Spatial and temporal variability of precipitation in a mountainous watershed using weighted interpolation by distance and elevation

Variabilidade espaço-temporal de precipitaões em uma bacia hidrográfica montanhosa utilizando interpolação ponderada por distancia e elevação

David de Andrade Costa[1,2], Yared Bayissa[2], Dhiego da Silva Sales[1], Rennan Mendes de Moraes dos Santos Dias[2,3], Jader Lugon Junior[1], Antônio J. Silva Neto[3], Raghavan Srinivasan[2]

1 Instituto Federal Fluminense, Campos, RJ, Brazil; david.costa@iff.edu.br

2 Department of Ecology and Conservation Biology, Texas A&M University, College Station, USA;

3 Department of Mechanical Engineering and Energy, Universidade do Estado do Rio de Janeiro, Nova Friburgo, RJ, Brazil;

Abstract

The aim of this study is to discretize precipitation through interpolation utilizing the inverse square of distance and elevation as weighting elements to represent the spatial and temporal variation of rainfall. This technique has proven to be highly effective in capturing precipitation variability by integrating distance and elevation data from 15 stations built from a network of 39 rain gauges clustered by subbasins, generating continuous time series from 2006 to 2022. Validation through the double mass method has confirmed the accuracy of these time series, all exhibiting a coefficient of determination greater than $R^2 = 0.99$. This approach has provided a detailed representation of rainfall patterns across the entire region, significantly advancing our understanding of spatial hydrological dynamics.

Keywords: regionalization; time series; data gap filling

Resumo

O objetivo deste estudo é discretizar as precipitações por meio de interpolação, utilizando o inverso do quadrado da distância e da elevação como elementos de ponderação para representar a variação espacial e temporal das chuvas. Esta técnica revelou-se altamente eficaz na representação da variabilidade das precipitações, ao integrar dados de distância e elevação de 15 estações construídas a partir de uma rede de 39 pluviômetros agrupados por sub-bacias para geração de séries temporais contínuas de 2006 a 2022. A validação por meio do método de dupla massa confirmou a precisão das séries temporais construídas, todas apresentando um coeficiente de determinação maior que $R^2 = 0{,}99$. Este enfoque proporcionou uma representação detalhada dos padrões de chuva em toda a região, contribuindo significativamente para o avanço de nossa compreensão da dinâmica hidrológica espacial.

Palavras-chave: *regionalização; séries temporais; preenchimento de falhas*

1. Introduction

Mountainous regions are crucial for natural resources, influencing water dynamics, vegetation distribution, soil erosion, and regional climate, primarily through precipitation and topography interaction. However, the topographic complexity of these areas presents unique challenges in understanding and modeling precipitation, requiring a detailed approach to capture its temporal and spatial variability [1, 2, 3]. One reason for studying precipitation in mountainous regions in detail is its direct influence on the availability of freshwater. Mountains act as large natural reservoirs, capturing and storing atmospheric moisture that turns into precipitation. This distribution varies greatly due to factors such as elevation, slope orientation, wind patterns, and local microclimates, being essential for estimating water availability [4, 5].

Furthermore, detailed analysis of precipitation is crucial for managing natural disaster risks. The rugged topography of mountains makes them susceptible to extreme events such as floods and landslides, triggered by intense precipitation, especially on saturated or unstable soils [6, 7]. This understanding is equally important for biodiversity since vegetation and wildlife depend on water, whose distribution is influenced by precipitation. Thus, the objective of this study is to discretize precipitation through interpolation using the inverse of the square of the distance and elevation as weighting elements to represent the spatial variability of precipitation for the period 2006-2022. Capturing the complexity of precipitation patterns in these areas enhances modeling and forecasting systems, enabling the development of sustainable strategies to promote the resilience of mountain communities and preserve these ecosystems.

2. Methods

2.1. Study area

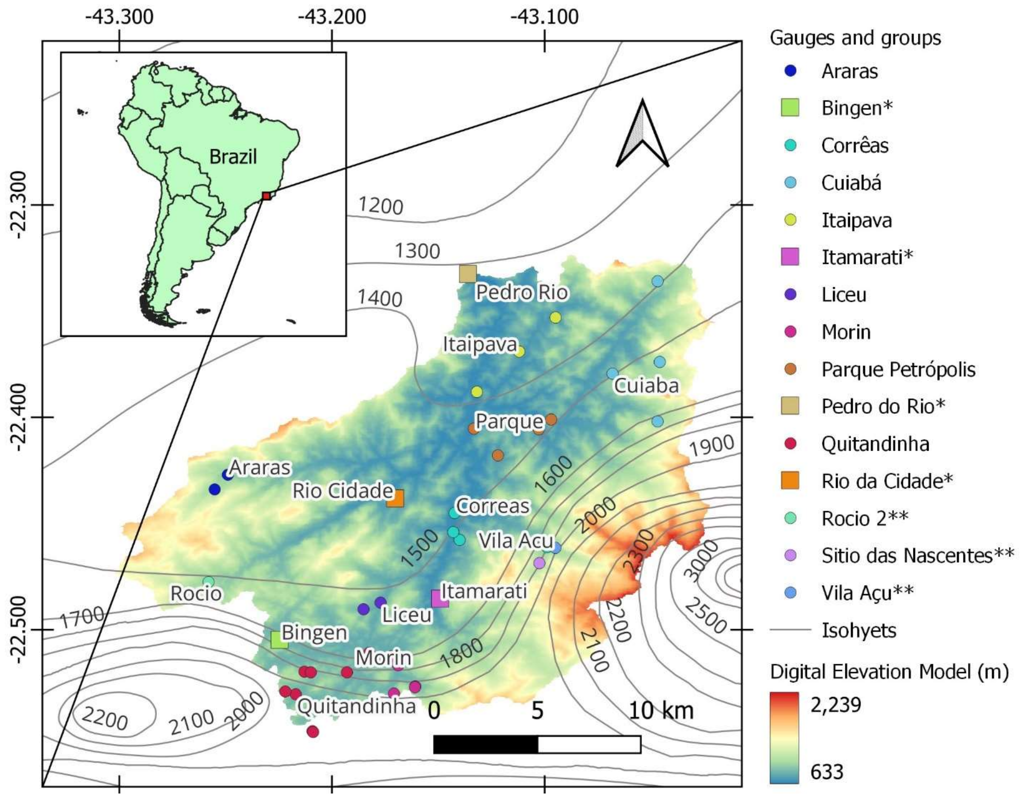

Figure 1: Precipitation gauges in the upper part of the Piabanha river basin. Source: prepared by the authors. * Reference stations not grouped. ** Forest and agricultural stations not grouped.

The upper portion of the Piabanha watershed (Figure 1), located in Brazil, is a study area rich in physical elements that play a crucial role in regulating the local ecosystem. The region is marked by a mountainous topography with deep valleys and steep slopes. This complex topography directly influences the flow of water in the basin, determining drainage patterns. The climate is predominantly tropical of high-altitude, characterized by hot and humid summers and mild and drier winters. However, due to the influence of altitude and topography, there are significant microclimatic variations within the basin. Higher areas tend to be cooler and more humid, while lower areas are warmer and drier. These climatic variations have a direct impact on the distribution of vegetation in the region. Additionally, the water coming from the mountains is an important water supply source for the local population, agriculture and other economic activities in the area.

2.2. Precipitation discretization

There are many institutions monitoring the study region. The precipitation and flow gauges are mainly operated by the Brazilian National Water and Sanitation Agency (ANA). The Geological Survey of Brazil (SGB) operates the stations from EIBEX project (Integrated Studies in Experimental Watersheds) [8]. The State Environmental Institute (INEA) of Rio de Janeiro operates the precipitation and water level gauges related to their Flood Alert program. Additionally, the Brazilian National Center for Monitoring and Alerts of Natural Disasters (CEMADEM) operates precipitation and water level gauges for natural disaster alerts. The processing and consistency of the precipitation time series followed the methodological flow represented in Figure 2. In total, data from 39 pluviometers were used.

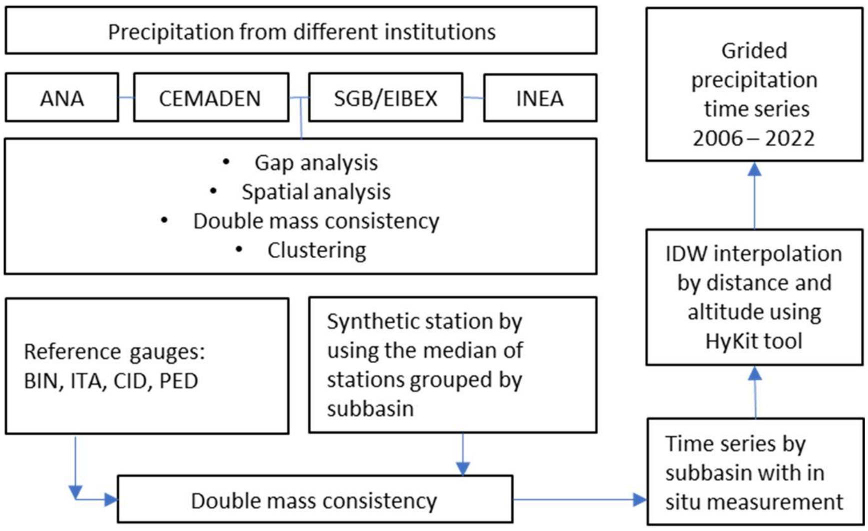

Figure 2: Methodological flow. Source: prepared by the authors.

Among the precipitation stations, four gauges operated by ANA stand out with the shortest periods of failure and were carefully consisted to serve as reference stations as they are situated within a homogeneous climatic region according to the Brazilian pluviometry atlas [9]. In order to check their consistency, the double-mass method was employed since it remains the most used technique for precipitation consistency [10]. According to Searcy & Hardison [11], this technique involves chronologically accumulating precipitation throughout the year and calculating the average accumulation of the stations under evaluation. Consequently, the accumulated precipitation of each station can be graphically plotted against the average accumulated precipitation. A linear regression line is expected to be observed in the dataset if the stations are within a homogeneous region. If deviations indicating a change in slope are detected along the line, the reason for the data inconsistency must be investigated, and a new line should be adjusted for the detected period. If necessary, data adjustment for the inconsistent period should be performed by applying a correction factor based on the ratio of the slopes of the adjusted lines.

Synthetic stations were generated by the median of stations within the same subbasin (Figure 1) to mitigate measurement errors, where stations in the same location may exhibit discrepant readings. This approach also enables the construction of temporal series with fewer gaps, considering that rarely do all stations experience gaps simultaneously. The tool "HyKit: A Tool for Grid-based Interpolation of Hydrological Variables" [12, 13, 14] was employed to interpolating the stations using distance and elevation weighting (Equation 1).

$$\hat{p}_k = W_D \sum_{i=1}^{N} \frac{1}{D} w(d)_i p_i + W_z \sum_{i=1}^{N} \frac{1}{Z} w(z)_i p_i \tag{1}$$

$$w(d) = {}^{1}/_{d^a} \quad \text{for d} > 0 \tag{2}$$

$$w(z) = \begin{cases} {}^{1}/_{z_{min}^{b}} & \text{for } z \leq z_{min} \\ {}^{1}/_{z^{b}} & \text{for } z_{min} < z < z_{max} \\ 0 & \text{for } z > z_{max} \end{cases} \tag{3}$$

Where, $\hat{p}$ is the interpolated precipitation for a grid cell, W_D and W_z are the importance factors for distance and elevations, respectively, p_i is the precipitation value in mm/d of the i^{th} gauge station and *N* is the number of gauges that are used in the interpolation for the current grid cell. Similarly, $w(d)_i$ and $w(z)_i$ are the individual gauge weighting factors for distance and elevation, respectively, and *D* and *Z* are the normalization quantities given by the sum of individual weighting factors $w(d)$ and $w(z)$, respectively, for all the gauges used. The weighting factors $w(d)_i$ and $w(z)_i$ based on inverse of distance and elevation are given by Equation 2-3, where, *d* is the distance in kilometers between the current grid and the gauge station used for interpolation, *z* is the absolute elevation difference in meters between the current grid cell and the gauge station used for interpolation, *a* and *b* are exponent factors for distance and elevation weightings, respectively. The exponents (*a* and *b*) and the weighting factors are dimensionless numbers, z_{min} and z_{max}, expressed in meters, are the minimum and maximum limiting values of elevation differences for computing elevation weightings.

The digital elevation model (DEM) was obtained from NASA [15] and resampled to 90 meters using the resampling tool available in SAGA [16] (System for Automated Geoscientific Analyses). Raster subbasins were created using QGIS software, as the HyKit tool enables the creation of temporal series based on the mean pixel values contained within subbasins. Both the DEM and subbasin files were subsequently converted to ASC files using the translate tool (convert format) in GDAL (Geospatial Data Abstraction Library), also available in QGIS software. The stations were interpolated using the inverse distance squared and elevation method, resulting in a precipitation dataset without gaps for all subbasins. Subsequently, the synthetic stations were validated using the double mass method [11], with reference to previously consisted stations: Bingen, Itamarati, Rio da Cidade, and Pedro do Rio.

3. Results and Discussions

The hydrological stations Pedro do Rio, Bingen, Itamarati, and Rio da Cidade, operated by the National Water Agency (ANA), exhibited a strong Pearson linear correlation (r), ranging from r = 0.78 between Bingen and Pedro do Rio, to r = 0.93 between Bingen and Itamarati, as well as between Rio da Cidade and Pedro do Rio. Precipitation variability was very well represented using the HyKit tool. The method applied both distance and elevation for 15 stations derived from a network of 39 precipitation gauges clustered by subbasins. The stations are situated at distances ranging from 1 to 20 kilometers from each other, with elevations ranging between 700 and 1500 meters above sea level. These stations were individually validated through the double mass method, all exhibiting a coefficient of determination greater than $R^2 = 0.99$ (Figure 3). This approach has provided a detailed depiction of rainfall patterns across the region, significantly enhancing our understanding of hydrological dynamics, as illustrated in Figure 4 and 5. Our results corroborate with Sirisena et al. [13], who employed the HyKit tool as a preprocessing step for modeling streamflow and sediment supply under climate changing conditions, achieving improved discretization of precipitation, thereby enabling a more accurate representation of hydrological modeling.

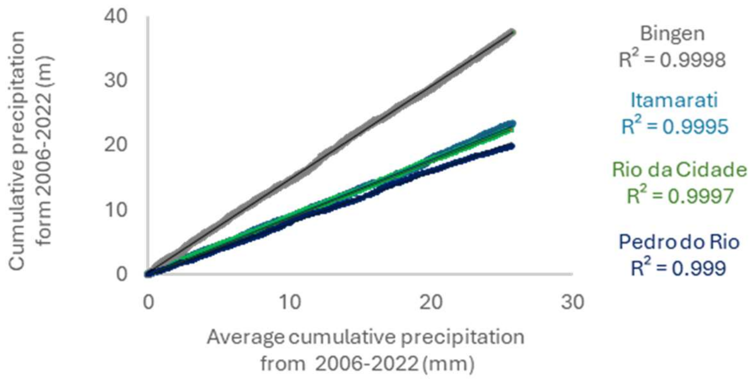

Figure 3: Double mass plot comparing cumulative precipitation among the reference stations: Bingen, Itamarati, Rio da Cidade, and Pedro do Rio. Horizontal axis denotes average cumulative precipitation across all stations, while vertical axis represents cumulative precipitation for individual stations. Source: prepared by the authors.

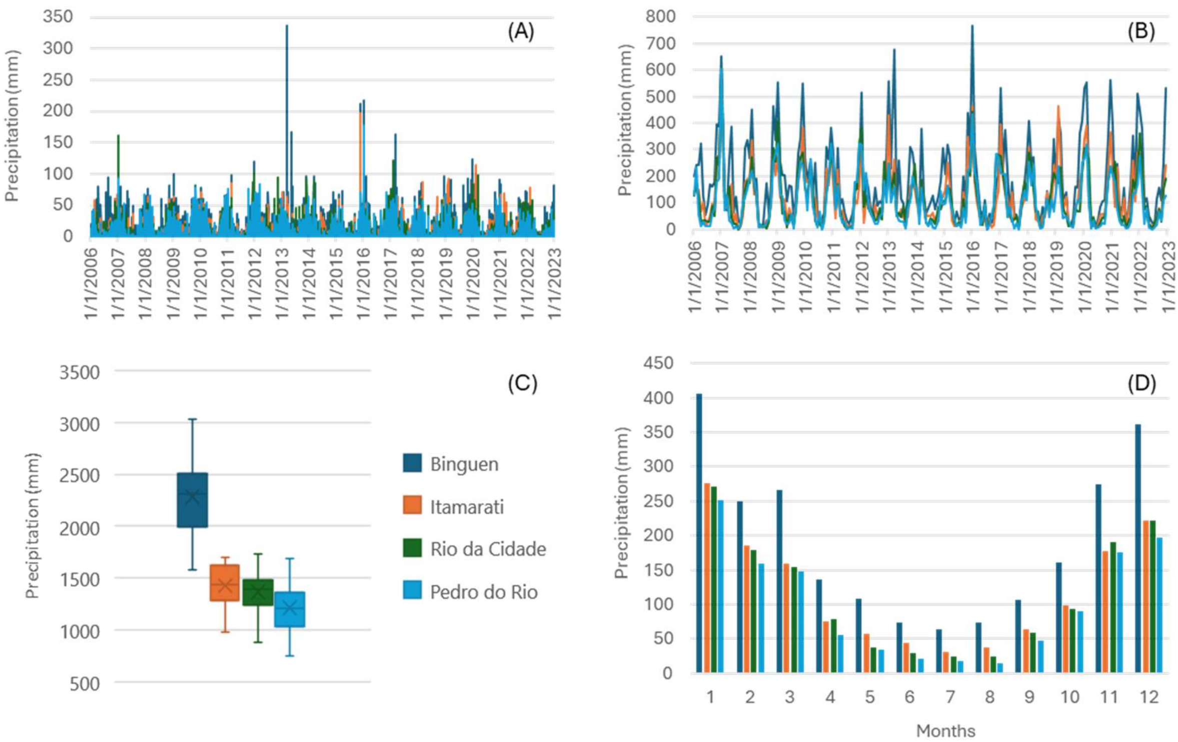

Figure 4: Reference stations. A) Daily precipitation time series; B) Monthly precipitation time series; C) Annual precipitation; and D) Average monthly precipitation. Source: prepared by the authors.

During the consistency process, some extreme values were identified. At the reference station Bingen/ANA, a precipitation of 335.7 mm was recorded on 03/18/2013, while the other reference stations, Itamarati, Rio da Cidade, and Pedro do Rio, recorded 63.7 mm, 36.3 mm, and 7.9 mm, respectively. This suggests widespread precipitation across the basin, with more intensity in the higher regions. To validate these extreme values, data from neighboring stations were also examined. It was found that the Coronel Veiga station, operated by INEA, recorded a precipitation of 347.7 mm on the same date, corroborating the extreme values recorded at both stations. However, on 07/02/2013, Bingen/ANA recorded a precipitation of 188.6 mm, while neighboring stations, Bingen/INEA and Coronel Veiga/INEA, recorded 20.7 mm and 34.7 mm, respectively. In this case, the arithmetic mean of these gauges was used to correct the value of 188.6 mm to 27.7 mm. On 07/02/2013, Bingen/ANA recorded 211.0 mm, consistent with records of 154.6 mm in Alto da Serra/CEMADEN and 198.6 mm in Itamarati/ANA. Additionally, on 01/15/2016, a precipitation of 218.3 mm was recorded in Bingen/INEA, consistent with a record of 136.7 mm in Alto da Serra/CEMADEN. Consequently, the four ANA stations were considered consistent and selected as reference for the other stations in the basin. Daily, monthly, and annual precipitation can be observed in Figure 4.

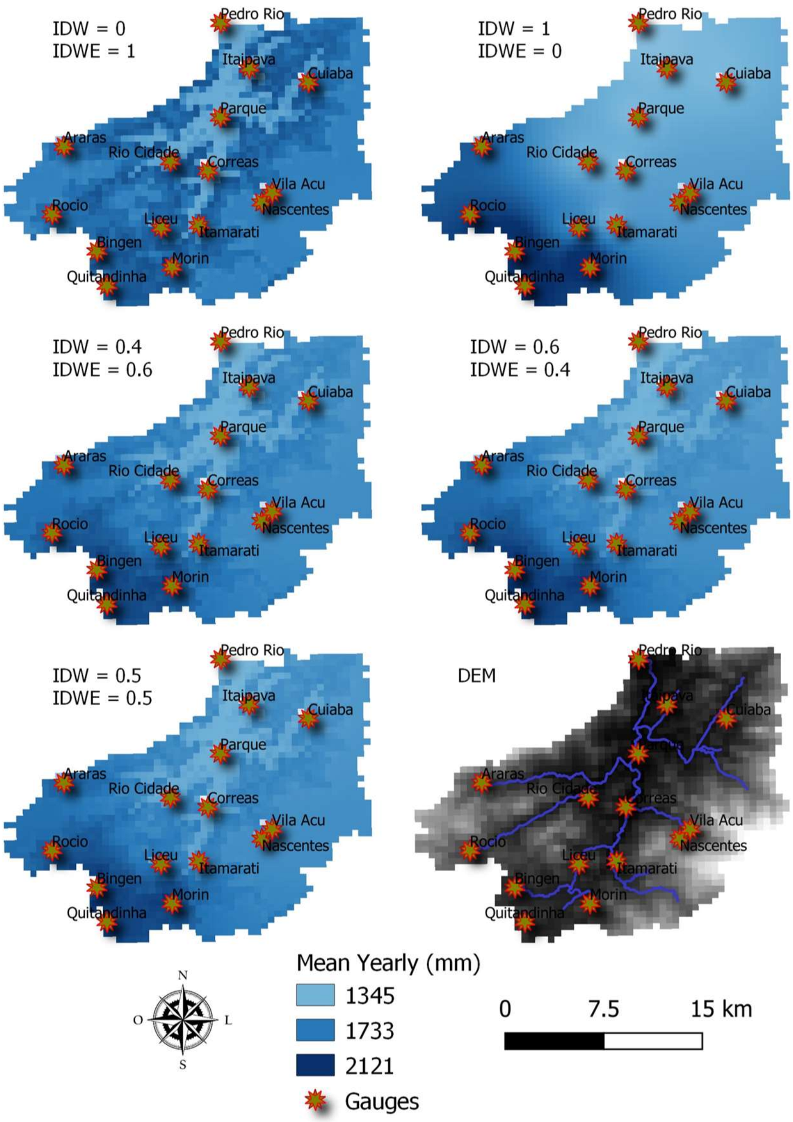

Figure 5: Interpolated annual precipitation using different weights for distance (IDW) and elevation (IDWE). Source: prepared by the authors.

Among 39 available precipitation stations, 32 were utilized to construct 8 synthetic stations (Figure 4) by calculating the median precipitation value from stations within the same subbasin. Additionally, 4 stations were used as reference stations, and 3 stations from experimental basins were used without grouping. This resulted in a total of 15 precipitation time series used as input data for the HyKit tool. For instance, the Quitandinha River subbasin has 7 precipitation gauges, but their data period ranges from 2011 to 2022, with all stations exhibiting significant non-coincident data gaps. Therefore, by adopting the median of the stations, a gap-free series starting from 2011 could be constructed. However, the reference stations have more extensive data series, some dating back to the 1930s, but we limited our analysis from the year 2006 onwards.

The HyKit tool was tested by assigning different weights for interpolation between distance (IDW) and elevation (IDWE) (Figure 5). However, based on literature [8, 9, 10], the configuration that assigns equal weights to distance and elevation was selected. The spatial distribution of precipitation shows higher rainfall in the higher regions of the basin, consistent with the adopted reference stations, indicating a good fit of the interpolation. Similarly, the lower regions of the basin exhibit lower precipitation. Thus, the inclusion of elevation as a weighting factor proves to be important for mountainous regions. This corroborates with literature [17], since the distance weighting method has demonstrated superior performance compared to several other standard methods of regionalizing precipitation.

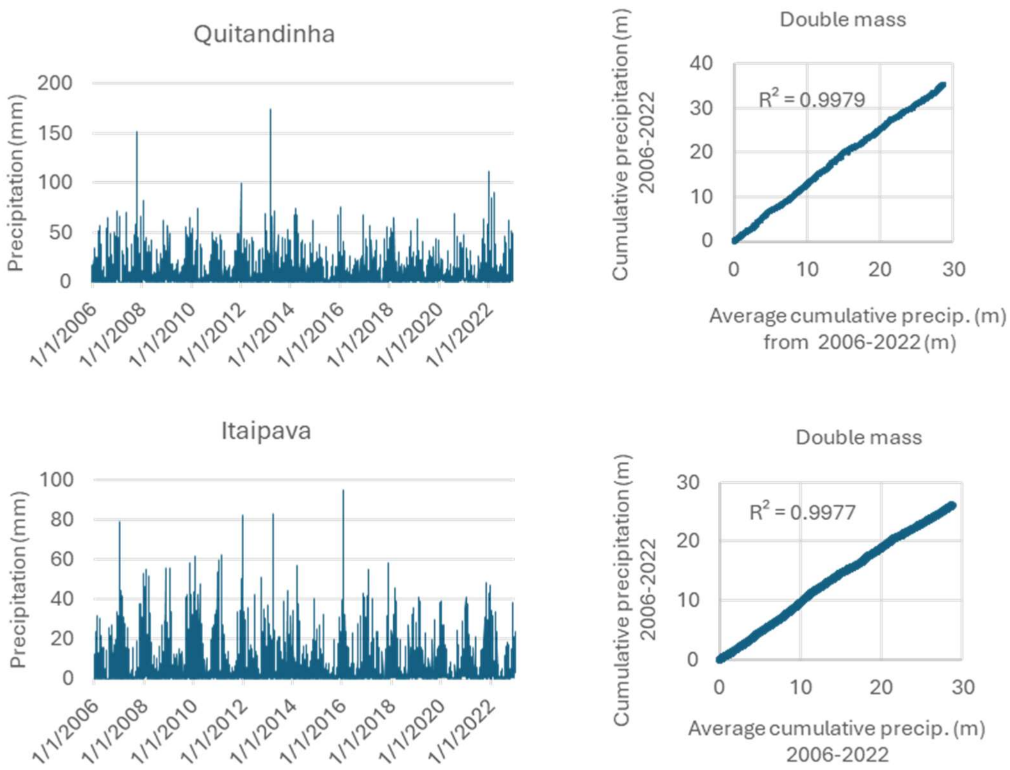

Figure 6: Synthetic stations interpolation by distance and altitude. Source: prepared by the authors.

After the interpolation process, all stations were validated using the double mass method in comparison with the reference stations mean. For instance, the interpolation for the Quitandinha River basin and a subbasin in the Itaipava region (Figure 6) generated a continuous series from 2006 to 2022, and their consistency through the double mass method showed coefficients of determination exceeding $R^2 = 0.99$, indicating an excellent fit. For the generated time series, all precipitation events exceeding 100.0 mm were individually inspected and confirmed based on other stations. The same analysis was conducted for all other stations (Figure 5), which also exhibited coefficients of determination exceeding 0.99. Our study is consistent with the approach of Arikan and Kahya (2019) [11], who utilized double mass curves to assess consistency and apply corrections to their time series data. They employed a minimum of four stations to calculate the average cumulative precipitation, which aligns with our study.

3. Conclusion

The main objective of this study was to discretize precipitation through interpolation, utilizing the inverse square of distance and elevation as weighting elements to represent the spatial variation of rainfall. We applied the HyKit tool to perform the interpolation process, this technique proved highly effective in representing precipitation variability. By incorporating both distance and elevation data from 15 stations built from a network of 39 precipitation gauges clustered by subbasins, this method yielded accurate results. Validation via the double mass method confirmed the precision of these stations, with all exhibiting a coefficient of determination greater than $R^2 = 0.99$. This approach offered a detailed depiction of rainfall patterns across the region, significantly advancing our understanding of spatial hydrological dynamics. In addition, constructing synthetic stations by averaging precipitation data from different gauges within the same subbasin has proven effective in building time series with fewer gaps and greater reliability. Our findings underscore the importance of precipitation discretization, particularly in mountainous regions. Hydrologists and practitioners stand to benefit significantly from this approach to perform precipitation interpolation based on distance and elevation, facilitating more accurate assessments and predictions in complex terrain. Future work could focus on employing alternative interpolation techniques for comparison and performance analysis.

Acknowledgements

The authors acknowledge the financial support provided by the following Brazilian agencies: FAPERJ, Carlos Chagas Filho Foundation for Research Support of the State of Rio de Janeiro; CNPq, National Council for Scientific and Technological Development; and CAPES, Coordination for the Improvement of Higher Education Personnel (Finance Code 001). We extend our acknowledgement to Texas A&M University for hosting the first author.

References

[1] L. C. Gomes, F. J. J. A. Bianchi, I. M. Cardoso, R. P. O. Schulte, R. B. A. Fernandes, and E. I. Fernandes-Filho, "Disentangling the historic and future impacts of land use changes and climate variability on the hydrology of a mountain region in Brazil," *J. Hydrol.*, vol. 594, p. 125650, Mar. 2021, doi: 10.1016/j.jhydrol.2020.125650.

[2] A. T. Silva, J. Lugon Junior, and W. R. Telles, "Estimating Extreme Rainfall Equation Parameter in Southeast Brazil Using Machine Learning," *Rev. Gestão Soc. e Ambient.*, vol. 18, no. 4, p. e05153, Mar. 2024, doi: 10.24857/rgsa.v18n4-097.

[3] L. N. N. de Souza, J. Lugon, and V. de Paulo Santos de Oliveira, "Influence of Climate Change on Precipitation and the Formulation of Rain Design in Brazil," *Fronteiras*, vol. 12, no. 1, pp. 10–21, 2023, doi: 10.21664/2238-8869.2023V12I1.P10-21.

[4] X. Zhang, M. Xu, S. Kang, H. Wu, and H. Han, "The spatiotemporal variability in

precipitation gradients based on meteorological station observations in mountainous areas of Northwest China," *Theor. Appl. Climatol.*, vol. 155, no. 1, pp. 163–186, Jan. 2024, doi: 10.1007/s00704-023-04618-7.

[5] D. S. Sales, J. L. Junior, V. de P. S. de Oliveira, N. S. Ferreira, and A. J. Silva Neto, "Sensitivity analysis of atmospheric phenomena models for precipitation assessment on the Paraíba do Sul River Watershed," *Ciência e Nat.*, vol. 44, p. e55, 2022, doi: 10.5902/2179460x66757.

[6] D. S. Sales, J. Lugon Junior, V. P. S. Oliveira, and A. J. Silva Neto, "Rainfall Input From Wrf-Arw Atmospheric Model Coupled With Mohid Land Hydrological Model for Flow Simulation in the Paraíba Do Sul River-Brazil," *J. Urban Environ. Eng.*, vol. 15, no. 2, pp. 188–203, 2021, doi: 10.4090/juee.20219.v15n2.188203.

[7] J. A. Marengo, L. M. Alves, T. Ambrizzi, A. Young, N. J. C. Barreto, and A. M. Ramos, "Trends in extreme rainfall and hydrogeometeorological disasters in the Metropolitan Area of São Paulo: a review," *Ann. N. Y. Acad. Sci.*, vol. 1472, no. 1, pp. 5–20, Jul. 2020, doi: 10.1111/nyas.14307.

[8] M. D. Villas-Boas, F. Olivera, and J. P. S. de Azevedo, "Assessment of the water quality monitoring network of the Piabanha River experimental watersheds in Rio de Janeiro, Brazil, using autoassociative neural networks," *Environ. Monit. Assess.*, 2017, doi: 10.1007/s10661-017-6134-9.

[9] CPRM, "Isoietas Médias Anuais de 1977 a 2006 na escala 1:5.000.000. In: Atlas Pluviométrico do Brasil.," Maceio, 2011. [Online]. Available: https://www.sgb.gov.br/publique///Mapas-e-Publicacoes/Atlas-Pluviometrico-do-Brasil-1351.html

[10] A. A. dos Reis, A. Weerts, M. H. Ramos, F. Wetterhall, and W. dos S. Fernandes, "Hydrological data and modeling to combine and validate precipitation datasets relevant to hydrological applications," *J. Hydrol. Reg. Stud.*, vol. 44, p. 101200, 2022, doi: 10.1016/j.ejrh.2022.101200.

[11] J. K. Searcy and C. H. Hardison, *Double-Mass Curves*. Washington: United States Government Print Office, 1960. [Online]. Available: http://dspace.udel.edu:8080/dspace/handle/19716/1592

[12] T. A. J. G. Sirisena, S. Maskey, R. Ranasinghe, and M. S. Babel, "Effects of different precipitation inputs on streamflow simulation in the Irrawaddy River Basin, Myanmar," *J. Hydrol. Reg. Stud.*, vol. 19, no. April, pp. 265–278, Oct. 2018, doi: 10.1016/j.ejrh.2018.10.005.

[13] T. A. J. G. Sirisena, S. Maskey, J. Bamunawala, E. Coppola, and R. Ranasinghe, "Projected streamflow and sediment supply under changing climate to the coast of the kalu river basin in tropical sri lanka over the 21st century," *Water (Switzerland)*, vol. 13, no. 21, 2021, doi: 10.3390/w13213031.

[14] S. Maskey, "HyKit: A Tool for Grid-based Interpolation of Hydrological Variables," Delft, The Netherlands, 2013.

[15] JPL/NASA, "NASADEM Merged DEM Global 1 arc second V001," California, 2020. doi: https://doi.org/10.5067/MEaSUREs/NASADEM/NASADEM_HGT.001.

[16] O. Conrad *et al.*, "System for Automated Geoscientific Analyses (SAGA) v. 2.1.4," *Geosci. Model Dev.*, vol. 8, no. 7, pp. 1991–2007, Jul. 2015, doi: 10.5194/gmd-8-1991-2015.

[17] I. Masih, *Understanding hydrological variability for improved water management in the Semi-Arid Karkheh basin, Iran*. The Netherlands: CRC Press/Balkema, 2011.