UNIVERSIDADE ESTADUAL PAULISTA
"JÚLIO DE MESQUITA FILHO"
Campus de São José dos Campos
Instituto de Ciência e Tecnologia

Cemaden
Centro Nacional de Monitoramento e Alertas de Desastres Naturais
UNIDADE DE PESQUISA DO MCTI

Ministério da Ciência, Tecnologia e Inovações

GOVERNO FEDERAL
BRASIL
UNIÃO E RECONSTRUÇÃO

BELIANA CAVALCANTE SAWADA DE CARVALHO

ANÁLISE DE SUSCETIBILIDADE A INUNDAÇÃO EM GRANDES CIDADES ATRAVÉS DO MÉTODO DE FLORESTAS ALEATÓRIAS: Estudo de caso no município de São Paulo (SP)

2024

BELIANA CAVALCANTE SAWADA DE CARVALHO

ANÁLISE DE SUSCETIBILIDADE A INUNDAÇÃO EM GRANDES CIDADES ATRAVÉS DO MÉTODO DE FLORESTAS ALEATÓRIAS: Estudo de caso no município de São Paulo (SP)

Dissertação apresentada ao Instituto de Ciência e Tecnologia, Universidade Estadual Paulista (Unesp), Campus de São José dos Campos; Centro Nacional de Monitoramento e Alertas de Desastres Naturais (Cemaden), como parte dos requisitos para obtenção do título de MESTRA pelo Programa de Pós-Graduação em DESASTRES NATURAIS.

Área: Desastres Naturais. Linha de pesquisa: Instrumentação e Análise de Dados.

Orientador: Prof. Dr. Silvio Jorge Coelho Simões

Coorientador: Profa. Dra. Tatiana Sussel Gonçalves Mendes

São José dos Campos

2024

Instituto de Ciência e Tecnologia [internet]. Normalização de tese e dissertação [acesso em 2024]. Disponível em http://www.ict.unesp.br/biblioteca/normalizacao

Apresentação gráfica e normalização de acordo com as normas estabelecidas pelo Serviço de Normalização de Documentos da Seção Técnica de Referência e Atendimento ao Usuário e Documentação (STRAUD).

Carvalho, Beliana Cavalcante Sawada de
Análise se suscetibilidade a inundação em grandes cidades através do método de florestas aleatórias: Estudo de caso no município de São Paulo (SP) / Beliana Cavalcante Sawada de Carvalho. - São José dos Campos : [s.n.], 2024.
56 f. : il.

Dissertação (Mestrado) - Pós-Graduação em Desastres Naturais - Universidade Estadual Paulista (Unesp), Instituto de Ciência e Tecnologia, São José dos Campos, 2024.
Orientador: Silvio Jorge Coelho Simões
Coorientador: Tatiana Sussel Gonçalves Mendes

1. Aprendizado em máquina. 2. Florestas Aleatórias. 3. Desastres Naturais. 4. Inundações. 5. São Paulo. I. Simões, Silvio Jorge Coelho, orient. II. Mendes, Tatiana Sussel Gonçalves, coorient. III. Universidade Estadual Paulista (Unesp), Instituto de Ciência e Tecnologia, São José dos Campos. IV. Universidade Estadual Paulista 'Júlio de Mesquita Filho' - Unesp. V. Universidade Estadual Paulista (Unesp). VI. Título.

Ficha catalográfica elaborada pela Biblioteca Prof. Achille Bassi e Seção Técnica de Informática, ICMC/USP
com adaptações - STATI, STRAUD e DTI do ICT/UNESP.
Renata Aparecida Couto Martins CRB-8/8376

IMPACTO POTENCIAL DESTA PESQUISA

A espacialização dos dados de chuva com o uso da geoestatística é significativo porque contribui para o entendimento dos padrões de precipitação em uma determinada região a partir de análises temporais e espaciais dos dados coletados, permitindo identificar áreas propensas a eventos extremos e a tomada de decisões sobre uso da terra e gestão de recursos, bem como contribui para o mapeamento de suscetibilidade a inundação em grandes cidades. Mapa de suscetibilidade a inundação é uma ferramenta de extrema importância para gestores na avaliação e gerenciamento de riscos. Os mapas gerados neste estudo fornecem informações detalhadas e precisas para identificar áreas favoráveis aos eventos de inundação, permitindo que a cidade de São Paulo possa implementar medidas preventivas e de preparação para diminuir os efeitos nas comunidades sujeitas aos riscos de inundação.

POTENTIAL IMPACT OF THIS RESEARCH

The spatialization of rainfall data with the use of geostatistics is significant because it contributes to the understanding of precipitation patterns in a given region based on temporal and spatial analyses of the data collected, allowing the identification of areas prone to extreme events and the making of decisions about land use and resource management, as well as contribute for flood susceptibility mapping in large cities. Flood susceptibility map is an extremely important tool for managers in the assessment and management of hazards. The maps generated in this study provide detailed and accurate information to identify areas favorable to flood events, allowing the city of São Paulo to implement preventive and preparedness measures to reduce the effects on communities subject to flood risks.

BANCA EXAMINADORA

Prof. Dr. Silvio Jorge Coelho Simões
Universidade Estadual Paulista "Júlio de Mesquita Filho"
Instituto de Ciência e Tecnologia
Campus de São José dos Campos

Prof. Dr. Rogério Galante Negri
Universidade Estadual Paulista "Júlio de Mesquita Filho"
Instituto de Ciência e Tecnologia
Campus de São José dos Campos

Profa. Dra. Mariana Ferreira Benessiuti Motta
Universidade Estadual Paulista "Júlio de Mesquita Filho"
Faculdade de Engenharia e Ciências
Campus de Guaratinguetá

São José dos Campos, 02 de fevereiro de 2024.

DEDICATÓRIA

Dedico este trabalho aos meus pais, irmãs, filha e marido por todo amor e carinho.

AGRADECIMENTOS

Ao Professor Silvio Jorge Coelho Simões, pela sua orientação, valiosas sugestões e ensinamentos.

À Professora Tatiana Sussel Gonçalves Mendes, pela sua orientação, permanente colaboração e acompanhamento.

À banca de qualificação, Professor Rogério Galante Negri e Professora Mariana Benessiuti Motta, pelas sugestões de melhoria ao meu trabalho.

Aos professores do Programa de Pós-graduação de Desastres Naturais que contribuíram para o meu aprendizado. Aos funcionários da secretária técnica da Pós-graduação.

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – CAPES pela bolsa de mestrado.

Ao meu pai e à minha irmã Taisa, que sempre me incentivaram nos estudos.

Aos meus amores, minha afilhada Loli e minha filha Marcela, por alegrarem os meus dias.

Ao meu marido Adriano, por todo apoio e cuidado com a nossa família.

Finalmente, à Deus, por estar sempre me guiando.

“É maravilhoso, Senhor, ter tão pouco a pedir e tanto para agradecer”. Autor desconhecido

RESUMO

CARVALHO, B. C. S. C. **Análise de suscetibilidade a inundação em grandes cidades através do método de Florestas Aleatórias:** Estudo de caso no município de São Paulo (SP). 2024. Dissertação (Mestrado em Desastres Naturais) - Universidade Estadual Paulista (Unesp), Instituto de Ciência e Tecnologia; Centro Nacional de Monitoramento e Alertas de Desastres Naturais (Cemaden), São José dos Campos, 2024.

O aumento da frequência de chuvas intensas devido aos efeitos das mudanças climáticas e o acelerado processo de expansão urbana, como compactação e impermeabilização do solo, são fatores que contribuem para a incidência de desastres naturais como inundações, alagamentos e deslizamentos de encostas. Atualmente, os eventos de inundação são a principal causa de danos em todo o mundo. Esses eventos apresentam tendência de aumento quando comparados à média anual de eventos no passado e podem ser agravados em cenários de variação climática. Considerando uma das cidades mais importantes do Brasil, São Paulo tem características peculiares relacionadas a áreas densamente ocupadas sobre e margeando os rios. Este trabalho tem como objetivo a análise suscetibilidade a inundação usando o método Floresta Aleatória em uma região do município de São Paulo. O método proposto usa como base dois inventários, um com 75 pontos de inundação e outro com 578 pontos de inundação e alagamento. O mapa de precipitação é um dos fatores condicionantes usados como dados de entrada no modelo. Para estimar a precipitação média anual para o município, foi utilizado o método geoestatístico de krigagem ordinária. Um MDT, derivado de dados LiDAR, foi usado para gerar a maioria dos fatores condicionantes. Ao todo, 10 fatores condicionantes foram usados, sendo: altitude, amplitude do relevo, aspecto, curvatura, curvatura em perfil, curvatura plana, declividade, litologia, pluviométrico e TWI. Os resultados foram comparados com um mapa geotécnico disponível para a área. O modelo gerado a partir dos pontos de inundação e alagamento apresentou melhor desempenho em termos de acurácia, AUC, sensibilidade e especificidade. Os resultados dos modelos revelam que as áreas de suscetibilidade à inundação vão além das classes de planície de inundação e área sujeita a inundação apresentadas no mapa geotécnico.

Palavras-chave: Aprendizado em máquina; florestas aleatórias; desastres naturais; inundações, São Paulo.

ABSTRACT

CARVALHO, B. C. S. C. ***Analysis of susceptibility to flooding in large cities using the Random Forest method:*** *a case study in the city of São Paulo (SP). 2024. Dissertation (Master's degree in Natural Disaster) - São Paulo State University (Unesp), Institute of Science and Technology, National Center for Monitoring and Early Warning of Natural Disasters (Cemaden), São José dos Campos, 2024.*

The increase in the frequency of heavy rains due to the effects of climate change and the accelerated process of urban expansion, such as soil compaction and sealing, are factors that contribute to the incidence of natural disasters such as flooding, flooding and landslides on slopes. Currently, flood events are the main cause of damage worldwide. These events present an upward trend when compared to the annual average of events in the past and can be worsened in climate variation scenarios. Considering one of the most important cities of Brazil, São Paulo has peculiar characteristics related to densely occupied areas over and bordering rivers. This work aims to analyze flood susceptibility using the Random Forest method in a region of the city of São Paulo. The proposed method uses two inventories, one with 75 flooding points and the other with 578 flooding and flash flood points. The precipitation map is one of the conditioning factors used as input data in the model. To estimate the average annual precipitation in the municipality of São Paulo, the geostatistical method of ordinary kriging was used. DTM derived from LiDAR data was used to generate most of the conditioning factors. In all, 10 conditioning factors were used, namely: altitude, relief amplitude, aspect, curvature, profile curvature, plan curvature, slope, lithology, rainfall and TWI. *The results were compared with a geotechnical map available for the area. The model generated from the flood and flash flood points better showed the highest performance in terms of accuracy, AUC, sensitivity and specificity. The results of the models reveal that the flooding susceptibility areas go beyond the classes of floodplain and area subject to flooding classes presented on the geotechnical map.*

Keywords: *Machine learning; random forests; natural disasters; flooding; Sao Paulo.*

SUMÁRIO

1 INTRODUÇÃO

De acordo com o relatório do Painel Internacional de Mudanças Climáticas (IPCC), entre 1970 e 2019, os eventos relacionados a inundação foram responsáveis por 44% de todos os desastres e 31% de todas as perdas econômicas (IPCC, 2022). As cidades estão cada vez mais enfrentando os efeitos das mudanças na cobertura natural do solo para áreas construídas, aumentando as superfícies impermeáveis (Tomás et al., 2022). Mudanças no uso e cobertura da terra e redirecionamento do fluxo de água, associados aos efeitos das mudanças climáticas, têm sido responsáveis pelo aumento dos eventos de inundação (Zope et al., 2016). O aumento da temperatura global acelera o ciclo hidrológico, intensificando as chuvas e contribuindo para as diversas ocorrências de eventos extremos de inundação (Hirabayashi et al., 2013; Wu et al., 2024). Localmente, a urbanização intensa pode aumentar a precipitação média, resultando no aumento da intensidade do escoamento superficial (Du et al., 2019; Feng et al., 2021; IPCC, 2021).

As inundações muitas vezes destroem as casas e propriedades das pessoas, causam deslocamento, perda de vidas, representam riscos à saúde e afetam negativamente os meios de subsistência das pessoas (Mambele et al., 2022). No Brasil, devido às suas características sociais e econômicas, as grandes cidades caracterizam-se pela alta densidade demográfica e pela histórico desordenado de ocupação do território. A expansão dos grandes centros urbanos, acompanhada pelo aumento da impermeabilização do solo, ocupação de várzeas, assoreamento de rios e eventos pluviométricos cada vez mais severos, intensificou os eventos de inundação. Muitas cidades cresceram sobre áreas de várzea (Aiyelokun, 2023), canalizando rios e ocupando suas nascentes, resultando em espaços potencialmente perigosos, especialmente no período chuvoso (Valverde et al., 2023).

Um exemplo é a cidade de São Paulo, que tem o maior PIB do Brasil e é uma das áreas mais densamente ocupadas do país (IBGE, 2022). Quando as enchentes atingem a população da capital paulista, as perdas econômicas e sociais são incalculáveis. Nessa perspectiva, é necessário que o município tenha uma gestão de riscos eficiente e uma política de monitoramento efetiva, para desenvolver ações e planos de prevenção e redução de riscos.

A cidade de São Paulo é caracterizada pela grande área construída e impermeabilizada, fatores que associados aos períodos chuvosos, típicos dos verões, contribuem para a ocorrência dos alagamentos e das inundações (Haddad e Teixeira, 2015). A disposição de lixo nas proximidades do curso d’água é também um fator para que estes

eventos ocorram com mais frequência. A primeira inundação conhecida no município de São Paulo ocorreu no Vale do Anhangabaú, na região central da cidade, após uma "tromba d'água" no dia primeiro de janeiro de 1850. Os danos registrados foram a destruição de 12 casas e ocorrência de três mortes (Governo do Estado de São Paulo, 2014). Em 1879, as inundações atingiram a várzea do Rio Tietê, a Rua 25 de março e a Ilha dos Amores, como consequência casas próximas ao Gasômetro foram atingidas pela água. No ano de 1905, as águas do Rio Tietê e Tamanduateí causaram estragos em edifícios e grandes prejuízos econômicos (Governo do Estado de São Paulo, 2014).

A maior inundação registrada na capital Paulista foi em 18 de fevereiro de 1929, após várias inundações terem sido ocorridas em diversas regiões da cidade entre 6 e 8 de fevereiro e com fortes chuvas nos dias 12 e 13 desse mesmo mês (Governo do Estado de São Paulo, 2014).

Entre os anos de 1991 e 2012, foram registrados oito eventos de inundação e 16.950 habitações e sistemas de infraestrutura afetados. Durante este período, as águas da bacia hidrográfica de contribuição da região da subprefeitura de São Miguel Paulista foram responsáveis por inundações graduais que provocaram o aumento do nível do rio Tietê sobre a área urbana. Em 2010, foram registrados os seguintes números: 622 desabrigados, quatro óbitos e 41.571 habitantes afetados (CEPED, 2013).

O ano de 2012 foi marcado pela quantidade de vezes que a prefeitura de São Paulo decretou situação de emergência devido aos danos por inundações, foram seis vezes no decorrer deste ano (CEPED, 2013). No ano de 2013 foram registrados 14 óbitos, dois feridos, três desalojados e uma unidade habitacional deteriorada (S2iD, 2022).

Mapas que indicam a suscetibilidade a inundação são ferramentas importantes na elaboração de planos de gestão de risco mais eficientes (Norallahi e Kaboli, 2021), indicando áreas de maior perigo e fornecendo suporte para a elaboração de mapas de risco. No entanto, experiências anteriores de gerenciamento de risco mostram que os mapas de suscetibilidade frequentemente utilizados não acompanham o processo rápido e desordenado de ocupação urbana.

Por outro lado, com a maior disponibilidade de dados de Sensoriamento Remoto e o auxílio de Sistemas de Informações Geográficas (SIG), a qualidade dos mapas de suscetibilidade à inundação tem sido melhorada (Pourghasemi et al., 2020). Nas últimas décadas, com o avanço da Inteligência Artificial (IA), métodos de predição utilizando algoritmos de Aprendizado de Máquina (*Machine Learning* - ML) têm apresentado resultados

confiáveis no mapeamento da suscetibilidade a inundação (Bui et al., 2019; Taromideh et al., 2022). A capacidade desses modelos em fornecer resultados altamente precisos se deve principalmente à forma como esses algoritmos são treinados, pois consideram locais onde o fenômeno analisado já ocorreu, estabelecendo uma relação entre a presença do fenômeno e as características dos fatores geográficos (Abedi et al., 2022).

Os algoritmos de ML têm sido amplamente aplicados no campo das ciências hidrológicas, como métodos baseados em dados (*data-driven*), os algoritmos são predominantemente para estimar o perigo de inundação a partir de registros históricos (Mobley et al., 2021). Os principais algoritmos utilizados no ML são Redes Neurais Artificiais (*Artificial Neural Network* - ANN), Máquina de Vetores de Suporte *(Support Vector Machine* - SVM) e Floresta Aleatória *(Randon Forest* - RF). Estudos anteriores mostraram que a RF se mostrou eficiente em relação à acurácia em modelos de suscetibilidade a inundação quando comparada a outros algoritmos (Abedi, 2022; Abu El-Magd. 2022; Taalab et al., 2018).

Nos modelos de suscetibilidade baseados em ML, diversos fatores condicionantes são usados. A maioria deles são derivados do Modelo Digital de Terreno (MDT). Hawker et al. (2018) salientam que a qualidade dos modelos de suscetibilidade a inundação aumenta com o uso de MDT mais detalhados, principalmente quando se trata de ambientes urbanos. Nesse sentido, a tecnologia *Light Detection and Ranging* (LiDAR) permite obter um MDT de alta resolução (menor que um metro) e precisão. Outro fator condicionante importante para o modelo de suscetibilidade a inundação é a precipitação. O mapa de precipitação, anual ou sazonal, contribui para o entendimento dos padrões de chuva em uma determinada região a partir de análises temporais e espaciais de dados coletados. Para a interpolação dos dados, a krigagem ordinária usando método geoestatístico podem apresentar melhores resultados quando comparado com métodos tradicionais de interpolação, como por exemplo, o Inverso do Quadrado da Distância (*Inverse Distance Weighted – IDW)* (Yanto et al., 2022; Aparecido et al., 2022; Simões et al., 2016; Mello et al., 2003). O método geoestatístico estima valores em locais não amostrados com base em observações disponíveis em locais amostrados (Landim, 2002), além de possibilitar a quantificação da incerteza para a área interpolada (Xavier Junior et al., 2020). A técnica é útil quando se trata de dados georreferenciados, nos quais a variabilidade espacial é uma consideração importante.

Dessa forma, a motivação do presente estudo advém da necessidade de existir novas metodologias de análises de suscetibilidade de inundação em áreas urbanas e densas, como a

cidade de São Paulo, onde as alterações das canalizações de rios e córregos realizadas no passado, não atendem atualmente o escoamento superficial das águas. Assim, diante do contexto exposto, este estudo tem como principal objetivo a aplicação de ML, por meio do algoritmo RF, na análise de áreas suscetíveis a inundação em uma região da cidade de São Paulo, utilizando registros históricos de inundação e alagamentos no período de 2013 a 2023, um MDT derivado de dados LiDAR de alta resolução para representar a superfície topográfica da região e um mapa da distribuição da precipitação média anual gerado utilizando geoestatística. Os objetivos específicos são: i) Verificar os fatores condicionantes que mais contribuem para o modelo de suscetibilidade a inundação; ii) Analisar o desempenho do modelo de suscetibilidade à inundação considerando conjuntos diferentes de registros de inundação e alagamentos e restrição para a distribuição de pontos de não inundação/alagamentos e; iii) Comparar os mapas de suscetibilidade à inundação gerados com os mapas geotécnicos disponíveis.

2 ARTIGOS

2.1 Artigo – Carvalho, B.C.S.; Mendes, T. S. G.; Simões, S. J. C. Distribuição Espacial das Chuvas na Cidade de São Paulo utilizando a Geoestatística / *Rainfall Spatial Distribution in the city of Sao Paulo using geostatistics.*

Artigo elaborado de acordo com as normas da Revista Brasileira de Ciências Ambientais. Enviado para publicação em: 13/01/2024 conforme Anexo A.

ABSTRACT

In the study of susceptibility to flooding and landslides, spatialized rainfall data, in raster format, is necessary as input data on the conditioning factors of these events. To this end, the study is based on historical precipitation series from meteorological stations located within a 20 km radius of the city of São Paulo. Average annual rainfall data from a period of thirty years (1985 – 2014) were used, which were obtained from the Hydrological Information System (HidroWEB) of the National Water Agency (ANA). In order to estimate the average annual rainfall in the capital of São Paulo, the geostatistical method of ordinary kriging was used. The spatialization and interpolation of the data show that the highest precipitation values are located in the southern region of the city, while the lowest are located in the northwest and northeast regions.

Keywords: *Average annual precipitation, Kriging, Spatialization.*

RESUMO

No estudo de suscetibilidade a inundação e deslizamento de terra são necessários dados espacializados de chuvas, no formato raster, como dados de entrada dos fatores condicionantes destes eventos. Para tanto, o estudo baseia-se nas séries históricas de precipitação das estações meteorológicas localizadas dentro de um raio de 20 km da cidade de São Paulo. Utilizou-se dados de precipitação pluviométrica média anual de um período de trinta anos (1985 – 2014) que foram obtidos no Sistema de Informações Hidrológicas (HidroWEB) da Agência Nacional de Águas (ANA). Com o objetivo de estimar a precipitação média anual da capital paulista foi empregado o método geoestatístico de krigagem ordinária. A espacialização e interpolação dos dados mostram que os maiores valores de precipitação estão localizados na região sul da cidade, enquanto os menores estão localizados na região noroeste e nordeste.

Palavras-chaves: Precipitação média anual, krigagem. Espacialização.

1. INTRODUCTION

The study of rainfall stations contributes to the understanding of precipitation patterns in a given region based on temporal and spatial analyses of the collected data. The data collected is essential for understanding seasonal and annual rainfall patterns in a region, this allows you to identify long-term trends.

The increase in global temperature accelerates the hydrological cycle, intensifying rainfall and contributing to the various occurrences of extreme events (Heo et al., 2023). Extreme rainfall events cause serious injuries, deaths, and significant socio-economic impacts, such as damage to property and infrastructure of energy and transportation systems, affecting people's health and lives, and threatening the ecosystem.

The city of São Paulo, the most populous in Brazil (IBGE, 2022), is characterized by a large built-up and impervious area, factors that, associated with rainy periods, contribute to the occurrence of flooding, flooding, and landslides (Marengo et al., 2020). When these events affect the population of the city of São Paulo, the economic and social damage is incalculable.

Understanding the spatialization of precipitation contributes to large cities becoming more resilient and adapted to disaster risks. Thus, the integration of rainfall data with the Geographic Information System (GIS) allows for a more comprehensive spatial analysis, facilitating the identification of geographic patterns, areas prone to extreme events, and decision-making regarding land use and resource management.

Geostatistics is widely used in many areas of science and engineering. Belkhiri et al. (2020) evaluated the distribution of the groundwater quality index in the El Mila plain using kriging and co-kriging interpolations. Through kriging, Pouladi et al. (2019) mapped the soil organic matter content in the Vindum field in Denmark and Kumar et al. (2020) mapped the spatial distribution of traffic-induced pollution criteria in Delhi, India. De Assis et al. (2021) used kriging to analyze spatio-temporal rainfall variability in the Pajeú river basin. Silva Neto et al. (2020) mapping of intense rainfall in the State of Tocantis using disaggregation and geostatistics by the kriging method.

Previous studies show that the geostatistical kriging method showed better results when compared to Inverse Distance Weighted (IDW) statistical methods (Yanto et al., 2022; Aparecido et al., 2022; Simões et al., 2016; Mello et al., 2003). Therefore, in this study, ordinary kriging was used as a geostatistical method of spatial interpolation. This method

estimates values at unsampled sites based on observations available at sampled sites (Landim, 2002). The technique is useful when dealing with georeferenced data, in which spatial variability is an important consideration. The possibility of quantifying uncertainty for an interpolated area is also particularly useful (Xavier Junior et al., 2020).

In view of the above, the work aims to analyze the spatial distribution of rainfall in the city of São Paulo, using the geostatistical method of kriging for a representation closer to reality. The precipitation map obtained in this article is important data for use in the studies of flood and landslide susceptibility in the city of São Paulo.

2. MATERIALS AND METHODS

2.1 Study area

The city of São Paulo is in the State of São Paulo (Figure 1B) in southeastern Brazil (Figure 1A). With great economic importance, it is the Brazilian federative unit with the highest GDP (IBGE, 2022). Although the city of São Paulo has great economic capacity, the losses resulting from the rains are greater when compared to other cities in Brazil.

The territorial area is 1,521.202 km², and 92.6% of households have adequate sanitation, 74.8% of urban households on public roads are wooded and 50.3% of urban households on public roads have adequate urbanization (presence of culverts, sidewalks, paving and curbs) (IBGE, 2022). The municipality has a predominantly urban population and is estimated at 12,396,372 people and a high population density of 7,398.26 inhabitants/km² (IBGE, 2022).

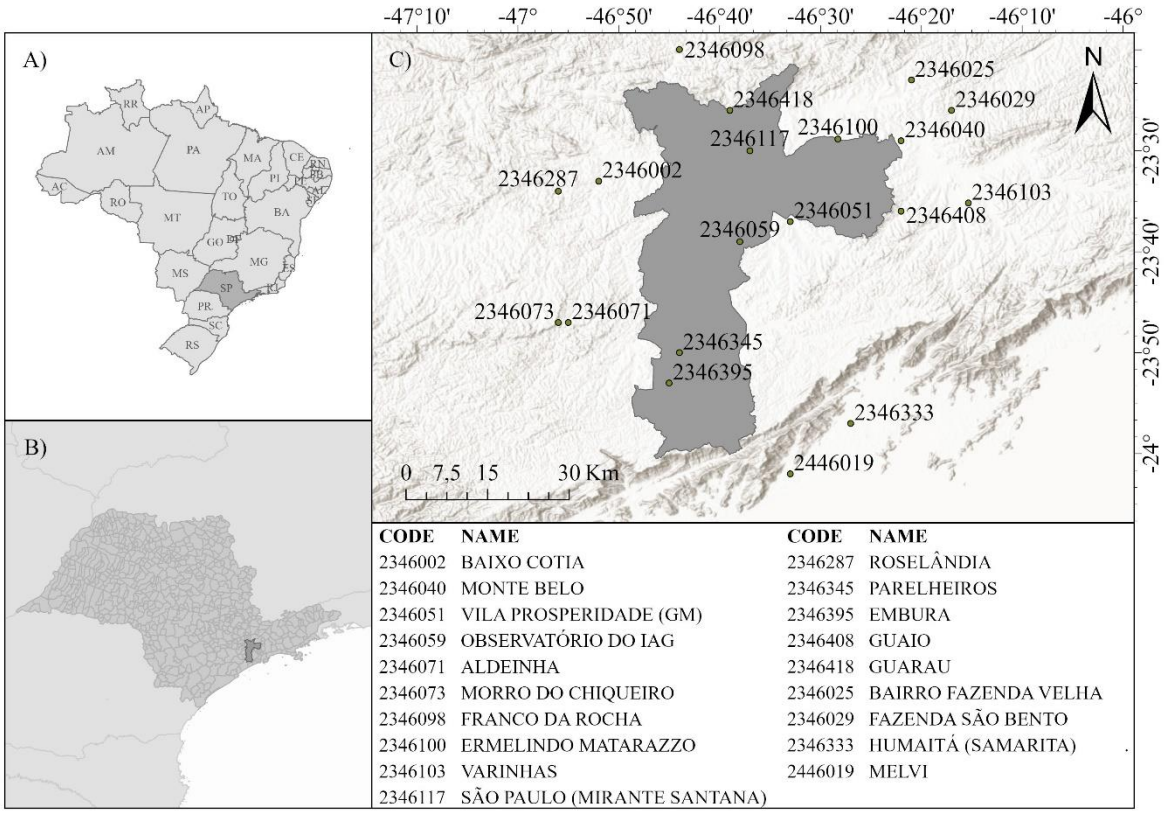


Figure 1 - *A) Location of the state of São Paulo in Brazil. B) Location of the city of São Paulo. C) Location of rainfall stations in study area considering the municipality and a buffer of 20 km.*

2.2 Rainfall data

The rainfall data were obtained from the website of the Hydrological Information System (HidroWEB, 2022) of the National Water Agency (ANA) and correspond to a series of monthly rainfall referring to the rainfall stations located within a buffer of 20 km from the city of São Paulo. The data correspond to the period from 1985 to 2014 and represent a 30-year time series. In view of the lack of continuous and consistent rainfall data, the stations that contain more than 7% of missing data were discarded. Thus, of a total of 243 stations within the study area, only 19 stations can be used, as shown in Figure 1C.

2.3 Geoestatistics

The term geostatistics was introduced by Matheron in 1963 (Cressie, 1988) and his main contributions were the development of the concept of semivariogram, which is a tool for modeling spatial dependence and autocorrelation in georeferenced data.

According to Camargo (1998), the variogram allows quantitatively representing the variation of a regionalized phenomenon in space and can be calculated through a sample $z(x_i)$*, i=1, 2, ..., n, as follows (Equation 1):*

$$2\hat{\gamma}(h) = \frac{1}{N(h)}\sum_{i=1}^{N(h)}[z((x_i) - z(x_i + h)]^2 \quad (1)$$

Where: $2\hat{\gamma}(h)$ *is the estimated variogram; N(h) is the number of pairs of measured values,* $z((x_i)$ and $z(x_i + h)$*, separated by a distance vector h; and* $z((x_i)$ and $z(x_i + h)$ *are values of the i-th observation of the regionalized variable, collected at the points* x_i *and* $x_i + h$*, separated by the distance h (Figure 2A).*

A) B)

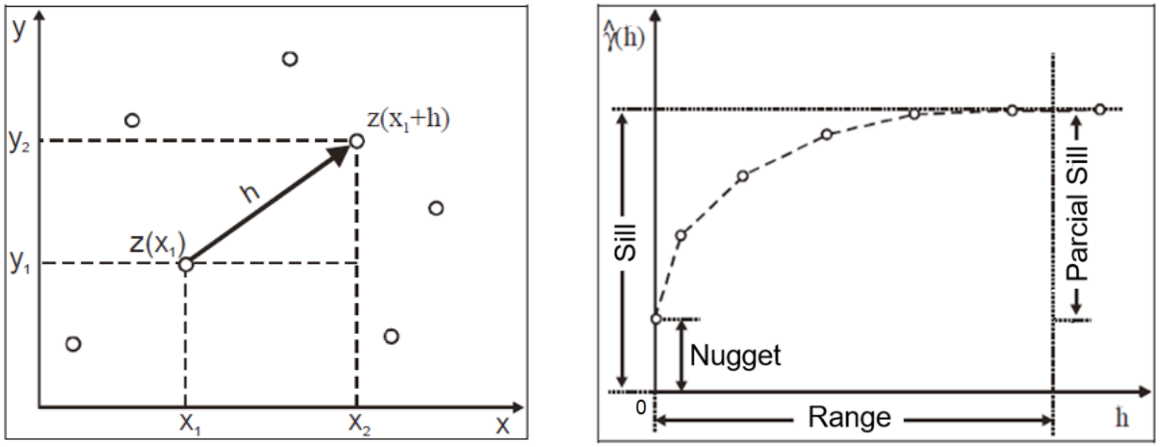


Figure 2 *- A) Sampling in two dimensions. B) Semivariogram parameters. Source: Camargo et al. (2004).*

According to Camargo (1998), it is expected that geographically closer observations are more similar and $\hat{\gamma}$ *increase with distance h. Figure 2B illustrates the parameters of the semivariogram, as follow: Range is the distance where the model first flattens and the samples are spatially correlated; Sill is the value on the y-axis that reaches the range, where*

there is no more spatial dependence on the samples; Nugget, in practice, is the measure that h tends to zero; and Partial sill is the difference between sill and nugget.

The value of the variable to be interpolated for a given point p is given by Equation 2.

$$Z(x_p) = \sum_{i=1}^{n} \lambda_i \cdot Z(x_i) \quad (2)$$

Where, $Z(x_p)$ *is the variable to be interpolated at the point* x_p*;* λ_i *is the is the weight of the i-th observation;* $Z(x_i)$ *is the value of the variable for the i-th observation collected at point* x_i*; and n is the number of neighboring observations used to interpolate the point.*

In the interpolation process, the weights are determined based on the distance between the observation and the point of interest, the spatial continuity and the geometric arrangement of the sampling set (Pereira et al., 2018). The geostatistics methods not only provide the interpolated values, but also the measurement of uncertainty. Thus, the geostatistical method used in the study differs from other interpolation methodologies because it provides the error associated with the estimated values (Landim, 2002).

According to Esri (2023), this measurement provides information about the possible outcomes for each location and not just an interpolated value. Thus, one of the main characteristics of geostatistics is to assume values, which are not necessarily the measured values, everywhere in the study area (Esri, 2023).

The Geostatistical Analyst Tool available in the ArcGIS Pro software was used for the proposed methodology, which consists of the following steps: statistical analysis, semivariogram generation and modeling, model validation, and kriging interpolation.

The statistical analysis of the variable was not performed in detail, this was done because the methodology would become impractical. The analysis of the linear regression graph was performed by selecting the best visual fit for later application of geostatistics.

The semivariogram was modeled and to choose the appropriate theoretical semivariogram it was necessary to adjust the curve to the empirical data, and a model type stable was adopted for this study, where the three parameters: nugget effect, the sill and the range, were adjusted by the software.

The evaluation of interpolation errors was performed by cross-validation, where the estimates are compared with the actual values at unsampled sites. This tool evaluates the

statistical model and consists of leaving out one data point at a time to determine how well that point can be estimated from the other data (Dubrule, 1983).

After the validation and choice of the model, i.e., after the calculation of the experimental semivariogram, the ordinary kriging was performed to perform the interpolations and the rainfall prediction map was constructed. According to Wackernagel (2003), ordinary kriging is the widely used kriging method, which consists of estimating a value at a point in a region using data in the vicinity of the estimation site. Finally, the error map was elaborated, which allowed the quantification of the uncertainty associated with the estimate.

3. RESULTS AND DISCUSSION

Figure 3 shows the linear regression between the measured and estimated values of the average annual rainfall in each rainfall station. The spatially dispersed point of the set shown in the graph refers to the Melvi rain station, number 2446019, located in the municipality of Praia Grande, on the south coast of São Paulo State.

The graph shows that the estimates are close to the true values and the gray (reference) and blue (simple linear regression) lines roughly agree. The intercept value equal to 10.9179 mm is considered small when compared to the attribute values.

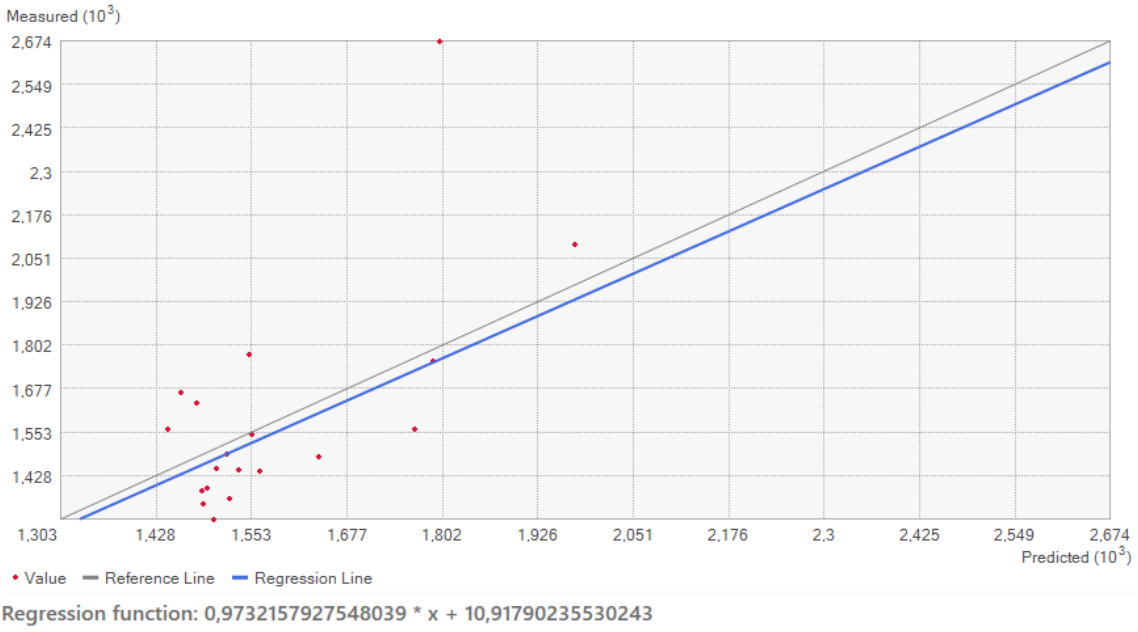


Figure 3 - *Linear regression graph between measured and estimated values.*

The parameters used as input data for the methodology and the estimates of errors for the choice of the theoretical semivariogram are presented in Table 1.

Table 1. *Method report and criteria used to choose the theoretical semivariogram through cross-validation.*

Method report	***Setting***
Method type	*Kriging Ordinary*
Output type	*Prediction*
Searching neighborhood	*Smooth*
Model type	*Stable*
Anisotropy	*Yes*
Error Estimates	***Values***
Mean	*-16.616*
Root-Mean-Square	*241.999*
Mean Standardized	*-0.061*
Root-Mean-Square Standardized	*1.063*
Average Standard Error	*216.974*

For Root-Mean-Square Standardized, the value was estimated at 1.063 (ideal values are close to one) and for Mean Standardized the estimated value was -0.061 (ideal values are close to zero). The optimal values for Root-Mean-Square and Root-Mean-Square Standardized occur when there is little difference between these errors.

From the spatialization and interpolation of the data (Figure 4A) it was observed that the highest precipitation values are in the southern region of the city, while the lowest are located in the northwest and northeast regions.

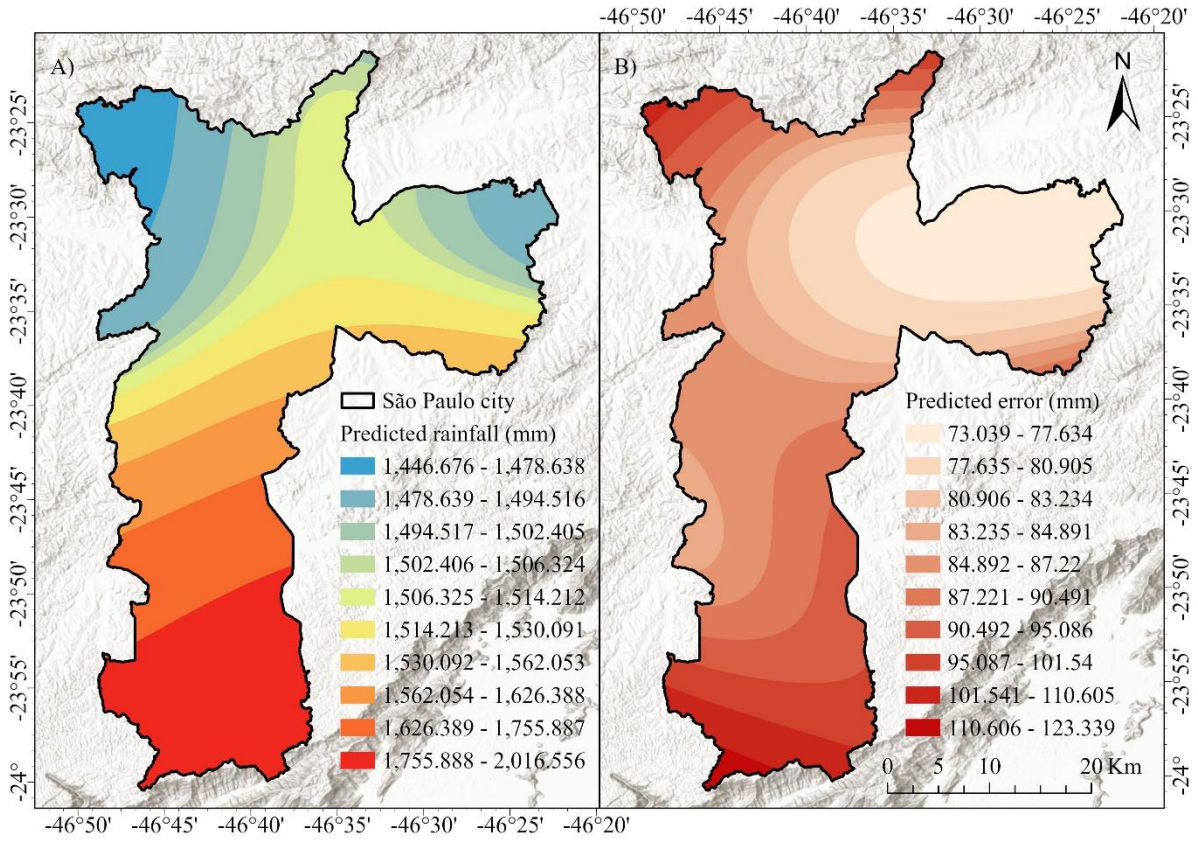


Figure 4 - *A) Map of the spatial distribution of the average annual rainfall of the city of São Paulo. B) Map of interpolation error estimates*

The southern region of the city of São Paulo is near the Serra do Mar, a stretch along the coast of São Paulo, which is characterized by the influence of orography on rainfall totals, being responsible for the high volume of rainfall in the region and is among the highest in Brazil (Pellegatti and Galvani, 2010).

Based on the map of interpolation error estimates (Figure 4B), it was found that the largest errors are associated with both areas with low and high rainfall densities, located, respectively, in the northwest and south regions of the city of São Paulo. When observing Figure 1C, it appears that those regions don't have rainfall stations with sufficient data for the analysis, creating areas with voids of sampled precipitation data. Aparecido et al. (2020) point out that Brazil still has few there are meteorological stations spread across the country and approximately 30% of the installed stations need maintenance for an accurate collection. Other problems are flaws, gaps and discontinuities in their historical time series (Oliveira Junior et al., 2022) that difficult climatology analysis in certain regions.

4. CONCLUSIONS

The advanced technology of geographic information systems extends the capacity for data analysis, contributing to a more comprehensive understanding of precipitation patterns. The generated precipitation raster can be used to represent a conditioning factor in the study of susceptibility to flooding and landslides.

The distribution of rainfall stations with consistent data in the city of São Paulo is not homogeneous and of the 19 stations studied, only six belong to the municipality. Therefore, rainfall prediction maps are a significant challenge in certain locations, as their accuracy can be affected by the lack of rainfall stations.

Thus, the error estimation map shows that the city of São Paulo needs densification of rainfall stations to improve the representativeness of the variable and thus contribute to the reduction of the dispersion of errors in rainfall prediction.

5. REFERENCES

Aparecido, L.E.D.O.; Moraes, J.R.D.S.C.D., Lima; R.F.D.; Torsoni, G.B., 2022. Spatial Interpolation Techniques to Map Rainfall in Southeast Brazil. Revista Brasileira de Meteorologia, v.37, 141-155. https://doi.org/10.1590/0102-77863710015

Belkhiri, L., Tiri, A., Mouni, L., 2020. Spatial distribution of the groundwater quality using kriging and Co-kriging interpolations. Groundwater for Sustainable Development, v. 11. https://doi.org/10.1016/j.gsd.2020.100473.

Camargo, E.C.G. Geoestatística: Fundamentos e Aplicações. Geoprocessamento em projetos ambientais (Accessed November 30, 2023) at: http://www.dpi.inpe.br/gilberto/tutoriais/gis_ambiente/5geoest.pdf.

Camargo, E.C.G.; Fucks, S.D.; Câmara, A.G., 2004. Análise espacial de superfícies. Brasília: Embrapa.

Cressie, N., 1988. Spatial Prediction and Ordinary Kriging. Mathematical Geology, v. 20, (4).

De Assis, J.O.M.; Menezes, A.F.; de Souza, W.M.; Sobral, M.D.C.M., 2021. Methods to analyze spatio-temporal rainfall variability: application to the Pajeú river basin, Pernambuco, Brazil. Brazilian Journal of Environmental Sciences (RBCIAMB), v. 56(4), 577-588.

Dubrule, O., 1983. Cross validation of kriging in a unique neighborhood. Mathematical Geology, v.15, 687–699. https://doi.org/10.1007/BF01033232

ESRI. (Accessed June 20, 2023) at: https://pro.arcgis.com/en/pro-app/latest/help/analysis/geostatistical-analyst.

Heo, S.; Park, S.; Lee, D.K., 2023. Multi-hazard exposure mapping under climate crisis using random forest algorithm for the Kalimantan Islands, Indonesia. Sci Rep, v. 13, 13472. https://doi.org/10.1038/s41598-023-40106-8.

Instituto Brasileiro de Geografia e Estatística. IBGE. Open data portal (Accessed June 20, 2022) at: https://cidades.ibge.gov.br/brasil/sp/sao- paulo/panorama.

Kumar, A., Mishra, R. K., Sarma, K., 2020. Mapping spatial distribution of traffic induced criteria pollutants and associated health risks using kriging interpolation tool in Delhi, Journal of Transport & Health, v. 18, 100879. https://doi.org/10.1016/j.jth.2020.100879.

Landim, P.M.B., Sturaro, J.R., 2002. Krigagem Indicativa aplicada à elaboração de mapas probabilísticos de riscos. Geomatemática, Texto Didático 6, DGA,0 IGCE, UNESP/Rio Claro, 2002. (Accessed April 16, 2022) at http://www.rc.unesp.br/igce/aplicada/textodi.html.

Marengo, J. A., Alves, L. M., Ambrizzi, T. et al., 2020. Trends in extreme rainfall and hydrogeometeorological disasters in the Metropolitan Area of São Paulo: a review. Ann N Y Acad Sci. https://doi. org/10.1111/nyas.14307.

Mello, C.R.; Lima, J.M.; Silva, A.M.; Mello, O.J.M.; Oliveira, M.S., 2003. Krigagem e inverso do quadrado da distância para interpolação dos parâmetros da equação de chuvas intensas. Revista Brasileira de Ciência do Solo, v. 27, 925-933.

Oliveira-Júnior, J.F.; Correia Filho, W.L.F.; Monteiro, L.S.; Shah, M.; Hafeez, A.; de Gois, G.; Lyra, G.B.; de Carvalho, M.A.; Santiago, D. B.; de Souza, A.; Mendes, D.; Costa, C.E.A.S.; Blanco, C.J.C.; Zeri, M.; Pimentel, L.C.G.; Jamjareegulgarn, P.; da Silva, E. B., 2022. Urban rainfall in the Capitals of Brazil: Variability, trend, and wavelet analysis. Atmospheric Research,v. 267, 105984. https://doi.org/10.1016/j.atmosres.2021.105984.

Pellegatti, C. H. G.; Galvani, E., 2010. Avaliação da precipitação na Serra do Mar – SP em eventos de diferentes intensidade e duração. GEOUSP Espaço e Tempo (Online), v. 14, (1), 147-158. (Accessed December 20, 2023) at: https://www.revistas.usp.br/geousp/article/view/74160.

Pereira, V.A.S.; Pugliesi, E.A.; Flores, E.F.; Camargo, P.D.O., 2018. Krigagem ordinária e visualização de incertezas aplicadas no monitoramento de irregularidades ionosféricas no Brasil. Revista Brasileira de Cartografia, v. 70(3), 967-996.

Pouladi, N., Møller, A. B., Tabatabai, S., Greve, M. H., 2019. Mapping soil organic matter contents at field level with Cubist, Random Forest and kriging. Geoderma, v. 342, 85-92. https://doi.org/10.1016/j.geoderma.2019.02.019.

Silva Neto, V.L.; Viola, M.R.; Mello, C.R.; Alves, M.V.G.; Silva, D.D.; Pereira, S.B., 2020. Mapeamento de chuvas intensas para o Estado do Tocantins. Revista Brasileira de Meteorologia, v. 35, 1-11. http://dx.doi.org/10.1590/0102-7786351017

Simões, S.J.C.; Gomes, L.; Mendes, R.M.; Mendes, T.S.G., 2016. SIG e modelos de escorregamentos: avaliando métodos para reduzir as incertezas de dados de solo e precipitação. Revista Brasileira de Cartografia, v. 9, 1737-1746. https://doi.org/10.14393/rbcv68n9-44440

Sistema de Informações Hidrológicas. HidroWEB. Open data portal (Accessed September 18, 2022) at: www.snirh.gov.br/hidroweb/serieshistoricas/.

Wackernagel, H., 2003. Ordinary Kriging. In: Multivariate Geostatistics. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-662-05294-5_11

Xavier Júnior, S.F.A.; Jale, J.D.S.; Stosic, T.; Santos, C.A.C.D.; Singh, V.P., 2020. Precipitation trends analysis by Mann-Kendall test: a case study of Paraíba, Brazil. Revista Brasileira de Meteorologia, v.35, 187-196.

Yanto, Apriyono, A., Santoso, P.B. et al., 2022. Landslide susceptible areas identification using IDW and Ordinary Kriging interpolation techniques from hard soil depth at middle western Central Java, Indonesia. Natural Hazards 110, 1405–1416. https://doi.org/10.1007/s11069-021-04982-5

2.2 Artigo – Carvalho, B. C. S.; Mendes, T. S. G.; Simões, S. J. C.; Makita, F. H. Mapeamento de suscetibilidade a inundações e alagamentos utilizando Floresta Aleatória na cidade de São Paulo, Brasil / ***Flooding and flash flood susceptibility mapping using Random Forest in São Paulo city, Brazil.***

Artigo elaborado de acordo com as normas do Periódico Natural Hazards. Enviado para publicação em: 16/01/2024 conforme Anexo B.

ABSTRACT

The increase in the frequency of heavy rains due to the effects of climate change and the accelerated process of urban expansion, such as soil compaction and sealing, are factors that contribute to the incidence of natural disasters such as flooding, flooding and landslides on slopes. Currently, flood events are the main cause of damage worldwide. These events present an upward trend when compared to the annual average of events in the past and can be worsened in climate variation scenarios. Considering one of the most important cities of Brazil, São Paulo has peculiar characteristics related to densely settlement areas over and bordering rivers. In this sense, this work aims to evaluate the predictive capacity of the Random Forest algorithm to flood susceptibility model in a region of São Paulo municipality based on two flood inventories (one with 75 flood points and the other with 578 flood and flash flood points) and to compare the two models with an available geotechnical map. MDT-derived LiDAR data was used to generate most of the conditioning factors. A total of 10 conditioning factors were used. The model generated from the flood and flash flood points better showed the highest performance in terms of accuracy, AUC, sensitivity and specificity. The results of the models reveal that the flooding susceptibility areas go beyond the classes of floodplain and area subject to flooding classes presented on the geotechnical map.

Keywords: Machine learning. Random Forests. Natural Disasters. Flooding. Sao Paulo.

1. INTRODUCTION

According to the International Panel of Climate Change (IPCC) report, between 1970 and 2019, flood-related events accounted for 44% of all disasters and 31% of all losses economic (IPCC 2022). Cities are increasingly facing the effects of changes in natural land cover for built-up areas by increasing impervious surfaces (Tomás et al. 2022). Changes in land cover and redirection of water flow, associated with the effects of climate change, have been responsible for the increase in flood events (Zope et al. 2016). The increase in global temperature accelerates the hydrological cycle, intensifying rainfall and contributing to the various occurrences of extreme flooding events (Hirabayashi et al., 2013). Locally, intense urbanization can increase average precipitation resulting in increased intensity of surface runoff (IPCC 2021).

Floods often destroy people's homes and properties, cause displacement, loss of life, pose health risks, and negatively affect people's livelihoods (Mambele et al. 2022). In Brazil, due to their social and socioeconomic characteristics, large cities are characterized by high demographic density and the disordered history of occupation of the territory. The expansion of large urban centers, accompanied by increased soil sealing, occupation of floodplains and river basin areas, silting of rivers and increasingly severe rainfall events, intensified flooding events. Many cities grew over floodplain areas, channeling rivers and occupying their sources, resulting in potentially dangerous spaces, especially in the rainy season (Valverde et al. 2023).

An example is the city of São Paulo, which has the largest GDP in Brazil and is one of the most densely occupied areas in the country (IBGE 2022). When floods hit the population of the capital of São Paulo, the economic and social losses are incalculable. From this perspective, it is necessary for the city to have efficient risk management and an effective monitoring policy, to develop actions and plans to prevent and reduce risks.

Maps that indicate susceptibility to flooding are important tools in developing more efficient risk management plans (Norallahi and Kaboli 2021), indicating areas of greatest danger and providing support for the preparation of risk maps. However, previous risk management experiences show that frequently used susceptibility maps do not keep up with the rapid and disorderly process of urban occupation.

On the other hand, with greater availability of remote sensing data and the aid of Geographic Information Systems (GIS), the quality of flood susceptibility maps has been

improved (Pourghasemi et al. 2020). In recent decades, with the advancement of Artificial Intelligence (AI), prediction methods using Machine Learning (ML) algorithms have presented reliable results in mapping flood susceptibility (Bui et al. 2019; Taromideh et al. 2022). The ability of these models to provide highly accurate results is mainly due to the way these algorithms are trained, as they consider locations where the analyzed phenomenon has already occurred, establishing a relationship between the presence of the phenomenon and the characteristics of geographic factors (Abedi et al. 2022).

ML is an alternative to traditional methods for susceptibility prediction models, this allows the analysis of the territory subjected to frequent changes, the development of susceptibility maps with greater reliability and precision and consequently the improvement of response for a system natural disaster.

ML algorithms have been applied in the field of hydrological sciences, as ML can be employed to estimate flood danger through data with historical records (Mobley et al. 2021).

The main algorithms used in ML are Artificial Neural Network (ANN), Support Vector Machine (SVM) and Random Forest (RF). Previous studies have shown that RF has been shown to be efficient in relation to accuracy in flood prediction models when compared to other ML algorithms (Abedi 2022; Abu El-Magd. 2022; Taalab et al. 2018).

It is in this sense that this study has as its main objective the application of ML, through the RF algorithm in mapping areas susceptible to flooding in a region of the city of São Paulo, using historical records of floods in the period from 2013 to 2023 and a Digital Terrain Model (DTM) derived from high-resolution LiDAR data to represent the topographic surface of the region. The specific objectives are: i) Verify which conditioning factors contribute most to flood events in the susceptibility model; ii) Analyze the performance of the flood susceptibility model considering a restriction for the distribution of non-flood points; and iii) Compare the generated flood susceptibility maps with geotechnical map available.

2. MATERIALS AND METHODS

2.1 Study area

The city of São Paulo, located in the State of São Paulo, grew close to the plains of the Tietê, Pinheiros and Tamanduateí rivers in the soil of the São Paulo sedimentary basin. Even though urban occupation depended on the presence of rivers nearby, from the 1970s onwards, the city began to spread into the Precambrian rock terrain, also occupying the hilly reliefs.

Currently, due to the intense occupation of the basins, the Tietê, Pinheiros and Tamanduateí rivers cross regions of high demographic density and intense and undisciplined land occupation, such as the city of São Paulo and adjacent municipalities, favoring the occurrence of floods (DAEE 2009).

The territorial area (2021) is 1,521.202 km², 92.6% of households have adequate sanitary sewage, 74.8% of urban households on public roads have trees and 50.3% of urban households on public roads have adequate urbanization (presence of manholes, sidewalks, paving and curbs) (IBGE 2022).

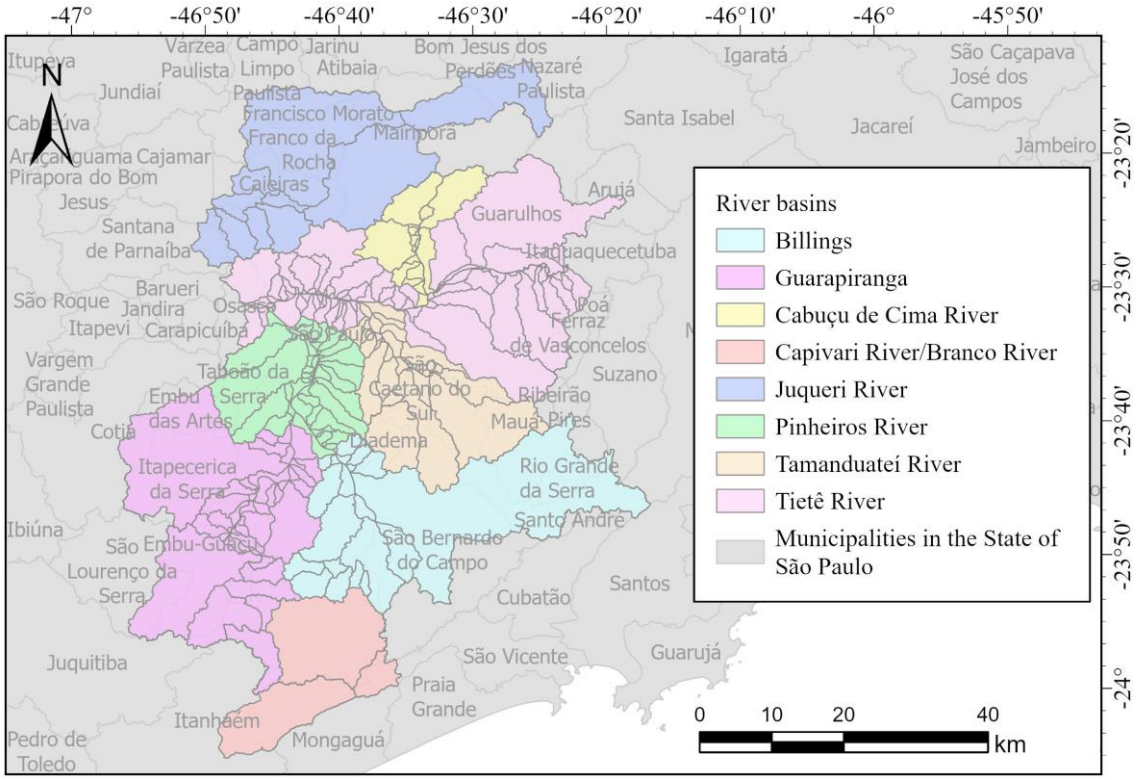
The municipality has a predominantly urban population and is estimated at 12,396,372 people and a high demographic density of 7,398.26 inhabitants/km² (IBGE 2022). According to IBGE data, the estimated population of the capital of São Paulo is more than 12 million, with 674,329 people exposed in areas at risk to floods and landslides.

Floods are one of the main causes of losses in the city of São Paulo, the total material damages recorded from 1995 to 2019 are approximately 4.7 million reais, with private losses being U$167,628.35 and public losses being U$ 924,026.06 (CEPED 2013).

Given these data, for this work, an urban area in the municipality was selected based on the following criteria:

- *i) Selection of the river basin (Figure 1) close to the banks of the Pinheiros and Tietê rivers, since major highways border these rivers, and the interruption of these routes causes great economic and social damage that has repercussions in other regions of the country;*
- *ii) Areas with high population density that generally have high soil sealing and urban drainage failures.*

Figure 1 –Map of the river basins in the city of São Paulo



Source: Prepared by the author.

2.2 Materials and dataset

The proposed methodology was developed using ArcGIS Pro (ESRI proprietary) software for producing maps, allowing the creation and management of geographic data in a single project; Whitebox Geospatial Inc. (Lindsay 2014) is software for geospatial analysis using Whitebox library (WbT); Scikit-learn library that is an open source machine learning library, in Python language, for predictive data analysis (Pedregosa et al., 2011); and the code editor Visual Studio Code (Microsoft proprietary).

A GeoSampa database of the municipality of São Paulo (available at: https://geosampa.prefeitura.sp.gov.br/PaginasPublicas/_SBC.aspx) that brings together georeferenced data on various topics in the municipality was used in this work. Data in GeoSampa is open and available in vector, raster, and point cloud file formats. Data referring to the Digital Terrain Model (MDT) is available as a point cloud with 3D coordinates obtained from Light Detection and Ranging (LiDAR) technology. The themes

used to characterize the study area are presented in Table 1 and are separated into groups and subgroups.

Table 1 – Georeferenced data available in GeoSampa database used in the methodology

Group	*Subgroup*	
Administrative limits	*District*	
	Subprefectures	
Physical environment	*Hidrography*	*River Basin*
		Floodable area
	Geothechnical map	
Protection and civil defense	*Flood occurrences*	
	Rain gauges	
LiDAR	*MDT*	

Source: Prepared by the author.

Moreover, for the history of daily rainfall, tables available at the National Water and Basic Sanitation Agency (ANA) were collected.

2.2 Methodology

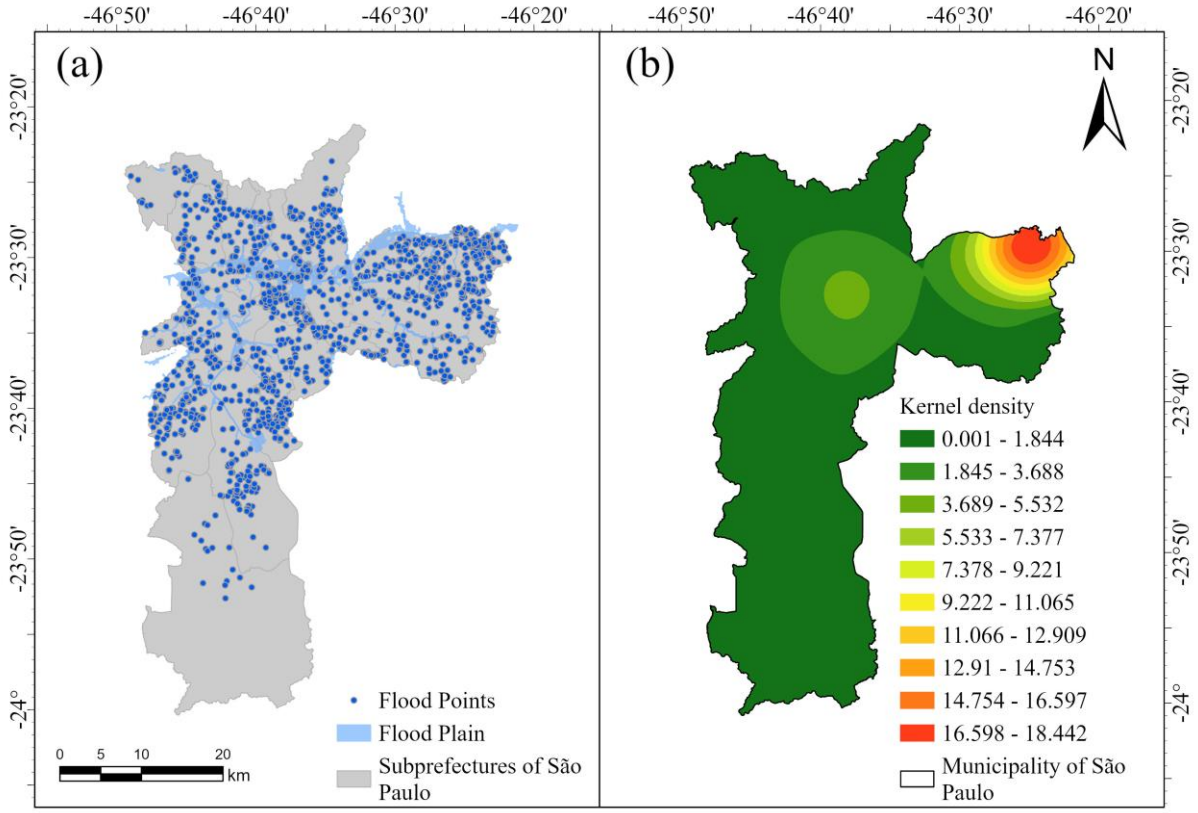
The methodological structure comprises the six steps described below.

2.2.1 Delimitation of the study area in municipality of São Paulo

In the period between 2013 and 2023, 4,161 flooding occurrences were recorded in the municipality of São Paulo (Figure 2a). These data are georeferenced points in SHP format available in the GeoSampa database (Table 1). To represent and analyze the spatial distribution of these flood events a density map by kernel estimator was carried out (Figure 2b). A radius of 10 km was used as a parameter to control the "smoothing" of the surface.

The density map allowed the identification that the central and eastern regions of the municipality are more critical areas that have the highest densities of flood points (Figure 2b).

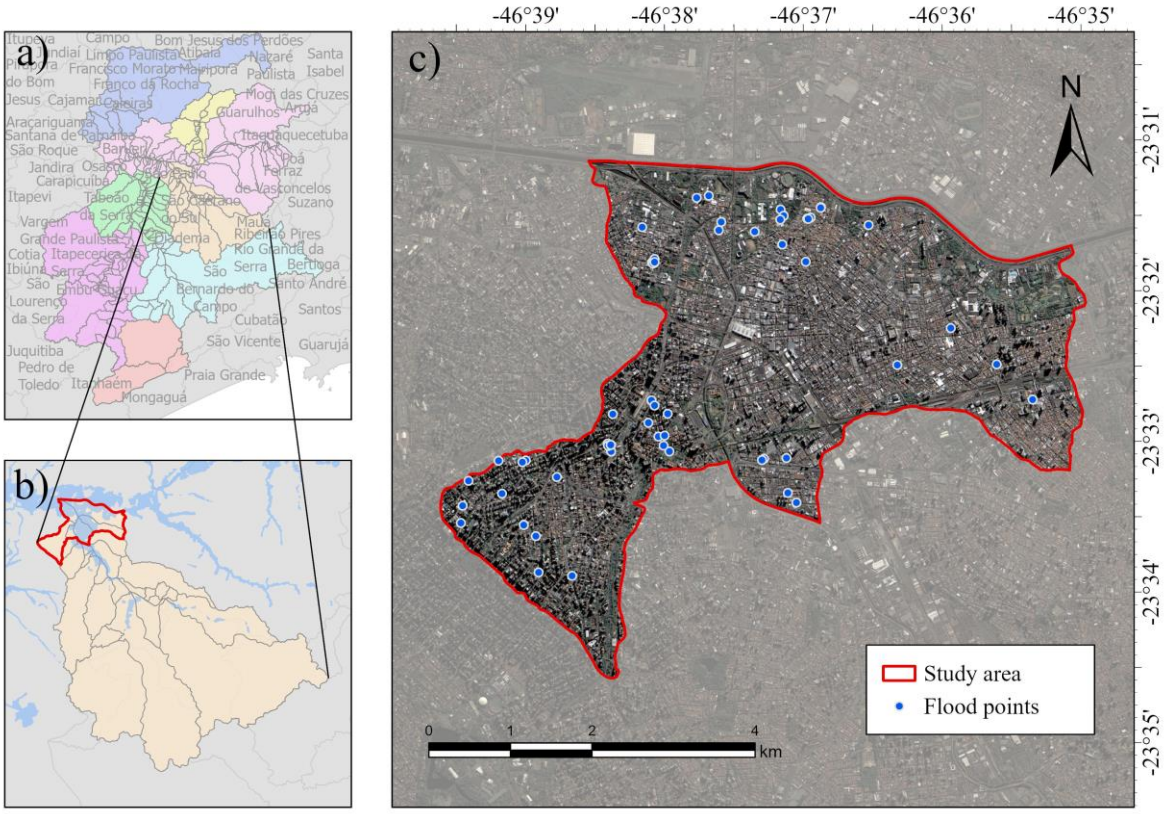
Figure 2 – (a) Flood records and (b) flood point density map using the kernel estimator for the municipality of São Paulo.



Source: Prepared by the author.

Although Figure 2b shows a higher density of points in the eastern region of the municipality, this area comprises the Tietê Ecological Park, an environmental protection area situated in the floodplain of the Tietê River. Thus, it was decided to select a basin in the central region of the municipality, also with a high density of points. The selected study area is in a densely occupied urban area in the city of São Paulo (Figure 3), which comprises six sub-basins of the downstream part of the Tamanduateí River Basin and was defined based on the presence of flood points at the interface of the Tamanduateí River flow, which is a tributary of the Tietê River.

Figure 3 – Location of flood points in the study area (a) Hydrographic sub-basins of the city of São Paulo, (b) Tamanduateí River Basin and (c) Study area delimited.



Source: Prepared by the author.

2.2.2 Extraction of conditioning factors

In this study, the conditioning factors were selected based on an analysis of the literature on flood susceptibility models (Chen et al. 2020; Ha et al. 2021; Islam et al. 2021) and the availability of existing data.

The MDT was generated from LiDAR data using 3D point cloud manipulation tools available in ArcGIS Pro. The conditioning factors were derived from the MDT with a spatial resolution of 0.5 m. The conditioning factors Aspect, Slope, and Topographic Wetness Index (TWI) were obtained using the Whitebox Geospatial Analysis Tools (GAT) library with Python language. The other conditioning factors (Altitude, Relief amplitude, Curvature, Plan curvature, Profile curvature) were generated in ArcGIS Pro.

Relief amplitude is the difference between the highest and lowest altitude in a given region (Christofoletti 1980) and was obtained with a 3x3 dimension mask.

Aspect is the azimuth tilt in degrees clockwise from North. Its hydrological

importance is related to solar irradiation (Moore et al. 1991) and the direction of surface flow. Shaded areas maintain high ground humidity, affecting water flow in flooded areas (Islam et al., 2021).

Curvature is the shape of the soil surface and depending on its characteristics, it can contribute to the flooding event in each area (Khosravi et al. 2019). Curvature, when combined with other surface parameters, contributes to the identification and classification of relief forms, as well as understanding the impacts of gravity and erosion.

Plan curvature is contouring curvature. This type of curvature is associated with flow and water content in the soil (Moore et al. 1991). When the curvature is convergent, the water flow is concentrated and when it is divergent, the flow is dispersed.

Profile curvature has concave or convex inclinations. This type of curvature is related to the acceleration and deceleration of water flow and the rate of erosion and soil deposition on slopes (Moore et al. 1991). Positive curvature indicates that the surface is convex in the direction of the slope, favoring the acceleration of surface flow and erosion. Negative curvature indicates that the surface is concave contributing to slow surface flow and deposition. Values equal to zero indicate that the surface is flat.

Slope corresponds to the slope gradient and is one of the most important factors for studying susceptibility to floods (Abedi et al. 2022; Costache and Bui 2019), since the steeper the slope, the greater the water flow and flood threats in the plains (Antzoulatos et al. 2022; Tehrany et al. 2015; Wang et al. 2015).

The TWI indicates the conditions of soil cover (Bui et al. 2020) and is defined as the ability of an area to be saturated up to the surface, that is, the effect of topography on flow direction (Collins et al 2022) through the contribution area and local slope characteristics. The TWI is given by Equation 1 (Beven and Kirkby 1979):

$$TWI = Ln\left(\frac{A_s}{tan\ (slope)}\right), \quad (1)$$

where: As is the specific catchment area (the total area contributing runoff to a particular location). It is obtained through the FD8 flow accumulation algorithm of the WbT library in a hydrologic modeling applied to MDT; and Slope is the rate of variation of the surface in the horizontal and vertical directions.

The Lithological map provides geological information about the study area. The map was extracted from the Geological Map of the Metropolitan Region of São Paulo in the CPRM - Geological Survey of Brazil database with a scale of 1:100,000. The information contained in the geotechnical and lithological maps makes it possible to evaluate the potential for flood events.

The rainfall data were obtained from the ANA Hydrological Information System (HidroWEB) website and correspond to a series of monthly precipitation referring to the rainfall stations located within a 20 km radius of the city of São Paulo. The data correspond to the period from 1985 to 2014 and represent a historical series of 30 years. Given the lack of continuous and consistent rainfall data, stations containing more than 7% of missing data were discarded, totaling data analysis from 19 stations in this study. The rainfall map was generated from the interpolation of average annual precipitation data using the kriging method using the ArcGIS Pro software and cropped to the study area.

After creating each of the raster files that represent the conditioning factors, the classes were defined as follows: Relief amplitude and TWI were classified using the natural breaks classification method. The Curvature, Profile curvature and Plane curvature were classified according to Chen et al. (2020), Aspect and Slope were classified according to Embrapa (2009), Lithological map were classified considering the geological categories; and the Precipitation were classified based in equal intervals. Tables 2 and 3 (Subsection 3.2) presents intervals for each conditioning factor.

2.2.3 Flood Event Inventory

The flood inventory is the main step for modeling the flood susceptibility map, as the RF prediction model statistically assumes that future floods occur in areas with conditions similar to those in the past (Bui et al. 2020; Costache et al. 2020).

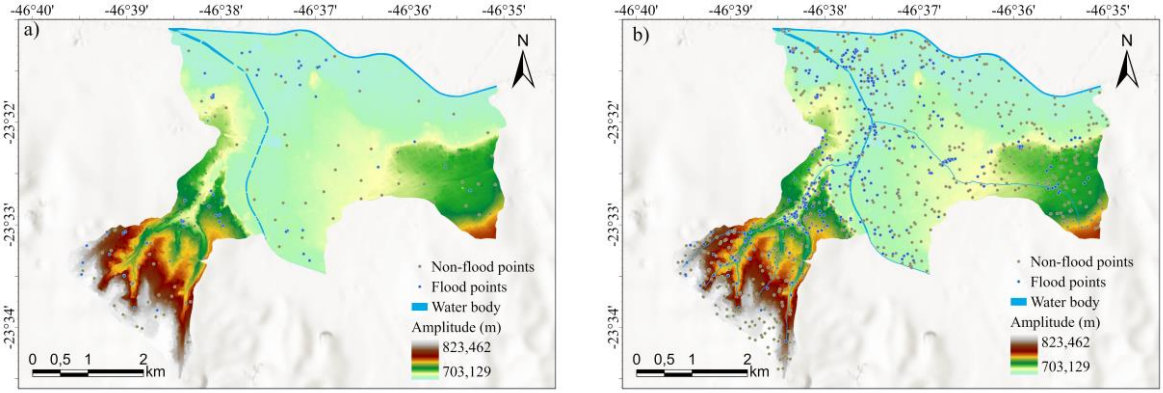
In this step, flood occurrence points in the delimitated area were used to generate two flooding inventory datasets. Inventory A contains 75 flood points and inventory B with 578 flood and flash flood points, both with points recorded between the period 2013 and 2023.

Within the selected area, non-flood points on slopes greater than 3° were randomly generated. For inventory A, 75 points of non-flooding points were randomly generated, and for inventory B, 578 points. Thus, the dataset of inventory A (Figure 4a) contains a total of

150 points (75 flood and 75 non-flood) and inventory B (Figure 5b) contains a total of 1156 points (578 flood and 578 non-flood). Each inventory is divided into training and validation samples. For the parameterization of this step, the samples have the following proportion: 70% are used for model training and 30% are used to analyze the performance of the final susceptibility model.

Figure 4 – Flood point and non-flood point dataset: a) dataset of inventory A and b) dataset of inventory B

a) b)



Source: Prepared by the author.

2.2.4 Frequency ratio (FR)

Frequency ratio (FR) is a bivariate statistical analysis method (Ullah and Zhang 2020), that is, the probability of occurrence and non-occurrence of any event in a delimitated area.

Often used in various studies such as earth sciences, environmental sciences, risk assessments and physical sciences (Rehman et al. 2022), FR allows for analyzing the relation between conditioning factors and flood events (Sarkar and Mondal, 2020) through the attributes (subclasses) of a conditioning factor (Amplitude, Aspect, Slope etc.).

Thus, FR is defined as the ratio between the flood area and the total study area (Equation 2) (Arora et al. 2019; Rahman et al. 2019):

$$FR = \frac{AD}{BC} \quad (2)$$

Where: A is the number of flood pixels for each subclass conditioning factor; B is the total number of flood pixels in the study area; C is the number of pixels in the subclass area of each conditioning factor; and D is the total number of pixels in the study area.

The Normalized Frequency Ratio (NFR) is obtained by Equation 3. The NFR values reflect the probability of a flood occurring in a pixel and vary from 0 to 1, thus, the probability of a flood occurring increases proportionally the higher the value of NFR.

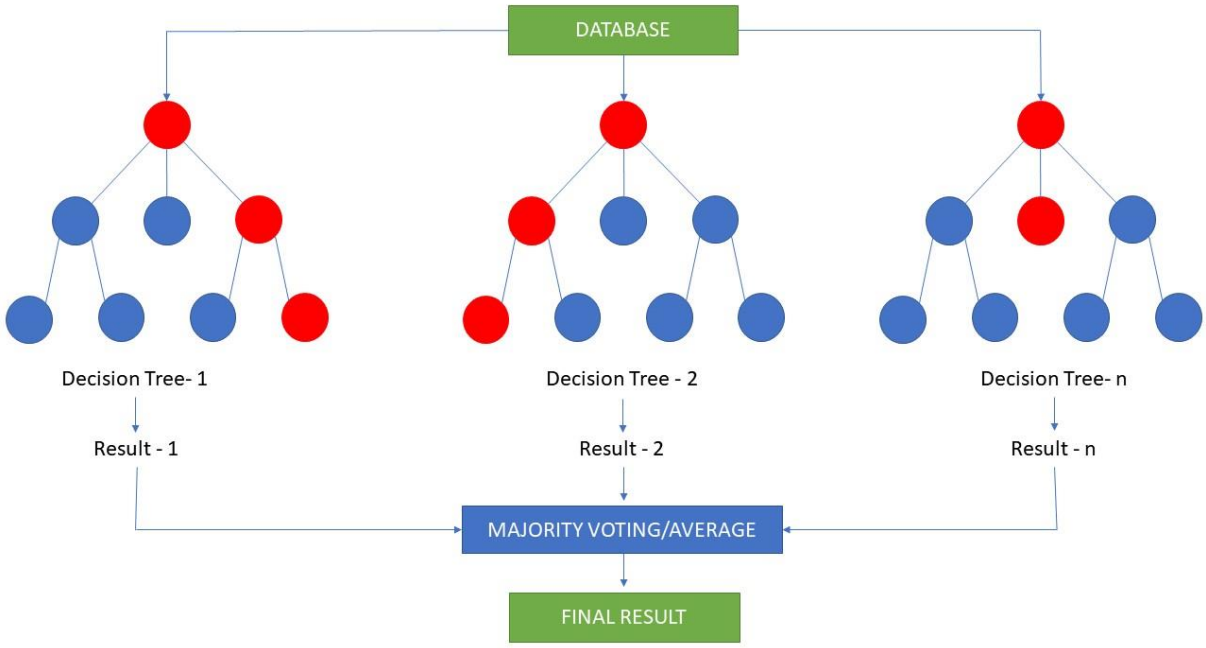
$$\text{NFR} = \frac{\text{FR factor class}}{\sum \text{FR factor class}} \quad (3)$$

2.2.5 Flood susceptibility map

The flood susceptibility model was performed by machine learning method through the Random Forest (RF) algorithm. To this, a scikit-learn library, in Python language, in the VsCode environment were used.

RF is defined by Breiman (2011) as a "combination of tree predictors such that each tree depends on the values of a random vector sampled independently and with the same distribution for all trees in the forest". The RF algorithm (Figure 5) consists of several individual decision trees. The construction of each tree is done by requested selection of a sample taken with selection (i.e., a bootstrap sample) from the original data set creating several phases of tree training (Pedregosa et al., 2011). In this way, subsets are generated with sizes equal to the original set (Chen et al. 2020). Unselected (out-of-bag) samples are used to evaluate model performance (Abedi et al. 2022).

Figure 5 – RF Model



Source: Prepared by the author.

Data classification is estimated based on voting results from multiple classifiers, i.e., the highest predicted probability in each class (Islam et al. 2021; Mobley et al. 2021). Thus, the results are obtained by merging the prediction or classification results of all decision trees (Chen et al. 2020).

Individual decision trees have only one outcome and tend to overfit. Therefore, with a greater number of groups and decisions, the average of the predictions can cancel out the errors, ensuring a more accurate result and an overall better model (Pedregosa et al. 2011).

Among the parameters to be defined in the RF algorithm, according to Abedi et al. (2022), one of the most important parameters to be adjusted is the size of the random subsets of resources (mtry) that will be considered in the node division. Another important parameter is the number of trees in the forest (ntree). A small value for ntree can lead to errors (Chen et al. 2020). Thus, increasing the number of trees in the forest tends to better results. On the other hand, there will be a higher computational cost.

The definition of the ntree and mtry parameters is done based on hyperparameterization tuning through a search grid, varying the parameter values to return the best combination of values, that is the one that obtains the highest performance. In this work, accuracy was used as a metric to measure performance to adjust of the parameters.

2.2.6 Model performance evaluation

The evaluation of RF performance will be performed using the Receive Operating Characteristic (ROC) curve, the area under the ROC curve (AUC), and metrics generated from the confusion matrix, which is a crosstab for training and validation of datasets where four types of results are possible: True positive (TP), True negative (TN), False positive (FP), and False negative (FN). FP and FN are misclassified pixel numbers.

The ROC curve determines the performance of statistical methods with binary classification (Manfreda, 2015), and is adopted to evaluate the accuracy of spatial prediction models for flood susceptibility modeling (Ahmadlou et al. 2021; Dodangeh et al. 2019). The AUC is a method used to compare different types of binary classifiers and reduce the ROC performance to a single scalar value (Manfreda, 2015), i.e., the value ranges from zero to one. Performance will be more robust when the AUC value is close to one (Bui et al., 2020; Norallahi, 2021).

The positive predictive rate (Equation 4) or precision is the ratio over the amount of TP and the total amount of positive values.

$$\text{Positive predictive rate} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{4}$$

The negative predictive rate (Equation 5) is the ratio of the amount of negative TN to the total amount of negative values.

$$\text{Negative predictive rate} = \frac{TN}{TN + FN} \tag{5}$$

Sensitivity (Equation 6) is used to successfully detect results classified as positive, while specificity (Equation 7) is used to detect negative ones.

$$Sensitivity = \frac{TP}{TP + FN} \tag{6}$$

$$Specificity = \frac{TN}{TN + FP} \tag{7}$$

Accuracy (Equation 8) evaluates the percentage of correct answers, it is the relationship between the number of correct predictions and the total number of entries.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (8)$$

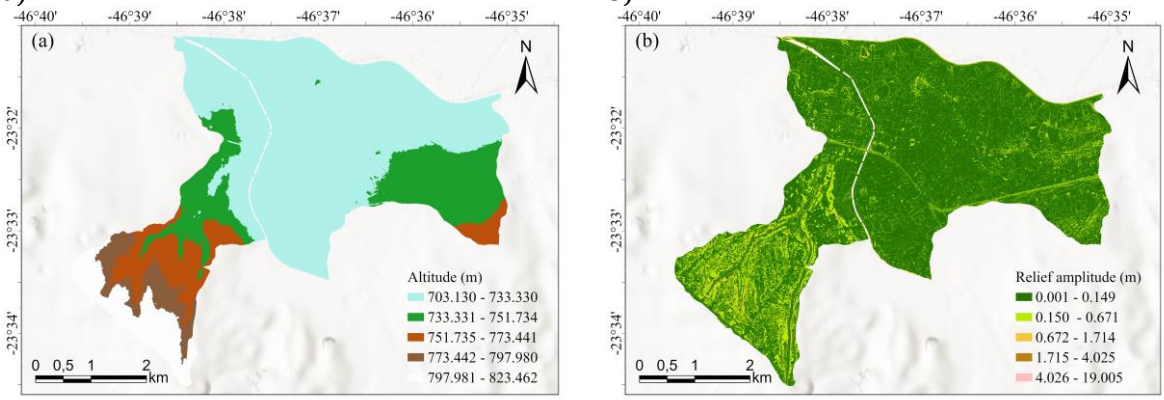
3 RESULTS

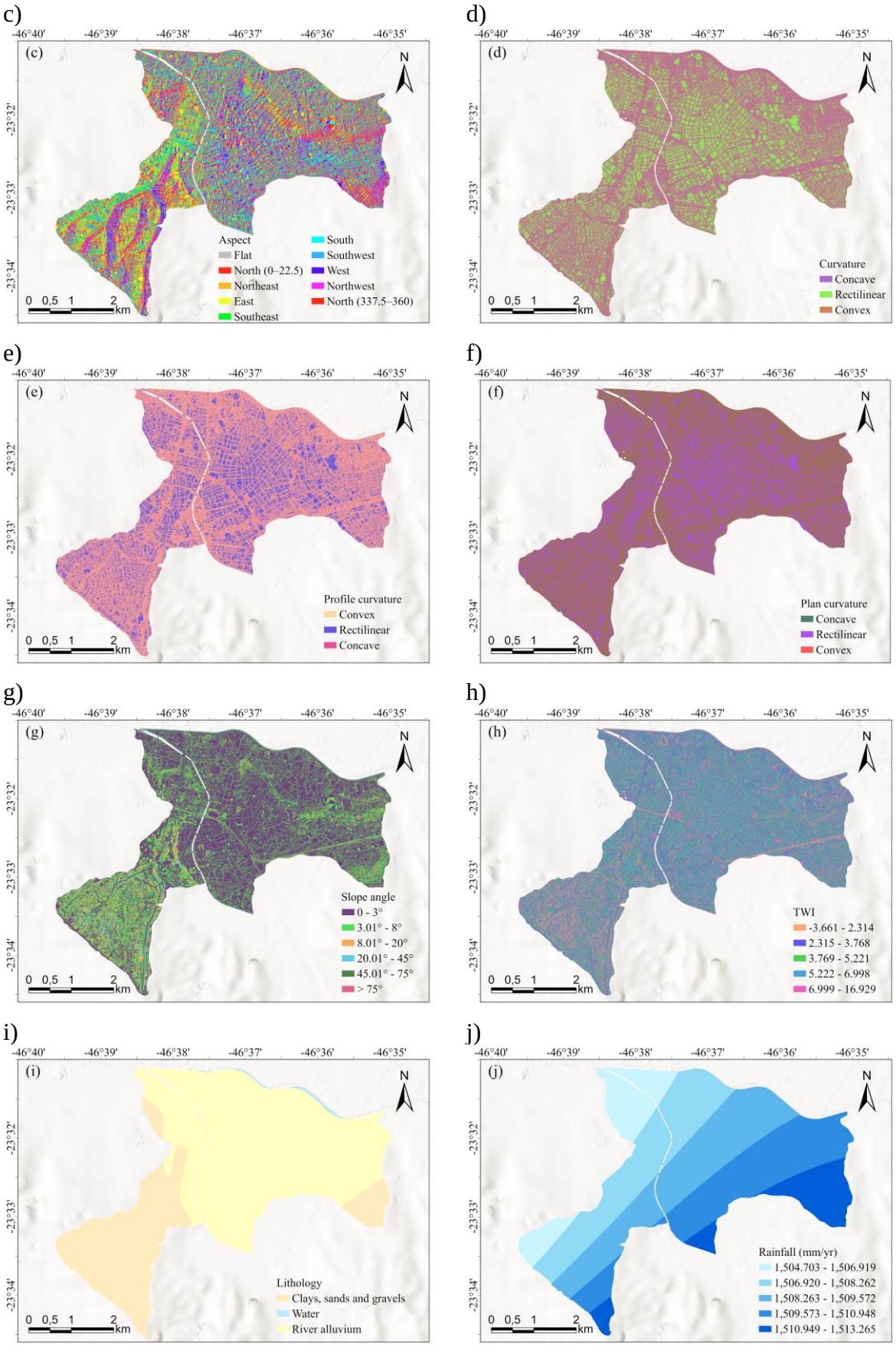
3.1 Conditioning factors

Figure 6 shows the conditioning factors used in the flood susceptibility model: Altitude, Relief amplitude, Aspect, Curvature, Profile curvature, Plan curvature, Slope, TWI, Lithology and Rainfall. Nodata area in the central region corresponds to the Tamanduateí River, which doesn't was considered in the analysis. The conditioning factors were reclassified as mentioned in Subsection 2.2.1.

Figure 6 – Conditioning factors: (a) Altitude, (b) Relief amplitude, (c) Aspect, (d) Curvature, (e) Profile curvature, (f) Plan curvature, (g) Slope angle, (h) TWI, (i) Lithology and (j) Rainfall.

a) b)





Source: Prepared by the author.

3.2 *FR analysis*

The occurrence frequency analysis of the flood points for classes of each conditioning factor was carried out using the FR method as described in Subsection 2.2.4 and is shown in Table 2 (Inventory A) and Table 3 (Inventory B).

The normalized results show that the highest value of NFR (0.75) is in Inventory A (Table 2) in the lithological class clays, sands and graves. In Inventory B (Table 3), the highest NFR value (0.70) is in the lowest altitude class (703,130 – 733,330).

Table 2- FR analysis between flood occurrence and conditioning factors (Inventory A)

Conditioning Factors	*Classes*	*% pixels*	*% flooding*	*FR*	*NFR*
Altitude	*703.130 – 733.330*	*60.08*	*40.00*	*0.67*	*0.14*
	733.331 – 751.734	*18.90*	*24.00*	*1.27*	*0.27*
	751.735 – 773.441	*8.94*	*17.33*	*1.94*	*0.41*
	773.130 – 797.980	*4.72*	*8.00*	*1.69*	*0.36*
	797.981 – 823.462	*7.36*	*10.67*	*1.45*	*0.30*
Relief amplitude	*0.001 – 0.149*	*80.24*	*69.33*	*0.86*	*0.33*
	0.150 – 0.671	*17.43*	*30.67*	*1.76*	*0.67*
	0.672 – 1.714	*1.97*	*0.00*	*0.00*	*0.00*
	1.715 – 4.025	*0.32*	*0.00*	*0.00*	*0.00*
	4.026 – 19.005	*0.04*	*0.00*	*0.00*	*0.00*
Aspect	*Flat (–1)*	*0.00*	*0.00*	*0.00*	*0.00*
	North (0–22.5)	*7.21*	*18.67*	*2.59*	*0.28*
	Northeast (22.5–67.5)	*14.39*	*17.33*	*1.20*	*0.13*
	East (67.5–112.5)	*13.07*	*13.33*	*1.02*	*0.11*
	Southeast (112.5–157.5)	*10.77*	*12.00*	*1.11*	*0.12*
	South (157.5–202.5)	*10.44*	*6.67*	*0.64*	*0.07*
	Southwest (202.5–247.5)	*10.70*	*1.33*	*0.12*	*0.01*
	West (247.5–292.5)	*12.24*	*17.33*	*1.42*	*0.15*
	Northwest (292.5–337.5)	*13.77*	*8.00*	*0.58*	*0.06*
	North (337.5–360)	*7.40*	*5.33*	*0.72*	*0.08*
Curvature	*Concave*	*36.59*	*36.00*	*0.98*	*0.33*
	Rectilinear	*26.81*	*21.33*	*0.80*	*0.27*
	Convex	*36.60*	*42.67*	*1.17*	*0.40*
Profile curvature	*Convex*	*38.08*	*53.33*	*1.40*	*0.50*
	Rectilinear	*23.73*	*12.00*	*0.51*	*0.18*
	Concave	*38.19*	*34.67*	*0.91*	*0.32*

Table 2 – Continuous

Plan curvature	*Concave*	*37.97*	*49.33*	*1.30*	*0.46*
	Rectilinear	*23.71*	*13.33*	*0.56*	*0.20*
	Convex	*38.32*	*37.33*	*0.97*	*0.34*
Slope in degrees	*0 - 3*	*57.80*	*50.67*	*0.88*	*0.16*
	3.01 - 8	*27.99*	*22.67*	*0.81*	*0.15*
	8.01 - 20	*10.14*	*21.33*	*2.10*	*0.39*
	20.01 - 45	*3.41*	*5.33*	*1.56*	*0.29*
	45.01 - 75	*0.63*	*0.00*	*0.00*	*0.00*
	> 75	*0.02*	*0.00*	*0.00*	*0.00*
TWI	*-3.661 – 2.314*	*16.06*	*26.67*	*1.66*	*0.35*
	2.315 – 3.768	*35.29*	*26.67*	*0.76*	*0.16*
	3.769 – 5.221	*26.75*	*34.67*	*1.30*	*0.27*
	5.222 – 6.998	*16.43*	*9.33*	*0.57*	*0.12*
	6.999 – 16.929	*5.47*	*2.67*	*0.49*	*0.10*
Lithology	*Clays, sands and gravels*	*31.16*	*57.33*	*1.84*	***0.75***
	Mass of water	*0.61*	*0.00*	*0.00*	*0.00*
	River alluvium	*68.24*	*42.67*	*0.63*	*0.25*
Rainfall (mm/yr)	*1,504.703 – 1,506.919*	*12.26*	*21.33*	*1.74*	*0.34*
	1,506.92 – 1,508.262	*24.18*	*41.33*	*1.71*	*0.33*
	1,508.263 – 1,509.572	*27.26*	*25.33*	*0.93*	*0.18*
	1,509.573 – 1,510.948	*25.21*	*6.67*	*0.26*	*0.05*
	1,510.949 – 1,513.265	*11.09*	*5.33*	*0.48*	*0.09*

Table 3 - FR analysis between flood and flash flood occurrence and conditioning factors (Inventory B)

Conditioning Factors	*Classes*	*% pixels*	*% flooding*	*FR*	*FRN*
Altitude	*703.130 – 733.330*	*16.06*	*60.03*	*3.74*	***0.70***
	733.331 – 751.734	*35.29*	*27.16*	*0.77*	*0.14*
	751.735 – 773.441	*26.75*	*8.48*	*0.32*	*0.06*
	773.130 – 797.980	*16.43*	*2.25*	*0.14*	*0.03*
	797.981 – 823.462	*5.47*	*2.08*	*0.38*	*0.07*
Relief amplitude	*0.001 – 0.149*	*80.24*	*86.33*	*1.08*	*0.48*
	0.150 – 0.671	*17.43*	*12.80*	*0.73*	*0.33*
	0.672 – 1.714	*1.97*	*0.87*	*0.44*	*0.19*
	1.715 – 4.025	*0.32*	*0.00*	*0.00*	*0.00*
	4.026 – 19.005	*0.04*	*0.00*	*0.00*	*0.00*

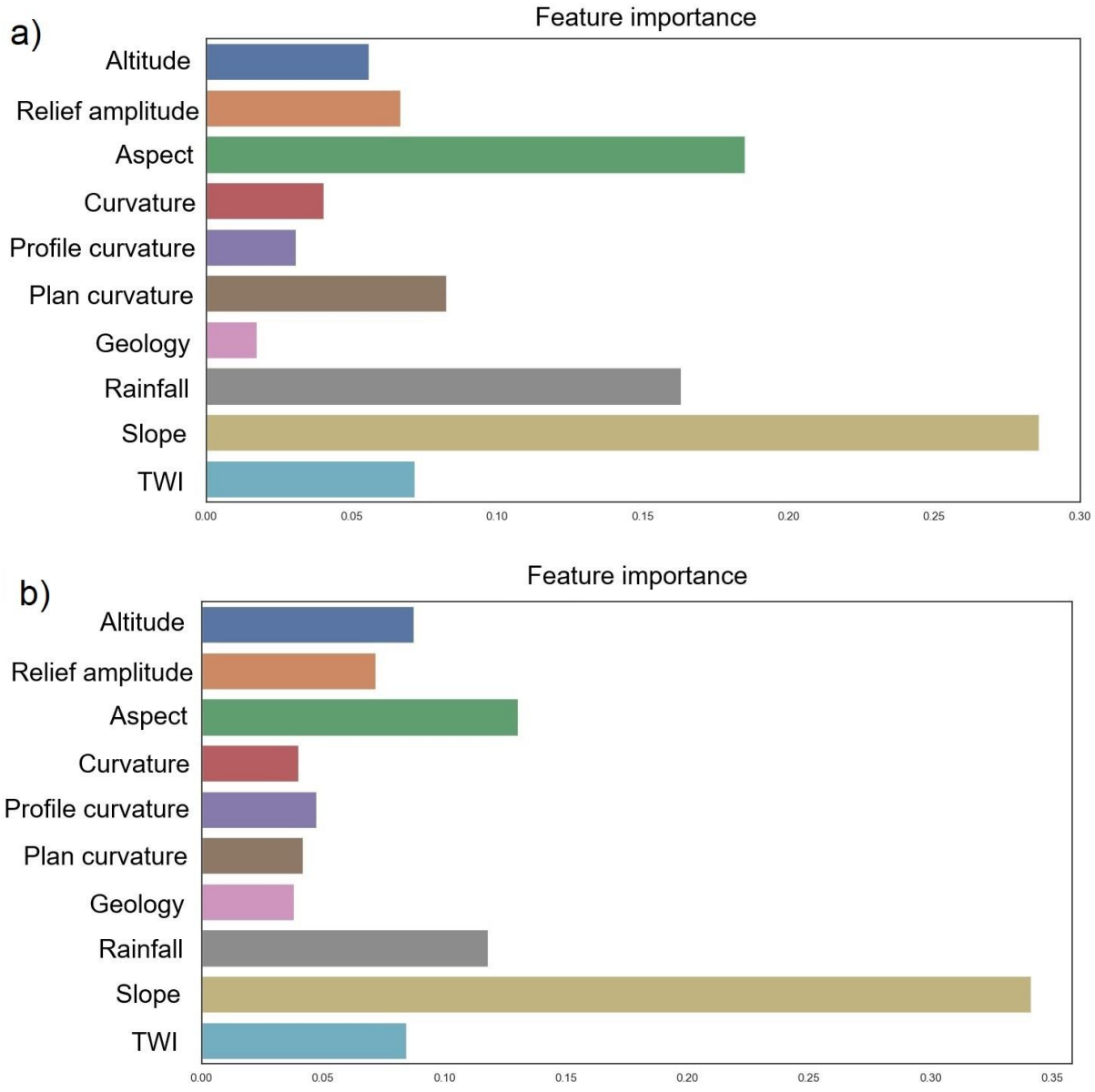
Table 3 - Continuous

Aspect	*Flat (–1)*	*0.00*	*0.00*	*0.00*	*0.00*
	North (0–22.5)	*7.21*	*9.34*	*1.30*	*0.15*
	Northeast (22.5–67.5)	*14.39*	*19.03*	*1.32*	*0.15*
	East (67.5–112.5)	*13.07*	*11.07*	*0.85*	*0.10*
	Southeast (112.5–157.5)	*10.77*	*10.38*	*0.96*	*0.11*
	South (157.5–202.5)	*10.44*	*8.82*	*0.85*	*0.10*
	Southwest (202.5–247.5)	*10.70*	*8.48*	*0.79*	*0.09*
	West (247.5–292.5)	*12.24*	*17.30*	*1.41*	*0.16*
	Northwest (292.5–337.5)	*13.77*	*11.07*	*0.80*	*0.09*
	North (337.5–360)	*7.40*	*4.50*	*0.61*	*0.07*
Curvature	*Concave*	*36.59*	*34.43*	*0.94*	*0.32*
	Rectilinear	*26.81*	*16.78*	*0.63*	*0.22*
	Convex	*36.60*	*48.79*	*1.33*	*0.46*
Profile curvature	*Convex*	*38.08*	*52.77*	*1.39*	*0.49*
	Rectilinear	*23.73*	*12.63*	*0.53*	*0.19*
	Concave	*38.19*	*34.60*	*0.91*	*0.32*
Plan curvature	*Concave*	*37.97*	*39.45*	*1.04*	*0.36*
	Rectilinear	*23.71*	*14.19*	*0.60*	*0.21*
	Convex	*38.32*	*46.37*	*1.21*	*0.42*
Slope in degrees	*0 - 3*	*57.80*	*61.42*	*1.06*	*0.28*
	3.01 - 8	*27.99*	*27.68*	*0.99*	*0.26*
	8.01 - 20	*10.14*	*8.65*	*0.85*	*0.23*
	20.01 - 45	*3.41*	*2.08*	*0.61*	*0.16*
	45.01 - 75	*0.63*	*0.17*	*0.28*	*0.07*
	> 75	*0.02*	*0.00*	*0.00*	*0.00*
TWI	*-3.661 – 2.314*	*16.06*	*16.09*	*1.00*	*0.23*
	2.315 – 3.768	*35.29*	*42.39*	*1.20*	*0.28*
	3.769 – 5.221	*26.75*	*29.76*	*1.11*	*0.26*
	5.222 – 6.998	*16.43*	*9.17*	*0.56*	*0.13*
	6.999 – 16.929	*5.47*	*2.60*	*0.47*	*0.11*
Lithology	*Clays, sands and gravels*	*31.16*	*41.70*	*1.34*	*0.44*
	Mass of water	*0.61*	*0.52*	*0.86*	*0.28*
	River alluvium	*68.24*	*57.79*	*0.85*	*0.28*
Rainfall (mm/yr)	*1,504.703 – 1,506.919*	*12.26*	*9.86*	*0.80*	*0.18*
	1,506.92 – 1,508.262	*24.18*	*47.58*	*1.97*	*0.43*
	1,508.263 – 1,509.572	*27.26*	*22.66*	*0.83*	*0.18*
	1,509.573 – 1,510.948	*25.21*	*16.09*	*0.64*	*0.14*
	1,510.949 – 1,513.265	*11.09*	*3.81*	*0.34*	*0.07*

3.3 Selection of flood conditioning factors

The importance of each conditioning factor in the flood susceptibility model is shown as an average of the decrease of impurities in the model, according to the feature importance in Figure 7. All flood conditioning factors contribute to the models. The conditioning factor Slope is considered the most important within the study area in both models.

Figure 7 – Feature importance: a) model A and b) model B



3.4 Model performance

In the validation step, the performance of the flood susceptibility model was verified firstly based on Accuracy reached with hyperparameters adjusting (Subsection 2.2.5). In this work, the parameters mtry and ntree were adjusted, as shown in Table with values of 0.733 and 0.879 for models A and B, respectively. Other metrics are also presented in Table 4.

Table 4 – Parameters (mtry and ntree) adjusted and performance analysis metrics of the flooding susceptibility model.

Metrics/Parameters	***Model A***	***Model B***
	mtry = 6 *ntree = 100*	*mtry = 4* *ntree = 300*
Accuracy	*0.733*	*0.879*
Positive predictive rate	*0.727*	*0.880*
Negative predictive rate	*0.696*	*0.878*
Sensitivity	*0.696*	*0.870*
Specificity	*0.727*	*0.888*
AUC	*0.821*	*0.925*

Source: Prepared by the author.

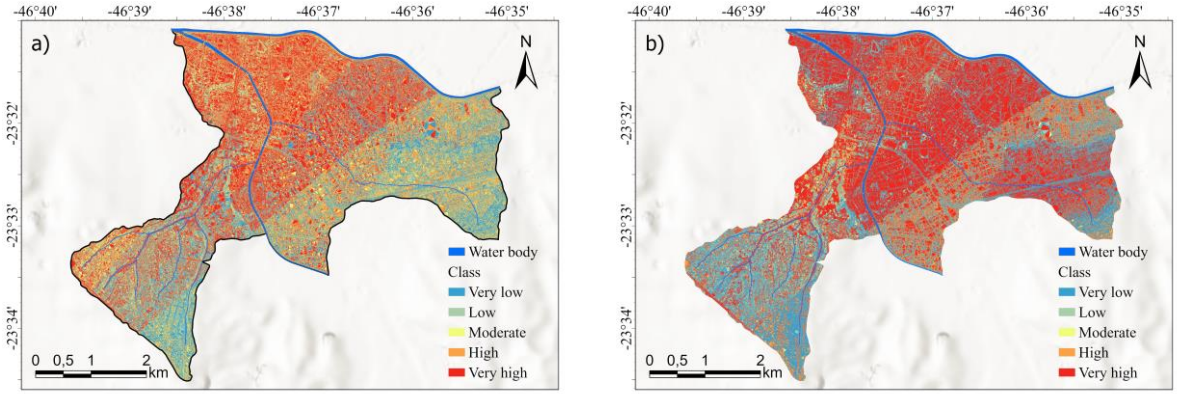
For flood and non-flood pixel classification, model B performs best for validation datasets, with values equal to 0.870 and 0.888, respectively. The results show that model B (that considers flooding and flash flooding occurrences) produced better results, with higher Accuracy (0.879) and AUC (0.925) values than model A. Therefore, it can be concluded that model B outperforms model A in modeling flood susceptibility.

3.5 Flood susceptibility map

Flood susceptibility maps were generated based on the training dataset of Inventory A and Inventory B, which correspond to models A (Figure 8a) and B (Figure 8b), respectively. The susceptibility map indicates to each pixel the probability of flooding based on the conditioning factors and the susceptibility model adopted. Then, susceptibility maps

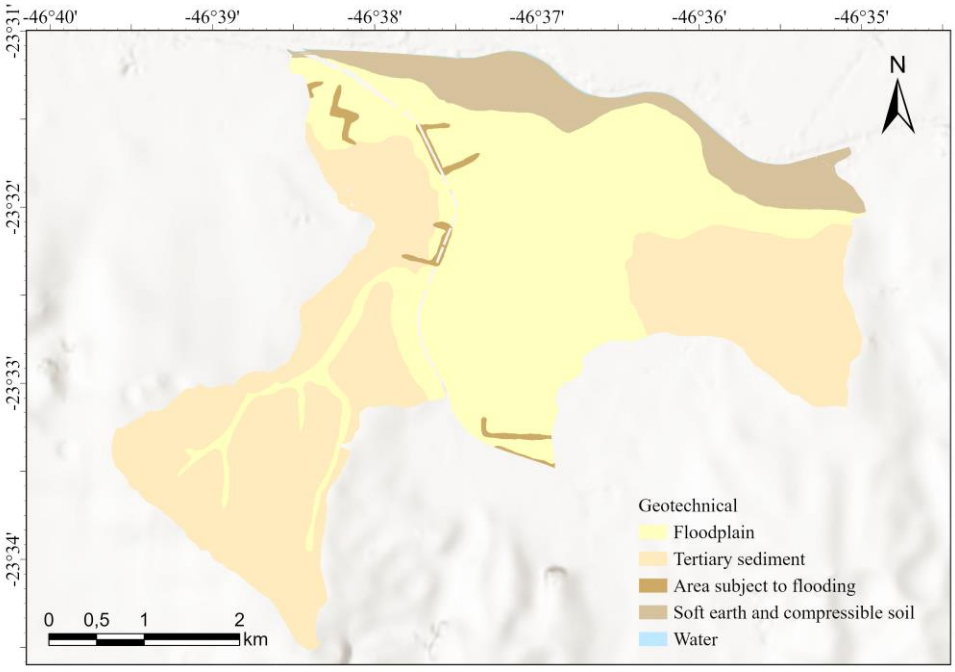
were classified into five susceptibility classes, very low, low, moderate, high, and very high, using the equal interval classification method.

Figure 8 - Flood susceptibility maps: a) Model A, b) Model B.



The results obtained by the proposed methodology can be compared with the geotechnical map (Figure 9) available at the GeoSampa database, on a scale of 1:10,000. This map provides data on soils, rocks, relief morphology and special occurrences such as landslides and floods.

Figure 9 - Geotechnical map



When comparing the geotechnical map with the results of the flood susceptibility models and flood occurrences it is possible to notice that the area susceptible to flooding goes beyond what is shown on the geotechnical map (classes floodplain and area subject to flooding).

4 DISCUSSIONS

The motivation for this study comes from the need for new technologies for flood susceptibility analysis in urban and dense areas, such as the city of São Paulo, where the changes in the channeling of rivers and streams carried out in the past do not currently meet the surface runoff of water.

The field of ML has had great notoriety in recent decades due to the increase in computational capacity and Artificial Intelligence techniques. ML comprises a vast set of algorithms and modeling tools for learning or training data (Carleo et al. 2019; Zhou 2021). The larger the set of quality data available for training, the better the performance of the data analysis method used by ML (Antzoulatos et al. 2022; Bentivoglio et al. 2021). The main advantage of using ML is automating the process of building analytical models after the algorithm learns from the training data (Janiesch et al. 2021).

The main objective of this study is to produce flood susceptibility maps using two flood point inventories and compare them with an available geotechnical map. The maps were generated using the random forest algorithm. Model B, generated with a larger data set that considers flooding and flash flood occurrences, showed higher performance when compared to model A, with a smaller data set (only flooding).

Analyzing the importance of conditioning factors provides important information about predictors (Collins et al. 2022). In this study, slope is the most important predictor for both models, indicating that this variable reveals the topographic features with more susceptibility to flooding.

The models presented in this study produced acceptable results in the prediction of susceptibility, proven by the evaluation metrics. Model B, when compared with similar studies using RF (Collins et al. 2022; Mobley et al. 2021), has high sensitivity in flood prediction (AUC = 0.925). Thus, the generated maps can be used to establish early warning systems and responses to disasters caused by flood events.

5 CONCLUSIONS

In this study, flood susceptibility prediction maps were developed in an area of the Tamanduateí River basin, located in the city of São Paulo. The RF algorithm, historical rainfall records, location records of flood points (model A) and flood and flash flooding points (model B), and topographic, geological and hydrological data were used. Validation methods using statistical metrics were used to verify the performance and compare models.

The validation results showed that the models showed good results. Model B, with flood and flash flooding, showed the highest performance in terms of accuracy, AUC, sensitivity and specificity. The NFR results showed that most of the floods, for this model, occurred in the low altitude class and the most significant conditioning factor is the Slope followed (in descending order) of the Aspect, Rainfall, Altitude, TWI, Amplitude, Profile curvature, Plan curvature, Curvature and Lithology. The results of the susceptibility models and the flood occurrences show that the area susceptible to flooding goes beyond what is shown on the geotechnical map.

The flood susceptibility maps analyzed in this study can contribute to the actions of planners and managers to prevent and mitigate the impacts caused by floods. In future research, the models can be applied to other flood areas considering the local characteristics of the areas to be studied.

REFERENCES

Abedi R., Costache R, Shafizadeh-Moghadam H, Pham QB (2022) Flash-flood susceptibility mapping based on XGBoost, random forest and boosted regression trees, Geocarto International 37:19 5479-5496. DOI: 10.1080/10106049.2021.1920636

Ahmadlou M, Al-Fugara A, Al-Shabeeb AR, Arora A, Al-Adamat R, Pham QB, Al-Ansari N, Linh NTT.; Sjedi, H (2021) Flood susceptibility mapping and assessment using a novel deep learning model combining multilayer perceptron and autoencoder neural networks. J Flood Risk Management 14:e12683. https:// doi.org/10.1111/jfr3.12683

Antzoulatos G, Kouloglou I-O, Bakratsas M, Moumtzidou A, Gialampoukidis I, Karakostas A, Lombardo F, Fiorin R, Norbiato D, Ferri M, Symeonidis A, Vrochidis S, Kompatsiaris I (2022) Flood Hazard and Risk Mapping by Applying an Explainable Machine Learning Framework Using Satellite Imagery and GIS Data. https://doi.org/10.3390/su14063251.

Arora A, Pandey M, Siddiqui MA, Hong H, Mishra VN (2019) Spatial flood susceptibility prediction in Middle Ganga Plain: comparison of frequency ratio and Shannon's entropy models, Geocarto International. https://doi.org/10.1080/10106049.2019.1687594

Bentivoglio R, Isufi E, Jonkman SN, Taormin R (2021) Deep Learning Methods for Flood Mapping: A Review of Existing Applications and Future Research Directions. Hydrology and Earth System Sciences. https://doi.org/10.5194/hess-2021-614

Beven KJ, Kirkby MJ (1979) A physically based, variable contributing area model of basin hydrology. Hydrological Sciences Bulletin. https://doi.org/10.1080/02626667909491834

Breiman L (2001) Random forests. Mach. Learn. https://doi.org/10.1023/A:1010933404324

Bui DT, Ngo PTT, Pham TD, Jaafari A, Minh NQ, Hoa PV, Samui P (2019) A novel hybrid approach based on a swarm intelligence optimized extreme learning machine for flash flood susceptibility mapping. Catena. https://doi.org/10.1016/j.catena.2019.04.009.

Bui QT, Nguyen QH, Nguyen XL, Pham VD, Nguyen HD, Pham VM (2020) Verification of novel integrations of swarm intelligence algorithms into deep learning neural network for flood susceptibility mapping. Journal of Hydrology. https://doi.org/10.1016/j.jhydrol.2019.124379.

Carleo G, Cirac I, Cranmer K, Daudet L, Schuld M, Tishby N, Vogt-Maranto L, Zdeborová L (2019) Machine learning and the physical sciences. Reviews of modern physics. https://doi.org/10.48550/arXiv.1903.10563

CEPED UFSC (2013) Atlas brasileiro de desastres naturais: 1991 a 2012 / Centro Universitário de Estudos e Pesquisas sobre Desastres. Florianópolis

Chen et al (2020) Modeling flood susceptibility using data-driven approaches of naïve Bayes tree, alternating decision tree, and random forest methods. Science of the Total Environment. https://doi.org/10.1016/j.scitotenv.2019.134979.

Christofoletti, A (1980) Geomorfologia. Blucher, São Paulo

Collins EL, Sanchez GM, Terando A, Stillwell C, Mitasova H, Sebastian A, Meetntemeyer R (2022) Predicting flood damage probability across the conterminous United States. Environmental Research Letters. https://doi.org/10.1088/1748-9326/ac4f0f

Costache R, Bui DT (2019) Spatial prediction of flood potential using new ensembles of bivariate statistics and artificial intelligence: A case study at the Putna river catchment of Romania. Science of The Total Environment. https://doi.org/10.1016/j.scitotenv.2019.07.197.

CRED (2022) 2021 Disasters in numbers. Brussels. This document is available at: https://cred.be/sites/default/files/2021_EMDAT_report.pdf

DAEE (2009) Plano de Macrodrenagem da Bacia do Alto Tietê.

Dodangeh E, Choubin B, Eigdir AN, Nabipour N, Panahi M, Shamshirband S, Mosavi A (2019) Integrated machine learning methods with resampling algorithms for flood susceptibility prediction. Science of the Total Environment. https://doi.org/10.1016/j.scitotenv.2019.135983

El-Magd A.; Ahmed S (2022) Random Forest and naïve Bayes approaches as tools for flash flood hazard susceptibility prediction, South Ras El-Zait, Gulf of Suez Coast, Egypt. Arabian Journal of Geosciences. https://doi.org/10.1007/s12517-022-09531-3

Hirabayashi Y, Mahendran R, Koirala S, KONOSHIMA L, YAMAZAKI D, WATANABE S, KIM, H, KANAE S. Global flood risk under climate change. Nature Clim Change 3, 816–821, 2013. https://doi.org/10.1038/nclimate1911

IBGE - Instituto Brasileiro de Geografia e Estatística (2022). https://cidades.ibge.gov.br/brasil/sp/sao-paulo/panorama. Accessed 20 June 2022.

IPCC (2021) Summary for Policymakers. In: Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change [Masson-Delmotte, V., P. Zhai, A. Pirani, S.L. Connors, C. Péan, S. Berger, N. Caud, Y. Chen, L. Goldfarb, M.I. Gomis, M. Huang, K. Leitzell, E. Lonnoy, J.B.R. Matthews, T.K. Maycock, T. Waterfield, O. Yelekçi, R. Yu, and B. Zhou (eds.)]. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, pp. 3−32, doi:10.1017/9781009157896.001.

IPCC (2022) Climate Change 2022: Impacts, Adaptation and Vulnerability. Contribution of Working Group II to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change [H.-O. Pörtner, D.C. Roberts, M. Tignor, E.S. Poloczanska, K. Mintenbeck, A. Alegría, M. Craig, S. Langsdorf, S. Löschke, V. Möller, A. Okem, B. Rama (eds.)]. Cambridge University Press. Cambridge University Press, Cambridge, UK and New York, NY, USA, 3056 pp., doi:10.1017/9781009325844.

Islam ARMT, Talukdar S, Mahato S, Kundu S, Eibek KU, Pham QB, Kuriqi A, Linh NTT (2021) Flood susceptibility modelling using advanced ensemble machine learning models. Geoscience Frontiers. https://doi.org/10.1016/j.gsf.2020.09.006.

Janiesch C, Zschech P, Heinrich K (2021) Machine learning and deep learning. Electron Markets 31:685–695. https://doi.org/10.1007/s12525-021-00475-2

Khosravi K, Shahabi H, Pham BT, Adamowski J, Shirzadi A, Pradhan B, Dou J, Ly H, Gróf G, Ho HL, Hong H, Chapi K, Prakash I (2019) A comparative assessment of flood susceptibility modeling using Multi-Criteria Decision-Making Analysis and Machine Learning Methods. Journal of Hydrology. https://doi.org/10.1016/j.jhydrol.2019.03.073.

Lindsay JB (2014) The Whitebox Geospatial Analysis Tools Project and Open-Acess GIS. Proceedings of the GIS Research UK 22nd Annual Conference, The University of Glasgow.

Manfreda S, Samela C, Gioia A, Consoli GG, Iacobellis V, Giuzio L, Cantisani A, Sole, A (2015) Flood-prone areas assessment using linear binary classifiers based on flood maps obtained from 1D and 2D hydraulic models. Natural Hazards. https://doi.org/10.1007/s11069-015-1869-5

Mobley W, Sebastian A, Blessing R, Highfield WE, Stearns L, Brody SD (2021) Quantification of continuous flood hazard using random forest classification and flood insurance claims at large spatial scales: a pilot study in southeast Texas. Natural Hazards Earth Systems Sciences, 21: 807-822. https://doi.org/10.5194/nhess-21-807-2021

Moore, ID, Grayson, RB, Ladson AR (1991) Digital terrain modelling: A review of hydrological, geomorphological, and biological, and biological applications. Hydrological processes. https://doi.org/10.1002/hyp.3360050103

Norallahi M, Kaboli HS (2021) Urban flood hazard mapping using machine learning models: GARP, RF, MaxEnt and NB. Natural Hazards. https://doi.org/10.1007/s11069-020-04453-3

Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Bondlel M et al (2011) Scikit-learn: Machine Learning in Python, JMLR 12:2825-2830. https://scikit-learn.org/stable/modules/ensemble.html?highlight=random+forest#forest. Accessed 28 August 2022.

Pourghasemi HR, Kariminejad N, Amiri M et al (2020) Assessing and mapping multi-hazard risk susceptibility using a machine learning technique. Scientific Report. https://doi.org/10.1038/s41598-020-60191-3

Rahman M, Ningsheng C, Islam MM, Dewan A, Iqbal J, Washakh RMA, Shufeng T (2019) Flood Susceptibility Assessment in Bangladesh Using Machine Learning and Multi-criteria Decision Analysis. Earth Syst Environ 3:585–601. https://doi.org/10.1007/s41748-019-00123-y

Rehman A, Song J, Haq F, Mahmood S, Ahamad MI, Basharat M, Sajid M, Mehmood, M.S (2022) Multi-Hazard Susceptibility Assessment Using the Analytical Hierarchy Process and Frequency Ratio Techniques in the Northwest Himalayas, Pakistan. Remote Sens. 14:554. https://doi.org/ 10.3390/rs14030554

Sarkar D, Mondal P (2020) Flood vulnerability mapping using frequency ratio (FR) model: a case study on Kulik river basin, Indo-Bangladesh Barind region. Appl Water Sci. https://doi.org/10.1007/s13201-019-1102-x

Taalab K, Cheng T, Zhang Y (2018) Mapping landslide susceptibility and types using random forest. Big Earth Data. https://doi.org/10.1080/20964471.2018.1472392

Taromideh F, Fazloula R, Choubin B, Emadi A, Berndtsson R (2022) Urban Flood-Risk Assessment: Integration of Decision-Making and Machine Learning. Sustainability 14: 4483. https://doi.org/ 10.3390/su14084483

Tehrany MS, Pradhan B, Mansor S, Ahmad N (2015) Flood susceptibility assessment using GIS-based support vector machine model with different kernel types. CATENA 125:91-101. https://doi.org/10.1016/j.catena.2014.10.017.

Tomás LR, Soares GG, Jorge AAS, Mendes JF, Freitas VLS, Santos LBL (2022) Flood risk map from hydrological and mobility data: A case study in São Paulo (Brazil). Transactions in GIS 26:2341–2365. https://doi.org/10.1111/tgis.12962

Ullah K, Zhang J (2020) GIS-based flood hazard mapping using relative frequency ratio method: A case study of Panjkora River Basin, eastern Hindu Kush, Pakistan. PLoS ONE 15(3): e0229153. https://doi.org/10.1371/journal. pone.022915

Valverde MC, Calado BN, Calado GG, Kuroki LY, Brambila R, Sousa AR (2023) Climate projections of precipitation and temperature in cities from ABC Paulista, in the Metropolitan Region of São Paulo-Brazil. Frontiers in Climv. 5:1127026. https://doi.org/10.3389/fclim.2023.1127026

Zhou Z-H (2021) Machine Learning. Springer Nature.

Zope PE, Eldho TI, Jothiprakash V (2016) Impacts of land use–land cover change and urbanization on flooding: A case study of Oshiwara River Basin in Mumbai, India. Catena 145:142-154. https://doi.org/10.1016/j.catena.2016.06.009.

Wang, Z.; Lai, C.; Chen, X.; Yang, b.; Zhao S.; Bai, X (2015) Flood hazard risk assessment model based on random forest. Journal of Hydrology 527:1130-1141. https://doi.org/10.1016/j.jhydrol.2015.06.008.

3 CONSIDERAÇÕES FINAIS

A partir do século XIX, o desenvolvimento das cidades decorrente da industrialização e da mecanização da agricultura acionou a procura de terras nas planícies de inundações. A ocupação nas imediações das margens dos rios promoveu alterações nos canais, bem como a redução da sua capacidade original de reduzir os períodos de cheias. Nas grandes cidades brasileiras, como por exemplo na cidade de São Paulo, os sistemas fluviais dos rios foram modificados através do aumento da área da seção transversal para assumir o crescimento populacional e de estruturas urbanas da cidade. Diante deste modelo de ocupação, as ocorrências de inundação e seus impactos na capital paulista, principalmente na região de convergência entre os Rios Pinheiros e Tietê, tornou-se um grande desafio no processo de gestão desta área.

As marginais dos Rios Pinheiros e Tietê, interligam as regiões norte e sul da cidade de São Paulo e são responsáveis pelo acesso de várias rodovias importantes do país. Nos eventos de inundação provocados por estes rios, os principais sistemas viários são paralisados e refletem no abastecimento de produtos essenciais para diversas regiões do Brasil e, com isso, a economia brasileira é diretamente afetada.

Neste estudo, os mapas de suscetibilidade à inundação foram desenvolvidos com resolução espacial de 50 cm em uma porção da bacia do rio Tamanduateí, localizada na cidade de São Paulo. Na metodologia, o modelo de suscetibilidade a inundação a partir do algoritmo de RF, usou registros de localização de pontos de inundação (modelo A) e de pontos de inundação e alagamento (modelo B) para compor as amostras de treinamento e validação. Fatores condicionantes topográficos, hidrológicos, geológicos e de precipitação foram usados no modelo de suscetibilidade a inundação. O mapa de precipitação média anual do município foi gerado a partir registros históricos de precipitação e interpolação dos dados usando uma abordagem geoestatística. Métricas estatísticas foram utilizadas para verificar o desempenho e comparar modelos de suscetibilidade a inundação gerados.

Os resultados mostraram que o modelo B, com registros de inundação e alagamentos, apresentou o melhor desempenho em termos de acurácia (0,879), AUC (0,925), sensibilidade (0,870) e especificidade (0,888) quando comparado com o modelo A, que usou apenas registros de inundação. Na análise de Frequência Relativa, tem-se que a maioria dos eventos de inundação ocorreram na classe de baixa Altitude. O fator condicionante mais significativo nos modelos de suscetibilidade foi a Declividade, seguido (em ordem decrescente) dos fatores

condicionantes Aspecto, Precipitação, Altitude, TWI, Amplitude, Curvatura de perfil, Curvatura plana, Curvatura e Litologia.

O fator condicionante pluviométrico (Precipitação) ocupou a terceira posição nos modelos A e B, indicando a importância dessa variável. No entanto, obter um mapa de precipitação baseados em dados de estações pluviométricas nem sempre é uma tarefa simples e um dos maiores problemas é a disponibilidade dos dados. No presente estudo verificou-se que a distribuição das estações pluviométricas com dados consistentes nas proximidades da cidade de São Paulo não é homogênea. Apenas 19 estações puderam ser consideradas para a análise, sendo somente seis delas pertencentes ao município de São Paulo. O mapa de estimativa de erros gerado pela abordagem geoestatística mostrou que a região necessita de adensamento de estações pluviométricas com séries históricas suficientes e dados consistentes para melhorar a representatividade da variável. Portanto, a elaboração de mapas de precipitação, anual ou sazonal, é um desafio significativo para determinadas localidades, pois sua precisão pode ser afetada pela falta de dados.

Os resultados dos modelos de suscetibilidade mostraram que a área suscetível à inundação vai além do que é mostrado no mapa geotécnico, que apresentam as classes de planície de inundação e áreas suscetíveis a inundação. Tal fato também pode ser observado pela localização dos eventos de inundação e alagamentos, a partir do inventário. Por isso que a metodologia proposta deste estudo, baseada em *data-driven*, propõe um avanço na coleta e processamento do conjunto de dados para a área de estudo, uma vez que os mapas de suscetibilidade a inundação disponíveis para o município de São Paulo não usam métodos de ML e geralmente possuem baixa precisão e pouco detalhamento, dado à baixa resolução espacial (10 a 30 metros) do MDT utilizado.

A atualização das áreas suscetíveis a inundação e de planícies de inundação do município deve ser realizada, no entanto é uma tarefa dispendiosa. Para obter uma compreensão abrangente das áreas sujeitas a inundação, seria interessante estudos que envolvessem uma abordagem baseada em modelo de chuva-vazão, tendo como resultado a simulação da mancha de inundação. No entanto, essa abordagem também é limitada devido à ausência de dados de vazão disponíveis na região, necessários para a calibração do modelo. Dessa forma, métodos baseados em ML, como proposto no presente estudo, proporcionam uma alternativa potencial para a atualização dessas áreas.

Espera-se que os mapas de suscetibilidade a inundação analisados neste estudo sejam utilizados como uma fonte de dados confiável, com a inclusão da determinação de áreas

prioritárias para evitar os danos decorrentes por este evento e, dessa maneira, o trabalho possa contribuir para as ações de planejadores e gestores na prevenção e mitigação de desastres causados pela inundação.

Este estudo é relevante na avaliação de inundação na cidade de São Paulo, porque a técnica de ML, aplicada através do algoritmo RF, apresentou ótima sensibilidade quando comparado a resultados de outros estudos. Todavia, a área deste estudo pertence a região central do município e diferentes resultados podem ser obtidos de acordo com as características regionais. Em pesquisas futuras, os modelos poderão ser aplicados a outras áreas inundáveis considerando as características locais das áreas a serem estudadas.

REFERÊNCIAS

ABEDI, R.; COSTACHE, R.; SHAFIZADEH-MOGHADAM, H.; PHAM, Q. B. Flash-flood susceptibility mapping based on XGBoost, random forest and boosted regression trees. **Geocarto International**, 37:19, p. 5479-5496, 2022.

AIYELOKUN, O.O., AIYELOKUN, O.D., AGBEDE, O.A. Application of random forest (RF) for flood levels prediction in Lower Ogun Basin, Nigeria. **Natural Hazards**, v.119, p. 2179–2195, 2023.

APARECIDO, L.E.D.O.; MORAES, J.R.D.S.C.D., LIMA; R.F.D.; TORSONI, G.B. 2022. Spatial Interpolation Techniques to Map Rainfall in Southeast Brazil. **Revista Brasileira de Meteorologia**, v.37, p. 141-155, 2022

Atlas Brasileiro de Desastres Naturais: 1991 a 2012 / Centro Universitário de Estudos e Pesquisas sobre Desastres. 2. ed. rev. ampl. – Florianópolis: CEPED UFSC, 2013.

BUI, Q.T; NGUYEN, Q.H.; NGUYEN, X. L.; PHAM, V. D.; NGUYEN, H. D; PHAM, V. M. Verification of novel integrations of swarm intelligence algorithms into deep learning neural network for flood susceptibility mapping. **Journal of Hydrology**, v. 581, 124379, 2020.

COLLINS, E.L.; SANCHEZ, G.M.; TERANDO, A.; STILLWELL, C.; MITASOVA, H.; SEBASTIAN, A.; MEETNTEMEYER, R. Predicting flood damage probability across the conterminous United States. **Environmental Research Letters**, 17, 2022.

DU, J., CHENG, L., ZHANG, Q. ET AL. Different Flooding Behaviors Due to Varied Urbanization Levels within River Basin: A Case Study from the Xiang River Basin, China. **Int J Disaster Risk Sci,** v.10, p. 89–102, 2019.

HADDAD, E. A.; TEIXEIRA, E. Economic impacts of natural disasters in megacities: The case of floods in São Paulo, Brazil. **Habitat International**, v. 45, p. 106-113, 2015.

HAWKER. L.; BATES, P.; NEAL, J.; ROUGIER, J. Perspectives on Digital Elevation Model (DEM) simulation for flood modeling in the absence of a high-accuracy open access global DEM. **Frontiers in Earth Science**, v. 6, p. 233, 2018.

HIRABAYASHI, Y.; MAHENDRAN, R.; KOIRALA, S.; KONOSHIMA, L.; YAMAZAKI, D.; WATANABE, S.; KIM, H.; KANAE, S. Global flood risk under climate change. **Nature Clim Change**, v. 3, p. 816–821, 2013

Histórica/Revista Online do Arquivo Público do Estado de São Paulo - Ano 10 - São Paulo: Governo do Estado de São Paulo, 2014.

Instituto Brasileiro de Geografia e Estatística (IBGE). Disponível em: <https://cidades.ibge.gov.br/brasil/sp/sao-paulo/panorama>. Acesso em: 20 de jun. de 2022.

IPCC, 2021: Summary for Policymakers. In: Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change [Masson-Delmotte, V., P. Zhai, A. Pirani, S.L. Connors, C. Péan, S. Berger, N. Caud, Y. Chen, L. Goldfarb, M.I. Gomis, M. Huang, K. Leitzell, E. Lonnoy, J.B.R. Matthews, T.K. Maycock, T. Waterfield, O. Yelekçi, R. Yu, and B. Zhou (eds.)]. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, pp. 3−32, doi:10.1017/9781009157896.001.

IPCC, 2022: Climate Change 2022: Impacts, Adaptation and Vulnerability. Contribution of Working Group II to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change [H.-O. Pörtner, D.C. Roberts, M. Tignor, E.S. Poloczanska, K. Mintenbeck, A. Alegría, M. Craig, S. Langsdorf, S. Löschke, V. Möller, A. Okem, B. Rama (eds.)]. Cambridge University Press. Cambridge University Press, Cambridge, UK and New York, NY, USA, 3056 pp., doi:10.1017/9781009325844.

EL-MAGD, A.; AHMED, S.. Random forest and naïve Bayes approaches as tools for flash flood hazard susceptibility prediction, South Ras El-Zait, Gulf of Suez Coast, Egypt. **Arabian Journal of Geosciences**, v. 15, n. 3, 2022.

FENG, B., ZHANG, Y., BOURKE, R. Urbanization impacts on flood risks based on urban growth data and coupled flood models. **Natural Hazards**, v. 106, p. 613–627, 2021.

LANDIM, P.M.B., STURARO, J.R.. Krigagem Indicativa aplicada à elaboração de mapas probabilísticos de riscos. Geomatemática, Texto Didático 6, DGA,0 IGCE, UNESP/Rio Claro, 2002. Disponível em: <http://www.rc.unesp.br/igce/aplicada/textodi.html>. Accesso em 16 de abril de 2022.

MELLO, C.R.; LIMA, J.M.; SILVA, A.M.; MELLO, O.J.M.; OLIVEIRA, M.S. Krigagem e inverso do quadrado da distância para interpolação dos parâmetros da equação de chuvas intensas. **Revista Brasileira de Ciência do Solo**, v. 27, p. 925-933, 2003.

MEMBELE, G. M.; NAIDU, M.; MUTANGA, O. Examining flood vulnerability mapping approaches in developing countries: A scoping review. **International Journal of Disaster Risk Reduction**, v. 69, p. 102766, 2022.

MOBLEY, W.; SEBASTIAN, A.; BLESSING, R.; HIGHFIELD, W.E.; STEARNS, L.; BRODY, S.D. Quantification of continuous flood hazard using random forest classification and flood insurance claims at large spatial scales: a pilot study in southeast Texas. **Natural Hazards Earth Systems Sciences**, v. 21: 807-822, 2021.

NORALLAHI, M.; KABOLI, H. S. Urban flood hazard mapping using machine learning models: GARP, RF, MaxEnt and NB. **Natural Hazards**, v.106, p. 119–137, 2021.

POURGHASEMI, H.R.; KARIMINEJAD, N.; AMIRI, M. et al. Assessing and mapping multi-hazard risk susceptibility using a machine learning technique. **Scientific Report**, v.10, p. 3203, 2020.

SIMÕES, S.J.C.; GOMES, L.; MENDES, R.M.; MENDES, T.S.G. SIG e modelos de escorregamentos: avaliando métodos para reduzir as incertezas de dados de solo e precipitação. **Revista Brasileira de Cartografia**, v. 9, 1737-1746, 2016

Sistema Integrado de Informações Sobre Desastres - S2iD. Ministério da Integração – MI. Disponível em: <http://S2iD.mi.gov.br/>. Acesso em: 01 de maio de 2022.

TAALAB, K.; CHENG, T.; ZHANG, Y. Mapping landslide susceptibility and types using random forest. **Big Earth Data**, v. 2, n. 2, p. 159–178, 2018.

TAROMIDEH, F.; FAZLOULA, R.; CHOUBIN, B.; EMADI, A.; BERNDTSSON, R. Urban Flood-Risk Assessment: Integration of Decision-Making and Machine Learning. **Sustainability**, v. 14, 4483, 2022.

TOMÁS, L.R.; SOARES, G.G.; JORGE, A.A.S.; MENDES, J.F.; FREITAS, V.L.S.; SANTOS, L.B.L. Flood risk map from hydrological and mobility data: A case study in São Paulo (Brazil). **Transactions in GIS**, v. 26, p. 2341–2365, 2022.

VALVERDE, M. C.; CALADO, B. N.; CALADO, G. G.; KUROKI, L.Y.; BRAMBILA, R.; SOUSA, A. R. Climate projections of precipitation and temperature in cities from ABC Paulista, in the Metropolitan Region of São Paulo-Brazil. **Frontiers in Clim,** v. 5, p.1127026, 2023.

XAVIER JÚNIOR, S. F. A.; JALE, J. D. S.; STOSIC, T.; SANTOS, C. A. C. D.; SINGH, V. P. Precipitation trends analysis by Mann-Kendall test: a case study of Paraíba, Brazil. **Revista Brasileira de Meteorologia**, v.35, p. 187-196, 2020.

WU, Y., ZHANG, W., HU, X. ET AL. Unravelling increasing flood hazard and influential factors in a tidal river. **Natural Hazards**, 2024.

ZOPE, P.E.; ELDHO, T.I.; JOTHIPRAKASH, V. Impacts of land use–land cover change and urbanization on flooding: A case study of Oshiwara River Basin in Mumbai, India. **Catena**, v. 145, p. 142-154, 2016.

YANTO, APRIYONO, A., SANTOSO, P.B. ET AL. Landslide susceptible areas identification using IDW and Ordinary Kriging interpolation techniques from hard soil depth at middle western Central Java, Indonesia. **Natural Hazards**, v. 110, p. 1405–1416, 2022.

ANEXO A – Artigo enviado à Revista Brasileira de Ciências Ambientais

Gmail **Beliana Cavalcante Sawada de Carvalho <beliana.carvalho@gmail.com>**

[RBCIAMB] Thanks for submission

1 mensagem

Revista Brasileira de Ciências Ambientais (RBCIAMB) via <rbciamb@abes-dn.org.br> 13 de janeiro de 2024 às 18:12
Responder a: "Revista Brasileira de Ciências Ambientais (RBCIAMB)" <rbciamb@abes-dn.org.br>
Para: Beliana Cavalcante Sawada de Carvalho <beliana.carvalho@gmail.com>

Dear Beliana Cavalcante Sawada de Carvalho,

We appreciate the submission of the paper "RAINFALL SPATIAL DISTRIBUTION IN THE CITY OF SÃO PAULO USING GEOSTATISTICS" to Brazilian Journal of Environmental Sciences (RBCIAMB).

Follow the progress of your submission through the system administration interface, available at:

Submission URL: https://www.rbciamb.com.br/Publicacoes_RBCIAMB/authorDashboard/submission/1914

Login: beliana

To complete the submission, the fee payment is required. To do so, access the link (Submission fee pay) and follow the payment procedure via PayPal for BRL 200.00.

Authors who are associated members to ABES, send an email to rbciamb@abes-dn.org.br informing the registration number to proceed with the fee exemption.

Editorial Team

Brazilian Journal of Environmental Sciences

ANEXO B – Artigo enviado à Revista Natural Hazards

Gmail Beliana Cavalcante Sawada de Carvalho <beliana.carvalho@gmail.com>

NHAZ-D-24-00110 - Submission Confirmation

1 mensagem

Natural Hazards (NHAZ) <em@editorialmanager.com> 16 de janeiro de 2024 às 01:27
Responder a: "Natural Hazards (NHAZ)" <ragavi.parthasarathy@springer.com>
Para: Beliana Carvalho <beliana.carvalho@gmail.com>

Dear CARVALHO BCS Carvalho,

Thank you for submitting your manuscript, Flooding and flash flood susceptibility mapping using Random Forest in São Paulo city, Brazil, to Natural Hazards.

During the review process, you can keep track of the status of your manuscript by accessing the journal's Editorial Manager site.

Your username is: beliana
If you forgot your password, you can click the 'Send Login Details' link on the EM Login page at https://www.editorialmanager.com/nhaz/

The submission id is: NHAZ-D-24-00110

Please refer to this number in any future correspondence.

Should you require any further assistance please feel free to e-mail the Editorial Office by clicking on "Contact Us" in the menu bar at the top of the screen.

With kind regards,
Springer Journals Editorial Office
Natural Hazards

Now that your article will undergo the editorial and peer review process, it is the right time to think about publishing your article as open access. With open access your article will become freely available to anyone worldwide and you will easily comply with open access mandates. Springer's open access offering for this journal is called Open Choice (find more information on www.springer.com/openchoice). Once your article is accepted, you will be offered the option to publish through open access. So you might want to talk to your institution and funder now to see how payment could be organized; for an overview of available open access funding please go to www.springer.com/oafunding. Although for now you don't have to do anything, we would like to let you know about your upcoming options.

This letter contains confidential information, is for your own use, and should not be forwarded to third parties.

Recipients of this email are registered users within the Editorial Manager database for this journal. We will keep your information on file to use in the process of submitting, evaluating and publishing a manuscript. For more information on how we use your personal details please see our privacy policy at https://www.springernature.com/production-privacy-policy. If you no longer wish to receive messages from this journal or you have questions regarding database management, please contact the Publication Office at the link below.

In compliance with data protection regulations, you may request that we remove your personal registration details at any time. (Use the following URL: https://www.editorialmanager.com/nhaz/login.asp?a=r). Please contact the publication office if you have any questions.