LEPTOSPIROSIS IN 2024 FLOOD-AFFECTED AREAS OF RIO GRANDE DO SUL: EXPLORING THE ROLE OF GIS, REMOTE SENSING, AND DASYMETRIC MAPPING

Vinícius Lima Guimarães[1]

[1]National Institute for Space Research (INPE), São José dos Campos, SP, vinicius.lima@inpe.br

ABSTRACT

Leptospirosis, a bacterial disease acquired through contact with contaminated water or soil, is the focus of this study, which examines the effect of the 2024 floods in Rio Grande do Sul, Brazil, on the onset of the disease. Population distribution, land use, and disease prevalence were assessed using Geographic Information Systems (GIS) and remote sensing. Descriptive analyses compared the demographics of the affected low-income group to the state's general population. Using dasymetric mapping, the incidence of leptospirosis was estimated, and its correlation with land use, elevation, and socioeconomic variables was examined. The study also highlights discrepancies in the reported proportion of flooded areas. Despite employing advanced models like Random Forest and Gradient Boosting, the statistical analysis struggled to reliably link leptospirosis incidence with other factors associated with the disease. More accurate data are needed to further investigate leptospirosis transmission dynamics in flood-prone areas..

Key words – *Leptospirosis, Rio Grande do Sul, Flooding, Dasymetric Mapping, Remote Sensing.*

1. INTRODUCTION

Leptospirosis is a bacterial disease that humans can contract through direct contact with the urine of infected animals or indirect exposure to water and soil contaminated by their urine. For infection to occur, the pathogen *Leptospira* must penetrate the skin barrier, especially if the skin is damaged [1]. In humans, symptoms appear after an incubation period of 7 to 12 days, varying from three days up to a month [2].

Activities like swimming or rice cultivation can lead to indirect transmission through exposure to contaminated water and soil , making them potential reservoirs for leptospires from infected animals [1]. In urban environments, rats (*Rattus norvegicus* and *Rattus rattus*) carry *Leptospira icterohemorrhagiae* and shed it in their urine, while in rural areas, dogs and cattle serve as reservoirs for the serovars *L. canicola* and *L. hardjo*, which are dominant in these regions [3]. The transmission cycle of *Leptospira* is illustrated in Figure 1.

Historically, leptospirosis is a disease with high incidence in the state of Rio Grande do Sul, presenting about 10 cases per 100000 inhabitants, affecting primarily male gender (>80%) and rural population (>60%). More than 60% of the cases occur in farming areas, particularly low-altitude regions (<300 meters) with irrigated rice fields. The transmission occurs both at the workplace and at home. In addition, rodents (capybaras and nutria) but also cattle and horses serve as important amplifying hosts in the chain of disease transmission [3].

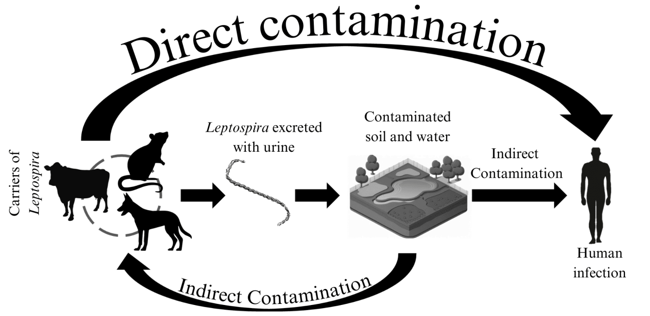

Figure 1: Transmission cycle of *Leptospira*.

In May 2024, 251 cases of leptospirosis were reported in the municipalities of Rio Grande do Sul, representing the greatest number of notified cases in 24 years [4]. This rise is potentially associated with the floods caused by the heavy rains from April 29, 2024, onwards, which raised the water levels of state's aquatic systems [5], affecting 485 municipalities. These cases, however, are linked not only to the historically disease-prone rural areas but also to urban flooded zones, where rats serve as disease carriers.

In floods, there is increased dispersion of the bacterium *Leptospira* due to the movement of contaminated water, raising cases among populations in vulnerable areas [6, 7].

Ecological studies have suggested that environmental features, such as regions experiencing flooding, increase the risk of leptospirosis outbreaks. The contribution of Geographic Information Systems (GIS) and remote sensing in these studies is essential to estimate flooded areas [5], identify environmental characteristics associated with leptospirosis, such as land use [3, 6, 7], and model the distribution of the population potentially affected by flooding and, consequently, exposed to *Leptospira* [8].

Thus, this study aimed to evaluate the relationship between population distribution, land uses, and leptospirosis incidence in the municipalities of Rio Grande do Sul following the natural disaster. Firstly, an exploratory analysis of the characteristics of low-income population affected by the flood was conducted. After that, it attempted at estimating the incidence rate in flooded regions, using dasymetric mapping. Moreover, the study sought to elucidate the relationship between land use and cover, territorial arrangement in rural and urban areas, and elevation with disease incidence.

Remote sensing and GIS techniques were used to model population distribution and identify land uses, elevation and slope within flooded areas. The exploratory analysis on the characteristics of the population affected by the floods focused on differences in sex, age, race, and socioeconomic conditions to understand how the natural disaster impacted distinct groups.

2. MATERIAL AND METHODS

cite amsmath,amssymb,amsfonts algorithmic graphicx textcomp tabularx amsmath cite cuted mathtools kantlipsum

The state of Rio Grande do Sul is located in the south Brazilian region, between latitude -32.03° and longitude -52.09º, covering an area of 52,80 square kilometers, bordering Argentina and Uruguay, and with Porto Alegre as its capital city. The state has 10882965 inhabitants within 499 municipalities [9], of which 485 were affected by the 2024 floods. Figure 2 presents its location .

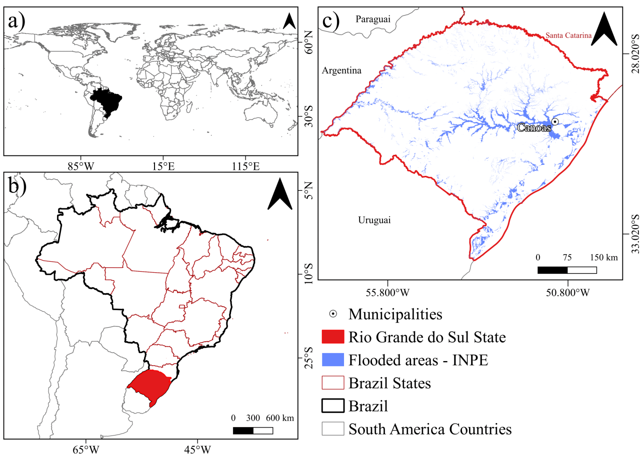

Figure 2: Study area location: (a) World Map and Brazil, (b) South America and Brazil, (c) Rio Grande do Sul State and the flooded areas of 2024.

The methodology is divided into four complementary parts, as illustrated in 3. These parts includes the exploratory analysis of official data on the affected population; the estimation of leptospirosis incidence in the flooded areas of the municipalities; the treatment of variables potentially associated with leptospirosis; and the statistical analysis of associations between disease incidence and the potentially related variables.

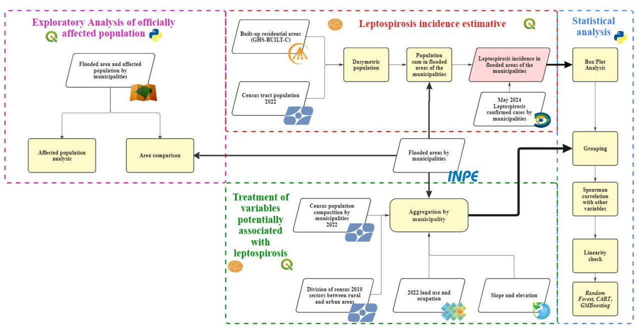

Figure 3: Flowchart of the Water Surface Extraction and Analysis Process.

2.1. Exploratory analysis of the officially affected population

The exploratory analysis of the officially affected population was based on the data from the Unique Map Plan Rio Grande (MUP RS) [10]. This is an integrated mapping system of the areas directly impacted by the natural disaster that occurred in May 2024 in Rio Grande do Sul, developed by the State Secretariat for Planning, Governance, and Management (SPGG). The system includes information on the areas consolidated by the SPGG as affected, based on the analysis of PlanetScope satellite images with a spatial resolution of 3 meters, as well as areas reported by each municipality Civil Defense, for which there is no public information about the methodology used for their definition.

To analyze the affected area defined by the MUP RS, a comparison was made with the area identified by the National Institute for Space Research (INPE), in the open-source software *QGIS 3.22*. The area defined by INPE will be used for the payment of the Reconstruction Aid, which is the official Federal Government program aimed at providing financial assistance to people affected by the floods, specifically in the areas where the water reached [11]. This is based on satellite image analysis and the contribution of information from other governmental and defense agencies at the federal, state, and municipal levels. This area has an official character regarding the definition of the flooded area boundaries.

The MUP RS also provides data on the composition by sex, race, education, age group, and income of families registered in the Cadastro Único para Programas Sociais (CadÚnico), which collects information from low-income families. Using *Python* programming language, a graphical analysis of these characteristics was conducted, comparing the characteristics of the low-income families affected by the floods with the characteristics of the entire population of the state of Rio Grande do Sul.

2.2. Estimation of leptospirosis incidence

The estimation of leptospirosis incidence, in turn, was conducted using the open-source software *TerraView* and *QGIS 3.22*. Initially, data collection was carried out for the population by census tract, based on the 2022 census data.

Data with 10-meter spatial resolution of built-up residential areas from the Global Human Settlement Layer (GHSL) [12] were also collected to obtain information on the spatial distribution of built surfaces. The GHS-BUILT-C dataset was derived from composite images from Sentinel-2, using advanced machine learning methods to estimate fractions of built surfaces (BUFRAC) and automatically classify built areas into residential (RES) and non-residential (NRES) domains. The GHS-BUILT-C classification methodology uses radiometric and morphological image descriptors in a symbolic machine learning (SML) approach, enabling identification of built surfaces at a 10-meter resolution. This product applies spatiotemporal interpolation of built surfaces, using a combination of data from multiple sensors and platforms, including Landsat and Sentinel-2, for the periods of 1975, 1990, 2000, 2014, and 2018 [13].

GHS-BUILT-C validation was performed by comparing the predictions of built surface fractions (BUFRAC) with reference data from building delineations available in vector format at a scale of 1:10,000. The test dataset, composed of approximately 50,000 globally representative test cases, showed a high correlation between the reference data and the GHSL predictions [13].

With the population data by census tract P and residential areas from GHS-BUILT-C, a binary dasymetric mapping (BDM) [14] was applied to redistribute the population from the census tracts into the built residential areas. The BDM method assumes that the population is exclusively present in the built-up areas and, therefore, redistributes the population proportionally to the area of the buildings.

In occupied areas, the redistribution of residents was done proportionally to the number of residential pixels, assuming that the population distribution is homogeneous within the census tract. The formula used to calculate the redistributed population in each destination zone was [14]:

$$\hat{P}_d = \frac{A_{o \cap d}}{A_d} \times P_o$$

where:

- $\hat{P}_d$ is the estimated population for each pixel d;
- P_o is the known population in the census tract o;
- $A_{o \cap d}$ is the area of residential pixels within the destination zone d that intersects with the census tract o;
- A_d is the total area of residential pixels in the destination zone d;
- n is the number of residential pixels in the destination zone.

The sum of the dasymetric population in the flooded areas of each municipality was calculated using the flood map provided by INPE. Using confirmed leptospirosis case data from the Notifiable Diseases Information System (SINAN) [4] the incidence rate of leptospirosis in the flooded areas of each municipality was calculated according to the formula below:

$$\mathrm{I}_f = \frac{N_c}{\Sigma_{\hat{P}_d,m}} \times 100{,}000$$

where:

- I_f is the incidence rate of leptospirosis in the flooded areas of each municipality;
- N_c is the number of confirmed cases of leptospirosis in May 2024, by municipality. Confirmed cases are defined as those that have laboratory confirmation with identification of the etiological agent, or clinical-epidemiological cases associated with the same epidemic or outbreak with laboratory-confirmed cases [15];
- $\Sigma_{\hat{P}_d,m}$ is the sum of the dasymetric population in the flooded areas of each municipality;

This calculation assumes that confirmed cases in the municipality in May 2024 are exclusively linked to the population exposed to the floods.

2.3. Treatment of variables associated with leptospirosis

To understand how population characteristics may impact flooding exposure and disease outcomes, population characteristics by municipality were obtained from the 2022 census population composition data [16].

The variables included in the analysis consisted of components of the total dependency ratio, the economically inactive male and female population across each age group (up to 14 years and over 65 years); the total economically active population; and the percentage distributions by different categories — male, female; and racial composition. These information were aggregated for each municipality.

Data from the 2010 census were also used, considering that they include the classification of areas as rural or urban at the census sector level. The census sectors were classified as urban or rural, firstly, based on municipal legislation related to urban perimeters, zoning, and taxes when such information could be accurately mapped. If no relevant legislation exists or if it is outdated, the classification relies on images, cartography, and field observations by IBGE staff to assess land use, density, and urban expansion. Additionally, physical features that are easily identifiable in the field are used to determine sector boundaries, which can lead to discrepancies with municipal legislation, especially when relying on hard-to-define lines [17]. In TerraView The flooded areas by municipalities were classified as rural, if more than 50% of its area was rural, and the same rule was used to urban class.

To assess how various land use characteristics influence flooding exposure and disease outcomes, 2022 MapBiomas data was gathered [18]. MapBiomas uses satellite images from the Landsat series (Landsat 5, 7, and 8) to produce annual land use and land cover maps in Brazil, with a spatial resolution of 30 meters per pixel. The images are processed on the Google Earth Engine platform through algorithms that create temporal mosaics, reduce spectral features, and classify land use classes using techniques like Random Forest. The results are integrated and filtered temporally to ensure consistency, producing detailed maps of the entire national territory.The maps are validated by comparing them with existing reference maps and performing accuracy analyses at sampling points, with an overall accuracy ranging from 73% to 80%.

From these data, was made an aggregation on the percentage of flooded areas within municipalities for different land uses: wetlands, grasslands, pastures, mosaics of uses, soybeans, rice, and other temporary crops. This includes areas historically inhabited by disease rural and urban carriers such as dogs, cattle, and rodents, which heightens the risk of indirect infection [3]. Additionally, the sum of the percentage of rural land uses associated with leptospirosis was analyzed. These factors were evaluated to understand their impact on disease transmission and public health outcomes.

Considering that elevation and slope are related not only to the distribution of flooding but also to the higher incidence of leptospirosis in lower altitude areas of Rio Grande do Sul

State [3], data from the ANADEM digital elevation model was collected [19]. The ANADEM is a digital elevation model developed to correct the bias caused by vegetation in global elevation models, such as the Copernicus DEM GLO-30 (COPDEM). This model utilizes a combination of altimetry data from the GEDI sensor, multispectral information from satellites like Landsat 8 and Sentinel 2, and advanced machine learning techniques, particularly the TreeBoost algorithm, to identify and remove the influence of vegetation cover on recorded elevation. It has a resolution of 30 meters. To validate the accuracy, ANADEM was compared with high precision measurements from ICESat-2. The validation results indicated that ANADEM has a root mean square error (RMSE) of 6.55 meters, demonstrating superior performance compared to other global which also corrects for vegetation bias.

2.4. Statistical analysis

In Python, by utilizing boxplot analysis and distinguishing between rural and urban flooded areas, different groups to examine the Spearman correlation with variables potentially associated with incidence were created. This approach helped assess linearity and the potential for linearization. For variables with non-linear associations, regression analysis was conducted using Random Forest, CART, and GMBoosting.

Random Forest is a machine learning ensemble method that builds many classification and regression trees and combines their output. It creates a "forest" of decision trees from random samples of the training data. The idea is that while individual decision trees often struggle with overfitting, the aggregation of multiple independent decision trees in Random Forest can reduce such risks and increase overall generalizability. The final prediction is then taken as the average of predictions (for regression) or the majority vote (for classification) across all trees.

CART (Classification and Regression Trees) is another machine learning algorithm that uses decision trees for classification and regression. Unlike Random Forest, which uses an ensemble of trees, CART constructs a single binary tree node that divides the data into two pieces based on an input variable. For classification tasks, CART uses metrics like Gini impurity or entropy to determine the best splits, while for regression, it minimizes the mean squared error (MSE). Although decision trees are simple and interpretable, they can be prone to overfitting if not properly pruned or regularized.

Building on the concept of combining models, Gradient Boosting Machine (GBM) is a machine learning method that produces a strong predictive model from a set of weak learners, usually decision trees. Unlike Random Forest and CART, GBM trains additional trees iteratively to correct the errors of previous models using a gradient-based optimization process to minimize a loss function. This approach is particularly effective at capturing complex nonlinear relationships in the data.

3. RESULTS

3.1. Officially affected population

The flooded area reported by MUP RS was 17890,60 square kilometers, including 16527,40 km² from areas consolidated by SPGG and 1363,19 square kilometers self-reported by municipal civil defenses. This represents an area approximately 771,33 km² smaller compared to the 18661,93 square kilometers reported by INPE. However, these differences do not imply that the areas will always be smaller in all municipalities, as illustrated in Figure 4.

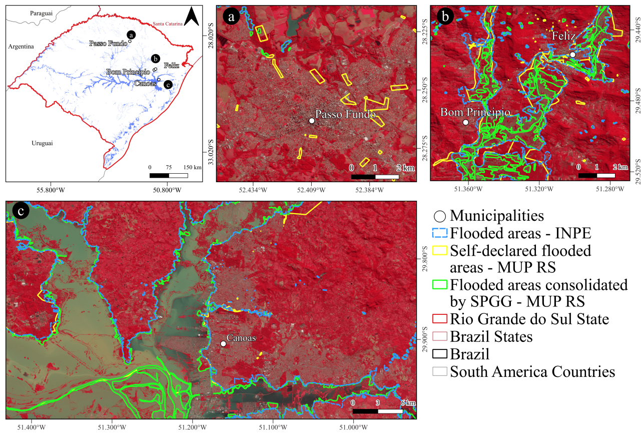

Figure 4: Example of the difference in flooded areas reported by the state of Rio Grande do Sul and INPE in the cities of Passo Fundo (a), Bom Princípio and Feliz (b), and Canoas (c), in a R5G3B2 PlanetScope composition.

Considering the area recognized by the state of Rio Grande do Sul, it's possible to observe that the characteristics of the affected population is similar to the overall population of the state, as shown in Figure 5.

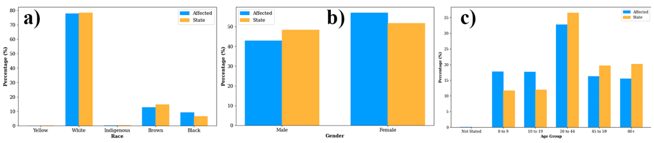

Figure 5: Comparison of the flood-affected population and the overall population of Rio Grande do Sul in terms of race (a), gender (b), and age group composition (c).

Differences in education level and monthly income among the low-income population affected by the flood can also be observed, as shown in Figure 6, where the most affected groups are those in extreme poverty and those with incomplete elementary education.

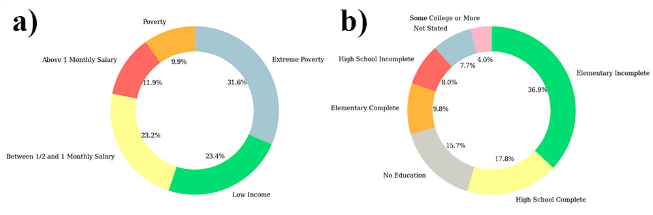

Figure 6: Composition of the affected population according to income (a) and education level (b).

3.2. Results of leptospirosis incidence estimation

When analyzing the distribution of residential areas in GHS-BUILT-C, it is apparent that the algorithm performs well in identifying both rural and urban residential areas, as shown in Figure 7.

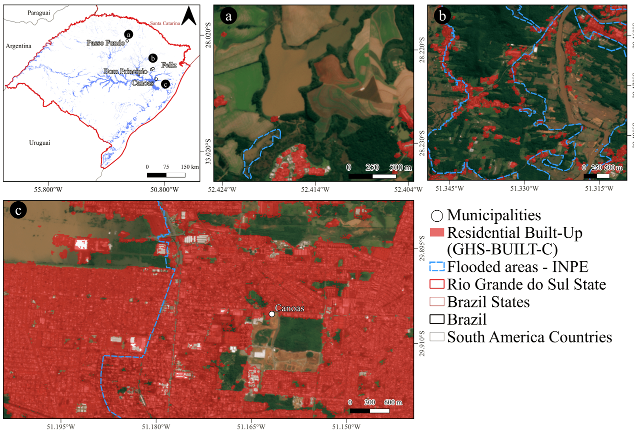

Figure 7: Examples of GHS-BUILT-C residential areas in rural regions of Passo Fundo (a) and Feliz (b), and urban areas in Canoas (c), in a true-color PlanetScope composition.

The disaggregation using the dasymetric method is shown in Figure 8. In areas with large urban concentrations, the disaggregation is less perceptible due to the size of the census sectors and the quantity of residential pixels. In contrast, in rural areas, the disaggregation is more noticeable, as large sectors are broken down into smaller amounts of residential pixels.

3.3. Results of statistical analysis

When analyzing the incidence of leptospirosis based on the dasymetric population aggregated by affected area in municipalities and distinguishing between predominantly urban or rural areas, six possible groupings were identified, as shown in Figure 9.

For these groups, Spearman's correlation and scatter matrix were checked with variables potentially associated with leptospirosis. It was found that there is no linear relationship or possibility of linearization between the dependent variable (leptospirosis incidence) and the independent variables, as illustrated in Figure 10. Thus, non-linear models should be applied.

When applying Random Forest, CART, and GMBoosting models to each group, as shown in Figure 11, it was observed that the models did not perform well in reducing errors between the dasymetric population-based incidence and the other variables potentially associated with leptospirosis.

4. DISCUSSION

The results indicate notable spatial and area differences between the flood-affected areas reported by INPE and those identified by the government of Rio Grande do Sul. Utilizing the data from Rio Grande do Sul assumes that these figures are approximations of the flooded area, with some areas being over- or underestimated. It is important to highlight that

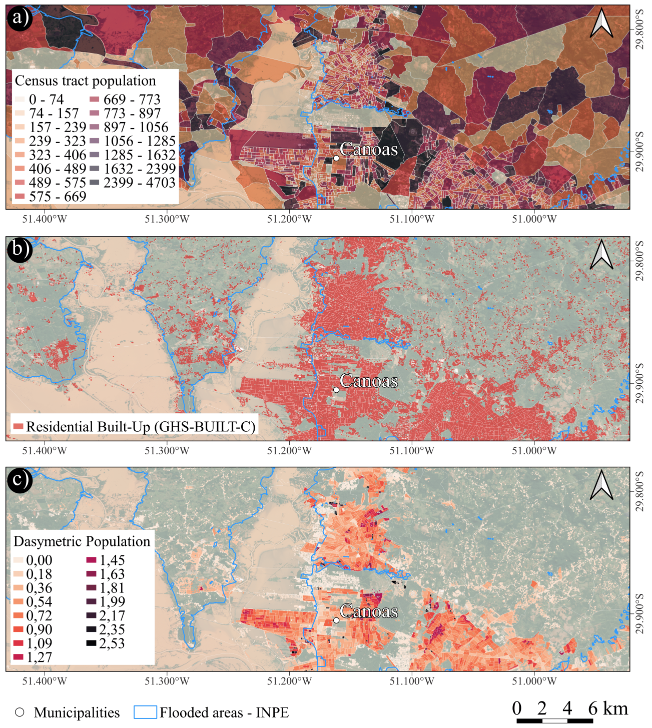

Figure 8: Example of results from the dasymetric disaggregation method in Canoas, showing aggregated population by sector (a), GHS-BUILT-C residential areas (b), and dasymetric population (c), in a true-color PlanetScope composition.

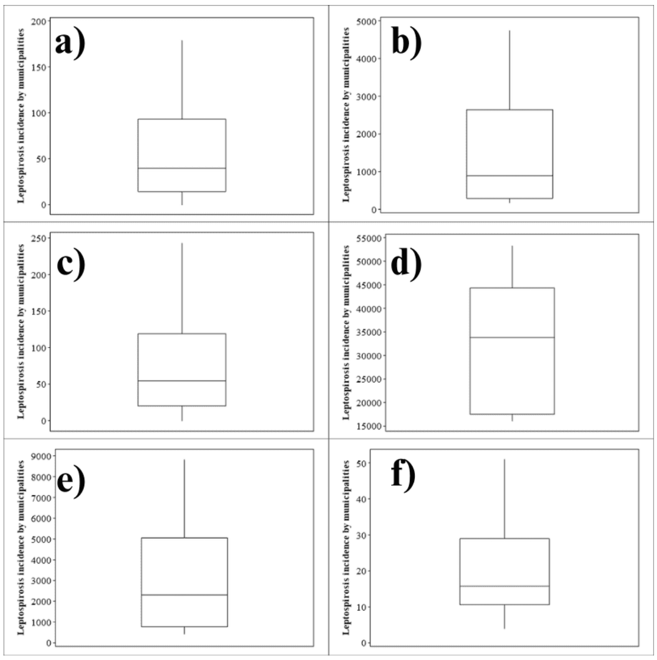

Figure 9: Groupings based on leptospirosis incidence in flooded areas. Group 1 (a) and 2 (b) show incidence considering all flooded areas; Group 1 (c), 2 (d), and 3 (e) focus on predominantly rural flooded areas; and Group 4 (f) represents predominantly urban flooded areas.

the methodology for self-reporting by civil defenses lacks transparency, potentially introducing additional errors into the analysis.

Exploratory data analysis reveals that, while the affected

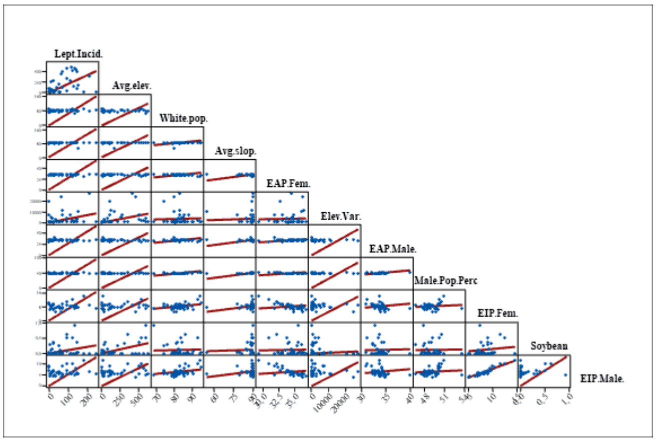

Figure 10: Example of a scatter plot between leptospirosis incidence and the ten variables with the highest Spearman correlation for Group 1 of predominantly rural flooded areas.

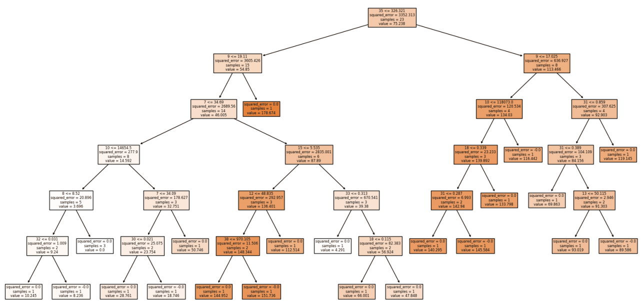

Figure 11: Example of a CART tree for Group 1 of predominantly rural flooded areas.

populations share gender, race, and age characteristics similar to the overall state population, they consist of historically more vulnerable social groups [20]. These groups are more susceptible to flooding, as evidenced by the higher number of low-income individuals affected and, consequently, exposed to flood risks. Future governmental actions are necessary to address this issue, alongside supplementary studies to analyze leptospirosis incidence among these groups.

Visually assessing examples of the GHS-BUILT-C's performance in identifying residential areas suggests that the model performs reasonably well. However, further efforts are needed to validate its performance in the study area.

Improvements in population estimation could be achieved through intelligent dasymetric mapping and dasymetric binary mapping incorporating data from the National Address Register for Statistical Purposes (CNEFE). These enhancements could potentially improve the accuracy of statistical analyses.

The statistical approach employed in this study, despite aiming to capture nonlinear relationships, did not robustly elucidate the connection between leptospirosis incidence in flooded areas and associated factors. Assumptions that confirmed cases within municipalities as of May 2024 were all related to the natural disaster, coupled with the chosen disaggregation methods and variables, may contribute to the observed discrepancies. Georeferenced data on likely infection locations and the application of alternative methods could provide further insights into the complex relationships between leptospirosis, flooding, and associated variables.

5. CONCLUSIONS

This study provides a comprehensive evaluation of the relationship between population distribution, land use, and leptospirosis incidence in Rio Grande do Sul following the 2024 floods. The analysis revealed notable discrepancies between the flood-affected areas reported by different agencies, indicating potential inconsistencies in data collection and reporting methods of Rio Grande do Sul state. The self-reported data from civil defenses, lacking transparency, may have introduced additional errors, highlighting the need for standardized and transparent methodologies for assessing flood impact.

The exploration of the affected population's demographics showed that although the characteristics of those affected are broadly similar to the overall state population, vulnerable groups such as low-income individuals with incomplete education were disproportionately impacted. This underscores the heightened susceptibility of these populations to flood risks and subsequent health issues like leptospirosis, pointing to a critical need for targeted public health interventions and support for these at-risk communities.

Despite the effective application of dasymetric mapping in redistributing population data within residential areas, the statistical models used to analyze leptospirosis incidence struggled to robustly identify significant relationships with associated variables. The limitations in the data assumptions and methodological approaches suggest that future studies should incorporate more precise geospatial data and explore alternative analytical methods to better understand the complex interactions between flooding, land use, and leptospirosis incidence.

6. ACKNOWLEDGMENTS

I extend my gratitude to the NICFI Satellite Data Program for providing the PlanetScope data and to the National Institute for Space Research for the flooded area data.

7. REFERENCES

[1] Younus M Nisa QU Ahmad HM Wazir N Akhtar T, Naeem MI and Tanveer K. Dog-mediated leptospirosis. *International Journal of Agriculture and Biosciences, Zoonosis*, 4:356–368, 2023.

[2] D. A. Haake and P. N. Levett. Leptospirosis in humans. *Current Topics in Microbiology and Immunology*, 387:65–97, 2014.

[3] Lammerhirt C. B. Almeida M. A. B. de Barcellos, C. and E. dos. Santos. Distribuição espacial da leptospirose no rio grande do sul, brasil: recuperando a ecologia dos estudos ecológicos. *Cadernos de Saúde Pública*, 19:1283–1292, 2003.

[4] CEVS. Tabulações da vigilância epidemiológica, 2024.

[5] ICSMD. Flooding in brazil, 2024.

[6] Smythe L. D. Craig S. B. Lau, C. L. and P Weinstein. Climate change, flooding, urbanisation and leptospirosis: fuelling the fire? *Transactions of the Royal Society of Tropical Medicine and Hygiene*, 104:631–638, 2010.

[7] Chalvet-Monfray K. Wiratsudakul A. Suwancharoen D. Cappelle J. Chadsuthi, S. A remotely sensed flooding indicator associated with cattle and buffalo leptospirosis cases in thailand 2011–2013. *BMC Infectious Diseases*, 18:602, 2018.

[8] Gil-Guirado S. Pérez-Morales, A. and V. Martínez-García. A method for population mapping and flood exposure assessment in touristic cities. *Applied Geography*, 142:102683, 2022.

[9] IBGE. Malha municipal, 2022.

[10] SPPG. Mapa Único plano rio grande, 2024.

[11] Brasil. Força-tarefa do governo trabalha para garantir auxílio reconstrução para famílias do rio grande do sul, 2024.

[12] Copernicus. Ghsl - global human settlement laye, 2024.

[13] European Commission and Joint Research Centre. Ghsl data package 2023, 2023.

[14] Filipe Silva. Modelação cartográfica e ordenamento do território : um ensaio metodológico de cartografia dasimétrica aplicado à região oeste e vale do tejo. *http://aleph.letras.up.pt/F?func=find-bfind*$_code = SYSrequest = 000190955, 012009$.

[15] Susanne Straif-Bourgeois, Julius L. Tonzel, Mirjam Kretzschmar, and Raoult Ratard. Infectious disease epidemiology, 2019.

[16] IBGE. Censo 2022, 2024.

[17] Caroline Krobath Luz Pera and Laura Machado de Mello Bueno. Revendo o uso de dados do ibge para pesquisa e planejamento territorial: reflexões quanto à classificação da situação urbana e rural. *Cadernos Metrópole*, 18(37):722–742, Sep 2016.

[18] Carlos Souza Jr and Tasso Azevedo. Mapbiomas general "handbook", 01 2017.

[19] Leonardo Laipelt, Bruno Comini de Andrade, Walter Collischonn, Alexandre de Amorim Teixeira, Rodrigo Cauduro Dias de Paiva, and Anderson Ruhoff. Anadem: A digital terrain model for south america. *Remote Sensing*, 16(13), 2024.

[20] Yi Qiang. Disparities of population exposed to flood hazards in the united states. *Journal of Environmental Management*, 232:295–304, 2019.