Universidade Estadual de Maringá
Departamento de Estatística
Programa de Pós-graduação em Bioestatística

ÍCARO DA COSTA FRANCISCO

The Urban Forest at Risk: Unveiling Windstorm-Induced Tree Fall Patterns through Spatial and Machine Learning Analyses in the City of Maringá, Brazil

Maringá – PR
2024

ÍCARO DA COSTA FRANCISCO

The Urban Forest at Risk: Unveiling Windstorm-Induced Tree Fall Patterns through Spatial and Machine Learning Analyses in the City of Maringá, Brazil

Dissertação apresentada ao Programa de PósGraduação em Bioestatística do Centro de Ciências Exatas da Universidade Estadual de Maringá, como requisito parcial para a obtenção do título de Mestre em Bioestatística.

Orientador: Prof. Dr. Diogo Francisco Rossoni, Universidade Estadual de Maringá
Membro externo: Prof. Dr. João Domingos Scalon, Universidade Federal de Lavras.
Membro Interno: Prof. Dr. Luciano de Andrade, Universidade Estadual de Maringá

Universidade Estadual de Maringá - UEM

Departamento de Estatística - DES

Programa de Pós-Graduação em Bioestatística

Maringá-PR

2024

Dados Internacionais de Catalogação-na-Publicação (CIP)
(Biblioteca Central - UEM, Maringá - PR, Brasil)

Francisco, Ícaro da Costa

F818u

The urban forest at risk : unveiling windstorm-induced tree fall patterns through spatial and machine learning analyses in the city of Maringá, Brazil / Ícaro da Costa Francisco. -- Maringá, PR, 2024.

37 f. : il. color., tabs., mapas

Orientador: Prof. Dr. Diogo Francisco Rossoni.

Dissertação (mestrado) - Universidade Estadual de Maringá, Centro de Ciências Exatas, Departamento de Estatística, Programa de Pós-Graduação em Bioestatística, 2024.

1. Queda de árvores - Maringá (PR). 2. Estatística espacial. 3. Árvores de regressão. 4. Arborização urbana - Maringá (PR). 5. Bioestatística. I. Rossoni, Diogo Francisco, orient. II. Universidade Estadual de Maringá. Centro de Ciências Exatas. Departamento de Estatística. Programa de Pós-Graduação em Bioestatística. III. Título.

CDD 23.ed. 519.5

Síntique Raquel Eleutério - CRB 9/1641

ÍCARO DA COSTA FRANCISCO

The Urban Forest at Risk: Unveiling Vendaval-Induced Tree Fall Patterns through Spatial and Machine Learning Analyses in the City of Maringá, Brazil

Dissertação apresentada ao Programa de PósGraduação em Bioestatística do Centro de Ciências Exatas da Universidade Estadual de Maringá, como requisito parcial para a obtenção do título de Mestre em Bioestatística.

BANCA EXAMINADORA

Prof. Dr. Diogo Francisco Rossoni

Universidade Estadual de Maringá – UEM

Documento assinado digitalmente
LUCIANO DE ANDRADE
Data: 24/06/2024 09:39:53-0300
Verifique em https://validar.iti.gov.br

Prof. Dr. Luciano de Andrade

Universidade Estadual de Maringá – UEM

Documento assinado digitalmente
JOAO DOMINGOS SCALON
Data: 20/06/2024 10:55:34-0300
Verifique em https://validar.iti.gov.br

Prof. Dr. João Domingos Scalon

Universidade Federal de Lavras – UFLA

Maringá, 12 de março de 2024

AGRADECIMENTOS

Gostaria de expressar minha profunda gratidão aos meus pais, Leonildo e Albanisa, e aos meus irmãos, Maria Rita e Hugo Cesar, pelo constante incentivo aos estudos, pelo apoio incondicional ao longo desses anos de mestrado e por terem sido fundamentais na minha formação como ser humano.

Agradeço aos meus amigos Rodrigo, Stéfani e Marcos da Silva Medeiros, pelo suporte acadêmico e pessoal e pela amizade durante esses anos.

Às professoras e colegas de trabalho, Professora Janete, Professora Nancy e Professora Robsmeire, por sempre apoiarem minhas decisões, incentivarem minha trajetória acadêmica e por serem como segundas mães para mim.

Um agradecimento especial ao Professor Dr. Oséias da Silva Matinuci, do Departamento de Geografia da UEM, por sua orientação durante a graduação, pelo papel crucial no meu desenvolvimento intelectual e acadêmico, e por me apoiar na decisão de explorar a bioestatística.

Ao meu orientador, Professor Diogo Francisco Rossoni, pela liberdade concedida na condução desta dissertação, pelo suporte no desenvolvimento do meu mestrado, pela orientação no campo da Estatística Espacial e, acima de tudo, por acreditar no meu sucesso em todos os momentos.

Minha gratidão também à coordenação e ao corpo docente do Programa de Pós-Graduação em Bioestatística, pelo conhecimento transmitido, pela dedicação e excelência que fazem deste programa um destaque entre os demais.

Agradeço ao Instituto Ambiental de Maringá e a todos os seus funcionários, especialmente ao Maurício, por disponibilizarem os dados essenciais para a realização deste trabalho. Estendo meus agradecimentos a Rafaela e Clayton pela organização dos dados brutos utilizados nesta pesquisa.

RESUMO

A arborização urbana desempenha um papel crucial na manutenção da segurança e resiliência dos ambientes urbanos, mas compreender as dinâmicas espaciais e os fatores subjacentes aos incidentes de queda de árvores continua sendo um desafio complexo. Neste estudo, realizamos uma análise abrangente dos incidentes de queda de árvores em Maringá, Paraná, Brasil, de 2015 a 2021, utilizando a estimativa de densidade de kernel, análise da função L não-homogênea e modelagem de árvore de regressão. Nossas descobertas revelam padrões espaciais intrigantes, com maiores concentrações de incidentes nas regiões norte e nordeste da cidade. Além disso, identificamos mudanças dinâmicas nas distribuições espaciais ao longo do tempo, enfatizando a necessidade de planejamento urbano proativo e estratégias de gerenciamento de riscos. A análise de árvore de regressão destacou fatores meteorológicos como contribuintes significativos para as quedas de árvores, fornecendo apontamentos para esforços de mitigação de riscos. No geral, nosso estudo contribui para uma melhor compreensão das dinâmicas espaciais dos incidentes de queda de árvores e defende métodos padronizados de coleta de dados e o desenvolvimento de ferramentas para aprimorar o gerenciamento da arboricultura urbana e promover ambientes urbanos mais seguros.
Palavras-chave: Queda de árvores. Estatística Espacial. Função L Não Homogênea. Árvores de Regressão.

ABSTRACT

Urban forestry plays a crucial role in maintaining the safety and resilience of urban environments, yet understanding the spatial dynamics and underlying factors of tree fall incidents remains a complex challenge. In this study, we conducted a comprehensive analysis of tree fall incidents in Maringá, Paraná, Brazil, from 2015 to 2021, using kernel density estimation, inhomogeneous L function analysis, and regression tree modeling. Our findings reveal intriguing spatial patterns, with higher concentrations of incidents in the northern and northeastern regions of the city. Moreover, we identified dynamic changes in spatial distributions over time, emphasizing the need for proactive urban planning and risk management strategies. Regression tree analysis highlighted meteorological factors as significant contributors to tree falls, providing actionable insights for risk mitigation efforts. Overall, our study contributes to a better understanding of the spatial dynamics of tree fall incidents and advocates for standardized data collection methods and the development of tools to enhance urban forestry management and promote safer urban environments.

Keywords: Tree Fall. Spatial statistics. Inhomogenous L function. Regressions Tree.

LIST OF FIGURES

LIST OF TABLES

CONTENTS

1. INTRODUCTION

Urban trees provide a myriad of benefits within urban landscapes. They significantly enhance the quality of life by imparting climatological benefits, mitigating the urban heat island effect, and reducing rainwater runoff. Additionally, trees contribute to the urban environment through financial, ornamental, and social advantages (van Haaften, et al., 2021).

However, despite the numerous advantages trees offer, the occurrence of tree falls presents a significant challenge, resulting in material damages and service disruptions in many cities worldwide.

Tree falls in urban areas are a global issue affecting cities from Hong Kong (Jim & Zhang, 2013) to Germany (Pretzsch, et al., 2015), from Brazil (Manfra, Massoca, Uras, Cavalari, & Locosselli, 2022) to Iran (Shabani & Akbarinia, 2017). As cities expand and the concern for ecological initiatives and urban greening to enhance the quality of life for residents increases, understanding the challenges associated with urban trees becomes paramount. Urban trees face various pressures not encountered in non-anthropogenic environments, with a probability of surviving beyond 35 years standing at only 35.1% (Smith, Dearborn, & Hutyra, 2019), making their life cycle relatively short. A total of 161 articles and studies have been published with the aim of studying this issue.

Scientific literature highlights wind and fungal decay as primary contributors to tree falls (van Haaften, et al., 2021) (Schmidt, Gaiser, & Dujesiefken, 2011). In their 2021 systematic review and meta-analysis, van Haaften et al. reported that, out of a total of 161 studies, 126 identified wind as the primary cause and 12 identified fungal decay.

These tree falls can result in significant material and economic losses, as well as human fatalities (Schmidlin, 2008). Understanding why and where these trees fall is crucial for the development of intervention measures focused on solving or mitigating this problem, aiming for better public policy. The “why” allows us to comprehend the reasons behind these tree falls, while the “where” points out the locations of the falls, enabling us to determine where we can take action and focus our efforts.

As is typical in many cities with abundant tree cover, Maringá faces challenges related to tree falls. Located in the southern region of Brazil, Maringá was honored as a “World Tree City” by the FAO-UN and the Arbor Day Foundation for the second consecutive year in 2023 (The Arbor Day Foundation, 2023). Globally, only 170 cities hold this prestigious designation, with the majority situated in North America and Europe. Maringá stands out as one of the 21 recognized cities in Brazil, alongside São Paulo and

Rio de Janeiro. Together, these cities have a combined population of 26,278,913 residents.

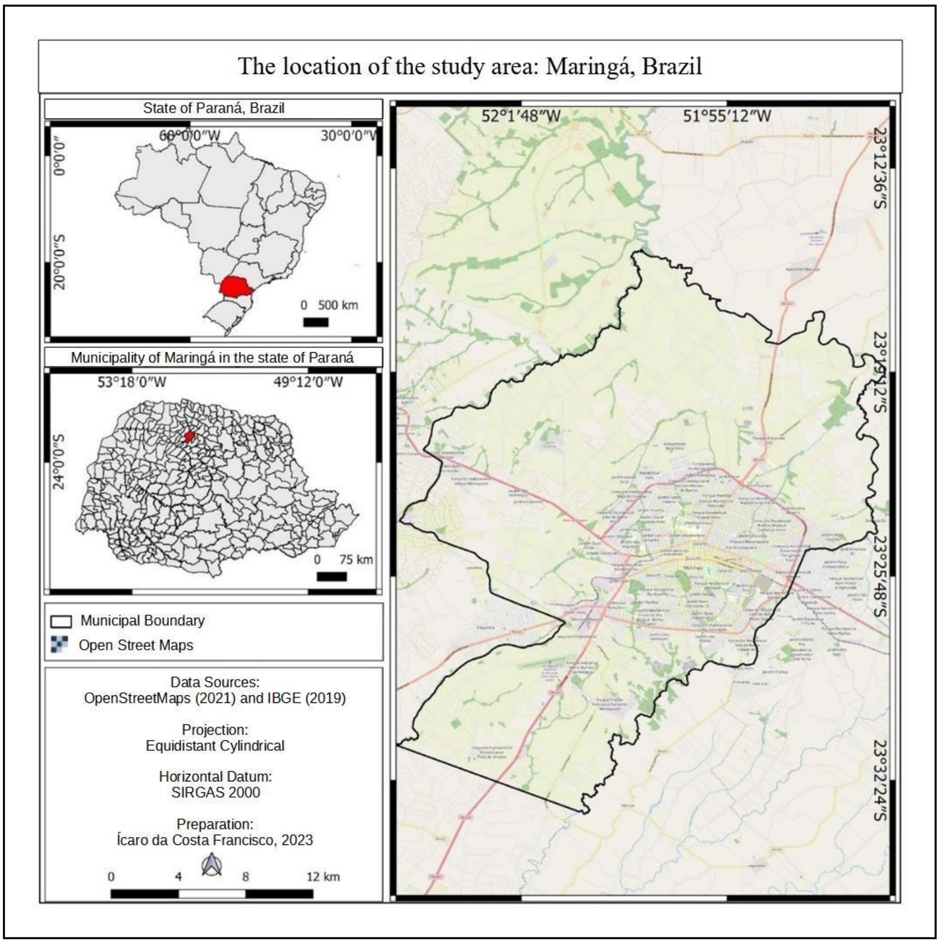
The occurrence of tree falls in Maringá presents a multifaceted problem that affects urban infrastructure and citizen welfare. Notably, between 2015 and 2021, there were 2,339 reported incidents of tree falls, with the most severe event occurring on September 14, 2021. This event affected 339 trees and resulted in widespread power outages, as reported by local authorities and local newspapers (RPC Maringá, 2021). This study aims to analyze the spatial patterns of tree fall incidents in Maringá, focusing on their distribution across the city and the contributing factors during the years under investigation. Statistical techniques will aid in identifying spatial patterns of tree fall incidents, while machine-learning algorithms will facilitate understanding the factors leading to these incidents. By comprehending these patterns and factors, policymakers and urban planners can devise effective mitigation strategies to minimize the impact of tree falls on urban infrastructure and public safety.

2. METHODOLOGY

2.1. Study Area

The city of Maringá is located in the northwest region of the state of Paraná, Brazil (Figure 1), within the phytogeographic domain of the Atlantic Forest and the phytoecological region of the Semideciduous Seasonal Forest (Zeidan & Ferreira, 2023). Maringa's climate is classified as subtropical, characterized by rainy summers and dry winters (Minaki, 2021). Founded as a municipality in 1951, Maringá is characterized by its urban planning inspired by the project conceived by Jorge de Macedo Vieira, based on the garden city proposals published by Ebenezer Howard in 1902 (Rego, 2001). With a population of 409,657 inhabitants and a high urbanization rate exceeding 90% [14], the city has approximately 163,000 trees in its urban landscape (The Arbor Day Foundation, 2023).

Figure 1: Maringá, Paraná, Brazil



Source: Open Street Maps (2021) and IBGE (2022)

2.2. Database

In this study, we utilized a dataset comprising 2,339 tree fall reports sourced from the Environmental Institute of Maringá (IAM). These reports included essential information such as address or geographical coordinates, and other details described in the table below (Table 1).

Table 1: Variables of the database

Variable	Code	Description
Identification	ID	Individual identification number
Date	Date	Date of the record
Species	SPECIES	Common name of the species
Address	Address	Address with Street Name, Number, and Neighborhood of the record
Latitude (°)	Lat	Latitude of the fall record measured in degrees
Longitude (°)	Long	Longitude of the fall record measured in degrees
Precipitation (mm)	PRECIPITATION	Accumulated precipitation in millimeters on the day of the record
Atmospheric Pressure (hPa)	PRESSURE AT	Atmospheric pressure in hectopascals recorded on the day
Maximum Temperature (°C)	MAX TEMPERATURE	Maximum temperature in degrees Celsius recorded on the day
Average Temperature (°C)	AV TEMPERATURE	Average temperature in degrees Celsius recorded on the day
Minimum Temperature (°C)	MIN TEMPERATURE	Minimum temperature in degrees Celsius recorded on the day
Maximum Wind Gust (m/s)	MAX WIND GUST	Maximum wind speed in meters per second recorded on the day
Average Wind Speed (m/s)	AV WIND SPEED	Average wind speed in meters per second on the day of the record
Neighborhood Age	NEIGHBORHOOD AGE	Age of the neighborhood where the tree is located

Source: Environmental Institute of Maringá (2021)

When geographical coordinates were absent, we undertook geocoding efforts to determine the tree fall locations. Between the years 2015 and 2021, 1,481 reports were successfully geocoded, representing approximately 63.32% of the dataset (Table 2).

Unfortunately, the spatial information for the remaining data was lost due to insufficient locational details, such as imprecise addresses, missing numbers, and incorrectly written street names.

Table 2: Geocoding Data Loss Percentage per year - 2015 to 2021

Year	Reports	Geocoded Reports	Missing data (%)
2015	99	51	48.48
2016	226	133	41.15
2017	419	216	48.45
2018	659	348	47.19
2019	228	109	52.19
2020	203	174	14.29
2021	505	450	10.89
2015-2021	2,339	1,481	36.68

Source: Author's own creation (2024)

Two distinct approaches were employed to analyze the spatial patterns of tree falls and the factors that contribute to their occurrence. The first approach involved understanding two spatial properties of the data: the first-order properties and the second-order properties. For this approach, the R Package 'spastats' (Baddeley & Turner, spatstat: An R Package for Analyzing Spatial Point Patterns, 2005) was utilized, which is available within the Comprehensive R Archive Network (CRAN). In the second approach, we used Regression trees to determine which factors contributed to these falls. For this purpose, the 'caret' package in R was employed (Kuhn, 2008). For the cartographic representations of the results, we employed the open Geographic Information System (GIS) software QGIS 3.28 ‘Firenze’ (QGIS Development Team, 2022).

The following section will provide a detailed explanation of these approaches, outlining the specific methodologies employed for each analysis. Through systematic application of these techniques, our goal is to thoroughly reveal the spatial patterns that characterize the occurrence of tree falls in Maringá and identify the contributing factors.

2.3. First approach – Spatial Analysis:

In the analysis of point patterns, a crucial focus lies on the precise location of phenomena, characterized by a pair of coordinates. This analytical approach centers on understanding the distribution of observed events and drawing inferences about the underlying processes generating them (Bivand, Pebesma, & Gómez-Rubio, 2013). To effectively unravel this, two key properties of point patterns are studied: first-order properties and second-order properties.

First-order properties measure the distribution of an event in the study area, involving the determination of the occurrence intensity of the process (Cressie, 1993).On the other hand, second-order properties provide information about the interaction between two points, indicating the tendency of events to be distributed in clustered, random, or regularly spaced patterns. It measures the spatial structure and types of interaction among events in point processes.

Some studies stand out for a spatial approach, seeking to identify where tree falls occur most frequently and mapping areas with higher risk. Among them, we highlight the work of Ribeiro and Lopes (2011) which aimed to map risk of tree fall in Lisbon, Portugal. Another relevant study is Shabani & Akbarinia (2017) investigation into spatial patterns of tree falls using logistic regression and machine learning for the Iranian province of Mazandran. This analysis primarily utilized variables from the physical environment to examine these patterns. Overall, these studies did not focus on or address the first and second-order properties of the spatial distribution of tree falls.

2.1.1. First Order Properties

2.1.1.1. Estimation of intensity through Kernel Smoothing

To identify first-order properties, it is necessary to estimate the intensity function. For this purpose, we employed the quartic kernel smoothing method. The kernel smoothing estimator used in this work is represented by the following equation (Bivand, Pebesma, & Gómez-Rubio, 2013):

$$\hat{\lambda}(x) = \frac{1}{h^2} \sum_{i=1}^{n} k \left(\frac{||x - x_i||}{h} \right) / q(||x||)$$

Where:

$k =$ *bivariate and symmetrical kernel function.*

$q =$ *border corrector used to account for missing observations near the border.*

$h =$ *bandwidth parameter estimated by the MSE.*

The edge correction method utilized was the Jones-Diggle improved edge correction (Jones, 1993). The bandwidth parameter was estimated using the Mean Square Error

(MSE). The kernel function used was the quartic kernel, whose expression in two dimensions is (Bivand, Pebesma, & Gómez-Rubio, 2013):

$$k(u) = \begin{cases} \frac{3}{\pi}(1 - \|u\|^2)^2 & if\ u \in (-1,1) \\ 0 & Otherwise \end{cases},$$

Where:

$\|u\|^2$ $is\ the\ squared\ norm\ of\ point\ u = (u_1, u_2)\ equal\ to\ u_1^2 + u_2^2.$

The kernel smoothing process enables the creation of a smooth intensity surface that represents the density of events across the study area. By applying this estimator, we effectively analyze the spatial distribution of tree fall incidents, revealing areas of high and low intensity.

Furthermore, the intensity was computed both for the entire study period and individually for each year. The resulting data were transformed into raster format and imported into QGIS 3.28 for the generation of cartographic products. This visualization approach aids in the portrayal of the spatial intensity variation over the years and offers a comprehensive view of the tree fall incident distribution across Maringá.

2.1.2. Second Order Properties

To estimate second order properties, two local functions were employed, the Inhomogenous K function and the Inhomogenous L function. n. We chose these inhomogeneous functions because the analysis of first order identified the spatial non-homogeneity of this distribution and the existence of second-order effects.

2.1.2.1. Inhomogenous K Function (K_{inhom}):

The Inhomogeneous K function ($\hat{K}_{inhom}$) is a generalization of the homogeneous K function, and its estimator, with the edge correction, was proposed as (Baddeley, Moller, & Waagepetersen, 2000):

$$\hat{K}_{inhom}(t) = \frac{1}{|A|}\sum_{i=1}^{n}\sum_{(j\neq i)=1}^{n}\frac{I_t(t_{ij})}{w_{ij}\hat{\lambda}(u_i)\hat{\lambda}(u_j)},\ \ t \geq 0,$$

Where:

$|A| =\ the\ area\ of\ region\ A$

$n = number\ of\ observed\ occurrences$

$w_{ij} = the\ correction\ factor\ for\ border\ effects$

$I_t(t_{ij}) = an\ indicator\ function\ that\ assumes\ 1\ when\ t_{ij}\ \leq t$

The w_{ij} corresponds to the proportion of the circumference of the circle centered on event u_i containing u_j that is within the study region $|A|$.

2.1.2.2. Inhomogenous L Function (L_{inhom}):

The estimator of the ($\hat{L}_{inhom}$) is obtained through the transformation of the $\hat{K}_{inhom}$. The formula for the estimator of the $\hat{L}_{inhom}$ from the $\hat{K}_{inhom}$, for a distance h, is expressed as:

$$\hat{L}_{inhom}(h) = \sqrt{\frac{\hat{K}_{inhom}(h)}{\pi}}.$$

The estimated $\hat{L}_{inhom}$ is obtained from the linearization of the $\hat{K}_{inhom}$, and to facilitate the understanding of the $\hat{L}_{inhom}$ outcomes, we present an illustrative model that aids in their interpretation within the context of spatial point pattern analysis (Figure 2).

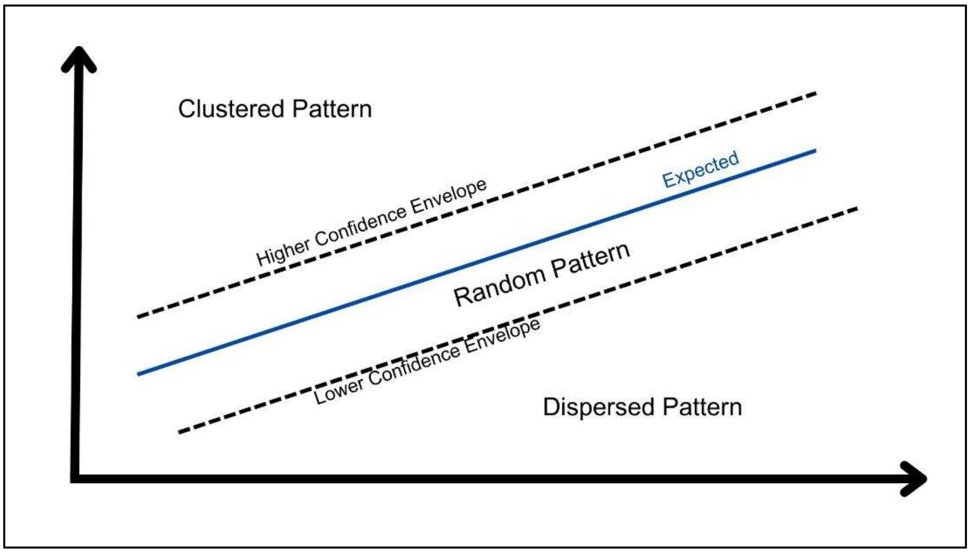
The functions in question evaluate the hypothesis of complete spatial randomness, which requires the spatial distribution of the data to be random. To test this hypothesis, it is necessary to conduct a Monte Carlo test, based on random samplings to obtain results. In the Monte Carlo test used in this study, 1,000 permutations were performed to construct an envelope containing the smallest and largest values for each point of the calculated function. Tests against the null hypothesis of complete spatial randomness within the $\hat{K}_{inhom}$ and $\hat{L}_{inhom}$ functions were conducted at a significance level of 0.002, a value corresponding to the number of permutations defined previously.

When the calculated $\hat{L}_{inhom}$ function value exceeds the upper envelope, suggests that the distribution of events in the spatial region is clustered. This signifies a cluster process, indicating that events are more closely packed together than would be expected under a random distribution. Such a pattern could indicate localized areas of high event occurrence. If the calculated value falls below the lower envelope, it signifies a regular pattern. This indicates a dispersed process, suggesting that events are more evenly spaced than expected under a random distribution. A regular pattern could imply a deliberate arrangement or a uniform distribution of events across the study area.

When the calculated values of the $\hat{L}_{inhom}$ function fall within the envelope, it indicates a random pattern, providing evidence for the null hypothesis of Complete

Spatial Randomness. In this scenario, events are distributed without any discernible clustering or regularity.

Figure 2: L Function graphic model



Source: Author's own creation (2023)

By analyzing the calculated values in relation to the envelope, we can ascertain whether the observed pattern of events leans towards clustering, dispersion, or randomness. This insightful model aids in comprehending the spatial arrangements of events and guides the interpretation of the $\hat{K}_{inhom}$ and $\hat{L}_{inhom}$ results.

In our study, this approach will be applied to discern the spatial characteristics of tree fall incidents in Maringá. By comparing the calculated values with the envelope, we will gain a deeper understanding of the underlying patterns and inform the formulation of effective mitigation strategies to address the tree fall problem within the urban landscape.

2.4. Second approach – Machine Learning Applications:

This approach was based on a study conducted in the city of São Paulo in 2022 (Manfra, Massoca, Uras, Cavalari, & Locosselli, 2022) which investigated the factors leading to tree falls in the city. This study employed a machine learning model known as Regression Trees, utilizing variables from both the natural and built environment, such as sidewalk width, building height, and neighborhood age. In our work, the Regression Tree technique was applied to identify the factors leading to tree falls during windstorms in Maringá. Some adaptations were made, with a greater focus on meteorological variables, as anthropogenic variables such as building height and sidewalk width were not available.

The Regression Tree method is a supervised learning technique used to construct predictive models from a dataset. Regression trees select splits that reduce variance and are divided into branches, nodes, and leaves. Target attribute values can be predicted from the mean value of the leaves to assess the quality of fit and prediction. In this algorithm, the response variable was defined as “Windstorm,” indicating whether the tree fell on a day with a recorded windstorm, with more than 30 falls on the same day. For the analysis, the database was split, with 20% of the data used for algorithm testing and 80% for training. The explanatory variables included in the model were the tree species, accumulated precipitation for the day, atmospheric pressure, maximum, minimum, and average temperature of the day, maximum wind gust, average wind gust, the age of the neighborhood where the tree is located, and the city region.

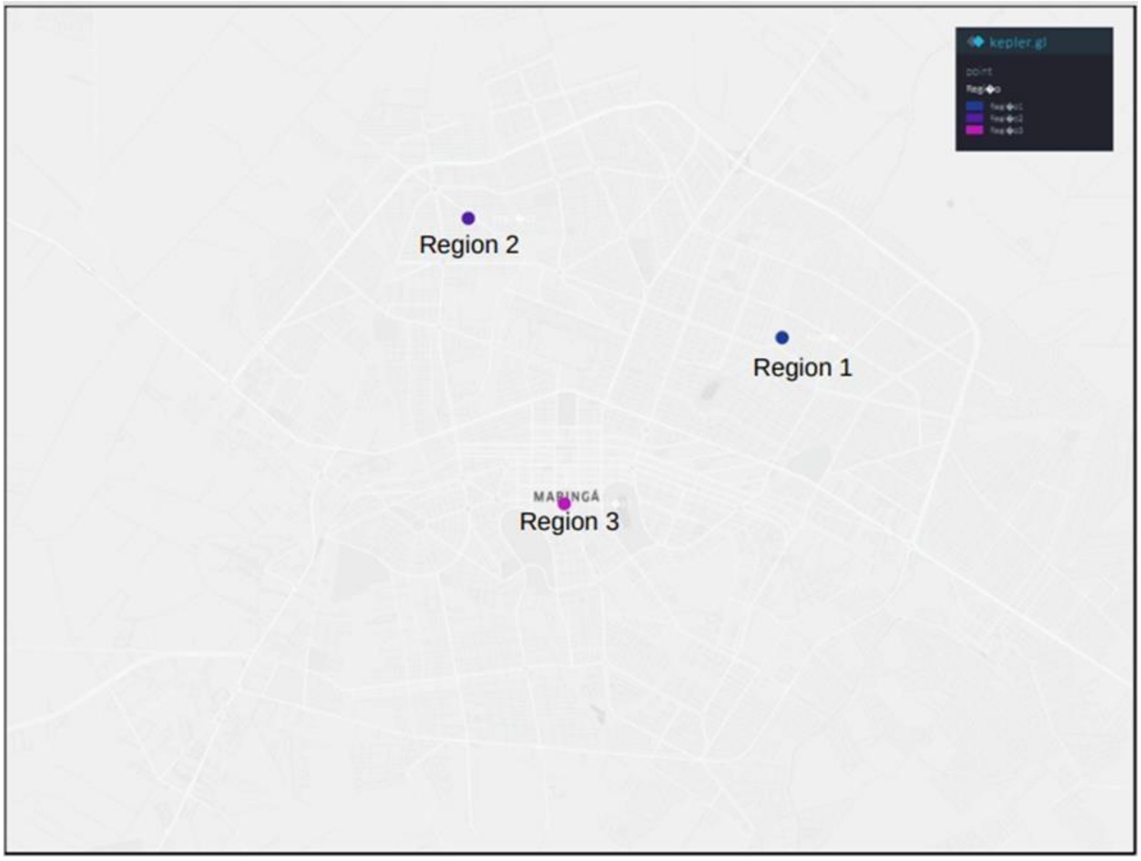
Since we did not have the age of the trees, we used the age of the neighborhoods where they were located, as there is no city-wide policy or plan for tree replacement. In most cases, trees are planted at the beginning of the neighborhood construction process.

For this purpose, the tree fall records were also grouped into three regions of the city. The criterion for grouping was the Haversine distance, a mathematical formula that calculates the distance between two points on the Earth’s surface using their latitude and longitude coordinates. The Haversine formula assumes the Earth is a perfect sphere and is based on the law of sines, considered one of the most accurate ways to calculate distance between two points on Earth. It is defined by the following equation (Azdy & Darnis, 2020):

$$d = 2 * R * \arcsin\left(\sqrt{\sin^2\left(\frac{\Delta lat}{2}\right) + \cos(lat_1) * \cos(lat_2) * \sin^2\left(\frac{\Delta long}{2}\right)}\right)$$

The two points considered were the coordinates of incidents and fall records, and one of the three delimited points in the city, as we can see below in figure 3.

Figure 3: Coordinates of the central points for the three regions



Region 1 (-23.403728443766564, -51.90655461940291) comprised records in the northeast region of the city. Region 2 (-23.38765853320574, -51.95255986574726) comprised records in the north and northwest region, and Region 3 (-23.426097031914328, -51.93848363365684) comprised records in the central region. he points for the construction of these three regions were chosen because they were the centroids of the regions where there were concentrations of tree falls in the city.

Two models were developed: Model I considering all variables and all reports; and Model II, considering only incidents recorded in regions 1 and 2. This was done to verify if these two regions stood out due to having the highest number of falls and if they exhibited a different behavior. To analyze the quality of these models, a confusion matrix will be constructed, and accuracy, precision, recall and specificity will be calculated for each model. The accuracy of a machine learning classification algorithm shows how often the model classifies a data point correctly.

$$\text{Accuracy} = \frac{Number\ of\ correct\ predictions}{Total\ number\ of\ predictions}$$

The precision represents the proportion of positive identifications that were actually correct.

$$\text{Precision} = \frac{True\ positives}{True\ positives + \ False\ positives}$$

The recall measures the proportion of positive identifications that were actually correct.

$$\text{Recall} = \frac{True\ positives}{True\ positives + \ False\ negatives}$$

In addition, specificity is described as the algorithm/model's ability to predict a true negative of each category available.

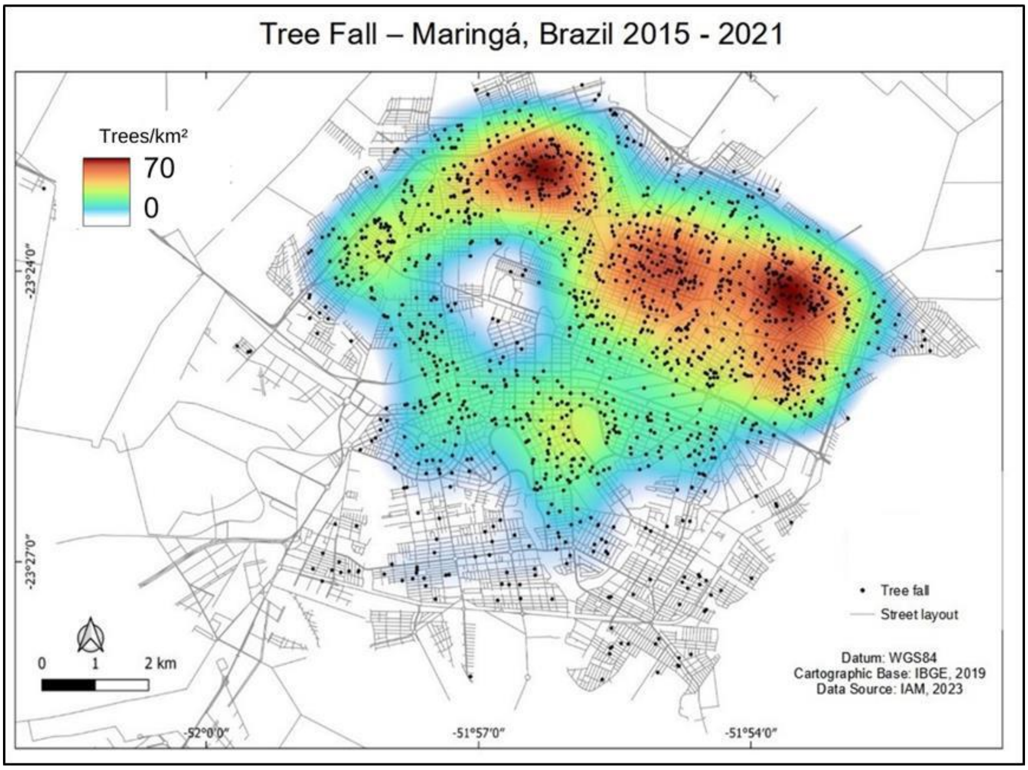
$$\text{Specificity} = \frac{True\ negatives}{True\ negatives\ +\ False\ positives}$$

3. RESULTS

3.1. Intensity Estimated

For the interval 2015-2021, the bandwidth calculated by the MSE was 250m.When analyzing this interval, the kernel analysis unveils a spatial distribution pattern wherein fewer reports occurred in the southern region of the city, while a higher number of reports were concentrated in the northern region (Figure 4). This spatial trend is particularly pronounced, with the most significant cluster of incidents visibly concentrated in the northern area, reaching intensities of up to o 70 incidents per square kilometer.

Figure 4: Tree fall - intensity estimation 2015-2021



Source: Author's own creation (2023)

The optimal bandwidth values that minimize the mean square error for the kernel density estimation for each year are as follows (Table 3):

Table 3: Calculated Bandwidth 2015-2021

Year	Calculated Bandwidth (h)
2015	390 m
2016	760 m
2017	890 m
2018	400 m
2019	530 m
2020	1910 m
2021	230 m
2015 - 2021	250 m

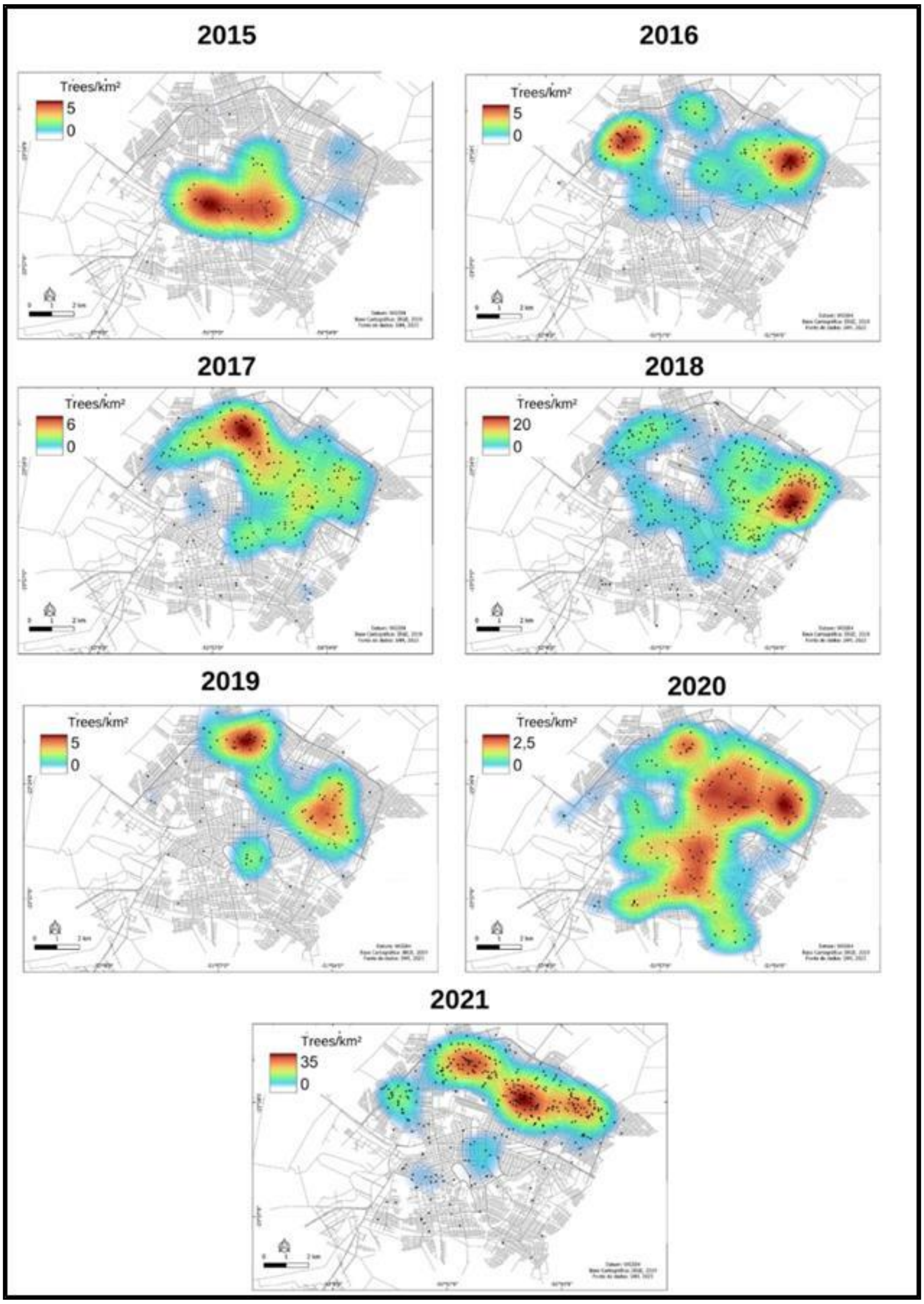
Source: Author's own creation (2024)

Over the span of several years, incidents of fallen trees in the city exhibited intriguing patterns (Figure 5). In 2015, most of these incidents occurred in the central part of the city, with an intensity of approximately 5 fallen trees per square kilometer, while the southern region remained incident-free. The subsequent year, 2016, witnessed a shift, with incidents concentrating primarily in the northern area. This time, two separate clusters emerged, one in the northeast and another in the northwest, solidifying the northern region as the focal point for these occurrences, with an intensity of approximately 5 incidents per square kilometer.

Moving to 2017, there was only one notable cluster with an intensity of 6 incidents per square kilometer, with most of the incidents occurring in the northern region of the city. Then, in 2018, the northeastern part of the city became the primary location for tree fall incidents, with an intensity of 20 fallen trees per square kilometer. The trend continued into 2019, where most incidents clustered once again in the northern area, accompanied by fewer reports in the southern region and a small cluster within the city center.

The year 2020 marked a noteworthy development as tree fall incidents spread across the entire city. Clusters of incidents occurred in both the northern and central areas, blurring the previously distinct patterns. This year also recorded the lowest intensity, registering 2.5 incidents per square kilometer in the areas of highest intensity. Finally, in 2021, the northern part of the city exhibited three distinct incident clusters with intensities exceeding 35 fallen trees per square kilometer, while the southern region experienced relatively fewer occurrences. These changing patterns over the years highlight the dynamic nature of tree fall incidents within different regions of the city.

Figure 5: Tree fall - evolution of the intensity estimation 2015-2021

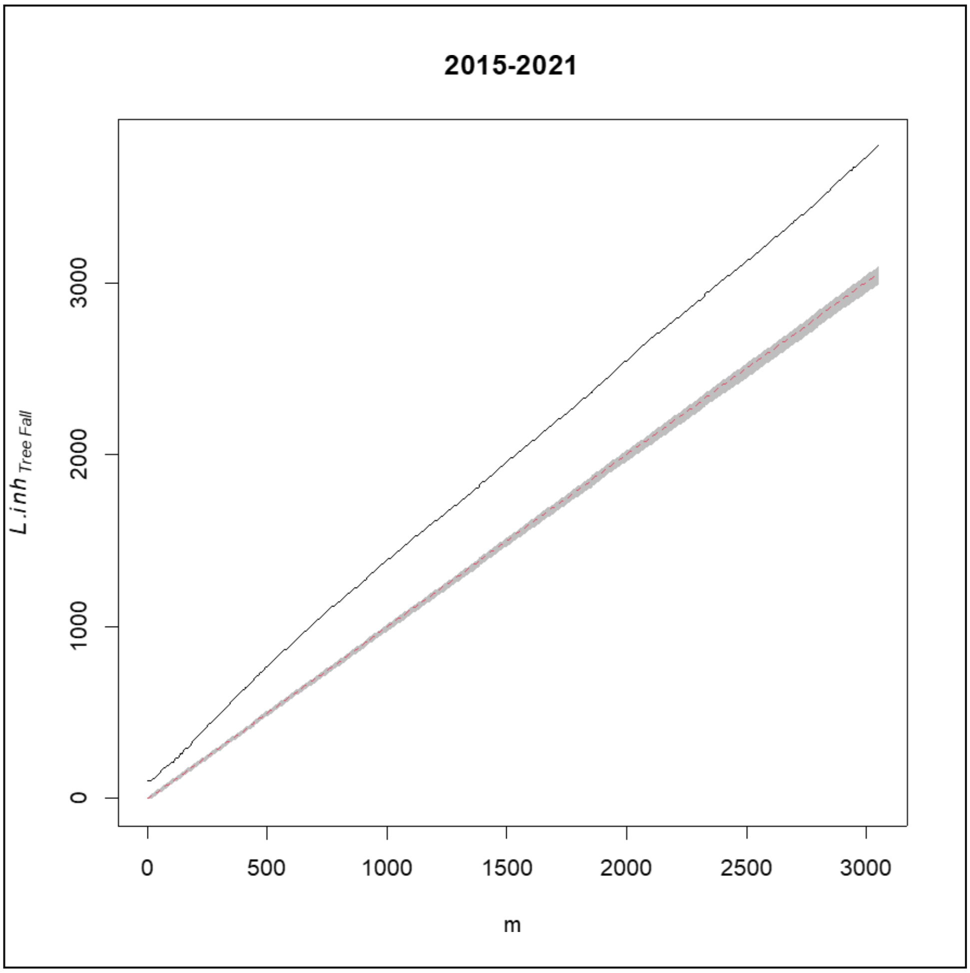


Source: Author's own creation (2024)

3.2. Inhomogenous L Function Results:

Utilizing the Inhomogenous L Function on the dataset, a discernible trend emerges due to the notably elevated calculated L_{inhom} values. This observation unequivocally indicates a clustered pattern within the spatial distribution of points across the 2015-2021 interval (Figure 6).

Figure 6: Inhomogenous L Function 2015-2021



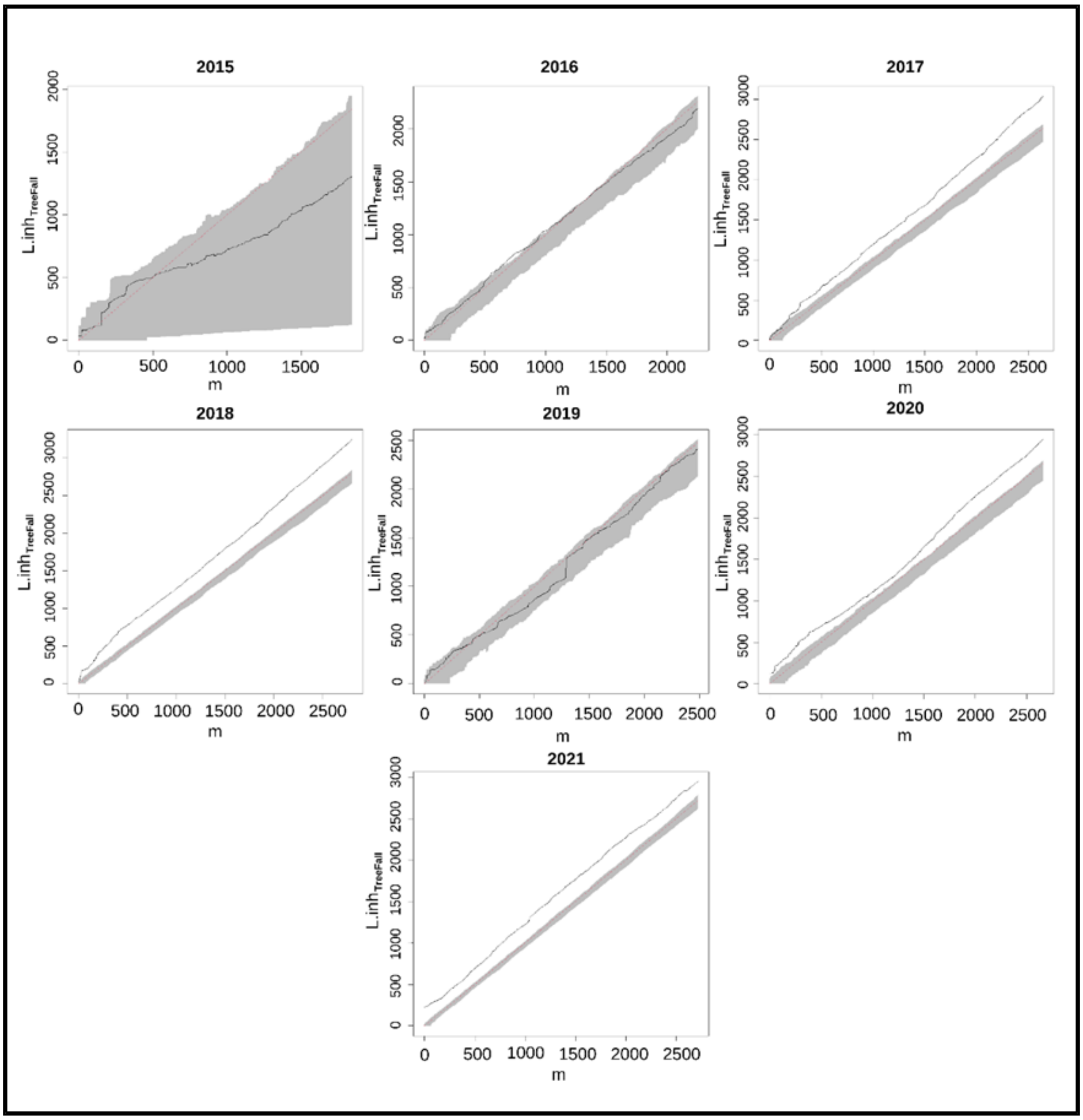
Source: Author's own creation (2023)

When conducting a more granular analysis by evaluating each year independently, a consistent outcome is evident. In 2015, the values observed at all assessed distances were within the envelope, indicating Complete Spatial Randomness for that year. Moving to 2016, a distinctive pattern emerged. Between 600 meters and 900 meters, the values exceeded the envelope, revealing clustering at these distances. Interestingly, at distances below 600 meters and above 900 meters, random patterns were observed.

In 2017, a shift was observed. At distances of approximately 250 meters, random patterns were identified. Beyond this distance, clustering patterns became more pronounced. The year 2018 exhibited a clear clustering pattern, with tree falls distributed in a clustered pattern. Contrastingly, 2019 recorded a random pattern across all distances.

In 2020, a distinct clustering pattern reappeared, evident at all distances assessed. Continuing this trend, 2021 also demonstrated a clustering pattern in the spatial distribution of tree falls.

Figure 7: Inhomogenous L Function applied to each year individualy



Source: Author's own creation (2024)

It’s interesting to note that despite the entire interval showing a cluster pattern, when observing each year separately, we see that this pattern did not repeat in all years, such as in 2015, 2016, and 2019. Also, these years had 48.48%, 41.15%, and 52.19% of data lost during geocoding. It is important to highlight that the two years which also had the lowest information loss during geocoding, in 2020 and 2021, presented a cluster distribution pattern.

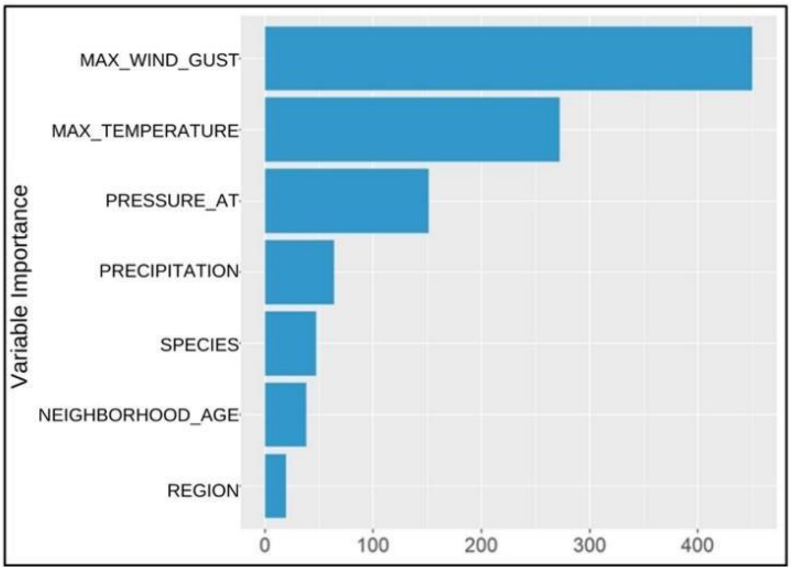
These annual assessments underscore the dynamic nature of tree fall incidents, with spatial patterns varying from complete spatial randomness to notable clustering, offering crucial insights for effective urban planning and risk management strategies.

3.3. Regression Tree:

3.3.1. Model I – All variables and all reports

In the first model, the explanatory variables included the tree species, accumulated precipitation on the day, atmospheric pressure, maximum temperature, maximum wind gust, neighborhood age, and the region where the tree fell. When applying this algorithm, the most important variables were identified (Figure 8). These variables are the maximum wind gust, maximum temperature, atmospheric pressure, precipitation, species, neighborhood age, and the region.

Figure 8: Variable Importance - Model I



Source: Author's own creation (2024)

It is observed that the four main variables are meteorological variables related to weather events, highlighting the importance and role of climate in influencing these events. The rules of this first model can be seen in Table 4.

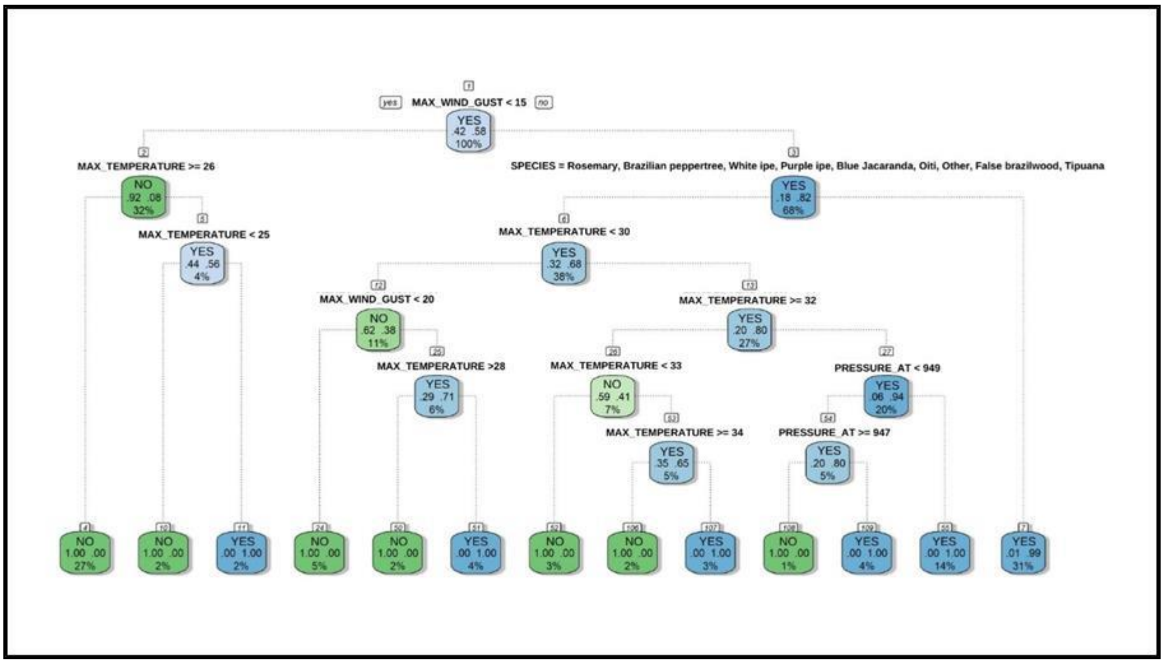
Table 4: Rules of the Model I

Rules	Cover
1.when MAX WIND GUST < 15 & MAX TEMPERATURE >= 26	27%
2.when MAX WIND GUST < 15 & MAX TEMPERATURE < 25	2%
3.when MAX WIND GUST is 15 to 20 & MAX TEMPERATURE < 30 & SPECIES is Rosemary or Brazilian peppertree or White ipe or Purple ipe or Blue jacaranda or Oiti or Others or False brazilwood or Tipuana	5%
4.when MAX WIND GUST >= 20 & MAX TEMPERATURE is 28 to 30 & SPECIES is Rosemary or Brazilian peppertree or White ipe or Purple ipe or Blue jacaranda or Oiti or Others or False brazilwood or Tipuana	2%
5.when MAX WIND GUST >= 15 & MAX TEMPERATURE is 32 to 33 & SPECIES is Rosemary or Brazilian peppertree or White ipe or Purple ipe or Blue jacaranda or Oiti or Others or False brazilwood or Tipuana	3%
6.when MAX WIND GUST >= 15 & MAX TEMPERATURE >= 34 & SPECIES is Rosemary or Brazilian peppertree or White ipe or Purple ipe or Blue jacaranda or Oiti or Others or False brazilwood or Tipuana	2%
7.when MAX WIND GUST >= 15 & MAX TEMPERATURE is 30 to 32 & SPECIES is Rosemary or Brazilian peppertree or White ipe or Purple ipe or Blue jacaranda or Oiti or Others or False brazilwood or Tipuana & PRESSURE AT is 947 to 949	1%
8.when MAX WIND GUST >= 15 & SPECIES is NA	31%
9.when MAX WIND GUST < 15 & MAX TEMPERATURE is 25 to 26	2%
10.when MAX WIND GUST >= 20 & MAX TEMPERATURE < 28 & SPECIES is Rosemary or Brazilian peppertree or White ipe or Purple ipe or Blue jacaranda or Oiti or Others or False brazilwood or Tipuana	4%
11.when MAX WIND GUST >= 15 & MAX TEMPERATURE is 33 to 34 & SPECIES is Rosemary or Brazilian peppertree or White ipe or Purple ipe or Blue jacaranda or Oiti or Others or False brazilwood or Tipuana	3%
12.when MAX WIND GUST >= 15 & MAX TEMPERATURE is 30 to 32 & SPECIES is Rosemary or Brazilian peppertree or White ipe or Purple ipe or Blue jacaranda or Oiti or Others or False brazilwood or Tipuana & PRESSURE AT < 947	4%
13.when MAX WIND GUST >= 15 & MAX TEMPERATURE is 30 to 32 & SPECIES is Rosemary or Brazilian peppertree or White ipe or Purple ipe or Blue jacaranda or Oiti or Others or False brazilwood or Tipuana & PRESSURE AT >= 949	14%

Source: Author’s own creation (2024)

Some interesting observations can be noted in the model (Figure 9): trees that fell during windstorms did so on days with maximum wind gusts greater than 15 m/s, and the species was unidentified, representing 31% of the classified fallen trees during windstorms. Another case of trees falling during windstorms occurred on days with wind gusts greater than or equal to 15 m/s, atmospheric pressure greater than or equal to 949 hPa, maximum temperature during the day between 30 ºC and 32 ºC, and the species was identified.

Figure 9: Regression Tree - Model I



Source: Author's own creation (2024)

Analyzing the confusion matrix of this model (Table 5), we observe that 106 cases were correctly classified as "NO." There were no false positives, indicating that the model did not mistakenly classify instances of the "NO" class as "YES." One case was mistakenly classified as "NO" when it was actually "YES." Additionally, 190 cases were correctly classified as "YES."

Table 5: Confusion Matrix - Model I

Model I	NO	YES
NO	106	0
YES	1	190

Source: Author's own creation (2024)

In this context, the model demonstrates exceptional performance with a precision of 100% (Table 6). This means that all instances classified as "YES" by the model were indeed positive. The results suggest that the model is highly accurate and capable of efficiently identifying positive instances with minimal false negatives and no false positives. The values associated with the model's performance can be observed in the table below. These numbers support the claim of high precision of the model, highlighting its ability to correctly identify positive instances while minimizing both false positives and false negatives.

Table 6: Model's performance - Model I

Test	Values
Recall	99.07%
Accuracy	99.66%
Specificity	100%
Precision	100%

Source: Author's own creation (2024)

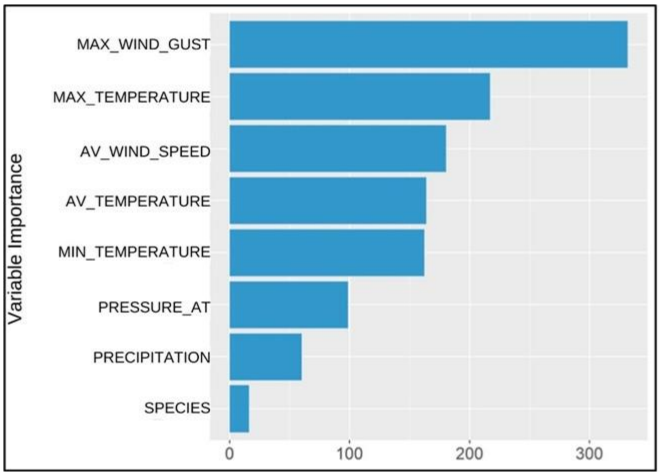
The findings of this regression tree model align with those observed in similar studies. For instance, the study “Factor Influencing Street Tree Hazard Condition in Rafaela, Argentina” (Castro, Alesso, Iaconis, Cerino, & Buyatti, 2019)highlights the significant influence of meteorological variables such as wind speed and temperature on tree falls. Similarly, the research “Climate drivers of tree fall on the streets of São Paulo, Brazil” (Locosselli, Miyahara, Cerqueira, & Buckeridge, 2021)emphasizes the critical role of maximum wind gusts and temperature in tree fall incidents, corroborating our results that identified maximum wind gust and temperature as primary factors. Additionally, the study “Average height of surrounding buildings and district age are the main predictors of tree failure on the streets of São Paulo, Brazil” (2022) mentions that urban characteristics like neighborhood age also play a crucial role, which is consistent with our finding that neighborhood age is an important variable in our model.

These studies collectively reinforce the notion that both climatic conditions and urban infrastructure significantly affect the likelihood of tree falls. The consistency of our findings with those of other research works adds robustness to the conclusion that maximum wind gusts, temperature, and urban characteristics are pivotal factors in predicting tree falls during windstorms. Such insights are crucial for urban planning and risk mitigation strategies to prevent tree falls and enhance urban safety.

3.3.2. Model II – Region 1 and Region 2

This model considered only the trees that fell in regions 1 and 2 of the city, which are the regions with the highest incidence of fallen trees. The explanatory variables included in the model were the tree species, accumulated precipitation of the day, atmospheric pressure, maximum, minimum, and average temperatures, maximum wind gust, and average wind speed. When applying this algorithm, the following variables were identified as the most important (Figure 10): maximum wind gust, maximum temperature, and average wind speed. In this model, we also observe that the main variables are related to the weather, particularly wind and temperature.

Figure 10: Variable Importance - Model II



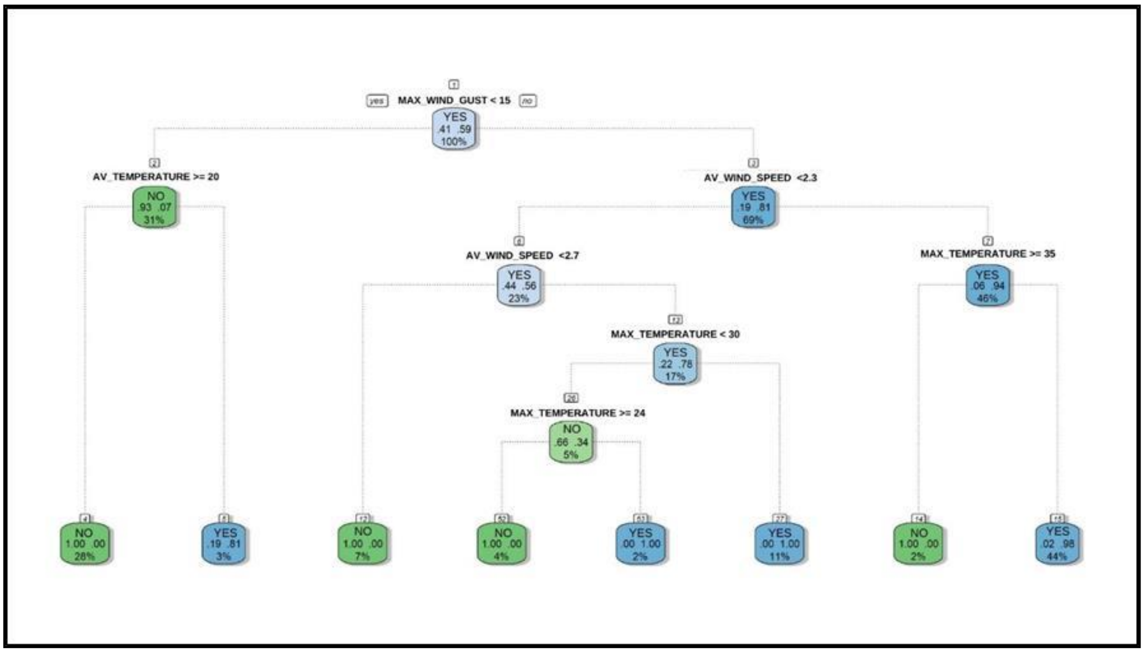
In table 7 there are the rules of the model and in the figure 11 the decision tree based on the calculated model.

Table 7: Rules of the Model II

Rules	Cover
1. When MAX WIND GUST < 15 & AV TEMPERATURE >= 20	28%
2. when MAX WIND GUST >= 15 & AV WIND SPEED is 2.3 to 2.7	7%
3. when MAX WIND GUST >= 15 & AV WIND SPEED >= 2.7 & MAX TEMPERATURE is 24 to 30	4%
4. when MAX WIND GUST >= 15 & AV WIND SPEED < 2.3 & MAX TEMPERATURE >= 35	2%
5. when MAX WIND GUST < 15 & AV TEMPERATURE < 20	3%
6. when MAX WIND GUST >= 15 & AV WIND SPEED < 2.3 & MAX TEMPERATURE < 35	44%
7. when MAX WIND GUST >= 15 & AV WIND SPEED >= 2.7 & MAX TEMPERATURE < 24	2%
8. when MAX WIND GUST >= 15 & AV WIND SPEED >= 2.7 & MAX TEMPERATURE >= 30	11%

Source: Author's own creation (2024)

Figure 20: Regression Tree - Model II



Source: Author's own creation (2024)

When assessing the effectiveness of this model (Table 8), we observe the following: 58 cases were correctly classified as "NO", 68 cases were falsely classified as "YES" when they were actually "NO", 49 cases were falsely classified as "NO" when they were actually "YES", and 122 cases were correctly classified as "YES".

Table 8: Confusion Matrix - Model II

Model II	NO	YES
NO	58	68
YES	49	122

Source: Author's own creation (2024)

When analyzing the test results for Model II, we observe that the model's ability to correctly identify cases of fallen trees during storms, expressed by Recall, is 54.21%. This highlights the efficiency of the model in capturing a considerable portion of true positive cases. The Accuracy, reflecting the overall precision of predictions, stands at 60.61%, indicating the percentage of correct predictions overall in both positive and negative classes. Specificity, assessing the model's precision in identifying cases where trees did not fall during storms, reaches 64%, emphasizing the accurate identification of cases without tree falls. Regarding Precision, examining the accuracy of positive predictions, we observe a value of 46.03%. This means that approximately 46% of positive predictions made by the model are correct among all positive predictions.

Table 9: Model's performance - Model II

Test	Values
Recall	54.21%
Accuracy	60.61%
Specificity	64%
Precision	46.03%

Source: Author's own creation (2024)

The performance of this model (Table 9) proved to be inferior to Model I; however, some highlights can be made. The maximum wind gust proved an important variable for both models, as did the maximum temperature, indicating that a wind gust above 15 m/s is a determining factor for tree falls during a windstorm. This finding is supported by the study “Climate drivers of tree fall on the streets of São Paulo, Brazil” (Castro, Alesso, Iaconis, Cerino, & Buyatti, 2019)which also identified wind gusts as a critical factor influencing tree falls. Furthermore, the relevance of average wind speed in our model aligns with the research “Factor Influencing Street Tree Hazard Condition in Rafaela, Argentina” (Locosselli, Miyahara, Cerqueira, & Buckeridge, 2021) where wind conditions significantly impacted tree stability.

By comparing these results with those from previous studies, we reinforce the importance of climatic variables, particularly wind speed and temperature, in predicting tree falls. These insights can inform urban management and planning strategies to mitigate risks associated with tree falls in urban areas.

4. CONCLUSION

This comprehensive analysis of tree fall incidents within the urban landscape has shed light on intriguing and evolving spatial patterns over the years. The identification of higher concentrations and risk of falls in the northern and northeastern regions of the city through kernel analysis is particularly noteworthy.

The examination of K_{inhom} and L_{inhom} functions has revealed dynamic spatial patterns in the distribution of tree falls over the years. Notably, the transition from complete spatial randomness to clustering patterns, as observed in different years, emphasizes the dynamic nature of these incidents. The northern and northeastern regions consistently emerge as areas with higher risks of tree falls, providing crucial insights for targeted urban planning and risk management strategies.

Furthermore, the prominence of the “Sibipiruna” species in tree falls emphasizes the need for species-specific management and mitigation efforts. The regression tree results, highlighting wind factors (maximum and average gusts), temperature, atmospheric pressure, and precipitation as significant contributors, offer valuable information for understanding the main factors influencing tree falls.

In summary, this multifaceted analysis underscores the dynamic spatial dynamics of tree fall incidents. The shifting clustering tendencies, concentrations in specific regions, and the influence of various factors contribute to a comprehensive understanding. Importantly, the findings advocate for the standardization of data collection methods, suggesting the implementation of a protocol or application based on the study’s findings. Such a tool could serve as a valuable asset for the management and planning of urban forestry, ultimately fostering a safer and more resilient urban environment.

BIBLIOGRAPHY

Azdy, R. A., & Darnis, F. (2020). Use of Harversine Formula in Finding Distance Between Temporary Shelter and Waste End Processing Sites. *Journal of Physics: Conference Series*, pp. 1-6.

Baddeley, A., & Turner, R. (2005). spatstat: An R Package for Analyzing Spatial Point Patterns. *Journal of Statistical Software, 12*(6), pp. 1-42. doi:10.18637/jss.v012.i06

Baddeley, A., Moller, J., & Waagepetersen, R. (2000). Non- and semiparametric estimation of interaction in Inhomogenous point patterns. *Statistica Neerlandica, 54*, pp. 329-350.

Baddeley, A., Rubak, E., & Turner, R. (2015). Spatial Point Patterns: Methodology and Applications with R. Chapman and Hall/CRC Press. Retrieved from https://www.routledge.com/Spatial-Point-Patterns-Methodology-and-Applications-with-R/Baddeley-Rubak-Turner/p/book/9781482210200/

Berman, M., & Diggle, P. J. (1989). Estimating Weighted Integrals of the Second-Order Intensity of a Spatial Point Pattern. *J. R. Statist Soc*, pp. 81-92.

Bivand, R. S., Pebesma, E. J., & Gómez-Rubio, V. (2013). Spatial Point Pattern Analysis. In R. S. Bivand, E. J. Pebesma, & V. Gómez-Rubio, *Applied Spatial Data Analysis with R* (pp. 155-190). New York, NY: Springer.

Câmara, G., & Carvalho, M. S. (2004). Análise de Eventos Pontuais. In S. Druck, M. S. Carvalho, & C. G, *Análise Espacial de Dados Geográficos* (pp. 1-15). Brasília: EMBRAPA.

Castro, D. C., Alesso, C. A., Iaconis, A., Cerino, M. C., & Buyatti, M. (2019). Factors influencing street tree hazard condition in Rafaela, Argentina. *Revista Arvore*, pp. 1-10. doi:10.1590/1806-90882019000400010

Cressie, N. E. (1993). Spatial Point Patterns. In N. E. Cressie, *Statistics for Spatial Data* (pp. 575-723). John Wiley & Sons.

da Silva, D., Francisco, I. C., Raniero , G. Z., Monteiro, C., Rossoni, D. F., & Monteiro, A. (2023). Application of NIR for Rapid Determination of Flour Quality by Machine Learning. *CHEMICAL ENGINEERING TRANSACTIONS*, pp. 193-198.

Dixon, P. M. (2002). Ripley's K funtion. *Encyclopedia of Environmetrics, 3*, pp. 1796-1803.

Esri. (n.d.). *How Multi-Distance Spatial Cluster Analysis (Ripley's K-function) works*. Retrieved from ArcGIS Pro: https://pro.arcgis.com/en/pro-app/latest/tool-reference/spatial-statistics/h-how-multi-distance-spatial-cluster-analysis-ripl.htm

IBGE. (2022). *Maringá.* Retrieved from Cidades@: https://cidades.ibge.gov.br/brasil/pr/maringa/panorama

Illian, J., Penttinen, A., Stoyan, H., & Stoyan, D. (2008). *Statistical Analysis and Modelling of Spatial Point Patterns.* John Wiley & Sons.

Jim, C., & Zhang, H. (2013). Defect-disorder and risk assessment of heritage trees in Urban Hong Kong. *Urban Forestry & Urban Greening*, pp. 585–596.

Jones, M. C. (1993). Simple boundary correction for kernel density estimation. *Statistics and Computing*, pp. 135-146. doi:10.1007/BF00147776

Kawamoto, M. T. (2012). Análise de técnicas de distribuição espacial com padrões pontuais e aplicação a dados de acidentes de trânsito e a dados de dengue de Rio Claro-SP. *Dissertação (mestrado)*. Botucatu, São Paulo, Brazil.

Kuhn, M. (2008). Building Predictive Models in R Using the caret Package. *Journal of Statistical Software, 28*(5), pp. 1–26. doi:10.18637/jss.v028.i05

Locosselli, G. M., Miyahara, A. A., Cerqueira, P., & Buckeridge, M. (2021). Climate drivers of tree fall on the streets of São Paulo, Brazil. *Trees*(35), pp. 1807–1815.

Manfra, R., Massoca, M. S., Uras, P. M., Cavalari, A., & Locosselli, G. M. (2022). Average height of surrounding buildings and district age are the main predictors of tree failure on the streets of São Paulo/Brazil. *Urban Forestry & Urban Greening*, pp. 1-8.

Minaki, C. (2021). Characteristics of nocturnal mobile transects of temperature and relative air humidity in a sample area of Maringá-PR. *GEOSUL*(36), pp. 610-634. doi:10.5007/2177-5230.2021.e72350

Oliveira, W. A. (2022). Métodos de processos pontuais para análise espacial de espécies arbóreas nativas de um fragemento florestal de Lavras-MG. 77 p. Lavras, Brazil: Universidade Federal de Lavras. Retrieved from http://repositorio.ufla.br/jspui/handle/1/55342

Pretzsch, H., Biber, P., Uhl, E., Dahlhausen, J., Rotzer, T., Caldentey, J., & Pauleit, S. (2015). Crown size and growing space requirement of common tree species in urban centres, parks, and forests. *Urban Forestry & Urban Greening*, pp. 466–479.

QGIS Development Team. (2022). QGIS Geographic Information System Developers Manual. *QGIS 3.28 'Firenze'*. Retrieved from http://www.qgis.org/

R Core Team. (2022). R: A language and environment for statistical computing. Vienna, Austria. Retrieved from https://www.R-project.org/

Rego, R. L. (2001). The city plant of Maringá and the concept of garden city. *Acta Scientiarum*, pp. 1569-1577.

Ribeiro, P., & Lopes, A. (2011). Modelação do risco de queda de árvores sobre as viaturas devido a ventos fortes em Lisboa. *VIII Congresso da Geografia Portuguesa*, pp. 1-9.

Ripley, B. D. (1977). Modelling Spatial Patterns. *Journal of the Royal Statistical Society: Series B (Methodological)*, pp. 172-192.

RPC Maringá. (2021, September 14). *Chuva forte provoca quedas de árvores e deixa 30 mil domicílios sem luz em Maringá*. Retrieved from g1: https://g1.globo.com/pr/norte-noroeste/noticia/2021/09/14/chuva-forte-provoca-quedas-de-arvores-e-deixa-30-mil-domicilios-sem-luz-em-maringa.ghtml

Schmidlin, T. W. (2008). Human fatalities from wind-related tree failures in the United States, 1995-2007. *Natural Hazards*(50), pp. 13-25. doi:10.1007/s11069-008-9314-7

Schmidt, O., Gaiser, O., & Dujesiefken, D. (2011, September 24). Molecular identification of decay fungi in the wood of urban trees. *European Journal of Forest Research*, pp. 885–891.

Shabani, S., & Akbarinia, M. (2017). Prediction spatial patterns of windthrow phenomenon in deciduous temperate forests using logistic regression and random forest. *CERNE*, pp. 387-394.

Smith, I. A., Dearborn, V. K., & Hutyra, L. R. (2019). Live fast, die young: Accelerated growth, mortality, and turnover in street trees. *PLOS ONE*, pp. 1-7. doi:https://doi.org/10.1371/journal.pone.0215846

Souza, S. B., & Scalon, J. D. (2018). Funções L Homogenea e Não Homogênea em Testes de Hipótese de Completa Aleatoriedade Espacial. *Revista de Estatística UFOP*, pp. 1-15.

The Arbor Day Foundation. (2023). *Recognised Cities.* Retrieved from Tree Cities of the World: https://treecitiesoftheworld.org/documents/2022/tcow-summary-2022-infographic.pdf

van Haaften, M., Liu, Y., Wang, Y., Zhang, Y., Gardebroek, C., Heijman, W., & Meuwissen, M. (2021, February 16). Understanding tree failure—A systematic review and meta-analysis. *PLOS ONE*, pp. 1-23.

Wei-Yin, L. (2011, January 06). Classification and Regression Trees. *WIREs - Data Mining and Knowledge Discovery, 1*, pp. 14-23.

Zeidan, D. N., & Ferreira, M. E. (2023). Floristic survey in a forested urban green area in Maringá-PR. In C. D. Silva, & R. O. Souza, *Meio Ambiente e Suas Interrelações Com O Desenvolvimento Sustentável* (pp. 29-44). Ponta Grossa: Atena Editora. doi:10.22533/at.ed.0162317103