

USO DA DISTÂNCIA EUCLIDIANA EM ESCORES DE COMPONENTES PRINCIPAIS NA IDENTIFICAÇÃO DE REGIÕES HOMOGÊNEAS DE TEMPERATURA MÁXIMA PARA O RIO GRANDE DO SUL.

Gilberto Barbosa Diniz¹, Solismar Damé Prestes², Julio R. Marques³, Simone V. De Assis⁴,

ABSTRACT – Works that has as a goal to determinate homogeneous regions of meteorological variables has been developed by many countries in researches to do climatological classifications. These Clusters have concentrated on the choice between methods, but insufficient attention has been given to the importance of the method of calculating the distance matrix. The calculus method of these matrixes can have an important effect on the clustering result. In this work, got climatological regions of Rio Grande do Sul using an Euclidean distance as a function of clustering using, like a cluster variable, scores of principal components in maximum medium temperature datas and compared their results

INTRODUÇÃO

Trabalhos objetivando determinar regiões homogêneas de elementos meteorológicos têm sido desenvolvidos em vários países em pesquisas visando classificações climatológicas (Richman e Lamb, 1985; Yu e Chang, 2001); no Brasil temos, dentre outros, os estudos para climatologia efetuados por Diniz (2002, 2003), Marques (2005). O método da análise de agrupamento consiste em determinar o nível de similaridade ou dissimilaridade entre indivíduos aplicando uma função de agrupamento a um conjunto de dados de uma determinada variável.

Conforme EVERITT (1974), classificação pode ser entendida como um processo para localizar entidades em classes inicialmente indefinidas, de modo que indivíduos da mesma classe sejam similares, entre eles. Essas classes de indivíduos similares serão os agrupamentos. Isso é feito através da técnica estatística multivariada chamada Análise de Agrupamento (AA). Nos processos de agrupamento envolvem a escolha de uma variável de agrupamento, uma função e uma técnica de agrupamento.

Segundo Mimmack et al. (2001), recomendações para métodos de agrupamentos têm se concentrado sobre a escolha entre métodos hierárquicos e não hierárquicos, e entre as várias opções de ligações; mas insuficiente atenção tem sido dada sobre a importância do método de cálculo da matriz de distância. Dado que todos os agrupamentos produzidos estão baseados, de alguma forma, nas distâncias medidas, o método de cálculo destas matrizes pode ter um importante efeito sobre o resultado do agrupamento. Neste trabalho objetiva-se determinar regiões climáticas do Rio Grande do Sul com a utilização da distância euclidiana como função de agrupamento usando como variável de agrupamento, escores de componentes principais em dados de temperatura média máxima de estações da rede meteorológica do Estado e comparar os resultados desses agrupamentos com outros obtidos com os dados brutos da mesma variável.

MATERIAL E MÉTODOS

Foram usados dados mensais para os meses de janeiro e julho, da temperatura máxima de 40 estações meteorológicas do Rio Grande do Sul, para o período de 1913 a 2002, com séries completas, cujas falhas existentes nos dados foram preenchidas utilizando técnica apropriada, pertencentes ao Instituto Nacional de Meteorologia – INMET (8º Distrito de Meteorologia – Porto Alegre) do Ministério da Agricultura e Abastecimento (MAPA) e da Fundação Estadual de Pesquisas Agropecuária (FEPAGRO), da Secretaria de Ciências e Tecnologia do RS. A matriz dos dados originais foi organizada da seguinte forma: (equação 1)

$$\mathbf{X}_{(n \times p)} = \begin{pmatrix} X_{11} & X_{12} & \cdot & \cdot & \cdot & X_{1p} \\ X_{21} & X_{22} & \cdot & \cdot & \cdot & X_{2p} \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ X_{n1} & X_{n2} & \cdot & \cdot & \cdot & X_{np} \end{pmatrix} \quad (1)$$

Nesta matriz $\mathbf{X}_{(n \times p)}$ os “p” vetores colunas indicam os valores da variável nos “p” anos e os “n” vetores linha, as estações meteorológicas, resultando uma matriz $\mathbf{X}_{(40 \times 90)}$

A este conjunto de dados foi aplicado Análise de Componentes Principais (ACP) para obter a matriz de escores padronizados, reduzida aos dois primeiros Componentes Principais (CP) e, sobre essa nova matriz $\mathbf{X}_{(40 \times 2)}$ aplicou-se o método de Agrupamento hierárquico aglomerativo da Ligação Completa, tendo como função de agrupamento a distância euclidiana (equação 2).

$$d_{ij} = \|\mathbf{x}_i - \mathbf{x}_j\| = \left[\sum_{k=1}^P (\mathbf{x}_{i,k} - \mathbf{x}_{j,k})^2 \right]^{1/2} \quad (2)$$

Optou-se por usar o mesmo método de agrupamento e o mesmo número de grupos, para facilitar uma comparação com regiões obtidas com a mesma variável, porém, usando uma matriz de dados brutos, método usado por DINIZ (2002).

Os grupos (regiões homogêneas) foram obtidos de forma subjetiva através de corte transversal no Dendrograma e as regiões no mapa seguiram-se os limites geográficos dos municípios cujas estações meteorológicas dos mesmos tiveram seus dados usados no trabalho.

¹ Dr. Prof. Adj. IV Dep. de Meteorologia, Fac. Met. Universidade Federal de Pelotas, Pelotas – RS, E-mail: gilberto@ufpel.edu.br.

² Meteorologista, Chefe do 8º DISME/INMET, Porto Alegre – RS, E-Mail: solismar@inmet.gov.br

³Dr. Meteorologista do CPPMet. Da Faculdade de Meteorologia da UFPel. E-mail: jmarques_fmnet@ufpel.edu.br.

⁴ Dra Profa. Adj. IV Dep. de Meteorologia, Fac. Met. Universidade Federal de Pelotas, Pelotas – RS, E-mail: assis@ufpel.edu.br

RESULTADOS E DISCUSSÃO

As regiões homogêneas, quatro no total, obtidas com o método proposto, usando os CPs que explicaram 87% e 2,4% da variância (janeiro) e 78,8% e 4,0% da variância (julho), estão mostradas nas figuras 1 e 2.

Nota-se nos grupos obtidos com essa técnica, uma semelhança entre as regiões obtidas tanto do mês de janeiro quanto do mês de julho. E, da mesma forma que MARQUES (2005), não existe uma separação física de uma mesma região homogênea, tão comum nos trabalhos de regionalização de DINIZ (2002). A depressão central (**R4**) aparece como uma região divisora entre as áreas montanhosas, de baixas temperaturas, do nordeste do Estado e planalto médio (**R2**), e a serra do nordeste e extremo sul (**R3**), também de baixas temperaturas influenciadas pela latitude e pela entrada das massas de ar vindas do Uruguai e Argentina. E por fim, A região (**R1**) localizada no norte do Estado no alto e médio vale do Uruguai

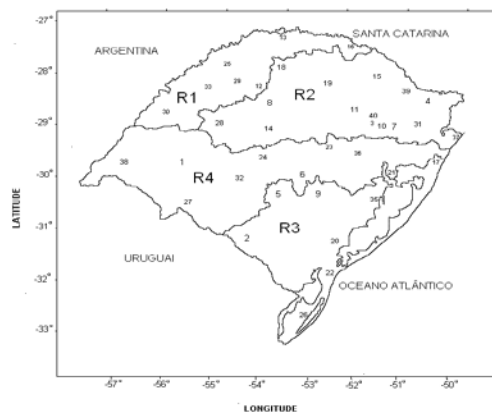


Figura 1. Regiões homogêneas da Temperatura máxima do mês de janeiro. Rio Grande do Sul, 1913 - 2002.

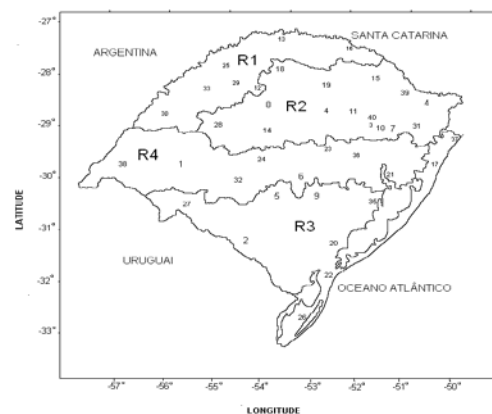


Figura 2. Regiões homogêneas da Temperatura máxima do mês de julho. Rio Grande do Sul, 1913 - 2002.

As Estações meteorológicas constituintes de cada região homogênea do Mês de janeiro são as seguintes: **R1**: {Irai, Marcelino Ramos, Santa Rosa, Santo Ângelo, São Borja e São Luiz Gonzaga}, **R2** {Caxias do Sul, Bento Gonçalves, Bom Jesus, Cruz Alta, Farroupilha, Guaporé, Júlio de Castilho, Lagoa Vermelha, Palmeira das Missões, Passo Fundo, Santiago, São Francisco de Paula, Vacaria e Veranópolis}, **R3** {Bagé, Caçapava do Sul, Encruzilhada do Sul, Osório, Pelotas, Rio Grande,

Santa Vitória do Palmar, e Tapes}, **R4** {Alegrete, Cachoeira do Sul, Ijuí, Porto Alegre, Santa Cruz do Sul, Santa Maria, Soledade, Taquari, Torres e Uruguiana}

As Estações meteorológicas constituintes de cada região homogênea do Mês de julho são as seguintes: **R1**: {Irai, Ijuí, Marcelino Ramos, Santa Rosa, Santo Ângelo, São Borja e São Luiz Gonzaga}, **R2** {Caxias do Sul, Bento Gonçalves, Bom Jesus, Cruz Alta, Farroupilha, Guaporé, Júlio de Castilho, Lagoa Vermelha, Palmeira das Missões, Passo Fundo, Santiago, São Francisco de Paula, Vacaria e Veranópolis}, **R3** {Bagé, Caçapava do Sul, Encruzilhada do Sul, Osório, Pelotas, Rio Grande, Santa Vitória do Palmar, Torres e Tapes}, **R4** {Alegrete, Cachoeira do Sul, Porto Alegre, Santa Cruz do Sul, Santa Maria, Soledade, Taquari, e Uruguiana}.

CONCLUSÕES

Neste trabalho, usando a técnica proposta, as regiões climaticamente homogêneas não se dividem, embora seja verdadeiro o fato de que a aproximação física não garante a similaridade climática. Não se pode com isso, dizer que esse é o melhor método. Uma forma de se verificar mais atentamente o melhor seria a determinação de matrizes de correlação dentro de cada grupo obtidos com diferentes métodos, para ver qual agrupa as estações mais similares.

REFERÊNCIAS

- Diniz, G. B. Preditores visando a obtenção de um modelo de previsão climática de temperaturas máximas e mínimas para a região homogêneas do Rio Grande do Sul: Porto Alegre: UFRGS, 2002. 167 f. Tese (Doutorado em Fitotecnia – Agrometeorologia) – Curso de Pós-Graduação em Fitotecnia, Universidade Federal do Rio Grande do Sul, Porto Alegre, 2002
- Marques, J R. Variabilidade espacial e temporal da precipitação pluvial no Rio Grande do Sul e sua relação com indicadores oceânicos.I: Porto Alegre: UFRGS, 20052. 209 f. Tese (Doutorado em Fitotecnia – Agrometeorologia) – Curso de Pós-Graduação em Fitotecnia, Universidade Federal do Rio Grande do Sul, Porto Alegre, 20052
- Mimmack, G. M. Choice of distance in cluster analysis: Defining Regions. Journal of Climate, Boston, v.14, n.12, p. 2790-2797, 2001
- Richman, M. B.; Lamb, P. J. Climatic Pattern Analysis of Three-and Seven -Day Summer Rainfall in the Central United States: Some Methodological Considerations and a Regionalization. Journal of Climate and Applied Meteorology, Boston, v.24, n.12, p. 1325-1343, 1985
- Yu, T. Y.; Chang, L. F. W. Delineation of air-quality basins utilizing multivariate statistical methods in Taiwan. Atmospheric Environment, New York, v.35, n 18, p.3155-3166, 2001.
- Everitt, B. Cluster Analysis. London: Heinemann Books, 1974. 135p.