Flood prediction model in Rio Grande do Sul based on data from the May 2024 flood

Carlos Henrique de Moura Silva [1]

[1] Data Scientist – São Paulo/SP Brazil – carlos-henrique.silva@serpro.gov.br

Keywords: Flooding Prediction, Rio Grande do Sul, Synthetic Aperture Radar, Hydrological Variables, Machine Learning.

1.1 Introduction

Floods represent a significant natural hazard, impacting communities, infrastructure, and ecosystems worldwide. The state of Rio Grande do Sul in Brazil is particularly susceptible to flooding due to its diverse topography and climatic variability. The recent major flood event in May 2024 underscored the urgency of developing robust predictive models to identify flood-prone areas. This study aims to create a predictive model for flood-prone areas in Rio Grande do Sul, leveraging Synthetic Aperture Radar (SAR) satellite data to identify flooded regions during the May 2024 event, and integrating hydrological bioclimatic data as predictors.

Flood prediction is a complex task that requires accurate and timely data. SAR satellite imagery offers a reliable method for detecting flood extents due to its ability to penetrate cloud cover and provide high-resolution data irrespective of weather conditions (Long, HU, Ding, Dong, Tian and Zeng, 2018). In this study, SAR data from the May 2024 flood event will be analyzed to delineate the affected areas. These delineations will serve as the dependent variable in our predictive modeling.

The predictive model was developed using statistical and machine learning techniques, enabling the identification of areas at heightened risk of flooding. The ultimate goal is to provide a tool that can aid in flood risk management and mitigation efforts, supporting decision-makers in the implementation of effective flood prevention and response strategies.

1.2 Methodology

This study generated a flood extent map to assess the affected areas using a methodical, step-by-step approach. The flood extent is determined through a change detection technique applied to Sentinel-1 Synthetic Aperture Radar (SAR) data. This technique involves a comparative analysis of images captured before (Figure 1) and after the flood event in May 2024 (Figure 2).

In this research, Sentinel-1 Ground Range Detected (GRD) imagery is employed. The preprocessing of GRD imagery comprises several crucial steps: Thermal-Noise Removal, Radiometric Calibration, and Terrain Correction. Following these steps, a Speckle filter is applied to enhance the image quality further.

The procedure is executed using Google Earth Engine. Within this platform, flood areas are delineated by marking specific points at various latitudes and longitudes within the inundated regions. These marked points serve as training data for the model (Figure 3).

To construct the model, it is essential to address the potential influence of geographical sampling bias on model outcomes, which can arise from aggregating data from multiple closely situated observations. We mitigate this by retaining only one randomly chosen occurrence record per pixel at the designated spatial resolution.

For model creation, we selected a combination of hydrological (GEE Community Catalog 2024) predictor variables (Table 1). The initial step involves creating a mask that delineates the presence of flooding data locations, preventing the generation of pseudo-absences in the same pixel locations as the actual presences. An issue related to the mask code, which is used to eliminate presence point locations from the pseudo-absence area, has been resolved. It was also necessary to exclude bodies of water from consideration since our area of interest includes aquatic regions.

Subsequently, random pseudo-absences were generated across the entire study area. After defining the locations of species presence and creating the pseudo-absences, we proceeded to elaborate on the distribution model, generating a raster file. A cross-validation approach known as "10 block repeated split-sample cross-validation" was employed (Giovanni, 2023). To ensure the integrity of our training and validation datasets, we maintained a balance between the number of pseudo-absences and presences. During each iteration, the spatial blocks were randomly divided into two sets: one for model fitting (70%) and the other for model validation (30%). Consequently, each of the 10 runs exhibited a distinct arrangement of presence and pseudo-absence points for both model fitting and validation.

By employing this methodology, the accuracy of flood mapping is significantly enhanced, yielding valuable data for disaster management and mitigation efforts.

1.3 Result

The outcome of this work is a web application developed using Google Earth Engine (Google Developers 2023), enabling users to interact with the flood prediction model created for the state of Rio Grande do Sul. This model was trained using data from the significant flood event in May 2024. By clicking on the map of the state, the web application returns the probability of flooding at the selected point (latitude/longitude) (Figure 6). The web application can be accessed at: https://ee-kikosmoura.projects.earthengine.app/view/probabilidade-de-enchente-no-rio-grande-do-sul

1.4 Figures

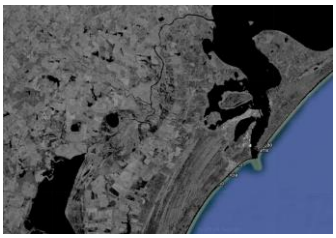

Figure 1. SAR image before the flood.

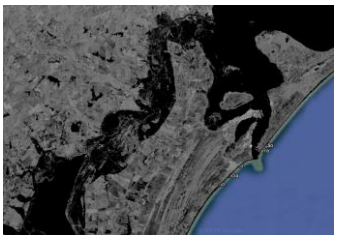

Figure 2. SAR image after the flood.

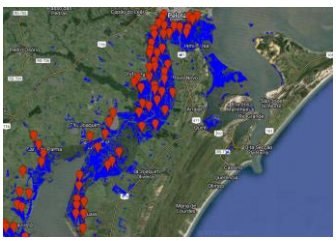

Figure 3. Points of presence in flood areas.

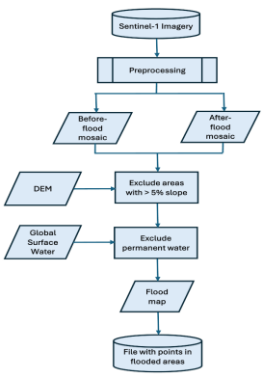

Figure 4. SAR image processing flow.

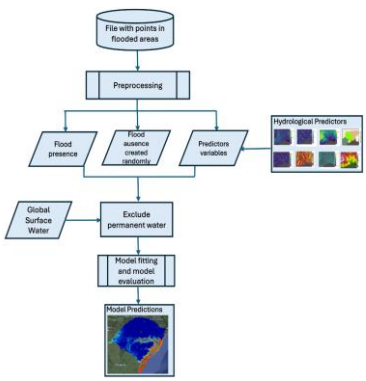

Figure 5. Predictive model processing flow.

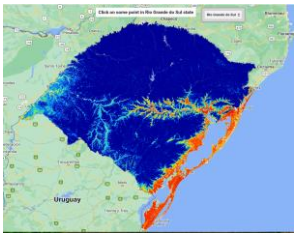

Figure 6. Flood prediction model generated.

1.5 Tables

Hydrological predictors
Flow accumulation Flow direction Drainage basin Depression Drainage direction Stream segment
Max curvature between highest upstream
Min curvature between highest upstream
Elevation difference between focal cell and downstream cell
Focal cell gradient
Elevation diff between focal grid cell and the outlet grid cell in the network
Elevation diff between focal grid cell and the downstream stream node grid cell
Dist between focal grid cell and the outlet grid cell in the network
Dist between focal grid cell and the downstream stream
Elevation difference between focal grid cell and its nearest downstream stream pixel
Elevation diff of the shortest path from focal grid cell to the sub-catchment drainage divide
Dist between focal grid cell and its nearest downstream stream grid cell
Euclidean dist between focal grid cell and the stream
Longest upstream dist between focal grid cell and the nearest sub-catchment drainage divide
Shortest upstream dist between focal grid cell and the nearest sub-catchment drainage divide
Compound topographic index (cti)
Stream transportation index (sti)
Stream power index (spi)

Table 1. Hydrological predictor variables.

References

Long, T., Hu, C., Ding, Z., Dong, X., Tian, W., Zeng, T., 2018: Geosynchronous SAR: System and Signal Processing. Tech Publications.

GEE Community Catalog. "Hydro90." Accessed June 17, 2024. https://gee-community-catalog.org/projects/hydro90/.

Giovanni, C., 2023: Fundamentals of Supervised Machine Learning With Applications in Python, R, and Stata

Google Developers. (2023). Creating web apps with Google Earth Engine. Google Developers. https://developers.google.com/earth-engine/tutorials/community/creating-web-apps. Accessed June 18, 2024.