

**UNIVERSIDADE ESTADUAL DO OESTE DO PARANÁ – UNIOESTE**  
**CENTRO DE CIÊNCIAS EXATAS E TECNOLÓGICAS – CAMPUS CASCAVEL**  
**PÓS-GRADUAÇÃO EM ENGENHARIA AGRÍCOLA**

**IDENTIFICAÇÃO DE REGIÕES HIDROLOGICAMENTE HOMOGÊNEAS POR**  
**AGRUPAMENTO FUZZY C-MEANS NO ESTADO DO PARANÁ**

**ANA CLAUDIA GUEDES SILVA**

**CASCAVEL**

**2018**

**ANA CLAUDIA GUEDES SILVA**

**IDENTIFICAÇÃO DE REGIÕES HIDROLOGICAMENTE HOMOGÊNEAS POR  
AGRUPAMENTO FUZZY C-MEANS NO ESTADO DO PARANÁ**

Dissertação apresentada ao Programa de Pós-Graduação em Engenharia Agrícola em cumprimento parcial aos requisitos para obtenção do título de Mestre em Engenharia Agrícola, área de concentração em Recursos Hídricos e Saneamento Ambiental.

Orientador: Prof. Dr. Benedito Martins Gomes.

**CASCADEL - PARANÁ - BRASIL**

**2018**

Dados Internacionais de Catalogação-na-Publicação (CIP)  
(Sistema de Bibliotecas - UNIOESTE)

S578i Silva, Ana Claudia Guedes.  
Identificação de regiões hidrologicamente homogêneas por agrupamento Fuzzy C-Means no estado do Paraná / Benedito Martins Gomes. --- Cascavel (PR), 2018.  
90 f.:il.

Orientador: Prof. Dr. Benedito Martins Gomes.  
Dissertação (Mestrado em Engenharia Agrícola) – Universidade Estadual do Oeste do Paraná, Campus de Cascavel, 2018, Programa de Pós-Graduação em Engenharia Agrícola.  
Inclui Bibliografia

1. Sistemas difusos. 2. Medidores de fluxo. 3. Análise por agrupamento. I. Gomes, Benedito Martins. II. Universidade Estadual do Oeste do Paraná. III. Título.

CDD 551.48

Rosângela A. A. Silva – CRB 9ª/1810

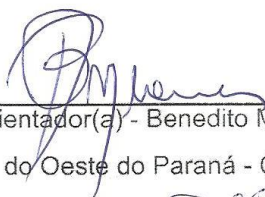
---

\*Revisão de Normas, Português e Inglês, pela Prof. Dhandara Soares de Lima Capitani, 20 de março de 2018

## ANA CLAUDIA GUEDES SILVA

Identificação de Regiões Hidrologicamente Homogêneas Por Agrupamento Fuzzy C-Means no Estado do Paraná

Dissertação apresentada ao Programa de Pós-Graduação em Engenharia Agrícola em cumprimento parcial aos requisitos para obtenção do título de Mestra em Engenharia Agrícola, área de concentração Recursos Hídricos e Saneamento Ambiental, linha de pesquisa Recursos Hídricos, APROVADO(A) pela seguinte banca examinadora:



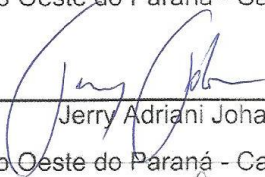
Orientador(a) - Benedito Martins Gomes

Universidade Estadual do Oeste do Paraná - Campus de Cascavel (UNIOESTE)



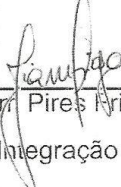
Eloy Lemos de Mello

Universidade Estadual do Oeste do Paraná - Campus de Cascavel (UNIOESTE)



Jerry Adriani Johann

Universidade Estadual do Oeste do Paraná - Campus de Cascavel (UNIOESTE)



Jiam Pires Frigo

Universidade Federal da Integração Latino-Americana (Unila)

Cascavel, 7 de fevereiro de 2018

## **BIOGRAFIA**

Ana Claudia Guedes Silva, nascida em 23 de maio de 1992 na cidade de Cascavel, Paraná, filha de Osni Silva e Claudécí Guedes da Cruz Silva.

No segundo semestre de 2010, ingressou no curso de Engenharia Ambiental e Sanitária da Universidade Federal de Santa Maria (UFSM), campus de Frederico Westphalen, Rio Grande do Sul, concluindo-o em julho de 2015.

Em 2016 iniciou o curso de Pós-Graduação em Engenharia Agrícola, nível mestrado, na área de concentração de Recursos Hídricos e Saneamento Ambiental, na Universidade Estadual do Oeste do Paraná.

“Não fui eu que ordenei a você?  
Seja forte e corajoso! Não se  
apavore nem desanime, pois, o  
Senhor, o seu Deus, estará com você  
por onde você andar”

- Josué 1:9

Dedico à minha família.

## AGRADECIMENTOS

Agradeço primeiramente a Deus, por ter me dado sabedoria e a vida que tenho;

Aos meus pais, Osni Silva e Claudécí Guedes, pelo empenho e a dedicação, não medindo esforços em proporcionar uma educação de qualidade. Eu amo vocês;

Ao meu irmão Junior pelo companheirismo. Te amo;

Ao Gabriel Trevisan, pela paciência, a compreensão, a motivação, a amizade, pelo amor, carinho e apoio irrestrito, em todos os momentos;

Aos meus professores e orientadores, Benedito Martins, Eloy de Mello e Josué de Castro, pela orientação, pela confiança, pelo exemplo de profissionalismo e seriedade, pelo incentivo e a ajuda na elaboração e na composição do trabalho;

A todos os professores que tive até o momento. Os conhecimentos passados colaboraram de forma expressiva para o meu desenvolvimento científico e pessoal;

Ao CNPq pelo apoio à pesquisa por meio da bolsa a mim concedida durante todo o período em que estive cursando o Mestrado;

Afinal, a todos que eu não citei, mas que direta ou indiretamente contribuíram para a realização desta Dissertação.



# IDENTIFICAÇÃO DAS REGIÕES HIDROLOGICAMENTE HOMOGÊNEAS POR AGRUPAMENTO FUZZY C-MEANS NO ESTADO DO PARANÁ

## RESUMO

O delineamento de regiões hidrologicamente homogêneas (RHH) é um procedimento essencial para provimento de informações indispensáveis aos trabalhos de modelagem, planejamento e gestão de recursos hídricos, principalmente quando se tem a necessidade de realizar a regionalização de vazões, visando estimar a disponibilidade hídrica em seções desprovidas de medições. A definição de estratégias de manejo e conservação dos recursos naturais depende de informações obtidas por meio da identificação de RHH, sendo também um dos passos de um estudo de regionalização de vazões. Assim, este trabalho tem como objetivo a identificação das RHH no estado do Paraná através do método de agrupamento Fuzzy C-Means. Foram utilizadas 9 variáveis, individualizadas para as 114 estações fluviométricas adotadas, sendo 4 variáveis dependentes referentes às vazões características (vazão média anual de longa duração ( $Q_{mld}$ ), vazão mínima anual com sete dias de duração e período de retorno de 10 anos ( $Q_{7,10}$ ), vazões associadas às permanências de 95% ( $Q_{95}$ ) e 90% ( $Q_{90}$ )) e 5 independentes referentes às características morfométrias da estação (área de drenagem ( $AD - m^2$ ), soma das drenagens ( $SD - m$ ), densidade de drenagem ( $DD - 1/m$ ) e a localização geográfica (latitude - Lat e longitude - Long). A partir da análise de componentes principais (ACP) identificou-se as variáveis  $Q_{mld}$ ,  $DD$ , Lat e Long como as menos representativas, sendo excluídas do estudo, dando procedência à análise de agrupamentos apenas com as variáveis  $AD$ ,  $SD$ ,  $Q_{90}$ ,  $Q_{95}$  e  $Q_{7,10}$ . Aplicou-se o Fuzzy C-Means para as variáveis escolhidas, sendo que a menor função objetiva encontrada foi para 4 Clusters no índice de fuzzificação ( $m$ ) 1,7. Separando as estações fluviométricas por clusters através dos graus de pertinência, obtivemos o maior número de estações no Cluster 3 (83 estações), seguidos do Cluster 4 (13 estações) e dos Clusters 1 e 2 (7 estações em cada cluster), e apenas 4 estações não foram inseridas em nenhum cluster, sendo classificadas como nebulosas, sendo que os grupos foram determinados praticamente pela distribuição das variáveis  $AD$  e  $SD$ . As menores áreas de abrangência, vazões analisadas e as menores quantidade de drenagens na área de cobertura das estações foram encontradas no Cluster 3, que estão bem espalhadas no estado do Paraná. Já os Clusters 1 e 4 ficaram intermediários entre os demais clusters em todos os parâmetros avaliados. O algoritmo Fuzzy C-Means se mostrou eficiente para o agrupamento das estações fluviométricas no estado do Paraná, onde foi possível encontrar as características de cada cluster formado, sem haver sobreposição de dados nos intervalos das variáveis analisadas.

**Palavras-chave:** Fuzzificação, vazão, análise de agrupamento.

# IDENTIFICATION OF HYDROLOGICALLY HOMOGENEOUS REGIONS BY FUZZY C-MEANS GROUP IN THE STATE OF PARANÁ

## ABSTRACT

The design of hydrologically homogeneous regions (RHH) is an essential procedure to provide information essential to the modeling, planning, and management of water resources, especially when it is necessary to perform the regionalization of flows, aiming to estimate the water availability in sections without measurements. The definition of strategies for the management and conservation of natural resources depends on information obtained through the identification of RHH, also being one of the steps of a study of regionalization of flows. Thus, this work has the objective of identifying the RHH in the state of Paraná through the grouping method Fuzzy C-Means. A total of 9 variables were used for the 114 fluviometric stations, with 4 dependent variables related to the characteristic flows (annual average long-term flow ( $Q_{mid}$ ), minimum annual flow with seven days duration and 10-year return period ( $Q_{7,10}$ ), flow rates associated to the 95% ( $Q_{95}$ ) and 90% ( $Q_{90}$ ) permanencies) and 5 independent variables related to the morphometric characteristics of the station (drainage area (AD -  $m^2$ ), sum of drainage (SD - m) (LA - Lat and longitude - Long). From the principal components analysis (PCA), the variables  $Q_{mid}$ , DD, Lat and Long were identified as the least representative, being discarded from the study, proceeding with the analysis using only the variables AD, SD,  $Q_{90}$ ,  $Q_{95}$ , and  $Q_{7,10}$ . The results were obtained using the Fuzzy C-Means for the chosen variables, and the smallest objective function was found for 4 Clusters in the study group, with index of and fuzzification (m) 1.7. Separating the fluviometric stations by clusters through degrees of pertinence, the largest number of stations were obtained in Cluster 3 (83 stations), followed by Cluster 4 (13 stations) and Clusters 1 and 2 (7 stations in each cluster), and only 4 stations were not inserted in any cluster, being classified as nebulae, where the groups were determined practically by the distribution of the AD and SD variables. The smaller areas of coverage, analyzed flows and the smaller amount of drainage in the coverage area of the stations were found in Cluster 3, considering they were well spread in the state of Paraná. Clusters 1 and 4 were intermediate among the other clusters in all parameters evaluated. The Fuzzy C-Means algorithm proved to be efficient for the grouping of fluviometric stations in the state of Paraná, where it was possible to find the characteristics of each cluster formed, without overlapping of data in the analyzed variables.

**Key-words:** Fuzzification, flow, cluster analysis.

## SUMÁRIO

<b>1</b>	<b>INTRODUÇÃO</b> .....	<b>1</b>
<b>2</b>	<b>OBJETIVOS</b> .....	<b>3</b>
2.1	Objetivo geral.....	3
2.2	Objetivo específicos.....	3
<b>3</b>	<b>REVISÃO BIBLIOGRÁFICA</b> .....	<b>4</b>
3.1	Vazões de referência.....	4
3.2	Regionalização Hidrográfica .....	4
3.3	Regiões Hidrologicamente Homogêneas.....	5
3.3.1	Conveniência Geográfica .....	6
3.3.2	Análise da distribuição de frequência.....	6
3.3.3	Híbridos .....	7
3.3.4	Árvore de classificação .....	7
3.3.5	Análise de correlação canônica .....	8
3.3.6	Entropia.....	9
3.3.7	Análise de agrupamentos de dados .....	9
3.4	Método de Agrupamento de dados.....	10
3.5	Métodos Hierárquicos .....	11
3.6	Métodos Não Hierárquicos .....	11
3.6.1	Métodos de Partição .....	12
3.6.2	Métodos Difusos .....	13
3.7	Algoritmo Fuzzy C-Means (FCM).....	14
3.8	Medidas de Similaridade.....	16
3.8.1	Distância Euclidiana.....	17
3.9	Análise de Componentes Principais – ACP.....	17

<b>4</b>	<b>MATERIAL E MÉTODOS.....</b>	<b>20</b>
4.1	Descrição da área de estudo .....	21
4.1.1	Hidrografia .....	21
4.2	Dados Utilizados .....	22
4.2.1	Obtenção das variáveis dependentes.....	23
4.2.1.1	Vazão média anual de longa duração (Q <sub>ml</sub> ).....	23
4.2.1.2	Vazões de permanência de 90 e 95% (Q <sub>90</sub> e Q <sub>95</sub> ).....	23
4.2.1.3	Vazão mínima com sete dias de duração .....	24
4.2.2	Obtenção das variáveis independentes.....	25
4.3	Análise de Componentes Principais – ACP.....	25
4.4	Identificação das Regiões Hidrologicamente Homogêneas – RHH.....	26
4.4.1	Transformação e Normalização .....	26
4.4.2	Mineração de Dados (Data Mining).....	27
4.4.3	Interpretação e Validação .....	27
<b>5</b>	<b>RESULTADOS E DISCUSSÃO .....</b>	<b>28</b>
5.1	Variáveis dependentes.....	28
5.2	Variáveis independentes.....	28
5.3	Aplicação da ACP aos dados dependentes e independentes.....	28
5.4	Análise de agrupamento Fuzzy C-Means .....	29
5.5	Caracterização das RHH .....	38
<b>6</b>	<b>CONCLUSÃO .....</b>	<b>43</b>
<b>7</b>	<b>CONSIDERAÇÕES FINAIS .....</b>	<b>43</b>
<b>8</b>	<b>REFERÊNCIAS .....</b>	<b>45</b>
	<b>APÊNDICES.....</b>	<b>57</b>
	<b>APÊNDICE A ESTAÇÕES FLUVIOMÉTRICAS DAS BACIAS HIDROGRÁFICAS DO ESTADO DO PARANÁ.....</b>	<b>58</b>
	<b>APÊNDICE B SCRIPT DA ROTINA PARA ACP.....</b>	<b>61</b>

<b>APÊNDICE C</b>	<b>MATRIZ DE ENTRADA COM AS VARIÁVEIS DEPENDENTES. ....</b>	<b>67</b>
<b>APÊNDICE D</b>	<b>MATRIZ DE ENTRADA COM AS VARIÁVEIS INDEPENDENTES....</b>	<b>70</b>
<b>APÊNDICE E</b>	<b>TABELA CONTENDO OS GRAUS DE PERTINÊNCIA PARA CADA ESTAÇÕES FLUVIOMÉTRICAS ESTUDADA NOS 4 CLUSTERS. ...</b>	<b>73</b>

## LISTA DE TABELAS

Tabela 1 Matriz de Correlação Simples entre as variáveis em estudo .....	29
Tabela 2 Variância explicada e cumulativa ao longo das componentes principais .....	30
Tabela 3 Autovetores das variáveis do Componente escolhido .....	31
Tabela 4 Valores das 176 funções objetivas obtidas por combinação cluster x índice de fuzzificação.....	34
Tabela 5 Graus de pertinência das estações fluviométricas para os 4 grupos formados .....	36
Tabela 6 Distribuição das estações fluviométrica por Cluster .....	37
Tabela 7 Características de cada região formada por meio do método de agrupamento Fuzzy C-Means.....	38
Tabela 8 Dados da AD nos clusters formados .....	40
Tabela 9 Dados da SD nos clusters formados .....	40
Tabela 10 Dados da Q90 e Q95 nos clusters formados.....	41
Tabela 11 Dados da Q7,10 nos clusters formados .....	41

## LISTA DE FIGURAS

Figura 1 Exemplo de uma árvore de decisão visando na compra de um computador.....	8
Figura 2 Fluxograma metodológico.....	20
Figura 3 Mapa com as 114 estações fluviométricas estudadas nas respectivas bacias hidrográficas.....	22
Figura 4 Biplot dos dados no espaço das duas primeiras PC, com pontos nomeados nas variáveis e nas 114 observações.....	32
Figura 5 Convergência da função objetivo para 4 agrupamentos. ....	35
Figura 6 Mapa das estações hidrologicamente homogêneas de vazão definidas no estudo através do método Fuzzy C-Means. ....	37
Figura 7 Distribuição de grupos pelas variáveis Soma das drenagens (m) x Área de drenagem (km <sup>2</sup> ), em escala logarítmica. ....	39

## 1 INTRODUÇÃO

Assim como todas as outras variáveis utilizadas para a caracterização de processos hidrológicos, a vazão possui comportamento aleatório, exigindo, para sua adequada avaliação, a análise das séries históricas que são obtidas nos postos fluviométricos instalados em bacias hidrográficas (COSTA et al., 2012).

O conhecimento das vazões e do comportamento hidrológico das bacias hidrográficas é fundamental para o planejamento dos recursos hídricos. Entretanto, nem sempre as informações necessárias para esse planejamento existem ou estão disponíveis, seja pelo escasso número de estações hidrométricas ou pelo curto período de observações.

Para tentar suprir essa deficiência, pesquisadores têm adotado a técnica de regionalização, que tem o objetivo produzir informações hidrológicas em locais com escassez de dados, com base na similaridade espacial de variáveis e parâmetros que permitem essa transferência. Contudo, antes da regionalização é necessária a identificação das regiões hidrologicamente homogêneas - RHH (TUCCI, 2002).

Há diversos estudos que apontam o uso vantajoso das RHH em escala regional heterogênea no que diz respeito aos processos hidrológicos, pois divide a região de interesse em sub-regiões agrupando as bacias que apresentam comportamento hidrológico homogêneo (MAMUN; HASHIM; DAOUD, 2010; BOOKER; SNELDER, 2012; TONGAL; DEMIRED; BOOIJ, 2013).

Porém, para que uma região seja considerada homogênea é necessário que existam evidências suficientes de que as diferentes amostras do grupo possuem similaridade de suas características hidrológicas, considerando, por exemplo, o comportamento das vazões e os aspectos físicos.

Essa etapa é a mais difícil e está sujeita a certa subjetividade, por exigir do analista e da metodologia empregada a capacidade de discernir se observações anômalas, eventualmente existentes em uma ou mais amostras do grupo (CALEGARIO, 2014).

Devido a essa subjetividade, é de grande importância a escolha das variáveis que serão utilizadas para identificar os comportamentos hidrológicos semelhantes. Em muitas situações pode ser incluído um grande conjunto de variáveis independentes e, na maioria dos casos, em uma análise estatística multivariada, essa inclusão é desnecessária, se tornando um obstáculo para a correta interpretação dos dados (MORAIS, 2011).

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 a escolha do método de identificação de RHH seja de fundamental importância na correta importância na correta



representatividade da variação, estatística e física, das vazões regionalizadas ao longo da hidrografia (MIRANDA, 2016).

Dentre os métodos de análise de agrupamento utilizados na hidrologia para obtenção de regiões homogêneas, destaca-se o método de agrupamento difuso Fuzzy C-Means por apresentar uma característica peculiar (graus de pertinência) que permite considerar ambiguidade e sobreposição dos grupos formados, aproveitando-se mais eficientemente as informações disponíveis.

Por conta disso, o método vem sendo utilizado em estudos internacionais para determinar regiões homogêneas a partir das vazões (RAO; SRINIVAS, 2006; SRINIVAS et al., 2008; TURAN; YURDUSEV, 2009; SADRI; BURN, 2011).

## **2 OBJETIVOS**

### **2.1 Objetivo geral**

Delimitar as regiões hidrologicamente homogêneas para o estado do Paraná.

### **2.2 Objetivo específicos**

- Reduzir a dimensionalidade do banco de dados;
- Aplicar o método de análise de agrupamento Fuzzy C-Means nas variáveis hidrológicas consideradas;
- Descrever as características de cada região hidrológica encontrada.

### 3 REVISÃO BIBLIOGRÁFICA

#### 3.1 Vazões de referência

A vazão é adotada quando se deseja simular as condições médias prevalentes, seja em escala anual, seja durante os meses chuvosos ou secos.

A vazão média de longa duração ( $Q_{mid}$ ) representa a disponibilidade hídrica de uma bacia para aproveitamento energético (PRUSKI; PRUSKI, 2011), sendo esta considerada a máxima vazão possível de ser regularizada, desconsiderando as perdas por evaporação e infiltração (TUCCI, 2002; CALEGARIO, 2014).

As vazões mínimas representam a disponibilidade hídrica natural ao longo da hidrografia, sendo a quantidade de água que permanece no leito do rio em épocas de estiagem ou de baixa precipitação. Elas asseguram a qualidade e a quantidade de água, no tempo e no espaço, necessárias para manter os componentes, as funções e os processos dos ecossistemas aquáticos (MMA, 2010).

Estas são úteis quanto ao aspecto qualitativo e quantitativo, tanto para a concessão de outorgas, quanto para a avaliação de projetos de navegação.

As vazões mínimas de referência mais utilizadas são: as vazões com permanência de 90% ( $Q_{90}$ ) e 95% ( $Q_{95}$ ); e a vazão mínima com sete dias de duração associado a um período de retorno de 10 anos ( $Q_{7,10}$ ) são vazões de referência porque representam o limite superior de utilização da água em um curso hídrico, se tornando um dos principais obstáculos de implementação da outorga, porém, constituem um procedimento adequado para a proteção dos rios (SOUZA, 2015; GRANZIERA, 2013).

A disponibilidade de informações destas variáveis ao longo da hidrografia é, em geral, deficiente ou até mesmo inexistente, havendo a necessidade da utilização da análise regional para a transferência de informações de outros locais cujo comportamento hidrológico seja semelhante ao do local de interesse (CALEGARIO, 2014).

#### 3.2 Regionalização Hidrográfica

Uma rede hidrológica raramente abrange todas as regiões de interesse em uma bacia hidrográfica devido ao alto custo que demanda a implantação e a manutenção de uma rede hidrométrica, gerando lacunas espaciais (PINTO, 2006).

É possível suprir parcialmente estas falhas pela transferência de informações provenientes de áreas monitoradas, que sejam hidrológicamente semelhantes. Este processo é denominado regionalização (WOLFF; DUARTE; MINGOTI, 2014).

A regionalização de vazões pode ser utilizada tanto para aumentar a confiabilidade dos valores estimados em uma seção já monitorada, como para estimar as vazões em áreas desprovidas de monitoramento (CALEGARIO, 2014).

Para Tucci (2002), a regionalização também pode ser usada para melhor explorar as amostras pontuais, verificando a consistência das séries hidrológicas e, como consequência, identificando a falta de postos de observações na área em estudo.

Há vários métodos citados na literatura sobre regionalização, sendo que os mais usuais envolvem a construção de modelos empíricos de regressão que relacionam a vazão de interesse e os aspectos físicos e climáticos da bacia. Existem outros métodos também, baseados, por exemplo, no uso de procedimentos de interpolação e calibração hídrica (VOGEL, 2005; CHAVES et al., 2002; AMORIM; NETTO; MENDIONDO, 2005).

A hierarquização do estudo da regionalização inicia-se com a divisão da área de estudo em regiões hidrologicamente homogêneas, na seleção do melhor modelo de regressão e da variável explicativa, sempre visando garantir maior confiabilidade preditiva e segurança no resultado final (IGAM, 2012).

### **3.3 Regiões Hidrologicamente Homogêneas - RHH**

A subdivisão em regiões homogêneas, conforme as variáveis de interesse, é prevista em estudos de regionalização na tentativa de se obter modelos mais adequados às realidades físicas e hidrológicas das bacias, devendo esta divisão seguir alguns critérios de similaridade.

Os critérios para defini-las não são rigidamente estabelecidos; entretanto, é comum considerar a similaridade das características geológicas, climáticas, hidrológicas e geográficas das sub-bacias (SOUSA, 2009).

Para Naghettini e Pinto (2007), uma região é homogênea se existem evidências suficientes de que as diferentes amostras do grupo possuem a mesma distribuição de frequências, considerando o fator de escala local.

É uma etapa essencial e considerada uma das mais importantes para a regionalização, visto que a delimitação incorreta dessas regiões compromete significativamente a qualidade das equações que serão obtidas.

Com o passar dos anos desenvolveram-se inúmeros métodos para definir e delimitar uma RHH a partir de parâmetros físicos e climáticos da bacia; porém, não existe concordância de uma técnica objetiva para ser aplicada em todas as situações (NAGHETTINI; PINTO, 2007).

Há trabalhos que abordam métodos existentes de maneira genérica, como é o caso dos estudos de Pruski et al. (2013) e Pereira, Kobiyama e Castro (2013). Já outros se

aprofundam na aplicação de apenas um método (MALEKINEZHAD; NACHTNEBEL; KLIK, 2011; WOLFF, 2013; PANSERA et al., 2013, 2015).

Contudo, no geral, os métodos existentes a fim de dividir uma área de drenagem em RHH são: conveniência geográfica, análise da distribuição de frequência, híbridos, árvore de classificação, análise de correlação canônica, entropia e análise de agrupamentos (MIRANDA, 2016), sendo descritos a seguir.

### 3.3.1 Conveniência Geográfica

Neste método, a homogeneidade é determinada a partir da qualidade do ajuste dos modelos de regressão múltipla das vazões com as características físicas e climáticas das sub-bacias.

Segundo Euclides et al. (2014), deve-se optar em avaliar os modelos linear, exponencial, potencial, logarítmico ou recíproco, como também a tendência e a classificação dos resíduos padronizados, além do erro percentual entre as vazões observadas e as estimadas pelo modelo que melhor se ajusta à distribuição das vazões.

É sugerida a aplicação desse método em bacias com rede de monitoramento de vazões que permitam o ajuste satisfatório de modelos de regionalização.

A principal limitação está na dependência de dados de vazão e na complexidade de aplicação, pois impõe a necessidade de se repetir todo o processo quando os resultados obtidos não são satisfatórios, envolvendo um processo de tentativa e erro até que se consigam modelos adequadamente ajustados.

### 3.3.2 Análise da distribuição de frequência

Este método considera que as distribuições de frequências das vazões adimensionalizadas das estações sejam similares em uma RHH (TUCCI, 2002).

Primeiramente, os valores de vazão ( $Q_i$ ) são dispostos em ordem decrescente, no caso das vazões máximas, ou crescente, no caso das mínimas, assinalando-se suas respectivas ordens ( $i$ ).

Em seguida, calculam-se as probabilidades de ocorrência dos eventos, através da função densidade de probabilidade (FDP), considerando que as distribuições de frequência das séries de vazões podem ser representadas por distribuições teóricas de probabilidades, tais como Gumbel (para vazões máximas) ou Weibull (para vazões mínimas).

Após gera-se o gráfico formado pelos pontos representativos dos valores de vazões, anotados em papel probabilístico adequado, obtendo-se uma linha reta. Desta forma, a equação da função de distribuição de frequências das vazões adimensionalizadas pode ser

obtida, por regressão linear simples, pelo método dos mínimos quadrados (FIOREZE, OLIVEIRA, FRANCO, 2008).

Assim, de acordo com este método, as estações que apresentam valores do coeficiente de regressão linear próximos entre si têm distribuições de probabilidades similares e provavelmente compõem uma RHH.

Destaca-se que este método pode incluir erros associados à subjetividade na comparação dos coeficientes nas curvas adimensionais, como também é dependente unicamente da existência de dados de vazão, não sendo aplicável a problemas com escassez ou indisponibilidade de dados de vazão (MIRANDA, 2016).

### 3.3.3 Híbridos

Os métodos híbridos são aqueles que utilizam, em conjunto, mais de um método a fim de dividir uma área de drenagem em RHH (SILVA, 2012).

O estudo de Rao e Srinivas (2006) utilizaram Single-Linkage, Complete-Linkage, Ward e K-means, realizando uma combinação de métodos de agrupamento hierárquicos e particional, obtendo resultados satisfatórios com relação à otimização da função objetivo para a regionalização de vazões.

A análise nesse trabalho mostrou que o agrupamento híbrido pode ser útil na redução do esforço necessário para identificar regiões homogêneas. É recomendado o uso dos algoritmos Ward e K-Means em conjunto, uma vez que ressaltaram que essa combinação proporciona uma flexibilidade suficiente e oferece perspectivas de melhoria em estudos de regionalização.

Já nos estudos de Euclides et al. (2001) foram utilizadas a Conveniência Geográfica e a Análise da Distribuição de Frequência das vazões adimensionalizadas, se mostrando com alta complexidade e subjetividade na identificação de RHH.

Os métodos híbridos têm a vantagem de fornecer resultados consistentes e serem mais versáteis do que os métodos correspondentes aplicados individualmente; contudo, devido a essa característica se tornam, geralmente, mais complexos de serem aplicados (MIRANDA, 2016).

### 3.3.4 Árvore de classificação

Uma árvore de classificação é um modelo lógico representado como uma árvore binária que é dividida em dois caminhos, mostrando como o valor da variável dependente pode ser previsto a partir de valores de uma série de variáveis explanatórias (BARROS; MARTINS; SOUZA FILHO, 2013).

Também chamada de árvore de decisão, possui uma “raiz”, que é o nó inicial a partir do qual todos os caminhos se derivam, um “ramo”, possível caminho na estrutura, e “folhas”, que são os nós intermediários ou finais.

O ponto divisor de maior aproveitamento (menor variância) entre os analisados é visto como o melhor ponto possível para aquela variável no nó em questão. Este processo é então repetido para cada uma das variáveis explanatórias disponíveis e o melhor ponto divisor encontrado entre todas as variáveis é utilizado como o realizador da atual divisão. Continuando o processo de ramificação dos nós, temos a árvore de decisão para os valores das variáveis dependentes de acordo com o intervalo das variáveis explanatórias (SATO et al., 2013).

De posse da árvore de decisão, a predição é feita pela entrada da estrutura da árvore, na raiz, e seguindo os ramos à esquerda ou à direita baseada nos valores das variáveis explanatórias até o nó final (Figura 1).

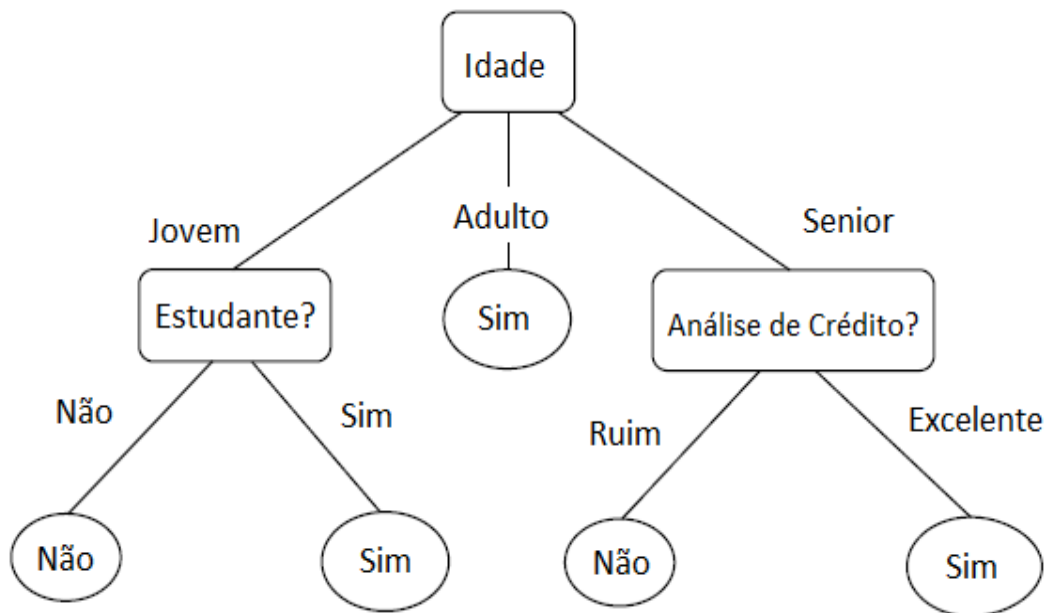


Figura 1 Exemplo de uma árvore de decisão visando na compra de um computador  
Fonte: Han et al. (2011).

### 3.3.5 Análise de correlação canônica

A análise de correlação canônica é um procedimento estatístico multivariado que permite verificar as correlações lineares existentes entre dois grupos ou conjuntos de variáveis.

Tem como objetivo principal explicar a relação entre o conjunto de dados, encontrando um pequeno número de combinações lineares para cada um dos conjuntos de variáveis, de modo a maximizar as correlações possíveis entre os grupos. O primeiro par de

variáveis estatísticas canônicas é obtido de forma a apresentar a maior correlação possível com os grupos de variáveis.

Pode-se obter um segundo par, não correlacionado com o primeiro, que contenha o máximo de informação remanescente que não tenha sido contemplada pela explicação do primeiro par de variáveis, e assim sucessivamente (HAIR JUNIOR et al., 2005; MINGOTI, 2005; FERREIRA, 2008; HAIR JUNIOR et al., 2009).

A correlação canônica simultaneamente prevê múltiplas variáveis dependentes a partir de múltiplas variáveis independentes (HAIR JUNIOR et al., 2009).

### 3.3.6 Entropia

Pode ser interpretada como uma medida de incerteza ou irregularidade de uma variável contida em um evento e, quanto maior a entropia, menos informações está inserida em uma série de observações (YANG; BURN, 1994).

Outra medida de 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 (SINGH, 2013).

A identificação de RHH considerando o conceito de entropia possui relativa vantagem sobre outras técnicas de agrupamento, pois provê 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.

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, Mishra e Singh (2011).

### 3.3.7 Análise de agrupamentos de dados

A análise de agrupamento (cluster analysis), também conhecida como análise de conglomerado, é o nome dado para um grupo de técnicas multivariadas inseridas na etapa de Mineração de Dados, principal etapa do processo KDD (BEM; WAISMANN, 2015).

Esta técnica visa extrair informações, dividindo uma área de estudo em regiões homogêneas, cuja finalidade primária é agregar objetos (estações fluviométricas) com base em uma medida de similaridade e nas características que eles possuem (HAIR; TATHAM; BLACK, 2005).

Realiza-se a segregação dos elementos ou variáveis em grupos homogêneos internamente, heterogêneos entre si e mutuamente exclusivos através das variáveis



utilizadas, que podem ser divididas em estatísticas locais e características locais (NAGHETTINI; PINTO, 2007).

Segundo Nascimento et al. (2013), o conjunto de dados é dividido em dois ou mais grupos, onde os mesmos apresentam uma elevada homogeneidade interna (similar dentro dos agrupamentos) e elevada heterogeneidade externa (entre agrupamentos). Devido à dificuldade de examinar todas as combinações de grupos possíveis em um grande volume de dados, desenvolveram-se diversas técnicas capazes de auxiliar na formação dos agrupamentos.

Geralmente essa ferramenta exploratória compreende cinco etapas: a) seleção das variáveis, b) seleção dos objetos a serem agrupados, c) padronização, d) seleção da medida de distância ou semelhança entre cada par de objetos, e) seleção do algoritmo de agrupamento (método). Aplicando isto, faz-se a escolha da quantidade de agrupamentos formados e a sua interpretação e validação (SOUZA et al., 2015)

Para Calegario (2014), as estatísticas locais são variáveis estimadas a partir dos dados objetos, tais como séries de vazões observadas nas estações fluviométricas. As características locais são as variáveis obtidas a partir de informações pontuais como, por exemplo, a área da bacia.

Segundo o mesmo autor, verificando estudos já realizados, identificaram-se alguns problemas relacionados à análise de cluster, dentre eles: realizar a escolha das variáveis que serão utilizadas para determinar a similaridade das RHH, pois diferentes tipos de combinações destas dão origem a regiões distintas; escolher corretamente o método de agrupamento e as medidas de similaridade disponíveis para aplicar ao estudo; possui a necessidade de determinar pelo índice de validação de agrupamento o número de RHH.

### **3.4 Método de Agrupamento de dados**

Diversas técnicas de análise de agrupamento têm sido amplamente empregadas na área da Hidrologia (FARSADNIA et al., 2014; PEREIRA; KOBIYAMA; CASTRO, 2013; MALEKINEZHAD; NACHTNEBEL; KLIK, 2011; WOLFF, 2013), e, em geral, método algum atende a todos os requisitos de um problema. Por isso é importante entender as características de cada um para a escolha do método adequado a cada tipo de dado (HALDIKI, 2001).

Assim, para saber se a escolha foi correta, após aplicado o pesquisador deve tomar a decisão de qual é a mais adequada ao seu propósito, analisando os elementos de um determinado conjunto onde deverão ser mutuamente similares e, preferencialmente, muito diferentes dos elementos de outros conjuntos, exigindo o conhecimento das propriedades particulares do método aliado aos objetos da pesquisa (HAIR; TATHAM; BLACK, 2005; MUÑOZ DIAS; RODRIGO, 2004).

Há duas grandes categorias que são utilizadas para a formação dos grupos, cada uma delas compreendendo diferentes tipos de algoritmos: os métodos hierárquicos e os não hierárquicos (PESSOA, 2015). Essas técnicas se distinguem, principalmente, pela metodologia utilizada na construção desses grupos.

### **3.5 Métodos Hierárquicos**

Os métodos hierárquicos são técnicas simples em que os dados são particionados sucessivamente, produzindo uma representação hierárquica dos agrupamentos, que podem ser identificados a cada fase. Permitem também a identificação do grau de semelhança entre os mesmos (GUELPELI, 2012).

Cada um dos objetos que participa do processo de análise de agrupamento é considerado como um grupo, que poderá se unir a outros elementos ou grupos de acordo com a medida de dissimilaridade.

Estes métodos são representados por um diagrama bi-dimensional chamado de dendograma ou diagrama de árvore, em que cada ramo representa um elemento e a raiz representa agrupamento de todos os elementos (MIRANDA et al., 2014).

Esse método requer uma matriz contendo as métricas de distância entre os agrupamentos em cada estágio do algoritmo, conhecida como matriz de similaridade entre agrupamentos; contudo, não requer que seja definido um número a priori de agrupamentos.

Segundo Fávero et al. (2009), os métodos hierárquicos são subdivididos em dois grupos: aglomerativos, em que ocorre a fusão de grupos, e os divisivos, em que há a separação dos grupos em cada etapa.

O processo de agrupamento nos métodos aglomerativos inicia com  $n$  grupos unitários formados por cada um dos elementos e, a cada iteração, os dois grupos mais próximos se unem até que todos os indivíduos constituam um único grupo. O algoritmo AGNES (AGglomerative NESTing) e o Ward podem ser citados como exemplo de métodos aglomerativos.

Nos métodos divisivos todos os elementos pertencem ao mesmo grupo e, nas iterações seguintes, os casos mais distantes vão sendo separados, até que existam tantos grupos quantos casos na última iteração. Como exemplo de método divisivo pode ser citado o algoritmo DIANA (DIvisive ANAlysis) (HAIR et al., 2005).

### **3.6 Métodos Não Hierárquicos**

Os métodos não hierárquicos estabelecem a separação de objetos de um conjunto por meio de suas características, que são definidas previamente e de forma arbitrária pelo

usuário, utilizando como princípio uma medida, em que seus resultados não constituem hierarquias.

O número de agrupamentos tem que ser escolhido inicialmente, o que pode gerar interpretações errôneas sobre o resultado caso o número de agrupamentos não seja o esperado pelo usuário, pois o algoritmo é sensível às condições iniciais, gerando resultados diferentes a cada tentativa (FUNG, 2001).

Esses métodos são mais rápidos quando comparados com os métodos hierárquicos porque não é necessário calcular e armazenar, durante o processamento, a matriz de dissimilaridade, sendo capazes de reagrupar os objetos em clusters diferentes daqueles em que foram colocados inicialmente. Outra vantagem é que os resultados não são tão afetados por outliers e podem ser utilizados para grandes conjuntos de dados (QUINTAL, 2006; CASTRO et al., 2010).

A escolha de um método particular de agrupamento depende do conhecimento de suas propriedades, do objetivo da pesquisa e das características do objeto (BUSSAB; MIAZAKI; ANDRADE, 1990; HAIR; TATHAM; BLACK, 2005).

Existem diferentes abordagens sobre as classes dos métodos não hierárquicos (PIMENTEL, 2013), podendo ser divididos em métodos de partição; em modelos; métodos difusos; e métodos de sobreposição. Eles se diferenciam, essencialmente, na forma como ocorre a primeira agregação dos objetos em clusters e no modo como as distâncias entre os centroides dos clusters e os objetos são medidos (QUINTAL, 2006). Nesse tópico, destacaremos os métodos de partição e os difusos.

### 3.6.1 Métodos de Partição

Conhecidos também por particionamento, foram desenvolvidos para agrupar elementos em K grupos, definidos previamente pelo pesquisador, a partir de uma série de fatores. Entretanto, nem todos os grupos possuem valores satisfatórios, havendo a necessidade de se aplicar o método várias vezes para diferentes valores de K, escolhendo os resultados que apresentem melhor interpretação dos grupos formados (DONI, 2004).

O objetivo principal do método é obter a melhor partição dos elementos nos grupos, de modo que a partição satisfaça dois requisitos básicos: coesão interna e separação dos grupos formados (MINGOTI, 2005).

Dentre os métodos de partição, o mais conhecido e utilizado é o algoritmo K-Means (FUNG, 2001), no qual, inicialmente, para a formação de grupos homogêneos selecionam-se X centroides aleatórios (podendo ser definidos pelo analista ou ser usadas as primeiras observações dos dados), em que o usuário deve fornecer ao algoritmo a quantidade de classes que ele deseja encontrar (Equação 1).

$$J = \sum_{j=1}^k \sum_{i=1}^n d(x_i, c_j) \quad \text{Eq.(1)}$$

Em que:

$J$  - função objetivo;

$n$  - número de elementos;

$d(x_i, c_j)$  - distância (a escolher) entre o objeto  $x_i$  e o centroide  $c_j$  do Grupo  $j$ .

Em seguida realiza-se a partição dos  $N$  objetos em  $K$  clusters através dos cálculos dos centroides (Equação 2), fazendo uma comparação entre cada posição dos objetos ao centroide por meio da distância.

$$y_k^j = \frac{1}{|C_j|} \sum_{x_i \in C_j} x_i^w \quad \text{Eq.(2)}$$

Em que:

$y_k^j$  - Função centroide;

$|C_j|$  - é o número de objetos pertencentes ao Grupo  $j$ , em que, assim, se calcula o centroide para a variável  $w$  no objeto  $x_i$ .

Após, agrupam-se esses objetos aos clusters de cujos os centroides se encontram mais próximos. Recalculam-se os centroides até que não haja mais variações significativas na distância mínima de cada objeto a cada um dos centroides dos  $k$  clusters – ou até que o número máximo de iterações, ou o critério de convergência, seja alcançado, sendo isso definido pelo analista.

### 3.6.2 Métodos Difusos

A teoria dos conjuntos difusos parte de uma generalização da teoria clássica dos conjuntos, substituindo a configuração binária por um intervalo de pertinência, de forma que um elemento passa a pertencer a um ou mais conjuntos com um determinado grau de pertinência no intervalo  $[0,1]$ , de tal forma que a soma das pertinências a conjuntos adjacentes equivale a uma unidade (ARAÚJO, 2013).

Os algoritmos do tipo difuso estendem o conceito de associação de cada elemento em um grupo, isto é, um indivíduo pode pertencer a diversas classes de acordo com uma função de pertinência capaz de associar cada padrão a cada um dos clusters, assumindo valores no intervalo  $[0,1]$ . Neste caso, cada classe é um conjunto nebuloso de todos os objetos (PIMENTEL, 2013).

Cada elemento do conjunto de dados possui um grau de pertinência para uma classe da partição, de forma que a soma de todos os graus relativos a esse elemento é necessariamente igual a 1, sendo que, por esse motivo, essa abordagem do algoritmo também pode ser chamada de probabilística. Já a soma de todos os graus de pertinência para um agrupamento estará no intervalo estabelecido entre 0 e o número de elementos ( $n$ ).

Segundo Quintal (2006), interpretação neste método de agrupamento é difícil pois uma parte do cluster pode sobrepor-se parcialmente ou totalmente a outro cluster, exigindo algoritmos complexos.

### **3.7 Algoritmo Fuzzy C-Means (FCM)**

O conceito de fuzzy set foi inicialmente explorado por Zadeh (1965) e posteriormente aplicado a métodos de agrupamento por Ruspini (1969). Mais tarde foi proposto o algoritmo Fuzzy C-Means (FCM) por Dunn (1973) e logo após generalizado por Bezdek (1981).

Atualmente, o método FCM é o mais conhecido da abordagem difusa em que os dados são não-exclusivos e sobrepostos, herdando características do K-Means, como a escolha do parâmetro que determina o número de grupos a ser definido (a priori) e o uso da distância Euclidiana para o cálculo da dissimilaridade entre objetos (PALOSCHI, 2013).

Contudo, a principal diferença entre esses dois métodos é que o FCM define agrupamentos – e não partições, como no K-Means. Além disso, o FCM aloca um elemento ao grupo cujo grau de pertinência seja o maior, sendo amplamente utilizado para as áreas de reconhecimento de padrões (OLIVEIRA, 2016).

O principal objetivo do FCM é avaliar cada dado segundo a sua proximidade com os representantes a cada agrupamento e, quanto mais próximo de um grupo o objeto estiver, maior será seu grau de pertinência para aquela classe, dependendo, assim, dos valores de graus de pertinência obtidos para o agrupamento final.

Os elementos mais próximos das bordas dos grupos possuem menor grau de pertinência, enquanto aqueles mais próximos ao centroide têm maior pertinência, e o centroide é obtido fazendo-se uma média ponderada do grau de todos os indivíduos daquele grupo (FERRARI, 2015).

A função objetivo desse algoritmo (Equação 3) é baseada na soma-de-erros-quadrados (Sum-of-Squared-Errors ou SSE), e, por ser um algoritmo iterativo, novos

centroides e graus de pertinência são calculados a cada iteração (PIMENTEL, 2013), visando sempre minimizar a soma de distâncias entre cada elemento do grupo e o respectivo representante. Quanto menor o valor da função, mais próximos os elementos estarão dos representantes da classe.

$$J = \sum_{j=1}^k \sum_{i=1}^n (u_{ji})^m d(x_i, c_j) \quad \text{q.(3)}$$

Em que:

$J$  - Função objetivo do FCM;

$K$  - número de grupos que se deseja encontrar;

$n$  - número de elementos;

$u_{ij}$  - é o grau de pertinência da amostra, assumindo valores no intervalo [0,1];

$m$  - parâmetro de Fuzzificação;

$d(x_i, c_j)$  - distância euclidiana entre o objeto  $x_i$  e o centroide  $c_j$  do Grupo  $j$ .

Segundo Hall e Mins (1999), o parâmetro  $m$ , conhecido também como índice de fuzzificação, controla o nível de difusividade no processo de classificação, possibilitando que elementos tenham diferentes graus de pertinência entre as classes, isto é, ele regula o grau de sobreposição entre as classes determinando a organização da partição final. Desta forma, para  $m = 1$  os agrupamentos apresentam limites rígidos equivalentes aos do K-means, e à medida que o valor cresce, os limites se tornam mais difusos.

Os representantes, chamados também de protótipos, são importantes para descrever os grupos de forma resumida e influenciam no cálculo do grau de pertinência de objetos. No tradicional K-Means, os protótipos são elementos centrais dos grupos medidos através de média; e no FCM, o parâmetro fuzzyficador  $m$  e os graus de pertinência são importantes para determinar os representantes.

Todos os resultados encontrados nas variáveis da Equação 4 são acumulados, fornecendo, assim, um valor de referência, que deverá ser minimizado a cada iteração do FCM.

Gomes (2014) utilizou o método Fuzzy C-Means para identificação de RHH no estado do Pará, a partir de variáveis explicativas de vazão, como a área de drenagem das bacias, vazão média, precipitação, comprimento e desnível do rio e tipo de solo, sendo definidas quatro regiões homogêneas de vazão e, segundo a autora, essas regiões poderão auxiliar nos estudos de regionalização hidrológica, reduzindo o tempo de análise de dados e

aumentando a confiabilidade dos resultados obtidos, além de fornecer subsídios ao Planejamento dos Recursos Hídricos no território paraense.

Já o estudo de Vargas et al. (2015) avaliou as técnicas K-means, Partitioned Around Medoids (PAM), K-Harmonic Means (KHM), Genetic K-Means (GKA) e Fuzzy C-Means (FCM), na determinação de RHH, visando à regionalização da  $Q_{90}$ , em que os métodos FCM, baseado em lógica fuzzy, e GKA, baseado em algoritmo genético, obtiveram o melhor desempenho em relação aos demais métodos testados, na estimativa da  $Q_{90}$  em locais desprovidos de monitoramento.

### 3.8 Medidas de Similaridade

A maioria dos métodos de análise de cluster requer uma medida de distância a escolher de acordo com o domínio do problema e os elementos a serem agrupados, operando com o conceito de distância ou métrica (ALBUQUERQUE et al., 2016).

Esse coeficiente possibilita a comparação objetiva entre variáveis do conjunto analisado, tornando possível verificar se um elemento A é mais parecido com B do que com C. Com base nessa medida, os indivíduos similares são agrupados e os demais são colocados em grupos separados (AAKER et al., 2001).

No que diz respeito à estimação dessa medida entre os indivíduos ou populações a serem agrupados, há dois tipos de coeficientes: similaridade e dissimilaridade, que variam conforme o tipo de variável estudada (quantitativa ou qualitativa).

Na medida de similaridade, quanto maior for o valor observado, mais parecidos serão os objetos. Já na medida de dissimilaridade, quanto maior for o valor observado, menos parecidos serão os objetos (MEYER, 2002).

A avaliação dessa medida entre objetos pode ser realizada por três maneiras (VIDAL, 2013): a) Medidas de distância: medem a distância ou a separação angular entre pares de pontos; b) Medidas correlacionais: os que medem a correlação entre pares de valores; c) Medidas de associação: os que medem a associação entre pares de caracteres qualitativos, baseando-se em tabelas de contingência.

As medidas baseadas em distância são os métodos mais frequentemente usados (HAIR; TATHAM; BLACK, 2005), sendo consideradas coeficientes de dissimilaridades representadas na forma de matriz, pois, objetos que possuem a maior distância entre si são menos semelhantes, um do outro, do que os objetos com a menor distância, e vice-versa (FÁVERO et al., 2009).

Kumar, Chhabra e Kumar (2014) descrevem uma série de medidas possíveis que podem ser utilizadas como medidas de dissimilaridade: Distância Euclidiana, Distância de Minkowski, Distância Absoluta (Bloco, City-Block ou Manhattan), Mahalanobis e Chebychev.

A medida utilizada neste estudo foi a distância euclidiana, um dos coeficientes mais conhecidos e frequentemente empregados quando todas as variáveis são quantitativas (PEREIRA; KOBIYAMA; CASTRO, 2013).

### 3.8.1 Distância Euclidiana

A distância euclidiana é interpretada como a distância entre dois indivíduos, cujas posições são determinadas em relação às suas coordenadas, definidas com referências a um grupo de eixos cartesianos, os quais possuem ângulos retos entes si.

Para Prado (2015), é a medida de distância mais utilizada para a análise de agrupamentos na qual é obtida mediante o teorema de Pitágoras para um espaço multidimensional.

A distância euclidiana entre duas observações (i e j) corresponde à raiz quadrada da soma dos quadrados das diferenças entre os pares de observações (i e j) para todas as p variáveis, conforme a Equação 4 (BUSSAB; MIAZAKI; ANDRADE, 1990).

$$d_{ij} = \sqrt{\sum_{k=1}^p (X_{ik} - X_{jk})^2} \quad \text{q.(4)}$$

Em que:

$d_{ij}$  - distância entre os elementos i e j;

$X_{ik}$  - valor assumido pela variável k do elemento i,

$X_{jk}$  - valor assumido pela variável k do elemento j.

### 3.9 Análise de Componentes Principais – ACP

É uma técnica multivariada usada para transformar um conjunto de variáveis observadas relacionadas (ou correlacionadas) numa nova série de variáveis que não se correlacionam e que sejam arranjadas em ordem decrescente de importância.

O objetivo principal do método é tentar reduzir a dimensionalidade do problema com a menor perda possível de informação, para encontrar as variáveis com maior significância, que expliquem a maior proporção possível da variabilidade total (KUCHLA, 2015; JOLLIFFE, 2002).

Originalmente a ACP foi introduzida por Pearson em 1901 e por Hotelling em 1933 (EVERITT; DER, 2009), sendo que transforma um sistema de eixos originais em um novo sistema que tem duas importantes propriedades: os novos eixos principais são ortogonais,



isto é, independentes um do outro, e as componentes principais decrescem em ordem, de modo que, a primeira variável absorverá a maior proporção possível da variação total e a segunda, absorverá a maior proporção da variação remanescente, e, assim sucessivamente (KERNKAMP et al., 2016).

Tem como vantagem de uso em avaliar a importância de cada característica estudada sobre a variação total, possibilitando o descarte das variáveis menos significativas, por estarem correlacionadas a outras variáveis, pela sua invariância ou por serem combinação linear de outras características (HONGYU; SANDANIELO; JUNIOR, 2015).

Contudo, é necessário observar se a matriz de dados é passível de fatoração, isto é, analisar se os dados podem ser submetidos ao processo de análise fatorial (MANSANO; PARRÉ; PEREIRA, 2017). Para isso, dois métodos de avaliação são mais comumente utilizados, a saber: o critério de Kaiser-Meyer-Olkin (KMO); e o Teste de Esfericidade de Bartlett.

O índice de KMO, também conhecido como índice de adequação da amostra, é um teste estatístico que sugere a proporção de variância dos itens que pode estar sendo explicada por uma variável latente (LORENZO-SEVA; TIMMERMAN; KIERS, 2011). Tal índice indica o quão adequado é a aplicação da análise fatorial para o conjunto de dados (HAIR; TATHAM; BLACK, 2005).

O KMO é calculado por meio do quadrado das correlações totais dividido pelo quadrado das correlações parciais, das variáveis analisadas (FIELD, 2005). Seu valor pode variar de zero a um. Valores iguais ou próximos a zero indicam que a soma das correlações parciais dos itens avaliados é bastante alta em relação à soma das correlações totais. Nesses casos, possivelmente a análise fatorial será inapropriada (PASQUALI, 1999).

O teste de esfericidade de Bartlett, por sua vez, avalia em que medida a matriz de (co)variância é similar a uma matriz-identidade (os elementos da diagonal principal têm valor igual a um), e os demais elementos da matriz são aproximadamente zero, ou seja, não apresentam correlações entre si (FIELD, 2005).

Segundo Hair, Tatham e Black (2005), esse teste avalia, também, a significância geral de todas as correlações em uma matriz de dados. Valores do teste de esfericidade de Bartlett com níveis de significância  $p < 0,05$  indicam que a matriz é fatorável (TABACHNICK; FIDELL, 2007), rejeitando a hipótese nula de que a matriz de dados é similar a uma matriz-identidade. Em geral, os resultados dos testes de KMO e de esfericidade de Bartlett tendem a ser uniformes, aceitando ou negando a possibilidade de fatoração da matriz de dados (DZIUBAN; SHIRKEY, 1974).

Existem na literatura vários critérios para escolha do número adequado de Componentes Principais (PC's) que representem fielmente a variância presente nos dados originais. Os critérios do número adequado de PC's para resumir o conjunto de dados

utilizados nesta pesquisa foram: descartar os PCs cujas cargas tenham variância explicada inferior a 50% (O'LENIC; LIVEZEY, 1998); excluir os PC's cujos autovalores associados são inferiores a um, de acordo com o critério de Kaiser ou critério mínimo (GARAYALDE et al., 1986).

Minucci (2014) utilizou a ACP para determinar quais estações fluviométricas seriam usadas em bacias que contemplavam com mais de 4 unidades das mesmas. Foi optado por selecionar as variâncias superiores a 90%.

A técnica da ACP também foi utilizada por Eastman e Fulk (1993) para analisar as condições ambientais e relaciona-los com eventos periódicos e não periódicos na África. Já Ceballos & Braga (1995) aplicaram a ACP em regiões homogêneas de radiação global diária para estimar valores em locais onde não se disponham de dados. Fragoso e Gomes (2008) aplicaram a ACP a dados de chuva para identificar padrões de precipitação associados aos eventos intensos de chuva no sul do Portugal, dentre outros.

Pereira, Braga e Paz (2010) aplicaram a ACP para determinar os padrões espaciais e correlações da precipitação do estado de Goiás, enquanto Mendonça et al., (2017) utilizou a mesma técnica multivariada no estudo da distribuição da precipitação no Rio Grande do Sul a partir de registros de chuvas.

#### 4 MATERIAL E MÉTODOS

Para alcançar os objetivos propostos, foram necessárias etapas contínuas que caracterizam o desenvolvimento metodológico. O desenvolvimento de cada uma dessas etapas e os aplicativos computacionais utilizados em cada uma delas podem ser visualizados na Figura 2. Nos itens subsequentes será apresentada a metodologia aplicada em cada etapa.

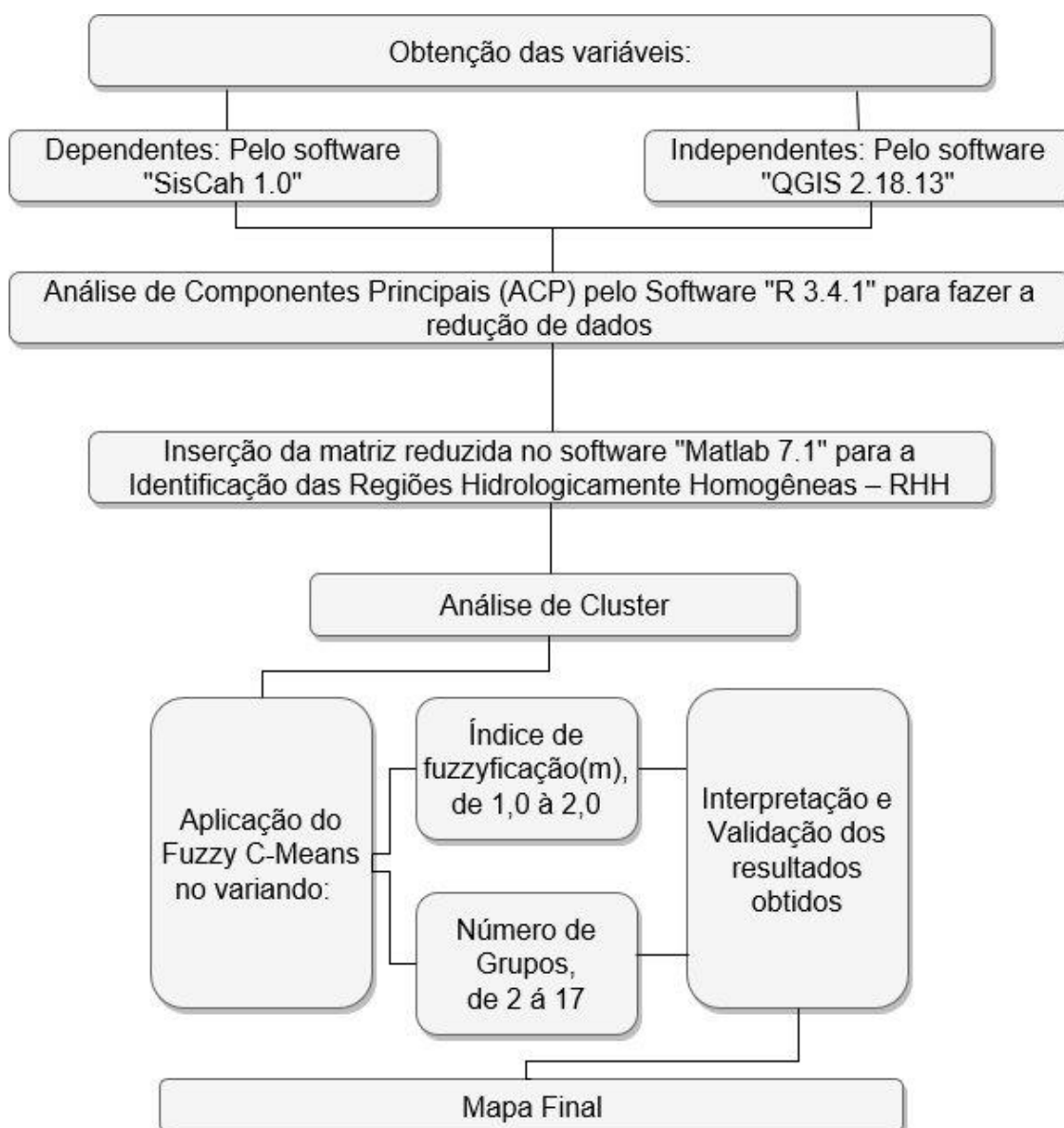


Figura 2 Fluxograma metodológico

#### 4.1 Descrição da área de estudo

O Paraná possui 399 municípios e ocupa uma área de 199.880 km<sup>2</sup> (PARANÁ, 2006). É um dos estados de maior altitude do Brasil: 52% do território fica acima dos 600 m, formado em sua maior parte por planaltos, sendo os primeiro, segundo e terceiro planaltos paranaenses os principais da região, e apenas 3% abaixo dos 200m, formado por algumas planícies, como a baixada litorânea (SANTOS et al., 2006).

Em decorrência de sua localização e extensão, as terras paranaenses situam-se sob a Zona do Trópico, ocupando uma posição de transição entre regiões tropicais e subtropicais, sendo caracterizando como um território de climas variados e de influências físicas diversas. No estado, se destacam quatro massas de ar atuantes: massa Equatorial Continental, massa Tropical Continental, massa Tropical Atlântica e massa Polar Atlântica, as quais fazem as precipitações pluviométricas terem distribuição espacial com variação de 1.100 a 1.920 mm por ano (NIMER, 1979; APARECIDO et al., 2016).

Esses aspectos climáticos são decorrentes da forte influência da compartimentação geomorfológica local que interfere na dinâmica dos subsistemas como: clima, relevo, hidrografia, solos, vegetação, altitude e latitude, propiciando diferentes culturas agrícolas e núcleos populacionais (SILVA et al., 2015).

##### 4.1.1 Hidrografia

A hidrografia do Paraná é constituída de rios que fluem diretamente para o litoral, e os que fluem para a região oeste, tributários do Rio Paraná, possuindo como características serem mais extensos. Entre estes últimos estão o Rio Paranapanema, que faz limite com o estado de São Paulo, e o Rio Iguaçu, que faz, em parte, limite com Santa Catarina e Argentina. O Rio Paraná faz a divisa com o estado do Mato Grosso do Sul e a fronteira com o Paraguai.

O Paraná possui 16 bacias hidrográficas instituídas através da Resolução n.º 024/2006/Sema, sendo elas: Bacia Litorânea, Bacia do Ribeira, Bacia do Cinza, Bacia do Iguaçu, Bacias do Paraná 1, 2 e 3, Bacias do Tibagi, Bacia do Ivaí, Bacia do Piriquire, Bacia do Pirapó, Bacia do Itararé, Bacias do Paranapanema 1, 2, 3 e 4 (SEMA, 2010).

Para o gerenciamento de recursos hídricos no estado do Paraná, foram definidas 12 unidades hidrográficas que comportam as 16 bacias hidrográficas do estado. A abrangência dessas unidades pode ser a bacia hidrográfica na sua totalidade ou parte dela, visando promover o planejamento territorial dos recursos hídricos, sendo que essas unidades representam a área de atuação dos comitês de bacia hidrográfica (SUDERHSA, 2006).

## 4.2 Dados Utilizados

As séries históricas de vazão foram obtidas da Rede Hidrometeorológica Nacional (HidroWeb) da Agência Nacional de Águas (ANA).

Cada estação apresenta um período de funcionamento diferente, tornando necessária a organização dos arquivos individualmente, aplicando como critério de escolha apenas as estações em funcionamento até o momento da pesquisa, como também as com os dados das variáveis disponíveis, totalizando 114 estações fluviométricas distribuídas pelo estado do Paraná (Figura 3).

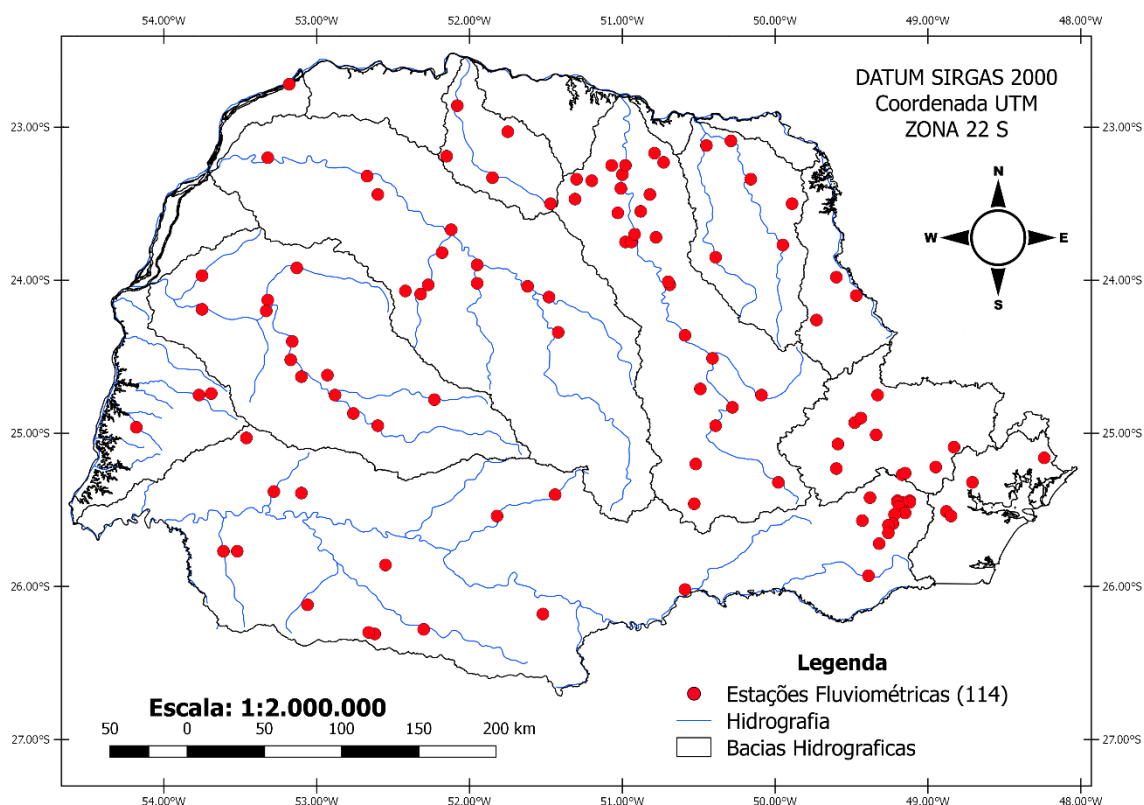


Figura 3 Mapa com as 114 estações fluviométricas estudadas nas respectivas bacias hidrográficas

Fonte: Rede de Drenagem ANA (2011).

Seguindo as recomendações da ANA (2011), foram eliminados os meses que apresentavam mais de 5% de registros sem informação e as demais falhas não sendo preenchidas.

As informações como código, coordenadas e bacias hidrográficas a que pertence cada estação são apresentadas no Apêndice A.

#### 4.2.1 Obtenção das variáveis dependentes

As variáveis escolhidas foram: vazão média anual de longa duração ( $Q_{mid}$ ) - ( $m^3/s$ ), vazão mínima anual com sete dias de duração e período de retorno de 10 anos ( $Q_{7,10}$ ) - ( $m^3/s$ ), vazões associadas às permanências de 95% ( $Q_{95}$ ) e 90% ( $Q_{90}$ ) - ( $m^3/s$ ), calculadas com apoio do Sistema Computacional para Análise Hidrológica (“SisCAH 1.0”) (SOUSA et al., 2009).

O software SisCAH é estruturado em módulos, capazes de realizar a importação de arquivos de séries históricas de vazão; pré-processamento de séries; preenchimento e extensão de séries; obtenção da curva de permanência de vazões; obtenção da curva de regularização de vazões e estimativa de vazões mínimas, médias e máximas.

##### 4.2.1.1 Vazão média anual de longa duração ( $Q_{mid}$ )

A  $Q_{mid}$  é definida como a média de todas as vazões diárias da série histórica (IGAM, 2012).

##### 4.2.1.2 Vazões de permanência de 90 e 95% ( $Q_{90}$ e $Q_{95}$ )

As vazões com permanência de 90 e 95% foram obtidas com auxílio da curva de permanência para cada série histórica. As curvas foram elaboradas a partir da divisão das vazões em classes e, posteriormente, as frequências de cada classe foram contabilizadas.

Para cada série foram definidos 50 intervalos de classe, baseados na escala logarítmica, devido à grande variação de magnitude das vazões. O intervalo de cada classe ( $\Delta X$ ) foi obtido pela Equação 5.

$$\Delta X = \frac{[\ln(Q_{m\acute{a}x}) - \ln(Q_{m\acute{i}n})]}{50} \quad \text{q.(5)}$$

Em que:

$Q_{m\acute{a}x}$  - vazão máxima da série;

$Q_{m\acute{i}n}$  - vazão mínima da série.

A vazão mínima da série é o limite inferior da primeira classe. O limite superior é definido pela Equação 6. Uma vez definido o limite superior da primeira classe, ele passa a ser o limite inferior da segunda classe, e assim sucessivamente.

$$Q_{i+1} = \exp [\ln(Q_i) + \Delta X] \quad \text{q.(6)}$$

Em que:

$Q_{i+1}$  - vazão do limite superior do intervalo  $i$ ;

$Q_i$  - vazão do intervalo  $i$ ;

$\Delta X$  - intervalo de classe.

A frequência associada a cada intervalo de classe é calculada pela Equação 7:

$$f_i = \left( \frac{Nq_i}{NT} \right) 100 \quad \text{q.(7)}$$

Em que:

$f_i$  - frequência;

$Nq_i$  - número de vazões de cada intervalo;

$NT$  - número total de vazões.

A curva de permanência pode ser visualizada graficamente plotando-se a frequência (no eixo das abcissas) pela vazão (no eixo das ordenadas).

#### 4.2.1.3 Vazão mínima com sete dias de duração

A vazão mínima com duração de sete dias e período de retorno de dez anos ( $Q_{7,10}$ ) foi obtida com auxílio da análise de frequência baseada em distribuições de probabilidade.

Na primeira etapa foram elaboradas as séries anuais de vazões mínimas com sete dias de duração ( $Q_7$ ) para cada série. Em seguida, visando estimar as vazões mínimas, empregaram-se os ajustes nas distribuições de probabilidade Log-Pearson 3, Pearson 3, Log-Normal 2, Log-Normal 3 e Weibull.

Depois do ajuste das distribuições de probabilidade e escolhida a mesma pelo menor erro padrão, foi realizada a estimativa da  $Q_7$  associada ao período de retorno de dez anos, resultando, assim, no índice  $Q_{7,10}$ . Esta estimativa foi feita com a aplicação do fator de frequência, conforme a Equação 8.

$$M = \mu + k \cdot \sigma \quad \text{Eq.(8)}$$

Em que:

$M$  - magnitude do evento para o período de retorno estabelecido;

$\mu$  - média dos eventos;

$k$  - fator de frequência;

$\sigma$  - desvio padrão dos eventos.

A metodologia utilizada para o cálculo do fator de frequência foi aquela apresentada por Kite (1988). Por fim, adotou-se a  $Q_{7,10}$  obtida pela distribuição de probabilidade que resultou no menor intervalo de confiança.

#### 4.2.2 Obtenção das variáveis independentes

Para a obtenção das características físicas da área de estudo, utilizou-se o Modelo Digital de Elevação (MDE) do estado do Paraná, disponibilizado por Valeriano (2004), em escala 1:250.000 e em resolução espacial de 30 metros.

As características físicas escolhidas das estações fluviométricas foram: área de drenagem (AD – m<sup>2</sup>), soma das drenagens (SD – m), densidade de drenagem (DD – 1/m) e a localização geográfica (latitude - Lat e longitude - Long). Todas essas variáveis foram obtidas por meio do software “QGIS 2.18.13”.

### 4.3 Análise de Componentes Principais – ACP

A Análise de Componentes Principais (ACP) foi realizada com apoio do software estatístico “R 3.4.1”. O script é apresentado no Apêndice B.

A ACP foi aplicada com o objetivo de analisar o comportamento dos dados e selecionar as variáveis com maior capacidade para explicar a variação das vazões. Com esta técnica se faz uma redução no número de variáveis a serem usadas modelagem dos fenômenos de interesse.

A verificação da adequação da amostra à análise fatorial foi avaliada pelo Teste de esfericidade de Bartlett e pelo índice de adequação da amostra de Kaiser-Meyer-Olkin (KMO).

Para o teste de KMO, os resultados podem variar entre 0 e 1. Quanto mais perto de 1, melhor. Os valores do índice KMO que indicam que a análise fatorial é apropriada varia de autor para autor. Para Hair Jr. et al., (2009) são valores aceitáveis entre 0,5 a 1,0. Pallant (2007) sugere 0,6 como um limite razoável.

Como regra para interpretação do índice de KMO nesse trabalho, seguiu-se o apresentado por Hutcheson e Sofroniou (1999), onde valores menores que 0,5 são considerados inaceitáveis, valores entre 0,5 e 0,7 são considerados medíocres; valores entre 0,7 e 0,8 são considerados bons; valores maiores que 0,8 e 0,9 são considerados ótimos e excelentes, respectivamente.

Os principais passos considerados foram: a) Análise para a escolha das variáveis a serem incluídas na matriz de dados original e se as mesmas necessitavam ser transformadas; b) O cálculo e análise da matriz de correlação (ou covariância), avaliando o coeficiente de correlação para um par de variáveis e verificando se na matriz de dados



originais havia dados correlacionados; c) Obtenção das Componentes Principais (PC) através das combinações lineares das variáveis originais descritas pelos autovalores; d) Escolha e análise dos PC's segundo o critério de Kaiser-Guttman.

Segundo Minucci (2014) existem duas interpretações diferentes que podem ser feitas ao estudar os resultados da ACP. Uma delas é analisar as variáveis obtidas e trabalhar com os PC's no novo sistema de coordenadas criado, esse método faz com que a nova amostra não tenha mais significado físico, mas sim matemático. O outro método consiste em avaliar quais são os indivíduos que, de fato, têm influência sobre o problema analisado, sendo essa interpretação usada na pesquisa.

Para isso, ordenou-se os autovalores e autovetores correspondentes em ordem decrescente de significância. A partir de então, buscou-se nos autovetores a linha  $n$  que detém o maior valor absoluto dentre todas as linhas do mesmo componente principal, essa linha  $n$  está relacionada ao  $n$ -ésimo indivíduo analisado e, portanto, esse indivíduo possui significância na pesquisa.

#### **4.4 Identificação das Regiões Hidrologicamente Homogêneas**

Nesse trabalho foi utilizado o processo chamado de Descoberta de Conhecimento em Bancos de Dados (Knowledge Discovery in Databases – KDD) para transformar os dados da série histórica das estações fluviométricas, que podem conter padrões e tendências, em conhecimento novo e útil (FAYYAD et al., 1996; SKUPIEN; CARVALHO, 2012).

O processo KDD pressupõe as seguintes etapas: Seleção dos dados; Pré-Processamento dos dados; Transformação; Mineração de Dados (Data Mining); Interpretação e Validação dos resultados (WAGNER; RIBEIRO; ZABADAL, 2016).

A Seleção e Pré-Processamento dos dados já foram descritas no item 4.2, sendo as demais etapas apresentadas a seguir.

##### **4.4.1 Transformação e Normalização**

A transformação dos dados consiste na padronização dos valores de entrada, evitando que grandes variações atrapalhem o treinamento e o aprendizado do algoritmo (MODENESI, 2008).

Contudo, o algoritmo escolhido, Fuzzy C-Means, é capaz de lidar com atributos sem a prévia transformação, como também sem a normalização, e como foram usados os dados no formato padrão e na mesma ordem de grandeza, não houve a necessidade de realizar a normalização.

#### 4.4.2 Mineração de Dados (Data Mining)

O método de agrupamento difuso utilizado foi o Fuzzy C-Means com o auxílio de ferramenta específica, implementada no software “MATLAB 7.1” por intermédio do “Fuzzy Logic Toolbox”.

O processo de agrupamento, pode ser resumido em duas etapas: estimativa da medida de similaridade entre as estações fluviométricas a serem agrupadas; e aplicação da técnica de agrupamento para a formação de grupos.

Para avaliar a similaridade entre as estações fluviométricas a serem agrupadas, organizou-se uma matriz de dados  $X_{(n \times p)}$ , cujas linhas correspondem a “n” unidades (estações fluviométricas), fornecendo “p” características (variáveis), ou seja, a primeira coluna seguiu com o número de identificação dos postos e as demais colunas com os valores de cada variável.

As similaridades entre os objetos descritos por variáveis escaladas em intervalos foram computadas com base na distância euclidiana entre cada par de estações.

Foram testados os valores de índice de fuzzificação ( $m$ ) variando entre 1,0 e 2,0, número máximo de iterações ( $t_{max}$ ) = 200 e número de grupos (Cluster), de 2 a 17, em razão de não saber quantos grupos homogêneos podem existir em todo o estado do Paraná. Como critério de parada, foi utilizado o erro mínimo de  $\epsilon = 0,0001$ , conforme indicado no trabalho de Pessoa (2015).

#### 4.4.3 Interpretação e Validação

Após a aplicação da metodologia FCM foram obtidos como resposta do processamento do algoritmo o valor da função objetivo e o número de interações, escolhendo-se a combinação Cluster e  $m$  com menor valor da função objetivo.

Após escolhido estes, obtiveram-se os graus de pertinência de cada elemento em cada cluster, sendo que para a formação dos grupos, foi utilizado o método de decisão pelo maior grau de pertinência, ou seja, para um dado objeto o maior grau de pertinência determinou qual agrupamento o objeto pertenceria.

Como não há uma base formal para determinar o grau de pertinência para tais situações, este foi escolhido empiricamente, sendo definida uma classificação para analisar os resultados.

O objeto que possuiu seu maior grau de pertinência acima de 0,7 foi considerado não confuso, mas se esse valor foi menor 0,7 então foi classificado como confuso ou nebuloso.

## **5 RESULTADOS E DISCUSSÃO**

Nesta seção discutiremos os resultados obtidos utilizando a análise multivariada, como também das configurações aplicadas sobre o algoritmo Fuzzy C-Means, descrevendo as particularidades encontradas em cada combinação.

Por fim, a descrição dos comportamentos de cada região hidrológica no estado do Paraná, obtidas através da configuração escolhida.

### **5.1 Variáveis dependentes**

A distribuição de probabilidade que se ajustou em 100% dos casos das  $Q_{7,10}$  e com o menor erro padrão foi a distribuição Log-Pearson tipo III. Assim, organizaram-se os dados em matriz em que as linhas representam as estações fluviométricas, e as colunas, as variáveis dependentes estudadas (Apêndice C).

### **5.2 Variáveis independentes**

Para as variáveis independentes organizaram-se os dados em matriz em que as linhas representam as estações fluviométricas, e as colunas, as variáveis estudadas (Apêndice D).

### **5.3 Aplicação da ACP aos dados dependentes e independentes**

Servindo-se da linguagem R e de pacotes específicos para exploração estatística, foi aplicada a Análise Fatorial em Componentes Principais (PC) nas variáveis dependentes e independentes para conhecer o comportamento dos dados e identificar quais eram mais significativos em relação à identificação das RHH (Tabela 1).

Tabela 1 Matriz de Correlação Simples entre as variáveis em estudo

VARIÁVEIS	AD	SD	DD	Q <sub>MLD</sub>	Q <sub>7,10</sub>	Q <sub>90</sub>	Q <sub>95</sub>	LAT	LONG
AD	1	0,99	0,21	0,8	0,93	0,97	0,96	0,3	-0,25
SD	0,99	1	0,25	0,81	0,92	0,96	0,95	0,32	-0,28
DD	0,21	0,25	1	0,22	0,15	0,18	0,17	0,21	-0,18
Q <sub>MLD</sub>	0,8	0,81	0,22	1	0,71	0,78	0,76	0,23	-0,24
Q <sub>7,10</sub>	0,93	0,92	0,15	0,71	1	0,98	0,99	0,28	-0,27
Q <sub>90</sub>	0,97	0,96	0,18	0,78	0,98	1	0,99	0,28	-0,27
Q <sub>95</sub>	0,96	0,95	0,17	0,76	0,99	0,99	1	0,28	-0,27
LAT	0,3	0,32	0,21	0,23	0,28	0,28	0,28	1	-0,2
LONG	-0,25	-0,28	-0,18	-0,24	-0,27	-0,27	-0,27	-0,2	1

AD = Área de Drenagem; SD = Soma das drenagens; DD: Densidade de Drenagem; Q<sub>mlD</sub> = Vazão média anual de longa duração; Q<sub>7,10</sub> = Vazão mínima anual com sete dias de duração e período de retorno de 10 anos; Q<sub>90</sub> e Q<sub>95</sub> = Vazões associadas às permanências de 90% e 95%; Lat = Latitude e Long = Longitude.

Com a matriz das potenciais correlações entre duas variáveis entre todas as variáveis envolvidas, foi possível identificar que AD, SD, Q<sub>mlD</sub>, Q<sub>7,10</sub>, Q<sub>90</sub> e Q<sub>95</sub> estão muito bem associadas entre si, considerando valores das correlações acima de 0,90. Já DD, Lat e Long não apresentam correlação com as demais variáveis.

Utilizando um nível de significância de 5% (BRAGA et al., 2012), foi rejeitado H<sub>0</sub> do teste de esfericidade de Bartlett, pois o p-valor encontrado no teste foi de 0,0001; corroborando a existência de correlação entre as variáveis, se mostrando significativo.

Já para o teste de KMO, os resultados podem variar entre 0 e 1. Quanto mais perto de 1, melhor. Os valores do índice KMO que indicam que a análise fatorial é apropriada varia de autor para autor. Para Hair Jr. et al. (2009) são valores aceitáveis entre 0,5 a 1,0. Pallant (2007) sugere 0,6 como um limite razoável.

Como regra para interpretação do índice de KMO nesse trabalho, seguiu-se o apresentado por Hutcheson e Sofroniou (1999), em que valores menores que 0,5 são considerados inaceitáveis, valores entre 0,5 e 0,7 são considerados medíocres; valores entre 0,7 e 0,8 são considerados bons; valores maiores que 0,8 e 0,9 são considerados ótimos e excelentes, respectivamente.

O KMO encontrado nesse trabalho foi 0,78, se enquadrando, portanto, como bom na classificação adotada.

Os resultados dos dois testes demonstram que as correlações parciais foram bastante satisfatórias permitindo dar prosseguimento a análise fatorial. Diante da

interpretação dos resultados do método de ACP para cada estação e para cada variável foi baseada na porcentagem da variação explicada para cada Componente Principal (PC) (Tabela 2).

Tabela 2 Variância explicada e cumulativa ao longo das componentes principais

	<i>PC1</i>	<i>PC2</i>	<i>PC3</i>	<i>PC4</i>	<i>PC5</i>	<i>PC6</i>	<i>PC7</i>	<i>PC8</i>	<i>PC9</i>
<b>1</b>	24.078	10.707	0.90731	0.88376	0.5894	0.2969	0.10284	0.07338	0.02053
<b>2</b>	0.6442	0.1274	0.09147	0.08678	0.0386	0.0098	0.00118	0.00060	0.00005
<b>3</b>	0.6442	0.7715	0.86300	0.94978	0.9884	0.9982	0.99935	0.99995	100.000
<b>4</b>	2.4077	1.0706	0.9073	0.8837	0.5894	0.2969	0.1028	0.0733	0.0205

1 – Desvio Padrão, 2 – Porcentagem da variância explicada pelas componentes, 3 – Variância explicada acumulada, 4 - Autovalor.

Segundo Lopes et al. (2010) espera-se, a princípio, que grande parte da variância (informação) dos dados seja explicada por um número pequeno de PC's. Analisando a variância de todos as PC's, o PC1 obteve a maior variância que os demais, sendo responsável por 64% da variância total explicada em relação aos demais componentes.

Dentre os critérios geralmente utilizados para a escolha de quais PC's que irá se utilizar, merece destaque a variância total explicada de cada componente (MINUCCI, 2014).

Além desse critério, também foi aplicado o de Kaiser-Guttman, mais conhecido como autovalor ou eigenvalue maior que 1 (PATIL et al., 2008). Ele propõe uma avaliação rápida e objetiva do número de PC a ser retido. A lógica por trás do critério de Kaiser-Guttman é simples: cada PC apresenta um autovalor que se refere ao total de variância explicada por ele.

Como o objetivo das análises fatoriais é reduzir um determinado número de variáveis observadas em um número menor de componentes, apenas os PC's com autovalor maior 1 foram analisados (FLOYD; WIDAMAN, 1995).

Diante da regra de Kaiser-Guttman, apenas dois PC's obtiveram autovalores maiores que 1, sendo eles o PC1 (2,4077) e o PC2 (1,0706). Minucci (2014) informa que não existe uma regra final definida que determine o número de componentes que devem ser selecionados para a explicação de um fenômeno.

O mesmo autor ainda afirma que dentre os critérios geralmente utilizados merece destaque aquele que estabelece a permanência das n primeiras componentes ( $n < p$ ), de acordo com a variação cumulativa que elas explicam. Uma variação cumulativa com um bom grau de explicação é superior a 75%, porém, fica a cargo do usuário julgar a necessidade de variação a ser explicada.

Diante disso, escolheu-se apenas o PC1 por se destacar nos dois critérios estabelecidos, variância total explicada e Kaiser-Guttman.

Desta forma, analisaram-se as linhas com os autovetores (Tabela 3) onde tais carregamentos podem ser considerados como uma medida da relativa importância de cada variável em relação às componentes principais e os respectivos sinais, se positivos ou negativos, indicam relações diretamente e inversamente proporcionais (BICUDO; BICUDO, 2004).

Tabela 3 Autovetores das variáveis do Componente escolhido

VARIÁVEIS	PC1
<b>AD</b>	0,4068
<b>SD</b>	0,4074
<b>DD</b>	0,1108
<b>Q<sub>MLD</sub></b>	0,3584
<b>Q<sub>7,10</sub></b>	0,3977
<b>Q<sub>90</sub></b>	0,4087
<b>Q<sub>95</sub></b>	0,4063
<b>LAT</b>	0,1543
<b>LONG</b>	0,1433

Para a escolha das variáveis significativas considerou-se apenas aquelas que obtiveram peso acima do maior valor decimal encontrado, ou seja, 0,40. Assim, analisando os autovetores obtidos do PC1 verificou-se que as variáveis escolhidas que detêm o maior valor absoluto dentre todas as linhas do mesmo componente principal foram: AD, SD, Q<sub>90</sub> e Q<sub>95</sub>.

Visto se tratar de uma variável importante para o estado do Paraná, já que é utilizada como referência no estabelecimento das vazões outorgáveis no estado, e por seu autovetor ser próximo ao valor critério, a Q<sub>7,10</sub> foi incluída dentre as variáveis escolhidas.

As variáveis DD, Q<sub>mid</sub>, Lat e Long apresentaram autovetores abaixo do critério estabelecido, desta forma, foram retiradas da matriz de dados por apresentarem baixo grau de significância.

Já era esperada a exclusão das variáveis Lat e Long do conjunto de dados por possuírem pouca representatividade em relação ao comportamento das vazões por serem características físicas e pontuais do local.

Calegario (2014) considerou as variáveis latitude e longitude na análise de cluster, contudo, isso resultou em maior convergência dos índices de validação de agrupamento em seu estudo.

Para ratificar a seleção das principais variáveis com relevância positiva, utilizou-se o biplot (Figura 4), em que os comprimentos destes vetores são proporcionais às variâncias das variáveis e os ângulos entre os vetores estão relacionados às correlações entre as variáveis, sendo que quanto menor o ângulo, mais correlacionadas estão (KLEFENS, 2009).

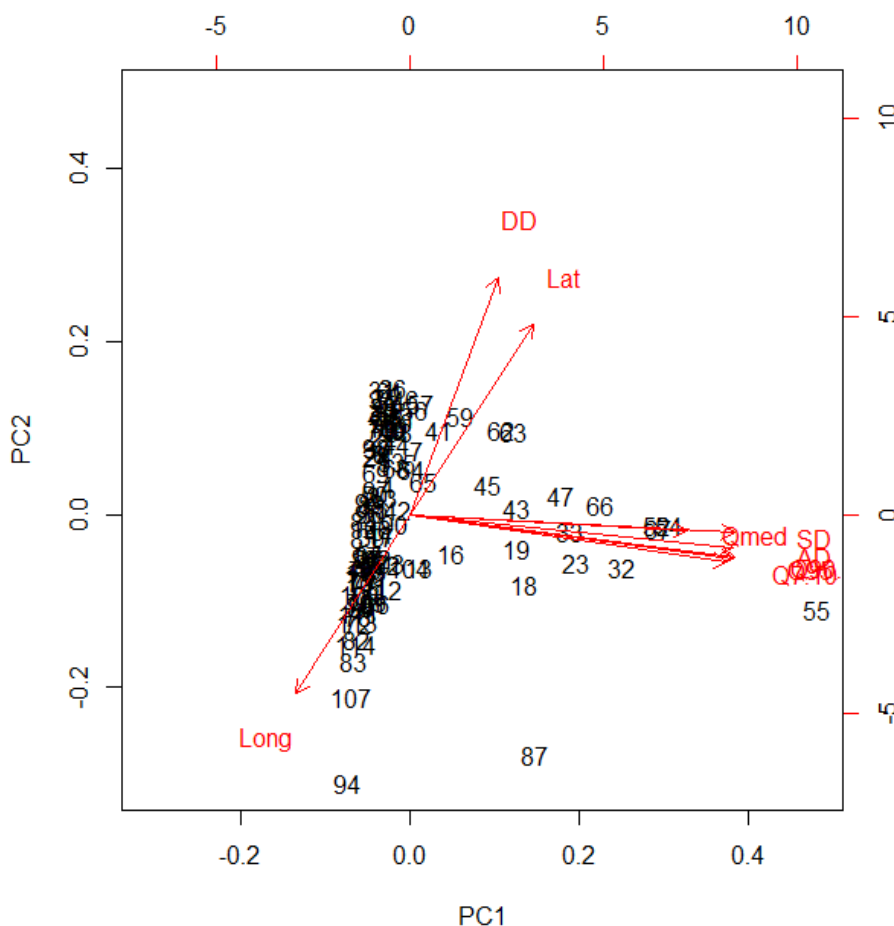


Figura 4 Biplot dos dados no espaço das duas primeiras PC, com pontos nomeados nas variáveis e nas 114 observações.

De acordo com Fritzsons et al. (2011), em uma figura que represente a ACP, vetores com medidas mais distantes de zero correspondem a variações com maior influência sobre o valor do PC, enquanto que vetores mais próximos de zero indicam que correspondem a uma variável com pequena influência sobre o PC.

Portanto, a Figura 4 reforça o que foi visualizado na matriz de correlação simples aplicado inicialmente na ACP, que não há relação positiva das variáveis DD, Lat e Long em relação às variáveis escolhidas e suas observações que mais se relacionam, em que tais variáveis ficaram bem distintas das variáveis selecionadas.

#### 5.4 Análise de agrupamento Fuzzy C-Means

Na aplicação do método Fuzzy C-Means utilizou-se a matriz de dados definida pós ACP, composta pelos valores das variáveis dependentes ( $Q_{90}$ ,  $Q_{95}$  e  $Q_{7,10}$ ) e independentes (AD e SD), das 114 estações fluviométricas, resultando em uma matriz de entrada de 114 x 5.

Realizaram-se 176 simulações para encontrar a melhor formação do agrupamento. O número de cluster e o m foram definidos por meio da minimização da função objetiva de cada simulação, conforme a Tabela 4.



Tabela 4 Valores das 176 funções objetivas obtidas por combinação cluster x índice de fuzzificação.

		m (índice de fuzzificação)					
		1	1,1	1,2	1,3	1,4	1,5
Cluster	2	8,061E+11	8,651E+11	1,176E+12	1,187E+12	8,155E+11	8,214E+11
	3	1,240E+12	1,029E+12	7,841E+11	8,994E+11	1,186E+12	8,732E+11
	4	1,146E+12	8,981E+11	1,168E+12	1,146E+12	1,220E+12	1,146E+12
	5	1,168E+12	8,061E+11	1,154E+12	6,670E+11	1,146E+12	1,164E+12
	6	8,242E+11	8,242E+11	5,494E+11	1,171E+12	1,208E+12	8,994E+11
	7	8,651E+11	8,651E+11	1,168E+12	8,268E+11	1,406E+12	1,178E+12
	8	1,162E+12	1,212E+12	1,176E+12	1,162E+12	8,420E+11	8,412E+11
	9	8,805E+11	8,242E+11	8,353E+11	1,212E+12	1,168E+12	1,212E+12
	10	8,061E+11	1,146E+12	1,176E+12	1,214E+12	1,156E+12	8,268E+11
	11	1,240E+12	1,240E+12	1,220E+12	1,156E+12	8,343E+11	1,140E+12
	12	1,403E+12	1,146E+12	8,866E+11	8,268E+11	1,176E+12	1,146E+12
	13	1,146E+12	1,230E+12	7,730E+11	8,061E+11	1,164E+12	1,168E+12
	14	8,901E+11	1,154E+12	8,651E+11	8,061E+11	1,240E+12	8,840E+11
	15	1,044E+12	8,268E+11	1,409E+12	1,220E+12	7,984E+11	8,732E+11
	16	8,242E+11	8,268E+11	7,958E+11	1,247E+12	8,268E+11	8,268E+11
	17	1,154E+12	8,651E+11	1,162E+12	8,151E+11	8,285E+11	8,061E+11
			m (índice de fuzzificação)				
		1,6	1,7	1,8	1,9	2	
Cluster	2	1,156E+12	1,168E+12	8,634E+11	1,206E+12	1,220E+12	
	3	1,442E+12	1,164E+12	8,242E+11	7,730E+11	8,242E+11	
	4	8,593E+11	<b>3,043E+11</b>	1,442E+12	8,061E+11	8,651E+11	
	5	1,220E+12	8,994E+11	8,651E+11	7,722E+11	8,810E+11	
	6	1,035E+12	8,214E+11	8,901E+11	1,234E+12	1,162E+12	
	7	8,468E+11	8,268E+11	7,662E+11	1,176E+12	8,155E+11	
	8	8,214E+11	1,147E+12	8,654E+11	1,206E+12	8,732E+11	
	9	1,154E+12	1,146E+12	8,061E+11	1,176E+12	1,146E+12	
	10	8,593E+11	8,268E+11	1,078E+12	8,268E+11	8,268E+11	
	11	8,214E+11	1,409E+12	8,654E+11	7,958E+11	8,155E+11	
	12	1,156E+12	1,140E+12	1,078E+12	8,732E+11	1,247E+12	
	13	8,840E+11	8,840E+11	8,651E+11	1,035E+12	1,178E+12	
	14	1,176E+12	1,240E+12	8,651E+11	1,202E+12	1,035E+12	
	15	1,162E+12	1,162E+12	7,841E+11	1,146E+12	1,221E+12	
	16	1,171E+12	1,162E+12	1,147E+12	1,154E+12	8,901E+11	
	17	8,268E+11	8,359E+11	1,154E+12	1,177E+12	8,071E+11	

A menor função objetiva encontrada foi para 4 Clusters no  $m$  1,7, correspondendo ao valor  $3,043E+11$ .

Segundo Yonamine et al. (2002), não existe uma base teórica para a escolha do valor de  $m$  pois esse parâmetro é acordado conforme as características do estudo como local, variáveis consideradas, objeto analisado, entre outras.

Em seu trabalho voltado para a região da Amazônia Legal, Pessoa (2015) variou o  $m$  entre 1,3 e 2,0, o número máximo de iterações em 200 e como critério de parada um erro mínimo de 0,0001. Variou o número de grupos entre 2 a 15, sendo que a formação mais adequada encontrada, nesse trabalho, foi de 14 grupos no  $m$  2,0.

Segundo Sadri e Burn (2011), não existe um número “correto” de clusters para um certo número de estações que satisfaça à condição de homogeneidade. No entanto, o objetivo é escolher o número de clusters onde dentro dos mesmo as características das variáveis variem pouco, de modo que em cada cluster haja uma distribuição interna equilibrada na frequência das variáveis.

Dando continuidade à pesquisa, dessa vez na formação escolhida, o algoritmo Fuzzy C-Means realizou 46 iterações. Para a primeira iteração do algoritmo, a função objetivo forneceu o valor  $11,9 \times 10^{13}$  e para a última iteração o valor calculado foi  $1,47 \times 10^{12}$ , sendo estabilizada a partir da iteração 9. A Figura 5 ilustra a convergência da função objetivo em relação ao número de iteração.

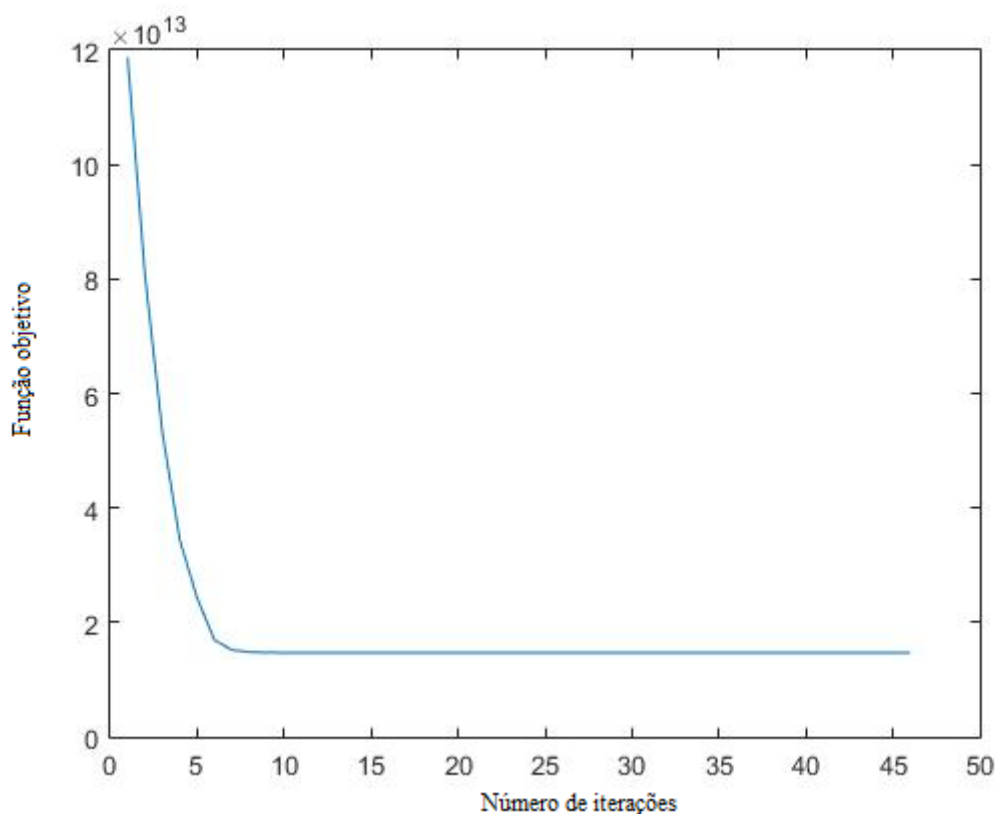


Figura 5 Convergência da função objetivo para 4 agrupamentos.

A partir disso, realizou-se a obtenção dos graus de pertinência para cada uma das 114 estações fluviométricas estudadas nos 4 Clusters (Tabela 5). A tabela completa está apresentada no Apêndice E.

Tabela 5 Graus de pertinência das estações fluviométricas para os 4 grupos formados

ID	Código	Cluster 1	Cluster 2	Cluster 3	Cluster 4
E1	64230500	0,00000	0,00000	0,99999	0,00001
E2	64240000	0,00006	0,00001	0,99759	0,00234
E3	64242000	0,00106	0,00018	0,92444	0,07432
E4	64360000	0,00187	0,00031	0,83362	0,16420
E5	64362000	0,00016	0,00002	0,00333	0,99649
E6	64365850	0,00005	0,00001	0,99854	0,00139
.	.	.	.	.	.
.	.	.	.	.	.
E114	82198000	0,00007	0,00001	0,99796	0,00195

A lógica Fuzzy permitiu que as variáveis não admitam valores precisos em apenas um cluster, necessariamente como 0 ou 1, mas sim que elas tenham graus de pertinência entre os elementos, em relação ao seu conjunto. Segundo Rignel, Chenci e Lucas (2011), isso possibilita a construção de várias regras, que facilitam a modelagem dos problemas, tornando-os, assim, menos complexos.

Sadri e Burn (2011) lembram que um objeto pode contribuir para dois ou mais clusters ao mesmo tempo, sendo uma vantagem do agrupamento difuso. Da mesma forma, para alcançar o tamanho efetivo de cada cluster, podem ser feitos ajustes e critérios para atender à homogeneidade interna.

Devido a isso, para a formação dos grupos, foi utilizado o método de decisão pelo maior grau de pertinência, ou seja, para um dado objeto o maior grau de pertinência determina a qual agrupamento o objeto pertencerá.

Todavia, tomaram-se alguns cuidados ao analisar os resultados, pois se um objeto possui seu maior grau de pertinência no intervalo 0,7 a 1,0, então é considerado não confuso; mas se esse valor for menor que 0,7 e maior 0,6, então este objeto poderá pertencer a um outro grupo.

No trabalho de Santos et al. (2016), a escolha dos cluster também se deu pelo grau de pertinência; contudo, optou-se apenas pelos maiores valores em cada cluster, não havendo critério de intervalo, devido à análise dos resultados por diferentes cenários.

Seguindo o critério determinado, obtivemos a maior concentração de estações no Cluster 3 (Tabela 6).

Tabela 6 Distribuição das estações fluviométrica por Cluster

Grupo	Número de estações	%
Cluster 1	7	6,1
Cluster 2	7	6,1
Cluster 3	83	72,8
Cluster 4	13	11,4
Nebulosas	4	3,5
Total	114	100

Observa-se que das 114 estações fluviométricas, 110 foram consideradas não confusas se enquadrando em algum Cluster, e apenas 4 foram designadas nebulosas (3,5%), pois possuem seu maior grau de pertinência menor que 0,6.

A distribuição espacial das estações fluviométricas as quais possuem RHH de vazões das estações fluviométricas encontradas no estado do Paraná por meio do método Fuzzy C-Means é apresentada na Figura 6.

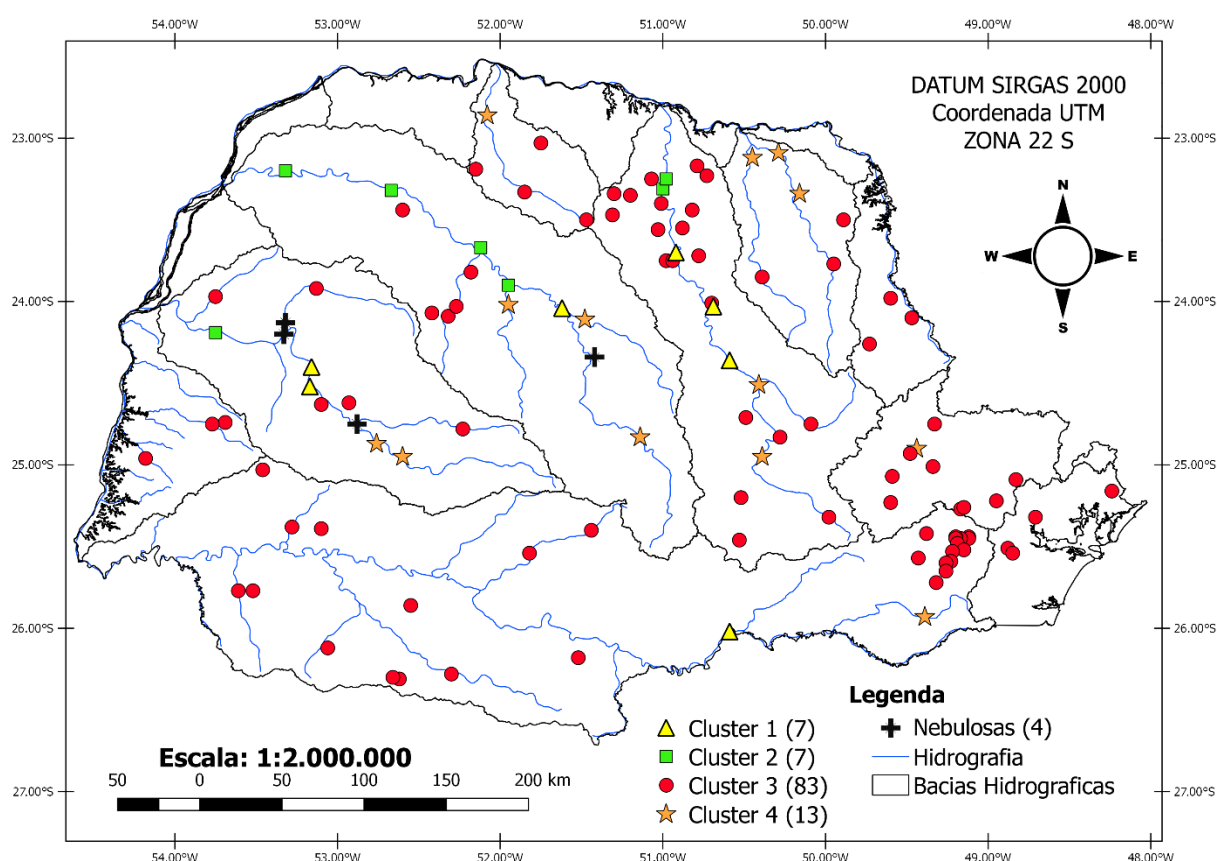


Figura 6 Mapa das estações hidrologicamente homogêneas de vazão definidas no estudo através do método Fuzzy C-Means

Sobre as estações Nebulosas, 3 delas se encontraram na Bacia Hidrográfica do Piquiri e 1 na Bacia Hidrográfica do Ivaí. Todas essas estações possuem mais de 27 anos

de dados de vazão; contudo, suas características não se enquadram em nenhum cluster segundo o agrupamento Fuzzy C-Means.

Um exemplo que representa muito bem um agrupamento nebuloso, no qual um objeto pode não pertencer exclusivamente a um grupo, é o caso da estação fluviométrica E43 – Cod. ANA 64645000, localizada na bacia hidrográfica Pirapó.

Observa-se no Apêndice E que esta estação tem 34% de chance de pertencer ao Cluster 1 e 56% ao Cluster 4. Mesmo apresentando uma grande chance de pertencer ao Cluster 4, a E43 não foi alocada em nenhum grupo, já que seu maior grau de pertinência foi menor que o critério estabelecido para a distribuição nos clusters, sendo alocada como nebulosa. O mesmo aconteceu nas estações E59, E65 e E66.

### 5.5 Caracterização das RHH

Os 4 clusters formados simbolizam 4 regiões hidrológicas encontradas para todo o estado do Paraná, onde cada uma delas possui características próprias, distintas uma da outra, segundo as variáveis analisadas.

Já as 4 regiões homogêneas geradas a partir do agrupamento das estações, de acordo com os maiores graus de pertinência, obtiveram as seguintes características (Tabela 7).

Tabela 7 Características de cada região formada por meio do método de agrupamento Fuzzy C-Means

Cluster	Área de Drenagem (Km <sup>2</sup> )	Soma das drenagens (m)	Q <sub>90</sub> (m <sup>3</sup> /s)	Q <sub>95</sub> (m <sup>3</sup> /s)	Q <sub>7,10</sub> (m <sup>3</sup> /s)
1	11219-18433	3213852-4714590	53,90-126,48	38,34-98,54	19,99-55,18
2	19391-34319	5555648-9173727	87,31-246,57	65,80-220,27	34,7-164,3
3	12-2212	100-574672	0,08-21,27	0,0001-19,48	0-15,59
4	2603-8759	718681-2198487	4,17-43,46	3,48-34,50	2,56-21,67

Os grupos foram determinados praticamente pela distribuição das variáveis AD e SD (Figura 7), havendo distinção entre os clusters pelas variáveis estudadas, tal como o Cluster 3, que apresentou as estações com menores áreas de drenagem e também as menores quantidades de drenagens na região de abrangência das estações. Já as estações

possuidoras das AD e SM, em um intervalo com valores maiores que 19000 Km<sup>2</sup> e 5500000 m, foram distribuídas e agrupadas no Cluster 2.

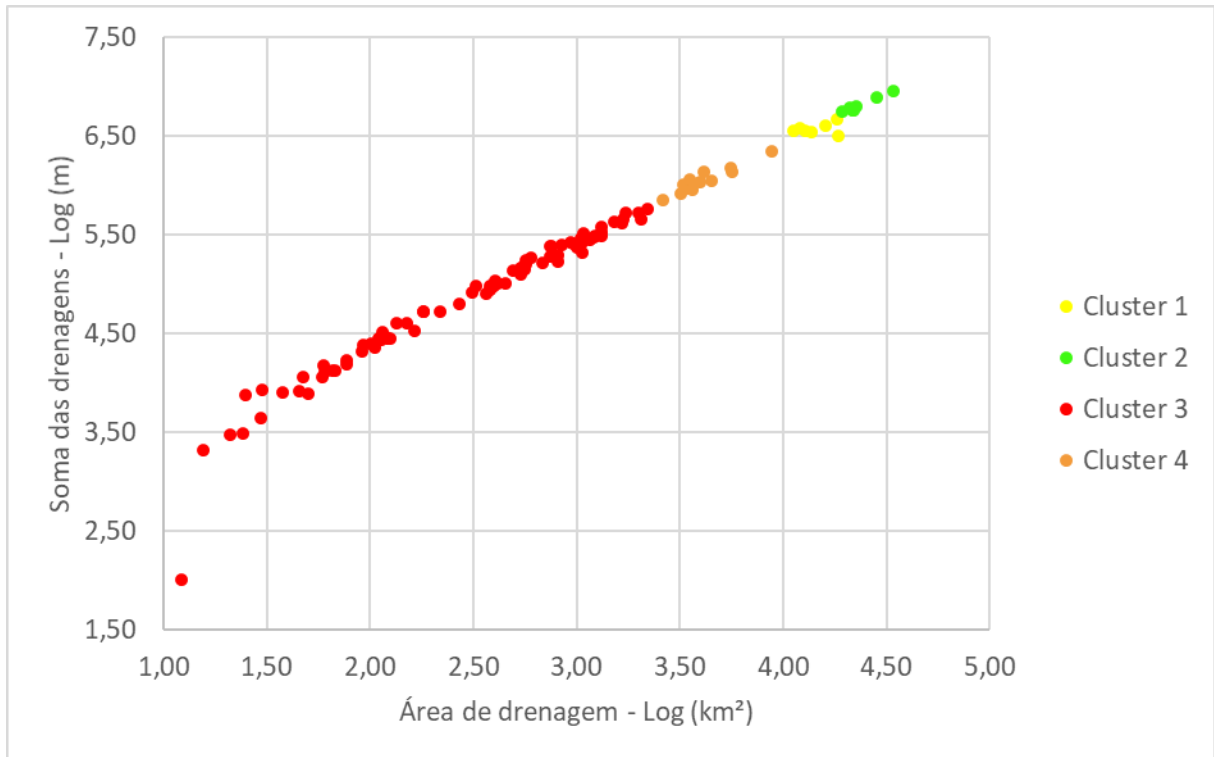


Figura 7 Distribuição de grupos pelas variáveis Soma das drenagens (m) x Área de drenagem (km<sup>2</sup>), em escala logarítmica

Sadri e Burn (2011) destacam que muitas das vezes são encontrados grupos que não atendem aos requisitos de falta de discordância, homogeneidade e tamanho suficientes para uma região efetiva e, portanto, as revisões dos clusters iniciais são inevitáveis.

Essa distribuição dos grupos pode ser considerada como uma das mais importantes contribuições do método Fuzzy C-Means, pois, segundo Pessoa (2015), se os grupos formados apresentarem alta similaridade interna e dissimilaridade entre eles, então pode-se utilizar o método para identificar a qual região homogênea uma determinada área carente de informações de vazões pertenceria, utilizando somente as variáveis estudadas.

Cada cluster obteve um intervalo nas variáveis estudadas, não havendo sobreposição de dados entre os clusters, formando assim características próprias entre eles.

Analisando a variável AD nos clusters formados, encontramos a maior média geral dessa variável no Cluster 2 (24132,49), como também o maior desvio padrão (4114,40) e valor de máximo (34319,74); seguidos dos Cluster 1 e Cluster 4. Já para o Cluster 3, encontraram-se os menores valores de todos os parâmetros analisados nessa variável (Tabela 8).

Tabela 8 Dados da AD nos clusters formados

Área de drenagem – AD (Km <sup>2</sup> )				
	Média	Desvio Padrão	Máximo	Mínimo
Cluster 1	14641,88	2512,37	18433,31	11219,45
Cluster 2	24132,49	4114,40	34319,74	19391,32
Cluster 3	560,95	453,19	2215,04	12,25
Cluster 4	4289,83	1124,71	8759,91	2603,09

Na SD obtivemos a maior média (3783542,57), desvio padrão (368162,25) e valor de máximo (4714590,54) no Cluster 1, seguidos do Cluster 4, Cluster 3 e Cluster 2. Entretanto, mesmo o Cluster 2 tendo dado os menores valores nos parâmetros citados anteriormente, o menor valor encontrado para o parâmetro mínimo foi no Cluster 3 (100,99) (Tabela 9).

Tabela 9 Dados da SD nos clusters formados

Soma das drenagens – SD (m)				
	Média	Desvio Padrão	Máximo	Mínimo
Cluster 1	3783542,57	368162,25	4714590,54	3213852,27
Cluster 2	24132,49	4114,40	34319,74	19391,32
Cluster 3	146370,13	121051,57	574672,22	100,99
Cluster 4	1166849,91	275948,79	2198487,42	718681,16

Analisando os parâmetros para a variável  $Q_{90}$ , encontramos a maior média, desvio padrão e valor máximo no Cluster 2 (Tabela 10). O mesmo foi encontrado para a  $Q_{95}$ . Já os valores para o parâmetro mínimo, para as mesmas, foram encontrados no Cluster 3.

Tabela 10 Dados da Q90 e Q95 nos clusters formados

Q90 (m <sup>3</sup> /s)				
	Média	Desvio Padrão	Máximo	Mínimo
Cluster 1	83,34	22,24	126,4769	53,9034
Cluster 2	146,67	37,83	246,57	87,31
Cluster 3	4,07	3,70	21,27	0,08
Cluster 4	18,55	9,76	43,46	4,17
Q95 (m <sup>3</sup> /s)				
	Média	Desvio Padrão	Máximo	Mínimo
Cluster 1	65,90	18,51	98,5401	38,3366
Cluster 2	120,29	36,13	220,27	65,80
Cluster 3	3,31	3,11	19,48	0,00
Cluster 4	14,64	8,22	34,50	3,48

O Cluster 2 foi novamente destaque nos maiores valores de média (76,68), desvio padrão (29,01) e máximo (164,30) para a variável  $Q_{7,10}$ , seguidos dos Cluster 1 e Cluster 4 (Tabela 11).

O Cluster 3 apresentou os menores valores para todos os parâmetros analisados, e avaliando o conjunto de dados desse cluster, destacamos a estação 64670012, onde para o parâmetro mínimo encontramos o valor 0,00. Isso se deu por não haver um período de 10 anos concluído para que se possa realizar o cálculo da  $Q_{7,10}$ , já que esta foi instalada em 2009 e encontra-se ativa até o presente momento.

Tabela 11 Dados da Q7,10 nos clusters formados

Q7,10 (m <sup>3</sup> /s)				
	Média	Desvio Padrão	Máximo	Mínimo
Cluster 1	37,92	11,45	55,18	19,99
Cluster 2	76,68	29,01	164,30	34,70
Cluster 3	2,18	2,14	15,59	0,00
Cluster 4	9,34	5,54	21,67	2,56



Ao todo, o Cluster 2 foi o que apresentou os maiores valores em todos os parâmetros analisados nas variáveis estudadas, situando-se as suas estações nas bacias hidrográficas do Ivaí (4 estações), Piquiri (1 estação) e Tibagi (2 estações).

Já o Cluster 3 caracterizou-se por possuir estações com as menores áreas de abrangência, vazões analisadas e as menores quantidade de drenagens em sua área de cobertura, não havendo uma aglomeração pontual de estações com essas características, mas sim bem espalhada na região estudada, havendo a presença das estações do Cluster 3 em todas as bacias hidrográficas do estado.

## 6 CONCLUSÃO

Após ajustar a distribuição de probabilidade para Log-Pearson tipo III, e organizar os dados das variáveis dependentes e independentes em matriz, foi possível aplicar a ACP para conhecer o comportamento dos dados e identificar quais variáveis eram mais significativas em relação à identificação das RHH.

Das 9 variáveis imposta na matriz inicial, a ACP apontou que apenas 4 delas eram relevantes significativamente, sendo a AD, SD,  $Q_{90}$  e  $Q_{95}$ . Contudo, foi inserida a  $Q_{7,10}$  devido a sua importância para o estado do Paraná.

Aplicando o algoritmo Fuzzy C-Means na matriz com as variáveis selecionadas pós ACP, definiu-se o m e o número de clusters por meio da menor função objetiva, em que esta foi encontrada para 4 Clusters no m 1,7.

Separando as estações fluviométricas por clusters através dos graus de pertinência, obtivemos o maior número de estações no Cluster 3 (83 estações), seguidos do Cluster 4 (13 estações) e dos Clusters 1 e 2 (7 estações em cada cluster). Apenas 4 estações não foram inseridas em nenhum cluster, sendo classificadas como nebulosas.

Os grupos foram determinados praticamente pela distribuição das variáveis AD e SD. Contudo, o Cluster 2 foi o que apresentou os maiores valores em média, desvio padrão, valor máximo e mínimo nas variáveis estudadas.

As menores áreas de abrangência, vazões analisadas e as menores quantidade de drenagens na área de cobertura das estações foram encontra no Cluster 3, que se encontraram bem espalhada no estado do Paraná.

Já os Clusters 1 e 4 ficaram intermediários entre os demais clusters em todos os parâmetros avaliados.

O algoritmo Fuzzy C-Means se mostrou eficiente para o agrupamento das estações fluviométricas no estado do Paraná, sendo que foi possível encontrar a características de cada cluster formado.

Deste modo, uma parte das etapas da regionalização de vazão, a identificação das RHH, foi realizada com êxito, podendo ser futuramente concluída ou até mesmo comparada com outros métodos de agrupamento diante das variáveis já organizadas.

## 7 CONSIDERAÇÕES FINAIS

Durante a elaboração do trabalho, a limitação metodológica de grande destaque foi a falta de dados atuais devido à localização das estações fluviométricas instaladas.

Não se trata de haverem poucas estações, mas sim pelo fato que não estarem distribuídas homoganeamente, havendo bacias hidrográficas onde não se encontra nenhuma estação, como é o caso da Paraná 1 e 2; Paranapanema 1, 2, 3 e 4.

Encontrou-se também dificuldade quanto à obtenção das mesmas variáveis em todas as estações e, devido a isso, das 484 estações ativas no Paraná, apenas 114 encontram-se com essas variáveis.

Os estudos abaixo são colocados como sugestões para a continuação de pesquisa em outras etapas:

Como sugestão, indica-se para trabalhos futuros a aplicação de Redes Neurais Artificiais (RNAs) como ferramenta de agrupamento, com a finalidade de identificar regiões homogêneas de vazão e comparar seus resultados ao método difuso Fuzzy C-Means; e também definir modelos de regionalização de vazão para o Paraná.

## 8 REFERÊNCIAS

AAKER, D. A.; KUMAR, V.; DAY, G. S. **Pesquisa de marketing**. São Paulo: Atlas, p. 745, 2001.

AGÊNCIA NACIONAL DE ÁGUAS – ANA. Outorga de direito de uso de recursos hídricos. **Cadernos de capacitação em recursos hídricos** v. 6, 2011.

ALBUQUERQUE, M. A.; SILVA, E. L.; BARROS, K. N. N. O.; JUNIOR, S. F. A. X. Comparação entre Coeficientes Similaridade uma aplicação em Ciências Florestais. **Revista Eletrônica Matemática e Estatística em Foco** v. 4, n. 2, p. 102-114, 2016.

AMORIM, E. L. C.; NETTO, A. P. O.; MENDIONDO, E. M. Estudo de métodos para regionalização de vazão. **Anais do ABRH**, v. 1, São Paulo, 2005.

APARECIDO, L. E. O.; ROLIM, G. S.; RICHETTI, J.; SOUZA, P. S.; JOHANN, J. A. Köppen, Thornthwaite and Camargo climate classifications for climatic zoning in the State of Paraná, Brazil. **Ciência e Agrotecnologia**, Lavras, v. 40, n. 4, 2016. Disponível em: <[http://www.scielo.br/scielo.php?script=sci\\_arttext&pid=S1413-70542016000400405](http://www.scielo.br/scielo.php?script=sci_arttext&pid=S1413-70542016000400405)>. Acesso em: 27 abr. 2017.

ARAÚJO, P. C. **Uma combinação de métodos de pesquisa em Educação Matemática: Método Bayesiano de Dados Difusos**. Tese (Doutorado em Matemática) Programa de Pós-Graduação em Matemática, Universidade Católica de São Paulo – PUC, 2013. Acesso em: <<https://sapiencia.pucsp.br/bitstream/handle/10971/1/Pericles%20Cesar%20de%20Araujo.pdf>>. Disponível em: 20 abr. 2017.

ASSANI, A. A.; CHALIFOUR, A.; LÉGARÉ, G.; MANOUANE, C.; LEROUX, D. Temporal regionalization of 7-day low flows in the St. Laurence watershed in Quebec (Canada). **Water Resources Management** v. 25, p. 3559-3574. 2009.

BARROS, F. V. F.; MARTINS, E. S. P. R.; SOUZA FILHO, F. A. **Regionalização de Parâmetros do Modelo Chuva-Vazão SMAP das Bacias Hidrográficas do Ceará**. Gerenciamento dos Recursos Hídricos no Semiárido. Fortaleza: FCPC/UFC, v. 1, p. 31-49, 2013.

BEM, J. S.; WAISMANN, M. Utilização da técnica da análise de clusters ao emprego da indústria criativa entre 2000 e 2010: estudo da Região do Consinos, RS. **Interações**, Campo Grande, v. 16, n. 1, p. 27-41, 2015.

BEZDEK, J. **Patter recognition with fuzzy objective function algorithms**. New York, Plenum Press, 1981.

BICUDO, C. E. M.; BICUDO, D. C. **Amostragem em Limnologia**. São Carlos: Editora RIMA Artes e Textos, p. 371, 2004.

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

BRAGA, C. C.; MACEDO, M. J. H.; SILVA, B. B.; BRAGA, R. C. Aplicação dos componentes principais na simulação e consistência de séries temporais. **Revista de Geografia (UFPE)** v. 29, n. 1, 2012.

BUSSAB, W. de O.; MIAZAKI, S. E.; ANDRADE, D. F. Introdução à análise de agrupamento. In: IX Simpósio Brasileiro de Probabilidade e Estatística, IME-USP. **Anais...** São Paulo, 1990.

CALEGARIO, A. T. **Identificação de regiões hidrologicamente homogêneas e comportamento das vazões mínima e média regionalizadas.** Dissertação (Mestrado em Engenharia Agrícola) - Programa de Pós-Graduação em Engenharia Agrícola. Universidade Federal de Viçosa, Viçosa, 2014. Disponível em: <<http://locus.ufv.br/bitstream/handle/123456789/8210/texto%20completo.pdf?sequence=1&isAllowed=y>>. Acesso em: 12 mar. 2017.

CASTRO, B. M.; FREITAS, S. M.; PRATA, B. A.; EVANGELISTA, G. L. **“Raio de Influência”**: um método de agrupamento alternativo para Análise de Cluster. 19º Sinape, São Pedro, SP, 2010. Disponível em: <<http://www2.ime.unicamp.br/sinape/sites/default/files/Bruno%20-%20Sinape%202010.pdf>>. Acesso em: 20 abr. 2017.

CEBALLOS, J. C.; BRAGA, C. C. Missing Data Assessment in a Solarimetric Network. **International Journal of Climatology** v.15, p. 325-340, 1995.

CHAVES, H. M. L. ROSA, J. W. C.; VADAS, R. G.; OLIVEIRA, R. V. T. Regionalização de vazões mínimas em bacias através de interpolação em sistemas de informações geográfica. **Revista Brasileira de Recursos Hídricos**, Porto Alegre, v.7, n. 3. p. 43-51, 2002.

COSTA, A. S.; CARIELLO, B. L.; BLANCO, C. J. C.; PESSOA, F. C. L. Regionalização de curvas de permanência de vazão de regiões hidrográficas do Estado do Pará. **Revista Brasileira de Meteorologia** v. 27, n. 4, 413-422, 2012. Acesso em: <<http://www.scielo.br/pdf/rbmet/v27n4/05.pdf>>. Disponível em: 25 fev. 2017.

DONI, M. V. **Análise de Cluster**: métodos hierárquicos e de particionamento. 93 f. Monografia de Pós-Graduação – Universidade Prebisteriana Mackenzie, São Paulo, 2004. Disponível em: <<http://meusite.mackenzie.com.br/rogerio/tgi/2004Cluster.PDF>>. Acesso em: 10 jan. 2017.

DOURADO, C. da S.; OLIVEIRA, S. R. de M.; AVILA, A. M. H. de. **Análise de zonas homogêneas em séries temporais de precipitação no Estado da Bahia.** Bragantia, Campinas, v. 72, n. 2, p.1 92-198, 2013.

DUNN, J. **A fuzzy relative of the Isodata process and its use in detecting compact, wellseparated clusters.** Journal of Cybernetics, v. 3, p 32-57, 1973. Disponível em: <<http://www.tandfonline.com/doi/abs/10.1080/01969727308546046>>. Acesso em: 17 mai. 2017.

DZIUBAN, C. D.; SHIRKEY, E. C. **When is a correlation matrix appropriate for factor analysis? Some decision rules.** Psychological Bulletin, 81(6), 358-361. 1974 Disponível em: <[https://www.researchgate.net/publication/232522477\\_When\\_Is\\_a\\_Correlation\\_Matrix\\_Appropriate\\_for\\_Factor\\_Analysis\\_Some\\_Decision\\_Rules](https://www.researchgate.net/publication/232522477_When_Is_a_Correlation_Matrix_Appropriate_for_Factor_Analysis_Some_Decision_Rules)>. Acesso em: 05 nov. 2017.

EASTMAN, J.R.; FULK, M. Long sequence time series evaluation using standardized principal components. **Photogrammetric Engineering & Remote Sensing**, v. 59, n. 9 p.1307-1312, 1993.

EUCLYDES, H. P.; FERREIRA, P. A.; RIBEIRO, C. A. A. S.; FARIA FILHO, R. F.; LEITE, C. V.; SANTOS, A. P.; ALTOÉ, D. R.; SANTOS, S. A.; OLIVEIRA, E. P.; NUNES, D. M.; FERREIRA, P. H. S.; VIANA, D. C. **Atlas Digital das Águas de Minas**. 3 Ed. Programa de pesquisa e desenvolvimento HIDROTEC, Fundação Rural Mineira (RURALMINAS), Universidade Federal de Viçosa (UFV), 2014.

EVERITT, B.S.; DER, G. A. **Handbook of statistical analyses using SAS**. London: Chapman & Hall, Book Review 2, v. 30, apr. 2009.

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

FÁVERO, L. P.; BELFIORE, P.; SILVA, F. L.; CHAN, B. L. **Análise de dados – Modelagem multivariada para tomadas de decisões**. Ed. Campus Elsevier, 2009.

FAYYAD, U. M.; PIATETSKY SHAPIRO, G.; SMYTH, P.; UTHURUSAMY, R. **Advances in Knowledge Discovery and Data Mining**. AAAIPress, The MIT Press, 1996. Disponível em: <<https://www.csd.uwo.ca/faculty/ling/cs435/fayyad.pdf>>. Acesso em: 17 maio 2017.

FERRARI, A. C. K. **Identificação de sistemas multiváriaveis por evolução diferencial otimizado por lógica nebulosa**. Dissertação (Mestrado em Engenharia Elétrica) - Departamento de Engenharia Elétrica, Universidade Federal do Paraná – UFPR, Curitiba, 2015.

FERREIRA, D. F. **Estatística Multivariada**. Lavras: UFLA, 2008. 662p.

FERREIRA, J. L.; LOPES, F. B.; ASSIS, A. S.; LÔBO, R. B. Espacialização climática e análise de dados longitudinais de bovinos Nelore criados nos Estados do Maranhão, Pará e Tocantins por meio de técnicas univariadas e multivariadas. **Semina: Ciências Agrárias**, Londrina, v. 35, n. 4, p. 2197-2210, 2014.

FIELD, A. **Discovering Statistics Using SPSS**. 2a ed., London: Sage, 2005.

FILHO, D. B. F.; PARANHOS, R.; ROCHA, E. C.; SILVA, A. H.; JUNIOR, J. A. S.; OLIVEIRA, L. E.; ALVES, D. P. Análise fatorial garantida ou o seu dinheiro de volta: uma introdução à redução de dados. **Revista Eletrônica de Ciência Política**, v. 5, n. 2, 2014.

FIGUEIREDO, A. P.; OLIVEIRA, L. F. C.; FRANCO, A. P. B. Avaliação do desempenho de equações de regionalização de vazões na Bacia Hidrográfica do Ribeirão Santa Bárbara, Goiás, Brasil. **Revista Ambiente e Água**, v. 3, n. 2, p. 68-82, 2008.

FLOYD, F. J.; WIDAMAN K. F. Factor analysis in the development and refinement of clinical assessment instruments. **Psychological Assessment**, v. 7, p. 286-299, 1995.

FRAGOSO, M.; GOMES, P.T. Classification of daily abundant rainfall patterns and associated large-scale atmospheric circulation types in Southern Portugal. **International Journal of Climatology**, v. 28, p. 537-544, 2008.

FRITZSONS, E.; MANTOVANI, L. E.; WREGGE, M. S.; CHAVES NETO, A. Análise da Pluviometria para Definição de Zonas Homogêneas no Estado do Paraná. **RaEGa**, n. 23, p. 555-572, 2011.

FUNG, G. A. **Comprehensive overview of basic clustering algorithms**. 2001. Disponível em:

<<http://citeseerx.ist.psu.edu/viewdoc/download;jsessionid=38CCA0A82B38450882DA0D9471D45509?doi=10.1.1.5.7425&rep=rep1&type=pdf>>. Acesso em: 06 abr. 2017.

GARAYALDE, E. J. G.; SILVA, M. G. A. da; TAVARES, A. de S. **Classificação meso-climática da região sul do Brasil pela análise de Componentes Principais**. In: Congresso Interamericano de Meteorologia, Congresso Brasileiro de Meteorologia, 1986, Brasília. **Anais...** Brasília: Sociedade Brasileira de Meteorologia, v. 1. p. 119-124, 1986.

GOMES, E. P.; BLANCO, C. J. C. **Identificação de regiões homogêneas de vazão via agrupamento Fuzzy C-Means no estado do Pará**. INICIAÇÃO CIENTÍFICA DA UFPA, Resumo Expandido, ISSN 2176-1213, 26<sup>o</sup> ed., 2014.

GRANZIERA, M, L. M. A Fixação de Vazões de Referência. **Revista de Direito Ambiental**, v. 70, n. 18, 2013. Disponível em: <<http://www.granziera.com.br/artigos/2013/7/17/a-fixao-de-vazes-de-referencia-adequadas-como-instrumento-de-segurana-juridica-e-sustentabilidade-ambiental-na-concesso-de-outorgas-de-direito-de-uso-de-recursos-hdricos>>. Acesso em: 17 mar. 2017.

GUELPELI, M. V. C. **Cassiopeia**: um modelo de agrupamento de textos baseado em sumarização. Tese (Doutorado em Computação) - Programa. de Pós-Graduação em Computação. Universidade Federal Fluminense, 2012.

HAIR JUNIOR, J.F.; ANDERSON, R.E.; TATHAM, R.L.; BLACK, W.C. **Análise multivariada de dados**. Porto Alegre: Bookman, 2005.

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

HAIR, J. F.; TATHAM, R. L.; BLACK, W. C. **Análise de regressão múltipla**. Tradução SCHELUP, A.; CHAVES, A. 5. ed., Porto Alegre: Bookman, 2005.

HALDIKI, M. On Clustering Validation Techniques. **Journal of Intelligent Information System**, v.17, n. 2-3, p.107-145, 2001.

HALL, M. J.; MINNS, W. A. The classification of hydrologically homogeneous regions. **Hydrological Sciences Journal**, v. 44, n. 5, p. 693-704, 1999. Disponível em: <<http://www.tandfonline.com/doi/pdf/10.1080/02626669909492268>>. Acesso em: 17 maio 2017.

HAN, J.; KAMBER, M.; PEI, J. **Data mining**: concepts and techniques. 3ed. San Francisco: Morgan Kaufmann Publishers, 2011.

HONGYU, K.; SANDANIELO, V. L. M.; JUNIOR, G. J. O. Análise de Componentes Principais: resumo teórico, aplicação e interpretação. **E&S - Engineering and Science**, v. 1, n. 5, 2015.

HUTCHESON, G. D.; SOFRONIOU, N. **The multivariate social scientist**: Introductory statistics using generalized linear models. London: Sage Publications, 1999.

INSTITUTO MINEIRO DE GESTÃO DAS ÁGUAS – IGAM. **Estudo de regionalização de vazão para o aprimoramento do processo de outorga no Estado de Minas Gerais.** Grupo de Pesquisas em Recursos Hídricos da UFV, Instituto Mineiro de Gestão das Águas, Belo Horizonte, 2012. Disponível em: <[http://www.igam.mg.gov.br/images/stories/cerh/ctig/36\\_RO\\_CTIG\\_11\\_10\\_2012/4.-apresentacao-igam-ufv-ctig-11102012.pdf](http://www.igam.mg.gov.br/images/stories/cerh/ctig/36_RO_CTIG_11_10_2012/4.-apresentacao-igam-ufv-ctig-11102012.pdf)>. Acesso em: 21 mar. 2017.

JOLLIFFE, I. T. **Principal Component Analysis.** 2nd ed. Springer Series in Statistics, New York: Springer-Verlag, New York, 2002.

KAHYA, E.; KALAYCI, S.; PIECHOTA, T. C. **Streamflow regionalization: case study of Turkey.** Journal of Hydrologic Engineering, v. 13. 2008.

KERNKAMP, C. L.; COSTA, C. K. F.; MASSUDA, E. M.; SILVA, E. S.; YAMAGUCHI, M. U.; BERNUCI, M. P. **Perfil de morbidade e gastos hospitalares com idosos no Paraná, Brasil, entre 2008 e 2012.** Cad. Saúde Pública, vol.32, n.7, Rio de Janeiro, 2016.

KITE, G. W. **Frequence and risk analyses in hydrology.** Highlands Ranch, Colorado: Water Resources Publications. 5. ed, p.257, 1988.

KLEFENS, P. C. de O. **O biplot na análise fatorial multivariada.** Dissertação, Escola Superior de Agricultura Luiz de Queiroz, Universidade de São Paulo, Piracicaba, 2009.

KUCHLA, M.; ARAÚJO, M. D. M.; SOARES, A. F.; QUINÁIA, S. P.; FELSNER, M. L. Classificação de Méis Silvestres de Diferentes Mesorregiões do Estado do Paraná, Brasil, pela Análise de Componentes Principais. **Revista Virtual de Química.** v. 7, n. 6, 2015.

KUMAR, V.; CHHABRA, J. K.; KUMAR, D. Performance Evaluation of Distance Metrics in the Clustering Algorithms. **INFOCOMP**, v. 13, n. 1, p. 38-51, 2014. Disponível em: <[www.dcc.ufla.br/infocomp/index.php/INFOCOMP/article/view/21+&cd=2&hl=pt-PT&ct=clnk&gl=br](http://www.dcc.ufla.br/infocomp/index.php/INFOCOMP/article/view/21+&cd=2&hl=pt-PT&ct=clnk&gl=br)>. Acesso em: 03 abr. 2017.

LORENZO-SEVA, U.; TIMMERMAN, M. E.; KIERS, H. A. The hull method for selecting the number of common factors. **Multivariate Behavioral Research**, v. 46, n. 2, p. 340-364, 2011.

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, 2011. Disponível em: <[http://ac.els-cdn.com/S1474706510001427/1-s2.0-S1474706510001427-main.pdf?\\_tid=9bed9608-0fd5-11e7-b446-00000aacb361&acdnat=1490279757\\_db9fad41203628622b566b46b98af636](http://ac.els-cdn.com/S1474706510001427/1-s2.0-S1474706510001427-main.pdf?_tid=9bed9608-0fd5-11e7-b446-00000aacb361&acdnat=1490279757_db9fad41203628622b566b46b98af636)>. Acesso em: 23 mar. 2017.

MAMUN, A. A.; HASHIM, A.; DAOUD, J. I. Regionalization of low flow frequency curves for the Peninsular Malaysia. **Journal of Hydrology**, v.381, p.174-180, 2010.

MANSANO, F. H.; PARRÉ, J. L.; PEREIRA, M. F. Análise fatorial aplicada para a classificação das incubadoras de empresas de base tecnológica do Paraná. Revista Brasileira de **Estudos Regionais e Urbanos (RBERU)**, v. 11, n. 2, p. 133-151, 2017.

MENDONÇA, N.; FERREIRA, V.; KALAS, F. A.; FIGUEIREDO, G.; DIAS, F. L. Estudo da Distribuição Pluviométrica do Rio Grande do Sul Através da Análise de Componentes Principais. 8º Salão Internacional de Ensino, Pesquisa e Extensão. Universidade Federal do Pampa. **Anais.. UNIPAMPA**, 2017.



MEYER, A. da S. **Comparação de coeficientes de similaridade usados em análises de agrupamento com dados de marcadores moleculares dominantes**. Dissertação (Mestrado em Engenharia Agrícola) - Escola Superior de Agricultura Luiz de Queiroz, ESALQ/USP, 2002.

MINGOTI, S. A. **Análise de Dados Através de Métodos de Estatística Multivariada: uma abordagem aplicada**. Belo Horizonte: Editora UFMG, 2005.

MINISTÉRIO DO MEIO AMBIENTE - MMA. **Vazão Ecológica e Vazão Ambiental**. Enfoque Eossistêmico Aplicado à gestão dos recursos hídricos. 41ª Reunião da Câmara Técnica de Controle e Qualidade Ambiental / CONAMA. 2010. Disponível em: <[http://www.mma.gov.br/port/conama/reuniao/dir1306/Apresentacao\\_Vazao+ecologica+vazao+ambiental.pdf](http://www.mma.gov.br/port/conama/reuniao/dir1306/Apresentacao_Vazao+ecologica+vazao+ambiental.pdf)>. Acesso em: 20 mar. 2017.

MINUCCI, L. R. **Regionalização De Série De Vazões Mensais utilizando Redes Neurais Artificiais**. Dissertação (Mestrado em Engenharia de Recursos Hídricos e Ambiental) Programa de Pós-Graduação em Engenharia de Recursos Hídricos e Ambiental, Universidade Federal do Paraná, 2014.

MIRANDA, A. C. R. **Regiões hidrologicamente homogêneas na Amazônia com base nas precipitações mensais**. Tese (Doutorado em Meteorologia) - Pós-Graduação em Meteorologia, Universidade Federal de Viçosa, Minas Gerais, 2016.

MIRANDA, C. A. S. F.; CARDOSO, M. G.; MACHADO, S. M.F.; GOMES, M. S.; ANDRADE, J.; TEXEIRA, M. L. Correlação entre composição química e eficácia antioxidante de óleos essenciais de plantas condimentares por Análise de Agrupamentos Hierárquicos (HCA). **exacta**, Belo Horizonte, v. 7, n.1, p. 65-74, 2014. Disponível em: <<http://revistas2.unibh.br/index.php/dcet/article/view/1223/670>>. Acesso em: 29 set. 2017.

MODENESI, M. V. **Análise de agrupamentos FCM utilizando processamento paralelo**. Dissertação (Mestrado). Pós-Graduação de Engenharia da Universidade Federal do Rio de Janeiro – UFRJ, 2008. Disponível em: <[http://wwwp.coc.ufrj.br/teses/mestrado/inter/2008/Teses/MODENESI\\_MV\\_08\\_t\\_M\\_int.pdf](http://wwwp.coc.ufrj.br/teses/mestrado/inter/2008/Teses/MODENESI_MV_08_t_M_int.pdf)>. Acesso em: 04 mai. 2017.

MORAES, M. R. B. **Modelagem Da Vazão Na Bacia Hidrográfica Do Rio Araguari, em Minas Gerais**. Dissertação (Mestrado em Qualidade Ambiental) - Programa de Pós-Graduação em Qualidade Ambiental, Universidade Federal de Uberlândia, 2016. Disponível em: <<https://repositorio.ufu.br/bitstream/123456789/17864/1/ModelagemVazaoBacia.pdf>>. Acesso em: 15 mar. 2017.

MORAIS, J. T. G. **Análise de componentes principais integrada a redes neurais artificiais para predição de matéria orgânica**. 2011. 68 f. Dissertação (Mestrado em Engenharia Industrial) – Programa de Pós-Graduação em Engenharia Industrial. Universidade Federal da Bahia, Salvador, 2011.

MUÑOZ DIAS, D.; RODRIGO, F. S. Spatio-temporal patterns of seasonal rainfall in Spain (1912-2000) using cluster and principal component analysis: comparison. **Annales Geophysicae**, v. 22, p. 1435-1448, 2004.

MWALE, D.; GAN, T. Y.; DEVITO, K. J.; SILINS, U.; MENDOZA, C.; PETRONE, R. **Regionalization of Runoff Variability of Alberta, Canada, by Wavelet, Independent Component, Empirical Orthogonal Function, and Geographical Information System Analysis**. Journal of Hydrologic Engineering, v. 16. 2010.

NAGHETTINI, M.; PINTO, E. J. A. **Hidrologia Estatística**. Serviço Geológico do Brasil. Belo Horizonte: CPRM, 2007. Acesso em: <[http://www.cprm.gov.br/publique/media/cap10-anal\\_reg.pdf](http://www.cprm.gov.br/publique/media/cap10-anal_reg.pdf)>. Disponível em: 25 fev. 2017.

NASCIMENTO, A. do.; ALMEIDA, R. M. V. R.; CASTILHO, S. R. C.; INFANTOSI, A. F. C. **Análise de correspondência múltipla na avaliação de serviços de farmácia hospitalar no Brasil**. Cadernos de Saúde Pública, vol. 29, n.6, Rio de Janeiro, 2013. Acesso em: <[http://www.scielo.br/scielo.php?script=sci\\_arttext&pid=S0102-311X2013000600013](http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0102-311X2013000600013)>. Disponível em: 23 mar. 2017.

NEVES, R. F. **Análise de agrupamento aplicado nas distâncias das cidades do Agreste Paraibano**. Trabalho de Conclusão de Curso, Universidade Estadual da Paraíba, 2013. Disponível em: <<http://dspace.bc.uepb.edu.br/jspui/bitstream/123456789/4313/1/PDF%20-%20Reginaldo%20Ferreira%20Neves.pdf>>. Acesso em: 23 mar. 2017.

NIMER, E. **Climatologia do Brasil**. Rio de Janeiro: IBGE, p. 422, 1979.

O'LENIC, E. A.; LIVEZEY, R. E. Practical considerations in the use of rotated principal component analysis (RPCA) in diagnostic studies of upper-air height fields, Mon. **Weather Review** v. 116, 1998. Disponível em: <[http://journals.ametsoc.org/doi/pdf/10.1175/1520-0493\(1988\)116%3C1682%3APCITUO%3E2.0.CO%3B2](http://journals.ametsoc.org/doi/pdf/10.1175/1520-0493(1988)116%3C1682%3APCITUO%3E2.0.CO%3B2)>. Acesso em: 12 nov. 2017.

OLIVEIRA, M. R. G. Mapping plaques Cisterns by Fuzzy grouping analysis. **IEEE Latin America Transactions** v. 14, n. 10, 2016. Disponível em: <<http://ieeexplore.ieee.org/abstract/document/7786317/>>. Acesso em: 20 abr. 2017.

PALLANT, J. **SPSS Survival Manual**. Open University Press, 2007.

PALOSCHI, R. A. **Aplicação do Software R na classificação de áreas associadas à produtividade da soja e variáveis agrometeorológicas por meio de agrupamento Fuzzy**. Superior de Tecnologia em Análise e Desenvolvimento de Sistemas – COADS, Universidade Tecnológica Federal do Paraná – UTFPR, Medianeira, Paraná, 2013. Acesso em: <[http://repositorio.roca.utfpr.edu.br/jspui/bitstream/1/2435/1/MD\\_COADS\\_2013\\_1\\_09.pdf](http://repositorio.roca.utfpr.edu.br/jspui/bitstream/1/2435/1/MD_COADS_2013_1_09.pdf)>. Disponível em: 20 abr. 2017.

PANSERA, W. A.; GOMES, B. M.; VILAS BOAS, M. A.; MELLO, E. L. Clustering rainfall stations aiming regional frequency analysis. **International Journal of Food, Agriculture and Environment (Print)** v. 11, p. 877-885, 2013.

PANSERA, W. A.; GOMES, B. M.; VILAS BOAS, M. A.; SAMPAIO, S. C.; MELLO, E. LEMOS; QUEIROZ, M. M. F. Regionalization of Monthly Precipitation Values in the State of Parana (Brazil) by using Multivariate Clustering Algorithms. **Irriga**, Botucatu, v. 20, p. 473-489, 2015.

PARANÁ. Resolução Sema – 024/2006. **Institui as Diretrizes para a Gestão das Bacias Hidrográficas**. Secretaria de Recursos Hídricos e Meio Ambiente – Sema/ PR, 2006. PARANÁ. Disponível em: <[http://celepar7.pr.gov.br/sia/atosnormativos/form\\_cons\\_ato1.asp?Codigo=1355](http://celepar7.pr.gov.br/sia/atosnormativos/form_cons_ato1.asp?Codigo=1355)>. Acesso em: 27 abr. 2017.

PASQUALI, L. **Análise fatorial: um manual teórico-prático**. Brasília: Editora UnB, 1999.

PATIL, V. H.; SINGH, S. N.; MISHRA, S.; DONAVAN, D. T. Efficient theory development and factor retention criteria: Abandon the eigenvalue greater than one criterion. **Journal of Business Research** v. 61, n. 2, p. 162-170, 2008.

PEREIRA, M. A. F.; KOBIYAMA, M.; CASTRO, N. M. R. Análise de homogeneidade da precipitação na bacia hidrográfica do rio Ijuí - RS. In. Simpósio Brasileiro de Recursos Hídricos, ed. 20, 2013. Bento Gonçalves, RS. **Anais...** Porto Alegre: ABRH, 2013. Disponível em: <<https://www.ufrgs.br/gpden/wordpress/wp-content/uploads/2014/08/AN%C3%81LISE-DE-HOMOGENEIDADE-DA-PRECIPITA%C3%87%C3%83O-NA-BACIA-IJU%C3%8D-RS.pdf>>. Acesso em: 23 mar. 2017.

PEREIRA, M. A. F.; KOBIYAMA, M.; CASTRO, N. M. R. Análise de variâncias pluviométricas na bacia hidrográfica do rio Taboão – RS. **Revista Brasileira de Meteorologia**, v. 29, n. 3, São Paulo, 2014. Disponível em: <[http://www.scielo.br/scielo.php?script=sci\\_arttext&pid=S0102-77862014000300009](http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0102-77862014000300009)>. Acesso em: 03 abr. 2017.

PEREIRA, R. C. G.; BRAGA, C. C.; PAZ, R. L. F. **Estudo da Pluviometria no Estado de Goiás**. Aplicando a Técnica de Análise Fatorial em Componentes Principais. SBMET, v.1, Belém-PA, 2010.

PESSOA, F. C. L. **Desenvolvimento de Metodologia para Regionalização de Curvas de Permanência de Vazões na Amazônia Legal**. Tese (Doutorado em Engenharia) - Programa de Pós-Graduação em Engenharia de Recursos Naturais da Amazônia, ITEC, Universidade Federal do Pará – UFP. Belém, 2015.

PIMENTEL, B. A. **Agrupamento de dados simbólicos usando abordagem Possibilistic**. Dissertação - Universidade Federal de Pernambuco, 2013. Disponível em: <<http://repositorio.ufpe.br/bitstream/handle/123456789/11420/BrunoAlmeidaPimentel.pdf?sequence=1&isAllowed=y>>. Acesso em: 17 maio 2017.

PINTO, J. A. O. de. **Avaliação de métodos para a regionalização de curva de permanência de vazões para a bacia do Rio das Velhas**. Dissertação (Mestrado em Engenharia) - Escola de Engenharia Universidade Federal de Minas Gerais, 2006. Acesso em: <<http://www.smarh.eng.ufmg.br/defesas/209M.PDF>>. Disponível em: 21 mar. 2017.

PRADO, B. Q. M. **Análise de Agrupamentos das Taxas de Incidência de Dengue nos Estados Brasileiros**. Trabalho de Conclusão de Curso (Especialização em Matemática) - Faculdade de Matemática, Universidade Federal de Uberlândia, 2015. Disponível em: <[http://www.famat.ufu.br/sites/famat.ufu.br/files/Anexos/Bookpage/TCC\\_BRUNA.pdf](http://www.famat.ufu.br/sites/famat.ufu.br/files/Anexos/Bookpage/TCC_BRUNA.pdf)>. Acesso em: 03 abr. 2017.

PRUSKI, F. F.; NUNES, A. D. A.; PRUSKI, P. L.; RODRIGUEZ, R. D. 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. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0022169412008797>>. Acesso em: 23 mar. 2017.

PRUSKI, F. F.; RODRIGUEZ, R. G.; NUNCES, A. A.; PRUSKI, P. L.; SINGH, V. P. Estimativa de vazões mínimas em regiões de extrapolação das equações de regionalização: um novo conceito. **Revista Engenharia Agrícola** v. 35, n. 5, 2015. Acesso em: <<http://submission.scielo.br/index.php/eagri/article/view/136995>>. Disponível em: 13 mar. 2017.

PRUSKI, F.F.; PRUSKI, P.L. Tecnologia e inovação frente a gestão de recurso 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, p. 27-58. 2011. Disponível em: <<https://pt.slideshare.net/MCTI/recursos-hdricos-em-regies-ridas-e-semiridas>>. Acesso em: 13 mar. 2017.

PRUSKI, F.F.; SILVA, D.D.; KOEZ, M. **Estudo da Vazão em Cursos d'Água**. Engenharia na Agricultura. Caderno didático: 43. Associação de Engenharia Agrícola, Universidade Federal de Viçosa, p. 152, 2006.

QUINTAL, G. M. C. C. **Análise de clusters aplicada ao Sucesso/Insucesso em Matemática**. 2006. 184 f. Dissertação (Mestrado em Matemática para o Ensino) - Departamento de Matemática e Engenharias, Universidade da Madeira, 2006. Disponível em: <<http://repositorio.uma.pt/bitstream/10400.13/224/1/GuidaCaldeiraMestrado.pdf>>. Acesso em: 06 abr. 2017.

RAJSEKHAR, D.; MISHRA, A; SINGH, V. P. **Drought Regionalization of Brazos River Basin Using an Entropy Approach**. Symposium on Data-Driven Approaches to Droughts. Proceedings, 2011.

RAO, A. R.; SRINIVAS, V. V. Regionalization of watersheds by fuzzy cluster analysis. **Journal of Hydrology** v. 318, n. 1-4, p. 57-79, 2006.

REGO, F. S. **Regionalização de vazões considerando a evapotranspiração real em seu processo de formação**. 2013. 62 f. Dissertação (Mestrado em Engenharia Agrícola) - Universidade Federal de Viçosa, 2013. Acesso em: <[http://www.cprm.gov.br/publique/media/diss\\_fernando\\_rego.pdf](http://www.cprm.gov.br/publique/media/diss_fernando_rego.pdf)>. Disponível em: 15 mar. 2017.

REIS, J. A. T.; CALADO, M. A. C.; BARBOSA, J. F.; MOSCON, M.; MENDONÇA, A. S. F. Análise Regional de Vazão Mínima de Referência na Região Centro-Sul do Estado do Espírito Santo. **Revista CIATEC, UPF** v. 5, n. 2, p.1-11, 2013.

RHODEN, A. C.; FELDMANN, N. A.; MUHL, F. R.; RITTER, A. F. S.; MOREIRA, A. A. Importância da Água e da Gestão dos Recursos Hídricos. **Revista Ciências Agroveterinárias e Alimentos** n. 1, 2016. Disponível em: <<http://revista.faifaculdades.edu.br:8080/index.php/cava/article/view/196>>. Acesso em: 13 mar. 2017.

RIGNEL, D. G. S.; CHENCI, G. P.; LUCAS, C. A. Uma introdução a lógica fuzzy. **Revista Eletrônica de Sistemas de Informação e Gestão Tecnológica** v. 01, n. 01, 2011.

RUSPINI, E. H. A new approach to clustering. **Information and Control**, Elsevier, v. 15, p.22-32, 1969. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0019995869905919>>. Acesso em: 17 maio 2017.

SADRI, S.; BURN, D. H. A Fuzzy C-Means approach for regionalization using a bivariate homogeneity and discordancy approach. **Journal of Hydrology** v. 401, p. 231-239. 2011. Acesso em: <[http://ac.els-cdn.com/S0022169411001429/1-s2.0-S0022169411001429-main.pdf?\\_tid=2695db68-252e-11e7-9e89-00000aacb35f&acdnat=1492626760\\_867e303ea08da2bbca8261f559df79cf](http://ac.els-cdn.com/S0022169411001429/1-s2.0-S0022169411001429-main.pdf?_tid=2695db68-252e-11e7-9e89-00000aacb35f&acdnat=1492626760_867e303ea08da2bbca8261f559df79cf)>. Disponível em: 17 mar. 2017.

SAMUEL, J.; COULIBALY, P.; METCALFE, R. A. Estimation of Continuous Streamflow in Ontario Ungauged Basins: Comparison of Regionalization Methods. **Journal of Hydrologic Engineering** v. 16, 2011.

SANTOS, L. J. C.; FIORI, C. O.; CANALI, N. E.; FIORI, A. P.; SILVEIRA, C. T.; SILVA, J. M. F.; ROSS, J. L. S. Mapeamento Geomorfológico do Paraná. *União da Geomorfologia Brasileira. Revista Brasileira de Geomorfologia*, v. 7, n. 2, p. 03-12, 2006. Disponível em: <[http://www.ugb.org.br/final/normal/artigos/SEPARATAS\\_RBG\\_Ano\\_7\\_n\\_2\\_2006/RBG\\_Ano\\_7\\_n\\_2\\_2006\\_03\\_12.pdf](http://www.ugb.org.br/final/normal/artigos/SEPARATAS_RBG_Ano_7_n_2_2006/RBG_Ano_7_n_2_2006_03_12.pdf)>. Acesso em: 27 abr. 2017.

SANTOS, S. H. L. dos; LOPES, A. A. S.; CHAGAS, P. F.; SOUZA, R. O. Análise de risco na concessão de outorga de lançamentos difusos de poluentes em rios, através de um modelo Fuzzy de transporte de massa. *Revista AIDIS de Ingeniería y Ciencias Ambientales* v. 9, p. 157, 2016.

SATO, L. Y.; SHIMABUKURO, Y. E.; KUPLICH, T. M.; GOMES, V. C. F. **Análise comparativa de algoritmos de árvore de decisão do sistema WEKA para classificação do uso e cobertura da terra.** Anais, XVI Simpósio Brasileiro de Sensoriamento Remoto - SBSR, INPE, Foz do Iguaçu, PR, 2013.

SECRETARIA DE ESTADO DO MEIO AMBIENTE E RECURSOS HÍDRICOS - SEMA. **Bacias Hidrográficas do Paraná. Série Histórica.** Curitiba - Paraná, 2010. Disponível em: <[http://www.meioambiente.pr.gov.br/arquivos/File/corh/Revista\\_Bacias\\_Hidrograficas\\_do\\_Parana.p](http://www.meioambiente.pr.gov.br/arquivos/File/corh/Revista_Bacias_Hidrograficas_do_Parana.p)>. Acesso em: 27 abr. 2017.

SILVA, G. N. S. **Regionalização automatizada de parâmetros de modelos chuva-vazão integrada a um sistema de informações geográficas.** 2012. 135f. Dissertação (Mestrado em Engenharia Urbana e Ambiental) - Universidade Federal da Paraíba, João Pessoa, 2012.

SILVA, W. L.; CHANG, M.; DEREZYNSKI, C. P.; FREITAS, M. Tendências observadas em indicadores de extremos climáticos de temperatura e precipitação no estado do Paraná. *Revista Brasileira de Meteorologia* v. 30, n. 2, p. 181-194, 2015.

SINGH, V. P. **Entropy Theory and its Application in Environmental and Water Engineering.** Chichester, UK, John Wiley & Sons, Ltd, 2013.

SKUPIEN, S. V.; CARVALHO, D. R. Data Mining on Nursing Consultation for the Postpartum Period. *Iberoamerican Journal of Applied Computing* v.2, p. 32-42, 2012. Disponível em: <<http://www.revistas2.uepg.br/index.php/ijac/article/view/5362>>. Acesso em: 17 maio 2017.

SOUSA, H. T. de. **Sistema computacional para regionalização de Vazões.** 2009. 86 f. Dissertação (Mestrado em Engenharia Agrícola) - Universidade Federal de Viçosa, Viçosa, MG, 2009.

SOUSA, H. T.; PRUSKI, F. F.; BOF, L. H. N.; SOUZA, R. C. **SisCAH 1.0:** Sistema computacional para análises hidrológicas. Brasília: Agência Nacional de Águas, Viçosa: Universidade Federal de Viçosa, 2009.

SOUZA, K. B. de. **Influência uso e ocupação do solo na vazão da Bacia do Rio Uruçuí-Preto, Piauí.** 2015. 80 f. Dissertação (Mestrado em Agronomia) - Programa de Pós-Graduação em Solos e Nutrição de Plantas, Universidade Federal do Piauí, Campus Professora Cinobelina Elvas Bom Jesus, PI, 2015. Disponível em: <<http://repositorio.ufpi.br/xmlui/bitstream/handle/123456789/239/Disserta%C3%A7%C3%A3o.pdf?sequence=1>>. Acesso em: 17 mar. 2017.

SOUZA, M. D. O.; RAINHA, K. P.; CASTRO, E. V. R.; CARNEIRO, M.; FERREIRA, R. D. Q. Análise exploratória das concentrações dos metais Na, Ca, Mg, Sr e Fe em extrato aquoso de petróleo, determinados por ICP OES, após otimização empregando planejamento de experimentos. *Química Nova* v. 38, n. 7, p. 980-986, 2015.

SRINIVAS, V. V.; TRIPATHI, S.; RAO, A. R.; GOVINDARAJU, R. S. Regional flood frequency analysis by combining self-organizing feature map and fuzzy clustering. **Journal of hydrology** v. 348, p. 148-166, 2008. Acesso em: <<http://www.sciencedirect.com/science/article/pii/S0022169407005537>>. Disponível em: 17 mar. 2017

SUPERINTENDÊNCIA DE DESENVOLVIMENTO DE RECURSOS HÍDRICOS E SANEAMENTO AMBIENTAL – SUDERHSA. **Bacias Hidrográficas do Paraná**. Paraná, 2006. Disponível em: <<http://www.aguasparana.pr.gov.br/modules/conteudo/conteudo.php?conteudo=82>>. Acesso em: 27 abr. 2017.

TABACHNICK, B. G.; FIDELL, L. S. **Using Multivariate Statistics**. 5 ed. Boston: Allyn and Bacon, 2007.

TONGAL, H.; DEMIREN, M. C.; BOOIJ, M. J. Seasonality of low flows and dominant processes in the Rhine River. **Stochastic Environmental Research and Risk Assessment** v.27, p.489-503, 2013.

TUCCI, C. E. M. **Regionalização de vazões**. Porto Alegre: UFRGS, vol. 1, 1 ed., p. 256, 2002.

TUCCI, C. E. M.; CLARKE, R.T. Regionalização hidrológica. In: PAIVA, J. B. D; PAIVA, E. M. C. D. **Hidrologia aplicada à gestão de pequenas bacias hidrográficas**. Porto Alegre, ABRH, Cap. 7, p. 169-222, 2001.

TURAN, M. E.; YURDUSEV, M. A. River flow estimation from upstream flow records by artificial intelligence methods. **Journal of Hydrology** v. 369, p. 71-77. 2009. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0022169409000948>>. Acesso em: 17 mar. 2017.

UNITED STATES GEOLOGICAL SURVEY – USGS. **Double-Mass Curves**. Manual of Hydrology: Part I. General Surface-Water Techniques, by James K. Searcy and Clayton H. Hardison, United States Department of the Interior, 1966.

VALERIANO, M. M. **Modelo digital de elevação com dados SRTM disponíveis para a América do Sul**. Ministério da Ciência e Tecnologia. Instituto Nacional de Pesquisas Espaciais - INPE, 2004.

VARGAS, M. M.; BESKOW, S.; CORRÊA, L. L. DURAES, M. F.; CALDEIRA, T. L.; MELLO, C. R.; AGUIAR, M. S. Técnicas de inteligência artificial para regionalização hidrológica: uma análise da Q90 no Rio Grande do Sul. XXI Simpósio Brasileiro de Recursos Hídricos. **Anais...** Brasília, DF, 2015.

VENDRUSCOLO, J. **Regionalização dos parâmetros de um modelo chuva-vazão para estimativa de vazões em bacias sem monitoramento sistemático**. 2005. 93 f. Dissertação (Mestrado em Engenharia Civil). Curso de Engenharia Civil, Recursos Hídricos e Saneamento Ambiental, Universidade Federal de Santa Maria, Santa Maria, 2005.

VIDAL, B. C. **Avaliação de Agrupamento em mistura de variáveis**. Dissertação (Mestrado em Estatística Aplicada e Biometria). 69 f. 2013. Universidade Federal de Viçosa, 2013. Disponível em: <<http://www.locus.ufv.br/bitstream/handle/123456789/4065/texto%20completo.pdf?sequence=1>>. Acesso em: 23 mar. 2017.

VOGEL, R. M. Regional calibration of watershed models. In: P. SINGH, V. P.; FREVERT, D. K. **Watershed Models**. Capítulo 3. Colorado: ed. CRC Press, 2005. Acesso em: <<https://engineering.tufts.edu/cee/people/vogel/documents/regional-calibration-chap3.pdf>>. Disponível em: 17 mar. 2017.

WAGNER, M. M.; RIBEIRO, V. G.; ZABADAL, J. Mineração de dados meteorológicos empregando dados de temperatura: o caso de uma cidade gaúcha. **Interfaces Científicas - Exatas e Tecnológicas**, Aracaju, v. 2, n. 2, p. 57-70, 2016.

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

WOLFF, W. **Avaliação e nova proposta de regionalização hidrológica para o Estado de São Paulo**. Dissertação (Mestrado em Engenharia Agrícola). Escola Superior de Agricultura "Luiz de Queiroz", 2013.

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

YANG, Y.; BURN, D. An entropy approach to data collection network design. **Journal of Hydrology** v. 157, n. 1-4, 1994. Disponível em: <<https://www.sciencedirect.com/science/article/pii/0022169494901112>>. Acesso em: 25 nov. 2017.

YONAMINE, F. S.; SPECIA, L.; CARVALHO, V. O.; NICOLETTI, M. C. **Aprendizado não supervisionado em domínios fuzzy Algoritmo Fuzzy C-Means**. eBook, São Carlos: UFSCAR, 2002.

YU, L.; WANG, S.; LAI, K. K. An Integrated Data Preparation Scheme for Neural Network Data Analysis. **IEE transactions on knowledge and data engineering** v. 18, p. 217-230. 2006.

ZADEH, L. A. Fuzzy sets. **Information and Control** v. 8, n. 3, p. 338-353, 1965. Disponível em: <[http://dx.doi.org/10.1016/S0019-9958\(65\)90241-X](http://dx.doi.org/10.1016/S0019-9958(65)90241-X)>. Acesso em: 17 maio 2017.

## APÊNDICES



**APÊNDICE A ESTAÇÕES FLUVIOMÉTRICAS DAS BACIAS HIDROGRÁFICAS DO ESTADO DO PARANÁ.**

Bacia Hidrográfica	Código	Latitude	Longitude
Itararé	64230500	-24,10	-49,47
	64240000	-24,26	-49,73
	64242000	-23,98	-49,60
Cinzas	64360000	-23,77	-49,95
	64365850	-23,50	-49,89
	64380000	-23,85	-50,39
Tibagi	64430200	-25,32	-49,98
	64441020	-25,46	-50,53
	64442800	-25,20	-50,52
	64453000	-24,83	-50,28
	64460000	-24,71	-50,49
	64477600	-24,75	-50,09
	64482000	-24,36	-50,59
	64491000	-24,03	-50,69
	64491260	-24,01	-50,70
	64494950	-23,75	-50,98
	64497000	-23,75	-50,94
	64498550	-23,70	-50,92
	64500000	-23,72	-50,78
	64501950	-23,55	-50,88
	64502000	-23,56	-51,03
	64504100	-23,44	-50,82
	64504450	-23,47	-51,31
	64504550	-23,40	-51,01
	64504581	-23,34	-51,30
	64506000	-23,31	-51,00
	64507000	-23,25	-50,98
	64507100	-23,25	-51,07
	64508020	-23,23	-50,73
64508500	-23,17	-50,79	
64529900	-23,50	-51,47	
64541000	-23,33	-51,85	
64544000	-23,19	-52,15	
64546900	-23,03	-51,75	
Pirapó	64625000	-24,83	-51,14
	64652000	-24,11	-51,48
	64550000	-22,86	-52,08

Ivaí	64659000	-24,02	-51,95
	64645000	-24,34	-51,42
	64655000	-24,04	-51,62
	64660500	-23,90	-51,95
	64670012	-24,09	-52,32
	646 71000	-24,03	-52,27
	64671950	-24,07	-52,42
	64673000	-23,82	-52,18
	64675002	-23,67	-52,12
	64682000	-23,44	-52,60
	64685000	-23,32	-52,67
	64693000	-23,20	-53,32
Piquiri	64773000	-24,78	-52,23
	64776100	-24,75	-52,88
	64785000	-24,62	-52,93
	64790000	-24,63	-53,10
	64795000	-24,52	-53,17
	64799500	-24,40	-53,16
	64810000	-23,92	-53,13
	64815000	-24,13	-53,32
	64820000	-24,20	-53,33
	64830000	-24,19	-53,75
64833000	-23,97	-53,75	
Paraná 3	64863360	-24,74	-53,69
	64863520	-24,75	-53,77
	64892500	-24,96	-54,18
Ribeira	81019350	-25,23	-49,60
	81080000	-25,07	-49,59
	81102000	-24,93	-49,48
	81120000	-25,01	-49,34
	81125000	-24,75	-49,33
	81290500	-25,27	-49,17
	81291000	-25,26	-49,15
	81299000	-25,22	-48,95
81303000	-25,09	-48,83	
Litorânea	82002000	-25,16	-48,24
	82121003	-25,32	-48,71
	82195002	-25,51	-48,88
	82198000	-25,54	-48,85
	81107000	-24,90	-49,44
	64362000	-23,34	-50,16
	64370000	-23,09	-50,29
	64447000	-24,95	-50,39
	64465000	-24,51	-50,41
	64504591	-23,35	-51,20
	64390000	-23,12	-50,45

	64767000	-24,95	-52,60
	64771500	-24,87	-52,76
	65003950	-25,44	-49,12
	65004995	-25,45	-49,12
	65006075	-25,45	-49,17
	65007045	-25,44	-49,20
	65007046	-25,45	-49,20
	65009000	-25,48	-49,19
	65010000	-25,52	-49,15
	65013005	-25,53	-49,22
	65015400	-25,59	-49,23
Iguaçu	65017006	-25,60	-49,26
	65017035	-25,65	-49,26
	65019400	-25,72	-49,32
	65021770	-25,42	-49,38
	65024000	-25,57	-49,43
	65135000	-25,93	-49,39
	65220000	-26,02	-50,59
	65365000	-26,18	-51,52
	65809000	-25,40	-51,44
	65815050	-25,54	-51,82
	65874990	-25,86	-52,55
	65925000	-26,28	-52,30
	65925770	-26,31	-52,62
	65925780	-26,30	-52,66
	65925800	-26,28	-52,59
	65970001	-25,39	-53,10
	65971010	-25,38	-53,28
	65974500	-25,77	-53,52
	65976001	-25,03	-53,46
	65981500	-25,77	-53,61

## APÊNDICE B SCRIPT DA ROTINA PARA ACP

Matriz de entrada ("MatrizParaACP")

Código	Área de Drenagem (Km <sup>2</sup> )	Soma das drenagens (m)	Densidad e da drenagem (1/Km)	Qmld (m <sup>3</sup> /s)	Q7,10 (m <sup>3</sup> /s)	Q90% (m <sup>3</sup> /s)	Q95% (m <sup>3</sup> /s)	Latitude	Longitude
a1	684,32	165744,59	0,24	20,51	1,88	5,14	3,97	-24,10	-49,47
a2	958,57	258916,96	0,27	3,55	0,34	0,86	0,60	-24,26	-49,73
a3	1672,38	454861,57	0,27	30,64	7,71	13,23	11,61	-23,98	-49,60
a4	2002,74	527597,01	0,26	28,22	3,64	7,49	6,07	-23,77	-49,95
a5	3957,92	1075673,26	0,27	62,07	8,74	16,48	13,31	-23,34	-50,16
a6	121,17	28049,79	0,23	7,73	0,11	0,27	0,21	-23,50	-49,89
a7	5597,01	1518641,78	0,27	63,07	8,03	14,30	11,47	-23,09	-50,29
a8	1056,30	299944,59	0,28	21,59	0,97	2,27	0,65	-23,85	-50,39
a9	3422,11	944727,37	0,28	38,81	3,47	6,92	5,59	-23,12	-50,45
a10	395,63	95687,80	0,24	36,66	2,24	4,09	3,21	-25,32	-49,98
a11	218,03	52404,60	0,24	4,13	0,40	0,91	0,69	-25,46	-50,53
a12	1315,37	308161,87	0,23	27,08	2,94	5,98	4,88	-25,20	-50,52
a13	5647,79	1382213,76	0,24	124,14	10,46	30,37	23,17	-24,95	-50,39
a14	1038,14	260249,28	0,25	25,73	4,41	8,43	6,66	-24,83	-50,28
a15	740,11	189622,75	0,26	12,00	1,46	3,09	2,37	-24,71	-50,49
a16	8759,91	2198487,42	0,25	159,50	21,28	43,46	34,50	-24,51	-50,41
a17	1205,61	301517,15	0,25	36,86	3,99	9,98	7,99	-24,75	-50,09
a18	13705,10	3478406,15	0,25	308,77	42,13	94,05	78,86	-24,36	-50,59
a19	15994,30	4109190,78	0,26	236,40	40,32	67,79	60,45	-24,03	-50,69
a20	362,34	80159,74	0,22	10,42	0,25	0,68	0,48	-24,01	-50,70
a21	493,85	139388,65	0,28	11,85	2,82	4,08	3,48	-23,75	-50,98
a22	532,37	146686,06	0,28	14,58	1,42	5,66	4,46	-23,75	-50,94
a23	18291,30	4714590,54	0,26	376,45	54,14	126,48	98,54	-23,70	-50,92
a24	59,13	15080,40	0,26	3,75	0,48	0,64	0,56	-23,72	-50,78
a25	565,12	174337,51	0,31	9,43	1,05	2,75	2,19	-23,55	-50,88
a26	748,96	241237,57	0,32	37,96	3,61	8,04	6,42	-23,56	-51,03
a27	29,64	4408,68	0,15	3,99	0,08	0,16	0,11	-23,44	-50,82
a28	112,70	26995,51	0,24	3,93	0,66	0,95	0,82	-23,47	-51,31
a29	326,87	95183,87	0,29	5,80	1,08	2,04	1,65	-23,40	-51,01
a30	47,58	11591,81	0,24	0,83	0,37	0,50	0,45	-23,34	-51,30
a31	133,54	40316,11	0,30	2,78	0,68	1,17	0,97	-23,35	-51,20
a32	21723,00	5751219,81	0,26	490,45	67,62	147,53	116,55	-23,31	-51,00

a33 21777,30 5764306,66 0,26 364,88 49,81 92,21 71,40 -23,25 -50,98

### Continuação Apêndice B

Código	Área de Drenagem (Km <sup>2</sup> )	Soma das drenagens (m)	Densidad e da drenagem (1/Km)	Qmld (m <sup>3</sup> /s)	Q7,10 (m <sup>3</sup> /s)	Q90% (m <sup>3</sup> /s)	Q95% (m <sup>3</sup> /s)	Latitude	Longitude
a34	180,98	53118,47	0,29	10,35	0,28	0,93	0,60	-23,25	-51,07
a35	941,24	264236,86	0,28	13,35	3,15	4,11	3,53	-23,23	-50,73
a36	1047,39	291826,00	0,28	15,80	2,03	4,88	3,91	-23,17	-50,79
a37	20,94	2996,43	0,14	0,58	0,13	0,29	0,22	-23,50	-51,47
a38	1229,26	312416,87	0,25	150,41	1,66	9,02	5,59	-23,33	-51,85
a39	24,23	3033,61	0,13	0,99	0,09	0,29	0,21	-23,19	-52,15
a40	1008,65	239377,82	0,24	18,84	3,97	6,84	6,16	-23,03	-51,75
a41	4465,22	1105335,18	0,25	71,55	21,67	34,67	29,21	-22,86	-52,08
a42	3554,57	939671,74	0,26	75,10	5,54	11,30	8,95	-24,83	-51,14
a43	8521,49	2369531,51	0,28	1188,53	19,08	52,18	34,15	-24,34	-51,42
a44	2603,09	718681,16	0,28	52,60	3,16	7,18	4,66	-24,11	-51,48
a45	12789,05	3573296,20	0,28	280,14	19,99	53,90	38,34	-24,04	-51,62
a46	3275,32	1019798,43	0,31	77,22	2,56	8,09	5,76	-24,02	-51,95
a47	19391,32	5555648,27	0,29	505,18	34,70	87,31	65,80	-23,90	-51,95
a48	571,08	159744,35	0,28	13,15	0,02	0,54	0,00	-24,09	-52,32
a49	849,30	249997,27	0,29	20,59	2,63	6,99	5,74	-24,03	-52,27
a50	77,37	15404,58	0,20	1,69	0,43	0,73	0,64	-24,07	-52,42
a51	1529,68	435682,82	0,28	34,40	7,17	14,51	11,99	-23,82	-52,18
a52	22446,60	6430884,58	0,29	1291,29	55,84	128,10	100,93	-23,67	-52,12
a53	807,98	171854,30	0,21	16,95	6,47	9,54	8,78	-23,44	-52,60
a54	28345,64	7914851,59	0,28	542,99	73,91	154,20	126,58	-23,32	-52,67
a55	34319,74	9173727,19	0,27	674,01	164,3	246,57	220,27	-23,20	-53,32
a56	3526,45	1158169,06	0,33	109,55	7,28	16,49	12,61	-24,95	-52,60
a57	4151,16	1361723,79	0,33	120,41	6,77	19,01	12,51	-24,87	-52,76
a58	759,46	239634,94	0,32	29,59	0,45	2,13	1,36	-24,78	-52,23
a59	7635,83	2502672,68	0,33	219,35	17,70	36,31	28,25	-24,75	-52,88
a60	1310,59	386235,96	0,29	29,62	8,34	12,82	9,48	-24,62	-52,93
a61	603,27	186855,69	0,31	17,79	3,33	5,76	4,82	-24,63	-53,10
a62	11219,45	3580047,80	0,32	298,95	26,34	62,17	46,72	-24,52	-53,17
a63	12060,63	3815414,27	0,32	350,62	27,34	71,65	53,33	-24,40	-53,16
a64	2038,41	450449,28	0,22	45,39	15,59	21,27	19,48	-23,92	-53,13
a65	2962,13	656477,29	0,22	61,44	19,71	30,68	26,68	-24,13	-53,32
a66	17400,82	5114924,01	0,29	412,49	71,25	128,30	107,95	-24,20	-53,33
a67	20923,81	6080735,72	0,29	506,95	90,58	170,79	140,47	-24,19	-53,75
a68	1058,82	208021,54	0,20	19,01	8,99	11,44	10,98	-23,97	-53,75

a69 60,18 13109,23 0,22 2,33 0,83 1,03 1,01 -24,74 -53,69

### Continuação Apêndice B

Código	Área de Drenagem (Km <sup>2</sup> )	Soma das drenagens (m)	Densidade da drenagem (1/Km)	Qmld (m <sup>3</sup> /s)	Q7,10 (m <sup>3</sup> /s)	Q90% (m <sup>3</sup> /s)	Q95% (m <sup>3</sup> /s)	Latitude	Longitude
a70	92,29	24417,97	0,26	4,06	0,80	1,07	0,99	-24,75	-53,77
a71	554,32	149558,30	0,27	45,25	0,82	2,57	1,81	-24,96	-54,18
a72	164,10	33777,18	0,21	2,77	0,34	0,66	0,55	-25,44	-49,12
a73	101,26	24865,53	0,25	6,84	0,81	1,10	0,99	-25,45	-49,12
a74	378,46	88410,71	0,23	7,38	0,63	1,82	0,84	-25,45	-49,17
a75	110,03	27921,32	0,25	2,25	0,48	0,81	0,69	-25,44	-49,20
a76	111,22	28544,47	0,26	24,82	0,00	1,05	0,61	-25,45	-49,20
a77	560,43	142164,95	0,25	14,35	1,06	2,14	1,49	-25,48	-49,19
a78	105,06	22924,02	0,22	2,99	0,66	1,10	0,93	-25,52	-49,15
a79	810,19	196791,01	0,24	17,33	1,66	3,62	2,61	-25,53	-49,22
a80	271,60	63329,46	0,23	6,04	0,94	1,79	1,51	-25,59	-49,23
a81	1152,60	280737,69	0,24	68,53	4,59	7,36	6,23	-25,60	-49,26
a82	67,47	13483,70	0,20	11,98	0,02	0,33	0,22	-25,65	-49,26
a83	45,66	8278,96	0,18	2,60	0,24	0,37	0,30	-25,72	-49,32
a84	24,81	7498,70	0,30	0,42	0,08	0,11	0,09	-25,42	-49,38
a85	180,09	52693,30	0,29	2,73	0,24	0,64	0,45	-25,57	-49,43
a86	3618,48	906548,56	0,25	11,55	2,57	4,17	3,48	-25,93	-49,39
a87	18433,31	3213852,27	0,17	383,59	55,18	107,31	85,09	-26,02	-50,59
a88	65,59	13336,05	0,20	4,65	0,40	1,04	0,76	-26,18	-51,52
a89	312,01	83670,70	0,27	9,38	0,86	2,03	1,61	-25,40	-51,44
a90	2215,04	574672,22	0,26	69,64	7,07	16,98	12,48	-25,54	-51,82
a91	37,83	7983,80	0,21	0,83	0,02	0,10	0,09	-25,86	-52,55
a92	120,00	29302,13	0,24	55,46	4,53	12,95	9,73	-26,28	-52,30
a93	29,90	8417,85	0,28	0,59	0,06	0,11	0,10	-26,31	-52,62
a94	12,25	100,99	0,01	1,07	0,09	0,11	0,11	-26,30	-52,66
a95	150,13	40772,89	0,27	9,44	0,23	0,80	0,55	-26,28	-52,59
a96	1080,22	330662,91	0,31	30,27	0,03	0,93	0,54	-25,39	-53,10
a97	385,35	93952,49	0,24	15,84	1,01	2,43	1,94	-25,38	-53,28
a98	114,39	33048,12	0,29	2,72	0,20	0,70	0,56	-25,77	-53,52
a99	50,46	7670,22	0,15	1,22	0,09	0,22	0,14	-25,03	-53,46
a100	1720,98	528081,70	0,31	60,03	0,15	2,26	1,16	-25,77	-53,61
a101	537,80	124412,26	0,23	11,08	2,83	4,57	3,87	-25,23	-49,60
a102	1320,84	333328,29	0,25	28,47	7,87	11,34	9,95	-25,07	-49,59
a103	1647,57	417383,81	0,25	30,47	9,47	13,56	11,87	-24,93	-49,48
a104	3188,70	839377,26	0,26	63,72	20,01	28,66	25,12	-24,90	-49,44

a105 417,43 101447,13 0,24 7,06 2,85 3,78 3,25 -25,01 -49,34

### Continuação Apêndice B

Código	Área de Drenage m (Km <sup>2</sup> )	Soma das drenagen s (m)	Densidad e da drenagem (1/Km)	Qmld (m <sup>3</sup> /s)	Q7,10 (m <sup>3</sup> /s)	Q90% (m <sup>3</sup> /s)	Q95% (m <sup>3</sup> /s)	Latitude	Longitude
a106	403,75	109878,04	0,27	7,58	2,81	3,74	3,42	-24,75	-49,33
a107	15,61	2104,65	0,13	0,41	0,02	0,08	0,08	-25,27	-49,17
a108	125,05	28543,28	0,23	8,02	0,89	1,15	1,02	-25,26	-49,15
a109	454,24	102390,82	0,23	13,68	3,65	5,63	4,80	-25,22	-48,95
a110	1067,50	263798,12	0,25	4,06	0,73	1,21	1,02	-25,09	-48,83
a111	91,88	20844,88	0,23	6,06	1,61	2,22	1,94	-25,16	-48,24
a112	378,41	96922,48	0,26	42,74	10,99	20,14	16,53	-25,32	-48,71
a113	77,16	16913,00	0,22	5,18	0,54	1,10	1,00	-25,51	-48,88
a114	58,24	11544,62	0,20	15,47	0,06	0,84	0,62	-25,54	-48,85

```
dados <- read.table("MatrizParaACP.txt", header = T)
dados
```

```
dados1 <- dados[2:10]
```

```
dados1
names(dados1)
```

```
is.data.frame(dados1)
```

```
dim(dados1)
```

```
colMeans(dados1)
```

```
S <- cov(dados1,dados1)
S
```

```
R <- cor(dados1,dados1)
R
```

```
cor.test(AD,SD)
cor.test(AD,DD)
cor.test(DD,SD)
```

### Continuação Apêndice B

```
## TESTE DE ESFERICIDADE DE BARTLETT (matriz R =/ da identidade)
n <- nrow(dados1)
p <- ncol(dados1)
chi2 <- -(n-1-((2*p+5)/6))*log(det(R))
ddl <- p*(p-1)/2
print(chi2)
print(ddl)
```

```

print(pchisq(chi2,ddl,lower.tail=F))

require(psych)
cortest.bartlett(R,n) ## teste de esfericidade de bartlett

require(psych)
KMO(R) ## Índice KMO pelo pacote psych ##

## COMANDO ESPECÍFICO DE COMPONENTES PRINCIPAIS ##

## USANDO A MATRIZ S ##

eigen(S)
cp <- prcomp(dados1) ## cria os componentes principais usando S ##
cp

cp <- prcomp(dados1, scale = T) ## cria os componentes principais usando R ##
cp

##pela regra de kaiser usa-se somente os coeficientes principais que forem maiores de 1.##

summary(cp) ## desvio padrão, proporção e proporção acumulada ##

screeplot(cp)
screeplot(cp, type = "lines") ## gráfico de cotovelo ##

names(cp)

cp$sdev ## desvio padrão dos CP's: raiz quadrada autovalores ##
cp$rotation ## coeficientes cada componente principal: autovetores ##
cp$center ## coordenada central: média amostral ##

cp$rotation[,1] ## coeficientes do 1º CP ##

score <- t(cp$rotation[,1]) %*% t(dados1)
score ## score para cada indivíduo no CP1 ##

cp$x ## scores com variáveis centradas ##

cbind(1:114,as.vector(score))

par(mfrow = c(1,2))

```

### Continuação Apêndice B

```

## no comando biplot podemos adicionar o subcomando choices = c(1,2), no qual escolhe
quais cps quer plotar ##

biplot(cp) ## gráfico biplot ##
## os vetores estão paralelos aos cp que melhor se ajustam. e o valor também é levado em
consideração

cor(as.vector(score), dados1)

```



```
require(graphics)

## USANDO A MATRIZ R ##

cp2 <- princomp(dados1, cor = T) ## usando a matriz de correlação ##
cp2

summary(cp2)

screepLOT(cp2, type = "lines")

names(cp2)
cp2$loa      ## coeficientes de cada componente principal ##
cp2$score    ## scores com variáveis centradas ##

cor(cp2$score[,1], AD)

par(mfrow = c(1,2))
plot(cp2$sco[,1], cp2$sco[,2])
biplot(cp2)
par(mfrow = c(1,2))
biplot(cp2, choices=1:2, scale=0.2, pc.biplot=FALSE, xlab="Comp.1", ylab="Comp.2")
biplot(cp2, choices=3:4, scale=0.2, pc.biplot=FALSE, xlab="Comp.3", ylab="Comp.4")
```

### APÊNDICE C MATRIZ DE ENTRADA COM AS VARIÁVEIS DEPENDENTES

Bacia Hidrográfica	Código	Qmed (m³/s)	Q7,10 (m³/s)	Q90% (m³/s)	Q95% (m³/s)
Itararé	64230500	20,51	1,88	5,14	3,97
	64240000	3,55	0,34	0,86	0,60
	64242000	30,64	7,71	13,23	11,61
Cinzas	64360000	28,22	3,64	7,49	6,07
	64365850	7,73	0,11	0,27	0,21
	64380000	21,59	0,97	2,27	0,65
Tibagi	64430200	36,66	2,24	4,09	3,21
	64441020	4,13	0,40	0,91	0,69
	64442800	27,08	2,94	5,98	4,88
	64453000	25,73	4,41	8,43	6,66
	64460000	12,00	1,46	3,09	2,37
	64477600	36,86	3,99	9,98	7,99
	64482000	308,77	42,13	94,05	78,86
	64491000	236,40	40,32	67,79	60,45
	64491260	10,42	0,25	0,68	0,48
	64494950	11,85	2,82	4,08	3,48
	64497000	14,58	1,42	5,66	4,46
	64498550	376,45	54,14	126,48	98,54
	64500000	3,75	0,48	0,64	0,56
	64501950	9,43	1,05	2,75	2,19
	64502000	37,96	3,61	8,04	6,42
	64504100	3,99	0,08	0,16	0,11
	64504450	3,93	0,66	0,95	0,82
	64504550	5,80	1,08	2,04	1,65
	64504581	0,83	0,37	0,50	0,45
	64506000	490,45	67,62	147,53	116,55
	64507000	364,88	49,81	92,21	71,40
	64507100	10,35	0,28	0,93	0,60
	64508020	13,35	3,15	4,11	3,53
	64508500	15,80	2,03	4,88	3,91
	64529900	0,58	0,13	0,29	0,22
	64541000	150,41	1,66	9,02	5,59
64544000	0,99	0,09	0,29	0,21	
64546900	18,84	3,97	6,84	6,16	
Pirapó	64625000	75,10	5,54	11,30	8,95
	64652000	52,60	3,16	7,18	4,66
	64550000	71,55	21,67	34,67	29,21
	64659000	77,22	2,56	8,09	5,76

Ivaí	64645000	1188,53	19,08	52,18	34,15
	64655000	280,14	19,99	53,90	38,34
	64660500	505,18	34,70	87,31	65,80
	64670012	13,15	0,02	0,54	0,00
	64671000	20,59	2,63	6,99	5,74
	64671950	1,69	0,43	0,73	0,64
	64673000	34,40	7,17	14,51	11,99
	64675002	1291,29	55,84	128,10	100,93
	64682000	16,95	6,47	9,54	8,78
	64685000	542,99	73,91	154,20	126,58
	64693000	674,01	164,30	246,57	220,27
Piquiri	64773000	29,59	0,45	2,13	1,36
	64776100	219,35	17,70	36,31	28,25
	64785000	29,62	8,34	12,82	9,48
	64790000	17,79	3,33	5,76	4,82
	64795000	298,95	26,34	62,17	46,72
	64799500	350,62	27,34	71,65	53,33
	64810000	45,39	15,59	21,27	19,48
	64815000	61,44	19,71	30,68	26,68
	64820000	412,49	71,25	128,30	107,95
	64830000	506,95	90,58	170,79	140,47
64833000	19,01	8,99	11,44	10,98	
Paraná 3	64863360	2,33	0,83	1,03	1,01
	64863520	4,06	0,80	1,07	0,99
	64892500	45,25	0,82	2,57	1,81
Ribeira	81019350	11,08	2,83	4,57	3,87
	81080000	28,47	7,87	11,34	9,95
	81102000	30,47	9,47	13,56	11,87
	81120000	7,06	2,85	3,78	3,25
	81125000	7,58	2,81	3,74	3,42
	81290500	0,41	0,02	0,08	0,08
	81291000	8,02	0,89	1,15	1,02
	81299000	13,68	3,65	5,63	4,80
	81303000	4,06	0,73	1,21	1,02
Litorânea	82002000	6,06	1,61	2,22	1,94
	82121003	42,74	10,99	20,14	16,53
	82195002	5,18	0,54	1,10	1,00
	82198000	15,47	0,06	0,84	0,62
	81107000	63,72	20,01	28,66	25,12
	64362000	62,07	8,74	16,48	13,31
	64370000	63,07	8,03	14,30	11,47
	64447000	124,14	10,46	30,37	23,17
	64465000	159,50	21,28	43,46	34,50
	64504591	2,78	0,68	1,17	0,97
	64390000	38,81	3,47	6,92	5,59
	64767000	109,55	7,28	16,49	12,61

## Continuação Apêndice C

	64771500	120,41	6,77	19,01	12,51
	65003950	2,77	0,34	0,66	0,55
	65004995	6,84	0,81	1,10	0,99
	65006075	7,38	0,63	1,82	0,84
	65007045	2,25	0,48	0,81	0,69
	65007046	24,82	0,00	1,05	0,61
	65009000	14,35	1,06	2,14	1,49
	65010000	2,99	0,66	1,10	0,93
	65013005	17,33	1,66	3,62	2,61
	65015400	6,04	0,94	1,79	1,51
Iguaçu	65017006	68,53	4,59	7,36	6,23
	65017035	11,98	0,02	0,33	0,22
	65019400	2,60	0,24	0,37	0,30
	65021770	0,42	0,08	0,11	0,09
	65024000	2,73	0,24	0,64	0,45
	65135000	11,55	2,57	4,17	3,48
	65220000	383,59	55,18	107,31	85,09
	65365000	4,65	0,40	1,04	0,76
	65809000	9,38	0,86	2,03	1,61
	65815050	69,64	7,07	16,98	12,48
	65874990	0,83	0,02	0,10	0,09
	65925000	55,46	4,53	12,95	9,73
	65925770	0,59	0,06	0,11	0,10
	65925780	1,07	0,09	0,11	0,11
	65925800	9,44	0,23	0,80	0,55
	65970001	30,27	0,03	0,93	0,54
	65971010	15,84	1,01	2,43	1,94
	65974500	2,72	0,20	0,70	0,56
	65976001	1,22	0,09	0,22	0,14
	65981500	60,03	0,15	2,26	1,16

**APÊNDICE D                    MATRIZ DE ENTRADA COM AS VARIÁVEIS INDEPENDENTES.**

Bacia Hidrográfica	Código	Área de Drenagem (Km <sup>2</sup> )	Soma das drenagens (m)	Densidade da drenagem (1/Km)	Latitude	Longitude
Itararé	64230500	684,32	165744,59	0,24	-24,10	-49,47
	64240000	958,57	258916,96	0,27	-24,26	-49,73
	64242000	1672,38	454861,57	0,27	-23,98	-49,60
Cinzas	64360000	2002,74	527597,01	0,26	-23,77	-49,95
	64365850	121,17	28049,79	0,23	-23,50	-49,89
	64380000	1056,30	299944,59	0,28	-23,85	-50,39
Tibagi	64430200	395,63	95687,80	0,24	-25,32	-49,98
	64441020	218,03	52404,60	0,24	-25,46	-50,53
	64442800	1315,37	308161,87	0,23	-25,20	-50,52
	64453000	1038,14	260249,28	0,25	-24,83	-50,28
	64460000	740,11	189622,75	0,26	-24,71	-50,49
	64477600	1205,61	301517,15	0,25	-24,75	-50,09
	64482000	13705,10	3478406,15	0,25	-24,36	-50,59
	64491000	15994,30	4109190,78	0,26	-24,03	-50,69
	64491260	362,34	80159,74	0,22	-24,01	-50,70
	64494950	493,85	139388,65	0,28	-23,75	-50,98
	64497000	532,37	146686,06	0,28	-23,75	-50,94
	64498550	18291,30	4714590,54	0,26	-23,70	-50,92
	64500000	59,13	15080,40	0,26	-23,72	-50,78
	64501950	565,12	174337,51	0,31	-23,55	-50,88
	64502000	748,96	241237,57	0,32	-23,56	-51,03
	64504100	29,64	4408,68	0,15	-23,44	-50,82
	64504450	112,70	26995,51	0,24	-23,47	-51,31
	64504550	326,87	95183,87	0,29	-23,40	-51,01
	64504581	47,58	11591,81	0,24	-23,34	-51,30
	64506000	21723,00	5751219,81	0,26	-23,31	-51,00
	64507000	21777,30	5764306,66	0,26	-23,25	-50,98
	64507100	180,98	53118,47	0,29	-23,25	-51,07
	64508020	941,24	264236,86	0,28	-23,23	-50,73
	64508500	1047,39	291826,00	0,28	-23,17	-50,79
	64529900	20,94	2996,43	0,14	-23,50	-51,47
	64541000	1229,26	312416,87	0,25	-23,33	-51,85
	64544000	24,23	3033,61	0,13	-23,19	-52,15
64546900	1008,65	239377,82	0,24	-23,03	-51,75	
	64625000	3554,57	939671,74	0,26	-24,83	-51,14
	64652000	2603,09	718681,16	0,28	-24,11	-51,48

Continuação Apêndice D				0,25	-22,86	-52,08
Ivaí	64659000	3275,32	1019798,43	0,31	-24,02	-51,95
	64645000	8521,49	2369531,51	0,28	-24,34	-51,42
	64655000	12789,05	3573296,20	0,28	-24,04	-51,62
	64660500	19391,32	5555648,27	0,29	-23,90	-51,95
	64670012	571,08	159744,35	0,28	-24,09	-52,32
	64671000	849,30	249997,27	0,29	-24,03	-52,27
	64671950	77,37	15404,58	0,20	-24,07	-52,42
	64673000	1529,68	435682,82	0,28	-23,82	-52,18
	64675002	22446,60	6430884,58	0,29	-23,67	-52,12
	64682000	807,98	171854,30	0,21	-23,44	-52,60
	64685000	28345,64	7914851,59	0,28	-23,32	-52,67
	64693000	34319,74	9173727,19	0,27	-23,20	-53,32
Piquiri	64773000	759,46	239634,94	0,32	-24,78	-52,23
	64776100	7635,83	2502672,68	0,33	-24,75	-52,88
	64785000	1310,59	386235,96	0,29	-24,62	-52,93
	64790000	603,27	186855,69	0,31	-24,63	-53,10
	64795000	11219,45	3580047,80	0,32	-24,52	-53,17
	64799500	12060,63	3815414,27	0,32	-24,40	-53,16
	64810000	2038,41	450449,28	0,22	-23,92	-53,13
	64815000	2962,13	656477,29	0,22	-24,13	-53,32
	64820000	17400,82	5114924,01	0,29	-24,20	-53,33
	64830000	20923,81	6080735,72	0,29	-24,19	-53,75
64833000	1058,82	208021,54	0,20	-23,97	-53,75	
Paraná 3	64863360	60,18	13109,23	0,22	-24,74	-53,69
	64863520	92,29	24417,97	0,26	-24,75	-53,77
	64892500	554,32	149558,30	0,27	-24,96	-54,18
Ribeira	81019350	537,80	124412,26	0,23	-25,23	-49,60
	81080000	1320,84	333328,29	0,25	-25,07	-49,59
	81102000	1647,57	417383,81	0,25	-24,93	-49,48
	81120000	417,43	101447,13	0,24	-25,01	-49,34
	81125000	403,75	109878,04	0,27	-24,75	-49,33
	81290500	15,61	2104,65	0,13	-25,27	-49,17
	81291000	125,05	28543,28	0,23	-25,26	-49,15
	81299000	454,24	102390,82	0,23	-25,22	-48,95
	81303000	1067,50	263798,12	0,25	-25,09	-48,83
Litorânea	82002000	91,88	20844,88	0,23	-25,16	-48,24
	82121003	378,41	96922,48	0,26	-25,32	-48,71
	82195002	77,16	16913,00	0,22	-25,51	-48,88
	82198000	58,24	11544,62	0,20	-25,54	-48,85
	81107000	3188,70	839377,26	0,26	-24,90	-49,44
	64362000	3957,92	1075673,26	0,27	-23,34	-50,16
	64370000	5597,01	1518641,78	0,27	-23,09	-50,29

	64447000	5647,79	1382213,76	0,24	-24,95	-50,39
	64465000	8759,91	2198487,42	0,25	-24,51	-50,41
	64504591	133,54	40316,11	0,30	-23,35	-51,20
	64390000	3422,11	944727,37	0,28	-23,12	-50,45
	64767000	3526,45	1158169,06	0,33	-24,95	-52,60
	64771500	4151,16	1361723,79	0,33	-24,87	-52,76
	65003950	164,10	33777,18	0,21	-25,44	-49,12
	65004995	101,26	24865,53	0,25	-25,45	-49,12
	65006075	378,46	88410,71	0,23	-25,45	-49,17
	65007045	110,03	27921,32	0,25	-25,44	-49,20
	65007046	111,22	28544,47	0,26	-25,45	-49,20
Iguaçu	65009000	560,43	142164,95	0,25	-25,48	-49,19
	65010000	105,06	22924,02	0,22	-25,52	-49,15
	65013005	810,19	196791,01	0,24	-25,53	-49,22
	65015400	271,60	63329,46	0,23	-25,59	-49,23
	65017006	1152,60	280737,69	0,24	-25,60	-49,26
	65017035	67,47	13483,70	0,20	-25,65	-49,26
	65019400	45,66	8278,96	0,18	-25,72	-49,32
	65021770	24,81	7498,70	0,30	-25,42	-49,38
	65024000	180,09	52693,30	0,29	-25,57	-49,43
	65135000	3618,48	906548,56	0,25	-25,93	-49,39
	65220000	18433,31	3213852,27	0,17	-26,02	-50,59
	65365000	65,59	13336,05	0,20	-26,18	-51,52
	65809000	312,01	83670,70	0,27	-25,40	-51,44
	65815050	2215,04	574672,22	0,26	-25,54	-51,82
	65874990	37,83	7983,80	0,21	-25,86	-52,55
	65925000	120,00	29302,13	0,24	-26,28	-52,30
	65925770	29,90	8417,85	0,28	-26,31	-52,62
	65925780	12,25	100,99	0,01	-26,30	-52,66
	65925800	150,13	40772,89	0,27	-26,28	-52,59
	65970001	1080,22	330662,91	0,31	-25,39	-53,10
	65971010	385,35	93952,49	0,24	-25,38	-53,28
	65974500	114,39	33048,12	0,29	-25,77	-53,52
	65976001	50,46	7670,22	0,15	-25,03	-53,46
	65981500	1720,98	528081,70	0,31	-25,77	-53,61

**APÊNDICE E TABELA CONTENDO OS GRAUS DE PERTINÊNCIA PARA CADA ESTAÇÕES FLUVIOMÉTRICAS ESTUDADA NOS 4 CLUSTERS.**

ID	Código	Cluster 1	Cluster 2	Cluster 3	Cluster 4
E1	64230500	0,00000	0,00000	0,99999	0,00001
E2	64240000	0,00006	0,00001	0,99759	0,00234
E3	64242000	0,00106	0,00018	0,92444	0,07432
E4	64360000	0,00187	0,00031	0,83362	0,16420
E5	64362000	0,00016	0,00002	0,00333	0,99649
E6	64365850	0,00005	0,00001	0,99854	0,00139
E7	64370000	0,00366	0,00036	0,01483	0,98115
E8	64380000	0,00014	0,00003	0,99344	0,00640
E9	64390000	0,00103	0,00014	0,03775	0,96108
E10	64430200	0,00000	0,00000	0,99985	0,00014
E11	64441020	0,00003	0,00001	0,99920	0,00076
E12	64442800	0,00016	0,00003	0,99220	0,00761
E13	64447000	0,00063	0,00007	0,00405	0,99526
E14	64453000	0,00006	0,00001	0,99750	0,00243
E15	64460000	0,00000	0,00000	0,99987	0,00013
E16	64465000	0,19566	0,01033	0,08934	0,70467
E17	64477600	0,00014	0,00003	0,99321	0,00662
E18	64482000	0,99560	0,00105	0,00084	0,00251
E19	64491000	0,99288	0,00380	0,00097	0,00235
E20	64491260	0,00001	0,00000	0,99969	0,00030
E21	64494950	0,00000	0,00000	1,00000	0,00000
E22	64497000	0,00000	0,00000	1,00000	0,00000
E23	64498550	0,84263	0,12727	0,00965	0,02045
E24	64500000	0,00007	0,00001	0,99810	0,00182
E25	64501950	0,00000	0,00000	0,99996	0,00004
E26	64502000	0,00003	0,00001	0,99859	0,00137
E27	64504100	0,00008	0,00002	0,99768	0,00222
E28	64504450	0,00005	0,00001	0,99851	0,00143
E29	64504550	0,00000	0,00000	0,99985	0,00015
E30	64504581	0,00007	0,00001	0,99797	0,00195
E31	64504591	0,00004	0,00001	0,99891	0,00105
E32	64506000	0,07153	0,91799	0,00372	0,00676
E33	64507000	0,06737	0,92264	0,00355	0,00644
E34	64507100	0,00003	0,00001	0,99922	0,00075
E35	64508020	0,00006	0,00001	0,99722	0,00271
E36	64508500	0,00012	0,00002	0,99451	0,00535
E37	64529900	0,00009	0,00002	0,99762	0,00228
E38	64541000	0,00017	0,00003	0,99150	0,00830
E39	64544000	0,00009	0,00002	0,99762	0,00227
E40	64546900	0,00003	0,00001	0,99868	0,00129
E41	64550000	0,00008	0,00001	0,00142	0,99849



E42      64625000      0,00108      0,00015      0,04051      0,95825  
 Continuação Apêndice E

ID	Código	Cluster 1	Cluster 2	Cluster 3	Cluster 4
E43	64645000	0,33933	0,01441	0,08847	0,55779
E44	64652000	0,00323	0,00050	0,37941	0,61687
E45	64655000	0,99872	0,00035	0,00024	0,00069
E46	64659000	0,00043	0,00006	0,01121	0,98830
E47	64660500	0,15893	0,82157	0,00682	0,01268
E48	64670012	0,00000	0,00000	1,00000	0,00000
E49	64671000	0,00004	0,00001	0,99815	0,00180
E50	64671950	0,00007	0,00001	0,99811	0,00181
E51	64673000	0,00089	0,00015	0,94040	0,05856
E52	64675002	0,00020	0,99975	0,00002	0,00003
E53	64682000	0,00000	0,00000	0,99997	0,00003
E54	64685000	0,03792	0,94609	0,00635	0,00964
E55	64693000	0,10377	0,83776	0,02409	0,03437
E56	64767000	0,00001	0,00000	0,00013	0,99986
E57	64771500	0,00043	0,00005	0,00301	0,99651
E58	64773000	0,00003	0,00001	0,99866	0,00130
E59	64776100	0,47485	0,01658	0,07863	0,42994
E60	64785000	0,00051	0,00009	0,96978	0,02961
E61	64790000	0,00000	0,00000	0,99989	0,00011
E62	64795000	0,99885	0,00032	0,00021	0,00061
E63	64799500	0,99996	0,00002	0,00001	0,00002
E64	64810000	0,00102	0,00018	0,92837	0,07044
E65	64815000	0,00315	0,00050	0,54489	0,45146
E66	64820000	0,52748	0,43407	0,01291	0,02555
E67	64830000	0,01131	0,98655	0,00077	0,00136
E68	64833000	0,00001	0,00000	0,99962	0,00037
E69	64863360	0,00007	0,00001	0,99802	0,00189
E70	64863520	0,00006	0,00001	0,99843	0,00151
E71	64892500	0,00000	0,00000	1,00000	0,00000
E72	65003950	0,00004	0,00001	0,99872	0,00123
E73	65004995	0,00006	0,00001	0,99844	0,00149
E74	65006075	0,00001	0,00000	0,99978	0,00021
E75	65007045	0,00005	0,00001	0,99854	0,00140
E76	65007046	0,00005	0,00001	0,99856	0,00138
E77	65009000	0,00000	0,00000	1,00000	0,00000
E78	65010000	0,00006	0,00001	0,99838	0,00155
E79	65013005	0,00001	0,00000	0,99979	0,00020
E80	65015400	0,00002	0,00000	0,99943	0,00055
E81	65017006	0,00009	0,00002	0,99576	0,00413
E82	65017035	0,00007	0,00001	0,99804	0,00188
E83	65019400	0,00008	0,00002	0,99784	0,00207
E84	65021770	0,00008	0,00002	0,99780	0,00210

E85      65024000      0,00003      0,00001      0,99921      0,00076  
 Continuação Apêndice E

ID	Código	Cluster 1	Cluster 2	Cluster 3	Cluster 4
E86	65135000	0,00144	0,00020	0,06275	0,93561
E87	65220000	0,96461	0,00537	0,00692	0,02310
E88	65365000	0,00007	0,00001	0,99803	0,00188
E89	65809000	0,00001	0,00000	0,99973	0,00026
E90	65815050	0,00243	0,00040	0,74549	0,25168
E91	65874990	0,00008	0,00002	0,99782	0,00208
E92	65925000	0,00005	0,00001	0,99858	0,00136
E93	65925770	0,00008	0,00002	0,99784	0,00206
E94	65925780	0,00009	0,00002	0,99750	0,00239
E95	65925800	0,00004	0,00001	0,99892	0,00104
E96	65970001	0,00024	0,00004	0,98792	0,01180
E97	65971010	0,00001	0,00000	0,99984	0,00016
E98	65974500	0,00005	0,00001	0,99870	0,00125
E99	65976001	0,00008	0,00002	0,99781	0,00209
E100	65981500	0,00187	0,00031	0,83283	0,16498
E101	81019350	0,00000	0,00000	0,99999	0,00001
E102	81080000	0,00025	0,00004	0,98732	0,01239
E103	81102000	0,00073	0,00013	0,95308	0,04606
E104	81107000	0,00223	0,00032	0,13519	0,86226
E105	81120000	0,00000	0,00000	0,99989	0,00010
E106	81125000	0,00000	0,00000	0,99994	0,00006
E107	81290500	0,00009	0,00002	0,99758	0,00231
E108	81291000	0,00005	0,00001	0,99856	0,00138
E109	81299000	0,00000	0,00000	0,99990	0,00010
E110	81303000	0,00006	0,00001	0,99725	0,00268
E111	82002000	0,00006	0,00001	0,99830	0,00162
E112	82121003	0,00000	0,00000	0,99986	0,00013
E113	82195002	0,00007	0,00001	0,99816	0,00176
E114	82198000	0,00007	0,00001	0,99796	0,00195