

# XI SIMPÓSIO DE RECURSOS HÍDRICOS DO NORDESTE

## REGIONALIZAÇÃO DA PRECIPITAÇÃO MENSAL UTILIZANDO ALGORITMOS DE AGRUPAMENTO HÍBRIDOS

Wagner Alessandro Pansera <sup>1</sup>; Benedito Martins Gomes <sup>2</sup> & Manoel Moises Ferreira de Queiroz <sup>3</sup>

**RESUMO** – Estimativas confiáveis de precipitação são vitais para o sucesso no planejamento hídrico de uma região. Para esta tarefa existem modelos probabilísticos que possibilitam estimativas de ocorrências na frequência e intensidade de chuvas, porém necessitam de dados no local alvo. Ocorre que uma rede de monitoramento hidrológico não cobre todos os locais de interesse, portanto há necessidade de uma ferramenta que transporte essa informação do local monitorado para o local de interesse. Isso pode ser efetuado pela metodologia de regionalização de Hosking e Wallis (1997) que permite estimar quantis e avaliar a homogeneidade de uma região, bem como identificar estações discordantes. Para encontrar regiões potencialmente semelhantes este trabalho utilizou técnicas de agrupamento multivariadas e submeteu as medidas de heterogeneidade e discordância. Os melhores resultados foram obtidos com a metodologia híbrida entre k-médias e Ward. Além disso, testou-se a distribuição gama com dois e três parâmetros, obtendo-se resultados satisfatórios.

**ABSTRACT**– Reliable estimates of precipitation are vital to success of regional water planning. For this task there are probabilistic models that provide estimates rainfall frequency and intensity, but they need data at the target site. However, hydrological monitoring network does not cover all sites, so there is need tools which carries the information from the monitored site to target site. This can be accomplished by Hosking and Wallis (1997) regionalization methodology, which estimates quantiles and evaluate regional homogeneity and identify discordant stations. To find potentially similar regions this study used multivariate clustering techniques and applied both measures of heterogeneity and discordancy. The best results were obtained with hybrid methodology between hierarchical Ward and k-means. In addition, it tested the Gamma distribution with two or three parameters, yielding satisfactory results.

**Palavras-Chave** – medida de heterogeneidade, medida de discordância, momentos-L.

### INTRODUÇÃO

Uma estimativa de precipitação deve obter o máximo de informações a partir dos dados, ser robusta no que diz respeito ao modelo de distribuição e outliers e ser consistente em diferentes locais. Análises de frequência local e regional são comumente utilizadas para as estimativas de precipitação de projeto. A análise de frequência local é comumente utilizada quando há dados históricos suficientes disponíveis nos locais de interesse. Porém, normalmente os registros são

---

1) Afiliação: Universidade Estadual do Oeste do Paraná, Rua Universitária - Jardim Universitário, Cascavel/PR, (45) 32203175, (45) 33247922, wapansera@yahoo.com.br

2) Afiliação: Universidade Estadual do Oeste do Paraná, Rua Universitária - Jardim Universitário, Cascavel/PR, (45) 32203175, (45) 33247922, bmgomes@unioeste.br

3) Afiliação: Universidade Federal de Campina Grande, Rua Jairo Viera Feitosa – Bairro dos Pereiros, Pombal/PB, (83) 3431-4016, (83) 34314016, ppgsa@ccta.ufcg.edu.br

insuficientes ou não há dados de precipitação nas estações pluviométrica da região. Neste caso a análise de frequência local não dá resultados confiáveis e faz estimativas imprecisas das chuvas de projeto. Devido a este problema, dependendo falta ou insuficiência de dados, análise de frequência regional é preferível à de análise de frequência local. Análise de frequência regional (RFA) é mais confiável e precisa que a análise de frequência local, mesmo com séries de dados curtas e apresentado falhas. Além disso, a transferência do conhecimento dos dados de locais relacionados na bacia hidrográfica é possível na RFA com base em regiões homogêneas que é um requisito básico para estimativas precisas e confiáveis (Saf, 2010).

A RFA envolve o uso de uma coleção de ferramentas multidisciplinares para estimação de valores de precipitação de projeto em locais onde não há estações pluviométricas, ou para melhorar estimativas obtidas das observações da estação de interesse. A informação necessária para estimação indireta da precipitação é normalmente transferida para locais não monitorados a partir de um conjunto de bacias similares através do processo de regionalização. Regionalização é uma ferramenta geográfica usada para identificar unidades espaciais que possuem atributos comuns e a sua separação de unidades que não possuem esses atributos. Se as unidades identificadas abrangem uma área geograficamente contígua chama-se de região, e o processo de criação de regiões é chamado regionalização (Cunderlik e Burn, 2006).

A RFA é comumente empregada utilizando o procedimento “index-flood”. A metodologia “index-flood” é baseada na identificação de grupos homogêneos para os quais os períodos de retorno podem ser expressos como um produto de dois termos. Esses termos são o fator de escala e o fator de crescimento, que descrevem as relações entre a variável adimensionalizada e períodos de retorno (conhecida como curva de crescimento). O fator de escala depende do local enquanto o fator de crescimento supõe-se válido para grupos homogêneos. Este procedimento permite estimar curvas de crescimento mais confiáveis devido a exploração de toda informação disponível no grupo e, ao mesmo tempo, permite delinear regiões homogêneas (Castellarin et al., 2008)

Um importante requisito para análise de frequência regional é a identificação de regiões que são usadas para transferência de informação hidrológica. Neste contexto, uma região, significa um conjunto de bacias hidrográficas, não havendo necessidade de contiguidade geográfica, que podem ser consideradas similares em termos de resposta hidrológica. O objetivo do processo de regionalização pode ser visto como a identificação de agrupamentos de regiões que são suficientemente similares para justificar a combinação e transferência de informação hidrológica de locais dentro da região (Burn, 1997).

Porém a simples delimitação de regiões homogêneas por metodologias de classificação não garante necessariamente que a região seja homogênea. Para verificar a veracidade da suposição de

homogeneidade Hosking e Wallis (1997) desenvolveram dois procedimentos: a medida de discordância e o teste de heterogeneidade.

O procedimento para utilização das técnicas de agrupamentos para delineamento de grupos de regiões homogêneas não é automático e o usuário deve intervir em cada passo fazendo varias escolhas, (1) o usuário deve selecionar as variáveis mais relevantes que serão usadas no calculo das distancias entre estações. Podem ser empregadas varias distancias nesse momento. (2) A escolha da função de ligação tem forte impacto na formação dos grupos. Finalmente, a escolha da distancia de corte na arvore hierárquica deve refletir os objetivos do usuário (ou seja, encontrar o numero ótimo de grupos).

Os objetivos deste estudo foram: (1) Investigar o potencial das metodologias de cluster na formação de regiões homogêneas de estações pluviométricas usando as séries históricas mensais como variável classificatória; (2) Avaliar a qualidade das regiões formadas.

## MATERIAL E MÉTODOS

### Etapas da pesquisa

A Figura 1 mostra em qual etapa foi aplicado cada método utilizado na pesquisa.

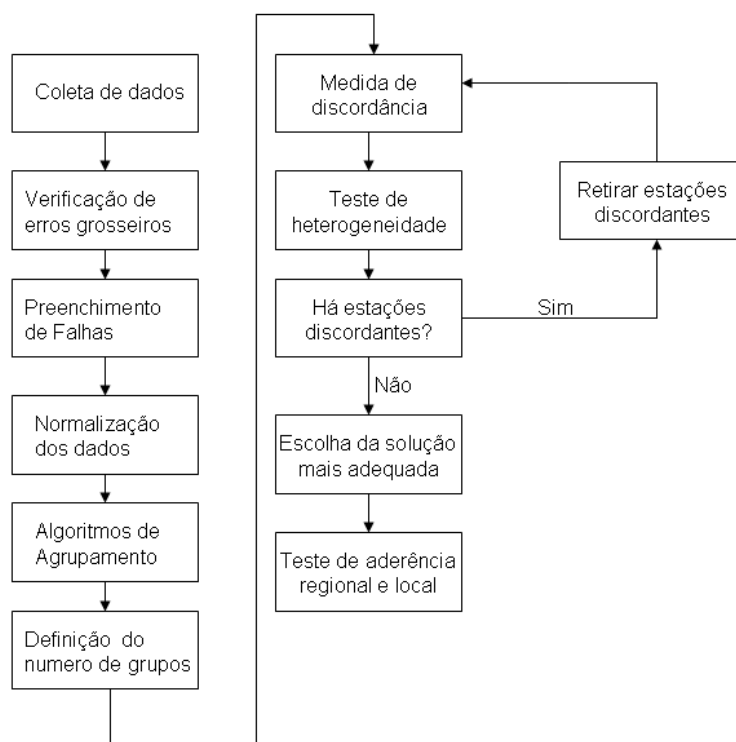


Figura 1 – Fluxograma da metodologia de pesquisa

### Dados utilizados neste estudo

A metodologia foi empregada em 227 estações pluviométricas localizadas no estado do Paraná, Brasil (Figura 2). Os dados foram obtidos no site da Agência Nacional das Águas (ANA)

via sistema de informações HIDROWEB. As estações foram selecionadas seguindo os seguintes critérios: (1) dados referentes ao período de 1976 a 2006, (2) menos de 18 falhas, (3) não possuir uma sequência de falhas superior a quatro consecutivas. Estes critérios ajudaram a obter estações com mesmo tamanho de série e possuírem poucas falhas. Tal procedimento foi adotado para testar as metodologias de agrupamento na obtenção de séries pluviométricas homogêneas que possam ser descritas pela mesma distribuição de probabilidades. Assim, foi criada uma matriz com 227 linhas e 372 colunas, onde as linhas eram as estações pluviométricas e as colunas os totais mensais em ordem cronológica.

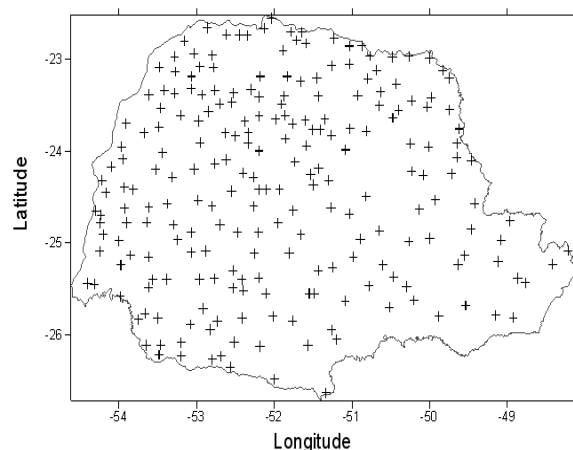


Figura 2 – Distribuição espacial das estações utilizadas neste estudo

### Analise de agrupamentos

Os dados foram transformados para ficarem no intervalo de [0,1]. A transformação é necessária por causa das diferenças de variância, magnitude e importância relativa (Rao e Srinivas, 2006a).

$$x_{ij} = \frac{y_{ij} - y_{j(\min)}}{y_{j(\max)} - y_{j(\min)}} \quad 1 \leq i \leq 227, 1 \leq j \leq 372 \quad (1)$$

Na Eq. (1),  $x_{ij}$  denota o elemento transformado,  $y$  é a matriz dados,  $i$  é a linha e  $j$  é a coluna.

Seja  $x_{ij}$  e  $x_{kj}$  denotem dados transformados das estações  $k$  e  $j$  ( $i = k = 1, 2, \dots, 227$ ) referente a  $j$ -ésima observação ( $j=1, 2, \dots, 372$ ), a distância euclidiana entre as estações foi estimada pela Eq.(2).

$$d(i, k) = \sqrt{\sum_{j=1}^{372} (x_{ij} - x_{kj})^2} \quad (2)$$

Quando a Eq.(2) é usada, é obtida a matriz das distâncias euclidianas entre as  $x_i$  estações, com dimensão 227x227 e diagonal nula.

Com a matriz de distâncias foram aplicadas as metodologias de agrupamentos hierárquicas: simples, completa, média, centroide e Ward. Maiores detalhes podem ser encontrados em Ferreira (2008).

Em seguida, foi utilizado o procedimento k-médias, sendo um algoritmo particional, é baseado no critério do erro quadrado. O objetivo geral é obter uma partição na qual, para um número fixado de grupos, sejam minimizados os erros quadrados (García; González, 2004). Nesta metodologia, os centroides iniciais do algoritmo k-médias são obtidos pelas técnicas hierárquicas de agrupamento. O agrupamento de duas etapas evita decisões subjetivas, e uma única solução pode ser obtida através de iterações e classificações para reduzir a duvida sobre os resultados (Cheng e Liao, 2009).

### Validação dos agrupamentos

O índice de Davies-Bouldin é baseado na ideia de que uma boa partição é aquela para a qual as medidas de separação entre classes e a densidade das classes apresentam valores altos. O índice é uma função da razão entre a soma da dispersão intra-classes e a separação entre-classes

$$DB = \frac{1}{n} \sum_{i=1}^n \max_{i \neq j} \left\{ \frac{S_n(Q_i) + S_n(Q_j)}{S(Q_i, Q_j)} \right\} \quad (3)$$

Na Eq.(3),  $n$  é o número de grupos,  $S_n$  é a distancia média de todos os objetos do grupo em relação ao seu centro,  $S(Q_i, Q_j)$  é a distância entre centros dos grupos. Em consequência disso, o índice de Davies-Bouldin terá valores baixos para bons agrupamentos, denotando grupos compactos e bem separados.

O índice Dunn baseia-se na ideia de identificar conjuntos de grupos que são compactos e bem separados

$$D = \min_{1 \leq i \leq K} \left\{ \min_{1 \leq i \leq K, j \neq i} \left\{ \frac{\delta(C_i, C_j)}{\max_{1 \leq k \leq K} \Delta(C_k)} \right\} \right\} \quad (4)$$

Na Eq.(4),  $\delta(C_i, C_j)$  representa a distância entre os agrupamentos  $C_i$  e  $C_j$ , calculada pela Eq.(5);  $\Delta(C_k)$  representa o diâmetro do grupo  $C_k$ , dado pela Eq.(6). O valor de  $k$  para os quais  $D$  é maximizado é tomado como o número ideal de agrupamentos.

$$\delta(C_i, C_j) = \max_{y_i \in C_i, y_j \in C_j} \{d(y_i, y_j)\} \quad (5)$$

$$\Delta(C_k) = \max_{y_i, y_j \in C_k} \{d(y_i, y_j)\} \quad (6)$$

Em que:  $d(y_i, y_j)$  é a distancia euclidiana entre os objetos  $y_i$  e  $y_j$ .

## Regionalização

Análise de frequência regional normalmente envolve três etapas: identificar grupos de estações pluviométricas homogêneas, descrever uma distribuição de probabilidade robusta regional, e aplicação de um método de estimativa regional dentro da região identificada homogênea. O pressuposto fundamental do método é que dados em diferentes locais em uma região homogênea têm a mesma distribuição. Exceto por um parâmetro de escala ou fator de índice (Dalrymple, 1960). Os quantis adimensionais das regiões homogêneas são estimados de acordo

$$P(F)_i = \mu_i p(F) \quad (7)$$

Na Eq.(7),  $p(F)$  quantil regional adimensionalizado de não exceder a probabilidade  $F$ ;  $P(F)$  é estimativa do quantil local  $i$ ;  $\mu_i$  é o fator de adimensionalização, assumindo que seja a média local. A média amostral local é estimada  $\mu = \sum P_i / n$  e os dados adimensionalizados  $p_{ij} = P_{ij} / \mu$ ,  $j=1,2,\dots,n_i$ ,  $i=1,2,\dots,N$ .

Hosking e Wallis (1997) propõem que os parâmetros da curva regional de quantis adimensionais sejam obtidos pela ponderação dos parâmetros locais estimados para cada local  $i$ , pelos respectivos tamanhos das amostras.

$$\hat{\theta}_k^R = \frac{\sum_{i=1}^N n_i \hat{\theta}_k^i}{\sum_{i=1}^N n_i} \quad (8)$$

Em que:  $N$  é o número de estações,  $\hat{\theta}_k^i$  é o momento-L de interesse, e  $n_i$  representa o tamanho da série.

Substituindo estas estimativas em  $p(F)$  produz a estimativa dos quantis regionais

$$p(F) = p(F; \theta_1^R, \dots, \theta_p^R) \quad (9)$$

## Medida de discordância

Hosking e Wallis (1997) sugeriram uma importante e funcional medida de discordância ( $D$ ) baseada na comparação dos quocientes-L amostrais (L-CV, L-skew, and L-kurt) das estações pertencentes à região homogênea. A medida de discordância identifica dados de precipitação que são grosseiramente discordantes com o grupo, no sentido geral para realização da RFA.

$$D = \frac{N_k}{3} (u_i - \bar{u})^T S^{-1} (u_i - \bar{u}) \quad (10)$$

Em que:  $u_i = (t^i, t_3^i, t_4^i)$  é o vetor dos quocientes-L do local  $i$  ( $i = 1, 2, \dots, N$ );  $\bar{u}$  são  $S$  calculados pelas Eq.(11) e (12).

$$\bar{u} = \frac{1}{N} \sum_{i=1}^N u_i \quad (11)$$

$$S = \frac{1}{N-1} \sum_{i=1}^N (u_i - \bar{u})(u_i - \bar{u})^T \quad (12)$$

Em que: N é o numero de estações,  $u_i$  é o vetor de momentos-L,  $\bar{u}$  e S são as matrizes de média e covariância, e T denota matriz transposta.

Hosking e Wallis (1997) forneceram os valores críticos da medida de discordância. Em muitos casos, a discordância pode aparecer devido à variabilidade amostral. Portanto, os dados em todos os locais com altos valores de discordância devem ser cuidadosamente analisados, independentemente da magnitude dos valores de discordância, antes de decidir se são discordantes (Rao e Srinivas, 2006b).

### Medida de heterogeneidade

O grau de homogeneidade hidrológica é muito importante. Hosking e Wallis (1997) sugeriram um teste estatístico como medida de heterogeneidade (H) para definir a homogeneidade regional. Tal medida é baseada nos momentos-L e na teoria que todas as estações da região têm a mesma população de momentos-L, avalia-se a região estudada é homogênea ou não. As regiões homogêneas podem ser analisadas através de um procedimento de simulação Monte Carlo, que é executado pelo ajuste da distribuição Kappa com quatro parâmetros para média regional dos quocientes-L.

No procedimento de simulação de Monte Carlo, 500 pontos de dados são gerados através de simulação numérica. As regiões sintéticas geradas são assumidas homogêneas com as mesmas características da região real, ou seja, mesmo tamanho de registros em cada estação, mesmos valores regionais dos momentos-L observados, e mesmo numero de estações. Após a simulação, a variabilidade das estatísticas-L da região real é comparada com a simulação:

$$V = \left[ \frac{\sum_{i=1}^N n_i (t^i - t^R)^2}{\sum_{i=1}^N n_i} \right]^{0,5} \quad (13)$$

Em que: t é L-CV,  $t^R$  valor regional dos L-CV.

Finalmente Hosking e Wallis (1997) descrevem a medida de heterogeneidade:

$$H = \frac{V - \mu_v}{\sigma_v} \quad (14)$$

Em que:  $\mu_v$  e  $\sigma_v$  são a média e o desvio-padrão de V ao final da simulação.

Os critérios estabelecidos por Hosking e Wallis (1997) para avaliar a heterogeneidade da região são:  $H < 1$  região aceitavelmente homogênea,  $1 \leq H < 2$  região possivelmente heterogênea,  $H \geq 2$  região definitivamente heterogênea.

### Estimação de quantis

A distribuição gama com dois e três parâmetros foi usada para estimar quantis de precipitação mensal, calculadas por:

$$f(x) = \frac{\beta^\alpha x^{\alpha-1} e^{-\beta x}}{\Gamma(\alpha)} \quad (15)$$

$$f(x) = \frac{1}{\beta \Gamma(\alpha)} \left( \frac{x - \gamma}{\beta} \right)^{\alpha-1} \exp\left( -\frac{x - \gamma}{\beta} \right) \quad (16)$$

Em que:  $f(x)$  é função de densidade de probabilidade,  $\beta$  parâmetro de escala,  $\alpha$  parâmetro de forma,  $\gamma$  parâmetro de posição,  $\Gamma(\ )$  função gama calculada pela Eq.(17).

$$\Gamma(\alpha) = \int_0^{\infty} x^{\alpha-1} e^{-x} dx \quad (17)$$

Os parâmetros das distribuições de probabilidade foram estimados usando a teoria dos momentos-L (Hosking e Wallis, 1997). Após isso, é necessário verificar o ajuste das frequências teóricas com a frequência empírica pelo teste de Kolmogorov-Smirnov:

$$D_N = \max_{-\infty < x < \infty} |F_N(x) - F_X(x)| \quad (18)$$

Em que:  $F_N$  é a frequência empírica,  $F_X$  é a frequência teórica estimada pela distribuição gama.

Foram estimados os parâmetros regionais de cada grupo, com os dados de cada estação adimensionalizados. Depois a frequência teórica regional foi confrontada a frequência empírica de cada estação pertencente ao grupo pelo teste de Kolmogorov-Smirnov.

$$D_R = \max \left[ \max_{-\infty < x < \infty} |F_i(x) - F_R(x)| \right] \quad i=1,2,\dots,N \quad (19)$$

Em que:  $F_i$  é a frequência empírica adimensional,  $F_R$  é a frequência teórica regional adimensional,  $N$  é o número de estações do grupo.

### RESULTADOS E DISCUSSÃO

Pode-se observar na Figura 3 que os métodos híbridos centroide, simples, completa e média possuem um comportamento muito semelhante na função objetivo. O método híbrido Ward e o método k-médias possuem os menores valores da função objetivo independentemente do número de grupos, além disso, suas curvas estão sobrepostas, apresentando visível monotocidade. Comparando



k-médias com método híbrido Ward observa-se grande semelhança nos grupos obtidos, bem como na função objetivo. Em contraste, as demais metodologias hierárquicas são profundamente modificadas pelo método k-médias (algoritmo híbrido). Conclusão semelhante aos autores Rao e Srinivas (2006b).

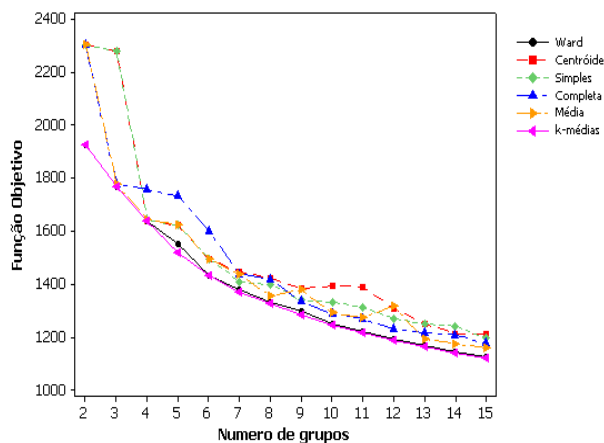


Figura 3 – Gráfico da função objetivo para a metodologia k-médias, bem como suas formas híbridas

A Figura 4 mostra que o índice de Dunn tem sua maior variação quando este sobe de 5 para 6 classes e seu máximo global esta situado entre 10 e 13 classes. O índice de Davies-Boulding tem um mínimo global para duas classes e um mínimo local para seis classes, bem como, seu máximo global é alcançado em 5 classes. Desta forma os índices evidenciam que seis classes é a melhor partição para o método híbrido de Ward.

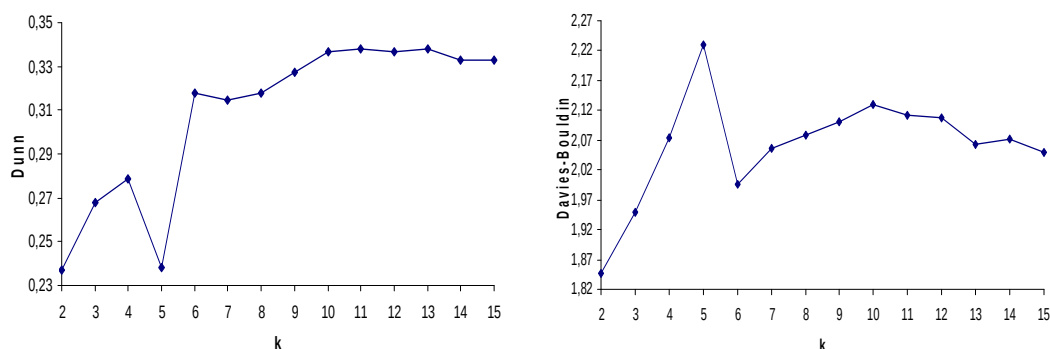


Figura 4 – Índices de Dunn e Davies-Bouldin em função do número de grupos (k) para o método híbrido Ward

### Teste de discordância e heterogeneidade

Observa-se na tabela 1 que para se obter grupos homogêneos foi necessário eliminar 13 estações, representando aproximadamente 6% das estações usadas neste estudo. Inicialmente os grupos 2, 3 e 4 eram classificados como aceitavelmente homogêneos, o grupo 1 como possivelmente homogêneo e o grupo 5 como definitivamente heterogêneo. Após a remoção das estações discordantes os grupos 1 e 5 tornaram-se aceitavelmente homogêneos, houve uma melhora

significativa nas medidas heterogeneidade destes grupos, principalmente do grupo 5. Já para os grupos 2, 3 e 4 não houve melhorias significativas na medida de heterogeneidade, pois estes grupos já eram classificados como aceitavelmente homogêneos e assim permaneceram após a remoção das estações discordantes. O grupo 6 pode ser classificado pela medida de heterogeneidade como possivelmente homogêneo e não foram observadas estações discordantes.

Tabela 1 – Teste de discordância e Heterogeneidade para a solução com seis grupos obtida pelo algoritmo híbrido de Ward

| Grupo | N  | Estação discrepante (ID) | D    | H     | Classificação               | Estação discrepante |
|-------|----|--------------------------|------|-------|-----------------------------|---------------------|
| 1     | 40 | x102                     | 3.07 |       |                             |                     |
|       |    | x103                     | 3.05 | 1.73  | Possivelmente homogênea     | Remover             |
|       |    | x38                      | 3.89 |       |                             |                     |
|       | 37 | -                        | -    | -0.09 | Aceitavelmente homogênea    | Não remover         |
| 2     | 58 | x77                      | 3.12 | 0.50  | Aceitavelmente homogênea    | Remover             |
|       | 57 | -                        | -    | 0.50  | Aceitavelmente homogênea    | Não remover         |
| 3     | 51 | x109                     | 6.06 | 0.32  | Aceitavelmente homogênea    | Remover             |
|       | 50 | x124                     | 3.01 | 0.30  | Aceitavelmente homogênea    | Remover             |
|       | 49 | x107                     | 3.02 | 0.20  | Aceitavelmente homogênea    | Remover             |
|       | 48 | x54                      | 3.27 | -0.31 | Aceitavelmente homogênea    | Remover             |
|       | 47 | -                        | -    | -0.33 | Aceitavelmente homogênea    | Não remover         |
|       |    |                          |      |       |                             |                     |
| 4     | 23 | x212                     | 3.02 | 0.64  | Aceitavelmente homogênea    | Remover             |
|       | 22 | x106                     | 3.01 | -0.40 | Aceitavelmente homogênea    | Remover             |
|       | 21 | -                        | -    | -0,53 | Aceitavelmente homogênea    | Não remover         |
| 5     | 50 | x182                     | 4.16 | 5.35  | Definitivamente heterogênea | Remover             |
|       |    | x196                     | 7.02 |       |                             |                     |
|       | 48 | x195                     | 3.96 | 0.57  | Aceitavelmente homogênea    | Remover             |
|       | 47 | -                        | -    | -0.47 | Aceitavelmente homogênea    | Não remover         |
| 6     | 5  | -                        | -    | 1.21  | Possivelmente homogênea     | Não remover         |

A figura 5 mostra o delineamento das seis regiões homogêneas identificadas pelo método híbrido entre Ward e k-médias.

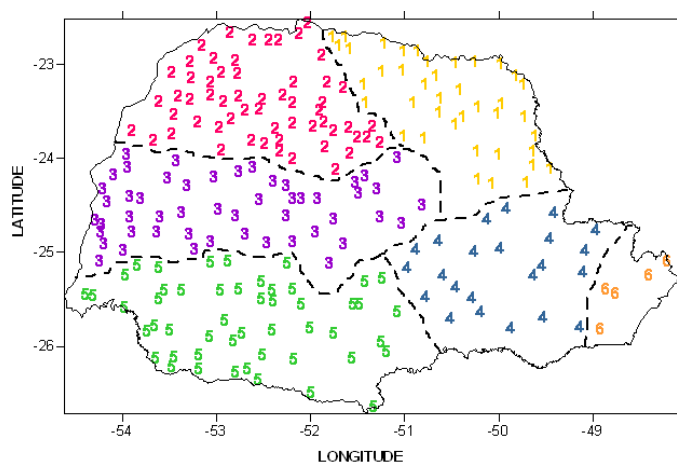


Figura 6 Delineamento das seis regiões homogêneas

Observa-se na figura 6 que a medida de discordância, associada à medida de heterogeneidade, foi eficiente em detectar estações que apresentam comportamento atípico, pois os diagramas de dupla massa mostram que há desvios entre a estação discordante e o grupo ao qual pertence. Saf (2010) aponta que tendências e mudanças das séries de dados de precipitação como importante causa de discordância.

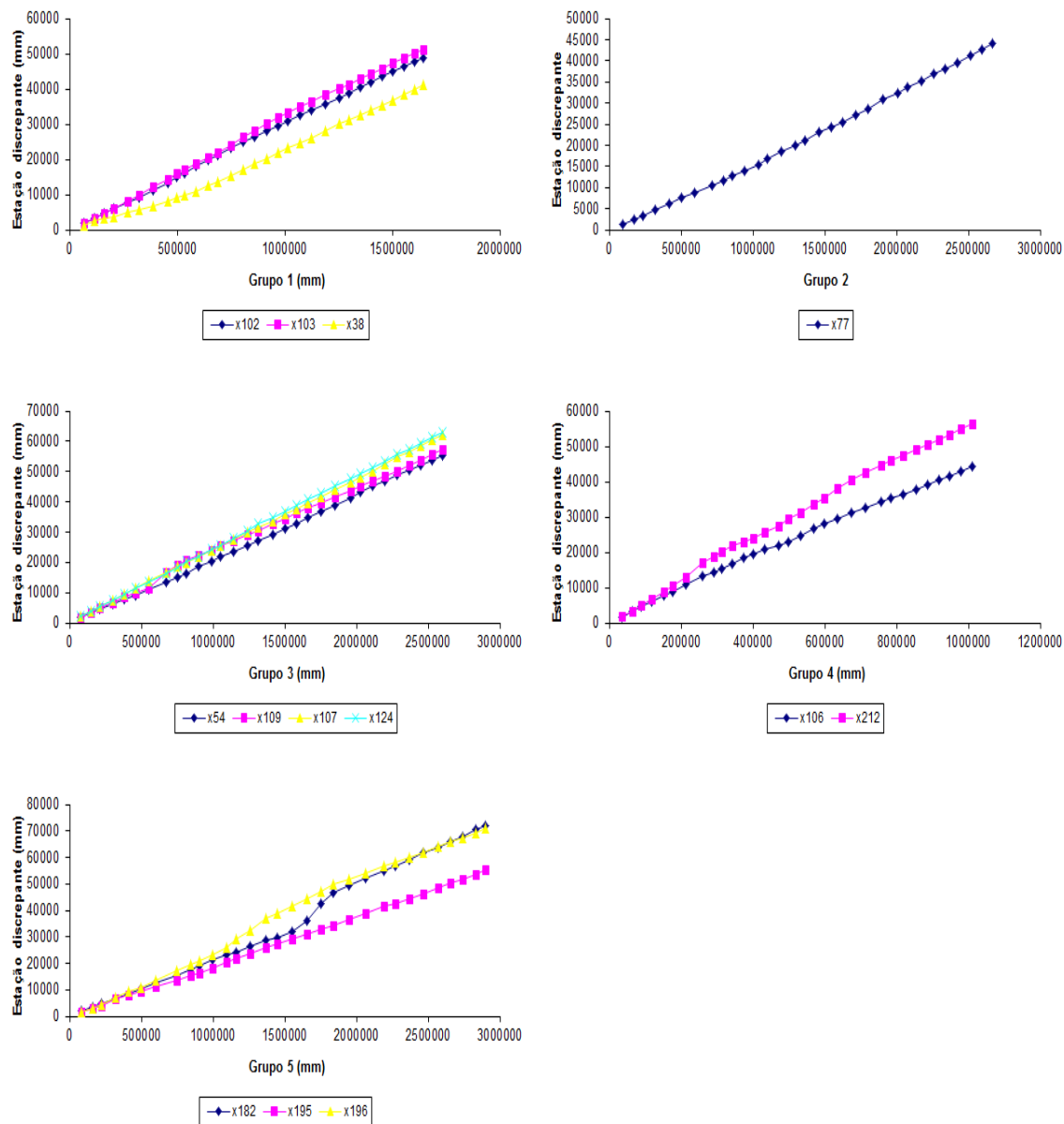


Figura 6 – Digrama de dupla massa das estações discordantes em relação ao grupo pertencente

### Teste de aderência

Na tabela 2 são verificadas apenas cinco situações em que a séries não passou pelo teste de aderência de Kolmogorov-Smirnov ao nível de 10% de significância usando a distribuição gama.

Isso mostra a versatilidade da distribuição gama em representar os dados locais através dos parâmetros regionais, portanto, ela pode ser usada na estimação de quantis regionais adimensionais.

Observa-se na tabela 3 que a distribuição Pearson tipo III se ajustou bem aos dados, pois apenas em três situações não houve aderência. Como a distribuição Pearson tipo III é a gama com três parâmetros, observa-se uma suavização nos valores do teste de aderência.

Pode-se concluir que mesmo para um nível de 10% de significância, um nível restritivo, que as duas distribuições de probabilidade modelam a distribuição de frequência de precipitação mensal no estado do Paraná.

Tabela 1 – Teste de Kolmogorov-Smirnov regional da distribuição gama

| Mês | Grupo  |               |        |               |               |        |
|-----|--------|---------------|--------|---------------|---------------|--------|
|     | 1      | 2             | 3      | 4             | 5             | 6      |
| jan | 0,1591 | 0,1787        | 0,2013 | 0,2019        | 0,1869        | 0,1597 |
| fev | 0,1905 | 0,2014        | 0,1985 | 0,2132        | <b>0,2249</b> | 0,1293 |
| mar | 0,2064 | 0,2048        | 0,1842 | 0,1533        | <b>0,2405</b> | 0,1372 |
| abr | 0,2137 | <b>0,2456</b> | 0,1743 | <b>0,2271</b> | <b>0,2260</b> | 0,1306 |
| mai | 0,1658 | 0,1689        | 0,1410 | 0,1380        | 0,1807        | 0,1255 |
| jun | 0,1793 | 0,1967        | 0,1611 | 0,1563        | 0,1965        | 0,1344 |
| jul | 0,2085 | 0,2002        | 0,1590 | 0,1362        | 0,1841        | 0,1623 |
| ago | 0,1932 | 0,1887        | 0,2058 | 0,1832        | 0,1962        | 0,1848 |
| set | 0,1696 | 0,1701        | 0,2058 | 0,1643        | 0,1684        | 0,1632 |
| out | 0,1675 | 0,1902        | 0,1954 | 0,155         | 0,1864        | 0,1739 |
| nov | 0,1612 | 0,1721        | 0,1784 | 0,2028        | 0,1917        | 0,1597 |
| dez | 0,1797 | 0,1782        | 0,1734 | 0,1767        | 0,1902        | 0,1820 |

$KS_{(10\%,31)}=0,214$

Tabela 1 – Teste de Kolmogorov-Smirnov regional da distribuição Pearson tipo III

| Mês | Grupo  |               |        |        |               |        |
|-----|--------|---------------|--------|--------|---------------|--------|
|     | 1      | 2             | 3      | 4      | 5             | 6      |
| jan | 0,1570 | 0,1751        | 0,1788 | 0,1875 | 0,1518        | 0,1595 |
| fev | 0,1499 | 0,1680        | 0,2033 | 0,1798 | 0,2078        | 0,1221 |
| mar | 0,2062 | 0,1985        | 0,1820 | 0,1552 | <b>0,2444</b> | 0,1386 |
| abr | 0,1617 | <b>0,2238</b> | 0,1586 | 0,1967 | <b>0,2207</b> | 0,1161 |
| mai | 0,1827 | 0,1628        | 0,1649 | 0,1382 | 0,1762        | 0,1277 |
| jun | 0,1560 | 0,1706        | 0,1456 | 0,1472 | 0,1694        | 0,1346 |
| jul | 0,2067 | 0,1755        | 0,1583 | 0,1398 | 0,1954        | 0,1408 |
| ago | 0,1434 | 0,1663        | 0,1568 | 0,1360 | 0,1628        | 0,1807 |
| set | 0,1489 | 0,1792        | 0,1749 | 0,1456 | 0,1684        | 0,1775 |
| out | 0,1639 | 0,1825        | 0,1836 | 0,1489 | 0,1786        | 0,1805 |
| nov | 0,1629 | 0,1815        | 0,1673 | 0,1920 | 0,1942        | 0,1602 |
| dez | 0,1729 | 0,1754        | 0,1731 | 0,1850 | 0,1893        | 0,1434 |

$KS_{(10\%,31)}=0,214$

## CONCLUSÃO

A metodologia híbrida entre os algoritmos k-médias e Ward foi capaz de agrupar estações em regiões homogêneas, sendo demonstrado pelas medidas de discordância e heterogeneidade.

Para se obtenção de uma região homogênea devem-se retirar as estações discordantes.

Obtém-se bom ajuste da distribuição gama com dois ou três parâmetros, possibilitando estimativas confiáveis para o planejamento hidrológico.

#### **BIBLIOGRAFIA**

- BURN, D.H. Catchment similarity for regional flood frequency analysis using seasonality measures. *Journal of Hydrology* 202, 212–230, 1997.
- CASTELLARIN, A.; BURN, D.H.; BRATH, A. Homogeneity testing: How homogeneous do heterogeneous cross-correlated regions seem?. *Journal of Hydrology* 360, 67-76, 2008.
- CHENG, C.L.; LIAO, M.C. Regional rainfall level zoning for rainwater harvesting systems in northern Taiwan. *Resources, Conservation and Recycling* 53, 421-428, 2009.
- CUNDERLIK, J.M.; BURN, D.H. Site-focused nonparametric test of regional homogeneity based on flood regime. *Journal of Hydrology* 318, 301-315, 2006.
- Ferreira, D. F. Estatística multivariada. 1ed, Lavras: Ed. UFLA, 2008.
- GARCIA, H. L.; GONZALEZ, I. M. Self-organizing map and clustering for wastewater treatment monitoring. *Engineering Applications of Artificial Intelligence* 17, 215-225, 2004.
- HOSKING, J.R.M., WALLIS, J.R. Regional frequency analysis – an approach based on L-moments. Cambridge University Press, New York, p. 224, 1997.
- RAO, A. R.; SRINIVAS, V.V. Regionalization of watersheds by fuzzy cluster analysis. *Journal of Hydrology* 318, 57-79, 2006(a).
- RAO, A. R.; SRINIVAS, V.V. Regionalization of watersheds by hybrid-cluster analysis. *Journal of Hydrology* 318, 37-56, 2006(b).
- SAF, B. Assessment of the effects of discordant sites on regional flood frequency analysis. *Journal of Hydrology* 380, 362-375, 2010.