
DISTRIBUTION OF ANNUAL RAINFALL IN HOMOGENEOUS REGIONS THROUGH DATA MINING TECHNIQUES IN SOUTHERN BRAZIL

A M H. de Avila⁽¹⁾ · D Ferreira Jr.⁽²⁾ · S R. M. Oliveira⁽³⁾

(1) Center of Research in Meteorological and Climate applied to Agricultural, University of Campinas-CEPAGRI/UNICAMP (avila@cpa.unicamp.br);(2) Institut of Mathematics, Statistics and Scientific Computation, University of Campinas (IMECC - UNICAMP), Campinas, SP; (3) Embrapa Agriculture Informatics at Campinas - Brazil

This study aimed to transform time series of annual data of precipitation of Rio Grande do Sul in homogeneous precipitation regions through clustering techniques (data mining). A hundred of weather stations were selected, from 1981 to 2010, obtained from the National Water Agency (ANA). Initially it was done filling in the missing data through the technique of imputing values by the weighted average of 10 closest weather stations using estimated precipitation values by satellite TRMM (Tropical Rainfall Measuring Mission) and data from surface weather stations available. With the data filled, we used the K-means method to detect homogeneous precipitation regions. Five clusters were found taking into account the measures internal cohesion and external cohesion that shows how homogeneous a sub-region is and how heterogeneous the sub-regions are with each other. The highest rainfall volumes were separated at north and northwest, however the northwest sub-region follows the same trend of the northern region, even with a volume of precipitation a little smaller. Cluster analysis grouped and captured the southern end of the state as a homogeneous region, which was expected. In this region, we note that the distribution of rain over the years tends to be smaller than the other ones. Finally, the east and northeast regions can be considered as intermediate to the others. There is also a similarity of these regions, however the east one is more humid, which was also expected to be a coastal region.

Keywords: times series, precipitation, clustering, climatic risk, missing data

DISTRIBUIÇÃO DA PRECIPITAÇÃO PLUVIAL ANUAL EM REGIÕES HOMOGÊNEAS POR MEIO DE MINERAÇÃO DE DADOS, NO SUL DO BRASIL

Este estudo teve como objetivo transformar as séries históricas de dados anuais de precipitação do Rio Grande do Sul em zonas pluviometricamente homogêneas por meio de técnicas de agrupamento (de Data Mining). Foram selecionadas 100 estações meteorológicas, período de 1981 a 2010, obtidos da Agência Nacional de Águas (ANA). Inicialmente foi feito o preenchimento dos dados faltantes por meio da técnica de imputar valores pela média ponderada de 10 estações meteorológicas mais próximas utilizando valores de precipitação estimados pelo satélite TRMM (Tropical Rainfall Measuring Mission) e dados de estações meteorológicas de superfície disponível. Com os dados preenchidos, foi aplicado o método do K-means para detectar regiões pluviometricamente homogêneas. Foram encontrados cinco clusters levando-se em conta as medidas de coesão interna e de coesão externa que mostra o quanto a sub-região é homogênea e o quanto as sub-regiões são heterogêneas entre si. Os maiores volumes de chuva foram separados ao norte e a noroeste. Sendo que a sub-região noroeste segue a mesma tendência da região norte, porém com um volume de precipitação um pouco menor. A análise de agrupamento captou e agrupou o extremo sul do Estado como uma região homogênea, o que era esperado. Nessa região, nota-se que a distribuição de chuva ao longo dos anos tende a ser menor do que as demais. Por último, as regiões leste e nordeste podem ser consideradas como intermediárias as outras. Há também uma similaridade dessas regiões, contudo a leste é mais úmida, o que também era esperado por ser região litorânea.

Palavras chave: Séries históricas, precipitação, técnica de agrupamento, risco climático, dados faltantes.