

MODELO DE CLASSIFICAÇÃO DE PROCESSOS EROSIVOS LINEARES AO LONGO DE FERROVIAS ATRAVÉS DE ALGORITMO DE ÁRVORE DE DECISÃO E GEOTECNOLOGIAS

Classification model of gully erosion process along of railway line through Decision Tree algorithm and geotechnology

Natália da Costa Souza ¹

Cira Pitombo ²

André Luiz Cunha ²

Ana Paula Camargo Larocca ²

Gerson Salviano de Almeida Filho ³

¹ Universidade de São Paulo – Escola Politécnica – Departamento de Engenharia de Transportes, Programa de Pós-Graduação em Engenharia de Transportes, Av. Prof. Luciano Gualberto, Travessa 03, nº 380, CEP 05508 - 010, São Paulo – SP – Brasil. Contato: nataliacostaptr@usp.br

² Universidade de São Paulo – Escola de Engenharia de São Carlos – Departamento de Engenharia de Transportes, Programa de Pós-Graduação em Engenharia de Transportes, Av. Trabalhador São-carlense, nº 380, CEP 13566-590, São Carlos – SP – Brasil. Contato: cira@sc.usp.br, acunha@sc.usp.br, larocca.ana@usp.br

³ Instituto de Pesquisas Tecnológicas – IPT – CTGEO – Laboratório de Riscos Ambientais - Av. Professor Almeida Prado, nº 532, CEP 05508-901, São Paulo – SP – Brasil. Contato: gersaf@ipt.br

Resumo:

A análise da dinâmica de desenvolvimento de processos erosivos no entorno das ferrovias são requisitos da legislação ambiental vigente para o setor ferroviário. Em abordagens capazes de subsidiar processos de análise, monitoramento e controle de riscos ambientais, algoritmos de Árvore de Decisão (AD), associados à utilização de sistemas de informações geográficas (SIG) têm sido ferramentas amplamente utilizadas. Tais estudos geram procedimentos de predição de cenários, medidas de probabilidade ou classificações dos terrenos quanto ao grau de suscetibilidade ou vulnerabilidade das áreas para ocorrência dos mesmos. No entanto, pressupõe-se que agentes desencadeadores de processos erosivos podem apresentar características específicas que contribuem para a formação de processos erosivos de diferentes tipologias, padrões e gravidade. Estudos classificatórios baseados nos agentes desencadeadores e nas características dos processos ainda são raros na literatura. O objetivo principal desta pesquisa é a combinação do uso de geotecnologias e AD para classificação dos processos erosivos lineares localizados na área de entorno da linha férrea Malha Paulista (SP). A análise procurou compreender como se comportam e agrupam os agentes desencadeadores de erosão em função do tipo de processo erosivo desenvolvido. Os resultados permitiram classificar a amostra de registros mapeados e identificar padrões que distinguem as tipologias de erosões lineares.

Palavras-chave: Erosão, Linha Férrea, Sistemas de Informação Geográfica (SIG), Árvore de Decisão, Classificação.

Abstract:

The dynamic analysis of erosion process development in the surroundings of the railways are requirements of environmental legislation for the railway sector. In approaches able to support process analysis, monitoring and control of environmental risks, Decision Tree algorithm (DT) associated with the use of geographic information systems (GIS) have been widely used tools. Such studies generate prediction scenarios procedures, probability measures or classifications of land in relation of degree of susceptibility or vulnerability of areas for their occurrence. However, it is assumed that triggering agents of erosion process may present specific features which contribute to the formation of erosive process of types different, patterns and severity. Classificatory studies based on triggering agents and the characteristics of the processes are still rare in the literature. The main objective of this research is the combination of the use of geotechnology and DT for classification of linear erosion located in the surrounding area of the railway Malha Paulista (SP). The analysis sought to understand how they behave and group the triggering agents of erosion in relation of the type of erosive process developed. The results permit classifies the mapped records sample and identify patterns that distinguish the types of linear erosions.

Keywords: Erosion, Railway, Geographic Information System (GIS), Decision Tree, Classification.

1. Introdução

A ocorrência de eventos erosivos acelerados associados à implantação de sistemas de transporte é um dos principais problemas socioambientais enfrentados pelos responsáveis pela operação das linhas férreas e rodoviárias do país. Feições erosivas do tipo linear e laminar são observadas tanto ao longo da plataforma, nos cortes e aterros, como fora dela, em caixas de empréstimo, áreas de jazidas exploradas, junto aos pés de aterros e jusante das obras de transposição. A imprecisão dos projetos de drenagem, principalmente por não levarem em conta a natureza dos terrenos quanto à suscetibilidade à erosão, o desinteresse pelos investimentos em obras complementares e a falta de manutenção das vias são as principais causas da grande incidência desses processos, que podem não só atingir as linhas como também induzir a ocorrência de novos processos, oferecendo risco às pessoas e à região localizada no entorno (Salomão, 1999). Particularmente em linhas férreas, a obstrução das vias causada pela dinâmica de desenvolvimento de processos erosivos e outros eventos de risco associados, envolve os responsáveis pela operação não só do ponto de vista econômico (paralisação das vias e impedimento do transporte de cargas e passageiros), mas principalmente pelo viés socioambiental (acidentes envolvendo passageiros e maquinistas, derramamento de cargas inflamáveis, contaminação do solo e lençóis freáticos, assoreamento de corpos d'água, exposição da fauna local, etc.).

De acordo com a Política Ambiental do Ministério dos Transportes (artigo 225, § 1º, inciso V), as atividades ferroviárias (construção, manutenção, duplicação, transporte) enquadram-se na definição de atividades potencialmente causadoras de significativa degradação do meio ambiente, sendo amplamente recepcionadas pelos instrumentos constitucionais. No Brasil, para processos de licenciamento ambiental e regularização dos empreendimentos ferroviários em operação, o Instituto Brasileiro do Meio Ambiente e dos Recursos Renováveis (IBAMA) prevê a implantação de planos e programas ambientais que alcancem o cumprimento das condicionantes necessárias à regularização e funcionamento adequado das atividades em operação no país. Tais

programas visam à minimização de danos provenientes da operação da malha, resultando no gerenciamento eficaz dos riscos de suas atividades, possibilitando a mitigação em caso de eventuais impactos aos meios físicos, bióticos e socioeconômicos. Dessa forma, um dos aspectos abordados dentre os programas exigidos é a análise das ocorrências de processos erosivos ao longo da faixa de domínio da malha, o diagnóstico da gravidade dos processos, sua espacialização e padrões de distribuição, ações de prevenção, recuperação das áreas afetadas e riscos oferecidos.

Em geral, na literatura, os estudos relativos às análises de dinâmica de desenvolvimento e à ocorrência de processos erosivos estão inseridos em abordagens ambientais teórico-metodológicas consagradas nas áreas de Ciências da Terra e Engenharias. Em diversas áreas do conhecimento e nas mais variadas abordagens capazes de subsidiar processos de análise, monitoramento e controle de riscos ambientais, técnicas estatísticas de análises multivariadas de dados associadas à utilização de sistemas de informações geográficas (SIG) têm sido ferramentas amplamente utilizadas. Os métodos de análises multicritérios são procedimentos metodológicos de cruzamento de informações adaptados para compor análises espaciais com SIG. No que tange ao desenvolvimento de metodologias e modelagens de análise ambiental e eventos de riscos associados à ocorrência e à dinâmica de processos erosivos, vários autores discutem a utilização de tais técnicas e dão ênfase ao caráter diverso da análise multivariada, sua capacidade analítica e preditiva, o potencial de manuseio de dados espaciais em SIG e a eficácia de análise e rapidez de processamento (Oh e Pradhan, 2011; Pradhan e Lee, 2010; Pourghasemi *et al.*, 2012; Devkota *et al.*, 2013; Bui *et al.*, 2014; Regmi *et al.*, 2014).

Estudos encontrados na literatura que utilizam técnicas de Árvore de Decisão, ou outras técnicas de mineração de dados, usualmente geram ferramentas de predição de cenários, medidas de probabilidade ou classificações qualitativas dos terrenos (áreas de estudo) quanto ao grau de suscetibilidade ou vulnerabilidade das áreas para ocorrência dos mesmos (Pradhan e Lee, 2010; Oh e Pradhan, 2011; Pourghasemi *et al.*, 2012; Devkota *et al.*, 2013; Bui *et al.*, 2014; Saito *et al.*, 2009; Bui *et al.*, 2012). No entanto, pressupõe-se que agentes desencadeadores de processos erosivos podem (ou não) apresentar características específicas que contribuem para a formação de processos erosivos de diferentes tipologias, padrões e gravidade. Estudos classificatórios baseados nos agentes desencadeadores e nas características dos processos ainda são raros na literatura.

O presente trabalho propõe a combinação do uso de geotecnologias (SIG e técnicas de geoprocessamento) e algoritmo de Árvore de Decisão para interpretação dos processos erosivos lineares (ravinas e boçorocas) localizados na área de entorno da linha férrea Malha Paulista (São Paulo, Brasil). Pressupõe-se que agentes desencadeadores de processos erosivos podem apresentar características específicas que contribuem para a formação de processos erosivos de diferentes tipologias, padrões e gravidade. Especificamente, a análise visa compreender como se comportam e como se agrupam as variáveis independentes (agentes desencadeadores de erosão) em função do tipo de processo erosivo desenvolvido (ravinas e boçorocas) – variável dependente.

A contribuição essencial do trabalho baseia-se no fato da pouca bibliografia relativa a estudos classificatórios dos processos erosivos baseados em fatores desencadeadores dos processos e de suas características. Além do uso de técnica de mineração de dados (AD), sem restrições prévias matemáticas e de fácil interpretação. Desta forma, há dois objetivos atuantes: (1) propor modelo de classificação dos processos segundo os agentes desencadeadores e; (2) Apresentar técnica não usual e de fácil interpretação para este propósito.

2. Área de Trabalho

A área de estudo, ilustrada pela Figura 1, pertence à região do traçado da linha ferroviária Malha Paulista, operada pela América Latina Logística – ALL. A linha, em operação, liga a Baixada Santista/Porto de Santos às cidades de Santa Fé do Sul, Panorâma e Colômbia (SP) e totaliza um percurso de 1.989 km que possibilita o escoamento da produção agropecuária e industrial dos estados de Mato Grosso, Mato Grosso do Sul e São Paulo. A região de interesse é restrita às análises e registros de processos erosivos diagnosticados no trecho que interliga as cidades de Mairinque (SP) a Botucatu (SP), totalizando um trajeto de aproximadamente 176 km.

Judicialmente, a faixa situada à margem da ferrovia, chamada faixa de domínio, é uma área de terreno com pequena largura em relação à extensão da linha e varia de acordo com cada trecho em todo o território nacional. Na impossibilidade de se trabalhar com uma faixa comum a todo trecho selecionado e na intenção de considerar principalmente os processos erosivos com desenvolvimento próximos às áreas que circundam a linha férrea, optou-se por analisar os processos erosivos pertencentes a um *buffer* (área de influência) de 5 km de abrangência ao longo do traçado da malha. A escolha da área de interesse se dá por conta da incidência de registros verificados no trecho, a importância econômica da linha e a atenção dada pela mídia local aos impactos negativos gerados pela ocorrência de fenômenos erosivos na região (principalmente dos municípios pertencentes à região de Bauru).

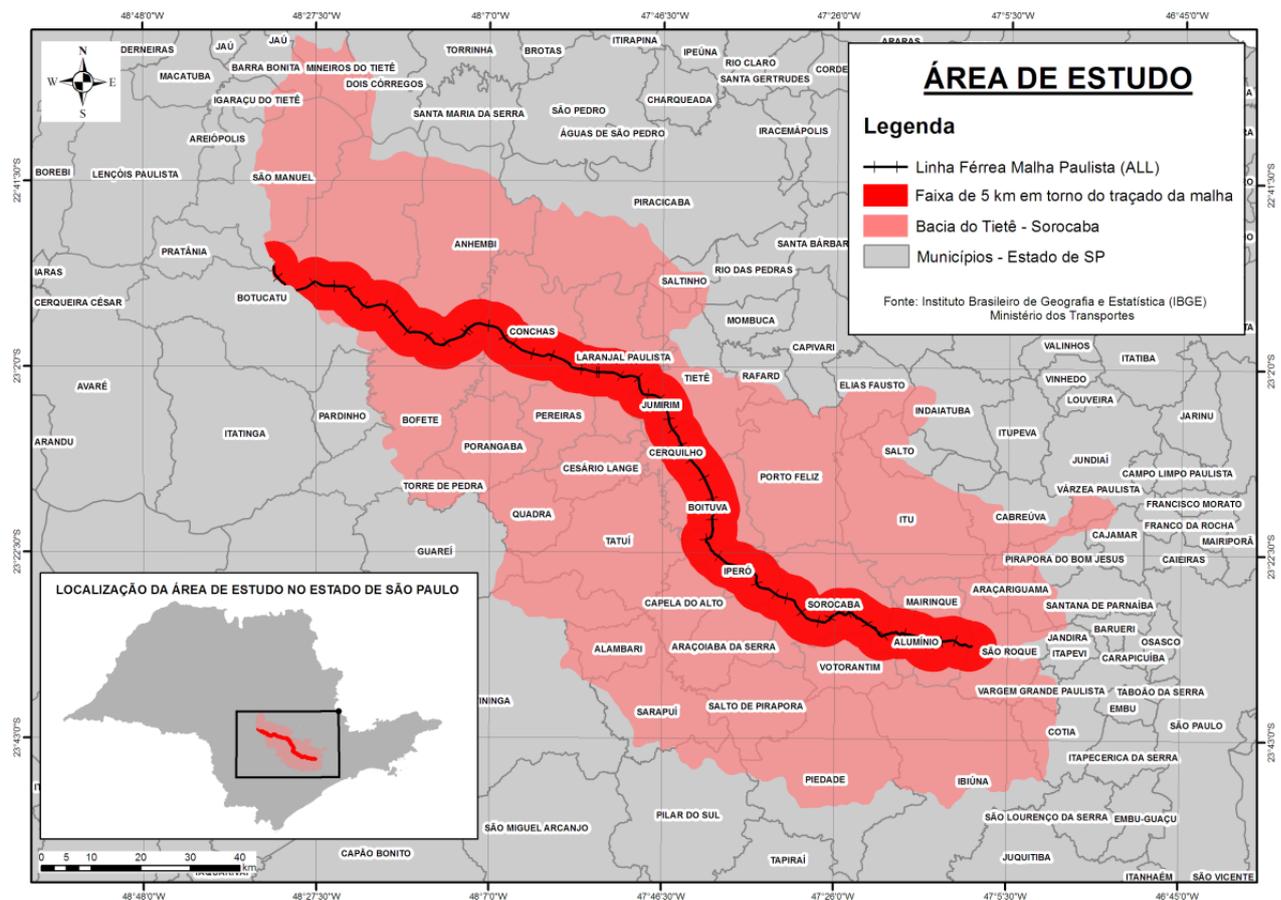


Figura 1: Mapa da região de estudo: Malha Paulista (recorte) – ALL.

3. Materiais e Método

O método para desenvolvimento da pesquisa baseou-se em três etapas distintas. A Figura 2 resume o procedimento realizado, os materiais utilizados e os produtos resultantes. As etapas propostas foram baseadas em Souza *et al.* (2014), que também utilizam ferramentas de geoprocessamento para caracterização dos processos erosivos contidos na base de dados utilizada e análises subsidiadas por técnicas estatísticas.

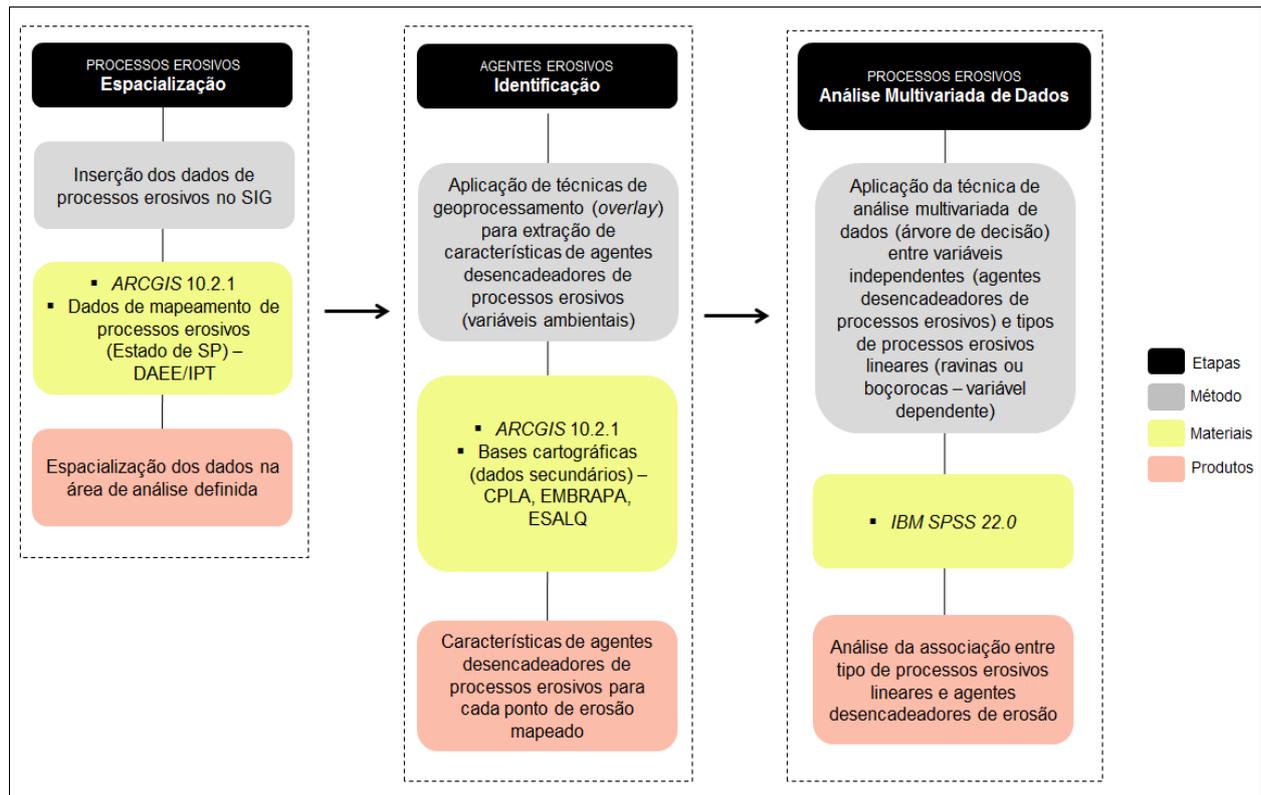


Figura 2: Fluxograma de procedimento metodológico.

3.1 Processos Erosivos – Espacialização e Definição

A base de dados de registro de processos erosivos da área de estudo foi fornecida pelo Instituto de Pesquisas Tecnológicas – IPT e pelo Departamento de Águas e Energia Elétrica do Estado de São Paulo – DAEE, fonte de um projeto cujo objetivo principal é fornecer elementos básicos para o planejamento de programas e ações voltadas ao equacionamento dos problemas causados pelas erosões lineares, urbanas e rurais, e pelas inundações/enchentes nas áreas urbanas de todo o território do estado de São Paulo para o ano de 2013. Para a área de trabalho selecionada, foram detectados 773 processos, sendo 301 ravinas e 472 boçorocas. Os registros foram coletados no sistema de projeção geográfica associados ao sistema geodésico WGS – 1984 e convertidos para SIRGAS 2000. A espacialização dos pontos utilizados foi feita através do *software ARCGIS 10.2.1* pela geração de um arquivo *shapefile* no formato de pontos.

Com relação à definição dos conceitos de erosão utilizados para classificar as erosões lineares (ravinas e boçorocas) da base dados, o IPT estabeleceu que a ravina é o aprofundamento das feições de sulco provocado pela ação erosiva da água de escoamento superficial concentrado, com profundidade mínima em torno de 50 cm e que não atingem o nível d'água subterrânea, onde atuam mecanismos de desprendimento de material dos taludes laterais e transporte de partículas do solo. Já as boçorocas são definidas como sendo as formas de erosão mais complexas e mais destrutiva no quadro evolutivo da erosão linear. Nas mais comuns, a profundidade vai de 15 a 30 metros, em comprimento podem alcançar centenas de metros, com perfil transversal em forma de “V”. A Figura 3 apresenta exemplos dos conceitos citados.



Figura 3: À esquerda: Processo de ravinamento em área rural, município de Itaquaquetuba (SP). À direita: Boçoroca de evolução acelerada, comum no planalto ocidental, município de Pardinho (SP). Fonte: Instituto de Pesquisas Tecnológicas – IPT

3.2 Agentes Desencadeadores de Erosão - Identificação

Esta etapa consiste na identificação de propriedades ambientais referentes a agentes desencadeadores de processos erosivos. Os agentes desencadeadores são aqueles que determinam as variações nas taxas de erosão. Por conta da interação desses fatores, certas áreas erodem mais do que as outras. De um modo geral, as pesquisas sobre erosão do solo consideram como sendo fatores dos processos erosivos a erosividade da chuva, as propriedades dos solos, cobertura vegetal e características das encostas (Morgan, 1986; Guerra, 1999). A intervenção do homem pode alterar esses fatores e, conseqüentemente, apressar ou retardar o aparecimento de processos erosivos.

Baseado em metodologias já difundidas sobre estudos de erosão (Crepani *et al.*, 2001; Ross, 1994 e 1995, Salomão, 1999), foram identificadas características ambientais da localização dos pontos mapeados. Para cada ponto mapeado foram extraídas, de bases de dados espaciais existentes e disponíveis, características ambientais de geologia, declividade, pedologia, criticidade da área quanto ao desenvolvimento do processo erosivo, uso do solo e características pluviométricas. Tais características formaram, posteriormente, o banco de dados e foram analisadas como variáveis independentes, enquanto o tipo de processo erosivo formado (ravinas ou boçorocas) caracterizava uma variável dependente binária.

No *software ARCGIS 10.2.1* foi feita uma sobreposição das informações selecionadas (geologia, declividade, pedologia, criticidade da área quanto ao desenvolvimento do processo erosivo, uso do solo e características pluviométricas) com o registro de erosão relacionado a cada ponto mapeado. Além das informações que foram consideradas para gerar o modelo de classificação das erosões (campo atributos), os dados referentes às bases cartográficas utilizadas (fonte, escala, etc.) para identificação das características ambientais consideradas estão detalhados na Tabela 1.

3.3 Processos Erosivos – Árvore de Decisão (AD)

Hair Jr. *et al.* (1998) conceituam Análise Multivariada de dados como um conjunto de técnicas estatísticas utilizadas com o objetivo de explicar e prever o grau de relações entre diversas variáveis independentes (inclusive entre si) e a variável dependente. Em linhas gerais, as técnicas de análise multivariada de dados podem ser definidas como ferramentas analíticas que auxiliam na investigação acerca de fenômenos complexos, envolvendo múltiplas dimensões, capazes de identificar padrões presentes nas diversas interações entre as variáveis analisadas. Técnicas de mineração de dados são classificadas, pelos mesmos autores, como técnicas de Análise Multivariada Emergentes.

A técnica de mineração de dados, Árvore de Decisão, é considerada uma forma simples de representação de relações existentes em um conjunto de dados. Os dados são divididos em subgrupos, com base nos valores das variáveis. O resultado é uma hierarquia de declarações do tipo "Se ... então ..." que são utilizadas, principalmente, para classificação (Breiman *et al.*, 1984). Uma árvore de decisão pode ser definida como um gráfico acíclico e direto que satisfaz às seguintes propriedades: (1) A hierarquia é denominada árvore e cada segmento é denominado nó; (2) Há um nó, chamado raiz, que contém todo o banco de dados; (3) Este nó contém dados que podem ser subdivididos dentro de outros sub-nós, chamados de nós filhos; (4) Existe um único caminho entre o nó raiz e cada nó; (5) Quando os dados do nó não podem ser mais subdivididos dentro de um outro subconjunto ele é considerado um nó terminal ou folha (Pitombo *et al.*, 2011).

O algoritmo usado para dividir os dados nos modelos de árvore visa identificar as variáveis independentes que fornecem máxima segregação dos dados segundo a variável dependente. Alguns dos algoritmos para criação de árvores de decisão são o C4.5 (Quilan, 1993), CHAID (Kass, 1980) e CART (Breiman *et al.*, 1984) e QUEST.

Neste trabalho, o software IBM SPSS 22.0 foi utilizado para o processamento da árvore. Utilizou-se o algoritmo *CART (Classification and Regression Tree)*, que estabelece relações binárias entre variáveis independentes e variável dependente. O critério de partição de dados do CART é o índice Gini, que mede o grau de heterogeneidade dos dados. Logo, pode ser utilizado para medir a impureza de um nó (Breiman *et al.*, 1984). Quando este índice é igual a zero, o nó é puro. Por outro lado, quando ele se aproxima do valor um, o nó é impuro, aumentando o número de classes uniformemente distribuídas neste nó.

Quando, nas árvores de classificação com partições binárias, se utiliza o critério de Gini tende-se a isolar num ramo os registros que representam a classe mais frequente. Portanto, neste estudo de caso, o objetivo é obter nós (classes), com características dadas pelas variáveis independentes, e maiores frequências de ocorrência de Ravinas ou Boçorocas.

A AD aqui foi gerada a partir da amostra final com 733 dados de registro de processos erosivos da variável dependente (tipos de processos erosivos: ravinas ou boçorocas), e adotando-se o mínimo de 20 observações por nós filhos. As variáveis independentes foram aquelas referentes

às características ambientais do local em que foi registrado o ponto de erosão, classificadas anteriormente como agentes desencadeadores de processos erosivos. Estabeleceu-se também o percentual de 70% da amostra total para treino e 30% para teste.

A aplicação da técnica de AD indicou como todas as variáveis independentes, relacionadas aos fatores desencadeadores do processo, se comportaram em cada classe de tipo de processo erosivo analisada (variável dependente) e se houve algum padrão que diferencie os dois tipos de classificação de erosões lineares mapeadas (ravinas e boçorocas).

A Tabela 1 fornece um resumo dos dados analisados (variáveis), fonte, escala de mapeamento (para dados espaciais vetoriais), descrição, atributos e tipo (variável dependente ou independente).

Levando-se em conta os dados espaciais selecionados, não foi estabelecida uma única escala cartográfica para todas as informações, uma vez que foram utilizadas bases disponíveis online e de forma gratuita, sempre com prioridade de escolha da escala de maior detalhe. Como dito anteriormente, quando sobrepostas à base cartográfica de processos erosivos, todas as bases espaciais de atributos ambientais cederam informações de características relativas aos pontos de erosões mapeados.

Tabela 1: Dados: variáveis analisadas – tipo, descrição, atributos e fonte

VARIÁVEL	TIPO	DESCRIÇÃO	ATRIBUTOS	FONTE	ESCALA
Processo erosivo	Variável dependente	Tipo de processo erosivo linear mapeado	Ravinas; Boçorocas	IPT/DAEE	não informado
Uso do solo	Variável independente	Tipologias de uso e ocupação do solo mapeadas ao longo do trecho de análise	Cobertura arbórea; Cobertura herbácea arbustiva; Solo exposto e Área construída	CPLA - Secretaria do Meio Ambiente do Estado de São Paulo	1:100.000
Geologia	Variável independente	Tipo de compartimento geológico presente no local em que se desenvolve o processo erosivo. Sua importância está na análise do tipo da rocha originária do terreno e o grau de suscetibilidade ao desenvolvimento de processos erosivos	Clástica; Metamorfismo regional; Plutônica e Vulcânica	CPRM - GEOBANK (mapa geológico estadual)	1:750.000
Declividade	Variável independente	Inclinação do terreno em relação ao horizonte; tem relação direta com a velocidade de transformação da energia potencial em energia cinética e, portanto, com a velocidade das massas d'água em movimento responsáveis pelo escoamento superficial	Plano (0 - 3%); Suave ondulado (3% - 8%); Ondulado (8% - 20%); Forte ondulado (20% - 45%); Montanhoso (45% - 75% e Escarpado (> 75%)	EMBRAPA	1:250.000
Pedologia	Variável independente	Tipo de solo presente no local em que se desenvolve o processo erosivo. O solo e suas propriedades são uns dos principais agentes desencadeadores do processo erosivo	Argissolos; Latossolos; Neossolos; Nitossolo e Neossolo	ESALQ - USP	1:500.000
Pluviometria	Variável independente	Representa a quantidade de energia potencial disponível para ser transformada em energia cinética e geração de processos erosivos. Os dados informam isoietas de precipitações médias anuais da região de interesse entre os anos de 1977 a 2006	1.300mm; 1.400mm e 1.500mm	CPRM - GEOBANK (Projeto atlas pluviométrico do Brasil)	1:5.000.000
Criticidade	Variável independente	Classificação do Instituto de Pesquisas Tecnológicas (IPT), realizada em 2013, quanto à localização do processo erosivo em bacias hidrográficas analisadas e categorizadas em áreas de muito alta, alta, média, muito baixa e baixa criticidade. O método de classificação considerou a frequência de ocorrência de processos erosivos e o tamanho da área da bacia	Áreas de Muito Alta, Alta, Média, Muito Baixa e Baixa Criticidade	IPT/DAEE	não informado

4. Resultados e Discussões

Pode-se afirmar que, em estudos de monitoramento e prevenção de riscos ambientais envolvendo o desenvolvimento de erosões, as aplicações de técnicas estatísticas de análises multivariadas combinadas com a utilização de sistemas de informações geográficas estão contextualizadas nas questões que contemplam os desafios de conciliar o desenvolvimento econômico à conservação do meio ambiente. Como dito anteriormente, é comum na literatura o emprego de técnicas de análise multivariada de dados associadas a um SIG. Como exemplo, podem ser citados brevemente pesquisas recentes de Pradhan e Lee (2010), Oh e Pradhan (2011), Pourghasemi *et al.* (2012), Devkota *et al.* (2013) e Bui *et al.* (2014). Esses estudos apresentam, como característica principal e comum a todos, resultados que geram indicadores de degradação ambiental e instrumentos de previsão e monitoramento de terrenos suscetíveis a processos erosivos e eventos de risco associados. Como complemento dessas análises, os resultados aqui alcançados podem servir de embasamento para estudos diversos de modelagem ambiental e desenvolvimento de ações norteadoras no âmbito do planejamento territorial.

As Figuras 4 e 5 são os resultados da aplicação da técnica de AD no banco de dados gerado (treinamento e teste, respectivamente), processados no *software* estatístico *IBM SPSS 22.0*. Os gráficos representam as hierarquias de escolhas das variáveis explicativas com objetivo de obtenção de classes (nós) homogêneos ou “puros” (índice Gini ou aprimoramento na Figura 4). A ideia é obter classes, caracterizadas por valores de variáveis explicativas, com maior percentual de frequência de ravinas ou boçorocas. Desta forma, é possível identificar fatores que influenciam a predominância de um processo erosivo em detrimento do outro. Neste trabalho foram obtidos sete nós terminais (ou folhas). Tais nós (Nó 3, Nó 4, Nó 5, Nó 8, Nó 9, Nó 11 e Nó 12) são as mencionadas classes homogêneas segundo processos erosivos. Quanto menores os valores obtidos de *Aprimoramento* (índice Gini), mais homogêneos ou puros serão os nós filhos resultantes. Além disso, a AD obtida apresentou 5 níveis, o que está relacionados com os critérios de partição do algoritmo – *stopping rules* (neste trabalho o critério foi mínimo de 20 observações por nó filho).

A variável de maior importância (que melhor explica a variabilidade dos dados segundo a variável dependente) é “Pedologia”. A partir da raiz, a árvore se ramifica em dois grupos principais: (1) Latossolos (Nó 1) e (2) Argissolos, Nitossolos e Neossolos (Nó 2). Tal variável está diretamente relacionada ao desenvolvimento de processos erosivos, sendo amplamente estudada por diversos autores como principal agente desencadeador de erosão. O modelo gerado apontou que há predominância de boçorocas (79,3%) em solos do tipo Latossolo (Nó 1). Isso porque, em termos gerais, solos desse tipo são mais profundos, apresentam estágio mais avançados de maturação e, portanto, facilitam o desenvolvimento de formações mais avançadas de erosão (como é o caso das boçorocas). Por outro lado, o grupo (2 – Nó 2) aponta um equilíbrio em relação à distribuição dos processos, indicando que formações pedológicas compostas por Argissolos, Nitossolos e Neossolos podem propiciar a formação tanto de ravinas (44,3%) quanto de boçorocas (55,7%).

Posteriormente ocorrem novas segmentações do conjunto de dados considerando as variáveis de “Uso do Solo”, “Criticidade”, “Declividade” e “Geologia”. Na ramificação que classifica os dados a partir dos usos do solo na região em que foram identificados os registros de processos erosivos, observa-se a predominância de boçorocas (75,7%) nas regiões de solo exposto (Nó 3) e áreas construídas (edificações) em regiões de Latossolos (Nó 4).

A variável “Criticidade”, que classifica a área de localização do processo erosivo, está representada pelas segmentações de nós 7 e 8: áreas de criticidade média, alta e muito alta

concentram processos erosivos lineares de ambos os tipos (ravinas – 56,4% e boçorocas – 43,6%), enquanto que em áreas classificadas como baixa e muito baixa criticidade há o predomínio de boçorocas (60,1%). Cabe reforçar que o conceito de crítico medido por essa variável (criticidade) não está relacionado à gravidade dos processos erosivos diagnosticados nas áreas analisadas, e sim à quantidade de erosões identificadas na área e o tamanho dela.

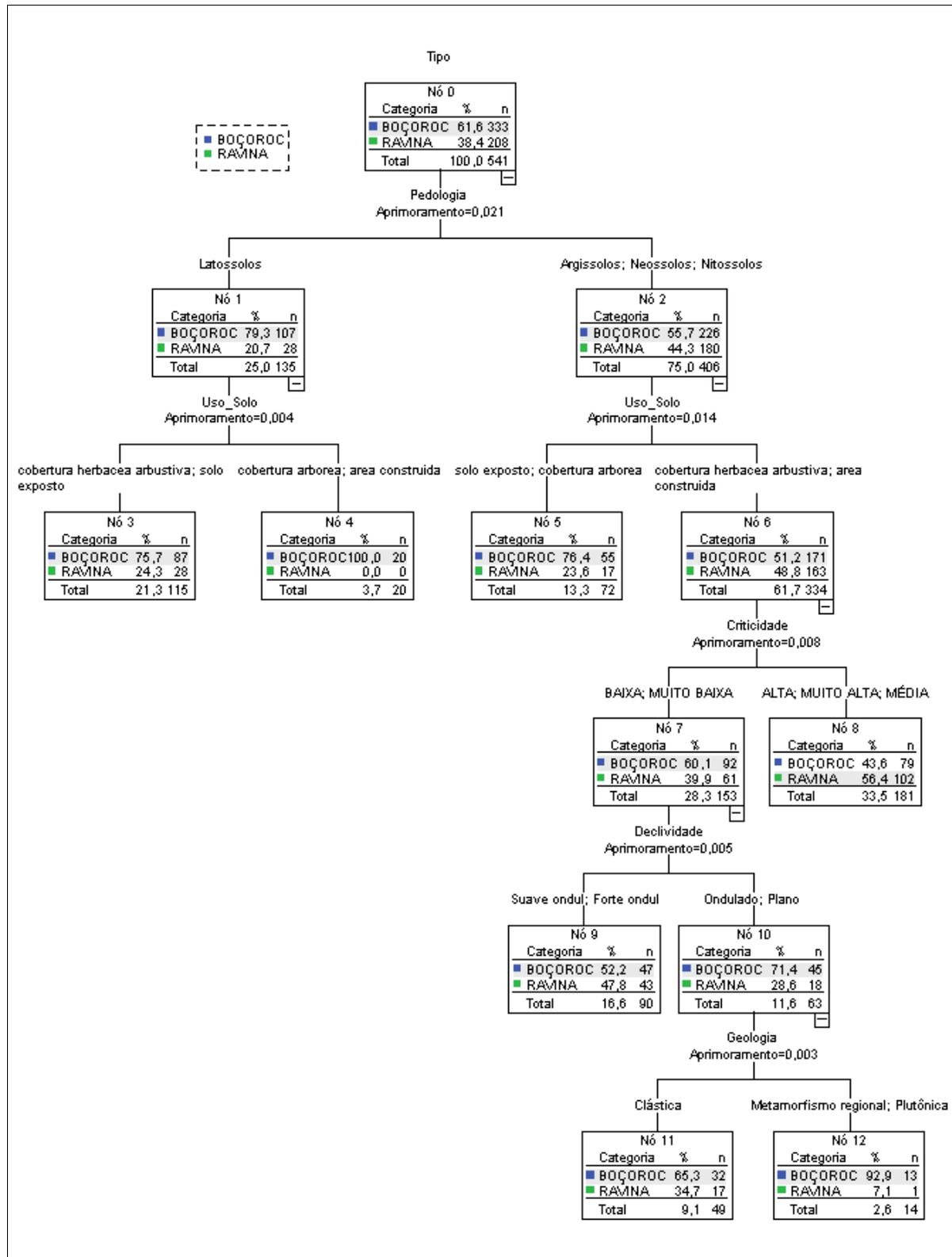


Figura 3: Árvore de Decisão – Tipos de processos erosivos (Amostradas de Treinamento)

Os atributos topográficos (“Declividade”) ou características das vertentes estão entre uma série de elementos controladores da distribuição dos processos erosivos. Sua importância para o desenvolvimento desses fenômenos advém do fato da topografia controlar a intensidade e direção dos fluxos hídricos pluviais, que por sua vez são os principais agentes erosivos no meio tropical úmido. Porém, para a presente pesquisa, os nós 9 e 10, representados pela variável “Declividade” são pouco conclusivos. Observa-se a maioria de processos erosivos do tipo boçoroca (71,4%) nos níveis mais baixos e intermediários de declives (plano e ondulado), ao passo em que em declividades mais acentuadas (forte ondulada) e intermediária (suave ondulado) há equivalência de registros de ambos os tipos.

Já no trabalho de Souza *et al.* (2014), as características topográficas (dentre elas, a declividade) apresentaram relação direta com o tipo de processo erosivo desenvolvido. Os autores realizaram uma pesquisa para testar a associação entre características topográficas e tipos de ocorrência de processos erosivos em linhas férreas. Através do uso de tecnologias geoespaciais e medidas estatísticas de associação não paramétrica, a pesquisa apresenta uma análise da relação entre diferentes tipos de processos erosivos (classificados de acordo com suas características próprias e graus de gravidade) e seus respectivos valores de declividade e hipsometria das vertentes as quais estão localizados. Para esse caso, os autores concluíram que a comparação dos resultados permitiu avaliar a importância da variável espacial na problemática tratada. A análise dos testes estatísticos não paramétricos apontou que existe associação estatisticamente significativa entre as variáveis estudadas.

Por fim, de acordo com Bui *et al.* (2014), em trabalhos recentes, a comparação entre os modelos gerados e a aplicação dos algoritmos de AD apontaram diferenças das capacidades de previsão e os resultados foram satisfatórios em relação a trabalhos já desenvolvidos, tais como Saito *et al.* (2009) e Bui *et al.* (2012). Porém, para a presente pesquisa, é possível afirmar, pela análise dos resultados, que o modelo de AD aplicado apontou dificuldades em diferenciar características relacionadas às ravinas levando-se em conta as variáveis explicativas utilizadas. Tal afirmação pode ser comprovada pelos totais de acertos e erros oriundos da aplicação do algoritmo: para os casos de ravinas, o percentual de erro/acerto ficou por volta dos 50% - Figura 6.

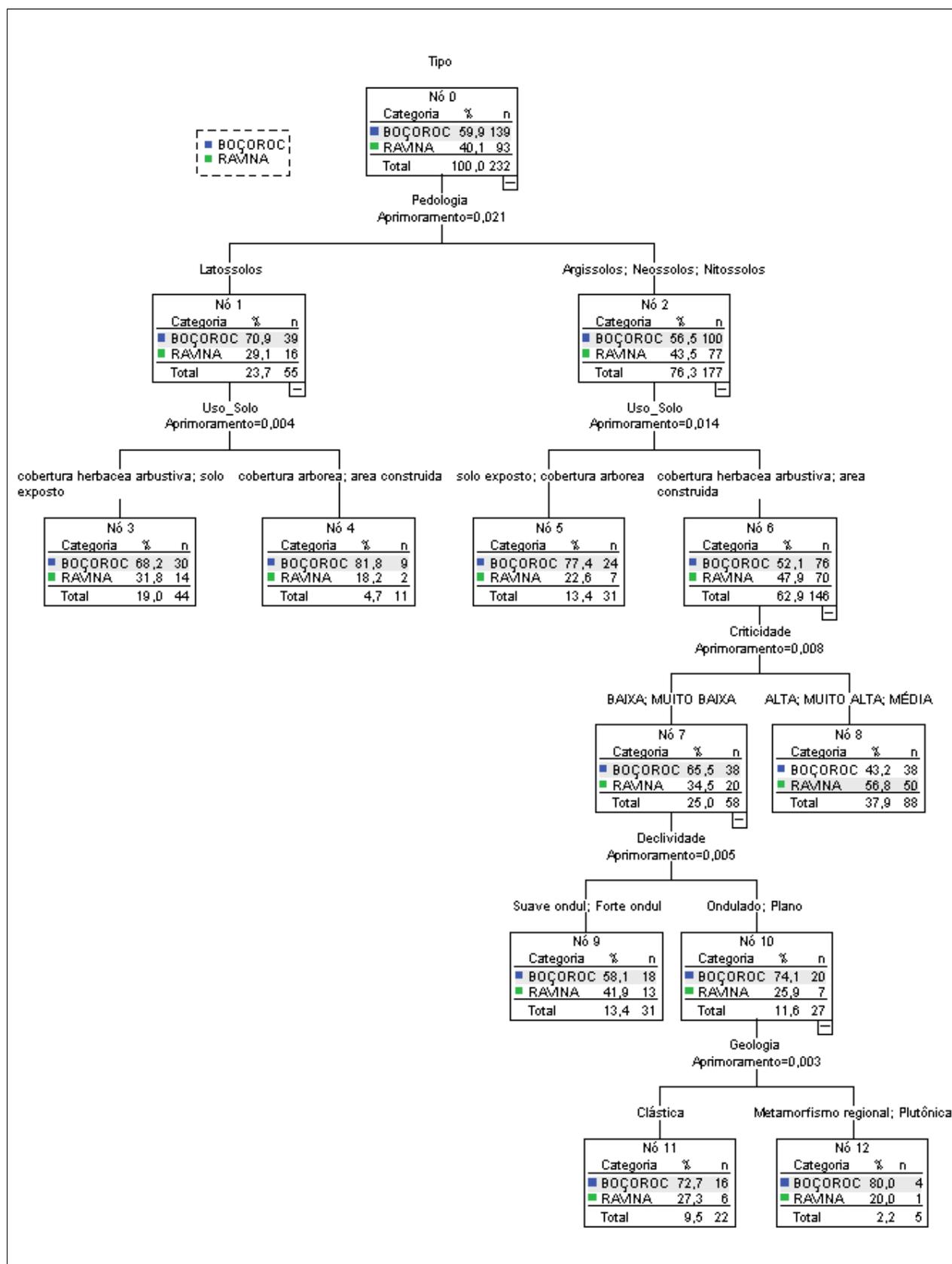


Figura 5: Árvore de Decisão – Tipos de processos erosivos (Amostras de Teste)

Amostra	Observado	Previsto		
		BOÇOROCA	RAVINA	Porcentagem Correta
Treinamento	BOÇOROCA	254	79	76,3%
	RAVINA	106	102	49,0%
	Porcentagem global	66,5%	33,5%	65,8%
Teste	BOÇOROCA	101	38	72,7%
	RAVINA	43	50	53,8%
	Porcentagem global	62,1%	37,9%	65,1%

Figura 6: Árvore de Decisão – Classificação

5. Considerações Finais

Apesar de saber que a dinâmica de desenvolvimento de processos erosivos requer bases mais sólidas e observações diretas, como checagem em campo, acompanhamento dos processos e análise detalhada da região, os resultados fornecidos pela técnica de mineração de dados (Árvore de Decisão), associada ao uso de geotecnologias, podem contribuir diretamente para a elaboração de indicadores ambientais de suscetibilidade de terrenos ao desenvolvimento de fenômenos erosivos e subsidiar métodos que objetivam o cumprimento da legislação ambiental vigente para o setor ferroviário no que diz respeito ao monitoramento, prevenção e controle de processos erosivos ao longo dos traçados das vias.

A análise multivariada, aplicada através da técnica de árvore de decisão, permitiu classificar a amostra de registros mapeados e identificar padrões que distinguem as tipologias de erosões lineares. Além disso, a coleta de dados utilizando bases cartográficas georreferenciadas mostrou-se eficaz no que tange as questões de praticidade e manuseio de banco de dados.

AGRADECIMENTOS

Os autores desejam expressar seus agradecimentos à CAPES pela bolsa de estudos concedida; à ALL (América Latina Logística) pelo apoio a pesquisa desenvolvida e ao DAEE (Departamento de Águas e Esgoto) pela autorização de utilização de dados; ao CNPq pela bolsa de produtividade, ao Instituto de Pesquisas Tecnológicas – IPT e a Fundação de Apoio ao Instituto de Pesquisas Tecnológicas – FIPT pela concessão da bolsa de pós-graduação do projeto Novos Talentos.

REFERÊNCIAS BIBLIOGRÁFICAS

Breiman, L.; Friedman, J.; Olshen, R.; Stone, C. “Classification and Regression Trees.” Chapman & Hall, Belmont. CA, 1984.

Bui, D.; Pradhan, B.; Lofman, O.; Revhaug, I.; Dick, O. "Landslide susceptibility assessment in the Hoa Binh province of Vietnam: a comparison of the Levenberg-Marquardt and Bayesian regularized neural networks." *Geomorphology* 171 – 172:12–29, 2012.

Bui, D. Tien.; H., Tien C.; Revhaug, I.; Pradhan, B.; Nguyen, D. B. "Landslide susceptibility mapping along the national road 32 of Vietnam using GIS-based j48 decision tree classifier and its ensembles." *Cartography from pole to pole*, pag. 303-317. Editora Springer Berlin Heidelberg, 2014.

Crepani, E.; Medeiros, J. S.; Duarte, V.; Hernandez Filho, P.; Florenzano, T. G.; Barbosa, C. C. F. "Sensoriamento Remoto e Geoprocessamento Aplicados ao Zoneamento Ecológico-Econômico e ao Ordenamento Territorial." Instituto Nacional de Pesquisas Espaciais - INPE, São José dos Campos, 2001.

Devkota, K. C.; Regmi, A. D.; Pourghasemi, H. R.; Yoshida, K.; Pradhan, B.; Ryu, I. C.; Dhital, M. R.; Althuwaynee, O. "Landslide susceptibility mapping using certainty factor, index of entropy and logistic regression models in GIS and their comparison at Mugling–Narayanghat road section in Nepal Himalaya." *Natural Hazards* 65 (1), pag. 135 – 165, 2013.

Guerra, A. J. T. O Início do Processo Erosivo. In: GUERRA, A. J. T.; SILVA, A. S.; BOTELHO, R. G. M. *Erosão e Conservação dos Solos*. Rio de Janeiro: Bertrand Brasil, 1999.

Hair Jr, J. F.; Anderson, R. E.; Tatham, R. L.; Black, W. C. *Multivariate Data Analysis*. 5ª ed. Prentice-Hall. Upper Saddle River, New Jersey, 730p, 1998.

Kass, G.V. (1980) An exploratory technique for investigating large quantities of categorical data. *Applied Statistics* , v. 29, p. 119–127.

Morgan, R.P.C. *Soil Erosion and Conservation*. Longman Group, Inglaterra, 1986, 298p.

Oh, T. C. e Pradhan, B. "Application of a neuro-fuzzy model to landslide-susceptibility mapping for shallow landslides in a tropical hilly area." *Computers & Geosciences* 37 (9), pag. 1264 – 1276, 2011.

Pourghasemi, H. R.; Pradhan, B.; Gokceoglu, C. "Application of fuzzy logic and analytical hierarchy process (AHP) to landslide susceptibility mapping at Haraz watershed, Iran." *Natural Hazards* 63 (2), 135 – 165, 2012.

Pradhan, B. e Lee, S. "Delineation of landslide hazard areas on Penang Island, Malasia, by using frequency ratio, logistic regression and artificial neural network models." *Environ Earth Sci*, vol. 60, pag. 1037 – 1054, 2010.

Pitombo, C. S.; Kawamoto, E.; Sousa, A. "An exploratory analysis of relationships between socioeconomic, land use, activity participation variables and travel patterns." *Transport Policy* (Oxford), v. 18, pag. 347 – 357, 2011.

Quinlan, R. *Learning Efficient Classification Procedures and their Application to Chess end-Games*. In: *Machine Learning: An Artificial Intelligence Approach*, p. 463-482. Tioga, Palo Alto, 1983.

Regmi, A. D.; Devkota, K. Chandra Y.; Pradhan, B.; Pourghasemi, H. R.; Kumamoto, T.; Akgun, A. "Application of frequency ratio, statistical index, and weights-of-evidence models and their comparison in landslide susceptibility mapping in Central Nepal Himalaya." *Arabian Journal of Geosciences* 7 (2), pag. 725 – 742, 2014.

Ross, J. L.S. “Análise Empírica da Fragilidade dos Ambientes Naturais e Antropizados.” Revista do Departamento de Geografia, v. 8, FFLCH - USP, São Paulo, 1994.

Ross, J. L. S. “Análise e Síntese na Abordagem Geográfica da Pesquisa para o Planejamento Ambiental.” Revista do Departamento de Geografia, v. 9, FFLCH-USP, São Paulo, 1995.

Saito, H.; Nakayama, D.; Matsuyama, H. “Comparison of landslide susceptibility based on a decision-tree model and actual landslide occurrence: the Akaishi Mountains, Japan.” *Geomorphology* 109, pag. 108 – 21, 2009.

Salomão, F. X. T. “Controle e Prevenção dos Processos Erosivos.” *Erosão e Conservação dos Solos: conceitos, temas e aplicações*. Bertrand Brasil, Rio de Janeiro, 1999.

Souza, N. C.; Gianotti, M. A.; Larocca, A. P. C.; Quintanilha, J. A.; Almeida Filho, G. S. “Análise da associação entre características topográficas e tipos de ocorrência de processos erosivos ao longo de linha férrea - estudo de caso: malha paulista – ALL (SP)”. *Bol. Ciênc. Geod., sec. Artigos, Curitiba*, v. 20, no 4, p.948-970, out-dez, 2014.

Licenças de Softwares

ARCGIS/ARCMAP10.2.1.: Maps throughout this paper were created using ArcGIS® software by Esri. ArcGIS® and ArcMap™ are the intellectual property of Esri and are used herein under license. Copyright © Esri. All rights reserved.

IBM SPSS Statistics Standard v.22 Série: 10101151048: Decision Trees, Neural Networks e Bootstrapping.

Recebido em 17 de novembro 2015.

Aceito em 11 de agosto de 2016.