UNIVERSIDADE FEDERAL DE LAVRAS

MARIANA GABRIELE MARCOLINO GONÇALVES

SOIL SURVEY AS SUPPORT TO PRECISION COFFEE CROP AND WINTER WINES DEVELOPMENT IN SOUTHEAST BRAZIL

LAVRAS - MG
2020

MARIANA GABRIELE MARCOLINO GONÇALVES

SOIL SURVEY AS SUPPORT TO PRECISION COFFEE CROP AND WINTER WINES DEVELOPMENT IN SOUTHEAST BRAZIL

Thesis presented to the Federal University of Lavras, as part of the requirements of the Postgraduate Program in Soil Science, area of concentration in Environmental Resources and Land Use, to obtain the title of Doctor.

Profa. Dra. Michele Duarte de Menezes
Advisor

Prof. Dr. Nilton Curi
Co-advisor

LAVRAS - MG
2020

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

Gonçalves, Mariana Gabriele Marcolino.
Soil survey as support to precision coffee crop and Winter wines development in Southeast Brazil / Mariana Gabriele Marcolino Gonçalves. - 2020.
118 p. : il.

Orientador(a): Michele Duarte de Menezes.
Coorientador(a): Nilton Curi.
Tese (doutorado) - Universidade Federal de Lavras, 2020.
Bibliografia.

1. soil survey. 2. coffee crop. 3. wine quality. I. Menezes, Michele Duarte de. II. Curi, Nilton. III. Título.

MARIANA GABRIELE MARCOLINO GONÇALVES

SOIL SURVEY AS SUPPORT TO PRECISION COFFEE CROP AND WINTER WINES DEVELOPMENT IN SOUTHEAST BRAZIL

Thesis presented to the Federal University of Lavras, as part of the requirements of the Postgraduate Program in Soil Science, area of concentration in Environmental Resources and Land Use, to obtain the title of Doctor.

APPROVED on August 28, 2020.

Dra. Renata Vieira da Mota	EPAMIG
Dr. Gabriel Araújo e Silva Ferraz	UFLA
Dra. Adélia Aziz Alexandre Pozza	UFLA
Dr. Sérgio Henrique Godinho da Silva	UFLA

Profa. Dra. Michele Duarte de Menezes
Advisor

Prof. Dr. Nilton Curi
Co-advisor

LAVRAS - MG
2020

Aos meus pais, Messias e Fatinha, a meus irmãos, Bruno e Diogo, aos meus avós Antônio e Aparecida, dedico.

AGRADECIMENTOS

Agradeço a Deus pela sabedoria e pelo dom da vida.

À Universidade Federal de Lavras e ao DCS pela excelente oportunidade de desenvolver essa pesquisa. À Fundação de Amparo à Pesquisa de Minas Gerais (FAPEMIG), à pela concessão da bolsa; à Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) e ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) pelo fomento para a realização deste estudo.

À minha orientadora Michele Duarte de Menezes pela atenção de sempre, pela orientação e ensinamentos, pela amizade.

Ao meu coorientador Nilton Curi pelos valiosos ensinamentos.

À Fazenda Itapuã, ao Milton Verdade e Pedro pela oportunidade e apoio no desenvolvimento desse estudo.

Ao pesquisador Murillo Albuquerque Regina pela estimada parceria. À Luciana, pela parceria e ajuda na interpretação dos resultados no decorrer deste estudo. Às pesquisadoras Cláudia Rita de Souza e Renata Vieira da Mota pela estimada ajuda na interpretação dos resultados e parceria na escrita de trabalhos técnico-científicos. Ao professor Alberto Inda pela ajuda nas análises mineralógicas.

Aos técnicos de laboratório Lívia, Geila, Mariene, Dulce, Téo, Bethânia e Alexandre pelo suporte e ajuda nas análises. À querida Dirce, pela ajuda e amizade de sempre.

Às amigas da república: Mari, Iara, Angélica, Marina e Letícia que se tornaram minha família em Lavras. Ao Marcelo, Diego e Fabio pela amizade e pelo apoio nas análises estatísticas e espaciais deste trabalho. Aos amigos que fizeram toda a diferença em todos os momentos: Fernanda, Maila, Gustavo Souza, Luiz, Gustavo Cocate, Luiza, Daniele, Giovana, Yasmmim, Sara, Maria Jéssica, Samara, Cristiane, Aline, Dione, Warley, Raquel, José Ferreira, Ana Paula, Monna Lysa, Jakeline, Ivan, César e os demais colegas.

Agradeço imensamente ao Josimar pelo apoio incondicional, nas discussões de resultados, incentivos, pelos bons momentos e por fazer parte da minha vida.

Agradeço especialmente aos meus irmãos e aos meus pais pela compreensão e por serem meus maiores exemplos de amor. Muito obrigada a todos!

ABSTRACT

Soil surveys provide subsidies for several applications, including decision making on the management of several crops. Thus, this work had as objectives: i) to define a method by means of grouping analysis for the management zones outline based on data from soils and coffee plantations, collected in areas with defined land parcels and adapt this method in a management of culture already implanted, ii) to characterize the soils, the climate, as well as to verify their relation with the composition of the Winter Wines produced in seven commercial vineyards of the Syrah cultivar, iii) search for areas similar to the soil mapping units that include vineyards of the Southern region of Minas Gerais and verify the relationship between the wines and grapes produced in two reference vineyards. In the first part of this study, a series of tests were carried out involving selection of variables by Random Forest, reduction of dimensionality by principal component analysis (PCA) and factor analysis for mixed data (FAMD), ending with the generation of clusters by hierarchical cluster analysis on principal components (HCPC). The most important variables to explain coffee yield and thus compose the management zones outline, classified by Random Forest, were crop age, crop density, silt fraction and soil organic matter content. The PCA explained total variance of 76.1% in the first two dimensions. Three clusters with a statistically significant difference in coffee production were outlined by the HCPC. In general, the following sequence of cluster generated (1→2→3) was found, increasing the crop age, and the content of the silt fraction, and decreasing in the yield and crop density. In the second part of this study, climatic and soil characterization was carried out in seven commercial vineyards, including soil classification, chemical, physical and mineralogical analyses, as well as the identification of the parent material of the vineyard soils. The qualitative profile of wines from the Syrah vine was also characterized. Four groups of vineyards were formed from their similarities in terms of edaphoclimatic characteristics: a) soils with high levels of sand on the surface, in places with high rainfall, originated wines with lower pH; b) Soils with homogeneous clay contents along the profile, in vineyards with high thermal amplitude, presented intermediate values for most wine compounds; c) shallow and young soils with high sand content, in a vineyard with low precipitation and high temperature, produced wine with the highest flavanol content. This wine also has high levels of most other evaluated compounds; d) deep soils, with basalt as the parent material, is related to wines with the highest levels of most compounds, however, this is due to the late harvest carried out in this vineyard. The edaphoclimatic conditions were important for the characterization of the typicity of the wines, and such conditions associated with the handling of double pruning allowed the production of quality wines, compared to wines from world traditional wine-growing regions. The third part of this work involves the extraction of climatic, geological and terrain information from soil mapping units that contain commercial vineyards in two municipalities in Minas Gerais. Such information was applied, using fuzzy logic and similarity vectors in an area of interest (provenance area), in order to verify areas with greater similarity in relation to the conditions extracted from the mapping units (soil map). Most of the variables of the two mapping units, mainly the mean temperature, rainfall and evapotranspiration are very similar. The mapping units differ mainly in terms of the higher altitude in Três Corações and by the different parent material of the soil.

Keywords: yield, hierarchical cluster on principal components, principal component analysis, wine quality, fuzzy logic

SUMMARY

1 FIRST PART

GENERAL INTRODUCTION

Soil surveys provide valuable information for soil management and decision making, including agriculture, livestock and silvicultural purposes, as well as or for non-agricultural activities (road allocation, waste disposal, mineral resource extraction, and recreation) (LEE; GRIFFITHS , 1986). Soil survey encompasses the soil identification based on its physical, chemical and morphological characteristics, including its occurrence throughout the landscape, and interpretations according to the intended objectives (RESENDE et al. 2014; ALPHEN; STOORVOGEL, 2000; IBGE, 2015). Thus, it is notorious the large amount of information obtained from a soil survey, which the soil map is one of main products, representation of the soil-landscape relationship, from the outline of the mapping units containing taxonomic units (soil type), and occasionally, phases of relief, stoniness, rockiness or vegetation.

Two important activities in the southeastern region of Brazil are coffee (ABIC 2017; CONAB 2019) and viticulture (AMORIM et al. 2005). The first one is a well consolidated activity especially in the southern Minas Gerais, being responsible for approximately 25% of national production (CONAB, 2019). The second activity is strongly evolving, made possible by a specific management of the vines called “double pruning”. This technique allows grapes to mature and harvest in winter, resulting in a higher concentration of solutes in berries, and harvests carried out in better phytosanitary conditions (FAVERO et al., 2011).

Considering the multiple interpretations from soil surveys, management zones establishment for assisting precision agriculture has great potential. Management zones are an effective and sustainable basis for the management of coffee growing, defined as: sub-regions that have relatively homogeneous soil-landscape characteristics (HAGHVERDI et al. 2015); or sub-areas with a combination of factors limiting plant yield (VRINDTS et al. 2005), generally defined by different variables (GAVIOLI et al. 2019). Although the establishment of management zones promotes increases in plant yield and environmental sustainability (ADHIKARI et al., 2009), results showing the feasibility of soil survey applications in its delineation are scarce in the scientific community. Often, the establishment of management zones is carried out from a large number of sampling points for fertility analysis purposes, which is quite costly. Thus, soil survey becomes an option for the design of management zones in agricultural crops.

While some worldwide regions have many years of tradition in fine wines production, such as the Bordeaux and Champagne region in France, Campania in Italy and Rioja and Galicia in Spain, new regions face the challenge in defining their characteristics in terms of product development (JONES et al., 2004). In this sense, information from the soil survey is of great relevance, since the characterization of soils adds value to the wine produced (WHITE, 2015). In addition, soil attributes are strong influencers in the development and typicity of wines (MORLAT; BODIN, 2006; PRIORI et al., 2019; VAN LEEUWEN et al., 2018). Although relatively recent, the wines produced from the Syrah cultivar in the South of Minas Gerais and northwest of São Paulo have their quality recognized by consumers and specialists, awarded nationally and internationally (BRANT; FIGUEIREDO; MOTA, 2018).

Regarding the evolution of fine wines production in the Southeast region of Brazil, it is also important to search for areas with greater production potential, taking into account soil-environmental characteristics. In this sense, fuzzy logic provides the integration of different layers of digital information, such as satellite images representing climatic or relief data, as well as geology and soil information. Also, when combined with similarity vectors (SHI, 2013), presents great potential to define areas for vine cultivation expansion, by extracting geographical information from commercial vineyards (here considered as successful cases), formalizing, and finding similar areas with low uncertainties.

This work aimed to apply the information obtained from soil surveys to assist the management of coffee crops, and to verify the relationships between soil characteristics and the quality of grape wines. For that, this dissertation is composed by three chapters in scientific articles format. The first article is entitled *Management zones quantitatively established from soil survey and crop management information: a study case in Brazilian coffee crops*, devoted to the definition of coffee management zones in commercial crops using semi-detailed soil survey and crop management information in the south of Minas Gerais, using machine learning techniques, cluster analysis, and principal components generated from two different methods. The second article entitled *Soil-environment and Syrah winter wine characterization to assist viticulture in southeastern Brazil* presents the characterization of soils, climate, and wines produced in the southeastern region of Brazil. The third article is entitled *Searching for similar terroir conditions for Syrah cv. with fuzzy logic and definition of potential new areas for viticulture in Minas Gerais, Brazil* aimed to search for similar areas to two soil mapping units that surround vineyards intended for the production of Syrah winter wines.

REFERENCES

ABIC. 2017. <http://abic.com.br/cafe-garante-recorde-as-lavouras-em-minas/> acessado em 20/08/2018

ADHIKARI, K. et al. **Site Specific Land Management: General Concepts and Applications**. 1st. ed. Luxembourg: European Commission, 2009.

ALPHEN, B. J. VAN; STOORVOGEL, J. J. A Functional Approach to Soil Characterization in Support of Precision Agriculture. **Soil Science Society of America Journal**, v. 64, p. 1706–1713, 2000.

AMORIM, D. A. et al. Produção extemporânea da videira, cultivar syrah, nas condições do sul de Minas Gerais. **Revista Brasileira de Fruticultura**, v. 4, p. 327–331, 2005.

BRANT, L. A. C.; FIGUEIREDO, G. M. DE; MOTA, R. V. DA. Vinhos de Inverno do Sudeste Brasileiro. **Territoires du vin**, v. 9, p. 1–4, 2018.

CONAB. Acompanhamento da safra brasileira de café- *Primeiro levantamento. Companhia Nacional de Abastecimento* (Vol. 6), 2019. http://www.conab.gov.br

FAVERO, A.C. et al. Double-pruning of 'Syrah' grapevines: a management strategy to harvest wine grapes during the winter in the Brazilian Southeast. **Vitis**, v. 50, p. 151-158, 2011.

GAVIOLI, A., et al. Identification of management zones in precision agriculture: An evaluation of alternative cluster analysis methods. **Biosystems Engineering**, v. 181, p. 86–102, 2019. doi:10.1016/j.biosystemseng.2019.02.019

HAGHVERDI, A., et al. Perspectives on delineating management zones for variable rate irrigation. **Computers and Electronics in Agriculture,** v. 117, p. 154–167, 2015.

IBGE. **Manual Técnico de Pedologia**. 3a ed. Rio de Janeiro: Instituto Brasileiro de Geografia e Estatística. v. 3, 2015.

JONES, G. V.; SNEAD, N.; NELSON, P. Geology and Wine 8. Modeling Viticultural Landscapes: A GIS Analysis of the Terroir Potential in the Umpqua Valley of Oregon Gregory. **Geoscience Canada**, v. 31, 2004.

LEE, E. M.; GRIFFITHS, J. S. The importance of pedological soil survey in land use planning, resource assessment and site investigation. **Geological Society,Engineering Geology Special Publications**, v. 4, n. 1, p. 453–466, 1986.

MORLAT, R.; BODIN, F. Characterization of viticultural terroirs using a simple field model based on soil depth - II. Validation of the grape yield and berry quality in the Anjou vineyard (France). **Plant Soil**, v. 281, p. 55–69, 2006. https://doi.org/10.1007/s11104-005-3769-z

PRIORI, S. et al. Scale effect of terroir under three contrasting vintages in the Chianti Classico area (Tuscany, Italy). **Geoderma** v. 334, p. 99–112, 2019.

https://doi.org/10.1016/j.geoderma.2018.07.048

RESENDE, M et al. **Pedologia: Base Para a Distinção de Ambientes**. Editora UFLA, Lavras. 6. ed. Lavras: UFLA, 322 p. 2014.

SHI, X. **ArcSIE User's Guide**. In *Spatial Inference Enterprise*, 2013.

VAN LEEUWEN, C.; ROBY, J.-P.; DE RESSÉGUIER, L. Soil-related terroir factors: a review. **OENO One**, v. 52, p. 173–188, 2018.

VRINDTS, E. et al. Management zones based on correlation between soil compaction, yield and crop data. **Biosystem Engeenering**, v. 92, p. 419–428, 2005.

WHITE, R. E. **Understanding Vineyard Soils**. 2nd. ed. New York: Oxford University Press, 2015.

2. SECOND PART – ARTICLES

ARTICLE 1

This article was submitted to the Precision Agriculture journal

Management zones quantitatively established from soil survey and crop management information: a case study in Brazilian coffee crop

Abstract

In Brazil, coffee crop is separated into land parcels, guided by visual analysis of relief and/or issues regarding farm logistics, in which the management is homogeneous. Management zones can be defined in an automated way, considering soil and crop management characteristics that affect coffee yield. Considering the importance of management zones and the peculiarity of Brazilian coffee crop, the objective was to define a method for management zones outline based on data collected in areas with defined land parcels, and to adapt this method in a crop management already implemented. Two initial datasets were used based on soil survey and/or coffee crop management information. Eight tests were developed, involving: ranking of the most important variables for coffee yield variations by Random Forest, reduction of data dimensionality through principal component analysis (PCA) or factorial analysis of mixed data (FAMD), generation of clusters with the hierarchical cluster on principal component (HCPC) and, evaluation of clusters generated by the statistical difference in yield. The most important variables ranked by Random Forest were crop age, crop density, silt fraction, and soil organic matter content. The PCA explained variance of 76.1% in the first two dimensions. Three clusters with statistically significant difference in coffee yield were outlined by HCPC, also in accordance with the general knowledge of soil-landscape and coffee management characteristics of the study region. In general, it was found the following sequence of clusters generated (1→2→3), increasing in crop age and silt fraction content, and decreasing in crop yield and crop density.

Keyword: hierarchical cluster on principal components; factor analysis for mixed data; principal component analysis; Random Forest; variables selection

Introduction

Brazil is responsible for the largest coffee production in the world, being Minas Gerais state responsible for 18% of this total, with almost only *Coffea arabica* L. (CONAB 2019). Considering environment, economic, and social characteristics, Minas Gerais is separated into different traditional coffee producer regions (Barbosa et al., 2010), in which the southern one has obtained great notoriety due to its highest coffee production (CONAB 2019). Considering their peculiarities, the coffee farms are traditionally separated into land parcels, defined through visual analysis of relief and/or issues regarding farm logistics. The average yield is then measured in each land parcel (one average value assigned for the whole parcel polygon), where soil and crop have been homogeneously managed. It is common for the cultivation of different coffee varieties in the same farm, containing different crop ages and densities. Thus, such variability of characteristics should be taken into consideration in precision agriculture implementation, especially for management zones outline.

Management zones could serve as a sustainable and effective basis for localizing coffee management, defined as subareas with relatively homogenous soil-landscape attributes (Haghverdi et al. 2015), or with relatively combination of yield-limiting factors (Vrindts et al. 2005), generally defined from different variables or attributes (Gavioli et al. 2019). It is wholesome that management zones as well as the variables used to define them must be stable over time, and correlated to yield. Besides, should also be simple and have a low cost of acquisition (Khosla et al. 2010; Li et al. 2013). Although cluster analysis is mostly used for delineating management zones (Gavioli et al. 2019), its application in precision coffee production is scarce.

Considering such practical aspects abovementioned, machine learning tools could assist management zones delineation by means of yield prediction. One possibility is the Random Forest (Breiman 2001) application, when combined with Recursive Feature Elimination (RFE) algorithm, decreasing even more the bias of prediction (Gregorutti et al. 2016). They present great potential not only to predict plant yield (Everingham et al. 2016; Smidt et al. 2016), but also to select (redundancy reduction) (Abdel-Rahman et al. 2013) and to rank the most important variables of prediction. With respect of dimension or redundancy reduction, principal component analysis (PCA) (Schemberger et al. 2017) has been applied in Precision Agriculture, contributing to summarize the main principal sources of variability in the datasets (Moral and Serrano 2019) in a new set of uncorrelated variables (King et al. 2005). Likewise PCA, factorial analysis of mixed data (FAMD) contributes to reduce the redundancy, not only for qualitative data, but also for quantitative ones (Lê et al. 2008). Such a technique has not yet been applied in precision agriculture.

By selecting the most important variables, or by reducing the data dimension, Random Forest, PCA, and FAMD contribute to the simplicity of management zones composition and data mining effectiveness, helping out the selection of variables related with crop yield. Thus, their combinations with cluster methods present great potential to improve even more the management zones outline. In addition, testing different cluster methods is important to ensure effectiveness, since different methods tend to generate different sizes and formats of zones (Gavioli et al. 2019). Keeping in mind the necessity of cluster methods advances, this work applied hierarchical cluster on principal component (HCPC) (Husson et al. 2017), a technique not yet applied for management zones outline purpose. Considering the most common cluster analysis, the strategies are based on hierarchical or partitioning cluster. HCPC has as an advantage the combination of three multivariate analysis: principal component (including PCA, FAMD), hierarchical clustering, and partition clustering (k-means method) (Josse and Husson 2016).

In respect to variables used for management zones delineation, the characteristics and crop requirements should be taken into consideration. Coffee is a long term perennial crop in which the root system explores greater soil depths (Ronchi et al. 2015), mainly in regions with defined dry periods, which is by far the dominant condition in Brazil. Thus, soil information in depth is important to understand better not only the fertility status, but also the root system development and water availability. The latter is an important characteristic, since one of the main risks listed for Brazilian agriculture is drought (Arias et al. 2015). Soil survey naturally supplies soil profile information. Subsurface characteristics are the main criteria for soil classification (Soil Survey Staff 2014) because they are less altered by the anthropic action. Although very informative, only a few studies applying soil survey information to management zones outline are found worldwide (Nawar et al. 2017).

Considering the well-recognized role of management zones in terms of agriculture sustainability as well as the coffee farms peculiarities, we hypothesize that based on the less subjectivity of data mining tool algorithms, not only parameters of soil survey, but also crop management could assist management zones outline in coffee crop. Thus, the objective of this study was to define a method for management zones outline based on data collected in areas with defined land parcels, and to adapt this method in a crop management already implemented. Considering coffee crop peculiarities and soil survey information, both were used as source of possible variables to compose management zones. Dimension reduction and variables selection were performed by means of Random Forest, PCA and FAMD prior to HCPC – a method for generating cluster. A total of 8 different sequences of tests were performed, ending with Tukey test performance, since statistical differences on crop yield are desirable to check the effectiveness of methods.

Material and methods

Study area

The study was carried out in coffee production farms in the municipalities of Alfenas, Alterosa, and Fama, Southern Minas Gerais State, Brazil (Fig. 1). According to Koppen's classification system, the climate is classified as Cwa, characterized by rainy and warm summers, and cold and dry winters (Alvares et al. 2013). The mean annual precipitation is 1,400 mm, the mean temperature in the coldest months (from May to July) is 18°C, while in the warmer months the average is 22°C (from December to February) (ANA, 2018). During the rainy season, short periods of drought are very common. The soils were formed mainly from gneiss (Zogheib et al. 2015), with soil texture varying from loam to clay, with slopes ranging from gentle undulated (slope of 3-8%) to strongly undulated (> 45% of slope). The total area includes six farms, totalizing 1,408.1 ha of coffee crop (*Coffea arabica*).

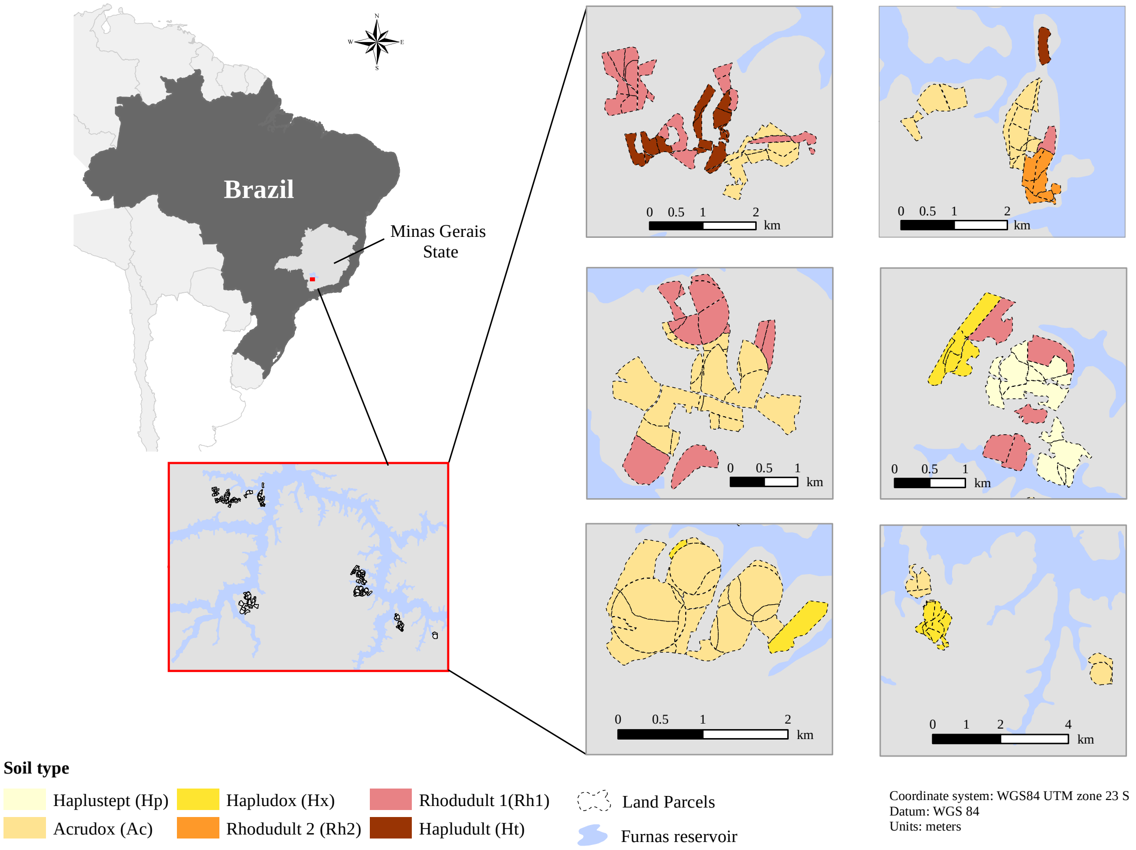

Fig. 1 Study area location and respective soil type maps of coffee farms, Minas Gerais State, Brazil.

In order to achieve the most suitable management zones outline, different tests regarding data input and the type of statistical method were developed. The complete flowchart as well as the statistical analysis assessment are presented in Fig. 2. The whole statistical analysis was developed in a freely available R software (R Core Team 2018).

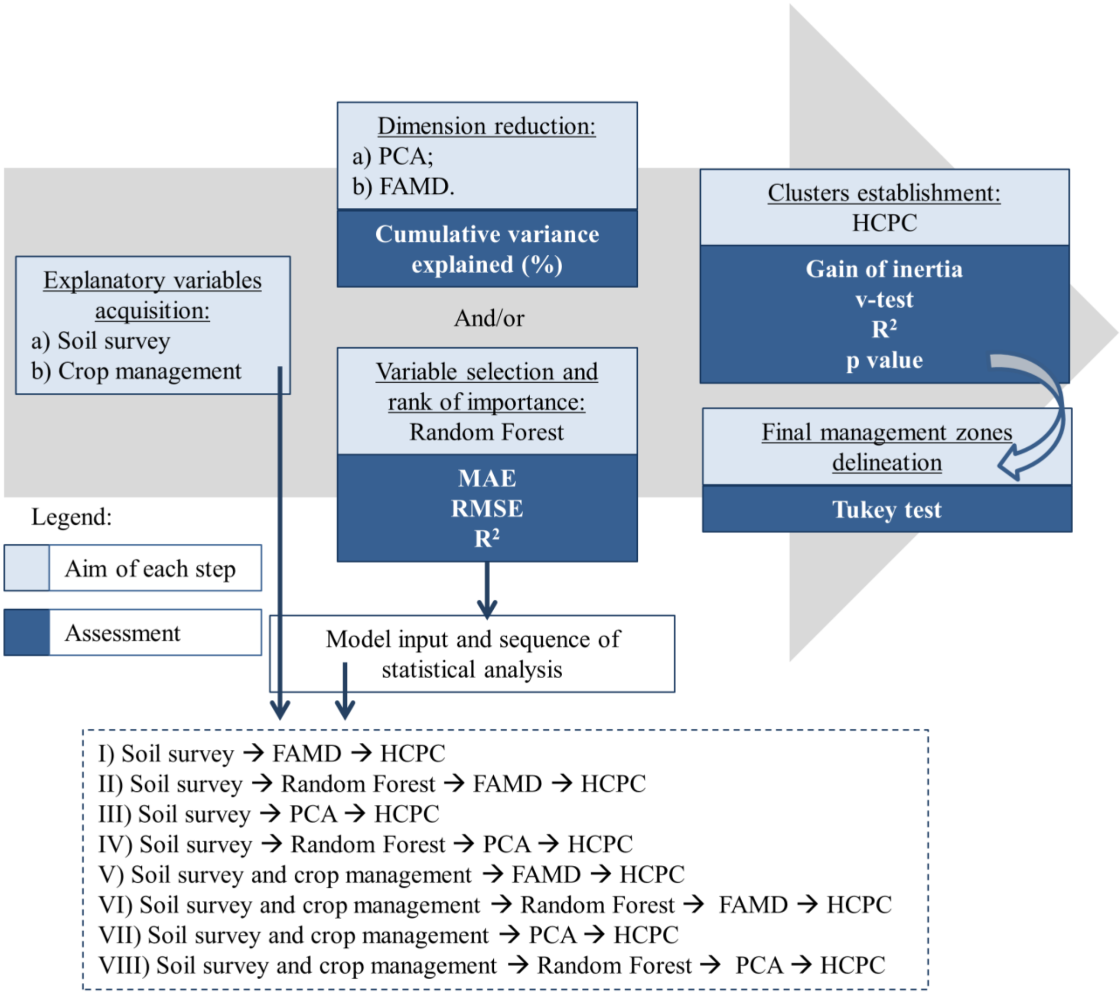

Fig. 2 Complete flowchart of the study: from data acquisition to final management zones assessment. FAMD: factor analysis for mixed data; PCA: principal component analysis; HCPC: hierarchical cluster on principal component; MAE: mean absolute error; RMSE: root mean square error; R^2: determination coefficient.

Explanatory variables acquisition

The explanatory variables or measurement methods considered in this study were those that potentially aid information for management zones establishment, influencing on coffee yield: a) soil survey; b) crop management. Details about the dataset are presented further.

a) Soil survey

With the aid of terrain attributes layers, an intensive field work generated a semi-detailed soil survey at a 1:25,000 scale (IBGE 2015). One of the main products of a soil survey is the soil map in a polygon-format, which in turn is composed of map units with taxonomical classes assigned. For such purpose, a total of 39 locations were: i) described *in situ* (morphological characterization) and sampled at the depths of 0-0.2 m, 0.4-0.7 m and 1.0-1.5 m (when possible) (Santos et al. 2015); ii) analyzed in a laboratory for chemical and physical characterization according to Embrapa (1997) and; iii) classified up to the great group taxonomic level, according to US Soil Taxonomy (Soil Survey Staff 2014). After the soils were classified, their spatial variability was delineated (soil mapping units) based on Brazilian technical reports (IBGE 2015) and Geographical Information System (GIS), and overlaid with slope maps to compose soil phases (additional information in soil taxa name), establishing a sound soil-landscape relationship. The slope was calculated from the satellite Alos1-Palsar Digital Elevation Model in a GIS software SAGA-GIS 7.1.0 (Conrad et al. 2015).

The following chemical analyses were carried out: pH in water; available P and K, extracted with Mehlich-1; exchangeable Ca^{2+}, Mg^{2+} and Al^{3+}, extracted with 1.0 mol L^{-1} KCl ; potential acidity ($H^+ + Al^{3+}$), extracted with 0.5 mol L^{-1} calcium acetate at pH 7.0; and soil organic matter (SOM) determined by Walkley and Black (1934) method; Remaining phosphorus (Alvarez et al. (2000); Cation exchange capacity at soil pH (effective CEC), cation exchange capacity at pH 7.0 (CEC pH 7.0), base saturation (BS), and aluminum saturation (AS) were then calculated.

Soil texture analysis was carried out by the pipette method (Gee and Bauder, 1986). The average of the quantitative soil properties was calculated considering all the depths collected, aiming to express general soil properties at greater depths, which is especially advantageous for long term perennial plants as coffee crop.

b) Crop management

Considering the crop management adopted in the study areas, information about plant density, crop age, coffee variety, and irrigation or no irrigation areas were also considered as explanatory variables, since they influence crop yield. These analyses considered the coffee yield of the 2016/2017 harvest. The inclusion of this information could also guide the next cultivations because, except for irrigation, the other crop management could be considered immutable throughout coffee growth stages. All the information was overlaid in GIS. Liming was applied in whole area. Fertilization was carried out following applications of phosphorus and slow release organic mineral compost to the planting furrow at a depth of 0.9 m and 0.6m, respectively. Phosphorus cover fertilizations were carried out between August and September. Nitrogen and potassium were applied in two or three installments from October to December.

Data analyses and statistical methods

In order to establish more accurate management zones and to make the methodology adaptable to different environmental conditions and crop management, different statistical methods and paths of analysis were explored, promoting variable selection, a rank of variables importance and reduction of dimension prior to cluster analysis. A total of eight different model inputs and sequences of statistical analyses were tested (model inputs I, II, III, IV, V, VI, VII, and VIII). More details are provided further.

Variables selection and rank of importance from Recursive Feature Elimination and Random Forest analysis

Many variables obtained from soil survey or crop management an influence crop yield (Rena et al. 2002; Tisdale et al. 1993; Chlingaryan et al. 2018). However, using all these variables in the definition of management zones would make their interpretation very complex or bring error or noise to the analysis. Thus, in order to select important variables that contribute most on crop yield, Random Forest analysis was performed with Recursive Feature Elimination (RFE) via R package caret (Classification and Regression Training) (Kuhn 2018) in R software 3.4.4.

Random Forest is an algorithm capable of handling multivariate data containing numeric and categorical variables, developed for both classification and regression analyses (Breiman 2001). The use of Random Forest associated with RFE is efficient in variables selection, presenting models suitable accuracy (Gomes et al. 2019; Menezes et al. 2020). RFE is an algorithm that iteratively eliminates the combinations of the least promising variables for the prediction model (Kuhn and Johnson 2013). To optimize the model for selecting the most important variables, 10-fold cross-validation was implemented. Its accuracy performance was assessed through the coefficient of determination (R^2), root mean square error (RMSE), and mean absolute error (MAE), according to the equations 1, 2 and 3:

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(mi-ei)^2}{\sum_{i=1}^{n}(mi-\acute{m\imath})^2} \quad (1)$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(ei - mi)^2} \qquad (2)$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|ei - mi| \qquad (3)$$

where: n is the number of crossings originated between soil maps and land parcels, mi is the yield corresponding to each crossing, $\acute{m\imath}$ is the average of the measured value, and ei is the estimated yield value. RMSE measures the spread of the error distribution: the smaller, the better the model (Isaaks and Srivastava 1989); MAE refers to the prediction bias: the lower this index, the better the model (Willmott 1982).

Dimension reduction

Dimension reduction is the first step of HCPC and it was obtained by two different methods of factorial analysis (Pagès 2015): (i) principal component analysis (PCA), and (ii) factor analysis for mixed data (FAMD) (FactoMineR package; Husson et al. 2007; Husson et al. 2017). This factorial analysis performs linear combinations between initial variables that are strongly related, generating synthetic variables called principal components or dimensions (Hou et al. 2017; Pagès 2015). The principal components hold the variables responsible for the maximum variance of the original data (Kassambara 2017a). It reduces database noise and therefore makes the clustering more stable than using the original variables (Husson et al. 2010; Praene et al. 2019).

For the calculation of the principal components, FAMD takes into account both quantitative variables, such as PCA, and qualitative variables such as multiple correspondence analysis (MCA) (Pagès 2004; Feuillet et al. 2012). According to Pagès (2004), FAMD is suitable when there are fewer qualitative variables than quantitative ones, as is the case of the present study. More details about FAMD can be found in Pagès (2015). To calculate linear combinations and thus generate the synthetic variables that summarize the variability of the data, the PCA only considers quantitative variables. Thus, the qualitative variables are illustrative only, being called supplementary variables (Feuillet et al. 2012; Husson et al. 2007; Husson et al. 2017; Husson et al. 2010). Both methods were tested to verify the effect of qualitative variables on the performance of the next step in the classification by HCPC (Fig. 2).

Hierarchical clustering on principal components (HCPC)

To define the homogeneous management zones as a function of the variables related to soil and crop management, HCPC was performed from FactoMineR package (Lê et al. 2008), following three standard steps in multivariate analysis (Kassambara 2017a): (i) computing principal components (as above-mentioned); (ii) computing hierarchical cluster using Ward's method (Ward, 1963) and; (iii) performing partitional clustering by k-means. After defining the number of dimensions (principal components), a hierarchical tree is formed using Ward's minimum variance method, which in turn consists of an agglomerative and hierarchical method widely used (Husson et al. 2017). In the agglomerative method, the algorithm begins by treating each object as an individual cluster, and then, pairs of clusters are merged until all clusters have been merged into a single cluster, containing all individuals with similar characteristics (Kassambara 2017b). Finally, the k-means algorithm starts from the separation performed by the hierarchical tree with Ward's method, based on Euclidian distance (Husson et al. 2010), and performs several iterations with the aggregation of clusters around a moving center. This latter procedure increases the robustness of Ward's classification and does not affect the number of clusters, but their final constitution (Husson et al. 2010; Kassambara 2017b).

The split criterion to define the number of clusters resulting from HCPC is based on inertia gain (Husson et al. 2010). The final goal is that the inertia between-clusters must be maximum, and within-cluster must be minimum as possible. The inertia inside characterizes the homogeneity of the cluster. The best number of Q clusters is one in which the increase in inertia between Q-1 and Q clusters is much greater than that increase between Q and Q + 1. For example, the gain in inertia when going from 2 to 3 clusters must be much greater than when going from 3 to 4 clusters. When the algorithm identifies this value, the partition is automatically concluded (Husson et al. 2010; Husson et al. 2017).

The HCPC also classifies, in general, the most important variables for each model input. Thus, concerning quantitative variables, the most important ones are those with the highest R^2. The p-value associated with R^2 corresponds to the test of the following null hypothesis: "the average of the variable is equal to the general average". The lower the p-value, the greater the significance, which indicates that the null hypothesis was denied (Husson et al. 2010). The X^2 test is performed between the cluster and the categorical variables. A p-value $<$ 0.05 means that the categorical variable is related to the cluster. The lower this value, the greater the association between them.

The v-test was applied to characterize the formed clusters and to assist in the management zones interpretation (Escofier and Pagès 2008). From the v-test, the clusters are described by qualitative and quantitative variables. The higher the v-test value for a given variable, more important it is to characterize the given cluster. The values of the v-test with negative signs indicate that the average of the variable in a given cluster is less than the general average (Husson et al. 2010). The v-test is calculated according to the following equation 4:

$$v - test = \frac{(\acute{x}_q - \acute{x})}{\sqrt{\frac{s^2}{I_q}\left(\frac{I - I_q}{I - 1}\right)}} \tag{4}$$

where: $\acute{x}_q$ represents the mean of variable *x* for individuals in cluster q, $\acute{x}$ is the mean of *x* for all individuals, s^2 is the variance, and I_q is the number of individuals within-cluster *q*. The v-test is used to test the following null hypothesis: “variable *x* does not characterize the cluster” (Husson et al. 2017).

Tukey test

Tukey test was performed at 10% significance to test the statistical difference of coffee yield of the 2016/2017 harvest (bags ha^{-1}) between clusters (*agricolae* package in R software version 3.4.4). Statistical differences are desirable, attesting the effectiveness of statistical methods in delineating management zones effectively different from each other. Besides that, a smaller number of clusters is sought in the establishment of the management zones, since more than three clusters would not be manageable (Moral and Serrano 2019).

Model input and a complete sequence of statistical analyses

In summary, by overlaying soil and crop management information, the following model inputs were taken to form the clusters, as shown in Fig. 2: (I) soil survey variables were submitted before the FAMD and then to HCPC; (II) soil survey variables were submitted prior to the Random Forest, and the selected variables were submitted to the FAMD and the HCPC; (III) soil survey variables were submitted prior to the PCA, and then to HCPC; (IV) soil survey variables were submitted to the PCA and then to HCPC; (V) soil survey and crop management variables dataset was submitted to a prior to the FAMD, and then to HCPC; (VI) soil survey and crop management variables were submitted to Random Forest, and the selected variables were submitted to FAMD and HCPC; (VII) soil survey and crop management variables dataset was submitted to prior to the PCA, and then to HCPC; (VIII) corresponds to test VI with PCA instead of FAMD, once Random Forest selected only quantitative variables

Results and discussion

Soil types, properties and crop management information

Six soil types were found in the farmlands: Haplustept (4.6%), Rhodudult1 (36.5%), Rhodudult2 (2.5%), Hapludult (4.7%), Hapludox (5.6%), and Acrudox (46.2%). Fig. 3 displays soil profile and morphological characteristics that better characterize soil types found in the farmlands. Fig. 3a is a Haplustept soil profile, with blocky structure in Bw horizon (moderate grade), with stoniness throughout the soil and the C horizon is closer to the surface. Fig. 3b is a blocky structure in Bt horizon surrounded clay skins, peculiar of Rhodudult1, Rhodudult2, and Hapludult. Rhodudults1 and 2 present clay content increasing in depth, and Hapludult presents clayey texture along the entire soil profile. Fig. 3a shows a Hapludox, a deeper soil with homogenous color, presenting B horizon structure intermediate between blocky and granules.

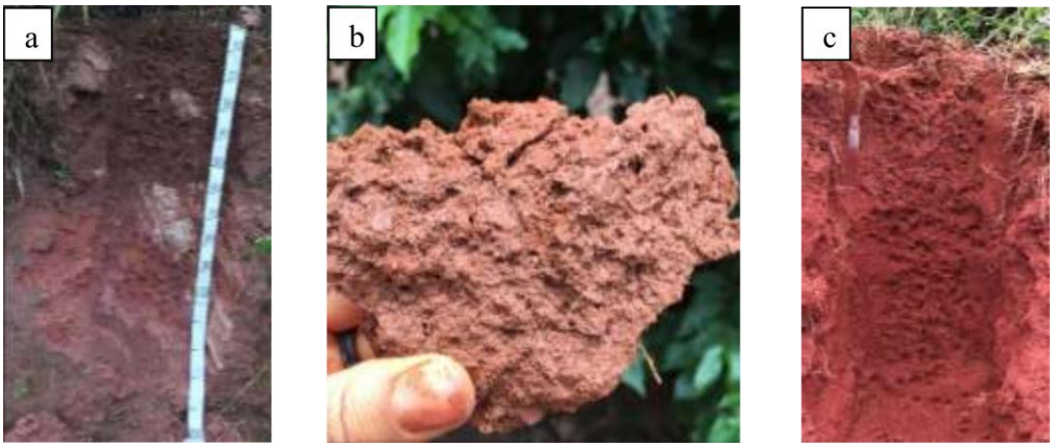

Fig. 3. Soil types found in farmlands: a) Haplustept soil profile; b) blocky structure with strong grade, and clay skins commonly found in the Bt horizon of Rhodudult1, Rhodudult2, and Hapludult; and c) Hapludox soil profile.

Fig. 3 also represents a soil weathering degree or chronological sequence, which is quite common under tropical conditions: Inceptisol (Haplustept) (Fig. 3a) → Ultisol (Rhodudult1, Rhodudult2, and Hapludult) (Fig. 3b) → Oxisol (Hapludox) (Fig. 3c). Such different development degrees (least weathered → intermediate weather degree → highly weathered, respectively) implying in contrasting characteristics of soils regarding morphology, water dynamics and natural fertility. While the weathering degree advances (arrow sequence), there is an occurrence of: soil thickness increasing; soil water storage capacity increasing; B horizon structure shifting from blocky to granular; permeability increasing; natural soil fertility decreasing (especially CEC and BS decreasing); erosion susceptibility decreasing (keeping all the erosion conditioners constant for comparing soils properly); silt/clay ratio decreasing (Resende et al. 2014); bulk density decreasing (Ajayi et al. 2009). Regarding soil texture, there is a general trend of clay increasing as soils evolve. Silt content tends to be lower in tropical conditions, however, its relation with clay content is used as a soil classification criterion (SiBCS 2018): the higher the silt content, the younger tends to be a given soil.

Keeping in mind that soil types information will be applied as environmental explanatory variables of models, Table 1 shows how such categorical information will be used as input model data. The term strong means that soil aggregates are more resistant to disaggregation, in turn, moderate means less resistance to disaggregation (Santos et al. 2015). The epipedon type represents a superficial diagnostic horizon used for soil identification and is mostly related to the color given by organic matter to this horizon, soil structure and depth (Weil and Brady 2017). Other categorical variables used in soil surveys and addressed here were the fertility class, stoniness, soil depth and soil type.

Table 1 Soil properties and types identified on the farms

Soil properties					
Shape and grade of structure[1]	Type and depth of epipedon	Fertility[2]	Soil Depth[3]	Stoniness	Soil type
Moderate blocky	Ochric (0 - 0.05 m)	Eutrophic	Shallow	Stony	Hp
Strong blocky	Ochric (> 0.05 m)	Dystrophic	Deep	Not stony	Ax
Between blocky and granules			Too deep		Hx
					Rh1
					Rh2
					Ht

[1] B horizon information; [2] fertility in the first 1.0 m of depth; Hp: Haplustept; Ax: Acrudox; Hx: Hapludox; Rh1: Rhodudult 1; Rh2: Rhodudult 2; Ht: Hapludult; [3] Shallow (≤ 0.5 m); Deep (> 1 and < or equal to 2 m); Too deep (> 2 m).

The quantitative mean values of soil chemical and physical properties, as well as the quantitative characteristics of crops, are shown in Table 2. Concerning the chemical characteristics, the mean pH was 5.6 and, compared to other soil properties, exhibits the lower CV (10.7%). The mean P content is 15.5 mg kg^{-1} (CV = 99.4%). The average effective CEC is 3.8 $cmol_c$ kg^{-1}. The mean BS is 47.7% (CV = 30.0%). The clay content ranged from 12.3 to 73.0% (CV = 25.7%) and, the lower values were determined in the Haplustept, the soil with lower pedogenic development. The high coefficients of variation in physical and chemical properties corroborate with the higher spatial variability and different weathering degrees of farmland soils, indicating the applicability of management zones separation (Moral et al. 2010; Caires et al. 2014; Moral and Serrano 2019).

In relation to the crop characteristics, the mean crop density is 4.746 plants ha^{-1} (CV = 23.0%), which is framed as a semi-dense crop. The mean crop age is 14.3 years (CV = 42.7%) and the average yield is 50.5 bags ha^{-1} (bags of 60 kg) , which is framed as very high (CV = 42.5%). According to Ribeiro et al. (1999), there are three systems of production-related to crop density in coffee plantations in Brazil: traditional (< 2,500 plants ha^{-1}), semi-dense (2,500 – 5,000 plants ha^{-1}), and dense (> 5,000 plants ha^{-1}). This is a relevant crop characteristic that influences the coffee yield per ha.

As for qualitative variables, four varieties of coffee plants were grown in crop areas: Rubi, Mundo Novo, Catuaí and Acaiá. Each one varieties has different yields potential, resistance to water deficit and different types of stress (Fazuoli et al. 2002). Among crops, there are areas that are irrigated and others that are not irrigated. There were two stages of development: areas in formation and areas in the production stage.

Table 2 Quantitative soil properties, terrain attributes, and crop management characteristics

	Mean	Sd	CV (%)	Minimum	Maximum
pH	5.6	0.6	10.7	4.7	7.2
K (mg kg^{-1})	100.9	53.2	52.7	38.2	346.7
P (mg kg^{-1})	15.5	15.4	99.4	1.0	55.0
Ca^{2+} ($cmol_c$ kg^{-1})	2.5	1.3	52.0	0.5	7.3
Mg^{2+} ($cmol_c$ kg^{-1})	0.8	0.4	50.0	0.2	2.0
Al^{3+} ($cmol_c$ kg^{-1})	0.3	0.4	133.3	0.0	1.5
H^+ +Al^{3+} ($cmol_c$ kg^{-1})	3.1	1.6	51.6	1.3	8.1
SB ($cmol_c$ kg^{-1})	3.5	1.7	48.6	0.9	10.2
Effective CEC ($cmol_c$ kg^{-1})	3.8	1.8	47.4	1.2	10.2
CEC pH 7.0 ($cmol_c$ kg^{-1})	6.6	2.4	36.4	3.4	13.6
BS (%)	47.7	14.3	30.0	15.8	78.2
AS (%)	12.0	14.0	116.7	0.6	50.5
SOM (dag kg^{-1})	1.7	0.5	29.4	0.7	3.0
Clay (%)	52.9	13.6	25.7	12.3	73.0
Silt (%)	21.1	9.3	44.1	8.3	37.0
Sand (%)	25.9	15.7	60.6	6.3	60.3
TR	1.1	0.1	9.1	0.8	1.4
Slope (%)	10.5	2.4	22.9	5.3	21.6
Crop density (plants ha^{-1})	4,746	1,091	23.0	1,667	11,126
Crop age (year)	14.3	6.1	42.7	3.0	32.0
Yield (bags ha^{-1})	50.5	21.5	42.5	6.5	106.0

Sd: standard deviation; CV: coefficient of variation; SB – sum of bases; CEC: cation exchange capacity; BS: base saturation; SOM: soil organic matter; TR: textural relationship (clay content in A horizon/clay content in B horizon); AS: Aluminum saturation.

Variables selection by Random Forest

Through Random Forest with RFE, the most important variables that explain crop yield were ranked, based on variables importance score (Fig. 4) (Kuhn 2012). In the soil dataset, the most important variables ranked by importance order were silt content, slope, epipedon type, SOM, and effective CEC, respectively (Fig. **Fig.** 4a) (R^2 of 0.16, MAE of 16.86 bags ha^{-1} and, RMSE of 21.63 bags ha^{-1}). For soil and crop management dataset the most important variables were crop age (years), crop density, silt content, and SOM (Figure 3b) (R^2 of 0.43, MAE of 12.44 bags ha^{-1}, and RMSE of 16.63 bags ha^{-1}). Based on statistical indexes, the combination of soil and crop management variables better explains yield variability, since the higher accuracy of models was found.

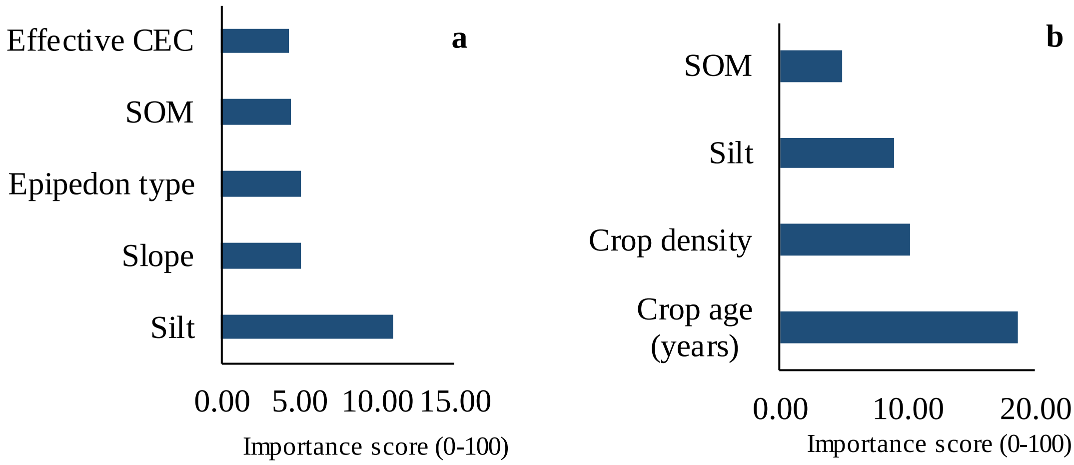

Fig. 4 Variables importance defined by Random Forest with Recursive Feature Elimination applied to verify the relationship between coffee yield and soil survey dataset, referent to model inputs II and IV (a); and between coffee yield and soil survey and crop management dataset referent to model inputs VI and VIII (b); effective CEC: effective cation exchange capacity; SOM: soil organic matter.

Economic and climatic factors, as well as the general crop management characteristics, are the main factors that influence not only coffee yield in Brazil (Rena et al. 2002), but also other crops in general (Chlingaryan et al. 2018; Tisdale et al. 1993). Since yield is an important parameter for management zones establishment, we incorporated crop management characteristics that govern yield by their intrinsic characteristics, such as age and crop density. Dissociating those characteristics that could lead to an incorrect management zones establishment, the application of this methodology created for established coffee farms sounds realistic.

In tropical regions, with ample predominance of clays that have low CEC (Fontes and Weed 1991), SOM assumes a great importance on generating negative superficial charges (Novais and Mello 2007). In addition to acting as a cementing agent, SOM promotes the stabilization of soil aggregates (Silva and Mendonça 2007) and substantially improves the water retention capacity of the soils (Dexter 2004). In addition, SOM has a great influence on soil porosity and therefore, affects water availability for plants (Dexter 2004).

Besides being a reliable soil weathering degree indicator, higher contents of silt may also promote soil susceptibility to water erosion (Weil and Brady 2017) by surface crusting and permeability decreasing. Regarding the variables chosen for the definition of management zones, it is interesting to note that silt, a stable variable, was selected as important for the definition of management zones. The choice of silt as a variable for defining management zones is in agreement with the desirable characteristic that these variables are stable (Ping et al. 2005).

Principal component analysis and Factor analysis of mixed data

The HCPC requires a reduction in the dimensionality of the data (Kassambara 2017b) that can be performed by different types of factor analysis. In this sense, FAMD must be performed in situations where at least one categorical variable is present (Pagès 2015). In those models that Random Forest selected only quantitative variables (Fig. 4b), FAMD performance of models VI and VIII was not possible.

In this study, PCA and FAMD were used to derive a small number of linear combinations with synthetic variables, called principal components (Pagès 2015), capable of explaining most of the data variability. The analysis of PCA or FAMD can be considered a pre-processing on the analysis of the HCPC, performing a reduction of the data into a few variables, containing the most important information (Kassambara 2017a). This makes the cluster division more robust, according to the number of dimensions considered, which must be the one that presents the most explained variance (Husson et al. 2017). Thus, as most of the variance was retained in the first 5 dimensions (Fig. 5), these were considered for the execution of the HCPC.

According to Fig. 5, which brings more details about cumulative variance explained by each model input, the models IV, VI and VIII enabled the explanation of 100% of the data variance in the first 4 principal components (the best performances among models), and the model input II presented 100% of the variance explained in the first 5 ones. Combining data on yield and soil and terrain attributes, Ping et al. (2005) also found a high percentage of variance explained in the first five components for management zones delineation, as the obtained in input models VI and VIII from soil and crop management variables dataset. It seems that the similarity in the nature of the soil and yield data used in the present study and in that one developed by Ping et al. (2005), promoted the explanation of most of the variability of the data in the same number of dimensions. Regarding HCPC, the use of 5 components is adequate for greater stability of the clustering, since it removed disregarded data that do not bring relevant information (noise) (Husson et al. 2010).

The model inputs III and VII were responsible for the second major variability explained by the PCA, in which Random Forest was not performed a prior. The lowest percentages of explained variability in first 5 dimension were found in the model inputs I (73.8%) and V (64.4%), in which the complete set of variables (soil and crop management variables dataset) were tested, without Random Forest selection and using FAMD. Similarly, Moral et al. (2010) also performed the PCA to summarize the variability of soil data before classification with the fuzzy c-means algorithm, in order to establish management zones. Unlike what was found in this study, and in accordance to Ping et al. (2005), only two dimensions were necessary to summarize most of the variability.

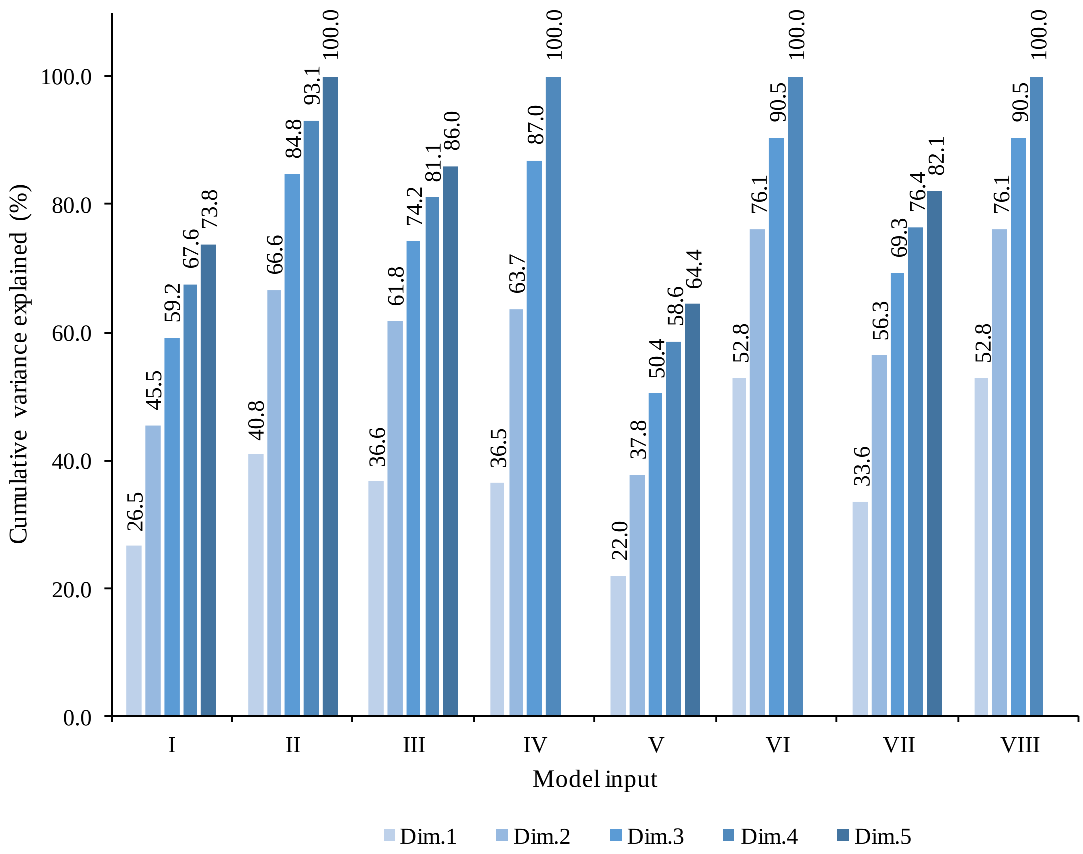

Fig. 5 Cumulative variance explained by principal components with different datasets inputs and variables selection. Dim – Dimension (principal component).

Hierarchical cluster on principal components (HCPC)

The best number of clusters is defined according to the inertia gain while the number of cluster increases (Husson et al. 2017). The inertia gain of each model input is presented in Fig. 6. These graphs were generated from the divisions of clusters produced with the different input models. The numbers in roman algorism indicate the model input. The size difference of the first bar of the graphs relative to the second one indicates the inertia gain by increasing the number of two clusters (first bar, q=1) to three (q=2) and so on. The change of color between dark blue and light blue bars indicates up to which cluster the number of the divisions remains significant and promotes differences between clusters. The dark blue columns become clear when the HCPC identifies the number of clusters (Q) that minimizes the gain of inertia when increasing a cluster. The number of clusters maintained is the one that enhances the inertia gain, or gain of variance between successive cluster numbers (Husson et al. 2010; Kassambara 2017b). Due to the characteristics of the data, different cluster numbers were generated by the HCPC in the input model tested. Models I and V resulted into 6 clusters: the highest number obtained among models. These two model inputs have in common the use of FAMD, as well as the least variability explained by PCA. It is noteworthy that models pairwise by similar statistical sequence of analysis I and V; II and VI, III and VII, IV and VIII presented the same scale of variation of inertia gain. Except for II and VI, the others presented similar number of clusters.

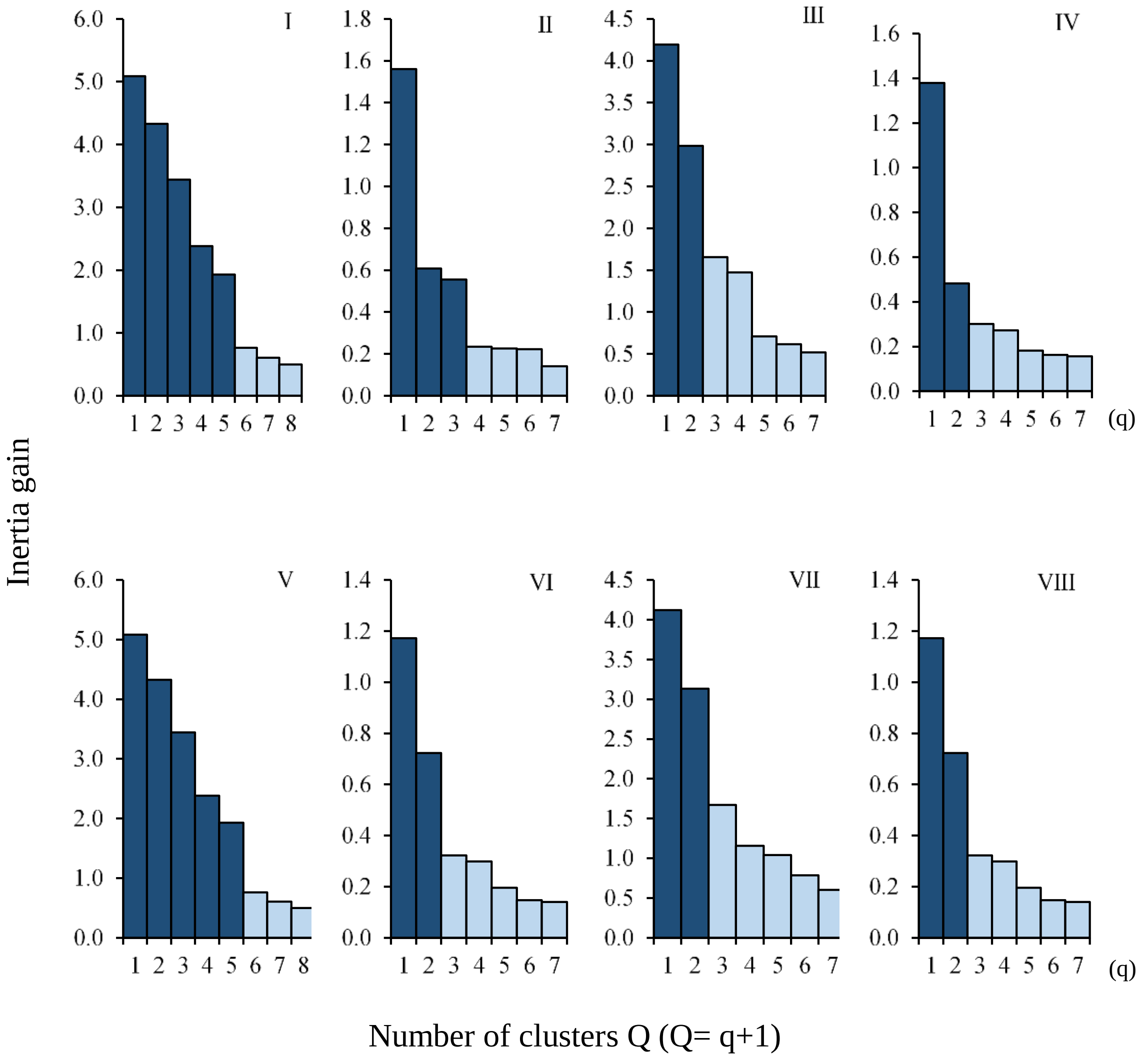

Fig. 6 Inertia gain obtained in the division of clusters.

The HCPC classifier ranks variables that most influence clusters generation, expressed by different quantitative (R^2) and qualitative (p-value of chi-square test) parameters (Table 3). For the quantitative ones, the higher the R^2, the greater the influence of a given variable on the division of clusters, as it was observed for soil fertility variables. Despite this, it is important to remind that soil fertility variables were not chosen neither by Random Forest as the most important ones nor in cluster separation.

With respect to qualitative variables, the lower the p-value, the more the categorical variable characterizes a given cluster (Husson et al. 2017). Thus, the soil type was the most important categorical variable (model inputs I, III, V, VII) due to its well-known importance on productive systems (Cavalli et al. 2020). Conversely, as already presented in the previous section, Random Forest did not select soil type variable in importance rank. Although very accurate (Biau and Scornet 2016), it has been reported that Random Forest (Breiman 2001) could present bias in variables importance if categorical predictors have different numbers of levels or if predictors are mixed categorical and continuous (Strobl et al. 2007; Boulesteix et al. 2012). In the soil survey and crop management variables dataset, the Random Forest selected only quantitative variables, so the PCA was performed and, consequently, the model inputs VI and VIII included the same variables and promoted the formation of equal management zones, from the delimitation by the HCPC. If some qualitative variables had been selected, it would be possible to compare the differences in size reduction and generation of management zones when PCA and when FAMD were applied (Pagès 2004).

Table 3 Variables that most influenced the divisions of the clusters formed

Model input*	Quantitative variables			Qualitative variables		
		R^2	p-value		p-value	df
I	SB ($cmol_c$ kg^{-1})	0.79	<0.01	Soil type	4.4e-71	25
(6)	Ca^{2+} ($cmol_c$ kg^{-1})	0.78	<0.01	Depth	1.3e-43	10
	CEC ($cmol_c$ kg^{-1})	0.78	<0.01	Structure	7.9e-34	10
	CEC pH 7.0 ($cmol_c$ kg^{-1})	0.78	<0.01	Stoniness	2.6e-32	5
	Clay (%)	0.71	<0.01	Fertility	1.8e-17	5
	$H+Al^{3+}$ ($cmol_c$ kg^{-1})	0.71	<0.01	Type and depth of epipedon	8.2e-17	5
II	CEC ($cmol_c$ kg^{-1})	0.68	<0.01	Type and depth of epipedon	3.5e-31	3
(4)	SOM (dag kg^{-1})	0.57	<0.01			
	Silt (%)	0.57	<0.01			
	Slope (%)	0.10	<0.01			
III	Al^{3+} ($cmol_c$ kg^{-1})	0.79	<0.01	Soil type	6.0e-20	10
(3)	$H+Al^{3+}$ ($cmol_c$ kg^{-1})	0.78	<0.01	Fertility	9.7e-16	2
	AS (%)	0.65	<0.01	Structure	7.6e-08	4
	BS (%)	0.64	<0.01	Depth	9.7e-07	4
	CEC pH 7.0 ($cmol_c$ kg^{-1})	0.63	<0.01	Type and depth of epipedon	2.2e-04	2
	Mg^{2+} ($cmol_c$ kg^{-1})	0.63	<0.01	Stoniness	9.1e-03	2
IV	Silt (%)	0.66	<0.01	Type and depth of epipedon	1.3e-09	2
(3)	CEC ($cmol_c$ kg^{-1})	0.66	<0.01			
	SOM (dag kg^{-1})	0.55	<0.01			
	Slope (%)	0.06	<0.05			
V	SB ($cmol_c$ kg^{-1})	0.77	<0.01	Soil type	3.6e-58	25
(6)	CEC ($cmol_c$ kg^{-1})	0.77	<0.01	Depth	6.8e-38	10
	CEC pH 7.0 ($cmol_c$ kg^{-1})	0.76	<0.01	Stoniness	2.6e-32	5
	Ca^{2+} ($cmol_c$ kg^{-1})	0.76	<0.01	Type and depth of epipedon	1.8e-28	5
	$H+Al^{3+}$ ($cmol_c$ kg^{-1})	0.70	<0.01	Structure	2.2e-21	10
	Mg^{2+} ($cmol_c$ kg^{-1})	0.69	<0.01	Stage of crop development	5.2e-20	5
				Irrigation area	7.8e-19	5
				Fertility	2.6e-16	5
				Variety	3.5e-12	20
VI	Crop age (years)	0.57	<0.01			
(3)	Silt (%)	0.53	<0.01			
	SOM (dag kg^{-1})	0.52	<0.01			
	Crop density (plants ha^{-1})	0.49	<0.01			
VII	CEC pH 7.0 ($cmol_c$ kg^{-1})	0.82	<0.01	Soil type	2.5e-15	10
(3)	CEC ($cmol_c$ kg^{-1})	0.72	<0.01	Stoniness	1.2e-08	2
	BS (%)	0.62	<0.01	Fertility	1.5e-08	2
	Ca^{2+} ($cmol_c$ kg^{-1})	0.62	<0.01	Irrigation area	1.6e-08	2
	Mg^{2+} ($cmol_c$ kg^{-1})	0.59	<0.01	Epipedon	2.3e-06	2
	Al^{3+} ($cmol_c$ kg^{-1})	0.55	<0.01	Depth	3.3e-04	4
				Structure	4.1e-04	4
				Variety	3.4e-03	8
				Stage of crop development	6.9e-03	2
VIII	Crop age (years)	0.57	<0.01			
(3)	Silt (%)	0.53	<0.01			
	SOM (dag kg^{-1})	0.52	<0.01			
	Crop density (plants ha^{-1})	0.49	<0.01			

*number of clusters between paranthesis; df: degree of freedom of the chi-square test between qualitative variables and the test; AS: aluminum saturation; SB: sum of bases; BS: base saturation; SOM: soil organic matter ; CEC: effective soil exchange capacity; CEC pH 7.0: cation exchange capacity at pH 7.0; ( ) the numbers in parentheses represent the number of clusters formed from each model input tested.

From the establishment of management zones by the models, the statistical differences of mean coffee yield within each HCPC cluster domain were calculated and are presented in Table 4. The purpose of this test was to choose model inputs with statistical differences among clusters, which was promoted by the models VI and VIII, totalizing 3 clusters. It is important to highlight that they presented the same mean values since they became equal models, because Random Forest did not ranked any categorical variable making impossible FAMD application. As already mentioned, both models are formed by 4 variables (crop age, crop density, silt content, SOM) that were responsible for the explanation of 43% of coffee yield variability.

Table 4 Mean coffee yield (2016/2017 harvest) within clusters formed by different models

Model input	Clusters					
	1	2	3	4	5	6
I	56.0 a	51.9 ab	47.6 ab	45.9 ab	44.2 b	43.6 b
II	58.6 a	47.0 b	45.2 b	45.2 b	-	-
III	53.0 a	46.9 a	46.4 a	-	-	-
IV	54.7 a	45.2 ab	44.3 b	-	-	-
V	63.6 a	59.2 a	45.9 ab	44.5 b	44.2 b	43.6 b
VI	56.6 a	46.4 b	34.3 c	-	-	-
VII	55.2 a	46.1 b	46.0 b	-	-	-
VIII	56.6 a	46.4 b	34.3 c	-	-	-

Means in the same row followed by different letters indicate significant differences according to Tukey's test ($p < 0.1$)

The best model input and the case study of coffee crop

Based on the whole statistical framework, the set of explanatory variables and the sequence of statistical models that best outlined management zones for the study area were the models VI and VIII. Thus, as already mentioned, they can be considered equals. Based on statistical performance, both models outperformed the others, since:

a) The models applied the combination of soil and crop management explanatory variables, which according to Random Forest variables selection, were those with higher accuracy and capability to explain coffee yield;
b) They were included in the group of PCA with higher % of variables explanation by the first 4 dimensions, suggesting consistency of analysis;
c) They presented higher inertia gain provided by 3 clusters composition;

d) They promoted statistical differences of mean coffee yield among all clusters created from HCPC, suggesting effectiveness of generating management homogenous zones that also differ from each other.

Although the explained variance of models were already presented in previous section (Fig. 5), in order to better understand the main drivers of coffee yield of this study case and specifically for models VI and VIII, Fig. 7 shows the first (Dim 1) and second (Dim 2) principal components of soil and crop management information, which account for most of the total variance (Kassambara 2017b). Based on silt content, SOM, crop density, and crop age information, the first principal component was able to explain 52.8% of total variance and, the second one 23.3%, totalizing 76.1%. The PCA shows that silt content and crop age are positively correlated, however, there is no cause and effect relationship between them. Conversely, crop age presented a negative correlation with crop density, due to the adoption of plantings with higher crop density.

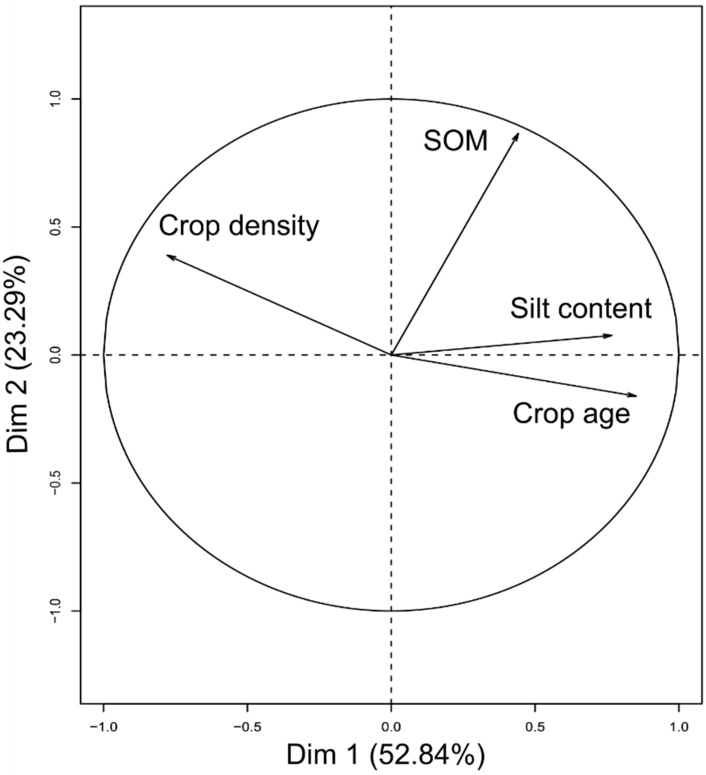

Fig. 7 Principal components analysis performed with four variables selected. SOM: soil organic matter.

Table 5 presents the final results of the HCPC, corresponding to the partitioning performed by Ward's method on the five dimensions considered in conjunction with the k-means partitioning. The description of the variables that are significantly ($p < 0.01$) correlated with the clusters is presented, starting with those that best characterize the partitioning (Husson et al. 2017). Based on that, Fig. 8 represents the spatial arrangement of the clusters formed, and Table 6, shows the soil types geographical expression within each cluster. Table 7 presents quantitative contents of soil properties for each depth of soil survey, as well as the coffee yield for three harvest years, providing support for the subsequent analyses of each cluster delineated. All together contributes to better comprehend the different scenarios or environments defined by the clusters created, as follows:

- Cluster 1: it is composed by the best combination of crop features for achieving higher yields – lower age of coffee plants (10.84 years) and higher crop density (5,174.05 plants ha^{-1}) (Table 5). The higher the plant density, the better use of water and nutrients occurs due to the greater density of roots in the soil (when water is not a limiting factor). Thus, higher coffee yields, especially in 2014/2015 and 2016/2017 years (51.81 and 56.58 bags ha^{-1}, respectively) were obtained (Table 7), representing values much above national Coffee Arabica yield average for such years (24 and 26 bags ha^{-1}, respectively) (Conab, 2019). Considering the most important soil characteristics ranked by Random Forest, the silt content was the lowest among clusters, corroborating the predominance of very highly weathered soils (Oxisols) (Table 5). Furthermore, another trace characteristic is the lower SOM among clusters.

- Cluster 2: although still above the national yield average, this cluster presents a slight decreasing of yield when compared with Cluster 1, outperforming crop yield only in 2012/1013 year (51.44 bags ha^{-1}) (Table 6). It is noteworthy that such highest coffee yield was obtained in a severe drought period in Brazil (Getirana 2016). Considering Random Forest variables ranked, the crop age is within average (16.24 years), as well as the silt content (Table 5). The latter, is in accordance with mature soils (Ultisols), along with the least proportion of young soils (Inceptisols) was found (Table 6). Higher clay content and SOM values among clusters were found (Table 7), which could be responsible for higher water retention during such drought periods.

- Cluster 3: this cluster presents the lowest plant yield for all vintages (values around 33 bags ha^{-1}). When we consider the standard deviation of values, some land parcels were below national average. Regarding Random Forest variables ranked, this cluster presents the higher age of crops (24.05 years) and lower crop density (2,904.09 plants ha^{-1}), an opposite trend of Cluster 1. The highest silt content was found, where there is predominance of young soils

(Inceptisols) among clusters.

Table 5 Correlation between three clusters and environmental variables related to soil and crop, obtained by the hierarchical cluster applied in land parcels from coffee plantations of southern Minas Gerais state, Brazil

Variables	v.test	Mean in category	Overall mean	Sd in category	Overall Sd	p-value
			Cluster 1			
Crop density	5.67	5,174.05	4,746.00	784.65	1,087.85	<0.01
SOM	-7.28	1.43	1.69	0.30	0.52	<0.01
Crop age	-8.07	10.84	14.25	3.21	6.09	<0.01
Silt content	-9.10	15.29	21.12	5.65	9.25	<0.01
			Cluster 2			
SOM	9.03	2.27	1.69	0.43	0.52	<0.01
Silt content	6.33	28.41	21.12	7.80	9.25	<0.01
Crop age	2.62	16.24	14.25	4.44	6.09	<0.01
			Cluster 3			
Crop age	8.10	24.05	14.25	5.60	6.09	<0.01
Silt content	4.72	29.77	21.12	5.55	9.25	<0.01
Crop density	-8.53	2,904.09	4,746.00	676.52	1087.85	<0.01

Sd: standard deviation; p-value <0.01 confirms the statistically significant correlation between variable and clusters.

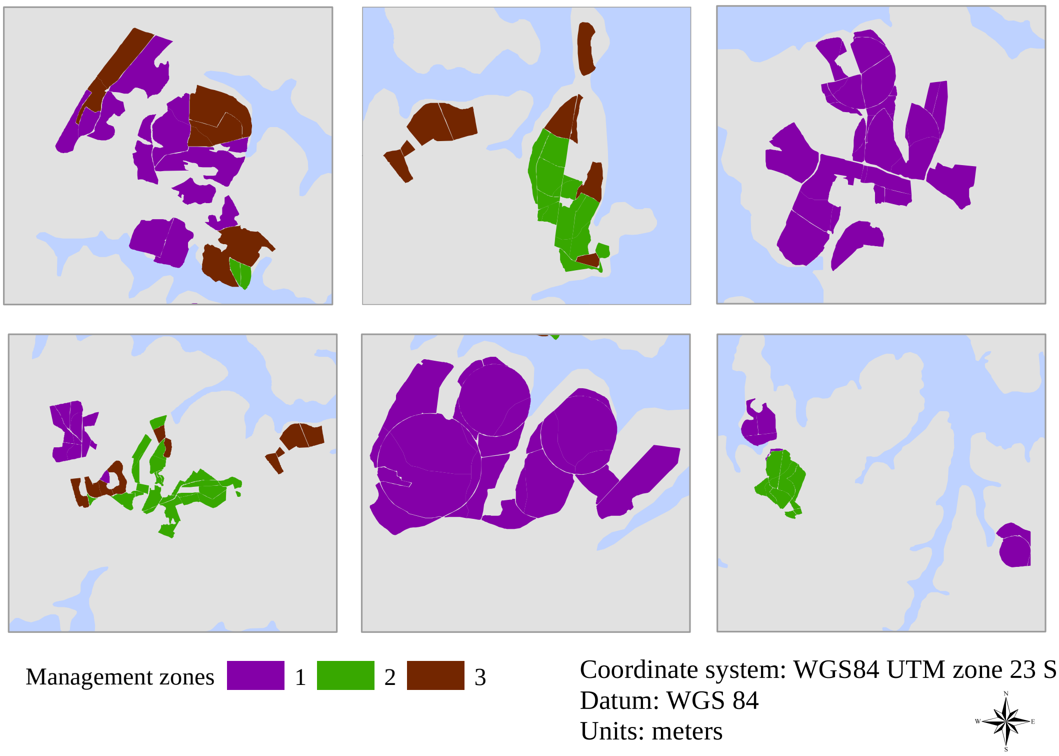

Fig. 8 Clusters obtained by hierarchical cluster on principal components, using soil properties (SOM and silt content) and crop characteristics (crop age and crop density).

Table 6 Soil types geographical expression (%) within different clusters

Clusters	% of soil types occurrence					
	Ax	Hx	Hp	Rh1	Ht	Rh2
1	36.5	1.3	1.8	23.7	-	-
2	5.5	3.9	0.3	6.5	2.3	2.2
3	4.2	0.4	2.5	6.3	2.4	0.3
Total	46.2	5.6	4.6	36.5	4.7	2.5

Ax - Acrudox; Hx- Hapludox ; Hp - Haplustept; Rh1 - Rhodudult 1; Ht – Hapludult; Rh2- Rhodudult 2.

Table 7 Quantitative properties (mean ± standard deviation) of soils at different depths, slope and coffee yield in management zones formed with hierarchical cluster analysis

	cluster 1 (n = 90)			cluster 2 (n = 46)			Cluster 3 (n = 22)		
	Epipedon	0.4-0.7 m	1.0-1.5 m	Epipedon	0.4-0.7 m	1.0-1.5 m	Epipedon	0.4-0.7 m	1.0-1.5 m
pH	5.91 ±0.89	5.41 ±0.83	5.28 ±2.91	6.46 ±0.65	5.49 ±0.74	4.94 ±1.25	6.1 ±0.97	5.71 ±0.72	5.04 ±1.91
K (mg kg^{-1})	184.2 ±76.91	51.6 ±37.5	30.59 ±27.96	215.89 ±71.8	69.82 ±51.66	43.07 ±36.77	196.13 ±82.19	80.71 ±49.59	50.44 ±34.33
P (mg kg^{-1})	31.06 ±37.78	1.32 ±3.26	0.39 ±0.26	59.63 ±36.13	1.07 ±0.73	0.6 ±0.53	48.46 ±51.05	0.82 ±0.22	0.3 ±0.14
Ca^{2+} ($cmol_c$ kg^{-1})	3.82 ±2.24	1.09 ±0.63	0.71 ±0.5	7.71 ±2.28	1.41 ±0.86	0.95 ±0.44	4.95 ±2.61	1.53 ±0.89	0.95 ±0.59
Mg^{2+} ($cmol_c$ kg^{-1})	1.14 ±0.74	0.46 ±0.43	0.25 ±0.18	2.21 ±0.89	0.46 ±0.13	0.36 ±0.14	1.45 ±0.84	0.57 ±0.46	0.35 ±0.23
Al^{3+} ($cmol_c$ kg^{-1})	0.22 ±0.35	0.28 ±0.43	0.19 ±0.3	0.06 ±0.02	0.56 ±0.87	0.55 ±0.9	0.29 ±0.56	0.22 ±0.37	0.16 ±0.36
H^+ +Al^{3+} ($cmol_c$ kg^{-1})	3.45 ±1.66	3 ±1.92	2.1 ±1.2	2.95 ±1.35	4.36 ±3.97	3.81 ±3.58	3.1 ±1.56	2.49 ±1.79	1.89 ±1.45
SB ($cmol_c$ kg^{-1})	5.44 ±3.06	1.68 ±0.91	1.06 ±0.69	10.48 ±3.11	2.05 ±1.07	1.42 ±0.63	6.91 ±3.48	2.32 ±1.17	1.43 ±0.9
Effective CEC ($cmol_c$ kg^{-1})	5.66 ±2.87	1.96 ±0.88	1.24 ±0.72	10.54 ±3.09	2.61 ±0.92	1.97 ±0.65	7.19 ±3.09	2.53 ±1.08	1.58 ±0.89
CEC pH 7.0 ($cmol_c$ kg^{-1})	8.89 ±2.23	4.67 ±1.76	4.62 ±7.78	13.42 ±2.3	6.41 ±3.45	5.23 ±3.23	10.01 ±2.7	4.81 ±1.71	3.31 ±1.6
BS (%)	57.95 ±22	38.74 ±18.75	30.28 ±18.18	76.65 ±12.56	40.47 ±20.79	35.25 ±20.01	65.22 ±24.34	49.39 ±18.73	40.26 ±23.58
AS (%)	7.69 ±14.94	15.49 ±18.29	12.38 ±16.68	0.71 ±0.44	19.19 ±27.55	20.7 ±30.91	10.23 ±22.76	10.09 ±13.79	9.78 ±16.43
SOM (dag kg^{-1})	2.89 ±0.76	0.95 ±0.42	0.67 ±1.1	4.82 ±1.29	1.27 ±0.23	0.65 ±0.21	3.11 ±1	0.96 ±0.38	0.43 ±0.22
Rem-P	31.79 ±9.96	16.57 ±10.21	9.03 ±8.34	34.1 ±6.17	11.24 ±4.54	6.21 ±4.14	34.02 ±10.6	16.72 ±12.13	7.26 ±9.8
Clay (%)	49.16 ±14.82	53.79 ±15.56	50.86 ±19.1	53.43 ±6.87	60.2 ±7.35	57.83 ±14.03	45.36 ±15.44	50.68 ±17.26	46.64 ±19.91
Silt (%)	16.89 ±6.96	15.49 ±6.57	14.78 ±9.01	32.2 ±6.58	26.39 ±8.38	25.67 ±11.21	31.82 ±5.36	27.91 ±6.06	27.68 ±12.89
Sand (%)	33.96 ±14.79	30.72 ±14.43	25.5 ±13.1	14.37 ±9.66	13.41 ±8.16	12.15 ±5.93	22.82 ±19.38	21.41 ±18.8	16.59 ±16.77
Crop density (years)	5,174.0			4,789.0			2,904.0		
Yield (2012/2013) (bags ha^{-1})	40.02 ±21.68			51.44 ±20.41			30.35 ±11.04		
Yield (2014/2015) (bags ha^{-1})	51.81 ±33.62			46.84 ±13.52			33.08 ±13.35		
Yield (2016/2017) (bags ha^{-1})	56.58 ±21.34			46.40 ±19.78			34.33 ±14.59		

SB: Sum of bases; CEC: cation exchange capacity; BS: base saturation; SOM: soil organic matter; TR: textural relationship (clay content in epipedon /clay content in B horizon); AS: Aluminum saturation; Rem- P: remaining P; n: number of polygons obtained from the crossing between plots and soil maps.

Conclusions

The design of quantitatively established management zones varied deferentially according to the model input and statistical analysis, in which the knowledge of the model metrics as well as the yield data were important to guide the choice of the most suitable ones. Soil and crop management variables selected by Random Forest, along with HCPC, were able to delineate clusters with contrasting crop yields. Such contrasts were also in accordance with the general knowledge of soil-landscape relations and coffee management characteristics of the study region. Summarizing the three clusters outlined (best method performance) (1→2→3), it was found a decreasing in values of coffee yield harvested in 2016/2017 (reference values for statistical analyses) as well as crop density. Following the sequence, it was also noticed the increasing values of crop age and of the silt content of the soils.

Although the models with best performance had as soil attributes ranked as the most important ones those normally acquired from routine soil fertility analyses (normally performed at 0-20 and 20-40 cm depths), it is important to highlight that the soil survey allowed the obtaining of soil unique information in depth (subsurface morphological, physical, and chemical constraints). In addition, soil survey provided management zones contours, whose soil mapping units represent different environments. Such environments could also guide in the decision making regarding water deficit, since it is frequent to occur short periods of drought during the rainy season in the study region. The support from field knowledge together with modern statistical approach provided secure and accurate coffee crop information.

References

Abdel-Rahman, E. M., Ahmed, F. B., & Ismail, R. (2013). 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. doi:10.1080/01431161.2012.713142

Ajayi, A. E., Dias Junior, M. de S., Curi, N., Araujo Junior, C. F., Souza, T. T. T., & Inda Junior, A. V. (2009). Strength attributes and compaction susceptibility of Brazilian Latosols. *Soil and Tillage Research*, 105(1), 122–127. doi:10.1016/j.still.2009.06.004

Alvares, C. A., Stape, J. L., Sentelhas, P. C., De Moraes Gonçalves, J. L., & Sparovek, G. (2013). Köppen's climate classification map for Brazil. *Meteorologische Zeitschrift*, 22(6), 711–728. doi:10.1127/0941-2948/2013/0507

Alvarez, V. H. V., Novais, R. F., Dias, L. E., Oliveira, J.A. 2000. Determinação e uso do fósforo remanescente. Boletim Informativo da Sociedade Brasileira de Ciência do Solo.Viçosa, 25, 1, 27-32.

ANA - Agência nacional das águas. Hidroweb: Sistema de Informações Hidrológicas (2018) http://www.snirh.gov.br/hidroweb/publico/medicoes_historicas_abas.jsfhttp://hidroweb. Accessed 28 September 2018.

Arias, D., Mundial, B., Mapa, P. M., Embrapa, P. A., Silva, F., Daniel, L., et al. (2015). *Revisão rápida e integrada da gestão de riscos agropecuários no Brasil*: Caminhos para uma visão integrada (1a.). Brasilia.

Barbosa, J. N., Borém, F. M., Alves, H. M. R., Volpato, M. M. L., Vieira, T. G. C., & de Souza, V. C. O. (2010). Spatial distribution of coffees from minas gerais state and their relation with quality. *Coffee Science*, 5(3), 237–250. doi:10.25186/cs.v5i3.340

Biau, G., & Scornet, E. (2016). A random forest guided tour. *Test*, 25(2), 197–227. doi:10.1007/s11749-016-0481-7

Boulesteix, A. L., Janitza, S., Kruppa, J., & König, I. R. (2012). Overview of random forest methodology and practical guidance with emphasis on computational biology and bioinformatics. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 2(6), 493–507. doi:10.1002/widm.1072

Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5–32. doi:10.1023/A:1010933404324

Caires, S. A., Wuddivira, M. N., & Bekele, I. (2014). Spatial analysis for management zone delineation in a humid tropic cocoa plantation. *Precision Agriculture*, 16(2), 129–147.

Cavalli, J. P., Reichert, J. M., Rodrigues, M. F., & de Araújo, E. F. (2020). Composition and functional soil properties of arenosols and acrisols: Effects on eucalyptus growth and productivity. *Soil and Tillage Research*, 196 (February 2019). doi:10.1016/j.still.2019.104439

Chlingaryan, A., Sukkarieh, S., & Whelan, B. (2018). Machine learning approaches for crop yield prediction and nitrogen status estimation in precision agriculture: A review. *Computers and Electronics in Agriculture*, 151(November 2017), 61–69. doi:10.1016/j.compag.2018.05.012

CONAB. (2019). Acompanhamento da safra brasileira de café- *Primeiro levantamento. Companhia Nacional de Abastecimento* (Vol. 6). http://www.conab.gov.br

Conrad, O., Bechtel, B., Bock, M., Dietrich, H., Fischer, E., Gerlitz, L., Wehberg, J., Wichmann, V., & Böhner, J. (2015). System for Automated Geoscientific Analyses (SAGA) v. 2.1.4. *Geoscientific Model Development*, 8, 1991–2007.

Dexter, A. R. (2004). Soil physical quality Part I. Theory, effects of soil texture, density, and organic matter, and effects on root growth. *Geoderma*, 120, 201–214. doi:10.1016/j.geodermaa.2003.09.005

Embrapa (1997). *Manual de Métodos de Análise de Solo*. Rio de Janeiro, RJ: Embrapa.

Escofier, B., & Pagès, J. (2008). *Analyses factorielles simples et multiples, Objectifs, méthodes et interprétation*. Paris: Dunod.

Everingham, Y., Sexton, J., Skocaj, D., & Inman-Bamber, G. (2016). Accurate prediction of sugarcane yield using a random forest algorithm. *Agronomy for Sustainable Development*, 36(27). doi:10.1007/s13593-016-0364-z

Fazuoli, L.C., Medina Filho, H.P., Gonçalves, W., Guerreiro Filho, O., & Silvarolla, M.B. (2002). Melhoramento do cafeeiro: variedades tipo arábica obtidas no Instituto Agronômico de Campinas. In: Zambolim, L. (Ed.) *O estado da arte de tecnologias na produção de café* (pp.63-215). Viçosa: Editora UFV.

Feuillet, T., Mercier, D., Decaulne, A., & Cossart, E. (2012). Classification of sorted patterned ground areas based on their environmental characteristics (Skagafjörður, Northern Iceland). *Geomorphology*, 139–140, 577–587. doi:10.1016/j.geomorph.2011.12.022

Fontes, M.P.F., & Weed, S.B. (1991). Iron oxides in selected Brazilian Oxisols. I. Mineralogy. *Soil Science Society of America Journal*, 55, 1143-1149.

Gavioli, A., de Souza, E. G., Bazzi, C. L., Schenatto, K., & Betzek, N. M. (2019). Identification of management zones in precision agriculture: An evaluation of alternative cluster analysis methods. *Biosystems Engineering*, 181, 86–102. doi:10.1016/j.biosystemseng.2019.02.019

Gee G. W., & J. W. Bauder. (1986). *Methods of Soil Analysis-Part 1: Physical and Mineralogical Methods*. Madison: Soil Science Society of America.

Getirana, A. (2016). Extreme Water Deficit in Brazil Detected from Space. *Journal of Hydrometeorology*, 17(2), 591–599. doi:10.1175/JHM-D-15-0096.1

Gomes, L. C., Faria, R. M., de Souza, E., Veloso, G. V., Schaefer, C. E. G. R., & Filho, E. I. F. (2019). Modelling and mapping soil organic carbon stocks in Brazil. *Geoderma*, doi:10.1016/j.geoderma.2019.01.007

Gregorutti, B., Michel, B., & Saint-Pierre, P. (2016). Correlation and variable importance in random forests. *Statistics and Computing*, 27(3), 659–678. doi:10.1007/s11222-016-9646-1

Haghverdi, A., Leib, B.G., Washington-Allen, R.A., Ayers, P.D., & Buschermohle, M.J. (2015). Perspectives on delineating management zones for variable rate irrigation, *Computers and Electronics in Agriculture*, 117, 154–167.

Hou, D., O'Connor, D., Nathanail, P., Tian, L., & Ma, Y. (2017). Integrated GIS and multivariate statistical analysis for regional scale assessment of heavy metal soil contamination: A critical review. *Environmental Pollution*, 231, 1188–1200. doi:10.1016/j.envpol.2017.07.021

Husson F., Josse J., Lê S., & Mazet, J. (2007). FactoMineR: Factor Analysis and Data Mining with R. R package version 1.04, URL http://CRAN.R-project.org/package=FactoMineR

Husson, F., Josse, J. & Pagès J. (2010). *Principal component methods - hierarchical clustering-partitional clustering: why would we need to choose for visualizing data?* Technical report.

Husson, F., Lê, S., & Pagès, J. (2017). *Exploratory Multivariate Analysis by Example Using R*. Boca Raton: CRC Press.

IBGE. (2015). *Manual Técnico de Pedologia*. Rio de Janeiro: Instituto Brasileiro de Geografia e Estatística.

Isaaks, E.H., & Srivastava, R.M. (1989). *An Introduction to Applied Geostatistics*. New York: Oxford University Press.

Josse, J., & Husson, F. (2016). missMDA: A package for handling missing values in multivariate data analysis. *Journal of Statistical Software*, 70(1). doi:10.18637/jss.v070.i01

Kassambara, A. (2017a). *Multivariate analysis II: practical guide go principal component methods in R: PCA, M (CA), FAMD, MFA, HCPC, factoextra*. STHD. http://www.analyticsvidhya.com/blog/2016/03/practical-guide-principal-component-analysis-python/

Kassambara, A. (2017b). *Practical guide to cluster analysis in R: Unsupervised machine learning.* STHD.

Khosla, R., Westfall, D. G., Reich, R. M., Mahal, J. S., & Gangloff, W. J. (2010). Spatial Variation and Site-Specific Management Zones. In M. A. Oliver (Ed.), *Geostatistical Applications for Precision Agriculture* (pp. 195–219). doi:10.1007/978-90-481-9133-8

King, J. A., Dampney, P. M. R., Lark, R. M., Wheeler, H. C., Bradley, R. I., & Mayr, T. R. (2005). Mapping potential crop management zones within fields: Use of yield-map series and patterns of soil physical properties identified by electromagnetic induction sensing. *Precision Agriculture*, 6(2), 167–181. doi:10.1007/s11119-005-1033-4

Kuhn, M. (2012). Variable selection using the caret package. Available: http://r-forge.r-project.org/scm/viewvc.php/*checkout*/pkg/caret/inst/doc/caretSelection.pdf?revision=77&root=caret&pathrev=90. (Accessed 19 January 2020).

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

Kuhn, M., & Johnson, K. (2013). *Applied Predictive Modeling*. New York: Springer.

Lê, S., Josse, J., & Husson, F. (2008). FactoMineR : An R Package for Multivariate Analysis. *Journal of Statistical software*, 25(1), 1–18.

Li, Y., Shi, Z., Wu, H.-X., Li, F., & Li, H.-Y. (2013). Definition of Management Zones for Enhancing Cultivated Land Conservation Using Combined Spatial Data. *Environmental Management*, 52(4), 792–806. doi:10.1007/s00267-013-0124-7

Menezes, M. D. de, Bispo, F. H. A., Faria, W. M., Gonçalves, M. G. M., Curi, N., & Guilherme, L. R. G. (2020). Modeling arsenic content in Brazilian soils : What is relevant ? *Science of the Total Environment*, 712, 136511. doi:10.1016/j.scitotenv.2020.136511

Moral, F. J., & Serrano, J. M. (2019). Using low - cost geophysical survey to map soil properties and delineate management zones on grazed permanent pastures. *Precision Agriculture*. doi:10.1007/s11119-018-09631-9

Moral, F. J., Terron, J. M., & Marques da Silva, J. R. (2010). Delineation of management zones using mobile measurements of soil apparent electrical conductivity and multivariate geostatistical techniques. *Soil and Tillage Research*, 106(2), 335–343.

Nawar, S., Corstanje, R., Halcro, G., Mulla, D., & Mouazen, A. M. (2017). *Delineation of Soil Management Zones for Variable-Rate Fertilization: A Review*. In Advances in Agronomy (pp. 175–245). Elsevier Inc. doi:10.1016/bs.agron.2017.01.003

Novais, R. F., & Mello, J. W. V. (2007). Relação solo-planta. In: Novais, R. F.; Alvarez, V. H.; Barros, N. F.; Fontes, R. L. F.; Cantarutti, R. B.; Neves, J. C. L. (Ed.). *Fertilidade do Solo* (pp.133-204). Viçosa: Sociedade Brasileira de Ciência do Solo.

Pagès, J. (2004). *Analyse factorielle de données mixtes. Revue de statistique appliquée*, 52(4), 93–111.

Pagès, J. (2015). *Multiple Factor Analysis by Example Using R*. Boca Raton: Taylor & Francis.

Ping, J. L., Green, C. J., Bronson, K. F., Zartman, R. E., & Dobermann, A. (2005). Delineating potential management zones for cotton based on yields and soil properties. *Soil Science*, 170(5), 371–385. doi:10.1097/01.ss.0000169904.56743.75

Praene, J. P., Malet-Damour, B., Radanielina, M. H., Fontaine, L., & Rivière, G. (2019). GIS-based approach to identify climatic zoning: A hierarchical clustering on principal component analysis. *Building and Environment*, 164, doi:10.1016/j.buildenv.2019.106330

R Core Team, 2018. R: a language and environment for statistical R Foundation for Statistical Computing. Available. https://www.R-project.org/ (verified 23 Aug. 2018).

Rena, A.B., & Da Matta, F.M. (2002). O sistema radicular do cafeeiro: estrutura e ecofisiologia. In: Zambolin, L. (Ed.). *O estado da arte de tecnologias na produção de café* (pp. 11-92). Viçosa: Editora UFV.

Resende, M., Curi, N., Rezende, S. B., Corrêa, G. F., & Ker, J. C. (2014). *Pedologia: Base para distinção de ambientes*. Lavras: Editora UFLA

Ribeiro, A. C., Guimarães, P. T. G., & Alvarez, V. V. H. (1999*). Recomendação para o uso de corretivos e fertilizantes em Minas Gerais – 5ª aproximação*. Viçosa, MG: Comissão de Fertilidade do Solo do Estado de Minas Gerais.

Ronchi, C. P., Sousa Júnior, J. M. de, Ameida, W. L. de, Souza, D. S., Silva, N. O., Oliveira, L. B. de, et al. (2015). Morfologia radicular de cultivares de café arábica submetidas a diferentes arranjos espaciais. *Pesquisa Agropecuária Brasileira*, 50(3), 187–195. doi:10.1590/S0100-204X2015000300001

Santos, R. D. dos, Santos, H. G. dos, Ker, J. C., Anjos, L. H. C. dos, & Shimizu, S. H. (2015). *Manual de descrição e coleta de solo no campo*. Viçosa, MG: Sociedade Brasileira de Ciência do Solo.

Schemberger, E. E., Fontana, F. S., Johann, J. A., & Souza, E. G. De. (2017). Data mining for the assessment of management areas in precision agriculture. *Engenharia Agrícola*, 37(1), 185–193. doi:10.1590/1809-4430-eng.agric. v.37, n.1, p.185-193/2017.

SiBCS. (2018). *Sistema Brasileiro de Classificação de Solos*. Brasília: Embrapa.

Silva I. R, & Mendonça E. S. (2007). Matéria orgânica do solo. In: Novais R. F., Alvarez V V. H., Barros N. F., Fontes R. L. F., Cantarutti R. B, & Neves J. C. L. (Eds.), *Fertilidade do solo*. (pp. 275-374). Viçosa: Sociedade Brasileira de Ciência do Solo.

Smidt, E. R., Conley, S. P., Zhu, J., & Arriaga, F. J. (2016). Identifying field attributes that predict soybean yield using random forest analysis. *Agronomy Journal*, 108(2), 637–646. doi:10.2134/agronj2015.0222

Soil Survey Staff. (2014). Keys to soil taxonomy. Washington. doi:10.1109/TIP.2005.854494

Strobl, C., Boulesteix, A. L., Zeileis, A., & Hothorn, T. (2007). Bias in random forest variable importance measures: Illustrations, sources and a solution. *Bio Med Central Bioinformatics*, doi:10.1186/1471-2105-8-25

Tisdale, S. L., Nelson, W. L., Beaton, J. D., & Halvlin, J. L. (1993). Soil fertility and fertilizers (5th ed.). New York: MacMillan Publishing.

Vrindts, E., Mouazen, A.M., Reyniers, M., Maertens, K., Maleki, M.R., & Ramon, H., De Baerdemaeker, J. (2005). Management zones based on correlation between soil compaction, yield and crop data. *Biosystems Engineering*, 92, 419–428.

Walkley, A. & Black, J.A. (1934). An examination of the Degtjareff method for determining soil organic matter, and proposed modification of the chromic acid titration method. *Soil Science*, 37, 29-38.

Ward, J. H. (1963). Hierarchical grouping to optimize an objective function. *Journal of the American Statistical Association*, 58(301), 236-244.

Weil, R. R., & Brady, N. C. (2017). *The nature and properties of soils*. USA: Pearson Education Limited.

Willmott, C.J. (1982). Some comments on the evaluation of model performance. *Bulletin of the American Meteorological Society*, 63, 1309–1313.

Zogheib, F. F., Novo, T. A., Degler, R., & Martins, L. C. D. (2015). Projeto Fronteiras de Minas Gerais. FOLHA NOVA RESENDE SF.23-V-D-I Escala 1:100.000.

ARTICLE 2

Article prepared according to the rules of Geoderma Regional

Soil-environment and Syrah winter wine characterization to assist viticulture in southeastern Brazil

1. Introduction

Southeastern Brazil has emerged as an important region for production of high quality fine wines (Favero et al., 2011). This evolution is closely related to a new approach of vineyard management, called double pruning that allows fruits to be harvested during the winter season (Regina et al. 2011). Two annual pruning are performed: the first at the end of the winter season (August or September), allowing the vegetative development of the shoot; and the second, called yield pruning, in January, performed to induce the grape harvest during the winter period. The climatic conditions of this period enable a greater accumulation of phenolic compounds and sugars compared to summer harvests, which is the traditional time of harvesting vines in Brazil (Amorim et al., 2005; Dias et al., 2012, 2017; Favero et al., 2011; Mota et al., 2011). Wines from this region are known as “Winter Wines” due to the season when the grapes are harvested (Brant et al., 2018). The Syrah variety has been the most suitable for these conditions (Amorim et al., 2005; Favero et al., 2011).

In wine terminology, the well-known term *terroir* is traditionally used to provide the notion of agricultural sites in the same geographic area that share a similar climate, soil, and management, and their association contributes to unique characteristics in the products (van Leeuwen and Seguin, 2006). The effect of soil and climate on grapevine development and the composition of grapes or wines has been demonstrated in important viticultural regions of the world, such as in Italy (Ferretti, 2019; Priori et al., 2019), Canada (Kotsaki et al., 2020a, 2020b), Spain (Perez-Alvarez et al., 2015), Portugal (Prata-Sena et al., 2018), and China (Wang et al., 2015). While some regions have hundreds of years of defining, developing, and understanding their terroirs, e.g. Bordeaux and Champagne in France, Campania in Italy and Rioja and Galicia in Spain, new regions still face the challenge of finding the most adapted varieties and better management to define their typicity (Jones et al., 2004).

Macroclimate (regional) and microclimate (local climate at fruit zone) conditions exert great influence on vine growth, yield, grape and wine quality attributes (Van Leeuwen & Seguin, 2006; Souza et al., 2019). Soil characteristics are important drivers of microclimate conditions, driving local water availability for vines. Along with its role of providing nutrients to plants, soils directly affect the vigor of the grapevines (Leeuwen et al., 2018; Morlat and Bodin, 2006), which in turn is directly related to the composition of the berries and wines (Cortell et al., 2008). Soils without any water restrictions might result in wine production with lower added value (Renouf et al., 2010). Also, it has been reported that clayey soils were more correlated to wines with higher pH (detrimental to wines) and lower levels of anthocyanins, alcohol, color intensity, and phenolics (Wang et al., 2015). Comparing soils with different degree of weathering, Morlat and Bodin (2006) found positive characteristics of wine quality for the lesser weathered ones: berries with smaller size, with higher anthocyanin content, and lower total acidity, which are good for wine quality.

Several studies regarding agronomic and physiological evaluations to confirm the suitability of double pruning (Amorim et al., 2005; Favero et al., 2011), influence of different rootstocks (Dias et al., 2012), as well as to compare the effect of the growing seasons (Favero et al., 2011), the maintenance of the quality of grape (Favero etal., 2008) and the aromatic profile of the wines (Mota et al., 2021) have been conducted. However, there is a lack of edaphoclimatic characterization in the vineyards studied to assist wine typicity characterization. Thus, the objectives of this study were to characterize at local scale the soils and the climate, as well as to verify their relationship with the Winter Wines composition produced in seven commercial vineyards of the Syrah cultivar in southeastern Brazil.

2. Material and methods

2.1. Study areas

The study areas comprises seven commercial vineyards of the Syrah cultivars located in areas traditionally cultivated with coffee in the southeast region (Fig. 1): Três Corações – MG (TC), Cordislândia – MG (COR), Andradas – MG (AND), São Sebastião do Paraíso – MG (SSP), Três Pontas – MG (TP), Espírito Santo do Pinhal – SP (PIN), and Itobí – SP (ITO). A soil profile sampling was carried out in each vineyard, at the depths of 0-20, 40-70 and 100-120, when possible, in the central part of crop areas, where the experiment was also carried out to evaluate plant breeding practices and from where the grapes were harvested for the wine production. The vineyards are between 10 and 15 years old, grown in vertical shoot position with bilateral cordons, with 4000 plant ha^{-1}. The management of double pruning was carried out as described in Favero et al. (2011). All the cultural practices of the vineyards, including fertilization and harvesting were carried out according to each viticulturist.

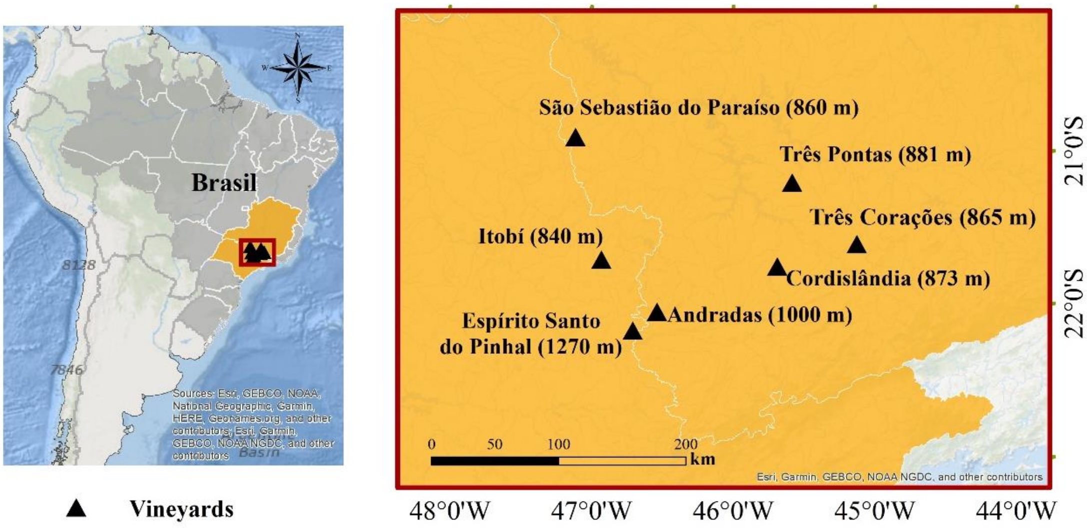

Fig. 1. The geographic location of the vineyards in Minas Gerais and São Paulo states and their respective local altitudes in parentheses.

2.2. Soil and climate characterization

The morphological description of the soil profiles was performed according to Santos et al. (2015), and soils were classified according to the Soil Survey Staff (2014). Physical, chemical, and mineralogical analysis were carried out in each soil profile following soil horizon depth. The following physical analyses were performed, according to Embrapa (1997): quantification of the gravel proportion; particle-size distribution with physical (vertical shaking) and chemical (NaOH 0,1 mol L^{-1}) dispersion on the air-dried fine earth (ADFE) ($\leq$ 2.0 mm). The particle size determination was performed by pipette method, enabling the silt/clay ratio calculation: an important value for tropical conditions that express the degree of soil weathering (the higher, the less weathered).

The following chemical analyses were carried out: pH in water; pH in KCL 1.0 mol L^{-1}; available P and K^+ exchangeable extracted by Mehlich-1; exchangeable Ca^{2+}, Mg^{2+} and Al^{3+} extracted with KCl 1.0 mol L^{-1}; potential acidity ($H^+ + Al^{3+}$) was extracted by 0.5 mol L^{-1} calcium acetate at pH 7.0; soil organic matter (SOM) was obtained by Walkley and Black (1934) method; remaining P (Rem-P) was obtained according Alvarez et al. (2000). Effective cation exchange capacity at soil pH (ECEC), cation exchange capacity at pH 7.0 (CEC pH 7.0), base saturation (BS) and aluminum saturation (AS) were then calculated. The determination of the well-crystallized Fe oxides (Fed) (Mehra and Jackson, 1960), and of low crystallinity (FeO) (McKeague and Day, 1966) was performed in order to characterize the degree of pedogenetic development of the studied soils, as well as the weathering intensity (Inda Junior and Kämpf, 2003).

Soil parent material information was assessed by the indication contained in geological map of the Minas Gerais State at the scale of 1: 1.000.000 (CPRM, 2003) and of the São Paulo State (Peixoto, 2010). The refinement of the geological map was carried out using information from the literature (Mancini et al., 2019a; Mancini, et al., 2019b; Grotzinger and Jordan, 2014), comparing with the levels of elements and oxides determined by portable x-ray fluorescence and the mineralogy of the soil fractions, as described below. The soil elemental and oxides contents in the soil samples determined by portable X-ray fluorescence (pXRF) (Bruker S1 Titan LE model). These samples were scanned in triplicate for 60 seconds using the Trace mode of Geochem software. The calibration of the pXRF was made by scanning standard reference materials 2710a and 2710b and check sample (CS). The minerals present in the clay, silt, and fine sand fractions of B horizons were identified with X-ray diffractometry, according to Embrapa (1997). For mineral identification it was used an X-ray Powder Diffraction with CuKα radiation (Ni filter and a current of 20 mA).

The climatic characterization of the grapes' maturation period (May to July) was performed based on average historical data (1982 to 2012) obtained from Climate.Data.org (https://pt.climate-data.org/).

2.3. Wine laboratory analysis

The grapes harvested from vineyards were vinified and their composition was evaluated in three harvests: 2016, 2017 and 2018. The total phenolics were analyzed by the Folin-Ciocalteau method based on a standard curve of gallic acid (Bergqvist et al., 2001). The ashes were determined by the gravimetric method and ashes alkalinity by the titrimetric method (Blouin 1992). The anthocyanin content was obtained by the differential pH method (Giusti and Wrolstad, 2001), and the result expressed in mg malvidin-3-glycoside per liter of wine. The dry extract was determined according to the AOAC method 920.62 (AOAC, 1995). Phenolics were quantified according to Amerine and Ough (1980), and flavanols content following Ribéreau-Gayon et al. (2006a). The color intensity was determined by the sum of the absorbance at 420, 520, and 620 nm (Curvelo-Garcia, 1988). The total polyphenols index (TPI) was determined at 280 nm in an UV/VIS spectrophotometer (Shimadzu UV-1800). The fixed acidity was determined according to OIV (2009) and the pH by a digital potentiometer (Micronal model B 474). The alcohol content by volume in a hydrostatic balance (Super Alcomat, Gibertini) after wine distillation (Super DEE Gibertini digital distilling unit) and sugars were determined by the Fehling method (Brasil, 1986).

2.4 Principal component analysis

The relationship among edaphoclimatic characteristics and the average composition of wines in three vintages (2016, 2017, and 2018) was assessed by the principal component analyses (PCA). The PCA was performed on two data sets: i) for A horizon soil attributes and climate data, and B horizons soil attributes and climate data. A horizon is chemically corrected, while B horizon expresses pedogenetic processes and is generally not corrected. Both are important to understand the influence of soils on wine composition. In the two data sets, soil information was used as active variables and the wine composition was plotted as supplementary variables (FactoMineR package, version 1.42) in R software (R Core Team 2018). Supplementary variables do not contribute to the construction of the principal components, they are also called illustrative variables and help in the interpretation of the plot (Husson et al., 2017).

3. Results and discussion

3.1. Soil types and physical properties

The availability of soil water is one of the most important soil characteristics that might influence terroir (Deloire et al., 2004; Van Leeuwen et al., 2018). In addition to soil texture, soil water retention depends on mineralogy (due to soil charge generation), soil depth, structure, and relief (Van Leeuwen et al., 2004; Resende et al., 2014). Some of those attributes are also criteria for soil types characterization (Santos et al., 2014; Soil Survey Staff, 2014), favoring interpretations on water availability and the behavior of plant roots. Thus, four soil types were found in the vineyards: Acrudox and Hapludox (differed by moisture regime, which in turn influence color) → Hapludult → Eutrudept. Following this chronological sequence, there is a decrease in soil depth, water storage, root penetration, and reduction of degree of weathering.

The variation in soil texture is due to differences in the parent material and, mainly, the degree of weathering of the soil (Resende et al., 2014). Soil texture classes varied between clay and sandy clay loam (Table 1). The most weathered soils (Acrudox and Hapludox), regardless of the parent material, presented homogeneous classes of texture throughout the soil profile. The Hapludult, less weathered in relation to Acrudox and Hapludox, showed a difference in textural class between the superficial and subsurface horizon. Eutrudept, however, despite the difference in textural class, has a very similar texture in the evaluated profiles.

The highest clay contents in the A horizon occur in TC and TP soils. The highest levels of sand on the surface occur in the soils of AND, PIN, and ITO which in turn are also one of those with the lowest levels of clay in the A horizon. Besides, these soils have gravel at all depths, contributing to reduce water storage. This is an important characteristic in vineyards, as it can reduce water storage at greater depths (Resende et al., 2014), and thus promote a certain water deficit to plants. In fact, evaluations carried out on plants in these vineyards showed moderate water stress in AND and weak water stress in PIN and ITO vineyards (Brant et al., 2021). Similarly, gravely soils of the viticultural region of Bordeaux region traditionally produces high-quality wines (cru classé), since this characteristics promote greater soil drainage (Seguin, 1986). TC, COR, AND, PIN, and TP presented higher levels of clay in depth. The SSP textural class is clay loam throughout the soil profile.

Table 1 Physical and morphological properties of the studied soils.

Hz.	Clay	Silt	Sand	CS	FS	TR	S/C	Gravel	Textural class	Structure*
	--------------------- dag kg^{-1} ----------------							g kg^{-1}		
P1 - ACRUDOX (TC)										
Ap	50	26	24	8	16	1.86	0.52	-	Clay	
Bo1	52	28	20	7	13		0.54	-	Clay	granular
Bo2	58	21	21	7	14		0.36	-	Clay	
P2 - ACRUDOX (COR)										
Ap	36	35	29	14	15	1.92	0.97	-	Clay loam	
Bo1	69	9	22	11	11		0.13	-	Clay	granular
Bo2	69	11	20	9	11		0.16	-	Clay	
P3 - HAPLUDULT (AND)										
Ap	30	20	50	35	15	1.77	0.67	52	Sandy clay loam	
Bt1	53	11	36	26	10		0.21	25	Clay	block
Bt2	53	13	34	25	9		0.25	163	Clay	
P4 -HAPLUDULT (PIN)										
A	40	9	51	38	13	1.25	0.23	36	Sandy clay	
Bt1	44	12	44	33	11		0.27	58	Clay	block
Bt2	50	12	38	28	10		0.24	96	Clay	
BC	49	12	39	30	9		0.24	131	Clay	
P5 - EUTRUDEPT (ITO)										
A	40	18	42	27	15	0.95	0.45	78	Clay	
Bw	38	21	41	26	15		0.55	109	Clay loam	block
P6 - ACRUDOX (SSP)										
Ap	38	24	38	7	31	1.00	0.63	1	Clay loam	
Bo1	38	26	36	7	29		0.68	-	Clay loam	granular
Bo2	40	27	33	6	27		0.68	-	Clay loam	
P7 - HAPLUDOX (TP)										
Ap	49	29	22	6	16	1.02	0.59	-	Clay	
Bo1	50	28	22	7	15		0.56	-	Clay	granular
Bo2	50	30	20	5	15		0.60	-	Clay	

Hz.: Soil horizon; CS: coarse sand; FS: fine sand; TR: textural relationship: clay content of B horizon/clay content of A horizon; S/C: relationship between silt and clay content; * B horizon structure.

3.2. Soil fertility analysis

It was found a wide variation of pH and fertility attributes between depths and vineyards, being in general a much greater suitability found in A horizon when compared with B horizon (higher pH, SOM, K, P, Ca^{2+}, and Mg^{2+}) (Table 2). It is important to highlight that previous soil use in TP (horticulture) might be responsible for the higher K content due to higher crop nutrients requirement. The COR presented the highest CEC for all horizons. The soil pH in H_2O ranged from 4.5 in the Bt1 horizon of the PIN soil to 7.60 in the Ap soil horizon in COR. In general, the acidity is lower in the A horizons than in B horizons, which is corroborated by both the low values of Al^{3+} and the high values of soil base saturation (BS). This is due to the practice of liming held in areas of vineyards without incorporation, since vines are perennial crop. In addition, TC and SSP presented low adequacy of base saturation for all depths, since 80% is considered ideal for grapevine crops (Ribeiro et al., 1999).

Since SOM is the largest reservoir of N for plants, and is even used as a basis for recommending nitrogen fertilization in some states in Brazil (Cantarella, 2007), as the present work seeks to perform a more general characterization with variables that are more consistent over time, we considered here that the SOM content reflects the N content. In this sense, the TC and COR soils have the highest N levels in the A and Bo horizons.

Table 2 Soil fertility analysis of the studied vineyards

Hz	pH (KCl)	pH (H_2O)	ΔpH	K	P	Na	Ca	Mg	Al	$H+Al^{3+}$	SB	ECEC	CEC	BS	AS	SOM	Rem P	Zn	Fe	Mn	Cu	B	S
				------- mg kg^{-1} -------			--------------------- cmolc kg^{-1} ---------------------							--- % ---		dag kg^{-1}	mg L^{-1}	--------------------- mg kg^{-1} ------------------					
										P1 - ACRUDOX (TC)													
Ap	4.8	5.6	-0.8	262.8	22.2	11.3	2.9	0.5	0.1	4.9	4.1	4.3	9.1	45.5	3.1	3.2	25.1	11.6	39.8	17.2	13.8	0.52	12.1
Bo1	5.4	5.3	0.0	14.8	1.4	5.1	0.8	0.3	0.1	2.6	1.1	1.2	3.7	30.5	5.8	1.7	3.7	0.3	25.0	1.3	1.8	0.12	43.8
Bo2	6.2	6.1	0.1	27.7	0.5	5.1	0.8	0.2	0.0	1.8	1.0	1.0	2.8	36.2	0.0	1.0	2.9	0.1	15.0	0.8	1.3	0.12	38.7
										P2 - ACRUDOX (COR)													
Ap	6.8	7.6	-0.8	247.0	44.1	10.3	7.7	2.8	0.1	1.4	11.1	11.2	12.5	89.1	0.5	3.6	20.9	16.1	25.8	20.8	10.8	0.15	5.7
Bo1	4.6	5.1	-0.5	52.4	0.8	6.1	1.3	0.6	0.2	4.9	2.0	2.2	7.0	29.2	8.1	1.5	9.8	0.1	36.0	3.6	1.5	0.14	49.5
Bo2	4.8	5.4	-0.6	21.7	0.7	8.2	1.3	0.9	0.1	4.2	2.2	2.3	6.5	34.5	4.3	1.2	9.5	0.1	20.7	3.0	1.3	0.10	52.8
										P3 - HAPLUDULT (AND)													
Ap	6.7	7.5	-0.9	188.7	158.8	32.0	7.8	1.7	0.0	1.3	9.9	10.0	11.2	88.3	0.3	2.6	35.2	13.6	39.0	110.8	6.4	0.31	6.2
Bt1	4.3	4.9	-0.6	60.3	0.9	6.1	1.3	0.5	0.4	3.9	2.0	2.4	5.9	33.5	16.6	0.6	11.0	0.2	50.5	3.7	2.5	0.08	44.4
Bt2	4.8	5.1	-0.3	56.3	0.9	7.2	0.9	0.4	0.1	2.9	1.4	1.5	4.3	32.1	9.2	0.4	6.2	0.2	34.2	4.3	2.1	0.08	48.4
										P4 -HAPLUDULT (PIN)													
A	6.1	6.8	-0.7	165.0	4.0	5.1	3.6	1.7	0.1	1.6	5.7	5.8	7.3	77.8	0.9	2.1	33.2	2.1	174.6	16.2	1.8	0.15	4.0
Bt1	4.2	4.6	-0.4	22.7	0.8	4.1	0.6	0.4	0.7	3.9	1.0	1.7	4.9	20.5	41.3	0.9	25.1	0.1	87.2	1.6	0.7	0.06	36.6
Bt2	4.2	4.5	-0.4	38.5	0.9	6.1	0.5	0.4	0.7	3.8	1.1	1.7	4.8	21.9	39.1	0.7	15.7	0.1	24.9	1.3	0.7	0.08	45.9
BC	4.9	4.8	0.1	49.4	0.5	11.3	0.4	0.3	0.1	2.1	0.9	1.0	3.0	29.4	8.4	0.3	14.2	0.1	15.4	2.8	0.3	0.07	44.9
										P5 - EUTRUDEPT (ITO)													
A	5.0	6.2	-1.2	168.0	156.3	11.3	3.7	1.2	0.0	3.3	5.3	5.4	8.6	62.0	0.7	1.5	39.6	10.2	43.8	30.2	3.5	0.20	3.3
Bw	5.6	6.4	-0.8	69.2	1.8	11.3	2.6	0.8	0.0	1.7	3.6	3.6	5.3	67.3	0.8	0.4	18.8	1.3	50.7	13.9	0.6	0.08	15.7
										P6 - ACRUDOX (SSP)													
Ap	4.5	5.6	-1.2	126.5	20.9	6.1	1.5	0.4	0.2	4.5	2.2	2.4	6.7	32.8	7.6	2.3	26.5	9.9	38.9	20.7	7.2	0.10	15.4
Bo1	6.6	7.4	-0.9	43.5	0.3	5.1	1.6	0.5	0.0	1.6	2.2	2.2	3.7	57.8	0.9	1.4	6.6	0.1	53.1	10.4	5.7	0.08	13.5
Bo2	6.3	6.7	-0.4	27.7	0.1	5.1	1.1	0.4	0.0	1.7	1.6	1.6	3.3	48.7	2.5	1.0	4.4	0.1	54.0	7.5	5.2	0.06	42.2
										P7 -HAPLUDOX (TP)													
Ap	5.9	6.5	-0.6	288.5	540.2	21.7	5.3	1.1	0.0	2.3	7.1	7.2	9.5	75.5	0.6	2.3	35.9	17.1	35.2	23.8	6.1	0.29	9.4
Bo1	4.7	5.0	-0.3	140.3	0.4	8.2	1.1	0.3	0.2	3.2	1.7	1.9	4.9	35.3	8.0	1.4	11.1	0.1	26.5	2.9	1.4	0.21	51.7
Bo2	5.9	6.1	-0.2	115.6	0.1	10.3	1.3	0.2	0.0	1.8	1.8	1.8	3.7	49.5	1.6	0.6	3.0	0.1	25.2	3.3	1.4	0.23	50.8

Hz - Soil Horizon; SB: sum of bases; ECEC - effective cation exchange capacity; CEC - cation exchange capacity at pH 7.0; BS - base saturation; AS - aluminum saturation; SOM - soil organic matter; Rem P - remaining phosphorus

3.3. Soil mineralogical composition

The mineral identification by means of X-ray diffraction (XRD) was performed in natural (Fig. 2) and treated clay particles for the concentration of oxides (Fig. 3) of B horizon and natural clay. The only soils with presence of primary minerals other than quartz in the silt and sand fractions are ITO and AND (Fig. 2). The sand fraction of ITO presents orthoclase (Fig. 2e), a primary mineral of feldspar group, whose presence denotes reserve of soil K^+ (Melo et al., 2009). The silt fraction of AND, in turn, presented peaks related to micas, which also denotes K^+ reserve (Fig. 2c).

The treatment enabled better oxide identification, a mineral group that tends to remain stable in soils under tropical conditions. It was found a marked presence of hematite (Hm) and goethite (Gt) in all of the soils. In addition to those Fe oxides, well-formed peaks of maghemite (Mh) were found in SSP, COR and TP (Fig. 3). It was not enough sample to perform this analysis on the clay fraction of TC, but from the natural clay analysis, it was possible to notice Hm and Gt. In addition to the occurrence of these Fe oxides, the occurrence of gibbsite (Gb) and kaolinite (Kt) was observed in all soils (Fig.2).

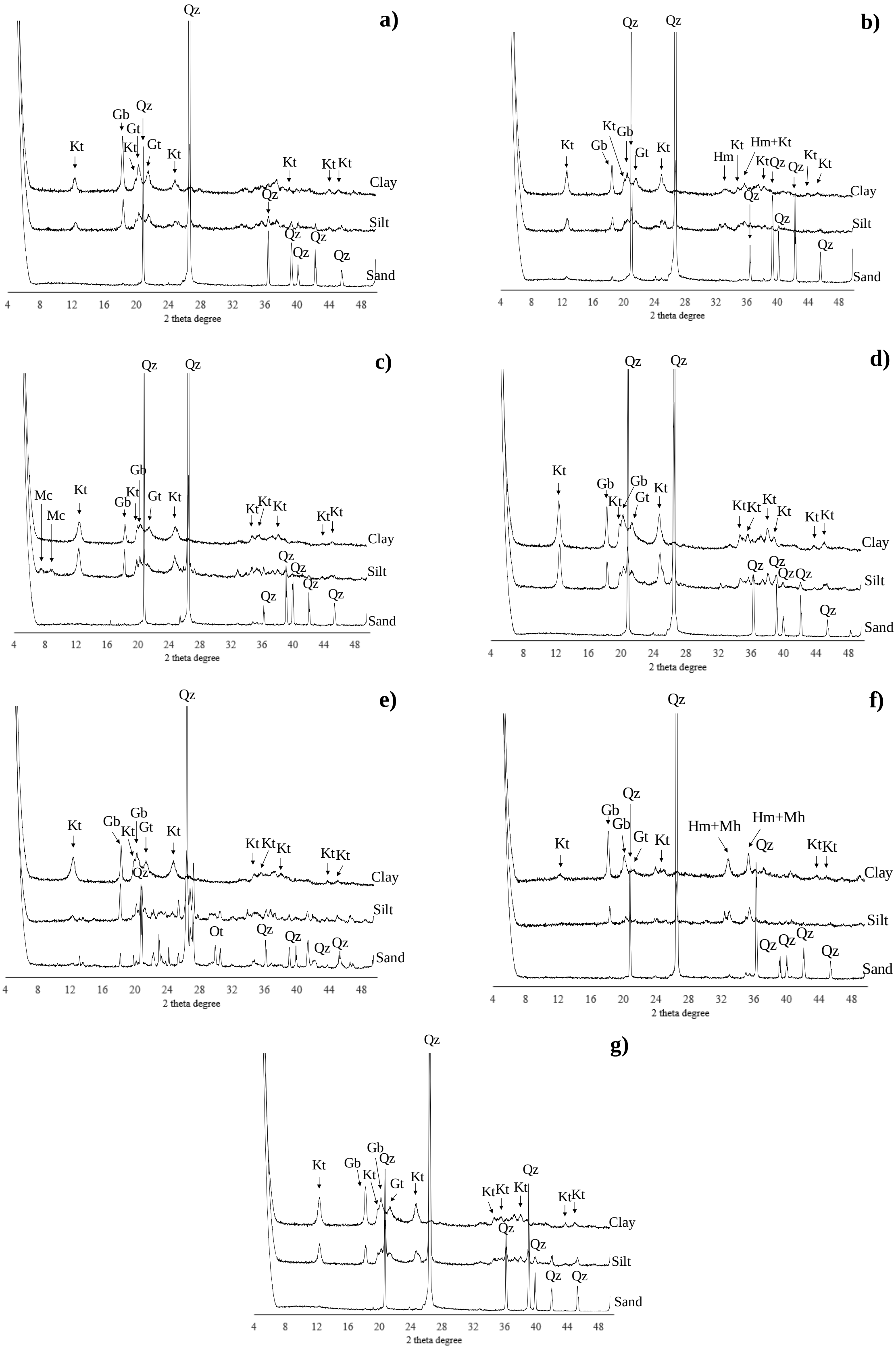

Fig. 2. Mineralogical composition of the B horizon of soils. a) TC-Acrudox; b) COR – Acrudox; c) AND – Hapludult; d) PIN – Hapludult; e) ITO – Eutrudept; f)SSP – Acrudox; g) TP – Hapludox. Kt: kaolinite; Gb: gibbsite; Gt: goethite; Hm: hematite; Mh: maghemite; Qz: quartz; Ot: orthoclase; Mc: mica.

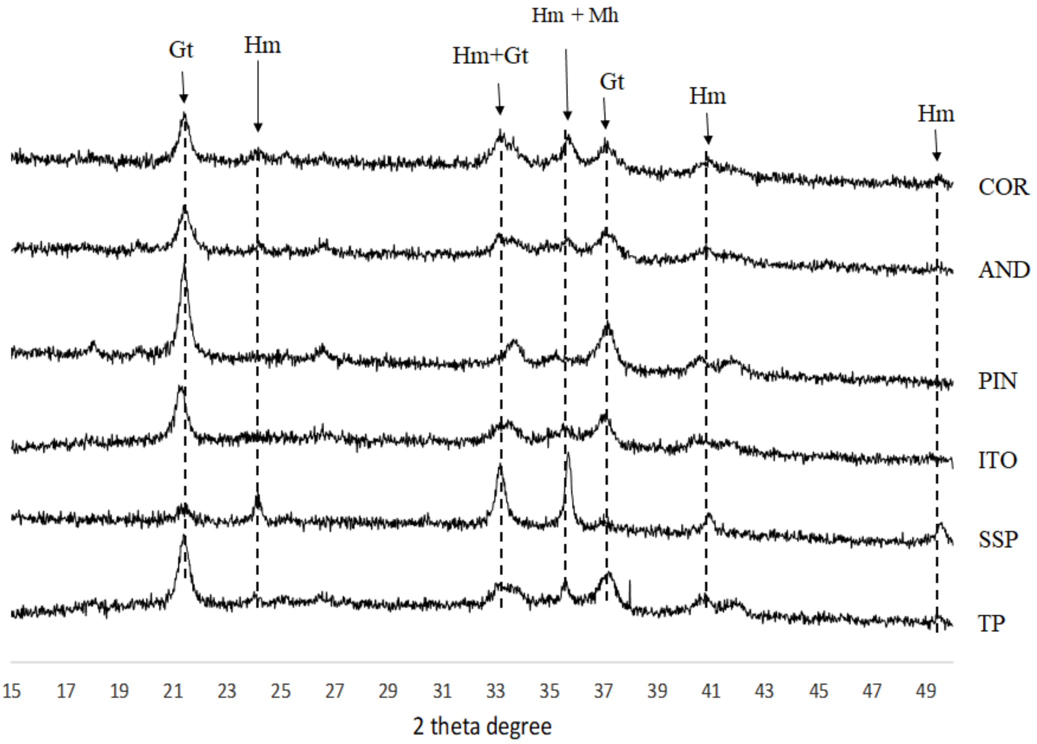

Fig. 3. X-ray diffractograms of iron oxides concentrated of clay from B horizons of the studied soils. COR – Acrudox; AND – Hapludult; PIN – Hapludult; ITO – Eutrudept; SSP – Acrudox; TP – Hapludox. Gt: goethite; Hm: Hematite; Mh: Maghemite.

3.4. Soil parent material and oxides as indicator of soil weathering degree

The soils of this study were formed *in situ*, in which the geology can also be considered the soil parent material. This information is important for viticulture, considering: a) it exerts influence on soil characteristics that govern water dynamics as well as nutrient availability, especially in soils less weathered whose characteristics still remain similar to those of their parent material (Bodin and Morlat, 2006; Huggett, 2006; Morlat and Bodin, 2006); b) wine market is traditionally ruled on geological/soil parent material, whose information is often found in wine bottle as a guide for consumers (Huggett, 2006).

Soils with different characteristics could be formed over the same parent material (Resende et al., 2014; Silva et al., 2019). Since soils are an open system, in which different pedogenic processes occurs, in order to better characterize soils by means of chemical elements that work as parent material tracers, pXRF spectrometry was applied. Also, such analysis could assist the *terroir* characterization. pXRF analysis has been successfully applied in Brazilian soils to tell soil parent materials apart with adequate accuracy (Mancini et al., 2019a; Mancini et al., 2019b). Figs. 4, 5, and 6 show the total chemical composition of soils obtained from pXRF.

Five different parent materials were found in the vineyard soils. The parent material of TC is biotite schist/gneiss, rich in biotite that is a primary mineral constituted by K, Mg, Fe, Cu, Mn, among other elements (Melo et al., 2009). However, the signs of occurrence of this easily weathered mineral and the reserve of nutrients that it can represent are not remarkable on the soil. This is due to the intense degree of weathering-leaching, corroborated by the lowest Feo (Inda Junior and Kämpf, 2003) (Table 4). The same effect of weathering on the parent material occurs in COR, AND, PIN, and TP (Melo et al., 2009).

Pyroxene Granulite, the soil parent material of the COR vineyard, consists of a metamorphic rock containing mainly felsic minerals such as quartz and feldspars, and the ferromagnesian mineral pyroxene (Haldar and Tišljar, 2014). Pyroxene is a mineral that is not very resistant to weathering, so much so that it was not observed in the mineralogical composition of the sand fraction of this soil (Fig. 2b). The pXRF results showed low levels of most trace elements and other elements originating from ferromagnesian minerals (Melo et al., 2009). It was found a predominance of SiO_2 and Al_2O_3.

The soils of AND and TP have gneiss as their parent materials. In these soils, in relation to the other elements, SiO_2 mostly constitutes this felsic rock. High levels of this oxide were found in both soils (Fig. 4). The AND soil presented higher levels of Sr and K_2O, while TP is among the highest levels of Al_2O_3. All of these elements and oxides are typical of gneiss rock (Grotzinger and Jordan, 2014; Mancini et al., 2019b).

The highest contents of K_2O, Ba, Ce, Mo, Nb, Rb, Y, Zn, and Zr were found in ITO, likely due to the low degree of soil weathering-leaching. Although PIN and ITO soils were formed by the same type of parent material (granite) they presented a quite different chemical composition and physical characteristics, due to greater pedogenetically development of PIN.

The SSP soil has as its parent material, a mixture of basalt and sandstone, with the predominance of the first rock. Basalt consists of ferromagnesian minerals such as pyroxene, hornblende, olivines, and Ca-plagioclases (Grotzinger and Jordan, 2014). The total content of Fe_2O_3, Cu, and Pb was higher in this soil. Also, the soil of SSP presented Fe content higher than the others (determined by pXRF and extracted by sodium dithionite (Fed) and ammonium oxalate (Feo) (Table 4).

Table 4 Fe content extracted by dithionite (Fed) and oxalate (Feo) in the clay fraction of the studied soils

Horizon	Fed (g kg^{-1})	Feo (g kg^{-1})	Feo/Fed
	P1 - ACRUDOX (TC)		
Ap	125.8	1.3	0.01
Bo1	101.8	1.1	0.01
Bo2	109.8	1.0	0.01
	P2 - ACRUDOX (COR)		
Ap	113.8	2.7	0.02
Bo1	120.5	2.6	0.02
Bo2	121.4	3.1	0.03
	P3 - HAPLUDULT (AND)		
Ap	84.8	4.1	0.05
Bt1	96.0	2.6	0.03
Bt2	95.3	2.5	0.03
	P4 -HAPLUDULT (PIN)		
Ap	70.2	2.6	0.04
Bt1	62.3	0.9	0.01
Bt2	54.1	0.8	0.01
BC	77.2	1.0	0.01
	P5 - EUTRUDEPT (ITO)		
Ap	82.6	3.1	0.04
Bw	106.4	1.5	0.01
	P6 - ACRUDOX (SSP)		
Ap	273.1	4.0	0.01
Bo1	238.6	3.8	0.02
Bo2	242.4	3.7	0.02
	P7 -HAPLUDOX (TP)		
Ap	78.2	2.1	0.03
Bo1	77.6	1.3	0.02
Bo2	76.3	1.0	0.01

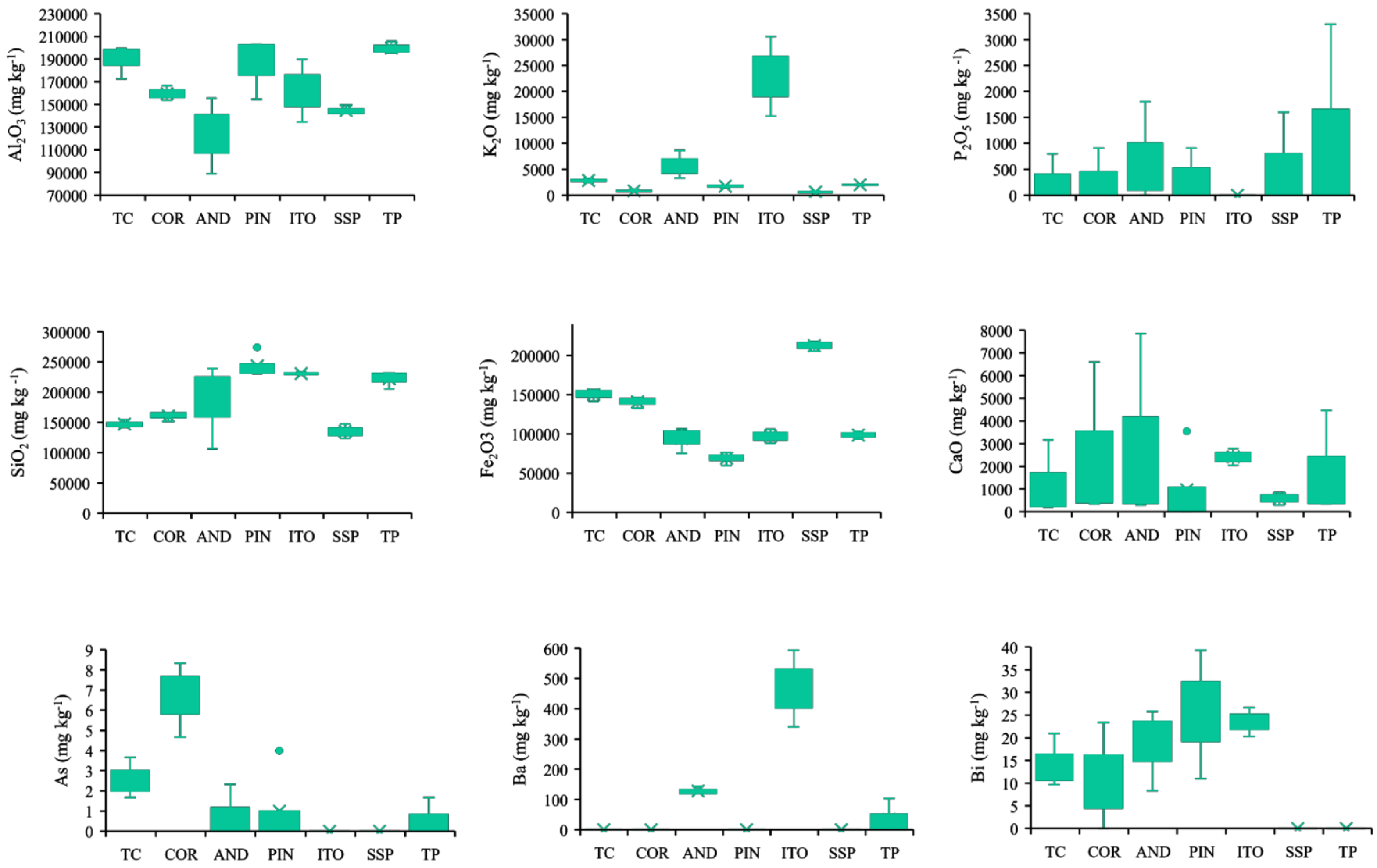

Fig. 4. Boxplot comparing elemental and oxides contents determined via portable X-ray fluorescence (pXRF) spectrometry in air-dried fine earth of the horizons of the soils of the vineyards in the southeastern region, Brazil. TC-Acrudox; COR – Acrudox; AND – Hapludult; PIN – Hapludult; ITO – Eutrudept; SSP – Acrudox; TP – Hapludox.

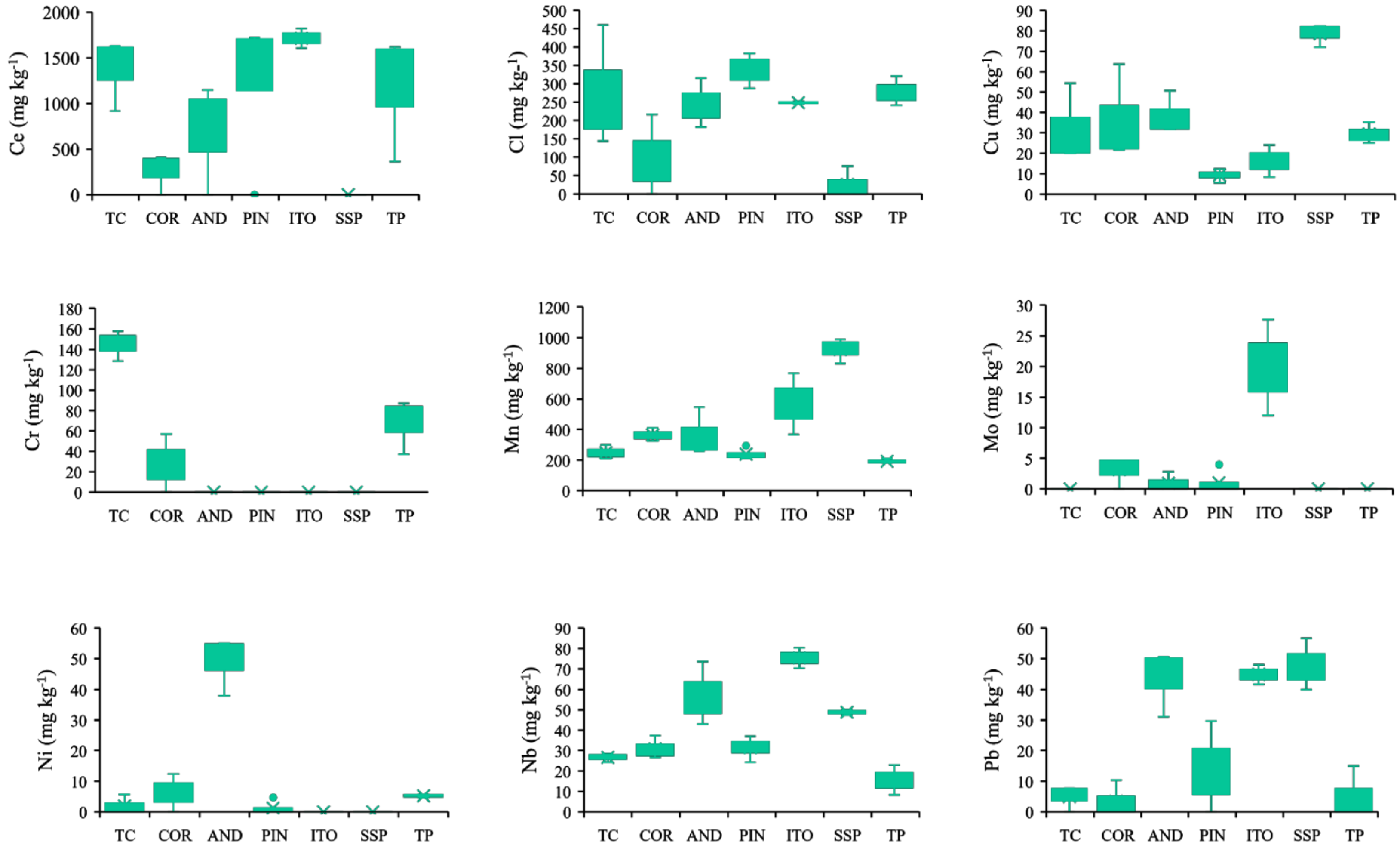

Fig. 5. Boxplot comparing elemental and oxides contents determined via portable X-ray fluorescence (pXRF) spectrometry in air-dried fine earth of the horizons of the soils of the vineyards in the southeastern region, Brazil. TC-Acrudox; COR – Acrudox; AND – Hapludult; PIN – Hapludult; ITO – Eutrudept; SSP – Acrudox; TP – Hapludox.

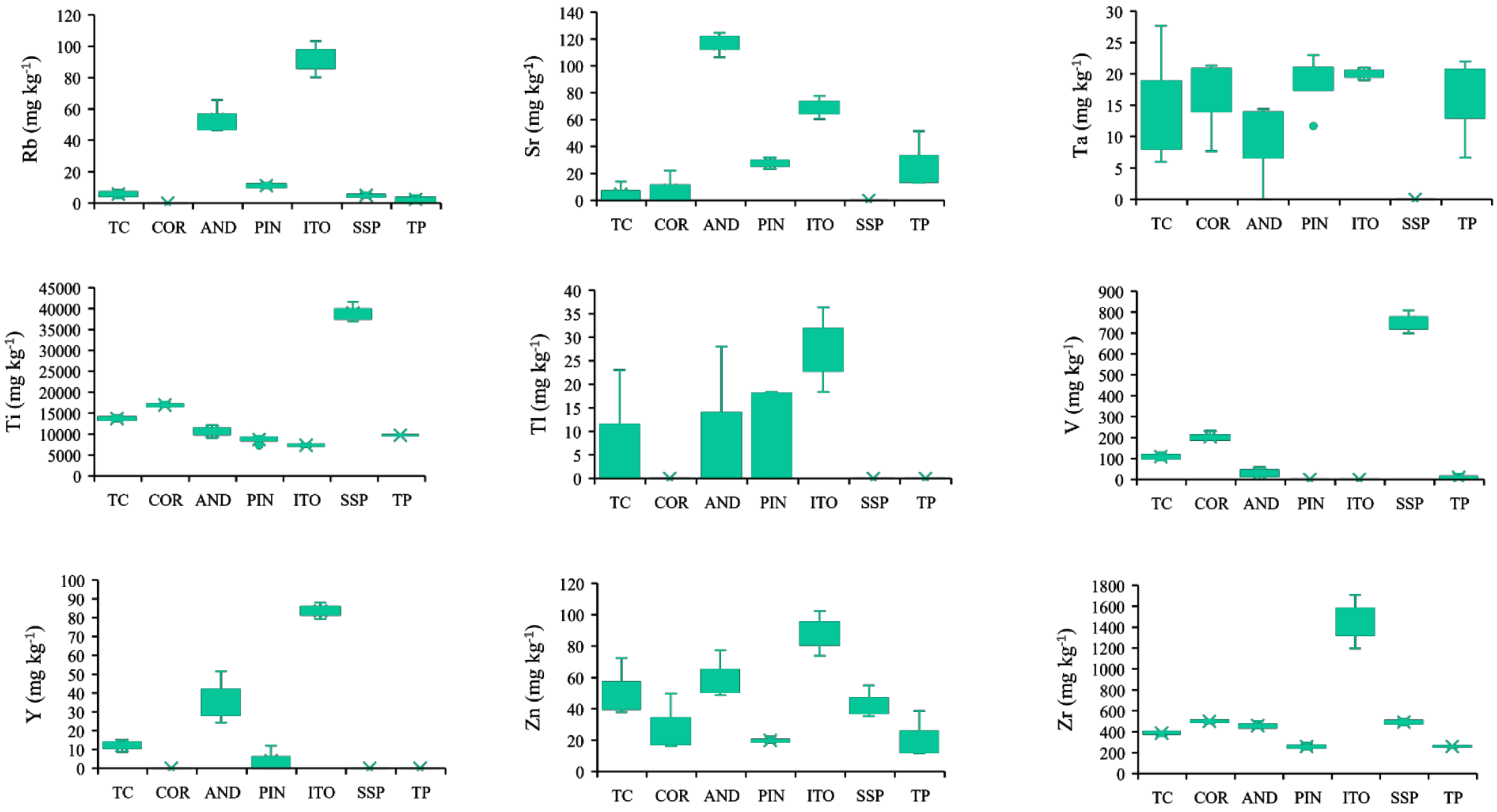

Fig. 6. Boxplot comparing elemental and oxides contents determined via portable X-ray fluorescence (pXRF) spectrometry in air-dried fine earth of the horizons of the soils of the vineyards in the southeastern region, Brazil. TC-Acrudox; COR – Acrudox; AND – Hapludult; PIN – Hapludult; ITO – Eutrudept; SSP – Acrudox; TP – Hapludox.

3.5. Climatic characterization

The climate influences the development of the vine, this in turn will affect the quality of the grapes and wines (Van Leeuwen et al., 2004). The low rainfall during the ripening period of the grapes is positive for the accumulation of sugars and phenolic compounds in the grapes (Amorim et al., 2005). In turn, high thermal amplitudes are favorable to the accumulation of phenolic compounds, directly related to the composition and quality of wines (Conde et al., 2007). Up to certain limits, higher temperatures during the maturation period can result in an increase in the pH of wines (Brant et al., 2021; Conde et al., 2007), favoring their susceptibility to oxidation and damage caused by microorganisms (Van Leeuwen et al., 2018).

The accumulated rainfall during the maturation period of vines grown under double pruning varies between 72 and 115 mm (Fig. 7a). The COR, AND, and PIN vineyards have the highest accumulated volume during the period considered. The thermal amplitude varies between 12.6 and 15.4 °C, with the highest thermal amplitudes for TC, COR, and TP (Fig. 7b). AND, PIN, and SSP vineyards have milder atmospheric temperatures with the maximum that do not reach 24 °C. ITO presents a warmer climate from May to July, with higher values of both the maximum temperature and the minimum temperature.

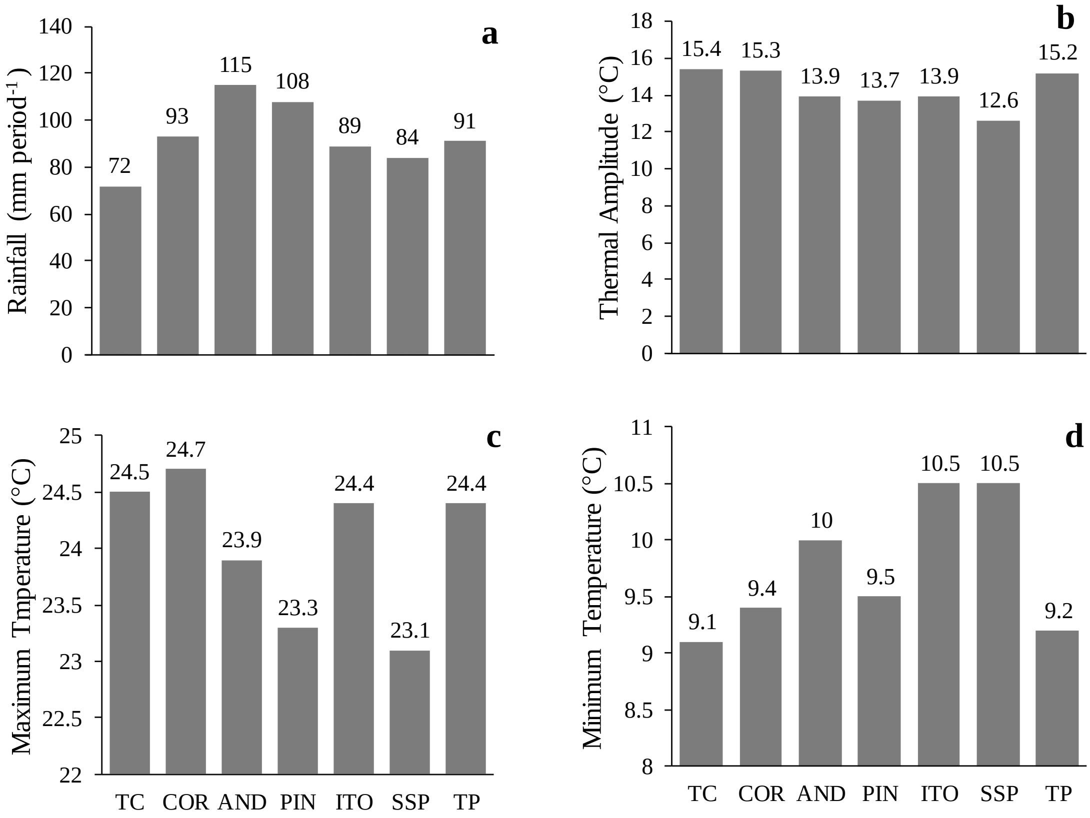

Fig. 7. Climatic characteristics of the vineyards studied during the maturation period (May to July).

3.6. Principal component analysis and the relationship between soil, climate and wine characteristics

As the combinations of physical, chemical, and mineralogical properties of soil and climate influence the quality of wines in different ways, PCA analysis was carried out to group the vineyards according to soil-environmental similarities. PCA analyses were performed from information of A and B horizon separately (Fig. 8 and 9, respectively), in order to best characterize soil-environment scenarios, since: a) soil chemical-physical characteristics markedly differ in depth, due to human (amendments application) and environmental factors (soil pedogenesis promote differentiation on soil structure, SOM, clay accumulation in depth, and others); b) most of the fine roots are found in the first 60 cm of depth, reaching A and B horizon (Smart et al., 2006); c) soil classification was an important source of information, and considering the soil types found, information from the subsurface horizons govern the classification in the first categorical level (order) (Soil Survey Staff, 2014); d) soil air is different from the air of the atmosphere. Even under the same atmospheric conditions, soil behave differently due its characteristics as color, water and coarse fragments content (govern soil temperature variations), or shape of structure, texture, charges or depth (govern water movement). Thus, we expected different relation and interaction between soil x climate on PCA analysis at either horizon. In order to facilitate understanding and, mainly, considering that B horizon govern the classification in the first categorical level (order) (Soil Survey Staff, 2014), the discussion of the groups of vineyards was carried out based on the groups formed by PCA with the data from this horizon.

Both climatic and physical and chemical characteristics of the soil were important for grouping the vineyards, as will be discussed below. The definition of the groups was based on the proximity of the vineyard principal components in the plots, represented by different formats (Fig. 8 and Fig. 9). It is possible to notice a differentiation of wine composition among groups formed from soil-environment characteristics (Fig. 10), pointing out a good performance of PCA on determined typical combination of soil-environment, that also tell apart wine tipicity.

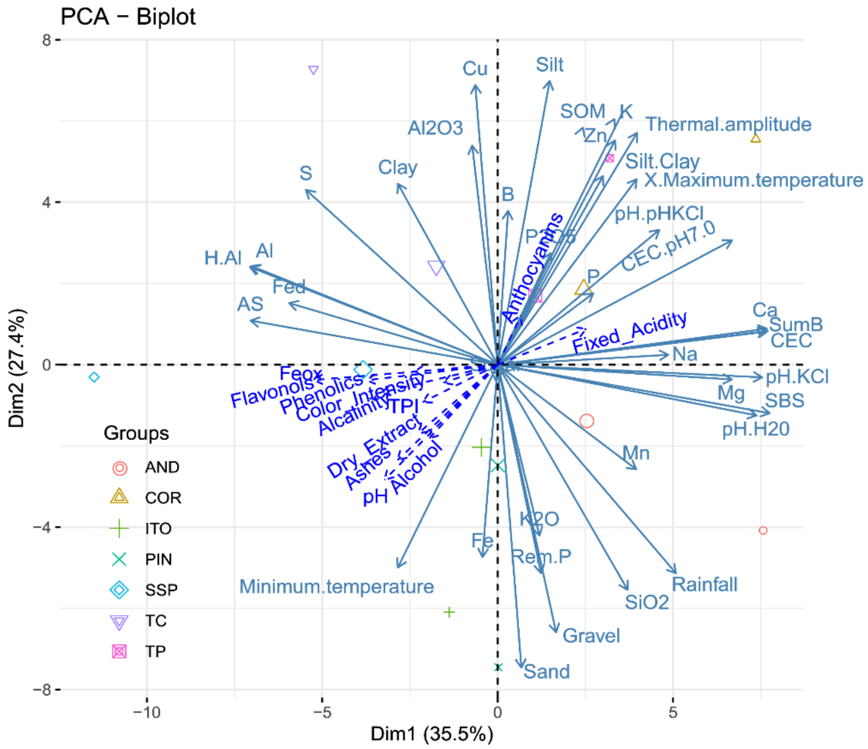

Fig. 8. Biplot of first principal component 1 (Dim1) and second principal component (Dim 2). A horizon data and climatic data as active variables and wine composition as supplementary variables.

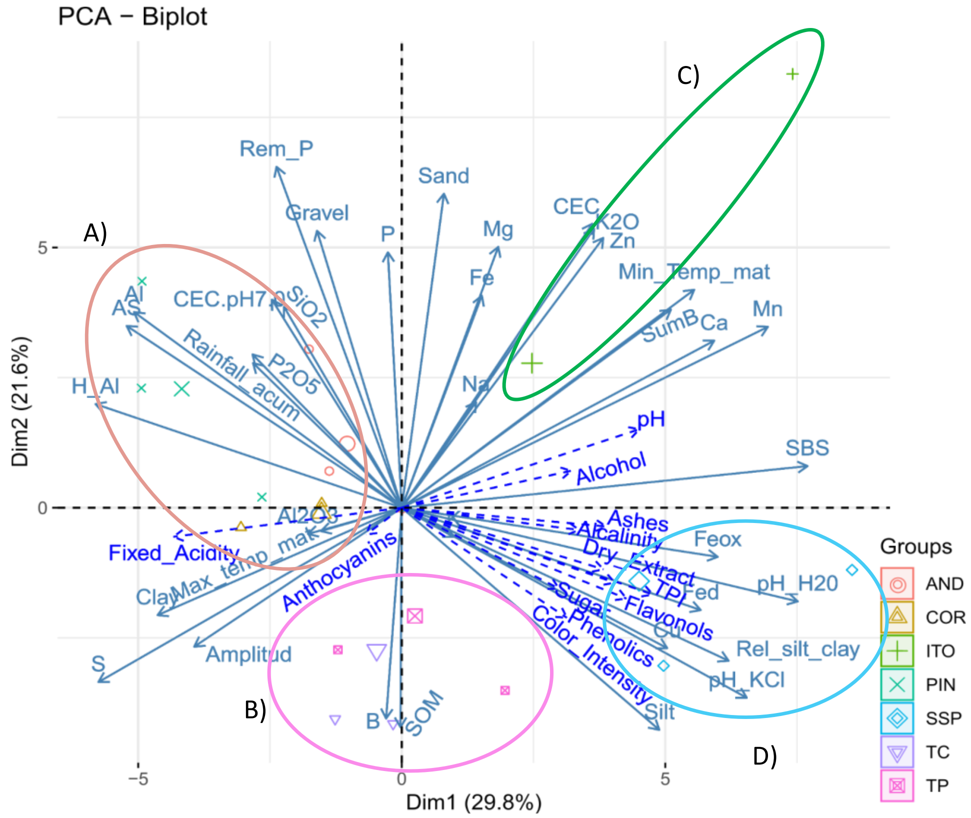

Fig. 9. Biplot of first principal component 1 (Dim1) and second principal component (Dim 2). B horizon data and climatic data as active variables and wine composition as supplementary variables.

Group A) Vineyards in COR, AND, and PIN: the soil characteristics of A horizons are quite different (Fig. 9), mainly due to differences in management. The B horizon presented greater correlation with elements related to Al^{3+} and AS, with higher content values observed in PIN. The climatic variable most related to this group was the accumulated rainfall, which is highest when compared with other groups (Fig. 8). Fig. 10 shows that the wines produced presented the lowest ashes alkalinity, and the lowest total polyphenols index (TPI), differing from other groups formed. In addition, the wines presented lower pH values and higher fixed acidity, which might be due to higher sand surface content leading to the less water retention in AND and PIN. Malic acid synthesis is higher when there is a limitation of water supply (Morlat and Bodin, 2006).

Group B) TP and TC vineyards: PCA shows that these vineyards have in common the higher levels of SOM and B in the subsurface. Higher SOM levels may have guaranteed greater vine yield, as demonstrated by phytotechnical work carried out in this area (Brant et al., 2021). Another similarity of soils belonging to this group is the clay textural class throughout the soil profile. These vineyards have a similar higher thermal amplitude (15.4 °C in TC and 15.2 ° C in TP) (Fig. 7b). Most parameters of wine composition showed great variation and intermediate values in relation to the other groups. The most similar parameters for TC and TP respectively are alcohol (13.85 and 14.20%), TPI (58.14 and 57.56), and sugars (2.94 and 3.31 g L^{-1}). TC wine has a fixed acidity of 6.07 and TP of 5.04 g L^{-1}, whereas the pH is 3.73 and 4.17 respectively. The lower acidity of TP wine is a result of the high levels of soil K^{+} (Van Leeuwen et al., 2018), since it tends to decrease the concentration of free acids in wines (Kodur, 2011). The color intensity is also lower in soils with high levels of such nutrient, as was verified in TP (Mpelasoka et al., 2003).

Group C) Vineyard in ITO: from PCA is possible to notice a clear differentiation of this vineyard soil, in which the marked differences of morphological attributes is the lesser thickness and the lesser pedogenetic development. This soil has high levels of subsurface sand, higher CEC (effective), higher Ca^{2+}, Mg^{2+}, and higher total K_2O content determined by pXRF. Contrary to the parent material of the other soils, the field work allowed the visualization of the altered granite in 70 cm of deep. The wine from this vineyard presented the highest flavanols content. Also, it was found high values of dry extract, color intensity, TPI, alcohol content, and sugars (Fig. 10). The high wine pH might be due to the high average atmospheric temperatures during the maturation period that favor the degradation of malic acid (Brant et al., 2021; Conde et al., 2007). The high residual sugar content, in addition to the high temperatures, can occur both due to the low precipitation (Amorim et al., 2005) and due to the lower soil water retention (Van Leeuwen et al., 2009). In addition, the high alcohol formation in this case inhibits the activity of the yeasts and a higher residual sugar content remains. Similar to this vineyard, grapes with high oenological potential also have showed high levels of TPI, produced in soils with less weathered material in France (Morlat and Bodin, 2006).

Group D) Vineyard in SSP: despite being an Acrudox, a soil type that was also found in other vineyards, it has a different parent material (mixture of basalt and sandstone). The subsurface horizons of this soil were correlated with higher pH values in water. Since the harvest was late in this vineyard, it cannot be said that the greatest correlation of wine compounds with this vineyard occurs due to edaphoclimatic factors. (Fig. 8 and Fig. 9). This is because the late harvest in this vineyard caused greater dehydration of the grapes and the consequent concentration of most of these compounds (Brant et al., 2021).

Comparatively, the winter wine composition related to wine quality is within the variation of the composition of wines produced in important viticultural regions of the world. Especially considering the alcohol content and pH similar to those found in California (Brillante et al., 2018); alcohol content, anthocyanins and TPI that occur in wines from

Greece (Koundouras et al., 2006); anthocyanins (Ristic et al., 2007) and alcohol and pH determined in wines from Italy (Priori et al., 2019).

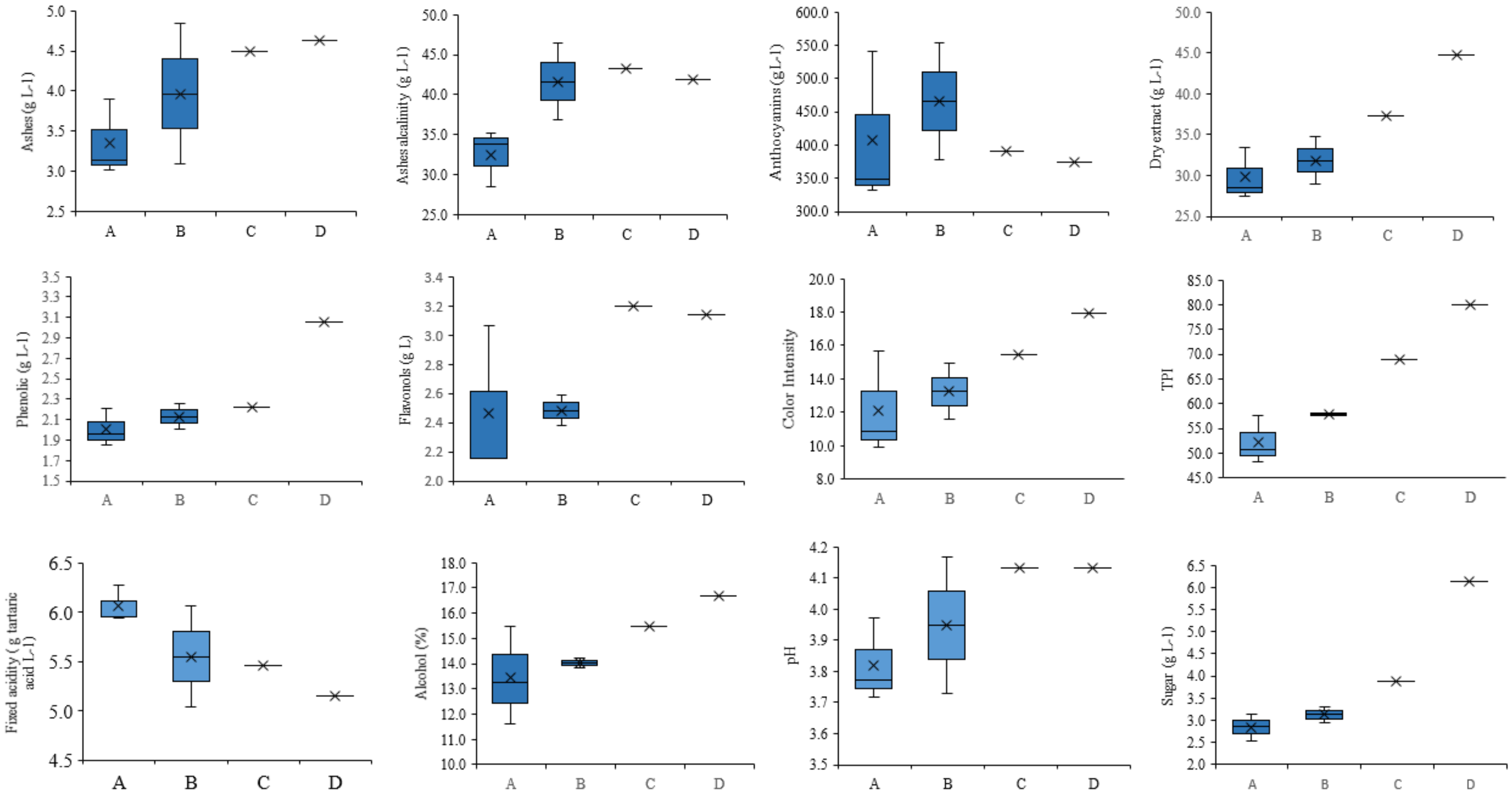

Fig.10. Boxplot with the composition of winter wines separated by groups of soils, according to the separation carried out by the PCA in the attributes of the subsurface horizons. A) Cordislândia (COR), Andradas (AND) and Espírito Santo do Pinhal (PIN); B) Três Corações (TC) and Três Pontas (TP); C) Itobí (ITO); D) São Sebastião do Paraíso (SSP).

4. Conclusions

The soil parent material of vineyards is quite diverse. AND and TP vineyards have gneiss as their parent material and PIN and ITO originate from granite. In such cases, the different degrees of weathering to which the parent material was submitted promoted the chemical, physical, morphological and mineralogical differences between the soils.

The ITO vineyard has a shallower soil, with less water storage capacity and also, it has the highest atmospheric temperature, which benefits the synthesis of anthocyanins, Flavanols, TPI, and alcohol. The soil and climate conditions of this vineyard are the most different from the others, which allowed clear visualization of the effect of the environment on the composition of the wines. A similar tendency was verified for the vineyard group formed by COR, AND and PIN, which have water conditions in the soil that allowed the formation of wines with less acidity.

The different types of soil, the climatic variations, and the management of double pruning, allowed the production of wines with high added value and quality when compared with the main viticultural regions in the world.

5. References

Alvarez, V. H. V., Novais, R. F., Dias, L. E., Oliveira, J.A. 2000. Determinação e uso do fósforo remanescente. Boletim Informativo da Sociedade Brasileira de Ciência do Solo.Viçosa, 25, 1, 27-32.

Amerine, M. A., Ough, C. S. Methods for analysis of musts and wines. New York: John Wiley & Sons, 1980. 341 p.

Amorim, D.A. de, Favero, A.C., Regina, M.D.E.A., 2005. Produção extemporânea da videira, cultivar syrah, nas condições do sul de minas gerais. Rev. Bras. Frutic. 27, 327–331.

AOAC - Association Of Official Analytical Chemists. Official methods of analysis 16. ed. Washington: AOAC, 1995.

Bergqvist, J., Dokoozlian, N., Ebisuda, N. 2001. Sunlight exposure and temperature effects on berry growth and composition of Cabernet Sauvignon and Grenache in the Central San Joaquin Valley of California. American Journal of Enology and Viticulture 52, 1-7.

Blouin, J. 1992. Techniques d'analyses des moûts et des vins. Paris: Dujardin – Salleron. 332p.

Bodin, F., Morlat, R., 2006. Characterization of viticultural terroirs using a simple field model based on soil depth I. Validation of the water supply regime, phenology and vine vigour, in the Anjou vineyard (France). Plant Soil 281, 37–54. https://doi.org/10.1007/s11104-005-3768-0

Brant, L. A. C., Souza, C. R., Mota, R. V., Fernandes, F. P., Gonçalves, M. G. M., Menezes, M. D.; Peregrino, I, Curi, N., Regina, M. A. 2021. Macro scale analysis of Syrah vineyards under winter growing cycles: Agronomical and ecophysiological responses**.** Scientia Agricola (in press).

Brant, L.A.C., Figueiredo, G.M. de, Mota, R.V. da, 2018. Vinhos de Inverno do Sudeste Brasileiro. Territ. du vin 9, 1–4.

Brasil. 1986. Ministerio da Agricultura Portaria no 76 de 26 de novembro de 1986. Dispõe sobre os metodos analiticos de Bebidas e Vinagre. Diario Oficial da Uniao, Brasilia 28 de Novembro de 1986 Seção 1, pt 2.

Brillante, L., Martínez-Lüscher, J., Kurtural, S.K., 2018. Applied water and mechanical canopy management affect berry and wine phenolic and aroma composition of grapevine (Vitis vinifera L., cv. Syrah) in Central California. Sci. Hortic. (Amsterdam). 227, 261–271. https://doi.org/10.1016/j.scienta.2017.09.048

Cantarella, H. Nitrogênio. In; Novais, R. F. et al. Fertilidade do solo. Viçosa, MG; Sociedade Brasileira de Ciência do Solo, 2007. 1017p. 375 a 470.

Conde, C., Silva, P., Fontes, N., Dias, A.C.P., Tavares, R.M., Sousa, M.J., Agasse, A., Delrot, S., Gerós, H. 2007. Biochemical changes throughout grape berry development and fruit and wine quality. Global Science Books. 1, 1-22.

Cortell JM, Sivertsen HK, Kennedy JA, Heymann H (2008) Influence of vine vigor on Pinot noir fruit composition, wine chemical analysis, and wine sensory attributes. American journal of enology and viti- culture 59: 1–10.

CPRM, 2003. Mapa geológico do estado de Minas Gerais - Escala 1:1.000.000. CPRM, Belo Horizonte

Curvelo-Garcia, A. S. 1988. Polifenóis. A cor dos vinhos. In: Controlo de qualidade dos vinhos. Lisboa: Instituto da Vinha e do Vinho. p. 311-347.

Deloire, A., Carbonneau, A., Wang, Z., Ojeda, H., 2004. Vine and water a short review. J. Int. des Sci. la Vigne du Vin 38, 1–13. https://doi.org/10.20870/oeno-one.2004.38.1.932

Dias, F.A.N., da Mota, R.V., de Souza, C.R., Pimentel, R.M. de A., de Souza, L.C., De Souza, A.L., Regina, M. de A., 2017. Rootstock on vine performance and wine quality of ‘syrah’ under double pruning management. Sci. Agric. 74, 134–141. https://doi.org/10.1590/1678-992X-2015-0384

Dias, F.A.N., Mota, R.V. da, Fávero, A.C., Purgatto, E., Shiga, T.M., Souza, C.R. de, Pimentel, R.M. de A., Regina, M. de A., 2012. Videira ’ Syrah ’ sobre diferentes porta - enxertos em ciclo de inverno no sul de Minas Gerais. Pesq. agropec. bras 47, 208–215.

Embrapa, 1997. Manual de Métodos de Análise de Solo. https://doi.org/1517-2627

Favero, A. C., Amorim, D. A., Mota, R. V., Souza, C. R. D, Regina, M. A. 2010. Physiological responses and production of 'Syrah' vines as a function of training systems. Scientia Agricola 67, 267-273.

Favero, A.C., Amorim, D.A. De, Mota, R.V. da, Souza, C.R. De, Regina, M.D.A., 2011. Double-pruning of ‘ Syrah ’ grapevines : a management strategy to harvest wine grapes during the winter in the Brazilian Southeast. Vitis 50, 151–158.

Ferretti, C.G., 2019. Relationship between the geology, soil assessment, and terroir of Gewürztraminer vineyards: A case study in the Dolomites of northern Italy. Catena 179, 74–84. https://doi.org/10.1016/j.catena.2019.03.044

Giusti, M.M., Wrosltad, R.E. 2001. Characterization and measurement of anthocyanins by uv-visible spectroscopy. Current Protocols in Food Analytical Chemistry. New York: John Willey & Sons.

Grotzinger, J., Jordan, T.H., 2014. Understanding Earth. W.H. Freeman, New York.

Haldar, S.K., Tišljar, J., 2014. Metamorphic Rocks, in: Introduction to Mineralogy and Petrology. Elsevier, pp. 213–232. https://doi.org/10.1016/B978-0-12-408133-8.00006-7

Huggett, J.M., 2006. Geology and wine: A review. Proc. Geol. Assoc. 117, 239–247. https://doi.org/10.1016/S0016-7878(06)80012-X

Husson, F., Lê, S., Pagès, J., 2017. Exploratory Multivariate Analysis by Example Using R, Second. ed. CRC PRess, Boca Raton.

Inda Junior, A. V., Kämpf, N., 2003. Avaliação de procedimentos de extração dos óxidos de ferro pedogênicos com ditionito-citrato-bicarbonato de sódio. R. Bras. Ci. Solo 27, 1139–1147. https://doi.org/10.1590/S0100-06832003000600018

Jones, G. V., Snead, N., Nelson, P., 2004. Geology and Wine 8. Modeling Viticultural Landscapes: A GIS Analysis of the Terroir Potential in the Umpqua Valley of Oregon Gregory. Geosci. Canada 31.

Kodur, S., 2011. Effects of juice pH and potassium on juice and wine quality, and regulation of potassium in grapevines through rootstocks (Vitis): a short review. Vitis 50–1, 1–6.

Kotsaki, E., Reynolds, A.G., Brown, R., Jollineau, M., Lee, H.-S., Aubie, E., 2020a. Proximal Sensing and Relationships to Soil and Vine Water Status, Yield, and Berry Composition in Ontario Vineyards. Am. J. Enol. Vitic. 71, 114–131. https://doi.org/10.5344/ajev.2019.19018

Kotsaki, E., Reynolds, A.G., Brown, R., Lee, H.S., Jollineau, M., 2020b. Spatial variability in soil and vine water status in ontario vineyards: Relationships to yield and berry composition. Am. J. Enol. Vitic. 71, 132–148. https://doi.org/10.5344/ajev.2019.19019

Koundouras, S., Marinos, V., Gkoulioti, A., Kotseridis, Y., Leeuwen, C. Van, 2006. Influence of Vineyard Location and Vine Water Status on Fruit Maturation of Nonirrigated Cv . Agiorgitiko (Vitis vinifera L .). Effects on Wine Phenolic and Aroma Components. J. Agric. food Chem. 54, 5077–5086.

Mancini, M., Weindorf, D.C., Chakraborty, S., Silva, S.H.G., dos Santos Teixeira, A.F., Guilherme, L.R.G., Curi, N., 2019a. Tracing tropical soil parent material analysis via portable X-ray fluorescence (pXRF) spectrometry in Brazilian Cerrado. Geoderma 337, 718–728. https://doi.org/10.1016/j.geoderma.2018.10.026

Mancini, M., Weindorf, D.C., Silva, S.H.G., Chakraborty, S., Teixeira, A.F. dos S., Guilherme, L.R.G., Curi, N., 2019b. Parent material distribution mapping from tropical soils data via machine learning and portable X-ray fluorescence (pXRF) spectrometry in Brazil. Geoderma 354, 113885. https://doi.org/10.1016/j.geoderma.2019.113885

McKeague, J.A., Day, J.H., 1966. Dithionite and Oxalate Extractable Fe and Al as Aids in Differentiating Various Classes Of Soils. Can. J. Soil Sci. 46, 13–22. https://doi.org/10.4141/cjss66-003

Mehra, J.P., Jackson, M.L., 1960. Iron Oxide Removal from Soils and Clays by a Dithionite-Citrate-Bicarbonate System Buffered with Bicarbonate Sodium. Clays Clay Miner. 7, 317– 327.

Melo, V. F.; Castilhos, R. M. V.; Pinto, L.F.S. Reserva Mineral do Solo In: Química e Mineralogia do Solo - Parte I - Conceitos básicos ed.Viçosa : SBCS, 2009, v.1, p. 251-332

Morlat, R., Bodin, F., 2006. Characterization of viticultural terroirs using a simple field model based on soil depth - II. Validation of the grape yield and berry quality in the Anjou vineyard (France). Plant Soil 281, 55–69. https://doi.org/10.1007/s11104-005-3769-z

Mota, R.V. da, Amorim, D.A. de, Favero, A.C., Purgatto, E., Murillo de Albuquerque Regina, 2011. Effect of trellising system on grape and wine composition of Syrah vines grown in the cerrado region of Minas Gerais. Ciência e Tecnol. Aliment. 31, 967–972.

Mota, R.V., Peregrino, I., Rivera, S.P.T., Hassimotto, N.M.A., de Souza, A.L., de Souza, C.R., 2021. Characterization of brazilian syrah winter wines at bottling and after ageing. Sci. Agric. 78. https://doi.org/10.1590/1678-992x-2019-0233

Mpelasoka, B.S., Schachtman, D.P., Treeby, M.T., Thomas, M.R., 2003. A review of potassium nutrition in grapevines with special emphasis on berry accumulation. Aust. J. Grape Wine Res. 9 (3), 154–168.

OIV - Organization Internationale de la Vigne et du Vin. 2009. Compendium of International Methods of Wine and Must Analysis, Paris, OIV, 1, 419p.

Peixoto, C.A.B., 2010. Geodiversidade do Estado de São Paulo, CPRM. ed. São Paulo.

Perez-Alvarez, E.P., Garcia-Escudero, E., Peregrina, F., 2015. Soil Nutrient Availability under Cover Crops: Effects on Vines, Must, and Wine in a Tempranillo Vineyard. Am. J. Enol. Vitic. 66, 311–320. https://doi.org/10.5344/ajev.2015.14092

Prata-Sena, M., Castro-Carvalho, B.M., Nunes, S., Amaral, B., Silva, P., 2018. The terroir of Port wine: Two hundred and sixty years of history. Food Chem. 257, 388–398. https://doi.org/10.1016/j.foodchem.2018.03.014

Priori, S., Pellegrini, S., Perria, R., Puccioni, S., Storchi, P., Valboa, G., Costantini, E.A.C., 2019. Scale effect of terroir under three contrasting vintages in the Chianti Classico area (Tuscany, Italy). Geoderma 334, 99–112.

https://doi.org/10.1016/j.geoderma.2018.07.048

Regina, M. A., Mota, R. V., Souza, C. R., & Favero, A. C. (2011). Viticulture for fine wines in Brazilian Southeast. Acta Horticulturae, 910, 113-120.

Renouf, V., Tregoat, O., Roby, J.P., Van Leeuwen, C., 2010. Soils, rootstocks and grapevine varieties in prestigious Bordeaux vineyards and their impact on yield and quality. J. Int. des Sci. la Vigne du Vin 44, 127–134. https://doi.org/10.20870/oeno-one.2010.44.3.1471

Resende, M.; Curi, N.; Rezende, S.B.; Corrêa, G.F.; Ker, J.C. Pedologia: Base para distinção de ambientes, 6 ed. Lavras: Editora UFLA, 2014, 378p.

Ribeiro, A. C., Guimarães, P. T. G., Alvarez A., A. H. (Eds.). Recomendações para o uso de corretivos e fertilizantes em Minas Gerais - 5° Aproximação. Viçosa: CFSEMG, 1999, 359p.

Ribéreau-Gayon, P., Glories, Y., Maujean, A., & Dubourdieu, D. (2006). Handbook of enology (2nd ed.). The chemistry of wine stabilization and treatments (Vol. 2). Chichester, England: John Wiley and Sons Ltd.

Ristic, R., Downey, M.O., Iland, P.G., Bindon, K., Francis, I.L., Herderich, M., Robinson, S.P., 2007. Exclusion of sunlight from Shiraz grapes alters wine colour, tannin and sensory properties. Aust. J. Grape Wine Res. 13, 53–65. https://doi.org/10.1111/j.1755-0238.2007.tb00235.x

Santos, H.G., Jacomine, P.K.T., dos Anjos, L.H.C., de Oliveira, V.A., de Oiveira, J.B. de, Coelho, M.R., Lumbrelas, J.F., Cunha, T.J.F. da, 2014. Sistema brasileiro de classificação de solos., 4th ed, Embrapa Solos-Livros técnicos (INFOTECA-E). Embrapa - Empresa Brasileira de Pesquisa Agropecuária, Rio de Janeiro. https://doi.org/ISBN 978-85-7035-198-2

Santos, R.D. dos, Santos, H.G. dos, Ker, J.C., Anjos, L.H.C. dos, Shimizu, S.H., 2015. Manual de descrição e coleta de solo no campo.

Seguin, G., 1986. “Terroir” and pedology of wine growing. Experientia 42, 861–873.

Silva, F.M., Weindorf, D.C., Silva, S.H.G., Silva, E.A., Ribeiro, B.T., Guilherme, L.R.G., Curi, N., 2019. Tropical Soil Toposequence Characterization via pXRF Spectrometry. Soil Sci. Soc. Am. J. 83, 1153–1166. https://doi.org/10.2136/sssaj2018.12.0498

Smart, D.R., Schwass, E., Lakso, A., Morano, L., 2006. Grapevine Rooting Patterns : A Comprehensive Analysis and a Review. Am. J. Enol. Vitic. 57, 89–104.

Soil Survey Staff, 2014. Keys to soil taxonomy, Twelfth. ed, USDA-NRCS. Washington. https://doi.org/10.1109/TIP.2005.854494

Souza, C.R. de, Mota, R.V. da, Silva, C.P.C., Raimundo, R.H.P., Fernandes, F. de P., Peregrino, I., 2019. Row orientation effects on Syrah grapevine performance during winter growing season. Rev. Ceres 66, 184–190. https://doi.org/10.1590/0034-737x201966030004

Van Leeuwen, C., Friant, P., Choné, X., Tregoat, O., Koundouras, S., Dubourdieu, D., 2004. Influence of Climate, Soil, and Cultivar on Terroir. Am. J. Enol. Vitic. 3, 207–217.

Van Leeuwen, C., Roby, J.-P., De Rességuier, L., 2018. Soil-related terroir factors: a review. OENO One 52, 173–188. https://doi.org/10.20870/oeno-one.2018.52.2.2208

Van Leeuwen, C., Seguin, G., 2006. The Concept of Terroir in Viticulture 17, 1–10. https://doi.org/10.1080/09571260600633135

Van Leeuwen, C., Trégoat, O., Choné, X., Bois, B., Pernet, D., Gaudillère, J.-P., 2009. Vine water status is a key factor in grape ripening and vintage quality for red Bordeaux wine. How can it be assessed for vineyard management purposes? OENO One 43, 121. https://doi.org/10.20870/oeno-one.2009.43.3.798

Walkley, A., Black, I.A., 1934. An examination of the Degtjareff method for deter- mining soil organic matter and a proposed modification of the chromic acid titration method. Soil Sci. 37, 29–38.

Wang, R., Sun, Q., Chang, Q., 2015. Soil types effect on grape and wine composition in Helan Mountain area of Ningxia. PLoS One 10, 1–12. https://doi.org/10.1371/journal.pone.0116690

ARTICLE 3

This article was prepared according to the rules of Applied Geography journal

Viticulture areas with similar terroir for Syrah cv. in Minas Gerais, Brazil with fuzzy logic approach

1. INTRODUCTION

Brazil had a total of 76,000 ha of vine production in 2018 (Mello, 2019), including the main producer states of Rio Grande do Sul, Pernambuco, Paraná, Santa Catarina, Bahia, Minas Gerais, and São Paulo (Mello, 2019). Although Rio Grande do Sul is the most consolidated producer state, Minas Gerais in southeastern Brazil presented the higher recent increase in cultivated areas (Mello, 2019). The Brazilian wine industry still has low competitiveness, since most of the fine wines consumed in Brazil are imported from Chile, Argentina, and European countries. In order to mitigate such effect, it is necessary to improve wine competitiveness, which could be reached with quality, price, and tipicity (Miele et al., 2010).

Typicity is a term in wine tasting, related to *terroir* concept (Regina et al., 2009), describing the degree to which the wine reflects its origins, as well as the signature characteristics of the vines from which it was produced. *Terroir* is a French term that expresses the relationship among wines, environmental, and human characteristics that are directly related to their origin (Vaudour, 2002). Regarding origin and typicity, our main focus in this study is related to a provenance area located in southern Minas Gerais state. The fine wines produced in this region were leveraged by a recent technique, called double pruning, which allows grapes development and ripening in the autumn-winter period. It results in the harvesting of grapes with higher quality and phytosanitary indices (Amorim et al., 2005; Favero et al., 2008, Favero et al., 2011; Mota et al., 2011; Dias et al., 2012). Because the grapes are harvested in winter, the name Winter Wines has been used. Currently, there are more than 152 ha of vineyards applying double pruning technique, widespread in 10 counties in the Southeast region of the country, with an average production of 740 t of wine per year (Brant et al., 2018). The cultivar Syrah showed the best agronomic and quality performance, compared to other evaluated cultivars (Amorim et al., 2005; Favero et al., 2011).

Thus, besides the success of this management on wine quality, the provenance study region is composed by counties with geographical similarities, in which Três Corações and Cordislândia counties have commercial vineyards (Regina et al.,2009), and several experiments with the Syrah vine and the management of double pruning (Amorim et al., 2005; Favero et al., 2011; Favero et al., 2008).

Keeping in mind the importance of typicity to assist wine *terroir* information, and that both characterizations are important to assist the expansion of vitiviniculture frontier in the provenance area, typical environmental variables in a geographical context that drive *terroir* should be analyzed and better understood. Besides soil, wine *terroir* is also influenced by environmental variables such as climate, relief, and geology (Vaudour, 2002). Since *terroir* has a spatial dimension (OIV, 2012), recent researches have seek for land classification in homogeneous land units considering different spatial information and tools for such purpose (Bonfante et al., 2011; Cardoso et al., 2019; Nowlin et al., 2019; Priori et al., 2019; Vianna et al., 2019).

Soil-environmental variables are mostly distributed in a continuous pattern (most are grid-based maps in this study case). Thus, such geographical phenomena, when combined to assist wine-terroir interpretations, imply in ambiguities and uncertainty. Furthermore, keeping in mind our goal to find areas with higher suitability for vine and wine production to assist vitiviniculture expansion, we applied fuzzy logics concepts (Zadeh, 1965). This method along with derived tools (Shi, 2013) generates statistical values of the environmental properties within a given soil mapping unit, working as a data mining tool, allowing: to explore the similarity between terroir-driven environmental variables and soil mapping units containing well-established vineyards (reference vineyards), by exploring Geographical Information System basis; to extract the core environmental concept by creating typical membership curves, which is in accordance to wine *terroir* concepts; to create fuzzy membership maps, where the more similar a local is to a prescribed soil mapping unit, the higher its similarity value (fuzzy membership).

In this sense, we hypothesize that fuzzy logics is a suitable tool to stratify the environment and define areas with greater pertinence of the production of fine Winter Wines according to environmental characteristics of soil mapping units that include commercial vineyards already established. The objective of this work was to search for soil mapping unit-like areas that encompass two commercial vineyards in the southern region of Minas Gerais and, to verify the differences and similarities of wines and grape produced in two reference vineyards. Therefore, attributes related to the terrain, climate, soil and geology of the soil mapping units were considered. The methodology of this work presented three main steps: i) obtaining the attribute layers and harmonization of the database; ii) extracting information from soil mapping units; iii) application, in the provenance area, of the information that was extracted from the soil mapping units.

2. MATERIALS AND METHODS

2.1. Study areas and analysis

2.1.1. Provenance area

The provenance area is located in southern region of Minas Gerais State (Fig. 1), encompassing 2,075 km² of area chosen based on environmental characteristics required for grape-growing and wine-production with high quality (Regina et al., 2006; Tonietto et al., 2006): a) average temperature between 18 and 10 °C in the coldest months, and the hottest month with a maximum of 22°C (Tonietto et al., 2006); b) driest season lasting three to four months and at least one month with a rainfall average of less than 60 mm; c) annual period of sunshine ranging from 1949 to 2050 hours. This area encompasses reference vineyards and its surrounded areas, which are characterized as local and regional approaches of this study.

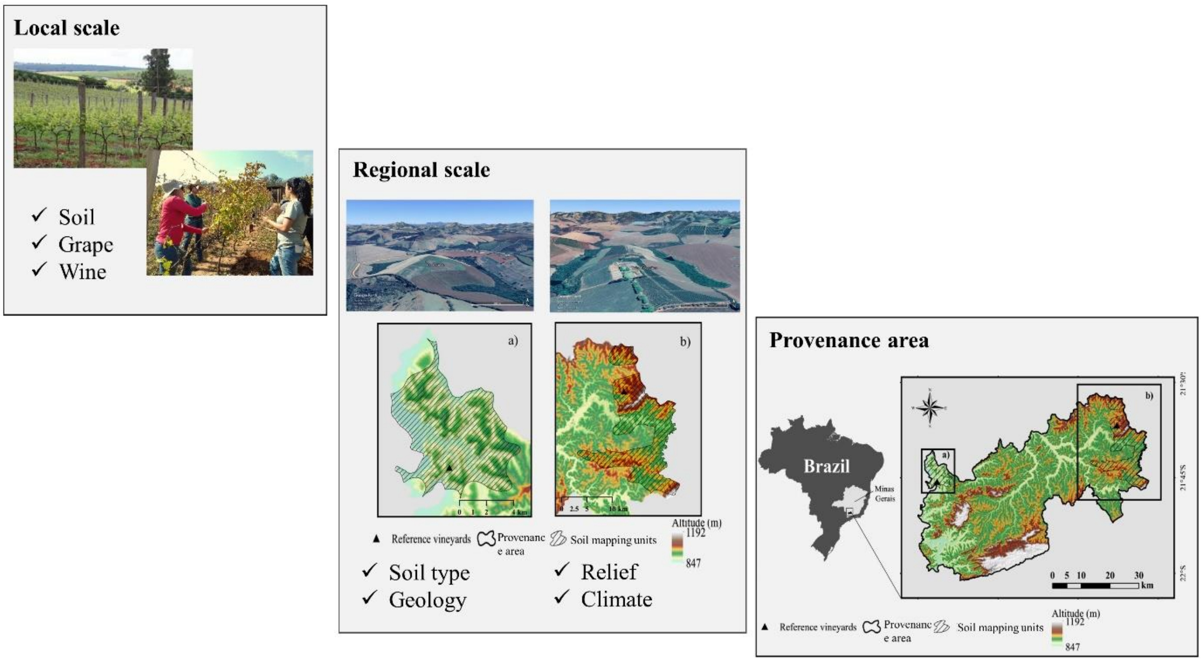

Fig. 1. Framework, different scales and study areas.

2.1.2. Local scale: reference vineyards

Besides the previous development of several scientific experiments for characterizing phytotechnica aspects, analysis of two commercial vineyards (reference vineyards) located in Três Corações and Cordislândia counties were performed for this study, including soil classification, soil mineralogy analysis, vineyards management information acquisition, grape, and wine composition analyses.

Soils classification: soil samples were collected and classified according to Soil Survey Staff (2014), based on morphological and physical-chemical characterizations and analysis (data not shown). Soil classification was carried out to refine the information of current soil map available for the State, which presents a small scale (1: 650,000) (more information will be discussed below). Soil associations occur within the mapping unit polygons due to scale characteristics, and there may be some divergence between the soil verified in the field and that contained in the map.

Soil mineralogy: X-ray diffraction (XRD) analysis was performed in samples collected at a depth of 40-70 cm on the sand, silt and clay fractions of the soils of the two reference vineyards. For mineral identification an X-ray Powder Diffraction with CuKα radiation (Ni filter and a current of 20 mA) was conducted. XRD indicates degree of soil weathering, and in the sand fraction works as a proxy for soil parent material allowing infer about nutrient reserves, soil water retention and physical stability of soil aggregates (Melo & Alleoni, 2009; Kämpf et al., 2012). Also, mineralogy exerts great influence on soil structure formation of the study area, since some minerals tend to arrange (forming mainly blocky structure in the B horizon) and others tends to disarrange (forming granular structure in the B horizon) soil particles. The latter causes an increasing of soil porosity as well as water permeability (Ferreira et al., 1999; Carducci et al., 2011).

Management details: the Syrah cultivar was managed with double pruning according to Favero et al. (2011), grafted onto the Paulsen 1103 rootstock. The first pruning was carried out in August for the formation of latent buds and constitutes the vegetative cycle, and the second pruning was carried out in January for the production of grapes and constituted the reproductive cycle. The spacing of 2.5 m x 1.0 m (4,000 plants ha^{-1}) was adopted and the two prunings maintained 20 latent buds per plant on average.

Grape quality: grape average compositions from the vintages of 2014, 2015, 2017, and 2018 for Três Corações, and vintages of 2016, 2017 and 2018 for Cordislândia were analyzed. For the grape composition, the following analyses were carried out in ten repetitions: pH, by means of digital potentiometer in the fresh grape juice (must); total soluble solids (°Brix) in reading on portable digital refractometer (model Pal 1, Atago); total titratable acidity (expressed in g L^{-1} of tartaric acid) by titration with NaOH 0.1mol L^{-1} using phenolphthalein as an indicator, and berry weight at harvest; anthocyanin content in skins by the differential pH method (Giusti and Wrolstad, 2000); total phenolics using the Folin-Ciocalteau method based on a standard curve of gallic acid (Amerine & Ough, 1980).

Concerning wine quality, analyses in triplicate following the same vintages abovementioned for grape quality were carried out: ash content obtained by gravimetric method, and ash alkalinity by titrimetric method (Amerine & Ough, 1980); dry extract obtained by evaporation of the wine according to the AOAC method (AOAC 1995); flavanols were determined by a colorimetric method (Blouin, 1992); intensity, shade, and total polyphenol index were determined according to Curvelo-Garcia (1988). Similarly, to grape, anthocyanin and total phenolics content were carried out.

In order to verify the relationship between the two reference vineyards regarding the composition of the grapes and the wines produced, the multivariate Principal Component Analysis (PCA) was performed (FactoMineR package, version 1.42) in R software (R Core Team, 2018). In addition, the averages of vintage of each reference vineyard were compared by means of Tukey test ($p < 0.05$) in Sisvar (Ferreira, 2011) software.

2.1.3. Regional scale: soil mapping units

Soil mapping units from a soil survey (1:650.000 scale; FEAM-CETEC-UFV-UFLA, 2010) that circumvent the vineyards were used in order to bring regional scale of analysis. Technically, a soil mapping unit is a collection of areas defined and named the same in terms of their soil components or miscellaneous areas, differing from the others by its uniquely characteristics in a soil map (Soil Science Division Staff, 2017). It is traditionally delineated based on the relation of the knowledge about the soil-landscape relationships recognized as a spatial entity, in which a given soil is mainly related to topography, geology or organisms (Hudson, 1992; Jenny, 1941; Hewitt, 1993; Cook et al., 1996). Considering the environmental aspects behind the soil mapping unit delineation and its significance under terroir state-of-art, it was explored as a case in a reasoning system. Case-based reasoning systems are based on previous experiences (soil mapping unit in this case) to solve new problems (find similar conditions within provenance area), supported by computer and GIS task to overlay and extract optimal values from the environmental data, based on soil mapping unit location (Begum et al., 2009; Shi et al., 2009). Since there are two commercial vineyards with production of high-quality wines in Cordislândia and Três Corações, their soil mapping units could be used as a basis for wine expansion in the provenance area. The matching of soil information between Local and Regional scales provides a suitable link of information still needed.

2.1.3.1. Spatial environmental variables

Environmental variations that are notoriously wine terroir-driver, as terrain, geology, and climate were obtained:

a) Soil type: information obtained from soil map of Minas Gerais state (1:650.000 scale) (FEAM, CETEC, UFV, UFLA, 2010) in a polygon format. There is a total of 13 mapping units, consisting of six soil types at the suborder level (according to Soil Survey Staff, 2014), and one mapping unit that corresponds to stone outcrops. Soil is one of the key components of the wine terroir, since it governs not only nutrients availability, but also root growth and water dynamics that vine plants depends on (Fayolle et al., 2019; Morlat & Bodin, 2006). Since the soils in the vineyards were described and sampled *in situ*, this fact ensures that the soil observed corresponded to the soil mapping polygon.

b) Geology: information taken from a geological map in a polygon format (CPRM, 2003), (1:1,000,000 scale). Geology, as indicative of the parent material, influences the soil type, as well as physical, chemical, mineralogical characteristics (Resende et al., 2014). These in turn will govern soil nutrients and water supply, as well as the development of the root system of grapevines (Huggett, 2006).

c) Relief: a digital elevation model (DEM) with a spatial resolution of 12.5 m was obtained from Alos 1-Palsar satellite image (download at https://www.asf.alaska.edu/, accessed in November 15, 2018). DEM quantitatively represents the continuous variation of relief through the landscape (Moore et al., 1993). From this raster map, the following terrain attributes were generated using SAGA GIS 6.4.0 (Conrad et al. 2015): slope, aspect, and vertical distance to channel network. Slope surface information is related with nutrient, sediments, and water movement (Weill & Brady, 2017; White, 2003). The aspect identifies the downslope direction, and according to Regina et al. (2006), the exposure facing north is preferred since the greater exposure of fruits to sunlight improves the quality of grapes and wines (Bergqvist et al., 2001), where the south face is less recommended. Jones et al. (2004) considered the proximity of the vineyard to water bodies as an important influencer on quality of the grapes, consisting of an information that is expressed by vertical distance above the channel network map. In addition, such proxies of relief are strongly related with soil spatial distribution (Milne, 1935), due to their conceptual relation with soils and less generalized level of detail (Costa et al., 2018).

d) Climate: rainfall and annual mean temperatures raster maps were acquired from WorldClim Version 2.0 (Hijmans et al., 2005) (1 km^2 resolution), one of the main global information database about the climate. The potential daily evapotranspiration was calculated in SAGA GIS 6.4.0 (Conrad et al., 2015) from minimum, mean and maximum temperature raster maps acquired from WorldClim as well. Since Wordclim spatial resolution is coarse considering the characteristics of this study, in order to downscale the spatial data bank taken into account the scale of analysis, ordinary kriging was performed in annual rainfall, and daily evapotranspiration dataset. For this was used Geostatistical Analyst tool within ArcGIS. Kriging follows the adjustment of the semivariograms, a function that describes the spatial relationship of a data set (Isaaks and Srivastava, 1989), according to the following equation:

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

where γ (h) is the estimated semivariance value at distance interval h; $N(h)$ is the number of experimental pairs within a distance h; $z(x_i)$ and $z(x_i + h)$ are the measured values of the variable z (evapotranspiration or precipitation) separated by the distance h.

Since altitude is strongly related with climate (Priori et al., 2019), the cokriging was performed for spatial interpolation of annual mean temperature using DEM, in order to take the advantage of using an auxiliary information, constituted by mean Worldclim temperature, to increase resolution and accuracy (Khosravi & Balyani, 2019). In this case, such downscaling is possible when there is a spatial correlation between the primary attribute (mean temperature) and auxiliary information (DEM) (Pardo-Igúzquiza et al., 2006). This spatial correlation was analyzed by the adjustment of parameters of the experimental cross variogram when cokriging was performed, according to the following equation:

$$\gamma_{12}(h) = \frac{1}{2N(h)} \sum_{i=1}^{N(h)} \{z_1(x_i) - z_1(x_i + h)\} * \{z_2(x_i) - z_2(x_i + h)\}$$

where, $N(h)$ is the number of observation pairs within distance h; $z_1(x_i)$, $z_1(x_i + h)$ e $z_2(x_i)$, $z_2(x_i + h)$ are the mean temperature and DEM values (Goovaerts, 1998).

The accuracy of kriging and cokriging was assessed by cross-validation, allowing the calculation of two statistical indices: the mean error (ME) and the root mean square error (RMSE), calculated as follows:

$$ME = \frac{1}{n} \Sigma_{i=1}^{n} (ei - mi)$$

$$RMSE = \sqrt{\frac{1}{n} \Sigma_{i=1}^{n} (ei - mi)^2}$$

where, *n*: number of observation points, *ei*: estimated value of the attribute (climatic variables or evapotranspiration) and, *mi*: measured values (pixels of the layers). The ME measures the bias of the prediction, and the best case is to have ME as close to zero as possible. The RMSE indicates the accuracy of the prediction and, the closer to zero, the greater the accuracy of the prediction.

2.1.3.1. Fuzzy logics and similarity vectors

In order to seek for areas with a higher similarity of the Reference Vineyards throughout provenance area, the fuzzy logic and similarity vectors concepts were applied by means of the software ArcSIE (Soil Inference Engine) version 10.1.002 (Shi et al., 2009). For that, the module Knowledge Discoverer was used to uncover the typical patterns of occurrence and creation of membership maps. In addition, such module is capable to convert case-based into rule-based reasoning (Shi, 2013), as a part of the process to extract and formalize information by means of optimality curves, as detailed below.

In the first step of Knowledge Discoverer, by the overlay of the soil mapping unit that circumvents Reference Vineyards on the environmental variables, information extraction was performed by means of mathematic functions or optimality curves built for each environmental variable. Thus, both soil mapping units are conceptually considered as a case, in which the curves represents the extraction and formalization of typical environmental covariates conditions within soil mapping units. Two types of optimality curves were built according to the environmental characteristics:

a) Gaussian curves – applied for numerical raster maps: all raster cell values were associated with soil mapping units contour to automatically generate optimality curves, to build Gaussian curves (bell-shapped) (Fig. 2a). The top of curve value (most optimal range) is represented by the median of pixel values, represented by the values *v1* and *v2* that are the lower and upper limits of a given environmental variable. In this case, the bell shape curve was used, where *v1 = v2*. The *w1* and *w2* values, in turn, define how optimal values will change as environmental characteristics deviate from typical characteristics. They correspond to which attribute value represents 50% of the optimal value for this attribute (generally speaking, it is equal to the standard deviation). The values of *r1* and *r2* control the flatness of the top and the steepness of the side parts of the curve (Zhu et al., 2010).

b) Nominal function – applied for categorical polygon map: the nominal information is numerically represented by a numerical code that is meaningful (e.g. number 4 represents a given soil type). In this case, *v1 = v2* since the most optimal value is the labelling only one type of information (Fig. 2b).

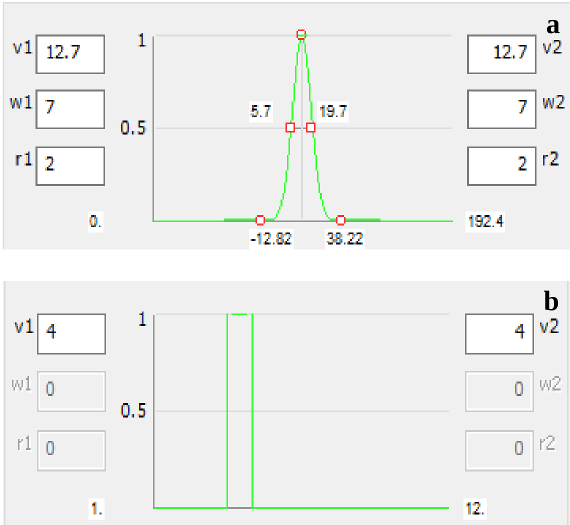

Fig. 2. a) Gaussian curve - continuous bell shape, and b) Nominal curve (for geology).

In the second step, after extracted and formalized information by means of optimality curves for each environmental variables, fuzzy membership maps were created for each soil mapping unit case. For this, the membership function was derived by Rule Based-Reasoning (RBR). The RBR sets similar values for each pixel, ranging from 0 (low similarity) to 1 (high similarity) (McKay et al., 2010), and the initial output from inferences is a fuzzy membership map in raster format. As two soil mapping units polygons were considered, two fuzzy membership maps were generated considering the two soil mapping units (Cordislândia and Três Corações) of Syrah in Southern Minas Gerais state. Thus, continuous fuzzy membership maps can represent a continuous variation of the typical values of the environmental variables of each soil mapping unit, according to the following formula (Zhu et al., 1997a):

$$V_{i,j} = \frac{\sum_{k=1}^{n} S_{ij}^{k} \ * \ V^{k}}{\sum_{k=1}^{n} S_{ij}^{k}}$$

where $V_{i,j}$ represents the estimated similarity to soil mapping unit at a given location (i, j), V^k is a typical value of the considered soil mapping unit (Cordislândia or Três Corações) and n is the total number of soil mapping units. The higher the membership value (closer to

1), the closer is a given location to the core or typical concept, the higher is the membership value assigned for that pixel (Shi, 2013).

The exaggeration index was calculated in order to access the uncertainty associated with the deviation from the definitions of a given soil mapping unit. Exaggeration is calculated according to the following equation:

$$E_{ij} = 1 - S_{ij}^{a}$$

where E_{ij} is the exaggeration uncertainty measure and S_{ij}^{a} is the similarity measured between the pixel (i,j) for a given soil mapping unit assigned (Zhu, 1997b). The higher the membership assigned for a given soil mapping unit (case), the less the exaggeration.

3. RESULTS AND DISCUSSION

3.1. Climate information downscaling from kriging and cokriging

The fitting parameters of semivariograms and cross-variogram, as well as the cross-validation accuracy indexes are presented in Table 1. Those models with lower ME and RMSE values were chosen for the spatial interpolation by means of kriging and cokriging (exponential model for cokriging annual mean temperature x DEM and evapotranspiration ordinary kriging, and spherical model for annual precipitation ordinary kriging). In general, values of ME as close to zero as possible, and RMSE as small as possible denoted suitable accuracy, supporting the application of such methods for downscale purposes, similarly to Pardo-Igúzquiza et al. (2006). It should also be emphasized that several viticulture zoning scientific works deal with spatial data in different scales or resolutions (Madruga et al., 2015; Nowlin et al., 2019; Vianna et al., 2019), according to the occurrence of environmental variables. Annual mean temperature cokriged with altitude, and rainfall and evapotranspiration kriging allowed the reduction of the layers from1 km² to a resolution of 250 m² through the ArcGIS default of Geostatistical Analyst.

Table 1. Models and adjusted parameters set for the semivariograms and cross variogram modeled for the downscaling of climatic variables.

Variable	Model	Adjustment parameters				
		ME	RMSE	Nugget	Range (m)	Partial sill
				Variogram		
Evapotranspiration (mm dia^{-1})	Exponential	-0.0000132	0.03 (mm dia^{-1})	0.004	147,461	0.02
Anual rainfall (mm)	Spherical	0.0060000	10.2 (mm)	227,384	44,491	1,093
				Cross variogram		
Mean temperature x altitude	Exponential	-0.00003	0.74	0.01; 2,109	47,342	0.2; 8,182.0
Mean temperature (°C)	Exponential	0.10000	0.67 (°C)	0.01	59,217.90	0.2
Altitude (m)	Exponential	2.83000	0.06 (m)	2,109.00	47,542.80	8,182.0

ME: mean error; RMSE: root mean square standardized

3.2. Environmental characterization of provenance area

The environmental characterization of provenance area is presented in Fig. 3, in which the soil mapping unit circumventing the Reference Vineyards is also displayed. The altitude is an important parameter for viticulture development (Oliveira et al., 2019) and it has averages 924 m with CV of 10%. This demonstrates that there are no major variations in altitude in the total area, with higher locations occupying only parts of the provenance area (Fig. 3). For the aspect map, the following intervals were considered: north (0-22.5 and 337.5 - 360°; northeast (22.5- 67.5°); east (67.5-112.5°); southeast (112.5-157.5 °); south (157.5-202.5°); southweast (202.5-247.5°); west (247.5-292.5); northwest (292.5-337.5°). The 62% CV indicates that there is a wide variation of exposure faces in the area.

The slope ranges from 0 to 95%, with an average of 15% and a standard deviation of 10 and CV = 67%. This high CV indicates that there is a great variation of this characteristic in the provenance area. Slope affects water storage dynamics vineyard management, as well as govern erosion susceptibility (steeper slopes are more likely to erode than gentle ones). Moreover, slope affects air movement, particularly cold air (Badr et al., 2018). Moderate slopes, between 5 and 15% are referred to as optimal for viticulture (Jones et al., 2004); however, as crops such as cereals and grains depend on flatter areas as well as with higher nutrient and water supplies, generally, farmers can also grow vines in regions with steeper slopes (Van Leeuwen & Seguin, 2006).

The mean temperature ranges from 16.95 to 20 °C, with a CV= 1.94%. The annual rainfall ranges from 1451 mm to 1678 mm, with a variation of 2% in the area and, evapotranspiration has too a small variation in the provenance area (3.56 to 4.00 mm day^{-1}, CV = 2.3%). In addition to demonstrating that there are no major climatic variations in the area, these low CVs are in accordance with the low altitude variation. Temperature, however, is a climatic factor very influenced by the topography and tends to decrease with increasing altitude (Van Leeuwen, 2010).

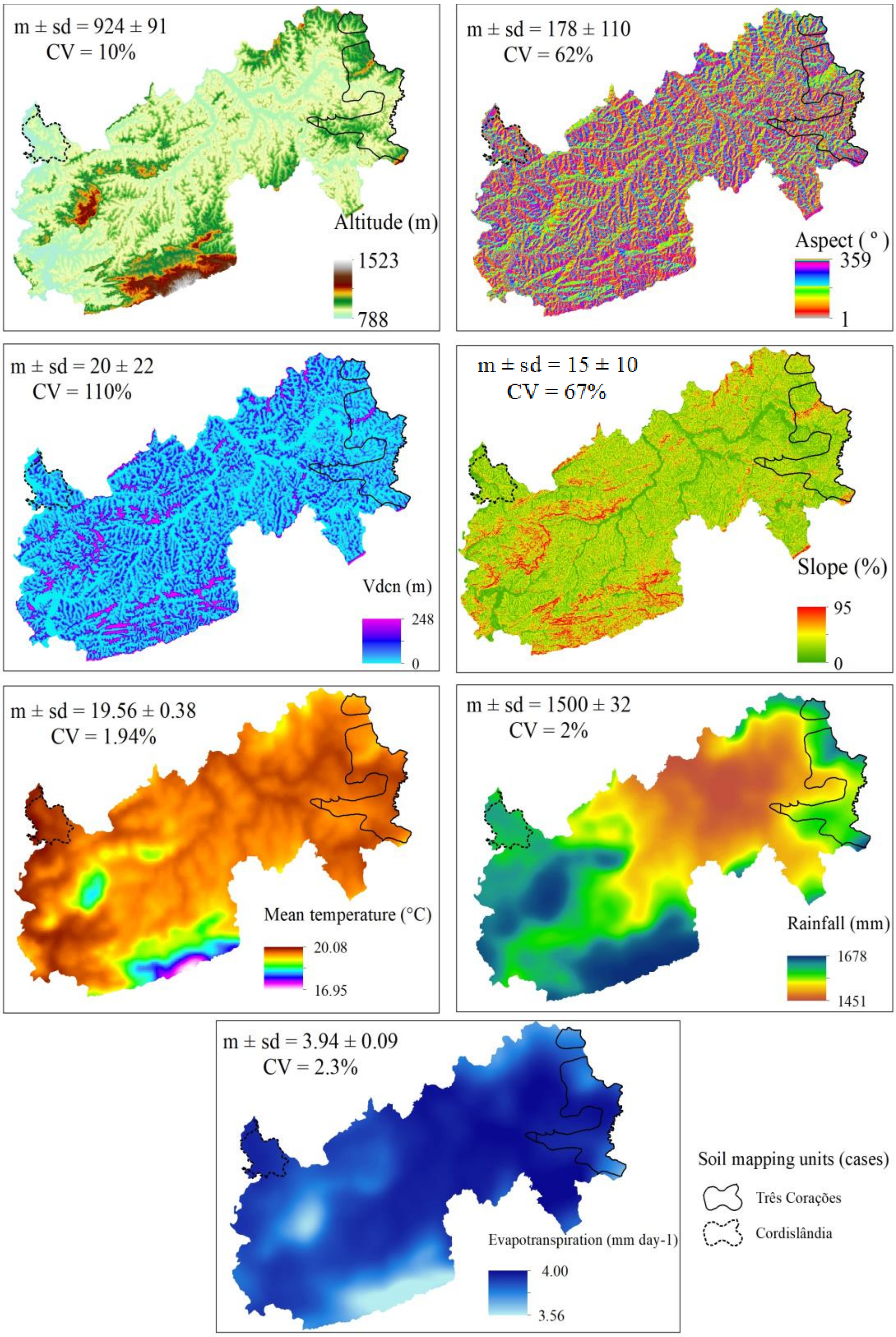

Fig. 3. Environmental continuous variables in the provenance area as well as soil mapping units (cases) in Cordislândia and Três Corações Reference vineyards. Vdcn: vertical distance to the channel network; m ± sd: mean ± standard deviation; CV: coefficient of variation.

Considering the whole provenance area, there are nine types of lithologies according to CPRM (2003) including biotite schist-gneiss, cyanite grenade, gneiss, metagrauwacke, orthogneiss, biotite-gneiss, pyroxene granulite, quartzite and unconsolidated sediments (Fig. 4).

The soils of Reference vineyards were classified as Oxisols (Soil Survey Staff, 2014), a predominant soil type in the whole provenance area (Acrudox and Hapludox) (Fig. 4). In general, Oxisols are thicker soils, with a little differentiation among horizons (considering soil texture or colors), finer-texture with little increase of clay content in depth, and are somewhat excessively drained (Soil Survey Staff, 2014; Santos et al., 2018).

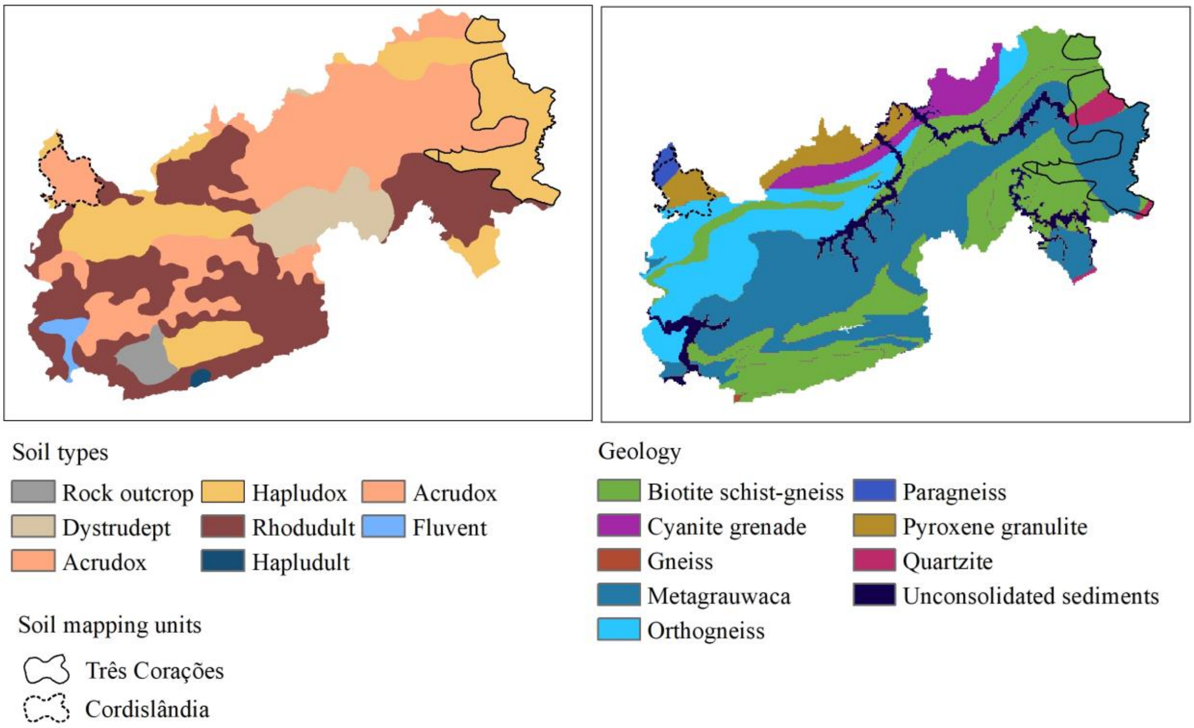

Fig. 4. Environmental nominal variables in the provenance area as well as soil mapping units (cases) in Cordislândia and Três Corações Reference vineyards

Considering the mineralogical aspects of Oxisols, the soil X-ray diffraction (Fig. 5) pointed out the presence of minerals that resulted from intense weathering-leaching of tropical conditions: goethite (iron oxide), hematite (iron oxide), gibbsite (aluminum oxide), and kaolinite (Resende et al., 2014). The soil mineral constitution exerts great influence on chemical and physical characteristics of soils (Ker, 1997), affecting the development of the vines, which in turn will be related to the quality of the wines produced (Van Leeuwen et al., 2018; Renouf et al., 2010).

Considering the chemical aspects, it is well-known that such minerals in clay fractions are responsible for the low cation exchange capacity of soils and low base saturation in the subsurface. Considering the coarser soil fraction, the predominance of quartz (SiO_2) was found, indicating the absence of easily weatherable minerals (Soil Survey Staff, 2014; Santos et al., 2018), as well as poor reserve of nutrients (Melo & Alleoni, 2009). However, the management of fertilization allows adequate nutrition of the vines. In addition, studies carried out in these same areas by Brant et al. (2021) showed that the management of double pruning, associated with the soil attributes and climatic conditions allowed a good agronomic performance in Syrah vines.

Regarding physical aspects, the higher the iron and aluminum oxides, the stronger and more stable tend to be soil micro-aggregates in the granular shape in B horizon (Wambeke, 1992), as it was found in the field campaign analysis. This shape of structure increases the macroporosity and soil hydraulic conductivity, much greater than normally predicted from clay content (Ker, 1997; Buol et al., 2011). It is important to highlight that granular structure shape in the B horizon is only reported in Brazilian Oxisols (Santos et al., 2018; Soil Survey Staff, 2014; FAO, 2015; Resende et al., 2014), and is therefore a distinctive characteristic of the provenance area and this characteristic can even be included as a differential in terms of requesting a geographical indication of wines. The suitability for viticulture of soils with higher macroporosity is well-recognized, for instance, in the traditional viticulture region of Burgundy (Seguin, 1986).

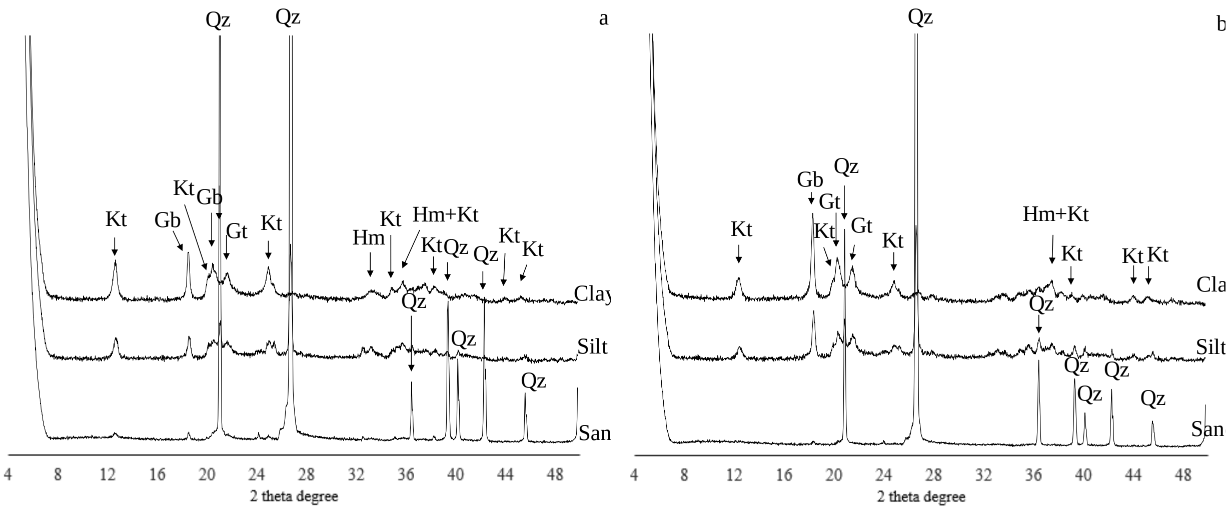

Fig. 5. Mineralogy of the sand, silt and clay fractions of the studied vineyard soils in the depth of 40-70 cm of the soils; a) Cordislândia; b) Três Corações; Kt: kaolinite, Gb: gibbsite; Hm: hematite; Gt: goethite; Qz: Quartz.

3.3. Optimality curves

From the overlay of environmental variables over soil mapping unit polygon, the optimality or inference curves were extracted by the Knowledge Discoverer module and statistics of the raster cells are calculated. The parameters of adjusting curves, which represents the core or central concepts of each area are shown in Table 2. For those with numerical variations, the top of curves (*v1*) refers to the median values of all the cell values that are enclosed by a soil mapping unit polygon, given the central tendency or the most optimal conditions.

The knowledge discoverer accessed the maximum, minimum and standard deviation values. Higher similarity between soil mapping units was found for the following variables, (minimum - maximum - standard deviation): slope (1 – 41 – 6% Cordislândia and 1.2 – 70 - 7.0% for Três Corações); aspect (-1 – 345- 79° Cordislândia and 4 – 352 - 80 ° for Três Corações); temperature (19.8 – 20.1 – 0.07 Cordislândia and 19.3 - 19.9 - 0.12 for Três Corações); rainfall (1503 – 1513 – 2 mm Cordislândia; 1472 – 1546 - 10 mm for Três

Corações); evapotranspiration (3.95 – 4.00 – 0.012 mm dia^{-1} Cordislândia and 3.76 - 4.03 - 0.07 mm day^{-1} for Três Corações), and soil type (Oxisols).

Also, higher similarity could be observed by the similarity median values (*v1*) of top of curves as well as their deviance (*w1*). Some of the characteristics abovementioned falls within the suitable conditions for production grapes or wine with quality: i) more gentle slopes are in accordance with the range of 5 to 15% that is considered as excellent for viticulture (Jones et al., 2004); ii) mean temperature is within the appropriate range for obtaining ideal levels of anthocyanins in grapes (Van Leeuwen, 2010) and in general, for obtaining adequate levels of TSS for ripening grapes (Van Leeuwen & Seguin, 2006). The wide variation in aspect does not constitute a limitation for viticulture. This is because, even though the north face is preferable in relation to the others (Regina et al., 2006), within a small area it is possible to choose slopes with the highest sun exposure side (north face) (Regina et al., 2006).

The areas differ from such features: i) altitude: Cordislândia presented lower altitudes (792-911-28 m) and v1= 843 when compared with Três Corações (849 - 1182 - 49) and v1= 931 m; ii) geology type. Regarding the influence of geological factors in the composition of the wines, where there are very deep soils, the influence of geology is minimal (Huggett, 2006). Thus, as a large part of the study area consists of deep and well-developed soils, to be inferred from the type of soil indicated in the soil map (Acrudox and Hapludox, Rhodudult and Hapludult) it is likely that the parent material does not have a direct effect on the composition of wines and grapes. However, according to Hugget (2006) in this situation, the parent material indirectly controls the composition of grapes, in terms of relief, soil composition, geomorphology, and water retention. There are many hypotheses about the influence of geology on wine typicity, however, some authors believe that little scientific evidence exists to establish how specifically geological parameters are actually involved in the typicity of wines (Priori et al., 2019).

Table 2. Parameters of inference curves for Cordislândia and Três Corações soil mapping units. Altitude, Slope, Aspect, Vdcn, Mean temperature, Evapotranspiration, Rainfall are numeric variables; Soil type and geology are categorical

		E function	
Environmental variable	Instance	v1	w1
Cordislândia			
Altitude (m)		843	28
Slope (%)		10.5	6.1
Aspect (°)		174.9	79
Vdcn (m)		10.6	14
Mean temperature (°C)		19.9	0.073
Evapotranspiration (mm day^{-1})		4	0.012
Rainfall (mm $year^{-1}$)		1508	1.8
Soil type	*e1*	Oxisol	
Geology	*e1*	Pyroxene granulite	
Três Corações			
Altitude (m)		931.1	48.9
Slope (%)		12.7	7
Aspect (degree)		179.6	79.7
Vdcn (m)		12.7	15.5
Mean temperature (°C)		19.7	0.12
Evapotranspiration (mm day^{-1})		4	0.07
Rainfall (mm $year^{-1}$)		1498	10.08
Soil type	*e1*	Oxisol	
Geology	*e1*	Biotite schist-gneiss	
	e2	Metagrauwaca	
	e3	Quartzite	

Vdcn: vertical distance to channel network. v: top of the curve (most optimal value, represented by median); w1: value referring to 50% of the optimal value

3.4. Classification of similarity of areas according to typical characteristics of each soil mapping unit polygons

Two fuzzy membership maps was generated from the information of two soil mapping unit polygons and their respective optimality curves (Fig. 6). In addition, three different ranges of similarities were proposed: low similarity (from 0 to 30% of membership); medium similarity (from 30 to 60%; of membership); and high similarity (from 60% to 100% of membership).

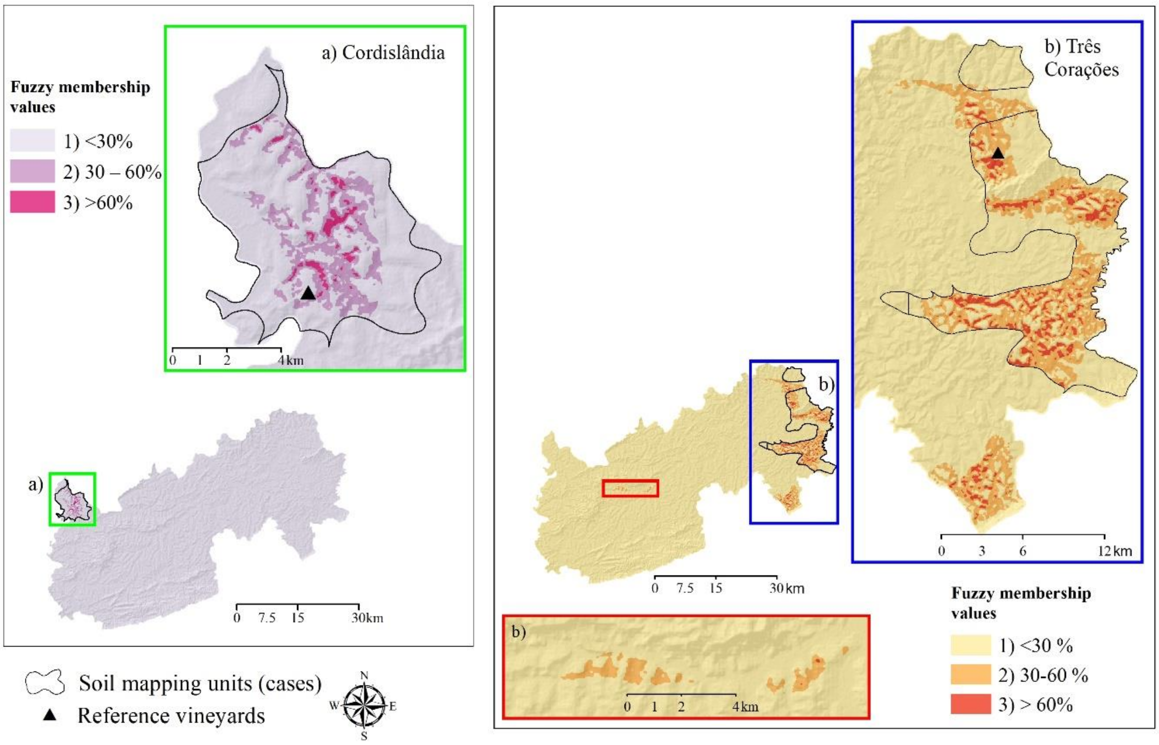

Fig. 6. Fuzzy membership obtained from Cordislândia and Três Corações reference soil mapping unit polygons.

Areas with highest similar conditions to Cordislândia were found geographically within the reference soil mapping unit polygon, since it is the only location with the typical combination of Acrudox with pyroxene granulite and gneiss. This is due to the fact that nominal functions assign total membership for the whole polygon. Thus, from a total of 46.5 km^2 reference area, 124 ha were found with higher similarities. Table 3 shows the mean ± standard deviation of each continuous environmental maps for Cordislândia. Overall, those areas classified as higher similarity presented intermediate altitudes (832 ± 7.0 m), the highest slope values (12.0 ± 2.0%), the highest aspect values (192 ± 46.0 degree), the lowest values of VDCN (6 ± 2.0 m). Regarding the climatic variables of mean temperature, evapotranspiration, and rainfall, no outstanding differences were found between the classes. These values are quite similar to those observed in Três Corações.

Regarding Três Corações, from Fig. 6 it is possible to notice that areas with higher similarities (higher membership values) are geographically dispersed throughout provenance area. This occurs since the combination of Hapludox and biotite schist-gneiss, metagrauwacke and quartzite, also occurs outside the mapping unit that contains the Três Corações vineyard. From 168.45 km^2, 1508 ha were found with higher similarities. Table 4 shows the mean ± standard deviation of each continuous environmental maps for Três Corações, where the areas that constitute higher similarity, similarly to Cordislândia showed intermediate altitudes (918.0 ± 18.0 m) compared to the low similarity (924 ± 26.0 m) and medium similarity (911 ± 28.0 m). The higher slopes were found in the higher similarity class (14.0 ± 3.0%). This class, as well as those defined according to Cordislândia soil mapping unit polygon, also had the smallest VDCN. The average annual temperature and precipitation also showed no differences in relation to the other classes.

The results of exaggeration, calculated for access the uncertainty of the procedure, are shown in Tables 3 and 4. Overall values of exaggeration are around 0.72 ± 0.19, suggesting lower conflict of cells with higher membership values for the same location. Mean values of exaggeration decrease from low → medium → higher similarity classes, decreasing the uncertainty of a class assignment of each cell.

Table 3. Topographic and climatic characteristics in similarity classes verified in the search for areas similar to Cordislândia soil mapping unit in the southern region of Minas Gerais. Mean ± standard deviation

	Similarity class			
	Low	Medium	High	Overall average
Altitude (m)	827.0 ± 18	848.0 ± 18.0	832.0 ± 7.0	839.0 ± 18.0
Slope (%)	11.0 ± 7.0	11.0 ± 4.0	12.0 ± 2.0	11.0 ± 4.0
Aspect (°)	154.0 ± 47.0	156.0 ± 76	192 ± 46.0	165 ± 65.0
VDCN (m)	7.0 ± 8.0	16 ± 8.0	6.0 ± 2.0	11.4 ± 8.68
Mean temperature (°C)	19.91 ± 0.02	19.9 ± 0.03	19.9 ± 0.03	19.9 ± 0.03
Evapotranspiration (mm daily^{-1})	3.99 ± 0.01	3.99 ± 0.01	3.99 ± 0.01	3.99 ± 0.01
Rainfall (mm)	1508.47 ± 0.83	1508.6 ± 1.31	1508.98 ± 0.83	1508.66 ± 1.12
Exaggeration	0.89 ± 0.05	0.73 ± 0.04	0.49 ± 0.06	0.71 ± 0.15
Area (ha)	103.0	1015.0	124.0	1241.0

VDNC: vertical distance to channel network; low: similarity between 0-30%; medium:30-60%; high: >60%. Mean ± standard deviation.

Table 4. Topographic and climatic characteristics in the similarities classes verified in the search for areas similar to Três Corações soil mapping unit in the southern region of Minas Gerais. Mean ± standard deviation

	Similarity class			
	Low	Medium	High	Overall average
Altitude (m)	924.0 ± 26.0	911.0 ± 28.0	918.0 ± 18.0	919.0 ± 24.0
Slope (%)	13.0 ± 5.0	12.0 ± 4.0	14.0 ± 3.0	13.0 ± 4.00
Aspect (°)	174.0 ± 79.0	174.0 ± 73.0	180.0 ± 45.0	176.0 ± 67.0
Vdcn (m)	16.0 ± 10.0	14.0 ± 8.0	11.0 ± 4.0	14.0 ± 8.0
Mean temperature (°C)	19.7 ± 0.07	19.7 ± 0.09	19.70 ± 0.06	19.71 ± 0.07
Evapotranspiration (mm daily^{-1})	3.97 ± 0.03	3.96 ± 0.05	3.98 ± 0.03	3.97 ± 0.03
Rainfall (mm)	1,497 ± 4.4	1,495 ± 7	1,497.69 ± 3.79	1,496.99 ± 4.91
Exaggeration	0.89 ± 0.04	0.73 ± 0.06	0.5 ± 0.08	0.72 ± 0.19
Area (ha)	1,634.0	8,296.0	1,583.0	11,514.0

VDCN: vertical distance to channel network; low: similarity between 0-30%; medium:30-60%; high: >60%; Mean ± standard deviation.

3.5. Linking environmental variables grape and wine composition

Environmental characteristics obtained on a local scale, from spatial information, extracted from the reference vineyards are shown in Table 5. Some environmental features of the reference vineyards follow the same pattern as those extracted by the Knowledge Discoverer in the two soil mapping units. However, unlike that observed in the soil mapping, Cordislândia's reference vineyard presented higher slope, and Vdcn than Três Corações.

Contrary to what was verified in the soil mapping unit, here in the reference vineyard, the aspects of the reference vineyards is different. Whereas in the Cordislândia vineyard the aspect of 224 ° indicates the southeast exposure face, that is, more facing South and in Três Corações, the aspect of 4.0 ° indicates the face of exposure facing North, resulting in greater insolation (Regina et al., 2006).

Table 5. Environmental characteristics of the Cordislândia and Três Corações Syrah reference vineyards in southeastern Minas Gerais

Vineyard	Cordislândia	Três Corações
Altitude (m)	869.00	987.00
Slope (%)	5.0	3.0
Aspect (°)	224.0	4.0
Vdcn (m)	40.0	25.0
Mean temperature (°C)	19.9	19.6
Evapotranspiration (mm day^{-1})	3.99	3.93
Rainfall (mm year^{-1})	1508	1501
Soil type	Acrudox	Acrudox
Geology	Pyroxene granulite	Biotite schist-gneiss

These parameters were taken from the layers used in spatial modeling with fuzzy logic. Vdcn: vertical distance to chanel network.

In addition to verifying the general behavior of the composition of the grapes and wines, the PCA was performed to understand the correlation between the variables and how important and remarkable they are for the typicity of each vineyard. PCA biplot showing both PC scores (triangles and circles) and PC loadings (vectors) on the grape and wine composition are shown in Figs. 8 and 9, respectively. To corroborate the results of the PCA, grapes and wines composition produced in the two vineyards addressing physic-chemical and agronomic, mean characteristics in three vintages (2016, 2017 and 2018) for Cordislândia and four vintages (2014, 2015, 2016 and 2017) for Três Corações are shown in Table 6.

The variance explained by the variables of the composition of the grapes was 83.6% (Fig. 8). The difference in the position of the grapes composition data in different quadrants shows that the PCA performed the separation of the two vineyards, based on these variables. The arrows in the PCA indicate that the composition of the grapes in Três Corações is more related to the TSS contents, in accordance with the higher values showed by Tukey test (Table 6). Moreover, the grapes produced in these vineyards have slightly higher values of total anthocyanin, pH and weight of the grapes compared to Cordislândia. The lower TSS in grapes from Cordislândia is due to the early harvest, which directly affects the composition of the wines, since it is more common for the grapes to be harvested with a higher TSS that indicate a higher level of ripeness (Ojeda et al., 2002; Ristic et al., 2007).

Anthocyanin and flavanols are phenolic compounds in the group of flavonoids. Flavonoids are important quality indicators due to their contribution to the appearance of wines (color), taste (bitterness) and mouthfeel (Tarara et al., 2008; Ristic et al., 2007). In general, phenolic compounds also contribute to the aging capacity of wines (Mota et al., 2021). Although there was no statistically significant difference in the levels of anthocyanins, the greater insolation, made possible by the north exposure face (aspect = 4 °) in Três Corações, may be responsible for the higher average content of this pigment in both grapes and wines (Table 6) (Bergqvist et al., 2001; Van Leeuwen, 2010). The same applies to phenolic content (Bergqvist et al., 2001) and flavanols (Oliveira et al., 2019; Ristic et al., 2007).

The PCA with the composition of the wines had an explained variance of 80.3% (Fig. 9). Cordislândia wines were related to higher values of fixed acidity. In addition to greater acidity, Cordislândia wine has a lower alcohol content (Table 6). This is also a consequence of the early harvest since the lower TSS results in lower alcohol content in wines (Mota et al., 2011; Conde et al., 2007; Koundouras et al., 2006). Even so, the alcohol content of the two wines evaluated are in line with the range recommended by Brazilian beverage legislation (between 8.6 and 14%) (MAPA, 2018).

The composition of grapes, including pH, TSS, anthocyanins, total phenolics and total acidity found in this study are suitable for the production of fine wines since they are similar to those from important wine regions in the world (Keller et al., 2012; Koundouras et al., 2006; Morlat & Bodin, 2006; Priori et al., 2019), the same pattern occurs for most parameters related to wine composition (Priori et al., 2019; Koundouras et al., 2006).

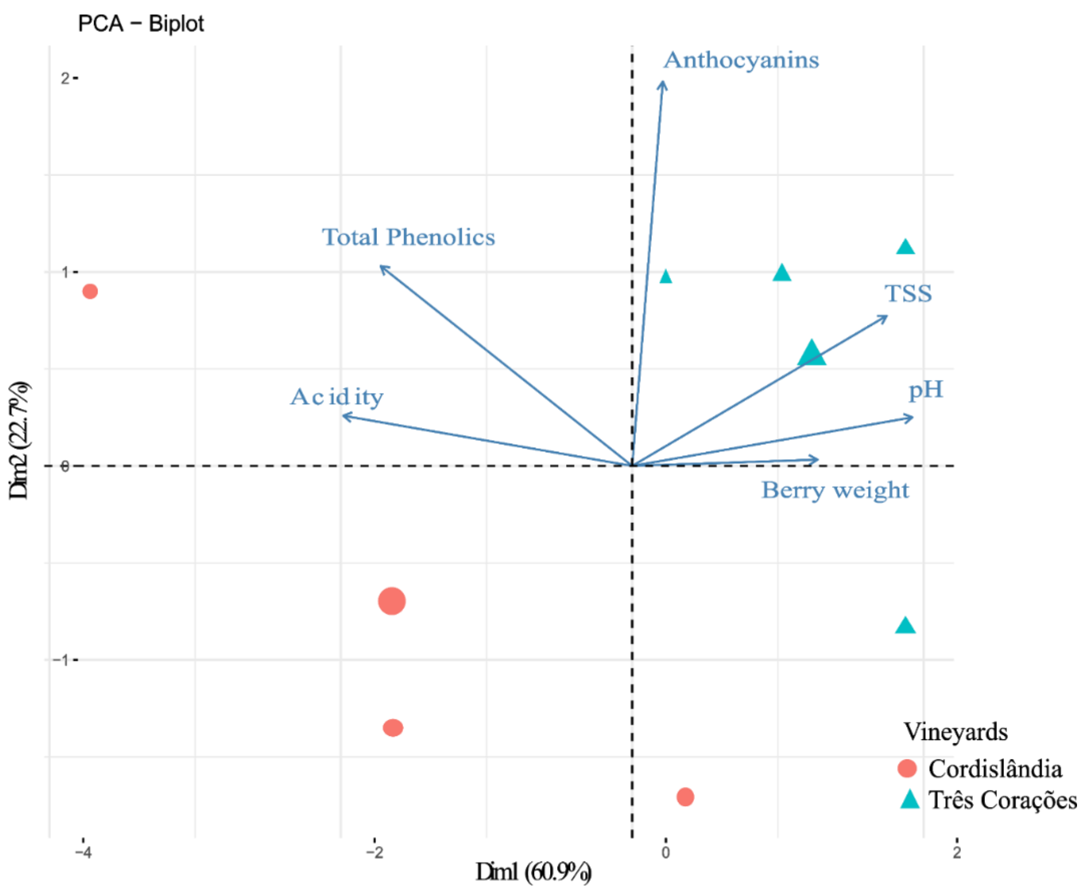

Fig. 8. Principal component analysis on grapes characteristics, produced in 2016, 2017 and 2018 vintages in the Cordislândia and 2014, 2015, 2017 and 2018 for Três Corações; Dim – dimension.

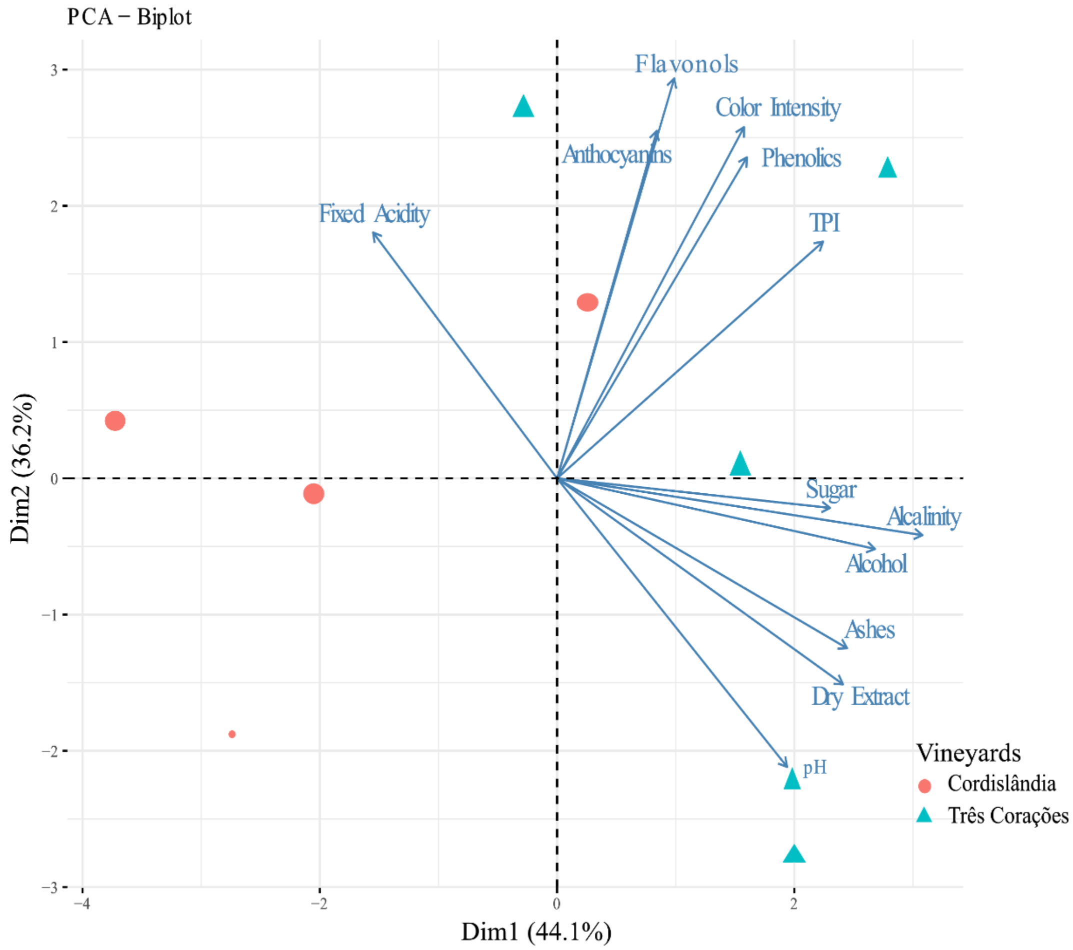

Fig. 9. Principal component analysis on wine composition produced in 2016, 2017 and 2018 vintages in the Cordislândia and 2014, 2015, 2017 and 2018 for Três Corações; Dim – dimension.

Table 6. Grape and wine mean features from two reference vineyards

	Cordislândia	Três Corações
Grape		
Berry Weight (g)	1.23 ± 0.13 a	1.35 ± 0.10 a
pH	3.31 ± 0.15 a	3.53 ± 0.10 a
TSS (°Brix)	18.7 ± 0.61 b	20.88 ± 0.68 a
Total acidity (g L^{-1})	7.19 ± 0.80 a	6.27 ± 0.36 a
Total anthocyanins (mg g $berry^{-1}$)	1.08 ± 0.12 a	1.21 ± 0.09 a
Total phenolics (mg g $berry^{-1}$)	3.43 ± 0.73 a	3.06 ± 0.24 a
Wine		
Alcalinity (g L^{-1})	28.44 ± 6.94 a	38.98 ± 4.67 a
Anthocyanins (mg L^{-1})	332.08 ± 71.83 a	400.28 ± 181.97 a
Ashes (g L^{-1})	3.02 ± 0.53 a	3.36 ± 0.54 a
Dry Extract (g L^{-1})	27.43 ± 1.18 a	31 ± 2.42 a
Phenolics (mg mL^{-1})	1.85 ± 0.20 a	2.01 ± 0.33 a
Flavanols (g L^{-1})	2.16 ± 0.29 a	2.27 ± 0.36 a
Color intensity (OD_{420}+OD_{520}+OD_{620})	9.87 ± 2.29 a	12.67 ± 2.64 a
TPI	48.24 ± 9.25 a	55.63 ± 3.91 a
Fixed acidity (g L^{-1})	6.28 ± 1.34 a	5.64 ± 0.62 a
Alcohol (%)	11.62 ± 0.45 b	14.38 ± 1.02 a
pH	3.72 ± 0.21 a	3.86 ± 0.18 a
Sugar (g L^{-1})	2.53 ± 0.58 a	3.17 ± 0.39 a

TSS: total soluble solids. TPI – total phenolic index; Different letters in the same line indicate statistical difference (Tukey test, $p < 0.05$) between the characteristics of the grapes or wines, produced in the 2016, 2017 and 2018 vintages in the Cordislândia, and in 2014, 2015, 2017 and 2018 for Três Corações.

4. CONCLUSIONS

Environmental conditions, including climate, soil and relief and the composition of grapes and wines, indicate that the southern region of Minas Gerais has a high suitability for producing wine. Especially if we consider that these wines produced there have a similar composition to quality wines produced in important wine-growing regions of the world.

The fuzzy logics was an adequate tool for the search for similar environmental conditions, since the areas with the highest values of similarity with the soil mapping unit were exactly where the uncertainty was lower. Lower uncertainty along higher membership values or similarities is an important metric for decision makers. In addition, the fuzzy logic tool provided an overview of the potential for the expansion of viticulture in the coffee region of southern Minas Gerais.

In Cordislândia, thicker soils, with high permeability, with an average altitude of 832 m, with undulated relief with 12.0% slope, whose geology is predominantly constituted by pyroxene granulite, with annual precipitation of 1508 mm, are more likely (greater than 60% probability) for the wine to have 332.08 ± 71.83 g L^{-1} of anthocyanins and phenolic between 1.85 ± 0.20 mg mL^{-1} making up the total 124.0 ha in the provenance area.

For Três Corações, thicker soils with high permeability developed from biotite schist-gneiss as parent material were found; however, in areas with altitudes of 917.56 m, slopes of 14.15%, average temperatures of 19.7 ° C, and annual precipitation of 1497.69 mm are more likely to produce wines with anthocyanin contents of 400.28 ± 181.97 mg L^{-1} and total phenolic of 2.01 ± 0.33 mg mL^{-1}, totaling an area of 1,583.0 ha in the provenance area.

Grape and wine characteristics were quite similar in the two reference vineyards. Considering environmental factors only and excluding human factors (related to the management of the vineyards and harvest decision), this may indicate the existence of one homogeneous wine terroir in the provenance area.

5. REFERENCES

Amerine, M. A., Ough, C. S. (1980). *Methods for analysis of musts and wines*. New York: John Wiley & Sons, 341 p.

Amorim, D. A. de, Favero, A. C., & Regina, M. D. E. A. (2005). Produção extemporânea da videira, cultivar syrah, nas condições do sul de minas gerais. *Revista Brasileira de Fruticultura*, 27(2), 327–331.

AOAC - Association Of Official Analytical Chemists. *Official methods of analysis* 16. ed. Washington: AOAC, 1995.

Badr, G., Hoogenboom, G., Moyer, M., Keller, M., Rupp, R., & Davenport, J. (2018). Spatial suitability assessment for vineyard site selection based on fuzzy logic. *Precision Agriculture*, 19(6), 1027–1048. https://doi.org/10.1007/s11119-018-9572-7

Begum, S., Ahmed, M. U., Funk, P., Xiong, N., & Schéele, B. V., (2009). A case-based decision support system for individual stress diagnosis using fuzzy similarity matching. *Computational Intelligence*, 25(3), 180-195.

Bergqvist, J., Dokoozlian, N., & Ebisuda, N. (2001). Sunlight Exposure and Temperature Effects on Berry Growth and Composition of Cabernet Sauvignon and Grenache in the Central San Joaquin Valley of California. *American Journal of Enology and Viticulture*, 1(April), 3–9.

Blouin, J. (1992*). Techniques d'analyses des moûts et des vins*. Paris: Dujardin – Salleron. 332p.

Bonfante, A., Basile, A., Langella, G., Manna, P., & Terribile, F. (2011). A physically oriented approach to analysis and mapping of terroirs. *Geoderma*, 167–168, 103–117. doi:10.1016/j.geoderma.2011.08.004.

Brant, L. A. C., Figueiredo, G. M. de, & Mota, R. V. da (2018). Vinhos de Inverno do Sudeste Brasileiro. *Territoires du vin*, 9, 1–4.

Brant, L. A. C., Souza, C. R., Mota, R. V., Fernandes, F. P., Gonçalves, M. G. M., Menezes, M. D.; Peregrino, I, Curi, N., Regina, M. A. 2021. Macro scale analysis of Syrah vineyards under winter growing cycles: Agronomical and ecophysiological responses. *Scientia Agricola*, 78(6), 2-9.

Buol, S.W., Southard, R.J., Graham, R.C., McDaniel, P.A., 2011. *Soil Genesis and Classification*. 6th ed. Wiley-Blackwell, 544p.

Cardoso, A. S., Alonso, J., Rodrigues, A. S., Araújo-Paredes, C., Mendes, S., & Valín, M. I. (2019). Agro-ecological terroir units in the North West Iberian Peninsula wine regions. *Applied Geography*, 107(May), 51–62. https://doi.org/10.1016/j.apgeog.2019.03.011

Carducci, C. E., Oliveira, G. C. de, Severiano, E. da C., & Zeviani, W. M. (2011). Modelagem da curva de retenção de água de Latossolos utilizando a Equação Duplo

Van Genuchten. Revista Brasileira de Ciência Do Solo, 35(1), 77–86. https://doi.org/10.1590/S0100-06832011000100007

Conde, C., Silva, P., Fontes, N., Dias, A.C.P., Tavares, R.M., Sousa, M.J., Agasse, A., DELROT, S., GERÓS, H. (2007). Biochemical changes throughout grape berry development and fruit and wine quality. *Global Science Books*, v. 1, n.1, p1-22

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

Cook, S. E., Corner, R. J., Grealish, G., Gessler, P. E., and Chartres, C. J. (1996). A rule-based system to map soil properties. *Soil Science Society America Journal*. 60, 1893–900.

Costa, E. M., Samuel-rosa, A., & Anjos, L. H. C. dos. (2018). Digital elevation model quality on digital soil mapping prediction accuracy. *Ciência e Agrotecnologia*. 42(6), 608–622. https://doi.org/http://dx.doi.org/10.1590/1413-70542018426027418

CPRM - Serviço geológico do Brasil. (2003). *Mapa geológico do estado de Minas Gerais*. Belo Horizonte: CPRM. Escala 1:1.000.000.

Curvelo-Garcia, A. S. Controle de qualidade dos vinhos: Química Enológica. *Métodos Analíticos*. Lisboa: Instituto da Vinha e do Vinho, 1988. 420 p.

Dias, F. A. N., Mota, R. V. da, Fávero, A. C., Purgatto, E., Shiga, T. M., Souza, C. R. de, et al. (2012). Videira ' Syrah ' sobre diferentes porta - enxertos em ciclo de inverno no sul de Minas Gerais. *Pesquisa agropecuária brasileira* 47, 208–215.

FAO - Food and Agriculture Organization of the United Nations. (2015*). Guidelines for soil description.* 4th ed rev. Rome. Available at http://www.fao.org/docrep/019/a0541e/a0541e.pdf

Favero, A. C., Amorim, D. A. De, Mota, R. V. da, Souza, C. R. De, & Regina, M. D. A. (2011). Double-pruning of ' Syrah ' grapevines : a management strategy to harvest wine grapes during the winter in the Brazilian Southeast. *Vitis,* 50(4), 151–158.

Favero, A. C., Amorim, D. A. de, Mota, R. V., & Regina, M. de A. (2008). Viabilidade de produção da videira 'syrah', em ciclo de outono inverno, na região sul de minas gerais 1. *Revista Brasileira de Fruticultura*, 30(2008), 685–690.

Fayolle, E., Follain, S., Marchal, P., Chéry, P., & Colin, F. (2019). Identification of environmental factors controlling wine quality: A case study in Saint-Emilion Grand Cru appellation, France. *Science of the Total Environment*, 694, 133718. https://doi.org/10.1016/j.scitotenv.2019.133718

FEAM-CETEC-UFV-UFLA. (2010). *Mapa de solos do Estado de Minas Gerais*: legenda expandida. Belo Horizonte: Fundação Estadual do Meio Ambiente. 49p.

Ferreira M. M, Fernandes B., & Curi N. (1999) Mineralogia da fração argila e estrutura de Latossolos da região sudeste do Brasil. *Revista Brasileira de Ciência* Solo, 23, 507-514.

Ferreira, D. F, Sisvar: A Computer statistical analysis system. (2011). *Ciência e Agrotecnologia*, 35(6):1039-1042

Giusti, M.M.; Wrosltad, R.E. (2000). *Characterization and measurement of anthocyanins by uv-visible spectroscopy*. Current Protocols in Food Analytical Chemistry. New York: John Willey & Sons

Goovaerts, P. (1998). Geostatistical tools for characterizing the spatial variability of micro-biological and physico-chemical soil properties. *Biology and Fertility of Soils*, 27, 315–334.

Hewitt, A. E. (1993). Predictive modelling in soil survey. *Soils and Fertilizers*, 56, 305–14.

Hijmans. R.J., Cameron. S.E., Parra. J.L., Jones. P.G., Jarvis. A. (2005). Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology,* 25, 1965–1978. https://doi.org/10.1002/joc.1276.

Hudson, B.D. (1992). The soil survey as paradigm-based science. *Soil Science Society of American Journal*, 56, 836–841.

Huggett, J. M. (2006). Geology and wine: A review. *Proceedings of the Geologists' Association*, 117(2), 239–247. https://doi.org/10.1016/S0016-7878(06)80012-X

Isaaks, E.H.; Srivastava, R.M. (1989). *An Introduction to Applied Geostatistics*. Oxford University Press, New York, NY, USA.

Jenny, H. (1941). *Factors of Soil Formation: A System of Quantitative Pedology*. McGraw-Hill Book Co., Inc., New York.

Jones, G. V., Snead, N., & Nelson, P. (2004). Geology and Wine 8. Modeling Viticultural Landscapes: A GIS Analysis of the Terroir Potential in the Umpqua Valley of Oregon Gregory. *Geoscience Canada,* 31(4).

Kämpf, N.; Marques, J. J., & Curi, N. (2012). Mineralogia de solos brasileiros. In: *Pedologia – Fundamentos* (pp. 81-146). SBCS.

Keller, M., Mills, L. J., & Harbertson, J. F. (2012). Rootstock Effects on Deficit-Irrigated Winegrapes in a Dry Climate : Vigor , Yield Formation , and Fruit Ripening. *American Journal of Enology and Viticulture*, 1, 29–39. https://doi.org/10.5344/ajev.2011.11078

Ker, J. C. (1997). Latossolos Do Brasil: Uma Revisão. *Geonomos,* 5(1), 17–40. https://doi.org/10.18285/geonomos.v5i1.187

Khosravi, Y., & Balyani, S. Spatial Modeling of Mean Annual Temperature in Iran: Comparing Cokriging and Geographically Weighted Regression. *Environmental Model Assess* 24, 341–354 (2019). https://doi.org/10.1007/s10666-018-9623-5

Koundouras, S., Marinos, V., Gkoulioti, A., Kotseridis, Y., Leeuwen, C. Van, 2006. Influence of Vineyard Location and Vine Water Status on Fruit Maturation of Nonirrigated Cv . Agiorgitiko (Vitis vinifera L .). Effects on Wine Phenolic and Aroma Components. *Journal of Agricultural and Food Chemistry*, 54, 5077–5086.

Madruga, J., Azevedo, E. B., Sampaio, J. F., Fernandes, F., Reis, F., & Pinheiro, J. (2015). Analysis and definition of potential new areas for viticulture in the Azores (Portugal). *Soil*, 1, 515–526. doi:10.5194/soil-1-515-2015.

MAPA-Ministério da Agricultura e Pecuária. (2018). *Instrução Normativa N° 14*, de 08 de fevereiro de 2018. https://doi.org/10.1051/matecconf/201712107005

McKay, J., Grunwald, S., Shi, X., Long, R.F. (2010). Evaluation of the transferability of a knowledge-based soil-landscape model. In: Boettinger, J.L., Howell, D.W., Moore, A.C., Hartemink, A.E., Kienast-Brown, S. (Eds.), *Digital Soil Mapping*. Springer, Netherlands, Dordrecht

Mello, L. M. R. (2019). *Vitivinicultura brasileira*: panorama 2018. Bento Gonçalves: Embrapa.

Melo, V. de F., & Alleoni, L. R. F. (2009). *Química e mineralogia do solo*: parte I - Conceitos básicos. Viçosa: SBCS

Miele, A., Rizzon, L. A., & Zanus, M. C. (2010). Discrimination of Brazilian red wines according to the viticultural region, varietal, and winery origin. *Ciência e Tecnologia de Alimentos*, 30(1), 268–275. https://doi.org/10.1590/S0101-20612010000100039

Milne, G. (1935). Some suggested units of classification and mapping particularly for East African Soils. *Soil research*, 4, 183–198.

Moore, I.D., Gessler, P.E., Nielsen, G.A., & Peterson, G.A. (1993). Soil attribute prediction using terrain analysis. *Soil Science Society of American Journal*, 57: 443-452

Morlat, R., & Bodin, F. (2006). Characterization of viticultural terroirs using a simple field model based on soil depth - II. Validation of the grape yield and berry quality in the Anjou vineyard (France). *Plant and Soil*, 281(1–2), 55–69. https://doi.org/10.1007/s11104-005-3769-z

Mota, R. V. da, Amorim, D. A. de, Favero, A. C., Purgatto, E., & Regina M. A. (2011). Effect of trellising system on grape and wine composition of Syrah vines grown in the cerrado region of Minas Gerais. *Ciência e Tecnologia de Alimentos*, 31(4), 967–972.

Mota, R.V., Peregrino, I., Rivera, S.P.T., Hassimotto, N.M.A., de Souza, A.L., de Souza, C.R., 2021. Characterization of brazilian syrah winter wines at bottling and after ageing. *Scientia Agricola*. 78. https://doi.org/10.1590/1678-992x-2019-0233

Nowlin, J. W., Bunch, R. L., & Jones, G. V. (2019). Viticultural site selection: Testing the effectiveness of North Carolina’s commercial vineyards. Applied Geography, 106(August 2017), 22–39. https://doi.org/10.1016/j.apgeog.2019.03.003

OIV. (2012). Oiv Guidelines for Vitiviniculture Zoning Methodologies on a Soil and Climate Level.

Ojeda, H., Andary, C., Kraeva, E., Carbonneau, A., Deloire, A. (2002). Influence of pre and postveraison water deficit on sintesis and concentration of skin phenolic compounds during berry growth of Vitis vinifera L., cv. Shiraz. *American Journal Enology and Viticulture*, 53 (4), 261-267.

Oliveira, J. B., Egipto, R., Laureano, O., de Castro, R., Pereira, G. E., & Ricardo-da-Silva, J. M. (2019). Climate effects on physicochemical composition of Syrah grapes at low and high altitude sites from tropical grown regions of Brazil. *Food Research International*, 121(October 2018), 870–879. https://doi.org/10.1016/j.foodres.2019.01.011

Pardo-Igúzquiza, E., Chica-Olmo, M., & Atkinson, P. M. (2006). Downscaling cokriging for image sharpening. *Remote Sensing of Environment*, 102(1–2), 86–98. https://doi.org/10.1016/j.rse.2006.02.014

Priori, S., Pellegrini, S., Perria, R., Puccioni, S., Storchi, P., Valboa, G., & Costantini, E. A. C. (2019). Scale effect of terroir under three contrasting vintages in the Chianti Classico area (Tuscany, Italy). *Geoderma*, 334(January 2018), 99–112. https://doi.org/10.1016/j.geoderma.2018.07.048

R Core Team. (2018). R: a language and environment for statistical R Foundation for Statistical Computing, Vienna, Austria.Available. https://www.R-project.org/ (verified 23 Aug. 2018).

Regina, M. A., Fráguas, J.C., Alvarenga, A. A., Souza, C.R, Amorim D.A., Mota R.V., & Favero, A.C. Implantação e manejo do vinhedo para produção de vinhos de qualidade. (2006). *Informe Agropecuário*, 27, 16-31.

Regina, M. de A., Mota, R. V. da, & Amorim, D. A. de (2009). Vinhos finos: novos horizontes em Minas Gerais. *Informe Agropecuário*, 30, 158-167.

Renouf V., Trégoat O., Roby J.-P. and van Leeuwen C. (2010). Soils, rootstocks and grapevine varieties in prestigious Bordeaux vineyards and their impact on yield and quality. *Journal International des Sciences de la Vigne et du Vin*, 44, 127-134. doi:10.20870/oeno-one.2010.44.3.1471

Resende, M., Curi, N., Rezende, S.B., Corrêa, G.F., Ker, J.C. (2014). *Pedologia: Base para distinção de ambientes*, 6th ed. Lavras: Editora UFLA.

Ristic, R., Downey, M.O., Iland, P.G., Bindon, K., Francis, I.L., Herderich, M., Robinson, S.P. (2007). Exclusion of sunlight from Shiraz grapes alters wine colour, tannin and sensory properties. *Australian Journal of Grape and Wine Research*, 13, 53–65. https://doi.org/10.1111/j.1755-0238.2007.tb00235.x

Santos, H. G. dos, Jacomine, P. K. T., Anjos, L. H. C. dos, Oliveira, V. A. de, Lumbreras, J. F., Coelho, M. R et al. (2018). *Sistema Brasileiro de Classificação de Solos* (5th ed.). Brasília: Emrabapa.

Seguin, G. (1986). “Terroir” and pedology of wine growing. *Experientia*, 42, 861–873.

Shi, X. (2013). ArcSIE User’ s Guide. In Spatial Inference Enterprise.

Shi, X., Long, R., Dekett, R., & Philippe, J. (2009). Integrating different types of knowledge for digital soil mapping. *Soil Science Society of America Journal*, 73, 1682–1692. http://dx.doi.org/10.2136/ sssaj2007.0158.

Soil Science Division Staff. 2017. *Soil survey manual*. C. Ditzler, K. Scheffe, and H.C. Monger (eds.). USDA Handbook 18. Government Printing Office, Washington, D.C.

Soil Survey Staff. (2014). *Keys to soil taxonomy*. In USDA-NRCS (Ed.), USDA-NRCS (Twelfth, Vol. 12). https://doi.org/10.1109/TIP.2005.854494

Tarara, J.M., Lee, J., Spayd, S.E., Scagel, C.F. (2008). Berry temperature and solar radiation alter acylation, proportion, and concentration of anthocyanin in Merlot grapes. *American Journal of Enology and Viticulture*, 59, 235–247.

Tonietto, J., Vianello, R. L., & Regina, M. de A. (2006). Caracterização macroclimática e potencial enológico de diferentes regiões com vocação vitícola de Minas Gerais. In *Informe Agropecuário*, 27(234), 32–55.

Van Leeuwen C. (2010). Terroir: the effect of the physical environment on vine growth, grape ripening and wine sensory attributes, p. 273-315. In: *Managing wine quality*, volume 1: Viticulture and wine quality. Reynolds A. (Ed.), Woodhead Publishing Ltd., Oxford, UK. doi:10.1533/9781845699284.3.273

Van Leeuwen, C., & Seguin, G. (2006). *The Concept of Terroir in Viticulture*. 17(1), 1–10. https://doi.org/10.1080/09571260600633135

Van Leeuwen, C., Roby, J.-P., De Rességuier, L. (2018). Soil-related terroir factors: a review. *OENO One* 52, 173–188. https://doi.org/10.20870/oeno-one.2018.52.2.2208

Vaudour, E. (2002). The Quality of Grapes and Wine in Relation to Geography : Notions of Terroir at Various Scales. *Journal of Wine Research*, 1264(October), 117–141. https://doi.org/10.1080/095712602200001798

Vianna, L. F. de N., Massignan, A. M., Pandolfo, C., & Dortzbach, D. (2019). Evaluating environmental factors, geographic scale and methods for viticultural zoning in the high-altitude region of Santa Catarina, Brazil. *Remote Sensing Applications: Society and Environment*, 13, 158–170. https://doi.org/10.1016/j.rsase.2018.10.018

Wambeke, A. *Soils of the tropics*. New York, McGraw-Hill, 1992. 342p.

Weill, R. R., & Brady, N. C. (2017). *The nature and properties of soils* (15th ed.).

White, R. E. (2003). *Soils for Fine Wines*. New York: Oxford University Press.

Zadeh, L. A. (1965). Fuzzy sets. *Information and Control*, 8, 338–353.

Zhu, A. (1997a). A similarity model for representing soil spatial information. *Geoderma*, 77, 217–242.

Zhu, A. (1997b). Measuring uncertainty in class assignment for natural resource maps under fuzzy logic. *Photogrammetric Engineering & Remote Sensing*, 63 (10), 1195–1202.

Zhu, A.-X., Yang, L., Li, B., Qin, C., Pei, T., & Liu, B. (2010). Construction of membership functions for predictive soil mapping under fuzzy logic. *Geoderma* 155, 164–174.