STRATEGIES FOR GROUND-BASED VALIDATION OF HOURLY AND DAILY GPM-IMERG DATA OVER FLOOD PRONE AREAS

Marcos Roberto Benso[1]; Antonio José Homsi Goulart[2] & Ali Tokay[3]

Resumo

As áreas urbanas são altamente suscetíveis aos impactos de eventos de chuva intensa, incluindo inundações e perdas e danos humanos e materiais. O conhecimento preciso da precipitação é crucial para a compreensão e previsão desses eventos; no entanto, a precisão das estimativas ainda é pouco compreendida em agregações horárias e diárias. Este artigo aborda a validação de produtos GPM-IMERG para chuvas intensas em áreas urbanas brasileiras. O estudo examinou o desempenho das estimativas de precipitação do GPM-IMERG comparando-as com observações de pluviômetros com frequência sub-horária, agregadas em escalas horárias e diárias. Diversos métodos de validação, incluindo métricas de eficiência, tabelas de contingência e classificação por limiar, são empregados para avaliar a precisão e a confiabilidade dos produtos GPM-IMERG.

Palavras-Chave – risco de desastres, chuvas intensas, sensoriamento remoto.

Abstract

Urban areas are highly susceptible to the impacts of heavy rainfall events, including flooding and human and material loss and damage. Accurate precipitation knowledge is crucial for understanding and predicting these events, however, the accuracy of estimations remains poorly understood at hourly and daily aggregations. This paper focuses on the validation of GPM-IMERG products specifically for heavy rainfall in Brazilian urban areas. The study examines the performance of GPM-IMERG precipitation estimates by comparing them with ground-based observations from sub-hourly rain gauges aggregated at hourly and daily scales. Various validation methods, including efficiency metrics, contingency tables, and threshold classification, are employed to assess the accuracy and reliability of the GPM-IMERG products.

Keywords – disaster risk, heavy rainfall, remote sensing.

[1]) "Luiz de Queiroz" College of Agriculture, University of São Paulo, marcosbenso@usp.br
[2]) "Luiz de Queiroz" College of Agriculture, University of São Paulo, antonio.goulart@usp.br
[3]) Goddard Space Flight Center, National Aeronautics and Space Administration, ali.tokay-1@nasa.gov

INTRODUCTION

Short duration heavy rainfalls can cause extreme losses and damages and unveil complex hazards. From 2002 to 2022 on average 121 million people were affected by storms, floods and landslides causing a financial damage of 141 billion USD and causing approximately 16541 deaths (EM-DAT, 2023). Moreover, the frequency of extreme precipitation is expected to increase due to impacts of climate change (Seneviratne et al., 20211). Ayat et al. (2022) reported significant positive trends in sub-hourly rainfall that were not captured by daily rain gauges and climate models.

Extreme rainfall events have been reported in South America in recent years (Marengo et al. , 2023). This problem is aggravated in poorly gauged sites that are vulnerable to floods and landslides. Satellite remotely sensed precipitation represents great potential to improve the heavy precipitation statistics, however the reliability of these datasets is dependent on the availability of ground-based data (Sun et al., 2018). Moreover, some studies have reported biases towards the estimation of deep convective storms in South America (Rasmussen et al., 2013) .

Given the challenge of validating precipitation products over undergauged areas, this work aims to validate Global Precipitation Measurement Integrated Multi-satellite Retrievals for GPM (GPM-IMERG) (Hoffman et al., 2015) against ground-based hourly and daily precipitation over an area prone to floods located in the State of São Paulo, Brazil.

BACKGROUND

NASA's Global GPM-IMERG is a precipitation product derived from the merge, calibration and interpolation of multi-satellite measurements using passive microwave (PMW) and microwave-calibrated infrared (IR) precipitation estimates (Huffman et al., 2015). The GPM-IMERG products are made available as near-real time data by Early and Late products and the Final product that is only released after monthly ground-based rain gauges are taken into consideration (Foelsche et al., 2017).

GPM-IMERG products have been tested in some regions of Brazil. For instance, Amorin et al. (2020) evaluated the level three late GPM-IMERG product over significant agricultural basins using daily ground-based data. Their validation method involved assessing the NSE of the SWAT model forced by satellite data. The study concluded that GPM-IMERG outperformed ground-based simulations in terms of accuracy.

Teodoro et al. (2020) investigated the GPM-IMERG (Early) product in the Bacia Hidrográfica do Rio Sapucaí (BHRS) region utilizing ground-based observations from ANA. The daily data was validated using efficiency metrics and contingency tables. However, this study found that the performance of GPM-IMERG in this area was relatively low. Another study conducted by (Rozante et al., 2018) focused on the GPM-IMERG Level 3 product in Brazil. Ground-based observations from CPTEC/INPE were employed, and the data was analyzed on a daily basis. Through the use of efficiency metrics, contingency tables, and threshold classification, the study revealed that IMERG tended to overestimate precipitation in this region.

Furthermore, Oliveira et al. (2018) examined the GPM-IMERG Level 3 product in the middle of the Amazon basin, specifically around the city of Manaus in the state of Amazonas, Brazil. They utilized ground-based observations from the S-band weather radar of the Amazon Protection National System (SIPAM) and INMET. The data, collected hourly, was validated using

the PUSH model. The findings showed that while the estimated error demonstrated a lower standard deviation than the observed error, there was a strong correlation between satellite and radar rainfall. This study successfully captured the systematic error along the Negro, Solimões, and Amazon rivers, particularly during the wet season.

METHODS

Ground-reference data was obtained from Brazil's National Center for Natural Disaster Monitoring and Warning (CEMADEN) (red boxes in Fig. 1; http://www.cemaden.gov.br/) for the period of January to December of 2022. The data were obtained from tipping bucket gauges and have a resolution of 10 minutes during rainy periods and 60 minutes during non-rainy periods. The gauge platforms are located in São Paulo Metropolitan Region (SPMR) (purple crosses in Fig. 1) with a reasonably uniform distribution within the selected GPM's cell (red box in Fig. 1). In total, 12 stations fall within the study area. These gauges are not part of the reference networks used for calibration in IMERG, making them an independent dataset for verification purposes.

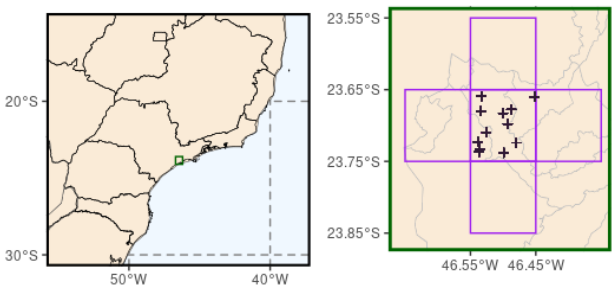


Figure 1. Location of study area. The pixel in the middle is referred to as the origin and the other four pixels are north, south, east, and west positions in relation to the origin pixel. Fonte: autoria própria.

A selection of 20 candidate stations were considered for establishing the basis of calculation of average precipitation within the IMERG pixel area. A leave-one-out cross-validation strategy was applied to calculate the percent absolute error according to different numbers of stations. This analysis is intended to evaluate the minimum number of stations within a pixel area that yield acceptable precipitation average for ground-validation. A percentage absolute error lower than 10% was adopted as threshold for station cutoff.

The classic metrics like probability of detection (POD), false alarm ratio (FAR), bias in detection (BID), and Heidke skill score (HSS), correlation coefficient and root mean square error (RMSE) were employed to validate GPM-IMERG precipitation (Tan et al., 2021).

RESULTS

Firstly, the number of ground-based stations was set to 12 in order to guarantee the quality criteria while minimizing the number of stations (figure and more details will be provided in the final version). The ground-validation considered two schemes for temporal aggregation: hourly and daily.

Validation presents different performance metrics according to pixel position in relation to the origin. The north pixel presented the best balance between the metrics. The lowest FAR indicating minimum false negative numbers and the lowest BID indicating lower bias towards ground-observations. Despite not having the best performance in terms of POD, CC, and RMSE,

the north pixel presented a higher HSS, which measures the skill of the estimate in indicating when a precipitation event occurred or not. Ultimately, the north pixel presented the best trade-off between POD and FAR.

As expected, the daily aggregation improved the correlation metrics, while not significantly altering the error metrics and the detection metrics, with the exception of FAR and BID. The number of false positives tended to be smaller in hourly than in daily. The bias tended to be positive in daily and negative in hourly, indicating that daily aggregates overestimated precipitation while hourly underestimated it, likely due to the sub-hourly convection patterns in the region.

Table 2 – Performance metrics for daily aggregates

Pixel	Hourly aggregates						Daily Aggregates					
	POD	FAR	BID	HSS	CC	RMSE	POD	FAR	BID	HSS	CC	RMSE
Origin	0.42	0.48	0.95	0.35	0.29	3.90	0.47	0.57	1.17	0.38	0.31	4.42
South	0.44	0.49	0.99	0.36	0.34	3.59	0.49	0.58	1.21	0.40	0.38	3.72
West	0.39	0.44	0.87	0.33	0.25	4.00	0.43	0.55	1.12	0.36	0.27	4.52
East	0.42	0.47	0.93	0.36	0.31	3.83	0.48	0.55	1.11	0.40	0.36	4.11
North	0.41	0.43	0.85	0.36	0.28	4.26	0.46	0.54	1.10	0.40	0.35	4.28

CONCLUSÕES

This work presented ground-validation strategies for GPM-IMERG precipitation estimates in flood prone areas. Considering a desired error rate for an application, it is possible to determine the minimum number of stations in order to calculate average precipitation within a pixel. This reveals the fragility of point to pixel analyses that consider only one precipitation measurement within the pixel area. This is exacerbated with the aggregation from hourly to daily that probably overlooks complex atmospheric patterns moving moisture flow in the atmosphere.

REFERÊNCIAS

AYAT, H.; EVANS, J. P.; SHERWOOD, S. C.; SODERHOLM, J. (2022). “Intensification of subhourly heavy rainfall”. Science, v.378, n.6620, pp.655–659.

HUFFMAN, G. J.; BOLVIN, D. T.; NELKIN, E. J.; WOLFF, D. B.; ADLER, R. F.; GU, G.; HONG, Y.; BOWMAN, K. P.; STOCKER, E. F. (2007). “The TRMM multisatellite precipitation analysis (TMPA): Quasi-global, multiyear, combined-sensor precipitation estimates at fine scales”. Journal of Hydrometeorology, v.8, n.1, pp.38–55.

RASMUSSEN, K. L.; CHOI, S. L.; ZULUAGA, M. D.; HOUZE JR., R. A. (2013). “TRMM precipitation bias in extreme storms in South America”. Geophysical Research Letters, v.40, n.13, pp.3457–3461.

SUN, Q.; MIAO, C.; DUAN, Q.; ASHOURI, H.; SOROOSHIAN, S.; HSU, K.-L. (2018). “A review of global precipitation data sets: Data sources, estimation, and intercomparisons”. Reviews of Geophysics, v.56, n.1, pp.79–107.

TAN, J.; PETERSEN, W. A.; TOKAY, A. (2016). “A novel approach to identify sources of errors in IMERG for GPM ground validation”. Journal of Hydrometeorology, v.17, n.9, pp.2477–2491.