolhido como número ótimo de grupos.

$$D = \frac{\min_{k \neq k'} d(x_i^k; x_j^{k'})}{\max_{i \neq j} d(x_i^k; x_j^k)}, \quad (6)$$

em que: $d(x_i^k; x_j^{k'})$ = distância entre os vetores característicos 'i' e 'j', pertencentes aos grupos 'K' e 'K''; e $d(x_i^k; x_j^k)$ = distância entre vetores característicos pertencentes ao cluster 'K'.

O índice de validação Silhouette Width (SW) é estimado pela equação 7 e o número de clusters que maximiza o índice é escolhido como número ótimo de grupos.

$$SW = \frac{1}{K} \sum_{k=1}^K \sum_{i=1}^N \frac{1}{n_k} \frac{\min[b(i)] - a(i)}{\max\{a(i), b(i)\}}, \quad (7)$$

em que: K = número de clusters; N = número de vetores característicos pertencentes ao cluster 'K'; $a(i)$ = média das distâncias do vetor de características 'i', pertencente ao cluster 'K', aos demais vetores pertencentes ao mesmo cluster; e $b(i)$ = média das distâncias do vetor característico 'i', pertencente ao cluster 'K', aos vetores característicos pertencentes ao cluster 'W' ($W = 1, \dots, K-1; W \neq K$).

O índice de validação Xie-Beni (XB) é estimado pela equação 8 e o número de clusters que minimiza o índice é escolhido como número ótimo de grupos.

$$XB = \frac{\sum_{k=1}^K \sum_{i=1}^N d(v_k; x_i)}{N \min_{k \neq k'} [d(v_k; v_{k'})]}, \quad (8)$$

em que: K = número de clusters; N = número de vetores característicos pertencentes ao cluster ' K '; $d(v_k; x_i)$ = distância entre o centroide do cluster ' k ' e o vetor característico ' i ', pertencente a ' k '; $d(v_k; v_{k'})$ = distância entre o centroide do cluster ' k ' e ' k' ';

A análise de cluster foi realizada para diferentes combinações das variáveis disponíveis visando caracterizar a influência da escolha das variáveis na identificação das RHH e, conseqüentemente, nas vazões regionalizadas.

A rotina de trabalho relativo a análise de cluster foi implementada no programa R com auxílio dos pacotes 'DAAG' (MAINDONALD; BRAUN, 2013) e 'ClusterCrit' (DESGRAUPES, 2013).

3.2.3 Entropia

A entropia Shannon para uma variável Z , ou entropia marginal de Z , é definida pela equação 9 e é uma quantidade de informação contida em Z . A estimativa da entropia envolvendo variáveis contínuas necessita, a priori, a estimação das probabilidades associada a cada evento, a qual foi realizada através de histograma em que o número de classes (NC) foi estimado pela regra de Sturges (1926), equação 10.

$$H(Z) = - \sum_{i=1}^{NC} P_i \log_2 P_i, \quad (9)$$

em que: ' m ' = número de classes do histograma de ' Z '; e P_i = probabilidade associada a cada classe, ou probabilidade marginal de Z no intervalo i ;

$$NC = 1 + 3,3 \log_{10} N, \quad (10)$$

em que N é o número de eventos associados a série.

Se duas variáveis (Z,W) são consideradas, $T(Z,W)$ é a informação mútua, ou quantidade de informação que uma variável aleatória contém à cerca da outra, e é mensurada por

$$T(Z, W) = H(Z) - H(Z/W), \quad (11)$$

em que: $H(Z/W)$ = perda de informação durante a transmissão ou quantidade de informação não compartilhada entre estações.

A estimativa de $H(Z/W)$ é feita pela equação 12.

$$H(Z/W) = \sum_{i,j}^m P(Z_i, W_j) \log_2 \frac{P(Z_i, W_j)}{P(W_j)}, \quad (12)$$

em que: $P(Z_i, W_j)$ = probabilidade conjunta; $P(W_j)$ = probabilidade marginal da variável W no intervalo j ; i = número de intervalos de classe de Z ; e j = número de intervalos de classe de W .

A probabilidade conjunta $P(Z_i, W_j)$, utilizada no cálculo da informação mútua, foi estimada, analogamente à probabilidade marginal, pela construção de um histograma bivariado.

No contexto da hidrologia, a variável aleatória Z pode ser, por exemplo, a série anual de vazões (médias, mínimas ou máximas) obtida nas séries históricas das estações fluviométricas. Dessa maneira, a entropia $H(Z)$ pode ser interpretada como a quantidade de informação contida na estação Z e $H(W)$ a quantidade de informação contida na estação W . A informação mútua $T(Z,W)$ é a quantidade de informação de Z inferida por W , ou vice-versa (YANG e BURN, 1994). Nesse estudo, as variáveis aleatórias utilizadas foram a vazão mínima com 7 dias de duração (Q_7) e a vazão média ($Q_{\text{média}}$), as quais constituíram a série anual para cada estação em estudo.

A informação mútua é simétrica (i.e. $T(Z,W) = T(W,Z)$) e não negativa. Quando duas estações são estatisticamente independentes, a informação mútua trocada é zero e, quando as duas são totalmente dependentes, a informação de uma estação pode ser completamente transmitida para a outra estação (i.e. $H(Z) = T(Z,W)$). Qualquer situação intermediária à apresentada remete a valores de T entre 0 e H , sendo que quanto maior T , maior a informação trocada entre estações. Dessa maneira, T é um indicador de capacidade de troca de informação e dependência entre estações (YANG; BURN, 1994).

Apesar de T ser um indicador de dependência entre duas estações, este não é um bom indicador, pois o limite superior dos valores de informação mútua entre duas variáveis varia de estação para estação (RAJSEKHAR et al., 2011). A fim de transpor tal problema, a informação mútua foi padronizada, dando origem à informação mútua direcional (IMD), cujos valores variam de 0 a 1. A IMD denota a fração da informação transferida de uma variável para outra e é expressa por

$$IMD_{zw} = \frac{T(Z, W)}{H(W)}, \quad (13)$$

$$IMD_{wz} = \frac{T(Z, W)}{H(Z)}, \quad (14)$$

em que: IMD_{zw} = fração de informação de W inferida por Z ; e IMD_{wz} = fração de informação de Z inferida por W .

O agrupamento das estações é realizado com base na matriz quadrada contendo os valores de IMD_{zw} e IMD_{wz} (i.e. Tabela 1) e o valor de IMD escolhido como limite (threshold). Quando ambos valores são superiores ao limite estabelecido, as estações Z e W são consideradas fortemente dependentes, pois as informações podem ser mutuamente inferidas, logo devem permanecer num mesmo grupo. Quando ambos valores são inferiores ao limite, as estações devem ser mantidas em grupos separados. Quando apenas um IMD é elevado, por exemplo, IMD_{zw} , a estação Z , cuja informação

pode ser inferida por W, pode ser agrupada com Z, caso ela não pertença a nenhum outro grupo. Caso contrário, ela permanece no grupo em que se encontra. Contudo, a estação Z não pode ser adicionada no grupo da estação W (YANG; BURN, 1994).

Tabela 1 - Matriz quadrada contendo valores de IMD_{zw} e IMD_{wz} utilizados no agrupamento por entropia, no qual as linhas correspondem a W e as colunas a Z

Estações	Z=1	Z=2	Z=3	Z=4	Z=5	Z=6	Z=7	Z=8
W=1	1	0,54	0,20	0,13	0,12	0,20	0,21	0,47
W=2	0,45	1	0,52	0,17	0,15	0,32	0,25	0,32
W=3	0,25	0,50	1	0,28	0,15	0,49	0,19	0,29
W=4	0,18	0,15	0,21	1	0,43	0,25	0,48	0,22
W=5	0,14	0,15	0,17	0,40	1	0,21	0,29	0,15
W=6	0,22	0,3	0,50	0,22	0,19	1	0,19	0,23
W=7	0,26	0,27	0,22	0,50	0,31	0,23	1	0,20
W=8	0,41	0,28	0,26	0,16	0,14	0,20	0,21	1

Fonte: Rajsekhar et al. (2011).

O número de regiões é controlado pelo valor de IMD escolhido como limite. Elevados valores de IMD dividem o conjunto de dados em um maior número de grupos, com poucas estações por grupo. Valores baixos de IMD produzem pequenos grupos com elevado número de estações por grupo.

Na matriz de IMD apresentada na Tabela 1, caso o valor limite de IMD fosse 0,35, seriam formados dois grupos, sendo um constituído pelas estações 1, 2, 3, 6 e 8 e outro pelas estações 4, 5 e 7. Se o valor limite de IMD fosse 0,20, todas as estações pertenceriam ao mesmo grupo. Caso o valor limite fosse 0,45, as estações 1, 2 e 3 formariam um grupo e as estações 4 e 7 outro. Como as estações 5 e 8 não apresentaram ambos valores de IMD_{zw} e IMD_{wz} maior que 0,45, deve-se checar se apenas um dos valores de IMD é maior que o limite. Como a estação 8 não pertence a nenhum grupo e IMD_{18} é maior que 0,45, a estação 8 pode ser agrupada no grupo 1. Para a estação 5, nenhum dos valores de IMD é maior que 0,45, dessa maneira esta estação não pertence a nenhum dos grupos.

A escolha do IMD é a etapa subjetiva do método e foi norteada com base em um aspecto: o número de estações preditivas (IMD_{zw} e $IMD_{wz} > \text{limite}$) por grupo deve ser, no mínimo, quatro.

As estações que foram utilizadas a fim de calibrar as equações de regionalização, dentro de cada grupo, foram somente aquelas cujo ambos IMD_{zw} e IMD_{wz} foram maiores que o limiar escolhido, visto que as demais não podem inferir sobre o comportamento de todo o grupo, como explicado anteriormente.

A análise de entropia foi implementada no programa R com auxílio do pacote 'entropy' (HAUSSER; STRIMMER, 2013). O algoritmo para cálculo da entropia necessita que os dados sejam emparelhados, ou seja, mesmo número de anos, e comuns, na série histórica de vazão. Contudo, as séries anuais de vazões continham falhas as quais foram preenchidas, quando possível, pelo método da regressão linear utilizando as estações fluviométricas em análise. Os anos que não foram preenchidos, devido à falta de correlação entre estações, foram desconsiderados da série anual de todas as estações.

3.3 Avaliação do comportamento estatístico e físico das vazões regionalizadas

3.3.1 Regionalização das vazões

3.3.1.1 Variáveis dependentes

As variáveis dependentes utilizadas para a regionalização das vazões na bacia em estudo foram a $Q_{7,10}$ e a Q_{mld} , pois a primeira é a vazão mínima de referência utilizada no estado de Minas Gerais e a segunda representa a disponibilidade hídrica da bacia, sendo que ambas foram disponibilizadas por GPRH e IGAM (2012). A $Q_{7,10}$ foi calculada com base em ajuste de modelo de distribuição teórica, com auxílio do programa SisCAH 1.0.

3.3.1.2 Variável independente

A variável independente utilizada no estudo foi a vazão equivalente ao volume precipitado menos um fator de abstração da precipitação correspondente a 750 mm (Peq_{750}), representada pela equação 15.

$$Peq_{750} = \frac{(P - 750)A}{k}, \quad (15)$$

em que: Peq_{750} = vazão equivalente ao volume precipitado menos 750 mm de abstração, m^3s^{-1} ; P = precipitação média anual na área de drenagem considerada, mm; A = área de drenagem, km^2 ; e k = fator de conversão, o qual é igual a 31.536.

As equações de regressão foram obtidas utilizando-se o programa SisCoRV 1.0 (SOUSA, 2009) desenvolvido pelo Grupo de Pesquisa em Recursos Hídricos – GPRH, vinculado ao Departamento de Engenharia Agrícola da Universidade Federal de Viçosa – UFV.

3.3.2 Análise do comportamento estatístico da regionalização

A análise do ajuste do modelo de regressão aos dados amostrais foi realizada para cada RHH com base no coeficiente de determinação (R^2), estimado pela equação 16.

$$R^2 = \frac{\sum_{i=1}^n (P_i - \bar{O})^2}{\sum_{i=1}^n (O_i - \bar{O})^2}, \quad (16)$$

em que: n = tamanho amostra; P_i = valor estimado da variável dependente; O_i = valor observado da variável dependente; e \bar{O} = média da variável dependente.

O coeficiente de determinação indica o quanto da variância amostral é explicada pelo modelo de regressão. Varia de zero a um, sendo que quanto maior seu valor, maior a parcela da variância amostral é explicada pelo modelo de regressão (NAGHETTINI E PINTO, 2007). O programa SisCoRV disponibiliza, como um dos resultados, o valor de coeficiente de determinação.

A comparação da eficiência dos métodos de identificação de RHH na estimativa das vazões foi baseada em 6 índices estatísticos que são: (1) amplitude do erro relativo; (2) erro relativo médio; (3) eficiência de Nash-Sutcliffe com dados logaritmados; (4) eficiência de Nash-Sutcliffe relativizada; (5) eficiência de Nash-Sutcliffe modificada; e (6) validação cruzada do modelo de regressão.

A amplitude do erro relativo (AE) é constituída pelo maior e menor valor do erro relativo (equação 17) nas estações fluviométricas, considerando toda área em estudo. O erro relativo médio (ER_m) constitui a média dos erros relativos e é dado pela equação 18. Quanto menor o erro relativo médio e a amplitude do erro relativo, melhor é a capacidade preditiva do modelo.

$$ER_i = \frac{P_i - O_i}{O_i}, \quad (17)$$

$$ER_m = \frac{\sum_{i=1}^N ER_i}{N}, \quad (18)$$

em que: P_i = valor estimado da variável dependente; O_i = valor observado da variável dependente; e N = número de valores observados.

A eficiência de Nash-Sutcliffe (E) foi proposta por Nash e Sutcliffe (1970) e é definida pela equação 19.

$$E = 1 - \frac{\sum_{i=1}^n (O_i - P_i)^2}{\sum_{i=1}^n (O_i - \bar{O})^2}, \quad (19)$$

em que: \bar{O} = média da variável dependente observada; P_i = valor estimado da variável dependente; e O_i = valor observado da variável dependente;

A desvantagem da eficiência de Nash-Sutcliffe é que a diferença entre os valores observados e estimados é elevada ao quadrado. Assim, valores de maior magnitude possuem grande influência no índice enquanto valores menores são negligenciados (LEGATES; MCCABE, 1999). Para diminuir tal sensibilidade (KRAUSE et al., 2005), a eficiência de Nash-Sutcliffe (E) foi calculada utilizando os logaritmos dos valores observados e preditos (E_1). Outras maneiras de ponderar tal problema, é utilizar a eficiência de Nash-Sutcliffe a partir dos desvios relativos (E_2), equação 19, e utilizar a equação 20 com coeficiente $j = 1$ (E_3).

$$E_2 = 1 - \frac{\sum_{i=1}^n \left(\frac{O_i - P_i}{O_i} \right)^2}{\sum_{i=1}^n \left(\frac{O_i - \bar{O}}{\bar{O}} \right)^2}, \quad (20)$$

$$E_3 = 1 - \frac{\sum_{i=1}^n |O_i - P_i|^j}{\sum_{i=1}^n |O_i - \bar{O}|^j}, \quad (21)$$

em que: \bar{O} = média da variável dependente observada; P_i = valor estimado da variável dependente; e O_i = valor observado da variável dependente;

A eficiência de Nash-Sutcliffe varia de $-\infty$ a 1, sendo que 1 é o ajuste perfeito e valores negativos indicam que a média dos valores observados é um melhor preditor que o modelo (KRAUSE et al., 2005).

O método de validação cruzada excluindo um elemento (leave-one-out cross-validation) foi utilizado nesse estudo e constitui em: (1) remover a estação 'i' do grupo 'j'; (2) regionalizar a vazão no grupo 'j'; (3) utilizar a equação obtida em (2) para

estimar a vazão em ‘i’; (4) repetir os passos 1-3 para todas as estações do grupo ‘j’; (5) repetir os passos 1-4 para todas as RHH identificadas; e (6) calcular o erro preditivo para cada estação do banco de dados.

A variância do erro preditivo (V_{cv}) e o coeficiente de determinação baseado na validação cruzada (R_{cv}^2) são estimados pelas equações 22 e 23.

$$V_{cv} = \frac{1}{n} \sum_{i=1}^n (P_i^{(-i)} - O_i)^2, \quad (22)$$

$$R_{cv}^2 = \frac{V_q - V_{cv}}{V_q}, \quad (23)$$

em que: V_q = variância espacial da vazão observada em toda área de estudos; O_i = vazão observada para estação ‘i’ e $P_i^{(-i)}$ a vazão estimada para estação ‘i’ quando esta é excluída da análise.

Quanto maior o valor do coeficiente de determinação da validação cruzada, melhor é a capacidade preditiva dos modelos obtidos pelo método de identificação de RHH.

A validação cruzada foi implementada no programa R, enquanto as demais análises da capacidade preditiva do modelo foram realizadas no programa LibreOffice Calc.

3.3.3 Análise de comportamento das vazões regionalizadas

Os modelos de regionalização obtidos para cada RHH permitem estimar a Q_{mld} e a $Q_{7,10}$ ao longo da hidrografia. Contudo, realizar uma análise física das vazões absolutas estimadas é uma tarefa difícil devido à escala de variação das mesmas. Para contornar tal problema, a análise do comportamento físico das Q_{mld} e das $Q_{7,10}$ regionalizadas foi realizada considerando o coeficiente de escoamento (CE) e a vazão específica mínima

($q_{7,10}$), respectivamente, os quais são obtidos pelas equações 24 e 25. A utilização desses índices permite melhor compreensão do comportamento das vazões regionalizadas, pois a escala de variação dos mesmos é reduzida.

$$CE = \frac{Q_{mld}}{Peq}, \quad (24)$$

$$q_{7,10} = \frac{Q_{7,10}}{A}, \quad (25)$$

em que: A = área de drenagem (km^2); $Q_{7,10}$ = vazão mínima (L s^{-1}); Q_{mld} = vazão média de longa duração (L s^{-1}); e Peq = precipitação equivalente (L s^{-1}).

3.3.3.1 Distribuição espacial do CE e da $q_{7,10}$

A avaliação do comportamento físico foi feita pela análise da distribuição espacial do CE, para a Q_{mld} , e da $q_{7,10}$, para as vazões mínimas, verificando-se a tendência do comportamento destas variáveis em relação ao mapa de precipitação.

3.3.3.2 Análise de segurança

A análise de segurança associada às estimativas obtidas para os diferentes métodos de identificação de RHH visa atenuar a insegurança associada à regionalização de vazões, principalmente em zonas de extrapolação dos limites de validade da equação de regressão.

A análise de segurança foi realizada por meio de um histograma de segurança, no qual foi construído em duas etapas. A primeira refere-se a construção de um histograma individual para cada uma das RHH identificadas por um dos métodos de agrupamento

utilizados. Cada classe do histograma recebe uma classificação de risco relacionada a gestão e planejamento dos recursos hídricos, a qual será explicada posteriormente. A segunda etapa consiste na fusão dos histogramas individuais, tendo como base a classificação de risco, formando um histograma final para cada agrupamento.

Os intervalos de cada histograma individual foram definidos com base nos valores mínimo e máximo do CE e da $q_{7,10}$ observados nas estações fluviométricas da RHH em análise. Os intervalos dos histogramas e classificação dos cursos d'água gerados na análise de segurança relativo à Q_{mld} e $Q_{7,10}$, para cada RHH, são apresentados nas Tabelas 2 e 3, respectivamente.

O limite físico máximo do CE, desconsiderando a existência de escoamento subterrâneo de bacias adjacentes, ou seja, quando os limites hidrogeológicos e topográficos são coincidentes, é 1, indicando que todo volume precipitado na bacia é escoado (equação 24) e, por isso, também foi incluído na análise de segurança. Sabe-se que, na prática, valores de CE superiores a 1 é possível, contudo, devido à complexidade da análise e ao fato de que em nenhuma das estações fluviométricas em análise foi observado valor superior a 1, tais valores foram considerados, neste trabalho, fisicamente inaceitáveis.

Os valores estimados de CE e $q_{7,10}$ inferiores, respectivamente, ao menor valor de CE e $q_{7,10}$, observados nas estações fluviométricas, são fisicamente possíveis e, além disso, provêm maior segurança em relação ao planejamento e a gestão de recursos hídricos, visto que, comparados com os valores observados nas estações fluviométricas, estes subestimam o CE e a $q_{7,10}$.

O intervalo constituído pelos valores máximo e mínimo do CE observados nas estações fluviométricas é fisicamente possível e confiável em relação ao planejamento e a gestão dos recursos hídricos, pois nesse intervalo está contido todos os valores observados nas estações fluviométricas. A mesma análise é feita para a $q_{7,10}$ (Tabela 3).

Tabela 2 - Intervalos do histograma da análise de segurança e respectiva classificação dos CE estimados em toda hidrografia

Intervalo	Classificação
$CE_e < CE_{\min_obs}$	Aceitável - Potencial de subestimativa
$CE_{\min_obs} < CE_e < CE_{\max_obs}$	Provável - Confiável
$CE_{\max_obs} < CE_e < 1$	Aceitável - Potencial de superestimativa
$1 < CE_e$	Inaceitável

em que: CE_e = coeficiente de escoamento estimado; CE_{\min_obs} = mínimo valor de coeficiente de escoamento observado nas estações fluviométricas da RHH em análise; e CE_{\max_obs} = máximo valor de coeficiente de escoamento observado nas estações fluviométricas da RHH em análise.

Tabela 3 - Intervalos do histograma da análise de segurança e respectiva classificação das $q_{7,10}$ estimadas em toda hidrografia

Intervalo	Classificação
$q_e < q_{\min_ob}$	Aceitável - Potencial de subestimativa
$q_{\min_ob} < q_e < q_{\max_ob}$	Provável - Confiável
$q_{\max_ob} > q_e$	Aceitável - Potencial de superestimativa

em que: q_e = coeficiente de escoamento estimado; q_{\min_ob} = mínimo valor de coeficiente de escoamento observado nas estações fluviométricas da região hidrologicamente homogênea em análise; e q_{\max_ob} = é máximo valor de coeficiente de escoamento observado nas estações fluviométricas.

Os CE estimados cujos valores estejam entre o máximo valor observado nas estações fluviométricas e o máximo valor de CE fisicamente possível ($CE=1$) caracterizam um intervalo cujos valores são considerados inseguros para o planejamento e gestão de recursos hídricos, pois podem levar à superestimativa do CE na seção considerada. Procedimento semelhante é realizado para a $q_{7,10}$, contudo, como mencionado anteriormente, é inexistente um limite superior para esta.

O intervalo constituído pelo valor de CE maior que a unidade compõe a quarta classe do histograma de risco relativo à Q_{mld} , sendo que valores de CE maiores que a unidade caracterizam um comportamento definido, neste trabalho, como fisicamente inaceitável.

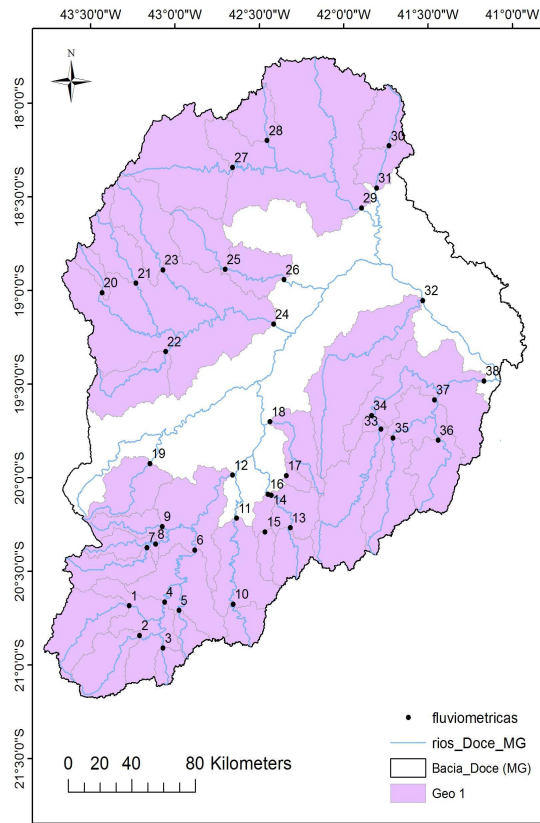
4 RESULTADOS E DISCUSSÃO

Os resultados foram divididos em três subtópicos. O primeiro relativo à identificação de RHH, o segundo inerente à regionalização da Q_{mld} e a análise do comportamento das vazões regionalizadas e o terceiro referente à regionalização da $Q_{7,10}$ e a análise do comportamento das vazões regionalizadas.

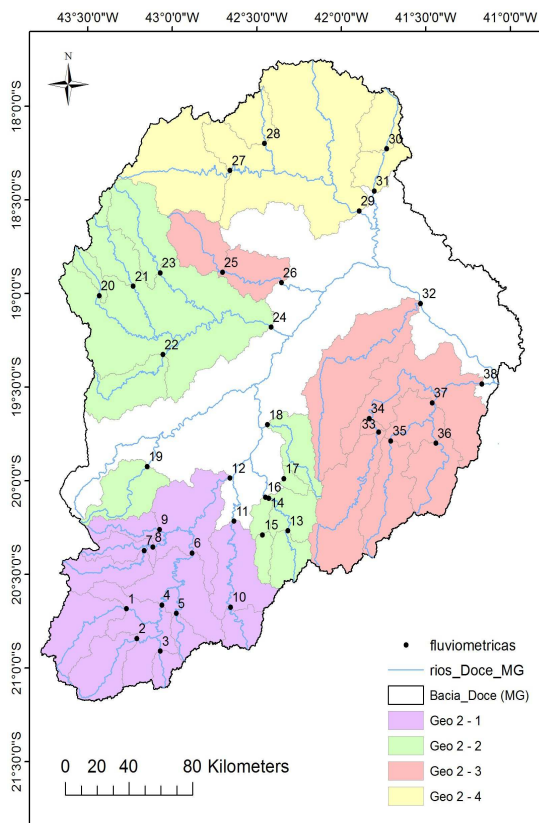
4.1 Identificação de regiões hidrologicamente homogêneas

4.1.1 Conveniência geográfica

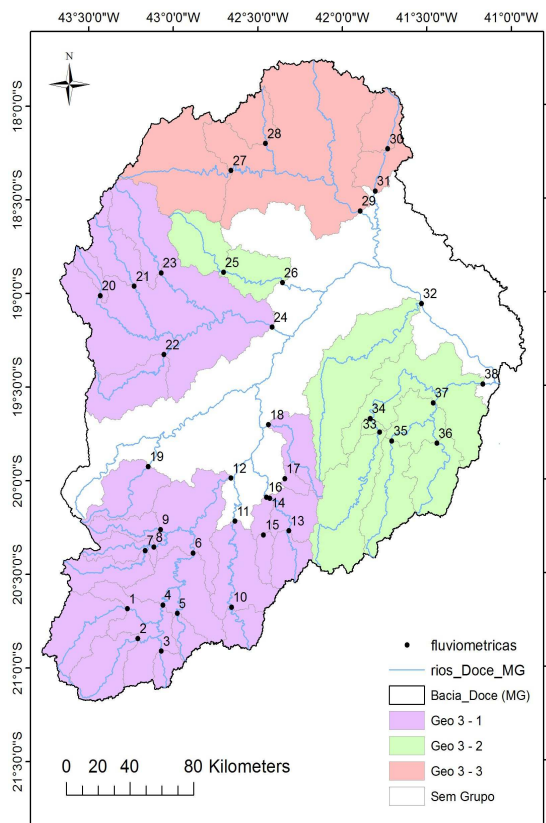
A identificação de RHH pelo método de conveniência geográfica resultou em três agrupamentos, denominados 'Geo 1', 'Geo 2' e 'Geo 3' e apresentados na Figura 6. O agrupamento apresentado na Figura 6a foi proposto com o intuito de caracterizar o comportamento estatístico e físico das vazões regionalizadas quando toda a área em estudo é considerada hidrologicamente homogênea, enquanto os agrupamentos apresentados nas Figuras 6b e 6c foram propostos por apresentarem os melhores desempenhos estatísticos entre os agrupamentos testados pelo método de conveniência geográfica.



(a)



(b)



(c)

Figura 7 - Regiões hidrologicamente homogêneas definidas pelo método de conveniência geográfica. 'Geo 1' (a), 'Geo 2' (b) e 'Geo 3' (c).

4.1.2 Análise de cluster

Na Tabela 4 são apresentados os índices de validação referentes às análises de cluster utilizando as variáveis: precipitação e evapotranspiração (Cluster 1); e precipitação (Cluster 2). Os valores em negrito indicam o número ótimo de clusters para cada índice de validação.

Tabela 4 - Índices de validação do número recomendável de grupos para 'Cluster 1' e 'Cluster 2'

Índices de validação		Número de regiões			
		2	3	4	5
Cluster 1 (P e ETR)	Calinski Harabasz (máx)	42,7	38,3	49	46,4
	Dunn (mín)	0,16	0,16	0,24	0,25
	Sillhouette (máx)	0,45	0,415	0,45	0,43
	Xie_Beni (mín)	3,6	2,47	1,42	1,52
Cluster 2 (P)	Calinski Harabasz (máx)	166	193	215	325
	Dunn (mín)	0,41	0,16	0,2	0,15
	Sillhouette (máx)	0,72	0,63	0,59	0,62
	Xie_Beni (mín)	0,44	2,63	2,56	3,03

Os índices de validação para os agrupamentos 'Cluster 1' e 'Cluster 2' são divergentes em relação ao número recomendado de clusters em que as estações devem ser divididas. Dessa maneira, a escolha do número de clusters foi realizada considerando a predominância dos quatro índices de validação.

Para o 'Cluster 1' os índices de validação indicam, predominantemente, que o número de clusters recomendável é quatro. Em relação ao agrupamento 'Cluster 2', não há predominância do número de clusters, pois os índices SW e XB conduzem a uma recomendação de dois clusters enquanto os índices CH e D indicam um número de clusters igual a cinco. Assim, para o agrupamento 'Cluster 2', foram consideradas ambas as opções, denominadas 'Cluster 2/2' e 'Cluster 2/5'.

Na Tabela 5 são apresentados os índices de validação referentes às análises de cluster considerando as combinações de variáveis: latitude, longitude, precipitação e evapotranspiração real (Cluster 3); e latitude, longitude e precipitação (Cluster 4).

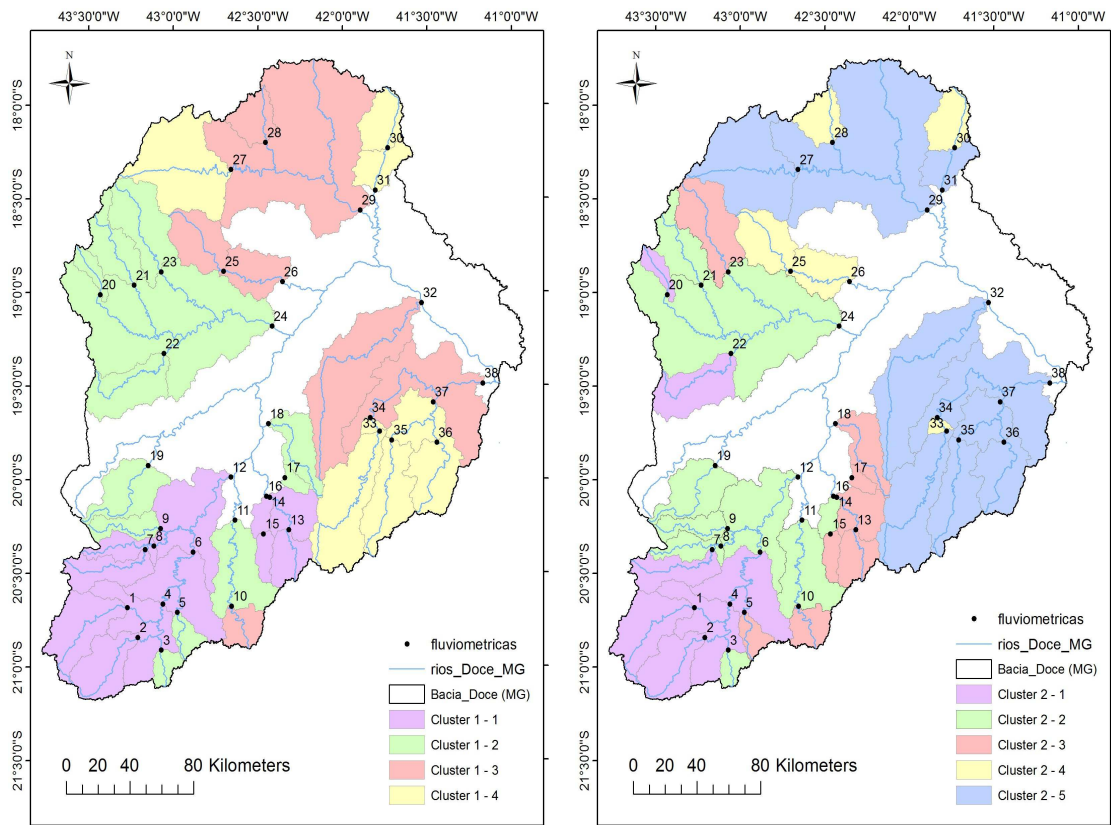
Tabela 5 - Índices de validação do número recomendável de grupos para 'Cluster 3' e 'Cluster 4'

Índices de validação		Número de regiões			
		2	3	4	5
Cluster 3 (Lat. Long. P e ETR)	Calinski Harabasz (máx)	37,2	30,2	28,9	27,4
	Dunn (mín)	0,30	0,19	0,22	0,23
	Sillhouette (máx)	0,43	0,33	0,37	0,39
	Xie_Beni (mín)	1,44	3,31	1,77	1,79
Cluster 4 (Lat. Long. e P)	Calinski Harabasz (máx)	58,6	46,5	46,8	55,6
	Dunn (mín)	0,37	0,21	0,30	0,29
	Sillhouette (máx)	0,52	0,42	0,44	0,43
	Xie_Beni (mín)	0,95	2,45	1,03	1,33

Os índices de validação para os agrupamentos 'Cluster 3' e 'Cluster 4', assim como para os agrupamentos 'Cluster 1' e 'Cluster 2', são discordantes em relação ao número de clusters em que as estações devem ser divididas, mas indicam, predominantemente, que o número de clusters recomendável para os dois agrupamentos é dois. Portanto, nas análises de cluster nas quais as variáveis latitude e longitude foram incluídas (Tabela 5), observou-se maior concordância dos índices de validação quando comparadas às análises em que as mesmas não foram incluídas (Tabela 4).

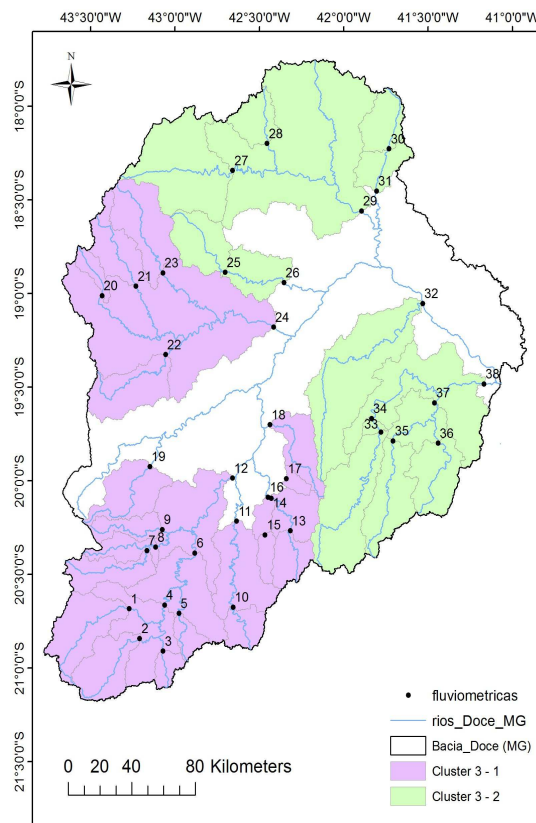
A escolha das variáveis a serem utilizadas na análise de cluster, assim como a escolha do número de grupos e dos índices de validação de agrupamento, são as etapas subjetivas dessa metodologia.

A distribuição das estações fluviométricas para os grupos formados é apresentada no Apêndice C. Os agrupamentos 'Cluster 2/2', 'Cluster 3' e 'Cluster 4' apresentam a mesma classificação para as estações fluviométricas, ou seja, as RHH formadas para estes agrupamentos são idênticas. Dessa maneira os três agrupamentos foram representados pelo agrupamento 'Cluster 3'. O agrupamento 'Cluster 2/5' foi renomeado para 'Cluster 2'. Resumidamente, a análise de cluster gerou três resultados distintos, denominados 'Cluster 1', 'Cluster 2' e 'Cluster 3' e são apresentados na Figura 7



(a)

(b)



(c)

Figura 8 - Regiões hidrologicamente homogêneas definidas pela análise de cluster. 'Cluster 1' (a), 'Cluster 2' (b) e 'Cluster 3' (c).

4.1.3 Entropia

A delimitação das RHH pela entropia é realizada individualmente para Q_{mld} e $Q_{7,10}$. Dessa maneira, os resultados obtidos são apresentados em tópicos distintos.

4.1.3.1 Identificação de regiões hidrologicamente homogêneas para regionalização da Q_{mld}

A escolha do valor limite de IMD para definir as RHH é a fase subjetiva do método da entropia. Diversos valores foram testados e o valor limite de IMD escolhido, considerando a Q_{mld} , foi 0,54.

Foi observado que para valores superiores a 0,54, as RHH formadas eram constituídas de poucas estações, o que inviabilizava a análise de regressão, visto que algumas regiões apresentavam duas e/ou três estações. Além disso, o número de estações excluídas da análise era elevada, por apresentar IMD_{wz} e IMD_{zw} menor que o limite escolhido. Para valores inferiores ao limite escolhido, toda área em estudo era considerada hidrologicamente homogênea, sendo que a diferença dos valores testados eram quanto ao número de estações excluídas da análise (IMD_{wz} e $IMD_{zw} < 0,54$). As RHH delimitadas pelo método da entropia, considerando a Q_{mld} , são apresentadas na Figura 9.

Pela análise de entropia, para fins de regionalização da Q_{mld} , dividiu-se a área em estudo em duas RHH, sendo que as estações 3, 5, 15, 17 e 33 foram excluídas das análises posteriores por apresentarem valores de IMD inferior a 0,54.

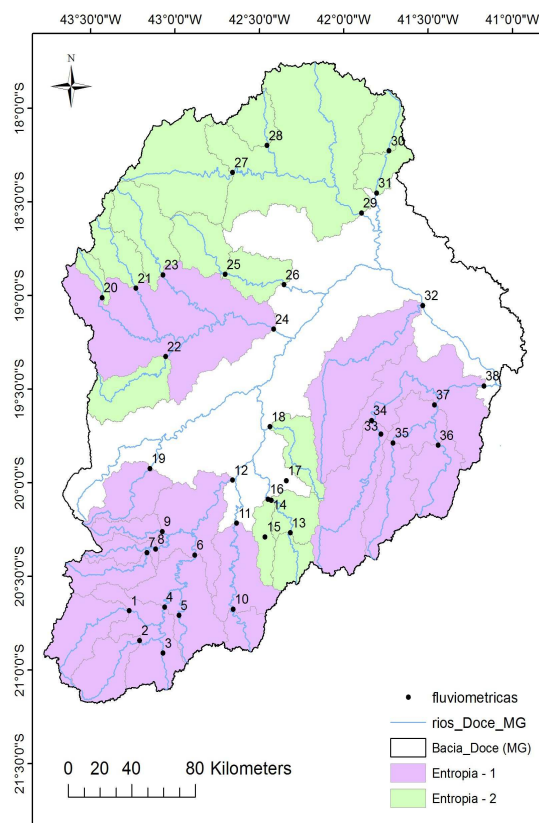


Figura 9 - Regiões hidrologicamente homogêneas definidas pela análise de entropia considerando a vazão média de longa duração.

4.1.3.2 Identificação de regiões hidrologicamente homogêneas para regionalização da $Q_{7,10}$

A escolha do valor limite de IMD para definir as RHH com intuito de regionalizar a $Q_{7,10}$ seguiu os mesmos critérios citados no item 4.1.3.1.

O valor limite de IMD foi 0,64. A escolha de limites superiores e inferiores à 0,64 levaram as mesmas conclusões apontadas no item 4.1.3.1. As RHH delimitadas pelo método da entropia para regionalização da $Q_{7,10}$ são apresentadas na Figura 10.

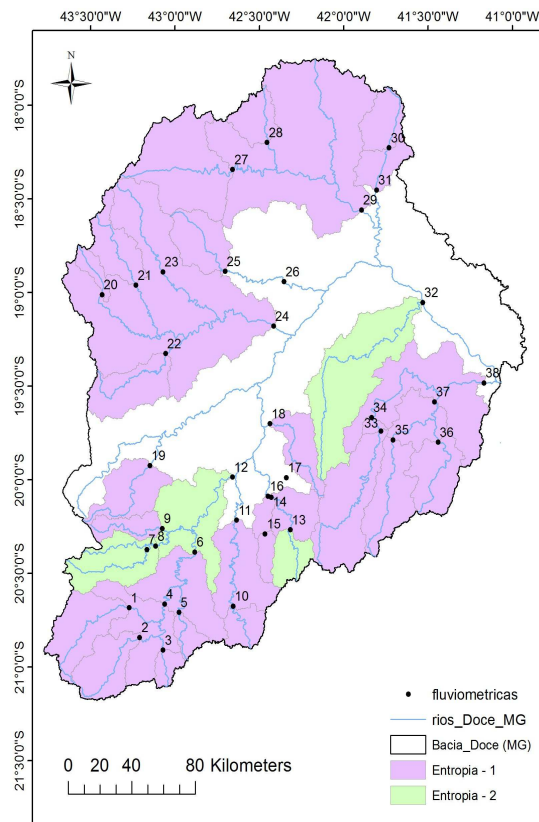


Figura 10 Regiões hidrologicamente homogêneas definidas pela análise de entropia - considerando a vazão mínima.

Pela análise de entropia, para fins da regionalização da $Q_{7,10}$, dividiu-se a área em estudo em duas RHH, sendo que as estações 7, 17, 23, e 26 foram excluídas das análises posteriores por apresentarem valores de IMD inferior a 0,64.

4.2 Análise do comportamento estatístico e físico das Q_{mld} regionalizadas

4.2.1 Regionalização da Q_{mld}

Na Tabela 6 são apresentadas as equações de regressão e o coeficiente de determinação obtidos para cada RHH identificada pelos métodos de agrupamento.

Tabela 6 - Equações de regressão da vazão média de longa duração e coeficientes de determinação obtidos para cada região hidrologicamente homogênea definida pelos métodos de agrupamento

Agrupamento	Região hidrologicamente homogênea				
	1	2	3	4	5
Geo 1	$0,946 \text{ Peq}_{750}^{1,002}$ $R^2 = 0,95$	-	-	-	-
Geo 2	$1,161 \text{ Peq}_{750}^{0,948}$ $R^2 = 0,97$	$0,979 \text{ Peq}_{750}^{0,993}$ $R^2 = 0,97$	$1,109 \text{ Peq}_{750}^{0,992}$ $R^2 = 0,91$	$0,439 \text{ Peq}_{750}^{1,146}$ $R^2 = 0,96$	-
Geo 3	$1,161 \text{ Peq}_{750}^{0,948}$ $R^2 = 0,97$	$1,109 \text{ Peq}_{750}^{0,992}$ $R^2 = 0,91$	$0,439 \text{ Peq}_{750}^{1,146}$ $R^2 = 0,96$	-	-
Cluster 1	$1,190 \text{ Peq}_{750}^{0,939}$ $R^2 = 0,94$	$0,988 \text{ Peq}_{750}^{0,995}$ $R^2 = 0,95$	$0,839 \text{ Peq}_{750}^{1,017}$ $R^2 = 0,98$	$0,176 \text{ Peq}_{750}^{1,557}$ $R^2 = 0,85$	-
Cluster 2	$1,442 \text{ Peq}_{750}^{0,890}$ $R^2 = 0,98$	$1,376 \text{ Peq}_{750}^{0,915}$ $R^2 = 0,95$	$0,913 \text{ Peq}_{750}^{0,988}$ $R^2 = 0,96$	$0,184 \text{ Peq}_{750}^{1,557}$ $R^2 = 0,96$	$0,964 \text{ Peq}_{750}^{0,997}$ $R^2 = 0,87$
Cluster 3	$1,057 \text{ Peq}_{750}^{0,972}$ $R^2 = 0,97$	$0,592 * \text{Peq}_{750}^{1,136}$ $R^2 = 0,90$	-	-	-
Entropia	$1,437 \text{ Peq}_{750}^{0,958}$ $R^2 = 0,96$	$0,859 \text{ Peq}_{750}^{0,988}$ $R^2 = 0,90$	-	-	-

As metodologias utilizadas na identificação das regiões hidrológicas geraram modelos de regressão com coeficientes de determinação ajustados maiores que 0,85. Considerando o período de dados fluviométricos e pluviométricos utilizados, a ETR de uma bacia pode ser mensurada pela equação 26 e o deflúvio médio (D) pela equação 27.

$$ETR = P - D , \quad (26)$$

$$D = P - ETR , \quad (27)$$

em que: ETR = evapotranspiração real (mm); P = precipitação total na área considerada (mm); e D = deflúvio médio da área considerada (mm).

No caso de bacias que possuem uma ETR próxima a 750 mm, a Peq_{750} pode apresentar valores bastante semelhantes à Q_{mld} (equação 28), podendo ocorrer um caso de correlação espúria.

$$Peq_{750} = \frac{(P - 750)A}{K} = \frac{DA}{K} \simeq Q_{mld} , \quad (28)$$

em que: P = precipitação total na área considerada (mm); e D = deflúvio médio da área considerada (mm).

4.2.2 Análise do comportamento estatístico das Q_{mld} regionalizadas

Os índices estatísticos utilizados para análise do comportamento das Q_{mld} regionalizadas são apresentados na Tabela 7.

Tabela 7 - Índices estatísticos obtidos para cada agrupamento

Agrupamento	AE (%)	E_1 (adimensional)	E_2 (adimensional)	E_3 (adimensional)	ER_m (%)	R^2_{cv} (adimensional)
Geo 1	-32 a 84	0,52	0,33	0,32	18,43	0,94
Geo 2	-31 a 31	0,72	0,68	0,41	14,7	0,95
Geo 3	-30 a 33	0,71	0,67	0,41	14,7	0,95
Cluster 1	-30 a 48	0,66	0,60	0,39	15,85	0,95
Cluster 2	-34 a 86	0,67	0,53	0,44	15,48	0,95
Cluster 3	-33 a 58	0,56	0,46	0,34	18,6	0,94
Entropia	-31 a 62	0,52	0,33	0,41	16,34	0,95

Pela análise dos índices estatísticos apresentados na Tabela 7 se evidenciou que estes são discordantes em mostrar qual método de agrupamento produziu melhor ajuste estatístico das Q_{mld} regionalizadas.

O coeficiente de determinação da validação cruzada é praticamente igual para todos os agrupamentos e o ER_m variou de 14,7% a 18,43%, o que pode ser considerado como valores muito próximo entre si. Dessa maneira, a análise do comportamento estatístico foi realizada considerando os demais quatro índices.

Os agrupamentos 'Geo 2' e 'Geo 3' foram os que apresentaram melhor desempenho estatístico no que diz respeito à AE, E_1 e E_2 . Dessa maneira, foram considerados os agrupamentos que obtiveram melhores resultados estatísticos.

O agrupamento 'Geo 1' foi o que apresentou pior desempenho estatístico no que diz respeito à AE, E_1 , E_2 e E_3 , sendo considerado o agrupamento com pior resultado estatístico. Entre os demais agrupamentos, os índices estatísticos não apresentaram convergência, o que dificultou a classificação dos mesmos quanto ao seu desempenho estatístico.

4.2.3 Análise do comportamento físico das Q_{mld} regionalizadas

A distribuição espacial do CE e as RHH dos agrupamentos 'Geo 1', 'Cluster 2', 'Entropia', 'Cluster 1', 'Geo 2', 'Geo 3' e 'Cluster 3' são apresentadas, respectivamente, nas Figuras 11, 12, 13, 14, 15, 16 e 17. Os resultados da análise de segurança são apresentados no Apêndice D.

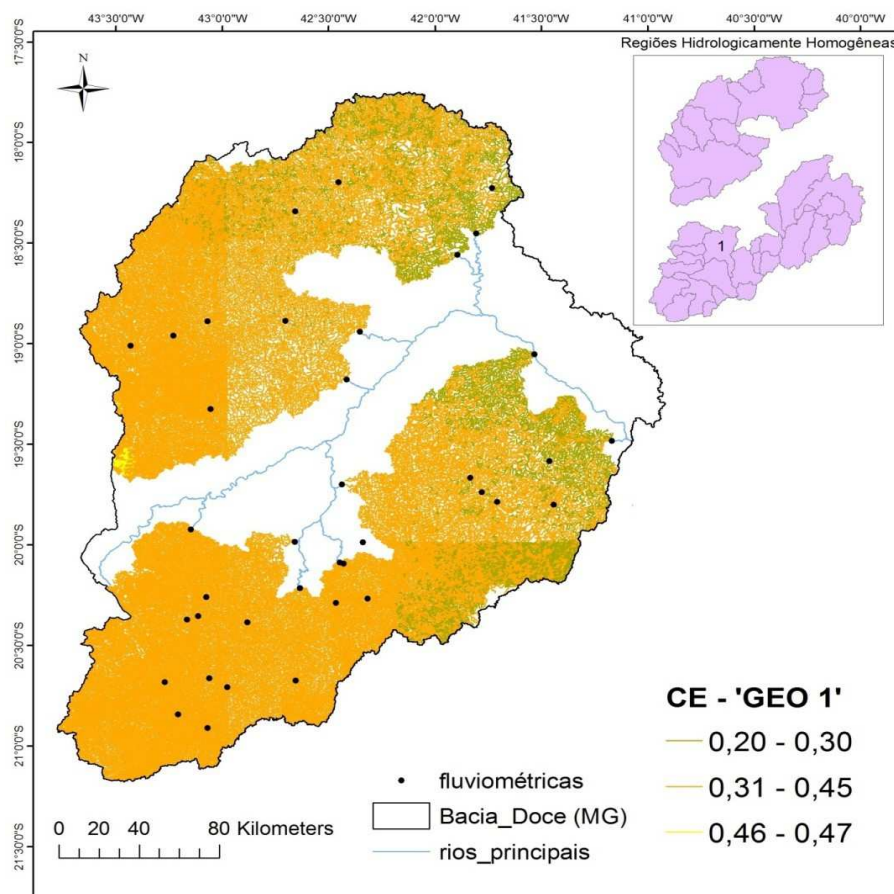


Figura 11 - Distribuição espacial do CE e RHH do agrupamento 'Geo 1'.

No caso do agrupamento 'Geo 1' (Figura 10) se evidenciou que a amplitude dos valores de CE estimados varia de 0,20 a 0,47, enquanto que os valores de CE observados nas estações fluviométricas variam de 0,17 a 0,60, ou seja, a consideração da área em estudo como uma RHH acarretou em uma redução da amplitude dos valores de CE estimados quando comparados com os valores observados, fazendo com que, conseqüentemente, todos os segmentos da hidrografia fossem classificados na faixa aceitável/provável na análise de segurança. Tal comportamento é inconsistente sob o

ponto de vista físico, uma vez que a amplitude dos valores de CE estimados, mesmo considerando regiões de extrapolação da equação de regionalização, é bem inferior à dos valores observados. Com base nesse comportamento se considerou o agrupamento 'Geo 1' não satisfatório para a regionalização da Q_{mld} .

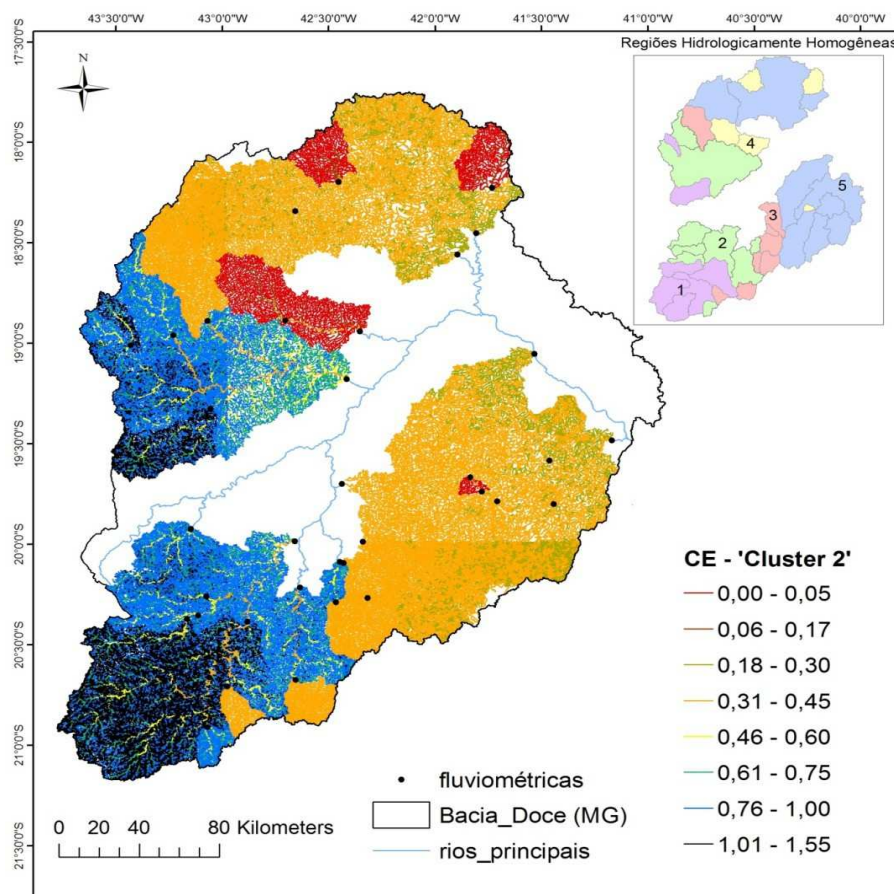


Figura 12 - Distribuição espacial do CE e RHH do agrupamento 'Cluster 2'.

Pela análise da distribuição espacial dos CE estimados relativos ao agrupamento 'Cluster 2' (Figura 11) se evidenciou que em 17,8% da hidrografia os valores de CE são superiores a 1, comportamento considerado inaceitável sob o ponto de vista físico e motivo para que o agrupamento também fosse considerado não satisfatório para a regionalização da Q_{mld} .

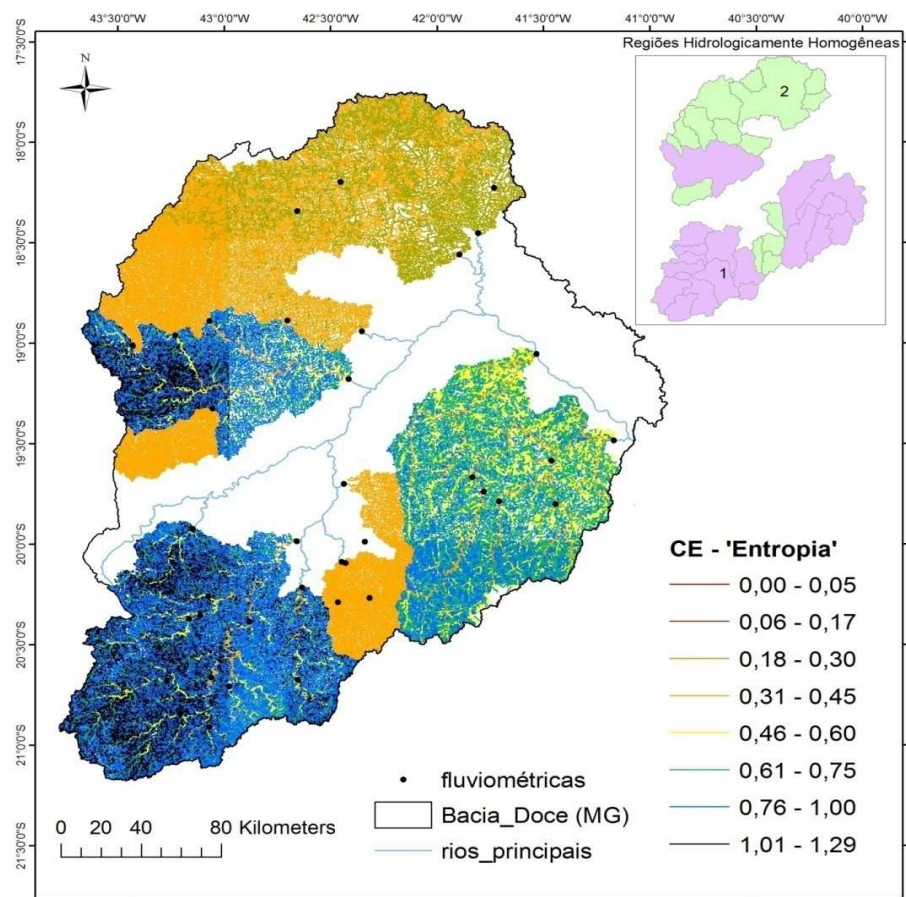


Figura 13 - Distribuição espacial do CE e RHH do agrupamento 'Entropia'.

Comportamento similar à 'Cluster 2' foi evidenciado para o agrupamento 'Entropia' (Figura 12), no qual 18,3% da hidrografia apresentou valores de CE superiores a 1, motivo pelo qual os resultados obtidos para esse agrupamento também fossem considerados não satisfatórios.

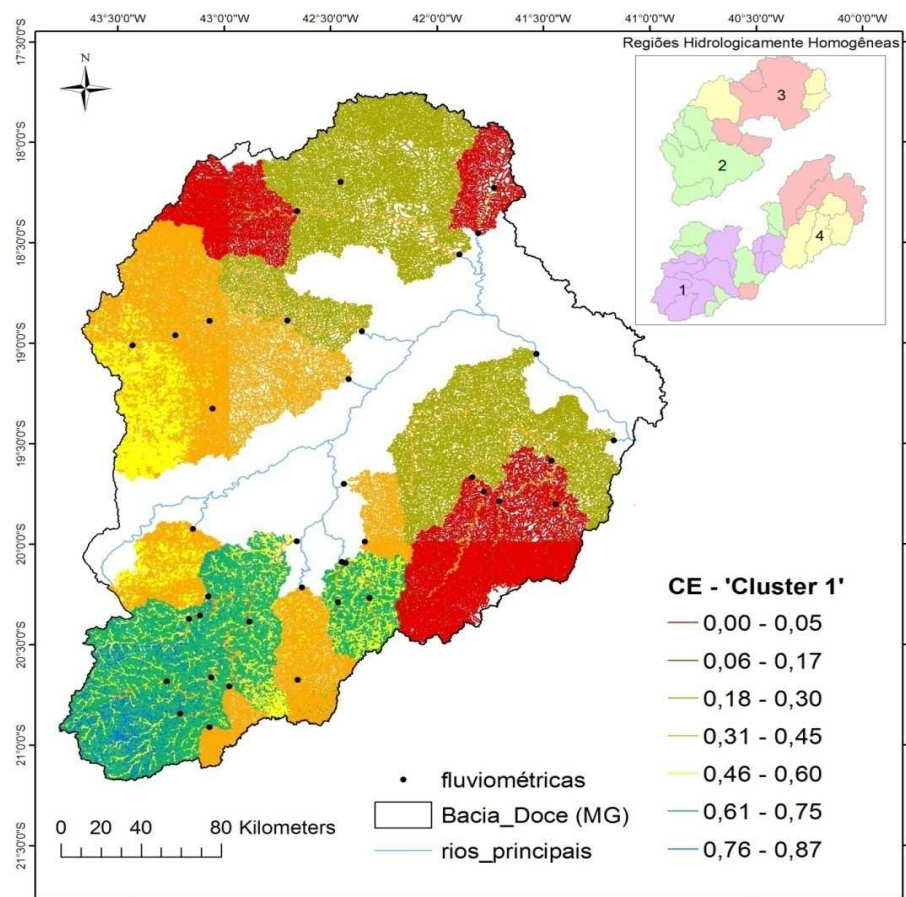


Figura 14 - Distribuição espacial do CE e RHH do agrupamento 'Cluster 1'.

A regionalização da Q_{mld} considerando o agrupamento 'Cluster 1' (Figura 13) apresentou, para a RHH 4, 76% da hidrografia com valores de CE estimados inferiores a 0,01, enquanto o menor valor de CE observado nessa região é de 0,17. Esse comportamento foi observado em segmentos da hidrografia com pequena área de drenagem, enquanto que naqueles associados a maiores áreas de drenagem, os CE estimados alcançaram valores na faixa de 0,31 a 0,45, próximos aos CE observados nas estações fluviométricas. Tal comportamento é fisicamente contestável visto que cursos d'água com valores de CE estimados em torno de 0,01 dão origem a rios com valores de CE estimados maiores que 0,31. Uma possível justificativa física para ocorrência desse comportamento seria o aumento da precipitação no sentido de jusante, contudo o mapa de precipitação mostra uma tendência contrária a tal fato, ou seja, a precipitação diminui no sentido de jusante. Com base nesse comportamento considerou-se o agrupamento 'Cluster 1' não satisfatório para a regionalização da Q_{mld} .

Os agrupamentos 'Geo 2', 'Geo 3' e 'Cluster 3', ao contrário dos agrupamentos 'Geo 1', 'Cluster 1', 'Cluster 2' e 'Entropia', foram considerados aceitáveis do ponto de

físico. Contudo, para alguns trechos da hidrografia, foram identificados comportamentos físicos contestáveis, os quais são descritos na sequência.

A distribuição espacial do CE para o agrupamento 'Geo 2' (Figura 14) apresentou valores de CE variando de 0,03 a 0,78, havendo um aumento da amplitude dos valores de CE estimados quando comparado a faixa de variação dos valores observados nas estações fluviométricas (0,17 a 0,60), fazendo com que 23,2% da hidrografia fosse enquadrada na classe de potencial superestimativa e 13,4% da hidrografia classificada na zona de potencial subestimativa.

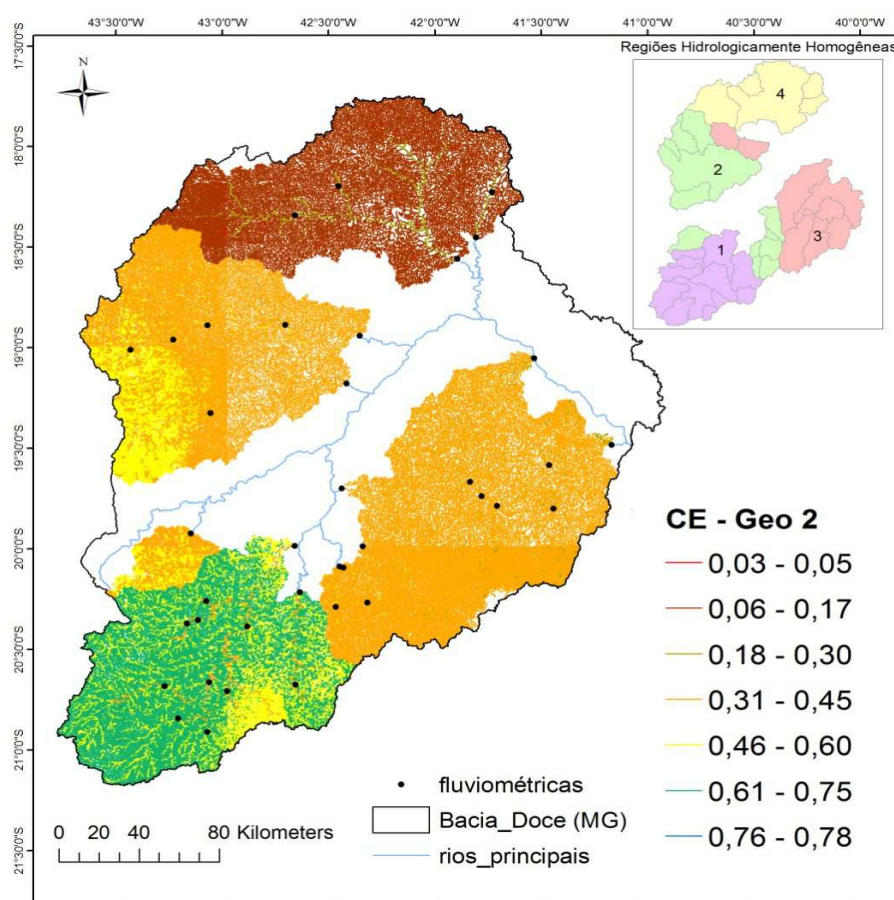


Figura 15 - Distribuição espacial do CE e RHH do agrupamento 'Geo 2'.

A amplitude dos CE estimados ao longo da hidrografia e observados nas estações fluviométricas, assim como a média dos valores de CE estimados nos trechos da hidrografia e a precipitação total (PT), em cada RHH do agrupamento 'Geo 2', são apresentados na Tabela 8.

Tabela 8 - Amplitude dos CE, estimados ao longo da hidrografia e observados nas estações fluviométricas, média dos valores de CE estimados nos trechos da hidrografia e PT nas RHH do agrupamento 'Geo 2'

RHH	Amplitude do CE		Média dos valores estimados de CE	PT (mm)
	Observado	Estimado		
1	0,31 a 0,60	0,38 a 0,78	0,62	1317
2	0,31 a 0,6	0,32 a 0,49	0,43	1288
3	0,28 a 0,5	0,26 a 0,52	0,38	1125
4	0,17 a 0,31	0,03 a 0,29	0,08	1165

A regionalização da Q_{mld} a partir do agrupamento 'Geo 2' gerou: a) 63% dos trechos da hidrografia com valores de CE superiores ao máximo observado nas estações fluviométricas (RHH 1); (b) média dos valores de CE estimados ao longo da hidrografia de 0,62, enquanto que o máximo valor observado nas estações fluviométricas foi de 0,60 (RHH 1). (c) valor mínimo dos CE estimados ao longo da hidrografia superior ao mínimo observado nas estações fluviométricas (RHH 1 e 2). Esses comportamentos podem gerar falsas expectativas quanto à disponibilidade hídrica nas respectivas regiões.

A regionalização da Q_{mld} a partir do agrupamento 'Geo 2' apresentou algumas inconsistências entre o mapa de distribuição do CE e o mapa de isoietas, sendo que embora as precipitações nas RHH 1 e RHH 2 sejam semelhantes, a diferença dos valores médios de CE estimados ao longo da hidrografia é expressiva (Tabela 8). Comportamento semelhante foi observado para a RHH 3 e a RHH 4.

A RHH 4 do agrupamento 'Geo 2' apresentou em segmentos da hidrografia à montante CE estimados menores que em trechos da hidrografia à jusante. Sabe-se que esse comportamento é fisicamente contestável, mas, diferentemente do agrupamento 'Cluster 2', o valor mínimo de CE estimado foi 0,03 e a média dos valores estimados nos trechos da hidrografia foi 0,08.

A regionalização da Q_{mld} a partir do agrupamento 'Geo 3' (Figura 15) apresentou valores de CE variando de 0,03 a 0,63. A porcentagem da hidrografia enquadrada na zona de potencial subestimativa foi de 13,17% e na zona aceitável/provável foi de 86,63%. As RHH 2 e 3 de 'Geo 3' são idênticas, respectivamente, às RHH 3 e 4 de 'Geo 2', portanto as considerações são as mesmas relatadas anteriormente.

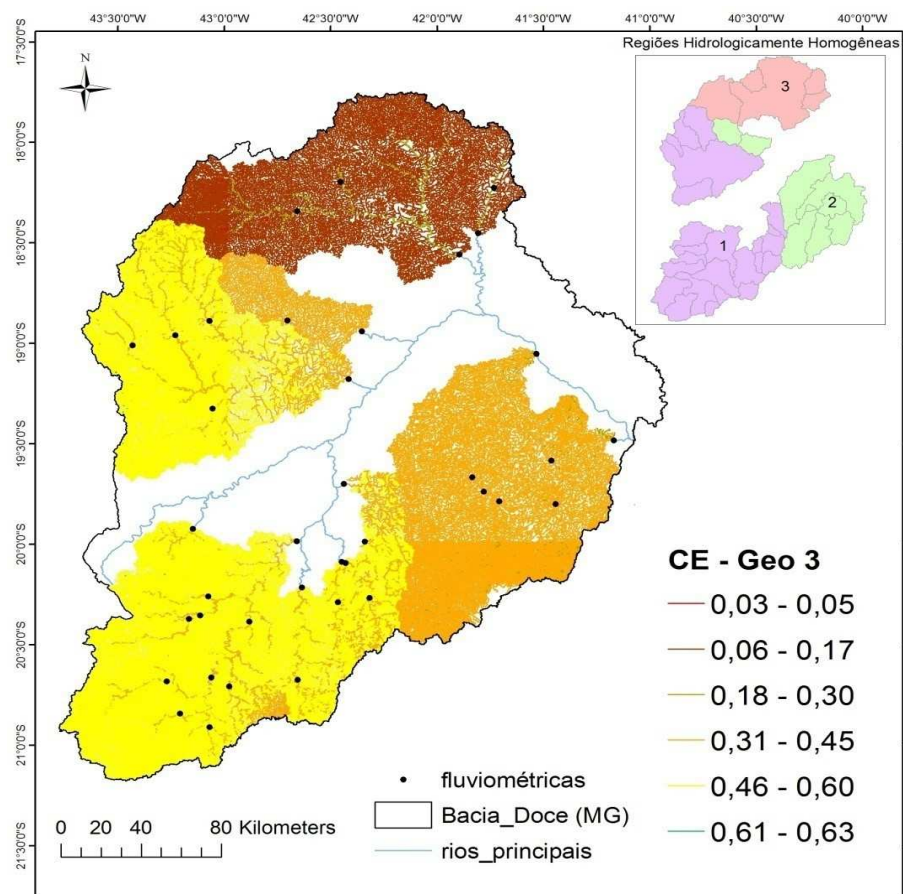


Figura 16 - Distribuição espacial do CE e RHH do agrupamento 'Geo 3'.

A amplitude dos CE estimados ao longo da hidrografia e observados nas estações fluviométricas, assim como a média dos valores de CE estimados nos trechos da hidrografia e a precipitação total (PT), em cada RHH do agrupamento 'Geo 3', são apresentados na Tabela 9.

Tabela 9 - Amplitude dos CE, estimados ao longo da hidrografia e observados nas estações fluviométricas, média dos valores de CE, estimados nos trechos da hidrografia, e PT nas RHH do agrupamento 'Geo 2'

RHH	Amplitude do CE (adimensional)		Média dos valores estimados de CE (adimensional)	PT (mm)
	Observado	Estimado		
1	0,31 a 0,6	0,37 a 0,63	0,50	1302
2	0,28 a 0,5	0,26 a 0,52	0,38	1125
3	0,17 a 0,31	0,03 a 0,29	0,08	1165

A regionalização da Q_{mld} na RHH 1 apresentou valor mínimo dos CE estimados ao longo da hidrografia superior ao mínimo observado nas estações fluviométricas, o

que leva a crer que tal modelo superestima a Q_{mld} em trechos da hidrografia e, conseqüentemente, podem gerar falsas expectativas de disponibilidade hídrica.

A regionalização da Q_{mld} a partir do agrupamento 'Cluster 3' (Figura 16) apresentou valores de CE variando de 0,04 a 0,63, semelhante a 'Geo 3'. A porcentagem da hidrografia enquadrada na zona de potencial subestimativa foi de 30%, e na zona aceitável/provável foi de 70%. A RHH 1 do agrupamento 'Cluster 3' é idêntica à RHH 1 do agrupamento 'Geo 3', portanto as observações são as mesmas relatadas anteriormente.

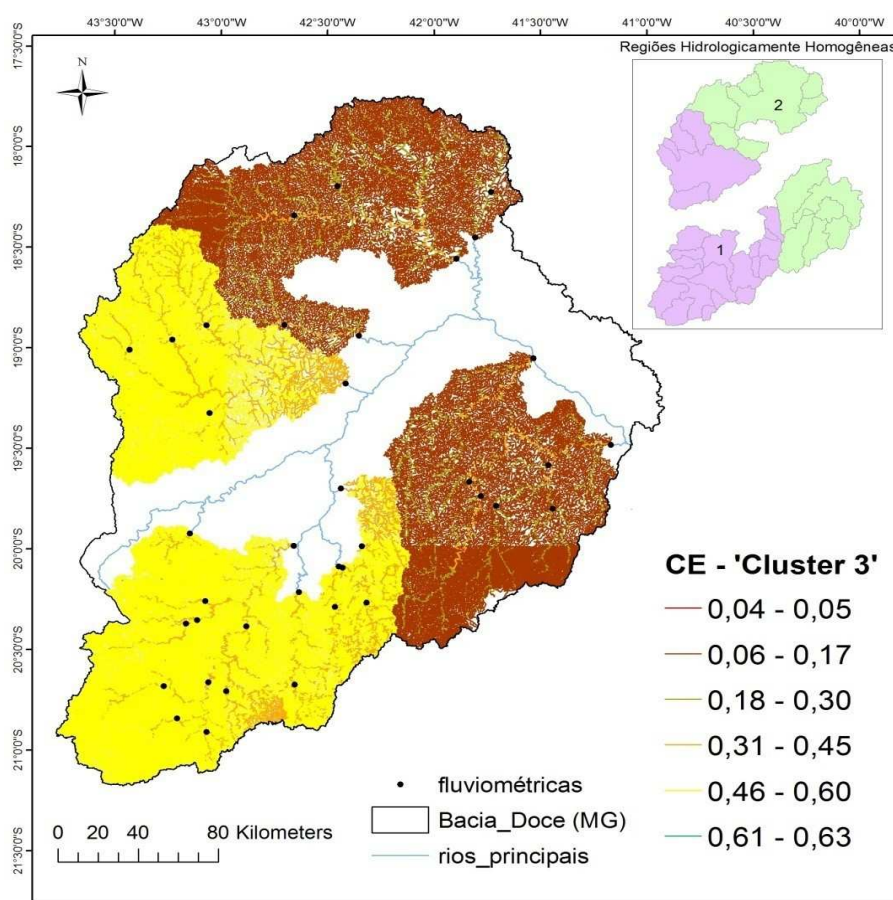


Figura 17 - Distribuição espacial do CE e RHH do agrupamento 'Cluster 3'.

A amplitude dos CE estimados ao longo da hidrografia e observados nas estações fluviométricas, assim como a média dos valores de CE estimados nos trechos da hidrografia e a precipitação total (PT), em cada RHH do agrupamento 'Cluster 3', são apresentados na Tabela 10.

Tabela 10 - Amplitude dos CE, estimados ao longo da hidrografia e observados nas estações fluviométricas, média dos valores de CE, estimados nos trechos da hidrografia, e PT nas RHH do agrupamento 'Cluster 3'

RHH	Amplitude do CE (adimensional)		Média dos valores estimados de CE (adimensional)	PT (mm)
	Observado	Estimado		
1	0,31 a 0,6	0,37 a 0,63	0,50	1302
2	0,17 a 0,5	0,04 a 0,37	0,12	1145

A RHH 2 do agrupamento 'Cluster 3' apresentou em segmentos da hidrografia à montante CE estimados menores que em trechos da hidrografia à jusante. Sabe-se que esse comportamento é fisicamente contestável, mas, diferentemente do agrupamento 'Cluster 2', o valor mínimo de CE estimado foi 0,04.

As regionalizações realizadas considerando os agrupamentos 'Geo 2', 'Geo 3' e 'Cluster 3' foram consideradas aceitáveis do ponto de vista físico, apesar dos comportamentos contestáveis discutidos. Dessa maneira, realizou-se uma análise comparativa entre os agrupamentos com intuito de identificar em qual deles os comportamentos estatístico e físico apresentam menor risco à regionalização da Q_{mld} .

Pela análise comparativa entre os agrupamentos 'Geo 2' e 'Geo 3' se evidenciou que os resultados obtidos na regionalização da Q_{mld} considerando o agrupamento 'Geo 3' apresentaram melhor desempenho nos seguintes aspectos: a) maior semelhança entre o mapa de isoietas e o comportamento do CE espacializado; (b) menores valores dos CE estimados na zona sul da área em estudo, fazendo com que os valores classificados na zona de superestimativa passassem para faixa considerada provável/aceitável, reduzindo o risco associado à regionalização. No que diz respeito aos resultados obtidos na regionalização da Q_{mld} considerando o agrupamento 'Geo 2', observou-se que o valor mínimo de CE estimado na RHH 2 (0,32) foi superior ao observado nas estações fluviométricas (0,31), contudo menor que o valor estimado quando considerado 'Geo 3' (0,37), fazendo com que, para essa região, o agrupamento 'Geo 2' apresentasse menor risco de superestimativa da Q_{mld} . Dessa maneira, não se pode determinar, de uma forma conclusiva, qual dos dois agrupamentos representa menor risco de superestimativas na regionalização da Q_{mld} .

Pela análise comparativa entre os agrupamentos 'Geo 2', 'Geo 3' e 'Cluster 3' se evidenciou que os resultados obtidos na regionalização da Q_{mld} , considerando os

agrupamentos 'Geo 2' e 'Geo 3', apresentaram melhor desempenho estatístico. No que diz respeito ao agrupamento 'Cluster 3', este apresentou desempenho estatístico entre os piores, contudo o comportamento físico foi o mais consistente.

A análise comparativa entre os agrupamentos 'Geo 2', 'Geo 3' e 'Cluster 3' não evidenciou qual dos agrupamentos representa menor risco à regionalização da Q_{mld} , pois, enquanto que os agrupamentos 'Geo 2' e 'Geo 3' apresentaram desempenho estatístico superior, o agrupamento 'Cluster 3' demonstrou-se mais satisfatório quando considerado o comportamento físico da vazão média regionalizada.

4.3 Análise do comportamento estatístico e físico das $Q_{7,10}$ regionalizadas

4.3.1 Regionalização da $Q_{7,10}$

Na Tabela 11 são apresentadas as equações de regressão e o coeficiente de determinação obtidos para cada região hidrologicamente homogênea obtidas por cada método de agrupamento.

Tabela 11 - Equações de regressão da $Q_{7,10}$ e coeficiente de determinação obtidos para cada região hidrologicamente homogênea definida pelos métodos de agrupamento

Agrupamento	Região hidrologicamente homogênea				
	1	2	3	4	5
Geo 1	$0,174 \text{ Peq}_{750}^{1,057}$ $R^2 = 0,7$	x	x	x	x
Geo 2	$0,415 \text{ Peq}_{750}^{0,884}$ $R^2 = 0,92$	$0,208 \text{ Peq}_{750}^{0,935}$ $R^2 = 0,93$	$0,217 \text{ Peq}_{750}^{1,007}$ $R^2 = 0,97$	$0,021 \text{ Peq}_{750}^{1,457}$ $R^2 = 0,74$	x
Geo 3	$0,26 \text{ Peq}_{750}^{0,95}$ $R^2 = 0,88$	$0,217 \text{ Peq}_{750}^{1,007}$ $R^2 = 0,97$	$0,021 \text{ Peq}_{750}^{1,457}$ $R^2 = 0,74$	x	x
Cluster 1	$0,189 \text{ Peq}_{750}^{1,052}$ $R^2 = 0,88$	$0,321 \text{ Peq}_{750}^{0,856}$ $R^2 = 0,84$	$0,199 \text{ Peq}_{750}^{0,976}$ $R^2 = 0,92$	$0,0021 \text{ Peq}_{750}^{2,397}$ $R^2 = 0,72$	x
Cluster 2	$0,189 \text{ Peq}_{750}^{1,058}$ $R^2 = 0,97$	$0,368 \text{ Peq}_{750}^{0,878}$ $R^2 = 0,77$	$0,414 \text{ Peq}_{750}^{0,663}$ $R^2 = 0,85$	$0,006 \text{ Peq}_{750}^{2,162}$ $R^2 = 0,75$	$0,098 \text{ Peq}_{750}^{1,159}$ $R^2 = 0,62$
Cluster 3	$0,258 \text{ Peq}_{750}^{0,952}$ $R^2 = 0,88$	$0,047 \text{ Peq}_{750}^{1,381}$ $R^2 = 0,7$	x	x	x
Entropia	$0,173 \text{ Peq}_{750}^{1,018}$ $R^2 = 0,74$	$0,181 \text{ Peq}_{750}^{1,087}$ $R^2 = 0,82$	x	x	x

O coeficiente de determinação para a RHH 5 do agrupamento 'Cluster 2' foi 0,62. Elesblon (2012) considerou como 0,70 o valor mínimo do coeficiente de determinação para que a equação de regressão fosse considerada satisfatória, entretanto, para o presente estudo, a equação de regressão relativa a RHH 5 do agrupamento 'Cluster 2' foi considerada satisfatória.

4.3.2 Análise do comportamento estatístico das $Q_{7,10}$ regionalizadas

Os resultados dos índices de eficiência utilizados para análise estatística das $Q_{7,10}$ regionalizadas são apresentados na Tabela 12.

Tabela 12 - Índices estatísticos obtidos considerando cada agrupamento

Agrupamento	AE (%)	E_1 (adimensional)	E_2 (adimensional)	E_3 (adimensional)	ER_m (%)	R^2_{cv} (adimensional)
Geo 1	-64 a 523	0,27	-4,64	0,17	52,1	0,76
Geo 2	-57 a 102	0,71	0,52	0,42	26,9	0,82
Geo 3	-57 a 103	0,60	0,32	0,24	33,7	0,80
Cluster 1	-64 a 150	0,57	0,26	0,25	33,7	0,80
Cluster 2	-50 a 302	0,54	-0,59	0,34	35,6	0,84
Cluster 3	-55 a 273	0,42	-0,96	0,17	44,3	0,75
Entropia	-61 a 581	0,30	-3,69	0,19	53,8	0,76

Pela análise dos índices estatísticos apresentados na Tabela 12 se evidenciou que estes são discordantes em mostrar qual método de agrupamento produziu melhor ajuste estatístico das Q_{mld} regionalizadas.

A eficiência de E_2 foi negativa para 4 dos 7 agrupamentos testados, indicando que a média das vazões observadas seria melhor preditor das vazões estimadas do que o próprio modelo. Contudo, E_1 e E_3 não apresentaram índices negativos, contrapondo E_2 .

Pela análise dos índices estatísticos apresentados na Tabela 12 se evidenciou que o agrupamento 'Geo 2' apresentou melhor desempenho estatístico em cinco dos seis índices utilizados. Dessa maneira, o agrupamento 'Geo 2' apresentou melhor desempenho estatístico dentre todos os agrupamentos.

Para os demais agrupamentos, 'Geo 3' apresentou melhor desempenho estatístico em quatro dos seis índices avaliados. Assim, foi considerado o agrupamento com segundo melhor desempenho estatístico.

Os agrupamentos 'Geo 1' e 'Entropia' apresentaram limite superior do erro relativo maior que 500% e foram considerados os agrupamentos com pior desempenho estatístico. Entre os demais agrupamentos, os índices estatísticos não apresentaram convergência, o que dificultou a classificação dos mesmos quanto ao seu desempenho estatístico.

4.3.3 Análise do comportamento físico das $Q_{7,10}$ regionalizadas

A distribuição espacial da $q_{7,10}$ e as RHH dos agrupamentos 'Geo 1', 'Entropia', 'Cluster 1', 'Cluster 2', 'Geo 2', 'Geo 3' e 'Cluster 3' são apresentadas, respectivamente, nas Figuras 18, 19, 20, 21, 22, 23 e 24.

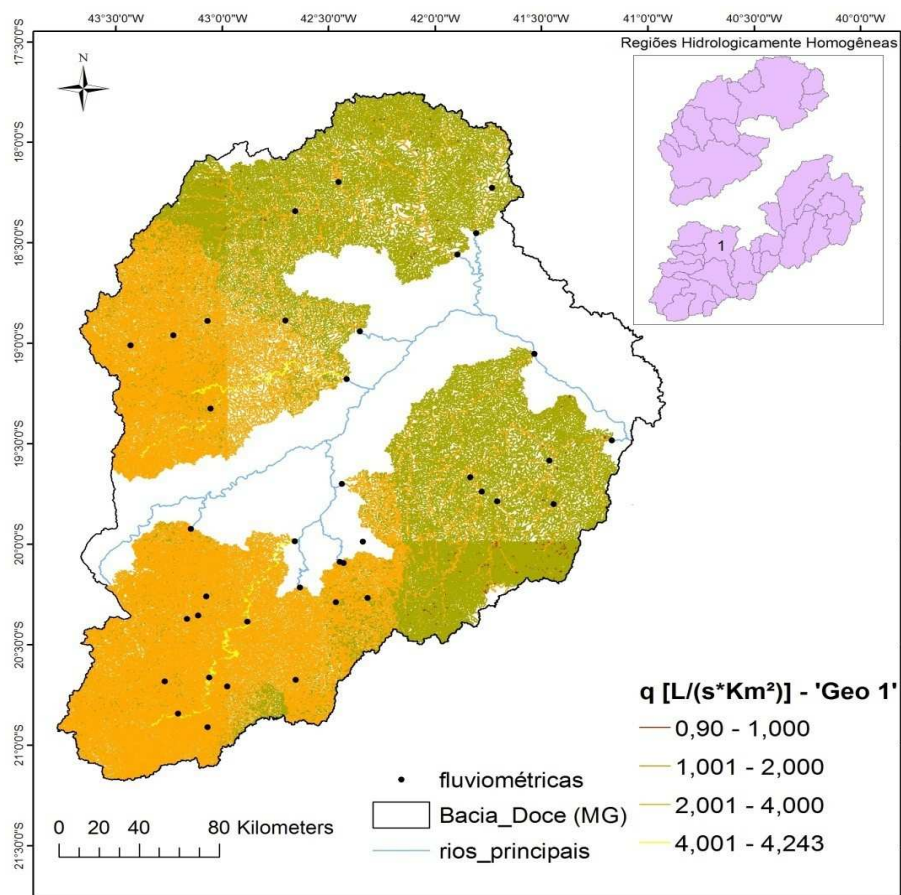


Figura 18 - Distribuição espacial da $q_{7,10}$ e RHH do agrupamento 'Geo 1'.

No caso do agrupamento 'Geo 1' (Figura 17) se evidenciou que a amplitude dos valores de $q_{7,10}$ estimados varia de 0,90 a 4,24 $L s^{-1} km^{-2}$ enquanto que os valores de $q_{7,10}$ observados nas estações fluviométricas variam de 0,37 a 9,15 $L s^{-1} km^{-2}$, ou seja, a consideração da área em estudo como uma região hidrologicamente homogênea acarretou em uma redução da amplitude dos valores de $q_{7,10}$ estimados quando comparados com os valores observados, fazendo com que, conseqüentemente, todos os segmentos da hidrografia fossem classificados na faixa aceitável/provável na análise de segurança. Tal comportamento é inconsistente sob o ponto de vista físico, uma vez que a amplitude dos valores de $q_{7,10}$ estimados, mesmo considerando regiões de extrapolação da equação de regionalização, é bem inferior a dos valores observados. Com base nesse

comportamento se considerou o agrupamento 'Geo 1' não satisfatório para a regionalização da $Q_{7,10}$.

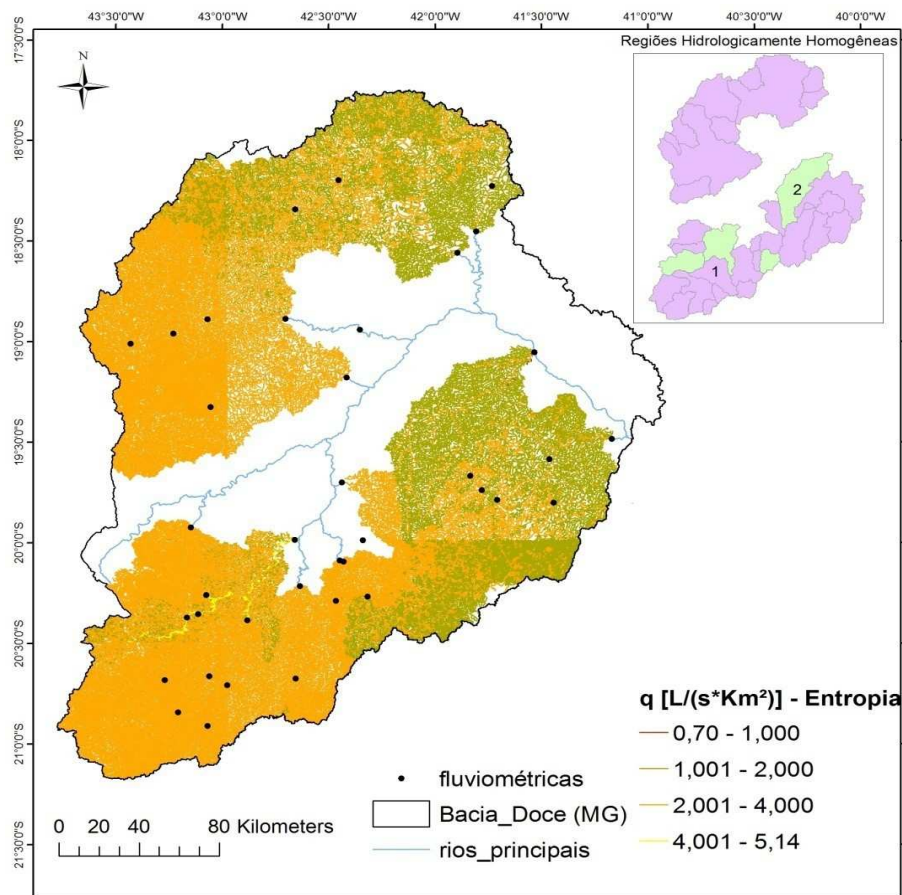


Figura 19 - Distribuição espacial da $q_{7,10}$ e RHH do agrupamento 'Entropia'.

Comportamento similar a 'Geo 1' foi evidenciado para o agrupamento 'Entropia' no qual os valores de $q_{7,10}$ estimados variaram de 0,70 a 5,14 $L s^{-1} km^{-2}$. Dessa maneira, se considerou também o agrupamento 'Entropia' não satisfatório para a regionalização vazão mínima.

A análise de segurança para o agrupamento 'Cluster 1' evidenciou que 24% da hidrografia foi classificada na zona de potencial superestimativa, sendo que o maior valor de $q_{7,10}$ estimado foi 2 vezes maior que o máximo valor observado nas estações fluviométricas.

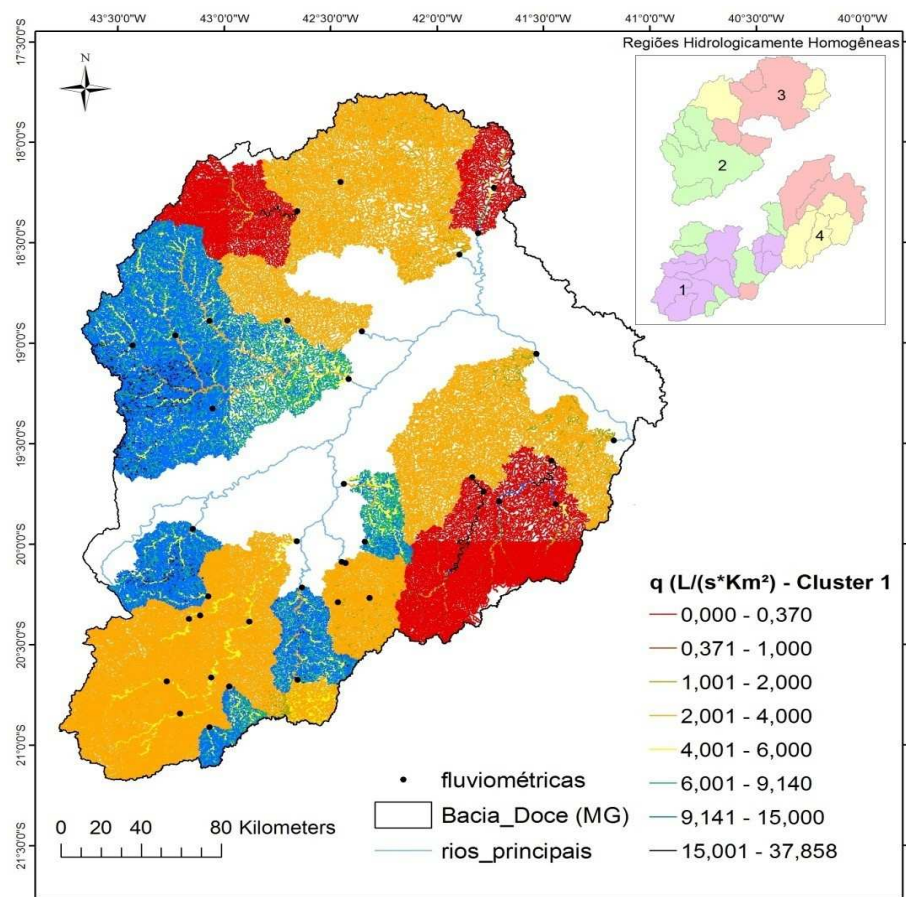


Figura 20 - Distribuição espacial da $q_{7,10}$ e RHH do agrupamento 'Cluster 1'.

A regionalização da vazão mínima considerando o agrupamento 'Cluster 1' (Figura 19) apresentou, para a RHH 4, 92% da hidrografia com valores de $q_{7,10}$ estimados inferiores a $0,01 \text{ L s}^{-1} \text{ km}^{-2}$, enquanto o menor valor de $q_{7,10}$ observado nessa região é de $0,37 \text{ L s}^{-1} \text{ km}^{-2}$. Esse comportamento foi observado nos trechos da hidrografia com pequena área de drenagem, enquanto que em trechos associados a maiores áreas de drenagem, as $q_{7,10}$ estimados alcançaram valores na faixa de 2 a $4 \text{ L s}^{-1} \text{ km}^{-2}$, próximos aos valores de $q_{7,10}$ observados nas estações fluviométricas. Tal comportamento é fisicamente contestável visto que cursos d'água com valores de $q_{7,10}$ estimados em torno de $0,01 \text{ L s}^{-1} \text{ km}^{-2}$ dão origem a rios com valores de $q_{7,10}$ estimados maiores que $2 \text{ L s}^{-1} \text{ km}^{-2}$. Uma justificativa física para ocorrência desse comportamento seria o aumento da precipitação no sentido de jusante, contudo o mapa de precipitação mostra uma tendência contrária a tal fato, ou seja, a precipitação diminui no sentido de jusante. Com base nesses comportamentos se considerou o agrupamento 'Cluster 1' não satisfatório para a regionalização da $Q_{7,10}$.

O agrupamento 'Cluster 2' apresentou 29,2% da hidrografia classificada na zona de potencial subestimativa e 34,4% da hidrografia enquadrada na zona de potencial superestimativa. O maior valor de $q_{7,10}$ estimado foi $85,9 \text{ L s}^{-1} \text{ km}^{-2}$, quase 10 vezes maior que o máximo valor observado dentre todas as estações fluviométricas ($9,15 \text{ L s}^{-1} \text{ km}^{-2}$).

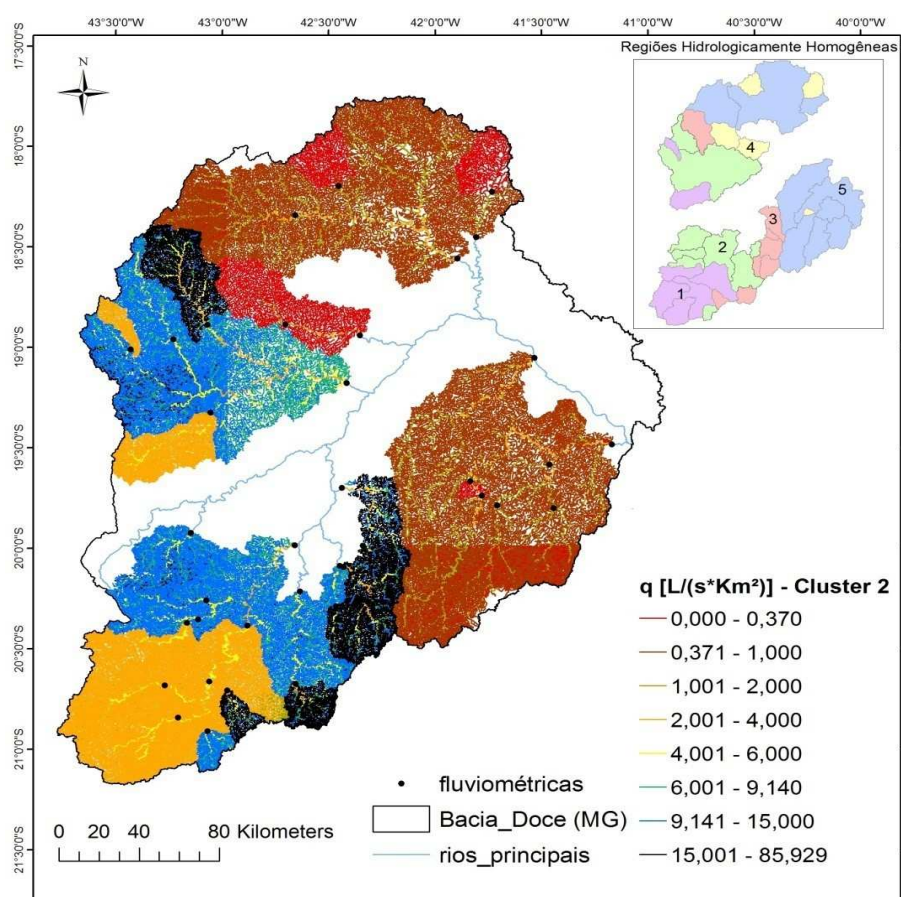


Figura 21 - Distribuição espacial da $q_{7,10}$ e RHH do agrupamento 'Cluster 2'.

A regionalização da vazão mínima considerando o agrupamento 'Cluster 2' (Figura 20) apresentou, para a RHH 4, 50% da hidrografia com valores de $q_{7,10}$ estimados inferiores à $0,01 \text{ L s}^{-1} \text{ km}^{-2}$. Assim como 'Cluster 1' esses valores foram evidenciados em trechos da hidrografia com menor área de drenagem, enquanto que em trechos associado a maior área de drenagem os valores estimados alcançam $3,7 \text{ L s}^{-1} \text{ km}^{-2}$. Com base nos comportamentos evidenciados, considerou-se a regionalização da $Q_{7,10}$ a partir do agrupamento 'Cluster 2' não satisfatório.

Os agrupamentos 'Geo 2', 'Geo 3' e 'Cluster 3', ao contrário dos agrupamentos 'Geo 1', 'Cluster 1', 'Cluster 2' e 'Entropia', foram considerados aceitáveis do ponto de físico. Contudo, para alguns trechos da hidrografia, foram identificados comportamentos físicos contestáveis, os quais são descritos na sequência.

A distribuição espacial da $q_{7,10}$ para o agrupamento 'Geo 2' (Figura 21) apresentou valores de $q_{7,10}$ variando de 0,0026 a 20,9 $L s^{-1} km^{-2}$, havendo um aumento da amplitude dos valores de $q_{7,10}$ estimados quando comparado a faixa de variação dos valores observados nas estações fluviométricas (0,37 a 9,15 $L s^{-1} km^{-2}$), fazendo com que 40% da hidrografia fosse enquadrada na classe de potencial superestimativa e 13,4% da hidrografia classificada na zona de potencial subestimativa (Apêndice X).

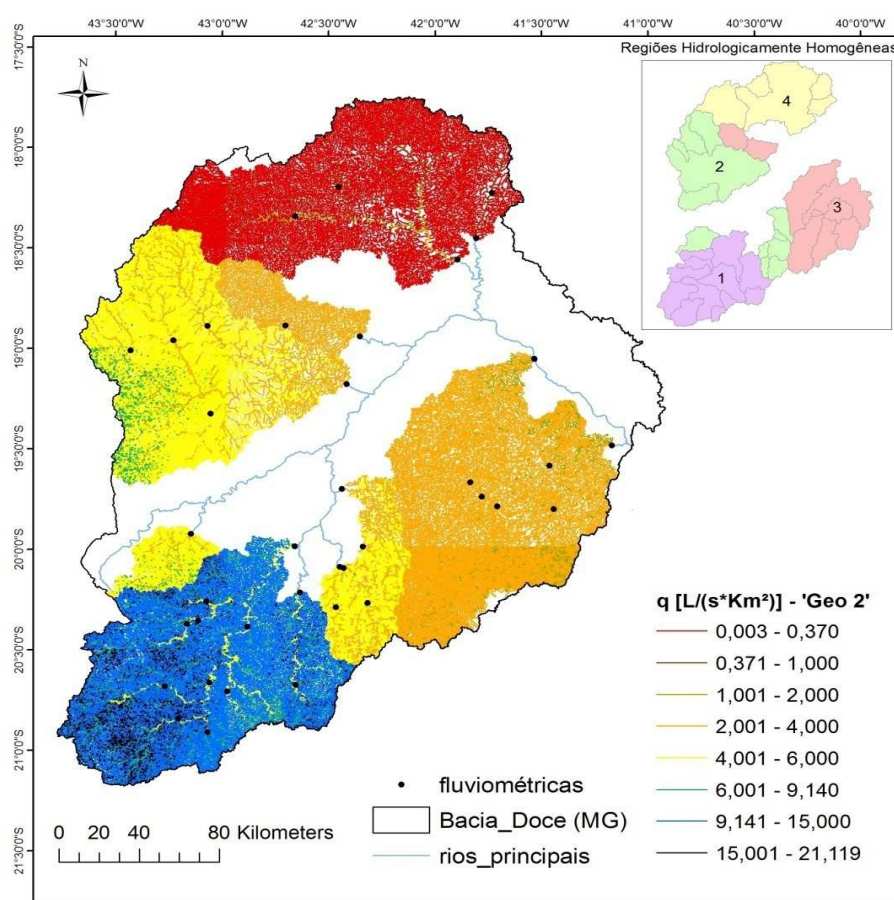


Figura 22 - Distribuição espacial da $q_{7,10}$ e RHH do agrupamento 'Geo 2'.

A amplitude das $q_{7,10}$ estimados ao longo da hidrografia e observados nas estações fluviométricas, assim como a média dos valores de $q_{7,10}$ estimados nos trechos da hidrografia e a PT, em cada RHH do agrupamento 'Geo 2', são apresentados na Tabela 13.

Tabela 13 - Amplitude das $q_{7,10}$, estimadas ao longo da hidrografia e observadas nas estações fluviométricas, média dos valores de $q_{7,10}$, estimados nos trechos da hidrografia, e PT nas RHH do agrupamento 'Geo 2'

RHH	Amplitude do $q_{7,10}$ (L s ⁻¹ km ⁻²)		Média dos valores estimados de $q_{7,10}$ (L s ⁻¹ km ⁻²)	PT (mm)
	Observada	Estimada		
1	2,84 a 9,15	4,10 a 20,9	12,32	1317
2	2,04 a 5,13	2,58 a 7,43	4,77	1288
3	1,81 a 3,78	1,38 a 4,08	2,50	1125
4	0,37 a 2,46	0,0026 a 2,22	0,08	1165

A regionalização da $Q_{7,10}$ a partir do agrupamento 'Geo 2' gerou: a) 85% dos trechos da hidrografia com valores de $q_{7,10}$ superiores ao máximo observado nas estações fluviométricas (RHH 1); (b) média dos valores de $q_{7,10}$ estimados ao longo da hidrografia de 12,32 L s⁻¹ km⁻², enquanto que o máximo valor observado nas estações fluviométricas foi de 9,15 L s⁻¹ km⁻² (RHH 1). (c) valor mínimo dos $q_{7,10}$ estimados ao longo da hidrografia superior ao mínimo observado nas estações fluviométricas (RHH 1 e 2). Esses comportamentos podem gerar falsas expectativas quanto à disponibilidade hídrica nas respectivas regiões.

A regionalização da $Q_{7,10}$ a partir do agrupamento 'Geo 2' apresentou algumas inconsistências entre o mapa de distribuição da $q_{7,10}$ e o mapa de isoietas, sendo que embora a precipitação na RHH 1 e RHH 2 seja semelhante, a diferença dos valores de $q_{7,10}$ estimados é expressiva (Tabela 13). Comportamento semelhante foi observado para a RHH 3 e a RHH 4.

A RHH 4 do agrupamento 'Geo 2' apresentou, em segmentos da hidrografia a montante, valores de $q_{7,10}$ estimados menores que em trechos da hidrografia a jusante. Sabe-se que esse comportamento é fisicamente contestável, mas, diferentemente dos agrupamentos 'Cluster 1' e 'Cluster 2', apenas 1,5% da hidrografia apresentou valores de $q_{7,10}$ estimada menor que 0,01 L s⁻¹ km⁻².

A regionalização da $Q_{7,10}$ a partir do agrupamento 'Geo 3' (Figura 22) apresentou valores variando de 0,003 a 8,16 L s⁻¹ km⁻². A porcentagem da hidrografia classificada zona de potencial subestimativa foi de 13,8%, semelhante a 'Geo 2', e na zona aceitável/provável foi de 86,2%. As RHH 2 e 3 de 'Geo 3' são idênticas,

respectivamente, às RHH 3 e 4 de 'Geo 2', portanto as considerações são as mesmas relatadas anteriormente.

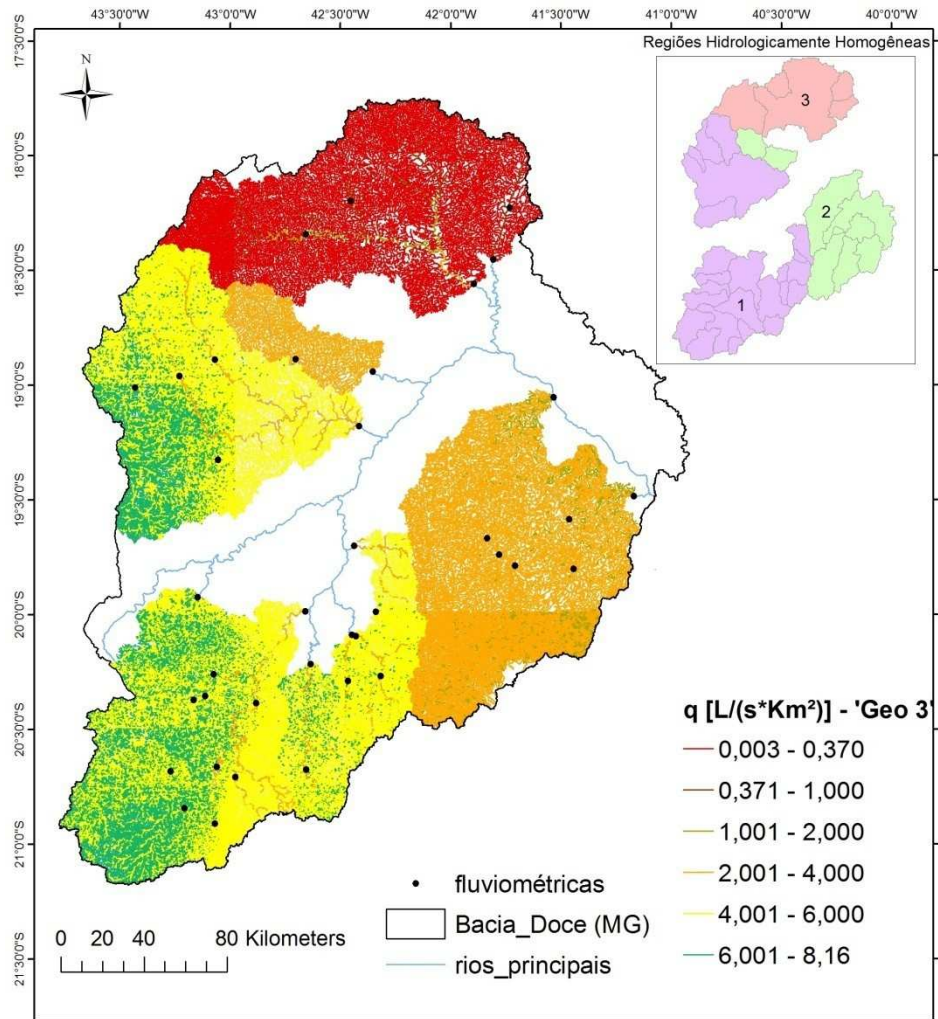


Figura 23 - Distribuição espacial da $q_{7,10}$ e RHH do agrupamento 'Geo 3'.

A amplitude das $q_{7,10}$ estimadas ao longo da hidrografia e observadas nas estações fluviométricas, assim como a média dos valores de $q_{7,10}$ estimados nos trechos da hidrografia e a PT, em cada RHH do agrupamento 'Geo 3', são apresentados na Tabela 14.

Tabela 14 - Amplitude das $q_{7,10}$, estimadas ao longo da hidrografia e observadas nas estações fluviométricas, média dos valores de $q_{7,10}$, estimados nos trechos da hidrografia, e PT nas RHH do agrupamento 'Geo 3'

RHH	Amplitude do $q_{7,10}$ (L s ⁻¹ km ⁻²)		Média dos valores estimados de $q_{7,10}$ (L s ⁻¹ km ⁻²)	PT (mm)
	Observada	Estimada		
1	2,04 a 9,15	3,38 a 8,19	5,60	1302
2	1,81 a 3,78	1,38 a 4,08	2,50	1125
3	0,37 a 2,46	0,003 a 2,22	0,08	1165

A regionalização da $Q_{7,10}$ na RHH 1 de 'Geo 2' apresentou valor mínimo das $q_{7,10}$ estimadas ao longo da hidrografia superior ao mínimo observado nas estações fluviométricas (Tabela 1), o que leva a crer que tal modelo superestima a $Q_{7,10}$ em trechos da hidrografia e, conseqüentemente, podem gerar falsas expectativas de disponibilidade hídrica.

A regionalização da $Q_{7,10}$ a partir do agrupamento 'Cluster 3' apresentou valores de $q_{7,10}$ variando de 0,01 a 8,16 L s⁻¹ km⁻², semelhante a 'Geo 3'. A porcentagem da hidrografia enquadrada na zona de potencial subestimativa foi de 29,9%, e na zona aceitável/provável foi de 70,1%. A RHH 1 do agrupamento 'Cluster 3' é idêntica à RHH 1 do agrupamento 'Geo 3', portanto as observações são as mesmas relatadas anteriormente.

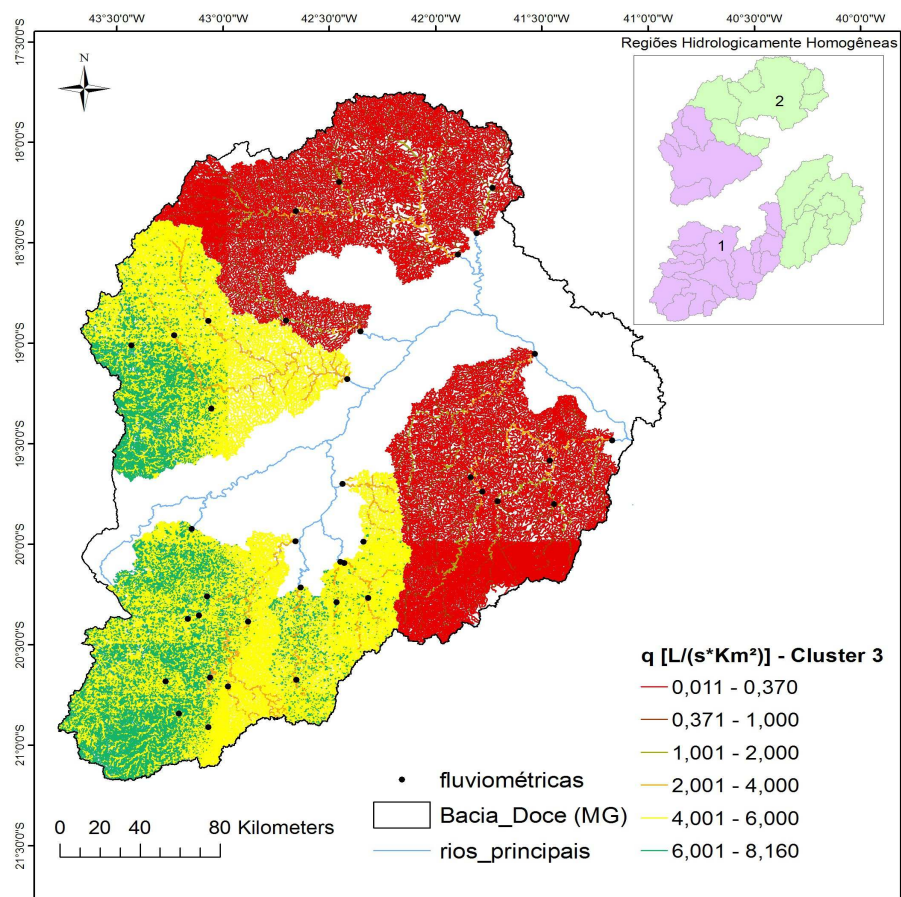


Figura 24 - Distribuição espacial da $q_{7,10}$ e RHH do agrupamento 'Cluster 3'.

A amplitude das $q_{7,10}$ estimadas ao longo da hidrografia e observadas nas estações fluviométricas, assim como a média dos valores de $q_{7,10}$ estimados nos trechos da hidrografia e a PT, em cada RHH do agrupamento 'Geo 3', são apresentados na Tabela 15.

Tabela 15 - Amplitude das $q_{7,10}$, estimadas ao longo da hidrografia e observadas nas estações fluviométricas, média dos valores de $q_{7,10}$, estimados nos trechos da hidrografia, e PT nas RHH do agrupamento 'Cluster 3'

RHH	Amplitude da $q_{7,10}$ ($L s^{-1} km^{-2}$)		Média dos valores estimados de $q_{7,10}$ ($L s^{-1} km^{-2}$)	PT (mm)
	Observado	Estimado		
1	2,04 a 9,15	3,38 a 8,19	5,60	1302
2	0,37 a 3,78	0,01 a 3,43	0,20	1145

A RHH 2 do agrupamento 'Cluster 3' apresentou em segmentos da hidrografia a montante valores de $q_{7,10}$ estimados menores que em trechos da hidrografia a jusante.

Sabe-se que esse comportamento é fisicamente contestável, mas, diferentemente do agrupamento 'Cluster 1' e 'Cluster 2', o valor mínimo de CE estimado foi $0,01 \text{ L s}^{-1} \text{ km}^{-2}$.

A regionalização a partir dos agrupamentos 'Geo 2', 'Geo 3' e 'Cluster 3' foram considerados aceitáveis do ponto de vista físico apesar dos comportamentos contestáveis apresentados anteriormente. Dessa maneira, realizou-se uma análise comparativa entre os agrupamentos com intuito de identificar em qual deles os comportamentos estatístico e físico observados apresentam menor risco à regionalização da $Q_{7,10}$.

Pela análise comparativa entre os agrupamentos 'Geo 2' e 'Geo 3' se evidenciou que os resultados obtidos na regionalização da $Q_{7,10}$ considerando o agrupamento 'Geo 3' apresentaram melhor desempenho nos seguintes aspectos: a) maior semelhança entre o mapa de isoietas e o comportamento do $q_{7,10}$ espacializado; (b) menores valores dos $q_{7,10}$ estimados na zona sul da área em estudo, fazendo com que os valores classificados na zona de superestimativa passassem para faixa considerada provável/aceitável, reduzindo o risco associado à regionalização. No que diz respeito aos resultados obtidos na regionalização da $Q_{7,10}$, considerando o agrupamento 'Geo 2', se evidenciou melhor desempenho nos seguintes aspectos: (a) estatístico; (b) o valor mínimo de $q_{7,10}$ estimado na RHH 2 ($2,84 \text{ L s}^{-1} \text{ km}^{-2}$) foi superior ao observado nas estações fluviométricas ($2,04 \text{ L s}^{-1} \text{ km}^{-2}$), contudo menor que o valor estimado quando considerado 'Geo 3' ($3,44 \text{ L s}^{-1} \text{ km}^{-2}$), assim para essa região o agrupamento 'Geo 2' apresenta menor risco de superestimativas da $Q_{7,10}$. Dessa maneira, não se constata nitidamente qual dos dois agrupamentos representa menor risco de superestimativas na regionalização da $Q_{7,10}$.

Pela análise comparativa entre os agrupamentos 'Geo 2', 'Geo 3' e 'Cluster 3' se evidenciou que os resultados obtidos na regionalização da $Q_{7,10}$, considerando os agrupamentos 'Geo 2' e 'Geo 3', apresentaram melhor desempenho estatístico. No que diz respeito ao agrupamento 'Cluster 3', este apresentou desempenho estatístico entre os piores, contudo o comportamento físico foi o mais consistente, pois houve maior semelhança entre o mapa de isoietas e o comportamento da $q_{7,10}$ espacializada.

A análise comparativa entre os agrupamentos 'Geo 2', 'Geo 3' e 'Cluster 3' não evidenciou qual dos agrupamentos representa menor risco à regionalização da $Q_{7,10}$, pois, enquanto que os agrupamentos 'Geo 2' e 'Geo 3' apresentaram desempenho estatístico superior, o agrupamento 'Cluster 3' demonstrou-se mais satisfatório quando considerado o comportamento físico da vazão mínima regionalizada.

5 CONCLUSÕES

Com base nos resultados obtidos no presente trabalho conclui-se que:

- a consideração das variáveis latitude e longitude na análise de cluster resultou em maior convergência dos índices de validação de agrupamento;
- os agrupamentos 'Geo 2' e 'Geo 3' foram os que apresentaram melhor desempenho estatístico na regionalização da Q_{mid} e da $Q_{7,10}$, enquanto que o agrupamento 'Geo 1' e 'Entropia' foram os que apresentaram os piores desempenhos estatísticos.
- os agrupamentos 'Geo 2', 'Geo 3' e 'Cluster 3' foram os que apresentaram desempenho físico aceitável, enquanto que para os demais agrupamentos foram evidenciados comportamentos físicos considerados inaceitáveis;
- o agrupamento 'Cluster 3' foi o que apresentou desempenho físico mais consistente;

6 REFERÊNCIAS

ALLEN, R. G.; PEREIRA, L. S.; RAES, D.; SMITH, M. **Crop evapotranspiration: guidelines for computing crop water requirements**. Roma: FAO (Irrigation and Drainage, n.56). 1998. 300p.

ALLEN, R. G. **Ref-ET- Reference evapotranspiration calculator**. Idaho: University of Idaho Research and Extension Center Kimberly, 2000.

BAENA, L. G. N.; SILVA D. D.; PRUSKI, F. F.; CALIJURI, M. L. *Regionalização de vazões com base em modelo digital de elevação para a bacia do rio Paraíba do Sul. Revista Engenharia na Agricultura, Jaboticabal, v. 24, n.3, p. 612-624, set./dez. 2004, doi: <http://dx.doi.org/10.1590/S0100-69162004000300013>*

BARI, M. F.; SADEK, S. Regionalization of low-flow frequency estimates for rivers in northwest Bangladesh. *Friend 2002 - Regional Hydrology: Bridging the Gap between Research and Practice*(274): 257-263. 2002.

CALINSKI, T.; HARABASZ, J. A dendrite method for cluster analysis. *Communications in Statistics*, 3, no. 1:1-27, 1974.

DIKBAS, F.; FIRAT, M.; KOC, A. C.; GUNGOR, M.; Defining homogeneous regions for streamflow processes in Turkey using a k-means clustering method. *Arabian Journal for Science and Engineering*. 38, pg. 1313-1319. 2013.

CHEN, Y. D.; H.UANG, G.; SHAO, Q.; XU, C. Regional analysis of low flow using L-moments for Dongjiang basin, South China, *Hydrological Sciences Journal*, 51:6, 1051-1064. 2006,

DALRYMPLE, T.; **Flood frequency analysis. Manual of Hydrology: Part 3. Flood-flow techniques**, U.S. Geological Survey Water Supply Paper 1543-A, Washington, D.C. 1960.

DESGRAUPES, B. **clusterCrit: Clustering Indices**. R package version 1.2.2. 2013. Disponível em: <http://CRAN.R-project.org/package=clusterCrit>

DINPASHOH, Y.; FAKHERI-FARD, A.; MOGHADDAM, M.; JAHANBAKHS, S.; MIRNIA, M. Selection of variables for the purpose of regionalization of Iran's

precipitation climate using multivariate methods. **Journal of Hydrology**. 297, 109–123. 2004.

DUNN, J. Well separated clusters and optimal fuzzy partitions. **Journal of Cybernetics**, 4:95 {104, 1974.

ELESBON, A. A. A. **GESTÃO DE RECURSOS HÍDRICOS: ANÁLISES ESTATÍSTICAS MULTIVARIADAS EM SUPORTE À REGIONALIZAÇÃO DE VAZÕES E PROPOSTA METODOLÓGICA PARA AVALIAÇÃO, REARRANJO E OTIMIZAÇÃO DE REDES DE MONITORAMENTO HIDROMÉTRICO**. 144 f. Tese (Doutorado) - Curso de Engenharia Agrícola, Departamento de Departamento de Engenharia Agrícola, Universidade Federal de Viçosa, Viçosa, 2011.

EVERITT, B. S.; LANDAU, S.; LEESE, M. **Cluster analysis**, Oxford University Press, London. 2001.

EUCLYDES, H.P.; SOUZA, E.F.; FERREIRA, P.A. **RH 3.0 – Regionalização hidrológica; manual do programa** - Viçosa: UFV, DEA; Brasília: DF: MMA; Belo Horizonte, MG: RURALMINAS, 1999. 149pg.

GAN, K.C.; MCMAHON, T.A.; O’Neill, I.C. Errors in estimated streamflow parameters and storage for ungauged catchments. **Water Resources Bulletin**, 26 (3), 443–450. 1990.

GPRH - Grupo de Pesquisas em Recursos Hídricos da UFV, IGAM - Instituto Mineiro de Gestão das Águas. **Estudo de regionalização de vazão para o aprimoramento do processo de outorga no Estado de Minas Gerais**. Belo Horizonte: IGAM. 2012. 415p.

HANN, C.T; BARFIELD, B.J; HAYES, J.C. **Design Hydrology and Sedimentology for Small Carchments**. Londres: Academic Press, 1994

HARTIGAN, J. A.; WONG, M. A. A K-means clustering algorithm. **Applied Statistics**. 28. 1979.

HAUSSER, J.; STRIMMER, K. **entropy: Estimation of Entropy, Mutual Information and Related Quantities**. R package version 1.2.0. 2013. Disponível em: <http://CRAN.R-project.org/package=entropy>

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

ISIK, S.; SINGH, V. P. Hydrologic regionalization of watersheds in Turkey. **Journal Of Hydrologic Engineering**, 13(9), 824-834. 2008.

KRAUSE, P.; BOYLE, D. P.; BASE, F. Comparison of different efficiency criteria for hydrological model assessment, **Adv. Geosci.**, 5, 89-97, 2005.

- LAAHA, G.; BLÖSCHL, G. A comparison of low flow regionalisation methods - catchment grouping. **Journal of Hydrology**. 323, p.193-214. 2006.
- LEGATES, D. R.; McCABE, G. J. Evaluating the use of “goodness-of-fit” measures in hydrologic and hydroclimatic model validation, **Water Resource Research**, 35, 1, 233–241, 1999.
- MAINDONALD, J.; BRAUM, W. J. **DAAG: Data Analysis And Graphics data and functions**. R package version 1.16. 2013. Disponível em: <http://CRAN.R-project.org/package=DAAG>
- MALEKINEZHAD, H.; NACHTNEBEL, H. P.; KLIK, A. Comparing the index flood and multiple regression methods using L-moments. **Physics and Chemistry of the Earth**. 36, p.54-60. 2011.
- MAMUM, A. A.; HASHIM, A.; DAOUD, J. I. Regionalisation of low flow frequency curves for the Peninsular Malaysia. **Journal of Hydrology** **381**(1-2): 174-180. 2010.
- NAGHETTINI, M.; PINTO, É. J. de A. **Hidrologia estatística**. Belo Horizonte, CPRM, 2007. 552p.
- NATHAN R.J.; McMAHON T.A. Estimating low-flow characteristics in ungauged catchments. **Water Resources Management**. 6(2): 85–100. 1992.
- NATHAN. R. J.; McMAHON, T. Identification of homogeneous regions for the purpose of regionalization. **Journal of Hydrology**, V. 121, p. 217-238, 1990.
- NASH, J. E.; SUTCLIFFE, J. V. River flow forecasting through conceptual models, Part 1 - A discussion of principles, **Journal of Hydrology**, v. 10, p. 282-290, 1970.
- NOVAES, L. F. Modelo para a quantificação da disponibilidade hídrica na Bacia Hidrográfica do Rio Paracatu. Viçosa: UFV. 2005. 104p. Dissertação Mestrado.
- PANDEY, G. R.; NGUYEN, V.T.V. A comparative study of regression based methods in regional frequency analysis. **Journal of Hydrology**. v.225, p.92-101. 1999.
- PRUSKI, F.F.; NUNES, A.A.; PRUSKI, P.L.; RODRIGUEZ, R. del G. Improved regionalization of streamflow by use of the streamflow equivalent of precipitation as an explanatory variable. **Journal of Hydrology**, v.476, p. 52-71, 2013, doi: <http://dx.doi.org/10.1016/j.jhydrol.2012.10.005>
- 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.; PRUSKI, P.L. Tecnologia e inovação frente a gestão de recursos hídricos. In: Medeiros, S. de S., Gheyi, H.R, Galvão, C. de O., Paz, V.P.S. (Eds), **Recursos**

- hídricos em regiões áridas e semiáridas.** Instituto Nacional do Semiárido, Campina Grande, 27-58. 2011.
- RAJSEKHAR, D.; MISHRA, A; SINGH, V. P. Drought Regionalization of Brazos River Basin Using an Entropy Approach. **Symposium on Data-Driven Approaches to Droughts.** 2011.
- RAO, A.R.; SRINIVAS, V.V. Regionalization of watersheds by fuzzy cluster analysis, **Journal of Hydrology.** 318, p.57-79. 2006^a.
- RAO, A.R.; SRINIVAS, V.V. Regionalization of watersheds by hybrid-cluster analysis, **Journal of Hydrology.** 318, p.37-56. 2006^b.
- RENCHER, A. C. **Methods of multivariate analysis,** Wiley, NewYork. 2002.
- R CORE TEAM. **R: A language and environment for statistical computing.** R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0. 2012. Disponível em: URL <http://www.R-project.org/>.
- RIGGS, H.C. Estimating flow characteristics at ungauged sites. In: Regionalisation in Hydrology, Proceedings of the Ljubljana Symposium, April 1990, **IAHS Publication No.** 191.
- ROUSSEEUW, P. J. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. **Journal of Computational and Applied Mathematics,** 20:53-65, 1987.
- SHANNON, C. E. A Mathematical Theory of Communication. **Bell System Technical Journal,** 27(3): 379-423. 1948.
- SILVA JÚNIOR, O.B.; BUENO, E.O.; TUCCI, C.E.M.; CASTRO, N.M.R. Extrapolação espacial na regionalização de vazão. **Revista Brasileira de Recursos Hídricos,** Porto Alegre: ABRH. 8 (1) Jan./Mar., 21-37. 2003.
- SINGH, V. P. **Entropy Theory and its Application in Environmental and Water Engineering.** Chichester, UK, John Wiley & Sons, Ltd, 2013. 662p.
- SIVAPALAN, M.; TAKEUCHI, K.; FRANKS, S.W.; GUPTA, V.K.; KARAMBIRI, H.; LAKSHIM, V.; LIANG, X.; MCDONNELL, J.J.; MENDIONDO, E.M.; CONNELL, O.; OKI, T.; POMEROY, J.W.; SCHERTZER, D.; UHLENBROOK, S.; ZEHE, E. IAHS Decade on predictions in ungauged basins (PUB), 2003–2012: Shaping an exciting future for the hydrological sciences. **Hydrology Sciences Journal** 48 (6), 857–880. 2003.
- SMAKHTIN, V.U. Low flow hydrology: a review. **Journal of Hydrology.** 240, 147-186. 2001.
- SOUSA, H.T. **Sistema computacional para regionalização de vazões.** Viçosa: UFV. 2009. 86p. Dissertação Mestrado.

- STURGES, H. A. The Choice of a Class Interval. **Journal of the American Statistical Association**. v. 21, p. 65-66. Mar. 1926.
- SYPERRECK, V. L. G.; KLOSOWSKI, E. S; GRECO, M.; FURLANETTO, C. Assessment of modeling approaches for estimates of reference evapotranspiration for the region of Palotina, Parana State, Brazil. **Acta Scientiarum – Agronomy**, 30: 603-609. 2008.
- THORNTHWAITE, C. W.; MATHER, J. R. **The water balance**. Publications in Climatology, New Jersey: Drexel Institute of Technology, 1955. 104p.
- TSAKIRIS, G.; NALBANTIS, I.; CAVADIAS, G. Regionalization of low flows based on canonical correlation analysis. **Advances in Water Resources**. 34, 865–872. 2011.
- TUCCI, C.E. **Regionalização de vazões**. Porto Alegre, UFRGS/IPH , 2002.
- VEZZA, P.; COMOGLIO. C.; ROSSO, M. VIGLIONE, A. Low Flows Regionalization in North-Western Italy. **Water Resources Management** 24(14): 4049-4074. 2010.
- VOGEL, R. M. **Regional calibration of watershed models**. Capítulo 3 In: SINGH, V. P.; FREVERT, D. F. (Ed.) *Watershed Models*. Colorado: ed. CRC Press, p. 47-71. 2005.
- XIE, X.L., BENI, G. **A validity measure for fuzzy clustering**. IEEE Transactions on Pattern Analysis and Machine Intelligence 13 (8), 841–847. 1991.
- YANG,Y.; BURN, D.H. An entropy approach to data collection network design, **Journal of Hydrology**, 157,307-324. 1994.

APÊNDICE A

Quadro 1A - Estações fluviométricas utilizadas no estudo, seus códigos, códigos simplificados, coordenadas geográficas, áreas de drenagem, precipitação média, vazão média de longa duração, coeficiente de escoamento e vazão mínima específica.

Estação	Número	Latitude (°)	Longitude (°)	Área (km ²)	PT (mm)	Q _{mdl} (m ³ /s)	CE	q _{7,10} (L s ⁻¹ km ⁻²)
56028000	1	-20,691	-43,299	1395	1347	24,07	0,40	4,076
56055000	2	-20,848	-43,242	1089	1382	19,02	0,40	5,140
56065000	3	-20,912	-43,097	291	1303	5,55	0,46	5,900
56075000	4	-20,67	-43,088	4251	1342	73,39	0,41	4,480
56090000	5	-20,714	-43	324	1239	4,11	0,32	5,256
56110005	6	-20,384	-42,903	6132	1341	101,96	0,39	4,495
56240000	7	-20,388	-43,183	857	1336	19,34	0,53	5,611
56335001	8	-20,361	-43,139	1371	1297	31,39	0,56	8,314
56337000	9	-20,267	-43,101	531	1321	13,43	0,60	9,147
56385000	10	-20,7	-42,673	534	1250	8,36	0,39	4,199
56415000	11	-20,226	-42,65	2036	1292	25,75	0,31	2,841
56425000	12	-19,994	-42,674	10079	1319	166,44	0,39	5,036
56460000	13	-20,277	-42,326	615	1243	10,08	0,42	2,681
56484998	14	-20,104	-42,44	1347	1257	19,96	0,37	2,232
56500000	15	-20,299	-42,478	272	1256	4,63	0,43	3,522
56510000	16	-20,098	-42,459	1800	1303	23,82	0,32	2,042
56520000	17	-19,999	-42,349	162	1270	2,06	0,32	4,086
56570000	18	-19,709	-42,446	814	1246	10,22	0,32	2,457
56610000	19	-19,932	-43,173	1163	1327	29,47	0,60	5,126
56750000	20	-19,014	-43,446	301	1381	7,37	0,56	3,558
56765000	21	-18,961	-43,243	972	1296	16,54	0,41	2,111
56787000	22	-19,332	-43,07	1280	1348	25,58	0,47	3,670
56800000	23	-18,895	-43,083	1521	1263	19,12	0,31	2,271
56825000	24	-19,188	-42,423	10170	1301	171,65	0,41	3,536
56845000	25	-18,895	-42,713	1064	1180	12,53	0,31	2,236
56846000	26	-18,952	-42,359	1965	1175	26,88	0,37	3,493
56860000	27	-18,883	-41,95	2610	1129	28,81	0,31	2,463
56870000	28	-18,202	-42,455	670	1148	6,15	0,25	1,663
56891900	29	-18,575	-41,918	10200	1079	86,11	0,25	1,558
56900000	30	-18,239	-41,731	732	1155	4,84	0,18	0,372
56915500	31	-18,459	-41,808	1390	1096	8,25	0,17	0,414
56940002	32	-19,069	-41,537	3250	1106	32,22	0,28	1,808
56976000	33	-19,743	-41,785	2290	1136	39,13	0,47	3,635
56978000	34	-19,683	-41,837	2287	1181	43,21	0,50	3,783
56988500	35	-19,799	-41,706	1420	1126	20,41	0,40	3,638
56989001	36	-19,811	-41,438	1187	1116	14,18	0,34	2,025

continua...

Quadro 3A - Cont.

Estação	Número	Latitude (°)	Longitude (°)	Área (km ²)	PT (mm)	Q _{mld} (m ³ /s)	CE	Q _{7,10} (L s ⁻¹ km ⁻²)
56989400	37	-19,595	-41,458	3190	1097	38,88	0,35	2,337
56990000	38	-19,493	-41,162	8810	1113	99,14	0,32	2,356

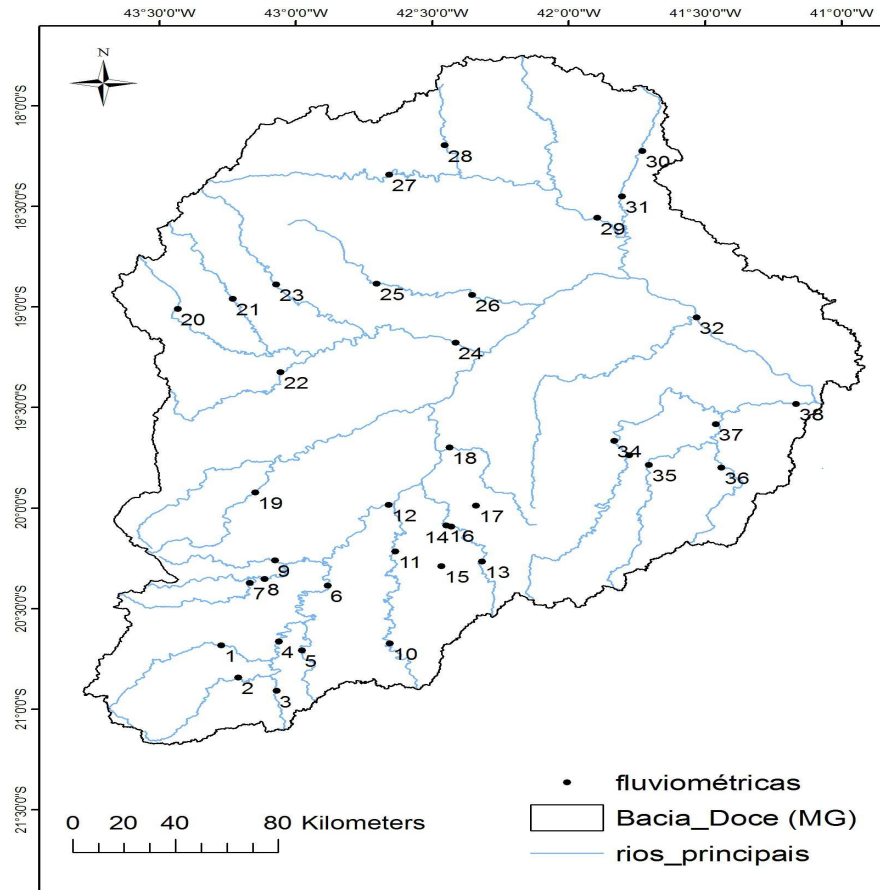


Figura 1A - Estações fluviométricas utilizadas no estudo, sua localização geográfica e códigos simplificados.

Quadro 2A - Estações climatológicas utilizadas no estudo, seus códigos e coordenadas geográficas

Código	Longitude	Latitude
83597	-40,07	-19,42
83648	-40,32	-20,32
83646	-41,1	-20,85
83492	-41,5	-17,85
83543	-41,93	-18,85
83442	-42,05	-16,83
83592	-42,15	-19,8
83594	-42,53	-19,48
83027	-42,68	-21,38
83488	-42,85	-17,85
83642	-42,85	-20,75
83591	-43,12	-19,83
83037	-43,25	-21,57
83692	-43,35	-21,77
83589	-43,43	-19,02
83538	-43,6	-18,25
83689	-43,77	-21,25
83587	-43,93	-19,93
83632	-44,05	-20,02
83586	-44,25	-19,47
83688	-44,27	-21,3
83581	-44,42	-19,88
83536	-44,45	-18,75

Quadro 3A - Estações pluviométricas utilizadas no estudo, seus códigos e coordenadas geográficas

Código	Latitude (°)	Longitude (°)
01741006	-17,6294	-41,7372
01742014	-17,6917	-42,5261
01840004	-18,7536	-40,8936
01840007	-18,5481	-40,9764
01840010	-18,8147	-40,6908
01841006	-18,6811	-41,1967
01841009	-18,9089	-41,1217
01841011	-18,9764	-41,6403
01841018	-18,7631	-41,3083
01841019	-18,7772	-41,4819
01841020	-18,88	-41,9508
01842004	-18,3583	-42,5956
01842005	-18,6119	-42,2786
01842007	-18,7722	-42,9311
01842008	-18,2011	-42,4553
01842020	-18,5528	-42,7642
01843002	-18,4656	-43,7431
01843011	-18,5928	-43,4125
01843012	-18,2797	-43,0006
01940005	-19,6922	-40,3986
01940006	-19,5333	-40,6458
01940009	-19,2203	-40,8533
01940012	-19,6628	-40,8353
01940013	-19,2375	-40,5914
01940016	-19,0594	-40,5156
01940022	-19,5889	-40,1983
01940023	-19,2742	-40,3208
01940025	-19,2953	-40,5175
01940036	-19,4	-40,0667
01940043	-19,8667	-40,8833
01941003	-19,5181	-41,0144
01941004	-19,3394	-41,2406
01941005	-19,0617	-41,5328
01941006	-19,5931	-41,4564
01941008	-19,9011	-41,0597
01941009	-19,6919	-41,0194
01941010	-19,4917	-41,1617
01941011	-19,6783	-41,8361
01941012	-19,0592	-41,0275
01941018	-19,1758	-41,8703
01941019	-19,8167	-41,4436
01942002	-19,8336	-42,3178

continua...

Quadro 3A - Cont.

Código	Latitude (°)	Longitude (°)
01942006	-19,9989	-42,3475
01942008	-19,3731	-42,1044
01942021	-19,8	-42,15
01942029	-19,5247	-42,6442
01942031	-19,8158	-42,4769
01942032	-19,1886	-42,4228
01943003	-19,2503	-43,0144
01943004	-19,5206	-43,7444
01943007	-19,9453	-43,4011
01943008	-19,4419	-43,1178
01943023	-19,6639	-43,6881
01943024	-19,6897	-43,5856
01943025	-19,2175	-43,3742
01943027	-19,8808	-43,3675
01943035	-19,2189	-43,5881
02040008	-20,1433	-40,9753
02041008	-20,1081	-41,7283
02042008	-20,1036	-42,44
02042010	-20,2989	-42,4781
02042011	-20,2261	-42,65
02042014	-20,7694	-42,3011
02042015	-20,7261	-42,9172
02042016	-20,6825	-42,8067
02042017	-20,2772	-42,3256
02042018	-20,3847	-42,9028
02042027	-20,9633	-42,0372
02042031	-19,9947	-42,6742
02043009	-20,3625	-43,1439
02043010	-20,6906	-43,2994
02043011	-20,39	-43,1803
02043013	-20,5219	-43,83
02043014	-20,6703	-43,0881
02043018	-20,9558	-43,8008
02043025	-20,5167	-43,0167
02043026	-20,8475	-43,2419
02043027	-20,2856	-43,0989
02142009	-20,9114	-42,3494
02143007	-21,1928	-43,6139

APÊNDICE B

Quadro 1B - Estações fluviométricas utilizadas no estudo, seus códigos simplificados e variáveis utilizadas na análise de cluster

Estação	Número	Latitude (°)	Longitude (°)	PT (mm)	ETR (mm)
56028000	1	-20,691	-43,299	1347	1005
56055000	2	-20,848	-43,242	1382	995
56065000	3	-20,912	-43,097	1303	970
56075000	4	-20,67	-43,088	1342	996
56090000	5	-20,714	-43	1239	971
56110005	6	-20,384	-42,903	1341	993
56240000	7	-20,388	-43,183	1336	998
56335001	8	-20,361	-43,139	1297	993
56337000	9	-20,267	-43,101	1321	969
56385000	10	-20,7	-42,673	1250	936
56415000	11	-20,226	-42,65	1292	971
56425000	12	-19,994	-42,674	1319	988
56460000	13	-20,277	-42,326	1243	996
56484998	14	-20,104	-42,44	1257	989
56500000	15	-20,299	-42,478	1256	990
56510000	16	-20,098	-42,459	1303	988
56520000	17	-19,999	-42,349	1270	980
56570000	18	-19,709	-42,446	1246	967
56610000	19	-19,932	-43,173	1327	952
56750000	20	-19,014	-43,446	1381	976
56765000	21	-18,961	-43,243	1296	963
56787000	22	-19,332	-43,07	1348	975
56800000	23	-18,895	-43,083	1263	952
56825000	24	-19,188	-42,423	1301	961
56845000	25	-18,895	-42,713	1180	944
56846000	26	-18,952	-42,359	1175	950
56860000	27	-18,883	-41,95	1129	966
56870000	28	-18,202	-42,455	1148	947
56891900	29	-18,575	-41,918	1079	954
56900000	30	-18,239	-41,731	1155	970
56915500	31	-18,459	-41,808	1096	963
56940002	32	-19,069	-41,537	1106	935
56976000	33	-19,743	-41,785	1136	990
56978000	34	-19,683	-41,837	1181	989
56988500	35	-19,799	-41,706	1126	982
56989001	36	-19,811	-41,438	1116	967
56989400	37	-19,595	-41,458	1097	971
56990000	38	-19,493	-41,162	1113	928

Quadro 2B - Estações fluviométricas utilizadas no estudo, seus códigos simplificados e variáveis padronizadas utilizadas na análise de cluster

Estação	Número	Latitude	Longitude	P T	ETR
56028000	1	-1,27218	-1,28163	1,166966	1,710846
56055000	2	-1,47788	-1,18993	1,539733	1,155991
56065000	3	-1,56198	-0,95913	0,68286	-0,08098
56075000	4	-1,2456	-0,94451	1,109366	1,227257
56090000	5	-1,30349	-0,80364	0,000436	-0,01735
56110005	6	-0,87024	-0,64903	1,095401	1,069455
56240000	7	-0,87497	-1,09601	1,041537	1,329065
56335001	8	-0,84075	-1,02646	0,620489	1,079635
56337000	9	-0,71733	-0,96489	0,886837	-0,16242
56385000	10	-1,28419	-0,28268	0,109373	-1,83208
56415000	11	-0,66345	-0,24591	0,572465	-0,04534
56425000	12	-0,35981	-0,28489	0,865779	0,830205
56460000	13	-0,73044	0,271509	0,03821	1,206896
56484998	14	-0,50289	0,088996	0,187279	0,89129
56500000	15	-0,75883	0,028306	0,178108	0,901471
56510000	16	-0,49488	0,058429	0,690475	0,845477
56520000	17	-0,36564	0,234741	0,330137	0,397521
56570000	18	0,01482	0,080136	0,075677	-0,22351
56610000	19	-0,27753	-1,08006	0,952288	-1,03288
56750000	20	0,924637	-1,51553	1,533055	0,188814
56765000	21	0,995267	-1,1917	0,614847	-0,47294
56787000	22	0,508502	-0,91616	1,18033	0,153181
56800000	23	1,081552	-0,93565	0,257698	-1,01252
56825000	24	0,697091	0,116462	0,663317	-0,5493
56845000	25	1,081552	-0,34558	-0,64668	-1,41466
56846000	26	1,006917	0,217907	-0,69926	-1,08888
56860000	27	1,096916	0,869958	-1,19965	-0,29477
56870000	28	1,989912	0,065517	-0,98384	-1,25177
56891900	29	1,500963	0,921823	-1,73529	-0,90053
56900000	30	1,940763	1,219515	-0,91761	-0,10134
56915500	31	1,652053	1,097692	-1,55358	-0,42713
56940002	32	0,85255	1,529167	-1,44142	-1,88298
56976000	33	-0,03033	1,133131	-1,12197	0,932013
56978000	34	0,049042	1,050291	-0,62692	0,8862
56988500	35	-0,1035	1,259384	-1,22737	0,504419
56989001	36	-0,11879	1,687759	-1,33687	-0,22351
56989400	37	0,164089	1,654977	-1,53719	-0,04534
56990000	38	0,298068	2,127651	-1,36503	-2,24949

APÊNDICE C

Quadro 1C - Classificação das estações nas RHH identificas

Estação	Número	Cluster 1	Cluster 2/2	Cluster 2/5	Cluster 3	Cluster 4	Entropia (Q _{md})	Entropia (Q _{7,10})
56028000	1	1	1	1	1	1	1	1*
56055000	2	1	1	1	1	1	1	1
56065000	3	2	1	2	1	1	-	1
56075000	4	1	1	1	1	1	1	1
56090000	5	2	1	3	1	1	-	1
56110005	6	1	1	1	1	1	1	1
56240000	7	1	1	1	1	1	1	-
56335001	8	1	1	2	1	1	1	2
56337000	9	2	1	2	1	1	1	1
56385000	10	3	1	3	1	1	1	1
56415000	11	2	1	2	1	1	1	1
56425000	12	1	1	2	1	1	1	2
56460000	13	1	1	3	1	1	2	2
56484998	14	1	1	3	1	1	2	1
56500000	15	1	1	3	1	1	-	1
56510000	16	1	1	2	1	1	2	1
56520000	17	2	1	3	1	1	-	-
56570000	18	2	1	3	1	1	2*	1
56610000	19	2	1	2	1	1	1	1
56750000	20	2	1	1	1	1	2	1
56765000	21	2	1	2	1	1	2	1
56787000	22	2	1	1	1	1	2*	1
56800000	23	2	1	3	1	1	2	-
56825000	24	2	1	2	1	1	1	1
56845000	25	3	2	4	2	2	2	1
56846000	26	3	2	4	2	2	2	-
56860000	27	4	2	5	2	2	2	1
56870000	28	3	2	4	2	2	2	1
56891900	29	3	2	5	2	2	2	1
56900000	30	4	2	4	2	2	2	1
56915500	31	4	2	5	2	2	2	1
56940002	32	3	2	5	2	2	1	2
56976000	33	4	2	5	2	2	-	1
56978000	34	4	2	4	2	2	1	1
56988500	35	4	2	5	2	2	1	1*
56989001	36	4	2	5	2	2	1	1
56989400	37	4	2	5	2	2	1	1
56990000	38	3	2	5	2	2	1	1

Nota: o asterisco denota estações que não foram utilizadas na estimativa dos parâmetros da equação de regressão.

APÊNDICE D

Quadro 1D - Análise de segurança relativa à regionalização da Q_{mid}

Agrupamento	Porcentagem de trechos classificados em cada classe			
	Aceitável/ Potencial de subestimativa	Provável - Confiável	Aceitável - Potencial de superestimativa	Inaceitável
Geo 1	0	100	0	0
Geo 2	13,37	63,43	23,21	0
Geo 3	13,37	86,63	0	0
Cluster 1	3,3	44,2	34,6	17,8
Cluster 2	22,5	51,5	26	0
Cluster 3	30	70	0	0
Entropia	0	37,43	44,14	18,3

Quadro 2D - Análise de segurança relativa à regionalização da $Q_{7,10}$

Agrupamento	Porcentagem de trechos classificados em cada classe		
	Aceitável - Potencial de subestimativa	Provável - Confiável	Aceitável - Potencial de superestimativa
Geo 1	0	100	0
Geo 2	13,8	45,93	40,3
Geo 3	13,8	86,2	0
Cluster 1	19,4	56,6	24
Cluster 2	29,2	36,4	34,4
Cluster 3	29,9	70,1	0
Entropia	3	97	0