ISSN: 2447-3359

REVISTA DE GEOCIÊNCIAS DO NORDESTE

Northeast Geosciences Journal

v. 12, nº 1 (2026)

https://doi.org/10.21680/2447-3359.2026v12n1ID41844

Social vulnerability related to hydrogeological risks in the southeastern region of the state of São Paulo

Vulnerabilidade social relacionada a riscos de origem hidrogeológicos na região sudeste do estado de São Paulo

Tanara Vieira Pascoto; Anna Silvia Palcheco Peixoto

[1] São Paulo State University, Department of Civil and Environmental Engineering, Bauru/SP, Brazil. Email: tamara.pascoto@unesp.br
ORCID**:** https://orcid.org/0000-0001-9892-5973

[2] São Paulo State University, Department of Civil and Environmental Engineering, Bauru/SP, Brazil. Email: anna.peixoto@unesp.br
ORCID**:** https://orcid.org/0000-0001-5867-2684

Abstract: A region's disaster risk is intrinsically linked to its social vulnerability. This article aims to assess social vulnerability to hydrogeological disasters in the southeastern region of the state of São Paulo, an area characterized by a high incidence of such events. Data from the IBGE Demographic Census were processed using Factor Analysis to generate two GIS-based maps: one employing the standard deviation method and the other the equal count method. Validation was conducted by comparing these vulnerability maps with an inventory of hydrogeological disasters. It was observed that the map generated by the standard deviation method showed a better correspondence with the disaster inventory than the map generated by the equal count method. The resulting map serves as a valuable tool for directing actions towards more vulnerable regions, thereby mitigating the impact of extreme climatic events in the area.

Keywords: Social Vulnerability Index; Factor Analysis; Social Vulnerability Chart, Hydrogeological Disasters.

Resumo: O risco a desastres de uma região está relacionado à sua vulnerabilidade social. O artigo tem como objetivo avaliar a vulnerabilidade social à desastres hidrogeológicos na região sudeste do estado de São Paulo, área que possui um elevado número de ocorrência desses desastres. Os dados do Censo Demográfico do IBGE foram processados por meio da Análise Fatorial para gerar duas cartas em SIG (Sistema de Informação Geográfica): uma utilizando o método do desvio padrão e outra o método da igual contagem. A validação foi realizada através da comparação das cartas de vulnerabilidade com o inventário dos desastres de origem hidrogeológica. Observou-se que a carta gerada pelo método do desvio padrão apresentou uma melhor correspondência com o inventário de desastres, do que a carta gerada pelo método do igual contagem. A carta gerada auxilia na destinação de ações a regiões mais vulneráveis a fim de diminuir o impacto desses eventos climáticos extremos na região.

Palavras-chave: Índice de vulnerabilidade social; Análise fatorial; Carta de vulnerabilidade social; desastres hidrogeológicos

Received: 20/10/2025; Accepted: 15/05/2026; Published: 12/06/2026.

1. Intruduction

Extreme climate events have both direct and indirect consequences for the global population, as communities are directly affected by the destruction of infrastructure, loss of homes, damage to agriculture, and compromised livelihoods. In 2022, the EM-DAT Emergency Event Database recorded 387 natural hazards and disasters worldwide, resulting in the loss of 30,704 lives and affecting 185 million people. Economic losses totaled approximately $223.8 billion (USAID, 2023).

Therefore, developing susceptibility maps for regions with a high likelihood of disasters is of utmost importance for prevention. However, these must be analyzed in conjunction with precipitation rates—in the case of rain-induced disasters—as well as the vulnerability of the local population (UNITED NATIONS, 2015; UNDRR, 2022). This is because risk is directly related to these concepts, since risk results from the interaction of hazard with social vulnerability, and hazard, in turn, with environmental susceptibility and triggering agents, such as heavy rains and high precipitation rates (SANDES, 2023; IPCC, 2022).

Precipitation rates have varied over the years, with an increasing frequency of extreme weather events. Projections of changes in the climate regime indicate that this trend is likely to worsen, with a probable decrease in rainfall in northern Brazil and an increase in the South and Southeast (GREVE et al., 2018).

Vulnerable populations, particularly those living in low-income areas of developing countries and coastal regions, are disproportionately affected by disasters, as they possess fewer resources to adapt to and recover from their impacts (NEUMAYER and PLÜMPER, 2007; SAITO, 2011).

Numerous studies have been conducted to understand and assess social vulnerability across different contexts, aiming to identify social groups and geographic areas that are most susceptible to social and economic risks. Susan L. Cutter et al. (2003), pioneers in this field, developed the Social Vulnerability Index for the United States. The index was based on an additive model derived from factor analysis and initially incorporated 250 census variables. Guillard-Gonçalves et al. (2015) applied an adaptation of the SoVI methodology to assess vulnerability in a region of Portugal.

Several parameters have been used to explain social vulnerability, regardless of the methodology adopted to analyze these variables within a given area. These parameters include income and employment, age, race/ethnicity, education, health, housing conditions, food security, access to public services, and social participation (CUTTER et al., 2003). Depending on the study area, the influence of each parameter may vary considerably. For example, Susan L. Cutter et al. (2003), in their study conducted in the United States, identified population wealth and age as the most significant variables, followed by the density of the built environment. Marcelino et al. (2006) found that, in Santa Catarina, the most relevant variables were population density, poverty intensity, and the elderly population. Similarly, Guillard-Gonçalves et al. (2015), in a study carried out in Lisbon, identified the elderly population, buildings without sewage systems, and the proportion of foreign residents as the most influential variables.

Hummell et al. (2016) applied the model developed by Susan L. Cutter et al. (2003) to the entirety of Brazil at the municipal level. The study used census data from the Brazilian Institute of Geography and Statistics, the Ministry of Social Assistance, and the Ministry of Health. A total of 58 variables were considered in the analysis, allowing the country to be classified into five vulnerability categories.

Nascimento Jr. and Sant'Anna Neto (2020) developed a social vulnerability index based on the model proposed by Susan L. Cutter et al. (2003) for the city of Santos. For this purpose, census tracts were used to analyze the variables, as they represent the smallest administrative units for which Brazilian census data are available. A total of 117 initial variables were evaluated, of which 77 were selected for the development of the model. The resulting index was mapped using the standard deviation technique and classified into six vulnerability categories: very low, low, moderate, moderately high, high, and very high. Torres et al. (2021) analyzed vulnerability and conducted a risk assessment within the framework of Brazilian legislation for the metropolitan region of São Paulo.

Hummell et al. (2016), who analyzed Brazil as a whole, identified poverty, urban–rural development, and migration as the three variables exerting the greatest influence on vulnerability. Hader et al. (2021), in a study conducted in Cubatão, identified population density, followed by the elderly population and the poverty rate, as the most significant parameters, corroborating the findings of Marcelino et al. (2006). Similarly, Nascimento Jr. and Sant'Anna Neto (2020), in their study of the city of Santos, identified households without nominal monthly per capita income, literacy, and gender and race as the most influential variables.

The media has highlighted the increasing number of impacts caused by extreme high-intensity and heavy-precipitation events in the Southeastern region of Brazil (CORRÁ, 2023; ARREGUY, 2024). Between 1996 and 2020, a total of 950 geological disasters occurred in the administrative regions of Registro, São Paulo, Baixada Santista, and São José dos Campos (PASCOTO et al., 2022). These disasters included slope failures and landslides. However, this number may be even higher, since the authors quantified these disasters based on the data available on the website of the State Civil Defense, which may indicate an underreporting of such occurrences.

The assessment of social vulnerability at a high spatial resolution, particularly at the census tract level, together with the validation of these models using observed disaster data from the southeastern region of the state of São Paulo, highlights a gap in the availability of replicable methodologies. In this context, this study proposes the development of a social vulnerability map for four administrative regions of the state — São José dos Campos, Baixada Santista, São Paulo, and Registro — based on data from the 2010 Demographic Census conducted by the Brazilian Institute of Geography and Statistics at the census tract level, using the standard deviation and equal-count classification methods in QGIS.

2. Methodology

The study aims to determine a social vulnerability to disasters index (SoVI) for the administrative regions of São Paulo, São José dos Campos, Baixada Santista, and Registro, based on the methodology proposed by Hummell et al. (2016), who developed a social vulnerability index for all of Brazil at the municipal level. To this end, the authors used variables from the 2010 IBGE Census, the Ministry of Social Assistance (2010), and the Ministry of Health (2011), totaling 58 variables, of which 12 were discarded after a multicollinearity test. Using the remaining 45 variables, they applied Factor Analysis with principal component extraction and Varimax orthogonal rotation, identifying 10 factors responsible for explaining 67% of the data variability. The SoVI was then obtained by summing these factors for each municipality, without assigning different weights.

In contrast to the original study, which was conducted using municipal-level data, the present study employs data from the 2010 Demographic Census at the census tract level, enabling a finer spatial resolution of the analysis. After the exclusion of census tracts containing incomplete data, a total of 38,494 tracts were included in the analysis.

The 2010 Demographic Census was selected as the primary data source for this analysis, considering that the disaster inventory covers the period from 2000 to 2020 and that risk assessments are commonly grounded in the historical occurrence of disaster events. Within this framework, the vulnerability data employed must be temporally compatible with the period during which the disasters occurred, thereby ensuring an adequate representation of the socioeconomic conditions of the population exposed to such events.

A total of thirteen variables were considered in the analysis: poverty rate, housing conditions, female income, proportion of women, White population, Black population, mixed-race population, Asian population, Indigenous population, illiteracy rate, population under 14 years of age, elderly population, and population density.

To calculate the poverty index, as expressed in equation (1), variables from the Brazilian Institute of Geography and Statistics referring to private households without monthly income (V01), private households with a nominal monthly per capita household income of up to 1/8 of the minimum wage (V02), and private households with a nominal monthly per capita household income ranging from 1/8 to 1/4 of the minimum wage (V03) were used. These three variables were selected because an income of up to one-quarter of the minimum wage corresponds to approximately R$330.00 per month (US$60.50 per month). This selection was based on the international poverty line threshold of US$2.15 per day, adopted since September 2022 (DIAZ-BONILLA et al., 2022).

$$\text{Poverty Index} = V01 + V02 + V03 \quad (1)$$

The housing analysis aimed to identify the dwellings considered most socially vulnerable. For this purpose, the Brazilian Institute of Geography and Statistics variables “residents living in collective and private households” (V04) and “residents in permanent dwellings” (V05) were used. Permanent dwellings are defined as structures built exclusively for residential purposes, which may be classified as either private or collective. Collective dwellings are characterized by interpersonal relationships governed by rules of administrative subordination. Improvised dwellings, in turn, comprise structures not originally intended for residential use but occupied as such (IBGE, 2016). The housing parameter was calculated by subtracting V05 from V04 (equation 2), thereby isolating collective and improvised dwellings without producing negative values, since V05 (permanent dwellings) constitutes a subset of V04 (collective, permanent, and improvised dwellings).

$$\text{Dwelling} = \text{V04} - \text{V05} \quad (2)$$

Female income (equation 3) was defined as the sum of women with no income and those with an income of up to 1/2 the minimum wage, based on IBGE variables V06 and V07, respectively.

$$\text{Female income} = V06 + V07 \quad (3)$$

The analysis of the female population considered the ratio between the number of women in each census tract (V08) and the total population of the respective tract (V09), as expressed in equation (4).

$$\text{Female population} = \text{V08}/\text{V09} \quad (4)$$

The race/ethnicity variable was represented by the proportion of the main population groups — White (V10), Black (V11), Brown (V12), Asian (V13), and Indigenous (V14) — relative to the total population of each census tract (V09), as expressed in equations (5) through (9).

$$\text{White population} = \text{V10}/\text{V09} \quad (5)$$

$$\text{Black population} = \text{V11}/\text{V09} \quad (6)$$

$$\text{Brown population} = \text{V12}/\text{V09} \quad (7)$$

$$\text{Asian population} = \text{V13}/\text{V09} \quad (8)$$

$$\text{Indigenous population} = \text{V14}/\text{V09} \quad (9)$$

The illiteracy rate was estimated as the proportion of illiterate individuals in relation to the total population. t was calculated as the difference between the total number of inhabitants (V09) and the literate population aged 15 years and older (V15), as presented in equation (10). The age threshold of 15 years was adopted because elementary education—which should be completed by the age of 14—is the stage at which children and adolescents are expected to achieve full literacy in reading and writing (Law No. 9,394, 1996).

$$\text{Illiteracy rate} = (\text{V09} - \text{V15})/\text{V09} \quad (10)$$

The population age structure was characterized using the variables “population under 14 years of age” and “elderly population.” The former was defined as the proportion of individuals younger than 14 years relative to the total population of each census sector, calculated as the ratio between the population under 14 years of age (V16) and the total population (V09), as presented in equation (11). The adoption of 14 years as the threshold is supported by Brazilian labor legislation, which permits entry into the workforce as a minor apprentice from that age onward (Law No. 10,097, 2000). Consequently, individuals aged 14 years and older may exhibit greater socioeconomic autonomy and adaptive capacity, potentially contributing to increased resilience when compared with younger population groups.

$$\text{Population under 14 years old} = \text{V16}/\text{V09} \quad (11)$$

For the “Elderly” variable, the ratio between the number of people over 60 years of age (V17) and the total number of people in the sector (V09) was calculated, as shown in equation (12). The adoption of 60 years as the reference threshold is consistent with Brazilian national legislation, which legally defines individuals aged 60 years and older as elderly persons (Law No. 10,741, 2003).

$$\text{Elderly population} = \text{V17}/\text{V09} \quad (12)$$

Finally, population density was the last factor determined. It was calculated as the ratio of the number of inhabitants (V09) to the area of the census tract, in km² (equation 13).

$$\text{Population density} = V09/Área \tag{13}$$

All 13 variables were standardized according to equation (14) to ensure comparability among indicators by transforming their values to a common scale ranging from 0 to 1.

$$x_n = \frac{(x - x_{min})}{(x_{max} - x_{min})} \tag{14}$$

Where: X_n = normalized variable n, x = variable to be normalized, x_{min} = the minimum value obtained for this criterion across all census tracts, and x_{max} = the maximum value obtained for this criterion across all census tracts

The Social Vulnerability Index (SoVI) was determined using Principal Component Analysis (PCA). Accordingly, the dataset comprising the 13 variables was initially subjected to multicollinearity diagnostics and correlation analysis to assess the suitability of the variables for factor extraction. Based on the results of the factor analysis, three variables were excluded due to their low correlation with the underlying factor structure, resulting in a final set of 10 variables used in the PCA.

The suitability of the data for factor analysis was verified using Bartlett's test ($p < 0.05$) and the KMO test (KMO between 0.5 and 1.0), ensuring the consistency of the model (HAIR JR. et al., 2014). Next, PCA was applied using Varimax rotation, with criteria of eigenvalues greater than 1.0 and a minimum cumulative variance of 70%. The interpretation of the factors was based on the most significant factor loadings (>|0.5|), with positive values indicating increased vulnerability and negative values indicating decreased vulnerability.

Each factor was calculated separately using equation (14), where Fn represents the factor for each census tract, WP1 is the factor loading of the factor for a specific variable, and Pn is the variable for each tract, determined by equations (1) through (12), after normalization using equation (13). For a clearer visualization, each factor was mapped individually in QGIS.

$$F_n = W_{P1}P_1 + W_{P2}P_2 + \cdots + W_{Pn}P_n \tag{15}$$

Last, the SoVI was calculated as the sum of the factor values (F) for each census tract (Equation 16). The model used was additive, meaning it made no assumptions about the relative importance of each factor in the total index value.

$$\text{SoVI} = F_1 + F_2 \ldots + F_n \tag{16}$$

The vulnerability map was generated using two distinct vulnerability classifications in the GIS: the standard deviation method and the equal-counting method, both of which employ five vulnerability categories: very low vulnerability, low vulnerability, medium vulnerability, high vulnerability, and very high vulnerability.

To validate the best vulnerability map for the region, a quantitative survey was conducted on the number of disasters and the number of fatalities associated with each vulnerability class, considering events that occurred in the study area between 1996 and 2020

3. Study Area

The study area is located in the southeastern region of the state of São Paulo and comprises the administrative districts of São José dos Campos, Baixada Santista, Registro, and São Paulo. It covers approximately 38,672.7 km², of which 6.2% (2,417.9 km²) is in the Baixada Santista region, 41.8% (16,174.9 km²) of São José dos Campos, 31.4% (12,131.7 km²) of Registro, and 20.6% (7,948.2 km²) of São Paulo (Figure 1). This region was chosen because it accounted for 61.4% of all geological disasters that occurred in the state of São Paulo between 1996 and 2020 (PASCOTO et al., 2022).

Hummell et al. (2016) classified the vulnerability of this study area into different levels, with a predominance of areas of medium vulnerability, as well as regions classified as high (São Paulo administrative region), low, and very low (São José dos Campos administrative region). This distribution is corroborated by the inventory of disasters between 1996 and

2020 (Figure 2), compiled from IPMET data (PELLEGRINA, 2011), in which similar events, such as landslides and mudslides, as well as floods and inundations, were grouped together due to conceptual overlap in the records.

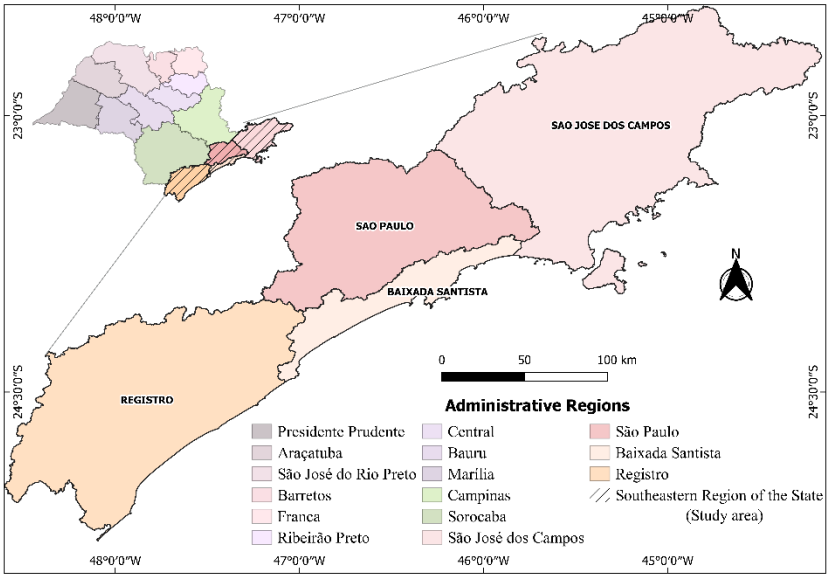

Figure 1 – Study area
Source: Elaborated by the authors (2025)

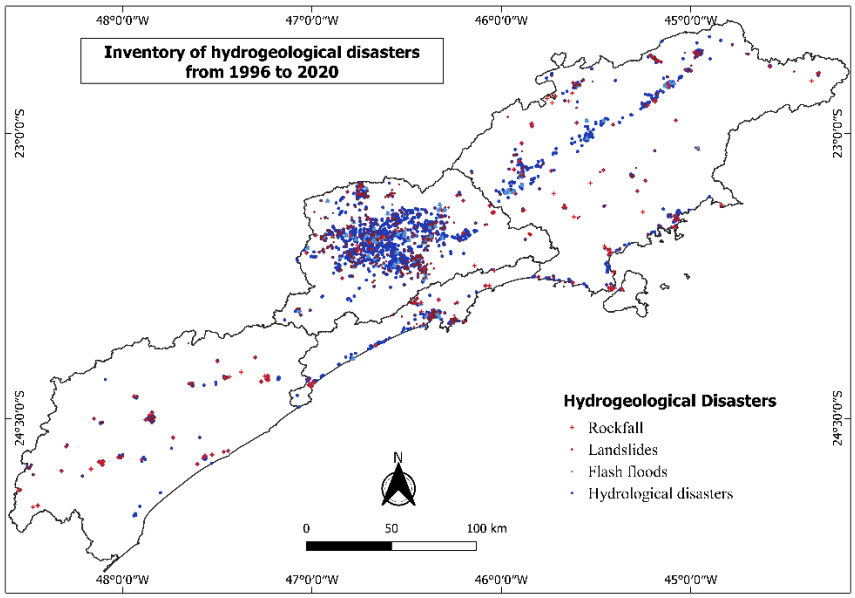

Figure 2 - Inventory of hydrogeological disasters in the study area from 1996 to 2020
Source: Elaborated by the authors (2025).

4. Results and Discussion

First, a multicollinearity test was conducted, which identified low correlations for the variables related to the Asian and Indigenous populations and the number of women; for this reason, these variables were excluded. The suitability of the data for factor analysis was confirmed by Bartlett's test ($p < 0.001$), indicating statistical significance, and the KMO test,

which yielded an overall MSA index of 0.689 and a minimum value of 0.545 for the housing variable (Table 1), both above the acceptable threshold, demonstrating the model's feasibility (HAIR et al., 1987).

Table 1 - KMO test sample adequacy measure

Variables	MSA
General	0,689
Poverty Index	0,690
Dwelling	0,545
Female income	0,560
White population	0,661
Black population	0,744
Mixed-race population	0,710
Illiteracy	0,805
Population under 14 years old	0,702
Elderly	0,848
Population density	0,615

Source: Elaborated by the authors (2025).

The number of retained factors was determined based on the eigenvalue criterion and the cumulative explained variance (Table 2). The first two factors exhibited eigenvalues greater than 1.0, whereas the third factor (0.996) was considered sufficiently close to the established threshold, in contrast to the fourth factor (0.732), which showed a markedly lower contribution. Moreover, the first three factors collectively explained 77.6% of the total variance in the dataset. Based on these results, the extraction of three factors was considered the most appropriate solution for the analysis.

Table 2 – Eigenvalues and cumulative variance of the factors

Factors	Eigenvalue	% of variance	Cumulative % of variance
1	4,5133	45,133	45,1
2	2,2486	22,486	67,6
3	0,9958	9,958	77,6
4	0,7324	7,324	84,6
5	0,4780	4,780	89,7
6	0,4390	4,390	94,1
7	0,3047	3,047	97,1
8	0,2061	2,061	99,2
9	0,0602	0,602	99,8
10	0,0219	0,219	100,0

Source: Elaborated by the authors (2025).

Table 3 presents the factor loadings associated with the three extracted factors. Factor 1 is primarily related to race/ethnicity, illiteracy, and age structure, exhibiting positive loadings for Black and Mixed-Race populations, illiterate individuals, and individuals under 14 years of age, as well as negative loadings for the White population and older adults. The Black and Mixed-Race population, which represents approximately 38.03% of the total population within the study area, may experience greater social vulnerability due to reduced access to resources necessary for disaster recovery and risk mitigation. The heightened vulnerability of these groups is associated with structural inequalities in resource distribution and processes of social marginalization (CUTTER et al., 2003; KOKS et al., 2015). In addition, illiteracy may limit the ability to respond effectively to disaster events by impairing the comprehension of risk communication and warning systems (PINHO et al., 2019; SAITO, 2011). Likewise, children under 14 years of age are generally considered more susceptible to the impacts of disasters due to their greater physical and social dependence (LIU et al., 2002; UNITED NATIONS, 2015).

Table 3 - Factor loadings of the extracted factors

Variable	Component	Singularity

	1	**2**	**3**	
Poverty Index		0,788		0,3005
Dwelling		0,947		0,0958
Female income		0,975		0,0362
White population	-0,882			0,1901
Black population	0,573			0,5229
Mixed-race population	0,892			0,1832
Illiteracy	0,846			0,2555
Population under 14 years old	0,874			0,1953
Elderly	-0,787			0,3312
Population density			0,920	0,1317

Source: Elaborated by the authors (2025)

Factor 2 is associated with the population's income and housing conditions. The poverty rate, housing conditions, and women's income contribute positively to this factor. This indicates that an increase in these parameters heightens the sector's vulnerability. People with lower incomes tend to be less resilient to the impacts of disasters than those with higher incomes. Furthermore, low-income individuals face greater difficulties in accessing resources allocated for prevention (NEUMAYER AND PLÜMPER, 2007; PINHO et al., 2019; SAITO, 2011). Regarding women's income, in 2019, women's wages were, on average, 77.7% of men's wages (IBGE, 2020). Therefore, women generally face more obstacles in recovering from disasters, in addition to being more vulnerable. Housing conditions are also related to the vulnerability of the location. Substandard housing influences the potential for disasters to occur, whether due to a lack of infrastructure or an inadequate location (MARQUES et al., 2020).

Factor 3 is positively and closely related to population density. This means that the higher the population density, the greater the vulnerability of the area. High population density can hinder evacuation efforts in the event of a disaster and generally results in a higher number of people affected (ALVINO-BORBA et al., 2020).

The three extracted factors were subsequently standardized according to Equations (17), (18), and (19), and spatially represented as shown in Figure 3. Figure 3a illustrates Factor 1, associated with race/ethnicity, illiteracy, and age structure; Figure 3b represents Factor 2, related to income and housing conditions; and Figure 3c depicts Factor 3, corresponding to population density. For each factor, vulnerability was classified into five categories using equal interval classification, defined by the following ranges: 0.0–0.2, 0.2–0.4, 0.4–0.6, 0.6–0.8, and 0.8–1.0.

$$Factor\ 1 = -0.882 * White\ Population + 0.573 * Black\ Population + 0.892 * Mixed - Race\ Population + 0.846 \times Illiteracy\ Rate + 0.874 \times Population\ Under\ 14\ Years\ of\ Age - 0.787 \times Elderly\ Population \quad (17)$$

$$Factor\ 2 = 0.788 \times Poverty\ Index + 0.947 \times Dwelling + 0.975 \times Female\ Income \quad (18)$$

$$Factor\ 3 = 0.920 \times Population\ density \quad (19)$$

A comparison of the three factors indicates that Factors 2 and 3 are associated with higher levels of social vulnerability, with more than 85% of the study area presenting values between 0.8 and 1.0 (Table 4). In contrast, Factor 1 reflects comparatively lower vulnerability levels, as more than 65% of the area exhibits values equal to or below 0.4 (Table 4). This pattern may be associated with the strong negative factor loading attributed to the White population (-0.882), which represents the predominant demographic group within the study area. Additionally, the negative loading associated with the elderly population (-0.787) also contributes to the lower vulnerability values observed for this factor.

The three factors were combined according to Equation (16) to obtain the SoVI. The SoVI was spatialized to generate the vulnerability map using two different methods for grouping the five classes: one using the standard deviation method (Figure 4) and the other using the equal-counting method (Figure 5).

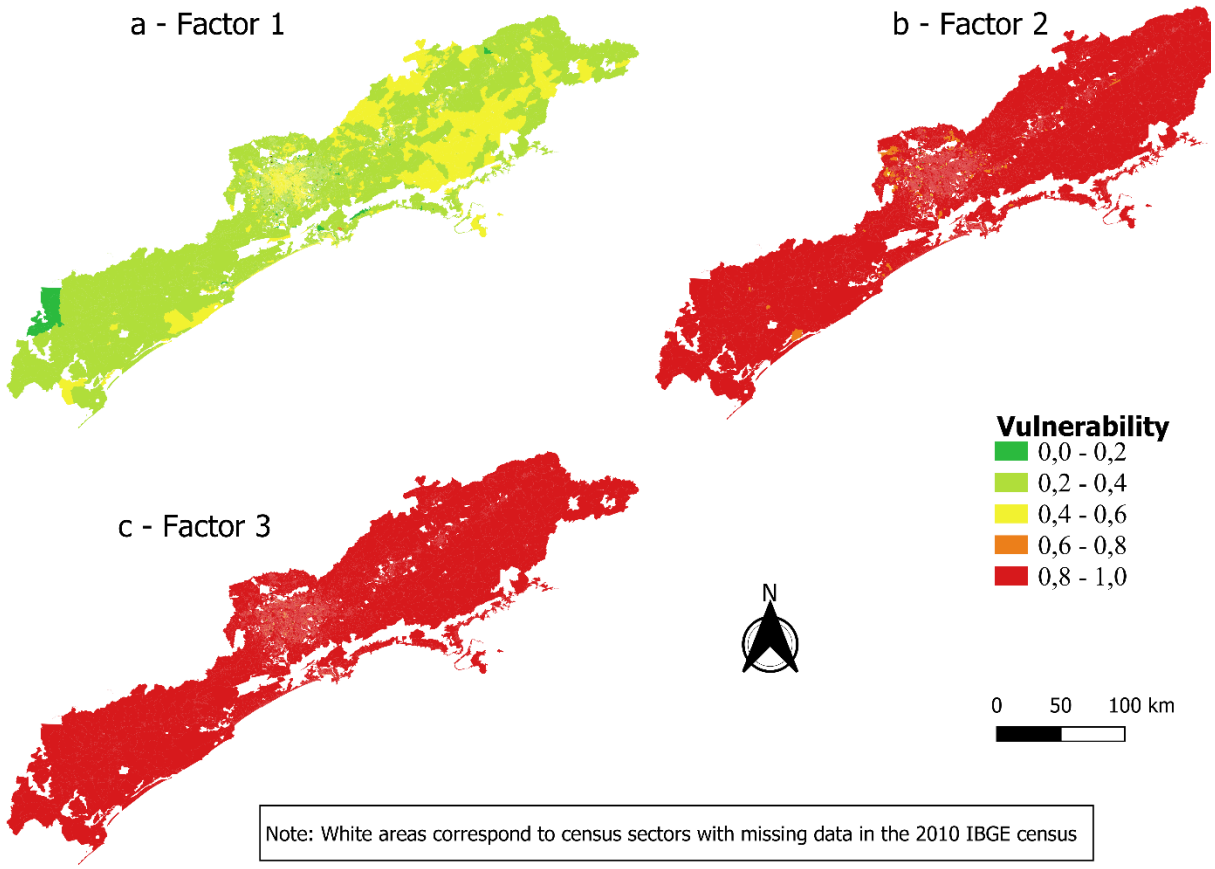

Figure 3 - Geographical distribution of the individual factors of the Social Vulnerability Index (SoVI). Source: Elaborated by the authors (2025).

Table 4 – Areas classified according to vulnerability level for each factor

Vulnerability	Factor 1		Factor 2		Factor 3	
	Area		Area		Area	
	Km²	%	Km²	%	Km²	%
0-0,2	653,98	1,69%	2,43	0,01%	0,03	0,00%
0,2-0,4	25.823,83	66,78%	4,40	0,01%	2,79	0,01%
0,4-0,6	7.797,94	20,16%	21,05	0,05%	9,10	0,02%
0,6-0,8	6,31	0,02%	427,56	1,11%	73,03	0,19%
0,8-1	0,05	0,00%	33.826,69	87,47%	34.197,18	88,43%
Unclassified	4.390,58	11,35%	4.390,59	11,35%	4.390,58	11,35%
In total	38.672,7	100,00%	38.672,7	100,00%	38.672,7	100,00%

Source: Elaborated by the authors (2025)

In the standard deviation method (Figure 4), the variation in the generated index is taken into account to classify areas into the five vulnerability classes more rigorously. This is because this classification method shows how the vulnerability values of a census tract compare to the average vulnerability value for the entire area. Thus, the map shows a large area classified as having high and very high vulnerability (66.13% and 5.69%, respectively), while the very low and low vulnerability classes are underrepresented on the map (Table 5).

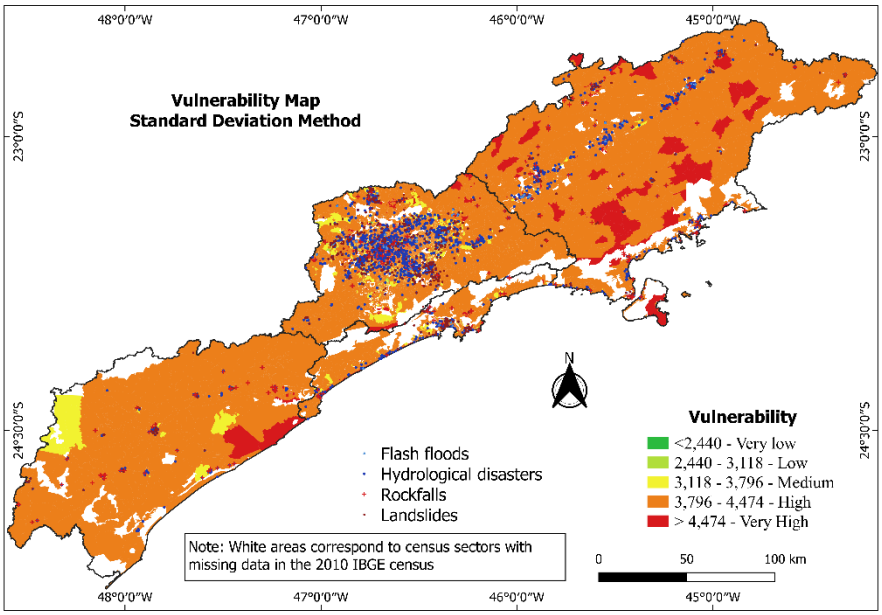

Figure 4 – Vulnerability map using the standard deviation method for the study area, by census tract. Source: Elaborated by the authors (2025)

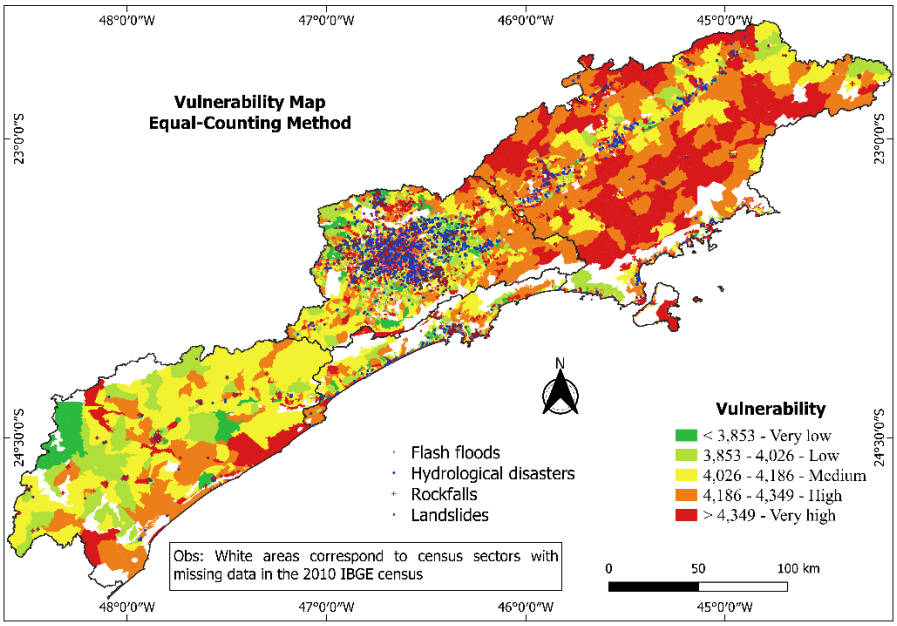

Figure 5 – Vulnerability map using the equal-counting method for the study area, by census tract. Source: Elaborated by the authors (2025)

Table 5 – Metrics for the developed maps in relation to the number of disasters and fatalities

Vulnerability	Area		Number of disasters		Fatalities	
	Standard deviation	Equal-counting	Standard deviation	Equal-counting	Standard deviation	Equal-counting
Very Low	0,01%	4,79%	0,03%	12,09%	0,00%	14,23%
Low	0,04%	12,98%	0,49%	18,49%	2,66%	12,52%

Medium	28,13%	30,07%	8,33%	21,69%	6,64%	13,66%
High	66,13%	30,03%	79,68%	20,78%	75,52%	26,57%
Very High	5,69%	22,13%	11,47%	26,95%	15,18%	33,02%

Source: Elaborated by the authors (2025)

The analysis, based on the overlay of the disaster inventory, revealed that most incidents occurred in regions classified as having high vulnerability (79.68%) and very high vulnerability (11.47%). In contrast, regions with low, very low, and medium vulnerability together accounted for only 8.85% of disaster occurrences (Table 5). These results suggest that the map is capable of adequately representing the region's vulnerability.

In the equal-count method (Figure 5), unlike the previous method that considers the SoVI, the vulnerability classes were distributed so that each class contained the same number of census tracts. However, this approach did not guarantee equal areas among classes, since census tracts vary in size. Consequently, areas classified as having very low and low vulnerability accounted for 4.79% and 12.98% of the study area, respectively. Areas with medium vulnerability represented 30.07%, while those classified as high and very high vulnerability accounted for 30.03% and 22.13%, respectively. Compared with the previous method, this map presents a much broader spatial distribution of vulnerability classes, whereas 66.13% of the area had previously been concentrated within a single vulnerability class (Table 5).

Regarding the number of disasters occurring in each vulnerability class, the equal-count method showed that areas classified as very low, low, and medium vulnerability accounted for 12.09%, 18.49%, and 21.69% of disaster occurrences, respectively. In addition, areas classified as high and very high vulnerability accounted for 20.78% and 26.95% of occurrences, respectively (Table 5). Although the very high vulnerability class recorded the largest proportion of occurrences, the distribution of disasters across the other classes was more widespread, which may indicate a limitation in the representativeness of this map.

This method also produced a map that differs from the vulnerability map proposed by Hummell et al.. Although an exact correspondence between the results was not expected, the objective was to generate a map that would corroborate the findings of those authors, which was achieved with the map presented in Figure 4. This outcome can be explained by the fact that the present study covers a smaller area and, consequently, provides a more detailed level of analysis, since it uses 2010 IBGE census data at the census-tract level rather than municipal-level data, as adopted in the previous study.

The standard deviation method was also evaluated by overlaying the number of fatalities associated with these disasters (Figure 6). During the analyzed period, the disasters resulted in 527 fatalities, of which 90.7% occurred in areas classified as having high or very high vulnerability (Table 5). Most disaster-related fatalities were concentrated in the Greater São Paulo region.

The distribution of fatalities was also analyzed using the equal-count method (Figure 7). Unlike the vulnerability map based on the standard deviation method, the equal-count method showed a less concentrated distribution of fatalities among the vulnerability classes. Although areas classified as high and very high vulnerability accounted for the largest proportion of fatalities (59.59%), regions classified as low and very low vulnerability also recorded a considerable share of fatalities (26.75%; Table 5). These results further support the conclusion that the standard deviation method is the most appropriate approach for developing the vulnerability map.

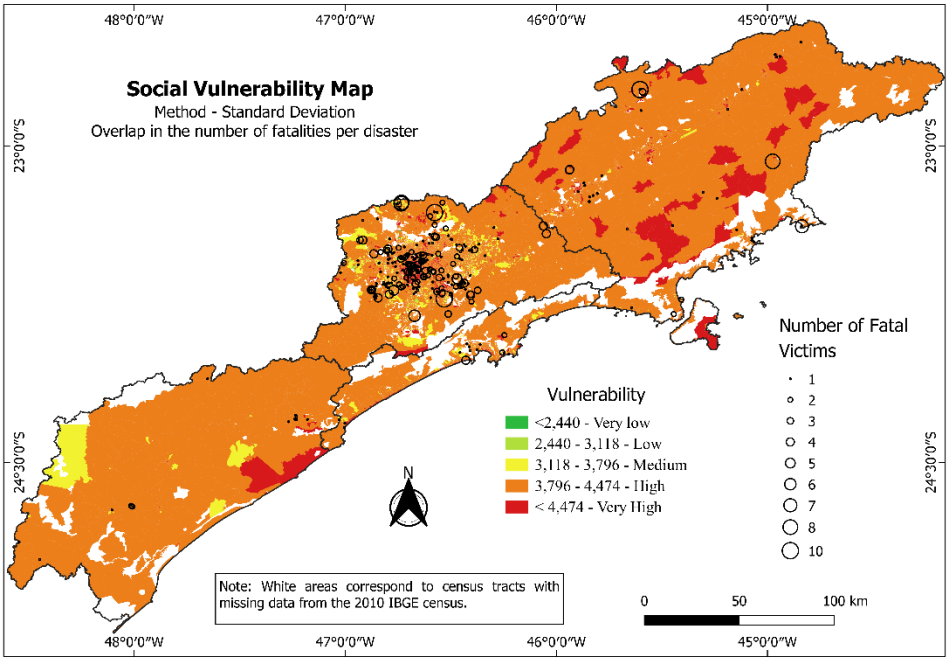

Figure 6 – Vulnerability map of the study area by census tract, developed using the standard deviation method and overlaid with the number of disaster-related fatalities.
Source: Elaborated by the authors (2025)

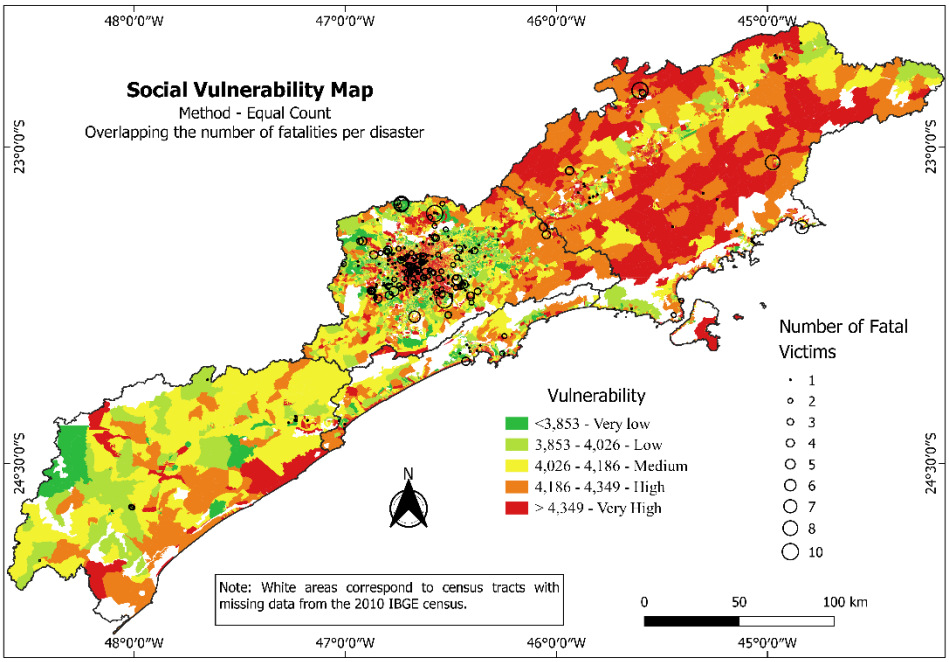

Figure 7 – Vulnerability map of the study area by census tract, developed using the equal-count method and overlaid with the number of disaster-related fatalities.
Source: Elaborated by the authors (2025)

5. Concluding Remarks

The present study aimed to identify areas with high social vulnerability to natural disasters within the administrative regions of São José dos Campos, Baixada Santista, and Registro. To achieve this objective, the analysis was based on data from the 2010 IBGE Demographic Census and employed Principal Component Analysis (PCA), a widely established methodological approach for vulnerability assessment.

The adopted methodology resulted in the development of two vulnerability maps. Among them, the map generated using the standard deviation method proved to be the most suitable for the analyzed areas, showing greater consistency with both previous studies and the inventory of natural disasters recorded in the region, unlike the map produced using the equal-count method.

The results indicated that the equal-count method introduced significant distortions between the mapped vulnerability levels and the actual spatial distribution of disasters. For example, in the Administrative Region of São José dos Campos, areas classified as having high and very high vulnerability did not correspond to the locations with the highest incidence of disasters, which, according to the inventory, were predominantly concentrated in areas classified as having medium vulnerability. Similarly, in the Administrative Region of São Paulo, the equal-count map classified several areas with a high incidence of disasters as having very low, low, and medium vulnerability.

In contrast, the map generated using the standard deviation method demonstrated greater accuracy by classifying these same areas as having medium and high vulnerability, thereby showing stronger agreement with the disaster inventory. Based on these findings, the authors strongly recommend the adoption of the standard deviation method for the spatial representation of vulnerability maps in the studied region.

The relevance of this assessment lies in its potential to support local and regional governments by providing an essential tool for strategically allocating financial resources and implementing targeted actions in the most vulnerable areas, thereby reducing the potential impacts and damages associated with climate change in the region.

Acknowledgments

The authors would like to thank CAPES for the doctoral fellowship awarded to the first author.

References

Alvino-Borba, A., Guerra, P. M., Moreira, L. A. G., Sacht, H. M., Almeida, J. A., & Mata-Lima, H. (2020). Desastres Naturais No Brasil E No Mundo: Uma Análise Holística Com Ênfase Nos Impactos Dos Eventos Hidrológicos E Meteorológicos/Natural Disasters in Brazil and Over the World: an Analysis Emphasizing Hydrological and Meteorological Events. *Brazilian Journal of Development*, *6*(9), 73718–73740. https://doi.org/10.34117/bjdv6n9-724

Arreguy, J. (2024, fevereiro 18). *Como está São Sebastião 1 ano após a tragédia provocada pelas chuvas*. Jornal Metrópoles.

Corrá, D. (2023). *Deslizamento de terra que devastou Caraguatatuba completa 50 anos*. G1 - Vale do Paraíba. https://g1.globo.com/sp/vale-do-paraiba-regiao/noticia/2017/03/deslizamento-de-terra-que-devastou-caraguatatuba-completa-50-anos.html

Cutter, S. L., Boruff, B. J., & Shirley, W. L. (2003). Social vulnerability to environmental hazards. *Social Science Quarterly*, *84*(2), 242–261. https://doi.org/10.1111/1540-6237.8402002

Diaz-Bonilla, C., Sabatino, C., Wu, H., & Nguyen, M. C. (2022). October 2022 Update to the Multidimensional Poverty Measure. *Global Poverty Monitoring Technical Note*, *26*(October). https://doi.org/10.1596/38203

Greve, P., Gudmundsson, L., & Seneviratne, S. I. (2018). Regional scaling of annual mean precipitation and water availability with global temperature change. *Earth System Dynamics*, *9*(1), 227–240. https://doi.org/10.5194/esd-9-227-2018

Guillard-Goncąlves, C., Cutter, S. L., Emrich, C. T., & Zêzere, J. L. (2015). Application of Social Vulnerability Index (SoVI) and delineation of natural risk zones in Greater Lisbon, Portugal. *Journal of Risk Research*, *18*(5), 651–674. https://doi.org/10.1080/13669877.2014.910689

Hader, P.R.P.; Reis, F.A.G.V.; Peixoto, A.S.P. Landslide risk assessment considering socionatural factors: methodology and application to Cubatão municipality, São Paulo, Brazil. **Natural Hazards** 110 (2), 1273-1304, 2022.

Hair, J. F., Anderson, R. E., & Tatham, R. L. (1987). *Multivariate data analysis with readings* (2º ed). Collier Macmillan.

Hair Jr., J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2014). Multivariate Data Analysis (MVDA). Em *Pharmaceutical Quality by Design: A Practical Approach* (7º ed). Pearson. https://doi.org/10.1002/9781118895238.ch8

Hummell, B. M. de L., Cutter, S. L., & Emrich, C. T. (2016). Social Vulnerability to Natural Hazards in Brazil. *International Journal of Disaster Risk Science*, *7*(2), 111–122. https://doi.org/10.1007/s13753-016-0090-9

Instituto Brasileiro de Geografia e Estatística (IBGE). (2016). *Metodologia do censo demográfico de 2010* (2º ed).

Instituto Brasileiro de Geografia e Estatística (IBGE). (2020). *Pesquisa nacional por amostra de domicíliuo contínua - 2019*.

IPCC. *Climate Change 2022: Impacts, Adaptation and Vulnerability*. Cambridge: Cambridge University Press, 2022.

Koks, E. E., Jongman, B., Husby, T. G., & Botzen, W. J. W. (2015). Combining hazard, exposure and social vulnerability to provide lessons for flood risk management. *Environmental Science and Policy*, *47*, 42–52. https://doi.org/10.1016/j.envsci.2014.10.013

Lei nº 9.394, de 20 de dezembro de 1996 (1996, 21 de dezembro). Estabelece as diretrizes e bases da educação nacional. Presidência da República. https://www.planalto.gov.br/ccivil_03/leis/l9 394.htm

Lei nº 10.741, de 1º de outubro de 2003 (2003, 2 de outubro). Dispões sobre o Estatuto da Pessoa Idosa e dá outras providências. Presidência da República. https://www.planalto.gov.br/ccivil _03/leis/2003/l10.741.htm#:~:text=LEI No 10.741%2C DE 1º DE OUTUBRO DE 2003.&text=Dispõe sobre o Estatuto do Idoso e dá outras providências.&text=Art.,a 60 (sessenta) anos.

Lei nº 10.097, de 19 de dezembro de 2000 (2000, 20 de dezembro). Altera dispositivos da Consolidação das Leis do Trabalho – CLT, aprovada pelo Decreto -Lei nº 5.452, de 1º de julho de 1943. Presidência da República. https://www.planalto.gov.br/ccivil_03/leis/l10097 .htm

Liu, X., Yue, Z. Q., Tham, L. G., & Lee, C. F. (2002). Empirical assessment of debris flow risk on a regional scale in Yunnan province, Southwestern China. *Environmental Management*, *30*(2), 249–264. https://doi.org/10.1007/s00267-001-2658-3

Marcelino, E. V., Nunes, L. H., & Kobiyama, M. (2006). Mapeamento De Risco De Desastres Naturais Do Estado De Santa Catarina. *Caminhos de Geografia*, *7*(17), 72–84.

Marques, J. M. da R., Queiroz de Lima, J. S., & Santos, J. D. O. (2020). Fragilidade ambiental, vulnerabilidade social e riscos de desastres no baixo curso do rio Maranguapinho, Fortaleza-Ceará-Brasil. *Territorium*, *27*(27(I)), 25–35. https://doi.org/10.14195/1647-7723_27-1_3

Nascimento JR, L., Sant'Anna Neto, J. L. (2020). Índice De Vulnerabilidade Social. *Revista de Geociências do Nordeste*, *6*, 8.

Neumayer, E., Plümper, T. (2007). The gendered nature of natural disasters: The impact of catastrophic events on the gender gap in life Expectancy, 1981-2002. *Annals of the Association of American Geographers*, *97*(3), 551–566. https://doi.org/10.1111/j.1467-8306.2007.00563.x

Pascoto, T. V., Rodrigues, J., & Peixoto, A. S. P. (2022). Carta de Suscetibilidade Geomorfológica a Movimentos de Massa para as Regiões Administrativas de São Paulo , São José dos Campos , Baixada Santista e Registro. *VIII Conferencia Brasileira sobre Estabilidade de Encostas*, *1*, 212.

Pellegrina, G. J. (2011). *Proposta de um procedimento metodológico para o estudo de problemas geoambientais com base em banco de dados de eventos atmosféricos severos* [Dissertação]. Universidade Estadual Paulista.

Pinho, G. M., Oliveira, L. C., Rocha, M. de M., & Barros, A. P. B. G. (2019). Mapeamento da Vulnerabilidade de Evacuação em Caso de Desastres Naturais Empregando a Sintaxe Espacial. *Revista Brasileira de Cartografia*, *71*(2), 328–366. https://doi.org/10.14393/rbcv71n2-45150

Saito, S. M. (2011). *Dimensão Socioambiental Na Gestão De Risco Dos Assentamentos Precários Do Maciço Do Morro Da Cruz, Florianópolis - Sc*. Universidade Federal de Santa Catarina.

Sandes, M. DOS S. **Mapeamento de risco de movimentos de massa em corredores rodoviários: estudo de caso na BR-116 entre os municípios de Jacareí e Queluz (SP)**. Dissertação (Mestrado). Universidade Estadual Paulista (Unesp), Faculdade de Engenharia de Bauru, 2023.

Torres, P. H. C., Gonçalves, D. A., Collaço, F. M. de A., Dos Santos, K. L., Canil, K., de Sousa Júnior, W. C., & Jacobi, P. R. (2021). Vulnerability of the São Paulo macro metropolis to droughts and natural disasters: Local to regional climate risk assessments and policy responses. *Sustainability (Switzerland)*, *13*(1), 1–16. https://doi.org/10.3390/su13010114

United Nations, UN. (2015). Sendai Framework for Disaster Risk Reduction 2015-2030. Em *Third United Nations World Conference on Disaster Risk Reduction*. Sendai.

United Nations Office for Disaster Risk Reduction (UNDRR). *Global Assessment Report on Disaster Risk Reduction 2022: Our World at Risk: Transforming Governance for a Resilient Future*. Geneva: UNDRR, 2022.

USAID. (2023). *Disasters year in Review 2022*. *70*, 2.