

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL

INSTITUTO DE PESQUISAS HIDRÁULICAS

ANÁLISE REGIONAL DE FREQUÊNCIAS  
APLICADA À PRECIPITAÇÃO PLUVIAL

TERESA LUISA LIMA DE CARVALHO

Dissertação submetida ao Programa de Pós-Graduação em Recursos Hídricos e Saneamento Ambiental da Universidade Federal do Rio Grande do Sul como requisito parcial para a obtenção do título de Mestre em Recursos Hídricos e Saneamento Ambiental.

Porto Alegre, outubro de 2007

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL

INSTITUTO DE PESQUISAS HIDRÁULICAS

ANÁLISE REGIONAL DE FREQUÊNCIAS APLICADA À PRECIPITAÇÃO PLUVIAL

TERESA LUISA LIMA DE CARVALHO

Dissertação submetida ao Programa de Pós-Graduação em Engenharia de Recursos Hídricos e Saneamento Ambiental da Universidade Federal do Rio Grande do Sul como requisito parcial para a obtenção do título de Mestre em Recursos Hídricos e Saneamento Ambiental.

ORIENTADOR: Prof. Dr. Olavo Correa Pedrollo

Banca Examinadora

Prof. Dr. André Luiz Lopes da Silveira                      IPH/UFRGS

Prof. Dr. Geraldo Lopes da Silveira                                      UFSM

Dr. Ronaldo Matzenauer    FEPAGRO

Porto Alegre, outubro de 2007

## APRESENTAÇÃO

Este trabalho foi desenvolvido no Programa de Pós-Graduação em Recursos Hídricos e Saneamento Ambiental da Universidade Federal do Rio Grande do Sul, sob orientação do Prof. Dr. Olavo Correa Pedrollo e apoio financeiro da CAPES.

Chegando ao fim dessa etapa, não me faltam agradecimentos que dirijo a todos aqueles que, de alguma forma, contribuíram para o desenvolvimento desse trabalho.

Início com o professor Olavo Pedrollo pela paciência e por dar um rumo a tantas e tantas idéias que tão teimosamente surgiram nessa caminhada.

À Márcia Pedrollo, extensivo à CPRM, pelo fornecimento do banco de dados e auxílio com os programas Hidro e Hidroplu.

Aos colegas do IPH e às amizades construídas ou prolongadas nesta estada em Porto Alegre: Angélica, Tati, Patrícia, Ana, Adriano; Ane e Roger, Chris e Dani, Ruberto e Regina, Marcus e Franci, Diego e Marilu, Uziel e Taci, Diogo, Nilson, Adriano, Fernando, Ferdnando, Lidiane e Geilma; aos companheiros da Avaí, 15: Rodrigo, Luís, Fernanda e Molly; e aos acolhedores amigos da Paróquia da Conceição.

Aos meus amigos e amigas que, na distância, torceram e rezaram por mim, durante todo este tempo.

Ao querido Petry e sua família, pelo apoio e afeto que tornaram meus dias mais suaves e alegres.

À minha família, minhas avós, tios, primos, minha sobrinha e meus irmãos, cada um a seu modo, por me prestigiarem com suas saudades e me apoiarem nessa jornada em terras distantes. E aos meus pais, Orlando e Heloisa, por respeitarem minhas opções e, com seu exemplo, seu carinho e seu incentivo, animarem meu espírito a cada dia.

Ao Sidnei e à Karina que, de coração aberto, no início e no fim, me colheram em seu lar. Vocês foram anjos no meu caminho (com asas alaranjadas) e eu serei eternamente grata. Obrigada, mesmo.

E, para terminar, agradeço a Deus simplesmente porque me ama assim, apaixonadamente.

## RESUMO

O conhecimento da probabilidade de ocorrência de precipitações é de grande importância para o planejamento de atividades como produção agrícola, sistemas de abastecimento de água e geração de energia. Neste sentido, a análise regional de frequências se apresenta como uma ferramenta para a redução de incertezas na estimativa das probabilidades, podendo também diminuir a dimensionalidade das informações. Desta forma, esta pesquisa tem por objetivo estudar uma metodologia para análise regional de frequências, aplicada à precipitação, na qual se admite que as séries que formam uma região homogênea seguem uma mesma distribuição, que dispensa a utilização de fatores de escala locais, e que seus parâmetros podem ser satisfatoriamente estimados pela média dos parâmetros de ajuste das séries locais. Para compor a metodologia foi construído um teste de homogeneidade regional e um teste para avaliar a qualidade do agrupamento, que leva em consideração os objetivos da análise. Os dois testes citados e a verificação da qualidade da obtenção da distribuição regional pela média dos parâmetros locais foram baseados na medida da máxima diferença de probabilidades entre duas séries e no teste de aderência de Kolmogorov-Smirnov (KS), auxiliando, então, numa melhor compreensão dos resultados, já que esse teste é referência na avaliação da qualidade do ajuste de funções probabilísticas. Como estudo de caso, a metodologia foi aplicada na análise de precipitações totais mensais e anuais no Rio Grande do Sul e em Santa Catarina, utilizando-se o método de agrupamento difuso *Fuzzy C-Means* para identificação das regiões. Verificou-se que o modelo de obtenção da distribuição regional proposto é satisfatório, embora necessite de um fator de correção para a distribuição Gama. O teste de aderência regional mostrou-se bastante útil para orientar a ponderação das variáveis e para auxiliar na determinação do melhor número de regiões, o que foi comprovado pelo alto grau de homogeneidade verificado, mesmo antes de serem aplicados ajustes nas regiões resultantes do agrupamento. Constatou-se ainda que a utilização dos graus de pertinência não produziu resultados satisfatórios para fazer estimativas pontuais e pouco contribuiu para a realização de ajustes no agrupamento. Todavia, o ajuste do modelo de alocação para novos pontos, através da matriz de pertinência, apresentou bons resultados, embora ainda seja necessária sua verificação em situação real. A análise regional da área de estudo resultou em nove regiões, que foram consideradas heterogêneas em apenas dois dos 117 períodos avaliados. Além disso, a partir do teste KS ( $\alpha = 20\%$ ), verificou-se que as distribuições regionais se mostraram válidas para 97,8% das séries, confirmando a possibilidade de utilizar a distribuição regional, sem a necessidade dos fatores de escala locais.

## ABSTRACT

Information on rainfall probability are helpful for some activities planning including agriculture, water supply and hydropower generation. For such purpose, the regional frequency analysis is a powerful tool to reduce uncertainties in probabilities estimation and also information's dimension. This research evaluates the possibility of taking rainfall regional frequency analysis considering that in a homogeneous region the site frequency distributions are identical, setting aside local scaling factors, and that its parameters may satisfactorily be estimated by average parameters of local series fitting. To address such goals, the methodology consisted of elaborating a regional homogeneity test and a test to assess the cluster quality. Both tests and the assessment over the regional distribution quality were based on the maximum probability differences between two series and on the Kolmogorov-Smirnov test (KS). As a case study, the methodology was applied on Rio Grande do Sul and Santa Catarina States monthly and annual precipitation data, using clustering method Fuzzy C-Means for identification of regions. It was verified that the regional distribution model is satisfactory, although it needs a correction factor for Gamma distribution. The test of regional adherence was useful to guide the variables weighting and to help on the determination of the optimum number of regions, once a high degree of homogeneity was obtained, even before the application of adjustments on the resultant regions from the cluster analysis. The use of membership degrees was not satisfactory on estimating at-site features and resulted on low contribution for cluster adjustments. However, the allocation model fitting for new points presented good results, though a real case assessment is still required. The regional analysis resulted in nine regions, which were heterogeneous in only two of the 117 evaluated periods. Moreover, from the KS test ( $\alpha= 20\%$ ), it was verified that the regional distributions were valid for 97.8% of the series, confirming the possibility to use regional distribution, without any local scaling factor.

## SUMÁRIO

<b>RESUMO</b> .....	iii
<b>ABSTRACT</b> .....	v
<b>SUMÁRIO</b> .....	vi
<b>LISTA DE TABELAS</b> .....	ix
<b>LISTA DE FIGURAS</b> .....	x
<b>LISTA DE SÍMBOLOS</b> .....	xi
<b>1 INTRODUÇÃO</b> .....	1
1.1 Motivação .....	1
1.2 Objetivos.....	4
<b>2 A PRECIPITAÇÃO E SUA VARIABILIDADE</b> .....	5
2.1 Aspectos gerais .....	5
2.2 Variabilidade Temporal.....	7
2.2.1 Fatores condicionantes .....	7
2.2.2 Distribuição de freqüência das séries de precipitação.....	8
2.3 Variabilidade espacial.....	10
<b>3 ANÁLISE REGIONAL</b> .....	11
3.1 Aspectos gerais .....	11
3.2 Consistência dos dados.....	13
3.3 Regiões homogêneas .....	14
3.3.1 Conceituação .....	14
3.3.2 Análise de agrupamento .....	15
3.3.2.1 Métodos hierárquicos e de partição.....	15
3.3.2.2 Número de grupos .....	18
3.3.3 Testes de homogeneidade.....	18
3.3.4 Escolha das variáveis e ajustes no agrupamento .....	20
3.4 Determinação da curva regional.....	21

3.5	Qualidade do ajuste da curva regional.....	23
3.6	Estimativa do fator de escala local .....	25
4	AGRUPAMENTO DIFUSO .....	26
4.1	Conceitos gerais.....	26
4.2	Método de agrupamento <i>Fuzzy C-Means</i> .....	27
4.3	Algoritmo de agrupamento .....	30
4.4	Padronização e ponderação das variáveis.....	30
4.5	Índices de validade do agrupamento .....	31
4.6	Delimitação dos grupos e defusão.....	34
4.7	Alocação de novos pontos .....	34
4.8	Estimativas pontuais a partir dos graus de pertinência.....	35
5	METODOLOGIA.....	36
5.1	Introdução.....	36
5.2	Recursos metodológicos.....	36
5.2.1	Verificação do modelo para obtenção da distribuição regional .....	36
5.2.2	Teste de aderência regional .....	38
5.2.3	Teste de homogeneidade .....	39
5.3	Etapas do método .....	42
5.3.1	Aspectos Gerais .....	42
5.3.2	Preparação dos dados.....	42
5.3.3	Ponderação das variáveis.....	43
5.3.4	Determinação do melhor número de grupos.....	44
5.3.5	Ajustes preliminares .....	44
5.3.6	Teste de homogeneidade .....	45
5.3.7	Ajuste de modelo para alocação de pontos não amostrados.....	45
5.3.8	Estimativa de características locais através dos graus de pertinência .....	46
6	ESTUDO DE CASO .....	47

6.1	Local da aplicação .....	47
6.2	Seleção das estações pluviométricas .....	48
6.3	Procedimentos aplicados .....	53
7	RESULTADOS E DISCUSSÃO .....	58
7.1	Verificação do método de obtenção das distribuições regionais.....	58
7.2	Identificação das regiões homogêneas .....	60
7.2.1	Ponderação das variáveis.....	60
7.2.2	Determinação do número de regiões .....	61
7.2.2.1	Análise dos Índices de validade.....	61
7.2.2.2	Resultados do teste de aderência regional .....	63
7.2.3	Ajustes no agrupamento .....	64
7.2.4	Teste de Homogeneidade.....	66
7.3	Caracterização das Regiões homogêneas .....	69
7.4	Distribuições regionais .....	79
7.5	Modelo para Alocação de pontos não amostrados .....	83
7.5.1	Determinação da matriz de pertinência .....	83
7.5.2	Estimativa do fator de escala local .....	84
8	CONCLUSÕES E RECOMENDAÇÕES .....	85
8.1	Conclusões.....	85
8.2	Recomendações .....	89
9	REFERÊNCIAS BIBLIOGRÁFICAS .....	90

## LISTA DE TABELAS

Tabela 3.1. Valor crítico do teste de aderência de Kolmogorov-Smirnov. ....	24
Tabela 6.1. Número de estações excluídas de acordo com o tamanho da série. ....	49
Tabela 6.2: Número identificador das estações pluviométricas e seus respectivos códigos, unidade da federação e ano de início da série. ....	52
Tabela 6.3: Fatores de mudança de escala. ....	54
Tabela 7.1: Valores de $\mu dp$ e $\sigma dp$ para a distribuição Normal. ....	58
Tabela 7.2: Valores de $\mu dp$ e $\sigma dp$ para a distribuição Gama. ....	58
Tabela 7.3: Média e desvio padrão da máxima diferença de probabilidades dos ajustes. ....	59
Tabela 7.4: Média de Sd (%) para diversos ponderadores. ....	60
Tabela 7.5 Resultado do teste de aderência regional. ....	63
Tabela 7.6: Valores de $\mu dp_s$ e $\sigma dp_s$ e os limites de homogeneidade, curva real. ....	66
Tabela 7.7: Valores de $\mu dp_s$ e $\sigma dp_s$ e os limites de homogeneidade, curva adimensional. ....	66
Tabela 7.8. Teste de homogeneidade $H_{dp}$ para a distribuição regional em escala real. ....	67
Tabela 7.9. Teste de homogeneidade $H_{dp}$ para a distribuição regional adimensional. ....	68
Tabela 7.10: Protótipos do agrupamento. ....	70
Tabela 7.11: Percentual pluviométrico por trimestre. ....	70
Tabela 7.12 Estações pertencentes à Região 1a. ....	75
Tabela 7.13 Estações pertencentes à Região 1b. ....	75
Tabela 7.14 Estações pertencentes à Região 2. ....	75
Tabela 7.15 Estações pertencentes à Região 3. ....	76
Tabela 7.16 Estações pertencentes à Região 4. ....	76
Tabela 7.17 Estações pertencentes à Região 5. ....	77
Tabela 7.18 Estações pertencentes à Região 6. ....	77
Tabela 7.19 Estações pertencentes à Região 7. ....	78
Tabela 7.20 Estações pertencentes à Região 8. ....	78
Tabela 7.21 Parâmetros e quantis Regionais. ....	80
Tabela 7.22: Séries discordantes no teste de aderência. ....	83
Tabela 7.23 Erros na estimativa das médias de precipitação locais. ....	84

**LISTA DE FIGURAS**

Figura 3.1: Métodos Hierárquicos. (Adaptado de Carlantonio (2001)).	16
Figura 4.1 Diferenças entre um conjunto clássico (a) e um conjunto difuso (b).	27
Figura 6.1 Localização e relevo do Rio Grande do Sul e Santa Catarina	47
Figura 6.2 Pré-seleção das estações pluviométricas do Rio Grande do Sul.	49
Figura 6.3: Distribuição espacial das estações pluviométricas utilizadas.	51
Figura 6.4 Ajuste da distribuição Gama para diferentes substituições do valor zero.	53
Figura 6.5 Frequência dos graus de pertinência máximos.	54
Figura 7.1 Curvas regionais Gama estimadas pela média dos parâmetros locais.	59
Figura 7.2: Percentual de séries discordantes na matriz não ponderada.	61
Figura 7.3: Índices de validade.	62
Figura 7.4 Relação entre $\mu_{max}$ e o resultado do teste de aderência regional.	64
Figura 7.5 Regiões homogêneas definidas para o Rio Grande do Sul e Santa Catarina.	69
Figura 7.6 Regiões de 1 a 8	72
Figura 7.7: Isoietas das precipitações médias anuais no Rio Grande do Sul (Sotério et al. 2005).	74

## LISTA DE SÍMBOLOS

$c$	Número de classes do agrupamento.
$CV-L$	Coeficiente de variação L
$Di$	Medida de discordância
$d_{ik}$	Distância do elemento $i$ ao centro do grupo $k$
$dm$	Diferença máxima entre probabilidade observada e estimada
$dp$	Média de $dm$ por região
$e_i$	Fator de escala do ponto $i$
$Ep$	Entropia de partição
$f$	Frequência de não-excedência
$F_G(x)$	Função distribuição de probabilidade da distribuição Gama
$F_N(x)$	Função distribuição de probabilidade da distribuição Normal
$G(x)$	Função de densidade de probabilidade da distribuição Gama
$H$	Medida de heterogeneidade
$H_{pd}$	Medida de heterogeneidade para a diferença máxima entre probabilidades
$m$	Coeficiente de difusividade
$N(x)$	Função de densidade de probabilidade da distribuição Normal
$p$	Probabilidade de não-excedência
$Pd$	Partição difusa
$Pds$	Partição difusa significativa
$Q$	Função de quantis
$q(F)$	Função de quantis adimensionais
$Sd$	Número de séries discordantes
$U$	Matriz de pertinência
$v$	Protótipo do agrupamento
$V$	Desvio padrão ponderado da estimativa de CV-L regional e local
$V_{FS}$	Índice de Fukuyama e Sugeno
$V_K$	Índice de Kwon
$V_{PB}$	Índice PBMF
$V_{XB}$	Índice de Xie-Beni
$w$	Ponderador
$Z$	Medida de aderência
$\Gamma$	Função Gama

$\alpha$	Parâmetro de posição da distribuição Gama
$\alpha$	Nível de significância
$\beta$	Parâmetro de escala da distribuição Gama
$\mu$	Parâmetro de posição da distribuição Normal
$\mu dp_s$	Média de $dp$ nas regiões simuladas
$\mu_{ik}$	Grau de pertinência do elemento $i$ ao grupo $k$
$\mu_{max}$	Grau de pertinência máximo
$\mu v$	Média dos $V$ 's das regiões simuladas
$\theta^i$	Parâmetro local
$\theta^R$	Parâmetro regional
$\sigma$	Parâmetro de escala da distribuição Normal
$\sigma dp_s$	Desvio de $dp$ nas regiões simuladas
$\sigma v$	Desvio dos $V$ 's das regiões simuladas

# 1 INTRODUÇÃO

## 1.1 Motivação

O conhecimento do comportamento da precipitação pluvial é de relevante importância, seja para planejamento de atividades relacionadas com a água, como geração de energia elétrica, sistemas de irrigação, exploração de aquíferos, sistemas de reserva e suprimento de água para abastecimento público, seja para estudos do comportamento hidrológico, como do escoamento superficial e da infiltração, que em geral utilizam a precipitação como variável condicionante.

Entretanto, como outros fenômenos naturais, a precipitação está sujeita a uma componente aleatória, tanto espacial quanto temporal, não sendo permitido saber exatamente quando, onde e como ela irá ocorrer e isto representa um empecilho às atividades de planejamento citadas acima.

De acordo com Lanna (2004) a estatística e a teoria das probabilidades surgiram exatamente para contornar esse tipo de deficiência de informação. Ou seja, a partir da observação, coleta e análise de fenômenos ocorridos, com posterior ajuste a modelos probabilísticos teóricos que possam representar o fenômeno em questão, é possível obter estimativas dos riscos de ocorrência futura.

Embora tenham existido avanços na construção de modelos probabilísticos e de métodos de ajuste visando a obtenção de estimativas mais eficientes, a qualidade da análise de frequências em cada local depende extremamente do tamanho e da consistência das séries observadas, que na maioria dos casos não são satisfatórios.

Entretanto, quando existem séries observadas em vários pontos disseminados no espaço e há similaridade nas frequências observadas, o conjunto de séries pode ser reunido em uma região, dita homogênea, e ser tratado conjuntamente, através da análise regional de frequências.

Assim, as estimativas de probabilidades obtidas pela análise regional, por levarem em conta um número maior de dados, poderão ser mais robustas que as produzidas pela análise de frequências de cada local, individualmente.

Dos métodos de análise regional de frequências (muitas vezes tratados em hidrologia como métodos de regionalização), destaca-se o procedimento do *index-flood* proposto por Dalrymple em 1960 (Davis e Naghetinni, 2001). Seu princípio básico é admitir que, numa região homogênea, as séries seguem uma mesma distribuição, à exceção de um fator de escala, que é próprio de cada local, denominado *index-flood*. Desta maneira, uma distribuição regional adimensional é

determinada e considerada válida para representar a forma da distribuição de probabilidades em todos os pontos da região, enquanto o valor das estimativas de quantis em um dado local é obtido multiplicando-se o quantil da distribuição adimensional pelo fator de escala próprio do local.

Entretanto, considerando-se que os fatores condicionantes do fenômeno observado são similares para os pontos que formam a região, parece admissível que as séries sejam igualmente distribuídas tanto em termos de forma, quanto de escala, desde que as possíveis diferenças de escala possam ser atribuídas à aleatoriedade presente no processo.

Essa hipótese merece ser investigada, pois se, na prática, for possível identificar regiões que seguem uma mesma distribuição, sem restrições, então se tornará desnecessário o conhecimento, ou modelagem, do fator de escala local para a obtenção das estimativas de quantis locais, o que representa uma vantagem, principalmente se as estimativas regionais forem mais robustas.

Todavia, caso seja do interesse da análise a manutenção da escala local, a distribuição regional poderá ser adimensionalizada por um fator de escala regional (que pode ser tomado como a média da distribuição regional) e redimensionalizada pelo fator de escala local, conforme usualmente se faz no método do *index-flood*.

A metodologia para análise regional proposta por Hosking e Wallis (1997), que utiliza os princípios do *index-flood*, tem se difundido bastante atualmente. Seu principal atrativo está na utilização de momentos-L para obtenção da distribuição regional de frequências e para a construção de estatísticas auxiliares, com destaque para a medida de heterogeneidade  $H$ , que inseriu maior objetividade no julgamento da homogeneidade das regiões.

A referida medida de heterogeneidade, porém, depende do conhecimento dos momentos-L, o que limita a sua aplicação. Seria desejável um recurso mais generalizável para a análise, já que a determinação dos parâmetros da distribuição regional pode ser executada por diferentes métodos.

Hosking e Wallis (1997), por exemplo, calculam os parâmetros da distribuição regional a partir dos momentos-L regionais, embora outras formas de determinação da distribuição regional sejam praticadas (Dalrympe, 1960 apud Hosking e Wallis (1997); Tucci, 2002 ).

Porém, seria útil verificar se um modelo simples, que utiliza a média dos parâmetros de ajuste das séries que formam uma região homogênea, produziria boas estimativas da distribuição regional.

Note-se que, para uma análise regional eficiente, é fundamental uma adequada identificação das regiões homogêneas.

Embora muitos métodos possam ser utilizados para a identificação das regiões, alguns autores, como Jingyi e Hall (2004) e Rao e Srinivas (2006), alegaram vantagens adicionais na utilização do método de agrupamento difuso *Fuzzy C-Means*, principalmente relacionadas ao uso dos graus de pertinência, tanto como auxiliar em possíveis ajustes nas regiões, como realocação de elementos, quanto como parte de um modelo para obtenção de características pontuais.

Porém, por considerarem suas investigações ainda inconclusivas, estes autores sugerem pesquisas adicionais da aplicação destes recursos.

Da mesma forma, a alocação de pontos com dados incompletos (variáveis independentes) aos agrupamentos é uma possibilidade ainda por experimentar, que abre a perspectiva de associação posterior de pontos não amostrados às regiões definidas.

Um procedimento importante na análise de agrupamento é a adimensionalização das variáveis a serem agrupadas, cujo objetivo é reduzir a importância que é indevidamente atribuída, pelos algoritmos de agrupamento, às variáveis numericamente maiores. Porém, algumas destas variáveis podem influenciar mais fortemente os fenômenos estudados, sendo pertinente, portanto, a aplicação de um sistema de pesos que aumente sua relevância no processo de agrupamento.

Embora a escolha desses pesos normalmente se baseie na experiência do analista, seria útil a utilização de algum critério objetivo, que orientasse a escolha destes pesos na direção de uma configuração de agrupamento mais adequada aos objetivos da análise.

Além disso, os métodos de agrupamento, em geral, apresentam como deficiência a necessidade de definição preliminar do número de classes. Usualmente, o procedimento adotado é iterativo, efetuando-se os agrupamentos para vários números de classes, e avaliando-se a melhor alternativa através de um índice de validade. No entanto, existem muitas alternativas de índices de validade disponíveis, sendo desejável que pelo menos um certo número destes índices apontem para uma mesma solução, atribuindo-lhe maior confiabilidade. Alguma forma de conciliação ou investigação conjunta destes índices, associada a outros recursos, que levem em consideração os objetivos específicos do agrupamento para a definição de regiões homogêneas, poderia auxiliar numa solução mais conclusiva.

## 1.2 Objetivos

Respondendo às motivações apresentadas, pretende-se investigar se é possível realizar a análise regional de frequências de precipitações com as condições das séries que formam uma região homogênea seguirem uma mesma distribuição de probabilidades, que permita dispensar a utilização de fatores de escala locais, e cujos parâmetros sejam satisfatoriamente estimados pela média dos parâmetros de ajuste das séries locais.

Os recursos planejados para o desenvolvimento da metodologia resultante constam, especificamente, de:

- Um método de verificação da homogeneidade regional;
- Um critério de avaliação da qualidade do agrupamento, que auxilie na escolha da ponderação das variáveis e na determinação do melhor número de regiões.

Como objetivo secundário, pretende-se verificar, durante um estudo de caso, possíveis contribuições do método de agrupamento difuso *Fuzzy C-Means*, ou, discriminadamente:

- se os graus de pertinência podem auxiliar em possíveis ajustes nas regiões após a fase de agrupamento, tais como a realocação de elementos para outras regiões;
- se características pontuais, como a média mensal em cada ponto, podem ser satisfatoriamente estimadas por um modelo baseado nos graus de pertinência.
- se um modelo de alocação para pontos não amostrados pode ser considerado eficiente.

No estudo de caso, a metodologia proposta é aplicada às precipitações totais mensais e anuais no Rio Grande do Sul e em parte de Santa Catarina, esperando-se contribuir para um melhor conhecimento sobre as probabilidades associadas à ocorrência de precipitações nesta área.

## 2 A PRECIPITAÇÃO E SUA VARIABILIDADE

### 2.1 Aspectos gerais

A precipitação é o fenômeno responsável pela reposição da água no continente. Embora aconteça sob a forma de neblina, orvalho, geada, neve, granizo e chuva, aqui o termo precipitação está associado exclusivamente à precipitação pluvial ou chuva.

Basicamente, a formação da precipitação deve-se ao processo de ascensão, resfriamento adiabático e condensação do ar quente e úmido, proveniente das camadas mais baixas da atmosfera, seguido do processo de crescimento das gotículas de água, até que tenham massa suficiente para gerar uma velocidade de queda superior à dos movimentos ascendentes verticais da atmosfera.

À medida que a massa de ar se eleva na atmosfera, ela se expande devido à diminuição da pressão atmosférica, mas sem que haja troca de calor. Então, o aumento do volume da massa de ar é compensado pela perda de temperatura, ocorrendo o resfriamento. Quando é atingido o ponto de saturação, a condensação se inicia sobre os núcleos de condensação existentes (Tubelis e Nascimento, 1987).

O processo de crescimento das gotículas ocorre por difusão do vapor de água e por coalescência e coalescência das gotículas entre si. A difusão ocorre na presença de um gradiente de tensão de vapor, havendo atração das partículas de maior tensão pelas de menor tensão. Quanto maior a temperatura e a salinidade, menor é a tensão de vapor. Assim, cristais de gelo em nuvens frias, e gotículas frias, partículas salinas e outras partículas higroscópicas em nuvens quentes constituem núcleos de condensação atraindo as gotículas de água pura com maior tensão de vapor.

No entanto, o processo de difusão por si só não consegue produzir gotas com tamanho suficiente para precipitar. Então, a partir de certo diâmetro, o crescimento se faz por coalescência e coalescência, ou seja, pelo choque direto entre as gotículas, decorrente das diferenças de velocidade de queda, seguido da incorporação das gotas menores pelas maiores. Este processo que inicialmente é lento se intensifica exponencialmente com o aumento do tamanho das gotas (Tubelis e Nascimento, 1987).

De acordo com a forma de ascensão do ar úmido as precipitações podem ser classificadas em convectivas, orográficas e frontais (Bertoni e Tucci; 2004).

O aquecimento do ar úmido próximo à superfície terrestre pode criar camadas de ar de diferentes densidades que se mantêm em equilíbrio instável. Quando ocorre uma perturbação desse equilíbrio, há uma ascensão brusca e localizada do ar quente e úmido, por ser menos denso em relação ao ar circundante. Se a elevação for suficiente para atingir a saturação e condensação há a formação de nuvens e possivelmente a formação da chamada precipitação convectiva.

Como está diretamente relacionada ao balanço positivo de radiação sobre a superfície, a precipitação convectiva é característica de regiões equatoriais, ocorrendo também em regiões temperadas durante o verão, devido ao aumento da radiação.

A precipitação orográfica, por sua vez, ocorre devido à interação entre o relevo e as massas de ar quente e úmido, geralmente vindas do oceano. As encostas de serras e montanhas funcionam como um anteparo que impede o fluxo horizontal da massa de ar, forçando o movimento vertical. Quando o nível de condensação é mais baixo que o topo de encosta sempre há formação de nuvens, que podem evoluir e dar origem à precipitação que, em geral, tem pouca intensidade, grande duração e cobre pequenas áreas.

Nas encostas situadas a barlavento, as chuvas são mais abundantes e diminuem muito rapidamente a sotavento, isso acontece tanto pela perda de umidade que ocorre a partir do nível de condensação, como pela quebra do mecanismo de subida ao atingir o topo da encosta. Sem a presença do anteparo a massa de ar descende, ocorrendo compressão adiabática, aumento da temperatura e dissolução da nuvem. Por conta disso, o lado oposto à encosta de subida é conhecido como região de sombra pluviométrica.

O processo frontal de deslocamento vertical de massas de ar é causado pela diferença de densidade de massas de ar em deslocamento horizontal. Uma massa de ar frio (mais densa) ao encontrar-se com uma massa quente (normalmente úmida) age como uma cunha, elevando a massa de ar quente que, por resfriamento adiabático, pode produzir chuva. As chuvas frontais são geralmente de grande duração, atingindo grandes áreas com intensidade média.

Em geral, a chegada de uma frente fria (uma massa de ar polar, por exemplo) forma nuvens mais profundas e chuvas mais intensas. O avanço das frentes quentes, ao contrário, é mais estável, gerando nuvens em camadas e precipitações fracas (Grimm, 1999).

Observa-se que os diferentes tipos de precipitação são fruto da interação de fatores estáticos (posição geográfica, altitude, configuração do relevo) e dinâmicos (movimentação das massas de ar, direção e intensidade dos ventos, incidência de radiação solar, etc) que imprimem à precipitação um caráter de variabilidade.

## 2.2 Variabilidade Temporal

### 2.2.1 Fatores condicionantes

Analisada pontualmente no espaço, a precipitação apresenta variações temporais intra e inter-anuais, com meses e anos mais úmidos ou menos úmidos, o que depende da variação temporal de fatores condicionantes da precipitação, como a incidência de radiação e a movimentação das massas de ar.

Efeitos periódicos, como a variação de incidência de radiação durante o ano, determinam as variações de temperatura na superfície do globo e condicionam a movimentação das massas de ar, podendo ocasionar configurações de precipitação específicas para determinadas épocas do ano.

Além disso, alterações na interação entre a superfície dos oceanos e a atmosfera (cuja periodicidade de ocorrência varia de 1 a 10 anos) podem afetar o comportamento climático, provocando anomalias de precipitação em algumas regiões ou até globalmente (Oliveira, 2001).

O fenômeno El-Niño é um exemplo disso e se caracteriza por um aumento anormal na temperatura da superfície do mar (TSM) no Pacífico Tropical com enfraquecimento dos ventos alísios na região equatorial. O mesmo fenômeno invertido, ou seja, resfriamento da superfície do mar é conhecido como La-Niña (ou fase fria do El-Niño) e geralmente ocasiona efeitos opostos ao do El-Niño, embora nem sempre uma mesma região seja influenciada efetivamente pelas duas fases do fenômeno (Oliveira, 2001).

Somam-se a esses fatores as condições naturais da circulação atmosférica, altamente aleatórias, que são responsáveis pela ação dos ventos, deslocamentos de umidade, etc.

Ou seja, a precipitação está cercada de condicionantes aleatórias, de forma que previsões de sua ocorrência não conseguem ser efetivadas com a antecedência necessária às atividades de planejamento e gestão dos recursos hídricos, que em geral são desenvolvidas para o longo prazo.

Para isso a estatística e a teoria das probabilidades têm contribuído, e, com base no tratamento das informações já existentes, oferece recursos para, a partir de modelos teóricos, serem estimadas as probabilidades ou os riscos de ocorrência de determinadas magnitudes de precipitação (bem como de outros eventos naturais).

### 2.2.2 Distribuição de freqüência das séries de precipitação

Embora seja comum a utilização da média amostral de séries de trinta anos como valor de referência na caracterização dos totais pluviométricos, sabe-se que o conhecimento da distribuição de freqüência desses valores traz informações muito mais esclarecedoras, realistas e úteis do comportamento pluviométrico.

A título de esclarecimento, uma distribuição de freqüências se refere ao percentual da ocorrência de um evento em um conjunto de amostras observadas, enquanto uma distribuição de probabilidades se refere ao percentual com que se espera que um evento venha a ocorrer. Pode-se de maneira geral afirmar que a distribuição de probabilidades é uma extrapolação da distribuição de freqüências. Para se passar de uma distribuição de freqüências para uma distribuição de probabilidades geralmente se recorre a ajustes de modelos teóricos de probabilidade ao conjunto de dados observados (Lanna, 2004).

Uma distribuição de probabilidades pode ser representada na forma de uma função de densidade de probabilidade, que ao ser integrada em um dado intervalo determina a probabilidade de um evento ocorrer nesse intervalo, e na forma da função cumulativa de probabilidades, ou função de probabilidades acumulada, cuja ordenada expressa a probabilidade de um evento não exceder um determinado valor.

De maneira geral, as séries de precipitação totais anuais se ajustam bem ao modelo de distribuição Normal, cuja função de densidade de probabilidade e a função de distribuição acumulada são dadas, respectivamente, pelas equações:

$$N(x) = \frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad (2.1)$$

$$p = F_N(x) = \frac{1}{\sigma \sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{(t-\mu)^2}{2\sigma^2}} dt \quad (2.2)$$

onde  $\mu$  = parâmetro de escala (mm);

$\sigma$  = parâmetro de forma (mm);

$x$  = limite superior da magnitude de precipitação (mm);

$p$  = probabilidade de não-excedência de  $x$ .

As séries mensais ou de menor período têm sido ajustadas com êxito à distribuição Gama incompleta (Varejão-Silva, 2006; Ávila, 1994), cuja função de densidade de probabilidade e a função de distribuição acumulada são dadas respectivamente pelas equações:

$$G(x) = \frac{1}{\beta^\alpha \Gamma(\alpha)} x^{\alpha-1} e^{-x/\beta} \quad (2.3)$$

$$P = F_G(x) = \frac{1}{\beta^\alpha \Gamma(\alpha)} \int_0^x t^{\alpha-1} e^{-t/\beta} dt \quad (2.4)$$

com  $\alpha > 0$ ,  $\beta > 0$ ,  $\Gamma(\alpha) > 0$  e  $G(x) = 0$  para  $x < 0$ , onde:

$\beta$  = parâmetro de escala (mm);

$\alpha$  = parâmetro de forma (adimensional);

$x$  = limite superior da magnitude de precipitação (mm);

$p$  = probabilidade de não-excedência de  $x$ ;

$(\Gamma)$  = função gama definida por:

$$\Gamma(\alpha) = \int_0^{\infty} t^{\alpha-1} e^{-t} dt \quad (2.5)$$

O ajuste dessas distribuições de probabilidades às distribuições de frequência, ou seja, a estimativa de seus parâmetros, pode ser realizada por diversos métodos: método dos momentos (tradicionais, momentos ponderados por probabilidades e momentos-L), método dos mínimos quadrados e método da máxima verossimilhança, cujas formulações podem ser encontradas em Lanna(2004); Hosking e Wallis (1997) e Clarke (2004);

Embora naturalmente a análise das frequências observadas e os ajustes às distribuições de probabilidade sejam realizados localmente, em condições físicas apropriadas é possível analisar conjuntamente vários postos de observação, com o propósito de melhoria das estimativas locais e, inclusive, redução da dimensionalidade das informações.

Em termos de precipitações, estas condições físicas estão intimamente relacionadas à sua variabilidade espacial.

### 2.3 Variabilidade espacial

A variabilidade espacial observada na precipitação deve-se, entre outros fatores, ao padrão de circulação atmosférica ao redor do globo, à interação das massas de ar com o relevo, à variação do balanço de radiação nas diferentes latitudes e à distância em relação ao mar, ou continentalidade.

O arranjo espacial das precipitações geralmente é avaliado através de mapas de isoietas, particularmente de precipitações médias. Entretanto, estes mapas podem ser elaborados para representar outras características das precipitações, como desvios com relação à média ou probabilidades de superarem um determinado valor, como pode ser apreciado no trabalho de Ávila (1994).

À parte a aleatoriedade local, e sendo mantidas as condições de relevo, é bastante presumível que, num dado raio de observação (onde a variação de latitude e de distância do mar possa ser considerada pequena), a distribuição de chuvas seja muito parecida, principalmente quando o local é caracteristicamente submetido a sistemas frontais de precipitação, que atingem grandes áreas ao mesmo tempo.

Por outro lado, o relevo, muitas vezes, impõe alterações nos padrões de comportamento, mesmo em distâncias relativamente pequenas, principalmente quando o regime de precipitações é suficientemente afetado pelos sistemas de natureza orográfica ou convectiva, ou mesmo pelos dois sistemas.

Em vista disso, existem situações onde seria mais prático, e até desejável, outras formas para sistematizar o conhecimento, com o objetivo de reduzir a dimensionalidade das informações.

Embora o mapeamento seja uma prática comum e muito útil para se observar cada comportamento em particular, perde um pouco de praticidade quando se deseja compreender um conjunto de características, como por exemplo, locais que apresentam mesma precipitação e mesma temperatura em determinadas épocas do ano.

Mesmo existindo ferramentas de geoprocessamento com esse propósito (Câmara et al., 2001), a alternativa mais indicada quando não há classes predefinidas e existe complexidade nas informações, é o uso de métodos multivariados de classificação de dados, como por exemplo, os métodos de análise de agrupamento (*cluster analysis*).

## 3 ANÁLISE REGIONAL

### 3.1 Aspectos gerais

O princípio de uma análise regional é agregar pontos com informações semelhantes a respeito de um fenômeno (função ou característica) de forma que possam ser tratados conjuntamente e as informações particulares de cada ponto possam auxiliar numa melhor compreensão do fenômeno.

Outra aplicação da análise regional é a possibilidade de expandir a informação para pontos não medidos, desde que possam ser considerados internos à região. Em hidrologia, esse procedimento é denominado regionalização hidrológica.

Na maioria das metodologias, o princípio básico é, a partir de um conjunto de observações, identificar regiões com semelhança de comportamento e ajustar um modelo de regressão da variável desejada, com base em variáveis explicativas do fenômeno.

Uma aplicação comum de análise regional é na análise de frequência de fenômenos naturais, mais comumente vazão e precipitação. De acordo com Davis e Naghettini (2001) a combinação de dados espacialmente disseminados, realizada numa análise regional de frequências, permite a obtenção de melhores estimativas de parâmetros e quantis.

De acordo com Tucci (2002), na análise regional de frequências, ou regionalização de funções estatísticas, três tipos básicos de procedimentos podem ser aplicados: regionalização dos parâmetros de uma distribuição estatística; dos valores correspondentes a um determinado tempo de retorno; e regionalização de uma curva de probabilidade adimensional e seu fator de adimensionalização.

O método de regionalização de parâmetros supõe que dentro da região existe um tipo de função teórica de distribuição estatística que ajusta bem todos os dados amostrados.

Assim, todas as séries da região são ajustadas a esse tipo de distribuição de probabilidades, com parâmetros locais  $\mu_i$  e  $\sigma_i$ . Estes parâmetros são utilizados para ajustar uma função de regressão cujas entradas são as variáveis explicativas e que é utilizado para estimar os parâmetros da distribuição em pontos sem dados.

Por outro lado, no método de regionalização de eventos associados a um determinado risco, cada série na região pode ser ajustada a uma distribuição diferente, determinando-se, a partir desta,

a magnitude do evento associada a um dado risco. Em seguida, é estabelecida uma regressão entre o evento e as variáveis explicativas. Para cada risco haverá uma regressão, que é considerada válida para todos os pontos dentro da região homogênea.

Por último, tem-se o método de regionalização de uma curva de probabilidade adimensional e seu fator de adimensionalização (ou fator de escala), o *index-flood*. Embora seja referente a cheias, o método e o termo *index-flood* têm uso consagrado em estudos de regionalização de frequências de qualquer tipo de variável hidrológica (Valverde, 2001).

Segundo Tucci (2002), este método é indicado para locais que possuem séries curtas e consiste na adimensionalização das curvas individuais de probabilidade com base no seu valor médio (fator de escala ou *index-flood*), reunião dos postos com a mesma tendência, estabelecendo-se, em seguida, uma curva adimensional regional média.

Uma vantagem de se estabelecer uma distribuição regional adimensional é que, a partir dos quantis adimensionais regionais, pode-se facilmente retornar aos quantis locais em escala verdadeira, ou redimensionalizá-los, simplesmente multiplicando-os pelo fator de escala local.

Como regra geral, para locais sem dados, o fator de escala local é estimado por uma função empírica ajustada por regressão.

Segundo Tucci (2002), as etapas de regionalização podem iniciar, para este método, por uma análise e seleção dos dados, seguida da determinação das curvas adimensionais locais e regional e finalizando com a regressão do fator de escala. A identificação de regiões homogêneas é citada como uma etapa posterior, indicada no caso de se verificar grande dispersão do conjunto de curvas individuais ou baixos coeficientes de correlação da equação de regressão.

De acordo com Hosking e Wallis (1997), porém, a identificação de regiões homogêneas é essencial ao bom desempenho do método *index-flood*, e por perceberem que os métodos de regionalização compreendem muitas decisões subjetivas, desenvolveram uma metodologia capaz de oferecer subsídios mais objetivos a este tipo de procedimento.

O método de análise regional de frequências, baseado em momentos-L (Hosking e Wallis, 1997), segue os princípios do *index flood*, com o diferencial de que, além das estimativas dos parâmetros da distribuição regional serem baseadas em momentos-L, é incorporado um conjunto de estatísticas auxiliares para diminuir a subjetividade das etapas da análise.

Os momentos-L constituem um conjunto de estatísticas derivadas dos momentos ponderados de probabilidade (MPPs) que, segundo seus autores, possuem vantagens teóricas em relação aos

momentos convencionais, sendo recomendáveis para o ajuste de várias distribuições, e robustos na presença de dados espúrios. Seus princípios e formulações podem ser encontrados em Hosking e Wallis (1997).

No método em questão, a análise regional de frequências segue as etapas de consistência dos dados, identificação de regiões homogêneas, escolha da distribuição regional e, finalmente, determinação de seus parâmetros (Davis e Naghettini, 2001). Em caso de se desejar obter estimativas em locais sem dados, pode-se incluir a etapa de modelagem do fator de escala, geralmente por regressão.

Cada uma das etapas acima será descrita, apontando-se os procedimentos indicados para cada uma delas, priorizando-se aspectos específicos mais relacionados ao estudo de precipitações totais.

### **3.2 Consistência dos dados**

De acordo com Bertoni e Tucci (2004), o primeiro passo antes de utilizar dados para tratamento estatístico é a identificação de erros nas séries. De maneira geral em um posto de observação podem ocorrer períodos sem informação ou com falhas de observação decorrentes de problemas nos aparelhos ou de erro humano.

Segundo Davis e Naghettini (2001) quando as condições de obtenção dos dados são alteradas ao longo do tempo, as séries podem apresentar tendências e não-estacionariedade, o que leva a amostra a ser considerada não-homogênea, não devendo ser utilizada para inferência estatística.

Os mesmos autores apresentam como técnicas para identificação de erros e heterogeneidade nas séries: a comparação dos totais pluviométricos mensais entre postos próximos ou entre um posto e postos vizinhos; a verificação de curvas de dupla-massa mensais/anuais (dupla acumulação) de um posto e a média de postos próximos; testes estatísticos para a verificação de independência, de homogeneidade e de pontos atípicos e a medida de discordância ( $D_i$ ) baseada em momentos-L.

Segundo Hosking e Wallis (1997), a medida de discordância tem por objetivo identificar pontos que apresentam estatísticas muito discrepantes em relação ao grupo, e consiste na relação entre os quocientes de momentos-L (coeficiente de variação-L (CV-L), assimetria-L e curtose-L), de forma que, num espaço tridimensional formado por estas três medidas, é considerado discrepante o ponto que se afasta muito do núcleo de concentração das amostras.

A pretensão, nesse caso, é diminuir a complexidade da investigação, que não necessitaria se estender a todos os postos, mas apenas aos de comportamento mais discrepante.

### **3.3 Regiões homogêneas**

#### **3.3.1 Conceituação**

Em Hidrologia, o termo regiões homogêneas está associado a regiões que possuem similaridade hidrológica. Para Lanna (1983), essa similaridade inclui fatores físicos, climáticos, biológicos, geológicos e efeitos antrópicos. Como há grande complexidade na consideração de todos esses fatores, o autor conceituou região homogênea como uma região na qual diversas características climáticas e fisiográficas teriam variabilidade mínima. A similaridade, nesse caso, seria observada com respeito aos fenômenos de maior interesse no processo hidrológico em estudo.

Para Tucci (2002), na regionalização hidrológica a homogeneidade é entendida como a semelhança na resposta das funções regionais obtidas.

Desta forma, a região homogênea não deve ser entendida necessariamente como um espaço geográfico contínuo (de 2 ou 3 dimensões), mas sim como uma região hiper-espacial, onde o número de dimensões é idêntico ao número de variáveis utilizadas da determinação da região.

Pode-se dizer que a identificação de regiões hidrológicamente homogêneas possui duas finalidades: uma descritiva, cujo objetivo é a caracterização hidrológica, e outra, como subsídio essencial na análise de frequência regional, especialmente quando se utiliza o método do *index-flood*.

De acordo com Tucci (2002), a regionalização hidrológica baseia-se no princípio de que existe similaridade de comportamento hidrológico entre alguns locais, e que o agrupamento desses locais em regiões hidrológicamente homogêneas permitiria uma interpolação ou extrapolação mais adequada e precisa da informação ao se aplicar a metodologia de regionalização escolhida.

Muitos autores consideram a identificação de regiões homogêneas como a etapa da regionalização que possui maior grau de dificuldade, por requerer muitas vezes julgamento subjetivo (Valverde, 2001; Jinguia e Rao, 2004). Para garantir uma identificação eficiente, medidas de homogeneidade e de discordância, baseadas em momentos-L, vêm sendo atualmente utilizadas (Pinto e Naghetinni, 1999; Dinpashoh et al., 2004; Jinguia e Rao, 2004; Rao e Srinivas, 2006).

A divisão de uma área de estudo em regiões homogêneas pode ser efetuada por vários métodos. Entre estes pode-se citar o método dos Resíduos (Bhaskar e O'Connor, 1989; Jingyi e Hall, 2004), a Análise de Componentes Principais (Nathan e McMahon, 1990; Dinpashoh et al., 2004; Marques, 2005 e Lopes, 2006) e os métodos de Análise de Agrupamento, sendo este último o mais aplicado.

Os métodos de Análise de Agrupamento, que são os métodos mais indicados para a classificação, têm sido amplamente empregados para reconhecimento de padrões nas mais diversas áreas da ciência, incluindo inúmeras aplicações em Engenharia, e mais especificamente em Hidrologia e Meteorologia, para obtenção de regiões homogêneas, tanto de vazões (Burn, 1989; Nathan e McMahon, 1990; Hall e Minns, 1999; Rao e Srinivas, 2006), como de precipitação (Valverde, 2001; Dinpashoh et al., 2004; Keller Filho et al. 2005; Lin e Chen, 2006; Lyra et al., 2006) e de temperatura (Becker, 1992).

### **3.3.2 Análise de agrupamento**

#### **3.3.2.1 Métodos hierárquicos e de partição**

O mecanismo básico dos métodos de agrupamento está na medida de similaridade. Nathan e MacMahon (1990) apresentam várias medidas de distância entre os grupos que podem ser adotadas, tais como a distância Euclidiana (simples e quadrática), o cosseno dos vetores das variáveis, a soma das diferenças absolutas de cada variável e a máxima diferença absoluta das variáveis. Baseando-se na medida de distância estabelecida, o método de agrupamento deverá ser capaz de unir os elementos próximos em um mesmo grupo.

De acordo com Carlantonio (2001), os métodos mais tradicionais de agrupamento são os métodos de partição e os métodos hierárquicos, embora existam ainda métodos baseados em grades, em densidade e em modelos, como é o caso do agrupamento através de Redes Neurais utilizado por Lin e Chen (2006) para regionalização de precipitações em Taiwan.

Os dois primeiros métodos serão detalhados em seguida, dada a sua maior importância. Os demais métodos podem ser consultados em Andrade (2004).

##### **a) Métodos hierárquicos**

Os métodos hierárquicos podem ser aglomerativos ou divisivos e se caracterizam por estabelecerem um diagrama de árvore ou dendograma composto pela sucessão hierárquica dos  $n$  grupos formados, tal que  $n$  é o número total de indivíduos do conjunto de dados (Figura 3.1).

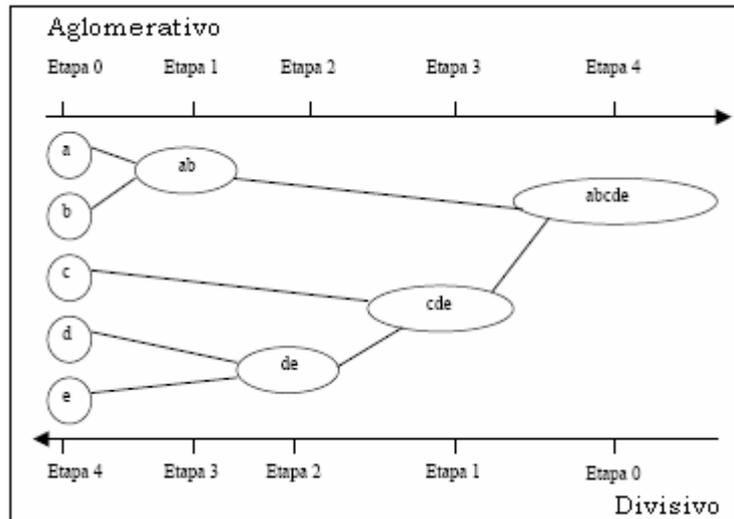


Figura 3.1: Métodos Hierárquicos. (Adaptado de Carlanonio (2001)).

Na abordagem aglomerativa, o processo de agrupamento inicia-se com  $n$  grupos unitários formados por cada um dos elementos  $e$ , a cada iteração, os dois grupos mais próximos se unem até que todos os indivíduos constituam um único grupo.

Já a abordagem divisiva executa o processo contrário, partindo de um único grupo composto por todos os elementos, vai se dividindo a cada iteração até os  $n$  grupos serem formados. Em qualquer das abordagens o processo de formação dos grupos pode ser interrompido antes do fim, desde que seja estabelecido o número de grupos desejado ou um critério de parada.

Em geral a abordagem aglomerativa é mais utilizada, devido à maior simplicidade de programação. Os algoritmos existentes se distinguem pelo método de ligação entre os elementos para formação dos grupos, bem como pela adoção de medidas de similaridade diferentes.

Andrade (2004) e Nathan e McMahon (1990) apresentam como métodos de ligação a Ligação Simples (vizinho mais próximo), Ligação Completa (Vizinho mais distante), Ligação Média (distância média entre todos os elementos de um grupo com o outro), Método do Centróide (distância entre os centróides de cada grupo) e Método de Ward, baseado na menor variância entre os grupos.

Destes, o Método de Ward tem sido mais largamente utilizado para aplicações em Hidrologia (Silva et al., 2001; Dinpashoh et al., 2004; Jinguia e Hall, 2004; Keller Filho et al., 2005; Souza et al., 2005; Lyra et al., 2006).

De acordo com Andrade (2004), a qualidade dos métodos hierárquicos puros pode ser prejudicada pela sua inabilidade interna de executar ajustes após uma fusão ou divisão terem sido executadas.

## b) Métodos de partição

Os métodos de partição supõem que o número de grupos  $k$  é um fator conhecido. Desta forma, são escolhidos  $k$  elementos (geralmente aleatórios) que constituem o centro de cada grupo. Em seguida, os elementos são alocados ao grupo mais próximo, ou seja, àquele em que a distância com relação ao centro seja mínima. Após a primeira alocação, um novo centro de cada grupo é determinado. A cada iteração os elementos são alocados para o centro mais próximo até que se chegue a uma alocação ótima, geralmente considerada como o menor somatório dos erros entre os pontos de cada grupo e suas médias (Novaes, 2001 e Carlantonio, 2001).

No método tradicional, conhecido como *k-means*, cada novo centro é estabelecido como a média dos elementos alocados no grupo. Em outra abordagem denominada *k-medoid*, o centro do grupo passa a ser o elemento cuja distância em relação à média do grupo é mínima.

De acordo com Andrade (2004), os métodos de partição são mais eficientes que os hierárquicos, uma vez que consistem no refinamento das posições dos elementos para um número de grupos pré-estabelecido.

O método de agrupamento difuso *Fuzzy C-Means (FCM)* pode ser considerado uma generalização do método *K-means*. Com a inserção da lógica difusa os elementos da base de dados passam a pertencer concomitantemente a todos os grupos, embora com graus diferentes de pertinência, que são proporcionais à sua distância em relação ao centro de cada grupo. Trata-se de uma generalização, porque, teoricamente, adotando-se apenas a máxima pertinência para cada ponto, se chegaria a um resultado semelhante ao do método *K-means* tradicional.

Esta característica peculiar, de atribuir graus de pertinência dos pontos a cada grupo, permite considerar ambigüidade e sobreposição dos grupos formados, aproveitando-se mais eficientemente as informações disponíveis. Por conta deste fator, o método vem sendo utilizado na determinação de regiões homogêneas para estudo de vazões por Hall e Minns (1999), Jingyi e Hall (2004), e Rao e Srinivas (2006); de solos em Amini et al. (2005), Goktepe et al. (2005); e de precipitações em Kulkarni e Kripalani (1998).

Kulkarni e Kripalani (1998) investigam os padrões temporais das chuvas de monções na Índia e explicitam como vantagem do uso do FCM em relação ao *K-means*, a existência dos graus de pertinência, cuja análise permite identificar os elementos que mais caracterizam cada grupo e as relações dos grupos entre si, revelando os que são mais assemelhados (por conterem elementos com graus de pertinência semelhantes) e os grupos totalmente opostos (cujos elementos possuem disparidade total dos graus de pertinência).

### 3.3.2.2 Número de grupos

Segundo Novaes (2002), o agrupamento pode ser visto como uma classificação onde as classes não foram definidas *a priori*, sendo a determinação do número de classes um problema ainda de difícil solução.

De maneira geral não há como determinar o número de grupos anteriormente à aplicação do método. Quando se utilizam métodos hierárquicos, geralmente o número de grupos é escolhido a partir da análise visual do dendograma, o que implica certo grau de subjetividade, muito embora a visualização da forma como os elementos se agrupam possa auxiliar na tomada de decisão.

Com os métodos de partição, a prática comum é aplicar o agrupamento para vários números de grupos e medir a qualidade dos resultados através dos denominados índices de validade do agrupamento.

Esses índices geralmente são baseados no critério de que o agrupamento ótimo é aquele cujas distâncias entre os elementos de um mesmo grupo são mínimas e as distâncias entre os centros de cada grupo são máximas.

Em geral o valor extremo (máximos ou mínimos, a depender do índice) é indicador do número ótimo de grupos. Entretanto, considerar a existência de um único número ótimo de grupos nem sempre é aceitável, uma vez que, a depender dos objetivos da classificação, detalhamentos maiores ou menores podem ser requeridos, o que aumenta ou diminui a faixa de similaridade. Nestes casos opta-se por extremos locais dentro da faixa desejada, uma vez que estes extremos são considerados bons agrupamentos. (Wu e Yang, 2005).

Alguns métodos utilizados para validar os resultados do agrupamento difuso são apresentados neste trabalho, enquanto que aqueles dirigidos a outros métodos de agrupamento podem ser encontrados em Andrade (2004).

### 3.3.3 Testes de homogeneidade

Embora os índices de validade tragam indícios da qualidade do agrupamento em si, não há segurança de que as regiões formadas sejam homogêneas em relação aos objetivos da análise.

Rao e Srinivas (2006) afirmam que quando não são utilizadas as estatísticas das séries, o agrupamento geralmente resulta em regiões não estatisticamente homogêneas e que quando as estatísticas fazem parte das variáveis agrupadas, os grupos resultantes poderiam ser apenas aparentemente homogêneos.

Por conta dessas observações, os autores recomendam a realização de testes de homogeneidade, mais especificamente as medidas de discordância ( $Di$ ) e de heterogeneidade ( $H$ ) baseadas em momentos-L (Jingyi e Hall, 2004; Rao e Srinivas, 2006).

Nesse caso, a medida de discordância traria indícios dos postos de comportamento mais discrepantes e, portanto, os primeiros candidatos à mudança ou exclusão de seus respectivos grupos, para torná-los mais homogêneos.

Hosking e Wallis (1997) comentam que muitos autores têm buscado medidas de verificação de homogeneidade que, de maneira geral, se baseiam na idéia de que, se uma região é homogênea, seu comportamento seria pouco diferente daquele encontrado em uma região virtual efetivamente homogênea.

Esse comportamento é expresso em termos da dispersão, dos valores locais em relação ao regional, de alguma característica da distribuição de frequência que pode ser o evento com tempo de retorno de 10 anos, o coeficiente de variação, ou algum dos quocientes-L (ou combinação deles).

A medida de heterogeneidade ( $H$ ), proposta por Hosking e Wallis (1997), quer certificar, para uma determinada região, se a dispersão dos quocientes-L amostrais (geralmente utiliza-se apenas o CV-L), é ou não equivalente àquela dispersão encontrada em uma região virtual que retrata a região real em termos de número e tamanho de séries.

Nesse caso, a região virtual e homogênea é obtida através da geração de séries aleatórias com número e tamanho das séries encontradas na região real. Os autores recomendam que as séries geradas sigam uma distribuição Kappa, cujos parâmetros são derivados dos momentos -L regionais da região real. A recomendação da distribuição Kappa é uma tentativa de tornar o teste menos comprometedor, já que essa distribuição é menos restritiva.

Para garantir a robustez do teste, é recomendado simular em média 500 regiões virtuais. Tanto para a região real, quanto para cada região virtual deve-se calcular a dispersão ( $V$ ) entre o quociente-L regional ( $t_R$ ) e o quociente-L em cada local ( $t_i$ ), ponderando pelo tamanho de cada série, conforme a formulação:

$$V = \frac{\sum_{i=1}^N n_i (t_i - t_R)^2}{\sum_{i=1}^N n_i} \quad (3.1)$$

Para determinação de  $H$ , calcula-se a média ( $\mu V_s$ ), e ao desvio padrão ( $\sigma V_s$ ) das dispersões nas regiões simuladas e compara-se com a dispersão na região real pela fórmula:

$$H = \frac{(V - \mu V_s)}{\sigma V_s}, \quad (3.2)$$

Hosking e Wallis (1997) consideram que se  $H < 1$  a região é aceitavelmente homogênea, se  $1 \leq H < 2$ , a região é possivelmente heterogênea e se  $H \geq 2$  a região é definitivamente heterogênea. De acordo com os autores, estes critérios são um tanto arbitrários e não necessitam ser encarados com rigor, pois mesmo regiões moderadamente heterogêneas podem produzir estimativas de quantis mais confiáveis que as estimativas locais.

Entretanto, a subdivisão e ou retirada de alguns pontos pode reduzir a heterogeneidade de uma região. Nesse caso, os critérios arbitrados podem servir de guia na redefinição das regiões, considerando-se que se  $1 \leq H < 2$  essa redefinição pode ser vantajosa, já para  $H > 2$  ela muito provavelmente será benéfica.

### 3.3.4 Escolha das variáveis e ajustes no agrupamento

De acordo com Dinpashoh et al. (2004), não há uma regra geral para determinar o número e o tipo de atributos a serem considerados na caracterização da precipitação. Em seu trabalho, cuja finalidade era a seleção de variáveis para a regionalização da precipitação anual no Irã, os autores utilizaram *Procrustes Analysis* (Krzanowski, 1987 apud Dinpashoh et al., 2004) para extrair as variáveis mais representativas de um total de 57 variáveis candidatas que incluía as coordenadas geográficas, os totais anuais e mensais, o número de dias com chuva, precipitação máxima diária, médias sazonais, etc. Do total foram selecionadas 12 variáveis: precipitação total anual; precipitação média dos meses secos; média dos meses secos do verão; média do inverno; média de julho, de outubro e de novembro; número de dias de chuva de agosto de setembro e de outubro; precipitação máxima diária de outubro e de novembro.

Nathan e McMahon (1990) também admitem que a escolha das variáveis utilizadas na determinação do grau de similaridade entre diferentes bacias (no caso da regionalização de vazões) é um problema comum em estudos dessa natureza, e aponta não só a necessidade de seleção, mas, possivelmente, a ponderação de determinadas variáveis, de acordo com sua relevância no problema.

Já Keller Filho et al. (2005), com o objetivo de identificar regiões homogêneas quanto à distribuição de probabilidades pentadiárias de chuvas no território brasileiro, utilizaram como variáveis classificatórias os parâmetros de posição, escala e forma das distribuições empíricas e a

proporção de pântadas secas. Becker et al. (1992) e Lyra et al. (2006), para definição de regiões homogêneas no Rio Grande do Sul e na Venezuela, respectivamente, utilizaram unicamente as precipitações médias mensais para definir regiões com o método de agrupamento hierárquico de Ward, como fizeram também Marques (2005) e Lopes (2006) no Rio Grande do Sul, embora tenham utilizado o Método das Componentes Principais.

Embora, claramente, não seja seguida uma regra para a seleção das variáveis que farão parte do agrupamento, Hosking e Wallis (1997) não recomendam a utilização das características das séries em análise, como desvios ou coeficientes de variação, para este fim.

Segundo eles, a não utilização dessas características permitiria a independência de testes de homogeneidade baseados no coeficiente de variação das séries amostrais.

Entretanto, os autores admitem a possibilidade de que as regiões assim obtidas venham a ser heterogêneas, sugerindo que sejam feitas adaptações no resultado do agrupamento para que sejam alcançadas as condições de homogeneidade, respeitando prioritariamente a coerência física das regiões. Os ajustes mencionados são do tipo: mover um ou alguns pontos de uma região para outra; retirar um ou alguns pontos da análise; subdividir regiões; distribuir os pontos de uma região entre outras; combinar regiões; obter mais dados e redefinir regiões.

Ou seja, ao fim da operação a configuração do agrupamento estará provavelmente desvirtuada e o teste não terá sido tão independente, tendo em vista que foi utilizado para obter a configuração final das regiões.

A tendenciosidade no agrupamento devida ao uso de estatísticas das séries pode ser atenuada ao se incluir características físicas e outras variáveis independentes na matriz de agrupamento, já que o resultado seria direcionado a atender todas as características ao mesmo tempo.

### **3.4 Determinação da curva regional**

Diversos modelos teóricos de distribuição de frequência podem se adaptar a um dado processo. Sua escolha, porém, pode ser baseada em experiências antecedentes para o mesmo tipo de fenômeno, plotando-se os dados observados e optando-se pelo modelo com forma mais semelhante, visualmente, ou por tentativa e erro de um conjunto de distribuições candidatas, selecionando-se a que apresentar melhor qualidade de ajuste.

Retomando-se o princípio do *index-flood*, numa região formada por  $N$  postos  $i$ , cada um deles com uma série de observações  $Q_{ij}$  de tamanho  $n_i$ , com  $i=1, 2, \dots, N$  e  $j=1, 2, \dots, n_i$ , tem-se que  $F_i(Q_{ij})$  é a função de distribuição acumulada da distribuição de freqüências no posto  $i$  e a sua inversa  $Q_i(F)$ ,  $0 < F < 1$ , é a função quantil da distribuição de freqüência de  $i$ , que numa região homogênea é dada por:

$$Q_i(F) = e_i q(F), \quad i = 1, 2, \dots, N. \quad (3.3)$$

onde  $q(F)$  é a função de quantis adimensionais da distribuição de freqüência regional e  $e_i$  o fator de escala do ponto  $i$ .

Note-se, entretanto, que  $q(F)$  será a melhor aproximação das distribuições adimensionais locais  $q_i(F) = Q_i / e_i$ .

Considerando-se que o tipo de distribuição de freqüências é conhecido (é a distribuição candidata) e seus  $\theta_p$  parâmetros não o são, tem-se que a função de quantis adimensionais regionais é definida por  $q(F; \theta_1, \dots, \theta_p)$ . Lembrando que se  $e_i$  é estimado pela média de  $Q_i$ , o que geralmente é adotado, então  $q(F)$  tem média unitária.

De acordo com Tucci (2002), a  $q(F)$  pode ser estimada pela média das  $q_i(F)$  existentes na região, sendo que a idéia é obter uma série de quantis adimensionais determinados pela média dos quantis adimensionais locais que caem dentro de cada intervalo  $j$  da variável reduzida de uma distribuição candidata, que é predeterminado e de mesmo tamanho. Ou seja,  $q_j(F) = \bar{q}_{ij}(F)$ , com  $j = 1, \dots, M$ , sendo  $M$  o total de intervalos. A partir da distribuição média, ajusta-se o modelo teórico graficamente, pelo método dos momentos ou pelo método da máxima verossimilhança.

Outra possibilidade citada pelo mesmo autor é a de ajustar a distribuição estatística candidata (por qualquer dos métodos) ao conjunto de todos os elementos de cada série adimensional da região, reunidos em uma amostra única. Esse método é conhecido como método da estação-ano.

Hosking e Wallis (1997) não recomendam o método da estação-ano, alegando que a união das observações adimensionais de séries de tamanhos diferentes não constituiria uma amostra de variáveis igualmente distribuídas e que provavelmente os eventos de vários locais observados em um mesmo período, se reunidos, não atenderiam à hipótese de independência dos eventos.

Uma outra possibilidade de estimar  $q(F)$ , segundo Hosking e Wallis (1997), é considerar que esses parâmetros regionais ( $\hat{\theta}_k^R$ ) sejam as razões de momentos-L da distribuição regional estimadas pela média ponderada pelo tamanho da série ( $n_i$ ) dos  $\hat{\theta}_k^i$  parâmetros das distribuições de quantis

adimensionais locais (nesse caso as razões de momentos-L amostrais de cada ponto). A estimativa de  $\hat{\theta}_k^R$  é formalmente dada por:

$$\hat{\theta}_k^R = \frac{\sum_{i=1}^N n_i \hat{\theta}_k^i}{\sum_{i=1}^N n_i}, \quad k = 1, \dots, p. \quad (3.4)$$

Conhecidas as razões de momentos regionais, a distribuição candidata é ajustada pelo método de ajuste baseado em momentos-L.

Um modelo com resultados similares, citado pelos mesmos autores, considera que a partir da equação 3.3 é possível determinar iterativamente os parâmetros da distribuição de frequências de  $q(F)$ , utilizando-se o método da máxima verossimilhança, de forma que esses parâmetros atendam ao máximo e ao mesmo tempo as  $N$  equações 3.3 de cada região, levando-se em conta que  $e_i$  e  $Q_i$ , são conhecidos. No entanto, Hosking e Wallis (1997) acreditam que a estimativa baseada em momentos-L é mais simples de ser obtida.

Em qualquer dos procedimentos utilizados para o ajuste do modelo probabilístico recomenda-se o uso de testes para medir a qualidade do ajuste, conhecidos como testes de aderência.

### 3.5 Qualidade do ajuste da curva regional

Os testes mais conhecidos são o gráfico quantil-quantil, o teste Qui-Quadrado ( $\chi^2$ ) e o teste de Kolmogorov-Smirnov. Hosking e Wallis (1997), porém, apresentam um teste específico denominado medida de aderência ( $Z$ ), que, simplificada, sugere que a diferença entre a curtose-L da distribuição candidata e a curtose-L regional dividida pelo desvio entre as curtoses-L regionais e a curtose-L obtida em várias regiões simuladas com as mesmas características daquela que se quer ajustar, deve ser o mais próxima possível de zero, admitindo-se como limite  $|Z| \leq 1,64$ .

Entre os testes tradicionais, Costa Neto (1977) e Varejão-Silva (2006) afirmam ser o teste de Kolmogorov-Smirnov mais eficiente que o Qui-quadrado e este primeiro tem sido largamente utilizado em meteorologia para verificar a qualidade dos ajustes das distribuições teóricas, como em Cargnelutti et al. (2005) e Matzenauer et al. (2003) para testar a aderência das séries decendiais de precipitação à distribuição Gama, e em Ávila (1994) e Araújo et al. (2001), para testar a aderência tanto da distribuição Gama (para período mensal), quanto da distribuição Normal (para período anual).

Sendo  $F(x)$  o modelo de distribuição de probabilidade acumulada a ser testado e  $G(x)$  a função empírica de distribuição acumulada da amostra, o teste de Kolmogorov-Smirnov constitui-se da determinação da máxima diferença absoluta entre as probabilidades estimadas pelas funções  $F$  e  $G$ , dada por:

$$dm = \max | F(x) - G(x) |, \quad (3.5)$$

e sua comparação com um valor crítico que é função do nível de significância ( $\alpha$ ) e do tamanho da amostra. (Costa Neto, 1977).

A hipótese nula é de que as distribuições  $F$  e  $G$  podem ser consideradas iguais, e será rejeitada sempre que  $dm$  for maior que o valor crítico que rege o teste.

Note-se pela tabela 3.1 que quanto maior o nível de significância ( $\alpha$ ) e o tamanho da amostra, menor é o valor crítico e, conseqüentemente, maior é a exigência do teste, uma vez que se  $dm$  for maior que o valor crítico, a hipótese nula é rejeitada.

Tabela 3.1. Valor crítico do teste de aderência de Kolmogorov-Smirnov.

Elementos da Amostra(N)	Nível de Significância ( $\alpha$ )			
	0,2	0,1	0,05	0,01
10	0,322	0,368	0,400	0,486
11	0,307	0,352	0,391	0,468
12	0,295	0,338	0,375	0,450
13	0,284	0,325	0,361	0,433
14	0,274	0,314	0,349	0,418
15	0,266	0,304	0,338	0,404
16	0,258	0,295	0,328	0,391
17	0,250	0,286	0,318	0,380
18	0,244	0,278	0,309	0,370
19	0,237	0,272	0,301	0,361
20	0,231	0,264	0,294	0,352
25	0,210	0,240	0,264	0,320
30	0,190	0,220	0,242	0,290
35	0,180	0,190	0,230	0,270
Fórmula				
Assintótica	$1,07/N^{1/2}$	$1,22/N^{1/2}$	$1,36/N^{1/2}$	$1,63/N^{1/2}$

Fonte: Massey (1980) apud Varejão-Silva (2006)

Verificou-se ainda que alguns autores, como Cargnelutti et al. (2005) e Matzenauer et al. (2003), consideraram os modelos que eram atendidos pelo teste KS com nível de significância de 20%, como modelos bem ajustados, já que este nível de significância é mais exigente.

Informações sobre testes gráficos e o teste Qui-Quadrado ( $\chi^2$ ) podem ser consultadas em Costa Neto (1977).

### **3.6 Estimativa do fator de escala local**

Como regra geral, o fator de escala é tomado como a média da distribuição. Desta forma, para obter os quantis locais partindo de uma distribuição regional adimensional, basta multiplicar os quantis adimensionais regionais pela média da série de dados observados no local de interesse.

Tradicionalmente, a obtenção do fator de escala para locais sem dados é realizada através de modelos de regressão, utilizando as características físicas dos pontos, conforme se verifica em Valverde (2001) e Davis e Naghetinni (2001).

Quando a variável objeto da análise pode ser espacialmente mapeada, como é o caso de precipitações totais anuais ou mensais, é possível estimá-la por mapeamentos já existentes (Hosking e Wallis, 1997), ou aplicar a metodologia da elaboração desses mapeamentos para realizar a estimativa (o método do inverso do quadrado da distância, por exemplo).

## 4 AGRUPAMENTO DIFUSO

### 4.1 Conceitos gerais

Apesar dos primeiros conceitos da lógica difusa ("Fuzzy Logic") datarem de 1965, vindos da teoria dos conjuntos difusos, apresentada por Zadeh (1965), e serem considerados relativamente novos, sua utilização tem se difundido aceleradamente em aplicações de sistemas de controle nas mais diversas áreas do conhecimento. (Pedrollo, 2000; Barbalho, 2001)

Em sua primeira publicação, Zadeh (1965) estende os conceitos da lógica booleana, generalizando a teoria clássica dos conjuntos, de forma que a pertinência de um elemento a um conjunto ganha o caráter de uma função de pertinência, que varia no intervalo [0 1], em substituição ao padrão binário 0 ou 1 para a condição do elemento não pertencer ou pertencer ao dado conjunto, que é próprio dos conjuntos clássicos.

Para ilustrar essa diferença, a Figura 4.1(a) apresenta a configuração de um conjunto clássico "A" formado pelo intervalo [4;6] de uma variável  $x$  definida num domínio (ou universo de discurso)  $X$ , havendo pela lógica booleana apenas dois graus de pertinência possíveis  $\mu=0$  e  $\mu=1$ . Na Figura 4.1 Diferenças entre um conjunto clássico (a) e um conjunto difuso (b). O conjunto difuso "D" representa o mesmo intervalo, no entanto, os valores externos aos limites vão gradativamente deixando de pertencer a "D", obedecendo a uma função de pertinência que, neste caso, pode ser formalmente escrita conforme a equação (4.1), havendo, no entanto, uma infinidade de funções que podem descrever o comportamento de um conjunto difuso.

$$\mu(x) = \begin{cases} 0, & \text{se } x \leq 2 \text{ ou } x \geq 8 \\ \frac{x}{2} - 1, & \text{se } 2 < x < 4 \\ 4 - \frac{x}{2}, & \text{se } 6 < x < 8 \\ 1, & \text{se } 4 \leq x \leq 6 \end{cases} \quad (4.1)$$

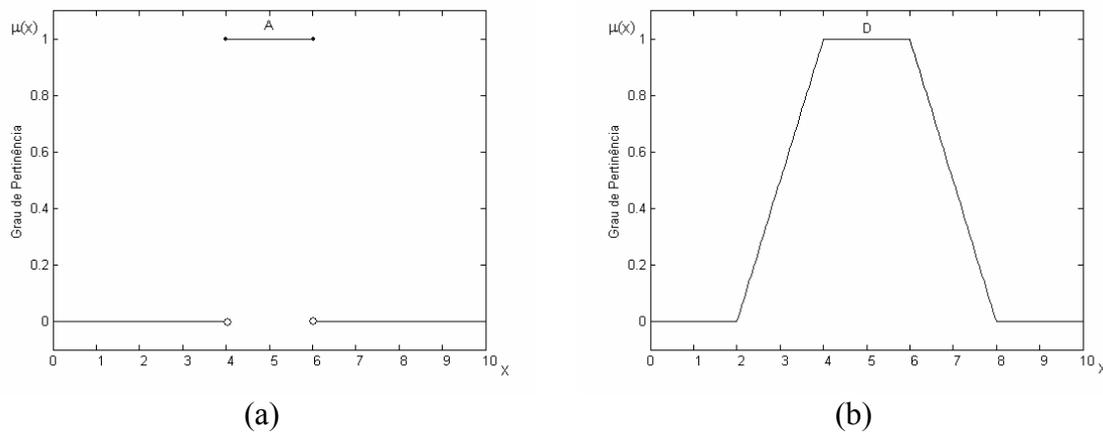


Figura 4.1 Diferenças entre um conjunto clássico (a) e um conjunto difuso (b).

O conceito de conjunto difuso permite, então, a representação da ambigüidade presente no processo humano de classificação dos elementos existentes no mundo real. Para Zadeh (1965), a forma difusa de classificação exerce um papel muito importante na maneira humana de pensar, particularmente no que diz respeito aos padrões de reconhecimento, comunicação e abstração.

Com esse novo conceito lógico, técnicas e metodologias tradicionais têm sido adaptadas ou reelaboradas sob a ótica dos conjuntos difusos, destacando-se os métodos de agrupamento multivariado e os sistemas especialistas.

## 4.2 Método de agrupamento *Fuzzy C-Means*

Os métodos de agrupamento multivariado possuem a capacidade de classificar uma série de dados de uma característica particular, considerando a influência de diversas outras características ou variáveis ao mesmo tempo. No entanto, nos métodos tradicionais os grupos formados possuem fronteiras rígidas ou bem definidas, desprezando a possibilidade de compartilhamento das informações entre os grupos na região de fronteira. Para inserir estes aspectos de ambigüidade e compartilhamento de informações entre grupos, Bezdek (1981) implementou um método de agrupamento difuso, o chamado método “*Fuzzy C-Means*” (FCM).

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

Partindo desse princípio, o método de classificação “*Fuzzy C-Means*”, proposto por Bezdek (1981), consiste em agrupar  $n$  pontos  $x_i$ , com  $i=1,2,\dots,n$ , de um espaço  $m$ -dimensional composto por  $M$  atributos ou características que direcionarão o agrupamento em  $c$  regiões difusas  $A_k$ , sendo  $k=1,2,\dots,c$ , de forma que cada ponto  $x_i$  pertence a todos os grupos  $A_k$  com diferentes níveis de pertinência  $\mu_{ik}$  que compõem a matriz de pertinência  $U=\mu_{ik}, \forall i,k$ . Deve-se respeitar ainda a condição de que não são formados grupos vazios e que o somatório dos  $\mu_{ik}$  de cada elemento  $i$  seja unitário. Estes princípios estão equacionados como segue :

$$\mu_{ik} = \mu_{A_k}(x_i) \in [0,1]$$

$$\text{onde } \sum_{k=1}^c \mu_{ik} = 1 \text{ para todo } i \quad (4.2)$$

$$0 < \sum_{i=1}^n \mu_{ik} < 1$$

Cada grupo formado possui um protótipo, centro ou centróide de classe  $v_k$  em torno do qual se agrupam os pontos. Quanto mais próximos os pontos se encontram do protótipo  $v_k$ , cujas coordenadas são dadas pela equação (4.3), maior é o grau de pertinência  $\mu_{ik}$  do ponto  $x_i$  ao grupo  $A_k$ , uma vez que  $\mu_{ik}$  resulta da relação entre a distância do ponto  $x_i$  ao centro grupo  $A_k$  e o somatório das distâncias de  $x_i$  a todos os centros, conforme a formulação (4.4).

$$v_{kj} = \frac{\sum_{i=1}^n (\mu_{ik})^m x_{kj}}{\sum_{i=1}^n (\mu_{ik})^m}, \quad j=1,2,\dots,M \quad (4.3)$$

$$\mu_{i,k} = \frac{d_{ik}^{-2/(m-1)}}{\sum_{k=1}^c (d_{ik}^{-2/(m-1)})} \quad (4.4)$$

A medida de proximidade adotada no método é a distância Euclidiana,  $d_{ik}$  (4.5), entre cada um dos pontos,  $x_i = [x_{i1}, x_{i2}, \dots, x_{im}]$ , e o protótipo (centróide),  $v_k = [v_{k1}, v_{k2}, \dots, v_{km}]$ , para cada grupo  $A_k$ .

$$d_{ik} = \sqrt{\sum_{j=1}^M (x_{ij} - v_{kj})^2} \quad (4.5)$$

Os pontos então se agrupam com o objetivo de resultarem distâncias mínimas entre o protótipo e os elementos do grupo e distâncias máximas dos protótipos entre si. Assim, o algoritmo deve minimizar a função objetivo dada por:

$$F_{\text{obj}} = \sum_{i=1}^n \sum_{k=1}^c (\mu_{ik})^m (d_{ik})^2 \quad (4.6)$$

Na equação (4.6), o coeficiente de difusividade ( $m$ ) é um parâmetro que controla o nível de difusividade no processo de classificação (Hall e Mins, 1999). Desta forma, para  $m=1$  os agrupamentos apresentam limites rígidos equivalentes aos do agrupamento booleano e, à medida que o valor cresce, os limites se tornam mais difusos, sendo indicado por Ross (1995) valores entre  $1,25 \leq m \leq 2$ .

Rao e Srinivas (2006) observam que, nas aplicações do FCM, é comum o uso indiscriminado de  $m=2$ , como se pode constatar nos trabalhos de Özekan et al. (1998) e Hall e Minns (1999).

Segundo eles, porém, dependendo das características dos dados, esse valor pode não ser adequado, gerando agrupamentos muito difusos, com pontos pertencendo a todas as classes e, por conseguinte, sem que o método consiga capturar a individualidade de cada grupo. Por outro lado, valores de  $m$  muito próximos a 1 podem acarretar em agrupamentos muito rígidos, em que os graus de pertinência seriam pouco sensíveis às semelhanças entre os grupos.

O efeito do valor de  $m$  pode ser avaliado observando-se o comportamento dos graus de pertinência máximos ( $\mu_{max}$ ) para todos os pontos, de forma que, para um número de grupos qualquer, o histograma dos  $\mu_{max}$  não deve se concentrar em torno de 1 (grupo muito rígido) nem em torno de  $1/c$  (grupo muito difuso), onde  $c$  é o número de grupos.

Assim, investigando-se o coeficiente de difusividade mais conveniente, é possível que sejam obtidos grupos mais bem definidos.

### 4.3 Algoritmo de agrupamento

O algoritmo de agrupamento iterativo consta das seguintes etapas:

1. Escolha do número de classe ( $c$ ) e do coeficiente de difusividade ( $m$ );
2. Inicialização da matriz de pertinência  $U^{(0)}$  e a cada passo de iteração  $p$  faz-se:
3. Cálculo dos protótipos do agrupamento  $v_{(kj)}^{(p)}$ , usando a equação (4.3);
4. Atualização da matriz de pertinência  $U^{(p)}$ , em que cada elemento é dado pela equação (4.4)
5. Comparação das diferenças dos elementos das matrizes  $U^{(p)}$  e  $U^{(p-1)}$  com o erro tolerado. Se o erro for maior que o tolerado, repetem-se os passos a partir de 3, caso contrário considera-se que o algoritmo tenha convergido.

De maneira geral, a matriz de partição é inicializada aleatoriamente. Por conta disso Rao e Srinivas (2006) apontam para a possível sensibilidade do algoritmo a essa aleatoriedade, levando-o a convergir para ótimos locais, o que significa que está sujeito a apresentar diferentes resultados mesmo quando conservados os demais parâmetros.

### 4.4 Padronização e ponderação das variáveis

Devido ao uso da distância Euclidiana como medida de similaridade, algumas variáveis, a depender da escala de valores utilizada, podem influenciar mais diretamente o resultado do agrupamento. Em detrimento disso a orientação de alguns autores (Burn, 1989; Hall e Minns, 1999 e Jingyi e Hall, 2004) é de que se adote a padronização (escalamento) das variáveis, uniformizando seus efeitos em termos de distância.

Duas possibilidades de padronização dos dados são apresentadas por Hall e Minns (1999), uma em que os dados são modificados para apresentarem média zero e variância unitária, de forma que cada elemento é modificado pela fórmula:

$$x_{ij} = \frac{y_{ij} - \bar{y}_j}{\sigma_j}, \quad (4.7)$$

onde  $x_{ij}$  é o elemento escalonado  $i$  da variável  $j$ ,  $y_{ij}$  o valor real do elemento  $i$  da variável  $j$  e  $\bar{y}_j$  e  $\sigma_j$  são a média e o desvio padrão amostrais da variável  $j$ .

E a outra em que os dados variam no intervalo  $[0,1]$ ,

$$x_{ij} = \frac{y_{ij} - y_{i(\min)}}{y_{i(\max)} - y_{i(\min)}}. \quad (4.8)$$

Na equação (4.8)  $x_{ij}$  é o elemento escalonado  $i$  da variável  $j$ ,  $y_{ij}$  o valor real do elemento  $i$  da variável  $j$  e  $y_{i(\min)}$  e  $y_{i(\max)}$  são os valores máximo e mínimo da variável  $j$ .

Carvalho e Pedrollo (2005) não detectaram diferenças no resultado dos agrupamentos utilizando um ou outro método, indicando que a escolha por um deles fica a critério do analista.

Posteriormente à padronização, Nathan e McMahon (1990) consideraram a possibilidade de ponderação das variáveis para direcionamento do agrupamento. Segundo os autores, nos casos em que as variáveis consideradas na matriz de dados são apenas explicativas de um fenômeno que se quer prever, pode-se utilizar análise de regressão ou de correlação dessas variáveis com o fenômeno, para nortear a ponderação.

A intenção da ponderação é elevar o valor das variáveis que se mostram mais representativas do fenômeno, de forma que sua importância seja controladamente considerada no agrupamento.

## 4.5 Índices de validade do agrupamento

Um ponto crítico na utilização do método de agrupamento é que ele parte da premissa de que o número de grupos é um parâmetro conhecido. Como na maioria das aplicações o número de grupos é uma incógnita, recorre-se a índices que identifiquem a qualidade do agrupamento que convencionalmente são chamados de índices de validade do agrupamento.

Para produzir uma escolha ótima do número de grupos, alguns índices têm sido utilizados para validar, *a posteriori*, o resultado do agrupamento. No entanto, não existe consenso quanto ao índice ou coeficiente de validade mais eficiente, havendo ainda pesquisas sendo desenvolvidas a esse respeito, como em Kim et al. (2004) e Wu e Yang (2005), numa tentativa de encontrar um índice que se adeque a qualquer estrutura de dados.

Rao e Srinivas (2006), em um estudo de regionalização de vazões, testaram o comportamento de cinco índices de validade de agrupamento. Dois deles consideraram apenas as características da matriz de pertinência  $U$  e foram propostos por Bezdek (1981): a partição difusa ( $Pd$ ) e a entropia de partição ( $Ep$ ) dados como:

$$Pd = \frac{1}{N} \sum_{i=1}^c \sum_{k=1}^N (\mu_{ik})^2 \quad (4.9)$$

$$Ep = -\frac{1}{N} \left[ \sum_{i=1}^c \sum_{k=1}^N (\mu_{ik} \log(\mu_{ik})) \right] \quad (4.10)$$

O coeficiente de partição difusa mede o sucesso da decomposição do conjunto de dados. O melhor agrupamento corresponde ao valor máximo de  $Pd$ . Entretanto, se o coeficiente resulta 1, o método difuso é desnecessário, indicando que poderia ter sido utilizada uma metodologia tradicional. Por outro lado, se resulta  $1/c$ , a classificação é irrelevante, pois os dados tendem a pertencer a todas as classes, com graus iguais e muito pequenos, aquém do grau significativo.

A principal dificuldade para se trabalhar com este coeficiente é a tendência de diminuir o valor com o aumento do número de classes. Há, porém, perturbações à medida que o número de classes cresce. Essas perturbações correspondem às situações em que ocorrem melhoras de desempenho com o acréscimo de um novo agrupamento.

Os outros três índices testados por Rao e Srinivas (2006) levam em conta o grau de compactação dos grupos e o quanto são separados entre si, incluindo, portanto, características da amostra nas formulações. O mais antigo é o índice de Fukuyama e Sugeno ( $V_{FS}$ ), que compara a variância dos centros em torno da média dos dados, e teoricamente atinge um valor mínimo para uma configuração de grupos mais compactos e bem separados:

$$V_{FS} = \sum_{k=1}^c \sum_{i=1}^n \mu_{ik}^m \left[ \sum_{j=1}^M (x_{ij} - v_{kj})^2 - \sum_{j=1}^M (v_{kj} - \bar{x}_j)^2 \right] \quad (4.11)$$

O índice de Xie-Beni ( $V_{XB}$ ) representa a razão entre a distância dos elementos do grupo com relação ao seu centro e a menor distância entre os centros dos grupos, de forma que quanto menor a razão melhor é a configuração do agrupamento:

$$V_{XB} = \frac{\sum_{k=1}^c \sum_{i=1}^n \mu_{ik}^m \sum_{j=1}^M (x_{ij} - v_{kj})^2}{N \min_{i \neq k} \sum_{j=1}^M (v_{kj} - v_{ij})^2} \quad (4.12)$$

Já o índice de Kwon ( $V_K$ ) é uma adaptação do índice de Xie-Beni apresentando um termo *ad hoc* para eliminar a monotonicidade por vezes verificada em  $V_{XB}$ :

$$V_K = \frac{\sum_{k=1}^c \sum_{i=1}^n \mu_{ik}^m \sum_{j=1}^M (x_{ij} - v_{kj})^2 + \frac{1}{c} \sum_{k=1}^c \sum_{j=1}^M (v_{kj} - \bar{x}_j)^2}{\min_{i \neq k} \sum_{j=1}^M (v_{kj} - v_{ij})^2} \quad (4.13)$$

Rao e Srinivas (2006) verificaram haver monotonicidade em  $Pd$ ,  $Ep$  e  $V_{FS}$ , mas, enquanto  $Pd$  e  $Ep$  resultavam sempre dois grupos como alternativa ótima,  $V_{FS}$  apresentava o número máximo de grupos testado como a melhor situação. No entanto, Hall e Mins (1999) utilizaram  $Pd$  para justificar o número de grupos adotado, não identificando monotonicidade no índice.

Embora Kwon tenha modificado  $V_{XB}$ , devido à monotonicidade, a mesma não foi verificada em  $V_{XB}$  com os dados de Rao e Srinivas (2006). O que pode significar interferência das características dos dados no comportamento dos índices.

Pakira et al. (2005) apresentaram o índice PBMF como de elevado desempenho na identificação do número de agrupamento ótimo, tanto em séries sintéticas, como em séries reais. Ao contrário dos índices anteriormente mencionados, este baseia-se na hipótese de que, o máximo da relação da distância máxima entre os centros e o somatório da distância entre os pontos e os centros, ponderado pelo grau de pertinência, é capaz de revelar o melhor número de grupos, de acordo com a formulação abaixo (Eq.(4.14))

$$V_{PB} = \frac{1}{c} \times \frac{\sum_{k=1}^1 \sum_{i=1}^n \mu_{ik}^m \sum_{j=1}^M (x_{ij} - v_{kj})^2 \times \max \sum_{j=1}^M (v_{kj} - v_{ij})^2}{\sum_{k=1}^c \sum_{i=1}^n \mu_{ik}^m \sum_{j=1}^M (x_{ij} - v_{kj})^2} \quad (4.14)$$

Note-se que o uso concomitante de um conjunto de índices diferentes pode representar um subsídio mais robusto na escolha do melhor número de grupos. Isso porque resultados de índices de origens diferentes e com interpretações diversas que apresentam recomendações coincidentes, reforçam as convicções para a escolha a ser feita.

## 4.6 Delimitação dos grupos e defusificação

Como, pelo método utilizado, todos os pontos tendem a pertencer simultaneamente a todos os grupos, mesmo que com grau de pertinência muito pequeno, há, teoricamente, a impossibilidade de delimitação absoluta dos grupos.

Por esta razão, Rao e Srinivas (2006) adotaram um critério de ponto de corte para delimitação das regiões, tal que, para pertencer a um grupo, o elemento deve apresentar grau de pertinência superior a um determinado valor, sugerindo para isso o valor de  $1/c$ , onde  $c$  equivale ao número de grupos.

De acordo com Ross (1995), os grupos difusos podem ser defusificados, ou seja, tornados rígidos ou booleanos, utilizando o método da máxima pertinência ou o método do centro mais próximo. No primeiro caso, para cada linha da matriz de pertinência  $U$  os graus de pertinência se tornarão binários com a substituição do máximo grau de pertinência por um e os demais graus de pertinência por zero (4.15).

$$\mu_{ij} = \max_{1 \leq i \leq c} \{\mu_{ik}\} = 1; \quad \mu_{ik} = 0 \quad \forall k \neq j \quad (4.15)$$

Pelo método do centro mais próximo as distâncias dos pontos aos centros de cada grupo são calculadas e adota-se grau de pertinência unitário ao grupo cujo centro é mais próximo ao ponto, enquanto as demais pertinências são nulas (equação (4.16)). O resultado final é equivalente ao que seria obtido pelo método *K-means*.

$$\text{Se } d_{ij} = \min_{1 \leq i \leq c} \{d_{ik}\} = \min_{1 \leq i \leq c} \|v_k - x_i\|$$

$$\text{Então } \mu_{ij} = 1; \quad \mu_{ik} = 0 \quad \forall k \neq j \quad (4.16)$$

## 4.7 Alocação de novos pontos

Segundo Rao e Srinivas (2006), a alocação de um novo ponto pode ser feita através da determinação de seu grau de pertinência a cada grupo, obtido pela equação (4.4), que depende da distância euclidiana entre o ponto e o centro de cada grupo.

As coordenadas dos centros são originárias do agrupamento e deve-se, preferencialmente, considerar todos os atributos utilizados no agrupamento.

Entretanto, se o agrupamento originalmente foi obtido utilizando-se atributos dependentes e independentes, e deseja-se alocar pontos não amostrados, é possível determinar os graus de pertinência utilizando-se apenas os atributos conhecidos.

Nesse caso, para o cálculo das distâncias, as coordenadas dos centros serão reduzidas aos atributos conhecidos no ponto a ser alocado.

Não foram encontradas, no entanto, referências na literatura com a aplicação do modelo de alocação, mas, tão somente, a sugestão de uso por Rao e Srinivas (2006).

#### 4.8 Estimativas pontuais a partir dos graus de pertinência

Hall e Mins (1999) sugerem a possibilidade estimar as coordenadas de um ponto  $x_{ij}$  através da média, ponderada pelos respectivos graus de pertinência ( $\mu_{ik}$ ), dos centros do agrupamento ( $v_{kj}$ )

Como a soma dos graus de pertinência de um ponto é unitária, a média ponderada é equivalente à soma ponderada, então o modelo de estimativa pode ser expresso pelo produto matricial:

$$x_{i,j} = \mu_{ik} * v_{kj} \quad (4.17)$$

Jingyi e Hall (2004), em uma análise regional de freqüências de cheias, utilizaram este modelo para obter estimativas locais de quantis adimensionais a partir dos quantis regionais adimensionais.

Os autores compararam os erros existentes entre os dados observados e os quantis estimados por uma curva de ajuste local, e os estimados pela curva regional e os estimados pelo modelo ponderado pelos graus de pertinência. Entretanto, consideraram o resultado inconclusivo, recomendando maiores estudos sobre esta técnica.

Observe-se, porém, que a tentativa de estimar quantis adimensionais locais é contraditória, uma vez que por principio os quantis regionais são mais robustos. Entretanto, a sugestão de Hall e Mins (1999) poderia ser utilizada, não para estimar os quantis locais (como fizeram Jingyi e Hall, 2004), mas sim para estimar o fator de escala local, substituindo a função de regressão que comumente é utilizada.

## **5 METODOLOGIA**

### **5.1 Introdução**

Este trabalho apresenta uma proposta metodológica para a análise regional de frequências baseada em duas hipóteses. Em primeiro lugar, que numa região homogênea as séries seguem uma mesma distribuição de probabilidades, dispensando o fator de escala local. Em segundo, que seus parâmetros podem ser obtidos pela média dos parâmetros de ajuste locais.

Nessas condições, é necessário adotar um modelo de distribuição que se adeque ao fenômeno, e que existam recursos disponíveis para o ajuste das séries locais a esse modelo de distribuição de probabilidades.

Com o objetivo de verificar se as estimativas produzidas podem ser consideradas satisfatórias, a segunda hipótese é testada através da simulação de regiões homogêneas.

Para compor a metodologia, alguns recursos metodológicos foram desenvolvidos, especificamente um teste para medir a qualidade do agrupamento, baseado no teste de aderência de Kolmogorov-Smirnov, e um teste de homogeneidade, pela adaptação e generalização da medida de heterogeneidade proposta por Hosking e Wallis (1997).

Durante a descrição do teste de homogeneidade, procura-se apresentar o embasamento teórico que justifique a primeira hipótese.

Após a exposição dos recursos metodológicos é apresentada uma descrição generalizada das etapas que constituem o método de análise regional de frequências proposto.

### **5.2 Recursos metodológicos**

#### **5.2.1 Verificação do modelo para obtenção da distribuição regional**

Os procedimentos a seguir foram realizados com o objetivo de verificar a validade da hipótese de que os parâmetros da distribuição regional podem ser determinados pela média dos parâmetros locais.

Como dito anteriormente, deve-se estabelecer inicialmente o tipo de distribuição que irá representar o fenômeno. Desta forma, considerando que o estudo é direcionado à análise de precipitações totais, aceitou-se que os modelos representados pelas distribuições Gama e Normal

são adequados para estimar as probabilidades de ocorrência de precipitações mensais e anuais, respectivamente. Portanto, apenas estas duas distribuições foram avaliadas.

O estudo consistiu na avaliação da qualidade da estimativa da distribuição verdadeira quando esta é obtida através da média dos parâmetros de ajustes locais, em regiões homogêneas hipotéticas.

Tomando-se como exemplo a avaliação referente à distribuição Gama, cada região foi formada pela simulação de  $N$  séries aleatórias  $i$  de tamanho  $n$ , seguindo uma distribuição de parâmetros  $\alpha$  e  $\beta$  arbitrários, chamada de distribuição original.

Para gerar uma série aleatória é suficiente gerar números aleatórios, entre 0 e 1, seguindo uma distribuição uniforme. Esses números são tomados como as probabilidades e, através da função inversa da distribuição de probabilidades desejada, com parâmetros predefinidos, são determinados os valores dos quantis, construindo-se, assim, a série aleatória.

Depois de geradas, essas séries locais foram ajustadas à distribuição Gama, pelo método da máxima verossimilhança, obtendo-se os parâmetros locais  $\alpha_i$  e  $\beta_i$ . Em seguida, foram estimados os parâmetros da distribuição original, através da média dos parâmetros locais.

Foram avaliadas algumas situações, nas quais variaram-se os valores de  $N$  (5, 10, 20 e 50 séries),  $n$  (20, 30 e 50 anos). Em cada situação o processo foi repetido 500 vezes mantendo-se constantes as variáveis. Ao final, cada uma das 500 distribuições estimadas foi comparada à distribuição original pela máxima diferença de probabilidades, calculando-se também o erro médio relativo dos parâmetros.

Para efeito comparativo, pesquisou-se qual é a média da máxima diferença de probabilidades encontrada no ajuste individual de séries aleatórias para diversos tamanhos de séries.

Desta forma, para cada tamanho de série e tipo de distribuição (Gama e Normal), 500 séries aleatórias foram geradas e ajustadas a distribuição de origem, verificando-se, sem seguida, a máxima diferença de probabilidades entre as amostras e os ajustes.

Portanto, se as diferenças referentes ao modelo de obtenção da distribuição regional forem semelhantes às diferenças encontradas no ajuste de séries consideravelmente longas, então se poderá aceitar que o modelo baseado na média dos parâmetros é satisfatório.

### 5.2.2 Teste de aderência regional

Para os efeitos dessa pesquisa considera-se que as regiões são obtidas pelo método de agrupamento multivariado *Fuzzy C-means*.

Considerando que, para cada condição de ponderação das variáveis e para cada opção de número de grupos, o método de agrupamento resultará uma configuração diferente das regiões, é proposto um critério de verificação da qualidade do agrupamento que tem por finalidade de evitar ou diminuir a necessidade de grande interferência, como divisões ou junções de regiões, para se atingir o critério de homogeneidade.

Embora a qualidade do agrupamento comumente seja avaliada pelos índices de validade apresentados no item 4.5, dificilmente se consegue comparar os diferentes resultados que ocorrem quando se aplicam modificações na matriz de dados, como, por exemplo, a ponderação das variáveis.

Assim, diferente dos índices de validade do agrupamento, o teste de aderência regional, aqui proposto, visa incorporar o objetivo do agrupamento que é, nesse caso, obter-se regiões constituídas por locais em que a distribuição de probabilidade do fenômeno analisado é semelhante.

A base do teste está no uso do teste de aderência de Kolmogorov-Smirnov (KS) para a comparação das séries observadas pontualmente em cada região com as respectivas distribuições regionais.

Assim, em cada região, são comparadas as séries observadas locais com a série resultante da função de quantis regional ( $Q$ ) para as frequências de não-excedência observadas ( $f$ ), dadas por:

$$Q(f_i) = F(x_i; \theta_{r1}, \theta_{r2} \dots)^{-1} \quad (5.1)$$

$$f_i = \frac{i}{n+1}, \quad i=1, 2, \dots, n, \quad (5.2)$$

onde  $Q$  é a inversa da função de distribuição de probabilidades ( $F$ ) com parâmetros regionais  $\theta_r$ , (obtidos pela média dos parâmetros locais),  $f_i$  é a frequência de não-excedência dos elementos  $x_i$  ordenados ascendentemente,  $i$  é o número de ordem, e  $n$  é o tamanho da série.

As séries significativas no teste KS, a 20% de significância, são denominadas séries discordantes e o resultado do teste de aderência regional é dado em termos do percentual de séries discordantes (Sd), estabelecendo-se como critério que quanto menor o valor de Sd, melhor é a configuração do agrupamento.

Quando a análise é dirigida a vários períodos de referência ao mesmo tempo (todos os meses, por exemplo), o valor de  $S_d$  é calculado para cada período. Nesse caso, deve-se buscar a condição de melhor média para  $S_d$ , mas com o melhor equilíbrio possível entre os valores de  $S_d$  em cada período.

Desta forma, o teste de aderência regional serve de subsídio para definição da melhor ponderação das variáveis e do melhor número de grupos.

Como valor de  $S_d$  pode tender a diminuir indefinidamente com o aumento do número de grupos, considera-se conveniente que, para definição do melhor número de grupos, o teste de aderência regional seja utilizado em conjunto com os índices de validade do agrupamento.

### 5.2.3 Teste de homogeneidade

Após a determinação da melhor configuração do agrupamento, convém a verificação da homogeneidade propriamente dita, ou seja, a comprovação de que as regiões podem ser consideradas homogêneas, uma vez que o teste de aderência regional avalia o agrupamento como um todo e não cada região em particular.

Nessa pesquisa considera-se que numa região homogênea as séries seguem uma mesma distribuição de probabilidades, mas não apenas em termos adimensionais.

O princípio de homogeneidade plena das regiões está baseado no fato de que, havendo pouca variação nos fatores físicos condicionantes do fenômeno dentro de uma região, as séries locais devem seguir uma mesma distribuição (a distribuição verdadeira), e conseqüentemente possuir uma mesma média. No entanto, por conta da aleatoriedade envolvida nas ocorrências e da restrição no tamanho das amostras, irão existir flutuações em relação à média da distribuição verdadeira.

Ora, se uma distribuição regional adimensionalizada pela média, por ser obtida a partir de várias distribuições locais, seria a estimativa mais próxima e robusta da distribuição verdadeira, como se aceita pelo *index-flood*, assim também a média da distribuição regional (fator de escala regional), por filtrar as diferenças relativas às observações, seria provavelmente mais próxima da média da distribuição verdadeira do que cada média local.

Por outro lado se a variabilidade dos fatores físicos condicionantes for suficiente para gerar diferenças significativas nas médias, a distribuição regional, adimensionalizada pela sua média (denominada aqui de fator de escala regional), ainda poderia ser válida.

Nesse caso, os quantis regionais, adimensionalizados pelo fator de escala regional, seriam redimensionalizados, multiplicando-os pelo fator de escala local, conforme o princípio do *index-flood*.

Para avaliar a condição de homogeneidade nas regiões resultantes do agrupamento, é proposta uma adaptação ao teste de heterogeneidade  $H$  apresentado por Hosking e Wallis (1997) e descrito no item 3.3.3, no qual são comparadas as características encontradas na região real com aquelas que devem existir numa região virtual e homogênea.

Nessas condições, adotou-se como critério que se a média das distâncias máximas, entre as probabilidades estimadas pela curva regional e as observadas nas séries locais, forem da mesma ordem da média dessas diferenças em uma região homogênea virtual, se poderia afirmar que a região avaliada é possivelmente homogênea.

Desta forma, simulando uma região virtual composta por séries aleatórias, mas que seguem uma mesma distribuição, é possível calcular qual a magnitude das diferenças que pode ser atribuída à aleatoriedade.

Assim, se as diferenças observadas na região real forem equivalentes às da região simulada, considera-se que a região é aceitavelmente homogênea, podendo-se assumir que a distribuição regional é plenamente válida.

Por outro lado, se essas diferenças forem muito superiores às encontradas na região virtual, haverá indícios de que elas sejam devidas principalmente a fatores físicos locais. Nesse caso, dois caminhos podem ser tomados: verificar se há indícios físicos para subdivisão das regiões, ou considerar a homogeneidade apenas em termos adimensionais, respeitando-se a média dos eventos locais, como costumeiramente é feito no método do *index-flood*.

Desta forma, para avaliar a homogeneidade de uma região  $r$ , composta por  $N_r$  séries  $i$ , com  $n_i$  elementos cada, deve-se simular um número suficientemente grande de regiões homogêneas virtuais que imitam a região  $r$ . Dessa forma, cada região homogênea virtual da região  $r$  é obtida simulando  $N_r$  séries aleatórias, cada uma com  $n_i$  elementos, que obedecem aos parâmetros da distribuição regional na região  $r$ , que são obtidos a partir dos parâmetros de ajuste em cada local.

O critério de homogeneidade é baseado na diferença máxima  $dm_i$  entre as probabilidades estimadas pela curva regional ( $G$ ) e as frequências observadas nas séries locais ( $F_i$ ), analogamente à equação (3.5) utilizada no teste de Kolmogorov-Smirnov.

$$dm_i = \max | F_i(x) - G(x) |, \quad (5.3)$$

Desta forma, na região real calcula-se a média ( $dp$ ) das diferenças máximas ( $dm_i$ ) entre as probabilidades estimadas pela curva regional ( $G$ ) e as frequências observadas nas séries locais ( $F_i$ ). Procedendo-se da mesma forma nas regiões simuladas, calcula-se  $dp_s$  que é a média das diferenças máximas entre as probabilidades estimadas pela curva regional e as frequências observadas nas séries locais simuladas. Ao fim das simulações (sugerindo-se que sejam simuladas 500 regiões homogêneas), são determinados a média ( $\mu dp_s$ ) e o desvio padrão ( $\sigma dp_s$ ) dos valores de  $dp_s$ .

Analogamente ao modelo de Hosking e Wallis (1997), uma medida de heterogeneidade ( $H_{dp}$ ) para a diferença máxima de probabilidades entre as séries foi estabelecida como:

$$H_{dp} = \frac{dp - \mu dp_s}{\sigma dp_s} \quad (5.4)$$

onde  $\mu dp_s$  e  $\sigma dp_s$  são a média e o desvio padrão da média das diferenças máximas de probabilidade do total de regiões simuladas

Acredita-se que  $\mu dp_s$ ,  $\sigma dp_s$  deverão ser dependentes do tamanho e número das séries que compõem a região, e do tipo de distribuição, mas não exatamente dos parâmetros.

Dessa forma, caso sejam avaliados vários períodos ao mesmo tempo dentro de uma região, a média e o desvio padrão das simulações, referentes a um período qualquer, serão válidos para todos os períodos, desde que estes sigam o mesmo tipo de distribuição.

Finalmente, considera-se válida a mesma proposição de Hosking e Wallis (1997), de que para  $H_{dp} < 1$ , a região é aceitavelmente homogênea;  $1 \leq H_{dp} < 2$ , a região é possivelmente heterogênea; e para  $H_{dp} \geq 2$ , a região é definitivamente heterogênea. De acordo com os autores, mesmo numa condição de relativa heterogeneidade, as estimativas geradas pela distribuição regional ainda serão mais robustas que as geradas pelas distribuições locais, e, portanto, são bem aceitas regiões com valores de  $H$  menores que 2, devendo-se buscar ajustes nas regiões que apresentam valores acima de 2.

Para a análise da homogeneidade sob o aspecto adimensional, a distribuição regional deve ser adimensionalizada, dividindo-se os quantis regionais pela média da distribuição regional e redimensionalizada para cada ponto  $i$ , multiplicando-se os quantis adimensionalizados pela média da respectiva série local. Este procedimento deve ser aplicado tanto na região real quanto nas regiões simuladas, mesmo que as séries aleatórias sejam geradas segundo uma distribuição de média unitária.

## 5.3 Etapas do método

### 5.3.1 Aspectos Gerais

No estudo de fenômenos naturais, como o de precipitações totais, por exemplo, geralmente há interesse na análise do comportamento das ocorrências com várias durações (quinzenais, mensais, bimestrais, anuais, etc.). É possível, nessa situação, definir separadamente uma configuração regional para cada caso, entretanto, a investigação de uma configuração única e que consiga atender satisfatoriamente todas as durações economiza trabalho, oferece resultados mais práticos e de mais fácil acesso, e permite uma visão integrada das ocorrências na área de interesse, sendo, portanto, mais indicada.

Cabe lembrar que a metodologia exige a adoção prévia do tipo de distribuição de frequências que irá representar o fenômeno estudado. No caso específico da análise de precipitações totais mensais e anuais, as distribuições Gama e Normal, respectivamente, são as distribuições geralmente utilizadas, sendo aqui sugeridas para esse tipo de estudo.

Por ser um objetivo adjacente desse trabalho, a exploração de possíveis contribuições do método de agrupamento difuso *Fuzzy C-Means* (FCM), o mesmo é aqui empregado para determinação das regiões, embora o resultado utilizado seja previamente defusificado pelo método da máxima pertinência (item 4.6). Entretanto, em outras aplicações, pode ser escolhido o método de agrupamento que for considerado mais conveniente.

No uso do FCM, sugere-se apenas que se avalie preliminarmente que valor do coeficiente de difusividade ( $m$ ) melhor se adequa aos dados, conforme os critérios citados no item 4.2.

A seguir são descritos, em linhas gerais, os procedimentos a serem aplicados na análise regional de frequências.

### 5.3.2 Preparação dos dados

Essa etapa inclui a escolha das estações, verificação da consistência das séries e preparação das variáveis para o agrupamento. Quanto ao período e tamanho das observações, sugere-se utilizar séries o mais longas possível, mas em um período coincidente, pelo menos para a etapa de identificação das regiões, a fim de minimizar possíveis distorções na configuração das regiões associadas à variabilidade temporal.

Com essa opção se deixa de aproveitar as informações contidas nas séries mais longas, entretanto, confirmando-se a homogeneidade de uma região para um período coincidente e de tamanho satisfatório, as informações das séries longas podem ser posteriormente incorporadas para a determinação das distribuições regionais, caso exista esse interesse.

Escolhidas as estações e consistidas as séries, devem ser determinados os parâmetros da distribuição adotada e as variáveis que serão utilizadas no agrupamento para cada uma das séries selecionadas.

As variáveis devem ser escalonadas por algum dos métodos apresentados no item 4.4 podendo-se ponderá-las com a finalidade de obter melhores configurações de agrupamento.

### **5.3.3 Ponderação das variáveis**

Com a matriz de dados preparada, seguem-se os procedimentos para escolha da ponderação aplicada às variáveis. Nesse caso, pode-se escolher uma ou algumas variáveis que terão sua importância aumentada pela multiplicação de um valor maior que 1, enquanto as demais variáveis permanecem com ponderação neutra (igual a 1).

A escolha das variáveis pode ser baseada em experiências anteriores, ou no resultado preliminar do teste de aderência regional, para o agrupamento sem ponderação, optando-se por ponderar os períodos com maior valor de  $S_d$ .

Escolhida a variável (ou variáveis) a ser ponderada devem ser testados alguns valores, que são as alternativas de ponderação, conforme se segue:

- Estabelecer que variáveis serão ponderadas e as alternativas de ponderação;
- Para cada alternativa de ponderação, aplicar o método de agrupamento considerando diferentes números de grupos (sugere-se de 2 a aproximadamente a raiz quadrada do total de pontos);
- Para cada resultado de agrupamento, calcular os parâmetros das distribuições regionais pela média dos parâmetros locais em cada região;
- Aplicar o teste de aderência regional, de acordo com o item 5.2.2, determinando o valor de  $S_d$  (médio e por período) para cada configuração (número de grupos) e a média desses resultados para cada alternativa de ponderação.
- Escolher a alternativa de ponderação que apresentar menor média de  $S_d$ .

### 5.3.4 Determinação do melhor número de grupos

Depois de estabelecida a melhor opção de ponderação, é necessário determinar o número de regiões mais adequado para os objetivos da análise. Para esta etapa, se propõe que inicialmente sejam avaliados os resultados dos índices de validade do agrupamento, apresentados no item 4.5, e as melhores opções, indicadas pelos índices, sejam comparadas com os resultados do teste de aderência regional.

Essa etapa pode ser resumida nos seguintes passos:

- Agrupar a matriz, devidamente ponderada, para os diferentes números de grupos, e aplicar os índices de validade do agrupamento;
- Identificar as melhores alternativas de agrupamento indicadas pelos índices de validade;
- Consultar os resultados de Sd (médio e por período), para as alternativas identificadas acima;
- Escolher a alternativa mais coerente, para nível de detalhamento desejado.

### 5.3.5 Ajustes preliminares

Quando se trabalha com várias durações (períodos de referência) ao mesmo tempo pode-se averiguar a adequação de cada ponto à sua região, contabilizando o número de séries discordantes (significativas no teste KS) com relação à distribuição regional.

Quando o ponto apresenta muitas séries discordantes (mais de duas ou três quando se avalia os 12 meses), é possível verificar se é um caso de má alocação. Para isso, verifica-se sua adequação às demais regiões, mantendo-se fixas as distribuições regionais. Se a diminuição no número de séries discordantes for significativa (de duas ou mais séries por ponto), o ponto pode ser realocado para a região à qual melhor se adaptou.

Note-se que quanto mais eficiente for a fase de agrupamento, menos situações de má alocação deverão existir.

Caso não se configure uma má alocação, ou seja, o ponto apresenta muitas séries discordantes, mas a mudança de região não traz melhorias, deve-se considerar a possibilidade de retirá-lo da análise, com a finalidade de retirar sua influência sobre a distribuição regional e, possivelmente, melhorar a estimativa regional.

### **5.3.6 Teste de homogeneidade**

Feitos os ajustes preliminares na configuração das regiões, a distribuição regional é recalculada aplicando-se, em seguida, o teste de homogeneidade descrito no item 5.2.3.

Nas regiões em que não é observada a homogeneidade para muitos períodos de referência (durações), deve-se investigar a possibilidade de maiores ajustes, como a subdivisão das regiões. Nesses casos, devem ser procurados fatores físicos que justifiquem os ajustes realizados.

Se prevalecer a condição de não homogeneidade para algum dos períodos, independentemente da região, pode-se estudar a possibilidade de determinar uma configuração regional própria, capaz de caracterizar a particularidade espacial da distribuição de frequências nesse período.

Após todas as intervenções, são determinadas as distribuições regionais considerando a configuração final das regiões.

### **5.3.7 Ajuste de modelo para alocação de pontos não amostrados**

Note-se que as regiões identificadas pelo método de agrupamento são definidas num espaço multidimensional, que é o espaço das variáveis utilizadas no agrupamento.

Desta forma, para estender a análise regional a pontos que não pertencem à matriz de dados utilizada originalmente no agrupamento, uma alocação visual desses pontos, apenas por proximidade geográfica, poderá produzir resultados inadequados.

Utilizando o método de agrupamento FCM, porém, é possível realizar a alocação pela determinação dos graus de pertinência dos novos pontos em relação aos centros dos grupos, através da equação (4.4).

Este modelo de alocação pode ser utilizado quando todas as variáveis participantes do agrupamento original são conhecidas, e também quando apenas uma parte delas estiver disponível, embora, neste caso, seja necessário verificar se apenas essas variáveis serão suficientes para alocar adequadamente os pontos.

No caso de se utilizar apenas uma parte das variáveis, deve-se considerar, durante o processo de alocação, apenas estas variáveis como coordenadas do centro do agrupamento.

Nos dois casos, a matriz dos dados a serem alocados, bem como o centro do agrupamento, devem ser adimensionalizados, com o mesmo critério de escalonamento e o mesmo fator de mudança de escala utilizados no agrupamento original.

Depois de adimensionalizados, o centro e a matriz de dados devem ser ponderados com a mesma alternativa de ponderação aplicada anteriormente. Entretanto, quando se utiliza apenas parte das variáveis, sugere-se que seja verificada, preliminarmente, qual a ponderação que deve ser aplicada para se obter resultados, no teste de aderência regional, próximos aos encontrados no agrupamento original.

A matriz de pertinência dos novos pontos é então calculada, pela equação 4.4, e os pontos deverão ser alocados na região à qual pertencerão com o máximo grau.

Espera-se que, caso o resultado do teste de aderência regional seja aceitável para os pontos que geraram o agrupamento, pontos não amostrados possam ser satisfatoriamente alocados, utilizando-se apenas as variáveis disponíveis.

### **5.3.8 Estimativa de características locais através dos graus de pertinência**

Caso se considere pertinente utilizar a distribuição regional apenas em termos adimensionais, faz-se necessário conhecer o fator de escala no ponto de interesse (aqui tomado como a média das observações). Para os pontos não amostrados esse valor deve ser estimado por uma função de regressão ou por mapas de isolinhas (no caso da análise de precipitações totais, por exemplo).

Outra alternativa é utilizar o modelo da média ponderada pelos graus de pertinência, conforme a equação (4.17), embora ainda se necessite de maiores investigações a respeito da qualidade dessas estimativas.

## 6 ESTUDO DE CASO

### 6.1 Local da aplicação

A região de estudo abrange o estado do Rio Grande do Sul e parte do estado de Santa Catarina, correspondendo a toda parte brasileira da bacia hidrográfica do rio Uruguai e as bacias que drenam para o mar, de todo o Rio Grande do Sul até o litoral sul de Santa Catarina. A área de estudo é então delimitada, aproximadamente, pelas coordenadas de 26° a 34° de latitude Sul e 48° a 58° de longitude Oeste, com uma área total de cerca de 330 mil Km<sup>2</sup>.

O estado do Rio Grande do Sul possui um território de 281.748km<sup>2</sup> e está localizado entre os paralelos 27° e 34° S e meridianos 50° e 57° O. As altitudes variam de 0m, na faixa litorânea marcada por grandes massas d'água, a 1200m na Serra do Nordeste, estando a maior parte do Estado entre 200m e 800m. As terras baixas da Depressão Central Gaúcha, na direção do paralelo 30°, dividem as áreas altas deixando ao norte o planalto Meridional e ao sul o Planalto Sul-Riograndense e o Planalto da Campanha (mais a sudoeste) (Figura 6.1).

A porção que envolve Santa Catarina corresponde à parte catarinense da bacia do rio Uruguai e o Litoral Sul do Estado com uma área total de aproximadamente 50 mil km<sup>2</sup>. O relevo é marcado por altitudes de até cerca de 1800m na Serra Geral, perdendo altitude para oeste. Já na planície litorânea, as altitudes variam de 0 a 300m.

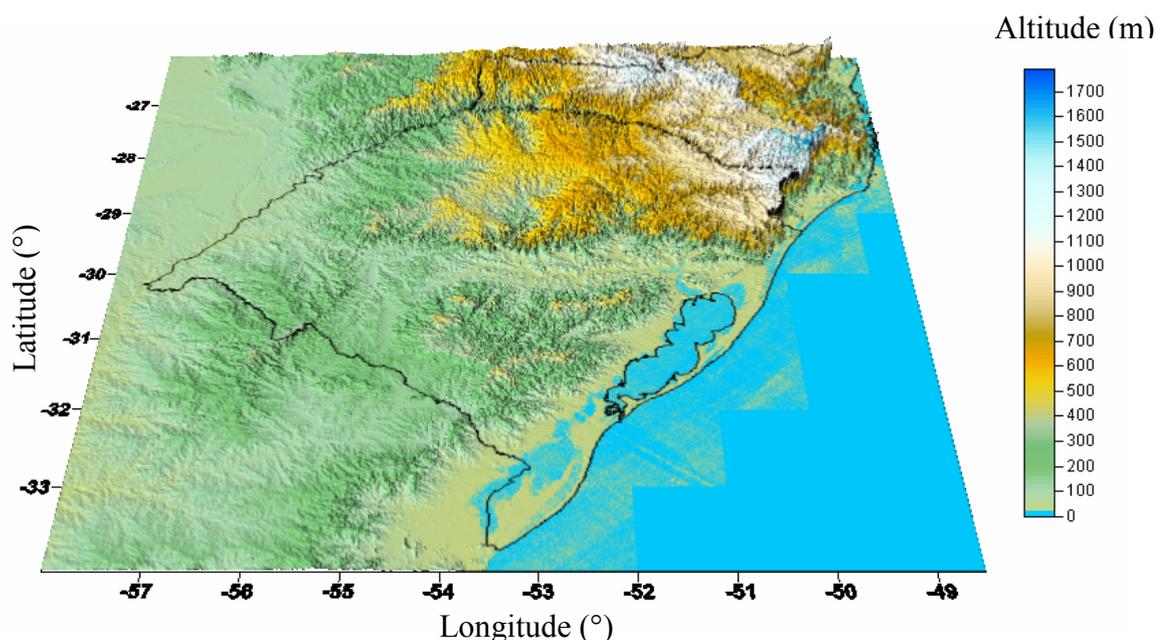


Figura 6.1 Localização e relevo do Rio Grande do Sul e Santa Catarina

De acordo com Moreno (1961), o clima do Rio Grande do Sul, dentro do sistema Koeppen de classificação climática, enquadra-se no tipo Temperado Úmido com áreas morfo-climáticas subtropical (Cfa) e temperada (Cfb). O regime pluviométrico caracteriza-se por chuvas durante todos os meses do ano, no entanto a gênese das chuvas de inverno é diferente da gênese de verão. No inverno ela está associada ao deslocamento de anticiclones (Massas Polar Atlântica e Pacífica) que têm ação duradoura, ocasionando chuvas mais prolongadas e em maior número de dias. No verão, as chuvas ocorrem pela ação dos ventos alísios úmidos gerados pela frente Tropical Sul-Atlântica, ocasionando chuvas de alta intensidade.

Ainda segundo Moreno (1961), especialmente o regime pluviométrico é influenciado pela orografia, chovendo mais nas encostas dos planaltos. Além disso, apresenta um gradiente negativo de volume no sentido norte-sul que poderia estar associado à mudança na direção dos ventos nas imediações do paralelo 30°, a partir do qual, teoricamente, os ventos alísios (úmidos) dariam lugar aos ventos de oeste (secos).

De acordo com Lopes (2006), os principais mecanismos causadores da precipitação pluvial no Rio Grande do Sul são as passagens de sistemas frontais e o desenvolvimento de aglomerados convectivos.

Além disso, a região Sul sofre a influência das interações oceano-atmosféricas, que ocasionam anomalias de precipitação. Observações sugerem que a ação do fenômeno El-Niño ocasiona na região chuvas abundantes na primavera e intensas de maio a junho, enquanto que as estiagens são associadas a ocorrências de La-Niña. (Oliveira, 2001).

## **6.2 Seleção das estações pluviométricas**

De posse do banco de dados com as estações pluviométricas da ANA (Agência Nacional de Águas), operadas pela CPRM (Companhia de Pesquisa em Recursos Minerais, hoje Serviço Geológico do Brasil), e pela CEEE (Companhia de Energia Elétrica do Estado), realizou-se uma pré-seleção das estações a serem utilizadas no trabalho, tendo como critério inicial a utilização de séries coincidentes de pelo menos 20 anos, na determinação das regiões homogêneas, para minimizar possíveis distorções na configuração das regiões associadas à variabilidade temporal.

Com essa opção se deixa de aproveitar as informações contidas nas séries mais longas, entretanto, confirmando-se a homogeneidade de uma região para um período coincidente e de tamanho satisfatório, as informações das séries longas podem ser posteriormente incorporadas, caso exista esse interesse.

O gráfico de barras da Figura 6.2 contribuiu para a escolha do período a ser adotado. Tomando-se inicialmente apenas o Rio Grande do Sul, as estações foram separadas por localização Norte/Sul, tendo como delimitante o paralelo 30°, para que a escolha do período não compromettesse a distribuição espacial. Como se vê, tanto pelo gráfico (Figura 6.2) como pela Tabela 6.1, a região norte possui cerca de 2,5 vezes mais estações que a região sul.

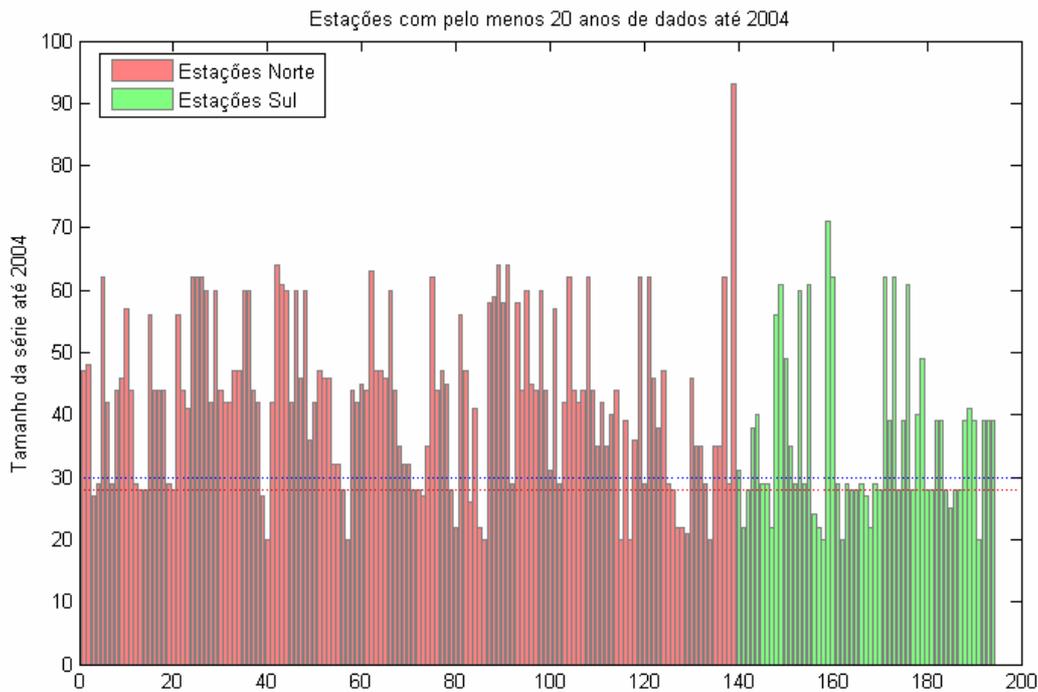


Figura 6.2 Pré-seleção das estações pluviométricas do Rio Grande do Sul.

Embora o tamanho estatisticamente recomendado para construção de séries não tendenciosas seja de 30 anos, sua escolha implicaria numa perda preliminar de 62 das 194 estações pré-selecionadas, onde 29 seriam da região sul, reduzindo em mais da metade o número de estações nessa região. Assim, pela análise da Tabela 6.1 e Figura 6.2 adotou-se o período de 1977 a 2004 como o mais favorável, dentro dos critérios estabelecidos, tomando-se um total de 169 estações para uma análise mais rigorosa, com referência a falhas e consistência.

Tabela 6.1. Número de estações excluídas de acordo com o tamanho da série.

Tamanho da Série	Estações Excluídas / Total de Estações		
	Norte	Sul	Total
30 anos	33 / 139	29 / 55	62 / 194
29 anos	23 / 139	21 / 55	44 / 194
28 anos	15 / 139	10 / 55	25 / 194

Após a análise da qualidade dos dados disponíveis, 6 estações foram excluídas devido à grande extensão de falhas nas séries, sendo ainda incorporadas 2 estações localizadas em Santa Vitória do Palmar e Uruguaiana, pertencentes ao INMET (Instituto Nacional de Meteorologia), além de 35 estações localizadas no Estado de Santa Catarina, totalizando um conjunto de 200 estações.

As falhas mensais, existentes no período de 1977 a 2004, foram preenchidas utilizando-se o programa Hidro-Plu - Programa de Homogeneização de Dados Pluviométricos da ANA, adotando-se, como referência, postos com correlação sempre superior a 60%.

Após o preenchimento das falhas foi verificada a consistência dos totais anuais de todas as estações através do método da dupla massa (Bertoni e Tucci, 2004), adotando-se a média das quatro estações mais próximas como valor de referência para o posto a ser testado, desde que as diferenças de altitude fossem inferiores a 300m. Com esse critério nenhum dos postos foi excluído.

O mapa da Figura 6.3 apresenta a distribuição final das 200 estações utilizadas no trabalho, associadas a um número de identificação presente na tabela 6.2, que contém também o código da estação, a unidade da Federação a que pertence e o ano de início das séries. Salienta-se que os primeiros quatro dígitos dos códigos originais das estações pluviométricas correspondem aos graus de latitude (dois primeiros) e graus de longitude (dois seguintes) da localização da estação. Informações mais precisas da localização de cada posto, bem como o nome da estação e sub-bacia a qual pertence, são encontradas no anexo A.

Apesar dos esforços para garantir uma boa cobertura espacial, ainda se verificou carência de dados na região central do Rio Grande do Sul (em torno do paralelo 30°). No entanto, algumas estações foram implantadas nesses locais a partir de 1986 e poderão ser utilizadas em análises futuras.

Verificou-se também menor concentração de estações no sul e no oeste do Estado. Entretanto, considerando-se apenas o Estado do Rio Grande do Sul (165 estações), foi possível trabalhar com uma densidade média de um posto para cada 1.700 km<sup>2</sup>, representando uma cobertura sensivelmente melhor que outras análises de precipitação anual ou mensal no Estado. Por exemplo, em Ávila (1994) e Lopes (2006), a densidade é de aproximadamente um posto para cada 8.000 km<sup>2</sup> e 7.000 km<sup>2</sup>, respectivamente.

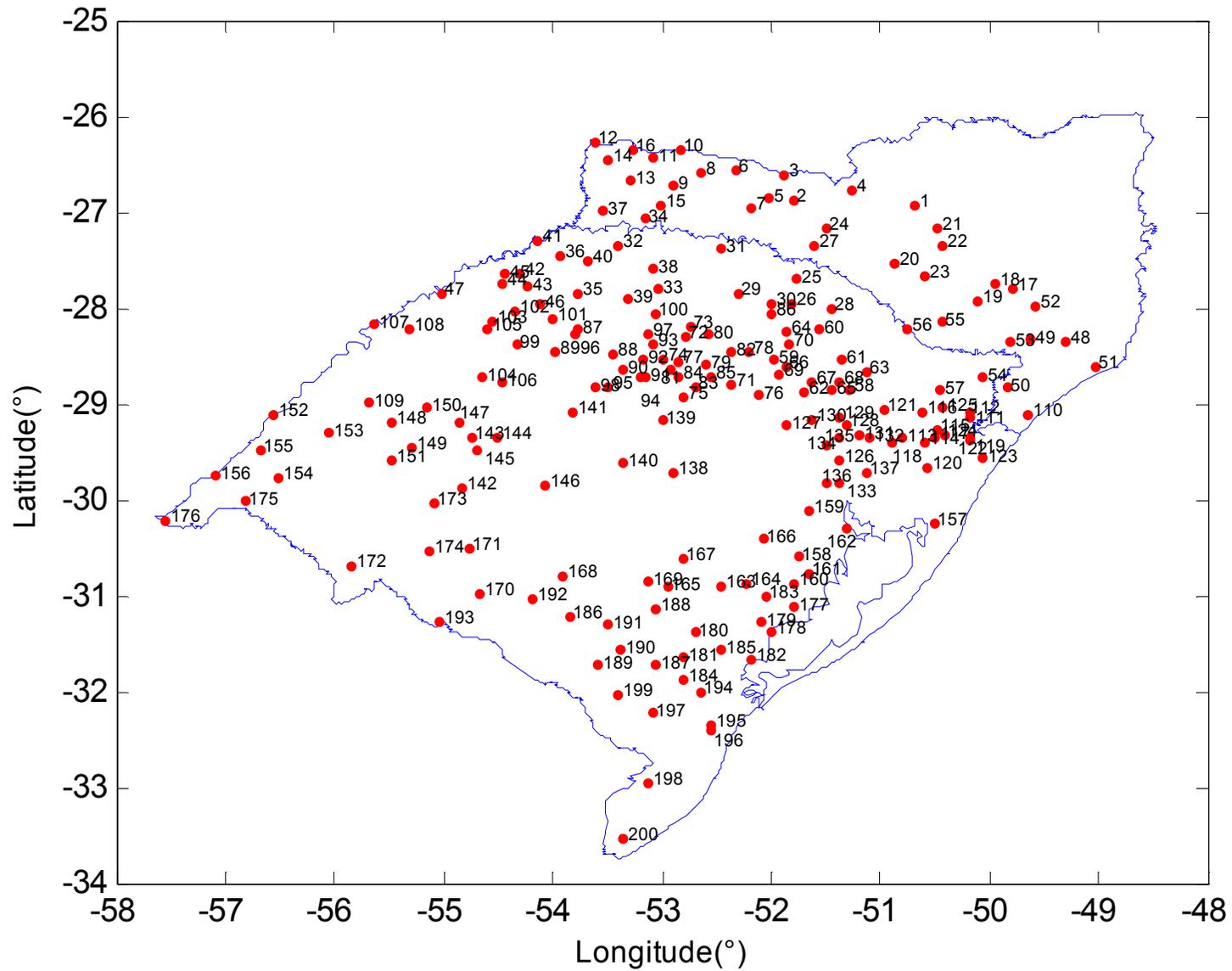


Figura 6.3: Distribuição espacial das estações pluviométricas utilizadas.

Tabela 6.2: Número identificador das estações pluviométricas e seus respectivos códigos, unidade da federação e ano de início da série.

ID	Código	UF	Ano	ID	Código	UF	Ano	ID	Código	UF	Ano	ID	Código	UF	Ano
1	2650019	SC	1976	51	2849020	SC	1976	101	2854001	RS	1969	151	2955008	RS	1976
2	2651001	SC	1943	52	2849021	SC	1943	102	2854003	RS	1943	152	2956005	RS	1970
3	2651022	SC	1972	53	2849023	SC	1976	103	2854004	RS	1961	153	2956006	RS	1969
4	2651036	SC	1976	54	2850002	RS	1961	104	2854005	RS	1957	154	2956007	RS	1943
5	2651040	SC	1976	55	2850004	SC	1958	105	2854006	RS	1959	155	2956008	RS	1976
6	2652000	SC	1957	56	2850008	RS	1943	106	2854012	RS	1976	156	2957003	RS	1977
7	2652001	SC	1969	57	2850009	RS	1943	107	2855001	RS	1949	157	3050002	RS	1974
8	2652002	SC	1972	58	2851003	RS	1943	108	2855002	RS	1957	158	3051004	RS	1976
9	2652021	SC	1976	59	2851005	RS	1944	109	2855005	RS	1963	159	3051005	RS	1966
10	2652031	SC	1976	60	2851007	RS	1962	110	2949003	SC	1976	160	3051016	RS	1976
11	2653001	SC	1969	61	2851008	RS	1944	111	2950003	RS	1947	161	3051017	RS	1976
12	2653002	SC	1972	62	2851009	RS	1961	112	2950008	RS	1946	162	3051023	RS	1976
13	2653004	SC	1969	63	2851010	RS	1962	113	2950009	RS	1941	163	3052007	RS	1944
14	2653005	SC	1972	64	2851020	RS	1962	114	2950010	RS	1947	164	3052009	RS	1943
15	2653007	SC	1954	65	2851021	RS	1957	115	2950011	RS	1941	165	3052010	RS	1955
16	2653013	SC	1976	66	2851022	RS	1957	116	2950019	RS	1941	166	3052011	RS	1969
17	2749031	SC	1957	67	2851024	RS	1944	117	2950023	RS	1960	167	3052012	RS	1976
18	2749035	SC	1976	68	2851026	RS	1944	118	2950026	RS	1944	168	3053007	RS	1944
19	2750007	SC	1958	69	2851027	RS	1960	119	2950027	RS	1960	169	3053015	RS	1944
20	2750008	SC	1959	70	2851028	RS	1962	120	2950028	RS	1961	170	3054002	RS	1969
21	2750010	SC	1959	71	2852004	RS	1962	121	2950033	RS	1944	171	3054016	RS	1976
22	2750012	SC	1961	72	2852006	RS	1941	122	2950034	RS	1960	172	3055003	RS	1976
23	2750020	SC	1976	73	2852007	RS	1944	123	2950038	RS	1974	173	3055004	RS	1976
24	2751004	SC	1943	74	2852008	RS	1944	124	2950041	RS	1947	174	3055005	RS	1976
25	2751006	RS	1957	75	2852009	RS	1962	125	2950048	RS	1976	175	3056004	RS	1976
26	2751007	RS	1957	76	2852014	RS	1944	126	2951005	RS	1962	176	3057002	RS	1976
27	2751012	SC	1976	77	2852015	RS	1959	127	2951010	RS	1943	177	3151002	RS	1976
28	2751017	RS	1976	78	2852016	RS	1944	128	2951013	RS	1961	178	3151003	RS	1943
29	2752006	RS	1943	79	2852023	RS	1969	129	2951017	RS	1963	179	3152002	RS	1966
30	2752014	RS	1962	80	2852024	RS	1962	130	2951019	RS	1961	180	3152003	RS	1943
31	2752017	RS	1976	81	2852029	RS	1958	131	2951022	RS	1943	181	3152005	RS	1976
32	2753002	RS	1961	82	2852031	RS	1959	132	2951023	RS	1961	182	3152008	RS	1966
33	2753004	RS	1959	83	2852032	RS	1959	133	2951024	RS	1970	183	3152011	RS	1944
34	2753006	SC	1959	84	2852036	RS	1963	134	2951026	RS	1962	184	3152013	RS	1976
35	2753007	RS	1948	85	2852037	RS	1972	135	2951027	RS	1970	185	3152016	RS	1964
36	2753009	RS	1961	86	2852046	RS	1976	136	2951028	RS	1964	186	3153003	RS	1956
37	2753013	SC	1976	87	2853001	RS	1961	137	2951036	RS	1961	187	3153004	RS	1976
38	2753014	RS	1976	88	2853002	RS	1963	138	2952003	RS	1965	188	3153006	RS	1976
39	2753015	RS	1976	89	2853003	RS	1959	139	2953007	RS	1969	189	3153007	RS	1966
40	2753016	RS	1976	90	2853007	RS	1961	140	2953008	RS	1943	190	3153008	RS	1965
41	2754001	RS	1949	91	2853008	RS	1941	141	2953030	RS	1976	191	3153017	RS	1976
42	2754003	RS	1960	92	2853011	RS	1957	142	2954001	RS	1943	192	3154003	RS	1976
43	2754007	RS	1960	93	2853012	RS	1959	143	2954004	RS	1959	193	3155001	RS	1976
44	2754008	RS	1960	94	2853014	RS	1944	144	2954005	RS	1966	194	3252005	RS	1966
45	2754009	RS	1976	95	2853015	RS	1960	145	2954007	RS	1945	195	3252006	RS	1964
46	2754010	RS	1976	96	2853016	RS	1969	146	2954019	RS	1976	196	3252008	RS	1965
47	2755001	RS	1949	97	2853018	RS	1972	147	2954020	RS	1976	197	3253001	RS	1965
48	2849001	SC	1939	98	2853020	RS	1972	148	2955002	RS	1959	198	3253003	RS	1966
49	2849009	SC	1969	99	2853023	RS	1976	149	2955006	RS	1969	199	3253004	RS	1965
50	2849019	SC	1976	100	2853026	RS	1976	150	2955007	RS	1969	200	3353007	RS	1961

### 6.3 Procedimentos aplicados

Este estudo de caso consiste na aplicação da metodologia proposta no item 5.3, para a análise regional de precipitações mensais e anuais na área de estudo anteriormente descrita.

A seguir, são apresentadas as particularidades dessa aplicação, procurando-se justificar as decisões tomadas em cada etapa.

#### a) Preparação da base de dados

De acordo com a metodologia seguida, as séries de precipitações mensais e anuais foram ajustadas à distribuição Gama e à distribuição Normal, respectivamente, determinando-se os parâmetros de ajuste pelo método da máxima verossimilhança.

Preliminarmente ao ajuste, foi necessário resolver o problema da ocorrência de precipitações nulas. Para evitar o uso de funções mistas, verificou-se a influência da substituição do valor zero por um número que tendesse a zero a partir de 1. O que se detectou foi que, à medida que os valores se aproximam de zero, maior é a deterioração do ajuste, o que conduz a uma crescente subestimação dos quantis inferiores e super-estimação dos superiores (Figura 6.4).

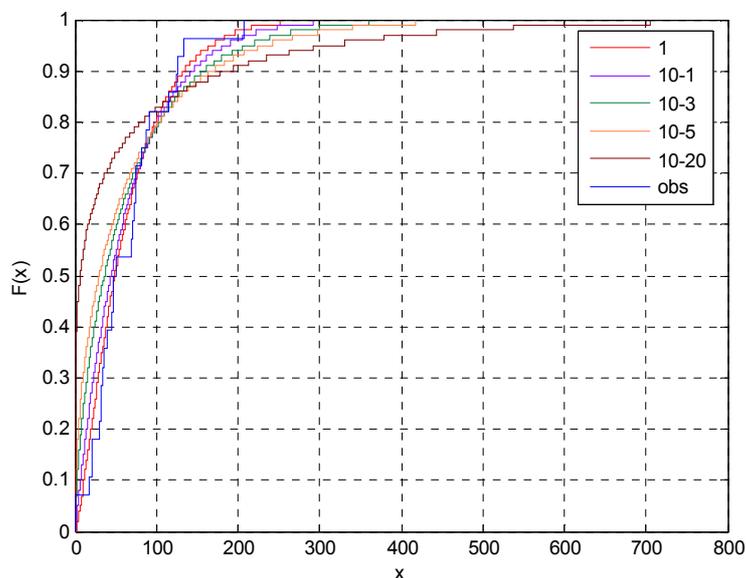


Figura 6.4 Ajuste da distribuição Gama para diferentes substituições do valor zero.

Desta forma, como, em termos práticos a ocorrência de precipitação mensal na ordem de 1mm ou a ausência de precipitação podem ser consideradas equivalentes, a precipitação nula foi substituída por 1mm. Isso não prejudicou o ajuste da cauda superior, enquanto que na cauda inferior entendeu-se ser irrelevante conhecer o tempo de retorno para frações do milímetro de chuva. É

possível, porém, que para durações menores, como a pentadial, seja mais prudente a utilização de uma função mista, ou a substituição do valor zero por 0,1mm, que é a menor altura efetivamente medida.

### b) Variáveis utilizadas no agrupamento

Foram adotadas como variáveis do agrupamento as coordenadas geográficas latitude, longitude e altitude, e as médias das séries (mensais e anual). A matriz de dados, constituída então de 16 variáveis e 200 pontos, foi escalonada aplicando-se a equação (4.8), de forma que todos os valores passaram a variar no intervalo [0,1]. Os fatores de mudança de escala, necessários para uma possível alocação de pontos externos à matriz, encontram-se na Tabela 6.3.

Tabela 6.3: Fatores de mudança de escala.

	Longitude	Latitude	Altitude
	(°)	(°)	(m)
Mínimo	-57,552	-33,533	2
Máximo	-49,033	-26,265	1250

	Precipitação média												
	Jan	Fev	Mar	Abr	Mai	Jun	Jul	Ago	Set	Out	Nov	Dez	Anual
Mínimo	90,4	96,4	83,1	86,3	81,0	64,0	56,6	37,4	70,6	95,0	91,2	82,7	1076,9
Máximo	235,0	268,8	199,1	217,6	202,1	192,9	207,9	160,3	205,4	254,1	210,2	202,9	2187,3

### c) Algoritmo de agrupamento

O método FCM, descrito no item 4.2, foi aplicado para o agrupamento da matriz de dados composta pelas variáveis citadas anteriormente.

O valor de  $m$  foi previamente investigado em um grupamento de quatro grupos (Figura 6.5), adotando-se  $m=1,5$  por apresentar melhor distribuição dos máximos graus de pertinência, condizente com as observações colocadas no item 4.2.

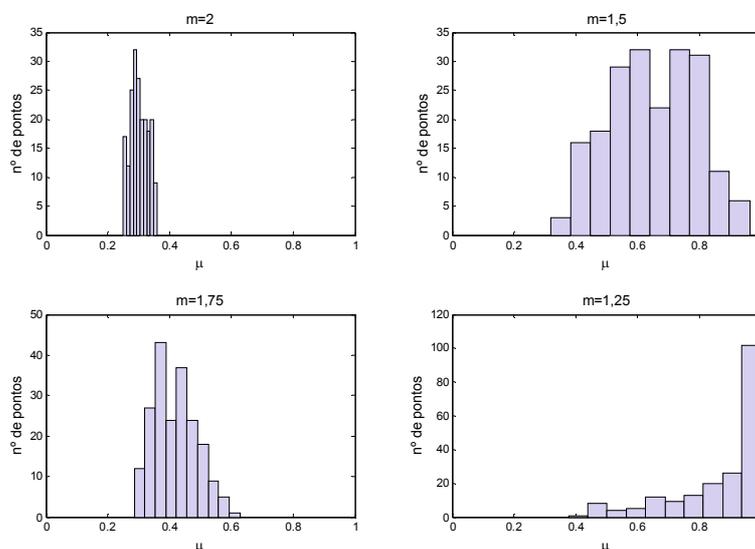


Figura 6.5 Frequência dos graus de pertinência máximos.

#### d) Ponderação das variáveis

As alternativas de ponderação aqui testadas se restringiram à aplicação de um ponderador ( $w$ ) ao atributo equivalente ao período (um dos meses ou o ano) com maior número de séries discordantes no teste de aderência regional.

Foram testados valores de  $w$  iguais a 1, 2, 3, 4 e 5. Cada matriz ponderada foi agrupada com número de grupos de 2 a 14, e cada resultado foi submetido ao teste de aderência regional.

Foi escolhida a alternativa de ponderação que apresentasse melhoria na média dos valores de  $S_d$  (médio e de cada período).

#### e) Determinação do número de grupos

Para determinação do melhor número de regiões, foram avaliados os comportamentos dos índices de validade  $Pd$ ,  $V_{XB}$ ,  $V_K$ ,  $V_{FK}$  e  $V_{PB}$ , cujas formulações encontram-se no item 4.5.

Numa tentativa de eliminar os efeitos do compartilhamento mínimo ( $1/c$ ), identificado no índice de partição difusa ( $Pd$ ), foi experimentada uma adaptação desse coeficiente, subtraindo-se o valor de  $1/c$ , numa tentativa de identificar as flutuações e tornar mais fácil a identificação das configurações mais favoráveis.

Essa nova configuração foi denominada Partição Difusa significativa ( $Pds$ ), cujo valor máximo é o indicativo do melhor agrupamento. A formulação é apresentada na equação (6.1).

$$Pds = \frac{1}{N} \sum_{i=1}^c \sum_{k=1}^N (\mu_{ik})^2 - 1/c \quad (6.1)$$

Procurou-se selecionar as melhores opções de grupos indicadas pelos graus de pertinência e comparar com o resultado do teste de aderência regional para se chegar à determinação do número de grupos que a ser adotado.

#### f) Verificação de falhas de alocação

A verificação de falhas de alocação seguiu as prescrições do item 5.3.5 da metodologia, avaliando-se a possibilidade de má alocação apenas para os pontos com duas ou mais séries discordantes no teste de aderência regional, realocando-se apenas os que apresentaram, devido à mudança de região, diminuição em pelo menos duas séries do número total de séries discordantes.

Procurou-se também verificar graficamente a existência de relações entre o comportamento dos graus de pertinência e o número total de séries discordantes em cada estação.

### **g) Verificação da homogeneidade**

Após a determinação das regiões e de efetuados os ajustes preliminares devidos à má alocação, aplicou-se o teste de homogeneidade descrito no item 5.2, averiguando-se o nível de qualidade dos ajustes regionais.

A condição de homogeneidade também foi verificada em termo de distribuição adimensional, mas apenas a título de ilustração, dando-se prioridade à análise da condição de homogeneidade da distribuição não adimensional.

Nessas circunstâncias, as situações de não homogeneidade mais críticas foram investigadas mais profundamente, verificando-se a viabilidade de subdivisão das regiões envolvidas.

Após todas as interferências, é apresentado o produto final da análise regional constituído da descrição climatológica de cada uma das regiões e da apresentação dos parâmetros das distribuições regionais em cada período, expondo-se também os correspondentes quantis regionais para os tempos de retorno mais usuais.

### **h) Modelo para alocação de pontos não amostrados**

Como não foram separadas amostras para essa verificação, o objetivo aqui é apenas ajustar o modelo de alocação para novos pontos, que seria baseado apenas nas coordenadas geográficas e na precipitação média anual, considerando que essa última pode ser bem estimada por métodos de espacialização (como o método do inverso da distância ponderada e o método de Kriging).

A matriz de dados e o centro do agrupamento foram padronizados pela equação (4.4), utilizando-se, como fator de escala, os dados da Tabela 6.3, correspondentes às quatro variáveis. A ponderação das variáveis foi, então, pesquisada para que a nova alocação atingisse bons resultados no teste de aderência regional, aplicando-se esta ponderação tanto à matriz de dados, quanto ao centro do agrupamento.

A nova matriz de pertinência foi calculada e os pontos foram alocados no grupo ao qual pertencessem com o máximo grau. Foi aplicado o teste de aderência regional, conservando-se as distribuições regionais do agrupamento original e comparando-se, em seguida, seus resultados com os do agrupamento original.

### **i) Estimativa do fator de escala local através dos graus de pertinência**

Caso seja considerado pertinente a utilização da distribuição regional apenas em termos de escala, faz-se necessário conhecer o fator de escala (nesse caso a média das observações) no ponto de interesse.

Procedeu-se, então, a verificação da qualidade das estimativas do fator de escala em cada local, a partir da média dos protótipos (centros do agrupamento), ponderadas pelos respectivos graus de pertinência do ponto.

Desta forma, a precipitação média em cada mês e a anual, para cada um dos 200 pontos, foi estimada pela equação (3.5), e comparada com as médias observadas, através da raiz do erro médio quadrático relativo (REMQR), dado pela fórmula abaixo:

$$REMQR = \left( \frac{1}{N} \sum_{n=1}^N ((x_i - x_{ei}) / x_i)^2 \right)^{1/2} \quad (6.2)$$

onde,  $x_i$  é o valor observado em  $i$ ,  $x_{ei}$  é o valor estimado de  $x_i$  e  $N$  é o total de séries.

O erro na estimativa por graus de pertinência foi comparado com o erro existente ao se adotar o protótipo de cada região como estimativa pontual. Nesse caso, foi utilizada a mesma formulação (Eq. 6.1), onde  $x_{ei}$  é o valor do protótipo da região à qual pertence o ponto  $i$ .

## 7 RESULTADOS E DISCUSSÃO

### 7.1 Verificação do método de obtenção das distribuições regionais

A análise da estimativa da distribuição regional média dos parâmetros de ajuste das séries locais, realizado anteriormente à aplicação, revelou que, para a distribuição Normal, o erro médio na estimativa dos parâmetros ficou em torno de zero e as máximas diferenças de probabilidades se mostraram dependentes do número e tamanho das séries na região, apresentando para regiões de 20 séries com 30 anos de dados,  $\mu dp$  igual a 0,036.

Tabela 7.1: Valores de  $\mu dp$  e  $\sigma dp$  para a distribuição Normal.

n° de séries	Distribuição Normal					
	Tamanho das séries					
	20		30		50	
	$\mu dp$	$\sigma dp$	$\mu dp$	$\sigma dp$	$\mu dp$	$\sigma dp$
5	0,069	0,028	0,050	0,021	0,036	0,015
10	0,057	0,174	0,041	0,014	0,028	0,011
20	0,052	0,009	0,036	0,009	0,023	0,008
50	0,050	0,003	0,033	0,002	0,021	0,003

Entretanto, quando a distribuição é do tipo Gama, a média dos parâmetros locais gera uma sub-estimativa em relação à função de distribuição de probabilidades original, que ocorre, provavelmente, devido à não linearidade envolvida na determinação dos parâmetros de ajuste.

Observou-se que os valores de  $\mu dp$  e  $\sigma dp$ , para a distribuição Gama, foram mais elevados que os apresentados pela distribuição Normal. (Tabela 7.2)

Tabela 7.2: Valores de  $\mu dp$  e  $\sigma dp$  para a distribuição Gama.

n° de séries	Distribuição Gama						Distribuição Gama corrigida					
	Tamanho das séries						Tamanho das séries					
	20		30		50		20		30		50	
	$\mu dp$	$\sigma dp$	$\mu dp$	$\sigma dp$	$\mu dp$	$\sigma dp$	$\mu dp$	$\sigma dp$	$\mu dp$	$\sigma dp$	$\mu dp$	$\sigma dp$
5	0,104	0,055	0,073	0,038	0,044	0,024	0,076	0,031	0,053	0,025	0,039	0,017
10	0,110	0,049	0,073	0,032	0,044	0,021	0,061	0,022	0,045	0,017	0,031	0,013
20	0,108	0,037	0,073	0,025	0,042	0,015	0,055	0,015	0,037	0,010	0,027	0,010
50	0,111	0,025	0,069	0,017	0,040	0,011	0,051	0,008	0,033	0,002	0,022	0,005

Verificou-se, ainda, que método superestima o parâmetro alfa e subestima o parâmetro beta, com erros equivalentes a 40% e 20% dos desvios dos parâmetros alfa e beta locais, respectivamente, devendo-se calcular estes parâmetros pela forma corrigida:

$$\hat{\alpha}_R = \left( \frac{1}{N} \sum_{i=1}^N \hat{\alpha}_i^R \right) - 0,4 \left( \sqrt{\frac{1}{N} \sum_{i=1}^N (\hat{\alpha}_i^R - \bar{\alpha}^R)^2} \right), \quad (7.1)$$

$$\hat{\beta}_R = \left( \frac{1}{N} \sum_{i=1}^N \hat{\beta}_i^R \right) + 0,2 \left( \sqrt{\frac{1}{N} \sum_{i=1}^N (\hat{\beta}_i^R - \bar{\beta}^R)^2} \right), \quad (7.2)$$

onde  $\hat{\alpha}_R$  e  $\hat{\beta}_R$  são os parâmetro alfa e beta da distribuição Gama regional,  $\hat{\alpha}_i$  e  $\hat{\beta}_i$  os parâmetros alfa e beta ajustados para a série local  $i$ , e  $\bar{\alpha}$  e  $\bar{\beta}$  a média dos  $\hat{\alpha}_i$  e  $\hat{\beta}_i$ , respectivamente.

Assim, para regiões com 20 séries de 30 anos, a média das diferenças máximas de probabilidade entre a distribuição original e as estimadas foi de 0,073, caindo para 0,037 após a correção dos parâmetros (ver ilustração na Figura 7.1, onde a curva em vermelho é a distribuição original e as em azul são as 500 estimativas em uma das situações simuladas).

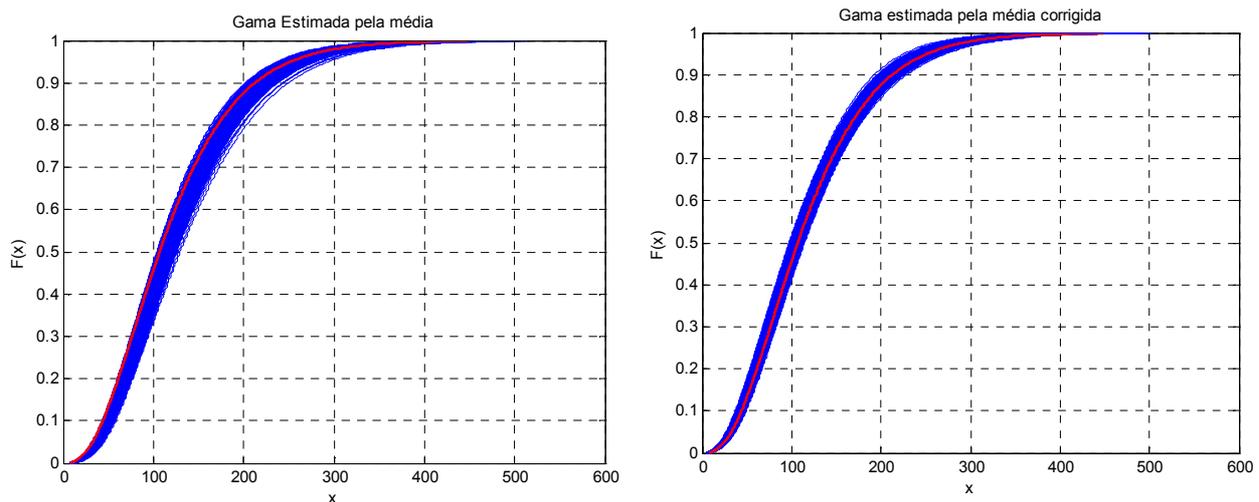


Figura 7.1 Curvas regionais Gama estimadas pela média dos parâmetros locais.

A Tabela 7.3 apresenta os valores da média e do desvio padrão da máxima diferença de probabilidades verificada no ajuste, de séries de tamanho  $n$ , às distribuições Normal e Gama.

Tabela 7.3: Média e desvio padrão da máxima diferença de probabilidades dos ajustes.

n	Normal		Gama		n	Normal		Gama	
	média	desvio	média	desvio		média	desvio	média	desvio
28	0,115	0,029	0,118	0,031	400	0,032	0,008	0,032	0,008
30	0,114	0,029	0,117	0,029	450	0,030	0,007	0,030	0,008
50	0,090	0,022	0,089	0,021	500	0,028	0,007	0,028	0,007
100	0,062	0,015	0,063	0,015	600	0,026	0,007	0,026	0,007
150	0,051	0,012	0,050	0,012	700	0,024	0,006	0,025	0,006
200	0,045	0,010	0,046	0,011	800	0,022	0,005	0,023	0,005
250	0,039	0,009	0,042	0,009	900	0,021	0,005	0,021	0,005
300	0,036	0,008	0,037	0,010	1000	0,020	0,005	0,020	0,005
350	0,034	0,008	0,034	0,008					

Observa-se que, em regiões homogêneas com 20 séries de 30 anos, os valores de  $\mu_{dp}$  equivalem, respectivamente, a um ajuste local de 300 anos de dados, independentemente da distribuição.

Como este tamanho de série é razoavelmente grande, pode-se considerar que a média dos parâmetros locais fornece uma estimativa satisfatória da distribuição verdadeira, desde que sejam aplicadas as devidas correções na distribuição Gama.

Desta forma, em todas as situações do estudo de caso em que foi necessário determinar a distribuição regional, seus parâmetros foram calculados pela média dos parâmetros de ajuste locais para a distribuição Normal, e, para a distribuição Gama, os parâmetros foram calculados pelas fórmulas 7.1 e 7.2.

## 7.2 Identificação das regiões homogêneas

### 7.2.1 Ponderação das variáveis.

Observados os resultados preliminares dos testes de aderência regional (Figura 7.2), verificou-se que as distribuições anuais apresentaram maior dificuldade de aderência à distribuição regional que as distribuições mensais.

Na tentativa de obter alguma melhoria no ajuste regional, foi considerada a ponderação da média anual com os valores inteiros de 1 a 5, obtendo-se como melhor resultado geral no teste de aderência regional o ponderador igual a 3.

A Tabela 7.4 apresenta a média do percentual de séries discordantes nos agrupamentos de 2 a 14 grupos, para cada um dos períodos de referência, e os diversos ponderadores testados. Verificando-se que, para  $w=3$ , ocorre uma pequena melhoria para todos os meses, e que, a partir daí, os meses de fevereiro a agosto ficam prejudicados.

Tabela 7.4: Média de Sd (%) para diversos ponderadores.

w	Séries Discordantes (%)												ano	média
	jan	fev	mar	abr	mai	jun	jul	ago	set	out	nov	dez		
1	3,85	4,46	4,12	3,15	0,31	4,92	6,62	4,62	3,46	1,42	0,50	4,92	8,23	3,89
2	3,81	3,73	4,31	3,46	0,42	5,27	6,73	4,27	3,08	1,15	0,38	4,50	5,92	3,62
3	3,77	3,62	4,62	3,42	0,54	5,23	7,54	4,69	2,54	1,23	0,46	4,58	3,65	3,53
4	3,50	3,73	5,04	4,15	0,62	4,88	8,50	5,65	2,19	1,27	0,35	4,15	2,65	3,59
5	3,62	3,65	5,19	4,58	0,58	5,35	8,77	5,65	2,35	1,42	0,38	4,92	2,42	3,76

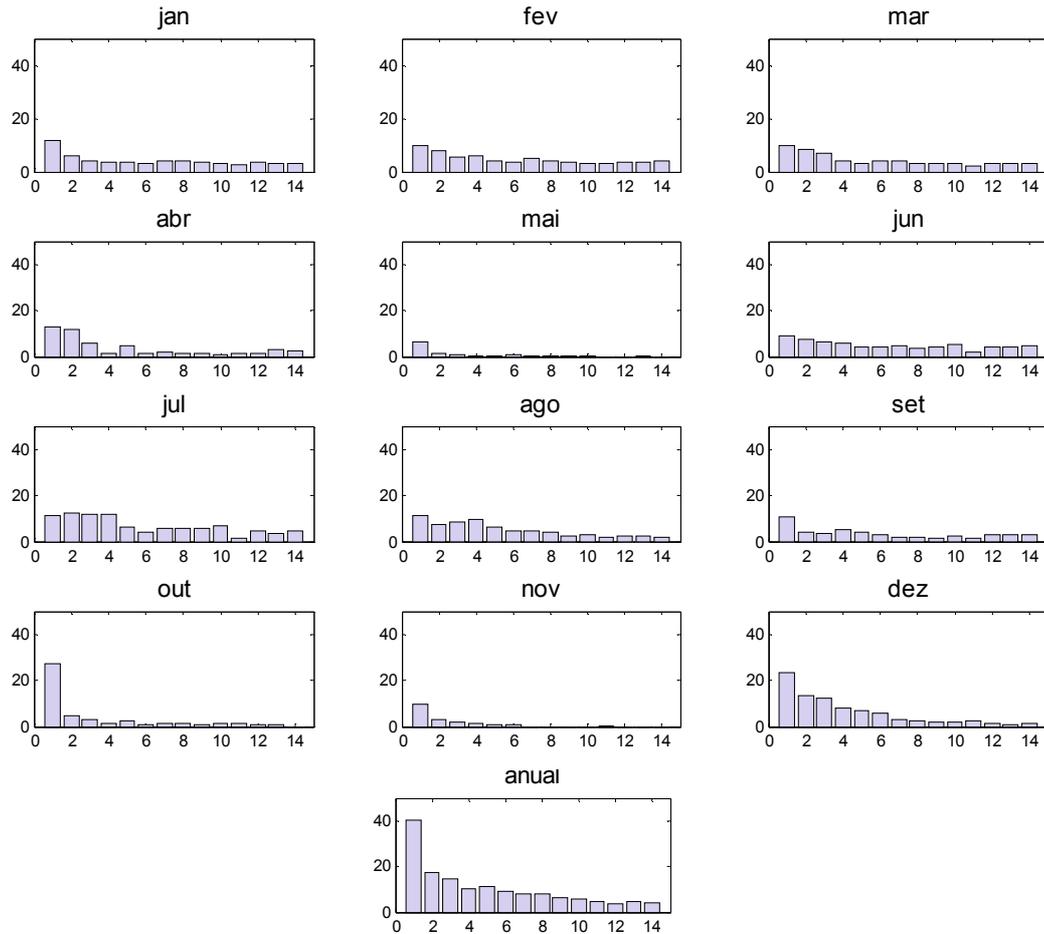


Figura 7.2: Percentual de séries discordantes na matriz não ponderada.

## 7.2.2 Determinação do número de regiões

### 7.2.2.1 Análise dos Índices de validade.

Foi avaliado o desempenho dos índices de validade Fukuyama-Sugeno ( $V_{FS}$ ), Xie-Beni ( $V_{XB}$ ) e Kwon ( $V_K$ ), cujos valores mínimos indicam bons agrupamentos, e os índices de Partição difusa ( $Pd$ ), PBMF ( $V_{PB}$ ) e Partição Difusa significativa ( $Pds$ ), nos quais os bons agrupamentos são indicados por valores máximos de cada índice.

De maneira geral, o extremo global é indicativo do número ótimo de grupos, embora extremos locais possam ser considerados como bons agrupamentos, podendo ser adotados a depender da finalidade da análise e do nível de detalhamento requerido.

Percebe-se pela Figura 3.1 que os índices  $Pd$ ,  $V_{FS}$  e  $V_{PB}$  apresentaram comportamento decrescente com o aumento do número de grupos. No entanto, enquanto  $V_{FS}$  aponta sempre para o máximo número de grupos testado (14 grupos),  $Pd$  e  $V_{PB}$  apontam a menor divisão (2 grupos) como a melhor configuração para o agrupamento.

Considerando-se também os máximos locais,  $V_{PB}$  também aponta discretamente 6 e 8 grupos como boa configuração de agrupamento.

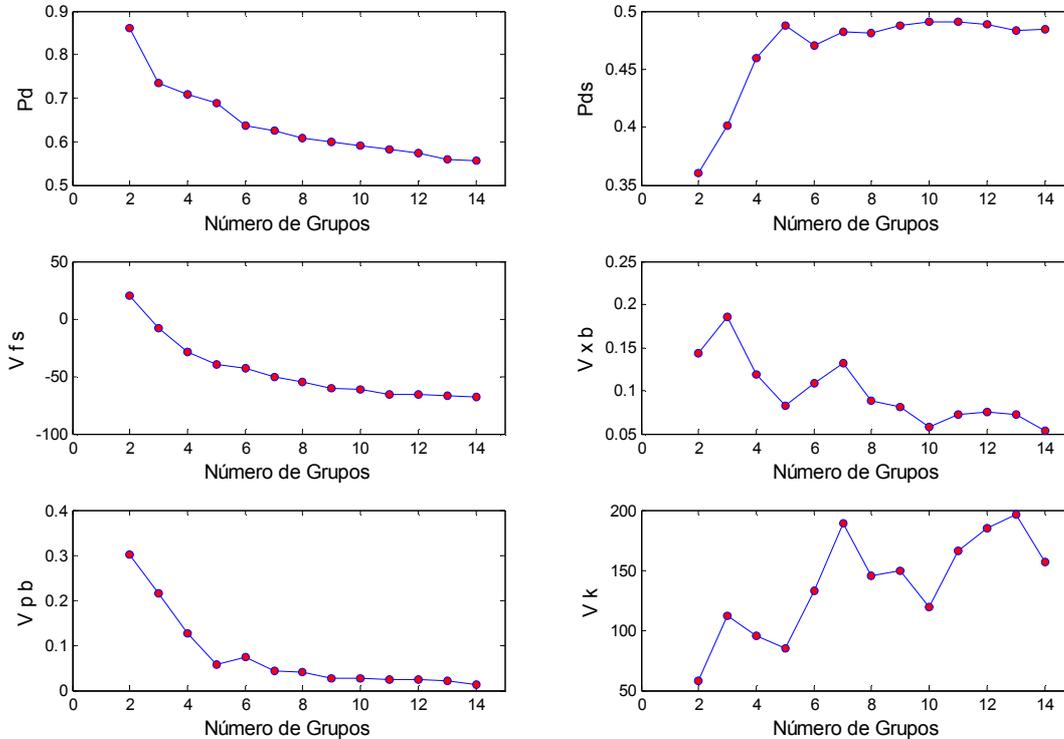


Figura 7.3: Índices de validade.

Ao contrário dos três índices anteriores, os índices  $V_{XB}$ ,  $V_K$  e  $Pds$  apresentaram resultados mais oscilantes. Isso significa que, além de um valor ótimo global, os índices conseguiram indicar algumas alternativas ao número de classes (extremos locais), que podem ser consideradas como bons agrupamentos.

Como ótimo global, cada um dos três índices apontou um resultado diferente, sendo dois grupos para o índice de Kwon (mesmo resultado de  $V_{PB}$  e  $Pd$ ), treze grupos para o índice de Xie-Beni e quatro grupos para a Partição Difusa Significativa.

Apesar de diferentes ótimos globais (para os limites estabelecidos), os índices  $V_{XB}$ ,  $V_K$  e  $Pds$  apresentaram alguns ótimos locais coincidentes.

Sendo 2, 4, 6, e 13 apontados ao mesmo tempo por  $V_{XB}$  e  $V_K$ , entretanto, para o  $Pds$ , 2 grupos é apontada como uma configuração de agrupamento ruim, 4 como uma configuração ótima, 6 grupos, embora apresente alto valor, não constitui um pico, e 13 grupos é a melhor opção se os objetivos requerem mais de 10 grupos.

Resultados semelhantes entre  $V_{XB}$ ,  $V_K$  são esperados, uma vez que o índice de Kwon surgiu de uma tentativa de eliminar a monotonicidade, verificada no índice de Xie-Beni para alguns conjuntos de dados. A Partição Difusa Significativa apresentou pontos de coerência com  $V_{XB}$  e  $V_K$ , revelando a significância de cada configuração de agrupamento.

### 7.2.2.2 Resultados do teste de aderência regional

Percebe-se pelos resultados da Tabela 7.5, que a distribuição da precipitação mensal possui menor variação espacial que a anual. Com exceção dos meses de outubro e dezembro, a adoção de um único grupo e, conseqüentemente, uma única distribuição (para cada mês), para toda a área de estudo, seria aceitável para mais de 80% das estações. Isso significa que, para esse nível de aceitação, não haveria necessidade de agrupamento dos dados.

Já para a distribuição anual menos de 60% das estações passariam no teste de aderência regional no caso de se adotar apenas uma distribuição para a totalidade da área de estudo.

Tabela 7.5 Resultado do teste de aderência regional.

Nº de Grupos	Sd (%)												ano KS	média
	jan	fev	mar	abr	mai	jun	jul	ago	set	out	nov	dez		
1	12,0	11,0	9,0	12,5	6,0	9,0	13,0	11,5	9,0	28,0	9,0	20,0	40,5	14,7
2	8,0	7,5	8,5	9,5	1,0	7,5	14,0	8,5	4,0	4,5	2,0	15,0	14,5	8,0
3	5,0	7,0	6,5	6,5	1,0	7,0	12,0	8,0	3,0	3,0	2,0	11,0	7,5	6,1
4	3,5	3,5	5,5	3,5	0,5	6,0	11,5	9,5	3,5	2,0	1,0	7,0	5,5	4,8
5	3,0	4,0	4,5	5,5	0,5	5,0	10,0	6,5	3,0	2,0	1,0	6,0	5,5	4,3
6	3,5	4,0	3,5	1,5	0,5	5,0	7,5	5,5	3,0	1,0	0,5	6,0	5,0	3,6
7	3,0	3,5	3,0	1,0	0,5	5,0	7,0	5,5	3,0	1,0	0,5	4,5	4,0	3,2
8	3,5	3,5	3,5	1,0	0,5	6,0	5,5	3,5	2,0	0,5	0,0	2,0	1,5	2,5
9	3,5	2,0	4,0	1,5	0,5	4,5	7,0	3,5	2,0	0,0	0,0	1,5	1,0	2,4
10	3,0	2,0	4,5	1,5	0,5	3,0	5,5	3,0	1,5	0,5	0,0	1,5	1,5	2,2
11	3,5	2,0	4,0	2,5	0,5	5,0	6,0	2,0	2,5	0,5	0,0	1,5	1,0	2,4
12	3,0	2,5	4,0	3,0	0,5	5,5	4,5	2,0	2,5	0,0	0,0	1,5	1,0	2,3
13	2,5	2,0	3,5	3,5	0,5	5,5	4,0	2,0	2,5	0,0	0,0	1,5	1,5	2,2
14	2,5	2,5	3,5	3,0	0,5	4,0	3,0	1,5	2,0	0,0	0,0	1,5	2,0	2,0

As linhas destacadas na Tabela 7.5 indicam as opções de número de regiões indicadas como bons agrupamentos pelos índices de validade. Note-se que as opções de 8, 10 e 14 não apresentaram diferenças efetivas no teste de aderência regional.

Desta forma, adotou-se a configuração de 8 regiões para representar as distribuições de precipitação na área de estudo, com uma média de 2,2 séries discordantes em cada período analisado, sendo os meses de junho e julho os que apresentam maior discordância (6,0% e 5,5%, respectivamente).

### 7.2.3 Ajustes no agrupamento

De acordo com o critério de qualidade de alocação adotado, foram encontradas 12 estações com duas ou mais séries discordantes no teste de aderência regional. Com elas, verificou-se o grau de diminuição do total de séries discordantes, caso houvesse mudança de região, sendo encontradas duas estações cujo grau de diminuição foi superior a um (nesse caso igual a dois), o que caracterizou as estações como má alocadas.

Assim, essas duas estações (51 e 120) foram realocadas, sendo a estação 51 da região 8 para a região 7 e a 120 da região 8 para a região 6.

Isso significa que o método de agrupamento foi eficiente, na distribuição dos elementos nos grupos. Caso o critério de má alocação considerasse qualquer melhoria com a mudança de grupo, ela seria dirigida a 12 estações (6%), o que continua indicando o bom desempenho do método de agrupamento.

O gráfico da Figura 7.4 apresenta a relação entre os máximos graus de pertinência ( $\mu_{max}$ ) e o número de séries discordantes em cada estação pluviométrica, sendo identificadas as 12 com duas ou mais séries discordantes.

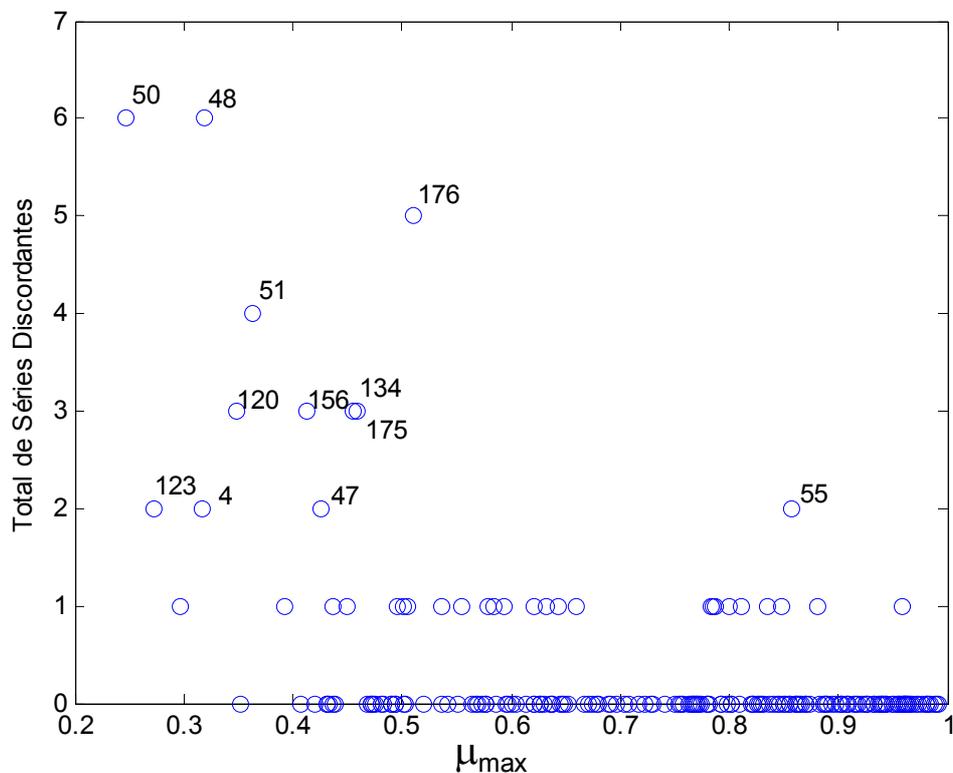


Figura 7.4 Relação entre  $\mu_{max}$  e o resultado do teste de aderência regional.

Nota-se que maioria das estações com elevado número de séries discordantes possui máximo grau de pertinência de baixo valor. Isso significa que esses pontos tiveram dificuldade de alocação, dividindo-se entre os vários grupos existentes.

Poderia se deduzir que os pontos com baixo máximo grau de pertinência estão mais sujeito a erros de alocação, ou sem similaridade com os demais pontos. Entretanto, há casos em que baixos valores de  $\mu_{max}$  podem significar apenas que o ponto é bem adaptado a vários grupos, o que seria comum nas fronteiras das regiões.

As estações 48, 50 e 176 foram as que apresentaram o maior número de séries discordantes, devendo ser analisadas mais de perto. Essas estações localizam-se, respectivamente, nos extremos leste (48, 50) e oeste (176) da área de estudo e apresentam regime de precipitação diferenciado das demais estações.

A estação Barra do Quaraí (176) é a que apresenta a menor precipitação de toda a área de estudo, além disso possui distribuição ao longo do ano também diferenciada. Apesar das diferenças com relação à sua região (Região1), seu máximo grau de pertinência não é baixo (em relação aos demais pontos problemáticos), pelo fato de não haver concorrência de outras regiões por este ponto, já que é um mínimo extremo. A estação Fazenda Junco (175) segue as tendências da estação 176, em termos da distribuição de chuvas durante o ano, entretanto os volumes precipitados já são consideravelmente maiores, o que leva a uma aproximação das regiões adjacentes, diminuindo um pouco o grau de pertinência.

As estações Orleans-montante (48) e Timbé do Sul (50) em Santa Catarina, não apresentaram similaridade suficiente aos demais pontos da região em estudo e possivelmente constituiriam outra região nas escarpas do planalto voltadas para o mar, entretanto o reduzido número de pontos nas mesmas condições provavelmente impede o algoritmo de detectá-la, quando se tenta elevar o número de grupos desejado.

A estação 51, embora tenha apresentado menos séries discordantes ao ser alocada na região 7, a mudança para essa região não é fisicamente coerente, já que a região é caracteristicamente de elevadas altitudes, e a estação 51 localiza-se no litoral de Santa Catarina, com 10m de altitude.

Desta forma, as estações 48, 50 e 51 foram excluídas da análise regional, tendo em vista que podem ser melhor analisadas em conjunto com outras estações próximas a elas e que, embora existentes, não foram consideradas nesse estudo.

A estação 176, contudo, foi preservada, pois unida às estações 175 e 156, caracteriza o regime pluviométrico do extremo oeste do Rio Grande do Sul. Estas estações também poderiam

formar um grupo independente, mas, para este grupo ser detectado no agrupamento, seria necessário utilizar informações do território uruguaio, que são menos disponíveis.

Deixou-se, portanto, para a fase do teste de homogeneidade a verificação da necessidade de separação desses pontos.

Após a realocação e exclusão dos pontos aqui citados, as curvas regionais foram recalculadas, seguindo-se com o teste de homogeneidade.

#### 7.2.4 Teste de Homogeneidade

As tabelas 7.6 e 7.7 apresentam, para a condição de curva real e adimensional respectivamente, os valores de  $\mu dp_s$  e  $\sigma dp_s$  para as distribuições Gama e Normal, bem como os limites para as condições de homogeneidade aceitável (até  $\mu + 1\sigma$ ), de possível heterogeneidade (até  $\mu + 2\sigma$ ) e de heterogeneidade definitiva (após  $\mu + 2\sigma$ ), obtidos a partir das 500 regiões simuladas de cada uma das oito regiões identificadas.

Tabela 7.6: Valores de  $\mu dp_s$  e  $\sigma dp_s$  e os limites de homogeneidade, curva real.

Região	nº de pontos	dps				Limites			
		Gama		Normal		Gama		Normal	
		$\mu$	$\sigma$	$\mu$	$\sigma$	$\mu + 1\sigma$	$\mu + 2\sigma$	$\mu + 1\sigma$	$\mu + 2\sigma$
1	22	0,161	0,010	0,161	0,011	0,171	0,181	0,172	0,182
2	12	0,160	0,014	0,162	0,014	0,175	0,189	0,176	0,190
3	30	0,161	0,009	0,160	0,009	0,170	0,179	0,169	0,178
4	27	0,160	0,010	0,161	0,010	0,170	0,180	0,170	0,180
5	13	0,160	0,014	0,162	0,013	0,174	0,189	0,175	0,188
6	43	0,161	0,007	0,160	0,008	0,168	0,176	0,168	0,176
7	25	0,161	0,010	0,161	0,010	0,171	0,181	0,171	0,181
8	26	0,160	0,010	0,161	0,010	0,170	0,180	0,170	0,180

Tabela 7.7: Valores de  $\mu dp_s$  e  $\sigma dp_s$  e os limites de homogeneidade, curva adimensional.

Região	nº de pontos	dps				Limites			
		Gama		Normal		Gama		Normal	
		$\mu$	$\sigma$	$\mu$	$\sigma$	$\mu + 1\sigma$	$\mu + 2\sigma$	$\mu + 1\sigma$	$\mu + 2\sigma$
1	22	0,128	0,008	0,123	0,007	0,135	0,143	0,130	0,137
2	12	0,128	0,010	0,123	0,009	0,138	0,148	0,132	0,140
3	30	0,128	0,006	0,123	0,006	0,134	0,140	0,129	0,136
4	27	0,128	0,006	0,123	0,006	0,135	0,141	0,129	0,135
5	13	0,128	0,009	0,123	0,009	0,137	0,146	0,132	0,140
6	43	0,128	0,005	0,123	0,005	0,134	0,139	0,127	0,132
7	25	0,128	0,006	0,123	0,006	0,135	0,141	0,130	0,136
8	26	0,128	0,006	0,123	0,006	0,135	0,141	0,129	0,135

Em cada região os resultados de  $\mu dp_s$  e  $\sigma dp_s$  foram extremamente semelhantes nos diferentes meses, apresentando em média coeficiente de variação de 0,25%. Ou seja, o critério de

homogeneidade mostrou-se independente dos parâmetros da distribuição, mas dependente das condições regionais de número de séries e do tipo da distribuição. Dessa forma, foi adotado um mesmo critério de homogeneidade ( $\mu dp_s$  e  $\sigma dp_s$ ) para todos os meses.

Nota-se que os valores de  $\mu dp_s$  foram bastante próximos para todas as regiões, enquanto os desvios de  $dp_s$  foram inversamente proporcionais ao número de séries em cada uma delas.

Observa-se também que, na análise das curvas adimensionais, os valores de  $\mu dp_s$  são muito menores que os apresentados para as curvas reais. Ou seja, para se atingir a condição de homogeneidade deve haver maior proximidade entre as curvas adimensionais do que seria necessário na análise não adimensional. O que parece coerente já que com a adimensionalização são eliminadas as flutuações de média.

Salienta-se que, nas duas condições, os valores de limites para a possível heterogeneidade ( $\mu + 2\sigma$ ) são sempre menores que o valor crítico do teste KS para o nível de significância de 20%, indicando um nível mais rigoroso para o teste de homogeneidade.

A Tabela 7.8 apresenta o resultado do teste de homogeneidade em cada região, na qual os elementos hachurados indicam os casos que apresentaram  $H_{dp} > 1$ , que somaram 13 das 104 condições analisadas (13 períodos e 8 regiões). Percebe-se que o mês de julho foi o que apresentou maior nível de heterogeneidade, com  $H_{dp} > 1$  em cinco das oito regiões. Com relação à homogeneidade regional, destacam-se as regiões 1 e 7, que apresentaram mais de um período com  $H_{dp} > 1$ .

Tabela 7.8. Teste de homogeneidade  $H_{dp}$  para a distribuição regional em escala real.

Região	Teste de homogeneidade $H_{dp}$												
	Jan	Fev	Mar	Abr	Mai	Jun	Jul	Ago	Set	Out	Nov	Dez	Ano
1	-1,78	0,47	-0,50	-1,46	-1,30	6,27	4,01	2,24	1,92	-0,98	-0,01	-0,33	0,43
2	-2,25	0,02	1,46	-2,25	-1,84	-0,60	0,63	0,02	0,84	-1,84	-2,25	-2,04	-1,12
3	-1,42	-1,80	0,25	-0,13	-3,86	-1,29	1,28	-0,65	-1,16	0,77	-2,83	0,00	-1,31
4	-2,21	-1,95	-1,82	-2,21	-2,08	-3,13	2,39	0,29	-2,60	-2,74	-2,87	-1,16	-3,23
5	-1,82	-0,85	-2,80	-1,82	-2,41	-0,65	1,10	-1,04	-1,43	-1,24	-1,24	-1,82	-0,80
6	0,58	1,38	-2,51	-1,83	-1,83	-2,28	0,81	0,46	-4,12	-3,31	-2,97	0,69	-2,94
7	-0,86	-0,42	-1,15	1,92	-1,88	-0,28	-1,30	-2,47	-2,61	-1,01	-2,03	1,33	1,16
8	-0,83	-1,22	-0,31	0,07	-1,09	-1,09	1,37	0,20	-0,31	-1,48	-1,61	0,72	-1,19
	H < 1	1 ≤ H < 2	H ≥ 2										

Para a distribuição adimensional (Tabela 7.9), ocorreram 15 casos com  $H > 1$ , entretanto existiram 8 situações com  $H > 2$ , enquanto para a distribuição não adimensional esta condição foi observada em apenas 4 casos. Ou seja, na análise da curva adimensional houve maior dificuldade de se atingir a condição de homogeneidade.

Tabela 7.9. Teste de homogeneidade  $H_{dp}$  para a distribuição regional adimensional.

Região	Teste de homogeneidade $H_{dp}$												
	Jan	Fev	Mar	Abr	Mai	Jun	Jul	Ago	Set	Out	Nov	Dez	Ano
1	0,71	0,28	2,87	-1,02	-0,80	4,38	1,36	-0,37	0,28	-1,24	2,65	0,92	-1,95
2	-1,25	0,25	0,85	-2,15	0,25	0,55	-1,85	0,85	-1,55	-1,55	-3,05	-0,95	-1,82
3	1,65	0,47	0,08	0,86	-2,67	-1,29	-1,68	-0,70	-0,51	-1,29	-1,10	0,28	-1,10
4	0,27	-1,52	-1,13	-0,53	-1,13	-2,52	0,67	2,67	-1,32	-1,92	-0,13	0,87	-1,58
5	0,06	0,06	-1,04	-0,77	-1,04	-0,21	2,00	-0,49	-0,77	-0,77	-1,60	-0,77	0,39
6	2,38	0,24	-0,09	0,73	-0,91	-0,58	-0,09	3,04	-1,40	-1,73	1,56	2,05	-1,14
7	-0,88	-0,44	-1,32	5,73	0,44	-1,10	0,88	-2,42	-3,08	-1,10	-0,88	0,66	-1,37
8	0,24	-1,73	-0,35	1,22	-1,53	0,04	0,83	1,02	-1,53	-2,12	1,02	0,24	-2,45

$H < 1$	$1 \leq H < 2$	$H \geq 2$
---------	----------------	------------

Note-se também que houve apenas 4 casos de coincidência para  $H > 1$  nas duas análises, e apenas 1 caso coincidindo com  $H > 2$ . Ou seja, possivelmente as causas de uma e outra condição de homogeneidade são diferenciadas.

Embora tenham existido diferenças entre as duas condições, esta discussão irá se ater apenas à análise das distribuições não adimensionais.

Como, de acordo com Hosking e Wallis (1997), as distribuições regionais obtidas em regiões parcialmente heterogêneas ainda produziriam estimativas mais robustas que as distribuições ajustadas localmente, não foram dispensados maiores esforços para a elevação do nível de homogeneidade em regiões com  $H < 2$ .

Desta forma, o caso crítico se encontra na região 1, na qual os meses de junho, julho e agosto apresentaram  $H > 2$ . Tal região já tinha merecido destaque durante a fase de ajustes preliminares, pela presença das estações 156, 175 e 176, localizadas no extremo oeste do Rio Grande do Sul, que apresentaram elevado número de séries discordantes no teste de aderência regional.

Assim, a região 1 foi subdividida em regiões 1a e 1b, essa última contendo as estações 156, 175 e 176, e cujos valores de  $dp$ ,  $\mu dp_s$  e  $\sigma dp_s$  (região com 3 elementos) foram de 0,161 e 0,029, para a distribuição Gama, e para a distribuição Normal, de 0,157 e 0,029.

Com isso, a região 1a passa a apresentar  $H > 1$  apenas no mês de fevereiro, embora na região 1b ainda se registrem 3 casos com  $H > 1$ , (meses de março e agosto, e a precipitação anual). No entanto, apenas a distribuição anual apresenta  $H > 2$ , podendo-se, nesse caso, utilizar os fatores de escala locais.

Devido à ponderação das precipitações anuais, foram observados alguns pontos deslocados do centro geográfico das regiões. Para maior harmonia sob o aspecto geográfico, alguns desses pontos foram deslocados para regiões de maior proximidade geográfica e altitudes mais

semelhantes, priorizando as regiões às quais pertencessem com o segundo maior grau de pertinência.

Tais mudanças só foram efetivadas após constatação de que a homogeneidade da região era garantida, e, desta forma, foram efetuadas nove trocas que envolveram as regiões 3, 4, 5, e 6, correspondendo em todos os casos à região de segundo maior grau de pertinência, embora, na metade dos casos, as diferenças entre o primeiro e o segundo graus de pertinência fossem bastante elevadas.

Ao final, dos 117 períodos analisados (9 regiões com 13 períodos cada), verificou-se que, em 88,8% desses, a região foi considerada aceitavelmente homogênea, em 9,4%, possivelmente heterogênea, e em 1,7%, heterogênea, sendo estas correspondentes ao mês de agosto da região 4 e à precipitação anual da região 1b.

### 7.3 Caracterização das Regiões homogêneas

As nove regiões homogêneas, quanto à distribuição de probabilidade pluviométrica, definidas para a área de estudo, estão representadas na Figura 7.5. Os respectivos centros são apresentados na Tabela 7.10.

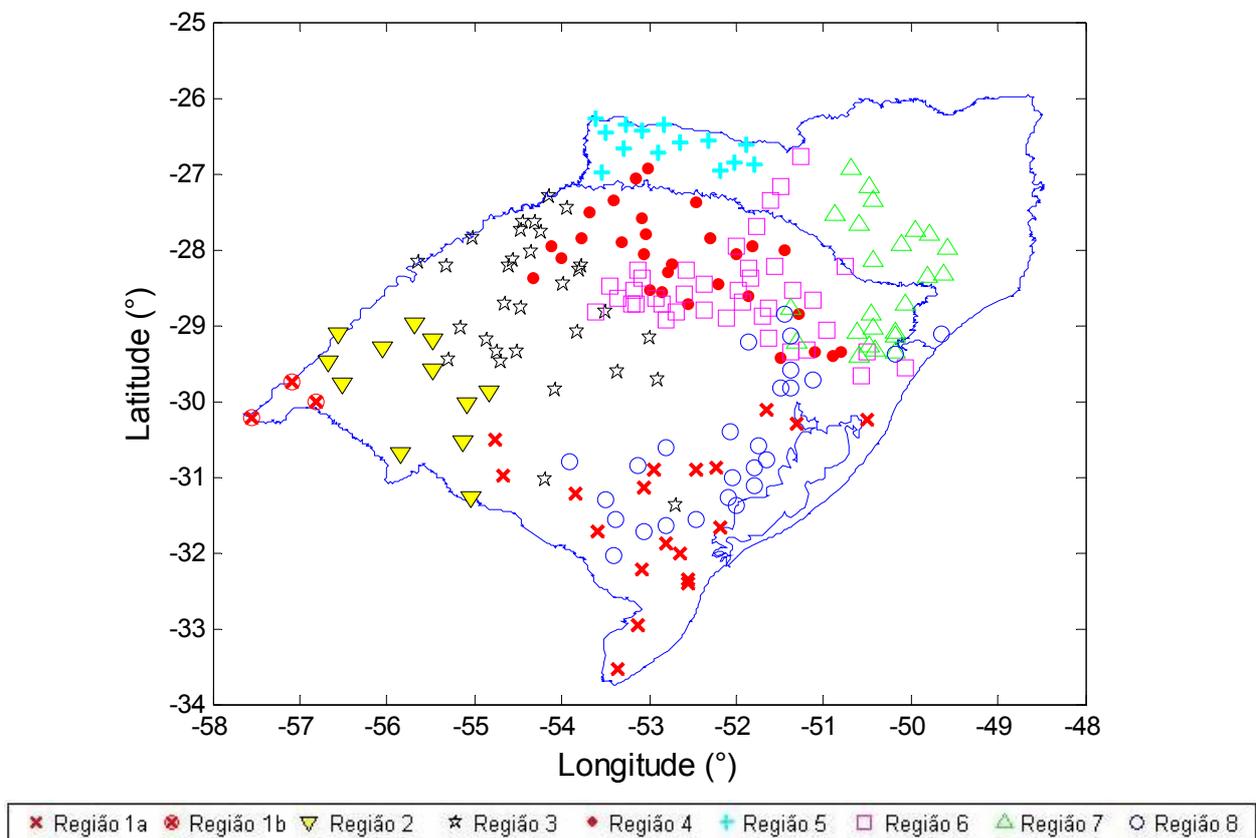


Figura 7.5 Regiões homogêneas definidas para o Rio Grande do Sul e Santa Catarina.

Tabela 7.10: Protótipos do agrupamento.

Região	n° Séries	Longitude (°)	Latitude (°)	Altitude (m)	Precipitação média (mm)												
					Jan	Fev	Mar	Abr	Mai	Jun	Jul	Ago	Set	Out	Nov	Dez	Anual
1	22	-52,9538	-31,4026	76	108	130	106	132	112	117	135	100	120	120	107	99	1384
2	12	-55,5217	-29,7089	131	130	146	144	185	134	115	118	81	133	164	146	116	1612
3	30	-54,0411	-28,6938	274	150	155	133	174	146	139	155	118	160	197	159	136	1822
4	27	-52,7297	-28,1362	515	164	162	127	153	154	145	166	132	171	213	159	155	1901
5	13	-52,6923	-26,8575	787	181	196	140	163	180	162	160	129	176	236	183	173	2079
6	43	-52,2244	-28,4944	543	151	149	114	139	134	136	162	123	165	188	150	142	1753
7	25	-50,3841	-28,4535	880	155	156	112	113	109	119	148	118	141	161	130	141	1602
8	26	-52,1037	-30,5744	153	123	138	112	136	118	134	153	113	139	139	122	121	1548

Observa-se que as regiões apresentam coerência física, embora não necessariamente com continuidade espacial. A média da precipitação anual para toda a área de estudo foi de 1673 mm, sendo 1645 mm para as estações do RS e 1800 mm para as estações amostradas em Santa Catarina.

Do ponto de vista intra-anual, praticamente não se observa variação, e em ambos os estados o trimestre mais chuvoso é o de setembro a novembro e o menos chuvoso o de março a maio, correspondendo respectivamente a 26% e 23% da precipitação anual no RS e 28% e 23% em SC. Destaca-se o mês de outubro com maior pluviometria, equivalendo a 10% da precipitação anual no RS e 11% em SC.

Porém, avaliando-se cada região separadamente, observa-se, pela Tabela 7.11, que nas regiões 2 e 1b existe um regime intra-anual diferenciado do regime geral, destacadamente na região 1b, onde se verificam evidências de sazonalidade, sendo mais chuvoso no outono (março a maio), com 31% do total anual, e menos chuvoso no inverno (junho a agosto), com 16% do total anual, observando-se nesse último mês um média menor que 60mm.

Tabela 7.11: Percentual pluviométrico por trimestre.

Região	Precipitação trimestral (% da anual)			
	Primavera	Verão	Outono	Inverno
1a	25,0	24,0	25,1	25,9
1b	26,1	27,1	30,8	16,0
2	27,4	24,3	29,1	19,2
3	28,3	24,1	25,2	22,4
4	28,6	25,1	22,6	23,7
5	28,6	26,7	23,3	21,4
6	28,6	25,3	21,9	24,2
7	27,1	28,1	20,8	24,1
8	25,9	24,7	23,6	25,9

Na região 2 o comportamento pluviométrico intra-anual acompanha com menor intensidade a sazonalidade identificada na região 1b, com baixas precipitações no inverno (19% do total anual), e altas no outono (29% do total anual), embora os volumes pluviométricos sejam mais elevados que na região 1b.

Observa-se ainda que nas regiões 1b e 2, o mês de abril é o que apresenta maior precipitação, com valor elevado também na região 3, onde o volume de chuva é menor apenas que o registrado no mês de outubro.

Nota-se que a partir da região 3 a precipitação se distribui mais igualmente durante o ano, destacando-se a primavera com a maior precipitação, e a menor precipitação variando entre o outono e o inverno. De fato as regiões 3, 4, 5 e 6 caracterizam-se por apresentar maior precipitação relativa no mês de outubro, o que contribui para a maior parcela de chuva durante a primavera.

A Figura 7.6 apresenta cada uma das regiões identificadas separadamente, cujas principais características são:

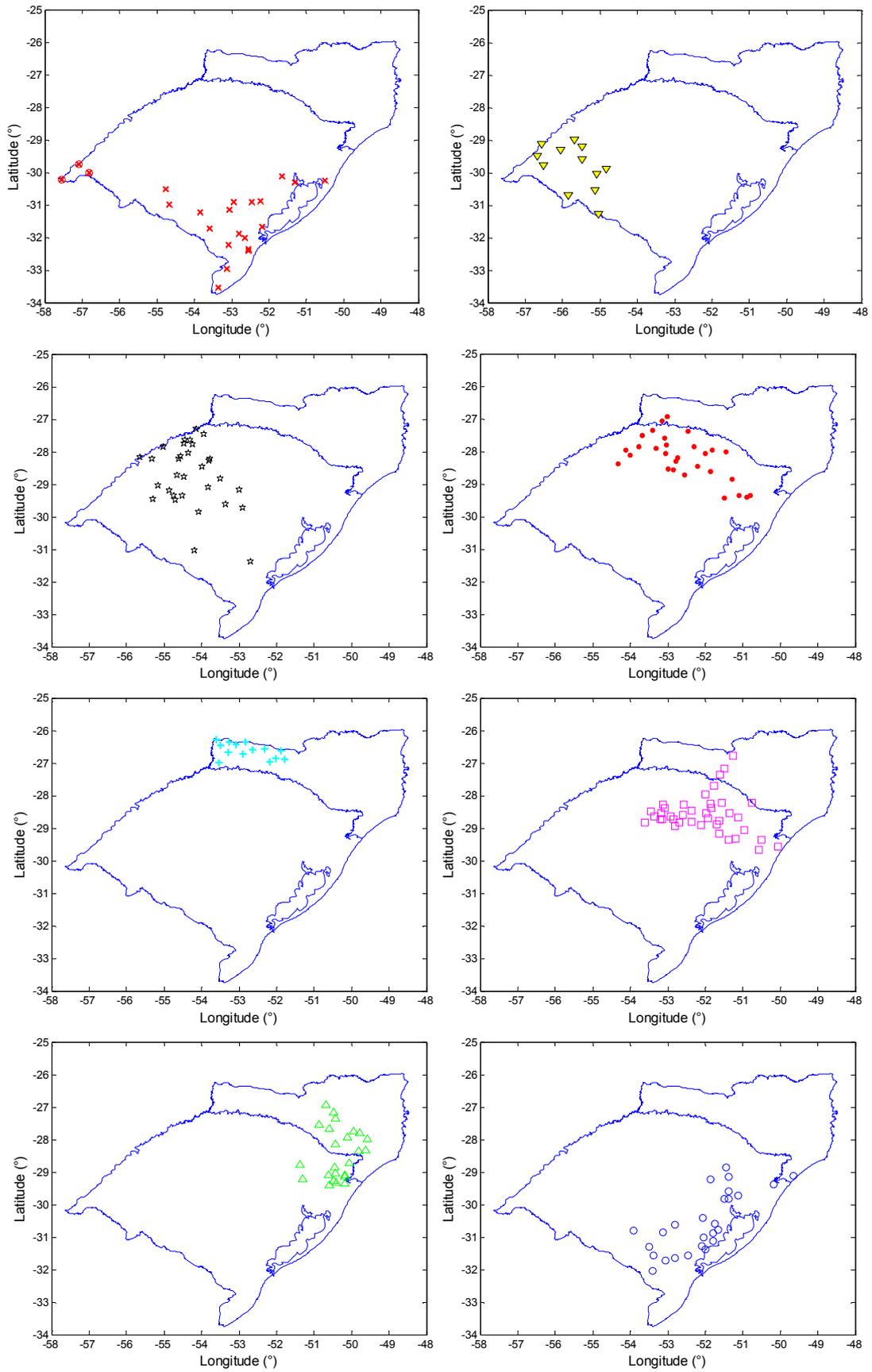
**Região 1:** Dividida em região 1a (19 estações) e 1b (3 estações), engloba a parte sul do RS se estendendo ao extremo oeste. É caracterizada pelas mais baixas precipitações da área de estudo. Os postos possuem altitude média de 76m, sendo que 18% deles estão acima de 100m de altitude. As estações do extremo oeste, reunidas na região 1b, apresentam evidências de sazonalidade, sendo mais chuvoso no outono e menos no inverno. A região 1a, que reúne estações abaixo do paralelo 30°, possui regime de precipitações bem distribuídas durante o ano.

**Região 2:** É composta por apenas 12 estações, reunidas no oeste do Rio Grande do Sul, sendo ainda predominantes as baixas altitudes, com média de 131m. Possui diferenças sazonais menos marcantes que a região 1b, mas ainda distingue-se do comportamento mais bem distribuído das demais regiões. O mês de abril apresenta as maiores precipitações, com média de 185mm.

**Região 3:** Com 31 estações, a região 3 cobre o centro e noroestes do Rio Grande do Sul, com duas estações isoladas ao sul (180 e 192), as quais, devido à elevada precipitação, não se adequaram às regiões mais próximas. A altitude da região é de aproximadamente 270 m, e a precipitação total anual fica em torno dos 1800 mm.

**Região 4:** Apresenta-se espacialmente mais dispersa, concentrando-se na parte norte do Rio Grande do Sul, mas englobando as estações de elevada precipitação, a leste, nos bordos do planalto. A altitude também é variável, indo desde 360m a 950m, ficando a precipitação anual em torno dos 1900mm.

**Região 5:** Totalmente localizada em Santa Catarina, compreende as estações do oeste desse Estado, caracterizadas por apresentar as maiores precipitações da área de estudo, com média superior a 2000mm anuais. Embora possua estações com mais de 1000m de altitude, não é a região de maior altitude média, ficando atrás da região 7.



× Região 1a   × Região 1b   ▾ Região 2   ☆ Região 3   ● Região 4   + Região 5   □ Região 6   ▲ Região 7   ○ Região 8

Figura 7.6 Regiões de 1 a 8

**Região 6:** Como a maior das regiões, compreende 39 estações que estão situadas basicamente no planalto meridional, com uma média de aproximadamente 540m de altitude, embora sejam observados alguns pontos isolados, situados ao sopé das encostas ou em vales, e que sofrem influência do efeito orográfico. O total pluviométrico anual fica em torno de 1750 mm, com pouca variação.

**Região 7:** Localizada no planalto meridional, caracteriza-se por apresentar elevadas altitudes (média próxima a 900m). A precipitação total anual, entretanto, é a mais baixa das regiões da porção norte da área de estudo, com médias anuais de apenas 1600mm, que é semelhante à encontrada na região 2, onde a altitude média é de 130m. Esta baixa precipitação pode estar relacionada a uma possível posição desfavorável com relação à entrada das massas de ar úmido na área de estudo.

**Região 8:** Agrupa pontos situados a leste da área de estudo, apresentando grande dispersão em termos de altitude, que varia de 2m a 500m. A divergência das altitudes, entretanto, não se reflete nos totais pluviométricos, apresentando média anual em torno de 1550 mm, sem grande dispersão. São identificadas, porém, algumas diferenças de volume entre o vale do rio Jacuí e a planície da Lagoa dos Patos, embora as diferenças identificadas não tenham sido suficientes para influenciar a homogeneidade da região.

A figura 7.7 apresenta um mapa de isoietas das precipitações médias anuais no estado do Rio Grande do Sul elaborado por Sotério et al. (2005).

Observe-se no mapa que as faixas de precipitação apresentam grande sinuosidade, o que explicaria a sobreposição visual entre as regiões identificadas pelo agrupamento (figura 7.5). Além disso, as baixas precipitações médias anuais apresentadas na região 5 (que possui elevada altitude) também são identificadas na área do mapa de isoietas a ela correspondente.

Assim, verifica-se que os resultados obtidos apresentam coerência com outras investigações efetuadas na área de estudo.

Nas tabelas 7.12 a 7.20 são apresentadas as estações pertencentes a cada região, suas precipitações médias em cada período e a identificação das séries que não passaram no teste KS, a 20% de significância, com relação à distribuição regional.

Cabe salientar que apenas 2,2% das séries não passaram no teste KS a 20% de significância, que é um nível de significância bastante restritivo. Já com relação à distribuição regional adimensional, todas as estações passaram no teste KS para esse mesmo nível.

Verifica-se, ainda, que as séries que não passaram no teste de KS geralmente ocorrem isoladamente, e na maioria dos casos não coincidem com os períodos que apresentaram altos valores de  $H_{dp}$  no teste de homogeneidade.

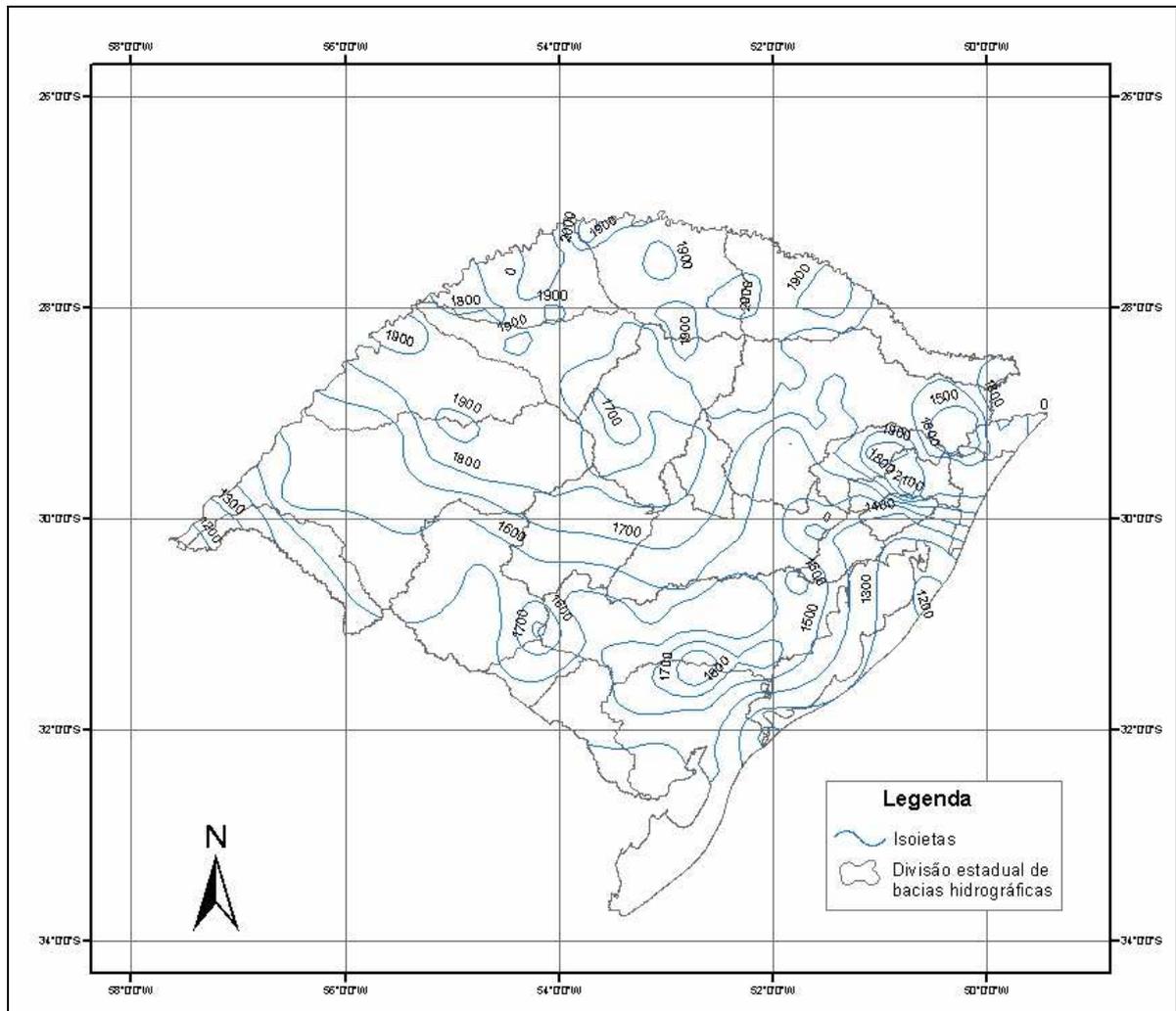


Figura 7.7: Isoietas das precipitações médias anuais no Rio Grande do Sul (Sotério et al. 2005).

Tabela 7.12 Estações pertencentes à Região 1a.

ID	Alt (m)	Precipitação Média (mm)												Anual
		J	F	M	A	M	J	J	A	S	O	N	D	
157	3	90	96 *	87	103	96	130	138	104	121	114	98	83	1259 *
159	40	99	100	86	111	92	134	146	111	128	127	114	105	1354
162	5	96	106	93	120	102	145	148	122	132	128	111	106	1411
163	60	114	103	101	132	109	116	149	98	127	114	100	101	1363
164	50	102	99	95	119	109	114	137	96	127	114	103	101	1313
165	80	123	128	113	145	117	131	149	91	130	123	99	111	1462
170	120	116	131	114	165	130	118	119	85	105	129	115	98	1426
171	100	116	143	113	165	127	111	126	81	126	143	116	98	1466
182	3	99	139	104	127	104	109	138	100	125	112	106	101	1363
184	27	107	143	114	135	116	115	146	105	121	118	112	107	1440
186	368	116	113	104	138	124	120	128	86	111	138	119	102	1397
188	120	121	135	101	137	127	128	139	97	128	131	105	111	1461
189	380	111	131	112	156	116	120	135	94	124	124	115	103	1441
194	20	112	157	112	135	111	112	145	107	127	114	111	104	1445
195	4	103	149	113	128	124	114	145	119	120	112	92	93	1413
196	12	112	138	106	132	113	117	147	115	127	117	91	83	1399
197	3	91	141	101	130	113	113	136	106	115	111	109	84	1350
198	3	98	135	99	117	106	121	131	106	120	110	95	86	1323
200	24	95	153	118	110	110	116	138	106	103	95	95	87	1326
média	75	106	129	105	132	113	120	139	102	122	120	106	98	1390

\*Significativa no teste KS .

Tabela 7.13 Estações pertencentes à Região 1b.

ID	Alt (m)	Precipitação Média (mm)												Anual
		J	F	M	A	M	J	J	A	S	O	N	D	
156	74	121	139	146	173	108	85	72	61	87	129	118	90	1329
175	80	120	132	131 *	169	114	83	77	64	96	136	121	99	1341
176	40	106	113	99	133	81	64	57	37	71	115	106	97	1077 *
média	65	115	128	125	158	101	77	69	54	85	127	115	95	1249

\*Significativa no teste KS .

Tabela 7.14 Estações pertencentes à Região 2.

ID	Alt (m)	Precipitação Média (mm)												Anual
		J	F	M	A	M	J	J	A	S	O	N	D	
109	140	124	135	146	195	139	126	115	90	142	168	166	119	1665
142	100	123	131	129	177	137	119	141	88	153	169	142	112	1621
148	100	132	147	153	199	140	120	118	82	138	179	163	118	1688
151	80	134	143	152	181	140	115	126	84	135	175	155	111	1651
152	35	130	153	156	190	131	111	97	69	124	172	159	117	1608
153	60	137	149	165	200	135	112	115	71	141	170	152	131	1681
154	120	142	166	183 *	203	129	102	102	72	119	181	143	117	1659
155	90	125	148	143	192	125	110	96	58	108	145	140	105	1495
172	300	129	164	126	171	139	103	97	70	109	147	128	120	1504
173	90	124	120	132	178	135	120	130	83	152	160	143	112	1590
174	200	125	159	133	193	133	123	131	92	140	151	135	110	1626
193	240	123	127	117	175	138	109	107	87	114	134	126	116	1474
média	130	129	145	145	188	135	114	115	79	131	163	146	116	1605

\*Significativa no teste KS .

Tabela 7.15 Estações pertencentes à Região 3.

ID	Alt (m)	Precipitação Média (mm)												Anual
		J	F	M	A	M	J	J	A	S	O	N	D	
36	450	152	156	133	158	151	137	140	108	151	215	136	155	1793
41	120	152	171	133	148	141	124	119	112	155	218	131	146	1751
42	325	132	155	118	160	145	135	130	126	153	208	145	153	1760
43	403	135	149	112	147	142	125	128	103	148	211	142	139	1679
44	330	135	154	132	171	161	142	131	113	169	219	152	162	1840
45	120	156	153	135	161	153	140	128	127	164	228	143	156	1844
47	100	146	155	117	176	139	110 *	112 *	105	144	212	147	152	1716
87	250	152	149	116	147	151	138	144	121	167	201	154	149	1789
89	160	157	147	120	165	154	136	152	124	161	206	155	144	1820
95	350	142	132	115	163	137	132	167	116	172	182	171	126	1755
96	270	156	143	110	152	156	134	151	125	171	202	147	139	1786
102	400	152	160	119	177	149	132	145	125	161	225	163	150	1857
103	200	139	143	116	190	137	132	132	120	149	208	171	141	1778
104	160	140	145	140	183	136	133	152	112	145	187	157	129	1759
105	160	150	154	127	189	151	148	141	132	156	218	170	140	1875
106	300	152	147	138	194	139	149	164	125	159	194	163	132	1854
107	60	161	169	162	208	139	127	133	111	150	212	191	152	1916
108	80	157	167	150	218	141	131	138	122	145	222	174	152	1919
138	80	152	152	129	164	134	161	166	130	166	168	147	136	1805
139	440	151	159	127	175	138	161	193 *	128	161	185	152	130	1860
140	25	157	147	139	170	147	148	180	123	169	173	151	129	1835
141	469	142	152	128	165	150	157	180	129	174	191	151	122	1841
143	300	136	162	159 *	185	153	133	160	99	157	191	175	114	1823
144	300	141	152	160	180	151	145	172	108	166	191	164	113	1843
145	100	161	159	155	199	154	130	169	102	168	186	169	121	1872
146	408	153	154	135	167	149	143	177	106	164	170	145	116	1778
147	420	157	170	163	193	162	144	170	109	167	191	180	120	1925
149	100	140	161	142	208	149	122	146	82	156	188	168	106	1767
150	380	146	172	147	199	167	136	159	109	161	188	176	126	1886
180	400	161	158	132	168	150	161	198	133	178	162	141	133	1876
192	390	145	151	138	186	154	149	159	107	140	160	134	117	1738
média	260	149	155	134	176	148	139	153	116	160	197	157	135	1818

\*Significativa no teste KS .

Tabela 7.16 Estações pertencentes à Região 4.

ID	Alt (m)	Precipitação Média (mm)												Anual
		J	F	M	A	M	J	J	A	S	O	N	D	
15	280	153	188	120	140	134	143	135	108	165	203	146	141	1776 *
26	687	168	150	121	135	157	140	185	128	174	213	163	147	1883
28	950	183	160	112	135	170	139	191	138	181	224	165	153	1951
29	763	179	182 *	125	156	187	148	179	159	205 *	236	164	173	2094
31	350	147	160	113	132	157	140	173	119	169	211	152	141	1813
32	530	151	159	116	151	154	146	148	127	153	211	152	151	1818
33	350	161	154	121	146	160	142	162	125	161	212	141	151	1836
34	400	148	201	127	152	157	145	154	119	158	215	153	161	1888
35	520	160	144	134	169	166	145	159	130	161	233	156	158	1916
38	378	153	171	133	164	171	147	170	131	170	226	160	163	1958
39	610	146	132	117	138	157	133	144	125	170	225	136	157	1779
40	502	158	165	130	158	169	135	153	121	162	231	149	164	1895
46	400	161	167	114	176	160	147	147	133	168	227	153	155	1906
58	630	173	169	115	140	128	142	184	141	168	178	151	166	1855
66	380	172	159	127	145	136	137	182	146	184	213	162	163	1927
72	570	176	148	140	158	161	151	179	141	188	224	175	161	2001
73	593	154	147	135	148	160	137	170	136	182	221	164	146	1900
74	460	161	151	128	152	152	143	177	135	176	206	155	149	1883
77	490	162	153	119	149	147	138	184	135	185	202	168	139	1882
78	580	168	159	104	137	149	133	174	137	183	203	162	149	1857
85	550	165	154	133	177	158	159	196 *	148	188	207	172	145	2004
86	672	161	159	125	137	155	132	174	133	175	212	147	146	1857
99	440	168	149	120	162	164	146	163	137	193	226	166	166	1961
100	450	164	150	132	142	178	142	159	129	176	223	163	162	1919
101	447	169	162	125	172	164	147	147	135	167	226	167	147	1928
113	850	168	169	133	155	126	165	182 *	153	180	184	159	172	1946
118	660	172	175	134	153	131	169	189	149	169	191	165	163	1959
132	550	160	162	133	152	135	175	194	155	168	191	159	165	1948
134	486	176	165	127	166	150	193 *	208 *	160 *	197	199	182	178	2102
média	535	163	161	124	152	155	147	171	136	175	213	159	156	1912

\*Significativa no teste KS .

Tabela 7.17 Estações pertencentes à Região 5.

ID	Alt (m)	Precipitação Média (mm)													Anual
		J	F	M	A	M	J	J	A	S	O	N	D		
2	1000	182	193	127	160	153	146	163	132	166	225	173	183	2003	
3	1250	191	178	133	156	181	179	165	128	171	232	189	167	2069	
5	1000	206	200	145	176	182	165	178	134	181	242	196	184	2187	
6	760	165	177	139	155	178	160	157	136	168	229	169	169	2004	
7	600	174	182	133	147	149	150	151	134	163	220	162	160	1926 *	
8	550	171	182	127	149	169	147	138	115	159	238	154	167	1917 *	
9	360	183	190	130	150	161	144	135	120	152	228	173	160	1926	
10	720	187	198	137	163	202	166	153	129	181	246	181	178	2120	
11	920	174	186	137	155	177	156	149	114	167	212	177	165	1971	
12	800	180	219	153	172	193	165	149	127	180	250	210	174	2171	
13	320	171	199	143	160	170	162	142	127	167	250	173	185	2049	
14	720	179	213	151	172	187	167	153	113	174	254	189	174	2126	
16	850	188	217	148	164	194	167	156	116	178	250	210	167	2156	
37	557	156	206	137	168	171	153	150	127	168	233	159	168	1997	
média	743	179	196	139	161	176	159	153	125	170	236	180	172	2044	

\*Significativa no teste KS .

Tabela 7.18 Estações pertencentes à Região 6.

ID	Alt (m)	Precipitação Média (mm)													Anual
		J	F	M	A	M	J	J	A	S	O	N	D		
4	1133	174 *	181 *	141	116	143	126	131	106	140	197	162	152	1769	
24	560	163	187 *	113	125	140	125	147	111	149	210	143	164	1778	
25	600	162	159	100	140	150	133	170	127	153	203	151	145	1795	
27	498	156	167	108	132	145	130	153	109	156	198	155	147	1757	
30	640	147	163	106	122	143	124	157	116	160	197	145	143	1722	
56	640	162	162	109	125	136	115	155	126	167	180	141	140	1714	
59	600	171	148	115	144	146	140	162	137	181	196	143	149	1832	
60	900	147	167	107	129	134	130	160	117	162	186	143	144	1726	
61	720	171	151	126	124	130	134	166	128	160	179	130	155	1754	
62	650	169	131	104	138	124	141	179	129	163	172	150	140	1739	
63	800	171	181	100	132	123	137	167	129	177	171	142	146	1775	
64	750	154	148	111	138	153	129	160	120	179	214	145	151	1801	
67	680	147	151	113	134	129	135	180	131	171	183	143	143	1761	
69	620	152	152	115	138	123	133	176	127	172	183	145	150	1767	
70	650	130	144	106	127	136	128	157	124	168	186	134	146	1687	
71	700	143	136	110	162	127	142	189	125	171	190	162	127	1782	
75	600	135	137	124	154	108	151	168	113	162	171	163	130	1716	
76	650	168	147	115	146	127	148	188 *	138	172	189	167	144	1848	
79	430	152	153	121	136	132	137	173	128	174	178	168	136	1789	
80	600	145	133	113	136	145	127	155	116	171	194	151	144	1730	
81	360	142	136	116	146	131	137	166	114	176	202	169	137	1771	
82	510	146	135	108	134	140	132	158	128	180	190	160	141	1752	
83	535	143	146	112	175 *	137	148	168	126	177	188	161	138	1819	
84	380	140	133	109	154	130	133	163	118	168	191	157	137	1733	
88	504	152	126	124	143	148	143	158	121	176	189	165	144	1787	
90	300	126	135	106	164	132	131	158	119	168	191	168	129	1725	
91	250	155	131	108	148	127	133	155	109	164	189	144	119	1680	
92	350	148	133	112	137	130	126	162	113	164	186	157	143	1712	
93	550	151	141	118	137	150	141	165	118	174	187	159	145	1786	
94	390	147	123	100	147	135	143	148	120	166	191	153	128	1700	
97	490	151	134	112	139	144	128	151	124	170	202	151	134	1742	
98	405	128	134	109	160	129	130	143	108	167	178	153	99 *	1638	
117	840	155	178	131	131	124	150	167	147 *	155	178	142	158	1816	
120	50	153	149	116	128	106	168	172	145 *	132	159	130	138	1696	
121	880	159	149	109	139	119	145	177	136	153	161	132	139	1715	
123	40	177	203 *	160 *	130	104	127	134	132	154	157	143	156	1777	
130	550	140	136	109	131	109	150	165	133	160	172	154	136	1695	
131	80	155	158	115	134	116	168	167	138	145	181	141	136	1755	
135	140	142	144	108	134	125	157	170	135	155	177	151	149	1749	
média	539	152	149	114	139	132	137	163	124	164	186	151	141	1751	

\*Significativa no teste KS .

Tabela 7.19 Estações pertencentes à Região 7.

ID	Alt (m)	Precipitação Média (mm)												Anual
		J	F	M	A	M	J	J	A	S	O	N	D	
1	1000	159	168	124	99	130	122	151	109	144	204 *	144	149	1703
17	900	148	147	93	100	95	99	124	105	135	157	119	128	1450 *
18	900	172	161	120	109	113	111	154	112	154	173	138	150	1666
19	1196	159	167	111	116	116	116	162	122	144	168	139	133	1651
20	720	159	167	105	122	127	115	151	110	141	195	139	133	1664
21	980	157	158	111	109	130	116	149	106	140	179	134	153	1642
22	940	153	167	111	110	117	112	142	110	138	175	130	161 *	1625
23	920	159	168	108	123	129	117	151	110	141	186	139	146	1677
49	1200	153	154	117	90	100	94	132	102	127	145	115	142	1471
52	997	167	162	108	98	107	102	144	114	143	142	127	147	1561
53	1080	182	172 *	124	104	116	108	159	124	146	166	147	157	1703
54	1150	169	162	126	121	114	128	161	144	156	167	135	150	1734
55	930	142	147	99	113	107	103	149	107	143	144	126	116 *	1495 *
57	640	161	160	120	107	104	124	154	126	143	163	134	145	1643
68	700	151	122	85 *	114	123	131	170	105	156	177	127	159	1620
111	970	159	157	112	113	97	126	140	129	137	152	123	131	1576
112	970	149	155	103	107	95	118	133	123	128	144	120	131	1506
114	920	168	153	119	126	115	147	163	133	142	158	134	164	1712
115	910	129	141	116	119	100	130	141	125	129	152	118	125	1525
116	900	151	152	108	126	107	139	161	125	152	158	135	139	1655
122	470	165	157	123	114	92	115	134	111	129	144	123	157	1563
124	950	140	152	118	113	90	127	146	124	137	150	115	133	1546
125	920	137	137	102	110	100	125	141	124	137	153	126	122	1512
128	750	124	138	110	133	114	148 *	159	132	155	164	138	135	1652
média	917	155	155	111	112	110	120	149	118	141	163	130	142	1606

\*Significativa no teste KS .

Tabela 7.20 Estações pertencentes à Região 8.

ID	Alt (m)	Precipitação Média (mm)												Anual
		J	F	M	A	M	J	J	A	S	O	N	D	
65	400	136	138	106	123	102	115	165	116	151	153	124	144	1574
110	16	154	159	138	113	109	109	119	126	129	139	128	134	1557
119	40	154 *	157	113	106	91	113	120	115	122	140	124	145	1500
126	50	135	130	105	130	107	152	153	126	136	151	129	127	1580
127	60	123	114	83	113	103	139	145	121	130	148	132	105	1457
129	500	124	132	91	120	106	137	166	123	146	155	129	133	1564
133	15	117	106	88	126	96	139	163	125	144	144	133	115	1496
136	20	122	112	94	126	99	148	164	127 *	134	148	123	119	1516
137	50	116	124	98	125	95	150	157	124	130	148	123	119	1503
158	120	131	147	126	124	128	151	165	121	148	146	125	141	1653
160	65	124	143	104	126	118	130	153	114	143	129	117	128	1531
161	25	111	135	113	123	114	137	154	119	154	127	128	129	1544
166	300	120	122	111	127	123	152	170	118	144	140	119	127	1573
167	400	129	130	118	164	131	145 *	170	115	158	147	131	111	1651
168	300	121	134	131	165	129	127	139	92	122	142	116	107	1525
169	340	128	132	119	159	122	128	146	97	134	141	110	113	1529
177	5	106	139	119	136	122	140	159	118	142	140	106	129	1556
178	2	99	157	117	147	132	129	167	111	153	131	118	114	1575
179	120	119	178	124	148	136	134	173	113	162	137	129	133	1686
181	250	124	150	119	150	128	136	168	114	143	127	131	108	1598
183	40	119	141	108	132	111	131	155	106	139	134	113	120	1509
185	40	110	152	123	145	131	123	142	104	135	126	115	118	1524
187	200	127	153	116	155	132	124	155	109	140	133	139	114	1598
190	440	123	139	124	177	135	129	151	104	138	133	118	106	1575
191	420	115	142	121	151	139	132	149	103	136	131	117	114	1550
199	260	124	140	114	149	131	127	148	106	118	132	124	100 *	1514
média	172	124	139	112	137	118	134	154	114	140	139	123	121	1555

\*Significativa no teste KS .

## 7.4 Distribuições regionais

Na Tabela 7.21 a seguir são fornecidos, para cada uma das regiões, os parâmetros ( $\theta_1$  e  $\theta_2$ ) da distribuição regional em cada período (12 mensais e 1 anual). São apresentados também os quantis para os tempos de retorno mais usuais, sendo estes adimensionalizados pela média da distribuição regional.

Esta alternativa dá a opção de redimensionalizá-los pelo fator de escala regional (média da distribuição regional), que é o recomendado nesta pesquisa, ou pelo fator de escala local, principalmente nos casos em que a série local não passa no teste KS e que foram identificados nas tabelas 7.12 a 7.20.

Nota-se que os quantis adimensionais da região 1b apresentam comportamento diferenciado das demais regiões, sendo esta região mais suscetível a variações extremas, principalmente nos meses de janeiro, e de julho a setembro, em que a precipitação, num tempo de retorno de 20 anos, excede a média em 2,7 vezes nos anos úmidos ou não ultrapassa 10% da média em períodos secos.

Nas outras regiões, a maior variabilidade acontece destacadamente no mês de agosto, no qual, para Tr de 20 anos, a média é excedida aproximadamente em 2,4 vezes, ou a chuva não ultrapassa 15% a 20% da média.

Tabela 7.21 Parâmetros e quantis Regionais.

Região 1a																
Período	Parâmetros		Média	Quantis Adimensionais não-excedentes							Quantis Adimensionais excedentes					
	01	02		Tr (anos)							Tr (anos)					
				2	5	10	20	50	75	100	5	10	20	50	75	100
Jan	1,800	58,737	106	0,82	0,38	0,24	0,15	0,09	0,07	0,06	1,52	1,99	2,45	3,04	3,30	3,48
Fev	1,852	71,377	132	0,83	0,39	0,25	0,16	0,09	0,07	0,06	1,51	1,98	2,43	3,01	3,26	3,43
Mar	1,803	59,605	107	0,82	0,38	0,24	0,15	0,09	0,07	0,06	1,52	1,99	2,45	3,04	3,30	3,48
Abr	2,304	57,155	132	0,86	0,45	0,30	0,21	0,13	0,11	0,10	1,47	1,88	2,27	2,76	2,97	3,12
Mai	1,714	65,628	112	0,81	0,37	0,23	0,14	0,08	0,06	0,05	1,52	2,02	2,49	3,10	3,37	3,56
Jun	3,000	39,387	118	0,89	0,51	0,37	0,27	0,19	0,16	0,15	1,43	1,77	2,10	2,51	2,68	2,80
Jul	2,300	60,320	139	0,86	0,45	0,30	0,21	0,13	0,11	0,10	1,47	1,88	2,27	2,76	2,98	3,12
Ago	2,092	46,966	98	0,85	0,42	0,28	0,19	0,12	0,09	0,08	1,49	1,92	2,34	2,87	3,09	3,25
Set	3,455	35,067	121	0,91	0,54	0,40	0,31	0,22	0,19	0,17	1,40	1,72	2,02	2,39	2,54	2,65
Out	3,156	37,868	120	0,90	0,52	0,38	0,28	0,20	0,17	0,16	1,42	1,75	2,07	2,46	2,63	2,75
Nov	2,924	37,086	108	0,89	0,51	0,36	0,27	0,18	0,16	0,14	1,43	1,78	2,11	2,53	2,71	2,83
Dez	1,978	49,451	98	0,84	0,41	0,26	0,18	0,11	0,08	0,07	1,50	1,95	2,38	2,93	3,17	3,34
Ano	1390,21	311,39	1390	1,00	0,81	0,71	0,63	0,54	0,50	0,48	1,19	1,29	1,37	1,46	1,50	1,52

Região 1b																
Período	Parâmetros		Média	Quantis Adimensionais não-excedentes							Quantis Adimensionais excedentes					
	01	02		Tr (anos)							Tr (anos)					
				2	5	10	20	50	75	100	5	10	20	50	75	100
Jan	1,294	95,086	123	0,76	0,29	0,16	0,09	0,04	0,03	0,03	1,57	2,16	2,74	3,49	3,82	4,06
Fev	2,363	52,946	125	0,86	0,45	0,31	0,22	0,14	0,12	0,10	1,47	1,87	2,25	2,73	2,94	3,09
Mar	2,127	57,105	121	0,85	0,43	0,28	0,19	0,12	0,10	0,08	1,49	1,92	2,33	2,85	3,07	3,23
Abr	1,617	95,437	154	0,80	0,36	0,21	0,13	0,07	0,06	0,05	1,53	2,05	2,54	3,18	3,46	3,65
Mai	1,549	66,360	103	0,79	0,34	0,20	0,12	0,07	0,05	0,04	1,54	2,07	2,58	3,24	3,52	3,73
Jun	1,748	41,560	73	0,82	0,38	0,23	0,15	0,08	0,07	0,06	1,52	2,01	2,48	3,08	3,34	3,52
Jul	1,424	45,448	65	0,78	0,32	0,18	0,11	0,05	0,04	0,03	1,56	2,11	2,65	3,35	3,66	3,88
Ago	1,294	39,087	51	0,76	0,29	0,16	0,09	0,04	0,03	0,03	1,57	2,16	2,74	3,49	3,82	4,06
Set	1,345	64,504	87	0,77	0,30	0,17	0,10	0,05	0,04	0,03	1,57	2,14	2,70	3,43	3,76	3,98
Out	2,329	54,387	127	0,86	0,45	0,30	0,21	0,14	0,11	0,10	1,47	1,88	2,26	2,75	2,96	3,11
Nov	1,836	63,651	117	0,83	0,39	0,24	0,16	0,09	0,07	0,06	1,51	1,98	2,44	3,02	3,27	3,45
Dez	2,046	48,823	100	0,84	0,42	0,27	0,18	0,11	0,09	0,08	1,49	1,93	2,36	2,89	3,12	3,29
Ano	1248,80	286,89	1249	1,00	0,81	0,71	0,62	0,53	0,49	0,47	1,19	1,29	1,38	1,47	1,51	1,53

Região 2																
Período	Parâmetros		Média	Quantis Adimensionais não-excedentes							Quantis Adimensionais excedentes					
	01	02		Tr (anos)							Tr (anos)					
				2	5	10	20	50	75	100	5	10	20	50	75	100
Jan	1,654	77,301	128	0,81	0,36	0,22	0,14	0,08	0,06	0,05	1,53	2,04	2,52	3,15	3,42	3,61
Fev	2,055	70,197	144	0,84	0,42	0,27	0,18	0,11	0,09	0,08	1,49	1,93	2,35	2,89	3,12	3,28
Mar	2,449	58,828	144	0,87	0,46	0,32	0,22	0,15	0,12	0,11	1,46	1,86	2,23	2,70	2,90	3,04
Abr	2,297	80,838	186	0,86	0,45	0,30	0,21	0,13	0,11	0,10	1,47	1,88	2,27	2,76	2,98	3,13
Mai	1,670	80,463	134	0,81	0,36	0,22	0,14	0,08	0,06	0,05	1,53	2,03	2,51	3,14	3,41	3,60
Jun	2,610	42,859	112	0,88	0,48	0,33	0,24	0,16	0,13	0,12	1,45	1,83	2,19	2,64	2,83	2,96
Jul	2,253	49,400	111	0,86	0,44	0,30	0,20	0,13	0,11	0,09	1,48	1,89	2,29	2,79	3,00	3,15
Ago	1,654	47,466	79	0,81	0,36	0,22	0,14	0,08	0,06	0,05	1,53	2,03	2,52	3,15	3,42	3,61
Set	3,083	41,332	127	0,89	0,52	0,37	0,28	0,20	0,17	0,15	1,42	1,76	2,08	2,48	2,65	2,77
Out	3,169	51,486	163	0,90	0,52	0,38	0,29	0,20	0,17	0,16	1,42	1,75	2,07	2,46	2,62	2,74
Nov	2,063	69,706	144	0,84	0,42	0,27	0,18	0,11	0,09	0,08	1,49	1,93	2,35	2,88	3,11	3,27
Dez	1,946	59,261	115	0,83	0,41	0,26	0,17	0,10	0,08	0,07	1,50	1,96	2,39	2,95	3,19	3,36
Ano	1605,20	396,47	1605	1,00	0,79	0,68	0,59	0,49	0,45	0,43	1,21	1,32	1,41	1,51	1,55	1,57

Região 3																
Período	Parâmetros		Média	Quantis Adimensionais não-excedentes							Quantis Adimensionais excedentes					
	θ1	θ2		Tr (anos)							Tr (anos)					
				2	5	10	20	50	75	100	5	10	20	50	75	100
Jan	2,011	73,601	148	0,84	0,41	0,27	0,18	0,11	0,09	0,08	1,50	1,94	2,37	2,91	3,15	3,31
Fev	2,115	72,079	152	0,85	0,43	0,28	0,19	0,12	0,10	0,08	1,49	1,92	2,33	2,85	3,08	3,24
Mar	2,359	56,086	132	0,86	0,45	0,31	0,22	0,14	0,11	0,10	1,47	1,87	2,25	2,74	2,95	3,09
Abr	2,256	77,634	175	0,86	0,44	0,30	0,21	0,13	0,11	0,09	1,48	1,89	2,28	2,78	3,00	3,15
Mai	1,892	78,314	148	0,83	0,40	0,25	0,17	0,10	0,08	0,07	1,51	1,97	2,41	2,98	3,23	3,40
Jun	4,397	32,002	141	0,93	0,59	0,46	0,36	0,28	0,25	0,23	1,36	1,64	1,89	2,20	2,34	2,43
Jul	2,598	56,567	147	0,88	0,48	0,33	0,24	0,16	0,13	0,12	1,45	1,83	2,19	2,64	2,83	2,97
Ago	1,781	63,763	114	0,82	0,38	0,24	0,15	0,09	0,07	0,06	1,52	2,00	2,46	3,06	3,31	3,50
Set	4,045	39,403	159	0,92	0,58	0,44	0,34	0,26	0,23	0,21	1,38	1,67	1,93	2,26	2,40	2,50
Out	2,960	66,226	196	0,89	0,51	0,36	0,27	0,19	0,16	0,14	1,43	1,78	2,11	2,52	2,69	2,82
Nov	2,055	76,074	156	0,84	0,42	0,27	0,18	0,11	0,09	0,08	1,49	1,93	2,35	2,89	3,12	3,28
Dez	2,388	55,944	134	0,86	0,46	0,31	0,22	0,14	0,12	0,10	1,47	1,87	2,25	2,72	2,93	3,08
Ano	1817,52	413,34	1818	1,00	0,81	0,71	0,63	0,53	0,50	0,47	1,19	1,29	1,37	1,47	1,50	1,53

Região 4																
Período	Parâmetros		Média	Quantis Adimensionais não-excedentes							Quantis Adimensionais excedentes					
	θ1	θ2		Tr (anos)							Tr (anos)					
				2	5	10	20	50	75	100	5	10	20	50	75	100
Jan	2,899	55,790	162	0,89	0,50	0,36	0,26	0,18	0,15	0,14	1,43	1,79	2,12	2,54	2,72	2,84
Fev	2,631	61,277	161	0,88	0,48	0,33	0,24	0,16	0,14	0,12	1,45	1,83	2,18	2,63	2,82	2,95
Mar	2,868	43,358	124	0,89	0,50	0,36	0,26	0,18	0,15	0,14	1,43	1,79	2,13	2,55	2,73	2,85
Abr	2,428	61,948	150	0,87	0,46	0,31	0,22	0,14	0,12	0,11	1,46	1,86	2,23	2,71	2,91	3,05
Mai	2,142	72,606	155	0,85	0,43	0,28	0,19	0,12	0,10	0,08	1,49	1,91	2,32	2,84	3,06	3,22
Jun	4,537	32,489	147	0,93	0,60	0,46	0,37	0,28	0,25	0,23	1,36	1,63	1,88	2,18	2,31	2,40
Jul	2,657	62,307	166	0,88	0,48	0,34	0,24	0,16	0,14	0,12	1,45	1,82	2,17	2,62	2,81	2,94
Ago	1,915	69,519	133	0,83	0,40	0,25	0,17	0,10	0,08	0,07	1,50	1,96	2,40	2,97	3,21	3,38
Set	5,114	33,948	174	0,94	0,62	0,49	0,40	0,31	0,28	0,26	1,34	1,59	1,82	2,10	2,22	2,30
Out	3,962	53,964	214	0,92	0,57	0,43	0,34	0,25	0,22	0,20	1,38	1,67	1,94	2,28	2,42	2,52
Nov	3,221	50,797	164	0,90	0,53	0,38	0,29	0,20	0,18	0,16	1,41	1,75	2,06	2,44	2,61	2,73
Dez	3,673	43,204	159	0,91	0,56	0,42	0,32	0,23	0,21	0,19	1,39	1,70	1,98	2,34	2,49	2,59
Ano	1911,88	378,32	1912	1,00	0,83	0,75	0,67	0,59	0,56	0,54	1,17	1,25	1,33	1,41	1,44	1,46

Região 5																
Período	Parâmetros		Média	Quantis Adimensionais não-excedentes							Quantis Adimensionais excedentes					
	θ1	θ2		Tr (anos)							Tr (anos)					
				2	5	10	20	50	75	100	5	10	20	50	75	100
Jan	3,632	50,470	183	0,91	0,55	0,41	0,32	0,23	0,20	0,18	1,39	1,70	1,99	2,34	2,50	2,60
Fev	2,743	72,808	200	0,88	0,49	0,35	0,25	0,17	0,14	0,13	1,44	1,81	2,15	2,59	2,77	2,90
Mar	3,455	40,365	139	0,91	0,54	0,40	0,31	0,22	0,19	0,17	1,40	1,72	2,02	2,39	2,54	2,65
Abr	1,834	86,440	159	0,83	0,39	0,24	0,16	0,09	0,07	0,06	1,51	1,99	2,44	3,02	3,27	3,45
Mai	1,707	103,048	176	0,81	0,37	0,23	0,14	0,08	0,06	0,05	1,52	2,02	2,50	3,11	3,38	3,56
Jun	3,078	50,864	157	0,89	0,52	0,37	0,28	0,19	0,17	0,15	1,42	1,76	2,08	2,48	2,65	2,77
Jul	1,918	78,583	151	0,83	0,40	0,26	0,17	0,10	0,08	0,07	1,50	1,96	2,40	2,97	3,21	3,38
Ago	2,156	57,487	124	0,85	0,43	0,28	0,19	0,12	0,10	0,09	1,48	1,91	2,32	2,83	3,06	3,21
Set	3,969	42,017	167	0,92	0,57	0,43	0,34	0,25	0,22	0,20	1,38	1,67	1,94	2,28	2,42	2,52
Out	5,590	42,061	235	0,94	0,64	0,51	0,42	0,33	0,30	0,28	1,33	1,57	1,78	2,05	2,16	2,24
Nov	3,711	48,520	180	0,91	0,56	0,42	0,32	0,24	0,21	0,19	1,39	1,70	1,98	2,33	2,48	2,58
Dez	4,219	39,512	167	0,92	0,59	0,45	0,35	0,27	0,24	0,22	1,37	1,65	1,91	2,23	2,37	2,46
Ano	2044,38	486,08	2044	1,00	0,80	0,70	0,61	0,51	0,47	0,45	1,20	1,30	1,39	1,49	1,53	1,55

Região 6																
Período	Parâmetros		Média	Quantis Adimensionais não-excedentes							Quantis Adimensionais excedentes					
	01	02		Tr (anos)							Tr (anos)					
				2	5	10	20	50	75	100	5	10	20	50	75	100
Jan	2,721	55,361	151	0,88	0,49	0,34	0,25	0,17	0,14	0,13	1,44	1,81	2,16	2,60	2,78	2,91
Fev	2,945	49,869	147	0,89	0,51	0,36	0,27	0,19	0,16	0,14	1,43	1,78	2,11	2,52	2,70	2,82
Mar	2,897	39,106	113	0,89	0,50	0,36	0,26	0,18	0,15	0,14	1,43	1,79	2,12	2,54	2,72	2,84
Abr	2,306	60,466	139	0,86	0,45	0,30	0,21	0,13	0,11	0,10	1,47	1,88	2,27	2,76	2,97	3,12
Mai	1,890	69,710	132	0,83	0,40	0,25	0,17	0,10	0,08	0,07	1,51	1,97	2,42	2,98	3,23	3,40
Jun	4,021	34,168	137	0,92	0,58	0,44	0,34	0,26	0,23	0,21	1,38	1,67	1,94	2,27	2,41	2,51
Jul	3,067	51,951	159	0,89	0,52	0,37	0,28	0,19	0,17	0,15	1,42	1,77	2,09	2,49	2,66	2,78
Ago	1,766	68,278	121	0,82	0,38	0,23	0,15	0,09	0,07	0,06	1,52	2,00	2,47	3,07	3,33	3,51
Set	4,488	36,370	163	0,93	0,60	0,46	0,37	0,28	0,25	0,23	1,36	1,63	1,88	2,19	2,32	2,41
Out	3,406	55,274	188	0,90	0,54	0,40	0,30	0,22	0,19	0,17	1,40	1,73	2,02	2,40	2,56	2,67
Nov	2,935	53,851	158	0,89	0,51	0,36	0,27	0,18	0,16	0,14	1,43	1,78	2,11	2,53	2,70	2,83
Dez	3,656	37,954	139	0,91	0,56	0,42	0,32	0,23	0,20	0,19	1,39	1,70	1,99	2,34	2,49	2,60
Ano	1751,05	330,60	1751	1,00	0,84	0,76	0,69	0,61	0,58	0,56	1,16	1,24	1,31	1,39	1,42	1,44

Região 7																
Período	Parâmetros		Média	Quantis Adimensionais não-excedentes							Quantis Adimensionais excedentes					
	01	02		Tr (anos)							Tr (anos)					
				2	5	10	20	50	75	100	5	10	20	50	75	100
Jan	3,400	44,963	153	0,90	0,54	0,40	0,30	0,22	0,19	0,17	1,41	1,73	2,03	2,40	2,56	2,67
Fev	4,296	35,977	155	0,92	0,59	0,45	0,36	0,27	0,24	0,22	1,37	1,65	1,90	2,22	2,35	2,45
Mar	3,722	29,750	111	0,91	0,56	0,42	0,32	0,24	0,21	0,19	1,39	1,69	1,98	2,33	2,47	2,58
Abr	2,368	47,580	113	0,86	0,46	0,31	0,22	0,14	0,12	0,10	1,47	1,87	2,25	2,73	2,94	3,09
Mai	2,127	54,534	116	0,85	0,43	0,28	0,19	0,12	0,10	0,08	1,49	1,92	2,33	2,85	3,07	3,23
Jun	3,364	36,124	122	0,90	0,54	0,40	0,30	0,21	0,19	0,17	1,41	1,73	2,03	2,41	2,57	2,68
Jul	3,003	48,848	147	0,89	0,51	0,37	0,27	0,19	0,16	0,15	1,43	1,77	2,10	2,50	2,68	2,80
Ago	2,037	56,783	116	0,84	0,42	0,27	0,18	0,11	0,09	0,08	1,49	1,94	2,36	2,90	3,13	3,29
Set	4,062	34,388	140	0,92	0,58	0,44	0,35	0,26	0,23	0,21	1,38	1,67	1,93	2,26	2,40	2,50
Out	4,154	39,425	164	0,92	0,58	0,44	0,35	0,26	0,23	0,21	1,37	1,66	1,92	2,24	2,38	2,48
Nov	3,953	33,114	131	0,92	0,57	0,43	0,34	0,25	0,22	0,20	1,38	1,67	1,94	2,28	2,42	2,52
Dez	4,396	31,863	140	0,93	0,59	0,46	0,36	0,28	0,25	0,23	1,36	1,64	1,89	2,20	2,34	2,43
Ano	1606,41	313,68	1606	1,00	0,84	0,75	0,68	0,60	0,57	0,55	1,16	1,25	1,32	1,40	1,43	1,45

Região 8																
Período	Parâmetros		Média	Quantis Adimensionais não-excedentes							Quantis Adimensionais excedentes					
	01	02		Tr (anos)							Tr (anos)					
				2	5	10	20	50	75	100	5	10	20	50	75	100
Jan	2,176	55,453	121	0,85	0,43	0,29	0,20	0,12	0,10	0,09	1,48	1,91	2,31	2,82	3,04	3,20
Fev	2,613	53,527	140	0,88	0,48	0,33	0,24	0,16	0,13	0,12	1,45	1,83	2,18	2,63	2,83	2,96
Mar	2,284	50,602	116	0,86	0,45	0,30	0,21	0,13	0,11	0,10	1,47	1,89	2,28	2,77	2,98	3,13
Abr	2,744	50,738	139	0,88	0,49	0,35	0,25	0,17	0,14	0,13	1,44	1,81	2,15	2,59	2,77	2,90
Mai	1,925	62,675	121	0,83	0,40	0,26	0,17	0,10	0,08	0,07	1,50	1,96	2,40	2,96	3,20	3,38
Jun	3,397	39,810	135	0,90	0,54	0,40	0,30	0,22	0,19	0,17	1,41	1,73	2,03	2,40	2,56	2,67
Jul	3,036	50,609	154	0,89	0,51	0,37	0,28	0,19	0,16	0,15	1,42	1,77	2,09	2,49	2,67	2,79
Ago	2,220	51,131	114	0,85	0,44	0,29	0,20	0,13	0,10	0,09	1,48	1,90	2,30	2,80	3,02	3,17
Set	3,798	35,972	137	0,91	0,56	0,42	0,33	0,24	0,21	0,19	1,39	1,69	1,97	2,31	2,46	2,56
Out	3,583	38,174	137	0,91	0,55	0,41	0,32	0,23	0,20	0,18	1,40	1,71	2,00	2,36	2,51	2,62
Nov	3,102	41,477	129	0,89	0,52	0,38	0,28	0,20	0,17	0,15	1,42	1,76	2,08	2,48	2,65	2,77
Dez	2,779	44,191	123	0,88	0,49	0,35	0,25	0,17	0,15	0,13	1,44	1,80	2,15	2,58	2,76	2,89
Ano	1555,24	317,20	1555	1,00	0,83	0,74	0,66	0,58	0,55	0,53	1,17	1,26	1,34	1,42	1,45	1,47

## 7.5 Modelo para Alocação de pontos não amostrados

### 7.5.1 Determinação da matriz de pertinência

Nesta etapa, uma nova matriz de pertinência foi calculada, considerando-se como centro do agrupamento as 3 coordenadas geográficas e a média da precipitação anual.

O ponderador que gerou, no teste de aderência regional, os resultados mais próximos dos apresentados no agrupamento original, continuou sendo  $w=3$ , aplicado apenas à média da precipitação anual.

Com isso, embora 12 estações tenham sido alocadas em regiões diferentes daquelas que haviam sido alocadas originalmente, esse fato não ocasionou diferença significativa no resultado do teste de aderência regional.

Comparando-se este resultado com o resultado do agrupamento original, percebe-se que os valores são extremamente próximos (Tabela 7.22).

Tabela 7.22: Séries discordantes no teste de aderência.

Alocação	Séries Discordantes (%)												Ano	Média
	Jan	Fev	Mar	Abr	Mai	Jun	Jul	Ago	Set	Out	Nov	Dez		
Original	3,5	3,5	3,5	1,0	0,5	5,0	6,0	3,5	2,0	0,5	0,0	2,5	1,5	2,5
Do modelo	3,5	3,5	4,0	3,0	0,5	6,0	5,0	4,0	2,0	1,0	0,0	2,5	2,5	2,9

Isso significa que, possivelmente, pontos não amostrados venham a ser adequadamente alocados em uma região, com grande possibilidade de a curva regional ser uma boa estimativa da distribuição de probabilidade local, para cada um dos pontos.

A hipótese de que pontos com baixo máximo grau de pertinência estão mais suscetíveis à mudança de grupo não fica bem caracterizada. Tomando-se, por exemplo, os 14 pontos com  $\mu_{max} < 0,3$ , apenas 2 deles apresentariam alguma melhora com a realocação para outro grupo, e apenas metade apresenta discordância em relação a algum dos 13 períodos.

O que se verifica é que pontos com  $\mu_{max}$  elevado dificilmente apresentarão séries discordantes com a curva regional, entretanto, com relação aos pontos com baixo  $\mu_{max}$ , não há como tirar conclusões, uma vez que máximos graus de pertinência baixos podem estar associados, tanto à má adequação ao grupo, quanto a semelhanças a com mais de um grupo ao mesmo tempo.

### 7.5.2 Estimativa do fator de escala local

Avaliou-se a qualidade da estimativa do fator de escala local através da média dos fatores de escala regionais, ponderados pelos graus de pertinência.

A Tabela 7.23 abaixo apresenta a raiz do erro médio quadrático relativo (Remqr) da estimativa da precipitação baseada nos graus de pertinência. Calculou-se o mesmo erro, ao adotar a média da distribuição regional e o centro do agrupamento difuso, como estimativa dos valores locais.

Nota-se que os erros do modelo que utiliza os graus de pertinência são mais elevados que quando se utiliza o protótipo (centro do agrupamento), ou a média da distribuição regional, sendo estes dois casos com erros semelhantes.

Tabela 7.23 Erros na estimativa das médias de precipitação locais.

Estimativa	REMQR (%)												Ano	Média
	Jan	Fev	Mar	Abr	Mai	Jun	Jul	Ago	Set	Out	Nov	Dez		
graus de pertinência	12,5	14,4	13,9	14,0	14,0	15,1	19,4	19,5	14,1	12,6	12,2	13,2	10,9	14,3
centro do agrupamento	7,7	10,4	10,0	9,8	10,0	11,0	16,5	16,5	9,3	8,4	7,6	8,9	4,1	10,0
média da distribuição	7,7	10,5	10,1	9,8	10,0	10,6	15,3	14,8	9,2	8,4	8,0	8,8	4,0	9,8

O que ocorre é que a estimativa local, sendo obtida pela média de outros valores médios, ainda que ponderada pelos graus de pertinência, tende a resultar uma perda ainda maior dos detalhes que fazem parte da caracterização local, detalhes estes que já haviam sido parcialmente perdidos pela modelagem regional.

## 8 CONCLUSÕES E RECOMENDAÇÕES

### 8.1 Conclusões

Pesquisou-se nesse trabalho a possibilidade de utilizar uma metodologia para análise regional de frequências, admitindo-se que as séries que formam uma região homogênea sigam uma mesma distribuição de probabilidades, que permita a dispensa da utilização de fatores de escala locais, e cujos parâmetros sejam estimados pela média dos parâmetros de ajuste das séries locais.

Foram construídos dois recursos de suporte a essa metodologia: um critério para avaliar a qualidade do agrupamento e um teste de homogeneidade.

Tanto estes dois recursos, como a avaliação da qualidade do uso das médias dos parâmetros de ajuste das séries locais para estimar a distribuição regional, foram baseados na medida da máxima diferença de probabilidades entre duas séries.

Esta medida foi escolhida por ser a base do teste de aderência de Kolmogorov-Smirnov (teste KS), que é um teste largamente utilizado para verificar a qualidade do ajuste de uma amostra ou para identificar se duas séries são igualmente distribuídas. Desta forma, foi possível obter maior segurança na análise dos resultados.

Com base nas investigações realizadas, verificou-se que a obtenção dos parâmetros regionais através da média dos parâmetros locais é adequada para as distribuições Normal e Gama, desde que sejam aplicadas as devidas correções aos parâmetros alfa e beta da distribuição Gama, uma vez que, provavelmente devido a não linearidade do ajuste, ocorre uma distorção desses parâmetros.

Considerando os termos de correção, pode-se dizer que, se a região é homogênea, o uso da média dos parâmetros locais traz resultados satisfatórios para os dois modelos de distribuição avaliados, uma vez que, em regiões com trinta anos de dados, a qualidade da estimativa foi equivalente à qualidade que seria resultante do ajuste individual de séries de aproximadamente 300 anos.

Salienta-se, porém, que a adoção deste modelo requer uma investigação preliminar de possíveis distorções nos parâmetros obtidos, quando forem utilizados outros tipos de distribuição.

Para avaliar a qualidade do agrupamento, foi desenvolvido o teste de aderência regional, que utiliza o teste KS, ao nível de 20% de significância, para verificar, em cada região, se a distribuição

regional e as séries locais são igualmente distribuídas. O teste avalia o percentual de séries que não passam no teste KS, estabelecendo-se, como critério, que quanto menor esse percentual, melhor é a configuração do agrupamento.

O teste de homogeneidade, por sua vez, consiste no cálculo da medida de heterogeneidade  $Hdp$ , que tem por princípio comparar a região real com regiões simuladas, em termos da média das máximas diferenças de probabilidades entre as séries locais e sua distribuição regional.

Verificou-se que os valores limites para a condição de homogeneidade são praticamente insensíveis aos parâmetros da distribuição, mas dependentes do número de séries nas regiões, sendo mais rigoroso na medida em que o número de séries é maior.

Note-se que, por ser baseado na medida da máxima diferença de probabilidades, o teste de homogeneidade independe do método com que se determina a distribuição regional, e também pode ser utilizado para testar a homogeneidade nos casos de regionalização da distribuição adimensional, devendo-se, nesse caso, adimensionalizar as séries locais das regiões simuladas, mesmo que sejam geradas a partir de uma distribuição com média unitária.

No estudo de caso, estes recursos foram utilizados para a análise regional de precipitações totais mensais e anuais, utilizando-se agrupamento difuso para a determinação das regiões.

Observou-se que o teste de aderência regional mostrou-se útil no processo de identificação das regiões, sendo fundamental na fase de ponderação das variáveis e no julgamento entre as opções de melhor número de regiões apontadas pelos índices de validade do agrupamento.

O uso do teste de aderência regional permitiu a obtenção de elevado índice de homogeneidade nas regiões e reduziu a necessidade de intervenções subjetivas na configuração das regiões originadas pelo agrupamento.

Ao avaliar os resultados dos índices de validade do agrupamento, verifica-se que não é razoável a busca por uma única alternativa ao número de grupos, como normalmente se pratica, mas sim que sejam apontados alguns bons resultados, indicados por máximos ou mínimos locais (dependendo do índice), ficando a escolha final sujeita ao nível de detalhamento necessário à análise ou ao confronto com outros índices, que considerem os objetivos da análise, como é o caso do índice de aderência regional.

O índice de validade  $Pds$  (partição difusa significativa), proposto nesse trabalho, mostrou-se coerente com os índices  $V_{XB}$  e  $V_K$ , sendo mais revelador que a partição difusa ( $Pd$ ), conforme esperado. O fato destes índices terem apontado resultados semelhantes, e da configuração do

agrupamento ter apresentado resultados satisfatórios, reforça também a sugestão da utilização conjunta de vários índices para a definição das alternativas de número de grupos mais adequadas.

Com relação às contribuições do método de agrupamento difuso *Fuzzy C-Means*, destaca-se a possibilidade de alocação de pontos sem registro de precipitação mensal, a partir das variáveis latitude, longitude, altitude e média da precipitação anual (que pode ser obtida por isoietas). Os resultados do ajuste sugerem que, possivelmente, pontos sem dados possam ser corretamente alocados, embora isto dependa fortemente de uma boa estimativa da precipitação anual.

O uso dos graus de pertinência para auxiliar os ajustes no resultado final do agrupamento não foi conclusivo, uma vez que se verificou casos de má alocação associados a máximos graus de pertinência ( $\mu_{max}$ ) baixos, embora em outros casos valores baixos de  $\mu_{max}$  estavam apenas relacionados à boa adequação do ponto a mais de um grupo.

Já a utilização dos graus de pertinência para fazer estimativas das médias de precipitação pontuais apresentou erros superiores aos erros verificados ao adotar-se o centro do agrupamento ou a média da distribuição regional como estimativa pontual.

Ou seja, o modelo de estimativa utilizando graus de pertinência se mostrou ineficiente, e a explicação provavelmente está no fato de que a ponderação exercida pelos graus de pertinência filtra excessivamente as estimativas, ocasionando uma perda ainda maior dos detalhes inerentes a cada local.

A partir do estudo de caso, esta pesquisa contribuiu para a compreensão do comportamento da precipitação nos estados do Rio Grande do Sul e Santa Catarina, tendo sido identificadas 9 regiões com elevado percentual de homogeneidade e fisicamente coerentes, fornecendo subsídios importantes para as atividades de planejamento que requeiram o conhecimento das probabilidades de ocorrência de precipitações totais anuais e mensais.

Note-se que apenas 9 distribuições de probabilidades regionais, em cada um dos 13 períodos, puderam substituir satisfatoriamente 197 estações, observando-se que as distribuições regionais foram validadas para 97,9% das séries, segundo o teste KS a 20% de significância, verificando-se que é possível dispensar o conhecimento dos fatores de escala locais.

Em resumo, pode-se concluir que:

- A obtenção dos parâmetros da distribuição regional pela média dos parâmetros de ajuste das séries locais apresentou resultados satisfatórios para as distribuições Normal e Gama, esta última necessitando de um fator de correção;
- O teste de aderência regional mostrou-se útil para auxiliar no processo de ponderação das variáveis e de definição do número de regiões;
- Foi possível construir um teste de homogeneidade que independe do método de determinação da distribuição regional;
- O uso dos graus de pertinência foi insatisfatório para fazer estimativas pontuais e pouco contribuiu na etapa de ajustes no agrupamento;
- O ajuste do modelo de alocação para novos pontos, por sua vez, apresentou bons resultados, o que indica que pontos sem dados observados possivelmente serão bem alocados com o uso deste modelo;
- Os resultados do estudo de caso demonstram que é possível considerar uma mesma distribuição na região homogênea, desobrigando a determinação do fator de escala local;
- Esta pesquisa contribuiu para a compreensão do comportamento da precipitação nos estados do Rio Grande do Sul e parte de Santa Catarina, constituindo um importante subsídio para as atividades de planejamento que requeiram o conhecimento das probabilidades de ocorrência de precipitações totais anuais e mensais.

## 8.2 Recomendações

Para dar continuidade a essa pesquisa recomenda-se:

- Verificar se é possível obter resultados satisfatórios com séries de diversos tamanhos na mesma região, sugerindo-se nesse caso que a média dos parâmetros seja ponderada pelo tamanho da série. Entretanto, é recomendável se certificar de que as séries utilizadas possam ser consideradas estacionárias.
- Verificar a existência de melhoria de qualidade na análise regional ao se utilizar regiões sobrepostas, ou seja, considerar a colaboração de um ponto na obtenção de curvas regionais em mais de uma região.
- Pesquisar as vantagens da incorporação de outras variáveis na identificação das regiões (como direção preferencial de ventos, mês mais chuvoso...).
- Comparar os resultados da metodologia proposta com o resultado de outras metodologias de análise regional, como a que é baseada em momentos-L.
- Estudar a eficiência do modelo de alocação para pontos não amostrados, separando uma parte das amostras para agrupamento e ajuste do modelo e outra para a sua validação.
- Avaliar a validade da metodologia para a análise regional de eventos extremos.

## 9 REFERÊNCIAS BIBLIOGRÁFICAS

- AMINI, M.; AFYUNI, M.; FATHIANPOUR, N.; KHADEMIA, H.; FLÜHLER, H. Continuous soil pollution mapping using fuzzy logic and spatial interpolation. **Geoderma**, 124, p. 223–233, 2005.
- ANDRADE, L. P. Procedimento Interativo de Agrupamento de dados. 2004. Dissertação de Mestrado. Universidade Federal do Rio de Janeiro, COPPE. Rio de Janeiro, 2004
- ARAÚJO, W. F.; ANDRADE JÚNIOR, A. S.; MEDEIROS, R. D.; SAMPAIO, R. A. Precipitação pluviométrica mensal provável em Boa Vista, Estado de Roraima, Brasil. **Revista Brasileira de Engenharia Agrícola e Ambiental**, v.5, n.3, p.563-567, 2001
- ÁVILA, A.M.H. **Regime de precipitação no Estado do Rio Grande do Sul com base em séries de longo prazo**. 1994. Dissertação de Mestrado – Faculdade de Agronomia, Universidade Federal do Rio Grande do Sul, Porto Alegre, 1994.
- BARBALHO, V.M.S. **Sistemas baseados em conhecimento e lógica difusa para simulação do processo chuva-vazão**. 2001. Tese de Doutorado - Universidade Federal do Rio de Janeiro, COPPE, Rio de Janeiro, 2001.
- BECKER, C. T. et al. Regionalização da precipitação e temperatura no Estado do Rio Grande do Sul a partir da análise de agrupamento. In: CONGRESSO BRASILEIRO DE METEOROLOGIA, 7., 1992, São Paulo. **Anais...** Rio de Janeiro: SBMet, 1992. p. 225 - 229.
- BERTONI, J.C. e TUCCI, C.E.M. **Precipitação**. In: Hidrologia: Ciência e Aplicação. 3<sup>a</sup> ed., Porto Alegre, Editora da UFRGS/ABRH, 2004.
- BEZDEK, J.C. Pattern recognition with fuzzy objective function algorithms. **Plenum Press**, New York. 1981.
- BHASKAR, N.R.; e O’CONNOR, C.A. Comparison of Method of Residuals and Cluster Analysis for flood Regionalization. **Journal of Water Resources Planning and Management**, vol. 115, n. 6, November, 1989.
- BURN, D.H. (1989): Cluster Analysis as Applied to Regional Flood Frequency. **Journal of Water Resources Planning and Management**, vol. 115, n. 5, September, 1989.
- CÂMARA, G.; BARBOSA, C. C.; CORDEIRO, J. P.; LOPES, E.; FREITAS, U. M.; LUCENA, I. **Álgebra de Mapas**. In: Introdução à ciência da geoinformação. São José dos Campos, INPE, 2001

- CARGNELUTTI FILHO, A.; MATZENAUER, R.; ANJOS, C. S.; SAMPAIO, M. R. Probabilidade de ocorrer precipitação pluvial decendial igual ou superior à evapotranspiração de referência, no Rio Grande do Sul. **Revista Brasileira de Agrometeorologia**, Santa Maria, v. 13, n. 2, p. 219-224, 2005.
- CARLANTONIO, L. M. **Novas Metodologias para Clusterização de Dados**. 2001. Dissertação de Mestrado. Universidade Federal do Rio de Janeiro, COPPE. Rio de Janeiro, 2001
- CARVALHO, T. L. L. e PEDROLLO, O.C. Caracterização Pluviométrica utilizando agrupamento difuso. In: SIMPÓSIO BRASILEIRO DE RECURSOS HÍDRICOS, 16. 2005. João Pessoa. **Anais...** Associação Brasileira de Recursos Hídricos, 2005. CD-ROM.
- CLARKE, R. T. **Hidrologia Estatística**. In: Hidrologia: Ciência e Aplicação. 3<sup>a</sup> ed., Porto Alegre, Editora da UFRGS/ABRH, 2004.
- COSTA NETO, P. L. O. **Estatística**. São Paulo, Edgard Blücher, 1977.
- DAVIS, E. G. E NAGHETTINI, M. C. Estudo de Chuvas intensas no Estado do Rio de Janeiro. Belo Horizonte: CPRM, 2001.
- DINPASHOH A, Y. et alli Selection of variables for the purpose of regionalization of Iran's precipitation climate using multivariate methods. **Journal of Hydrology**, 297, p.109–12, 2004.
- GOKTEPE, A.B.; ALTUN, S.; SEZER, A. Soil Clustering by Fuzzy C-Means Algorithm. **Advances in Engineering Software**, 36, p. 691–698, 2005
- GRIMM. A.M. **Meteorologia Básica** – Notas de aula. Primeira versão eletrônica – setembro 1999. <http://fisica.ufpr.br/grimm/aposmeteo/introd.html>
- HALL, M.J.; e MINNS, W.A. The classification of hydrologically homogeneous regions. **Hydrological Sciences Journal**, 44(5), October 1999.
- HOSKING, J.R.M., WALLIS, J.R., 1997. **Regional Frequency Analysis: An approach based on L-Moments**. Cambridge University press, Cambridge
- JINGYI, Z.; HALL, M.J. Regional flood frequency analysis for the Gan-Ming River basin in China. **Journal of Hydrology**, 296, p. 98–117, 2004.
- KELLER FILHO, T.; ASSAD, E.D.; LIMA, P.R.S.R. Regiões pluviometricamente homogêneas no Brasil. **Pesquisa agropecuária brasileira**, Brasília, v.40, n.4, p.311-322, abr. 2005
- KIM, D-W; LEE, K.H.; LEE, D. On cluster validity index for estimation of the optimal number of fuzzy clusters. **Pattern Recognition**, 37 (2004) 2009 – 2025
- KULKARNI, A.; KRIPALANI, R. H. Rainfall Patterns over India: Classification with Fuzzy c-Means Method. **Theoretical and Applied Climatology**. 59, 137±146, 1998.

- LANNA, A.E. Províncias hidrológicas e bacias representativas. In: Simpósio Brasileiro de Hidrologia e Recursos Hídricos, 5, 1983. Blumenau. **Anais...** Associação Brasileira de Recursos Hídricos, Florianópolis, 1983. vol. 1, p. 3-24.
- LANNA, A. E. **Elementos de Hidrologia Estatística**. In: Hidrologia: Ciência e Aplicação. 3<sup>a</sup> ed., Porto Alegre, Editora da UFRGS/ABRH, 2004.
- LIN, G.; CHEN L. Identification of homogeneous regions for regional frequency analysis using the self-organizing map. **Journal of Hydrology**, 324 (2006) 1–9
- LOPES, F. Z. **Relação entre o MEI (Multivariate Enso Index) e a precipitação pluvial no Estado do Rio Grande do Sul**. 2006. Dissertação de Mestrado - Universidade Federal de Pelotas, Pelotas, 2006.
- LYRA, G. B.; GARCIA B. I. L.; PIEDADE S.M.S.; SEDIYAMA G. C.; SENTELHAS P.C. Regiões homogêneas e funções de distribuição de probabilidade da precipitação pluvial no Estado de Táchira, Venezuela. **Pesquisa agropecuária brasileira**, Brasília, v.41, n.2, p.205-215, fev. 2006
- MARQUES, J. Q. R. **Variabilidade espacial e temporal da precipitação pluvial no Rio Grande do Sul e sua relação com indicadores oceânicos**. 2005. Tese de Doutorado – Faculdade de Agronomia, Universidade Federal do Rio Grande do Sul, Porto Alegre, 2005.
- MATZENAUER, R.; CARGNELUTTI FILHO, A; ANJOS, C. S. Probabilidade de ocorrer precipitação pluvial igual ou superior à evapotranspiração máxima na cultura da soja, no Rio Grande do Sul. **Revista Brasileira de Agrometeorologia**, Santa Maria, v. 11, n. 2, p. 313-320, 2003.
- MORENO, J, A. **Clima do Rio Grande do Sul**. Secretaria da Agricultura, Porto Alegre-RS, 42p. 1961.
- NATHAN, R.J. E MCMAHON, T.A. Identification of Homogeneous Regions for the Purposes of Regionalisation. **Journal of Hydrology**, 121, p.217-238. , 1990.
- NOVAES, U. R. **Agrupamento de Dados através de algoritmos Swarm** 2002. Dissertação de Mestrado - Universidade Federal do Rio de Janeiro, COPPE , Rio de Janeiro, 2002.
- OLIVEIRA, G. S. O. **El Niño e Você - o fenômeno climático**. São José dos Campos, Editora Transtec, 2001.
- ÖZELKAN, E.C.; GALAMBOSI, A.; DUCKSTEIN, L.; BÁRDOSSY, A. A multi-objective fuzzy classification of large scale atmospheric circulation patterns for precipitation modeling. **Applied Mathematics and Computation**, 91, p.127-142, 1998.

- PAKHIRA, M.K.; BANDYOPADHYAY, S.; MAULIK, U A study of some fuzzy cluster validity indices, genetic clustering and application to pixel classification. **Fuzzy Sets and Systems** 155 (2005) 191–214.
- PEDROLLO, O.C. **Previsão em tempo atual de cheias com uso de sistema especialista difuso**. 2000. Tese de Doutorado – Instituto de Pesquisas Hidráulicas, Universidade Federal do Rio Grande do Sul, Porto Alegre, 2000.
- PINTO, E.J.A.; e NAGHETTINI, M. Definição de Regiões Homogêneas e a Regionalização de Precipitações Máximas Diárias da Bacia do Alto São Francisco. In: Simpósio Brasileiro de Recursos Hídricos, 13. 1999. Belo Horizonte. **Anais...** Associação Brasileira de Recursos Hídricos, 1999. CD-ROM.
- RAO, A. R.; SRINIVAS, V.V. Regionalization of watersheds by fuzzy cluster analysis **Journal of Hydrology** 318 (2006) 57-79
- ROSS, T. J. **Fuzzy Logic with Engineering Applications**. McGraw –Hill, New York, 1995.
- SILVA, J. B.; CUNHA, F. B. E GAVIÃO NETO, W. P. Modelagem das chuvas trimestrais por regiões homogêneas no Estado do Ceará. **Revista Brasileira de Agrometeorologia**, Santa Maria, v. 9, n. 2, p. 317-324, 2001
- SOTÉRIO, P. W.; PEDROLLO, M. C.; ANDRIOTTI, J. L. Mapa de isoietas do rio grande do sul. In: SIMPÓSIO BRASILEIRO DE RECURSOS HÍDRICOS, 16. 2005. João Pessoa. **Anais**. Associação Brasileira de Recursos Hídricos, 2005. CD-ROM.
- SOUZA, I. A.; AZEVEDO, J. R. G.; MACIEL NETTO, A. Variabilidade espacial das chuvas na bacia hidrográfica do rio una-pe. In: SIMPÓSIO BRASILEIRO DE RECURSOS HÍDRICOS, 16. 2005. João Pessoa. **Anais**. Associação Brasileira de Recursos Hídricos, 2005. CD-ROM.
- TUBELIS, A.; NASCIMENTO, T. **Meteorologia descritiva: fundamentos e aplicações**. São Paulo, Nobel, 1987.
- TUCCI, C.E.M. **Regionalização de Vazões**. Ed. Universidade/UFRGS, 2002.
- VALVERDE, A. E. L. **Regionalização de Chuvas Intensas para a Bacia do Rio Doce**. 2001. Dissertação de Mestrado – Universidade Federal de Viçosa, Viçosa-MG, 2001.
- VAREJÃO-SILVA, M. A **Meteorologia e Climatologia**. Versão digital 2. Recife, 2006. disponível em: <http://www.agritempo.gov.br>
- WU, K-L; YANG, M-S. A cluster validity index for fuzzy clustering. **Pattern Recognition Letters**, vol. 26, n.9 julho 2005.
- ZADEH, L. Fuzzy sets. **Information and control**, 8(3): 338-353, 1965.

## Anexo A

### Localização das Estações Pluviométricas

ID	Código	Latitude (°)	Longitude (°)	Altitude (m)	Ano inicial	OP**	Sub Bacia	Nome da Estação	UF	ID	Código	Latitude (°)	Longitude (°)	Altitude (m)	Ano inicial	OP**	Sub Bacia	Nome da Estação	UF
1	2650019	-26:55:48	-50:41:17	1000	jul/76	1	71	LEBON REGIS	SC	51	2849020	-28:36:24	-49:01:59	10	mar/76	3	84	JAGUARUNA	SC
2	2651001	-26:52:24	-51:47:47	1000	abr/43	1	73	CAMPINA DA ALEGRIA	SC	52	2849021	-27:59:19	-49:34:39	997	ago/76	1	71	URUBICI	SC
3	2651022	-26:36:31	-51:52:54	1250	mar/72	1	73	SANTO AGOSTINHO	SC	53	2849023	-28:22:05	-49:48:15	1080	dez/76	1	70	DESPRAIADO	SC
4	2651036	-26:46:33	-51:15:46	1133	abr/76	1	72	QUILOMETRO 30	SC	54	2850002	-28:44:00	-50:04:00	1150	jan/61	2	86	CAP. S. JOSE DOS AUSENTES	RS
5	2651040	-26:51:42	-52:01:09	1000	set/76	1	73	PONTE SERRADA	SC	55	2850004	-28:09:08	-50:26:27	930	jun/58	1	70	COXILHA RICA	SC
6	2652000	-26:33:22	-52:19:51	760	out/57	1	73	ABELARDO LUZ	SC	56	2850008	-28:12:39	-50:45:31	640	jun/43	1	70	PASSO SOCORRO *	RS
7	2652001	-26:57:09	-52:10:57	600	ago/69	1	73	BONITO	SC	57	2850009	-28:52:00	-50:27:14	640	abr/43	1	86	PASSO TAINHAS	RS
8	2652002	-26:35:02	-52:38:25	550	mar/72	1	73	MARATA	SC	58	2851003	-28:51:12	-51:17:01	630	abr/43	1	86	ANTONIO PRADO	RS
9	2652021	-26:44:10	-52:54:01	360	mar/76	1	73	JARDINOPOLIS	SC	59	2851005	-28:33:00	-51:58:00	600	ago/44	2	86	CASCA I	RS
10	2652031	-26:21:01	-52:50:17	720	set/76	1	73	SÃO LOURENÇO DO OESTE	SC	60	2851007	-28:14:00	-51:33:00	900	nov/62	2	86	ENCRUZILHADA II	RS
11	2653001	-26:26:48	-53:04:49	920	nov/69	1	74	CAMPO ERE DNAEE-EMPASC	SC	61	2851008	-28:33:00	-51:21:00	720	jan/45	2	86	ITUIM (ENTRE RIOS)	RS
12	2653002	-26:15:55	-53:37:15	800	mar/72	1	74	DIONÍSIO CERQUEIRA	SC	62	2851009	-28:53:00	-51:42:00	650	fev/61	2	86	FAGUNDES VARELA	RS
13	2653004	-26:40:53	-53:17:17	320	nov/69	1	74	PONTE DO SARGENTO	SC	63	2851010	-28:40:00	-51:08:00	800	out/62	2	86	FAZENDA ROSEIRA	RS
14	2653005	-26:27:16	-53:29:48	720	mar/72	1	74	SÃO JOSE DO CEDRO	SC	64	2851020	-28:15:00	-51:51:00	750	jan/63	2	86	PASSO DAS PEDRAS	RS
15	2653007	-26:55:36	-53:00:28	280	jul/54	1	73	SAUDADES	SC	65	2851021	-28:52:03	-51:26:44	400	nov/57	1	86	PASSO DO PRATA	RS
16	2653013	-26:21:22	-53:16:34	850	mar/76	1	74	PALMA SOLA	SC	66	2851022	-28:37:12	-51:51:58	380	out/57	1	86	PASSO MIGLIAVACA	RS
17	2749031	-27:48:09	-49:46:41	900	jun/57	1	71	VILA CANOAS	SC	67	2851024	-28:46:07	-51:37:12	680	ago/44	1	86	PRATA	RS
18	2749035	-27:44:39	-49:56:39	900	ago/76	1	71	BOCAINA DO SUL	SC	68	2851026	-28:46:00	-51:22:00	700	ago/44	2	86	SEGREDO	RS
19	2750007	-27:55:17	-50:05:55	1196	mai/58	1	71	PAINEL	SC	69	2851027	-28:42:00	-51:56:00	620	jan/61	2	86	SERAFINA CORREIA	RS
20	2750008	-27:32:16	-50:51:35	720	mai/59	1	71	PASSO CARU	SC	70	2851028	-28:23:00	-51:50:00	650	out/62	2	86	TRINTA E CINCO	RS
21	2750010	-27:09:45	-50:28:00	980	out/59	1	71	PONTE ALTA DO NORTE	SC	71	2852004	-28:48:00	-52:22:00	700	dez/62	2	86	AULER	RS
22	2750012	-27:20:41	-50:26:02	940	ago/61	1	71	PONTE DO RIO ANTINHAS	SC	72	2852006	-28:17:38	-52:47:18	570	abr/41	1	74	CARAZINHO	RS
23	2750020	-27:39:38	-50:34:59	920	abr/76	1	71	SÃO JOSE DO CERRITO	SC	73	2852007	-28:11:18	-52:44:43	593	jan/44	1	74	COLONIA XADREZ	RS
24	2751004	-27:10:18	-51:30:01	560	abr/43	1	72	JOACABA	SC	74	2852008	-28:32:00	-52:59:00	460	set/44	2	85	COLORADO	RS
25	2751006	-27:42:14	-51:46:03	600	out/57	1	72	PAIM FILHO	RS	75	2852009	-28:56:00	-52:48:00	600	jan/63	2	85	DEPOSITO	RS
26	2751007	-27:57:01	-51:48:52	687	jul/57	1	72	SANANDUVA	RS	76	2852014	-28:55:00	-52:07:00	650	ago/44	2	86	ILOPOLIS	RS
27	2751012	-27:20:32	-51:36:30	498	abr/76	1	72	CAPINZAL	SC	77	2852015	-28:34:00	-52:51:00	490	abr/59	2	85	LAGOA TRES CANTOS	RS
28	2751017	-28:00:17	-51:27:09	950	mai/76	1	72	CLEMENTE ARGOLO	RS	78	2852016	-28:27:00	-52:12:00	580	ago/44	2	86	MARAU	RS
29	2752006	-27:51:15	-52:18:14	763	abr/43	1	73	EREBANGO	RS	79	2852023	-28:36:00	-52:36:00	430	abr/69	2	85	PONTE JACUI	RS
30	2752014	-27:57:00	-52:00:00	640	jan/63	2	72	CHARRUA	RS	80	2852024	-28:17:00	-52:34:00	600	dez/62	2	85	PULADOR	RS
31	2752017	-27:23:20	-52:27:16	350	mai/76	1	73	ITATIBA DO SUL	RS	81	2852029	-28:38:00	-52:55:00	360	jun/58	2	85	USINA COLORADO	RS
32	2753002	-27:21:00	-53:24:00	530	fev/61	2	74	FREDERICO WESTPHALEN	RS	82	2852031	-28:28:00	-52:22:00	510	mai/59	2	85	VILA TRES PASSOS	RS
33	2753004	-27:48:42	-53:01:40	350	jun/59	1	74	LINHA CESCO	RS	83	2852032	-28:49:00	-52:42:00	535	abr/59	2	85	VOLTA ALEGRE	RS
34	2753006	-27:04:16	-53:09:32	400	mai/59	1	74	PALMITOS	SC	84	2852036	-28:43:00	-52:51:00	380	set/72	2	85	ESPUMOSO	RS
35	2753007	-27:51:00	-53:46:00	520	jan/48	2	74	SANTO AUGUSTO	RS	85	2852037	-28:43:00	-52:33:00	550	set/72	2	85	DESPRAIADO	RS
36	2753009	-27:27:00	-53:56:00	450	dez/60	2	74	TRES PASSOS	RS	86	2852046	-28:03:24	-51:59:46	672	set/76	1	72	TAPEJARA	RS
37	2753013	-26:59:51	-53:31:55	557	mar/76	1	74	IPORA	SC	87	2853001	-28:14:00	-53:46:00	250	dez/60	2	75	AJURICABA	RS
38	2753014	-27:35:54	-53:04:15	378	mai/76	1	74	LIBERATO SALZANO	RS	88	2853002	-28:29:00	-53:27:00	504	jan/63	2	75	BELISARIO	RS
39	2753015	-27:54:48	-53:18:39	610	ago/76	1	74	PALMEIRA DAS MISSOES	RS	89	2853003	-28:27:18	-53:58:15	160	ago/59	1	75	CONCEICAO	RS
40	2753016	-27:30:06	-53:40:55	502	set/76	1	74	MIRAGUAI	RS	90	2853007	-28:39:00	-53:21:00	300	dez/60	2	85	PASSO DA DIVISA	RS
41	2754001	-27:18:11	-54:08:04	120	jun/49	1	74	ALTO URUGUAI	RS	91	2853008	-28:44:00	-53:09:00	250	nov/41	2	85	PASSO DO LAGOAO	RS
42	2754003	-27:38:00	-54:18:00	325	dez/60	2	74	HORIZONTAL	RS	92	2853011	-28:33:00	-53:10:00	350	out/57	2	85	PONTE SANTO ANTONIO	RS
43	2754007	-27:47:00	-54:14:00	403	dez/60	2	74	TRES DE MAIO	RS	93	2853012	-28:23:00	-53:05:00	550	abr/59	2	85	SALDANHA MARINHO	RS
44	2754008	-27:45:00	-54:28:00	330	dez/60	2	74	TUPARENDI	RS	94	2853014	-28:43:43	-53:11:09	390	set/44	1	85	SANTA CLARA DO INGAI	RS
45	2754009	-27:39:14	-54:26:32	120	jun/76	1	74	TUCUNDUVA	RS	95	2853015	-28:49:00	-53:30:00	350	jan/61	2	85	TRES CAPOES	RS
46	2754010	-27:58:05	-54:06:59	400	set/76	1	74	ESQUINA ARAUJO	RS	96	2853016	-28:16:00	-53:48:00	270	ago/69	2	75	USINA AJURICABA	RS
47	2755001	-27:51:16	-55:01:21	100	mar/49	1	74	PORTO LUCENA	RS	97	2853018	-28:16:00	-53:07:00	490	out/72	2	85	MAQUINISTA SEVERO	RS
48	2849001	-28:21:32	-49:17:42	90	mai/39	3	84	ORLEANS-MONTANTE*	SC	98	2853020	-28:50:00	-53:36:00	405	set/72	2	85	PASSO DOS ALEMAES	RS
49	2849009	-28:20:00	-49:37:00	1200	out/69	1	70	BOM JARDIM DA SERRA	SC	99	2853023	-28:23:21	-54:19:55	440	ago/76	1	75	CONDOR	RS
50	2849019	-28:50:06	-49:50:16	115	abr/76	3	84	TIMBE DO SUL	SC	100	2853026	-28:03:31	-53:03:58	450	jan/77	1	74	CHAPADA	RS

\* PCD - INPE \*\* Operadora = 1 - CPRM, 2 - CEEE, 3 - EPAGRI, 4 - INMET

ID	Código	Latitude (°)	Longitude (°)	Altitude (m)	Ano inicial	OP**	Sub Bacia	Nome da Estação	UF	ID	Código	Latitude (°)	Longitude (°)	Altitude (m)	Ano inicial	OP**	Sub Bacia	Nome da Estação	UF
101	2854001	-28:06:25	-53:59:33	447	nov/69	1	75	BOA VISTA	RS	151	2955008	-29:35:26	-55:28:55	80	jul/76	1	76	MANOEL VIANA *	RS
102	2854003	-28:01:35	-54:20:37	400	abr/43	1	74	GIRUA	RS	152	2956005	-29:07:05	-56:33:28	35	mar/70	1	75	ITAQUI	RS
103	2854004	-28:08:00	-54:33:00	200	fev/61	2	75	GUARANI DAS MISSOES	RS	153	2956006	-29:18:30	-56:03:17	60	out/69	1	76	PASSO MARIANO PINTO *	RS
104	2854005	-28:44:01	-54:38:52	160	out/57	1	75	PASSO MAJOR ZEFERINO	RS	154	2956007	-29:46:12	-56:30:57	120	abr/43	1	76	PLANO ALTO	RS
105	2854006	-28:12:38	-54:36:08	160	ago/59	1	75	PASSO VIOLA	RS	155	2956008	-29:28:28	-56:41:01	90	mai/76	1	76	JOAO ARREGUI	RS
106	2854012	-28:47:16	-54:27:07	300	set/76	1	75	COIMBRA	RS	156	2957003	-29:45:23	-57:05:37	74	jan/63	4	76	URUGUAIANA-IPAGRO	RS
107	2855001	-28:10:57	-55:38:36	60	mar/49	1	75	GARRUCHOS	RS	157	3050002	-30:15:11	-50:30:23	3	jan/74	1	87	PALMARES DO SUL	RS
108	2855002	-28:12:43	-55:19:11	80	out/57	1	75	PASSO DO SARMENTO	RS	158	3051004	-30:35:53	-51:45:16	120	set/76	1	87	CERRO GRANDE	RS
109	2855005	-28:59:42	-55:40:49	140	nov/63	1	75	FAZ. S. CECILIA DO BUTUI	RS	159	3051005	-30:06:24	-51:38:56	40	dez/66	1	87	GUAIBA COUNTRY CLUB	RS
110	2949003	-29:07:02	-49:38:20	16	mar/76	3	84	SOMBRIO	SC	160	3051016	-30:52:14	-51:47:41	65	jun/65	1	87	CAMAQUA	RS
111	2950003	-29:09:00	-50:10:00	970	mai/47	2	86	AZULEGA	RS	161	3051017	-30:46:18	-51:39:35	25	jul/76	1	87	FAZENDA DA BOA VISTA	RS
112	2950008	-29:06:00	-50:10:00	970	mar/46	2	86	CAMISAS	RS	162	3051023	-30:17:51	-51:18:51	5	mar/76	1	87	BARRA DO RIBEIRO	RS
113	2950009	-29:22:00	-50:48:00	850	mai/41	2	87	CANELA	RS	163	3052007	-30:54:00	-52:27:00	60	jan/49	2	87	PASSO DA GUARDA	RS
114	2950010	-29:24:00	-50:35:00	920	jun/47	2	87	CAPOA DOS COXOS	RS	164	3052009	-30:53:00	-52:14:00	50	abr/44	2	87	PASSO SAO JOSE	RS
115	2950011	-29:16:00	-50:29:00	910	abr/41	2	87	CERRITO	RS	165	3052010	-30:55:00	-52:57:00	80	out/55	2	87	PORTO TARUMA	RS
116	2950019	-29:05:00	-50:37:00	900	mar/47	2	86	LAJEADO GRANDE	RS	166	3052011	-30:25:09	-52:04:24	300	nov/69	1	87	QUITERIA	RS
117	2950023	-29:21:00	-50:30:00	840	nov/60	2	87	PONTE SANTA CRUZ	RS	167	3052012	-30:37:19	-52:48:35	400	jul/76	1	87	SERRA DOS PEDROSAS	RS
118	2950026	-29:25:00	-50:54:00	660	ago/44	2	87	RENANIA	RS	168	3053007	-30:48:00	-53:54:00	300	ago/44	2	87	LAVRAS DO SUL	RS
119	2950027	-29:23:00	-50:11:00	40	jul/60	2	87	RIO DOS CARVALHOS	RS	169	3053015	-30:52:00	-53:07:00	340	jul/44	2	87	SANTANA DA BOA VISTA	RS
120	2950028	-29:40:00	-50:34:00	50	jan/61	2	87	ROLANTE	RS	170	3054002	-30:58:41	-54:40:33	120	jan/34	1	76	DOM PEDRITO	RS
121	2950033	-29:04:00	-50:58:00	880	set/44	2	86	SECA	RS	171	3054016	-30:30:52	-54:46:13	100	jun/76	1	76	GRANJA UMBU	RS
122	2950034	-29:22:00	-50:11:00	470	out/60	2	87	SERRA DO PINTO	RS	172	3055003	-30:41:44	-55:50:28	300	mai/76	1	76	FAZENDA ENCERRA	RS
123	2950038	-29:34:20	-50:03:23	40	jan/74	1	87	TERRA DE AREIA	RS	173	3055004	-30:01:46	-55:05:32	90	out/76	1	76	SAICA	RS
124	2950041	-29:19:00	-50:25:00	950	jan/48	2	87	VARZEA SAO JOAO	RS	174	3055005	-30:31:59	-55:07:36	200	set/76	1	76	SANTA RITA	RS
125	2950048	-29:03:00	-50:26:00	920	jun/76	2	86	FAXINAL DOS PELUCIOS	RS	175	3056004	-30:01:05	-56:49:14	80	mai/76	1	77	FAZENDA JUNCO	RS
126	2951005	-29:35:00	-51:22:00	50	jan/63	2	87	CAI	RS	176	3057002	-30:12:50	-57:33:07	40	jul/76	1	77	BARRA DO QUARAI	RS
127	2951010	-29:14:03	-51:51:15	60	abr/43	1	86	ENCANTADO	RS	177	3151002	-31:07:49	-51:47:19	5	nov/76	1	87	PACHECA	RS
128	2951013	-29:13:00	-51:19:00	750	abr/61	2	87	FARROUPILHA	RS	178	3151003	-31:22:10	-51:59:12	2	jan/43	1	87	SAO LOURENCO DO SUL	RS
129	2951017	-29:08:00	-51:22:00	500	nov/62	2	86	JANSEN	RS	179	3152002	-31:16:52	-52:04:45	120	jan/66	1	87	BOQUEIRAO	RS
130	2951019	-29:10:00	-51:38:00	550	abr/61	2	86	MONTE BELO	RS	180	3152003	-31:23:28	-52:41:50	400	jan/43	1	88	CANGUCU	RS
131	2951022	-29:20:05	-51:11:19	80	abr/43	1	87	NOVA PALMIRA	RS	181	3152005	-31:39:15	-52:47:48	250	set/76	1	88	VILA FREIRE	RS
132	2951023	-29:22:00	-51:06:00	550	abr/61	2	87	NOVA PETROPOLIS	RS	182	3152008	-31:40:07	-52:10:49	3	jan/66	1	88	GRANJA SAO PEDRO	RS
133	2951024	-29:49:02	-51:22:40	15	fev/70	1	87	PORTO GARIBALDI	RS	183	3152011	-31:00:01	-52:02:59	40	abr/44	1	87	PASSO DO MENDONCA	RS
134	2951026	-29:26:00	-51:30:00	486	dez/62	2	87	SAO SALVADOR	RS	184	3152013	-31:52:46	-52:48:31	27	jan/77	1	88	PEDRO OSORIO	RS
135	2951027	-29:21:59	-51:22:16	140	jan/70	1	87	SAO VENDELINO	RS	185	3152016	-31:34:23	-52:27:43	40	ago/64	1	88	PONTE CORDEIRO DE FARIAS	RS
136	2951028	-29:49:16	-51:29:43	20	set/64	1	87	SAPUCAIA DO SUL	RS	186	3153003	-31:13:00	-53:51:00	368	fev/56	2	87	PARAISO	RS
137	2951036	-29:43:00	-51:08:00	50	abr/61	2	86	VILA SCHARLAU	RS	187	3153004	-31:44:11	-53:03:06	200	set/76	1	88	FERRARIA	RS
138	2952003	-29:43:20	-52:53:38	80	dez/65	1	85	BOTUCARAI	RS	188	3153006	-31:08:01	-53:03:16	120	jan/77	1	87	PASSO DA CAPELA	RS
139	2953007	-29:11:00	-53:00:00	440	mai/69	2	85	COLONINHA	RS	189	3153007	-31:44:10	-53:35:20	380	jan/66	1	88	PEDRAS ALTAS	RS
140	2953008	-29:37:39	-53:21:12	25	abr/43	1	85	DONA FRANCISCA *	RS	190	3153008	-31:34:39	-53:22:37	440	out/65	1	88	PINHEIRO MACHADO	RS
141	2953030	-29:05:08	-53:49:19	469	jul/76	1	75	TUPANCIRETA	RS	191	3153017	-31:18:51	-53:29:52	420	set/76	1	87	TORRINHAS	RS
142	2954001	-29:52:40	-54:49:25	100	abr/43	1	76	CACEQUI	RS	192	3154003	-31:01:43	-54:11:04	390	set/76	1	76	TORQUATO SEVERO	RS
143	2954004	-29:21:57	-54:44:05	300	jul/59	1	76	ERNESTO ALVES	RS	193	3155001	-31:16:30	-55:02:16	240	ago/76	1	76	TRES VENDAS	RS
144	2954005	-29:21:32	-54:30:03	300	ago/66	1	76	FURNAS DO SEGREDO	RS	194	3252005	-32:00:20	-52:39:10	20	jan/66	1	88	GRANJA CEL PEDRO OSORIO	RS
145	2954007	-29:29:25	-54:41:20	100	out/57	1	76	JAGUARI	RS	195	3252006	-32:20:58	-52:32:25	4	jul/64	1	88	GRANJA CERRITO	RS
146	2954019	-29:51:07	-54:04:03	408	mai/76	1	76	QUEVEDOS	RS	196	3252008	-32:24:16	-52:33:21	12	set/65	1	88	GRANJA SANTA MARIA	RS
147	2954020	-29:11:12	-54:51:21	420	ago/76	1	76	SANTIAGO	RS	197	3253001	-32:14:11	-53:05:16	3	dez/65	1	88	ARROIO GRANDE	RS
148	2955002	-29:11:46	-55:28:28	100	jun/59	1	76	CACHOEIRA SANTA CECILIA	RS	198	3253003	-32:57:09	-53:07:08	3	jan/66	1	88	GRANJA OSORIO	RS
149	2955006	-29:27:32	-55:17:25	100	set/69	1	76	PONTE DO MIRACATU	RS	199	3253004	-32:01:40	-53:23:55	260	jan/66	1	88	HERVAL	RS
150	2955007	-29:02:51	-55:09:04	380	set/69	1	76	UNISTALDA	RS	200	3353007	-33:31:57	-53:20:58	24	out/12	4	88	SANTA VITORIA DO PALMAR	RS

\* PCD - INPE \*\* Operadora = 1 - CPRM, 2 - CEEE, 3 - EPAGRI, 4 - INMET