

ROSANA DE FÁTIMA COLAÇO GIBERTONI



**REGIONALIZAÇÃO DE PARÂMETROS HIDROLÓGICOS  
UTILIZANDO ANÁLISE DIFUSA**

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

Orientadora: Miriam Rita Moro Mine

Co-orientador: Eloy Kaviski

CURITIBA

2002

ROSANA DE FÁTIMA COLAÇO GIBERTONI

## **REGIONALIZAÇÃO DE PARÂMETROS HIDROLÓGICOS UTILIZANDO ANÁLISE DIFUSA**

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

Orientadora: Miriam Rita Moro Mine

Co-orientador: Eloy Kaviski

CURITIBA

2002

**REGIONALIZAÇÃO DE PARÂMETROS HIDROLÓGICOS  
UTILIZANDO ANÁLISE DIFUSA**

por

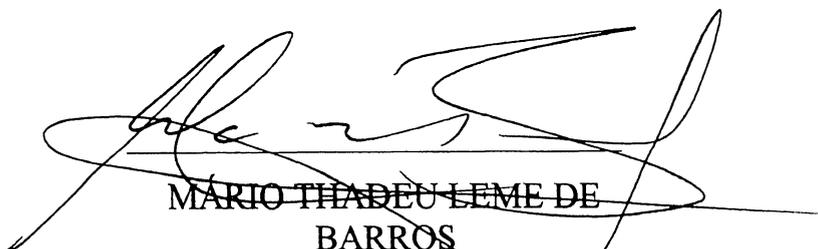
**ROSANA DE FÁTIMA COLAÇO GIBERTONI**

Dissertação aprovada como requisito parcial para a obtenção do título de Mestre no Programa de Pós-graduação em Engenharia de Recursos Hídricos e Ambiental do Setor de Tecnologia da Universidade Federal do Paraná, pela comissão formada pelos professores:

PRESIDENTE:

  
MIRIAM RITA MORO MINE  
UNIVERSIDADE FEDERAL DO PARANÁ  
UFPR

MEMBROS:

  
MÁRIO THADEU LEME DE  
BARROS  
UNIVERSIDADE DE SÃO PAULO - USP

  
HEINZ DIETER FILL  
UNIVERSIDADE FEDERAL DO PARANÁ  
UFPR

  
ELOY KAVISKI  
UNIVERSIDADE FEDERAL DO PARANÁ  
UFPR

Curitiba, 18 de outubro de 2002.

Ao meu esposo Gilbert,  
meu filho Eros,  
meus pais, Luiza e Nelson,  
e meus irmãos, Rosângela e Luciano.

## AGRADECIMENTOS

Durante a elaboração desta pesquisa de dissertação pude contar com a colaboração de algumas pessoas e entidades que muito me ajudaram, tornando possível a conclusão deste trabalho. Gostaria de deixar registrado neste espaço os meus sinceros agradecimentos:

À professora Miriam Rita Moro Mine, pela orientação desta pesquisa de dissertação.

Ao professor Eloy Kaviski, pela orientação e incentivo no desenvolvimento deste trabalho e pelos conhecimentos transmitidos no convívio profissional.

À Universidade Federal do Paraná, que através do Programa de Pós-Graduação em Engenharia de Recursos Hídricos e Ambiental, possibilitou a realização desta pesquisa.

Aos professores do Programa de Pós-Graduação em Engenharia de Recursos Hídricos e Ambiental, pelos conhecimentos e experiência compartilhados durante a realização do curso.

Ao Centro de Hidráulica e Hidrologia Professor Parigot de Souza, pela infra-estrutura dada ao curso, pelo apoio e pela compreensão da importância da conclusão deste trabalho.

Ao amigo Henrique Luiz Scremin, pelo auxílio na elaboração dos desenhos.

À amiga Silvia Maria Maximiano, pelas diversas colaborações no desenvolvimento do trabalho.

Às amigas Mara Bufrem Bossan Senna e Helena Maria Oliveira Vita pela correção das referências bibliográficas e auxílio nas pesquisas.

Aos meus amigos, pelo incentivo e carinho em todos os momentos.

Aos meus familiares, em especial ao meu esposo e meu filho, por compreenderem os momentos em que estive ausente e pelo amor que nos une.

## SUMÁRIO

<b>LISTA DE TABELAS .....</b>	<b>vii</b>
<b>LISTA DE FIGURAS.....</b>	<b>viii</b>
<b>RESUMO.....</b>	<b>ix</b>
<b>ABSTRACT.....</b>	<b>x</b>
<b>1. INTRODUÇÃO .....</b>	<b>1</b>
1.1. OBJETIVO GERAL .....	1
1.2. MOTIVAÇÃO .....	2
1.3. DESCRIÇÃO GERAL DAS TÉCNICAS UTILIZADAS.....	3
1.3.1. Regionalização hidrológica.....	3
1.3.2. Análise difusa .....	3
1.3.2.1. Geral .....	3
1.3.2.2. Classificação difusa.....	4
1.3.2.3. Regressão difusa.....	5
1.4. ESTRUTURA DO TRABALHO.....	5
<b>2. MÉTODOS DE REGIONALIZAÇÃO E ANÁLISE DIFUSA .....</b>	<b>6</b>
2.1. GERAL.....	6
2.2. REGIONALIZAÇÃO HIDROLÓGICA.....	6
2.2.1. Introdução .....	6
2.2.2. Regionalização de parâmetros por equações de superfície.....	9
2.2.3. Análise de agrupamentos .....	10
2.2.3.1. Geral .....	10
2.2.3.2. Técnicas para análise dos resultados .....	12
2.2.4. Análise discriminante .....	13

2.3. ANÁLISE DIFUSA .....	15
2.3.1. Introdução .....	15
2.3.2. Histórico e aplicações .....	17
2.3.3. Conceitos básicos e terminologia da análise difusa.....	19
2.3.4. Operações matemáticas envolvendo conjuntos difusos.....	27
2.3.5. Classificação difusa .....	30
2.3.5.1. Geral .....	30
2.3.5.2. Descrição do método .....	31
2.3.6. Regressão difusa .....	34
2.3.6.1. Geral .....	34
2.3.6.2. Regressão linear difusa.....	36
2.3.6.3. Regressão por metas difusas.....	43
<b>3. MÉTODO PROPOSTO.....</b>	<b>47</b>
3.1. INTRODUÇÃO .....	47
3.2. REGIONALIZAÇÃO DE PARÂMETROS DE SÉRIES DE VAZÕES MÉDIAS DIÁRIAS.....	48
3.2.1. Introdução .....	48
3.2.2. Descrição geral do método .....	49
3.2.3. Dados utilizados.....	50
3.2.4. Definição das regiões homogêneas.....	59
3.2.4.1. Desenvolvimento .....	59
3.2.4.2. Resultados.....	60
3.2.5. Determinação das equações de regressão que estimam $\bar{Q}$ , $CV[Q]$ , $MP_1[Q]$ , $MP_2[Q]$ e $MP_3[Q]$ .....	62
3.2.5.1. Desenvolvimento .....	62
3.2.5.2. Resultados.....	63
3.2.6. Análise discriminante .....	65

3.3. REGIONALIZAÇÃO DE PARÂMETROS DE SÉRIES DE VAZÕES MÁXIMAS ANUAIS .....	66
3.3.1. Introdução .....	66
3.3.2. Descrição geral do método .....	67
3.3.3. Dados utilizados.....	68
3.3.4. Definição das regiões homogêneas.....	74
3.3.4.1. Desenvolvimento.....	74
3.3.4.2. Resultados.....	74
3.3.5. Determinação das equações de regressão que estimam $\bar{Q}_{\max}$ e $CV[Q_{\max}]$ .....	75
3.3.5.1. Desenvolvimento.....	75
3.3.5.2. Resultados.....	76
<b>4. ANÁLISE DOS RESULTADOS.....</b>	<b>79</b>
4.1. GERAL.....	79
4.2. REGIONALIZAÇÃO DE PARÂMETROS DE SÉRIES DE VAZÕES MÉDIAS DIÁRIAS.....	80
4.3. REGIONALIZAÇÃO DE PARÂMETROS DE SÉRIES DE VAZÕES MÁXIMAS ANUAIS .....	84
<b>5. CONCLUSÕES E RECOMENDAÇÕES .....</b>	<b>89</b>
5.1. GERAL.....	89
5.2. PRINCIPAS CONCLUSÕES .....	89
5.3. RECOMENDAÇÕES .....	92
<b>REFERÊNCIAS BIBLIOGRÁFICAS .....</b>	<b>94</b>

## LISTA DE TABELAS

3.1 Estações fluviométricas e parâmetros básicos utilizados no estudo .....	53
3.2 Estações pluviométricas e parâmetros $\bar{P}$ e $CV[P]$ .....	56
3.3 Agrupamentos determinados na classificação difusa.....	61
3.4 Equações de regressão para estimativa do valor lognormalizado de $\bar{Q}$ .....	64
3.5 Equações de regressão para estimativa do valor lognormalizado de $MP_1[Q]$ .....	65
3.6 Equações de regressão para estimativa do valor lognormalizado de $MP_2[Q]$ .....	65
3.7 Equações de regressão para estimativa do valor lognormalizado de $MP_3[Q]$ .....	65
3.8 Estações fluviométricas e parâmetros regionalizados $s$ , $\bar{Q}_{max}$ , $CV[Q_{max}]$ , $\bar{P}_{max1}$ e $CV[P_{max1}]$ .....	70
3.9 Estações pluviométricas e parâmetros $\bar{P}_{max1}$ e $CV[P_{max1}]$ .....	71
3.10 Agrupamentos determinados na classificação difusa.....	75
3.11 Equações de regressão para estimativa de $\ln(\bar{Q}_{max})$ .....	77
3.12 Equações de regressão para estimativa de $\ln(CV[Q_{max}])$ .....	77
4.1 Estações fluviométricas analisadas na aplicação do estudo de caso .....	80
4.2 Parâmetros estimados pelos dados amostrais.....	83
4.3 Parâmetros estimados pelo método proposto .....	83
4.4 Parâmetros estimados pelo método utilizado para comparação.....	83
4.5 Estimativa de $\bar{Q}$ através da classificação difusa e interpolação ponderada.....	84
4.6 Estações fluviométricas analisadas na aplicação do estudo de caso .....	85
4.7 Parâmetros de entrada no método de regionalização proposto .....	85
4.8 Parâmetros estimados pelos dados amostrais.....	86
4.9 Parâmetros estimados pelo método proposto .....	87
4.10 Parâmetros estimados pelo método utilizado para comparação.....	87

## LISTA DE FIGURAS

2.1	Funções de pertinência em relação à idade de uma pessoa.....	16
2.2	Forma gráfica de algumas nomenclaturas para o conjunto difuso.....	20
2.3	Formas típicas de funções de pertinência.....	22
2.4	Exemplos de conjunto difuso convexo e não-convexo.....	23
2.5	Exemplos de conjunto difuso aberto e fechado.....	25
2.6	Função de pertinência de um número difuso L-R.....	26
2.7	Representação da função de pertinência de um número difuso L-R.....	27
2.8	Interpretação A da regressão difusa no espaço da função de pertinência de $y^*$ .....	37
2.9	Interpretação B da regressão difusa no espaço $x$ , $y$ e espaço da função de pertinência dos parâmetros.....	39
2.10	Representação das funções lineares de pertinência: (a) não-crescente; (b) não-decrescente.....	43
2.11	Funções de pertinência utilizadas no modelo de regressão de metas difusas: (a) não-crescente; (b) não-decrescente.....	45
3.1	Estações fluviométricas utilizadas no estudo de caso.....	55
3.2	Estações pluviométricas utilizadas na regionalização de $\bar{P}$ e $CV[P]$ .....	58
3.3	Estações pluviométricas utilizadas na regionalização de $\bar{P}_{\max 1}$ e $CV[P_{\max 1}]$ .....	73
4.1	Estações fluviométricas utilizadas na avaliação do método proposto para regionalização de parâmetros de séries de vazões médias diárias.....	82
4.2	Estações fluviométricas utilizadas na avaliação do método proposto para regionalização de parâmetros de séries de vazões máximas anuais.....	88

## RESUMO

O objetivo geral desta pesquisa de dissertação é analisar a aplicação da análise difusa em uma modelagem em recursos hídricos. A análise difusa foi escolhida como técnica para investigação por ser um método de multicritérios para impactos sob incerteza, sendo adequado à modelagem e ao gerenciamento de recursos hídricos. As incertezas estão ligadas principalmente às características das bacias hidrográficas e à imprecisão e escassez de dados hidrológicos. A área escolhida para investigação foi a regionalização de parâmetros hidrológicos em bacias hidrográficas de pequeno e médio porte.

Foram utilizadas as seguintes técnicas de lógica difusa: análise de agrupamento de dados pelo método “*c-means*” e regressão por metas difusas. Os parâmetros hidrológicos regionalizados são de séries de vazões médias diárias (média, coeficiente de variação e momentos com pesos probabilísticos); e de vazões máximas anuais (média e coeficiente de variação).

As conclusões do trabalho são baseadas na aplicação de índices de desempenho para os resultados e na comparação dos resultados desta dissertação com os resultados obtidos por outros métodos de regionalização. Através do estudo de caso foi possível atingir o objetivo desta dissertação, ou seja, verificar que a aplicação de ferramentas difusas como técnica de modelagem de recursos hídricos, no caso, regionalização de parâmetros hidrológicos, fornece resultados melhores que os obtidos com a utilização de métodos estatísticos padrões.

## ABSTRACT

This dissertation addresses the application of fuzzy sets theory in water resources modeling. The fuzzy analysis was chosen as a technique for investigation because of its adequacy for multi criteria modeling of impacts under uncertainty for water resources management. The uncertainties are mostly related to physiographic characteristics of the basins and also to imprecision and scarcity of hydrologic data. The investigation focused on regionalization of hydrologic parameters in small and medium basins.

The following techniques of fuzzy logic were used: analysis of clustering of numerical data by the "c-means" algorithm and fuzzy goal regression; regionalization of mean daily discharge (mean, coefficient of variation and probability weighted moments); and of maximum annual flow (mean and coefficient of variation).

The conclusions are based on performance indices of the results and comparison of the results of this dissertation with other methods of regionalization. Through the case study it was possible to verify that the application of fuzzy tools as a technique of water resources modeling, in this case, regionalization of hydrologic parameters, is justified by better results than those obtained with standard statistical methods.

# 1. INTRODUÇÃO

## 1.1. OBJETIVO GERAL

Na quase totalidade das nações a temática envolvendo recursos hídricos é de suma importância, pois lida com um dos elementos vitais à sobrevivência humana. A importância da água não está somente no fato de suprir algumas necessidades básicas do homem, mas, enquanto recurso natural a ser utilizado para diversos fins, atua diretamente no desenvolvimento econômico de uma sociedade. Apesar de haver problemas com relação à disponibilidade e/ou qualidade da água em praticamente todos os locais, em muitos, as ações planejadas não são tomadas a seu devido tempo, ou mesmo, não há sequer um planejamento ou uma atuação gerencial que auxilie a tomada de decisões. Atuando com a área gestora está a área técnica que fornece os subsídios para o correto manejo dentro das bacias hidrográficas. Neste âmbito, a área de recursos hídricos tem recebido nos últimos tempos a colaboração de muitos pesquisadores no sentido de ofertar novas técnicas na compreensão e solução dos problemas. Uma destas importantes contribuições é a aplicação da análise difusa no gerenciamento e modelagem em recursos hídricos.

O objetivo geral desta pesquisa de dissertação é analisar a aplicação da análise difusa em uma modelagem em recursos hídricos. A área escolhida para investigação foi a regionalização de parâmetros hidrológicos em bacias hidrográficas de pequeno a médio porte (50 a 5000 km<sup>2</sup>), e as técnicas utilizadas foram a classificação difusa e a regressão difusa. Os parâmetros hidrológicos regionalizados referem-se a séries de vazões médias diárias (média, coeficiente de variação e momentos de pesos probabilísticos); e de vazões máximas anuais (média e coeficiente de variação). As conclusões do trabalho são baseadas na aplicação de índices de desempenho para os resultados e comparação dos resultados desta dissertação com os

resultados obtidos pelo método de regionalização proposto em KAVISKI (1992).

## 1.2. MOTIVAÇÃO

A análise difusa foi escolhida como técnica para investigação por ser um método de multicritérios para impactos sob incerteza, adequado à modelagem e ao gerenciamento de recursos hídricos, dadas as incertezas sobre diversos aspectos característicos das bacias hidrográficas, além da imprecisão, escassez ou mesmo inexistência de dados hidrológicos.

No presente estudo a análise difusa foi aplicada na regionalização hidrológica, técnica que procura suprir uma das principais dificuldades que os hidrólogos brasileiros enfrentam, que é a insuficiência de dados em boa parte do território nacional. A maioria dos estudos envolvendo recursos hídricos necessita de observações hidrológicas de campo na região analisada. O principal motivo para a insuficiência e/ou inexistência de estações de coleta de dados hidrológicos é a falta de recursos financeiros dos órgãos responsáveis pela instalação e manutenção das redes de estações hidrométricas. Além disso, em nosso país falta uma definição clara das entidades responsáveis e da origem dos recursos necessários para instalação e, principalmente, manutenção das estações de coleta. Na verdade, a carência de dados hidrológicos é uma realidade em quase todas as regiões do mundo. Por este motivo, a regionalização hidrológica é um tema bastante difundido, porém ainda passível de receber contribuições no sentido de aprimorar as diversas técnicas utilizadas e obter estimativas mais eficientes dos parâmetros hidrológicos em questão.

Nesta pesquisa foram selecionadas para investigação bacias de pequena a média contribuição hídrica, pois além de serem as mais problemáticas em relação à indisponibilidade de dados, são muito visadas para a implantação de empreendimentos descentralizados, como PCH's (pequenas centrais hidrelétricas) e sistemas de

abastecimento de água e de irrigação.

### 1.3. DESCRIÇÃO GERAL DAS TÉCNICAS UTILIZADAS

#### 1.3.1. Regionalização hidrológica

A regionalização hidrológica é utilizada quando se necessita realizar estudos hidrológicos em um local onde não existem informações hidrológicas suficientes. Neste caso, pode-se fazer uso de técnicas de regionalização para transferir informações de locais com maior disponibilidade de dados para locais com poucos ou sem dados observados. O método de regionalização proposto neste estudo é constituído pelas seguintes etapas: i) definir regiões homogêneas através da análise de agrupamentos (classificação difusa) em termos dos parâmetros a serem regionalizados e em função dos parâmetros meteorológicos e fisiográficos dos locais estudados; ii) através dos resultados do melhor agrupamento, definir os parâmetros necessários para a aplicação de análise discriminante; iii) implementar a técnica de análise discriminante para obter as probabilidades de um local sem dados pertencer às regiões homogêneas definidas; iv) definir as equações de regressão para cada região homogênea através da análise de regressão difusa.

Os parâmetros regionalizados para um local sem dados são estimados através das equações de regressão e em função dos parâmetros descritivos da bacia.

#### 1.3.2. Análise difusa

##### 1.3.2.1. Geral

A análise difusa é classificada como uma das técnicas dos Sistemas

Inteligentes. Os Sistemas Inteligentes abrangem técnicas inspiradas no comportamento humano ou da natureza. A lógica difusa é baseada na linguagem natural da comunicação humana e os primeiros conceitos de conjunto difuso foram introduzidos, em 1965, por Lotfi A. Zadeh (ZADEH, 1965).

Um dos elementos principais associados à análise difusa é o conjunto difuso. A idéia do conjunto difuso pode ser tomada como uma generalização do conceito do conjunto clássico, no qual a função de pertinência pode assumir apenas os valores 1 ou 0 para cada indivíduo do conjunto universal, caracterizando o indivíduo como membro ou não membro do conjunto que está sendo analisado. No caso dos conjuntos difusos a função de pertinência é mais geral, podendo assumir valores dentro de uma determinada faixa, indicando um grau de pertinência variável dos elementos no conjunto em questão.

A análise difusa permite trabalhar tanto com valores numéricos como com termos lingüísticos, podendo ser representados, respectivamente, por um ou mais conjuntos difusos.

São diversas as ferramentas desenvolvidas relacionadas à lógica difusa: controladores difusos, classificação difusa, regressão difusa, entre outros. Na presente pesquisa foram utilizadas as técnicas da classificação e regressão difusa, que são mais convenientes para a utilização na regionalização.

#### 1.3.2.2. Classificação difusa

O objetivo da classificação é identificar grupos de elementos de uma amostra, tal que a estrutura dos seus dados possua maior similaridade dentro do mesmo agrupamento. Na classificação difusa cada local pode possuir um grau de aderência não nulo em mais de uma classe. Os valores numéricos do grau de aderência podem ser sumariados em termos de uma matriz particionada difusa. Dado que o número de valores da função de pertinência que são possíveis para descrever cada

ponto é infinito, é necessário definir um critério de classificação. No método proposto para classificação difusa “*c-means*” (HALL & MINNS, 1999), é definida uma função objetivo a ser minimizada, baseada na distância Euclidiana entre cada dado pontual e o centro da sua classe.

### 1.3.2.3. Regressão difusa

A regressão difusa é de extrema adequabilidade na aplicação em recursos hídricos, pois em hidrologia é muito comum determinar a relação entre variáveis com base em poucos e imprecisos dados. Nestes casos, a aplicação da análise estatística convencional não pode ser justificada, dada a escassez dos dados (BARDOSSY *et al.*, 1990).

A análise de regressão difusa descreve a variável dependente na forma de um número difuso, que é um tipo específico de conjunto difuso ligado somente a variáveis numéricas. Os parâmetros da regressão são também considerados como números difusos. As variáveis independentes podem ou não ser tomadas como números difusos. O caminho pelo qual a regressão difusa é conduzida é determinar os parâmetros da regressão que minimiza a imprecisão da variável dependente.

## 1.4. ESTRUTURA DO TRABALHO

Este trabalho apresenta a seguinte organização: no segundo capítulo descrevem-se os métodos de regionalização e análise difusa utilizados nesta dissertação; no terceiro capítulo são apresentados o desenvolvimento e aplicação do método proposto em um estudo de caso; o quarto capítulo trata da validação do método proposto com aplicação do estudo de caso e análise dos resultados; e finalmente no quinto capítulo são discutidas as conclusões e recomendações.

## 2. MÉTODOS DE REGIONALIZAÇÃO E ANÁLISE DIFUSA

### 2.1. GERAL

Neste capítulo são distintamente abrangidos os temas referentes à regionalização hidrológica e análise difusa. Assim, este capítulo é dividido em duas partes principais. A primeira parte trata da regionalização hidrológica, explanando sobre algumas técnicas de mapeamento regional a serem utilizadas na complementação do método proposto nesta dissertação. Na segunda parte desta revisão é apresentado o tema referente à análise difusa, onde são apresentados o histórico, terminologia, conceitos básicos, operações matemáticas com conjuntos difusos, bem como os conceitos de classificação difusa e regressão difusa.

### 2.2. REGIONALIZAÇÃO HIDROLÓGICA

#### 2.2.1. Introdução

Em hidrologia o termo regionalização tem sido utilizado para designar a transferência de informações de um local para outro dentro de uma área com similaridade no comportamento hidrológico (TUCCI, 2002). A regionalização hidrológica é utilizada quando deseja-se estudar um local onde não existem informações hidrológicas suficientes. Neste caso, pode-se fazer uso das ferramentas de regionalização para transferir as informações necessárias de locais com disponibilidade de dados para os locais sem ou com poucos dados observados.

A maioria dos estudos natureza envolvendo recursos hídricos necessita de informações hidrológicas na região analisada. Nem sempre dispõe-se de observações

hidrológicas em quantidade suficiente no local, quando esta falta deve ser suprida pela regionalização. Na verdade, por mais densa que seja uma rede de coleta de dados, nunca será possível cobrir todos os prováveis locais a serem estudados.

As técnicas de regionalização podem ser classificadas como subjetivas e objetivas (KAVISKI, 1992). Em geral, as técnicas subjetivas apresentam a desvantagem da ambigüidade, pois baseiam-se na subjetividade do usuário. Geralmente, na aplicação deste tipo de ferramenta são utilizados mapas e ábacos e estão sujeitos a algum tipo de interpretação pessoal. Os métodos de regionalização baseados em técnicas objetivas não dependem da inferência do usuário. Além disso, apresentam maior praticidade, pois, idealmente, estas técnicas são implementadas computacionalmente e dependem de poucos dados de entrada. Outra classificação apresentada em KAVISKI (1992) divide as técnicas de regionalização em quatro grupos: técnicas empíricas, extensão de séries, regressão e mapeamento regional.

FLEMING (1975) apresenta uma definição para os métodos empíricos, segundo a qual, estas técnicas dependem de equações matemáticas para exprimir o parâmetro desejado. Nestas equações, a relação entre as variáveis é unívoca, porém, os valores dos coeficientes são estimados empiricamente. Os métodos empíricos ainda apresentam sua importância, principalmente em bacias hidrográficas que drenam áreas bastante reduzidas, onde a escassez de estações fluviométricas condiciona o uso destas técnicas.

As técnicas de extensão de séries são utilizadas em locais com períodos de observação curtos para os quais se deseja melhorar a média e variância amostral, com base em locais similares com períodos mais extensos. Em geral, as variáveis envolvidas nesta técnica devem possuir séries com estrutura temporal semelhante.

Geralmente as técnicas de regressão utilizadas em hidrologia consistem em estimar parâmetros hidrológicos em função de parâmetros fisiográficos e climatológicos. O uso destes modelos em regionalização hidrológica é bastante difundido. A regressão é geralmente utilizada quando se deseja estimar parâmetros

estatísticos, tais como, a média, variância e quantis em locais sem dados situados em uma região para a qual são disponíveis as observações hidrológicas em outros pontos. A regressão também pode ser útil para melhorar a precisão de estimadores em locais com poucos dados. Em geral, as observações hidrológicas em locais distintos são assumidas como amostras de variáveis aleatórias independentes, mas a técnica da regressão GLS (mínimos quadrados generalizado) permite considerar locais correlacionados.

As técnicas de mapeamento regional consistem na delimitação de regiões que possuem comportamento hidrológico similar. As estimativas regionais dependem fortemente da técnica de mapeamento regional utilizada. Dado o extraordinário desenvolvimento dos computadores, o mapeamento regional por técnicas objetivas praticamente substituiu a elaboração de mapas manuais de regiões homogêneas.

As técnicas pertencentes aos quatro grupos definidos acima, em geral, são complementares em um estudo de regionalização.

Nesta dissertação é dada ênfase apenas às técnicas utilizadas pelo método proposto, que são: regressão, interpolação por equações de superfície, análise de agrupamentos e análise discriminante. Considerando a classificação apresentada em KAVISKI (1992) as técnicas utilizadas pelo método proposto consistem em técnicas de regressão e mapeamento regional (equações de superfície para interpolação, análise de agrupamentos e análise discriminante).

Os itens subsequentes descrevem a interpolação ponderada e por superfície “*spline*”, análise de agrupamentos e análise discriminante. A técnica de regressão utilizada nesta dissertação está descrita na segunda parte desta revisão bibliográfica que trata sobre análise difusa.

### 2.2.2. Regionalização de parâmetros por equações de superfície

Existem diversas técnicas de interpolação espacial, sendo que as que utilizam

equações de superfície primam pela praticidade e objetivismo. Neste item serão analisadas a interpolação por superfície “*spline*” e interpolação ponderada.

### *Superfície “spline”*

A função interpoladora “*spline*” pode ser representada por:

$$h^*(\underline{x}) = \sum_{i=1}^n \varphi_i K(\underline{x}, \underline{x}_i), \quad (2.1)$$

$$K(\underline{x}, \underline{x}_i) = f_i^2(\underline{x}) \ln[f_i^2(\underline{x})], \quad (2.2)$$

$$f_i(\underline{x}) = \sqrt{\frac{(x_i - x)^2}{s_x^2} + \frac{(y_i - y)^2}{s_y^2} + \frac{(z_i - z)^2}{s_z^2}}. \quad (2.3)$$

Na expressão (2.1)  $h^*(\underline{x})$  é a variável a ser interpolada,  $\underline{x}$  é um vetor com as componentes  $x$ ,  $y$  e  $z$  (latitude, longitude e altitude) e  $n$  é o número de locais com dados observados. Os coeficientes  $\varphi_i$  são determinados com a aplicação da condição de que a expressão (2.1) seja verdadeira para todos os  $n$  locais. Na expressão (2.3),  $x$ ,  $y$  e  $z$  representam, respectivamente, a latitude, longitude e altitude do local analisado;  $x_i$ ,  $y_i$  e  $z_i$  representam, respectivamente, as latitudes, longitudes e altitudes dos locais com dados observados; e  $s_x^2$ ,  $s_y^2$  e  $s_z^2$  representam a variância das latitudes, longitudes e altitudes, respectivamente, dos locais com dados observados.

### *Interpolação ponderada*

A interpolação ponderada considera a combinação linear em função dos valores observados dos  $n$  locais e sua função pode ser representada por:

$$h^*(\underline{x}) = \sum_{i=1}^n \lambda_i(\underline{x}) h(\underline{x}_i) \quad (2.4)$$

Os pesos  $\lambda_i(\underline{x})$  são definidos diretamente a priori, e assume-se que dependem exclusivamente da distância entre o vetor  $\underline{x}$  e os vetores dos locais com dados:

$$\lambda_i(\underline{x}) = \frac{1/f_i^2(\underline{x})}{\sum_{j=1}^n [1/f_j^2(\underline{x})]}. \quad (2.5)$$

Na expressão (2.5)  $f_i(\underline{x})$  é a distância Euclidiana normalizada pela variância, expressa em (2.3) e  $n$  é igual ao número de locais com dados observados.

### 2.2.3. Análise de agrupamentos

#### 2.2.3.1. Geral

Classificação ou agrupamentos podem ser caracterizados como uma técnica de análise estatística multivariada, através da qual efetua-se a divisão de uma região em sub-regiões que apresentem similaridade nos fenômenos considerados. Este particionamento é realizado através da obtenção da interdependência entre variáveis e entre indivíduos. As sub-regiões determinadas são chamadas de regiões homogêneas, para as quais são definidas relações únicas que expliquem os parâmetros a serem regionalizados em função das características descritivas do local (KAVISKI, 1992).

Na análise de agrupamentos o conjunto  $X = \{\underline{x}_i\}_{i=1}^N$  é decomposto em  $K$  sub-conjuntos. O conjunto  $X$  é, na verdade, formado por  $N$  vetores amostrais  $\underline{x}_i$ , sendo  $\underline{x}_i = (x_{i1}, x_{i2}, \dots, x_{in})$ ,  $N$  igual ao número de elementos a serem agrupados e  $n$  igual ao número de características analisadas. A análise de agrupamentos é realizada em função de todas as possíveis combinações dos parâmetros característicos dos elementos. Em função dos melhores resultados é definido o número de grupos, seus elementos e os parâmetros classificadores a serem utilizados na análise discriminante.

A diferença entre as diversas técnicas de análise de agrupamentos está na definição dos critérios para caracterizar a similaridade dentro dos grupos e a dissimilaridade entre os grupos. Conforme PARK *et al.*, citados por KAVISKI (1992,

p. 98), as técnicas podem ser divididas em duas categorias principais: técnicas hierárquicas e não-hierárquicas. As técnicas não-hierárquicas podem ser subdivididas nas seguintes classes: técnicas de particionamento ótimo, técnicas de pesquisa de densidade, técnicas de acumulação, técnicas mistas, entre outras.

Em KAVISKI (1992), a análise de agrupamentos foi realizada por meio do algoritmo “*K-means*”, uma técnica do particionamento ótimo do método de agrupamento não-hierárquico, cujo critério de otimização é a minimização do traço da matriz de dispersão total intra-grupos  $\text{tr}(\mathbf{W})$ . Outro critério comumente empregado é a maximização do traço da matriz de dispersão total inter-grupos  $\text{tr}(\mathbf{B})$ . O traço da matriz de dispersão  $\text{tr}(\mathbf{W})$  significa o erro gerado pela divisão de  $N$  objetos em  $K$  agrupamentos  $e(N,K)$ , que é obtido pela expressão abaixo:

$$\text{tr}(\mathbf{W}) = e(N, K) = \sum_{k=1}^K \sum_{l=1}^{L_k} d(z_l^{(k)}, \hat{\underline{\mu}}_k)^2. \quad (2.6)$$

Na expressão (2.6)  $L_k$  é o número de elementos classificados no grupo  $G^{(k)}$ ;  $\underline{z}_l^{(k)}$  é o vetor que contém as coordenadas, valores das características descritivas dos elementos que pertencem ao grupo  $k$ ;  $K$  é o número total de grupos;  $N$  é o número de objetos a serem agrupados;  $\hat{\underline{\mu}}_k$ , é o vetor que contém as coordenadas do centróide do grupo  $k$ ; e  $d(z_l^{(k)}, \hat{\underline{\mu}}_k)$  é denominada de distância de Mahalanobis, e é obtida por:

$$d(z_l^{(k)}, \hat{\underline{\mu}}_k) = \left[ (z_l^{(k)} - \hat{\underline{\mu}}_k)^T \hat{\underline{\Sigma}}^{-1} (z_l^{(k)} - \hat{\underline{\mu}}_k) \right]^{1/2}. \quad (2.7)$$

Na expressão (2.7)  $\hat{\underline{\Sigma}}$  é a matriz de transformação positiva-definida composta das variâncias e covariâncias dos componentes do vetor  $\underline{z}$  e  $^T$  representa a operação padrão de transposição de matrizes.

Outro método de análise de agrupamentos disponível é a classificação difusa “*c-means*” (HALL & MINNS, 1999). Na classificação difusa cada local possui um grau de aderência em mais de uma classe. Os valores do grau de aderência podem ser sumariados em termos de uma matriz particionada difusa. Dado que o número de

valores de aderência que são possíveis para descrever cada ponto é infinito, é necessário definir um critério de classificação. Na classificação difusa “*c-means*” é definida uma função objetivo a ser minimizada, baseada na distância Euclidiana entre cada dado pontual e o centro da sua classe.

A técnica de classificação “*c-means*” é apresentada posteriormente no item 2.3.5.

### 2.2.3.2. Técnicas para análise dos resultados

Os resultados da análise de agrupamentos podem ser avaliados em função de diversas relações. GALEATI *et al.*, citados por KAVISKI (1992, p. 144) propõem a relação abaixo para avaliar o valor de  $K$ :

$$F_1 = \left( \frac{e(N,K)}{e(N,K+1)} - 1 \right) (N - K + 1). \quad (2.8)$$

Os valores de  $e(N,K)$  e  $e(N,K+1)$  da expressão (2.8) podem ser obtidos por (2.6);  $K$  significa o número total de grupos; e  $N$  é o número de objetos a serem agrupados. Considera-se que os dados são distribuídos normalmente e que o resultado de (2.8) expressa a redução da variância ao se passar de  $K$  para  $K + 1$  agrupamentos. Empiricamente, é aceita a justificativa de adotar  $K + 1$  classes, caso  $F_1$  resulte maior que 10.

Conforme MARDIA *et al.*, citados por KAVISKI (1992, p. 144), outras relações podem auxiliar na definição do número de grupos, e são expressas por:

$$L_k \geq n + 1, \quad (2.9)$$

$$N \geq K(n + 1), \quad (2.10)$$

$$K \cong (N/2)^{1/2}. \quad (2.11)$$

Nestas expressões  $N$  é o número de elementos a serem agrupados,  $n$  é o número de parâmetros classificadores,  $K$  é o número de grupos e  $L_k$  é o número de elementos reunidos no grupo  $k$ .

Ainda, ACREMAN e SINCLAIR, citados por KAVISKI (1992, p. 144), apresentam um critério de avaliação dos agrupamentos, baseado no exame da consequência hidrológica da combinação de  $K$  grupos e  $n$  parâmetros classificadores. A técnica consiste em determinar a razão entre a variação entre agrupamentos,  $V_B$ , e a variação dentro dos agrupamentos,  $V_W$ , de cada um dos parâmetros a serem regionalizados. Considerando um parâmetro qualquer,  $\theta$ , as expressões para  $V_B$  e  $V_W$  são as seguintes:

$$V_B = \frac{1}{K-1} \sum_{k=1}^K (\bar{\theta}_{k\cdot} - \bar{\theta}_{\cdot\cdot})^2, \quad (2.12)$$

$$V_W = \frac{1}{N-K} \sum_{k=1}^K \sum_{l=1}^{L_k} (\theta_{kl} - \bar{\theta}_{k\cdot})^2, \quad (2.13)$$

$$\text{sendo: } \bar{\theta}_{k\cdot} = \sum_{l=1}^{L_k} \theta_{kl} / L_k, \quad \bar{\theta}_{\cdot\cdot} = \sum_{k=1}^K \sum_{l=1}^{L_k} \theta_{kl} / N.$$

Nestas expressões  $N$  é o número de elementos a serem agrupados;  $K$  é o número de grupos;  $L_k$  é o número de elementos reunidos no grupo  $k$ ; e  $\theta$  é o parâmetro que está sendo avaliado.

A passagem de  $K$  para  $K + 1$  agrupamentos se justifica, caso a relação  $V_B/V_W$  seja crescente.

#### 2.2.4. Análise discriminante

O objetivo da análise discriminante é alocar uma dada bacia hidrográfica em um dos  $K$  grupos definidos na análise de agrupamentos. A alocação é realizada com base nas características fisiográficas e/ou climatológicas (KAVISKI, 1992). Na análise

de agrupamentos são definidas as  $n$  características fisiográficas e/ou climatológicas a serem utilizadas na análise discriminante.

Supõe-se que a cada grupo está associada a função densidade de probabilidade  $f_k(\underline{x})$  em  $R^n$  de que o vetor  $\underline{x}$ , representado por  $n$  características  $x_i$ ,  $\underline{x}=(x_1, \dots, x_n)$ , pertença ao grupo  $k$ .

Existem diversas regras discriminantes que podem ser utilizadas. Neste trabalho é utilizada a mesma regra discriminante utilizada em KAVISKI (1992), ou seja, a regra discriminante de máxima verossimilhança para  $\underline{x}$ . Isto equivale a alocar  $\underline{x}$  no grupo  $k$ , tal que, seja mínimo o quadrado da distância de Mahalanobis:

$$d_k(\underline{x}) = (\underline{x} - \hat{\underline{\mu}}_k)^T \hat{\underline{\Sigma}}^{-1} (\underline{x} - \hat{\underline{\mu}}_k). \quad (2.14)$$

Na expressão (2.14) é suposto que as amostras do grupo sejam distribuídas normalmente com parâmetros  $\hat{\underline{\mu}}_k$  e  $\hat{\underline{\Sigma}}$ ,  $k = 1, \dots, K$  e com  $\hat{\underline{\Sigma}} > 0$ ;  $\hat{\underline{\mu}}_k$  é o vetor do centróide do grupo  $k$ ,  $\hat{\underline{\Sigma}}$  é a matriz positiva-definida de variâncias e covariâncias das componentes do vetor  $\underline{x}$  e  $^T$  representa a operação padrão de transposição de matrizes.

Conforme TASKER, citado por KAVISKI (1992, p. 146), as probabilidades a priori são proporcionais ao número de estações de cada agrupamento,  $L_k$ , assim a função a ser minimizada é:

$$d_k(\underline{x}) = \frac{(\underline{x} - \hat{\underline{\mu}}_k)^T \hat{\underline{\Sigma}}^{-1} (\underline{x} - \hat{\underline{\mu}}_k)}{2} - \ln(L_k). \quad (2.15)$$

A probabilidade a posteriori de que  $\underline{x}$  pertença a um grupo qualquer  $j$  pode ser obtida por:

$$p_j = \frac{\exp[-d_j(\underline{x})]}{\sum_{k=1}^n \exp[-d_k(\underline{x})]}. \quad (2.16)$$

Na expressão (2.16)  $n$  é o número de grupos definidos na análise de agrupamentos.

## 2.3. ANÁLISE DIFUSA

### 2.3.1. Introdução

A lógica difusa (nebulosa ou “fuzzy”) é baseada na linguagem natural da comunicação humana, onde os conceitos não são precisos, mas aproximados. Por isso, a lógica difusa pode utilizar variáveis lingüísticas no lugar de variáveis numéricas. Variáveis lingüísticas admitem como valores termos primários, como "muito pequeno", "altamente viável do ponto de vista econômico", "pouco frio", etc. Entretanto, a lógica difusa também permite trabalhar com variáveis numéricas, cujo exemplo mais difundido são os números difusos.

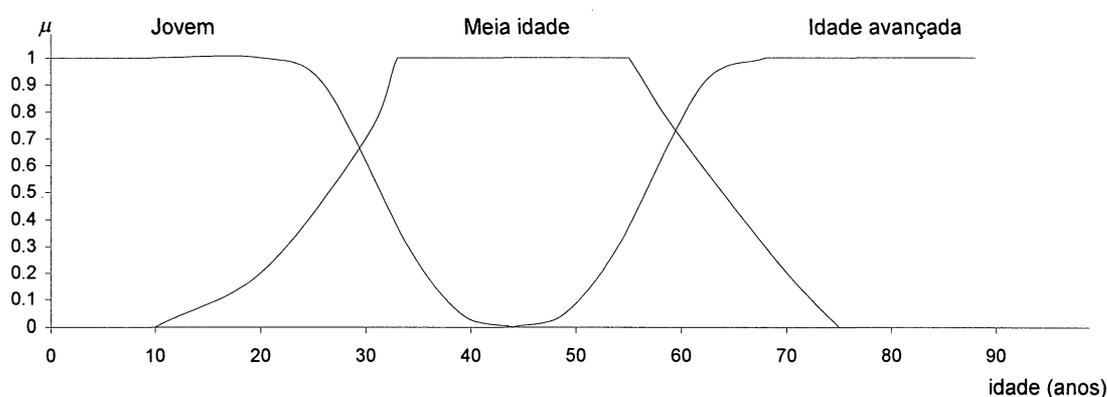
Exemplos de aplicação da análise difusa encontram-se em diversas áreas: economia, medicina, bancos de dados, produtos industriais, modelagem e gerenciamento de recursos hídricos, entre outros. O número cada vez maior de pesquisadores interessados no uso da lógica difusa deve-se principalmente às seguintes vantagens:

- Simplicidade - o método é fundamentado em conceitos matemáticos de fácil compreensão e na linguagem natural humana;
- Flexibilidade - permite a implementação de novas idéias em qualquer parte do sistema;
- Tolerância à imprecisão - a lógica difusa trabalha bem com a incerteza, fornecendo bons resultados;
- Facilidade de adaptação do conhecimento de especialistas;

A idéia básica do conjunto difuso é a função de pertinência que classifica um elemento como pertencente ou não ao conjunto. A função de pertinência de um conjunto difuso pode ser tomada como uma generalização dessa função no caso de conjuntos clássicos, nos quais pode assumir o valor 1 ou 0 para cada indivíduo do

conjunto universal, o que caracteriza o indivíduo como membro ou não membro do conjunto que está sendo analisado. Esta função pode ser vista como um caso especial de uma função mais geral para a qual os indivíduos do conjunto universal podem assumir valores dentro de uma determinada faixa, assim indicando um grau de aderência variável destes elementos no conjunto em questão. Neste caso, o conjunto é denominado conjunto difuso. A figura 2.1 traz, como ilustração, a representação difusa de três conjuntos difusos, cujas funções de pertinência expressam a percepção subjetiva ( $\mu$ ) de um indivíduo qualquer em relação à idade das pessoas; ou seja, as funções exprimem o grau que as diversas idades possuem em cada um dos conjuntos difusos.

FIGURA 2.1 – FUNÇÕES DE PERTINÊNCIA EM RELAÇÃO À IDADE DE UMA PESSOA In: AZEVEDO *et al.*, 2000, p. 52.



A especificação de funções de pertinência é subjetiva, pois um mesmo conceito (como por exemplo, *a idade ideal para se casar*) pode receber funções de pertinência muito diferentes se analisado por pessoas diferentes. Esta diferença, originada pela subjetividade, não pode ser confundida com aleatoriedade. A não aleatoriedade, inclusive, é a principal distinção entre a teoria dos conjuntos difusos e a teoria da probabilidade. Como exemplo, seja o conjunto difuso *jovem* relacionado à idade de uma pessoa e uma determinada pessoa de 40 anos. Supondo que o grau de

aderência obtido pela função de pertinência com a idade de 40 anos seja igual a 0,3, este grau de associação não é a probabilidade da pessoa ser uma pessoa jovem, pois sua idade já está definida.

### 2.3.2. Histórico e aplicações

O conceito de conjunto difuso foi introduzido, em 1965, por Lotfi A. Zadeh, da Universidade da Califórnia em Berkeley (ZADEH, 1965; AZEVEDO *et al.*, 2000). Até meados de 1960 os recursos computacionais disponíveis eram fundamentados na lógica booleana. Estes sistemas não eram capazes de resolver problemas de características ambíguas ocasionadas pela subjetividade ou incerteza na representação do sistema e imprecisão ou insuficiência de dados disponíveis.

Uma das primeiras aplicações da lógica difusa foi efetuada em 1974 pelo Prof. Mamdani, do Queen Mary College, Universidade de Londres. O Prof. Mamdani conseguiu êxito no controle de uma máquina a vapor utilizando os princípios da lógica difusa, após várias tentativas com outros sistemas de controle.

O sistema de controle difuso é uma das grandes aplicações da análise difusa. A principal característica do sistema de controle difuso é sua capacidade de reproduzir o comportamento do operador humano. Isto permite o grande êxito nas aplicações nos sistemas especialistas e de apoio à decisão (GALVÃO. *In*: GALVÃO e VALENÇA, 1999). A partir do início de 1980 os controladores difusos foram amplamente utilizados com aplicações em usinas nucleares, refinarias, processos biológicos e químicos, trocadores de calor, máquinas diesel, sistemas de tratamento de água, sistemas de operação automática de trens, entre outros. Como exemplo deste último caso, pode-se citar a estrada de ferro de Sendai, inaugurada em 1987, cujos projetistas utilizaram a lógica difusa no controle da aceleração, frenagem e parada do trem.

Outra aplicação interessante, executada em 1995 por Maytag, foi em uma máquina de lavar pratos "inteligente". A máquina combina um termistor, um sensor

condutivo para determinar o nível de detergente através da medição dos íons na água, um sensor de turvação para medir a sujeira na lavagem, e um sensor magnético para ler a taxa de giro. Através dos dados fornecidos por estes dispositivos, o sistema determina uma otimização no ciclo de lavagem, a fim de dispendir o mínimo de energia, detergente e água.

O Japão é o país que tem mais entusiasmo na aplicação da lógica difusa. Controladores difusos são comumente incorporados aos bens de consumo japoneses. Nos dias de hoje mais aplicações vêm surgindo. Dentre os produtos disponíveis no mercado que utilizam a lógica difusa pode-se citar: geladeiras (Sharp), ar condicionado (Mitsubishi), câmeras de vídeo (Cannon, Panasonic), máquinas de lavar roupa (Sanyo), na indústria automotiva: transmissões automáticas, injeção eletrônica, suspensão ativa, freios ABS. A National Aeronautics & Space Administration - NASA tem estudado controle difuso para ancorar suas naves automaticamente no espaço.

A área de recursos hídricos é uma das mais adequadas para implementação da lógica difusa, dada a incerteza nos processos envolvidos, escassez de dados e dada a existência de alternativas conflitantes na tomada de decisões no gerenciamento dos sistemas de recursos hídricos. A lógica difusa tem sido estudada e aplicada por diversos pesquisadores, principalmente na área de operação de sistemas de reservatórios (YIN *et al.*, 1999; FONTANE *et al.*, 1997; OWEN *et al.*, 1997; RUSSELL & CAMPBELL, 1996; SHRESTHA *et al.*, 1996). A análise difusa apresenta-se como um bom método objetivo para auxiliar na tomada de decisões em bacias hidrográficas, onde deve-se considerar a existência de conflitos sobre o uso dos recursos hídricos por diferentes grupos de usuários interessados.

Alguns pesquisadores aplicam a lógica difusa em conjunto com outras técnicas de controle, como redes neurais, criando sistemas conhecidos como algoritmos genéticos adaptativos, com o objetivo de construção de um sistema difuso capaz de aprender.

Tão promissora quanto as outras diversas técnicas que aplicam a análise

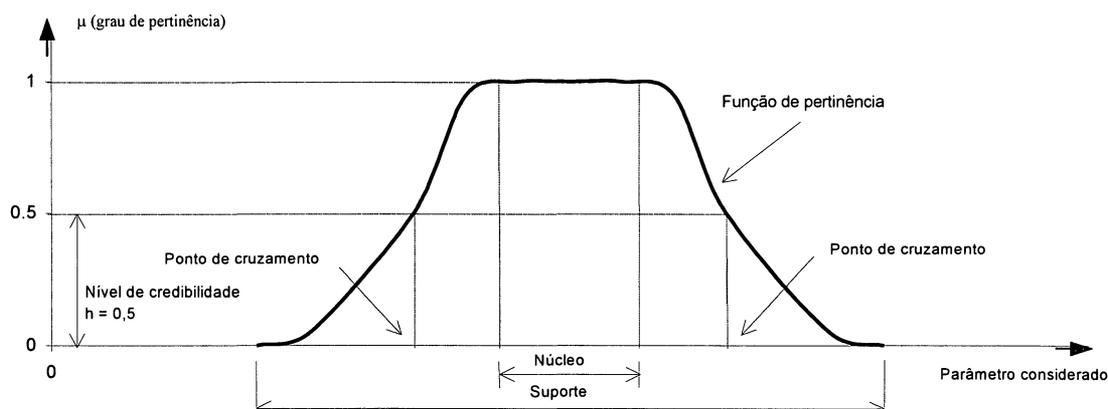
difusa em recursos hídricos é a regressão difusa. Neste particular, a regressão difusa é de extrema adequabilidade, pois em hidrologia é muito comum determinar a relação entre variáveis baseada em poucos e imprecisos dados (BARDOSSY *et al.*, 1990). Outra importante técnica desenvolvida é a classificação difusa. Esta ferramenta pode ser utilizada, por exemplo, na regionalização hidrológica, onde se aplica a técnica de agrupamentos na classificação de regiões hidrológicamente homogêneas (HALL & MINNS, 1999). Neste trabalho a regressão e classificação difusa serão utilizadas como técnicas da regionalização de parâmetros hidrológicos.

Muitas aplicações interessantes em recursos hídricos e ambientais podem ser encontradas em publicações técnico-científicas. YU *et al.* (2000) apresenta um modelo de previsão de chuva, cujos parâmetros foram estimados através da regressão por metas difusas. Em CAMPANA (1995) é apresentada a obtenção de um hidrograma unitário através de técnicas de regionalização baseadas na regressão linear difusa. GALVÃO (1995) apresenta o desenvolvimento de um método de previsão climática no nordeste para uso em operação de reservatórios, onde foram considerados conceitos representados por variáveis difusas. O trabalho desenvolvido em GATES *et al.* (1991) apresenta uma metodologia para tomada de decisões considerando um sistema de irrigação implantado em um contexto que possui diversos critérios de planejamento. Em BOGARDI *et al.* (1983) foi desenvolvida uma metodologia de apoio ao gerenciamento de um aquífero, onde objetivos de caráter ambiental são implementados em conjuntos difusos.

### 2.3.3. Conceitos básicos e terminologia da análise difusa

Para descrever os conceitos básicos utiliza-se neste trabalho a nomenclatura comumente usada na literatura. Na figura 2.2 pode ser visualizada parte da nomenclatura utilizada neste trabalho. Em seguida, são apresentados os principais conceitos relacionados à análise difusa.

FIGURA 2.2 – FORMA GRÁFICA DE ALGUMAS NOMENCLATURAS PARA O CONJUNTO DIFUSO



### *Conjuntos difusos*

Os conjuntos difusos são coleções de  $n$  variáveis do tipo  $i_1, i_2, \dots, i_n$ , que caracterizam, qualificam e alteram o objeto de discurso (BORELLI, 1999). Aos conjuntos difusos estão associadas termos lingüísticos e números difusos. Um número difuso é um tipo específico de conjunto difuso ligado somente a variáveis numéricas. A quantidade de termos lingüísticos ou números difusos define o número de funções de pertinência associadas ao conjunto, as quais deverão ser simultaneamente utilizadas no mapeamento de um valor de entrada. O exemplo da figura 2.1 tem como conjunto difuso a idade da pessoa e associados a este conjunto os termos lingüísticos jovem (pouca idade), meia idade e velho (muita idade). Na análise de um objeto de discurso vários conjuntos podem ser considerados e suas respectivas saídas combinadas para compor um valor final único. Utilizando a mesma referência anterior o objeto de discurso poderia ser, por exemplo, a “competitividade de uma pessoa no mercado de trabalho”, que poderia ter além do conjunto difuso *idade*, outros como, *sexo* e *grau de escolaridade*.

### *Função de pertinência*

Uma forma usual de representar a função de pertinência  $\mu_A$ , relacionada ao

conjunto difuso  $A$  (pertencente ao conjunto universal  $X$ , ou universo de discurso  $X$ ), pode ser escrita como:

$$\mu_A : X \rightarrow [0, 1].$$

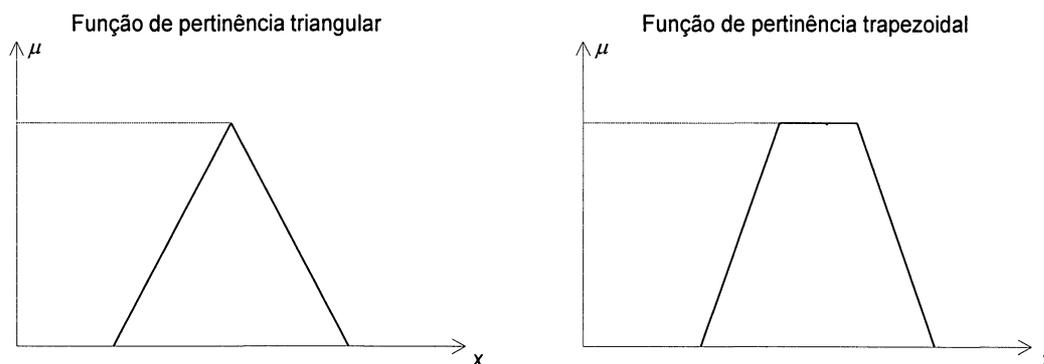
Na função acima,  $[0, 1]$  corresponde ao intervalo de números reais entre 0 e 1, inclusive.

O mais comum é utilizar o intervalo clássico de números reais entre 0 e 1, para exprimir graus de aderência, derivando para o conjunto difuso *normal* ou *normalizado*. Entretanto, qualquer intervalo arbitrário pode ser utilizado para esta finalidade.

O bom desempenho de um conjunto difuso para a representação de um termo linguístico ou número difuso, depende da apropriação de sua função de pertinência. Por isso, a adequada determinação desta função é de muita importância. Os métodos propostos para esta tarefa têm, geralmente, sido empíricos, e comumente envolvem questionários feitos para especialistas no assunto em questão, para avaliar quantitativa e qualitativamente as percepções subjetivas quanto aos graus de aderência de uma determinada classe conceitual ou valor numérico. Os resultados destas pesquisas podem ser implementados através de diversos métodos. As respostas podem ser processadas quantitativamente através da estatística do número de respostas, ou podem ser avaliadas qualitativamente, quando o tempo que o especialista leva para responder as questões é considerado na avaliação do grau de subjetividade.

As funções de pertinência dos conjuntos difusos podem assumir diversas formas sendo a trapezoidal e a triangular as mais utilizadas. Estas formas apresentam como principais vantagens a simplicidade na sua determinação e facilidade de implementação computacional. A figura 2.3 apresenta exemplos de função de pertinência triangular e trapezoidal.

FIGURA 2.3 – FORMAS TÍPICAS DE FUNÇÕES DE PERTINÊNCIA



É comum representar a função triangular através de três valores pertencentes ao eixo das abcissas (parâmetros). Estes valores estão relacionados aos vértices do triângulo, sendo que dois representam os limites mínimo e máximo do trecho que possui alguma pertinência, e um representa o vértice que possui pertinência máxima. De forma similar pode-se representar a função trapezoidal, porém, neste caso, dois dos quatro parâmetros condizentes à função representam a pertinência máxima.

#### *Conjuntos difusos convexos e não-convexos*

As funções de pertinência podem descrever conjuntos difusos convexos ou conjuntos difusos não-convexos. O conjunto difuso convexo tem nos dois lados da função em torno do seu ponto de pertinência máxima, valores sempre crescentes ou sempre decrescentes, o mesmo não ocorrendo para o conjunto difuso não-convexo. A figura 2.4 apresenta um exemplo de conjunto difuso convexo e um não-convexo.

#### *Cardinalidade*

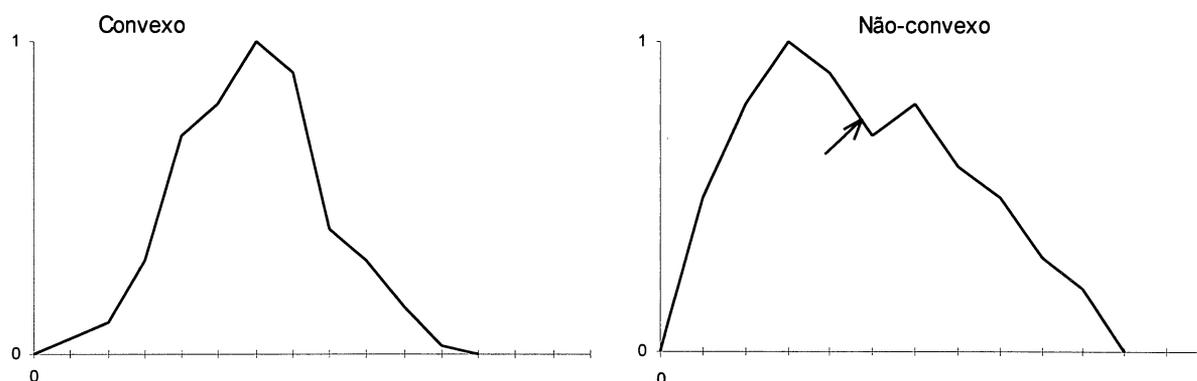
A cardinalidade de um conjunto difuso  $A$  para o universo de discurso  $X$  discreto é expressa como:

$$|A| = \sum_{x \in X} \mu_A(x). \quad (2.17)$$

A cardinalidade de um conjunto difuso  $A$  para o universo de discurso  $X$  contínuo é dada por:

$$|A| = \int_X \mu_A(x) dx. \quad (2.18)$$

FIGURA 2.4 – EXEMPLOS DE CONJUNTO DIFUSO CONVEXO E NÃO-CONVEXO



### *Suporte*

O suporte (“support”) de um conjunto difuso  $A$  é o intervalo de valores de  $x$  em  $X$ , tais que  $\mu_A(x) > 0$  (ver figura 2.2).

$$\text{suporte}(A) = \{x | \mu_A(x) > 0\}. \quad (2.19)$$

### *Nível de credibilidade*

O nível de credibilidade  $h$  é um valor pré-especificado correspondente ao menor grau de uma função de pertinência a ser aceito (ver figura 2.2). Um conceito ligado ao nível de credibilidade  $h$  é o “conjunto ao nível  $h$ ”, que corresponde ao intervalo de valores de  $x$  em  $X$ , tais que  $\mu_A(x) > h$ .

### *Núcleo*

O núcleo (“core”) de um conjunto difuso  $A$  é o conjunto de todos os pontos de  $x$  em  $X$ , tais que  $\mu_A(x) = 1$  (ver figura 2.2).

$$\text{núcleo}(A) = \{x \mid \mu_A(x) = 1\}. \quad (2.20)$$

### *Normalidade*

Um conjunto difuso  $A$  é dito normal se há pelo menos um ponto  $x \in X$  com pertinência máxima igual a 1.

### *Ponto de cruzamento*

O ponto  $x \in X$  no qual  $\mu_A(x) = 0,5$ , chama-se ponto de cruzamento (“crossover”). A representação do ponto de cruzamento pode ser vista na figura 2.2.

$$\text{cruzamento}(A) = \{x \mid \mu_A(x) = 0,5\}. \quad (2.21)$$

### *Simetria de um conjunto difuso*

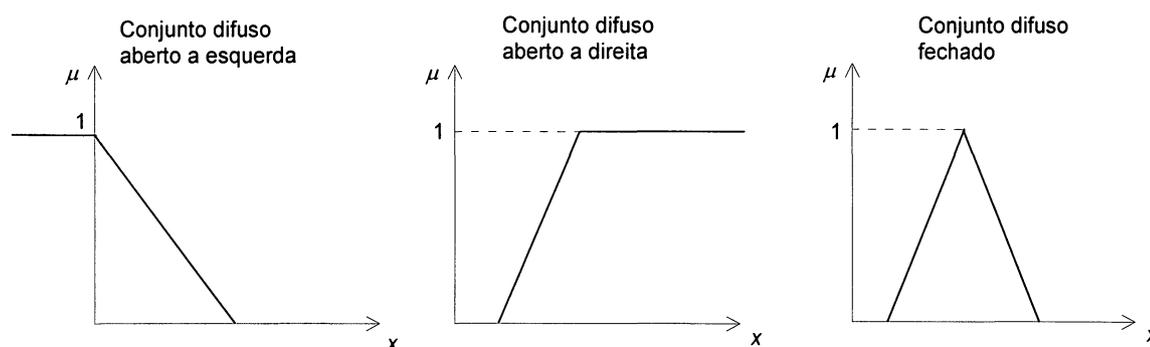
Se uma função de pertinência é simétrica em torno de um certo ponto  $x = c$ , então ela define um conjunto difuso simétrico. Esta função de pertinência é expressa por:

$$\mu_A(c + x) = \mu_A(c - x) \text{ para todo } x \in X. \quad (2.22)$$

### *Conjunto difuso aberto e fechado*

Um conjunto difuso é aberto a esquerda se  $\lim_{x \rightarrow -\infty} \mu_x = 1$  e  $\lim_{x \rightarrow +\infty} \mu_x = 0$ ; um conjunto é aberto a direita se  $\lim_{x \rightarrow -\infty} \mu_x = 0$  e  $\lim_{x \rightarrow +\infty} \mu_x = 1$ ; um conjunto difuso é fechado se  $\lim_{x \rightarrow -\infty} \mu_x = 0$  e  $\lim_{x \rightarrow +\infty} \mu_x = 0$ . A figura 2.5 apresenta exemplos de conjunto difuso aberto e fechado.

FIGURA 2.5 – EXEMPLOS DE CONJUNTO DIFUSO ABERTO E FECHADO



### Número difuso

Um conjunto difuso, contido no conjunto dos Números Reais ( $\mathbb{R}$ ), que possui as propriedades de normalidade e convexidade é denominado número difuso  $A^*$ . Neste trabalho, o asterisco associado a uma letra representa um número difuso. Os números difusos são amplamente utilizados nas aplicações práticas, constituindo o tipo mais básico de conjunto difuso, pois permitem a utilização de valores numéricos no lugar de termos lingüísticos, e são de fácil compreensão.

### Largura de um número difuso

Para um número difuso a largura é definida como a distância entre os valores limites do intervalo de  $x$ , cuja função de pertinência assume valores maiores ou iguais ao nível de credibilidade  $h$ .

$$\text{largura}(A^*) = |x_2 - x_1|, \quad (2.23)$$

$$\text{onde: } \mu_{A^*}(x_1) = \mu_{A^*}(x_2) = h.$$

### Número difuso médio

O número difuso médio  $FM(A^*)$  é uma das técnicas mais utilizadas para promover a “desfuzzificação” (“*defuzzification*”), processo em que o valor da saída difusa é convertido em um único valor numérico. A média de um número difuso  $A^*$  é

dada por:

$$FM(A^*) = \frac{\int_{-\infty}^{+\infty} x \mu_{A^*}(x) dx}{\int_{-\infty}^{+\infty} \mu_{A^*}(x) dx}. \quad (2.24)$$

### Número difuso L-R

Um conjunto difuso  $A$  é chamado número difuso L-R  $A^*$  se a sua função de pertinência é dada por (BARDOSSY *et al.*, 1990):

$$\mu_{A^*}(x) = L\left(\frac{S-x}{b}\right) \quad x \leq S \quad b > 0, \quad (2.25)$$

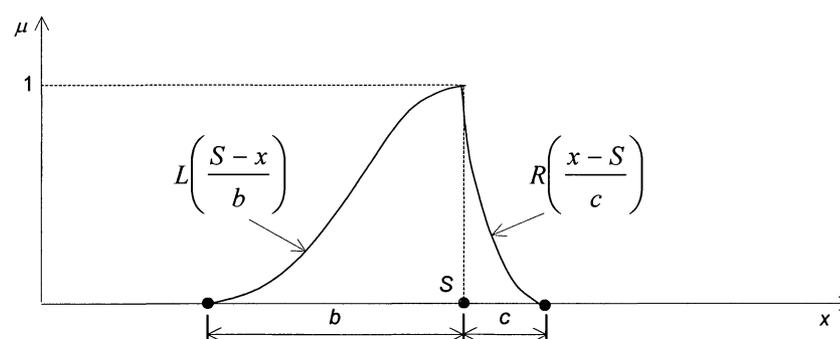
$$\mu_{A^*}(x) = R\left(\frac{x-S}{c}\right) \quad x > S \quad c > 0, \quad (2.26)$$

onde:

- $L$  – função de pertinência a esquerda do ponto onde  $\mu_{A^*}(x) = 1$ ;
- $R$  – função de pertinência a direita do ponto onde  $\mu_{A^*}(x) = 1$ ;
- $S$  – valor de  $x$  para o qual  $\mu_{A^*}(x) = 1$ ;
- $b$  – distância entre  $S$  e o maior valor de  $x$  para o qual  $\mu_{A^*}(x) = 0$ ;
- $c$  – distância entre  $S$  e o menor valor de  $x$  para o qual  $\mu_{A^*}(x) = 0$ .

Para a definição de um número difuso L-R são necessários, além das funções  $L$  e  $R$ , três parâmetros –  $S$ ,  $b$  e  $c$ . A figura 2.6 apresenta um exemplo de função de pertinência de um número difuso L-R.

FIGURA 2.6 – FUNÇÃO DE PERTINÊNCIA DE UM NÚMERO DIFUSO L-R. In: BARDOSSY *et al.*, 1990, p. 1498.

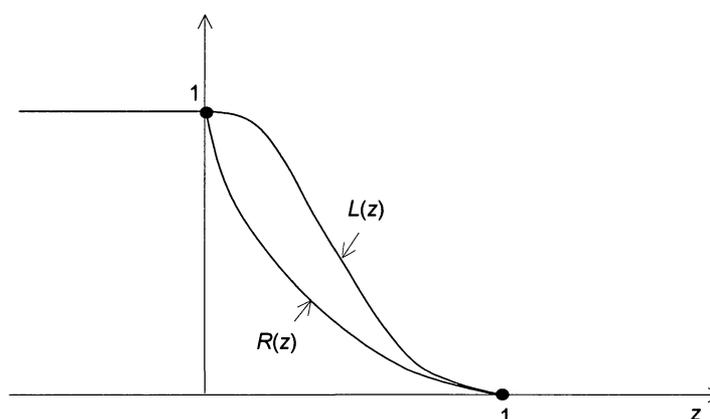


As funções  $L$  e  $R$  podem ser funções contínuas estritamente decrescentes ou não estritamente decrescentes. Neste trabalho, por simplicidade, não será abordada a definição mais geral das funções  $L$ - $R$ , que corresponde a funções não estritamente decrescentes. A figura 2.7 é a representação das funções  $L$  e  $R$ , continuamente decrescentes, definidas no intervalo  $[0, 1]$ , e como funções de  $z$ . Nesta figura podem ser verificadas as seguintes condições de contorno:

$$L(z) = R(z) = 1 \quad z \leq 0, \quad (2.27)$$

$$L(z) = R(z) = 0 \quad z > 1. \quad (2.28)$$

FIGURA 2.7 – REPRESENTAÇÃO DA FUNÇÃO DE PERTINÊNCIA DE UM NÚMERO DIFUSO  $L$ - $R$ . In: BARDOSSY *et al.*, 1990, p. 1498.



#### 2.3.4. Operações matemáticas envolvendo conjuntos difusos

Como os conjuntos clássicos, os conjuntos difusos também possuem operações matemáticas, inclusive, algumas similares entre as duas abordagens.

##### *União*

A união de dois conjuntos difusos  $A$  e  $B$  resulta em um conjunto difuso  $C$ , descrito por:

$$C = A \cup B \quad \text{ou} \quad C = A \text{ OU } B. \quad (2.29)$$

A função de pertinência de  $C$  resulta em:

$$\mu_C(x) = \max(\mu_A(x), \mu_B(x)) = \mu_A(x) \vee \mu_B(x). \quad (2.30)$$

### *Interseção*

A interseção de dois conjuntos difusos  $A$  e  $B$  resulta em um conjunto difuso  $C$ , descrito por:

$$C = A \cap B \quad \text{ou} \quad C = A \text{ E } B. \quad (2.31)$$

A função de pertinência de  $C$  resulta em:

$$\mu_C(x) = \min(\mu_A(x), \mu_B(x)) = \mu_A(x) \wedge \mu_B(x). \quad (2.32)$$

### *Complemento*

O complemento de um conjunto difuso  $A$  é descrito por:

$$\bar{A} \quad \text{ou} \quad \neg A \quad \text{ou} \quad \text{NÃO } A.$$

A função de pertinência de  $\bar{A}$  é designada por:

$$\mu_{\bar{A}}(x) = 1 - \mu_A(x). \quad (2.33)$$

### *Comutatividade*

Uma função  $F$  é dita comutativa se satisfaz a seguinte propriedade:

$$F(a,b) = F(b,a) \quad \text{para } \forall a, b \in [0,1]. \quad (2.34)$$

### *Associatividade*

Uma função  $F$  é dita associativa se satisfaz a seguinte propriedade:

$$F(a, F(b, c)) = F(F(a, b), c) \text{ para } \forall a, b, c \in [0, 1]. \quad (2.35)$$

### *Monotonicidade*

Uma função  $F$  é dita monotônica se satisfaz a seguinte propriedade:

$$F(a, b) \leq F(c, d) \text{ se } a \leq c \text{ e } b \leq d, \text{ para } \forall a, b, c \text{ e } d \in [0, 1]. \quad (2.36)$$

### *Normas triangulares (T-normas)*

Uma função  $T$  especifica, em geral, a interseção difusa de dois conjuntos difusos  $A$  e  $B$ :

$$\mu_{A \cap B}(x) = T(\mu_A(x), \mu_B(x)). \quad (2.37)$$

A função  $T$  é denominada T-norma se satisfaz as condições de comutatividade, associatividade e monotonicidade e se verifica a seguinte propriedade:

$$T(a, 1) = a \text{ para } \forall a \in [0, 1] \quad (1 \rightarrow \text{elemento neutro}). \quad (2.38)$$

### *Conormas triangulares (T-conormas)*

Uma função  $S$  especifica, em geral, a união difusa de dois conjuntos difusos  $A$  e  $B$ :

$$\mu_{A \cup B}(x) = S(\mu_A(x), \mu_B(x)). \quad (2.39)$$

A função  $S$  é denominada S-norma ou T-conorma se satisfaz as condições de

comutatividade, associatividade e monotonicidade e se verifica a seguinte propriedade:

$$S(a,0) = a \quad \text{para } \forall a \in [0,1] \quad (0 \rightarrow \text{elemento neutro}). \quad (2.40)$$

### *Dualidade*

A dualidade se verifica para uma T-norma T e uma S-norma S em relação a uma operação de negação  $\neg$ , se ambas satisfazem as Leis de De Morgan:

$$\neg(T(a,b)) = S(\neg a, \neg b) \quad \text{para } \forall a, b \in [0,1], \quad (2.41)$$

$$\neg(S(a,b)) = T(\neg a, \neg b) \quad \text{para } \forall a, b \in [0,1]. \quad (2.42)$$

### *Operação de negação*

A operação de negação pode ser obtida por:

$$\neg a = 1 - a. \quad (2.43)$$

### *Processo de “desfuzificação”*

O processo em que os valores da saída difusa final são convertidos em um valor ordinário, é conhecido como “desfuzificação”. Existem vários métodos de “desfuzificação”, sendo que, os mais utilizados são: método do máximo critério, método da média dos máximos e método do centro de massa (centróide).

## 2.3.5. Classificação difusa

### 2.3.5.1. Geral

A classificação (análise de agrupamentos de dados) pode ser caracterizada

como uma das técnicas de análise estatística multivariada, através da qual efetua-se a divisão de uma região em sub-regiões que apresentem similaridade nos fenômenos considerados. Este particionamento é realizado através da obtenção da interdependência entre variáveis e entre indivíduos. As sub-regiões determinadas são chamadas de regiões homogêneas, para as quais são definidas relações únicas que expliquem os parâmetros a serem regionalizados em função das características descritivas do local (KAVISKI, 1992).

Como já explanado no item 2.2.3, na análise de agrupamentos o conjunto  $X = \{\underline{x}_i\}_{i=1}^N$  é decomposto em  $K$  sub-conjuntos. O conjunto  $X$  é, na verdade, formado por  $N$  vetores amostrais  $\underline{x}_i$ , sendo  $\underline{x}_i = (x_{i1}, x_{i2}, \dots, x_{in})$  e  $n$  é igual ao número de parâmetros característicos dos elementos (parâmetros classificadores). A análise de agrupamentos é realizada em função de todas as possíveis combinações dos parâmetros característicos dos elementos. Em função dos melhores resultados é definido o número de grupos, seus elementos e os parâmetros classificadores a serem utilizados na análise discriminante.

#### 2.3.5.2. Descrição do método

O método proposto para classificação difusa, “*c-means*” (HALL & MINNS, 1999), permite trabalhar com variáveis ordinárias ou difusas. No caso de considerar que as divisões dos dados são abruptas (determinísticas), cada ponto é associado com somente uma classe. No caso da divisão difusa, cada ponto pode possuir um grau de aderência em mais de uma classe. Neste último caso é definida uma família de  $K$  conjuntos difusos  $A_k$ , onde  $k = 1, 2, \dots, K$ , no universo de discurso de  $X$ , onde cada conjunto representa uma classe  $k$ . O valor de aderência que cada ponto  $n$  pode possuir na classe  $k$  pode ser escrito como:

$$\mu_{kn} = \mu_{A_k}(\bar{x}_n) \in [0,1], \quad (2.44)$$

sendo que:

$$\sum_{k=1}^K \mu_{kn} = 1 \text{ para todo } n, \quad (2.45)$$

$$0 < \sum_{n=1}^N \mu_{kn} < N. \quad (2.46)$$

Na expressão acima  $N$  representa o número total de pontos. Os valores de aderência podem ser representados na forma de uma matriz, denominada matriz de partição difusa  $U$ , onde o número de linhas é igual a  $K$  e o número de colunas é igual a  $N$ . Em cada linha da matriz  $U$  representam-se os valores das aderências para cada um dos  $N$  pontos na classe  $k$ . Dado que o número de valores de aderência que cada ponto pode assumir em determinada classe é infinito, é necessário definir um critério de classificação ou função objetivo para determinar os elementos da matriz  $U$ . Na classificação difusa, a função objetivo é baseada na distância Euclidiana  $d_{nk}$  entre cada dado pontual  $n$  e o centro de classe  $k$ :

$$d_{nk} = d(\underline{x}_n - \underline{v}_k) = \sqrt{\sum_{c=1}^C (x_{nc} - v_{kc})^2}. \quad (2.47)$$

Na expressão (2.47)  $C$  representa o número total de características descritivas do local. A função objetivo é dada por:

$$F_{\text{obj}} = \sum_{n=1}^N \sum_{k=1}^K (\mu_{kn})^r (d_{nk})^2. \quad (2.48)$$

Na equação (2.48)  $r$  é o parâmetro que controla o nível de incerteza (“fuzziness”) no processo de classificação. Para  $r = 1$ , o agrupamento passa a ter aderências abruptas às classes; com valores de  $r$  aumentando, a classificação torna-se cada vez mais difusa. Conforme ROSS, citado por HALL & MINNS (1999, p. 698), o intervalo tido como razoável é  $1,25 \leq r \leq 2$ .

As coordenadas do centro da classe para determinada característica  $c$  descritiva do local na classe  $k$  podem ser dadas por:

$$U_{kc} = \frac{\sum_{n=1}^N (\mu_{kn})^r x_{nc}}{\sum_{n=1}^N (\mu_{kn})^r}. \quad (2.49)$$

Para se obter o valor difuso ótimo de  $K$ , minimiza-se a função objetivo  $F_{\text{obj}}$ . O nível de precisão  $r$  é pré-estabelecido, para o qual pode-se chegar à melhor solução através de um método iterativo. O processo pode ser descrito como segue (HALL & MINNS, 1999):

- a) Seleciona-se o número de classes  $K$  e o nível de precisão  $r$ . Cada passo pode ser denotado por um número  $(p) = 0, 1, 2, \dots$ ;
- b) Inicializa-se a matriz de partição,  $U^{(0)}$ ;
- c) Obtém-se os centros das classes,  $U_{kc}^{(p)}$ ;
- d) Obtém-se o valor de  $F_{\text{obj}}^{(p)}$ ;
- e) Atualiza-se os elementos da matriz difusa  $U^{(p)}$ , através da seguinte expressão:

$$\mu_{kn}^{(p+1)} = \left[ \sum_{j=1}^K \left( \frac{d_{nk}^{(p)}}{d_{nj}^{(p)}} \right)^{\frac{2}{r-1}} \right]^{-1}. \quad (2.50)$$

- f) Repete-se os passos (c) a (e), fazendo-se  $p = p + 1$ , até que a diferença entre  $F_{\text{obj}}^{(p+1)}$  e  $F_{\text{obj}}^{(p)}$ , seja menor que um valor limite  $\epsilon$  pré-determinado.

O algoritmo apresentado tende a ser mais lento para valores crescentes de  $r$ . A avaliação do sucesso é dada pelo coeficiente de partição difuso:

$$F_c = \frac{\text{tr}(U * U^T)}{N}. \quad (2.51)$$

Na equação (2.51)  $*$  e  $^T$  representam as operações padrões de multiplicação

e transposição de matrizes, e  $\text{tr}$  representa a soma dos elementos da diagonal da matriz entre parênteses (traço da matriz). Um valor unitário de  $F_c$  é obtido se todas as entradas em  $U$  são iguais a zero ou um, e um valor igual a  $1/K$  indica completa ambigüidade, isto é, todos os valores de aderência são iguais a  $1/K$ . Os elementos da diagonal da matriz são proporcionais às aderências não-compartilhadas dos conjuntos de dados entre as classes difusas.

Os dois métodos de “desfuzificação” que podem ser utilizados são o de máxima pertinência e o de centro de classe mais próximo. No método de máxima pertinência o ponto  $n$  é agrupado na classe que possui o valor de aderência máximo. No método do centro de classe mais próximo, cada ponto  $n$  é agrupado na classe cujo centro possui a menor distância Euclidiana até o referido ponto. Cabe lembrar que na minimização da função objetivo baseada na distância Euclidiana são considerados todos os pontos do conjunto de dados, o que não implica necessariamente, na utilização da mínima distância Euclidiana para cada ponto individualmente. Em geral, os agrupamentos gerados não apresentam grandes diferenças entre os dois métodos de “desfuzificação”.

A minimização da função objetivo expressa em (2.48) corresponde à minimização do traço da matriz de dispersão total intra-grupos (2.6), critério de otimização do método de classificação “*K-means*”. Porém, no método “*c-means*” as divisões dos dados são difusas dada a consideração do grau de aderência  $\mu$  de cada um dos pontos em mais de uma classe.

### 2.3.6. Regressão difusa

#### 2.3.6.1. Geral

A regressão difusa pode ser uma ferramenta muito útil em recursos hídricos,

considerando-se que em hidrologia geralmente necessita-se determinar a relação entre variáveis baseada em poucos e imprecisos dados. O campo de aplicação da modelagem da regressão difusa vem crescendo rapidamente. As possibilidades de aplicação são diversas, com exemplos importantes nas áreas de planejamento, projeto, operação, gerenciamento e controle (CHANG *et al.*, 1996). A regressão difusa possibilita a consideração da complexa inferência humana na avaliação do sistema. A medida do erro é o único tipo de incerteza enfatizado na regressão por mínimos quadrados, sendo que nenhuma interferência do especialista pode ser feita na análise. Neste tipo de abordagem não é possível considerar os erros que as variáveis independentes possam apresentar.

A análise de regressão difusa descreve a variável dependente  $y$  na forma de um número difuso. Em todos os casos de regressão difusa os parâmetros da regressão são considerados como números difusos. A variável independente  $x$  pode ou não ser tomada como um número difuso. O caminho pelo qual a regressão difusa é conduzida consiste em determinar os parâmetros da regressão que respeitem a minimização da imprecisão de  $y$ .

O modelo de regressão difusa pode ser genericamente representado por:

$$y^*(t) = \underline{x}(t)^T a^* + \varepsilon. \quad (2.52)$$

Na expressão (2.52)  $y^*(t)$  representa a variável dependente, sendo que  $t=1...T$  e  $T$  é o número total de pontos amostrais;  $\underline{x}(t)$  é o vetor de variáveis independentes,  $\underline{x}(t)=[x_1(t), \dots, x_{k-1}(t)]$ ;  $a^*$  é o vetor de parâmetros a serem estimados; e  $\varepsilon$  é o erro esperado. Na regressão difusa, considera-se que o vetor  $a^*$  é composto por números difusos, sendo representado por:

$$a^* = [a^*(1), \dots, a^*(K)]. \quad (2.53)$$

Na expressão (2.53)  $K$  é o número total de parâmetros a serem determinados. Além do vetor  $a^*$ , na regressão difusa, o vetor  $\underline{x}(t)$  ou a variável  $y^*(t)$ , ou ambos, podem ser tomados como um conjunto de números difusos.

Em TANAKA *et al.* (1982) é apresentado o primeiro trabalho de desenvolvimento da regressão difusa, onde foi considerado o modelo de regressão linear difusa. Neste trabalho, é demonstrada a construção de um modelo de programação linear sujeito a um grau mínimo de ajuste em cada observação, para resolver o processo de mapeamento onde deve ser considerada a menor incerteza possível. O problema consiste na otimização do número difuso previsto através da minimização de um determinado critério de incerteza.

BARDOSSY *et al.* (1990) considerou o modelo linear definido em TANAKA *et al.* (1982), apresentando porém, outros critérios de incerteza.

Em CHANG *et al.* (1996) é apresentado um modelo modificado de regressão linear difusa, denominado de modelo de regressão por metas difusas. O estudo de caso analisado demonstrou que este modelo utiliza um sistema mais simples e que geralmente fornece resultados melhores se comparados aos resultados da regressão linear difusa.

Em seguida são descritos os modelos de regressão linear difusa e regressão por metas difusas, tais como apresentados respectivamente em BARDOSSY *et al.* (1990) e CHANG *et al.* (1996).

#### 2.3.6.2. Regressão linear difusa

Em BARDOSSY *et al.* (1990) são apresentadas duas interpretações da regressão difusa. Para a compreensão das interpretações, por simplicidade, foi considerado um modelo de regressão linear com uma variável independente ( $K=2$ ) e apenas os parâmetros  $a^*$  foram considerados como números difusos. Logo, este modelo pode ser representado por:

$$y^*(t) = f[x(t), a^*] = a^*(1) + a^*(2) \cdot x(t) + \varepsilon, \quad t = 1, \dots, T. \quad (2.54)$$

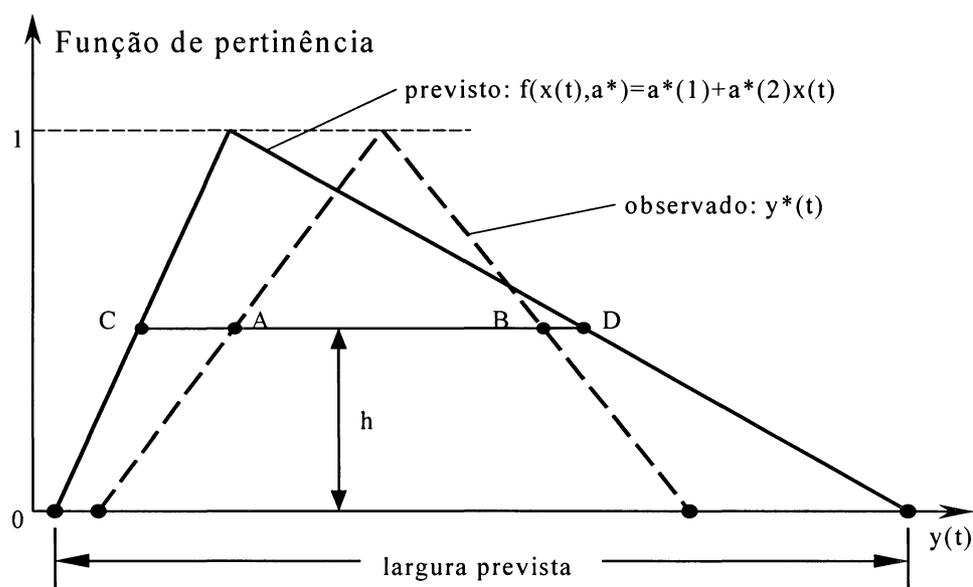
Na expressão (2.54)  $y^*(t)$  representa a variável dependente;  $x(t)$  é a variável

independente;  $a^*(1)$  e  $a^*(2)$  são os parâmetros a serem estimados;  $\varepsilon$  é o erro esperado; e  $T$  corresponde ao número total de pontos amostrais disponíveis para a análise de regressão. Na regressão linear difusa apresentada em BARDOSSY *et al.* (1990), considera-se que o vetor  $a^*$  é composto por números difusos do tipo  $L-R$  (item 2.3.3).

### Interpretação A

A figura 2.8 pode auxiliar na compreensão desta interpretação do modelo linear de regressão difusa. Esta figura ilustra as funções de pertinência do número difuso observado  $y^*(t)$  e do número difuso previsto  $f[x(t), a^*]$ . Para ambas as funções de pertinência foi considerado o mesmo nível de credibilidade  $h$  (ver item 2.3.3). Os intervalos de  $y(t)$  aceitos ao mesmo nível mínimo de aceitação são representados por  $\overline{AB}$  e  $\overline{CD}$ , respectivamente, para o número difuso observado  $y^*(t)$  e previsto  $f[x(t), a^*]$ .

FIGURA 2.8 – INTERPRETAÇÃO A DA REGRESSÃO DIFUSA NO ESPAÇO DA FUNÇÃO DE PERTINÊNCIA DE  $y^*$ . In: BARDOSSY *et al.*, 1990, p. 1499.



Considerando a interpretação A, o objetivo da análise de regressão difusa consiste em encontrar os números difusos  $a^*(1)$  e  $a^*(2)$ , tal que o intervalo ou largura do número difuso observado se encontre dentro do intervalo do número difuso

previsto, considerando para ambos os números difusos o mesmo nível de credibilidade  $h$ . Consequentemente, o segmento  $\overline{CD}$  deve conter  $\overline{AB}$ , para todo  $t = 1, \dots, T$ .

Como pode ser observado na figura 2.8, existem inúmeras soluções para este caso de análise de regressão difusa. Assim, deve ser adotado um critério de otimização do número difuso a se prever, dado pela função (2.52). O objetivo da otimização é determinar o vetor de parâmetros  $a^*$ , tal que a incerteza associada com  $f^*(t)$  seja a menor possível. Na interpretação A, a medida da incerteza  $V$  a ser minimizada é definida como a incerteza prevista, ou seja, a incerteza definida no domínio das variáveis independentes. Um dos caminhos para obter este tipo de incerteza é através da medida de imprecisão da regressão  $H(x(1), \dots, y)$ . O cálculo de  $H$  pode ser feito considerando-o igual à área sob a função de aderência  $\mu(x(1), \dots, x(I), y)$ , onde  $I$  é o número total de variáveis independentes consideradas. Utilizando como ilustração o exemplo da figura 2.8,  $H$  é a área sob o triângulo formado pela função de pertinência de  $f[x(t), a^*]$  (prevista):

$$H(x(1), \dots, x(I)) = \int_{-\infty}^{+\infty} \mu(x(1), \dots, x(I), y) \cdot dy, \quad (2.55)$$

$$\mu(x(1), \dots, x(I), y) = \max_{a^*} \{ \min_k [\mu_a(k; a^*(k)); y = f(x, a^*)] \}. \quad (2.56)$$

Na expressão (2.56)  $\mu_a(k, a^*(k))$  é a função de pertinência do número difuso  $a^*(k)$ . O domínio de  $x(i)$  é o intervalo que tem como limites mínimos e máximos, respectivamente, os valores  $x(i)^-$  e  $x(i)^+$ , que são respectivamente, os valores extremos inferior e superior das amostras de  $x$ .

Finalmente, a medida da incerteza  $V$ , é considerada como sendo a média de  $H(y)$ , expressa por:

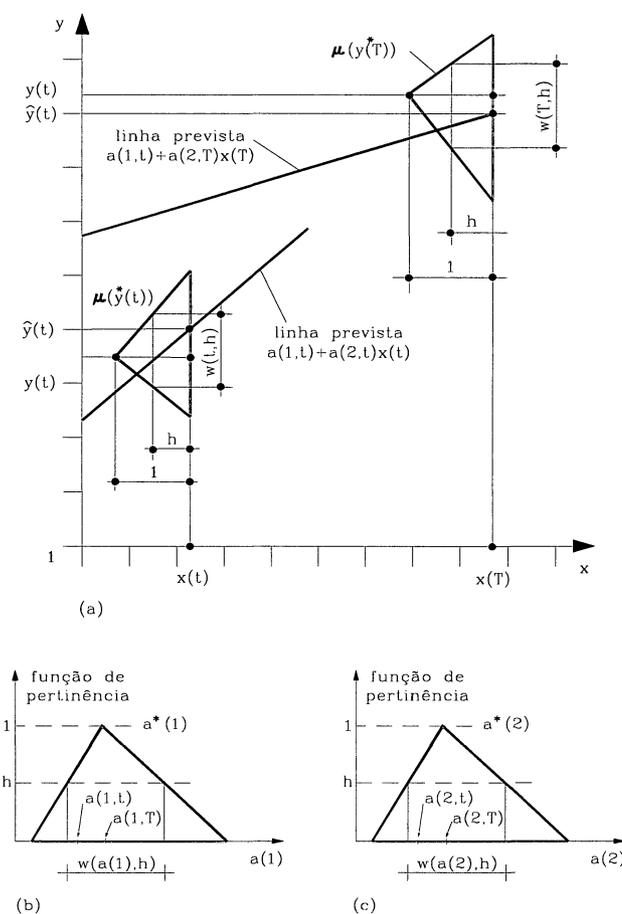
$$V = \int_{x(1)^-}^{x(1)^+} \dots \int_{x(I)^-}^{x(I)^+} H(x) \cdot dx(I) \dots dx(1). \quad (2.57)$$

Na expressão anterior,  $V$  corresponde ao hipervolume sob a função de pertinência  $\mu(y)$ .

### Interpretação B

A segunda interpretação apresentada em BARDOSSY *et al.* (1990) é ilustrada na figura 2.9, onde são considerados na parte (a) da figura, os conjuntos de 2 pontos  $(x(t), y(t))$  e  $(x(T), y(T))$ . Como pode ser observado, os números difusos  $y^*(t)$  e  $y^*(T)$ , para o mesmo nível de credibilidade  $h$ , possuem, respectivamente, as larguras  $w(t, h)$  e  $w(T, h)$ . As figuras 2.9(b) e 2.9(c) representam as funções de pertinência dos números difusos dos parâmetros  $a^*(1)$  e  $a^*(2)$ . Os intervalos  $w(a(1), h)$  e  $w(a(2), h)$  podem ser visualizados nestas figuras. Como pode ser verificado na figura 2.9(a), para o nível de credibilidade  $h$  adotado, existe um número infinito de funções  $a(1, t) + a(2, t)x(t)$ , tal que  $a(1, t)$  é um elemento de  $w(a(1), h)$  e  $a(2, t)$  é um elemento de  $w(a(2), h)$ .

FIGURA 2.9 – INTERPRETAÇÃO B DA REGRESSÃO DIFUSA NO ESPAÇO  $x, y$  E ESPAÇO DA FUNÇÃO DE PERTINÊNCIA DOS PARÂMETROS. In: BARDOSSY *et al.*, 1990, p. 1499.



A interpretação B pode ser assim resumida: para todo  $y(t)$  pertencente a  $w(t,h)$ , devem existir coeficientes  $a(1,t)$  e  $a(2,t)$ , pertencentes, respectivamente, a  $w(a(1),t)$  e  $w(a(2),t)$ , que satisfaçam a função  $y(t)=a(1,t)+a(2,t)x(t)$ . Como existem inúmeras linhas que assumem esta função e que satisfazem as condições apresentadas, deve ser adotado um critério que permita a otimização do número difuso previsto. A interpretação B pode estar associada com duas diferentes formas de medida de incerteza, largura média ou máxima largura de  $a^*(1)$  e  $a^*(2)$ . A medida da incerteza média dos parâmetros individuais a ser minimizada (TANAKA *et al.*, 1982), é expressa como:

$$V = \frac{1}{2K} \sum_{k=1}^K (b(k) + c(k)). \quad (2.58)$$

Na expressão (2.58) são consideradas as larguras dos números difusos correspondentes a cada um dos parâmetros  $a^*(k)$ . Os significados físicos de  $b(k)$  e  $c(k)$  podem ser conferidos na figura 2.6 (parâmetros  $b$  e  $c$ ).

A segunda medida de incerteza pode ser tomada como a máxima das medidas de incertezas verificadas nos parâmetros individualmente. A medida da incerteza neste caso é o maior comprimento do intervalo do número difuso entre os parâmetros  $a^*(k)$ , para o qual a função de pertinência é maior ou igual a zero:

$$V = \max_k (b(k) + c(k)). \quad (2.59)$$

#### *Obtenção do sistema a ser otimizado*

Para qualquer uma das três formas de obtenção da medida de incerteza, bem como para a regressão difusa linear e não-linear, o problema recai na programação matemática com restrições lineares. Nesta dissertação será discutido somente a regressão linear. O modelo não-linear é discutido por BARDOSSY e BARDOSSY *et al.*, conforme citado em BARDOSSY *et al.* (1990, p. 1500).

Opcionalmente, na análise de regressão pode ser considerado um ponto de referência  $\bar{x}$ . Este valor deve, se possível, ser selecionado tal que represente o ponto onde acredite-se existir maior precisão na análise de regressão. O ponto de referência pode ser a média, o máximo ou o valor médio entre os extremos das observações de  $x(t)$ , entre outros.

As expressões que formarão as restrições são:

$$\min_y Y(t, h) \geq A(1) - b(1)L^{-1}(1, h) + \sum_{x(k, t) \geq \bar{x}(k)} [(A(k) - b(k)L^{-1}(k, h))(x(k, t) - \bar{x}(k))] + \sum_{x(k, t) < \bar{x}(k)} [(A(k) + c(k)R^{-1}(k, h))(x(k, t) - \bar{x}(k))] \quad (2.60)$$

$$\max_y Y(t, h) \leq A(1) + c(1)R^{-1}(1, h) + \sum_{x(k, t) \geq \bar{x}(k)} [(A(k) + c(k)R^{-1}(k, h))(x(k, t) - \bar{x}(k))] + \sum_{x(k, t) < \bar{x}(k)} [(A(k) - b(k)L^{-1}(k, h))(x(k, t) - \bar{x}(k))] \quad (2.61)$$

Nas expressões acima  $Y(t, h)$  é a largura do intervalo do número difuso  $y^*(t)$  ao nível de credibilidade  $h$  e  $A(k)$  corresponde ao parâmetro  $S$  da figura 2.6. Considerando os critérios de medida de incerteza da interpretação B, a análise de regressão pode ser realizada através da programação linear. Para a medida de incerteza da interpretação A, a análise de regressão pode ser transformada em um modelo linear. Considerando os pontos extremos dos intervalos válidos para  $x(k)$  como sendo:

$$x(k)^- = \bar{x}(k) - d(k), \quad x(k)^+ = \bar{x}(k) + e(k).$$

A medida de incerteza (2.57) pode ser reescrita como uma função linear de  $b(k)$  e  $c(k)$ :

$$V = \int_{\bar{x}(k)-d(k)}^{\bar{x}(k)+e(k)} \dots \int_{\bar{x}(2)-d(2)}^{\bar{x}(2)+e(2)} \int_{-\infty}^{+\infty} \max_{a^*} \{ \min_k \mu_a(k; a(k)^*); y = f(x, a^*) \} dy dz(1), \dots, dx(K). \quad (2.62)$$

Os resultados da análise de regressão podem ser interpretados por 2 diferentes caminhos. Em problemas de tomada de decisão é comum utilizar os resultados como números difusos completos, ou seja, é considerada toda a função de pertinência. Se é necessário um valor ordinário, um valor único, o resultado pode ser

fornecido em termos do número difuso médio (ver item 2.3.3).

Em BARDOSSY *et al.* (1990) são apresentadas as seguintes observações:

- a) a medida de incerteza  $V$  da regressão depende do vetor de referência  $\bar{x}$  adotado, do nível de credibilidade  $h$  adotado e da função L-R adotada para representar as funções de pertinência;
- b) é importante que o pesquisador possua um conhecimento mínimo da relação existente entre a variável dependente e as variáveis independentes, a fim de definir corretamente o tipo de função que melhor represente a dependência entre  $x$  e  $y$ ;
- c) em geral, a medida de incerteza prevista (interpretação A) fornece melhores resultados que a máxima e média medida de incerteza (interpretação B);
- d) para níveis de credibilidade muito baixos (menores que 0,5) as variáveis previstas  $y(t)$  tornam-se muito imprecisas. Para níveis de credibilidade muito altos (maiores que 0,7 ou 0,8) a variável prevista  $y(t)$  se torna abrupta;
- e) em geral, as funções do tipo L-R, como por exemplo,  $1 - x^p$ , são adequadas para descrever os números difusos da regressão. Valores maiores de  $p$  indicam maior certeza acerca do valor central da função de pertinência ou a incerteza nos valores que se afastam do valor central.

No estudo de caso realizado por BARDOSSY *et al.* (1990) foi investigada a sensibilidade da análise de regressão linear difusa quanto à seleção do critério de incerteza e ponto de referência, chegando à seguinte conclusão: quando há maior credibilidade no valor do ponto de referência a ser adotado, o critério da medida de incerteza indicado é o da incerteza média (interpretação B). Caso contrário, ou seja, quando não existe nenhuma certeza sobre o ponto de referência, o critério da medida de incerteza prevista fornece melhores resultados.

### 2.3.6.3. Regressão por metas difusas

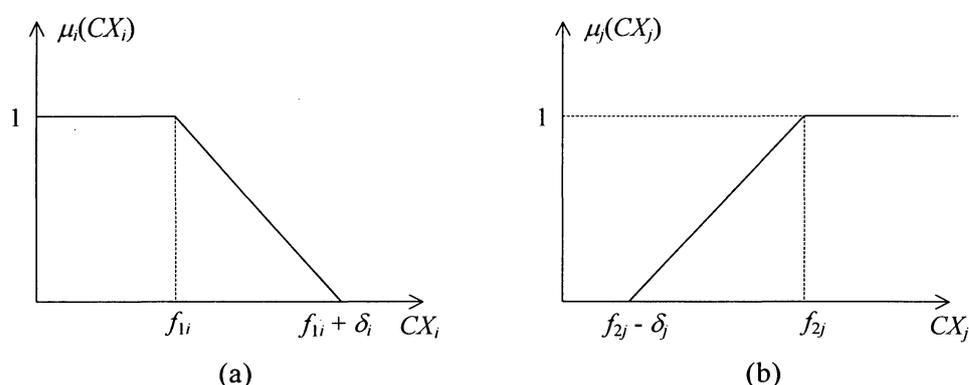
Ao contrário da regressão linear difusa, na qual é realizada a minimização da incerteza a partir de um nível mínimo de credibilidade  $h$  imposto, a regressão por metas difusas visa maximizar o nível de credibilidade, através da imposição da largura difusa determinada pelo conhecimento do especialista. O problema recai na programação objetiva com múltiplas metas difusas e pode ser simplificado por:

$$CX_i \lesseqgtr f_1, \quad (2.63)$$

$$CX_j \gtrseq f_2. \quad (2.64)$$

Nas expressões acima  $\lesseqgtr$  e  $\gtrseq$  têm, respectivamente, as interpretações lingüísticas ‘aproximadamente menor ou igual a’ e ‘aproximadamente maior ou igual a’. As equações (2.63) e (2.64) representam as restrições objetivas. É conveniente representar estas inequações por funções de pertinência lineares dos tipos não-crescente e não-decrescente. A figura 2.10 apresenta estes dois tipos de representação das funções de pertinência, onde as regiões de função crescente e decrescente são consideradas lineares. O intervalo de valores no eixo de  $CX_i$  para o qual os valores de pertinência  $\mu_i$  são decrescentes, denominado ‘intervalo de tolerância’, é igual a  $\delta_i$  (figura 2.10(a)). Analogamente, o intervalo de valores no eixo de  $CX_j$  para o qual os valores de pertinência  $\mu_j$  são crescentes é igual a  $\delta_j$  (figura 2.10(b)). Os intervalos de tolerância são, entre os valores afetados pela inferência do especialista, os de maior sensibilidade.

FIGURA 2.10 – REPRESENTAÇÃO DAS FUNÇÕES LINEARES DE PERTINÊNCIA: (a) NÃO-CRESCENTE; (b) NÃO-DECRESCENTE. In: CHANG *et al.*, 1996, p. 436.



As expressões das funções de pertinência não-crescente e não-decrescente são dadas, respectivamente, por (2.65) e (2.66):

$$\mu_i(CX_i) = \begin{cases} 1 & \text{se } CX_i \leq f_{1i}, \\ 1 - \frac{(CX_i - f_{1i})}{\delta_i} & \text{se } f_{1i} < CX_i < f_{1i} + \delta_i, \\ 0 & \text{se } CX_i \geq f_{1i} + \delta_i. \end{cases} \quad (2.65)$$

$$\mu_j(CX_j) = \begin{cases} 1 & \text{se } CX_j \geq f_{2j}, \\ 1 - \frac{(f_{2j} - CX_j)}{\delta_j} & \text{se } f_{2j} < CX_j < f_{2j} + \delta_j, \\ 0 & \text{se } CX_j \leq f_{2j} + \delta_j. \end{cases} \quad (2.66)$$

Quando o modelo de otimização torna-se muito complicado pela maximização de muitos níveis de credibilidade, o sistema pode ser simplificado pela introdução de um nível de credibilidade pré-definido.

Para o conjunto de  $n$  observações ( $x$  e  $y$ ) existem  $n$  metas difusas a serem analisadas no modelo de programação difusa. O problema é concebido como a maximização dos mínimos níveis de credibilidade das metas sujeitas às restrições. Cada restrição é representada por uma função de pertinência triangular, que é na verdade a interseção das funções de pertinência não-crescente e não-decrescente. Assim, cada meta restrita deve ser representada por duas inequações de restrições. O maior grau de ajuste ocorre quando o valor observado da variável dependente  $y$  é igual ao valor previsto de  $\sum_i a_i x_i$ .

Tomando como representação do modelo de regressão a ser resolvido a forma  $y_i = \sum_{i=1}^n a_i x_i + e_i$ , o modelo de programação de metas difusas é tomado como:

$$y_i = \sum_{i=1}^n a_i x_i + e_i, \quad \sum (h_{i1} + h_{i2}), \quad (2.67)$$

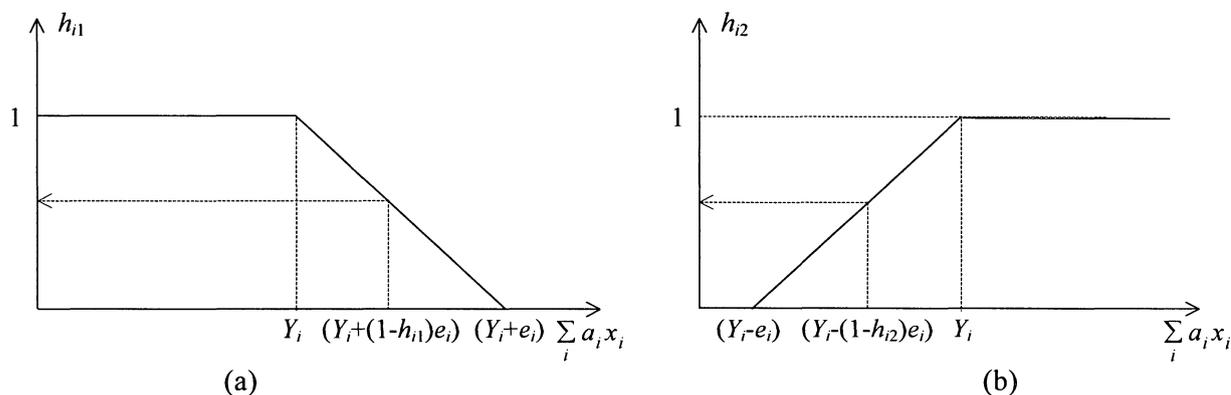
$$\text{restrito a:} \quad \sum_{j=1}^n a_{ij} x_{ij} + h_{i1} e_i \leq y_i + e_i \quad \forall i = 1, \dots, T, \quad (2.68)$$

$$\sum_{j=1}^n a_{ij} x_{ij} - h_{i2} e_i \geq y_i - e_i \quad \forall i = 1, \dots, T, \quad (2.69)$$

$$\begin{aligned} & 0 \leq h_{i1}, h_{i2} \leq 1. \\ & 0 \leq h_{i1}, h_{i2} \leq 1. \end{aligned} \quad (2.70)$$

Nas expressões anteriores  $y_i$  representa a variável dependente;  $x_{ij}$  representam as variáveis independentes, onde  $j=1\dots n$  e  $n$  representa o número total de variáveis independentes;  $a_{ij}$  são parâmetros a serem estimados;  $T$  corresponde ao número total de pontos amostrais;  $h_{i1}$  e  $h_{i2}$  são, respectivamente, os níveis de credibilidade da função de pertinência não-crescente e não-decrescente, a serem determinados na otimização; e  $e_i$  é o termo de incerteza ou intervalo de tolerância, para o qual a função de pertinência toma valores intermediários entre 0 e 1. As figuras 2.11(a) e 2.11(b) ilustram as funções de pertinência não-crescente e não-decrescente, respectivamente. A função de pertinência não-crescente (figura 2.11(a)) representa o primeiro grupo de restrições, representado por (2.68), enquanto a função de pertinência não-decrescente (figura 2.11(b)) explica o segundo grupo de restrições, representado por (2.69).

FIGURA 2.11 - FUNÇÕES DE PERTINÊNCIA UTILIZADAS NO MODELO DE REGRESSÃO DE METAS DIFUSAS: (a) NÃO-CRESCENTE; (b) NÃO-DECRESCENTE. In: CHANG *et al.*, 1996, p. 437.



Das expressões (2.67) a (2.70) e figura 2.11 podem ser verificadas as seguintes proposições (CHANG *et al.*, 1996):

*Proposição 1*

“Quando os valores de  $h_{i1}$  e  $h_{i2}$  são simultaneamente iguais a 1, o valor previsto de  $\sum_i a_i x_i$  é equivalente ao valor observado de  $y_i$ ”.

*Proposição 2*

“Quando  $h_{i1} = 1$  e  $0 \leq h_{i2} < 1$ , o valor previsto de  $\sum_i a_i x_i$  é menor que o valor observado de  $y_i$ ; analogamente, quando  $h_{i2} = 1$  e  $0 \leq h_{i1} < 1$ , o valor previsto de  $\sum_i a_i x_i$  é maior que o valor observado de  $y_i$ ”.

*Proposição 3*

“O exato grau de ajuste pode ser definido como:”

$$h_{i1} \cap h_{i2} = \min[h_{i1}, h_{i2}]. \quad (2.71)$$

*Proposição 4*

“O termo de incerteza  $e_i$  deve ser determinado antes da análise de regressão ser realizada, baseado em conhecimento do especialista. O valor de  $e_i$  pode ser assumido como uma razoável proporção de  $y_i$  de acordo com o padrão de distribuição da amostra.”

### 3. MÉTODO PROPOSTO

#### 3.1. INTRODUÇÃO

Neste capítulo descreve-se o desenvolvimento e estudo de caso do método proposto.

O objetivo geral deste trabalho de dissertação é verificar a aplicabilidade da análise difusa em recursos hídricos. A modelagem escolhida para aplicação foi regionalização hidrológica. A fim de possibilitar a avaliação desta aplicação em termos de melhoria nos resultados, selecionou-se para a análise de comparação um método de regionalização já proposto. O método escolhido para realizar a análise de comparação dos resultados foi implementado em KAVISKI (1992). O objetivo da análise foi verificar o comportamento de algumas técnicas de lógica difusa na regionalização, desta maneira, foram utilizados os mesmos dados de entrada, as mesmas técnicas não difusas complementares, os mesmos estudos de caso e as mesmas aplicações para análise dos resultados do método proposto em KAVISKI (1992). Esta atitude visou assegurar que as diferenças nos resultados se justificassem exclusivamente pelas técnicas de lógica difusa implementadas. O método desenvolvido por KAVISKI (1992) foi selecionado para comparação pelos seguintes motivos:

- trata-se de um método objetivo, de simples aplicação, que utiliza um número reduzido de dados de entrada e dispensa a utilização de mapas, gráficos e ábacos;
- foi comparado a outros métodos disponíveis no meio técnico-científico, comprovando ser mais eficaz;
- o estudo de caso envolveu bacias hidrográficas de pequenas áreas de drenagem, que são atualmente unidades hidrológicas de grande interesse para a implantação de empreendimentos ligados aos recursos hídricos;

O método apresentado para estudo de caso em KAVISKI (1992), foi subdividido em: i) método de regionalização de vazões de cheia por meio do uso de fórmulas empíricas; ii) regionalização de parâmetros de séries de vazões médias diárias; e iii) regionalização de parâmetros de séries de vazões máximas anuais. Neste trabalho de dissertação foram considerados os itens (ii) e (iii), para os quais foram aplicados estudos de caso.

As técnicas de lógica difusa consideradas mais adequadas para implementação na regionalização foram classificação difusa e regressão difusa.

## 3.2. REGIONALIZAÇÃO DE PARÂMETROS DE SÉRIES DE VAZÕES MÉDIAS DIÁRIAS

### 3.2.1. Introdução

Este método teve como premissa regionalizar parâmetros de séries de vazões médias diárias de maneira objetiva e que faça uso de dados básicos de fácil obtenção (fisiográficos e climatológicos). A unidade hidrológica focada foram bacias de pequena a média área de drenagem (entre 50 e 5000 km<sup>2</sup>). Os parâmetros regionalizados foram: média das vazões médias diárias  $\bar{Q}$ , coeficiente de variação das vazões médias diárias  $CV[Q]$  e momentos de pesos probabilísticos de ordem 1 a 3 das vazões médias diárias uniformes,  $MP_1[Q]$ ,  $MP_2[Q]$  e  $MP_3[Q]$ . Estes parâmetros são de grande importância para a obtenção de diversas características da bacia, tais como a curva de permanência, ferramenta fundamental para o desenvolvimento de estudos de PCH's (rentabilidade econômica e potência ótima de dimensionamento), sistemas de abastecimento de água, condições para navegação, entre outros. Os parâmetros regionalizados são determinados com as seguintes expressões:

$$\bar{Q} = \sum_{i=1}^n Q_i / n, \quad (3.1)$$

$$CV[Q] = \left[ \sum_{i=1}^n (Q_i - \bar{Q})^2 / (n-1) \right]^{1/2} / \bar{Q}, \quad (3.2)$$

$$MP_k[Q] = \sum_{i=1}^{n-k} Q_i (C_{n-i}^k / C_{n-1}^k) / n, \quad (3.3)$$

sendo que:

$$C_m^k = \frac{m!}{(m-k)!k!}. \quad (3.4)$$

Nestas expressões  $Q_i$  representa a vazão média diária;  $n$  é o número de elementos componentes da série histórica de vazões diárias; e  $k$  é a ordem do momento de pesos probabilísticos sendo considerado, sendo que  $k=1...3$ .

### 3.2.2. Descrição geral do método

A estrutura de regionalização utilizada nesta dissertação pode ser esquematizada pelas etapas descritas a seguir.

Foram coletados dados climatológicos, fisiográficos e fluviais, de locais situados na região de estudo, no caso, o Estado do Paraná. Os dados são referentes ao ponto de instalação das estações fluviométricas selecionadas ou às bacias hidrográficas representadas por estas estações. Os dados que não estavam sobre os locais de instalação das estações fluviométricas foram transferidos para estes locais por alguma técnica de regionalização.

A próxima etapa foi utilizar a análise de agrupamentos para definir  $K$  regiões homogêneas em termos das variáveis  $\bar{Q}$ ,  $CV[Q]$ ,  $MP_1[Q]$ ,  $MP_2[Q]$  e  $MP_3[Q]$ . Através da análise de agrupamentos foi possível determinar os parâmetros classificadores, ou seja, os parâmetros climatológicos e fisiográficos que foram utilizados em uma aplicação de caso na etapa denominada de análise discriminante. Os parâmetros climatológicos e fisiográficos utilizados foram: latitude  $x_F$ , longitude  $y_F$ , altitude  $z_F$ ,

área de drenagem  $A$ , precipitação total média anual de longo período  $\bar{P}$  e o coeficiente de variação das precipitações totais anuais  $CV[P]$ . A técnica utilizada para a definição das regiões homogêneas foi a classificação difusa.

Para cada uma das regiões homogêneas foram definidas as equações que determinam os parâmetros  $\bar{Q}$ ,  $CV[Q]$ ,  $MP_1[Q]$ ,  $MP_2[Q]$  e  $MP_3[Q]$  em função das características da bacia. Os dados utilizados na determinação das equações de regressão foram os dados observados ou regionalizados das estações do agrupamento de cada região homogênea. Nesta dissertação foi proposta a regressão difusa para a determinação das expressões que relacionam os parâmetros.

Após o desenvolvimento do método de regionalização foram implementadas as técnicas utilizadas para a aplicação do método em uma bacia hidrográfica sem informações hidrológicas situada na região do estudo. Os dados necessários que foram coletados referentes à bacia sem informações hidrológicas para aplicação do método são: latitude, longitude, altitude e área de drenagem. Em função da localização e altitude desta bacia, foram determinados por regionalização (interpolação), a precipitação total média anual de longo período e o coeficiente de variação das precipitações totais anuais. Assim, em função dos parâmetros classificadores determinados na análise de agrupamentos foi realizada a análise discriminante para definir as probabilidades de que o local em análise pertença a cada uma das  $K$  regiões homogêneas. Enfim, os parâmetros  $\bar{Q}$ ,  $CV[Q]$ ,  $MP_1[Q]$ ,  $MP_2[Q]$  e  $MP_3[Q]$  foram determinados pela equação de regressão definida para a região homogênea em que se enquadrava o local estudado.

### 3.2.3. Dados utilizados

Foram selecionadas tantas estações fluviométricas sobre a área de interesse quanto possível. Destas estações interessaram os dados de vazões médias diárias: média  $\bar{Q}$  ( $l/s/km^2$ ), coeficiente de variação  $CV[Q]$  e momentos de pesos

probabilísticos de ordem 1 a 3 das vazões médias diárias uniformes,  $MP_1[Q]$ ,  $MP_2[Q]$  e  $MP_3[Q]$ ; dados relacionados ao ponto de instalação da estação: latitude  $x_F$  (graus decimais), longitude  $y_F$  (graus decimais) e altitude  $z_F$  (m); e dados relacionados às bacias hidrográficas drenadas pelas estações: área de drenagem  $A$  (km<sup>2</sup>).

Foram também selecionadas diversas estações pluviométricas localizadas sobre e no entorno da área de interesse. Das informações pluviométricas foi extraída a precipitação total média anual de longo período  $\bar{P}$  (mm) e o coeficiente de variação das precipitações totais anuais  $CV[P]$ , além da latitude  $x_P$  (graus decimais), longitude  $y_P$  (graus decimais) e altitude  $z_P$  (m) da estação.

Outros tipos de dados característicos do local poderiam ser selecionados para a aplicação do método, desde que disponíveis nos locais desprovidos de informações fluviais. Exemplos destes dados são informações meteorológicas como temperatura, umidade, insolação, entre outros. Os dados utilizados nesta pesquisa são de fácil obtenção e acredita-se que possuem grande influência nos parâmetros de séries de vazões médias diárias.

Em KAVISKI (1992) foram selecionadas para a região de estudo 55 estações fluviométricas e 94 estações pluviométricas. Um dos critérios de seleção utilizados para a escolha das estações fluviométricas foi relativo à área de drenagem ser menor que 5000 km<sup>2</sup>. O grupo de estações é o mesmo utilizado em estudos de regionalização para pequenas bacias do Estado do Paraná desenvolvidos por KAVISKI e FIOR (1986). Os parâmetros regionalizados nestes estudos são de: vazões médias, mínimas e máximas. O último ano de dados utilizado destas estações é 1982. Cabe ressaltar que, nesta dissertação, não foi efetuada uma atualização dos dados a fim de assegurar que as diferenças nos resultados entre os métodos comparados possam ser avaliadas exclusivamente pela implementação de técnicas de análise difusa.

Dados tais como, área de drenagem, latitude, longitude e altitude das estações fluviométricas foram obtidos de BRASIL (1987). Os dados latitude, longitude e altitude das estações pluviométricas foram obtidos de BRASIL (1981). Os

parâmetros  $\bar{Q}$ ,  $CV[Q]$ ,  $MP_1[Q]$ ,  $MP_2[Q]$  e  $MP_3[Q]$  de cada uma das estações fluviométricas foram obtidos de KAVISKI e FIOR (1985). Os valores de  $\bar{P}$  e  $CV[P]$  de cada estação pluviométrica foram obtidos de HOLTZ, citado por KAVISKI (1992, p. 167). Estes valores foram obtidos dos dados observados entre os anos de 1954 e 1964.

Foram considerados os valores de  $\bar{P}$  e  $CV[P]$  regionalizados para as estações fluviométricas obtidos diretamente de KAVISKI (1992). Os valores de  $\bar{P}$  e  $CV[P]$ , disponíveis para as estações pluviométricas, foram transferidos para os locais de instalação das estações fluviométricas utilizadas no estudo, através do método de regionalização implementado em KAVISKI (1992). Neste trabalho, a regionalização de  $\bar{P}$  e  $CV[P]$  foi efetuada de acordo com os seguintes critérios: para estações fluviométricas localizadas em regiões de interpolação, os valores foram obtidos por superfície “*spline*”; e para estações fluviométricas localizadas em regiões de extrapolação foi utilizada a interpolação ponderada.

Na tabela 3.1 são apresentadas as estações fluviométricas utilizadas no estudo, cuja identificação é feita através do código ANEEL<sup>1</sup> e nome. Nesta tabela estão também disponíveis as seguintes informações: período de dados utilizado e os parâmetros básicos a serem utilizados na regionalização hidrológica:  $x_F$ ,  $y_F$ ,  $z_F$ ,  $A$ ,  $\bar{Q}$ ,  $CV[Q]$ ,  $MP_1[Q]$ ,  $MP_2[Q]$ ,  $MP_3[Q]$ ,  $\bar{P}$  e  $CV[P]$ . A figura 3.1 apresenta a localização das estações fluviométricas dentro do Estado do Paraná.

Na tabela 3.2 são apresentadas as estações pluviométricas utilizadas no estudo e os respectivos parâmetros básicos. Constam desta tabela: código ANEEL<sup>2</sup>, nome,  $x_P$ ,  $y_P$ ,  $z_P$ ,  $\bar{P}$  e  $CV[P]$ . A figura 3.2 apresenta a localização das estações pluviométricas dentro da área de estudo. Como pode ser observado na tabela 3.2, o dado  $CV[P]$  não é disponível para todas as estações pluviométricas da lista. Este parâmetro foi regionalizado para as estações fluviométricas em função de um número reduzido de dados amostrais.

TABELA 3.1 - ESTAÇÕES FLUVIOMÉTRICAS E PARÂMETROS BÁSICOS UTILIZADOS NO ESTUDO

continua

Nº	Código	Nome	Período		Latitude		Longit.		Altitude (m)	Área (km <sup>2</sup> )	$\bar{Q}$ (l/s/km <sup>2</sup> )	CV[Q]	MP <sub>1</sub> [Q]	MP <sub>2</sub> [Q]	MP <sub>3</sub> [Q]	$\bar{P}$ (mm)	CV[P]
			Início	Fim	°	'	°	'									
1	64248000	Salto do Itararé	1931	1970	23	36	49	37	470	5 000	11,616	0,9357	0,3154	0,1778	0,1214	1 099	0,1900
2	64360000	Tomazina	1926	1982	23	46	49	57	483	2 020	13,030	1,1752	0,2753	0,1424	0,0915	994	0,1760
3	64380000	Figueira	1953	1983	23	51	50	23	526	1 030	12,680	1,3116	0,2467	0,1233	0,0783	1 078	0,1785
4	64390000	Ponte Sta. Terezinha	1931	1982	23	06	50	27	340	3 450	10,833	1,6367	0,2098	0,0991	0,0615	1 283	0,1800
5	64440000	Santa Cruz	1937	1967	25	12	50	09	790	1 340	16,083	1,0291	0,2712	0,1385	0,0893	1 366	0,2600
6	64460000	Bom Jardim	1941	1982	24	42	50	29	750	722	15,617	1,3956	0,2589	0,1348	0,0874	1 443	0,1900
7	64475000	Castro	1930	1954	24	47	50	00	990	1 190	15,715	1,0977	0,2670	0,1326	0,0833	1 482	0,1904
8	64500000	Salto São Pedro	1957	1983	23	40	50	49	940	194	5,724	0,7107	0,3430	0,1957	0,1341	1 555	0,1446
9	64547000	Santa Fé	1950	1965	22	59	51	49	550	1 020	15,931	0,6499	0,3422	0,1908	0,1283	1 390	0,1868
10	64550000	Vila Silva Jardim	1967	1982	22	50	52	06	250	4 630	16,359	0,6300	0,3494	0,1981	0,1340	1 373	0,1923
11	64620000	Rio dos Patos	1930	1982	25	12	50	56	690	1 090	19,418	1,4347	0,2265	0,1087	0,0674	1 435	0,2000
12	64625000	Tereza Cristina	1956	1982	24	50	51	09	550	3 570	19,691	1,6953	0,2002	0,0924	0,0566	1 502	0,1798
13	64660000	Rib. Raposo Tavares	1966	1983	24	03	51	57	330	3 240	21,355	1,6859	0,1888	0,0807	0,0463	1 516	0,2106
14	64775000	Balsa do Cantú	1967	1982	24	45	52	42	360	2 510	21,067	1,6357	0,2048	0,0955	0,0590	1 735	0,1920
15	64780000	Ponte Tourinho	1966	1979	24	53	53	04	700	288	22,336	0,7117	0,3213	0,1729	0,1137	1 674	0,1865
16	64785000	Ponte do Goio-Bang	1967	1982	24	37	52	56	320	1 350	17,060	0,6608	0,3356	0,1847	0,1226	1 712	0,1933
17	64790000	Salto Sapucaí	1964	1983	24	38	53	06	730	535	32,128	1,4771	0,2549	0,1344	0,0882	1 577	0,1875
18	64810000	Balsa do Goio-Êre	1963	1982	23	55	53	08	320	2 420	16,973	0,5555	0,3621	0,2098	0,1448	1 493	0,1938
19	64890000	Santa Helena	1962	1975	24	53	54	21	250	1 660	24,087	1,6018	0,2481	0,1249	0,0789	1 588	0,1997
20	65010000	Fazendinha	1955	1983	25	34	49	04	910	110	25,438	0,8631	0,3082	0,1675	0,1112	1 638	0,2050
21	65015000	Campina do Taquaral	1963	1978	25	36	49	12	910	258	18,778	0,8378	0,2908	0,1507	0,0979	1 468	0,2037
22	65035000	Porto Amazonas	1935	1982	25	33	49	53	754	3 660	16,441	0,9219	0,2836	0,1443	0,0919	1 383	0,2010
23	65085000	Bateias de Baixo	1941	1967	26	04	49	16	800	490	22,732	0,7919	0,3249	0,1837	0,1256	1 402	0,1900
24	65090000	Fragosos	1967	1982	26	09	49	23	790	800	22,968	0,6962	0,3308	0,1837	0,1239	1 432	0,1871
25	65095000	Rio Preto do Sul	1951	1982	26	13	49	36	780	2 500	19,740	0,8161	0,3105	0,1690	0,1125	1 394	0,1852
26	65100000	Rio Negro	1930	1982	26	06	49	48	766	3 380	18,272	0,8765	0,3010	0,1625	0,1079	1 279	0,1784
27	65130000	Salto Baraça	1956	1981	25	55	49	16	920	294	22,496	1,0614	0,2604	0,1270	0,0786	1 391	0,2089
28	65135000	R. da Várzea dos Limas	1939	1982	25	57	49	23	810	602	17,809	0,8162	0,3119	0,1703	0,1138	1 305	0,1900

TABELA 3.1 - ESTAÇÕES FLUVIOMÉTRICAS E PARÂMETROS BÁSICOS UTILIZADOS NO ESTUDO

Nº	Código	Nome	Período		Latitude		Longit.		Altitude (m)	Área (km <sup>2</sup> )	$\bar{Q}$ (l/s/km <sup>2</sup> )	CV[Q]	MP <sub>1</sub> [Q]	MP <sub>2</sub> [Q]	MP <sub>3</sub> [Q]	conclusão	
			Início	Fim	°	'	°	'								$\bar{P}$ (mm)	CV[P]
29	65155000	São Bento	1930	1982	25	56	49	47	750	2 010	15,888	0,7199	0,3210	0,1747	0,1163	1 301	0,1810
30	65180000	Salto Canoinhas	1951	1982	26	22	50	17	765	793	20,960	1,3813	0,2309	0,1088	0,0662	1 494	0,1791
31	65200000	Pontilhão	1939	1965	25	52	50	33	750	1 960	17,848	1,0100	0,2513	0,1175	0,0709	1 489	0,1728
32	65365000	Porto Vitória	1945	1982	26	10	51	13	790	165	28,191	0,8780	0,2996	0,1576	0,1016	1 580	0,1600
33	65370000	Jangada	1945	1983	26	22	51	15	800	1 060	23,042	1,1873	0,2503	0,1222	0,0761	1 541	0,1800
34	65385000	Col. Augusto Loureiro	1955	1980	26	11	51	17	790	1 240	25,372	1,1955	0,2523	0,1245	0,0780	1 581	0,1615
35	65415000	Fazenda Maracanã	1945	1982	26	02	51	09	840	320	25,343	1,1716	0,2639	0,1347	0,0865	1 829	0,1800
36	65770000	Balsa do Pinhalzinho	1963	1980	25	56	51	24	750	1 650	23,171	1,1057	0,2650	0,1338	0,0853	1 627	0,1488
37	65810000	Guarapuava	1936	1967	25	27	51	27	950	731	17,491	1,2303	0,2445	0,1188	0,0744	1 766	0,1566
38	65815000	Salto Curucaca	1960	1983	25	31	51	47	1 000	2 230	28,195	1,0787	0,2731	0,1394	0,0887	1 807	0,1655
39	65825000	Santa Clara	1949	1983	25	38	51	58	740	3 910	26,032	0,9905	0,2790	0,1428	0,0912	1 793	0,2000
40	65855000	Usina Cavernoso	1951	1983	25	30	52	13	850	1 500	25,172	1,1207	0,2640	0,1304	0,0807	1 863	0,1935
41	65890000	Campo Novo	1940	1968	25	28	52	54	550	142	26,818	1,3817	0,2507	0,1263	0,0806	1 826	0,1500
42	65925000	Salto Claudelino	1965	1983	26	17	52	21	1 026	1 780	31,100	1,0746	0,2687	0,1347	0,0844	1 701	0,1805
43	65927000	Porto Palmeirinha	1955	1982	26	02	52	40	450	3 410	27,431	1,0939	0,2632	0,1306	0,0818	1 761	0,1789
44	65945000	Ponte do Vitorino	1956	1982	26	03	52	48	550	545	25,195	1,0511	0,2753	0,1402	0,0891	1 856	0,1815
45	65955000	Balsa do Santana	1956	1982	25	54	52	51	450	1 720	26,719	1,3688	0,2353	0,1132	0,0693	1 855	0,1700
46	81060000	Salto São Luiz	1952	1982	25	11	49	30	892	250	8,324	0,5111	0,3656	0,2082	0,1413	1 219	0,2090
47	81102000	Balsa do Jacaré	1936	1982	24	56	49	29	398	1 750	13,207	0,8546	0,3532	0,2083	0,1457	1 560	0,1713
48	81125000	Turvo	1945	1982	24	45	49	20	400	392	17,671	0,6727	0,3624	0,2144	0,1501	1 305	0,1810
49	81135000	Balsa do Cerro Azul	1930	1982	24	47	49	16	270	4 570	14,776	0,7507	0,3488	0,2030	0,1407	1 003	0,2263
50	81140000	Cerro Azul	1930	1982	24	49	49	16	320	435	10,120	0,8634	0,3572	0,2131	0,1501	1 150	0,1900
51	81300000	Praia Grande	1930	1969	25	10	48	53	750	920	18,337	0,7732	0,3355	0,1935	0,1344	1 262	0,1900
52	81710000	Jacupiranga	1951	1968	24	42	48	00	40	790	26,296	1,0781	0,2894	0,1606	0,1096	1 907	0,2029
53	82160000	Véu da Noiva	1930	1957	25	26	48	57	680	54	41,267	1,9120	0,2819	0,1500	0,0964	3 586	0,1800
54	82230000	Vossoroca	1931	1944	25	49	49	04	910	151	19,178	0,8289	0,3249	0,1843	0,1263	1 629	0,1930
55	82230800	Salto do Tanque	1931	1946	25	48	49	00	910	253	22,704	1,2594	0,3021	0,1704	0,1165	1 755	0,1898

FONTE: ANEEL

NOTA: Abreviaturas utilizadas:

Longit. – Longitude.



TABELA 3.2 - ESTAÇÕES PLUVIOMÉTRICAS E PARÂMETROS  $\bar{P}$  E CV[P]

continua

Nº	Código	Nome	Latitude		Longitude		Altitude (m)	$\bar{P}$ (mm)	CV[P]
			°	'	°	'			
1	02251021	Balsa do Paranapanema	22	39	51	22	250	1 154,1	...
2	02251024	Porecatu	22	46	51	23	350	1 245,8	...
3	02251026	Santa Fé	22	59	51	50	550	1 389,7	...
4	02252011	Fazenda Sasaki Irmãos	22	42	52	00	460	1 280,1	...
5	02349028	Jacarézinho	23	10	49	58	471	1 289,5	0,16
6	02349031	Salto do Itararé	23	36	49	37	470	1 098,9	...
7	02349066	Ponte Melo Peixoto	23	01	49	54	390	1 080,4	0,21
8	02350000	Jataizinho	23	15	50	59	340	1 220,1	0,19
9	02350001	Porto Santa Terezinha	23	06	50	27	340	1 283,2	0,18
10	02350002	Andirá	23	05	50	17	375	1 126,6	0,26
11	02350008	Nova Fátima	23	26	50	33	680	1 321,0	...
12	02350009	Figueira	23	51	50	23	526	1 077,5	...
13	02350015	Usina Apucarantina	23	42	50	56	450	1 472,1	0,16
14	02350016	Cachoeira - Salto Fiu	23	43	50	56	500	1 444,1	...
15	02351009	Londrina	23	21	51	09	610	1 437,7	0,18
16	02351012	Arapongas	23	24	51	26	810	1 480,0	0,15
17	02351014	Maringá	23	25	51	58	555	1 615,6	...
18	02351061	Apucarana	23	33	51	27	870	1 523,4	...
19	02352000	Porto Paraíso do Norte	23	19	52	40	250	1 326,8	...
20	02352018	Cianorte	23	39	52	39	560	1 374,1	...
21	02353001	Santa Isabel do Ivaí	23	00	53	11	400	1 327,9	...
22	02449000	Capela da Ribeira	24	39	49	00	180	1 136,2	0,30
23	02449005	Itararé	24	06	49	20	737	1 187,3	...
24	02449008	Cerro Azul	24	49	49	16	320	1 150,5	0,19
25	02449019	Jaguariaíva	24	14	49	42	891	1 383,5	0,18
26	02450002	Tibagi	24	30	50	24	750	1 354,5	0,20
27	02450005	Telêmaco Borba	24	17	50	38	650	1 333,5	0,15
28	02450006	Salto Mauá	24	03	50	42	550	1 402,5	0,19
29	02450009	Bom Jardim	24	42	50	29	750	1 443,0	0,19
30	02450010	Castro	24	47	50	00	996	1 488,4	0,19
31	02451004	Posto Indígena do Ivaí	24	32	51	38	790	1 489,5	...
32	02451014	Ivaiporã	24	15	51	39	650	1 660,9	...
33	02452006	Usina São João Auxiliar	24	05	52	19	630	1 535,3	...
34	02454000	Porto Mendes Gonçalves	24	30	54	20	150	1 552,5	0,21
35	02454001	Guaíra	24	04	54	15	218	1 464,4	0,21
36	02548000	Morretes	25	28	48	50	8	1 779,4	0,18
37	02548001	Praia Grande	25	10	48	53	750	1 262,3	0,19
38	02548002	Véu da Noiva	25	26	48	57	680	3 586,5	0,18
39	02548003	Colônia da Cachoeira	25	14	48	45	80	2 363,9	0,12
40	02548004	Matinhos	25	49	48	32	2	2 897,4	0,17
41	02548006	Guaraqueçaba	25	18	48	20	2	2 812,5	...
42	02548007	Mergulhão	25	18	48	43	9	2 261,7	...
43	02548008	Guaratuba	25	53	48	34	6	1 895,3	...
44	02548010	Paranaguá	25	31	48	31	9	2 017,5	0,13
45	02548013	Usina Chaminé	25	48	48	59	800	2 239,9	...
46	02548016	Alexandra	25	34	48	38	10	1 631,7	0,26
47	02548017	Saguarema	25	31	48	43	5	1 716,3	0,36

TABELA 3.2 - ESTAÇÕES PLUVIOMÉTRICAS E PARÂMETROS  $\bar{P}$  E CV[P]

Nº	Código	Nome	Latitude		Longitude		Altitude (m)	$\bar{P}$ (mm)	conclusão CV[P]
			°	'	°	'			
48	02548018	Bairro Alto	25	14	48	44	120	2 355,7	...
49	02548027	Marumbi	25	30	48	52	60	1 864,6	0,19
50	02548030	Rio Capivari	25	12	48	58	350	2 800,5	...
51	02549000	São Bento	25	58	49	49	750	1 283,3	0,18
52	02549001	Porto Amazonas	25	33	49	53	793	1 284,5	0,21
53	02549002	Curitiba	25	25	49	18	945	1 335,9	0,19
54	02549003	Rio da Várzea dos Lima	25	57	49	23	810	1 304,5	0,19
55	02549006	Curitiba	25	26	49	16	947	1 403,3	0,17
56	02549008	Vossoroca	25	49	49	05	805	1 549,2	0,14
57	02549010	Roça Nova	25	28	49	01	950	1 703,3	0,21
58	02549012	Votuverava – R. Branco do Sul	25	08	49	17	950	1 341,3	...
59	02549013	Engenheiro Bley	25	38	49	45	864	1 215,8	0,26
60	02549032	Colombo	25	20	49	14	960	1 278,8	0,17
61	02550000	Rio dos Patos	25	12	50	56	690	1 435,5	0,20
62	02550001	São Mateus do Sul	25	52	50	23	760	1 405,0	0,19
63	02550002	Sumidouro	25	02	50	05	850	1 379,6	0,20
64	02550003	Santa Cruz	25	12	50	09	790	1 366,3	0,26
65	02550004	Braço do Potinga	25	52	50	49	760	1 491,9	0,14
66	02550005	Rio Claro do Sul	25	56	50	41	750	1 567,4	0,16
67	02550006	Ivaí	25	01	50	51	768	1 517,6	0,22
68	02550007	Irati	25	28	50	38	910	1 434,7	...
69	02550008	Ponta Grossa	25	06	50	10	869	1 421,3	0,18
70	02550011	Usina Pitanguí	25	02	50	07	850	1 398,0	0,20
71	02550015	Usina Manoel Ribas	25	19	50	00	870	1 552,0	...
72	02551001	Leonópolis	25	41	51	12	960	1 782,4	0,14
73	02551004	Santa Clara	25	38	51	58	740	1 793,3	0,20
74	02551006	Guarapuava	25	24	51	28	1 108	1 662,5	0,17
75	02552000	Quedas do Iguaçu	25	28	52	54	550	1 826,0	0,15
76	02552001	Águas do Verê	25	46	52	56	390	1 930,0	0,19
77	02552002	Balsa do Santana	25	54	52	51	450	1 855,1	0,17
78	02552005	Usina Rio Cavernoso	25	30	52	13	600	1 774,5	...
79	02553000	Salto Osório (Rio Iguaçu)	25	32	53	02	380	1 991,1	0,23
80	02554001	Parque Nacional do Iguaçu	25	37	54	29	100	1 555,7	0,23
81	02554002	Salto Cataratas	25	41	54	26	152	1 749,5	0,22
82	02554004	Foz do Iguaçu	25	33	54	34	161	1 516,9	0,18
83	02649006	Rio Negro	26	06	49	48	770	1 276,2	0,18
84	02649021	Rio Negro	26	06	49	48	790	1 280,3	0,19
85	02649011	Bateias de Baixo	26	04	49	16	800	1 401,9	0,19
86	02651000	União da Vitória	26	14	51	04	736	1 463,1	0,13
87	02651003	Jangada	26	22	51	15	800	1 540,9	0,18
88	02651004	Porto Vitória	26	10	51	13	790	1 579,9	0,16
89	02651005	Fazenda Maracanã	26	02	51	09	840	1 828,6	0,18
90	02651006	Palmas	26	29	51	59	1 091	2 100,0	...
91	02651011	Matos Costa (São João)	26	29	51	10	1 201	1 604,9	...
92	02651013	Colônia Augusto Loureiro	26	11	51	17	890	1 651,6	...
93	02651015	Usina Bituruna	26	10	51	33	1 000	1 613,2	...
94	02652017	Usina Francisco Beltrão	26	03	52	56	600	1 920,4	...

FONTE: ANEEL

NOTA: Sinais convencionais utilizados:

... Dado não disponível.



### 3.2.4. Definição das regiões homogêneas

#### 3.2.4.1. Desenvolvimento

A análise de agrupamentos foi realizada usando-se classificação difusa “*c-means*” (HALL & MINNS, 1999). A técnica foi utilizada com o auxílio de ferramenta específica, implementada no MATLAB através do “*Fuzzy Logic Toolbox*” (MathWorks, 1998).

Os melhores agrupamentos dos parâmetros hidrológicos  $\bar{Q}$ ,  $CV[Q]$ ,  $MP_1[Q]$ ,  $MP_2[Q]$  e  $MP_3[Q]$  foram determinados em função dos parâmetros  $x_F$ ,  $y_F$ ,  $z_F$ ,  $A$ ,  $\bar{P}$  e  $CV[P]$ . Foram realizadas todas as combinações possíveis, envolvendo os parâmetros classificadores a fim de determinar os melhores agrupamentos.

O método proposto de classificação difusa consiste em determinar os elementos da matriz de partição difusa  $U$ , através da minimização da função objetivo dada por (2.48). Os elementos da matriz são os valores de pertinência que cada elemento pode assumir em cada grupo. O número de linhas da matriz  $U$  corresponde ao número de classes e o número de colunas corresponde ao número total de estações fluviométricas.

O método de “desfuzificação” implementado no MATLAB busca agrupar cada estação na classe em que possui a maior aderência.

As diversas características descritivas de cada um dos locais foram normalizadas através da equação:

$$x_{nc} = \frac{y_{nc} - \bar{y}_c}{\sigma_c}. \quad (3.5)$$

Na expressão (3.5)  $y_{nc}$  representa a característica  $c$  do local  $n$ ,  $\bar{y}_c$  é a média da característica  $c$  entre os conjuntos de dados; e  $\sigma_c$  é o desvio padrão da

característica  $c$ .

O método de classificação difusa “*c-means*” não permite a definição a priori do número ideal de classes. Logo, o número ideal de classes foi determinado iterativamente, analisando-se a vantagem de se passar de  $K$  para  $K+1$  grupos através da verificação da consequência hidrológica. Nesta técnica avalia-se a razão entre a variação entre agrupamentos  $V_B$  e a variação dentro dos agrupamentos  $V_W$ , dos parâmetros hidrológicos a serem regionalizados. Os valores de  $V_B$  e  $V_W$  são obtidos por (2.12) e (2.13), respectivamente. O número ideal de classes é considerado igual a  $K$ , quando a relação  $V_B/V_W$  passa a ser decrescente ao utilizar  $K+1$  grupos. Para evitar a formação de agrupamentos com o número de estações muito discrepantes entre si, o valor ideal de  $K$  foi determinado também em função de (2.9).

Em KAVISKI (1992) o número ideal de agrupamentos também foi definido em função da verificação da consequência hidrológica e que ao menos a equação expressa em (2.9) fosse satisfeita. O método utilizado pelo autor no referido trabalho para análise de agrupamentos foi o “*K-means*”. Este método utiliza como critério de otimização a minimização do erro devido ao particionamento de  $N$  objetos em  $K$  agrupamentos, através da distância de Mahalanobis. O erro devido ao particionamento e a distância de Mahalanobis são dados, respectivamente, por (2.6) e (2.7).

#### 3.2.4.2. Resultados

O agrupamento adotado foi o gerado para 7 grupos em função dos parâmetros classificadores  $x_F$ ,  $y_F$ ,  $\bar{P}$  e  $CV[P]$ . Este agrupamento obteve o maior valor da média de  $V_B/V_W$  entre os agrupamentos que verificam a equação expressa em (2.9). Para este agrupamento a equação (2.10) também foi verificada. Os valores da relação  $V_B/V_W$ , obtidos considerando o agrupamento adotado, para os parâmetros hidrológicos  $\bar{Q}$ ,  $CV[Q]$ ,  $MP_1[Q]$ ,  $MP_2[Q]$  e  $MP_3[Q]$  são, respectivamente, iguais a: 1.555, 0.454, 0.801, 0.935 e 1.025. Estes resultados são melhores se comparados aos valores obtidos

em KAVISKI (1992), onde o agrupamento adotado considerou  $x_F$  e  $y_F$  como parâmetros classificadores, o valor de  $K$  resultou igual a 5 e os valores de  $V_B/V_W$  determinados para os parâmetros  $\bar{Q}$ ,  $CV[Q]$ ,  $MP_1[Q]$ ,  $MP_2[Q]$  e  $MP_3[Q]$  são respectivamente iguais a: 1.260, 0.076, 0.152, 0.188 e 0.215.

Conforme indicação de ROSS, citado por HALL & MINNS (1999, p. 698), os valores de  $r$ , parâmetro que especifica o nível de incerteza e é utilizado na função expressa em (2.48), são recomendados para variar de 1,25 a 2,0. O valor de  $r$  foi adotado igual ao máximo indicado ( $r=2,0$ ), indicando assim um alto nível de incerteza a respeito da ambigüidade dos dados.

O coeficiente de partição difuso  $F_c$ , obtido pela expressão (2.51) efetua a avaliação do sucesso do agrupamento e resultou igual a 0,80. Este resultado revela que a classificação foi bem sucedida, pois está bem acima de 0,14 ( $1/K$ ). Quando  $F_c$  resulta igual a  $1/K$  indica extrema ambigüidade dos resultados.

A tabela 3.3 apresenta a relação de estações fluviométricas, identificadas através de seus códigos ANEEL, classificadas em cada um dos 7 grupos.

TABELA 3.3 – AGRUPAMENTOS DETERMINADOS NA CLASSIFICAÇÃO DIFUSA

GRUPOS						
Grupo 1	Grupo 2	Grupo 3	Grupo 4	Grupo 5	Grupo 6	Grupo 7
64248000	64440000	64660000	65010000	65180000	65810000	81060000
64360000	64460000	64775000	65015000	65200000	65815000	81125000
64380000	64475000	64780000	65035000	65365000	65825000	81135000
64390000	64620000	64785000	65085000	65370000	65855000	81140000
64500000	64625000	64790000	65090000	65385000	65890000	81300000
64547000	81102000	64810000	65095000	65415000	65925000	
64550000	81710000	64890000	65100000	65770000	65927000	
			65130000		65945000	
			65135000		65955000	
			65155000		82160000	
			82230000			
			82230800			

### 3.2.5. Determinação das equações de regressão que estimam $\bar{Q}$ , $CV[Q]$ , $MP_1[Q]$ , $MP_2[Q]$ e $MP_3[Q]$

#### 3.2.5.1. Desenvolvimento

O modelo de regressão difusa adotado nesta dissertação foi o por metas difusas, pois conforme explanado em CHANG *et al.* (1996), este modelo parece fornecer um melhor ajuste da variável dependente  $y$  se comparado ao modelo de regressão linear difusa. Os parâmetros hidrológicos a serem regionalizados foram determinados por expressões que relacionem estes parâmetros com as características fisiográficas e climatológicas da bacia estudada. Estas expressões foram determinadas uma para cada grupo homogêneo definido na análise de agrupamentos, através da análise de regressão por metas difusas. Todos os parâmetros considerados na regressão foram lognormalizados. O modelo de regressão difusa pode ser generalizado por:

$$\ln(y_i) = a_0^* + \sum_{j=1}^p a_j^* \ln(x_{i,j}) + \varepsilon, \quad i=1, \dots, L_k. \quad (3.6)$$

$$\ln(y_i) = a_0^* + \sum_{j=1}^p a_j^* \ln(x_{i,j}) + \varepsilon, \quad i=1, \dots, L_k.$$

nalizada descrita na forma difusa, assumida como  $\bar{Q}$ ,  $CV[Q]$ ,  $MP_1[Q]$ ,  $MP_2[Q]$  e  $MP_3[Q]$ ;  $x_j$ ,  $j = 1, \dots, p$ , representa as características fisiográficas e climatológicas da bacia e podem ser:  $x_F$ ,  $y_F$ ,  $z_F$ ,  $A$ ,  $\bar{P}$  e  $CV[P]$ ;  $a_j^*$ ,  $j = 0, \dots, p$ , são os parâmetros difusos a serem determinados na regressão difusa;  $\varepsilon$  é o erro esperado; e  $L_k$  é o número de estações do agrupamento considerado. O modelo de programação difusa é representado pelas expressões (2.67) a (2.70).

Os modelos de regressão ajustados foram avaliados por fim, pelo grau de ajuste médio  $\bar{h}$  obtido pela seguinte expressão:

$$\bar{h} = \frac{\sum_{i=1}^{L_k} \min[h_{i1}, h_{i2}]}{L_k}. \quad (3.7)$$

Os valores de  $h_{i1}$  e  $h_{i2}$  da expressão (3.7) correspondem aos níveis de

credibilidade otimizados na estimativa dos parâmetros da regressão (ver fig. 2.11). Foram analisados todos os possíveis casos de combinação entre parâmetros utilizados como variáveis independentes, sendo adotado o caso que apresentou o maior grau de ajuste médio  $\bar{h}$ .

### 3.2.5.2. Resultados

A análise de regressão difusa recai em um problema de programação por metas difusas de maximização de uma função objetivo sujeita a diversas restrições. O problema foi implementado com o auxílio da ferramenta de otimização do MATLAB através do “*Optimization Toolbox*” (MathWorks, 1997).

Foram realizados testes de sensibilidade com os dois maiores agrupamentos (Grupos 4 e 6) para definir qual valor do termo de incerteza  $e$ , expresso nas equações (2.38) e (2.39), representa melhor os valores previstos da variável dependente  $y$ . Este valor, em geral, é pré-definido pelo especialista. Nos testes de sensibilidade, sempre que possível, foram utilizados todos os parâmetros fisiográficos e climatológicos como variáveis independentes. Os valores de  $e$  considerados foram iguais a 30, 60 e 100% de  $y$ . Os ajustes foram analisados para todos os parâmetros a serem regionalizados, exceto para  $CV[Q]$ , cujos resultados demonstraram-se inconsistentes. O parâmetro para análise de decisão foi a soma das porcentagens dos erros  $SPE$ , obtido pela seguinte expressão:

$$SPE = \sum_{i=1}^{L_k} \left( \frac{|y_i - \hat{y}_i|}{\hat{y}_i} 100\% \right). \quad (3.8)$$

Na expressão (3.8)  $y_i$  significa a variável dependente observada;  $\hat{y}_i$  representa a variável dependente estimada; e  $L_k$  é o número de estações do agrupamento considerado. Através dos testes de sensibilidade, observou-se que a qualidade da estimativa dos parâmetros praticamente não é alterada pelos diferentes

valores de  $e$  utilizados. Para a obtenção das equações de regressão dos parâmetros o valor de  $e$  considerado é igual a 60% de  $y$ .

Não foi possível determinar as equações de regressão por metas difusas para  $\ln(\text{CV}[Q])$ , pois os resultados apresentaram-se inconsistentes (valores negativos ou demasiadamente superestimados).

As tabelas 3.4 a 3.7 apresentam os modelos de regressão obtidos através do método proposto para  $\ln(\bar{Q})$ ,  $\ln(\text{MP}_1[Q])$ ,  $\ln(\text{MP}_2[Q])$  e  $\ln(\text{MP}_3[Q])$ , respectivamente, que apresentaram o maior grau de ajuste médio  $\bar{h}$  expresso na equação (3.7). Na última linha destas tabelas é apresentado o grau de ajuste médio  $\bar{h}$ . Os melhores ajustes foram determinados para equações com termo independente, cujo grau de ajuste médio  $\bar{h}$  resultou igual a 1 na maioria dos ajustes. Porém estas equações não possuem bom comportamento na aplicação em locais que não foram utilizados na construção do modelo, estimando valores muito diferentes dos observados.

Pelos valores do grau de ajuste médio, observa-se que todas as equações de regressão que foram definidas tiveram um bom desempenho.

Em KAVISKI (1992) os parâmetros  $\bar{Q}$ ,  $\text{CV}[Q]$ ,  $\text{MP}_1[Q]$ ,  $\text{MP}_2[Q]$  e  $\text{MP}_3[Q]$  foram determinados por interpolação ponderada, onde é considerada a probabilidade de que o local sem dados pertença a um determinado grupo.

TABELA 3.4 – EQUAÇÕES DE REGRESSÃO PARA ESTIMATIVA DO VALOR LOGNORMALIZADO DE  $\bar{Q}$

Parâmetro	Coeficientes de regressão						
	Grupo 1	Grupo 2	Grupo 3	Grupo 4	Grupo 5	Grupo 6	Grupo 7
$A$	...	...	...	...	...	...	...
$x_F$	...	...	...	...	...	...	...
$y_F$	...	...	...	...	...	...	...
$z_F$	...	...	...	...	...	...	...
$\bar{P}$	1,382737	0,415806	0,410614	0,560754	0,450606	0,418548	0,291288
$\text{CV}[P]$	4,352804	0,166438	-0,034672	0,641649	0,099260	-0,076868	-0,457575
$\bar{h}$	0,8975	0,9191	0,9362	0,9495	0,9515	0,9483	0,8154

TABELA 3.5 – EQUAÇÕES DE REGRESSÃO PARA ESTIMATIVA DO VALOR LOGNORMALIZADO DE  $MP_1[Q]$ 

Parâmetro	Coeficientes de regressão						
	Grupo 1	Grupo 2	Grupo 3	Grupo 4	Grupo 5	Grupo 6	Grupo 7
$A$	0,059531	0,032944	-0,190526	0,000094	-0,085446	0,026604	-0,046872
$X_F$	...	...	...	...	...	...	...
$Y_F$	-2,120181	-0,839680	1,436730	-0,193749	0,691150	-0,642959	0,670079
$Z_F$	...	...	...	...	...	...	...
$\bar{P}$	0,945005	0,237092	-0,776316	-0,057171	-0,473780	0,137976	-0,473142
$CV[P]$	...	...	...	...	...	...	...
$\bar{h}$	0,8918	0,8697	0,7751	0,9349	0,9534	0,9591	0,9776

TABELA 3.6 – EQUAÇÕES DE REGRESSÃO PARA ESTIMATIVA DO VALOR LOGNORMALIZADO DE  $MP_2[Q]$ 

Parâmetro	Coeficientes de regressão						
	Grupo 1	Grupo 2	Grupo 3	Grupo 4	Grupo 5	Grupo 6	Grupo 7
$A$	0,072272	-0,079109	-0,248446	-0,003238	-0,127156	0,026985	-0,052838
$x_F$	...	...	...	...	...	...	...
$y_F$	-2,747671	-1,191433	2,065428	-0,414945	0,615400	-0,912008	0,686764
$z_F$	...	...	...	...	...	...	...
$\bar{P}$	1,194922	0,438386	-1,151989	-0,018313	-0,491602	0,188603	-0,553602
$CV[P]$	...	...	...	...	...	...	...
$\bar{h}$	0,9023	0,8775	0,8007	0,9339	0,9574	0,9608	0,9799

TABELA 3.7 – EQUAÇÕES DE REGRESSÃO PARA ESTIMATIVA DO VALOR LOGNORMALIZADO DE  $MP_3[Q]$ 

Parâmetro	Coeficientes de regressão						
	Grupo 1	Grupo 2	Grupo 3	Grupo 4	Grupo 5	Grupo 6	Grupo 7
$A$	0,079954	-0,156076	-0,275471	-0,007982	-0,153062	0,019100	-0,063332
$x_F$	...	...	...	...	...	...	...
$y_F$	-3,233353	-1,467178	2,166568	-0,617256	0,372047	-0,995517	0,726463
$z_F$	...	...	...	...	...	...	...
$\bar{P}$	1,397587	0,597113	-1,241952	0,039698	-0,403196	0,178125	-0,616237
$CV[P]$	...	...	...	...	...	...	...
$\bar{h}$	0,9067	0,8837	0,8157	0,9333	0,9585	0,9623	0,9769

### 3.2.6. Análise discriminante

O objetivo da análise discriminante é alocar uma dada bacia hidrográfica em um dos  $K$  grupos definidos na análise de agrupamentos. A alocação é realizada com

base nas características fisiográficas e/ou climatológicas (KAVISKI, 1992). Na análise de agrupamentos são definidas as  $n$  características fisiográficas e/ou climatológicas a serem utilizadas na análise discriminante.

Supõe-se que a cada grupo está associada a função densidade de probabilidade  $f_k(\underline{x})$  em  $R^n$  de que o vetor  $\underline{x}$ , representado por  $n$  características  $x$ ,  $\underline{x}=(x_1, \dots, x_n)$ , pertença ao grupo  $k$ .

Existem diversas regras discriminantes que podem ser utilizadas. Neste trabalho foi utilizada a mesma regra discriminante utilizada em KAVISKI (1992), ou seja, a regra discriminante de máxima verossimilhança para  $\underline{x}$ . Isto equivale a alocar  $\underline{x}$  no grupo  $k$ , tal que, seja mínimo o quadrado da distância de Mahalanobis.

Conforme citado no item (2.2.4), as probabilidades a priori são proporcionais ao número de estações de cada agrupamento,  $L_k$ . A função a ser minimizada é dada pela expressão (2.15). A probabilidade a posteriori de que  $\underline{x}$  pertença a um grupo qualquer  $j$  pode ser obtida por (2.16).

A análise discriminante é utilizada diretamente na aplicação de um estudo de caso. Cabe lembrar que, neste caso, o vetor  $\underline{x}$  é composto pelos valores das seguintes características:  $x_F, y_F, \bar{P}$  e  $CV[P]$  (item 4.2).

### 3.3. REGIONALIZAÇÃO DE PARÂMETROS DE SÉRIES DE VAZÕES MÁXIMAS ANUAIS

#### 3.3.1. Introdução

O objetivo desta aplicação foi regionalizar parâmetros de séries de vazões máximas anuais de forma direta e objetiva, com a utilização de poucos dados básicos de fisiografia, climatologia e fluviometria. A região estudada foi a mesma analisada na regionalização de parâmetros de séries de vazões médias diárias, ou seja, bacias de

pequena a média área de drenagem (entre 50 e 5000 km<sup>2</sup>), localizadas no Estado do Paraná. Os parâmetros regionalizados são: média das vazões máximas anuais  $\bar{Q}_{\max}$  e coeficiente de variação das vazões máximas anuais  $CV[Q_{\max}]$ . A importância destes parâmetros encontra-se na determinação de vazões de cheia associadas a determinados riscos em diversos tipos de obras hidráulicas e para o entendimento da ocorrência de inundações.

### 3.3.2. Descrição geral do método

A estrutura de regionalização para determinar  $\bar{Q}_{\max}$  e  $CV[Q_{\max}]$  seguiu a mesma seqüência de passos que no caso das vazões médias diárias (3.2.2).

Os parâmetros básicos utilizados na análise de agrupamentos foram: fisiográficos - latitude  $x_F$ , longitude  $y_F$ , altitude  $z_F$ , área de drenagem  $A$  e declividade média do talvegue principal  $s$ ; climatológicos - precipitação total média anual de longo período  $\bar{P}$ , coeficiente de variação das precipitações totais anuais  $CV[P]$ , média dos totais precipitados máximos anuais de um dia de duração  $\bar{P}_{\max 1}$  e coeficiente de variação dos totais precipitados máximos anuais de um dia de duração  $CV[P_{\max 1}]$ ; fluviais - vazão média diária de longo período  $\bar{Q}$  e coeficiente de variação das vazões médias diárias  $CV[Q]$ .

Foram determinados por regionalização (interpolação), em função da localização e altitude desta bacia, a precipitação total média anual de longo período, o coeficiente de variação das precipitações totais anuais e a média e o coeficiente de variação dos totais precipitados máximos anuais de um dia de duração. Através do método de regionalização implementado no item 3.2 foram determinados para o local sem informações hidrológicas, os valores da média e coeficiente de variação das vazões médias diárias. Desta maneira, foi necessário que o método de regionalização descrito no item 3.2 estivesse implementado para a mesma área de estudo, considerando as mesmas estações fluviométricas e pluviométricas.

### 3.3.3. Dados utilizados

Selecionou-se diversas estações fluviométricas localizadas dentro da região de estudo. Os dados coletados são referentes às séries de vazões: média das vazões médias diárias  $\bar{Q}$  (l/s/km<sup>2</sup>), coeficiente de variação das vazões médias diárias  $CV[Q]$ , média das vazões máximas anuais  $\bar{Q}_{\max}$  (l/s/km<sup>2</sup>) e coeficiente de variação das vazões máximas anuais  $CV[Q_{\max}]$ ; dados relacionados ao ponto de instalação da estação: latitude  $x_F$  (graus decimais), longitude  $y_F$  (graus decimais) e altitude  $z_F$  (m); e dados relacionados às bacias hidrográficas drenadas pelas estações: área de drenagem  $A$  (km<sup>2</sup>) e declividade média do talvegue principal  $s$  (m/km). As estações fluviométricas utilizadas neste trabalho são as mesmas do item 3.2 e as mesmas utilizadas em KAVISKI (1992). A tabela 3.1 apresenta estas estações e os parâmetros:  $x_F$ ,  $y_F$ ,  $z_F$ ,  $A$ ,  $\bar{Q}$ ,  $CV[Q]$ ,  $MP_1[Q]$ ,  $MP_2[Q]$ ,  $MP_3[Q]$ ,  $\bar{P}$  e  $CV[P]$ . O último ano de dados utilizado destas estações é 1982. Os parâmetros  $\bar{Q}_{\max}$ ,  $CV[Q_{\max}]$  e  $s$  de cada uma das estações fluviométricas foram obtidos de KAVISKI e FIOR, conforme citado em KAVISKI (1992, p. 180).

Foi selecionada uma rede de estações pluviométricas representativa da área de interesse e seu entorno. Os dados coletados foram: a precipitação total média anual de longo período  $\bar{P}$  (mm), o coeficiente de variação das precipitações totais anuais  $CV[P]$ , a média dos totais precipitados máximos anuais de um dia de duração  $\bar{P}_{\max 1}$ , e o coeficiente de variação dos totais precipitados máximos anuais de um dia de duração  $CV[P_{\max 1}]$ ; além da latitude  $x_P$  (graus decimais), longitude  $y_P$  (graus decimais) e altitude  $z_P$  (m) da estação. Os dados de  $\bar{P}$  e  $CV[P]$  regionalizados para os locais das estações fluviométricas foram obtidos diretamente do item 3.2, onde foi utilizado o grupo de estações pluviométricas listadas na tabela 3.2. Para a obtenção de  $\bar{P}_{\max 1}$  e  $CV[P_{\max 1}]$  foram selecionadas 69 estações pluviométricas localizadas sobre e no entorno da região estudada. Os dados latitude, longitude e altitude das estações foram obtidos de BRASIL (1981).

Os valores de  $\bar{P}_{\max 1}$  e  $CV[P_{\max 1}]$  de cada estação pluviométrica foram

obtidos de HOLTZ, citado por KAVISKI (1992, p. 180). O autor definiu para cada estação pluviométrica equações de regressão por mínimos quadrados ordinário (OLS – Ordinary Least Squares) expressas por:

$$h_{T,t} = \alpha T^\beta (t + \gamma)^\delta. \quad (3.9)$$

Na expressão (3.9)  $h_{T,t}$  é a altura da precipitação associada a  $T$  anos de recorrência ( $T=1, \dots, 100$ ) e  $t$  dias de duração ( $t=1, \dots, 6$ ),  $\alpha$ ,  $\beta$ ,  $\gamma$  e  $\delta$  são parâmetros estimados por OLS e  $\gamma$  é um parâmetro obtido por tentativas de ajuste em papel bi-log. As expressões que fornecem  $\bar{P}_{\max 1}$  e  $CV[P_{\max 1}]$  derivam da denominada distribuição de Pareto e são escritas como:

$$\bar{P}_{\max 1} = \alpha'(1 - \beta)^{-1}, \quad (3.10)$$

$$CV[P_{\max 1}] = \beta(1 - 2\beta)^{-1/2}. \quad (3.11)$$

Os valores de  $\bar{P}_{\max 1}$  e  $CV[P_{\max 1}]$  disponíveis para as estações pluviométricas, foram transferidos para os locais de instalação das estações pluviométricas utilizadas no estudo, através do método de regionalização implementado em KAVISKI (1992). A regionalização de  $\bar{P}_{\max 1}$  e  $CV[P_{\max 1}]$  foi efetuada de acordo com os seguintes critérios: para estações pluviométricas localizadas em regiões de interpolação, os valores foram obtidos por superfície “*spline*”; e para estações pluviométricas localizadas em regiões de extrapolação foi utilizada a interpolação ponderada. As regiões de interpolação são definidas através do limite inferior e superior da latitude para cada grau de longitude. Neste trabalho foram considerados os valores de  $\bar{P}_{\max 1}$  e  $CV[P_{\max 1}]$  regionalizados para as estações pluviométricas obtidos diretamente de KAVISKI (1992).

Na tabela 3.8 são apresentadas para cada estação pluviométrica utilizada no estudo, os valores de  $s$ ,  $\bar{Q}_{\max}$ ,  $CV[Q_{\max}]$ ,  $\bar{P}_{\max 1}$  e  $CV[P_{\max 1}]$ . A tabela 3.9 apresenta as estações pluviométricas utilizadas para a obtenção de  $\bar{P}_{\max 1}$  e  $CV[P_{\max 1}]$ . Constam desta tabela o código ANEEL, nome,  $x_P$ ,  $y_P$ ,  $z_P$ ,  $\bar{P}_{\max 1}$  e  $CV[P_{\max 1}]$ . A figura 3.3 apresenta a localização das estações pluviométricas da tabela 3.9.

TABELA 3.8 - ESTAÇÕES FLUVIOMÉTRICAS E PARÂMETROS REGIONALIZADOS  $s$ ,  $\bar{Q}_{\max}$ ,  $CV[Q_{\max}]$ ,  $\bar{P}_{\max 1}$ 

E $CV[P_{\max 1}]$							
Nº	Código	Nome	Declividade (m/km)	$\bar{Q}_{\max}$ (l/s/km²)	$CV[Q_{\max}]$	$\bar{P}_{\max 1}$ (mm)	$CV[P_{\max 1}]$
1	64248000	Salto do Itararé	3,068	71,790	0,48576	90,134	0,23911
2	64360000	Tomazina	5,556	119,461	0,52537	91,237	0,21771
3	64380000	Figueira	4,429	101,381	0,60616	100,870	0,22696
4	64390000	Ponte Santa Terezinha	2,164	121,946	0,55820	90,060	0,22000
5	64440000	Santa Cruz	1,846	89,117	0,48247	83,400	0,27000
6	64460000	Bom Jardim	1,429	135,929	0,56990	96,140	0,27000
7	64475000	Castro	3,571	104,038	0,58394	72,400	0,22976
8	64500000	Salto São Pedro	21,429	32,351	0,38997	94,251	0,22674
9	64547000	Santa Fé	4,681	73,722	0,50538	90,917	0,25966
10	64550000	Vila Silva Jardim	2,615	74,198	0,48036	91,800	0,24048
11	64620000	Rio dos Patos	3,182	183,331	0,43558	80,240	0,19000
12	64625000	Tereza Cristina	3,468	244,832	0,48587	92,505	0,23483
13	64660000	Ribeirão Raposo Tavares	3,778	272,747	0,48868	109,293	0,23460
14	64775000	Balsa do Cantú	3,655	290,298	0,38606	145,806	0,27613
15	64780000	Ponte Tourinho	5,000	101,667	0,37820	107,886	0,23038
16	64785000	Ponte do Goio-Bang	3,000	70,062	0,48930	148,118	0,28091
17	64790000	Salto Sapucaí	4,500	236,692	0,65149	101,115	0,24088
18	64810000	Balsa do Goio-Êre	3,333	59,119	0,37500	97,962	0,22817
19	64890000	Sta Helena	3,439	337,180	0,70417	111,070	0,22187
20	65010000	Fazendinha	5,789	148,600	0,47939	92,234	0,25482
21	65015000	Campina do Taquaral	2,759	70,252	0,47471	91,805	0,24409
22	65035000	Porto Amazonas	2,685	78,648	0,42659	89,224	0,21121
23	65085000	Bateias de Baixo	1,667	111,849	0,34722	92,910	0,27000
24	65090000	Fragosos	1,294	95,276	0,44600	91,017	0,25782
25	65095000	Rio Preto do Sul	0,902	88,188	0,32629	85,090	0,23407
26	65100000	Rio Negro	0,684	84,607	0,33882	74,106	0,20338
27	65130000	Salto Baraça	2,000	139,656	0,45281	83,076	0,24602
28	65135000	Rio da Várzea dos Limas	2,143	85,203	0,36450	89,460	0,24000
29	65155000	São Bento	0,966	55,661	0,37813	85,573	0,20004
30	65180000	Salto Canoinhas	1,000	135,187	0,60454	89,207	0,23295
31	65200000	Pontilhão	1,667	70,586	0,39152	83,879	0,25935
32	65365000	Porto Vitória	8,000	174,000	0,34147	100,330	0,21000
33	65370000	Jangada	4,271	192,815	0,40704	92,910	0,20000
34	65385000	Colônia Augusto Loureiro	4,310	214,289	0,34402	98,904	0,21213
35	65415000	Fazenda Maracanã	8,889	216,591	0,47510	94,540	0,25000
36	65770000	Balsa do Pinhalzinho	4,330	183,914	0,44444	113,365	0,19390
37	65810000	Guarapuava	4,213	160,241	0,47754	103,738	0,22918
38	65815000	Salto Curucaca	2,608	163,317	0,47464	94,496	0,23478
39	65825000	Santa Clara	2,500	143,368	0,48733	100,980	0,22000
40	65855000	Usina Cavernoso	4,444	201,166	0,50363	95,231	0,22849
41	65890000	Campo Novo	8,485	223,310	0,66491	145,340	0,22000
42	65925000	Salto Claudelino	3,373	213,823	0,30665	93,598	0,23130
43	65927000	Porto Palmeirinha	2,847	185,193	0,35949	103,261	0,23205
44	65945000	Ponte do Vitorino	3,846	198,158	0,43947	104,260	0,23038
45	65955000	Balsa do Santana	2,537	259,076	0,35013	106,860	0,23117
46	81060000	Salto São Luiz	4,545	25,640	0,41575	81,006	0,28307
47	81102000	Balsa do Jacaré	6,667	104,710	0,61771	82,490	0,19145
48	81125000	Turvo	14,222	100,513	0,48284	72,534	0,17484
49	81135000	Balsa do Cerro Azul	5,902	94,720	0,48156	72,739	0,22913
50	81140000	Cerro Azul	13,778	85,071	0,50930	69,860	0,19000
51	81300000	Praia Grande	9,583	111,543	0,31022	84,940	0,25000
52	81710000	Jacupiranga	12,600	192,663	0,30695	101,020	0,24807
53	82160000	Véu da Noiva	35,714	756,981	0,77104	144,060	0,27000
54	82230000	Vossoroca	15,385	126,828	0,26148	90,625	0,27512
55	82230800	Salto do Tanque	12,895	277,360	0,50037	92,020	0,27904

TABELA 3.9 - ESTAÇÕES PLUVIOMÉTRICAS E PARÂMETROS  $\bar{P}_{\max 1}$  E  $CV[P_{\max 1}]$ 

continua

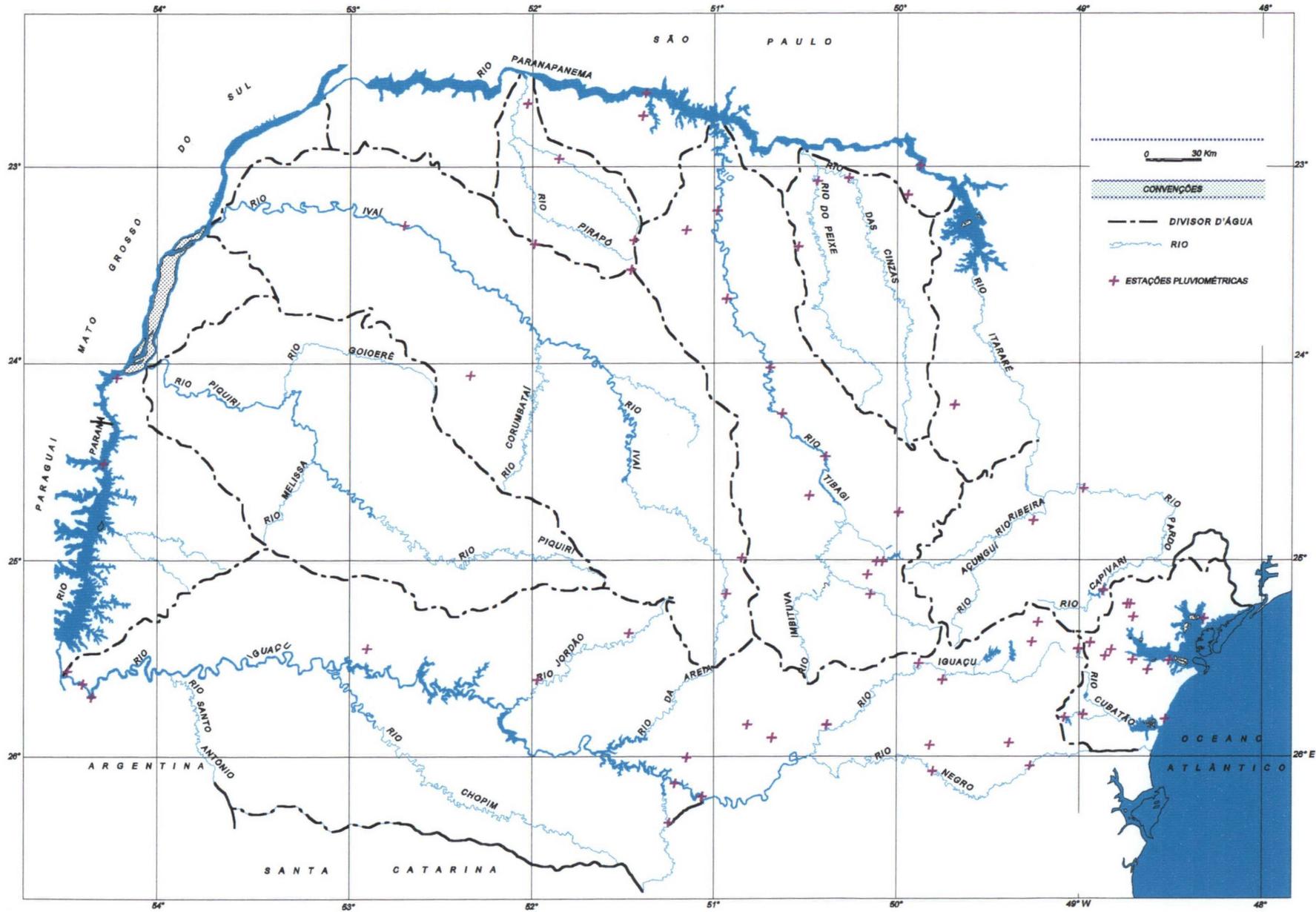
Nº	Código	Nome	Latitude		Longitude		Altitude (m)	$\bar{P}_{\max 1}$ (mm)	$CV[P_{\max 1}]$
			°	'	°	'			
1	02251021	Balsa do Paranapanema	22	39	51	22	250	92,75	0,26
2	02251024	Porecatu	22	46	51	23	350	89,61	0,26
3	02251026	Santa Fé	22	59	51	50	550	90,96	0,26
4	02252011	Fazenda Sasaki Irmãos	22	42	52	00	460	91,47	0,30
5	02349028	Jacarézinho	23	10	49	58	471	86,05	0,23
6	02349066	Ponte Melo Peixoto	23	01	49	54	390	70,83	0,31
7	02350000	Jataízinho	23	15	50	59	340	88,94	0,21
8	02350001	Porto Santa Terezinha	23	06	50	27	340	90,06	0,22
9	02350002	Andirá	23	05	50	17	375	87,62	0,22
10	02350008	Nona Fátima	23	26	50	33	680	88,53	0,22
11	02350015	Usina Apucarantina	23	42	50	56	450	96,78	0,23
12	02351009	Londrina	23	21	51	09	610	90,10	0,22
13	02351012	Arapongas	23	24	51	26	810	107,13	0,24
14	02351014	Maringá	23	25	51	58	555	92,52	0,21
15	02351061	Apucarana	23	33	51	27	870	103,14	0,24
16	02352000	Porto Paraíso do Norte	23	19	52	40	250	82,63	0,18
17	02449000	Capela da Ribeira	24	39	49	00	180	73,02	0,30
18	02449008	Cerro Azul	24	49	49	16	320	69,86	0,19
19	02449019	Jaguariaíva	24	14	49	42	891	90,05	0,24
20	02450002	Tibagi	24	30	50	24	750	79,33	0,16
21	02450005	Telêmaco Borba	24	17	50	38	650	84,74	0,20
22	02450006	Salto Mauá	24	03	50	42	550	106,79	0,25
23	02450009	Bom Jardim	24	42	50	29	750	96,14	0,27
24	02450010	Castro	24	47	50	00	996	71,84	0,23
25	02452006	Usina São João Auxiliar	24	05	52	19	630	98,28	0,25
26	02454000	Porto Mendes Gonçalves	24	30	54	20	150	129,53	0,18
27	02454001	Guaíra	24	04	54	15	218	121,37	0,23
28	02548000	Morretes	25	28	48	50	8	89,37	0,23
29	02548001	Praia Grande	25	10	48	53	750	84,94	0,25
30	02548002	Véu da Noiva	25	26	48	57	680	144,06	0,27
31	02548003	Colônia da Cachoeira	25	14	48	45	80	122,91	0,21
32	02548004	Matinhos	25	49	48	32	2	155,40	0,28
33	02548006	Guaraqueçaba	25	18	48	20	2	63,96	0,14
34	02548007	Mergulhão	25	18	48	43	9	128,86	0,25
35	02548010	Paranaguá	25	31	48	31	9	133,61	0,26
36	02548013	Usina Chaminé	25	48	48	59	800	112,36	0,35
37	02548016	Alexandra	25	34	48	38	10	115,72	0,33
38	02548017	Saquarema	25	31	48	43	5	96,71	0,27
39	02548018	Bairro Alto	25	14	48	44	120	107,45	0,28
40	02548027	Marumbi	25	30	48	52	60	122,87	0,40
41	02549000	São Bento	25	58	49	49	750	83,18	0,20
42	02549001	Porto Amazonas	25	33	49	53	793	86,33	0,21
43	02549003	Rio da Várzea dos Lima	25	57	49	23	810	89,46	0,24
44	02549006	Curitiba	25	26	49	16	947	84,25	0,20
45	02549008	Vossoroca	25	49	49	05	805	103,93	0,38
46	02549010	Rocha Nova	25	28	49	01	950	83,49	0,24
47	02549013	Engenheiro Bley	25	38	49	45	864	86,01	0,26
48	02549032	Colombo	25	20	49	14	960	71,95	0,24
49	02550000	Rio dos Patos	25	12	50	56	690	80,24	0,19
50	02550001	São Mateus do Sul	25	52	50	23	760	88,35	0,26

TABELA 3.9 - ESTAÇÕES PLUVIOMÉTRICAS E PARÂMETROS  $\bar{P}_{\max 1}$  E  $CV[P_{\max 1}]$ 

Nº	Código	Nome	Latitude		Longitude		Altitude (m)	$\bar{P}_{\max 1}$ (mm)	conclusão
			°	'	°	'			$CV[P_{\max 1}]$
51	02550002	Sumidouro	25	02	50	05	850	78,05	0,23
52	02550003	Santa Cruz	25	12	50	09	790	83,40	0,27
53	02550004	Braço do Potinga	25	52	50	49	760	110,94	0,20
54	02550005	Rio Claro do Sul	25	56	50	41	750	74,49	0,26
55	02550006	Ivaí	25	01	50	51	768	87,64	0,22
56	02550008	Ponta Grossa	25	06	50	10	869	97,79	0,23
57	02550011	Usina Pitangui	25	02	50	07	850	73,85	0,20
58	02551004	Santa Clara	25	38	51	58	740	100,98	0,22
59	02551006	Guarapuava	25	24	51	28	1108	85,98	0,22
60	02552000	Quedas do Iguaçu	25	28	52	54	550	145,34	0,22
61	02554001	Parque Nacional do Iguaçu	25	37	54	29	100	118,62	0,20
62	02554002	Salto Cataratas	25	41	54	26	152	137,03	0,32
63	02554004	Foz do Iguaçu	25	33	54	34	161	90,55	0,20
64	02649021	Rio Negro	26	06	49	48	790	70,48	0,20
65	02649011	Bateias de Baixo	26	04	49	16	800	92,91	0,27
66	02651000	União da Vitória	26	14	51	04	736	88,93	0,21
67	02651003	Jangada	26	22	51	15	800	92,91	0,20
68	02651004	Porto Vitória	26	10	51	13	790	100,33	0,21
69	02651005	Fazenda Maracanã	26	02	51	09	840	94,54	0,25

FONTE: ANEEL

FIGURA 3.3 - ESTAÇÕES PLUVIOMÉTRICAS UTILIZADAS NA REGIONALIZAÇÃO DE  $\bar{P}_{max1}$  E  $CV[P_{max1}]$



### 3.3.4. Definição das regiões homogêneas

#### 3.3.4.1. Desenvolvimento

O desenvolvimento da análise de agrupamentos é similar ao apresentado no item 3.2.4.1. As diferenças referem-se aos parâmetros regionalizados e classificadores.

A análise de agrupamentos dos parâmetros hidrológicos  $\bar{Q}_{\max}$  e  $CV[Q_{\max}]$  foi realizada em função dos parâmetros  $x_F, y_F, z_F, A, s, \bar{P}, CV[P], \bar{P}_{\max 1}, CV[P_{\max 1}], \bar{Q}$  e  $CV[Q]$ . Os melhores agrupamentos foram determinados através da implementação de todas as combinações possíveis envolvendo os parâmetros classificadores.

#### 3.3.4.2. Resultados

O melhor agrupamento dos parâmetros  $\bar{Q}_{\max}$  e  $CV[Q_{\max}]$  foi o gerado para 3 grupos em função dos parâmetros classificadores  $s, \bar{P}$  e  $\bar{P}_{\max 1}$ . Este agrupamento obteve o maior valor da média de  $V_B/V_W$  entre os agrupamentos que verificam a equação expressa em (2.9). Para este agrupamento a equação (2.10) também foi verificada. Os valores da relação  $V_B/V_W$  assumidos para os parâmetros hidrológicos  $\bar{Q}_{\max}$  e  $CV[Q_{\max}]$  são, respectivamente, iguais a: 2,231 e 0,692. Em KAVISKI (1992) os parâmetros classificadores são  $A, \bar{Q}$  e  $CV[Q]$ , o valor de  $K$  resultou igual a 4 e os valores de  $V_B/V_W$  determinados para os parâmetros  $\bar{Q}_{\max}$  e  $CV[Q_{\max}]$  são respectivamente iguais a: 1,710 e 0,754.

Os valores de  $r$ , parâmetro que especifica o nível de incerteza e é utilizado na função expressa em (2.48), são recomendados entre de 1,25 a 2,0 (ROSS, citado por HALL & MINNS, 1999, p. 698). O valor de  $r$  foi adotado igual ao máximo indicado ( $r=2,0$ ), indicando assim um alto nível de incerteza a respeito da ambigüidade dos dados.

O coeficiente de partição difuso  $F_c$  (equação 2.51) resultou igual a 0,72. Este resultado revela que a classificação foi bem sucedida, pois está bem acima de 0,33 ( $1/K$ ).

A tabela 3.10 apresenta a relação de estações fluviométricas, identificadas através de seus códigos ANEEL, classificadas em cada um dos 3 grupos.

TABELA 3.10 – AGRUPAMENTOS DETERMINADOS NA CLASSIFICAÇÃO DIFUSA

GRUPOS		
Grupo 1	Grupo 2	Grupo 3
64248000	64500000	64775000
64360000	64660000	64785000
64380000	64780000	65890000
64390000	64790000	82160000
64440000	64810000	
64460000	64890000	
64475000	65010000	
64547000	65365000	
64550000	65385000	
64620000	65415000	
64625000	65770000	
65015000	65810000	
65035000	65815000	
65085000	65825000	
65090000	65855000	
65095000	65925000	
65100000	65927000	
65130000	65945000	
65135000	65955000	
65155000	81710000	
65180000	82230000	
65200000	82230800	
65370000		
81060000		
81102000		
81125000		
81135000		
81140000		
81300000		

### 3.3.5. Determinação das equações de regressão que estimam $\bar{Q}_{\max}$ e $CV[Q_{\max}]$

#### 3.3.5.1. Desenvolvimento

O modelo de regressão difusa utilizado para estimar os parâmetros segue a

metodologia apresentada no item 3.2.5.1 (séries de vazões médias diárias).

O modelo de regressão difusa pode ser generalizado conforme a expressão (3.6). Nesta expressão  $y$  é a variável regionalizada descrita no forma difusa, assumida como  $\bar{Q}_{\max}$  e  $CV[Q_{\max}]$ ;  $x_j$ ,  $j = 1, \dots, p$ , representa as características fisiográficas, climatológicas e fluviais da bacia e podem ser:  $x_F$ ,  $y_F$ ,  $z_F$ ,  $A$ ,  $s$ ,  $\bar{P}$ ,  $CV[P]$ ,  $\bar{P}_{\max 1}$ ,  $CV[P_{\max 1}]$ ,  $\bar{Q}$  e  $CV[Q]$ ;  $a_j^*$ ,  $j = 0, \dots, p$ , são os parâmetros difusos a serem determinados na regressão difusa; e  $L_k$  é o número de estações do agrupamento considerado.

### 3.3.5.2. Resultados

A análise de regressão difusa recai em um problema de programação por metas difusas de maximização de uma função objetivo sujeita a diversas restrições. O problema foi implementado com o auxílio da ferramenta de otimização do MATLAB através do “*Optimization Toolbox*” (MathWorks, 1997).

Para definir qual valor de do termo de incerteza  $e$ , indicado nas expressões (2.38) e (2.39), representa melhor os valores previstos de  $y$ , foram realizados testes de sensibilidade com os dois maiores agrupamentos (Grupos 1 e 2). Os valores de  $e$  considerados foram iguais a 30, 60 e 100% de  $y$ . Os ajustes foram analisados para todos os parâmetros a serem regionalizados. O parâmetro para análise de decisão foi a soma das porcentagens dos erros  $SPE$ , obtido por (3.8).

Para a obtenção das equações de regressão dos parâmetros  $\bar{Q}_{\max}$  e  $CV[Q_{\max}]$ , o valor de  $e$  considerado é igual a 60% de  $y$ , pois este valor gerou os menores resultados de  $SPE$  para os Grupo 1 e 2.

As tabelas 3.11 e 3.12 apresentam os modelos de regressão obtidos através do método proposto para  $\ln(\bar{Q}_{\max})$  e  $\ln(CV[Q_{\max}])$ , respectivamente, que apresentaram o maior grau de ajuste médio  $\bar{h}$ , obtido pela equação (3.7). Na última linha destas tabelas é apresentado o grau de ajuste médio  $\bar{h}$ .

Tal como revelam os valores do grau de ajuste médio, observa-se que as equações de regressão definidas para  $\ln(\bar{Q}_{\max})$  tiveram um bom desempenho para todos os Grupos. Para a estimativa de  $\ln(\text{CV}[Q_{\max}])$  dos Grupos 2 e 3 a equação de regressão também foi bem definida. Para o Grupo 1 o desempenho da estimativa de  $\ln(\text{CV}[Q_{\max}])$  foi razoável.

TABELA 3.11 – EQUAÇÕES DE REGRESSÃO PARA ESTIMATIVA DE  $\ln(\bar{Q}_{\max})$

Parâmetro	Coeficientes de regressão		
	Grupo 1	Grupo 2	Grupo 3
$x_F$	0,658119	...	2,348600
$y_F$	...	...	...
$z_F$	...	...	...
$A$	...	...	...
$s$	...	0,173615	...
$\bar{Q}$	...	1,551947	...
$\text{CV}[Q]$	...	...	...
$\bar{P}$	...	...	...
$\text{CV}[P]$	...	-1,389156	...
$\bar{P}_{\max 1}$	0,548671	-0,487208	...
$\text{CV}[P_{\max 1}]$	...	0,158054	1,449501
$\bar{h}$	0,8906	0,9252	0,8012

NOTA: Sinais convencionais utilizados:

... Parâmetro não utilizado na análise de regressão.

TABELA 3.12 – EQUAÇÕES DE REGRESSÃO PARA ESTIMATIVA DE  $\ln(\text{CV}[Q_{\max}])$

Parâmetro	Coeficientes de regressão		
	Grupo 1	Grupo 2	Grupo 3
$x_F$	-0,562072	-0,2010125	...
$y_F$	...	3,1806229	...
$z_F$	...	...	0,540900
$A$	...	-0,0954916	...
$s$	...	-0,2081925	...
$\bar{Q}$	...	-0,313342	...
$\text{CV}[Q]$	...	0,7759554	...
$\bar{P}$	...	-0,3685427	...
$\text{CV}[P]$	...	0,8308353	...
$\bar{P}_{\max 1}$	0,239364	-1,3728395	-0,762994
$\text{CV}[P_{\max 1}]$	...	0,1961126	-0,003393
$\bar{h}$	0,6942	0,7963	0,8623

NOTA: Sinais convencionais utilizados:

... Parâmetro não utilizado na análise de regressão.

Os melhores ajustes foram determinados para equações com termo independente, cujo grau de ajuste médio  $\bar{h}$  resultou igual a 1 na maioria dos ajustes. Porém estas equações não possuem fornecem bons resultados na aplicação em locais que não foram utilizados na construção do modelo, estimando valores muito diferentes dos observados.

Em KAVISKI (1992) os parâmetros  $\ln(\bar{Q}_{\max})$  e  $\ln(\text{CV}[Q_{\max}])$  foram transferidos através de equações de regressão lineares determinadas por WLS (mínimos quadrados ponderado) para cada agrupamento. Na definição dos parâmetros todas as variáveis envolvidas foram lognormalizadas. Foi suposto que  $\bar{Q}_{\max}$  e  $\text{CV}[Q_{\max}]$  seguem o modelo log-normal a 2 parâmetros.

A análise discriminante foi utilizada na aplicação de um estudo de caso, através dos parâmetros classificadores  $s$ ,  $\bar{P}$  e  $\bar{P}_{\max 1}$  (item 4.3).

## 4. ANÁLISE DOS RESULTADOS

### 4.1. GERAL

O método de regionalização desenvolvido nesta dissertação foi aplicado em alguns locais, a fim de efetuar a análise dos resultados dos estudos de caso. Com a análise dos resultados foi possível verificar a aplicabilidade do método através da comparação com os dados observados e verificar se o método proposto fornece melhorias nos resultados através de comparações com o método apresentado em KAVISKI (1992).

Para possibilitar a análise dos resultados, os locais estudados são estações fluviométricas que possuem dados observados, porém não foram utilizadas no desenvolvimento dos estudos de casos.

Os índices estatísticos utilizados para avaliar os métodos que foram comparados são o desvio médio quadrático  $\bar{d}_q$ , e absoluto  $\bar{d}_a$ , e são aplicados em cada parâmetro regionalizado  $p$ :

$$\bar{d}_q(p) = \left\{ \sum_{i=1}^n [(y_i - \hat{y}_i) / y_i]^2 / n \right\}^{1/2}, \quad (4.1)$$

$$\bar{d}_a(p) = \left[ \sum_{i=1}^n (|y_i - \hat{y}_i| / y_i) \right] / n. \quad (4.2)$$

Nas expressões acima  $y_i$  corresponde ao valor do parâmetro  $p$  na estação  $i$  obtido através dos dados históricos observados em  $i$ ;  $\hat{y}_i$  corresponde ao valor do parâmetro  $p$  estimado pelo método de regionalização proposto nesta dissertação ou em KAVISKI (1992); e  $n$  é o número de locais analisados.

## 4.2. REGIONALIZAÇÃO DE PARÂMETROS DE SÉRIES DE VAZÕES MÉDIAS DIÁRIAS

A aplicação do método de regionalização proposto para a obtenção de parâmetros de séries de vazões médias diárias foi efetuada nos locais das estações fluviométricas apresentadas na tabela 4.1. A figura 4.1 apresenta a localização destas estações dentro do Estado do Paraná. Todas estas estações drenam áreas menores que 5000 km<sup>2</sup>. O conjunto de estações selecionadas apresenta distintos tamanhos de séries e qualidade. Na tabela 4.1 são apresentados dados obtidos diretamente de BRASIL (1987) - código, área de drenagem, coordenadas, altitude e período de observação considerado – e dados obtidos de KAVISKI (1992) – estimativas da precipitação total média anual de longo período e coeficiente de variação das precipitações totais anuais.

TABELA 4.1 – ESTAÇÕES FLUVIOMÉTRICAS ANALISADAS NA APLICAÇÃO DO ESTUDO DE CASO

Código	Nome	Período		Latitude		Longit.		Altitude (m)	Área (km <sup>2</sup> )	$\bar{P}$ (mm)	CV[P]
		Início	Fim	°	'	°	'				
64659000	Barbosa Ferraz	1975	1984	24	01	51	57	320	3294	1499	0,21128
64673000	Quinta do Sol	1975	1984	23	49	52	11	315	1534	1475	0,18989
64680000	Jussara	1977	1984	23	37	52	28	275	727	1404	0,19128
64682000	Japurá	1977	1984	23	26	52	35	285	807	1370	0,19134
64771500	Porto Guarani	1976	1990	24	51	52	45	360	4223	1769	0,19228
64815000	Fazenda Uberaba	1979	1984	24	07	53	19	310	2941	1553	0,19563
65190000	Balsa do Potinga	1942	1953	25	52	50	49	760	222	1492	0,14000
65295000	Sta. Cruz do Timbó	1974	1975	26	23	50	52	750	2614	1537	0,16405
65690000	Leonópolis	1958	1967	25	41	51	12	960	358	1782	0,14000

Primeiramente, foi efetuada a análise discriminante sobre as estações da tabela 4.1. Os parâmetros classificadores correspondem às seguintes características:  $x_F$ ,  $y_F$ ,  $\bar{P}$  e CV[P].

Para as nove estações da tabela 4.1 foram estimados, pelo método proposto os parâmetros: média das vazões médias diárias  $\bar{Q}$  e momentos de pesos probabilísticos de ordem 1 a 3 das vazões médias diárias uniformes,  $MP_1[Q]$ ,  $MP_2[Q]$  e  $MP_3[Q]$ . Como já explanado no item 3.2.5.2, não foi possível determinar as equações

de regressão por metas difusas para o coeficiente de variação das vazões médias diárias  $CV[Q]$ . A determinação deste parâmetro foi efetuada pelo método utilizado em KAVISKI (1992), ou seja, por interpolação ponderada, onde é considerada a probabilidade de que o local sem dados pertença a um determinado grupo. Nas tabelas 4.2, 4.3 e 4.4 são apresentados os parâmetros estimados pelos dados amostrais, pelo método de regionalização proposto nesta dissertação e pelo método apresentado em KAVISKI (1992), respectivamente. A tabela 4.3 apresenta também a qual dos sete agrupamentos pertence a estação analisada. O conjunto de estações que deu origem a cada uma das regiões homogêneas pode ser verificado na tabela 3.3 do item 3.2.4.2.

Para a maioria das estações não foi possível efetuar as comparações com dados amostrais em termos dos parâmetros  $CV[Q]$ ,  $MP_1[Q]$ ,  $MP_2[Q]$  e  $MP_3[Q]$ , pois estes dados não estão disponibilizados. Apenas uma única estação possui as estimativas amostrais destes parâmetros. A análise dos resultados considerou apenas o parâmetro  $\bar{Q}$ .

Para o método proposto nesta dissertação os valores de  $\bar{d}_q$  e  $\bar{d}_a$  resultaram iguais a 0,127 e 0,118, respectivamente. Para o método apresentado em KAVISKI (1992) os valores de  $\bar{d}_q$  e  $\bar{d}_a$  resultaram iguais a 0,168 e 0,160, respectivamente. Considerando estes dois índices, o método proposto nesta dissertação apresenta um bom grau de melhoria na regionalização do parâmetro  $\bar{Q}$ .

Em uma análise complementar efetuou-se a estimativa do parâmetro  $\bar{Q}$ , através da aplicação da classificação difusa e da interpolação ponderada, onde é considerada a probabilidade de que o local sem dados pertença a um determinado grupo. Esta última técnica foi utilizada em KAVISKI (1992) para as estimativas dos parâmetros  $\bar{Q}$ ,  $CV[Q]$ ,  $MP_1[Q]$ ,  $MP_2[Q]$  e  $MP_3[Q]$ . O objetivo desta análise foi verificar se houve alguma melhoria considerando somente a aplicação da classificação difusa. Os valores de  $\bar{Q}$  obtidos por este método estão apresentados na tabela 4.5.



TABELA 4.2 – PARÂMETROS ESTIMADOS PELOS DADOS AMOSTRAIS

Código	$\bar{Q}$ (l/s/km <sup>2</sup> )	CV[Q]	MP <sub>1</sub> [Q]	MP <sub>2</sub> [Q]	MP <sub>3</sub> [Q]
64659000	24,27	...	...	...	...
64673000	24,95	...	...	...	...
64680000	19,81	...	...	...	...
64682000	22,45	...	...	...	...
64771500	27,24	1,62	0,2003	0,0894	0,0527
64815000	22,45	...	...	...	...
65190000	26,80	...	...	...	...
65295000	27,07	...	...	...	...
65690000	27,32	...	...	...	...

FONTE: ANEEL (boletins do extinto DNAEE)

NOTA: Sinais convencionais utilizados:

... Dado não disponível.

TABELA 4.3 – PARÂMETROS ESTIMADOS PELO MÉTODO PROPOSTO

Código	Classe	$\bar{Q}$ (l/s/km <sup>2</sup> )	CV[Q]	MP <sub>1</sub> [Q]	MP <sub>2</sub> [Q]	MP <sub>3</sub> [Q]
64659000	3	21,25	1,17	0,2134	0,1026	0,0636
64673000	3	21,19	1,13	0,2516	0,1275	0,0809
64680000	3	20,76	1,13	0,3037	0,1643	0,1069
64682000	3	20,55	1,12	0,3044	0,1654	0,1076
64771500	3	22,82	1,18	0,1830	0,0822	0,0500
64815000	3	21,62	1,20	0,2202	0,1068	0,0665
65190000	5	22,15	1,09	0,2985	0,1553	0,0815
65295000	5	22,80	1,04	0,2386	0,1119	0,0561
65690000	5	24,00	1,11	0,2648	0,1346	0,0707

TABELA 4.4 – PARÂMETROS ESTIMADOS PELO MÉTODO UTILIZADO PARA COMPARAÇÃO

Código	$\bar{Q}$ (l/s/km <sup>2</sup> )	CV[Q]	MP <sub>1</sub> [Q]	MP <sub>2</sub> [Q]	MP <sub>3</sub> [Q]
64659000	21,29	1,67	0,1907	0,0821	0,0474
64673000	19,38	1,32	0,2461	0,1230	0,0783
64680000	18,19	1,07	0,2859	0,1523	0,1005
64682000	17,32	0,98	0,2998	0,1624	0,1079
64771500	22,06	1,32	0,2475	0,1244	0,0796
64815000	19,92	0,86	0,3195	0,1777	0,1199
65190000	21,34	1,05	0,2667	0,1334	0,0843
65295000	23,93	1,10	0,2671	0,1351	0,0859
65690000	22,84	1,11	0,2657	0,1342	0,0855

FONTE: KAVISKI (1992).

Utilizando a classificação difusa e interpolação ponderada na estimativa de  $\bar{Q}$ , os valores de  $\bar{d}_q$  e  $\bar{d}_a$  resultaram iguais a 0,154 e 0,140, respectivamente. Para o método apresentado em KAVISKI (1992) os valores de  $\bar{d}_q$  e  $\bar{d}_a$  resultaram iguais a 0,168 e 0,160, respectivamente. Considerando estes dois índices, pode-se concluir que mesmo com a utilização da técnica de lógica difusa apenas na classificação há um bom grau de melhoria na regionalização do parâmetro  $\bar{Q}$ .

TABELA 4.5 – ESTIMATIVA DE  $\bar{Q}$  ATRAVÉS DA CLASSIFICAÇÃO DIFUSA E INTERPOLAÇÃO PONDERADA

Código	$\bar{Q}$ (l/s/km <sup>2</sup> )	Código	$\bar{Q}$ (l/s/km <sup>2</sup> )	Código	$\bar{Q}$ (l/s/km <sup>2</sup> )
64659000	20,61	64682000	18,18	65190000	22,69
64673000	19,21	64771500	22,98	65295000	22,22
64680000	18,86	64815000	22,05	65690000	23,61

### 4.3. REGIONALIZAÇÃO DE PARÂMETROS DE SÉRIES DE VAZÕES MÁXIMAS ANUAIS

O método de regionalização proposto foi utilizado para estimar os parâmetros de séries de vazões máximas anuais nos locais das estações fluviométricas apresentadas na tabela 4.6. A figura 4.2 apresenta a localização destas estações dentro da região de estudo. As áreas drenadas por estas estações são menores que 5000 km<sup>2</sup>. O conjunto de estações selecionadas tem uma boa representatividade quanto ao tamanho e qualidade das séries. Na tabela 4.6 são apresentados os dados obtidos diretamente de BRASIL (1987): código, área de drenagem, coordenadas, altitude e período de observação considerado. Na tabela 4.7 são apresentados dados obtidos de KAVISKI (1992): estimativas da declividade média da bacia, precipitação total média anual de longo período  $\bar{P}$ , coeficiente de variação das precipitações totais anuais  $CV[P]$ , média dos totais precipitados máximos anuais de um dia de duração  $\bar{P}_{\max 1}$  e

coeficiente de variação dos totais precipitados máximos anuais de um dia de duração  $CV[P_{\max 1}]$ ; e os dados estimados pelo método proposto no item 4.2: vazão média diária de longo período  $\bar{Q}$  e coeficiente de variação das vazões médias diárias  $CV[Q]$ .

TABELA 4.6 – ESTAÇÕES FLUVIOMÉTRICAS ANALISADAS NA APLICAÇÃO DO ESTUDO DE CASO

Código	Nome	Período		Latitude		Longit.		Altitude (m)	Área (km <sup>2</sup> )
		Início	Fim	°	'	°	'		
64659000	Barbosa Ferraz	1975	1984	24	01	51	57	320	3294
64673000	Quinta do Sol	1975	1984	23	49	52	11	315	1534
64680000	Jussara	1977	1984	23	37	52	28	275	727
64682000	Japurá	1977	1984	23	26	52	35	285	807
64771500	Porto Guarani	1976	1990	24	51	52	45	360	4223
64815000	Fazenda Uberaba	1979	1984	24	07	53	19	310	2941
65190000	Balsa do Potinga	1942	1953	25	52	50	49	760	222
65295000	Sta. Cruz do Timbó	1974	1975	26	23	50	52	750	2614
65690000	Leonópolis	1958	1967	25	41	51	12	960	358
81100000	Passo do Assungui	1936	1945	25	01	49	29	390	1573
82111000	Mergulhão	1945	1966	25	19	48	43	9	344
82121000	Limoeiro	1930	1952	25	15	48	46	8	350
82170000	Morretes	1938	1982	25	29	48	50	3	208

TABELA 4.7 – PARÂMETROS DE ENTRADA NO MÉTODO DE REGIONALIZAÇÃO PROPOSTO

Código	Declividade (m/km)	$\bar{P}$ (mm)	$CV[P]$	$\bar{P}_{\max 1}$ (mm)	$CV[P_{\max 1}]$	$\bar{Q}$ (l/s/km <sup>2</sup> )	$CV[Q]$
64659000	3,82	1499	0,21128	108,577	0,23202	21,2538	1,16960
64673000	3,71	1475	0,18989	102,380	0,21693	21,1917	1,13170
64680000	5,46	1404	0,19128	95,835	0,19928	20,7615	1,12800
64682000	4,31	1370	0,19134	87,501	0,19146	20,5534	1,11860
64771500	2,68	1769	0,19228	150,412	0,28058	22,8239	1,17640
64815000	2,80	1553	0,19563	129,924	0,26289	21,6226	1,19750
65190000	4,83	1492	0,14000	110,940	0,20000	22,1494	1,09160
65295000	3,84	1537	0,16405	91,716	0,22162	22,8040	1,03960
65690000	7,14	1782	0,14000	101,492	0,24355	23,9951	1,11360
81100000	9,65	1720	0,16839	86,909	0,20520	20,7923	0,99300
82111000	30,15	2226	0,09970	127,394	0,25045	25,6463	1,10840
82121000	50,26	2289	0,08649	127,499	0,23488	25,6070	1,11580
82170000	37,34	1746	0,22103	85,852	0,22665	17,3361	0,97320

O primeiro passo foi efetuar a análise discriminante sobre as estações da tabela 4.6, através dos parâmetros classificadores correspondem às seguintes características:  $s$ ,  $\bar{P}$  e  $\bar{P}_{\max 1}$ .

Para as estações selecionadas foram estimados, pelo método proposto, os

parâmetros média das vazões máximas anuais  $\bar{Q}_{\max}$ , e coeficiente de variação das vazões máximas anuais  $CV[Q_{\max}]$ . Nas tabelas 4.8, 4.9 e 4.10 são apresentados os parâmetros estimados pelos dados amostrais, pelo método de regionalização proposto nesta dissertação e pelo método apresentado em KAVISKI (1992), respectivamente. Na tabela 4.9 é indicado a qual dos agrupamentos pertence a estação analisada. O conjunto de estações que deu origem a cada uma das regiões homogêneas pode ser verificado na tabela 3.12 do item 3.3.4.2.

Para a estimativa de  $\bar{Q}_{\max}$ , pelo método proposto nesta dissertação, os valores de  $\bar{d}_q$  e  $\bar{d}_a$  resultaram iguais a 0,372 e 0,278, respectivamente. Para o método apresentado em KAVISKI (1992) os valores de  $\bar{d}_q$  e  $\bar{d}_a$  resultaram iguais a 0,400 e 0,355, respectivamente. Considerando estes dois índices, o método proposto nesta dissertação apresenta um bom grau de melhoria na regionalização do parâmetro  $\bar{Q}_{\max}$ .

Considerando a estimativa de  $CV[Q_{\max}]$ , os valores de  $\bar{d}_q$  e  $\bar{d}_a$  resultaram pelo método proposto iguais a 0,334 e 0,252, respectivamente. Para o método apresentado em KAVISKI (1992) os valores de  $\bar{d}_q$  e  $\bar{d}_a$  resultaram iguais a 1,290 e 0,794, respectivamente. Por estes resultados, pode-se constatar que o método proposto nesta dissertação apresenta uma melhora significativa também na regionalização do parâmetro  $CV[Q_{\max}]$ .

TABELA 4.8 – PARÂMETROS ESTIMADOS PELOS DADOS AMOSTRAIS

Código	$\bar{Q}_{\max}$ (l/s/km <sup>2</sup> )	$CV[Q_{\max}]$	Código	$\bar{Q}_{\max}$ (l/s/km <sup>2</sup> )	$CV[Q_{\max}]$
64659000	422,83	0,4184	65295000	169,47	...
64673000	159,06	0,4837	65690000	322,31	0,3271
64680000	89,68	0,2734	81100000	75,29	0,5360
64682000	78,66	0,5824	82111000	262,66	0,0880
64771500	353,31	0,3580	82121000	229,93	0,0640
64815000	87,33	0,3436	82170000	605,42	0,3430
65190000	214,98	0,3295			

FONTE: ANEEL (boletins do extinto DNAEE)

NOTA: Sinais convencionais utilizados:

... Dado não disponível.

TABELA 4.9 – PARÂMETROS ESTIMADOS PELO MÉTODO PROPOSTO

Código	Grupo	$\bar{Q}_{\max}$ (l/s/km <sup>2</sup> )	CV[ $Q_{\max}$ ]	Código	Grupo	$\bar{Q}_{\max}$ (l/s/km <sup>2</sup> )	CV[ $Q_{\max}$ ]
64659000	2	101,61	0,5109	65295000	1	102,84	0,4687
64673000	1	102,12	0,5096	65690000	2	252,20	0,3573
64680000	1	97,94	0,5040	81100000	1	96,41	0,4767
64682000	1	92,69	0,4953	82111000	3	265,76	0,0817
64771500	3	299,93	0,5288	82121000	3	240,65	0,0766
64815000	2	102,83	0,4492	82170000	1	96,93	0,4704
65190000	2	193,19	0,3630				

TABELA 4.10 – PARÂMETROS ESTIMADOS PELO MÉTODO UTILIZADO PARA COMPARAÇÃO

Código	$\bar{Q}_{\max}$ (l/s/km <sup>2</sup> )	CV[ $Q_{\max}$ ]	Código	$\bar{Q}_{\max}$ (l/s/km <sup>2</sup> )	CV[ $Q_{\max}$ ]
64659000	310,16	0,4969	65295000	170,73	0,3217
64673000	177,61	0,2904	65690000	153,23	0,4883
64680000	133,97	0,6452	81100000	71,20	0,4883
64682000	108,07	0,6815	82111000	123,62	0,2885
64771500	197,60	0,3823	82121000	90,45	0,2848
64815000	128,92	0,3861	82170000	378,36	0,2157
65190000	139,26	0,4981			

FONTE: KAVISKI (1992).



## 5. CONCLUSÕES E RECOMENDAÇÕES

### 5.1. GERAL

Neste capítulo apresenta-se as conclusões e recomendações obtidas ao longo da pesquisa e no desenvolvimento do método proposto. O objetivo geral desta pesquisa de dissertação consta em analisar a aplicação da análise difusa em uma modelagem em recursos hídricos. A área escolhida para investigação foi regionalização de parâmetros hidrológicos em bacias hidrográficas de pequeno a médio porte (50 a 5000 km<sup>2</sup>), e as técnicas de lógica difusa selecionadas foram classificação difusa e regressão difusa. Os parâmetros hidrológicos regionalizados são de séries de vazões médias diárias (média, coeficiente de variação e momentos de pesos probabilísticos); e de vazões máximas anuais (média e coeficiente de variação). As conclusões do trabalho são baseadas na aplicação de índices de desempenho para análise dos resultados e comparação dos resultados desta dissertação com os resultados obtidos pelo método de regionalização proposto em KAVISKI (1992).

Através do estudo de caso foi possível atingir o objetivo desta dissertação, ou seja, verificar que a aplicação de ferramentas da lógica difusa em uma técnica de modelagem de recursos hídricos, no caso, regionalização de parâmetros hidrológicos, fornece resultados melhores que os obtidos com a utilização de métodos tradicionais.

### 5.2. PRINCIPAS CONCLUSÕES

O método proposto apresentou bastante simplicidade na sua conceituação matemática e desenvolvimento, confirmando uma das principais vantagens citadas da

análise difusa. Esta característica foi verificada em ambas as técnicas de lógica difusa utilizadas nesta dissertação. Outra vantagem comum aos sistemas difusos é a utilização da experiência do especialista no desenvolvimento do método. Esta característica foi também verificada em ambas as técnicas de lógica difusa implementadas.

Na classificação difusa o especialista pode intervir no parâmetro que controla o nível de incerteza  $r$  (“*fuzziness*”) no processo de particionamento, tomando como base o intervalo de variações nos parâmetros a serem classificados. Pode-se atribuir um valor menor para  $r$ , caso ocorram na região estudada regiões com características fisiográficas, meteorológicas e fluviais bastante diferenciadas entre si e/ou os parâmetros fluviais possuam forte dependência dos parâmetros fisiográficos e meteorológicos. Desta maneira, os agrupamentos formados possuem um maior grau de certeza, fato este já conhecido previamente.

No caso da regressão difusa, os intervalos de tolerância são, entre os parâmetros afetados pela inferência do especialista, os de maior sensibilidade. Porém, verificou-se no estudo de caso, que, para os parâmetros hidrológicos relacionados às séries de vazões médias diárias, as variações nos intervalos de tolerância não interferem na definição das equações.

A classificação difusa apresentou melhores resultados que a análise de agrupamentos pelo método “*K-means*” tanto no desenvolvimento do método proposto como na aplicação do estudo de caso. Os valores da relação de variâncias  $V_B/V_W$  assumidos para os parâmetros hidrológicos  $\bar{Q}$ ,  $CV[Q]$ ,  $MP_1[Q]$ ,  $MP_2[Q]$ ,  $MP_3[Q]$ ,  $\bar{Q}_{\max}$  e  $CV[Q_{\max}]$  para o método proposto são, respectivamente, iguais a: 1.555, 0.454, 0.801, 0.935, 1.025, 2.231 e 0.692. Em KAVISKI (1992) os valores de  $V_B/V_W$  determinados para os parâmetros  $\bar{Q}$ ,  $CV[Q]$ ,  $MP_1[Q]$ ,  $MP_2[Q]$ ,  $MP_3[Q]$ ,  $\bar{Q}_{\max}$  e  $CV[Q_{\max}]$  são respectivamente iguais a: 1.260, 0.076, 0.152, 0.188 e 0.215, 1.710 e 0.754. Pode-se observar pela comparação dos valores que há um significativo aumento da relação de variâncias para os parâmetros classificados pelo método “*c-means*”, exceto para  $CV[Q_{\max}]$ .

Para verificar se haveria, efetivamente, alguma melhoria na aplicação do estudo de caso pela utilização da classificação difusa, foi efetuada uma análise complementar para a estimativa de  $\bar{Q}$ , onde não foi considerada a regressão difusa e, sim, a interpolação ponderada. Esta última técnica foi utilizada em KAVISKI (1992). Utilizando a classificação difusa e interpolação ponderada na estimativa de  $\bar{Q}$ , os valores de  $\bar{d}_q$  e  $\bar{d}_a$  resultaram iguais a 0,154 e 0,140, respectivamente. Para o método apresentado em KAVISKI (1992) os valores de  $\bar{d}_q$  e  $\bar{d}_a$  resultaram iguais a 0,168 e 0,160, respectivamente. Considerando-se estes dois índices, pode-se concluir que mesmo com a utilização da técnica de lógica difusa apenas na classificação há uma melhoria na regionalização do parâmetro  $\bar{Q}$ .

A regressão difusa forneceu ajustes perfeitos (grau de ajuste médio  $\bar{h} = 1$ ) para a maioria dos parâmetros hidrológicos analisados, quando utilizado o termo independente. Porém estas equações não possuem bom comportamento na aplicação em locais que não foram utilizados na construção do modelo.

Para o método proposto nesta dissertação os valores de  $\bar{d}_q$  e  $\bar{d}_a$  resultaram iguais a 0,127 e 0,118, respectivamente, na estimativa do parâmetro  $\bar{Q}$ . Para o método apresentado em KAVISKI (1992) os valores de  $\bar{d}_q$  e  $\bar{d}_a$  resultaram iguais a 0,168 e 0,160, respectivamente. Para a estimativa de  $\bar{Q}_{\max}$ , pelo método proposto nesta dissertação, os valores de  $\bar{d}_q$  e  $\bar{d}_a$  resultaram iguais a 0,372 e 0,278, respectivamente. Para o método apresentado em KAVISKI (1992) os valores de  $\bar{d}_q$  e  $\bar{d}_a$  resultaram iguais a 0,400 e 0,355, respectivamente. Considerando a estimativa de  $CV[Q_{\max}]$ , os valores de  $\bar{d}_q$  e  $\bar{d}_a$  resultaram pelo método proposto iguais a 0,334 e 0,252, respectivamente. Para o método apresentado em KAVISKI (1992) os valores de  $\bar{d}_q$  e  $\bar{d}_a$  resultaram iguais a 1,290 e 0,794, respectivamente. Por estes resultados, pode-se constatar que o método proposto nesta dissertação apresenta uma melhora na regionalização dos parâmetros.

Com base nos resultados apresentados pode-se concluir que a análise difusa pode ser utilizada na estimativa de parâmetros hidrológicos em regionalização. As

principais vantagens são maior simplicidade, rapidez e bons resultados. A classificação difusa apresentou-se como um método bastante consistente e poderia substituir outras técnicas de análises de agrupamentos.

Como no caso do modelo de regressão por mínimos quadrados ordinário (OLS), a regressão difusa considera que os erros são independentes e possuem variâncias iguais. A regressão difusa carece de maiores investigações e pode ser utilizada na complementação de outros métodos. Sempre que possível deve-se tomar o cuidado de verificar as estimativas para locais com dados e que não foram utilizados na construção do modelo, pois a consistência dos resultados não pode ser conclusiva, dadas as limitações da abordagem desta dissertação.

### 5.3. RECOMENDAÇÕES

Os controladores difusos são amplamente utilizados em diversas aplicações em recursos hídricos. Esta técnica de lógica difusa é especialmente adequada no gerenciamento de recursos hídricos. Recomenda-se a investigação sobre a utilização desta técnica na operação de reservatórios ou compatibilização dos diversos usos hídricos da bacia.

A regressão linear difusa como apresentada em BARDOSSY *et al.* (1990) consiste em um método mais complicado se comparado à regressão por metas difusas. Recomenda-se avaliar esta técnica em um estudo de caso e comparar os resultados com outros métodos de regressão.

Recomenda-se investigar os motivos pelos quais não foi possível obter as equações de regressão para  $\ln(CV[Q])$ . As investigações efetuadas nesta dissertação não permitiram o conhecimento da causa das inconsistências geradas para este caso. Considera-se pertinente a investigação de como considerar na construção do modelo de regressão difusa a não independência e a não igualdade das variâncias dos erros.

Tal como sugerido em HALL & MINNS (1999), poderia ser investigada a construção de um sistema híbrido na análise de agrupamentos. No referido trabalho o autor sugere a utilização das redes neurais *Kohonen* para a determinação do número de grupos, que servirá como dado de entrada na classificação “*c-means*”. A utilização das redes neurais *Kohonen* na classificação não exige o conhecimento prévio do número de classes. A vantagem da utilização da técnica “*c-means*” no particionamento final dos dados, verifica-se pela atenção especial dada aos pontos que possuem significantes níveis de aderência em mais de uma classe.

A proposição de novos métodos para a realização da classificação de dados encontra-se em franca expansão. A realização de pesquisas para investigar a validade destes métodos em regionalização hidrológica são de grande importância. Como por exemplo, cita-se o método baseado no modelo magnético granular, proposto por BLATT, WISEMAN & DOMANY (1997). O método de classificação fundamenta-se em um modelo geralmente usado para descrever propriedades físicas de materiais ferromagnéticos não homogêneos (modelo de Potts). Este método é interessante porque nenhuma hipótese é imposta sobre a distribuição dos dados a serem classificados. Pelas análises apresentadas pelos autores do método pode-se concluir que se atingiu um relativo sucesso em relação aos métodos clássicos de análise estatística multivariada. Recomenda-se que em estudos futuros sejam realizadas análises comparativas com o método de classificação difusa.

## REFERÊNCIAS BIBLIOGRÁFICAS

AZEVEDO, Fernando Mendes de, BRASIL, Lourdes Mattos e OLIVEIRA, Roberto Célio Limão de. **Redes Neurais com aplicações em controles e em sistemas especialistas**. Florianópolis: Bookstore, 2000.

BARDOSSY, Andras, BOGARDI, Istvan & DUCKSTEIN, Lucien. “Fuzzy regression in hydrology”. **Water Resources Research**, Washington, 7(26):1497-1508, julho, 1990.

BLATT, Marcelo, WISEMAN, Shai & DOMANY, Eytan. “Data clustering using a model granular magnet”. **Cond-mat/9702072 v1**, 1997.

BOGARDI, Istvan, BARDOSSY, Andras & DUCKSTEIN, Lucien. “Regional management of na aquifer for mining under fuzzy environmental objectives”. **Water Resources Research**, Washington, 6(19):1394-1402, dezembro, 1983.

BORELLI, João. **Automação da inferência humana com o uso de lógica nebulosa: uma aplicação**. São Paulo: 06/12/1999 (consulta na internet, endereço <http://membro.intermega.globo.com/celsofr/fuzzy//artigos> em 31/05/2001).

BRASIL. DNAEE - DEPARTAMENTO NACIONAL DE ÁGUAS E ENERGIA ELÉTRICA. **Inventário das estações pluviométricas**. Brasília, 1981.

BRASIL. DNAEE - DEPARTAMENTO NACIONAL DE ÁGUAS E ENERGIA ELÉTRICA. **Inventário das estações fluviométricas**. Brasília, 1987.

CAMPANA, Néstor Aldo. “Regionalização de hidrograma unitário usando regressão fuzzy”. **Anais do XI Simpósio Brasileiro de Recursos Hídricos e II Simpósio de Hidráulica e Recursos Hídricos dos Países de Língua Oficial Portuguesa**, Volume I – Hidrologia, Recife, 1995, pp. 33-38.

CHANG, Ni-Bin, CHEN, Y. L. & YANG, H. H.. “A fuzzy goal regression model for the construction cost estimation of municipal waste incinerators”. **International Journal of Systems Science**, 5(27):433-445, 1996.

FLEMING, George. **Computer simulation techniques in hydrology**. New York: Elsevier, 1975.

FONTANE, Darrell G., GATES, Timothy K. & MONCADA, Enrique. “Planning reservoir operations with imprecise objectives”. **Journal of Water Resources Planning and Management**, Reston, 3(123):154-162, maio/jun., 1997.

GALVÃO, Carlos de Oliveira. “Controle de processos e sistemas de apoio à decisão”. *In*: GALVÃO, Carlos de Oliveira e VALENÇA, Mêuser Jorge Silva. **Sistemas inteligentes: aplicações a recursos hídricos e sistemas ambientais**. Porto Alegre: Ed. Universidade/UFRGS/ABRH, 1999, pp. 225-246. (Coleção ABRH de Recursos Hídricos; 7.)

GALVÃO, Carlos de Oliveira. “Modelagem fuzzy da previsão climática no nordeste para uso em operação de reservatórios”. **Anais do XI Simpósio Brasileiro de Recursos Hídricos e II Simpósio de Hidráulica e Recursos Hídricos dos Países de Língua Oficial Portuguesa**, Volume II – Água Subterrânea e Semi-Árido, Recife, 1995, pp. 57-62.

GATES, Timothy K., HEYDER, Walter E., FONTANE, Darrell G. & SALAS, José D. “Multicriterion strategic planning for improved irrigation delivery. I: Approach”. **Journal of Irrigation and Drainage Engineering**, Reston, 6(117):897-913, nov./dez., 1991.

HALL, M. J. & MINNS, A. W.. “The classification of hydrologically homogeneous regions”. **Hydrological Sciences Journal**, Wallingford, 44(5):693-704, outubro, 1999.

KAVISKI, Eloy. “Métodos de regionalização de eventos e parâmetros hidrológicos”. Dissertação de Mestrado, UFPR, Curso de Pós-Graduação em Engenharia Hidráulica, Curitiba: novembro, 1992.

KAVISKI, Eloy e FIOR, Maria T.A.B. “Regionalização de curvas de permanência de vazões médias diárias em pequenas bacias hidrográficas do Estado do Paraná. **Anais do VI Simpósio Brasileiro de Hidrologia e Recursos Hídricos**. São Paulo, 1985, v. 3, pp. 188-201.

\_\_\_\_\_. “Regionalização de vazões máximas em pequenas bacias hidrográficas do Estado do Paraná. **Anais do XII Congresso Latino Americano de Hidráulica**. São Paulo, 1986, v. 2, pp. 62-75.

MathWorks. **Optimization Toolbox – User’s Guide Version 5**, 1997 (Manual do Usuário do MATLAB).

MathWorks. **Fuzzy Logic Toolbox – User’s Guide Version 2**, 1998 (Manual do Usuário do MATLAB).

OWEN, William J., GATES, Timothy K. & FLUG, Marshall. “Variability in perceived satisfaction of reservoir management objectives”. **Journal of Water Resources Planning and Management**, Reston, 3(123):147-153, maio/jun., 1997.

RUSSELL, Samuel O. & CAMPBELL, Paul F.. “Reservoir operating rules with fuzzy programming”. **Journal of Water Resources Planning and Management**, Reston, 3(122):165-170, maio/jun., 1996.

SHRESTHA, Bijaya P., DUCKSTEIN, Lucien & STAKHIV, Eugene Z.. “Fuzzy rule-based modeling of reservoir operation”. **Journal of Water Resources Planning and Management**, Reston, 4(122):262-269, jul./ago., 1996.

TANAKA, Hideo, UEJIMA, Satoru & ASAI, Kiyoji. “Linear regression analysis with fuzzy model”. **IEEE Transactions on Systems, Man and Cybernetics**, 6(SMC-12):903-907, nov./dec., 1982.

TUCCI, Carlos E. M. **Regionalização de Vazões**. Porto Alegre: Ed. Universidade/UFRGS, 2002.

YIN, Y. Y., HUANG, G. H. & HIPEL, K. W. "Fuzzy relation analysis for multicriteria water resources management". **Journal of Water Resources Planning and Management**, Reston, 1(125):41-47, jan./fev., 1999.

YU, Pao-Shan, CHEN, Chia-Jung & CHEN, Shiann-Jong. "Application of gray and fuzzy methods for rainfall forecasting". **Journal of Hydrologic Engineering**, Reston, 4(5):339-345, outubro, 2000.

ZADEH, Lotfi A. "Fuzzy sets". **Information and Control**, 8(3):338-353, 1965.