Revista Brasileira de Cartografia
ISSN 1808-0936 | https://doi.org/10.14393/revbrascartogr
Sociedade Brasileira de Cartografia, Geodésia, Fotogrametria e Sensoriamento Remoto

Dynamics of linear erosion processes in urban areas: an approach based on GIS and CART algorithm

Dinâmica dos processos erosivos lineares em áreas urbanas: uma abordagem apoiada em SIG e algoritmo CART

Tatiane Ferreira Olivatto[1], José Augusto Di Lollo[2] and Denise Balestrero Menezes[3]

1 Postgraduate Program in Urban Engineering, Federal University of São Carlos (UFSCar), São Carlos-SP, Brazil. tatianeolivatto@ufscar.br. ORCID: https://orcid.org/0000-0002-5770-7088
2 Engineering Faculty, São Paulo State University (UNESP), Ilha Solteira-SP, Brazil. jose.lollo@unesp.br. ORCID: https://orcid.org/0000-0002-6703-5377
3 Postgraduate Program in Urban Engineering, Federal University of São Carlos (UFSCar), São Carlos-SP, Brazil. denisebm@ufscar.br. ORCID: https://orcid.org/0000-0003-2962-3028

Received: 08.2025 | Accepted: 02.2026

Abstract: Linear erosion is a significant geomorphological process that progresses through distinct stages—rills, gullies, and ravines—shaped by natural and anthropogenic factors. While extensively studied in rural environments, its occurrence and dynamics in urban areas remain understudied, despite their implications for infrastructure stability and urban planning. This study investigates the relative influence of substrate conditions, land use, rainfall erosivity, and geomorphometric variables on the occurrence of gullies and ravines in urban areas, using a case study in São Paulo State, Brazil. An inventory of 1,398 mapped urban erosion features was analyzed using Geographic Information Systems (GIS) and a Classification and Regression Tree (CART) model. Results indicate that anthropogenic factors—particularly triggering actions related to surface runoff and stormwater management—are the dominant controls on urban linear erosion, surpassing traditional natural predictors such as rainfall erosivity. Geological conditions and geomorphometric attributes, especially elevation and plan curvature, play a secondary but meaningful role by conditioning the susceptibility of terrain units. The results also shown a cluster of geological formations with a lower propensity for gully development. The final pruned CART model achieved balanced classification performance (overall accuracy of 72.7%, AUC = 0.78), demonstrating robust generalization and limited overfitting. These findings highlight the central role of urban hydrological alteration in erosion processes and emphasize the need to integrate erosion risk assessment into urban planning and stormwater management strategies, contributing to more resilient and sustainable cities.
Keywords: Gully. Ravine. Land Use. Geographic Information System. Decision Tree.

Resumo: A erosão linear é um importante processo geomorfológico que evolui por estágios distintos — sulcos, ravinas e voçorocas — sendo condicionada por fatores naturais e antrópicos. Embora amplamente estudada em ambientes rurais, sua ocorrência e dinâmica em áreas urbanas permanece pouco investigada, apesar de suas implicações para a estabilidade da infraestrutura e o planejamento urbano. Este estudo analisa a influência relativa das condições do substrato, do uso do solo, da erosividade das chuvas e de variáveis geomorfométricas na ocorrência de ravinas e voçorocas em áreas urbanas, a partir de um estudo de caso no estado de São Paulo, Brasil. Um inventário de 1.398 feições de erosão linear urbana mapeadas foi analisado usando Sistemas de Informação Geográfica (SIG) e um modelo de Árvores de Classificação e Regressão (CART). Os resultados indicam que fatores antrópicos — particularmente ações desencadeadoras associadas ao escoamento superficial e ao manejo das águas pluviais — constituem os principais controles da erosão linear urbana, superando preditores naturais tradicionais, como a erosividade da chuva. As condições geológicas e os atributos geomorfométricos, especialmente a altitude e o plano de curvatura, exercem um papel secundário, porém relevante, ao condicionarem a suscetibilidade das unidades do terreno. Os resultados também evidenciaram um agrupamento de formações geológicas com menor propensão ao desenvolvimento de voçorocas. O modelo CART final podado apresentou desempenho de classificação equilibrado (acurácia global de 72,7%, AUC = 0,78), indicando boa capacidade de generalização e baixa suscetibilidade ao sobreajuste. Esses achados reforçam o papel central das alterações hidrológicas induzidas pela urbanização nos processos erosivos e destacam a necessidade de integrar a avaliação do risco de erosão ao planejamento urbano e às estratégias de gestão das águas pluviais, contribuindo para cidades mais resilientes e sustentáveis.
Palavras-chave : Voçoroca. Ravina. Uso do Solo. Sistema de Informação Geográfica. Árvore de Decisão.

1 INTRODUCTION

Soil erosion is among the most damaging environmental degradation processes on the globe, reaching urban and rural populations. The consequences of erosion processes, besides damaging the environment (Chalise et al., 2019; Issaka & Ashraf, 2017), have economic and social reflexes, such as loss of agricultural areas, compromising food production and material and human losses in urban occupations in risk zones (Pierce & Lal, 2017; Pimentel, 2006; Rodrigo-Comino et al., 2018; Telles et al., 2011; van Zelm et al., 2018; Zdruli et al., 2016).

Estimates show that 50% of global soil erosion is caused by water (Blanco-Canqui & Lal, 2010). The Brazilian Southeastern region holds most of the country's economic activities (55.4%), as well as 42.6% of the Brazilian population (IBGE, 2010). Most erosion processes are observed in the Paraná River basin, where natural and anthropic conditions have intensified these processes (Anache et al., 2017; Costa et al., 2015; Dorici et al., 2016; Sparovek et al., 2007).

Several researchers rely on studies about accelerated soil erosion as core environmental degradation, which is triggered by activities such as changes in land cover, inappropriate agricultural practices, as well as wrong occupation of slope areas, water bodies' neighbourhoods' and land prone to erosion, besides associated socio-environmental issues (Mota, 1995; Zuquette et al., 2013; Di Lollo et al., 2018; Carvalho et al., 2019).

Understanding erosion susceptibility as inherent to the environment, it is possible to define different potentials of predisposition to erosion-processes development depending on a combination of several conditions. Cunha and Guerra (2009) categorized the factors triggering and/or controlling erosion processes as follows: geomorphometric features, land use/cover, geological material, and rainfall erosivity.

In this context, the spatial distribution and dynamics of erosion-conditioning factors are typically assessed using cartographic approaches, particularly through susceptibility maps. This can be achieved using different information treatment techniques, as observed in Pons et al.(2007), Di Lollo and Sena (2013), Galharte et al. (2014), Durães and Mello (2016), Javidan et al. (2019) and Prieto-Amparán et al. (2019).

Despite the increasing concerns over soil erosion in urban environments, most existing studies focus on rural areas, leaving a significant gap in understanding the specific drivers and mechanisms of linear erosion in cities. Traditional approaches often emphasize natural factors, such as geomorphology and rainfall erosivity, but fail to adequately incorporate anthropogenic influences, which are dominant in urban settings. This study advances the field by integrating both natural and human-induced variables—such as upstream and downstream land use—into an erosion susceptibility analysis, employing a Classification and Regression Tree (CART) model. While decision tree-based methods have been widely applied in environmental modeling, their use in urban erosion studies remains limited. By applying this methodological approach to a densely populated and highly urbanized environment, this research provides new insights into the role of land use dynamics and stormwater management in shaping urban erosion processes. Moreover, the framework developed is not limited to a specific geographic region and can be adapted to assess erosion susceptibility in various urban contexts worldwide.

Acknowledging the relevance of this topic for a better understanding of the triggering and evolution of erosion processes, and considering the scarcity of studies focusing on urban areas, the aim of the present study was to assess the influence of natural and anthropic conditions on linear erosion processes observed in urban areas of São Paulo State, Brazil.

1.1 Gully erosion in urban areas

Urban linear erosion is widely recognized as a direct outcome of anthropogenic disturbances to local hydrological systems. Rapid and frequently unplanned urban growth modifies natural runoff dynamics through vegetation removal, soil compaction, surface sealing, and the artificial concentration of flow along streets, plots, and drainage structures. These changes increase runoff volume and velocity, allowing surface water to exceed soil resistance and initiate incision processes that may rapidly evolve into larger erosional features (Júnior et al., 2010; De Albuquerque et al., 2020; Arantes et al., 2021; Mawe et al., 2025).

A consistent finding in the literature is the central role of roads, unpaved surfaces, and deficient stormwater drainage in triggering urban gullies. Streets, road embankments, drainage ditches, and concentrated

discharge points often function as preferential flow paths, promoting headcut initiation and channel deepening, particularly on steep or topographically convergent slopes (Gudiño-Elizondo et al., 2018; Rahmati et al., 2022; Imwangana et al., 2025). Evidence from both Global South cities and Brazilian case studies shows that even areas with moderate natural susceptibility can develop severe erosion once urban hydrology is profoundly altered (Zolezzi et al., 2018; Mawe et al., 2024).

Natural controls such as slope gradient, soil texture, lithology, and groundwater behavior strongly condition the spatial distribution and rate of erosion, but their influence is commonly amplified by human interventions. Sandy or deeply weathered soils, typical of many tropical and subtropical urban environments, are particularly vulnerable when combined with increased surface runoff and reduced infiltration capacity (Mahamba et al., 2023; Chen et al., 2024; IPT, 2012). Several studies indicate that erosion accelerates once channels intersect the groundwater table, intensifying headward retreat and lateral expansion (Frankl et al., 2020; Mawe et al., 2024).

Urban linear erosion produces substantial environmental, social, and economic impacts, including sediment delivery to downstream channels, reservoir siltation, infrastructure damage, housing loss, land devaluation, and population displacement (Kuhn et al., 2024; Mawe et al., 2025). These effects are most acute in urban peripheries, where informal or poorly serviced settlements frequently coincide with steep slopes and inadequate drainage infrastructure, a pattern long documented in Brazilian cities (Almeida Filho et al., 2001).

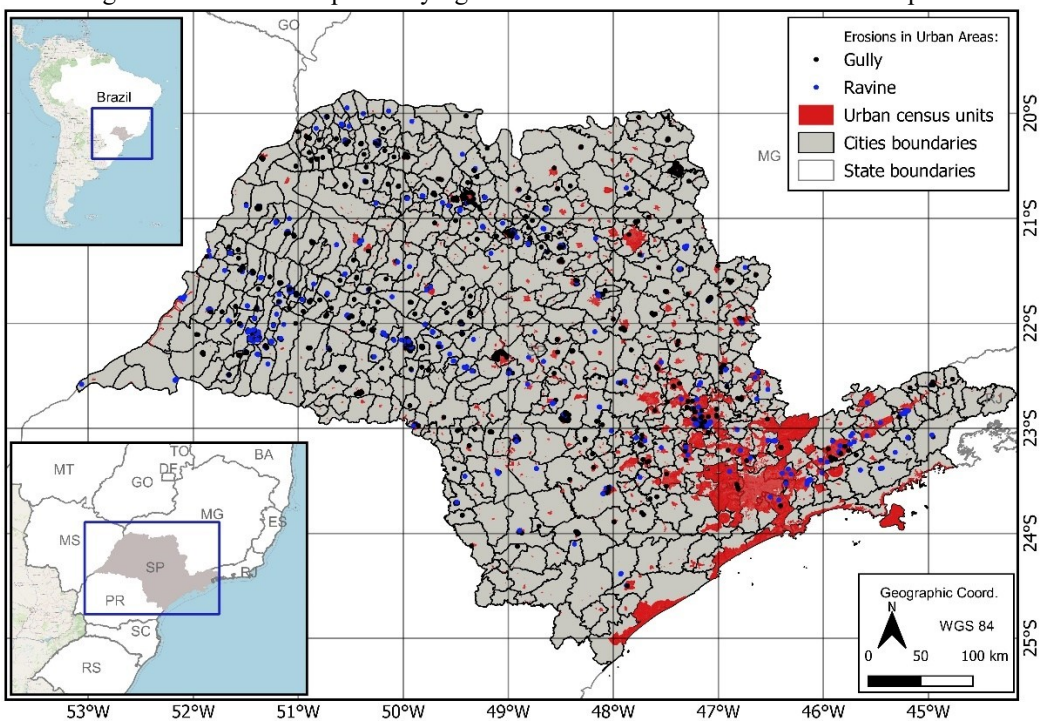
Overall, urban linear erosion reflects a structural mismatch between land-use practices and the physical capacity of the landscape. Contemporary studies converge in showing that effective mitigation relies less on isolated corrective works and more on integrated strategies combining land-use regulation, stormwater management, vegetation cover maintenance, and planning informed by geomorphological and hydrological constraints (Frankl et al., 2020; Mawe et al., 2025).

2 MATERIALS AND METHODS

São Paulo State was chosen as the study area, with a focus on urban regions defined by census-based urban-rural classifications (IBGE, 2010). The erosion features analyzed in this study were obtained from the official inventory produced by the Technological Research Institute (IPT). According to the methodological report (IPT, 2012), the identification of urban erosion processes was carried out through visual interpretation of high-resolution satellite images available in Google Earth (SPOT-5 and GeoEye-1), complemented by the analysis of 1:50,000 IBGE topographic maps. The erosion features identified through image interpretation were subsequently evaluated and systematized in standardized inventory forms, followed by field surveys, contributing to the improvement of database quality. Therefore, the present study did not perform the delimitation or validation of erosion features inventoried in 2010, relying instead on this previously established and validated official dataset. The location of São Paulo state, the identification of urban census sectors and the 1,398 analyzed urban erosion features (949 gullies; 449 ravines) are shown in Figure 1.

It is important to emphasize that the classification of linear erosion features adopted in this study follows the conceptual framework used by the IPT, which is consistent with a widely applied geomorphological tradition in Brazilian erosion studies. According to this approach, linear erosion developed from rills - shallow and narrow incisions formed by concentrated surface runoff – to ravines. Ravines are elongated erosive features, generally longer than wide, with variable depths, limited lateral branching, and without interception of the groundwater table. Their lateral expansion results from surface runoff within the channel, promoting undercutting at the base of the slopes and subsequent mass movements. Minimum depths reported in the literature for ravines range from approximately 30 cm (Tricart, 1977) to 50 cm (Imeson & Kwaad, 1980). When surface runoff processes evolve to intercept the groundwater table, a combination of surface and subsurface erosion mechanisms leads to the development of gullies (locally referred to as voçorocas), which are larger and more complex erosive forms associated with processes such as piping, sand liquefaction, and slope failures (Oliveira, 1994; Salomão, 1994).

Figure 1 – Location map identifying urban census sectors and urban erosion spots.



Source: Authors (2026).

This study therefore adopts the erosion typologies as defined in the IPT inventory, acknowledging that alternative classifications may exist in the international literature. The investigation aimed not only to understand this phenomenon but also to identify the data required to guide the analysis and methodological choices. Accordingly, two main steps were established: (i) data extraction and processing and (ii) data mining.

2.1 Data extraction and processing step

This stage consisted of extracting data of interest from several sources and processing them in the Geographic Information System (GIS) environment of ArcGIS Pro 2.8.3 software (Esri, 2021). A database of ravine and gully linear erosion types mapped by IPT (2012) was elaborated by associating features related to the four categories presented in the introduction section: substrate-related factors (IPT, 2012), land use (IPT, 2012; MapBiomas Project, 2010), rainfall erosivity (Ricardi & Lima, 2022) and geomorphometric features (CATI, 2016).

Each mapped erosion feature was treated as an individual analytical unit and used as the reference element for database construction. Attributes related to substrate characteristics and part of the land use information were already available in the original IPT (2012) erosion inventory and, therefore, did not require spatial overlay procedures, being subjected only to a preprocessing stage for the organization and standardization of categorical variables, as detailed in the subsequent sections. The remaining datasets—land use information derived from the MapBiomas Project, rainfall erosivity, and geomorphometric features obtained from the Digital Elevation Model (DEM)—were spatially intersected with the linear features representing gullies and ravines. As a result, a structured geodatabase was developed in which each erosion feature (row) was associated with a set of attributes (columns) corresponding to the dependent variables of the four analytical categories considered in this study. This procedure enabled the integration of heterogeneous data sources into a single, consistent database for subsequent data mining analyses.

2.1.1 SUBSTRATE-RELATED FACTORS

Data related to substrate characteristics were already available in feature ‘assessed database source standard’. They were the only ones not requiring reclassification or processing because they were organized into well-structured categories.

Substrate-related features were observed in features mapped by IPT, namely: geomorphology, geology and pedology. Possible occurrences in each feature are shown in Chart 1.

Chart 1 – Substrate-related features and occurrences linked to erosion processes.

Attributes	Occurrence
Geomorphology[1] (based on Ross & Moroz, 1997)	Lower Ribeira Depression; Middle Paraíba Depression; Paulista Peripheral Depression; Atlantic Plateau; São Paulo Plateau; Western Paulista Plateau; Coastal and Fluvial Plains.
Geology[2] (based on DAEE/UNESP, 1982)	Coastal Complex – Migmatites; Embu Complex – Migmatites; Pinhal Complex – Migmatites; Silvianópolis Complex – Embrechitic gneiss; Silvianópolis Complex – Granulites; Alluvial deposits; Colluvial deposits; Summit deposits; Caçapava Formation; Rio Claro Formation; São Paulo Formation; Tremembé Formation; Açungui Group – Mica schists; Amparo Group – Paragneiss; Bauru Group – Adamantina Formation; Bauru Group – Caiuá Formation; Bauru Group – Marília Formation; Bauru Group – Santo Anastácio Formation; Paraná Group – Furnas Formation; Passa Dois Group – Corumbataí Formation; São Bento Group – Botucatu Formation; São Bento Group – Pirambóia Formation; São Bento Group – Serra Geral Formation; São Roque Group – Phyllites; São Roque Group – Schists; Tubarão Group – Aquidauana Formation; Tubarão Group – Tatuí Formation; Tubarão Group – Itararé Subgroup; Basic suites; Granitoid suites.
Pedology[3] (based on Oliveira et. al., 1999)	Área urbana – solo não classificado (Urban soils / Technosols); Argissolos Vermelho-Amarelos (Acrisols); Argissolos Vermelhos (Acrisols); Cambissolos Háplicos (Cambisols); Gleissolos Melânicos (Gleysols); Latossolos Amarelos (Ferralsols); Latossolos Vermelho-Amarelos (Ferralsols); Latossolos Vermelhos (Ferralsols); Neossolos Litólicos (Leptosols); Neossolos Quartzarênicos (Arenosols); Nitossolos Vermelhos (Nitisols).

[1] Geomorphological units were referenced according to the official Brazilian geomorphological map, with translation limited to generic geomorphological terms while preserving original toponyms.
[2] Geological units follow the official nomenclature of the São Paulo State geological mapping. Formal lithostratigraphic names were preserved, while lithological terms were translated into English.
[3] Soil classes according to the Brazilian Soil Classification System (SiBCS). For international reference, the corresponding classes according to the World Reference Base for Soil Resources (FAO, 2022) are also indicated. The conversion to WRB was performed at the Reference Soil Group level, which does not explicitly account for color attributes used in the SiBCS.

Source: Authors (2026).

2.1.2 LAND USE

Land use data, although already belonged to the original source's features, refer to textual information and were not set into categories, which is a requirement for the CART decision tree algorithm. Therefore, it was necessary to organize the collected information into established categories to make the data mining technique application feasible.

These data's reorganization focused on originally-divergent textual data due to their plural and singular use, to terms' reversed ordering, to detailed featuring in some cases – but generalized in others (for example: surface runoff and rainwater runoff from the urban area discharged at the river head), among others. This process led to reduction ranging from 82 to 10 possible occurrences set for upstream use, from 89 to 8 for downstream use data, and from 134 to 5 for triggering action data. The land use attributes, divided into "upstream use", "downstream use" – both in comparison to the mapped erosion features – and "triggering action", as the established categories are shown in Chart 2.

Considering data descriptive nature, requiring categorization due to its large number of classes, option was made for incorporating into the analyses the land cover data extracted from the MapBiomas Project for the year 2010 (the year of data collection). This incorporation was justified by the greater standardization of classes and by enabling a possible comparison in future applications of this same methodology, as the case may be. This incorporation, included in Chart 2, also allowed the attenuation of any generalizations resulting from the categorization step.

For the upstream and downstream use attributes, occurrences categorized as Rural include agricultural use, pasture (with and without terraces) and exposed soil. Urban features include residential, commercial and industrial uses. It is also important to mention that, although the attribute "triggering action" does not specifically regard soil use, it was added to the analysis due to its association with anthropic activities that influence the environmental soil dynamics.

Chart 2 – Land use features and occurrences linked to erosion processes.

Attributes	Occurrence
Upstream use	Urban; Urban in consolidation; Urban and Rail/Highway (or Rail/Highway only); Urban in consolidation and Railway/Highway; Urban and Rural; Urban in consolidation and Rural; Rural; Rural and Railway/Highway (or Railway/Highway only); Drainage; Vegetation
Downstream use	Urban; Urban in consolidation; Urban and Rural; Urban in consolidation and Rural; Rural; Rural and Vegetation; Drainage; Vegetation
Triggering actions	Fluvial Erosion; Surface runoff; Stormwater disposal (pipe); Surface runoff and stormwater disposal; Railroad/Highway stormwater Runoff/Disposal
Land cover (Mapbiomas Project, 2010)	Farming; Urbanized Area; Forest; Non-Forest Natural Formation; Agriculture and Pasture Mosaic

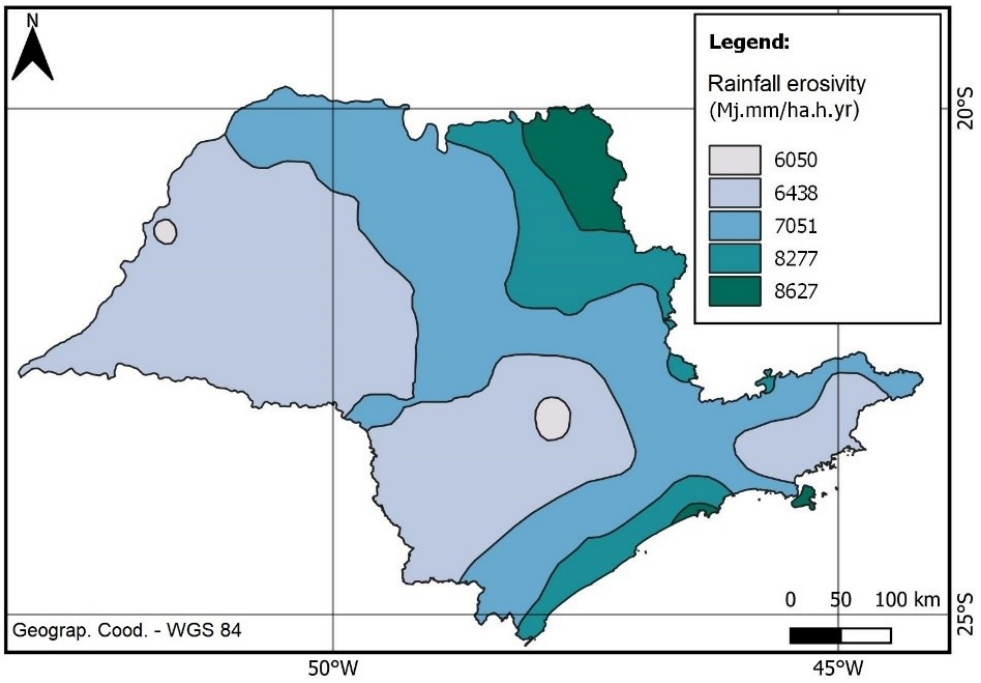
Source: Authors (2026).

2.1.3 RAINFALL EROSIVITY

Data on rainfall erosivity were obtained from the study conducted by Ricardi and Lima (2022), who updated rainfall erosivity estimates set for São Paulo State based on yearly and monthly averages between 1997 and 2017. These data resulted from the spatial extrapolation of measurements taken from the state's rain gauge stations during the aforementioned time interval – these datasets were made available by the Water and Electric Power Department (DAEE, 2019).

In order to add these data to the herein proposed analyses, the vectorization of the cartographic product about rainfall erosivity was carried out in GIS environment, in the form of polygons (see Figure 2), followed by the intersection with the erosion records - resulted in values ranging from 6,050 to 8,627 Mj.mm/ha.h.yr. In total, 10 of the 1,398 points were located in regions presenting erosivity of 6,050 Mj.mm/ha.h.yr, 718 were in regions with erosivity 6,438 Mj.mm/ha.h.yr, 520 were in places with erosivity 7,051 Mj.mm/ha.h.yr, 80 were in location with erosivity 8,277 Mj.mm/ha.h.yr, and 70 were in regions with erosivity 8,627 Mj.mm/ha.h.yr.

Figure 2 – São Paulo State Rainfall Erosivity Map for the 1997-2017.



Source: Authors (2026).

2.1.4 GEOMORPHOMETRIC VARIABLES

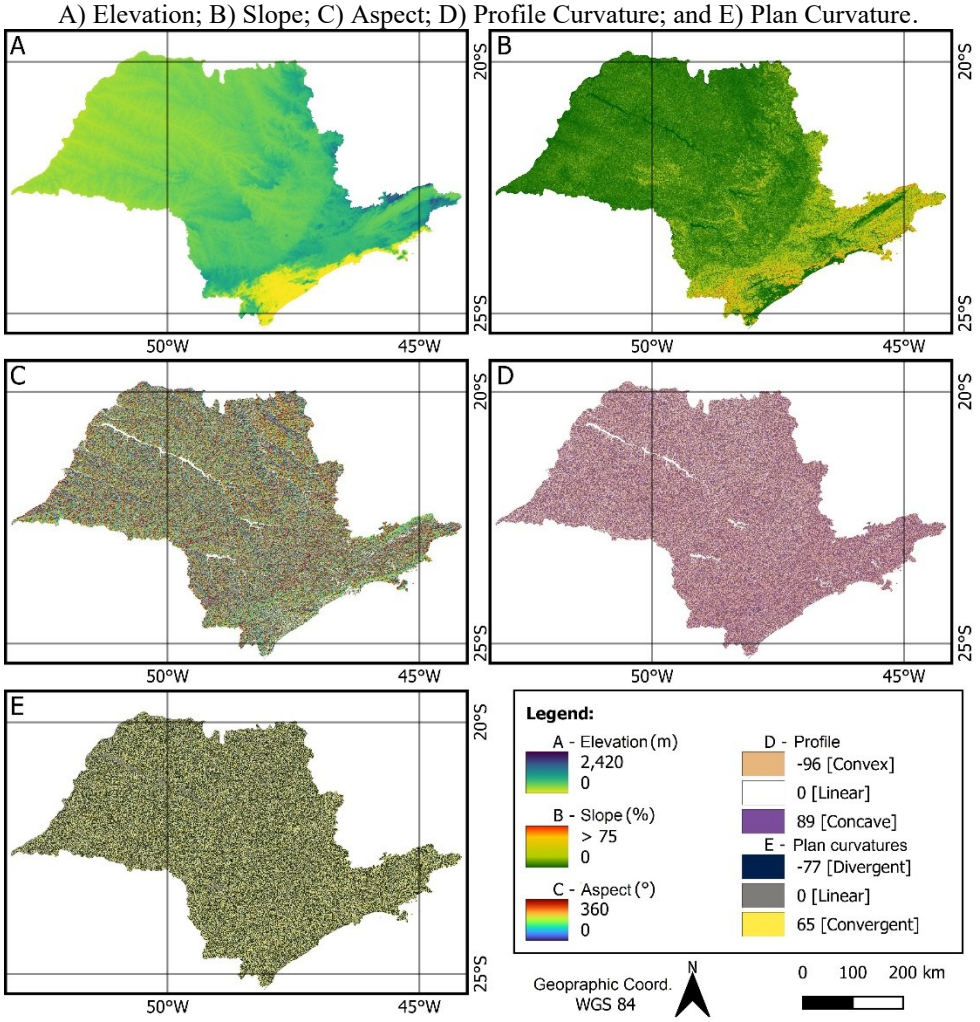
In the relief study the following features were analyzed: elevation, slope, aspect, profile and plan curvature. The Digital Elevation Model (DEM) prepared by the Brazilian Farming Information Center (CATI, 2016) was used for data extraction purposes. It was elaborated based on products from NASA's Shuttle Radar

Topography Mission (SRTM), with 30m spatial resolution; corrected with data from NASA and METI's ASTER GDEM V2 sensor, also with 30m spatial resolution, and geoid ripple for São Paulo State.

First of all, based on the aforementioned base, a hypsometric map was elaborated, as well as maps of slope (using the Slope GIS tool), aspect (using the Aspect GIS tool), profile and plan curvature (using the Curvature GIS tool). It was decided not to categorize these data, given their continuous numerical features. Altitudes were recorded in meters, slope in percentage, aspect in degrees (azimuth angle corresponding to the greatest slope of the terrain), profile curvature in negative values for convex, in positive values for concave and in null values for linear; and plan curvature in positive values for divergent, in negative values for convergent and in null values for linear.

Data forming geomorphometric characterization - DEM-based geoprocessing - led to 5 intermediate maps (Figure 3). As from erosive processes records, its locations were crossed with the produced geomorphometric layers – through the Intersect GIS tool. Subsequently, elevation, slope, aspect, profile curvature and plan curvature values recorded for each point were extracted.

Figure 3 – Maps of geomorphometric variables.
A) Elevation; B) Slope; C) Aspect; D) Profile Curvature; and E) Plan Curvature.



Source: Authors (2026).

Altitudes ranged from 0 to 2,420 meters, being the mean altitude 557.36 meters and the most frequent one 487.50 meters. Mean slope was close to 11%, presenting the highest occurrence at the range from 4% to 6% - slopes above 20% remained lower than 10% of all cases. Aspect, which ranged from 0° to 360°, recorded quite varying occurrences, with prevalence in Northeastern (45°), Southeastern (135°), Southwestern (225°) and Northwestern (314°). With respect to the profile curvature, 43.27% of collected data referred to positive values, assigned to concave profile; 49.49% to negative values, assigned to convex profile; and 7.24% to null values, which corresponds to curvature-free regions. Finally, regarding plan curvature, it was possible observing 39.32% positive values, which designate convergent planes; 48.49% negative values, designating

divergent planes; and 12.19% null values that underline linear planes.

2.2 Data mining step

The application of data mining methods in studies focused on linear erosion processing have shown overall upward trends, mainly from 2018 onwards (according to search in Scopus database carried out by the present authors in December 2022). Random Forest, Decision Tree and Artificial Neural Networks were the most often applied methods. In this work, the Decision Tree (DT) was chosen because it is a simplified model capable of representing associations between a given dataset (independent variables) and a dependent variable (Medeiros et al., 2014; Quinlan, 1983). IBM SPSS Statistics 20 software (2011) was used for method application purposes.

This technique runs a decision-making procedure (bipartition of occurrences) based on algorithms that classify data according to their homogeneity towards the investigated variable. CART is one of the most often applied algorithms (Breiman et al., 1984), which was also used in the current research. It adopts the Gini Index (GI) to achieve data binary partition – it can range from 0 to 1: the closer to 0, the purer the node, consequently, the more homogenous. Among other words, observations branch from independent variables' attributes - taking into consideration the highest occurrence of a single category of the dependent variable (Breiman et al., 1984).

The decision-making procedure is repeated across multiple levels of the tree. Therefore, several hyperparameter configurations were systematically tested in order to control tree complexity, optimize predictive performance, and address class imbalance between erosion types. Model tuning focused on the main stopping rules of the CART algorithm: maximum tree depth, minimum number of cases in parent and child nodes, and minimum improvement in node impurity based on the Gini Index.

All experiments were conducted using a 10-fold cross-validation procedure, with the maximum tree depth fixed at 10. The tuning strategy was organized into three stages. In the first stage, models without misclassification costs were tested using two alternative minimum node size configurations (49/7 and 70/14 cases in parent and child nodes, respectively), aiming to assess the trade-off between model complexity and classification balance. In the second stage, an asymmetric cost matrix was introduced to mitigate class imbalance, and the same node size configurations were compared. In the third stage, the impact of tree pruning was evaluated by varying the minimum Gini improvement threshold (0.001, 0.002, and 0.003) under the selected cost structure.

Preliminary tests also explored the use of class weights; however, these configurations resulted in reduced predictive performance and were therefore not retained. Instead, an asymmetric misclassification cost matrix was adopted, in which no cost was assigned to correct classifications, while higher penalties were imposed on misclassification errors. Misclassifying a gully as a ravine was assigned a cost of 1.0, whereas misclassifying a ravine as a gully was assigned a higher cost of 1.5. This configuration reflects the greater analytical and practical relevance of correctly identifying gullies, which generally represent more advanced and severe erosion processes.

Model performance was primarily evaluated using cross-validation risk and overall accuracy, with particular emphasis on class-specific accuracies in order to assess balance between ravines and gullies. The final model was selected based on a combination of (i) predictive performance, (ii) balance between classes, and (iii) model parsimony, favoring configurations that reduced overfitting while maintaining interpretability.

In accordance with the implementation available in IBM SPSS Statistics, the average overall accuracy across the folds was used as the primary metric for cross-validation-based model comparison. The Area Under the Receiver Operating Characteristic Curve (AUC–ROC) was computed directly within the software environment. Subsequently, based on the confusion matrix obtained for the selected model, additional performance metrics—including precision, sensitivity (recall), and F1-score—were calculated to provide a more detailed evaluation of classification performance, particularly with respect to class-specific behavior (Powers, 2011).

The attributes used in this investigation, as well as their likely occurrences, were described in item 2.1. The linear erosion type (with two associated categories: ravine and gully) was adopted as dependent variable,

and the other variables were the independent variables, as described in Chart 3 (which also indicates the variable's typology).

Chart 3 – Features of the investigated variables.

Type	Nature	Variable	Typology
Dependent	Linear erosion	Ravines	Textual
		Gullies	Textual
Independent	Substrate-related	Geomorphology	Textual
		Geology	Textual
		Pedology	Textual
	Land use	Upstream use	Textual
		Downstream use	Textual
		Triggering action	Textual
		Land cover (MapBiomas)	Textual
	Erosivity	Rainfall erosivity	Numeric
	Geomorphometric	Elevation	Numeric
		Slope	Numeric
		Aspect	Numeric
		Profile curvature	Numeric
		Plan curvature	Numeric

Source: Authors (2026).

3 RESULTS AND DISCUSSION

According to IPT data, 41,262 ravine and gully linear erosion processes spots were identified and characterized in São Paulo State, of which 1,398 were observed in urban areas. If one has in mind that 96,7% of the total population in the state lives in these areas (IBGE, 2022), then, one can see the relevance of learning more about the association between triggered erosion processes in urban areas.

Therefore, the herein suggested data mining methodology aimed to establish these associations by investigating the most important agents for erosion occurrences. The data extraction and processing stage was crucial for structuring the dataset to ensure compatibility with the software and for integrating attributes already recognized in the literature as influential. The data mining step could start once values are extracted from the target features of each of the 1,398 erosion process features.

3.1 Data Mining Step

To evaluate the sensitivity of the CART classifier to hyperparameter choices and to address class imbalance between ravines and gullies, a set of six models was trained and assessed using 10-fold cross-validation. The tested configurations varied the minimum number of cases in parent and child nodes, the minimum improvement in Gini impurity required for splitting, and the use of asymmetric misclassification costs. Table 1 summarizes the structural characteristics of each tree and their respective predictive performances, including class-specific accuracies, overall accuracy, and resubstitution and cross-validation risks.

Table 1 – Summary of CART hyperparameter tests results.

Models	Misclassification Costs		Minimum samples		Minimum Gini	Number of		TD	Accuracy (%)			Risk		
	R	G	PN	CN		N	TN		R	G	$\bar{x}$	Resubstitution	CV	Difference
1	0.00	0.00	49	7	0.001	49	25	9	51.2	88.7	76.7	0.233	0.336	0.103
2	0.00	0.00	70	14	0.001	31	16	8	57.5	82.6	74.5	0.255	0.300	0.045
3	1.00	1.50	49	7	0.001	49	25	9	74.2	76.0	75.4	0.288	0.418	0.130
4	1.00	1.50	70	14	0.001	31	16	8	75.1	72.4	73.2	0.308	0.407	0.099
5	1.00	1.50	70	14	0.002	29	15	7	79.5	69.0	72.4	0.309	0.413	0.104
6	1.00	1.50	70	14	0.003	19	10	5	80.8	65.0	70.1	0.330	0.399	0.069

Note: R = ravine; G = gully; PN = parent node; CN = child node; N = nodes; TN = terminal nodes; TD = tree depth; $\bar{x}$ = overall; CV = Cross-Validation. Source: Authors (2026).

In the first stage (Models 1 and 2), only the minimum number of cases in parent and child nodes was varied, while no misclassification costs were applied. Both models achieved relatively high overall accuracies (76.7% and 74.5%, respectively), but exhibited strong class imbalance, with substantially higher accuracy for gullies than for ravines. Model 1, with smaller minimum node sizes (49–7), produced a deeper and more complex tree (depth = 9), whereas Model 2 (70–14) resulted in a simpler structure (depth = 8) with slightly lower overall accuracy but reduced risk difference, suggesting improved generalization.

In the second stage (Models 3 and 4), an asymmetric misclassification cost matrix was introduced to mitigate class imbalance. The inclusion of costs markedly improved the balance between ravine and gully accuracies. Compared to their cost-free counterparts, both models exhibited a substantial increase in ravine accuracy, indicating that the cost structure effectively penalized misclassification of ravines. Model 4 (70–14) achieved a more compact tree while maintaining comparable class-specific accuracies, demonstrating that the cost matrix improved class balance without excessively increasing model complexity.

In the third stage (Models 4, 5, and 6), the minimum Gini improvement required for splitting was increased from 0.001 to 0.002 and 0.003 while maintaining the same node size and cost configuration. As the Gini threshold increased, tree complexity was progressively reduced, with a decrease in both depth and number of terminal nodes. However, this simplification came at the expense of predictive balance: although ravine accuracy continued to increase, gully accuracy declined substantially, leading to lower overall accuracy and reduced class equilibrium in Models 5 and 6. These results indicate that overly restrictive splitting criteria can oversimplify the model and degrade classification performance.

Based on the joint evaluation of predictive performance, class balance, and model complexity, Model 4 was selected as the basis for the final classifier. Although it did not achieve the highest overall accuracy among all tested configurations, it provided the best trade-off between ravine and gully accuracies, avoiding the strong class bias observed in cost-free models and the performance degradation associated with more restrictive Gini thresholds. To further enhance generalization and interpretability, this model was subsequently pruned to a maximum tree depth of five levels.

This pruning strategy substantially reduced structural complexity while preserving virtually the same predictive performance, resulting in a more parsimonious and robust representation of erosion processes. The final pruned model maintains a balanced classification behavior between erosion types and exhibits a moderate difference between resubstitution and cross-validation risks, indicating good generalization capacity and limited susceptibility to overfitting. This configuration therefore represents a robust and interpretable solution for mapping urban linear erosion processes.

Table 2 summarizes the metrics derived from the confusion matrix of the selected pruned model, as well as AUC values and confidence interval (CI) from ROC analysis based on predicted probabilities for each erosion class under 10-fold cross-validation. The final pruned CART model achieved an overall accuracy of 72.7%, with class-specific performance metrics indicate a balanced predictive behavior between erosion types. For ravines, the model obtained a recall of 71.3%, demonstrating good sensitivity in identifying this class, although with moderate precision (55.9%), reflecting residual confusion with gullies. For gullies, precision reached 84.4%, while recall was 73.4%, indicating a high reliability of positive predictions and a satisfactory detection rate.

Table 2 – Performance metrics of the final CART model.

Metric	**Ravine**	**Gully**
Accuracy	0.727	
Precision	0.559	0.844
Recall	0.713	0.734
F1-score	0.626	0.785
AUC (ROC)	0.780	
95% CI (AUC)	0.755–0.805	
Resubstitution Risk	0.319	
Cross-Validation Risk	0.397	

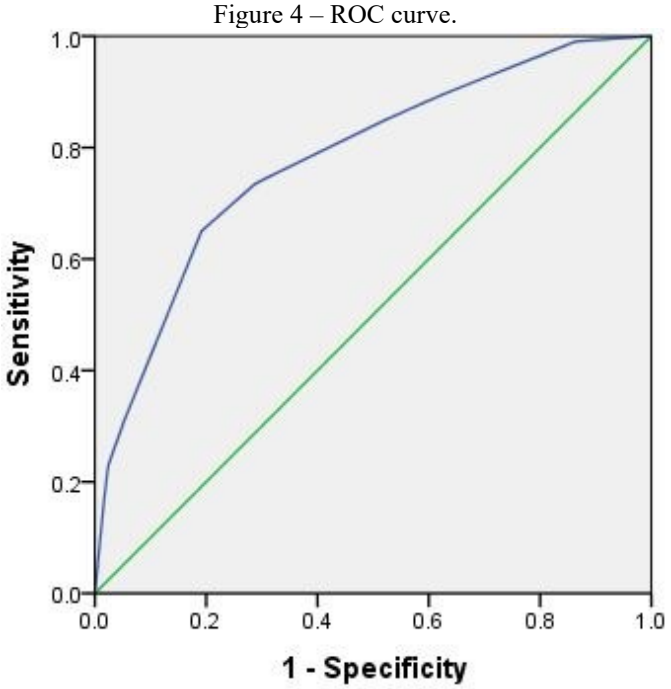
Source: Authors (2026).

The resulting F1-scores (0.626 for ravines and 0.785 for gullies) confirm that the adopted cost-sensitive

configuration effectively mitigated the strong class bias observed in preliminary models without misclassification costs. Importantly, the relatively small gap between resubstitution and cross-validation risks (0.078) further suggests improved generalization and reduced overfitting due to pruning.

ROC analysis further supports the robustness and generalization capacity of the final pruned CART model. The ROC curve yielded an AUC of 0.780, with a 95% confidence interval ranging from 0.755 to 0.805 ($p < 0.001$), indicating good discriminatory ability between erosion types. This result confirms that the classifier performs substantially better than random assignment and maintains consistent predictive power across different probability thresholds. The shape of the ROC curve (see Figure 4) reveals a stable trade-off between sensitivity and specificity over a broad range of cutoff values, suggesting that the model's performance is not overly dependent on a single decision threshold. This behavior is particularly relevant in urban erosion studies, where management priorities may favor either higher sensitivity (to avoid missing severe gully features) or higher specificity (to reduce false alarms), depending on planning and mitigation objectives. Therefore, the ROC analysis reinforces the suitability of the selected model for operational and decision-support contexts, complementing class-specific metrics and supporting the interpretation that the adopted pruning strategy effectively balances predictive performance and model generalization.

Figure 4 – ROC curve.



Note: diagonal segment produced by ties. Source: Produced by the authors at SPSS tool (2026).

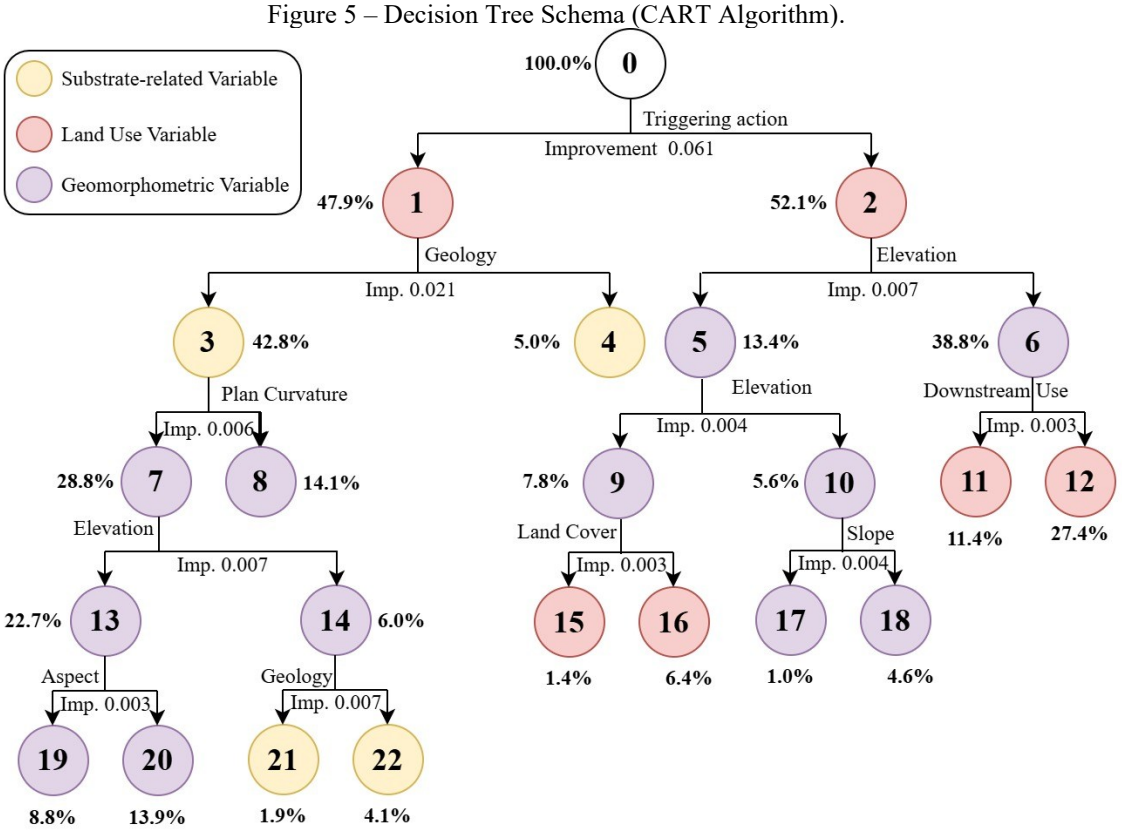
Figure 5 depicts the final CART decision tree obtained after hyperparameter optimization and the incorporation of asymmetric misclassification costs. The tree structure represents the hierarchical decision process based on the explanatory variables, with each node corresponding to a subset of erosion features that is increasingly homogeneous with respect to erosion type. Node purity is measured by the Gini Index, and each split is associated with an improvement value indicating the reduction in impurity. Colored nodes highlight the thematic group of each explanatory variable, distinguishing substrate-related variables, land-use factors, and terrain characteristics.

The model resulted in a tree with moderate depth (5) and 23 nodes - being 12 of them terminal nodes - ensuring both interpretability and generalization. The classification outcomes associated with each node are detailed in Table 3, which reports the relative frequencies of ravines and gullies within each partition. The algorithm calculates the improvement value (Gini Index) of each branching, which ranged from 0.003 to 0.061. These values measure how much the homogeneity related to the dependent variable in the child node was enhanced in comparison to the parent nodes.

The simultaneous analysis of both – Figure 5 and Table 3 - enables identifying: (i) by which values of the explanatory variables (independent) the resulting nodes are characterized, at different levels; and (ii) which

nodes presented the highest frequency percentages recorded for erosion types (ravine and gully). Thus, it is possible to establish the factors influencing the prevailing occurrence of a given erosion type in comparison to the others.

Figure 5 – Decision Tree Schema (CART Algorithm).



Source: Authors (2022).

At the root level (Node 0), the algorithm selected “Triggering Action” as the most informative variable, yielding the highest normalized importance (100%). This first bifurcation separates the dataset into two nearly balanced branches (47.9% and 52.1% of the observations), while already producing distinct class compositions: Node 1 presents a near equilibrium between ravines (48.4%) and gullies (51.6%), whereas Node 2 is strongly dominated by gullies (82.9%). This result confirms the central role of runoff concentration, stormwater disposal, and fluvial processes in controlling the spatial differentiation between ravines and gullies in urban environments.

Although this first level presents the highest Gini Index value (0.061), the variable responsible for data partitioning, “Triggering Action”, was identified by the algorithm as the most important variable for the classification, as pointed out in Table 4, which relates the independent variable according to its relevance for the classification process. This finding aligns with studies by Guo et al. (2019), which emphasize the importance of stormwater disposal in accelerating erosion processes in urbanized settings. However, unlike rural contexts, where agricultural practices dominate as erosion triggers, urban areas present unique challenges, such as concentrated runoff, inadequate stormwater infrastructure, and impervious surfaces. This highlights the necessity of urban-specific approaches to manage these processes effectively.

After “Triggering Action”, related to land use, variables “Geology” (related to substrate characteristics) and “Elevation” (related to geomorphometric variables) were the ones better explained data variability. These variables are accountable for data bipartition at the next level (2). Variable “Rainfall Erosivity” only appears in the 11th position.

By observing node 4, which is explained by variable “Geology”, it is possible to notice that the geological formations cluster in this node holds 87.1% ravine occurrences, indicating less appropriate formations to gullies’ development. The geological formations found in node 3 does not lead to a clear conclusion about erosion types, however, as the analyses proceed to nodes 7 and 8, it is possible to infer that plan curvatures >0.0036 (convergent) holds the highest ravine occurrences and ≤0.0036 (divergent or weakly

convergent forms) holds the highest gully occurrences.

Table 3 – List of Decision Tree nodes and divisions.

Node	Parent Node	Gully		Ravine		Explanatory variable	Resulting Cluster
		N	%	N	%		
0	(root)	449	32.1	949	67.9	Type of Erosion	
1	0	324	48.4	345	51.6	Triggering action	Surface runoff; Railroad/Highway stormwater Runoff/Disposal
2	0	125	17.1	604	82.9	Triggering action	Fluvial Erosion; Stormwater disposal (pipe); Surface runoff and stormwater disposal
3	1	263	43.9	336	56.1	Geology	Silvianópolis Complex – Embrechitic gneiss; Silvianópolis Complex – Granulites; Alluvial deposits; Colluvial deposits; Summit deposits; Caçapava Formation; Rio Claro Formation; Amparo Group – Paragneiss; Bauru Group – Adamantina Formation; Bauru Group – Marília Formation; Bauru Group – Santo Anastácio Formation; Paraná Group – Furnas Formation; São Bento Group – Botucatu Formation; São Bento Group – Pirambóia Formation; São Bento Group – Serra Geral Formation; São Roque Group – Phyllites; Tubarão Group – Aquidauana Formation; Tubarão Group – Tatuí Formation; Tubarão Group – Itararé Subgroup; Basic suites.
4	1	61	87.1	9	12.9	Geology	Coastal Complex – Migmatites; Embu Complex- Migmatites; Pinhal Complex- Migmatites; São Paulo Formation; Tremembé Formation; Açungui Group - Mica schists; Bauru Group – Caiuá Formation; Passa Dois Group – Corumbataí Formation; São Roque Group – Schists; Granitoid Suites.
5	2	56	29.9	131	70.1	Elevation	≤ 453.0
6	2	69	12.7	473	87.3	Elevation	> 453.0
7	3	156	38.8	246	61.2	Plan Curvature	≤ 0.0036
8	3	107	54.3	90	45.7	Plan Curvature	> 0.0036
9	5	22	20.2	87	79.8	Elevation	≤ 416.5
10	5	34	43.6	44	56.4	Elevation	> 416.5
11	6	7	4.4	152	95.6	Downstream Use	Urban in consolidation; Urban and Rural; Urban in consolidation and Rural; Rural and Vegetation
12	6	62	16.2	321	83.8	Downstream Use	Urban; Rural; Drainage; Vegetation
13	7	139	43.7	179	56.3	Elevation	≤ 660.0
14	7	17	20.2	67	79.8	Elevation	> 660.0
15	9	9	47.4	10	52.6	Land Cover	Farming
16	9	13	14.4	77	85.6	Land Cover	Urbanized Area; Forest; Non-Forest Natural Formation; Agriculture and Pasture Mosaic
17	10	1	7.1	13	92.9	Slope	≤ 2.76039
18	10	33	51.6	31	48.4	Slope	> 2.76039
19	13	43	35.0	80	65.0	Aspect	≤ 128.347438
20	13	96	49.2	99	50.8	Aspect	> 128.347438
21	14	14	51.9	13	48.1	Geology	Tubarão Group – Itararé Subgroup; Silvianópolis Complex – Granulites; São Bento Group – Pirambóia Formation; Alluvial deposits; Silvianópolis Complex – Embrechitic gneiss.
22	14	3	5.3	54	94.7	Geology	Colluvial deposits; Summit deposits; Caçapava Formation; Rio Claro Formation; Amparo Group – Paragneiss; Bauru Group – Adamantina Formation; Bauru Group – Marília Formation; Bauru Group – Santo Anastácio Formation; Paraná Group – Furnas Formation; São Bento Group – Botucatu Formation; São Bento Group – Serra Geral Formation; São Roque Group – Phyllites; Tubarão Group – Aquidauana Formation; Tubarão Group – Tatuí Formation; Basic suites.

Source: Authors (2022).

Table 4 – Importance relation of independent variables selected by the CART algorithm.

	Independent Variable	Importance	Normalized Importance (%)
1	Triggering action	0.062	100.0
2	Geology	0.049	78.9
3	Elevation	0.036	57.2
4	Geomorphology	0.034	53.7
5	Land cover	0.024	38.8
6	Slope	0.023	37.6
7	Downstream use	0.022	36.0
8	Plan curvature	0.013	20.2
9	Profile curvature	0.009	14.7
10	Erosivity	0.008	13.3
11	Aspect	0.007	10.5
12	Pedology	0.006	9.7
13	Upstream use	0.005	7.7

Note: Colors indicates variable type and are consistent with those used in Figure 5. Source: Authors (2026).

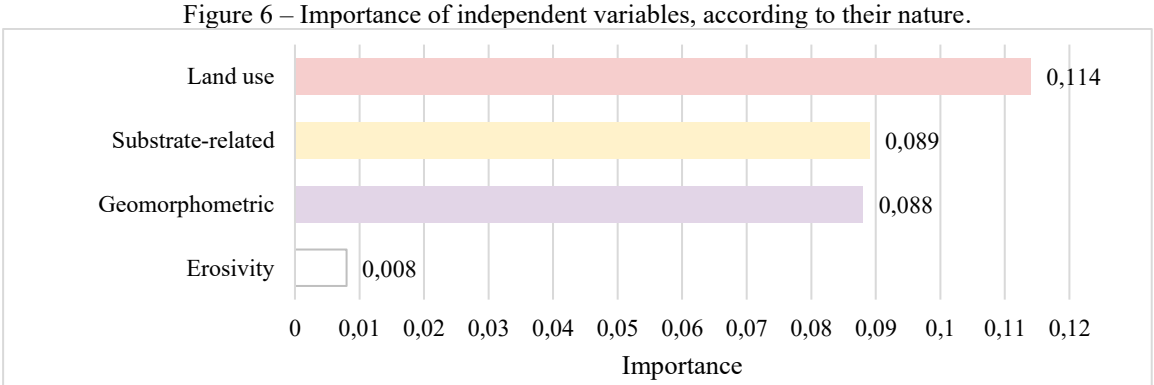
The identification of geologic formations and soil types less prone to gully formation offers valuable insights for land-use planning. These areas could serve as benchmarks for soil conservation practices, guiding interventions in more vulnerable regions. Moreover, understanding these stable conditions can inform urban development policies, ensuring that natural protective factors are preserved or even replicated in erosion-prone areas.

Although “Elevation” ranks among the most important predictors, its discriminatory capacity is limited. The elevation thresholds identified at multiple levels (453 m, 416.5 m, 660 m) generally lead to nodes still dominated by gullies, indicating that altitude alone is insufficient to clearly differentiate erosion types. This reinforces the interpretation that elevation primarily acts as a contextual variable rather than a direct driver of erosion typology.

In contrast, when elevation is examined in combination with slope, a more nuanced pattern emerges. For elevations above 416.5 m, areas characterized by slopes greater than 2.76% exhibit a slightly higher concentration of ravine features (51.6%), as observed in node 18. Although this difference is not pronounced, it suggests a modest tendency toward conditions less favorable to gully development under steeper topography at higher altitudes. Conversely, for elevations below 416.5 m associated with gentler slopes (≤ 2.76%), gully erosion becomes strongly dominant, accounting for 92.9% of the cases (node 17). This contrast indicates that, while elevation alone has limited discriminatory power, its interaction with slope plays a meaningful role in conditioning the susceptibility to the initiation and progression of gully processes.

Figure 6 summarizes the relative importance of variable classes by category (as an interpretative strategy to support overall interpretation and synthesis of results). Land-use-related variables were predominant in explaining erosion processes, highlighting human activities as the primary drivers of linear erosion in urban areas. After that, the main explicative variable is substrate-related, followed by geomorphometric variables and rainfall erosivity, as the less important factor.

Figure 6 – Importance of independent variables, according to their nature.



Note: Colors indicates variable type and are consistent with those used in Figure 5. Source: Authors (2026).

In general, the findings underline the critical role of stormwater management in mitigating urban erosion processes, as widely documented in the literature (Vanmaercke et al., 2021; Rahmati et al., 2022). This

pattern does not contradict the existing literature, but rather highlights a shift in the relative importance of controlling factors in urban settings. While rainfall, soil properties, and topography remain essential predisposing conditions for erosion development (Rahmati et al., 2022; Vanmaercke et al., 2021), the modification of hydrological processes by urban land use can be decisive in determining where and how linear erosion features evolve. Similar conclusions have been reported in studies showing that poorly planned urbanization amplifies erosion susceptibility by concentrating runoff and reducing infiltration, even under comparable climatic conditions (Frankl et al., 2020; Zolezzi et al., 2018; Mawe et al., 2025).

This perspective is particularly relevant under climate-change scenarios, in which increases in rainfall intensity are expected to exacerbate erosion hazards where urban drainage systems are inadequate (Kuhn et al., 2024; Chen et al., 2024). These results provide actionable insights for decision-makers, highlighting the need for sustainable practices that integrate urban hydrology with land-use planning.

4 CONCLUSIONS

Overall, the procedures applied to process data and the GIS tools incorporated to complement the target attributes investigation were effective in subsidizing the subsequent data mining technique. Based on the recorded results, the used method, decision tree supported by the CART algorithm, was efficient in identifying the variables better explaining the occurrence of each erosion type.

Substrate-related factors and terrain characteristics were shown to act as conditioning elements. The areas least prone to gully development were better explained by substrate-related factors, especially when combined with convergent plan curvature. Although elevation and slope alone showed limited discriminatory capacity, their interaction contributed to differentiating erosion behavior in specific contexts, reinforcing the importance of multivariate analysis.

The results indicate that anthropogenic factors dominate the differentiation between ravines and gullies in urban areas of São Paulo State. Among all explanatory variables, triggering action - closely linked to surface runoff (concentrated) and stormwater disposal (diffuse) - emerged as the most influential predictor, highlighting the central role of urban hydrological alteration in erosion processes. Land-use-related variables proved to be the most relevant for inciting the studied phenomenon since the combined explanatory importance values accounted for the highest percentage, followed by substrate-related factors, geomorphometric attributes and rainfall erosivity.

From a methodological perspective, the adoption of asymmetric misclassification costs proved essential for mitigating class imbalance and improving the identification of ravines without compromising gully detection. The subsequent pruning of the decision tree to a maximum depth of five levels resulted in a parsimonious and robust model, with balanced class performance, moderate overall accuracy (72.7%), good discriminatory ability (AUC = 0.78), and limited overfitting, as indicated by the small difference between resubstitution and cross-validation risks.

These outcomes underscore that effective mitigation of urban linear erosion depends less on isolated corrective measures and more on integrated strategies that combine land-use regulation, stormwater management, and planning informed by geomorphological constraints. Equally important is the continuous development and updating of erosion inventories such as the one used in this study, originally compiled in 2010. The maintenance of systematic, up-to-date databases would enable model validation against newly observed erosion features, strengthening its predictive reliability under real-world conditions. Moreover, advances in remote sensing, terrain modeling, and land-use mapping now allow the incorporation of higher-resolution geomorphometric and surface data, which can substantially refine susceptibility assessments. The proposed framework therefore offers a transferable and scalable methodological basis for erosion risk analysis in urban areas, while highlighting the need for sustained data acquisition efforts to support more accurate, dynamic, and resilient urban planning under ongoing urbanization and climate change scenarios.

5 BIBLIOGRAPHY

Almeida Filho, G. S., Gouveia, M., Júnior, J., & Canil, K. (2001). *Prevenção e controle da erosão urbana no*

Estado de São Paulo. In Anais do 21º Congresso Brasileiro de Engenharia Sanitária e Ambiental (pp. 1-12).

Anache, J. A. A., Wendland, E. C., Oliveira, P. T. S., Flanagan, D. C. & Nearing, M. A. (2017). Runoff and soil erosion plot-scale studies under natural rainfall: A meta-analysis of the Brazilian experience. *Catena*, 152, 29–39. https://doi.org/10.1016/j.catena.2017.01.003.

Arantes, L. T., Carvalho, A. C. P., Carvalho, A. P. P., Lorandi, R., Moschini, L. E. & Di Lollo, J. A. (2021). Surface runoff associated with climate change and land use and land cover in southeast region of Brazil. *Environmental Challenges*, 3, 100054. https://doi.org/10.1016/j.envc.2021.100054.

Blanco-Canqui, H. & Lal, R. (2010). *Principles of Soil Conservation and Management* (1st ed.). Springer Netherlands. https://doi.org/10.1007/978-1-4020-8709-7.

Breiman, L., Friedman, J., Olshen, R. & Stone, C. (1984). *Classification and Regression Trees*. Chapman & Hall.

Carvalho, A. P. P., Guerrero, J. V. R., Silva, E. V. E., Pinto, M. J. R., Vaz, R. M. G. F. C., Pereira, C. T., Lorandi, R., Di Lollo, J. A. & Moschini, L. E. (2019). Environmental Fragility to Erosion in an Anthropogenic Watershed in the Northeast of the State of São Paulo, Brazil. *Anuário Do Instituto de Geociências - UFRJ*, 42(3), 7–18. https://doi.org/10.11137/2019_3_07_18.

CATI. (2016). *Digital Elevation Model of São Paulo State*. Brazilian Farming Information Center. http://mapas.cati.sp.gov.br/mdecati.html.

Chalise, D., Kumar, L. & Kristiansen, P. (2019). Land Degradation by Soil Erosion in Nepal: A Review. *Soil Systems*, 3(1), 12. https://doi.org/10.3390/soilsystems3010012.

Chen, Y., Jiao, J., Yan, X., Li, J., Vanmaercke, M., & Wang, N. (2024). Response of gully morphology and density to the spatial and rainy-season monthly variation of rainfall at the regional scale of the Chinese Loess Plateau. *Catena*, 232, 107773. https://doi.org/10.1016/j.catena.2023.107773.

Costa, C. W., Piga, F. G., Moraes, M. C. P. de, Dorici, M., Sanguinetto, E. de C., Di Lollo, J. A., Moschini, L. E., Lorandi, R. & Oliveira, L. J. (2015). Environmental Fragility and water scarcity in catchment basins: Headwaters of Araras River – Araras, SP. *Revista Brasileira de Recursos Hídricos*, 20(4), 946–958. http://dx.doi.org/10.21168/rbrh.v20n4.p946-958.

Cunha, S. B. & Guerra, A. J. T. (2009). Environmental Degradation. In A. J. T. Guerra & S. B. Cunha (Eds.), *Geomorphology and Environment* (7th ed., pp. 337–379). Bertrand Brasil.

DAEE. (1982). *Geological map of the State of São Paulo* [Map; scale 1:250,000]. Joint project: Department of Water and Electricity and Institute of Geosciences and Exact Sciences - UNESP, Rio Claro.

DAEE. (2019). *Hydrological Database*. Department of Water and Electricity. http://www.hidrologia.daee.sp.gov.br/.

De Albuquerque, A. O., de Carvalho Júnior, O. A., Guimarães, R. F., Gomes, R. A. T., & Hermuche, P. M. (2020). Assessment of gully development using geomorphic change detection between pre- and post-urbanization scenarios. *Environmental Earth Sciences*, 79, 1–14. https://doi.org/10.1007/s12665-020-08958-9.

Di Lollo, J. A., Guerrero, J. V. R., Abe, A. C. P. & Lorandi, R. (2018). Land Change, Soil Degradation Processes, and Landscape Management at the Clarinho River Watershed, Brazil. In Shakoor, A. & Cato, K. (Eds.) *IAEG/AEG Annual Meeting Proceedings, San Francisco, California*, 2018 – Volume 2 (pp. 99–106). Springer, Cham. https://doi.org/10.1007/978-3-319-93127-2_15.

Di Lollo, J. A. & Sena, J. N. (2013). Establishing erosion susceptibility: analytical hierarchical process and traditional approaches. *Bulletin of Engineering Geology and the Environment*, 72(3–4), 589–600. https://doi.org/10.1007/s10064-013-0529-9.

Dorici, M., Costa, C. W., de Moraes, M. C. P., Piga, F. G., Lorandi, R., de Lollo, J. A. & Moschini, L. E. (2016). Accelerated erosion in a watershed in the southeastern region of Brazil. *Environmental Earth Sciences*, 75(19), 1301. https://doi.org/10.1007/s12665-016-6102-7.

Durães, M. F. & Mello, C. R. de. (2016). Spatial distribution of the potential and current soil erosion for the

Sapucaí River Basin, MG, Brazil. *Engenharia Sanitaria e Ambiental*, 21(4), 677–685. https://doi.org/10.1590/s1413-41522016121182.

Esri. (2021). *ArcGIS Pro* (2.8.3).

FAO. (2022). *World Reference Base for Soil Resources 2022*: International soil classification system for naming soils and creating legends for soil maps (4th ed.). FAO. https://doi.org/10.4060/cc3166en.

Frankl, A., Nyssen, J., Vanmaercke, M., & Poesen, J. (2020). Gully prevention and control: Techniques, failures and effectiveness. *Earth Surface Processes and Landforms*, 46, 220–238. https://doi.org/10.1002/esp.5033.

Galharte, C. A., Villela, J. M. & Crestana, S. (2014). Sediment yield estimation due to changes in land use and cover. *Revista Brasileira de Engenharia Agrícola e Ambiental*, 18(2), 199–201. https://doi.org/10.1590/S1415-43662014000200010.

Gudiño-Elizondo, N., Biggs, T. W., Castillo, C., Bingner, R. L., Langendoen, E. J., Taniguchi, K., Kretzschmar, T., Yuan, Y., & Liden, D. (2018). Measuring ephemeral gully erosion rates and topographical thresholds in an urban watershed using unmanned aerial systems and structure from motion photogrammetric techniques. *Land Degradation & Development*, 29, 1896–1905. https://doi.org/10.1002/ldr.2976.

Guo, Y., Peng, C., Zhu, Q., Wang, M., Wang, H., Peng, S. & He, H. (2019). Modelling the impacts of climate and land use changes on soil water erosion: Model applications, limitations and future challenges. *Journal of Environmental Management*, 250, 109403. https://doi.org/10.1016/J.JENVMAN.2019.109403.

IBGE. (2010). *Brazilian Demographic Census*. Brazilian Institute of Geography and Statistics. https://censo2010.ibge.gov.br/.

IBGE (2022). *São Paulo / Brasil: Panorama*. https://cidades.ibge.gov.br/brasil/sp/panorama.

IBM. (2011). *IBM SPSS Statistics* (20).

Imeson, A. C., & Kwaad, F. J. P. M. (1980). Gully types and gully prediction. *Geografisch Tijdschrift*, 14(5), 430-441.

Imwangana, F. M., Zola, B. J., Kangudia, J. M., & Sanghy, S. (2025). Gully erosion associated with peri-urbanization: Focus on the catchments around Kimwenza in the south of Kinshasa (DR Congo). *East African Journal of Environment and Natural Resources*, 8(2). https://doi.org/10.37284/eajenr.8.2.3160.

IPT. (2012). *Registration of Erosion and Flooding Points in the State of São Paulo* (Vol. 2, Issues 131.057-205). Technological Research Institute Technical Report.

Issaka, S. & Ashraf, M. A. (2017). Impact of soil erosion and degradation on water quality: a review. *Geology, Ecology, and Landscapes*, 1(1), 1–11. https://doi.org/10.1080/24749508.2017.1301053.

Javidan, N., Kavian, A., Pourghasemi, H. R., Conoscenti, C. & Jafarian, Z. (2019). Gully Erosion Susceptibility Mapping Using Multivariate Adaptive Regression Splines—Replications and Sample Size Scenarios. *Water*, 11(11), 2319. https://doi.org/10.3390/w11112319.

Júnior, O. A. C., Guimarães, R. F., Freitas, L. F., Gomes-Loebmann, D., Gomes, R. A. T., Martins, E. S., & Montgomery, D. R. (2010). Urbanization impacts upon catchment hydrology and gully development using multi-temporal digital elevation data analysis. *Earth Surface Processes and Landforms*, 35, 611–629. https://doi.org/10.1002/esp.1917.

Kuhn, C., Reis, F., Furegatti, S., Zarfl, C., & Peixoto, A. (2024). Economic impacts of an urban gully are driven by land degradation. *Natural Hazards*, 120, 13995–14026. https://doi.org/10.1007/s11069-024-06727-6.

Mahamba, J., Kayitoghera, G., Musubao, M., Chuma, G., & Sahani, W. (2023). Evolution of gully erosion and susceptibility factors in the urban watershed of the Kimemi (Butembo/DR Congo). *Geography and Sustainability*, 4(3), 268-279. https://doi.org/10.1016/j.geosus.2023.07.001.

MapBiomas Project. (2010). *Collection 8.0 of the Annual Series of Land Cover and Land Use Maps of Brazil*. https://brasil.mapbiomas.org/.

Mawe, G., Landu, E., Imwangana, F., Hubert, A., Dille, A., Bielders, C., Poesen, J., Dewitte, O., & Vanmaercke, M. (2024). What controls the expansion of urban gullies in tropical environments? Lessons learned from contrasting cities in DR Congo. *Catena*, 241, 108055. https://doi.org/10.1016/j.catena.2024.108055.

Mawe, G., Landu, E., Dujardin, E., Imwangana, F., Bielders, C., Hubert, A., Michellier, C., Nzolang, C., Poesen, J., Dewitte, O., & Vanmaercke, M. (2025). Mapping urban gullies in the Democratic Republic of the Congo. *Nature*, 644, 952–959. https://doi.org/10.1038/s41586-025-09371-7.

Medeiros, A. R. C., Araújo, Y. B. de, Vianna, R. P. de T. & Moraes, R. M. de. (2014). Decision support model applied to the recognition of non-adherent individuals to antihypertensive therapy. *Saúde em Debate*, 38(100). https://doi.org/10.5935/0103-104.20140016.

Mota, S. (1995). *Preservation and Conservation of Water Resources* (2nd ed.). ABES.

Oliveira, A. M. S. (1994). *Technogenic deposits and reservoir silting: The Capivara Reservoir, Paranapanema River* [Doctoral dissertation]. University of São Paulo, São Paulo.

Oliveira, J. B., Camargo, M. N., Rossi, M. & Calderano Filho, B. (1999). *Pedological map of the State of São Paulo* [Map; scale 1:500,000]. Instituto Agronômico de Campinas (IAC); Embrapa Solos.

Pierce, F. J. & Lal, R. (2017). *Monitoring the Impact of Soil Erosion on Crop Productivity*. In Soil Erosion Research Methods (pp. 235–263). Routledge. https://doi.org/10.1201/9780203739358-10.

Pimentel, D. (2006). Soil Erosion: A Food and Environmental Threat. *Environment, Development and Sustainability*, 8(1), 119–137. https://doi.org/10.1007/s10668-005-1262-8.

Pons, N. A. D., Pejon, O. J. & Zuquette, L. V. (2007). Use of geoprocessing in the study of land degradation in urban environments: the case of the city of São Carlos, state of São Paulo, Brazil. *Environmental Geology*, 53(4), 727–739. https://doi.org/10.1007/s00254-007-0685-y.

Powers, D. M. W. (2011). Evaluation: From Precision, Recall and F-Measure to ROC, Informedness, Markedness and Correlation. *International Journal of Machine Learning Technology*, 2(1), 37–63. https://doi.org/10.48550/arXiv.2010.16061.

Prieto-Amparán, Pinedo-Alvarez, Vázquez-Quintero, Valles-Aragón, Rascón-Ramos, Martinez-Salvador & Villarreal-Guerrero. (2019). A Multivariate Geomorphometric Approach to Prioritize Erosion-Prone Watersheds. *Sustainability*, 11(18), 5140. https://doi.org/10.3390/su11185140.

Quinlan, J. R. (1983). *Learning Efficient Classification Procedures and Their Application to Chess End Games*. In R. S. Michalski, J. G. Carbonell & T. M. Mitchell (Eds.), Machine Learning (pp. 463–482). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-662-12405-5_15.

Rahmati, O., Kalantari, Z., Ferreira, C. S. S., Chen, W., Soleimanpour, S., Kapović-Solomun, M., Seifollahi-Aghmiuni, S., Ghajarnia, N., & Kazemabady, N. (2022). Contribution of physical and anthropogenic factors to gully erosion initiation. *Catena*, 2010, 105925. https://doi.org/10.1016/j.catena.2021.105925.

Ricardi, A. de M. & Lima, C. G. da R. (2022). Spatial and temporal variability of rain erosivity (EI30) in the State of São Paulo, Brazil. *Geosciences*, 40(4), 965-985. https://doi.org/10.5016/geociencias.v40i04.15492.

Rodrigo-Comino, J., Keesstra, S. & Cerdà, A. (2018). Soil Erosion as an Environmental Concern in Vineyards. The Case Study of Celler del Roure, Eastern Spain, by Means of Rainfall Simulation Experiments. *Beverages*, 4(2), 31. https://doi.org/10.3390/beverages4020031.

Ross, J. L. S., & Moroz, I. C. (1997). *Geomorphological map of the State of São Paulo* [Map; scale 1:500,000]. Laboratório de Geomorfologia, Departamento de Geografia, FFLCH/USP; Laboratório de Cartografia Geotécnica, Geologia Aplicada, IPT; Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP).

Salomão, F. X. T. (1994). *Linear erosion processes in Bauru (SP): Cartographic regionalization applied to preventive urban and rural control* [Doctoral dissertation]. University of São Paulo, São Paulo.

Sparovek, G., Correchel, V. & Barretto, A. G. O. P. (2007). The risk of erosion in Brazilian cultivated pastures. *Scientia Agricola*, 64(1), 77–82. https://doi.org/10.1590/S0103-90162007000100012.

Telles, T. S., Guimarães, M. de F. & Dechen, S. C. F. (2011). The costs of soil erosion. *Revista Brasileira de*

Ciência do Solo, 35(2), 287–298. https://doi.org/10.1590/S0100-06832011000200001.

Tricart, J. (1977). *Ecodinâmica*. Secretaria de Planejamento da Presidência da República, Fundação Instituto Brasileiro de Geografia e Estatística, Diretoria Técnica, Superintendência de Recursos Naturais e Meio Ambient. https://books.google.com.br/books?id=41BgAAAAMAAJ.

van Zelm, R., van der Velde, M., Balkovic, J., Čengić, M., Elshout, P. M. F., Koellner, T., Núñez, M., Obersteiner, M., Schmid, E. & Huijbregts, M. A. J. (2018). Spatially explicit life cycle impact assessment for soil erosion from global crop production. *Ecosystem Services*, 30, 220–227. https://doi.org/10.1016/j.ecoser.2017.08.015.

Vanmaercke, M., Panagos, P., Vanwalleghem, T., Hayas, A., Foerster, S., Borrelli, P., … Poesen, J. (2021). Measuring, modelling and managing gully erosion at large scales: A state of the art. *Earth-Science Reviews*, 218, 103637. https://doi.org/10.1016/j.earscirev.2021.103637.

Zdruli, P., Karydas, C. G., Dedaj, K., Salillari, I., Cela, F., Lushaj, S. & Panagos, P. (2016). High resolution spatiotemporal analysis of erosion risk per land cover category in Korçe region, Albania. *Earth Science Informatics*, 9(4), 481–495. https://doi.org/10.1007/s12145-016-0269-z.

Zolezzi, G., Bezzi, M., Spada, D., & Bozzarelli, E. (2018). Urban gully erosion in sub-Saharan Africa: A case study from Uganda. *Land Degradation & Development*, 29, 849–859. https://doi.org/10.1002/ldr.2865.

Zuquette, L. V., Rodrigues, V. G. S. & Pejon, O. J. (2013). *Recovery of Degraded Areas*. In Environmental Engineering: concepts, technology and management. Elsevier. https://repositorio.usp.br/item/002472425.

Funding

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Process 88887.658530/2021-00. The research was also supported by National Council for Scientific and Technological Development (CNPq) - Process #311393/2021-7:

Author Contributions

T. F. Olivatto: Conceptualization, Methodology, Data curation, Formal analysis, Investigation, Visualization Writing- Original draft. J. A. Di Lollo and D. B. Menezes: Supervision, Validation, Writing-Reviewing and Editing.

Conflict of Interest

No conflict of interest to declare.

First Author's Biography

PhD and master degree in Urban Engineering at UFSCar, with a mobility period at the Universidad Nacional del Litoral and a sandwich program at the Universidad Nacional de Mar del Plata, both in Argentina. Graduated in Environmental and Sanitary Engineering from PUC Campinas, including a sandwich program at the University of South Australia. Specialized in Data Science through CeMEAI and ICMC/USP, with technical training in Geomatics from Colégio Técnico de Limeira (UNICAMP), and a specialization in University Teaching. Dedicated to research that integrates geospatial information and remote sensing with artificial intelligence, particularly machine learning.