São José dos Campos - SP
LV SBPO
LV Simpósio Brasileiro de Pesquisa Operacional
São José dos Campos, SP - 6 a 9 de novembro de 2023

Managing Uncertainties in Hydrological Event Forecasting with AI: An Approach Based on Optimized Allocation of Rain Gauges.

Roberval Gonçalves Moreira Filho
Universidade Federal de São Paulo
Avenida Cesare Mansueto Giulio Lattes, 1201 - Jardim Santa Ines I, São José dos Campos - SP,
roberval.filho@unifesp.br

Luiz Leduíno de Salles Neto
Universidade Federal de São Paulo
Avenida Cesare Mansueto Giulio Lattes, 1201 - Jardim Santa Ines I, São José dos Campos - SP,
luiz.leduino@unifesp.br

Leonardo Bacelar Lima Santos
Centro Nacional de Monitoramento e Alertas de Desastres Naturais
Estr. Dr. Altino Bondesan 500, São José dos Campos, SP, 12247-016
leonardo.santos@cemaden.gov.br

ABSTRACT

This work proposes a methodology for optimizing rain gauge station allocation in São José dos Campos/SP. The methodology consists of collecting data from existing rain gauge stations, using the inverse distance interpolation technique to estimate values at un-sampled points and grouping the data into groups using the K-Means algorithm. The goal is to maximize the coverage of the rain area and provide more accurate and reliable information about the amount and intensity of rainfall in the region. With this approach, we hope to contribute to improving rainfall monitoring in Brazil, representing an essential step towards reducing the damage caused by socio-environmental disasters associated with rainfall. The proposed methodology resulted in an average improvement of 11.6% in rainfall detection compared to the previous arrangement, making it an effective solution to improve the prevention and mitigation of the impacts of hydrological events.

KEYWORDS. optimized allocation, rain gauge stations, rainfall monitoring.

Computational intelligence, OR in agriculture and environment

http://live-proceedings.pantheonsite.io/p/174855

1. Introdution

Socioenvironmental disasters can have a significant impact on communities and a country's economy. In Brazil, hydrological events are one of the most impactful classes of socioenvironmental disasters. According to [Freitas et al. [2014]], they represent 32.7% of the events that occurred in Brazil between 1991 and 2010 and account for the highest number of deaths between 2011 and 2019, representing over 90% of the cases [Parizzi [2022]].

Rainfall monitoring is an important component in disaster risk management. One way to perform monitoring is with rain gauges. These stations measure the quantity and intensity of rainfall in a specific area and can automatically send the collected data to the cloud through data telemetry. Real-time monitoring of rainfall data is important for predicting and preventing the impacts of hydrological events [Sensors [2021]].

Thus, we aim the following overall objective:

• To propose an optimized allocation of rain gauges to maximize rainfall area coverage.

To achieve this objective, the following specific goals will be pursued:

1. Analyze the data from rain gauges in the city of São José dos Campos/SP to determine the amount of rainfall in each location.

2. Interpolate the data to estimate the amount of rainfall in new locations based on the information from existing rain gauges.

3. Apply a clustering algorithm to the resulting data to identify candidate regions for new rain gauge locations.

In this work, we employed artificial intelligence techniques, such as inverse distance interpolation and K-Means clustering, to manage uncertainties in the prediction of hydrological events. The optimized allocation of rain gauges proposed by our methodology can help reduce uncertainties in the prediction of hydrological events by providing more accurate and reliable information about the quantity and intensity of rainfall in different regions of the city.

2. Methodology

We aim at a propose a methodology for the optimized allocation of rain gauges in the city of São José dos Campos. The methodology consists of collecting data from existing rain gauges, interpolating values at unsampled points using inverse distance weighting (IDW) interpolation technique, and clustering the data into k groups using the K-Means algorithm. The objective is to maximize the coverage of the rainfall area and provide more accurate and reliable information about the quantity and intensity of rainfall in the region.

2.1. Data

The data used in this work were obtained from the Cemaden website [for Monitoring e of Socioenvironmental Disasters [2023]] through their interactive map. Information regarding the rain gauges in São José dos Campos between the years 2018 and 2022 was collected for the months of January, February, and March. The data was provided in spreadsheet format for each month, containing several columns, including the name of each station (*nomeEstacao*), its latitude (*latitude*) and longitude (*longitude*) coordinates, and the measured value (*valorMedida*) representing the rainfall measurement at regular intervals.

Only the relevant columns were selected: *nomeEstacao*, *latitude*, longitude, and valorMedida. The information from the rain gauges in the technology park was grouped into a single station by calculating the average of the sum of all measurements. The result was a dataframe containing each rain gauge station (with a single station representing the combination of all stations in the technology park) with the sum of rainfall at each location.

2.2. Inverse Distance Weighting Interpolation with KDTree

We aim employed the inverse distance weighting (IDW) interpolation technique, [Shepard [1968]], to estimate values at unsampled points. IDW assigns values to unknown points based on the weighted inverse distance from the closest known points. To optimize the search for the nearest neighbors, we used a data structure called KD-tree, [Bentley [1975]], which organizes points in a k-dimensional space. In addition to interpolation at specific points, we developed a function to create new geographical locations and interpolate values at those locations.

We implemented a function that utilizes the DataFrame of sampled points and a parameter to expand the latitude and longitude range of the new locations. We also implemented the *idw_kdtree* function, which uses the DataFrame of sampled points and a parameter for weight calculation. This function constructs a KD-tree and finds the nearest neighbors using the query method of the KD-tree. The interpolated values are obtained through a weighted sum of the values from the nearest sampled locations.

2.3. Clustering with K-Means

We employed the K-Means clustering algorithm to divide the data into k groups. This method is a clustering algorithm that aims to partition n observations into k groups, [MacQueen [1967]].

We implemented the *kmeans_clustering* function, which utilizes the K-Means algorithm to cluster the data in a DataFrame based on the latitude, longitude, and *valorMedida* columns. The group labels and centroids are returned by the function. This function is applied to a DataFrame divided into 11 groups, and the labels are stored in the *kmeans_labels* column, indicating the group to which each location belongs. Additionally, for this research, the centroid locations were considered as a new solution.

3. Results

This work was developed using Python in the Google Colaboratory environment. To evaluate the results, we calculated the sum of rainfall detected at locations with interpolated values within a 2-kilometer radius of any rain gauge and compared it with the sum of rainfall detected within a 2-kilometer radius of any centroid defined by the K-Means algorithm.

When comparing rainfall detection using the current arrangement of rain gauges with the detection using the centroids proposed by the K-Means algorithm within a 2-kilometer radius, we observed that the proposed arrangement showed an improvement in rainfall detection. This suggests that the new arrangement may be more effective in detecting the amount of rainfall compared to the current arrangement of rain gauges.

4. Conclusions

In this work, a methodology was developed and applied to optimize the allocation of rain gauges in the city of São José dos Campos/SP. The data from existing rain gauges were collected and analyzed. Subsequently, interpolation and clustering techniques were used to estimate values at unsampled points and identify candidate regions for new rain gauge locations. The goal was to maximize the coverage of the rainfall area and provide more accurate and reliable information about the quantity and intensity of rainfall in the region.

The results showed that the proposed methodology was effective in improving rainfall detection. The new arrangement of rain gauges proposed by the K-Means algorithm showed an average improvement of 11.6% in rainfall detection compared to the previous arrangement. This suggests that the proposed new arrangement may be more effective in detecting the amount of rainfall compared to the current arrangement of rain gauges.

However, it is important to note that this work has some limitations. For example, we only used data from existing rain gauges and did not consider other sources of information, such as satellite images or radar data. The quality of the data collected by the rain gauges can affect the results. If the collected data is not accurate or reliable, it can affect the effectiveness of the proposed methodology.

Although the results showed a suggestive improvement in rainfall detection, this work is still partially completed, and there is room for future improvements. Further steps will be taken to enhance the results, including testing new algorithms and conducting analyses in other cities.

I am grateful CAPES(Coordenação de Aperfeiçamento de Pessoal de Nível Superior), CNPq(Conselho Nacional de Desenvolvimento Científico e Tecnológico) and AGEVAP(associação pró-gestão das águas da bacia hidrográfica do rio paraíba do sul) for financial support, which has greatly contributed to my work. Their funding has provided access to resources and opportunities for research and professional development. I sincerely appreciate the commitment of these institutions to promote academic growth.

References

Bentley, J. L. (1975). Multidimensional binary search trees used for associative searching. *Communications of the ACM*, 18(9):509–517.

for Monitoring, N. C. e of Socioenvironmental Disasters, A. (2023). Mapa interativo. URL `http://www2.cemaden.gov.br/mapainterativo/`.

Freitas, C. M. d. et al. (2014). Desastres naturais e saúde: uma análise da situação do brasil. *Ciência & Saúde Coletiva*, 19(9):3645–3656. URL `https://doi.org/10.1590/1413-81232014199.00732014`.

MacQueen, J. B. (1967). Some methods for classification and analysis of multivariate observations. In *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability*, volume 1, p. 281–297.

Parizzi, M. G. (2022). Panorama dos desastres climatológicos, hidrológicos, meteorológicos e geológicos durante o período de 2010 â 2019. *Derbyana*, 43.

Sensors, S. (2021). Estação pluviométrica. URL `https://sigmasensors.com.br/estacao-pluviometrica`.

Shepard, D. (1968). A two-dimensional interpolation function for irregularly-spaced data. In *Proceedings of the 1968 23rd ACM national conference*, p. 517–524. ACM.