XXI SIMPÓSIO BRASILEIRO DE RECURSOS HÍDRICOS
Segurança Hídrica e Desenvolvimento Sustentável: desafios do conhecimento e da gestão
De 22 a 27 de novembro de 2015, Brasília – DF

ANÁLISE DE CLUSTER DA PRECIPITAÇÃO NA BACIA DO RIO IGUAÇU, REGIÃO SUL DO BRASIL

Patrícia Kazue Uda[1]& Ana Clara Lazzari Franco[2] & Gabriella Queen[3] & Nadia Bernardi Bonumá[4] & Masato Kobiyama[5]*

Resumo:O objetivo deste trabalho foi identificar regiões de precipitação homogênea na área da bacia do rio Iguaçu e proximidades por meio de técnicas de agrupamento de dados. Os dados utilizados neste estudo compreendem 36 anos de medidas de precipitação diária, de 86 estações pluviométricas, acumulados anualmente, do período de 1977 a 2012. Com base nos dados de precipitação e com o auxílio do agrupamento hierárquico seguido pelo algoritmo *k-means*, foram definidos 2 grupos de estações. O comportamento médio dos dois grupos seguema mesma tendência ao longo dos anos, mas a precipitação média anual do grupo 1 é em torno de 20% superior a do grupo 2.Para o grupo 1, a precipitaçção média foi de 1966 mm.ano^{-1}, já para o grupo 2, foi de 1572 mm.ano^{-1}. Para os dois grupos o ano mais chuvoso foi o de 1983 (3075 mm.ano^{-1} para o grupo 1, e 2243 mm.ano^{-1} para o grupo 2), correspondente a um ano sob influência do fenômeno El Niño.

Palavras-chave: Análise de agrupamento. Comportamento espacial da precipitação. Bacia do rio Iguaçu.

CLUSTER ANALYSIS OF PRECIPITATION IN IGUAÇU RIVER BASIN, SOUTHERN BRAZIL

Abstract: The objective of this study was to identify homogeneous rainfall areas within the Iguaçu river basin through data clustering techniques. The data used in this study cover 36 years of daily rainfall measurements from 86 precipitation-gauge stations, accumulated to annual scale, from 1977 to 2012. The hierarchical cluster method, followed by the k-means algorithm was applied to annual rainfall data, which showed the existence of 2 homogeneous precipitation regions in the study area. The mean behavior of both groups is similar, but the group 1 mean annual precipitation is about 20% greater than the group 2 mean annual precipitation. The mean values of the annual precipitation of the group 1 and 2 were 1966 mm.year^{-1} and 1,572 mm.year^{-1}, respectively. The rainiest year was 1983 for both groups (3,075 mm.year^{-1} for group 1 and 2,243 mm.year^{-1} for group 2), and it was influenced by El Niño phenomena.

Keywords: Cluster analysis. Spatial precipitation behavior. Iguaçu river basin.

[1] Doutoranda do Programa de Pós-Graduação em Engenharia Ambiental (PPGEA) da Universidade Federal de Santa Catarina (UFSC), Florianópolis/SC, e-mail: patricia.kazue@posgrad.ufsc.br
[2] Mestranda do PPGEA, UFSC, e-mail: ana.franco@posgrad.ufsc.br
[3]Acadêmica do curso de Engenharia Sanitária e Ambiental da UFSC, e-mail: gsaqueen@gmail.com
[4]Professora Adjunta do Departamento de Engenharia Sanitária e Ambiental da UFSC, e-mail: nadia.bonuma@ufsc.br
[5]Professor Titular do Instituto de Pesquisas Hidráulicas (IPH) da Universidade Federal do Rio Grande do Sul (UFRGS),e-mail: masato.kobiyama@ufrgs.br

INTRODUÇÃO

O estudo da precipitação tem papel importante no gerenciamento dos recursos hídricos e a compreensão de seu comportamento espaço-temporal traz importantes contribuições nesse sentido. A análise do comportamento espacial da precipitação pode ser ferramenta útil na organização e planejamento territorial, uma vez que apresenta as tendências espaciais da área de estudo.

Dentre as técnicas que podem ser utilizadas para a análise do comportamento espacial da precipitação, existem as análises estatísticas de agrupamento. Estas, consistem em técnicas de identificação de grupos homogêneos de dados dentro de um conjunto global, com o objetivo de que os dados dentro de um grupo sejam os mais parecidos possível, e os grupos entre si sejam os mais distintos possível. As técnicas de agrupamento de dados podem ser separadas em hierárquicas ou não-hierárquicas (discriminatórias). As primeiras, fazem agrupamentos sucessivos diminuindo ou aumentando o número de grupos (*clusters*). As não-hierárquicas, por sua vez, dividem os dados em um número pré-determinado de grupos (Hair *et al.,* 2009). As técnicas hierárquicas podem ser divididas quanto ao critério de agrupamento dos *clusters*, dentre os mais comuns podem ser citados os métodos de vizinho mais próximo (*single linkage*), vizinho mais distante (*complete linkage*) e o método de Ward (Rao & Srinivas, 2008). O método *k-means* consiste em um método não-hierárquico, que utiliza como critério de agrupamento o centróide de cada *cluster*. O método consiste em um procedimento onde, definido previamente o número de *clusters*, os elementos são agrupados com os elementos "sementes", que podem ou não ser pré-definidos. Cada elemento é agrupado ao *cluster* cujo centróide esteja mais próximo e cada novo agrupamento o centróide do grupo se modifica e é recalculado. Se ao longo do processo um elemento se torna mais próximo de outro *cluster*, ele é realocado, de modo que se tem um processo iterativo de otimização até que não haja alterações nos centróides dos grupos (Hair *et al.*, 2009).

Na hidrologia, o uso de técnicas de mineração de dados no estudo do comportamento espaço-temporal de variáveis meteorológicas, em especial, a precipitação, pode ser encontrado em diversos estudos. Souza*et al.*(2012) utilizaram técnicas de agrupamento hierárquico para análise da precipitação e temperatura no estado do Mato Grosso do Sul. Boschi *et al.* (2011) utilizaram o algoritmo *k-means* na determinação do comportamento espacial e temporal da precipitação no estado do Rio Grande do Sul, aplicando a metodologia a dois decênios subseqüentes, encontrando uma alteração no volume de precipitação entre os decênios. Conde *et al.*(2012), Silva *et al.*(2010)e Dourado *et al.* (2010) definiram regiões de precipitação homogêneas através do algoritmo *k-means.*

A bacia do rio Iguaçu, localiza-se no sul do Brasil, abrangendo o sul de estado do Paraná e o Planalto Norte de Santa Catarina. É uma importante região econômica, possui elevada concentração de hidrelétricas, além de ser a sub-bacia do rio Paraná com o maior potencial de geração de energia e ter uma importante participação na produção agrícola brasileira. A bacia apresenta grande interesse ecológico, com destaque para o parque nacional do Iguaçu, reconhecido pela UNESCO (SEMA, 2010). Por sua abrangência territorial e importância econômica, o conhecimento do comportamentoda precipitação, em diferentes escalas temporais e espaciais, pode dar subsídio à definição de ações e intervenções no planejamento territorial e de gestão hídrica. O objetivo deste estudo foi compreender o comportamento espacial da precipitação na bacia do rio Iguaçu, por meio de técnicas estatísticas de agrupamento de dados.

ÁREA DE ESTUDO

A bacia do Rio Iguaçu abrange uma área total de cerca de 70.800 km^2. Deste total, 80,4% encontra-se no estado do Paraná, 16,5% em Santa Catarina (totalizando 68.410 km^2) e 3% na Argentina. O rio Iguaçu, formado pelo encontro dos rios Irai e Atuba, é considerado o maior rio totalmente paranaense e possui comprimento de 1.320 km, correndo no sentido leste-oeste e desaguando no rio Paraná. A bacia do rio Iguaçu é a sub-bacia do rio Paraná com o maior potencial de geração de energia (CASTANHARO *et al*., 2007;SEMA, 2010).

Dentre as maiores cidades situadas na bacia, incluem-se Curitiba, Guarapuava e parte de Cascavel. Estima-se que sua população total seja de aproximadamente 4,5 milhões de habitantes. As indústrias predominantes da bacia são a automobilística, papel, frigoríficos, laticínios, entre outras. Possui grande área ocupada pela agropecuária. A região ainda é caracterizada por um grande número de Unidades de Conservação (SEMA, 2010).

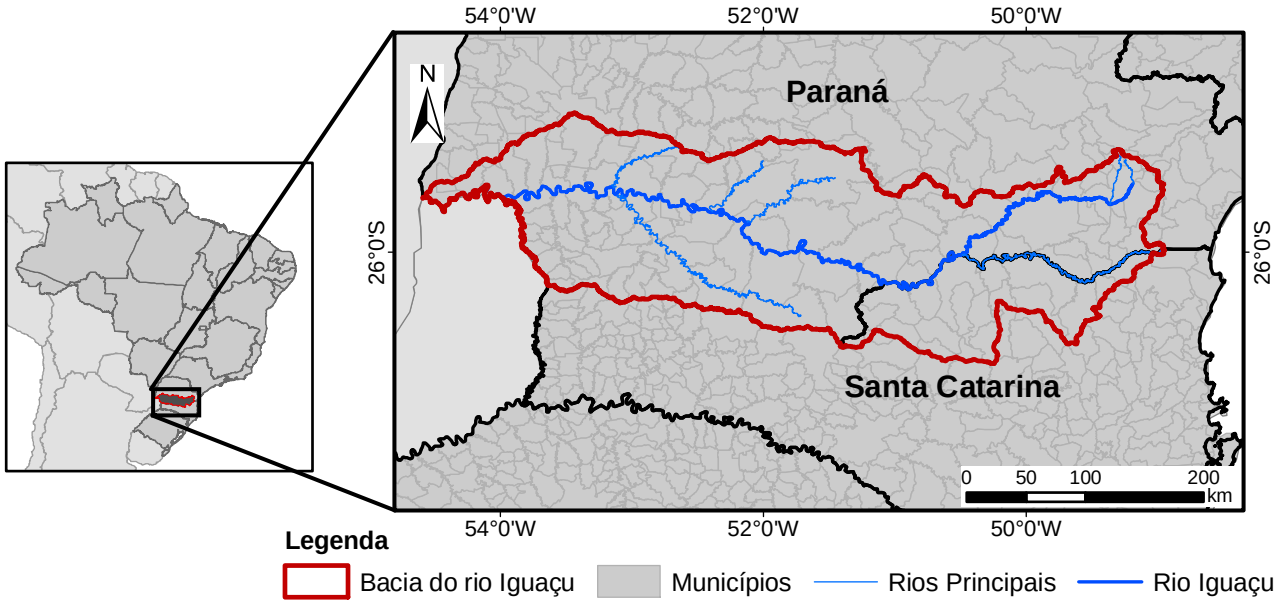

Figura 1: Localização da área de estudo.

Para fins de gestão e planejamento de recursos hídricos, a bacia do Iguaçu é dividida em três unidades hidrográficas: Alto Iguaçu, Médio Iguaçu e Baixo Iguaçu. Estas unidades hidrográficas têm relação com o Primeiro, Segundo e o Terceiro Planalto Paranaense, respectivamente. A altitude da bacia varia de 89 a 1.509 m, sendo os locais de maior altitude localizados na região do médio Iguaçu. Segundo a classificação de Köppen, maior parte da bacia é caracterizada por clima Cfb (Médio, Alto e parte do Baixo Iguaçu), clima temperado, temperatura média do mês mais frio abaixo de 18°C (mesotérmico), média no mês mais quente abaixo de 22°C, verões frescose sem estação seca definida. Em sua extremidade esquerda, a bacia possui clima Cfa, subtropical, temperatura média no mês mais frio inferior a 18°C, geadas pouco freqüentes, temperatura média do mês mais quente acima de 22°C, com verões quentes e sem estação seca definida (IAPAR, 1994).

METODOLOGIA

Os dados pluviométricos da região foram obtidos no site da Agência Nacional de Águas (ANA) pela plataforma online Hidroweb, disponível em <http://hidroweb.ana.gov.br>. As estações pluviométricas foram selecionadas de modo a se obter o maior período de dados diários sem falhas e uma distribuição homogênea dentro da área de estudo, totalizando 86 estações. A série histórica de dados utilizada compreende os anos de 1977 a 2012. Para a análise de *cluster* (agrupamento) os dados foram consistidos, preenchendo-se eventuais falhas, e acumulados anualmente.

Para definição do número de grupos foi aplicado preliminarmente o método de agrupamento hierárquico *complete linkage*, a fim de identificar possíveis *outliers*, e definir o número de *clusters* a serem obtidos com o método de agrupamento discriminatório. O método *complete linkage* calcula a proximidade entre dois grupos como a máxima distância, ou mínimo de similaridade, entre dois pontos em cada grupo. O algoritmo começa determinando a menor distância entre pontos de um mesmo grupo e constrói a matriz de ditsâncias. Assim, os grupos vão se juntando. Se dois elementos são definidos como pertencentes a um mesmo grupo (por exemplo, a e b, a distância entre este grupo e um outro (c)é dada pela Equação 1:

$$D = [(a, b), c] = máx\{D_{a,c}, D_{b,c}\} \quad (1)$$

A medida de distância (proximidade ou dissmilaridade) utlizada neste estudo foi a Distância Euclidiana, que é a distância geométrica (D) entre dois pontos (x_i e y_i) em um plano cartesiano:

$$D(x, y) = \sqrt{\sum_{i=1}(x_i - y_i)^2} \quad (2)$$

O resultado da aplicação do método *complete linkage* foi um dendograma, mostrando os agrupamentos realizados a cada passo. O número de grupos foi determinado de maneira subjetiva através de um corte transversal no dendrograma resultante. Visto que os métodos hierárquicos têm a desvantagem da impossibilidade de realocação dos elementos entre os grupos, ou seja, uma vez agrupados eles não podem ser separados, foi realizado um refinamento dos grupos pelo método *k-means*.

De posse da configuração inicial dos grupos definidos pelo dendograma, foi escolhido um elemento de cada um dos grupos, sendo designados como elementos "semente" na aplicação do método *k-means*. A função objetivo a ser minimizada no método *k-means* é descrita pela Equação 3:

$$E = \sum_{k=1}^{k} \sum_{x_i \in C_k} d(x_i, x_{0k}) \quad (3)$$

onde x_{0k} é o centróide do cluster C_k e $d(x_i, x_{0k})$ é o quadrado da distância euclidiana entre o ponto x_i e o centróide do cluster.

RESULTADOS E DISCUSSÃO

A Figura 2 mostra o dendograma obtido pelo método hierárquico de agrupamento *complete linkage*, com a distância euclidiana como medida de similaridade e como a variável de agrupamento os dados de precipitação anual, dos anos de 1977 a 2012, das 83 estações meteorológicas na área de estudo. O eixo das ordenadas mostra a distância euclidiana em porcentagem, em que, quanto maior seu valor, maior a heterogeneidade dos dados das estações agrupadas. O número de grupos foi determinado de maneira subjetiva através de um corte transversal no dendograma resultante do agrupamento. Considerando o nível de heterogeneidade, bem como a distribuição espacial das

estações meteorológicas, foi adotada a distância euclidiana de 70% para a definição do número de grupos, gerando-se 2 grupos. Nesta etapa, foram identificadas 3 estações pluviométricas com observações atípicas, que foram retiradas antes da última análise de agrupamento pelo método *complete linkage*.

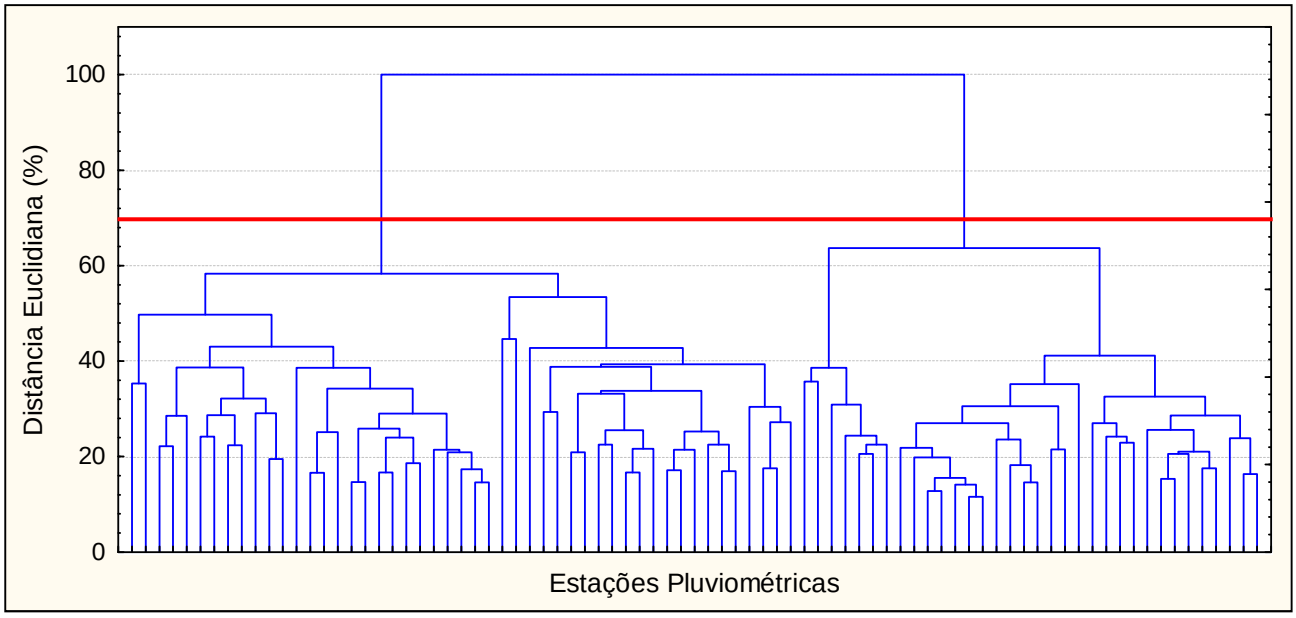

Figura 2: Dendograma obtido pelo método *complete linkage*.

A Figura 3 mostra o refinamento da análise e a disposição espacial dos dois grupos definidos após aplicado o método não-hierárquico *k-means*,com o número de grupos e a semente de cada grupo baseados na análise hierárquica. Observa-se uma concordância da delimitação dos grupos (linha pontilhada) em relação ao relevo da bacia do rio Iguaçu. O grupo 1 abrange todo o Terceiro Planalto Paranaense. Já o grupo 2 abrange o Primeiro e Segundo Planaltos Paranaenses, o que revela uma semelhança entre os climas destas duas regiões.

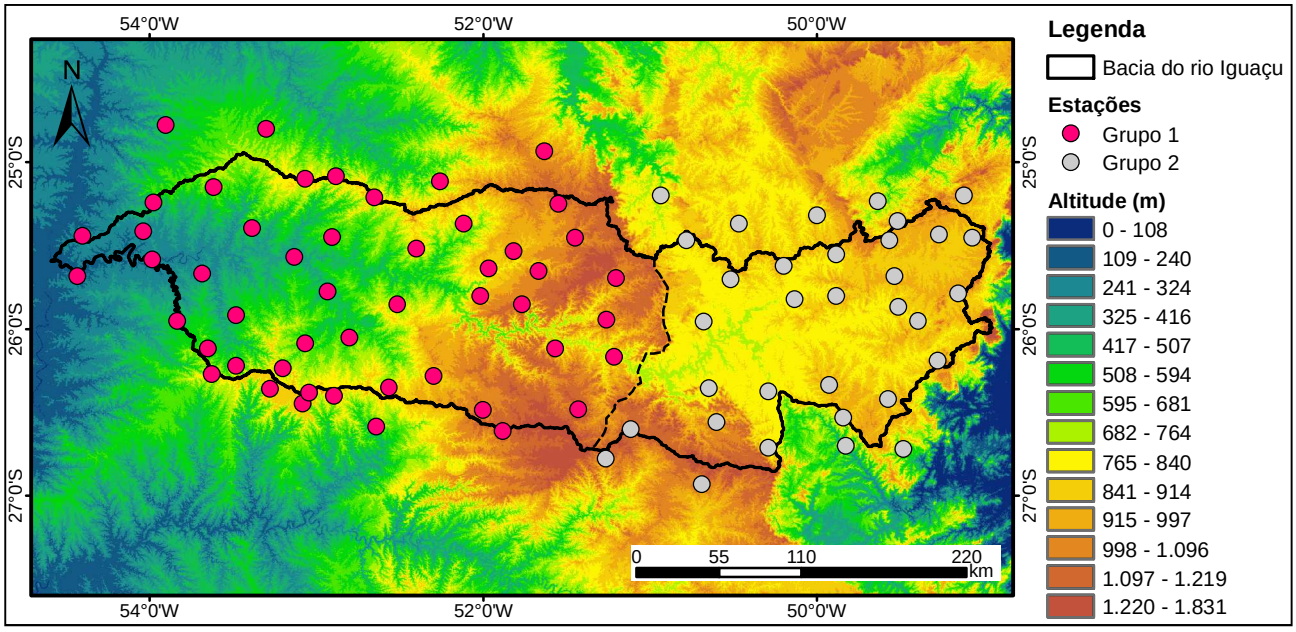

Figura 3: Agrupamento final das estações pluviométricas pelo método *k-means*.

A Figura 4 mostra a precipitação média anual de cada grupo, para o período estudado. Fica evidente que, apesar de o comportamento médio dos dois grupos seguirem uma mesma tendência ao longo dos anos, a precipitação média anual do grupo 1 é sempre superior a do grupo 2. Em média, a precipitação anual do grupo 1 é 20% maior que a do grupo 2. A maior diferença relativa entre os grupo ocorreu no ano de 1981 (35%) e, a menor, no ano de 2010 (4%). Para um melhor entendimento deste processo, seria necessário um estudo dos sistemas climáticos que influenciam a precipitação destas duas regiões.

No grupo 1, a precipitação mínima foi de 1287 mm.ano^{-1}, observada no ano de 1978, a máxima de 3075mm.ano^{-1}foi observada no ano de 1983 e a média é de 1966 mm.ano^{-1}. Para o grupo 2, o ano de 1983 também foi o ano mais chuvoso, com 2243mm.ano^{-1}, o ano mais seco foi em 1985, com 982 mm.ano^{-1} e a média de todos os anos foi de 1572 mm.ano^{-1}. Cabe destacar que especialmente o ano de 1983 foi marcado como um ano de influência forte do fenômeno El Niño, caracterizado como anômalo em 1982/1983, e que acarretou em grande aumento dos níveis pluviométricos nos meses de verão e inverno (Marengo *et al.*, 2006).

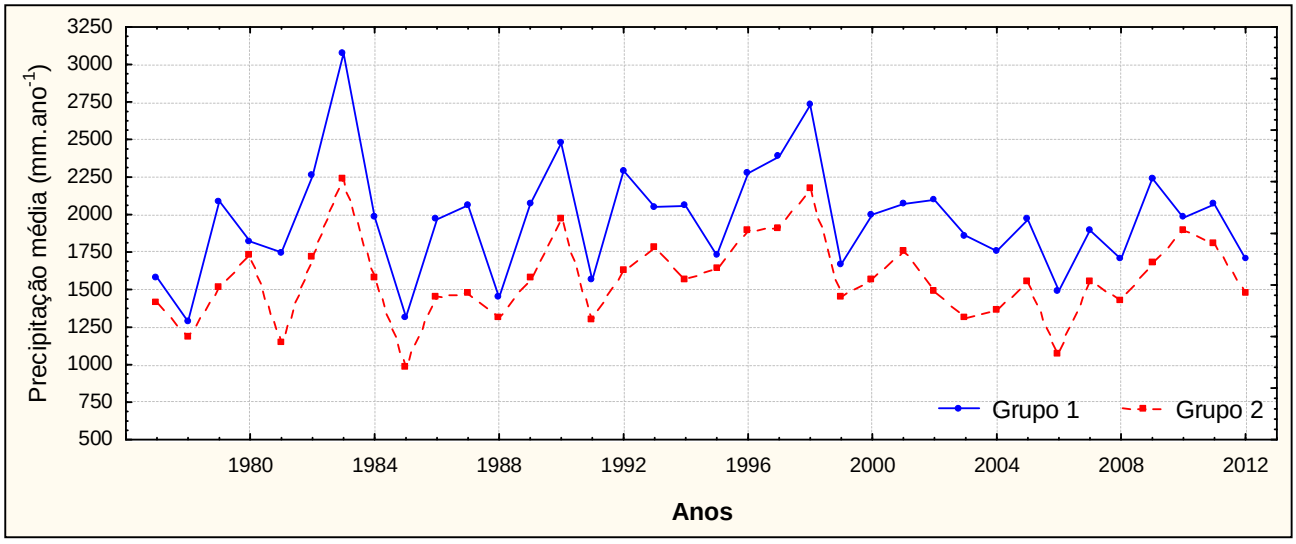

Figura 4: Precipitação média anual para os grupos 1 e 2.

Os grupos encontrados no presente estudo diferem daqueles encontrados por Silva *et al.*(2001) no seu estudo da mesma bacia utilizando de dados de precipitação de 37 estações, para o período de 1967-1997. Os autores identificaram três regiões distintas de precipitação homogênea, uma a leste, uma a oeste e uma a sul da bacia. As diferenças encontradas entre o presente trabalho e Silva *et al.*(2001) podem ser justificadas pela diferença na quantidade e localização das estações utilizadas, e também pelos diferentes períodos estudados.

CONCLUSÕES

As técnicas de agrupamento hierárquico e discriminatório apresentaram bons resultados ao serem aplicadas conjuntamente naidentificação de regiões de precipitação homogêneas na área de estudo. Foram identificadas 2 regiões de precipitação homogênea na região, distribuídas de modo coerente com a topografia da região.

A precipitação média anual do grupo 1 é sempre superior a do grupo 2, sendo em torno de 20%. A precipitação média anual do grupo 1 é de 1966 mm.ano^{-1}, e de 1572 mm.ano^{-1} para o grupo 2. Para os dois grupos, o ano de 1983 é apontado como o mais chuvoso.

Apesar de os resultados apontarem comportamentos distintos da precipitação anual dentro da bcia do rio Iguaçu, faz-se necessário um estudo futuro, com testes de significância, confirmando ou não a existência dessas áreas heterogêneas.

AGRADECIMENTOS

Os autores agradecem ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) e à Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) pelo fomento desta pesquisa. Em adição, agradecem ao Programa de Pós Graduação em Engenharia Ambiental da Universidade Federal de Santa Catarina pelo fomento, estrutura física e oportunidade de divulgação dos resultados desta pesquisa.

REFERÊNCIAS

BOSCHI, R. S.; OLIVEIRA, S. R. D. M.; ASSAD, E. Ténicas de Mineração de dados para análise da precipitação pluvial decenal no Rio Grande do Sul. *Engenharia Agrícola*, v. 31, n. 6, p. 1189–1201, 2011.

CASTANHARO, G.; GIBERTONI, R. DE F.C.; MÜLLER, I.I.; ANDRIOLO, M.V.; KAVISKI, E.; GUILHON, L.G.F.; ROCHA, V.F. Previsão de Vazões na Bacia do Rio Iguaçu Baseada no Modelo SMAP e com Incorporação de Informações de Precipitação. *Revista Brasileira de Recursos Hídricos*. v. 12, n.3, p. 57-68, Jul/Set 2007.

CONDE, F. C. et al. Identificação de Regiões Homogêneas de Precipitação Acumulada Mensal do Distrito Federal via Análise de Cluster. *Anais* do XVI Congresso Brasileiro de Meteorologia, 2010.

DOURADO, C. DA S.; ÁVILA, A. M. H. DE; OLIVEIRA, S. R. M. Regionalização da precipitação no estado da Bahia por meio de técnicas de mineração de dados. *Anais* do XVII Congresso Brasileiro de Meteorologia, 2012.

HAIR, J. F. et al. *Análise multivariada de dados*. 6. ed. ed.Porto Alegre: Bookman, 2009.

INSTITUTO AGRONÔMICO DO ESTADO DO PARANÁ. *Cartas climáticas do Estado do Paraná*. Londrina, IAPAR, 1994. 49 p. ilust. (IAPAR, Documento, 18).

MARENGO, José A. *Mudanças climáticas globais e seus efeitos sobre a biodiversidade: caracterização do clima atual e definição das alterações climáticas para o território brasileiro ao longo do século XXI.* Brasília: MMA, 2006. 212 p. Série Biodiversidade, v.26.ISBN 85-7738-038-6

NETO, A. S. R. *Análise da Conjuntura Agropecuária - Safra 2010/11 - Agrometeorologia*. Estado do Paraná Secretaria da agricultura e do abastecimento Departamento de Economia Rural, 2010.

RAO, A. R.; SRINIVAS, V. V. *Regionalization of watersheds an approach based on cluster analysis.* Dordrecht, Netherlands.: Springer Science+Business Media B.V., 2008.

SEMA - SECRETARIA DE ESTADO DO MEIO AMBIENTE E RECURSOS HÍDRICOS. B*acias Hidrográficas do Paraná.* Curitiba, Governo do Estado do Paraná, 2010.

SILVA, A. P. N. DA et al. Análise de clusters da precipitação do Estado do Piauí. *Anais* do XVI Congresso Brasileiro de Meteorologia, 2010.

SILVA, F. M.; KELLY, V.; TEIXEIRA, J. *Variabilidade interanual da precipitação na bacia do rio Iguaçu*. Maringá, v. 23, n. 6, p. 1439–1444, 2001.

SOUZA, A. DE et al. Análise de Agrupamento da Precipitação e da Temperatura no Mato Grosso do Sul. *Revista ACTA Geográfica*, p. 109–124, 2012.