LUANA SOUSA COSTA

SOIL-ENVIRONMENT DIGITAL INFORMATION TO PROVIDE SOLUTIONS FOR SOLID WASTE DISPOSAL AND COFFEE YIELD MODELING

LAVRAS – MG
2025

LUANA SOUSA COSTA

SOIL-ENVIRONMENT DIGITAL INFORMATION TO PROVIDE SOLUTIONS FOR SOLID WASTE DISPOSAL AND COFFEE YIELD MODELING

Dissertation presented to the Universidade Federal de Lavras, as part of the requirements of the Graduate Program in Soil Science, with a concentration in Environmental Resources and Land Use, for the attainment of the Doctoral degree.

Prof. Michele Duarte de Menezes, PhD.
Advisor

LAVRAS – MG
2025

Ficha catalográfica elaborada pelo Sistema de Geração de Ficha Catalográfica da Biblioteca Universitária da UFLA, com dados informados pelo(a) próprio(a) autor(a).

Costa, Luana Sousa.
Soil-environment digital information to provide solutions for solid waste disposal and coffee yield modeling / Luana Sousa Costa. – 2024.
66 p.: il.

Orientador(a): Michele Duarte de Menezes.

Tese (doutorado) - Universidade Federal de Lavras, 2024.
Bibliografia.

1. Digital soil mapping. 2. coffee system. 3. land suitability. I. de Menezes, Michele Duarte. II. Título.

LUANA SOUSA COSTA

SOIL-ENVIRONMENT DIGITAL INFORMATION TO PROVIDE SOLUTIONS FOR SOLID WASTE DISPOSAL AND COFFEE YIELD MODELING

Dissertation presented to the Universidade Federal de Lavras, as part of the requirements of the Graduate Program in Soil Science, with a concentration in Environmental Resources and Land Use, for the attainment of the Doctoral degree.

APPROVED on February 22, 2024.
PhD. Elvio Giasson – UFRGS (external)
PhD. Tiago Teruel Rezende – UFLA (external)
PhD. Renata Andrade (internal)
PhD. Sergio Henrique Godinho Silva (external)

Prof. Michele Duarte de Menezes, PhD.
Advisor

Lavras – MG
2025

AGRADECIMENTOS

A Deus por seguir me amparando, protegendo e cuidando.

Aos meus pais que sempre me apoiaram e depositaram toda a confiança em mim e em meus sonhos.

À Professora orientadora Michele, pelos ensinamentos, conversas, paciência e por ter oferecido todo suporte necessário para desenvolvimento do trabalho.

Aos professores do departamento de solos da UFLA por todo ensinamento durante as disciplinas e pela disponibilidade em nos auxiliar no desenvolvimento da pesquisa.

Ao professor Elpídio Filho – UFV que através da oferta da disciplina de Pedometria forneceu a base para que eu pudesse realizar parte das análises deste trabalho.

Aos colegas da salinha, Marcelo, Maria Eduarda e Fernanda por todo apoio, ensinamentos, cafezinhos e momentos de descontração.

Ao Consórcio de Saneamento Regional – Consane – nas pessoas de Daniela de Fatima Pedroso e Ivan Massimo Leite Leite pelo suporte e parceria.

Ao DCS/UFLA, pela oportunidade, suporte e ensinamentos.

Agradeço a todos que de alguma forma contribuíram e me ajudaram durante toda a trajetória.

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Código de Financiamento 001.

GENERAL ABSTRACT

Multiple digital information sources, field observations, or laboratory routines to characterize the soil and environment features coupled with powerful predictive models are vital for solving problems, offering new insights, discovering complex patterns, and optimizing land management and use. This dissertation presents two chapters involving soil-environment as a fundamental basis for analysis. The first chapter proposes a suitability system for solid construction waste disposal with further application involving a novel digital soil map of Nepomuceno municipality, Minas Gerais, as a study case. Despite extensive pedological knowledge, current environmental legislation and regulations have simplified aspects that may pose risks for solid construction waste disposal activities and human health. A digital soil mapping framework was developed to map soils and to provide environment characterization. In addition, a suitability system was established based on interpretations of the soil-landscape relationship through attributes listed in a guide table, discussing the potentialities and limitations associated with its position in the landscape. The second chapter aimed to develop farm-scale coffee yield predictive models from machine learning algorithms and digital terrain models, soil fertility information, magnetic susceptibility, airborne gamma-ray spectroscopy, monthly precipitation, and satellite vegetation indices. The random forest, gradient boosting machine (GBM), and radial support vector machine (SVM) algorithms retrieved a proper accuracy of predictions for soil classes (chapter one, only random forest applied) and coffee yield (chapter two). The attributes in the first chapter proposed as criteria for the suitability system complement the current state legislation. Topography and soil depth were the most limiting factors of the areas in the case study. A total of 236 ha closer to the urban perimeter connected by roads in good condition were classified as suitable for managing medium- and small-scale daily volume, whose destination might reduce transportation and installation costs in the study area. Although different predictive models for coffee yield were developed, considering both positive and negative biennial states, the model incorporating all four seasons outperformed the others. This superior performance is likely attributed to the larger volume of input information available, enabling the model to adjust across a wider range of values. Concerning the choice of algorithms, they demonstrated comparable performance, with the random forest algorithm exhibiting slightly better results - lower RMSE (12.35 bags ha^{-1}) and MAE (8.79 bags ha^{-1}), and a higher R^2 (0.87) compared to all other models in this study. Rainfall, NDVI, and soil properties related to fertility were ranked as the most important factors for forecasting yield.

Keywords: Digital soil mapping; environmental covariate; machine learning; coffee system; land suitability.

RESUMO GERAL

Múltiplas fontes de informação digital, observações de campo ou rotinas laboratoriais para caracterizar os atributos do solo e do ambiente e, juntamente com modelos preditivos poderosos, são vitais para resolver problemas, oferecer novos *insights*, descobrir padrões complexos e otimizar o manejo e uso da terra. Esta tese apresenta dois capítulos que envolvem o solo e seu ambiente de ocorrência como base fundamental. O primeiro capítulo propôs um sistema de aptidão para a disposição de resíduos sólidos da construção civil com posterior aplicação em um mapa digital de solos do município de Nepomuceno como caso de estudo. Apesar do amplo conhecimento pedológico, a legislação e regulamentação ambiental vigente simplifica aspectos que podem representar riscos para as atividades de disposição de resíduos sólidos de construção e para a saúde humana. Uma estrutura de mapeamento digital de solos foi desenvolvida para mapear solos e fornecer a caracterização ambiental. Além disso, foi estabelecido um sistema de aptidão baseado em interpretações da relação solo-paisagem através de atributos listados em uma tabela guia, discutindo as potencialidades e limitações associadas à sua posição na paisagem. O segundo capítulo teve como objetivo desenvolver modelos preditivos de produtividade de café em escala agrícola a partir de algoritmos de aprendizado de máquina e modelos digitais de terreno, informações sobre fertilidade do solo, suscetibilidade magnética, espectroscopia de raios gama aerotransportados, precipitação mensal e índices de vegetação obtidos via satélite. Os algoritmos de *random forest* (RF), *gradient boosting machine* (GBM) e radial support vector machine (SVM) recuperaram uma precisão adequada de previsões para classes de solo (capítulo um, apenas o algoritmo RF aplicado) e produtividade de café (capítulo dois). Os atributos do primeiro capítulo propostos como critérios para o sistema de aptidão complementam a legislação estadual vigente. A topografia e a profundidade do solo foram os fatores mais limitantes das áreas do estudo de caso. Um total de 236 ha mais próximos do perímetro urbano interligados por estradas em boas condições foram classificados como aptos para o manejo do volume diário de médio e pequeno porte, cuja destinação poderá reduzir custos de transporte e instalação na área de estudo. Embora diferentes modelos preditivos para a produtividade do café tenham sido desenvolvidos, considerando estados bienais positivos e negativos, o modelo que incorpora todas as safras estações superou os demais. Este desempenho superior é provavelmente atribuído ao maior volume de informações de entrada disponíveis, permitindo que o modelo se ajuste a uma gama mais ampla de valores. Quanto à escolha dos algoritmos, eles demonstraram desempenho comparável, com o algoritmo RF exibindo resultados ligeiramente melhores – RMSE (12,35 sacas ha^{-1}) e MAE (8,79 sacas ha^{-1}) mais baixos, e um R^2 mais alto (0,87) em comparação com todos os outros modelos deste estudo. A precipitação, o NDVI e as propriedades do solo relacionadas com a fertilidade foram classificadas como os factores mais importantes para a previsão da produtividade.

Palavras-chave: Mapeamento digital de solo; covariável ambiental; aprendizagem de máquina; cafeicultura; aptidão de terras.

Social, technological, economic, and cultural impacts

The first article was developed in collaboration with the Regional Consortium for Basic Sanitation (CONSANE) involving students, faculty, and municipal authorities. By developing a method to delimitate suitable areas for the disposal of construction and demolition waste, this work directly contributed to the Municipal Solid Waste Management Plan, offering a solution to the limited technical and financial resources in small municipalities, which hinder effective waste management and compliance with current regulations. By supporting integrated solid waste management (aligned with SDG 6 - Clean Water and Sanitation), the article provides a sustainable approach to managing construction and demolition waste, reducing environmental impact, and promoting responsible land use. The thematic areas of university extension covered include Environment and Health. The direct beneficiaries included municipal decision-makers and the local population of Nepomuceno municipality, in Minas Gerais state, totaling over 25,000 residents, with potential applications for neighboring municipalities. The second article presents a study carried out in collaboration with a commercial coffee farm in the Campos das Vertentes indication of origin region. It explored machine learning algorithms to predict coffee yield based on soil topography, parent material, vegetation indexes, climate data, and along with a dataset of historical yield values. This study enhances the understanding of factors influencing yield at site-specific, enabling data-driven decisions for sustainable resource management. It directly supports SDG 2 – Zero Hunger and Sustainable Agriculture by proposing methods to improve yield prediction, optimize soil inputs management, and implement precision agriculture strategies. The results impact local coffee farmers and the regional economy, with potential applications for broader agricultural sectors. Overall, this dissertation contributes to SDG 15 – Life on Land, focusing on sustainable land use and conservation practices. It exemplifies the integration of academic research, extension activities, and technological innovation to address pressing environmental and agricultural challenges.

Impactos sociais, tecnológicos, econômicos e culturais

O primeiro artigo foi desenvolvido em colaboração com o Consórcio Regional de Saneamento Básico (CONSANE), envolvendo ainda estudantes, professores e autoridades municipais. Ao desenvolver um método para delimitar áreas adequadas para o descarte de resíduos de sólidos da construção civil, este trabalho contribuiu diretamente para o Plano Municipal de Gestão de Resíduos Sólidos, oferecendo uma solução para os recursos técnicos e financeiros limitados dos pequenos municípios, que dificultam a gestão eficaz de resíduos e o cumprimento das regulamentações vigentes. Ao apoiar a gestão integrada de resíduos sólidos (alinhada com o ODS 6 - Água Limpa e Saneamento), o artigo oferece uma abordagem sustentável para a gestão de resíduos de construção e demolição, reduzindo o impacto ambiental e promovendo o uso responsável da terra. As áreas temáticas da extensão universitária abordadas incluem meio ambiente e saúde. Os beneficiários diretos foram os tomadores de decisão municipais e a população local do município de Nepomuceno, Minas Gerais, totalizando mais de 25.000 habitantes, com aplicações potenciais para municípios vizinhos. O segundo artigo apresenta um estudo realizado em colaboração com uma fazenda comercial de café na região de indicação geográfica Campos das Vertentes, Minas Gerais. O estudo explorou o uso de um algoritmo de aprendizado de máquina para prever a produtividade do café com base na topografia do solo, material de origem, índices de vegetação, dados climáticos, além de um conjunto de dados históricos de valores de produtividade. Este estudo aprimora a compreensão dos fatores que influenciam a produtividade para manejo de precisão, permitindo decisões baseadas em dados para o manejo sustentável dos recursos. Ele apoia diretamente o ODS 2 – Fome Zero e Agricultura Sustentável ao propor métodos para melhorar a previsão da produtividade do café, otimizar o manejo de insumos do solo e implementar estratégias de agricultura de precisão. Os resultados impactam os cafeicultores locais e a economia regional, com aplicações potenciais para setores agrícolas mais amplos. No geral, esta dissertação contribui para o ODS 15 – Vida Terrestre, com foco em práticas sustentáveis de uso da terra e conservação. Ela exemplifica a integração de pesquisa acadêmica, atividades de extensão e inovação tecnológica para enfrentar desafios ambientais e agrícolas urgentes.

SUMMARY

FIST PART

1. GENERAL INTRODUCTION

In addition to serving as a support and source of water and nutrients for plants, soil serves as the primary medium for waste disposal (STRECK et al., 2018). The knowledge about the soils and their spatial variability is crucial to ensure optimal uses, preventing degradation, environmental contamination, and loss of ecosystem functions. Concerning land use for waste disposal, the design of environmental solutions should take into consideration soil, since its suitability is variable depending on granulometry, structure, depth, and position in the landscape where they occur, among others (OLIVEIRA et al., 2016). In addition, the definition of soil spatial variability is pivotal for crop management (VISCARRA ROSSEL; LOBSEY, 2016). Information about soil, combined with other attributes such as landscape features, climate, and organisms captured through various analyses and sensors organized as a database serves as a source of information to aid proper decision-making.

There are several techniques for processing data and generating information to enhance the planning of environmental resource use. Machine learning algorithms present themselves as promising tools, given their significant potential in learning complex patterns (prediction) from a database (in the case of supervised classification) and facilitating to obtaining information from the data, namely, the discovery of new patterns (SHARMA et al., 2021). Through modeling, meaningful insights can optimize soil use and management strategies.

Although there is extensive knowledge about soils and environmental applications, current legislation and regulations are oversimplified posing risks to the disposal of solid construction waste and human health, as discussed in the first article. The variability of the environment across croplands can impose limitations on achieving higher and sustainable yields. Strategies for understanding the sources of yield variability are discussed in the second article, focusing on the coffee system. Therefore, this dissertation was structured into two chapters in the format of scientific articles, with soil-landscape as the fundamental basis for better land use planning and crop production.

In response to a demand from the Regional Basic Sanitation Consortium (CONSANE) for the co-creation and formulation of policies in municipalities, the first article, entitled *Disposal of solid waste from civil construction: a screening proposal for a suitability system and case study in Nepomuceno, Minas Gerais,* proposes a suitability system for the disposal of solid construction waste. This system was subsequently applied to a novel digital soil map of Nepomuceno municipality as a case study. Despite extensive pedological knowledge, current environmental legislation and regulations simplify aspects that may pose environmental and

human health risks during solid construction waste disposal activities. A digital soil mapping framework was developed to map soils and provide a better environmental characterization since there is a lack of soil information at a proper resolution. Furthermore, a suitability system was established based on interpretations of the soil-landscape relationship through attributes listed in a guide table, where the potential for waste disposal was defined.

Coffee is recognized as one of the world's most widely consumed beverages and plays a crucial role in the socio-economic of Brazil fabric. Brazil, as the leading global producer and supplier of coffee, significantly contributes to economic sustainability and social development by generating income and creating employment opportunities. Covering 2.25 million hectares of crops (CONAB, 2024) and offering a rich variety of coffee types, the Brazilian coffee industry has fostered the emergence of distinct geographical indications, which have notably influenced consumer preferences and reshaped coffee consumption patterns.

The recent attainment of a provenance indication by the Campos das Vertentes micro-region (INPI, 2020), located in the coffee-rich state of Minas Gerais, marks a noteworthy development. Despite the production of high-quality coffees in this region, a notable gap exists in information concerning the environmental factors influencing coffee production, particularly the drivers of coffee yield at a site-specific scale (Santana et al., 2021; Martello et al., 2022a). Recognizing the imperative for comprehensive yield monitoring, this study seeks to address the scarcity of information by investigating the intricate factors affecting coffee yield, crucial for effective harvest planning and sustainable coffee system evaluations. Thus, the second article, entitled *Long-term coffee yield prediction from machine learning algorithms at farm scale*, aimed to develop predictive models of coffee yield through machine learning algorithms and digital terrain models, soil fertility, magnetic susceptibility, airborne gamma-ray spectroscopy, monthly precipitation, and vegetation indexes. Factors influencing coffee yield are multifaceted, encompassing biotic, abiotic, and management practices, resulting in pronounced spatial variability. The absence of consensus on mathematical models or explanatory variables underscores the complexity of accurately predicting coffee yields. This study underscores the necessity of robust prediction methodologies, exploring the evolving landscape of predictive modeling in coffee systems. It delves into the underexplored realm of site-specific management studies and highlights the increasing prevalence of sophisticated machine learning models in predicting coffee yields, providing valuable insights into the complex dynamics of coffee production.

REFERENCES

CONAB. **Acompanhamento da safra Brasileira de café**. Available at: https://www.conab.gov.br/info-agro/safras/cafe/boletim-da-safra-de-cafe/item/download/51500_05d8a26dc91d95853fb934b03934bc4b. Acessed in January 22, 2024.

INPE – Instituto Nacional da Propriedade Industrial. Indicações geográficas. **Revista da Propriedade Industrial**, N° 2603, 24 de novembro de 2020. Available in http://revistas.inpi.gov.br/pdf/Indicacoes_Geograficas2603.pdf

MARTELLO, M. et al. Use of Active Sensors in Coffee Cultivation for Monitoring Crop Yield. **Agronomy**, v. 12, n. 9, p. 2118, 2022.

OLIVEIRA, J.M., et al. Land suitability for final waste disposal with emphasis on septic systems installation in southern Minas Gerais, Brazil. **Ciênc. grotec**. vol.40 no.1 Lavras Jan./Feb. 2016.

SANTANA, L. S. et al. Advances in precision coffee growing research: A bibliometric review. **Agronomy**, v. 11, n. 8, p. 1557, 2021.

SHARMA, A. et al. Machine Learning Applications for Precision Agriculture: A Comprehensive Review. ***IEEE Access***, vol. 9, p. 4843-4873, 2021.

STRECK, E.V et al**. Solos do Rio Grande do Sul**. 3. ed. rev. ampl. Porto Alegre: UFRGS: EMATER/RS-ASCAR, 2018. 251 p.

VISCARRA ROSSEL, R. A.; LOBSEY, C. **Scoping review of proximal soil sensors for grain growing**. CSIRO, Australia, July, p. 52, 2016.

SECOND PART – ARTICLES

2 DISPOSAL OF SOLID WASTE FROM CIVIL CONSTRUCTION: A SCREENING PROPOSAL FOR A SUITABILITY SYSTEM AND CASE STUDY IN NEPOMUCENO, MINAS GERAIS

* Corresponding author:
e-mail: michele.menezes@ufla.br

Received: May 11, 2023
Approved: October 04, 2023
Published: March 11, 2024

https://doi.org/10.36783/18069657rbcs20230044

HIGHLIGHTS

- Suitability system complements the current environmental legislation.
- Pedology tools support municipality management in the disposal of solid waste from civil construction.
- *Municipality of Nepomuceno can allocate landfills in areas totaling 236 ha.*

ABSTRACT: Most Brazilian municipalities do not have regulated areas for solid waste disposal in civil construction. Usually, residues are disposed of vacant lots and dumps, posing risks to the population health and the environment. Soils are the primary means for the disposal or recycling of waste, highlighting the importance of well-characterized soils and their respective landscape. This study aimed to establish a land suitability system for solid residues in civil construction and apply such information in a case study in Southeastern Brazil. An unprecedented digital soil map with a resolution of 30 m was created using the random forest classifier algorithm and soil field prospection information. A guide listing favorable soil-landscape attributes that most prevent soil erosion, water bodies or water table contamination was elaborated and discussed. Thus, such information was linked through a suitability system to classify areas with potential for receiving waste on a daily volume basis as follows: large

size: >500 m^3 day^{-1}, medium size: >100 and <300 m^3 day^{-1}, and small size <100 m^3 day^{-1}. Topography and soil depth were the most limiting factors of the areas in the case study. The attributes proposed as criteria for the suitability system complement the current state legislation. A total of 236 ha closer to the urban perimeter connected by roads in good condition were classified as suitable for managing medium- and small-scale daily volume, whose destination might reduce transportation and installation costs in the study area.

Keywords: land-use planning, soil survey, random forest, environmental legislation.

2.1 INTRODUCTION

Historically, suitability systems that applied soil-landscape information as a fundamental basis have mainly focused on food, fuel, or fiber production (Ramalho Filho and Beek, 1995; Costa et al., 2009; Amaral, 2011; Silva et al., 2013; Jamil et al., 2018; Taveira et al., 2019). However, soils have faced unprecedented pressures concerning degradation and urbanization (Viscarra-Rossel et al., 2016), increasing the challenge of environmental sustainability to reduce the impacts on ecosystem services (McBratney et al., 2014). In this sense, Pedron et al. (2006) developed a system considering potential urban use for waste disposal, urban construction, urban agriculture, and environmental preservation groups. Oliveira et al. (2016) developed a land-suitability system for waste disposal, emphasizing the installation of septic systems. Ghosh and Nanda (2016) combined favorable soil-landscape characteristics with low population density to create a suitability system for general waste disposal in India. With the rapid urbanization and growing population, Brazilian municipalities need a specific guide to regulate solid waste disposal, especially from civil construction.

Most Brazilian municipalities do not have regulated areas for solid waste disposal in civil construction. Usually, the waste is discarded in vacant lots or dumps, bringing risks to the population health and the environment (Abrelpe, 2019). In general, one of the significant impact is the large volume generated of such wastes, being managed as low-hazardous. However, they might contain organic materials, hazardous products, and various packaging that can accumulate water, favoring insect proliferation and other disease vectors, such as *Aedes aegypti* (Karpinski et al., 2008), a recurrent problem in Brazil.

The regulatory legislation concerning solid waste disposal of civil construction had an important milestone in 2002, through resolution No. 307/2002 of the National Environment Council – CONAMA, which established that the generating source of the waste was responsible for the proper disposal. Nevertheless, most Brazilian municipalities dispose of the waste in

irregular locations. Municipalities urban cleaning service is not considered legally responsible for collecting this waste. In 2018, 122,012 Mg day^{-1} of waste were abandoned on roads and other public places. In the southeastern region of Brazil, cleaning services have collected 63,679 Mg day^{-1} of waste, equivalent to a per capita collection of 0.726 kg $inhabitant^{-1}$ day^{-1} (Abrelpe, 2019).

In addition to federal regulations, the state of Minas Gerais has two essential guidelines for solid waste management: a) the normative deliberation of the State Environmental Policy Council – Copam No. 244 of 2022, which establishes guidelines for the definition of areas for the final disposal of the urban solid waste; and b) the Federal law No. 12,305 of 2010 (Brasil, 2010), consisting of the national solid waste management policies as a mandatory mean of integrating management regulations. The management regulations must identify the most appropriate areas for adequate waste disposal. Federal law No. 14,206 of 2020 set a deadline for implementing the environmentally appropriate final disposal of municipalities with a population of up to 50,000 until August 2, 2024. Additionally, the Brazilian Standard – NBR 15,113:2004 establishes the minimum requirements for designing, implementing, and operating landfills for solid waste from civil construction and inert waste. Notably, NBR 15,113:2004 does not recommend any waterproofing treatment for civil construction landfills, increasing the importance of soil characterization to ensure the suitability of facilities and environmental sustainability.

The current resolution Copam No. 244 of 2022 only suggests that solutions concerning waste disposal must be regionalized, but detailed guidance about the proper places for disposal is not provided. The soil surveys have stood out considering the tools supporting an adequate environmental and land-use diagnosis (Menezes et al., 2009; Stoorvogel et al., 2017). A complete soil survey report carries in-depth information about the morphological, physical, and chemical properties of a soil profile, attached to a typical landscape of their occurrence, encompassing the environment of each soil class (Resende et al., 2014). Such variation imposes different suitability for waste disposal directly on soil (Oliveira et al., 2016). More recently, digital soil mapping techniques have been applied based on the generalization of the traditional five factors of soil formation proposed by Jenny (1941), from which seven predictive factors (environmental covariates) in digital format are derived from soil, climate, organisms, relief, parent material, time, and spatial/geographical position (McBratney et al., 2003), constituting the so-called Scorpan model. From this framework, soil can be spatially predicted from its properties and environmental covariates, which are proxies of state factors for the spatial association. This additional model input is less connected to soil-forming processes but

improves spatial predictions using closely measured soil sample information (Wadoux et al., 2020).

Soils are the primary mean of disposal or recycling of waste due to their ability to promote: a) the modification of organic and inorganic compounds through chemical reactions and physical or biological processes (Streck et al., 2018); b) filtering, reducing contamination and preventing many pollutants from reaching groundwater (Oliveira et al., 2016); and c) the remediation process since its ecosystem attributes and services contribute to the immobilization, dissipation, and filtering of contaminants. In this context, carrying out a soil survey is a strategic step as it characterizes the spatial variability of soils since their properties influence the retention and leaching of elementals or molecules, erosion, and surface runoff potential. However, the only harmonized soil survey available for the state of Minas Gerais presents a small scale (1:650,000) (UFV/CETEC/UFLA/FEAM, 2010). Soil surveys at more detailed scales compatible with solid waste management purposes are necessary.

According to the national solid waste policy (No. 12,305 of 2010), all municipalities must plan for integrated solid waste management, including identifying environmentally appropriate areas for final disposal. Small and medium-sized municipalities do not have sufficient financial resources and technical staff for the efficient and sustainable management of solid waste. Thus, a consortia is an alternative tool to universalize services (Ventura et al., 2020). To answer a demand from Regional Basic Sanitation Consortium (CONSANE) for policy-making in co-creation and municipalities, this study aimed to develop a suitable system for solid waste disposal in civil construction. In addition, the system was applied in the municipality of Nepomuceno – Minas Gerais State as a study case.

2.2 MATERIALS AND METHODS

2.2.1 Study area

The selected study area was Nepomuceno Municipality in the state of Minas Gerais, Brazil. It is located between latitudes 7672934 and 7637880 m and longitudes 469927 and 484733 m, zone 23S (UTM coordinates, *datum* WGS 84) (Figure 1), with altitudes ranging from 756 to 1136 m. The municipality has a total of 583.78 km^2 of area and has 26,769 inhabitants (IBGE, 2020). The pilot area contains physiographic features representative of the Campos das Vertentes region, where the interpretations generated here could be extrapolated. Nepomuceno plays an essential sanitary role even for the surrounding municipalities due to sanitary and industrial landfills installed for receiving waste. The climate is characterized as Cwb according to the Köppen classification system (Alvares et al., 2013) (mesothermal with mild and humid

summers and dry winters). The mean annual temperature is 18.9 °C, and the annual rainfall is 1,630 mm. Granite, gneisses, migmatites, quartzites, and ultramafic rocks mainly represent geology. The native vegetation is represented by subperennial tropical forests and savannas (Cerrado) with pastures and cultivated areas consisting mainly of coffee plantations (Trouw et al., 2007). Thus, applying the suitability system in landscapes with similar conditions to those abovementioned is recommended.

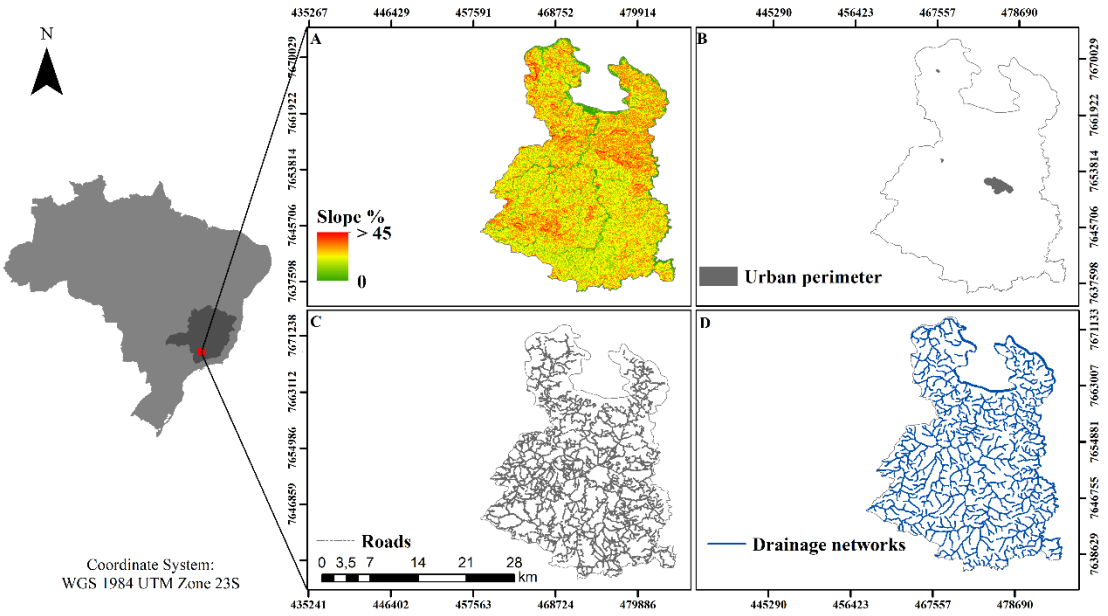

Figure 1. Location map of the municipality of Nepomuceno. (a) slope range; (b) urban perimeter; (c) road infrastructure; and (d) drainage networks. Adaptad from Spatial Data Infrastructure of the State Environment and Water Resources System (IDE- Sisema).

Initially, parameters were applied according to the normative deliberation of the State Environmental Policy Council – Copam No. 118/2008. Although revoked, the guidelines for the adequacy of the final disposal of urban solid waste in the State of Minas Gerais (Table 1). Based on these criteria, the spatial pre-definition of areas with the potential for allocating civil construction landfills was carried out.

Table 1. Criteria used to delimit areas with potential for allocating civil construction landfills according to Normative Deliberation Copam No. 118/2008

Weighted-overlay analysis parameters	**Criteria**
Distance from urban centers	1 km > suitable location <10 km
Distance from roads	Suitable location <100 m
Distance from water bodies	Suitable location >500 m
Slope	Suitable location <30 %

Slope was calculated from the digital elevation model (DEM) with 30 m of resolution obtained from the Shuttle Radar Topography Mission (SRTM) in the SAGA geographic information system (System for Automated Geoscientific Analysis) (Conrad et al., 2015). Information about the current land-use and coverage of the soil, hydrography, vegetation, and roads (including the local roads) (Figure 1) were obtained from the Spatial Data Infrastructure of the State Environment and Water Resources System (IDE- Sisema) (Sisema, 2019) at 1:50.000 scale.

Areas considered inapt or possessing any restriction according to Copam No. 118/2008 were firstly excluded (Table 1). For that, the weighted overlay tool from ArcGIS version 10.1 (ESRI®) was applied using the digital layers of the slope, location of urban centers, hydrography, and road infrastructure.

2.2.2 Digital soil mapping

Prospecting locations were previously selected for field campaigns from the Latin hypercube conditioned by the cost algorithm developed in the R software (R Development Core Team, 2009) through the *clhs* package (Minasny and McBratney, 2006). The following georeferenced digital information was used as input data in the algorithm: a) terrain attributes that have significance regarding soil-landscape (DEM, slope, surface texture of the terrain, and topographic wetness index); b) information that might affect the cost of access of a given location: the Euclidean distance function was applied to the road infrastructure layer in ArcGIS version 10.1 (ESRI®) to create weights based on the distances from the roads (the longer the distance, the higher the weight assigned, the higher the cost of access) (Roudier et al., 2012; Silva et al., 2015).

In 25 soil locations, morphological description and sample collections were performed according to Santos et al. (2015). Soil texture analysis was done using the pipette method at 0.00-0.20 , 0.40-0.70, and 1.00-1.20 m soil layers (Gee and Bauder, 1986). Soils were classified according to the Brazilian Soil Classification System (Santos et al., 2018) and US Soil Taxonomy (Soil Survey Staff, 2014). Additionally, 18 legacy soil profiles containing full descriptions were obtained from Villela (2020) to compose the complete soil database.

2.2.3 Environmental covariates

The DEM with 30 m resolution was obtained from the Shuttle Radar Topography Mission (SRTM). Images were pre-processed to fill gaps and make the model hydrologically consistent. From the DEM, the covariates related to topography were calculated in the SAGA GIS software

(Conrad et al., 2015): aspect, valley depth, channel network base level (cnbl), channel network distance (cnd), LS factor, multiresolution index of valley bottom flatness (mrvbf), SAGA topographic wetness index (stwi), slope, and terrain surface texture (texture).

Airborne gamma-ray spectrometry survey was used as a proxy of soil parent material (Lacoste et al., 2011; Weihermann et al., 2016). This sensor measures the natural radiation emanating from the Earth's surface from the decay series of K (%), Th (ppm), and U (ppm) in the first 0.30 m of the surface. Standard processing for aerial geophysical surveys was performed using OASIS MONTAJ 9.7 software (Smethurst, 2005). The interpolation of the dataset was calculated using minimum curvature in regular grids of 100 m (1/4 to 1/5 of the spacing of the flight lines), generating the K (%), Th, and U (ppm) concentration raster grids (IAEA, 2003). With the grid knitting tool, the resulting grids were fixed. Finally, the total gamma-ray flux (dose) was calculated based on the weighted additions of radioactive elements using the formula (Wilford, 2012): Dose = 13.078 K (%) + 5.67 U (ppm) + 2.49 Th (ppm). Th and U concentrations generally increase, and K decreases during bedrock weathering and soil formation (Wilford and Minty, 2006).

Annual precipitation data with a 30 m resolution was obtained from WorldClim 2.1 (Fick and Hijmans, 2017) to represent climate as a soil-forming factor. Soil organic carbon stock was obtained from Gomes et al. (2019) due to the influence of organisms for soil characteristics.

2.2.4 Random forest: input information, modeling, and accuracy assessment

First, a circular buffer with a 30 m radius around each soil prospection point was created. This buffer consisted of a polygon used to overlay a raster map to extract pixels within this area. Thus, it was assumed the same soil class for each pixel within this buffer polygon in order to increase the training dataset size. Results have suggested that this task increases the accuracy of random forest (RF) (Pelegrino et al., 2016; Machado et al., 2019). In this study, the buffer increased the database for 131 soil input information (Table 2) over the 41 field sampling.

Table 2. Soil database used for mapping, soil classes in the Brazilian Soil Classification System and U.S. Soil Taxonomy (this one between parentheses), number of field prospections, and samples number extracted from the buffer

Soil class	Number of field prospections	Number of samples extracted with the 30 m buffer
Cambissolo Háplico Tb Distrófico (Dystrustept)	5	12
Floodplain soils complex	5	10
Latossolo Amarelo Distrófico (Hapludox)	3	6
Latossolo Vermelho Distrófico (Acrudox)	20	43

Latossolo Vermelho-Amarelo Distrófico (Hapludox)	3	7
Nitossolo Vermelho Distrófico (Rhodudult)	1	3
Argissolo Vermelho Distrófico (Rhodudult)	2	5
Argissolo Vermelho-Amarelo Distrófico (Hapludult)	2	4
Total	41	131

Second, since the RF algorithm is sensitive to different training dataset strategies (Machado et al., 2019), the Spearman correlation test was performed to reduce the multicollinearity among environmental covariates (the ones with more than 80 % correlation were excluded) (Koreen and Richardson, 2015).

Furthermore, the soil map was obtained from spatial prediction with the RF algorithm using *randomForest* package (Breiman, 2001) in the R software (R Development Core Team, 2009). For this study, the *ntree* optimal parameter value found was 500, from the iterations in the caret package and better accuracy return. The *mtry* hyperparameter controlled the number of sample variables. Thus, the number of variables used for each tree was the square root of the total number of variables (*mtry* = 4) (Machado et al., 2019). The class of each sample is then determined by the most frequent type from the trees constructed. Two-thirds of the samples were used to build the trees (model), and the remaining third was used for validation and called out-of-bag (OBB) calculation. The samples were inserted into the decision tree, and a predicted class was assigned to each OOB sample. A single model was obtained with the RF result, accompanied by the aggregate error estimated and the overall OBB estimated error rate (Heung et al., 2014).

The accuracy of the digital soil map was assessed by overall accuracy (OA) and Kappa index (KI) from a confusion matrix. The OA was calculated from the sum of the main diagonal components of the confusion matrix divided by the total number of samples used in the validation. The KI, a measure of agreement, considered all elements of the confusion matrix, the number of soil classes, and the samples correctly classified. Values range from -1, where there is more significant disagreement, to 1, suggesting excellent agreement (Landis and Koch, 1977; Pelegrino et al., 2016).

2.2.5 Suitability system for solid waste disposal from civil construction

Suitability classes were defined according to the potential or not for receiving waste in daily volume: large size: >500 m^3 day^{-1}, medium size: >100 and <300 m^3 day^{-1}, small size <100

m^3 day^{-1}, and inadequate (lands that are not supposed to receive any waste). The greater the volume, the higher the potential risk of environmental contamination. The potential for receiving waste in daily volume was established by Copam No. 217/2017, which deals with aspects according to size criteria, polluting potential, and location criteria. Thus, the areas that bring together all the attributes favorable to waste disposal will be classified as those with the greatest volume reception capacity.

Soil and landscape properties obtained from soil survey and geoprocessing analysis related to water dynamics or soil erosion were analyzed. Favorable attributes are those that most prevent soil erosion and or water bodies or water table contamination. The increasing of limiting attributes decreases the volume of waste suitability. The most limiting attribute is the determinant to fit a given soil-landscape in each suitability class. A final guide table is suggested containing a broad combination of soil features applied in the study case.

2.3 RESULTS

2.3.1 Suitability system

Table 3 shows the suitability system of waste disposal for civil construction. The latter establishes criteria for classification according to the size and polluting potential of environmental licensing of facilities and activities that use environmental resources in Minas Gerais, Brazil. The values was adapted from Oliveira et al. (2016) and Streck et al. (2018). The general description of the selected soil and terrain characteristics of the proposed system is presented further:

- Soil depth: is related to the volume of soil available for the adsorption, attenuation, and stability of residues. This natural attenuation generally involves transferring and stabilizing the contaminant from one location to another (from the surface to the subsurface, for example). The main advantage is reducing and diluting the pollutants efficiently and continuously (Andrade et al., 2010). Studies show that some leaches from civil construction waste may contain significant metallic contaminants (Torgal and Jalali, 2011; Staunton et al., 2015). High concentrations of As, Ba, Cd, Co, Sb, Se, and Tl were found in slugs collected in civil construction waste disposal areas compared to slugs collected in control areas (Staunton et al., 2015). Although there are few studies about the polluting potential, little evidence reinforces the need to consider the remedial role of soil depth when characterizing sites with the potential to receive waste.

- Soil texture: consists of the proportions of soil mineral fractions with a diameter smaller than 2 mm, whose classifications based on the textural triangle are clayey or very clayey (clay content greater than 350 g kg^{-1}), loam (less than 350 g kg^{-1} of clay and more than 150 g kg^{-1} of

sandy, excluding sand fraction), and sandy (sand fraction). The tiniest fractions are responsible for the higher sorption, retention, and inactivation of residues.

- Textural gradient: consists of the accumulation of clay at the subsurface, resulting in a reduction of soil permeability at depth (B horizon), contributing to increasing the lateral flow and surface runoff of water, increasing susceptibility to erosion and risk of contamination of adjacent areas or watercourses (Streck et al., 2018).

- Soil drainage: refers to the speed of water flowing through the soil infiltration, affecting its moisture regime. Drainage provides an indication of soil permeability as well as the risk of groundwater contamination.

- Water table seepage: refers to the presence and depth of occurrence of the water table, where areas with its level close to the surface should be avoided, as transport of contaminants in solution may occur.

- Perched water table: its presence or absence is observed, being more frequent in soils with substantial physical impedance or abrupt textural gradient. Its occurrence can lead to the transport of contaminants in solution and lateral flow.

- Flood risk: areas subject to flooding are inadequate for waste disposal due to the risk of contaminants transportation by water. Redoximorphic soil colors (Kämpf and Curi, 2012) can help to indicate flood events or locals with a higher probability of their occurrence.

- Local relief and slope: related to the conformation of the soil surface. The steeper the slope, the greater the risks of erosion and contamination of adjacent areas and watercourses. Furthermore, this information is helpful from the point of view of adjustments for installing landfills or assessment improvement. The relief phases considered were: flat (0-3 % slope), gently undulated (3-8 %), undulated (8-20 %), and strongly undulated (20-30 %, according to current legislation, wastes could not be disposed in slopes greater than 30 %), and strongly undulated to mountainous (>30 %).

According to Copam deliberation No. 217/2017 for civil construction landfills, earthworks are not recommended, and the waste should be deposited directly on the soil surface, so stoniness was not included as an attribute in the suitability system.

Table 3. Guide table for the suitability system for disposal of civil construction waste

Soil and landscape attributes	Large, medium, and small size	Medium and small size	Small size	Inadequate

Soil depth	>1.50 m	1.00-1.50 cm	0.50-1.00 m	<0.50 m
Soil texture	Clayey	Loam	Loam	Sandy
Soil textural gradient	Without or gradual clay increase	With gradient, being abrupt at depth >1.00 m	With gradient, being abrupt at depth >1.00 m	Without
Drainage class	Excessive, strong or accentuated	Well to moderate	Imperfect	Poor or very poor
Water table seepage	Absent or >1.80 m	Absent or 1.00-1.80 m	<1.00 m	Superficial
Perched water table	Absent	Absent	Present	Present
Flood risk	Null	Null	Rare	From occasional to frequent
Local relief and slope	Flat to gently undulated (<8 %)	Undulated (8-20 %)	Strongly undulated (20-30 %)	Strongly undulated to steep (>30 %)

Large size: >500 m^3 day^{-1}, medium size: >100 and <300 m^3 day^{-1}, and small size <100 m^3 day^{-1} (Copam No. 217/2017). Values adapted from Streck et al. (2018) and Oliveira et al. (2016).

2.3.2 Digital soil mapping

With the performance of Spearman correlation, the following environmental covariates were excluded: channel network base level (cnbl), channel network distance (cnd), and valley depth. This previous selection, along with the iterative steps of the RF algorithm, ensures using the most adapted environmental covariates for the study area. Thus, the spatial prediction of soils in the Nepomuceno municipality can be viewed in figure 2. Estimated OBB error was 6.82 %, and both overall accuracy and kappa index were 0.95, indicating a good accuracy of the predictive model. The RF algorithm and environmental covariates have demonstrated their ability to stratify environments and predict soil classes with adequate accuracy (Machado et al., 2019; Carvalho Junior et al., 2020).

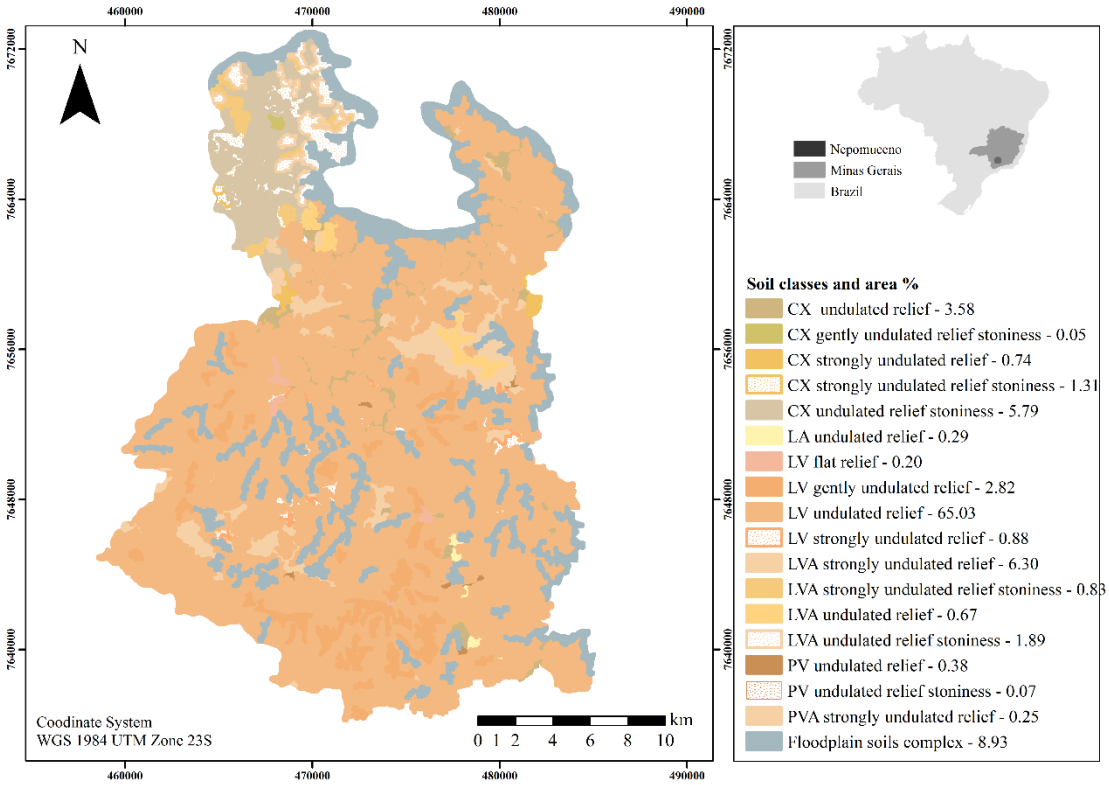

Figure 2. Soil map of the municipality of Nepomuceno. CX: *Cambissolo Háplico Tb Distrófico* (Dystrustept); LA: *Latossolo Amarelo Distrófico* (Hapludox); LV: *Latossolo Vermelho Distrófico* (Acrudox); LVA: *Latossolo Vermelho-Amarelo Distrófico* (Hapludox); PV: *Argissolo Vermelho Distrófico* (Rhodudult); PVA: *Argissolo Vermelho-Amarelo Distrófico* (Hapludult).

Latossolos Vermelhos Distróficos (Acrudox), *Latossolos Vermelho-Amarelos Distróficos* (Hapludox), and *Latossolos Amarelos Distróficos* (Hapludox) (78.90 % of the area) occurred mainly in undulated relief on more elongated smoothed slopes, which is the predominant landscape condition in the municipality. The *Cambissolos Háplicos Tb Distrófico* (Dystrustept) corresponded to the second-largest geographic expression in the municipality, totaling 11.47 % of the area, and occurred mainly in undulated and strongly undulated relief. Floodplain soils complex (8.93 %) corresponded to those environments closer to water bodies in flat or concave relief that leads to water stagnation, causing iron and manganese reduction (greyish soils) or organic carbon accumulation (organic soils). *Argissolos Vermelhos Distróficos* (Rhodudult) and *Argissolos Vermelho-Amarelos Distróficos* (Hapludult) occupied 0.7 % of the area, mainly in undulated relief. They present clay content increasing in depth, with B horizon presenting a blocky structure, decreasing the permeability compared to A horizon (granular structure).

Based on the guide table (Table 3), field campaigns, and the soil map of the municipality of Nepomuceno (Figure 3), the suitability system was applied in areas previously selected according to the normative deliberation Copam 108/2008 (Table 1). The weighted overlap

excluded 95 % of the municipality that does not comply with the legislation. The remaining areas were then classified according to the potential volume of waste received per day (Figure 3).

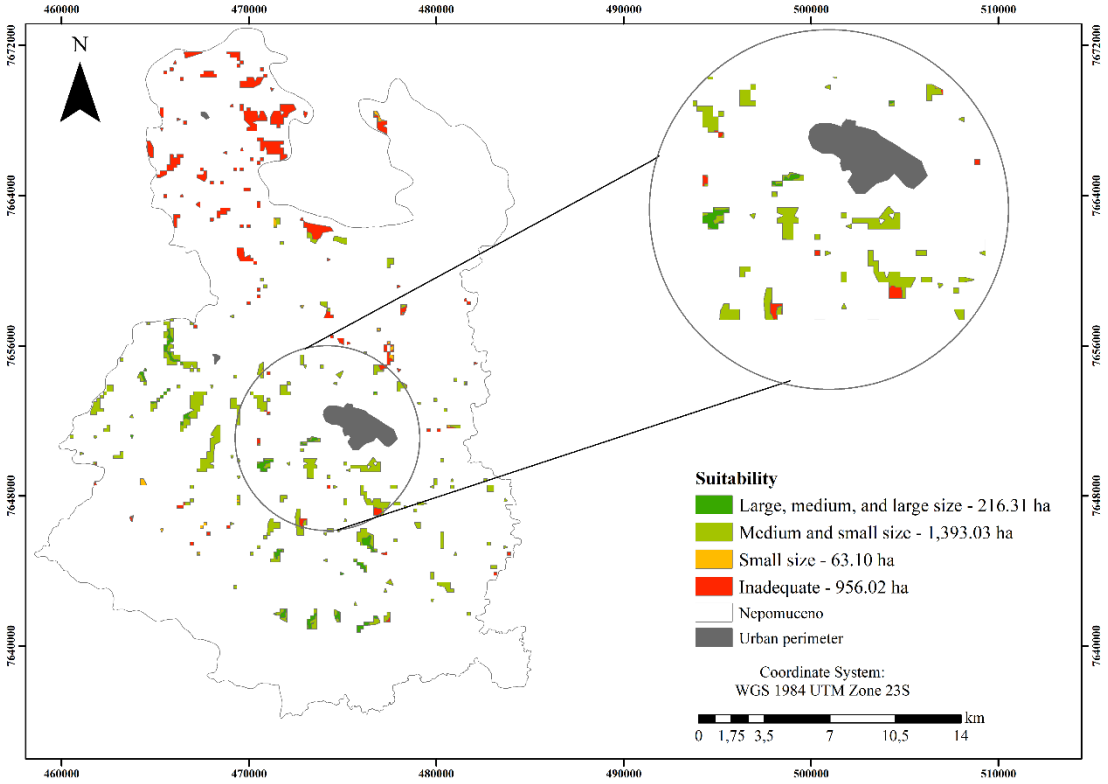

Figure 3. Map of suitability for solid waste disposal from civil construction for Nepomuceno.

2.4 DISCUSSION

Latossolos Vermelhos Distróficos (Acrudox) are very weathered-leached, thicker, have uniform clay content in-depth, and have granular structure in the B horizon. Thus, they are porous soils, mainly in the subsurface, which gives them high permeability (Ferreira, 1999). Their physical properties allied to gentle and stable relief of occurrence decrease susceptibility to erosion. In addition, their thickness contributes to a large volume of soil for waste remediation. *Cambissolos Háplicos Tb Distrófico* (Dystrustept) mainly occurred in the steepest portions of the relief. They present an incipient degree of development and higher values of silt/clay ratio (> 0.6 in the B horizon) that confer greater instability and increase erosion susceptibility. They presented a loam texture, with stoniness in some areas, and additionally, a small *solum* depth was observed, which are shallower soils in dissected reliefs.

Where the granite-gneiss emerges, there is an occurrence of steeper relief and the occurence of *Cambissolos Háplicos Tb Distróficos* (Dystrustept) is favored. The occurrence of ultramafic,

mafic, and schist rocks and gentle relief (Oliveira et al., 1983) favors the occurrence of *Latossolos Vermelhos Distróficos* (Acrudox). It has been reported the predominance of *Latossolos Vermelhos Distróficos* (Acrudox) formed over gabbro, whose rock is more easily weathered than other ones reported in the region (Resende et al., 2011; Curi et al., 2020), corroborating with the occurrence of Acrudox and Dystrustept in the digital soil map.

Argissolos Vermelhos Distróficos (Rhodudult) and *Argissolos Vermelho-Amarelos Distróficos* (Hapludult) are well-drained, and as highlighted on the soil map (Figure 3), they presented clay content increasing in depth, with a gradual or abrupt textural gradient at depths >1.00 m. Stoniness and rocky outcrops were observed in some areas where they occur. Although the *Nitossolos Vermelhos Distróficos* (Rhodudult) appear in the training dataset, the RF algorithm did not perform its spatial prediction likely caused by its occurrence in only one sampled place of the area, hampering the learning process of the algorithm. This limitation concerning the natural imbalance dataset has been previously reported by Sharififar et al. (2019) and Taghizadeh-Mehrjardi et al. (2020).

Most of the areas for solid waste disposal from civil construction were classified as medium and small (1393.03 ha), where the determining and limiting attribute was the undulated relief where *Latossolos Vermelhos Distróficos* (Acrudox) and *Latossolos Vermelho-Amarelos Distróficos* (Hapludox) predominated. The areas classified as inadequate (956.02 ha) were limited by their location in floodplain soils, where flood risk was the limiting factor, and by the occurrence of *Cambissolos Háplicos Tb Distróficos* (Dystrustept) in undulated relief (Table 4). Notably, although the legislation considers only the distance from water bodies (500 m) for water disposal, this distance was not enough to exclude all floodplain areas as indicated in the soil mapping. This fact highlights the importance of soil surveys on a more detailed scale for the accurate choice of disposal facilities since floodplain environments tend to occur in smaller mapping units.

Table 4. Soil mapping units and their respective suitability for civil construction of waste disposal in m^3 day^{-1}, geographical expression, and the most limiting factors

Soil mapping unit	**Suitability** m^3 day^{-1}	**Area** %	**Most limiting factor**
ARGISSOLO VERMELHO Distrófico (Rhodudult) undulated relief	Medium and small size	0.45	Relief
ARGISSOLO VERMELHO – AMARELO Distrófico (Hapludult) strongly undulated relief	Inadequate	0.25	Relief

CAMBISSOLO HÁPLICO Tb Distróficos (Dystrustept) - strongly undulated relief	Inadequate	2.05	Depth
CAMBISSOLO HÁPLICO Tb Distróficos (Dystrustept) - undulated relief	Inadequate	9.37	Depth
CAMBISSOLO HÁPLICO Tb Distróficos (Dystrustept) – gently undulated relief	Inadequate	0.05	Depth
LATOSSOLO AMARELO Distrófico (Hapludox) – undulated relief	Medium and small size	0.29	Relief
LATOSSOLO VERMELHO – AMARELO Distrófico (Hapludox) – strongly undulated relief	Inadequate	7.13	Relief
LATOSSOLO VERMELHO – AMARELO Distrófico (Hapludox) – undulated relief	Small size	2.56	Relief
LATOSSOLO VERMELHO Distrófico (Hapludox) – strongly undulated relief	Small size	0.88	Relief
LATOSSOLO VERMELHO Distrófico (Acrudox) – undulated relief	Medium and small size	65.03	Relief
LATOSSOLO VERMELHO Distrófico (Acrudox) – gently undulated relief	Large, medium and large size	2.82	Unrestricted
LATOSSOLO VERMELHO Distrófico (Acrudox) – flat relief	Large, medium and large size	0.20	Unrestricted
Floodplain soils complex	Inadequate	8.93	Flood risk

The areas classified as large, medium, and small (total of 216 ha) were those in an undulated relief with a predominance of *Latossolos Vermelhos Distróficos* (Acrudox), strongly drained soils with a granular structure in the B horizon, which in turn are very weathered-leached and thick. Those are the areas with the most favorable attributes to receive greater

volumes of waste daily. Finally, areas classified as small (63.10 ha) were limited by undulated relief.

Normative Copam No. 244 of 2022, which revoked Copam No. 118/2008, is the only one with aspects that surrogates applied Pedological information or environmental characteristics. It establishes that waste disposal must be at a minimum distance of 500 m from water bodies in slopes lesser than 30 % and prioritize soils with reduced water permeability. Based on insights promoted by the field campaign and mapping of the study case, as well as accumulated pedological knowledge, the simplicity of normative resolution might lead to potential environmental risks, which are worth noting:

a) Concerning the distance of 500 m from water bodies: floodplains present a complex hydrological dynamic of the extent of inundations during flood events, representing a risk considering only a fixed distance. In this sense, Mello and Curi (2012) suggest applications of pedological and geomorphological indicators as a step forward for hydrological phenomena interpretations.

b) Concerning soils that occur in slopes lesser than 30 %: considering the technical soil survey reports (IBGE, 2015), a broad range of slopes might contain landscapes with different risks of soil erosion. Thus, more specific slope ranges (relief phases, IBGE, 2015) would accurately guide the choice of lands suited for waste disposal facilities. In addition, landscape information should be combined and analyzed, including soil properties such as accumulation of clay content in depth, shallow depths, or greater silt/clay contents, since all of them also increase erosion susceptibility (Resende et al., 2014). Besides the environmental risks, once installed, erosional processes might decrease waste disposal facilities longevity and can increase additional costs of their adequacy.

c) Concerning soils with lower water permeability: these soils present water flowing less efficiently through the soil pores. Considering this soil property only, soils with high and low suitability for waste disposal are also included (Oliveira et al., 2016; Streck et al., 2018). One example is that instead of only preventing contaminated water from reaching the water table or aquifer, the water stagnation or accumulation - an undesirable attribute for waste disposal sites – increases the risk of contamination of water resources. Thus, for an adequate diagnosis, other soil properties, such as redoximorphic soil colors and the occurrence of perched water tables, must improve soil drainage and permeability diagnosis.

Waste displacement is often considered costly for waste disposal in appropriate locations (Doussoulin and Bittencourt, 2022). In this sense, figure 3 highlights the urban perimeter surrounded mostly by areas classified as having the potential to receive medium and small-

sized volumes. This fact, along with proper roads observed during the field campaigns, might reduce the costs of transportation as well as building disposal facilities. Once waste disposal has begun, future assessments are necessary to ensure the sustainability of the system from the maintenance of crucial soil functions, especially those concerning water purification and soil contaminant reduction. Due to the absence of technical guides, diagnostic qualifiers proposed by Costa et al. (2019) are suggested.

2.5 CONCLUSIONS

Digital soil map of the municipality of Nepomuceno supported the evaluation of the areas concerning the suitability system proposed. Steep relief and shallow soils were the most determining limitations in reducing the potential size of the areas in this case study.

The attributes proposed as criteria for classifying the suitability of areas for solid waste disposal from civil construction complemented the current legislation. In addition, more subsidies were offered to indicate the potential areas in the municipalities' integrated solid waste management plans. This approach may be used under similar environmental conditions as in this study for the southeastern region of Brazil. The suitability to receive medium- and small-sized volumes, combined with the proximity to the urban center and good-quality roads, enables the Nepomuceno municipality to have the potential to allocate landfills in areas that add up to 236 ha.

ACKNOWLEDGMENTS

Our thanks to National Council for Scientific and Technological Development - CNPq and Coordination of Superior Level Staff Improvement - CAPES for the scholarship financing; Regional Consortium of Basic Sanitation – CONSANE for the technical and financial support; Minas Gerais Economic Development Company - CODEMIG for providing the airborne gamma-ray spectroscopy survey information.

REFERENCES

Associação Brasileira de Empresas de Limpeza Pública e Resíduos Especiais - Abrelpe. Panorama dos resíduos sólidos no Brasil 2018/2019. São Paulo: Abrelpe; 2019. Available from: http://abrelpe.org.br/panorama/.

Associação Brasileira de Normas Técnicas - ABNT. Norma Brasileira 15,133, de 30 de junho de 2004. Resíduos sólidos da construção civil e resíduos inertes – Aterros Diretrizes para projeto, implantação e operação.

Alvares CA, Stape JL, Sentelhas PC, Gonçalves JDM, Sparovek G. Köppen's climate classification map for Brazil. Meteorol Z. 2013;22:711-28. https://doi.org/10.1127/0941-2948/2013/0507

Amaral FCS. Sistema brasileiro de classificação de terras para irrigação: Enfoque na região semiárida. Rio de Janeiro: Embrapa Solos; 2011.

Andrade JDA, Augusto F, Jardim ICSF. Biorremediação de solos contaminados por petróleo e seus derivados. Eclet Quim. 2010;35:17-43. https://doi.org/10.1590/S0100-46702010000300002

Brasil. Lei nº 12.305, de 2 de agosto de 2010: Institui a Política Nacional de Resíduos Sólidos; altera a Lei no 9.605, de 12 de fevereiro de 1998; e dá outras providências. Brasília, DF: Diário Oficial da União; 2010. Available from: https://www.planalto.gov.br/ccivil_03/_ato2007-2010/2010/lei/l12305.htm.

Brasil. Ministério do Meio Ambiente. Resolução Conama no 307, de 5 de julho de 2002. Estabelece diretrizes, critérios e procedimentos para a gestão dos resíduos da construção civil. Diário Oficial da União, Brasília, 17 jul. 2002.

Breiman L. Random forests. Mach Learn. 2001;45:5-32. https://doi.org/10.1023/A:1010933404324

Carvalho Junior W, Pereira NR, Fernandes Filho EI, Calderano Filho B, Pinheiro HSK, Chagas CS, Bhering SB, Pereira VR, Lawall S. Sample design effects on soil unit prediction with machine: randomness, uncertainty, and majority map. Rev Bras Cienc Solo. 2020;44:e0190120. https://doi.org/10.36783/18069657rbcs20190120

Conrad O, Bechtel B, Bock M, Dietrich H, Fischer E, Gerlitz L, Wehberg J, Wichmann V, Böhner J. System for Automated Geoscientific Analyses (SAGA) v. 2.1.4. Geosci Model Dev. 2015 8:1991-2007. https://doi.org/10.5194/gmd-8-1991-2015

Conselho Estadual de Política Ambiental - Copam. Deliberação Normativa Copam nº 118, 27 de julho de 2008: Altera os artigos 2º, 3º e 4º da Deliberação Normativa 52/2001, estabelece novas diretrizes para adequação da disposição final de resíduos sólidos urbanos no Estado, e dá outras providências. Diário do Executivo – Minas Gerais; 2008. Available from: http://www.siam.mg.gov.br/sla/download.pdf?idNorma=7976

Conselho Estadual de Política Ambiental - Copam. Deliberação Normativa Copam nº 217, 6 de dezembro de 2017: Estabelece critérios para classificação, segundo o porte e potencial poluidor, bem como os critérios locacionais a serem utilizados para definição das modalidades de licenciamento ambiental de empreendimentos e atividades utilizadores de recursos ambientais no Estado de Minas Gerais e dá outras providências. Belo Horizonte: Diário do Executivo – Minas Gerais; 2017. Available from: http://jornal.iof.mg.gov.br/xmlui/handle/123456789/192323

Conselho Estadual de Política Ambiental - Copam. Deliberação Normativa Copam nº 244, 27 de janeiro de 2022: Dispõe sobre os critérios para implantação e operação de aterros sanitários em Minas Gerais e dá outras providências. Belo Horizonte: Diário do Executivo – Minas Gerais; 2017. Available from: https://www.jornalminasgerais.mg.gov.br/?dataJornal=2022-02-17

Costa AM, Curi N, Araújo EF, Marques JJ, Menezes MD. Management units for Eucalyptus cultivation in four physiographical regions of Rio Grande do Sul, Brazil. Sci For. 2009;37:465-73.

Costa JR, Pedron FA, Dalmolin RSD, Schenato RB. Field description and identification of diagnostic qualifiers for urban soils in Brazil. Rev Bras Cienc Solo. 2019;43:e0180121. https://doi.org/10.1590/18069657rbcs20180121

Curi N, Silva E, Gomes FH, Menezes MD, Silva SHG, Teixeira AFS. Mapeamento de solos, aptidão agrícola e taxa de adequação do uso das terras do município de Lavras (MG). Lavras: Editora UFLA; 2020.

Doussoulin JP, Bittencourt M. How effective is the construction sector in promoting the circular economy in Brazil and France?: A waste input-output analysis. Struct Change Econ D. 2022;60:47-58. https://doi.org/10.1016/j.strueco.2021.10.009

Ferreira MM, Fernandes B, Curi N. Influência da mineralogia da fração argila nas propriedades físicas de Latossolos da região Sudeste do Brasil. Rev Bras Cienc Solo. 1999;23:515-24. https://doi.org/10.1590/S0100-06831999000300004

Fick SE, Hijmans RJ. WorldClim 2: New 1km spatial resolution climate surfaces for global land areas. Int J Climatol. 2017;37:4302-15. https://doi.org/10.1002/joc.5086

Gee GW, Bauder JW. Particle-size analysis. In: Klute A, editor. Methods of soil analysis: Part 1 Physical and mineralogical methods. Madison: SSSA; 1986. p. 383-411.

Ghosh S, Nanda S. Site suitability analysis for solid waste management using multi criteria analysis. In: Prashanthi M, Sundaram R, editors. Integrated waste management in India. Cham: Springer; 2016. https://doi.org/10.1007/978-3-319-27228-3_3

Gomes LC, Faria RM, Souza E, Veloso GV, Schaefer CEGR, Fernandes Filho EI. Modeling and mapping soil organic carbon stocks in Brazil. Geoderma. 2019;340:337-50. https://doi.org/10.1016/j.geoderma.2019.01.007

Heung B, Bulmer CE, Schmidt MG. Predictive soil parent material mapping at a regional-scale: A Random Forest approach. Geoderma. 2014;214-215:141-54. https://doi.org/10.1016/j.geoderma.2013.09.016

International Atomic Energy Agency - IAEA. Guidelines for radioelement mapping using gamma-ray spectrometry data. Austria: IAEA; 2003. Available from: https://www-pub.iaea.org/MTCD/Publications/PDF/te_1363_web.pdf

Instituto Brasileiro de Geografia e Estatística - IBGE. Manual técnico de pedologia. 3. ed. Rio de Janeiro: Fundação IBGE; 2015. Available from: https://agenciadenoticias.ibge.gov.br/media/com_mediaibge/arquivos/18acacacfd63204b29c6d820a430b8e4.pdf.

Instituto Brasileiro de Geografia e Estatística - IBGE. Pesquisa Nacional de Área da unidade territorial: Área territorial brasileira. Rio de Janeiro: IBGE; 2020. Available from: https://cidades.ibge.gov.br/brasil/mg/nepomuceno/panorama.

Jamil M, Ahmed R, Sajjad H. Land suitability assessment for sugarcane cultivation in Bijnor district, India using geographic information system and fuzzy analytical hierarchy process. GeoJournal. 2018;83:595-611. https://doi.org/10.1007/s10708-017-9788-5

Jenny H. Factors of soil formation: A system of quantitative pedology. New York: McGraw-Hill; 1941.

Kämpf N, Curi N. Formação e evolução do solo (Pedogênese). In: Ker JC, Curi N, Schaefer CEGR, Torrado PV, editors. Pedologia: Fundamentos. Viçosa, MG: Sociedade Brasileira de Ciência do Solo; 2012. p. 207-302.

Karpinski LA, Pandolfo A, Reinehr R, Rojas JWJ. Gestão diferenciada de resíduos de construção e demolição: Uma visão abrangente no município de Passo Fundo-RS. Estudos Tecnológicos. 2008;4:69-87.

Koreen M, Richardson M. On the importance of training data sample selection in random forest image classification: A case study in peatland ecosystem mapping. *Remote Sens. 2015;*7:8489-515. https://doi.org/10.3390/rs70708489

Lacoste M, Lemercier B, Walter C. Regional mapping of soil parent material by machine learning based on point data. Geomorphology. 2011;133:90-9. https://doi.org/10.1016/j.geomorph.2011.06.026

Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics. 1977;33:159-74. https://doi.org/10.2307/2529310

Machado DFT, Silva SHG, Curi N, Menezes MD. Soil type spatial prediction from random forest: Different training datasets, transferability, accuracy and uncertainty assessment. Sci Agric. 2019;76:243-54. https://doi.org/10.1590/1678-992X-2017-0300

McBratney A, Field DJ, Koch A. The dimensions of soil security. Geoderma. 2014;213:203-13. https://doi.org/10.1016/j.geoderma.2013.08.013

McBratney, AB, Mendonça Santos, ML, Minasny, B. On digital soil mapping. Geoderma. 2003; 117: 3–52. https://doi.org/10.1016/S0016-7061(03)00223-4.

Mello CR, Curi N. Hydropedology**.** Cienc Agrotec. 2012;36-2:137-46. https://doi.org/10.1590/S1413-70542012000200001

Menezes MDD, Curi N, Marques JJ, Mello CRD, Araújo ARD. Levantamento pedológico e sistema de informações geográficas na avaliação do uso das terras em sub-bacia hidrográfica de Minas Gerais. Cienc Agrotec. 2009;33:1544-53. https://doi.org/10.1590/S1413-70542009000600013

Minasny B, McBratney AB. A conditioned Latin hypercube method for sampling in the presence of ancillary information. Comput Geosci. 2006;32:1378-88. https://doi.org/10.1016/j.cageo.2005.12.009

Oliveira JMD, Silva SHG, Menezes MDD, Kämpf N, Curi N. Land suitability for final waste disposal with emphasis on septic systems installation in southern Minas Gerais, Brazil. Cienc Agrotec. 2016;40:37-45. https://doi.org/10.1590/S1413-70542016000100003

Oliveira V, Costa AMR, Azevedo WP, Camargo MN, Olmos J. Pedologia: Levantamento exploratório de solos. In: Projeto Radambrasil. Levantamento de Recursos Naturais V.32: Folhas SF.23/24 Rio de Janeiro/Vitória. Rio de Janeiro; 1983. p. 385-552.

Pedron FA, Dalmolin RSD, Azevedo AC, Poelkin EL, Miguel P. Utilização do sistema de avaliação do potencial de uso urbano das terras no diagnóstico ambiental do município de Santa Maria – RS. Cienc Rural. 2006;36:468-77. https://doi.org/10.1590/S0103-84782006000200017

Pelegrino MHP, Silva SHG, Menezes MDD, Silva ED, Owens PR, Curi, N. Mapping soils in two watersheds using legacy data and extrapolation for similar surrounding areas. Cienc Agrotec. 2016;40:534-46. https://doi.org/10.1590/1413-70542016405011416

Ramalho Filho A, Beeh KJ. Sistema de avaliação da aptidão agrícola das terras. 3. ed. rev. Rio de Janeiro: Embrapa-CNPS; 1995.

Resende M, Curi N, Ker JC, Rezende SB. Mineralogia dos solos brasileiros: Interpretação e aplicações. 2 ed. Lavras: Editora UFLA; 2011.

Resende M, Curi N, Rezende SD, Corrêa GF, Ker JC. Pedologia: Base para distinção de ambientes. 6. ed. rev. amp. Lavras: Editora UFLA; 2014.

Roudier P. Clhs: a R package for conditioned Latin hypercube sampling. Available from: <https://cran.r-project.org/web/packages/clhs/clhs.pdf> 23/03/2021.

Roudier P, Beaudette DE, Hewitt AE. A conditioned Latin hypercube sampling algorithm incorporating operational constraints. In: Minasny B, Malone BP, McBratney AB, editors.

Digital soil assessments and beyond. Proceedings of the 5th Global Workshop on Digital Soil Mapping, Sydney, Australia, 10–13. Boca Raton: CRC Press; 2012. P. 227-31.

Santos HG, Jacomine PKT, Anjos LHC, Oliveira VA, Lumbreras JF, Coelho MR, Almeida JA, Araújo Filho JC, Oliveira JB, Cunha TJF. Sistema brasileiro de classificação de solos. 5. ed. rev. ampl. Brasília, DF: Embrapa; 2018.

Santos RD, Santos HG, Ker JC, Anjos LHC, Shimizu SH. Manual de descrição e coleta de solo no campo. 7. ed. rev. ampl. Viçosa, MG: Sociedade Brasileira de Ciência do Solo; 2015.

Sharififar A, Sarmadian F, Malone BP, Minasny B. Addressing the issue of digital mapping of soil classes with imbalanced class observations. Geoderma. 2019; 350:84-92. https://doi.org/10.1016/j.geoderma.2019.05.016

Silva SHG, Owens PR, Silva BM, Oliveira GCD, Menezes, MDD, Pinto LC, Curi N. Evaluation of conditioned latin hypercube sampling as a support for soil mapping and spatial variability of soil properties**.** Soil Sci Soc Am J. 2015;79:603-11. https://doi.org/10.2136/sssaj2014.07.0299

Silva VA, Curi N, Marques JJG, Carvalho LMT, Santos WJR. Soil maps, field knowledge, forest inventory and ecological-economic zoning as a basis for agricultural suitability of lands in Minas Gerais elaborated in GIS. Cienc Agrotec. 2013;37:538-49. https://doi.org/10.1590/S1413-70542013000600007

Sistema Estadual de Meio Ambiente e Recursos Hídricos - Sisema. Infraestrutura de dados espaciais do sistema estadual de meio ambiente e recursos hídricos - IDE-Sisema. Belo Horizonte: Sisema; 2019. Available from: http://idesisema.meioambiente.mg.gov.br.

Smethurst MA. A software tool to rotate spatial data on the surface of the sphere(Earth) for Geosoft's Oasis Montaj. Norway: NGU Report; 2005.

Staunton J, Williams CD, Morrison L, Henry T, Fleming GT, Gormally MJ. Spatio-temporal distribution of construction and demolition (C&D) waste disposal on wetlands: A case study. Land Use Policy. 2015;49:43-52. https://doi.org/10.1016/j.landusepol.2015.06.023

Stoorvogel JJ, Bakkenes M, Temme AJ, Batjes NH, Brink BJ. S-world: A global soil map for environmental modeling. Land Degrad Dev. 2017;28:22-33. https://doi.org/10.1002/ldr.2656

Streck EV, Kämpf N, Dalmolin RSD, Klamt E, Nascimento PD, Schneider P, Pinto LFS. Solos do Rio Grande do Sul. 3. ed. rev. ampl. Porto Alegre: UFRGS, Emater/RS, Ascar; 2018.

Taghizadeh-Mehrjardi R, Schmidt K, Eftekhari K, Behrens T, Jamshidi M, Davatgar N, Toomanian N, Scholten, T. Synthetic resampling strategies and machine learning for digital soil mapping in Iran. Eur J Soil Sci. 2020;71:352-68. https://doi.org/10.1111/ejss.12893

Taveira LRS, Weindorf DC, Menezes MD, Carvalho TS, Motta PEF, Teixeira AFS, Curi N. Land use capability classification adaptation in low and intermediate technology farming

systems: A soil erosion indicator. Soil Use Manage. 2019;37:164-80. https://doi.org/10.1111/sum.12555

Torgal FP, Jalali S. Eco-efficient construction and building materials. London: Springer; 2011.

Trouw RAJ. Paciullo FVP, Ribeiro A, Cherman A, Chrispim S, Maciel RR. Nepomuceno - SF.23-V-D-III, escala 1:100.000: Nota explicativa. Minas Gerais: UFRJ/CPRM; 2007. Available from: https://rigeo.cprm.gov.br/handle/doc/10467.

Universidade Federal de Viçosa – UFV / Fundação Centro Tecnológico de Minas Gerais - CETEC / Universidade Federal de Lavras - UFLA / Fundação Estadual do Meio Ambiente - FEAM. Mapa de solos do Estado de Minas Gerais. Escala 1:650.000. Belo Horizonte: FEAM; 2010. Available from: http://www.feam.br/noticias/1/1355-mapa-de-solos.

Ventura KS, Suquisaqui ABV. Application of SWOT and 5W2H tools for analysis of solid urban waste intermunicipal consortia. Ambient Constr. 2020;20:333-49. https://doi.org/10.1590/s1678-86212020000100378

Villela BS. Proposta de conservação e recuperação da Microbacia do Córrego Sapé, Nepomuceno – MG. Lavras: Universidade Federal de Lavras; 2020.

Viscarra-Rossel RA, Behrens T, Ben-Dor E, Brown DJ, Demattê JAM, Shepherd KD, Shi Z, Stenberg B, Stevens A, Adamchuk V, Aïchi H, Barthès BG, Bartholomeus HM, Bayer AD, Bernoux M, Böttcher K, Brodský L, Du CW, Chappell A, Fouad Y, Genot V, Gomez C, Grunwald S, Gubler A, Guerrero C, Hedley CB, Knadel M, Morrás HJM, Nocita M, Ramirez-Lopez L, Roudier P, Campos EMR, Sanborn P, Sellitto VM, Sudduth KA, Rawlins BG, Walter C, Winowiecki LA, Hong SY, Ji W. A global spectral library to characterize the world's soil. Earth-Sci Rev. 2016;155:198-230. https://doi.org/10.1016/j.earscirev.2016.01.012

Weihermann JD, Ferreira FJF, Cury LF, Silveira CT. Gamma-ray spectrometry of granitic suites of the Paranaguá Terrane, Southern Brazil. J Appl Geophys. 2016;132:38-52. https://doi.org/10.1016/j.jappgeo.2016.06.017

Wadoux AMJC, Minasny B, McBratney AB. Machine learning for digital soil mapping: Applications, challenges and suggested solutions. Earth-Sci Rev. 2020;210:103359. https://doi.org/10.1016/j.earscirev.2020.103359

Wilford J. Weathering intensity index for the Australian continent using airborne gamma-ray spectrometry and digital terrain analysis. Geoderma. 2012;183-184:124-42. https://doi.org/10.1016/j.geoderma.2010.12.022

Wilford J, Minty B. The use of airborne gamma-ray imagery for mapping soils and understanding landscape processes. Dev Soil Sci. 2006;31:207-18. https://doi.org/10.1016/S0166-2481(06)31016-1

3 LONG-TERM COFFEE YIELD PREDICTION FROM MACHINE LEARNING ALGORITHMS AT FARM SCALE

Abstract

Coffee, a globally consumed beverage, holds immense economic and social significance, with Brazil emerging as the world's largest producer. Campos das Vertentes micro-region in Minas Gerais State, Brazil, recently obtained a provenance indication. There is still a lack of information regarding site-specific factors that influence coffee yield, especially considering new tools or digital information available, hindering more accurate yield predictions. This study aimed to predict coffee yield using machine learning algorithms (random forest, gradient boosting machine - GBM, and support vector machine - SVM), considering the influence of soil and environment on coffee yield of alternate bearing cycles. The study area comprises a 30.25-hectare commercial farm characterized in detail consistent with on-farm monitoring. The coffee yield dataset, spanning the harvest season from 2017/2018 to 2020/2021, included the monitoring of soil properties, relief, vegetation indexes, and rainfall. As expected, significant biennial yield fluctuations were observed, reflecting the coffee plant's phenological cycle. Such a fact justified the five model configurations suggested to predict coffee yield, exploring negative and positive biennial cycles and their combination. The accuracy of the machine learning models was assessed using an independent dataset to calculate the mean absolute error (MAE), root mean square error (RMSE), and coefficient of determination (R^2). The results indicated that models trained and validated for biennial cycles exhibited poor performance, emphasizing the need for considering alternate bearing cycles. The random forest algorithm outperformed SVM and GBM, presenting the most accurate predictions. Despite significant yield variability during the evaluation seasons, the model built from a dataset containing all harvest season information showed higher reliability among models, reaching RMSE of 12.35 bags ha^{-1}, MAE of 8,79 bags ha^{-1}, and R^2 of 0.87. Explanatory variables, including monthly rainfall, Normalized Vegetation Index (NDVI), and soil fertility (pH, exchangeable Ca, and sum of bases), were crucial in predicting coffee yield. This study contributes to precision agriculture management, providing insights into the key factors influencing coffee yield in the Campos das Vertentes region and, hence, into coffee yield predictions.

Keywords: *Coffea arabica*; remote sensing; proximal sensing; soil properties; recursive feature elimination.

3.1 INTRODUCTION

Coffee is one of the most widely consumed beverages in the world. Brazil stands out as the largest world coffee producer and is a top coffee supplier worldwide. As the second-largest consumer market of coffee beverages, the Brazilian coffee industry contributes to social and economic sustainable development by generating income and employment. Totaling 2.25 million hectares of crops (CONAB, 2024) with contrasting coffee beverage typicity, different geographical indications have been established and recognized by producers, markets, and customers, whose information has changed the way consumers buy and drink coffee (QUINTÃO et al., 2017).

Campos das Vertentes, a micro-region located in Minas Gerais State (the largest producer State in Brazil), recently obtained a provenance indication certification (INPI, 2020). This recognition was requested by the Campos das Vertentes Coffee Growers Association, which encompasses 17 municipalities in a region characterized by plateaus with elevations ranging from 500 to 1,000 m, including altitudes above 1,500 meters in eastern Serra da Mantiqueira. The climate characterized by cool, rainy summers, and cold winters at higher elevations, provides favorable conditions for coffee growing with beverages characterized by sweetness, balanced body, and notes of chocolate and nuts. A geographical indication (GI) is a certification system that uses a geographical name (or its synonym) to distinguish products or services originating from a specific region. Although high-quality coffees have been produced (RIBEIRO et al., 2020), information concerning the environment of coffee production is scarce, notably coffee yield drivers at a site-specific scale.

Yield monitoring is essential for different stakeholders, furnishing support to harvest planning (ZANELLA et al., 2024) and investigation concerning the profitability and sustainability of coffee systems (VICTORINO et al., 2016). Factors affecting coffee yield are multivariate, involving general factors such as biotic, abiotic, and management practices (WANG et al., 2011). Interactions among those factors have resulted in high spatial variability of coffee yield (SANTANA et al., 2021; MARTELLO et al., 2022a) without consensus regarding mathematical models or explanatory variables that primarily influence plant yield.

Accurate yield forecasting requires reliable predictions (KITTICHOTSATSAWAT et al., 2022). Some fundamental characteristics can be depicted from ultimate results encompassing coffee systems (without consociation) that applied fundamentals of prediction/forecasting: a) there has been a substantial rising application of complex machine

learning models recently (KOUADIO et al., 2018; BARBOSA et al., 2021a; MARTELLO et al., 2022b; KITTICHOTSATSAWAT et al., 2022; KITTICHOTSATSAWAT et al., 2023; SANTHOSH and UMESH, 2023); b) studies approaching the complexity of plant-yield driver for site-specific management have been overlooked. Most studies have been developed based on small plots, homogenous lands, or plant-yield legacy datasets from governmental or research centers without mentioning spatial information features (IDOL and YUKANA, 2019; VICTORINO et al., 2016; DINH et al., 2022; BARBOSA et al., 2021a; KITTCHOSATSAWAT et al., 2022). The studies that adopted a multivariate statistical approach applied as explanatory variables in the yield models: a) foliar or satellite-based vegetation indexes (BARBOSA et al., 2021a; MARTELLO et al., 2022b; THAO et al., 2022; ZANELLA et al., 2024); b) soil fertility information (KOUADIO et al. 2018); c) climate information (JAYAKUMAR et al., 2016; APARECIDO and ROLIM, 2018; DINH et al. 2022); and d) integrative information such as yield of previous year + climate (VICTORINO et al., 2016), management information + climate (KITTICHOTSATSAWAT et al., 2022 and 2023), and management information + soil fertility information (SANTOSH et al., 2023). Although machine learning models have proved capable of learning complex patterns from multiple sources of explanatory variables, the broad complexity of soil-environment features has been neglected even though crucial for precision agriculture management (GONÇALVES et al., 2022). Great emphasis has been put on climate information only, but one should remember that the plant might be under different soil drainage regimes even under the same rainfall since soil and landscape characteristics could govern water retention and movement differently (GOOD, NOONE, and BOWEN, 2015; COVINO, 2017). In this sense, Kouadio et al. (2021) argued that the integrative perspective might reduce prediction errors and broaden the capabilities of the prediction model.

In addition to the usage of explanatory variables, the training and validating characteristics are essential in machine learning models, especially relevant in coffee systems to optimize the experimental design due to plant alternate bearings. Physiologically, coffee plants present two cycles: the first year presents essentially vegetative activity and lower plant yield (negative biennially); the second year is marked by a reproductive phase leading to higher plant yield (positive biennially) (BERNARDES et al. 2012; GARCIA and ORIANS, 2022). Statistically, a long-term experiment involving *Coffea arabica*, the most cultivated species in Brazil, might be represented by a dataset with a contrasting range of plant yield values between harvest seasons, adding a temporal component to models. This feature has also been incorporated into agrometeorological models for estimating coffee yield (VICTORINO et al.,

2016; APARECIDO and ROLIM, 2018) and should be explored in machine learning models. Environmentally, studies have suggested that water deficits in high-yield seasons affect coffee's reproductive stage more, and in low-yield seasons, they affect the vegetative stage of the crop more (APARECIDO and ROLIM, 2018). Socially, yield fluctuations have been identified as a critical contributor to climate change vulnerability among coffee farmers, given their impacts on income, food security, health, and education (BACA et al., 2014), and thus, should be better investigated.

This study hypothesizes that models to predict coffee yield trained from datasets containing data regarding negative or positive seasons collected biennially could accurately model coffee yield and outperform models trained and validated by using coffee yields of all seasons (collected annually). Thus, we aimed to predict coffee yield using machine learning algorithms considering how alternate bearing might influence model accuracy in a long-term experiment. The relation of parent material-soil-landscape-plant-climate information was used to predict coffee yield by random forest, GBM, and SVM. In addition, we extensively discuss the new insights promoted by RFE-machine learning regarding the main drivers of coffee yield in decision-support systems at the farm scale adapted to a recent indication of origin region Campos das Vertentes.

3.2 MATERIALS AND METHODS

3.2.1 Study area characterization and plant management

The pilot area consists of a commercial farm located in the municipality of Santo Antônio do Amparo in the region of Campos das Vertentes indication of origin (INPI, 2020) (FIGURE 1), state of Minas Gerais, Brazil. The region’s relief is characterized by undulating plateaus, with altitudes ranging from 600 to 1,500m. The climate is classified as Cwb according to Köppen-Geiger, characterized as humid subtropical, featuring hot, humid summers and cold and dry winters (ALVARES et al., 2013). A total of 30.25 hectares of coffee crops (*Coffea arabica L.*, cultivar Acaiá IAC 474-19) were monitored during the 2017, 2018, 2019, and 2020 harvest seasons. Genotype and atmospheric climate are considered uniform within the study due to the size of the study area.

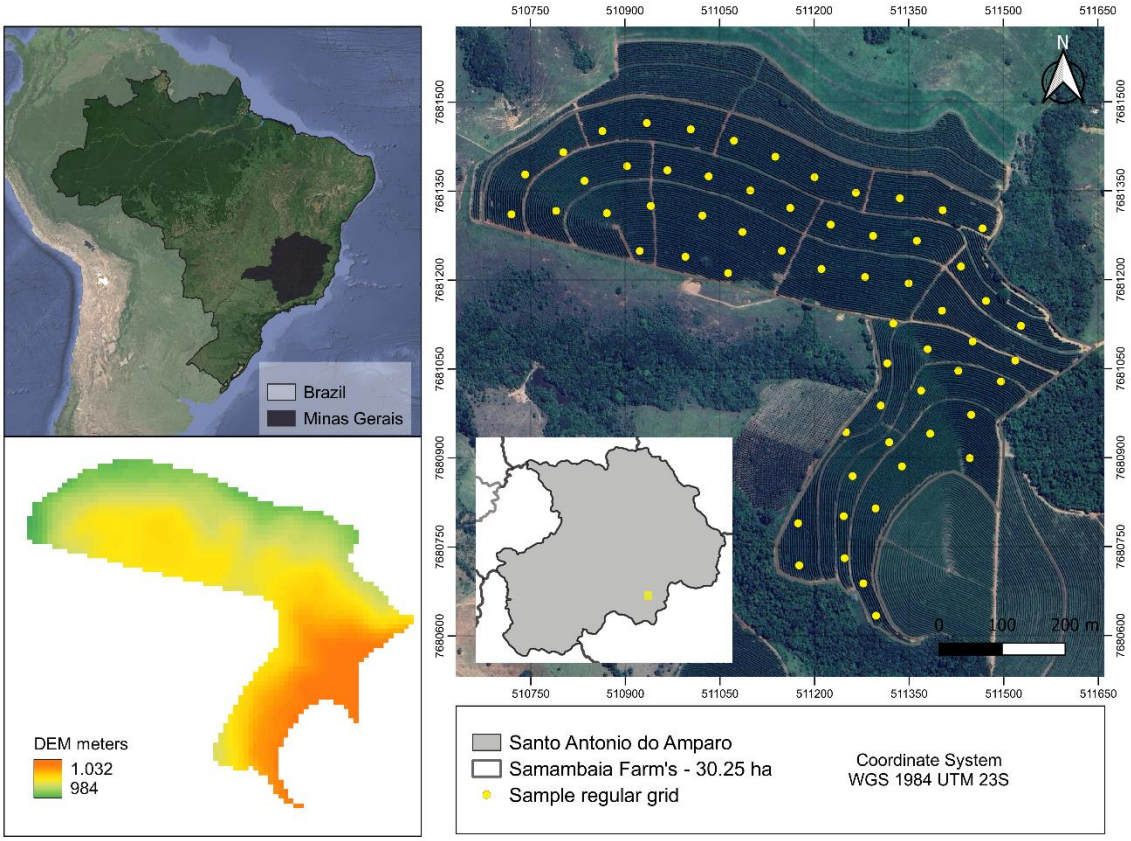

Figure 1. Study area location and plant yield sampling design. DEM – Digital elevation model.

Coffee plants have been sown and grown following the spatial scheme of 3.6 m between rows and 0.8 m between plants, consisting of an unshading intensified monoculture with *Brachiaria decumbens* between the rows. Coffee seedlings were planted in 2000. A drastic pruning was performed in 2015 to renew the coffee plants and retake their productive potential (THOMAZIELLO; PEREIRA, 2008). Pests and diseases are continuously monitored and ongoing phytosanitary management of the crop area. Chemical and biological control of pests and diseases are employed based on the specific needs identified through monitoring and interventions. Fertilizer was applied between November and October in each harvest year at a variable rate following agronomic recommendations based on the Brazilian technical reports devoted to coffee systems (FUNDAÇÃO PROCAFÉ, 2019).

A detailed soil survey was performed, and homogeneity of soil physic-hydric characteristics was observed. For such reason, they did not compose further models. The entire study area presented the occurrence of Oxisols, a highly weathered-leached, deep, with high clay content throughout the soil profile. Although very clayey, the granular structure in the

subsurface B horizon increases soil porosity, thereby ensuring high permeability and favoring plant root growth (FERREIRA, 1999; GONÇALVES et al., 2022).

3.2.2 Dataset acquisition

The complete framework of this study is displayed in Figure 2. The predictive models were trained to predict crop yield in different harvest seasons (dependent variables) by exploring soil properties, relief attributes, plants, and rainfall as independent variables. Those variables, in turn, were chosen following the state-of-the-art coffee yield predictions (JAYAKUMAR et al., 2016; APARECIDO & ROLIM, 2018; SANTOS et al., 2022; BENTO et al., 2022; ZANELLA et al., 2024), adapted to tropical conditions and based on successfully applied sensing (GONÇALVES et al., 2022). It is noteworthy that some variables remain stable over time (such as relief, magnetic susceptibility, and airborne gamma-ray spectrometry), while others were monitored annually (including plant yield, soil fertility, vegetation indices, and rainfall). More details are provided below.

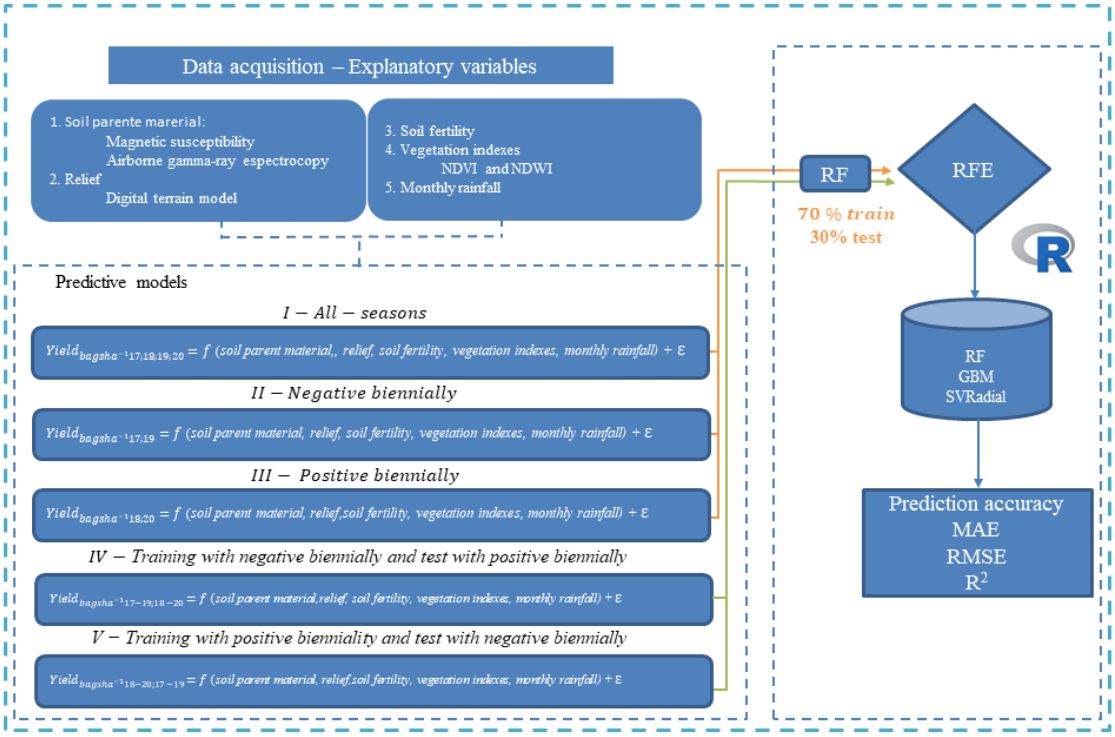

Figure 2. Flowchart of the study from data acquisition to evaluation of predictive models from MAE - mean absolute error, RMSE - root mean square error, R^2 - coefficient of determination.

3.2.3 On-farm monitoring

The sampling scheme of plant yield consists of a total of 61 field observations (FIGURE 1), following a 70 x 70 meters spacing. Each point represents a central georeferenced point

where four plants were harvested: two to the right and two to the left, according to Colaço and Molin (2015) and Oliveira et al. (2011). The coffee yield was determined for each point when the plant reached physiological maturity, typically between May and June. Professional and experienced staff carefully and manually harvested the coffee plants, allowing the calculation of the yield per central point (plant L^{-1}). The volume of coffee harvested from each plant was measured using a graduated container in L. The average yield for the four plants harvested at each sampling point (L $plant^{-1}$) was measured, and then yield values per sampled point were converted into bags of 60 kg of processed coffee per hectare (bags ha^{-1}) using the equation proposed by Matiello et al. (2007):

$$Yield_{bags\ ha^{-1}} = \frac{Yield_{L\ plant^{-1}} \times (plant\ ha^{-1})}{480}$$

In addition, soil samples were taken at 0-20 cm of depth following the sampling scheme presented in Figure 1 for the laboratory (from 2017 to 2020) and proximal sensing analysis (in 2020, the final season of analysis). The soil chemical properties analyzed included P (mg dm^{-3}) and K ($cmol_c\ dm^{-3}$) extracted by Mehlich-1, Ca ($cmol_c\ dm^{-3}$), Mg ($cmol_c\ dm^{-3}$) and Al ($cmol_c\ dm^{-3}$) extracted with 1.0 mol l^{-1} KCl, H + Al ($cmol_c\ dm^{-3}$) extracted with 0.5 mol l^{-1} calcium acetate at pH 7.0, soil pH in water. Thus, base saturation (BS%), sum of bases (SB), ($cmol_c\ dm^{-3}$), and cation exchange capacity at pH 7 (CEC) ($cmol_c\ dm^{-3}$) were calculated.

The air-dried fine earth soil samples were submitted to the magnetometer Bartington MS2B (Bartington Instruments Ltd., England) at a low frequency (0.47 kHz) (BARBOSA et al., 2021b). Such a proximal sensor retrieves values of the induced magnetization of an object relative to the magnetization field intensity (MULLINS, 1977). Since soil survey samples were taken in a different sampling scheme of plant yield, magnetic susceptibility values were spatially predicted across the entire area through ordinary kriging (PEREIRA et al., 2022; KRAVCHENKO, 2003) to harmonize the yield dataset.

3.2.4 Remote sensing and GIS-based dataset

The NDVI (ROUSE et al., 1974) and normalized difference water index (NDWI) (GAO, 1996) were calculated from Landsat 8 sensor satellite images (30 m resolution) for all images in the Surface Reflectance Tier 1 collection encompassing the study area from the 2017 to 2020 harvest seasons. First, the atmospheric correction was performed and the median was systematically calculated based on a stackle of vegetation index per year. Thus, we ended up with a raster-based map containing ranges of yearly median values for NDVI and NDWI. The

image processing was performed on the Google Earth Engine platform (GORELICK et al., 2017). Xie et al. (2019) observed an increasing accuracy of land use and land cover predictions when employing the median to extract information and generate products from the Landsat 8 satellite bands. The median was chosen to eliminate noise and invalid values in the time series.

A digital elevation model (DEM) (30 m of resolution) was obtained from the Shuttle Radar Topography Mission (SRTM). The image was firstly pre-processed to fill gaps and make the model hydrologically consistent. Using the SAGA GIS geographic information system (CONRAD et al., 2015), operated through the *RSAGA*, and *raster* package in R software version 4.1.1, the following morphometric variables were calculated from the DEM: aspect, slope (TRAVIS, 1975), gradient (HJERDT, 2004), multiresolution index of ridge top flatness, multiresolution index of valley bottom flatness (GALLANT and DOWLING, 2003), normalized height, eal surface area (OLAYA, 2004), Saga wetness index (BOEHNER, 2006), standardized height, valley depth (BOEHNER and SELIGE, 2006), terrain ruggedness index (RILEY et al., 1999), terrain surface classification, terrain surface texture (CONRAD, 2012), and topographic position index (WILSON & GALLANT, 2000).

Airborne gamma-ray spectrometry measures the natural radiation emanating from the Earth's surface, resulting from the decay series of K (%), Th (ppm), and U (ppm) in the first 0.30 m of the surface. Although results reflect the soil surface, such information has been broadly applied in geological surveys and functions as a proxy of soil parent material since those radionucleotides are driven mainly by the mineralogy and geochemistry of rocks, soils, and sediments (LACOSTE et al., 2011; WEIHERMANN et al., 2016). In this study, these data were used to derive information about the parent material, representing the parent material component of the soil-forming factors. The raw information was provided by the Geological Survey of Brazil (CPRM) and the Company of Economic Development of Minas Gerais (CODEMIG). Standard processing for aerial geophysical surveys was conducted using OASIS MONTAJ 9.7 software (SMETHURST, 2005). The dataset interpolation was calculated using minimum curvature in regular grids of 100 m (1/4 to 1/5 of the spacing of the flight lines), generating K (%), Th, and U (ppm) concentration raster grids (IAEA, 2003). Subsequently, the resulting grids were finalized using the grid knitting tool. Finally, the total gamma-ray flux (dose) was calculated based on the weighted additions of radioactive elements using the formula (Wilford, 2012): Dose = 13.078 K (%) + 5.67 U (ppm) + 2.49 Th (ppm). Th and U concentrations generally increase, while K decreases during bedrock weathering and soil formation (WILFORD and MINTY, 2006).

3.2.5 Predictive models

Machine learning algorithms were employed to relate coffee and explainable variables following further general model:

$$\begin{aligned} &\textit{Coffee yield} = f(\textit{soil parent material, soil fertility, relief,} \\ &\textit{vegetation indexes, monthly rainfall)} + \textit{error} \end{aligned} \qquad \textit{(Eq. 1)}$$

Random forest, SVM, and GBM algorithms were implemented, respectively, in the *RandomForest* (BREIMAN et al., 2018), *caret* (MEYER et al., 2019), and *gbm* package within the R software environment (R CORE TEAM, 2019). Random forest was developed as an extension of the CART (Classification and Regression Trees) method. It is based on a non-parametric technique that combines predictions made by multiple decision trees, and each tree is generated based on the values of an independent set of random vectors. Each of these sets is created by a type of sampling called bootstrap. Three parameters were defined for this purpose: the number of trees (*ntree*), the minimum number of data in each terminal node (*nodesize*), and the number of variables used in each tree (*mtry*) (LIAW and WIENER, 2002; HU and SZYMCZAK, 2023). The *mtry* requires special consideration (BREIMAN, 2002) since its parametrization leads to model optimization in the *caret* package (KUHN et al., 2020).

The radial basis function SVM employs supervised learning techniques based on the use of hyperplanes for optimal separation between classes in a dataset (CORTES and VAPNIK, 1995; HASTIE et al., 2009), where the margin between the points of the two closest classes, referred to as support vectors, is maximized, leading to a higher probability of better generalization. The parameters used to tune the model in the caret package (KUHN et al., 2020) include the penalty (cost), which controls the trade-off between margin errors and training errors, and the kernel width (sigma), which controls the degree of non-linearity of the model.

The GBM utilizes the concept of additive modeling and combines gradient descent with boosting (FRIEDMAN, 2001). Each decision tree is built by taking a random subsample of the training dataset. The algorithm builds new base learners to be maximally correlated with the negative gradient of the loss function associated with the entire set (CARVALHO JUNIOR et al., 2020). The objective of GBM is to improve model performance by combining a large number of simple trees. The learning procedure consecutively fits new models to provide a more accurate estimate of the response variable.

To understand the tradeoffs of alternate bearing on coffee yield prediction, five different models permuting input training and validation datasets were proposed (FIGURE 2), keeping the general model presented above ((Eq. 1) as the fundamental basis:

- Model I: the coffee yield values from all the harvest seasons were applied to training and validating models, consisting of the most extensive model input dataset.
- Model II: only coffee yield values in the negative biennial harvest season were applied for training and validating the models.
- Model III: only coffee yield values in positive biennial harvest season were applied for training and validating the models.
- Model IV: models were trained using yield values in a negative biennial harvest season and validated with values obtained in a negative biennial harvest season to simulate the farm routine data acquisition sequence in one year and the necessity to forecast yield for the next year.
- Model V: contrary to Model IV, models were trained with positive and validated with negative biennial harvest season datasets. The optimization of each hyperparameter mentioned above for each model was tested using five randomly selected values (*tuneLength*) performed by the *caret* package, evaluated through 10-fold cross-validation. For each model I, II, and III, the process was repeated 50 times with its subset of variables and compared by the mean values of the accuracy variables. For models IV and V, the repetition process (repeated 50 times) could not be established due to the previous definition in the dataset for training and testing established. Different groups of training and validation datasets can yield varying accuracy results. Hence, conducting several repetitions is crucial for determining prediction variability (KUHN and JOHNSON, 2013).

Finally, each model underwent recursive feature elimination (RFE), which iteratively eliminates less promising variable combinations for the prediction model (KUHN and JOHNSON, 2013). Model optimization, focusing on the most influential variables, occurred through 10-fold cross-validation. The caret package (Classification and Regression Training) (KUHN, 2018) in R software version 4.1.1 executed the RFE function for each model. Based on the covariates selected with the RFE function, models were created for each algorithm tested in this study. Selecting covariates with RFE is a strategy to reduce the complexity of the model, making it more straightforward to interpret the relationship between the response variable and the predictor covariates. Furthermore, selecting the most promising covariates improves the accuracy of the model, reducing overfitting and expanding the possibility for generalizations (GREGORUTTI et al. 2017).

From the full coffee yield dataset, 70% was used for model training and 30% was devoted only to the accuracy assessment of models from MAE, RMSE, and R^2 indexes. RMSE indicates the spread of the error distribution, and MAE represents prediction bias: the smaller,

the better the model's performance. The closer the R^2 to 1.0, the better the fit between predicted and estimated values, and the better the models.

3.3 RESULTS AND DISCUSSION

3.3.1 Plant yield

Table 1 displays the descriptive statistics of coffee yield through different harvest seasons. In addition, by analyzing the boxplots (FIGURE 3), it is possible to visualize the distribution and skewness of the dataset. Descriptive statistics of all the explanatory variables are presented in the Appendix. There is a notable yield variability within and between years, revealing datasets with different statistical features. The most pronounced difference across years is related to the synchronous alternate bearing. Coffee plants take two years to complete the phenological cycle of fructification. In the first phenological year, the plant branches only grow to produce beans in the second phenological year. This phenomenon leads to biennial yields: low yields are achieved when the plants are devoted to developing new branches (negative biennially); high yields in the next year are then achieved when plants are producing beans in the new branches (positive biennially). Similar results have been found Nonato et al. (2021) and Faneli et al. (2020) for Coffeea arabica in Brazil. Notably, there was a greater dispersion of yield values, especially for the seasons 2017 and 2019, with negative biennially, when the coefficient of variation found was 58.5% and 30.2%, respectively.

Table 1. Descriptive statistics of coffee yield through the harvest seasons.

Coffee yield (bags ha^{-1})	Minimum	Maximum	Mean ± STD	CV (%)
Season 2017	1.0	25.5	11.1 ± 6.5	58.5
Season 2018	37.0	64.3	51.5 ± 7.3	14.2
Season 2019	11.6	65.2	32.4 ± 9.8	30.2
Season 2020	48.9	139.2	93.7 ± 20.0	21.3

STD – standard deviation; CV – coefficient of variation.

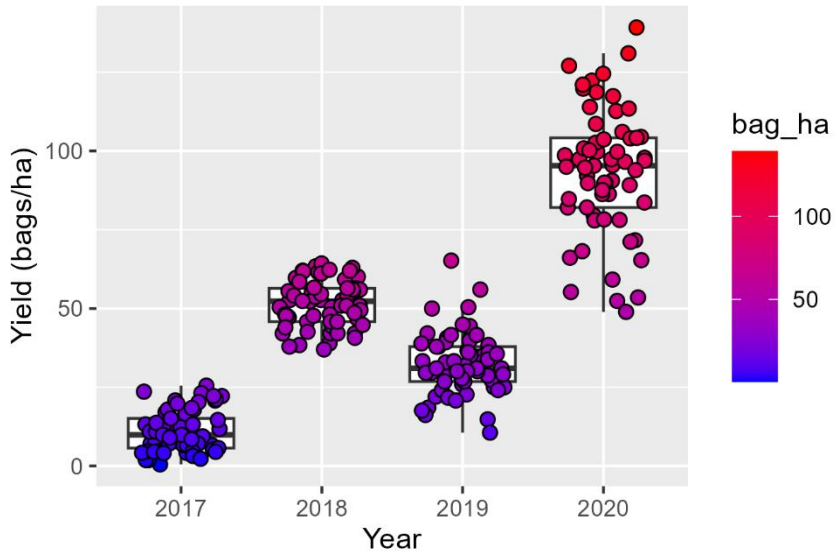

Figure 3. Boxplots of coffee yield (bags ha⁻¹) through harvest seasons.

Yield values and stability are equally crucial for farmers' livelihoods (profitability) and market supply (BACSI et al., 2022). Thus, according to Conab (2024), the average yield reached for the State of Minas Gerais was 23.1, 31.7, 23.7, and 32.2 bags ha^{-1} for 2017, 2018, 2019, and 2020, respectively. The average yield of this study area followed the same biennial trend. Average values above the State average were found, except for 2017. The plants were subjected to severe pruning in 2015, reflected in the yield values by the end of the phenological cycle in 2017. The season of 2020 presented remarkable yields when compared to 2018 (both on positive biennially), whose values were way higher than the regional coffee yield. Proper management might provide increased yield and value stability over time.

3.3.2 Machine learning models accuracy

Table 2 presents the accuracy indexes for all the predictive models. Once RMSE and MAE represent averages of predicted versus estimated dataset, the violin plots (FIGURE 4) were presented in addition to depicting the dispersion of all estimations retrieved by 50 repetitions. Overall, within a similar dataset of training/modeling, the algorithms presented quite similar performances, with slight deviations among RMSE, MAE, and R^2 and the similarity of density curves in the violin plots. Although previous studies have found contrasting performances of different algorithms to predict coffee yield (KOUADIO et al., 2018; BARBOSA et al., 2021), the results from this study suggest that the configuration of model training and validation has driven the accuracy rather than algorithm performance.

Table 2. Accuracy assessment of coffee yield predictive models.

Machine learning algorithm	RMSE	MAE	R^2
Model I - All harvest seasons			
Random forest	12.35	8.79	0.87
Support vector machine	12.46	8.90	0.86
Gradient boosting machine	12.43	8.82	0.86
Model II - Negative biennially			
Random forest	8.13	6.44	0.65
Support vector machine	8.55	6.75	0.62
Gradient boosting machine	8.33	6.62	0.63
Model III - Positive biennially			
Random forest	16.80	12.52	0.60
Support vector machine	16.90	12.75	0.59
Gradient boosting machine	16.54	12.05	0.61
Model IV - Training with negative biennially and test with positive biennially			
Random forest	58.83	49.88	0.36
Support vector machine	60.56	54.76	0.12
Gradient boosting machine	58.32	48.97	0.41
Model V - Training with positive biennially and test with negative biennially			
Random forest	68.47	67.57	0.33
Support vector machine	52.35	52.64	0.00
Gradient boosting machine	54.95	53.37	0.12

RMSE – root mean square of error; MAE – mean absolute error; R^2 – coefficient of variation.

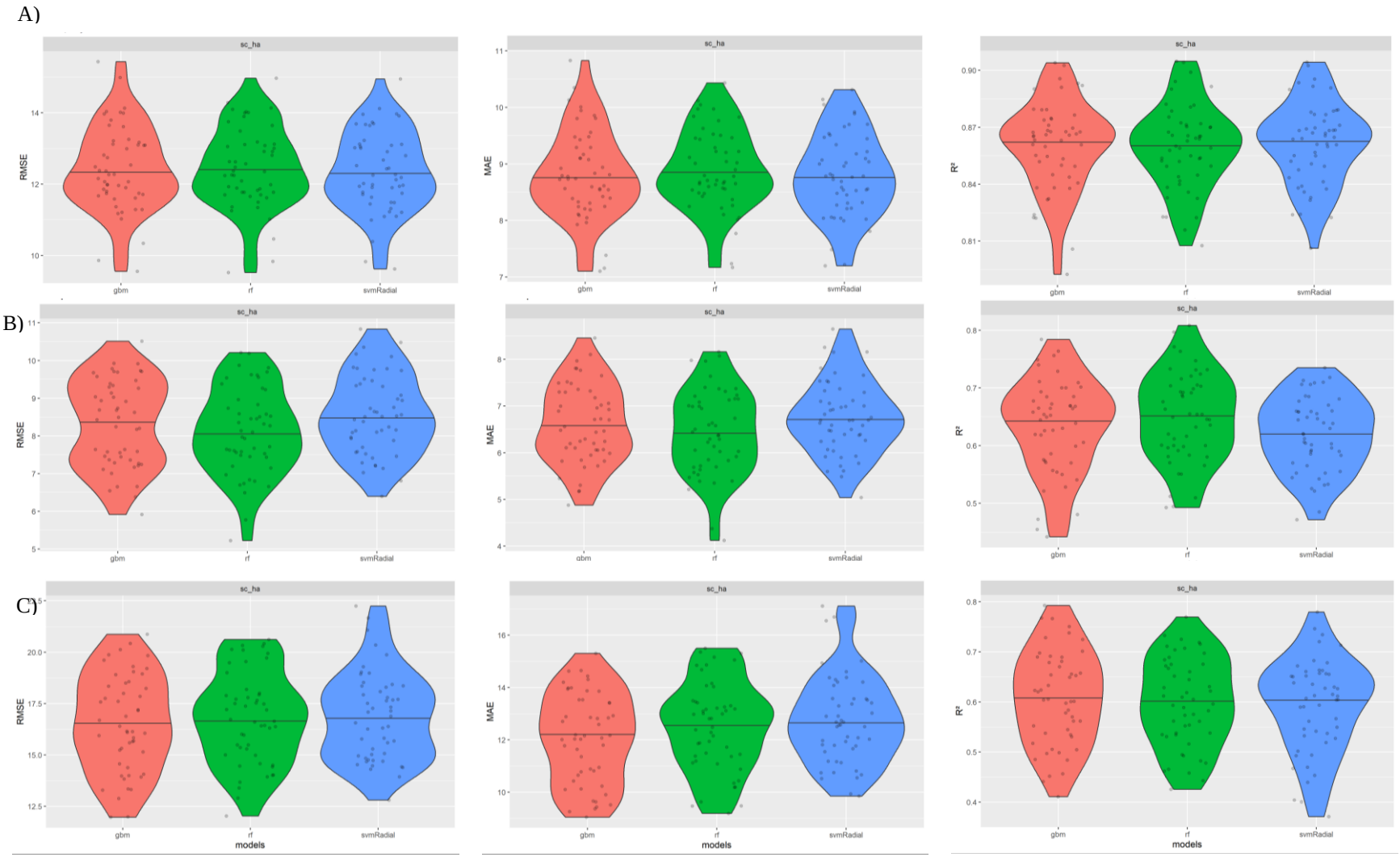

Figure 4. The model's performance obtained from 50 repetitions displaced in violin plots. A) All harvest seasons; B) negative biennially; C) Positive biennially.

Since the predictive models were developed to attend to precision agriculture management, the models should be discussed not only from a statistical accuracy perspective but also from the feasibility of tangible implementation. The range of values used for training models based on one year differs greatly from the next biennial year. Thus, Models IV and V presented the worst performance, being considered unreliable by stakeholders. This fact also shines a light on the precautions that should be taken by managers who are initiating a precision agriculture management program interested in forecasting plant yield, obtaining a dataset from one harvest season, and applying or recalibrating the model based on the subsequent season.

Another practical option would be the development of two different models based only on negative or positive biennially (Models II and III), and both should be applied together for managers to cover the different plant biennial stages. However, Model III had poorer performance prediction, with values of RMSE (16.54 – 16.80 bags ha^{-1}) and MAE (12.01 – 12.75 bags ha^{-1}) twice as high as Model II (RMSE of 8.13 – 8.55 bags ha^{-1}; MAE of 6.44 – 6.62 bags ha^{-1}), and both reaching only a moderate R^2 (from 0.59 to 0.65).

We suggested for end-users the application of Model I that could be applied for any harvest season and presented with proper accuracy. Although yield values contrasted throughout harvest seasons, the algorithm presented excellent capability to predict yield. Concerning the type of algorithms, they performed similarly, with a slightly better performance obtained from the random forest with lower RMSE (12.35 bags ha^{-1}) and MAE (8.79 bags ha^{-1}) and higher R^2 (0.87) among all other models of this study. Also, the GBM model presented extreme or outlier RMSE, MAE, and R^2 values on violin plots (FIGURE 3), which could be related to underestimating or overestimating the predictions.

Random forest algorithm has presented greater accuracy when compared to generalized linear models. Martello et al. (2022b) coupled coffee yield with spectral bands shown through random forest and reached an R^2 of 0.93 and lower prediction errors. Santos et al. (2023) found a precision of 91% with good results based on performance metrics, considering the random forest an effective and versatile machine-learning method compared to other algorithms tested.

3.3.3 Ranking of the most important explanatory variables and interpretations

Figure 5 shows the ranking of the most important explanatory variables of Models I, II, and III. The contrast of explanatory variables ranked for Models II and III (RFE-random forest) is remarkable, which might reveal the effects of biennially. Soil properties related to fertility have increased in importance, being selected more frequently for the construction of most

models for models trained based on plants in positive biennially, whose cause-effects should be further investigated.

The explanatory variables of Model I RFE-random forest will be extensively discussed since it consists of the most reliable and accurate model. This study departed from 41 explanatory variables, and from RFE-random forest only 14 were ranked as most important. Thus, besides the adequacy of accuracy, this model consists of the most parsimonious. The RFE selected a subset of explanatory variables, which is advantageous in promoting increasing accuracy, reducing complexity, and preventing multicollinearity and model over-fitting (Wadoux et al., 2020).

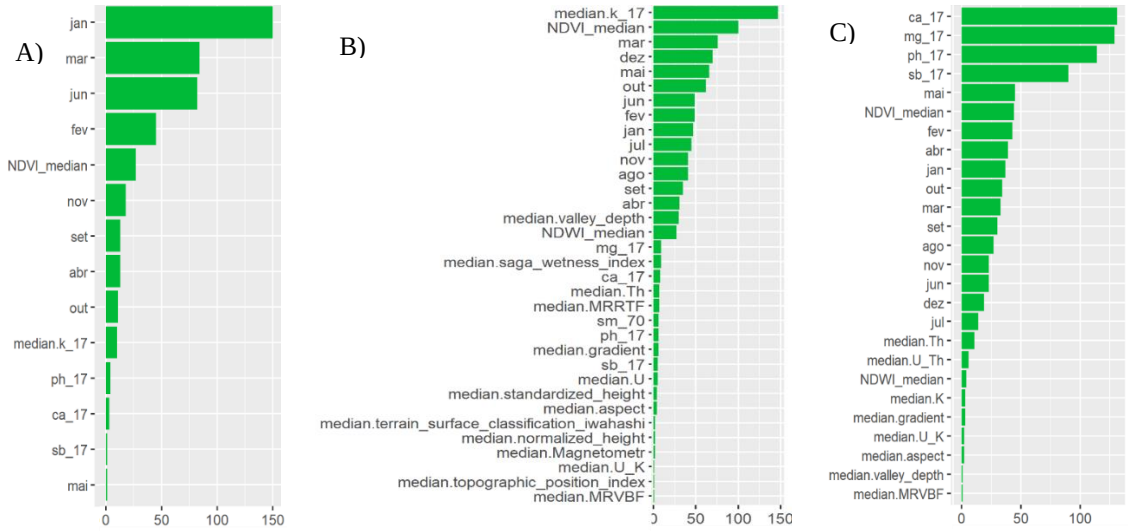

Figure 5. Each round of the model. A) All-seasons; B) Negative biennially; C) Positive biennially.

Machine learning techniques, particularly the random forest algorithm, show promise in pattern recognition. However, the development of knowledge extends beyond mere pattern recognition and successful predictions (WADOUX et al., 2020b). Therefore, it is crucial to analyze the significance of variable explanations. The most important explanatory variables for the most proper model (FIGURE 4a) could be grouped as monthly rainfall, NDVI, and soil fertility (pH, exchangeable Ca, and SB) information.

The total annual rainfall during the years of this study was consistent, with values of 1,447 mm, 1,560 mm, 1,473 mm, and 1,445 mm for the years 2017, 2018, 2019, and 2020, respectively. However, variations in the distribution of rainfall were observed throughout these years, particularly in the months of November and December 2019 and January and February

2020 (FIGURE 6). These variations correlate with the occurrence of positive biennially and greater productivity in the historical series under analysis.

The most frequently selected covariates for building the models were rainfall in the months of January, March, June, and February. Notably, precipitation between November and April exhibited varying frequencies of importance, suggesting a greater sensitivity of the models to this rainy period (Figure 5). This rainy period coincides with the flowering and anthesis phase, as observed by Victorino et al. (2016). Using a predictive modeling approach for municipalities in the Southern region of Minas Gerais through backward selection, they noted a similar sensitivity, reinforcing the significance of water availability during flowering. Aparecido et al. (2017) observed that average coffee yield was influenced by both water and energy variables. They identified that water deficit during flowering was the most crucial variable during periods of positive biennially. Conversely, during times of negative biennially, they noted that yield tends to be affected when the water deficit occurs in April.

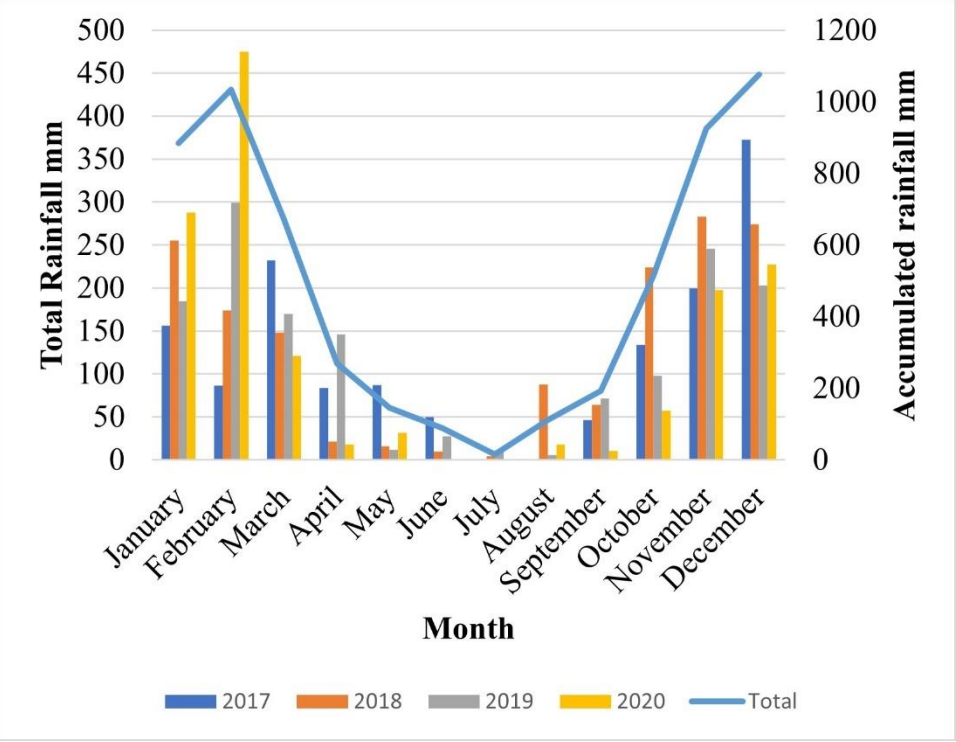

Figure 6. Monthly rainfall distribution through coffee harvest season.

Soil properties were agronomically classified based on technical reports for soil fertility analysis interpretation adapted to the study region (RIBEIRO et al., 2019). Although all mineral elements are involved in different degrees of crop growth and development processes (KOUADIO et al., 2018), we provided further agronomical investigation as a surrogate of

cause-effect analysis once random forest provided an overall interpretation (ranking of the most important variables) lacking local interpretations because the algorithm is not explainable, and decisions are quite complex (so-called black-box method) (GRIMM et al., 2008).

Except for soil exchangeable K^+ (FIGURE 7), all the most important variables presented significative spatial-temporal variability (FIGURES 8, 9, and 10). Soil K^+ is required and accumulated in significant quantities by the coffee plants, and thus, it is massively applied mainly in farms with higher technological levels, such as the commercial farm of this study. K^+ participates in enzyme activation of metabolic processes in the plant, such as photosynthesis, synthesis of proteins and carbohydrates, and maintenance of cell turgidity (MALAVOLTA, 2006). Exchangeable soil K^+, which was classified as "good" in the 2017 harvest season (FIGURE 7A) may have affected yield in 2018. Although positive biennially in 2018 (FIGURE 2), the average values were lower than in 2020 (also positive biennially).

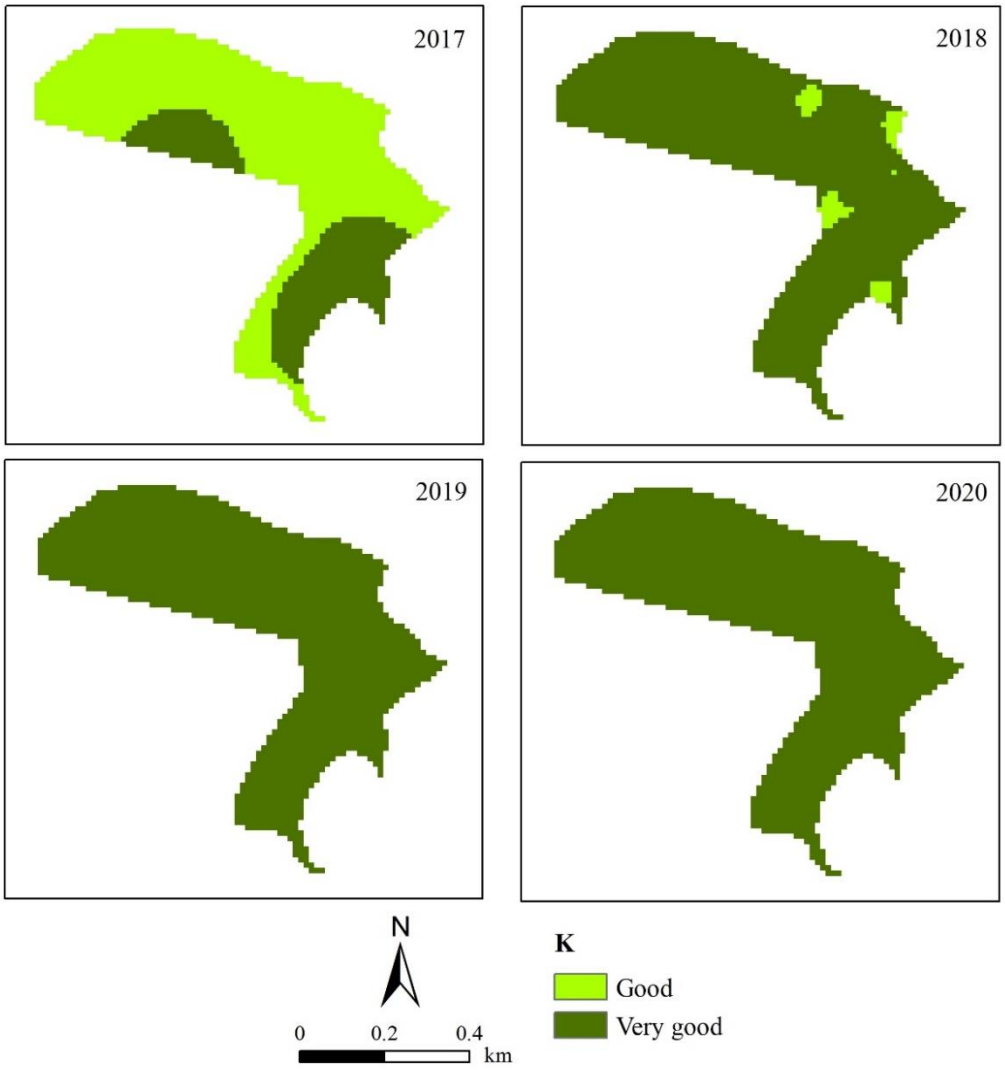

Figure 7. Soil exchangeable K^+ classified agronomically according to Ribeiro et al. (1999). Good: $71 \leq K^+$ content ≤ 120 mg dm^{-3}; very good: K^+ content >121 mg dm^{-3}.

Soil pH presented more significant temporal variability, which might be related to their importance in the model since both low and high values affect plant nutrient availability. The soil pH of the 2017, 2018, and 2019 harvest seasons reflected soil correction with lime since weathered-leached soils tend to be naturally acidic. The consequences of lower soil pH in 2020 might affect yield probably in the next season.

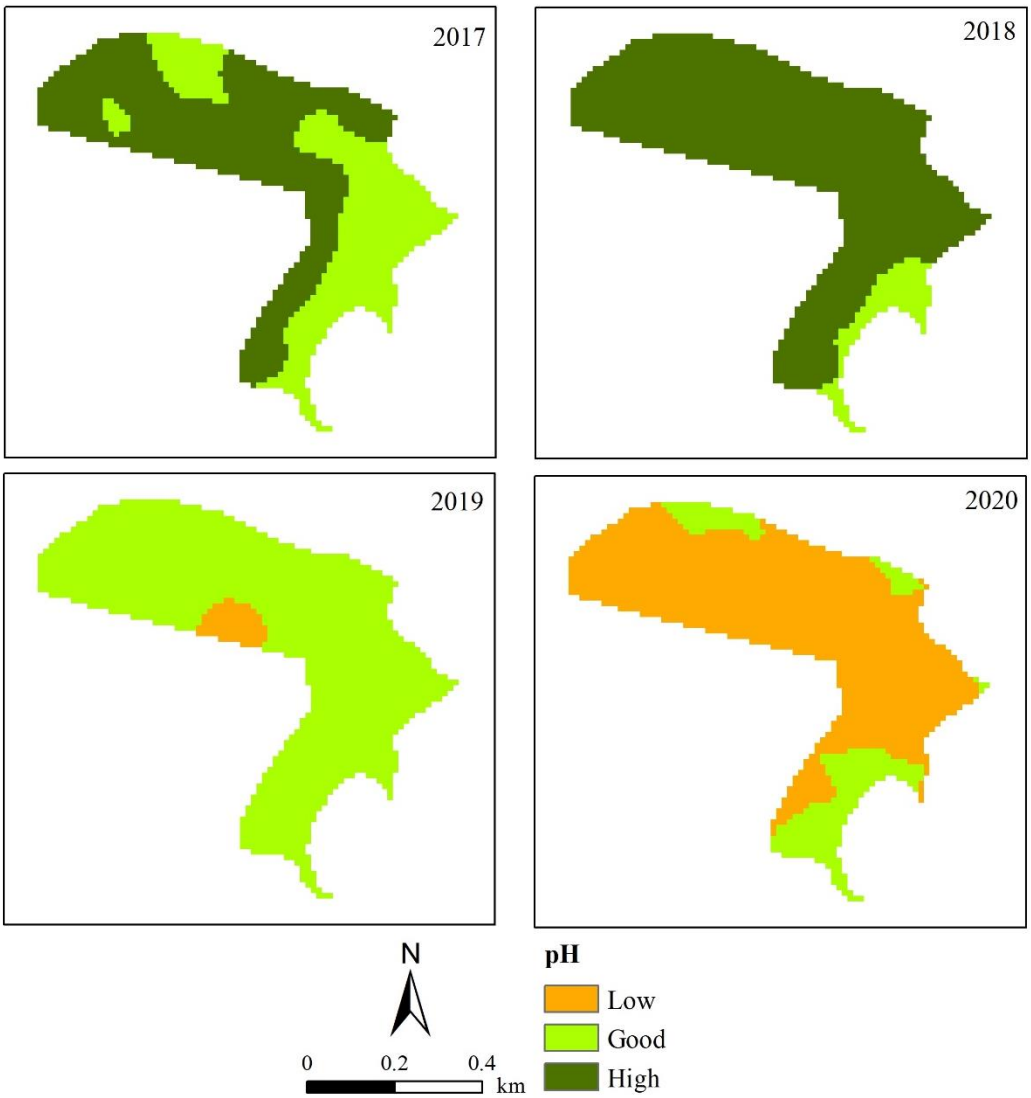

Figure 8. Soil pH classified agronomically according to Ribeiro et al. (1999). Low: 4.5 ≤ pH ≤ 5.5; good 5.6 ≤ pH ≤ 6.0; high: 6.1 ≤ pH ≤7.0.

Calcium is essential for roots and leaf growth due to structural functions in plants and increases tolerance of hydric stress. Another benefit of lime inputs consists of the increase in exchangeable Ca^{2+}. However, except for 2018, the spatial distribution of pH and Ca maps reveals quite a contrasting pattern when compared yearly. Although a spatial similarity between maps was expected, such spatial patterns demonstrated a non-multicollinear model.

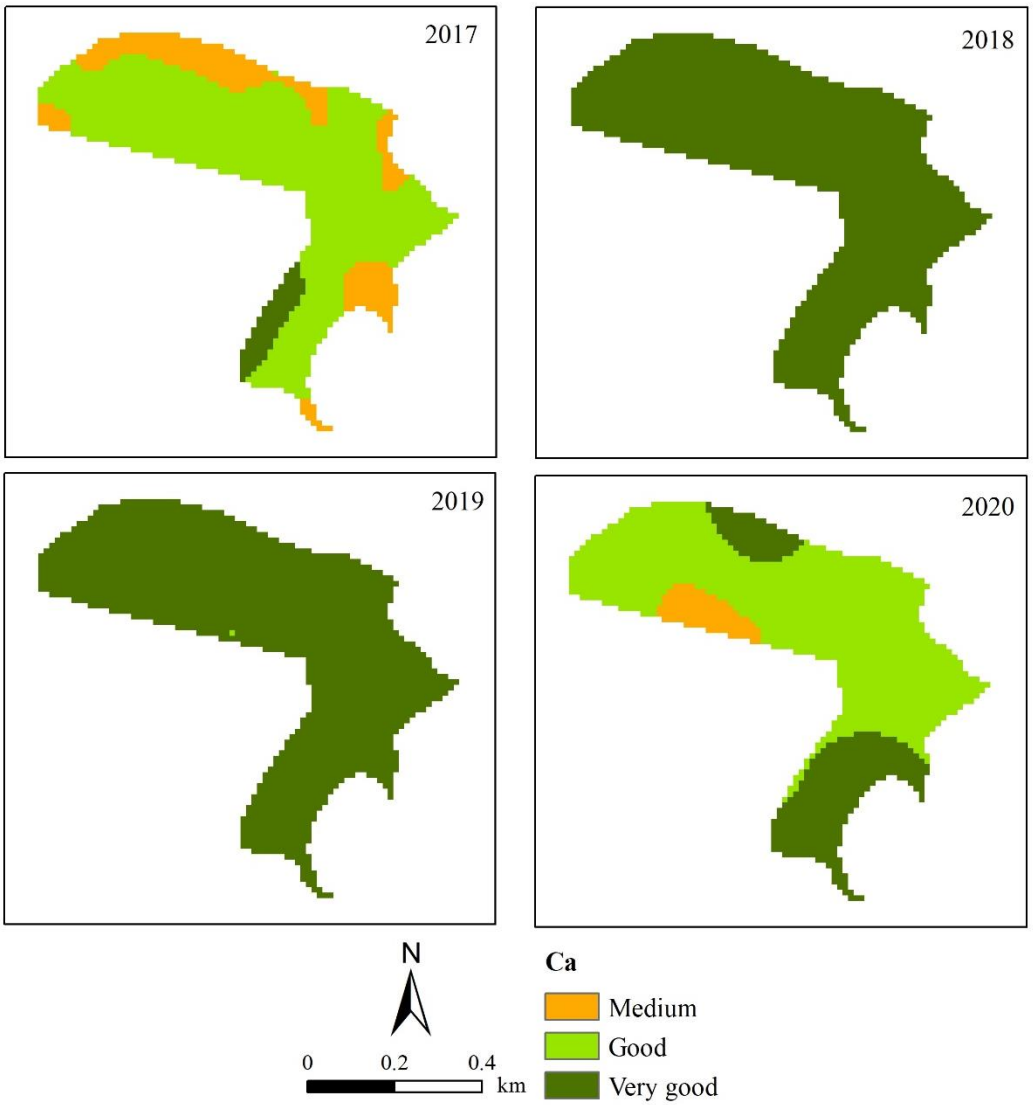

Figure 9. Soil exchangeable Ca^{2+} classified agronomically according to Ribeiro et al. (1999). Medium: $1.21 \leq Ca^{2+} \leq 2.40$ cmolc dm^{-3}; good $2.41 \leq Ca^{2+} \leq 4.00$ cmolc dm^{-3}; high: $Ca^{2+} >$ 4.00 cmolc dm-3.

The SB is used in soil to express exchangeable bases, which is essential for calculating other soil fertility indexes, such as cation exchange capacity and base saturation (AZEVEDO and MANNING, 2023). Considering the most limiting scenario, there is no spatial coincidence of lands where “medium” SB were found, suggesting an effect of variable rate application rather than landscape influence. Once again, there is no similarity of yearly spatial patterns of SB (FIGURE 10) with K^{+}, Ca^{2+}, and Mg^{2+}.

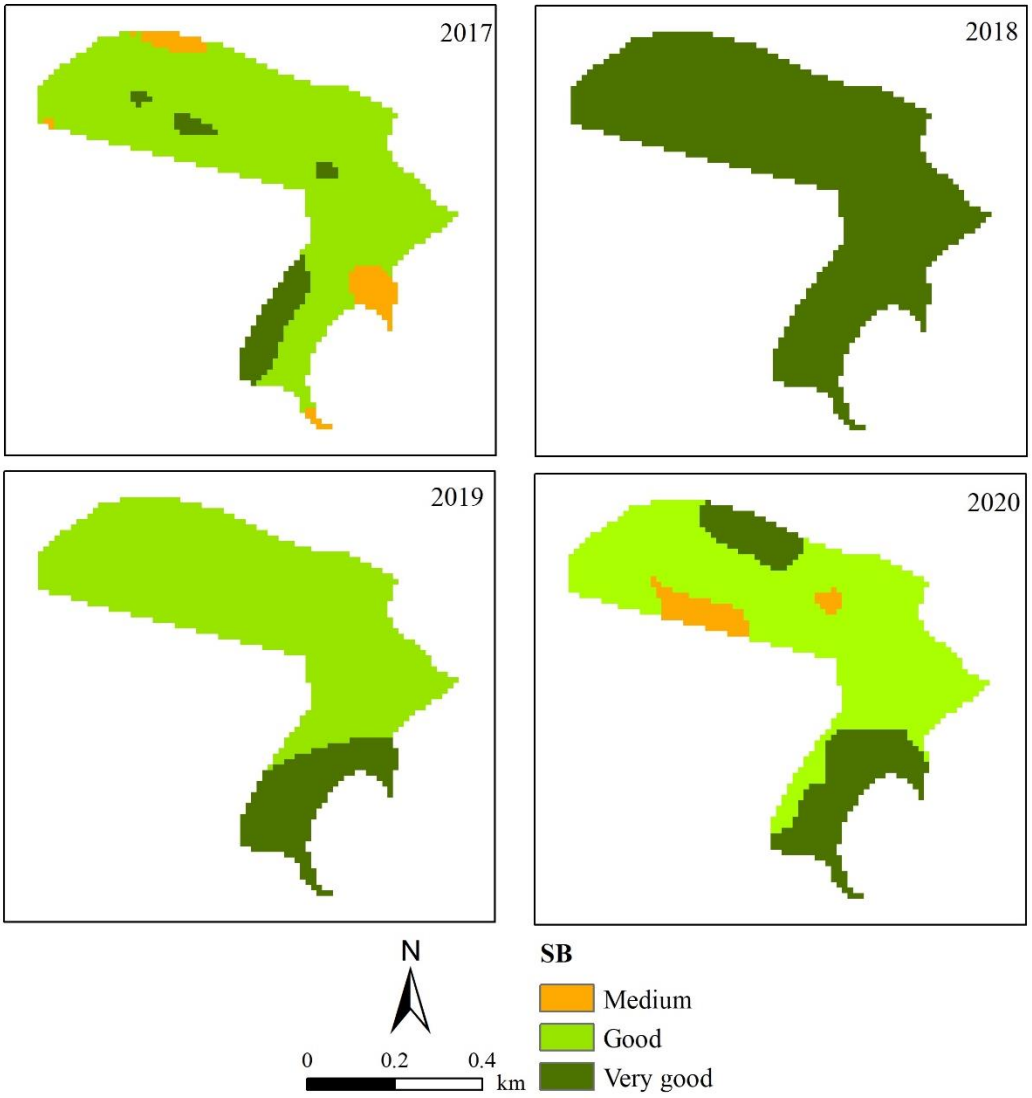

Figure 10. Sum of basis (SB) classified agronomically according to Ribeiro et al. (1999). Medium: 1.81 ≤ SB ≤ 3.60 $cmol_c$ dm^{-3}; good 3.61 ≤ SB ≤ 6.00 $cmol_c$ dm^{-3}; high: Ca > 6.00 $cmol_c$ dm^{-3}.

3.4 CONCLUSIONS

Contrary to the stated hypothesis, coffee yield predictive models from machine learning algorithms containing long-term datasets were more proper than models based only on negative or positive biennially. RFE-random forest proved to be a model with higher accuracy and parsimoniousness that elected meaningful variables for coffee yield prediction. Rainfall, NDVI, and fertility-related soil properties were ranked as the most important variables to forecast yield. Their meaningfulness was consistent with what was expected based on accumulated agronomical knowledge, reinforcing the importance of soil-environment information at farm-scale studies to assist precision agriculture management.

3.5 FINAL CONSIDERATIONS

In this study, the entire area consists of Oxisols, presenting insignificant soil physical and morphological variations across the study area except for soil color, as observed in the detailed soil survey that was carefully performed. Attention should be taken to the future applications of the models developed based on this study for the surrounding region. Gonçalves et al. (2022) reported the soil silt content as one of the most critical attributes to define management zones with great significance on coffee yield in the Southern region of Minas Gerais State. Authors attributed this to the fact that under tropical conditions, silt is a reliable indicator of the degree of soil weathering since contrasting soil types occurred in coffee farms. Thus, one of the strengths of yield models that are machine learning-based is that they are easily updated if a new and significant explanatory variable is discovered. In addition, Aragão et al. (2020) found that biological attributes were capable of discriminating contrasting coffee yield at local scale experiments in Minas Gerais State. Soil quality was strongly related to coffee yield, and biological attributes stand out as a soil quality indicator. Thus, these findings suggest that further yield model improvement should consider biological aspects, although such analyses are still neglected by large or smallholders, according to the authors.

REFERENCES

ABDEL-RAHMAN, E. M.; AHMED, F. B.; ISMAIL, R. Random forest regression and spectral band selection for estimating sugarcane leaf nitrogen concentration using EO-1 Hyperion hyperspectral data. International Journal of Remote Sensing, 34(2), 712–728, 2013.

ALVARES, C. A. et al. Köppen's climate classification map for Brazil. Meteorologische zeitschrift, 22(6), 711-728, 2013.

APARECIDO, et al Agrometeorological Models for Forecasting Coffee Yield. Agronomy Journal, v. 109, n. 1, 2017.

APARECIDO, L. E. D. O., & ROLIM, G. D. S. Forecasting of the annual yield of Arabic coffee using water deficiency. Pesquisa Agropecuária Brasileira, 53, 1299-1310, 2018.

BACA, M. et al. An Integrated Framework for Assessing Vulnerability to Climate Change and Developing Adaptation Strategies for Coffee Growing Families in Mesoamerica. PLoS ONE, 9, e88463, 2014.

BARBOSA, B. D. S. et al. UAV-based coffee yield prediction utilizing feature selection and deep learning. Smart Agricultural Technology, 1, 2021a. https://doi.org/10.1016/j.atech.2021.100010

BARBOSA, J. Z. et al. National-scale spatial variations of soil magnetic susceptibility in Brazil. Journal of South American Earth Sciences, 108, 103191, 2021b.

BENTO, N. L. et al. Characterization of Recently Planted Coffee Cultivars from Vegetation Indices Obtained by a Remotely Piloted Aircraft System. Sustainability, v. 14, n. 3, p. 1446, 2022. https://doi.org/10.5424/sjar/2022203-18808

BERNARDES, T. et al. Monitoring biennial bearing effect on coffee yield using MODIS remote sensing imagery. Remote Sensing, v. 4, n. 9, p. 2492-2509, 2012. https://doi.org/10.3390/rs4092492

BOEHNER, J. AND SELIGE, T. Spatial prediction of soil attributes using terrain analysis and climate regionalisation. In: Boehner, J., McCloy, K.R., Strobl, J. [Ed.]: SAGA - Analysis and Modelling Applications, Goettinger Geographische Abhandlungen, Goettingen: 13-28, 2006.

BREIMAN, L. Random forest. Machine learning, v. 45, p. 5–32. 2001.

CAMARGO, Â. P.; DE CAMARGO, M. B. P. Definition and outline for the phenological phases of arabic coffee under brazilian tropical conditions. Bragantia, v. 60, n. 1, p. 65–68, 2001.

CARVALHO JUNIOR, W. D. et al. Sample design effects on soil unit prediction with machine: randomness, uncertainty, and majority map. Revista Brasileira de Ciência do Solo, 44, 2020. https://doi.org/10.36783/18069657rbcs20190120

CODEMGE. Levantamento Aerogeofísico. Available in http://www.codemge.com.br/atividades-em-destaque/mineracao/levantamento-aerogeofisico/. 11/12/2022.

COELHO, F. F., et al. Digital soil class mapping in Brazil: A systematic review. Scientia Agricola, v. 78, n. 5, p. 1–11, 2021.

COLAÇO, A. F.; MOLIN, J. P. Agricultura de Precisão - Boletim Técnico 02 Piracicaba- SP, 2015.

CONAB. Acompanhamento da safra Brasileira de café. Available at: https://www.conab.gov.br/info-agro/safras/cafe/boletim-da-safra-de-cafe/item/download/51500_05d8a26dc91d95853fb934b03934bc4b. Acessed in January 22, 2024.

CONRAD, O. Module Terrain Surface Convexity / SAGA-GIS Module Library Documentation (v2.2.5), 2012. http://www.saga-gis.org/saga_tool_doc/2.2.5/ta_morphometry_20.html

CONRAD, O. et al. System For Automated Geoscientific Analyses (SAGA) V. 2.1.4, Geosci. Model Dev., V. 8, P. 1991-2007. 2015

COVINO, T. Hydrologic connectivity as a framework for understanding biogeochemical flux through watersheds and along fluvial networks. Geomorphology, v. 277, p. 133-144, 2017. https://doi.org/10.1016/j.geomorph.2016.09.030

CORTES, C.; VAPNIK. V. Support-vector networks. Mach. Learn., 20 pp. 273-297, 1995

CRAPARO, A.C.W., et al. Coffea arabica yields decline in tanzania due to climate change: global implications. Agricultural and forest meteorology, V.207, P.1-10, 2015. J.Agr and For. Met. 2015.

CURI, N. et al. Mapeamento de solos e magnetismo no campus da UFLA como traçadores ambientais. Editora UFLA. 147p. 2017.

DINH, T. L. A., AIRES, F., & RAHN, E. Statistical Analysis of the Weather Impact on Robusta Coffee Yield in Vietnam. Frontiers in Environmental Science, 10, 2022. https://doi.org/10.3389/fenvs.2022.820916

EVERINGHAM, Y. et al. Accurate prediction of sugarcane yield using a random forest algorithm. Agronomy for Sustainable Development, 36(27), 2016.

FANELLI CARVALHO H. et al. The effect of biennially on genomic prediction of yield in arabica coffee. Euphytica, 216(6), 101. 2020. https://doi.org/10.1007/s10681-020-02641-7

FERRAZ, G. A. S., et al. Variabilidade espacial dos atributos da planta de uma lavoura cafeeira. Revista Ciencia Agronomica, v. 48, n. 1, p. 81–91, 2017.

FERRAZ, G. A. S., et al. Variabilidade espacial e temporal do fósforo, potássio e da produtividade de uma lavoura cafeeira. Engenharia Agrícola, v. 32, n. 1, p. 140–150, 2012.

GALLANT, J. C.; WILSON, J. P. Terrain analysis: principles and applications. [s.l.] : John Wiley & Sons, Ltd, 2000.

GALLANT, J. C.; DOWLING, T. I. A multiresolution index of valley bottom flatness for mapping depositional areas. Water resources research, v. 39, n. 12, 2003.

GAO, B. C. NDWI—A normalized difference water index for remote sensing of vegetation liquid water from space. Remote sensing of environment, v. 58, n. 3, p. 257-266, 1996.

GARCIA, G. M., & ORIANS, C. M. Reproductive tradeoffs in a perennial crop: Exploring the mechanisms of coffee alternate bearing in relation to farm management. Agriculture, Ecosystems & Environment, 340, 108151, 2022. https://doi.org/10.1016/j.agee.2022.108151

GAVIOLI, A. et al. Identification of management zones in precision agriculture: An evaluation of alternative cluster analysis methods. Biosystems Engineering, 181, 86–102, 2019.

GONÇALVES, M.G.M. et al. Pedology-based management class establishment: a study case in Brazilian coffee crops. Precision Agric 23, 1027–1050, 2022. https://doi-org.ez26.periodicos.capes.gov.br/10.1007/s11119-021-09873-0

GOOD, S. P.; NOONE, D.; BOWEN, G. Hydrologic connectivity constrains partitioning of global terrestrial water fluxes. Science, v. 349, n. 6244, p. 175-177, 2015. DOI: 10.1126/science.aaa5931

GORELICK, N. et al. Google Earth Engine: Planetary-scale geospatial analysis for everyone. Remote Sensing of Environment, v. 202, 18-27, 2017.

GREGORUTTI, B., MICHEL, B. & SAINT-PIERRE, P. Correlation and variable importance in random forests. Stat Comput 27, 659–678, 2017. https://doi-org.ez26.periodicos.capes.gov.br/10.1007/s11222-016-9646-1

GRIMM, R. et al. Soil organic carbon concentrations and stocks on Barro Colorado Island - Digital soil mapping using Random Forests analysis. Geoderma, 146(1-2):102-113, 2008.

HJERDT, K. N. et al. A new topographic index to quantify downslope controls on local drainage. Water resources research, 40(5), 2004.

HU, J., & SZYMCZAK, S. A review on longitudinal data analysis with random forest. Briefings in Bioinformatics, 24(2), bbad002. 2023

HUSSON F. et al. FactoMineR: Factor Analysis and Data Mining with R. R package version 1.04,< http://CRAN.R-project.org/package=FactoMineR> 2007.

HUSSON, F., LÊ, S., & PAGÈS, J. Exploratory Multivariate Analysis by Example Using R. Boca Raton: CRC Press. 2017.

IDOL, T. W., & YOUKHANA, A. H. A rapid visual estimation of fruits per lateral to predict coffee yield in Hawaii. Agroforestry Systems, 94(1), 81–93, 2020. https://doi.org/10.1007/s10457-019-00370-y

INPE – Instituto Nacional da Propriedade Industrial. Indicações geográficas. Revista da Propriedade Industrial, N° 2603, 24 de novembro de 2020. Available in http://revistas.inpi.gov.br/pdf/Indicacoes_Geograficas2603.pdf

JAYAKUMAR, M., RAJAVEL, M., & SURENDRAN, U. Climate-based statistical regression models for crop yield forecasting of coffee in humid tropical Kerala, India. International journal of biometeorology, 60, 1943-1952, 2016.

KHOSLA, R. et al. Spatial Variation and Site-Specific Management Zones. In M. A. Oliver (Ed.), Geostatistical Applications for Precision Agriculture, p. 195–219, 2010.

KITTICHOTSATSAWAT, Y. et al. Forecasting arabica coffee yields by auto-regressive integrated moving average and machine learning approaches. AIMS Agriculture and Food, 8(4), 1052–1070, 2023. https://doi.org/10.3934/AGRFOOD.2023057

KITTICHOTSATSAWAT, Y., TIPPAYAWONG, N., & TIPPAYAWONG, K. Y. Prediction of arabica coffee production using artificial neural network and multiple linear regression techniques. Scientific Reports, 12(1), 2022. https://doi.org/10.1038/s41598-022-18635-5

KOUADIO, L. et al. Artificial intelligence approach for the prediction of Robusta coffee yield using soil fertility properties. Computers and Electronics in Agriculture, 155, 324–338, 2018. https://doi.org/10.1016/j.compag.2018.10.014

KOUADIO, L. et al. Performance of a process-based model for predicting robusta coffee yield at the regional scale in Vietnam. Ecological Modelling, 443, 2021. https://doi.org/10.1016/j.ecolmodel.2021.109469

KRAVCHENKO, A.N. Influence of spatial structure on accuracy of interpolation methods. Soil Science Society of American Journal, v.67, p.1564-1571, 2003.

KUHN, M. Package ‘caret”. Journal of Statistical Software. 28 (5) 1-26. 2018.

KUHN, M., & JOHNSON, K. Applied Predictive Modeling. New York: Springer. 2013.

LI, Y. et al. Definition of Management Zones for Enhancing Cultivated Land Conservation Using Combined Spatial Data. Environmental Management, 52(4), 792–806, 2013.

MALAVOLTA, E. Manual de nutrição mineral de plantas. São Paulo: Agronômica Ceres, 638p, 2006.

MARTELLO, M. et al. Coffee-Yield Estimation Using High-Resolution Time-Series Satellite Images and Machine Learning. AgriEngineering, 4(4), 888–902, 2022b. https://doi.org/10.3390/agriengineering4040057

MARTELLO, M. et al. Use of Active Sensors in Coffee Cultivation for Monitoring Crop Yield. Agronomy, 12(9), 2022a. https://doi.org/10.3390/agronomy12092118

MULLINS C. E. Magnetic susceptibility of the soil and its significance in soil science—a review. Eur J Soil Sci. 28: 223–246, 1977.

OLAYA, V. A Gentle Introduction to SAGA GIS. Göttingen University, Göttingen, 2004.

OLIVEIRA, R. B. DE et al. Levantamento do tipo de malha amostral, tamanho de área e número de pontos utilizados em análise geoestatísticaII Simpósio de Geoestatística Aplicada em Ciências Agrárias. UNESP, Botucatu-SP, 2011.

PEREIRA, G.W. et al.Smart-Map: An Open-Source QGIS Plugin for Digital Mapping Using Machine Learning Techniques and Ordinary Kriging. Agronomy 12, 1350, 2022. https://doi.org/10.3390/agronomy12061350

POPPIEL, R. R. et al. High resolution middle eastern soil attributes mapping via open data and cloud computing. Geoderma, v. 385, p. 114890, 1 mar. 2021.

QUINTÃO, R.T., BRITO, E.P.Z., BELK, R.W. The taste transformation ritual in the specialty coffee market. Rev. Adm. Empresas 57, 483–494, 2017. https://doi.org/ 10.1590/S0034-759020170506.

RIBEIRO, A. C.; GUIMARÃES, P. T. G.; ALVAREZ V., V. H. (Ed.). Recomendação para o uso de corretivos e fertilizantes em Minas Gerais: 5. Aproximação. Viçosa: Comissão de Fertilidade do Solo do Estado de Minas Gerais, 359p., 1999.

RIBEIRO, B. et al. Sensory evaluation of coffee cultivars in the Campo das Vertentes Mesoregion, Minas Gerais. African Journal of Agricultural Research, 2020. DOI: 10.5897/AJAR2019.14556.

RILEY, S. J.; DEGLORIA, S. D.; & ELLIOT, R. Index that quantifies topographic heterogeneity. Intermountain Journal of sciences, 5(1-4), 23-27, 1999.

RODRIGUES JUNIOR, F. A. et al. Geração de zonas de manejo para cafeicultura empregando-se sensor SPAD e análise foliar. Rev. bras. eng. agríc. ambient. v.15 n.° 8, 2011.

ROSA et al. Estimativa da produtividade de café com base em um modelo agrometeorológico-espectral. Pesq. agropec. bras. 45 (12), 2010.

ROUSE JR, J. W. et al. Monitoring the vernal advancement and retrogradation (green wave effect) of natural vegetation. 1974.

SANTANA, L.S. et al. Advances in Precision Coffee Growing Research: A Bibliometric Review. Agronomy, 11, 1557, 2021.

SANTHOSH, C. S., & UMESH, K. K. An ensemble approach for coffee crop yield prediction based on agronomic factors. ASEAN Engineering Journal, 13(3), 29–38, 2023. https://doi.org/10.11113/aej.V13.18846

SANTOS, S. A. et al. Supervised classification and NDVI calculation from remote piloted aircraft images for coffee plantations applications. Coffee Science - ISSN 1984-3909, [S. l.], v. 16, p. e161978, 2022. DOI: 10.25186/.v16i.1978.

SENA, N. C. et al. Soil sampling strategy in areas of difficult acess using the cLHS method, Geoderma Regional, Volume 24, 2021.

SHARMA, A. et al. Machine Learning Applications for Precision Agriculture: A Comprehensive Review. IEEE Access, vol. 9, p. 4843-4873, 2021.

SILVA, S. de A.; LIMA, J. S. de S.; SOUZA, G. S. de. Estudo da fertilidade de um Latossolo Vermelho-Amarelo húmico sob cultivo de café arábica por meio de geoestatística. Revista Ceres, v. 57, n. 4, p. 560–567, ago. 2010.

SMITH, M.J. GOODCHILD, M.F.; LONGLY, P.A. Geospatial analysis: a comprehensive guide. 2021. Available at https://www.spatialanalysisonline.com/. Accessed in February 2024.

SOTT, M. K. et al. Precision techniques and agriculture 4.0 technologies to promote sustainability in the coffee sector: state of the art, challenges and future trends. IEEE access, v. 8, p. 149854–149867, 2020.

TEIXEIRA, A. F. S. et al. Tropical soil pH and sorption complex prediction via portable X-ray fluorescence spectrometry, Geoderma, v. 361, 2020.

Thao, N. T. T. et al. Early Prediction of Coffee Yield in the Central Highlands of Vietnam Using a Statistical Approach and Satellite Remote Sensing Vegetation Biophysical Variables. Remote Sensing, 14(13), 2022. https://doi.org/10.3390/rs14132975

THOMAZIELLO, R. A.; PEREIRA, S.P. Poda e condução do cafeeiro arábica. Campinas: IAC, 2008. 39p.

TRAVIS, M. R. VIEWIT: computation of seen areas, slope, and aspect for land-use planning (Vol. 11). Department of Agriculture, Forest Service, Pacific Southwest Forest and Range Experiment Station, 1975.

VAN KLOMPENBURG, T.; KASSAHUN, A.; CATAL, C. Crop yield prediction using machine learning: A systematic literature review. Computers and Electronics in Agriculture. Elsevier B.V, 1 out. 2020.

VICTORINO, E. C., DE CARVALHO, L. G., & FERREIRA, D. F. Agrometeorological modeling for coffee productivity forecast in the south region of minas gerais state; [Modelagem agrometeorológica para a previsão de produtividade de cafeeiros na região sul do estado de minas gerais]. Coffee Science, 11(2), 211–220, 2016.

VISCARRA ROSSEL, R. A.; LOBSEY, C. Scoping review of proximal soil sensors for grain growing. CSIRO, Australia, July, p. 52, 2016.

VRINDTS, E. et al. Management zones based on correlation between soil compaction, yield and crop data. Biosystems Engineering, 92, 419–428, 2005.

WADOUX, A.; MINASNY, B.; McBRATNEY, A. B. Machine learning for digital soil mapping: Applications, challenges and suggested solutions. 2020.

WADOUX, Alexandre MJ-C. et al. A note on knowledge discovery and machine learning in digital soil mapping. European Journal of Soil Science, v. 71, n. 2, p. 133-136, 2020b.

WANG, N. et al. Evaluating coffee yield gaps and important biotic, abiotic, and management factors limiting coffee production in Uganda. European Journal of Agronomy, 63, 1–11, 2015. https://doi.org/10.1016/j.eja.2014.11.003

Wilson, J. P., & Gallant, J. C. (Eds.). Terrain analysis: principles and applications. John Wiley & Sons, 2000.

XIE, S. et al. Automatic land-cover mapping using landsat time-series data based on google earth engine. Remote sensing, 11(24), 3023, 2019.

ZANELLA, M. A. et al. Coffee yield prediction using high-resolution satellite imagery and crop nutritional status in Southeast Brazil. Remote Sensing Applications: Society and Environment, 33, 101092, 2024. https://doi.org/10.1016/j.rsase.2023.101092.

APPENDICES

Table 1. Descriptive statistics of auxiliary variables stable over time.

Explanatory variables	Minimum	Mean ± STD	Maximum	CV (%)
Soil parent material				
Magnetic Susceptibility at 70 cm soil depth	5.43	23.98 ± 15.61	55.81	65.12
K (%)	0.01	0.11 ± 0.05	0.22	48.44
Magnetometry	- 49.29	- 11.85 ± 20.43	22.93	-172.37
Th (ppm)	12.04	14.96 ± 1.46	16.94	9.77
U (ppm)	65.21	181.33 ± 106.58	435.66	58.78
Ratio Th/K	0.75	1.15 ± 0.20	1.46	17.59
Ratio U/K	3.51	14.22 ± 9.15	36.41	64.31
Ratio U/Th	0.05	0.08 ± 0.01	0.10	18.31
Terrain attributes				
Aspect	31.03	227.66 ± 110	353.83	48.32
Gradient	0.09	0.23 ± 0.11	0.56	48.65
MRRTF	0.00	0.37 ± 0.88	2.99	241.29
MRVBF	0.00	0.64 ± 0.47	1.92	75.38
Normalized height	0.14	0.56 ± 0.24	0.89	42.74
Real surface area	156.39	157.54 ± 0.84	160.81	0.54
Saga wetness index	2.99	3.94 ± 0.41	4.72	10.50
Slope	2.37	6.93 ± 2.39	13.67	34.49
Standardized height	985.18	999.05 ± 9.82	1020.79	0.98
Terrain ruggedness index	0.45	1.02 ± 0.31	1.93	30.38
Terrain surface classification	65.07	86.59 ± 16.19	113.64	18.70
Terrain surface texture	0.17	2.58 ± 2.47	12.43	95.94
Topographic position index	-5.66	-0.24 ± 2.47	4.01	1013.59
Valley depth	1.00	4.74 ± 3.11	13.42	65.60

STD – standard deviation; CV – coefficient of variation.

Table 2. Descriptive statistics of explanatory variables unstable over time.

Explanatory variables	Minimum	Mean ± STD	Maximum	CV (%)
2017 harvest season				
Ca ($cmol_c\ dm^{-3}$)	1.88	3.00 ± 0.59	4.37	19.75
K ($cmol_c\ dm^{-3}$)	0.24	0.28 ± 0.03	0.33	9.25
Mg ($cmol_c\ dm^{-3}$)	1.06	1.62 ± 0.34	2.63	21.07
pH	5.11	5.97 ± 0.31	6.57	5.25
SB	3.22	4.91 ± 0.84	7.16	17.16
NDVI	479.52	562.72 ± 22.94	602.67	4.08
NDWI	165.31	206.94 ± 14.96	240.65	7.23
2018 harvest season				
Ca ($cmol_c\ dm^{-3}$)	4.96	5.48 ± 0.25	5.85	4.50
K ($cmol_c\ dm^{-3}$)	0.29	0.38 ± 0.07	0.65	17.99
Mg ($cmol_c\ dm^{-3}$)	2.16	2.55 ± 0.21	3.00	8.17
pH	5.87	6.31 ± 0.19	6.55	2.95
SB	7.46	8.42 ± 0.45	9.24	5.33
NDVI	570.70	661.57 ± 23.57	717.92	3.56
NDWI	213.81	286.72 ± 36.86	345.07	12.86
2019 harvest season				
Ca ($cmol_c\ dm^{-3}$)	2.64	3.39 ± 0.52	4.49	15.30
K ($cmol_c\ dm^{-3}$)	0.38	0.42 ± 0.03	0.49	6.02
Mg ($cmol_c\ dm^{-3}$)	1.08	1.38 ± 0.21	1.87	14.94
pH	5.38	5.68 ± 0.13	5.90	2.34
SB	4.13	5.20 ± 0.74	6.86	14.33
NDVI	572.90	661.06 ± 30.91	710.10	4.68
NDWI	180.72	301.05 ± 55.04	398.17	18.28
2020 harvest season				
Ca ($cmol_c\ dm^{-3}$)	2.10	3.39 ± 0.77	5.12	36.42
K ($cmol_c\ dm^{-3}$)	0.26	0.35 ± 0.05	0.47	20.33
Mg ($cmol_c\ dm^{-3}$)	0.86	1.37 ± 0.31	2.13	36.17
pH	4.77	5.23 ± 0.20	5.75	4.11
SB	3.12	5.13 ± 1.18	7.94	37.85
NDVI	420.14	516.21 ± 45.35	653.43	10.79
NDWI	101.43	165.17 ± 41.84	296.21	41.25

STD – standard deviation; CV – coefficient of variation.