

ISSN 0104-1347

Filling in missing rainfall data in the Andes region of Venezuela, based on a cluster analysis approach

Preenchimento de falhas em séries de precipitação pluvial na região dos Andes, Venezuela, baseado na análise de agrupamento

Beatriz Ibet Lozada Garcia¹, Paulo Cesar Sentelhas²,
Luciano Tapia³, Gerd Sparovek⁴

Abstract: Several agrometeorological studies require daily rainfall, mainly those which have as objective to model or simulate water budget, crop development and yield, and occurrence of crop pests and diseases. However, the presence of missing data is a problem that normally occurs, which limits these studies. The origin of this problem varies, but they are more related to improperly devices functioning or lack of technical officers to make observations. Such problems are common in the different institutions which control weather-stations networks in Venezuela. Simple and feasible alternatives to improve the quality of rainfall database are required. We established the hypothesis that daily rainfall data from a weather station can be used to fill in missing data from another surrounding weather station. Data used to test our hypothesis were obtained from 106 weather stations in the Andes region, Venezuela, considering a period of 31 years (1967-1997). The original rainfall database presented 17.3% of missing data (207,534 days). Using a cluster analysis (Ward's method, with Euclidean distance), the proposed method, named as *Closest Station*, reduced the percentage of missing data to 2.5% (29,495 days). The performance of our proposed method was evaluated by mean absolute error (MAE), which ranged from 1.7 to 4.0 mm day⁻¹, and by Willmott agreement index (d), which was 0.57 for daily basis and 0.83 for monthly basis. The contingency analysis showed that our proposed method overestimated rainfall events for daily data, which resulted in a smaller fraction of correct estimates (FC = 0.48) and a larger false alarm ratio (FAR = 0.53), limiting their use. For the other time scales, from 7 to 30 days, FC was greater than 0.88 and FAR smaller than 0.07, which allow the use of this technique for several purposes in agrometeorological studies.

Key-words: *Closest Station* method, precipitation, climatology, database

Resumo: Vários estudos agrometeorológicos exigem o uso de dados diários de chuva, especialmente os que têm como objetivo a modelagem e simulação do balanço hídrico, do crescimento, desenvolvimento e rendimento das culturas, da ocorrência e proliferação de pragas e doenças. Entretanto, a presença de dados faltantes nas séries de dados é um problema que normalmente ocorre, limitando assim tais estudos. As origens de tais problemas são diversas, mas estão principalmente relacionadas ao mau funcionamento dos equipamentos e à falta de observadores. Tais problemas são muito comuns nas redes pluviométricas das mais diversas instituições venezuelanas. Sendo assim, há a necessidade de se desenvolver técnicas simples e factíveis, que possibilitem o preenchimento das falhas existentes nas séries históricas, melhorando a qualidade dos bancos de dados de chuva. Baseado nisso, estabeleceu-se a hipótese de que os dados diários de chuva de uma dada estação podem ser usados para preencher as falhas de uma estação vizinha próxima. Os dados utilizados para testar a hipótese formulada foram obtidos de 106 estações meteorológicas da região dos Andes, Venezuela,

¹ Instituto Nacional de Investigaciones Agrícolas (INIA), Bramón, Táchira, Venezuela. E-mail: blozada@inia.gov.ve

² Departamento de Ciências Exatas – ESALQ, Universidade de São Paulo, Piracicaba, Brasil. E-mail: pcsentel@esalq.usp.br

³ Centro de Informática na Agricultura (CIAGRI) - Universidade de São Paulo, Piracicaba, SP, Brasil.
E-mail: lrtapia@esalq.usp.br

⁴ Departamento de Ciência do Solo – ESALQ, Universidade de São Paulo, Piracicaba, SP, Brasil. E-mail: gerd@esalq.usp.br

