Agronomy Research **17**(2), 418–429, 2019
https://doi.org/10.15159/AR.19.114

Principal components in the study of soil and plant properties in precision coffee farming

G.A.S. Ferraz[1], P.F.P. Ferraz[1], F.B. Martins[2], F.M. Silva[1], F.A. Damasceno[1] and M. Barbari[3]

[1]Federal University of Lavras – UFLA, Departament of Agricultural Engineering, University Campus, BR37200-000 Lavras-MG, Brazil
[2]Rural Federal University of Rio de Janeiro – UFRRJ, BR-465, Km 7, BR 23.897-000 Seropédica- RJ, Brazil
[3]Department of Agriculture, Food, Environment and Forestry (DAGRI), Università degli Studi di Firenze, Via San Bonaventura, 13, IT50145 Firenze, Itália
[*]Correspondence: gabriel.ferraz@ufla.br

Abstract. In this work, a principal component analysis was performed to evaluate the possibility of discarding obsolete soil and plant variables in a coffee field to eliminate redundant and difficult-to-measure information in precision coffee farming. This work was conducted at Brejão Farm in Três Pontas, Minas Gerais, Brazil, in a coffee field planted with 22 ha of Topázio cultivar. The evaluated variables were the yield, plant height, crown diameter, fruit maturation index, degree of fruit maturation, leafing, soil pH, available phosphorus (P), remaining phosphorus (Prem), available potassium (K), exchangeable calcium (Ca2+), exchangeable magnesium (Mg2+), exchangeable acidity (Al3+), potential acidity (H + Al), aluminium saturation ($N_{(Al)}$), potential CEC (CECp), actual CEC (CECa), sum of bases (SB), base saturation (BS) and organic matter (OM). The data were evaluated by a principal component analysis, which generated 20 components. Of these, 7 representing 88.98% of the data variation were chosen. The variables were discarded based on the preservation of the variables with the greatest coefficients in absolute values corresponding to the first component, followed by the variable with the second highest absolute value corresponding to the second principal component. Based on the results, the variables V, OM, fruit maturity index, plant height, yield, leafing and P were selected. The other variables were discarded.

Key words: Multivariate analysis, coffee plant, precision agriculture, fertility, management.

INTRODUCTION

Precision agriculture in coffee production has been called Precision Coffee Farming (Ferraz et al., 2012a), and it can be defined as the set of techniques and technologies based on the spatial variability of soil and plant properties capable of assisting the coffee farmer in crop management to maximize profitability and increase fertilizer, spraying and harvesting efficiency, thereby leading to increased yield and final quality of the product (Ferraz et al., 2012b).

However, the use of precision coffee farming often requires the use of several variables for more accurate decision-making, which makes the work more complex and, in some cases, more costly (Almeida et al., 2011). With a high number of variables, many of them can contribute to the characteristics under evaluation. Thus, redundant and difficult-to-measure variables can be eliminated, thus reducing the time and costs of experiments (Leite et al., 2009).

According to Olive (2017) principal component analysis is used to explain the dispersion structure with a few linear combinations of the original variables. Jolliffe & Cadima (2016) affirms that large datasets are increasingly widespread in different areas and in order to interpret such datasets, methods are required to drastically reduce their dimensionality in an interpretable way. Principal component analysis (PCA) is used to obtain a small number of linear combinations (principal components) of a set of variables that retain as much information on the original variables as possible (Jolliffe, 1972; Jolliffe & Cadima, 2016; Olive, 2017). According to Morais (2011), PCA helps to reduce the size of a data set by selecting the principal components (PCs), discarding the original variables and excluding possible outliers and Jolliffe & Cadima (2016) affirms that PCA is one of the oldest and most widely used methods to perform it.

The principal components were used for different aspects of coffee farming, such as evaluating characters related to the vegetative growth of Arabica coffee cultivars (Freitas et al., 2007), discriminating between maturation stages and types of post-harvest processing (Arruda et al., 2011), performing multiple linear regression modelling of coffee crop yield (Carvalho et al., 2004), evaluating morphological characters of Arabica coffee (Teixeira et al., 2013) and studying the spatial variability of chemical attributes of a soil cultivated with coffee plants (Silva et al., 2010b; Silva & Lima 2012). But all of these studies had just focused on soil or on plant features separately. It is known that plant and soil can affect the coffee production, so they need to be studied together in order to obtain more and better information about the coffee production.

Thus, the aim of this work was to perform a principal component analysis of data obtained by precision coffee farming and evaluate the possibility of discarding soil and plant variables in a coffee field to eliminate redundant and difficult-to-measure information in precision coffee farming, so making the precision coffee farming more feasible to farmers.

MATERIALS AND METHODS

The experiment was carried out at the Brejão farm, which is located in the Três Pontas Municipality in southern Minas Gerais, Brazil. The study area was cultivated with 22 hectares of coffee (*Coffea arabica* L.) of the Topazio cultivar, which was transplanted in December 2005 at a spacing of 3.8 m between rows and 0.8 m between plants, for a total of 3289 plants ha^{-1}. The geographical coordinates of the central point of the area are 21° 25' 58" south latitude and 45° 24' 51" west longitude.

The local climate is characterized as mild, tropical of altitude, with moderate ambient temperatures and hot and rainy summer, classified by Köppen as Cwa (Sá Junior et al., 2012). The soil was classified as Haplustox (EMBRAPA, 2016).

In this study, 20 variables were used, and 14 of these variables were soil related, and 6 were coffee plant related. Thus, sampling was carried out in the study area, in which a regular 57 x 57 m sampling grid was delimited, and a total of 64 georeferenced points (average of 2.9 points per hectare) were sampled using the Topcon FC-100 GPS data collector (Topcon Positioning Systems Inc, Livermore, Calif., USA), whose mean error was 10 cm. Within this grid, another four regular sampling grids (called zoom) were created, and the points were spaced at 3.8 x 3.8 m. These grids were positioned at four different points of the main grid. Each magnification will correspond to 10 georeferenced sampling points (one point of the main grid and nine points of the new grid). Therefore, the grid was composed of 100 georeferenced sampling points (Fig. 1).

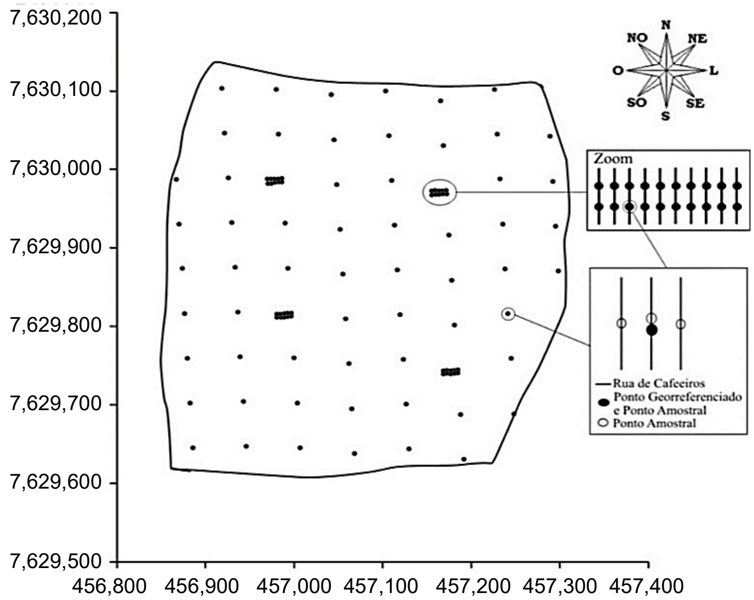

Figure 1. Sampling grid and procedure.

Each sampling point corresponded to four plants: two plants located in the coffee row where the point was georeferenced and the other two plants located in each row to the side of the reference point.

The soil samples were collected by subsampling in the canopy projection of the coffee plants at a depth of 0 to 20 cm using a Dutch auger. The traditional soil sampling diffused for coffee cultivation, recommends that soil sampling should be performed from 0–20 cm (Guimarães et al., 2002; Matiello et al., 2010) depth and it was performed by many studies (Ferraz et al., 2012b; Silva et al., 2010a; Silva et al., 2013; Silva & Lima 2012). At each sampling point, a sub-sample of each of the four plants that make up this point was collected. The subsamples of each sampling point were homogenized to form a composite sample for the point.

The evaluated soil chemical properties were as follows: soil pH, available phosphorus (P) (Mehlich 1), remaining phosphorus (Prem), available potassium (K) (Mehlich 1), exchangeable calcium (Ca^{2+}) (extractor: KCL – 1 mol L^{-1}), exchangeable magnesium (Mg^{2+}) (extractor: KCL – 1 mol L^{-1}),exchangeable acidity (Al^{3+}) (extractor: KCL – 1 mol L^{-1}), potential acidity (H + Al) (extractor: SMP), aluminium saturation ($N_{(Al)}$), potential cation exchange capacity (CECp) actual cation exchange capacity (CECa), sum of bases (SB), base saturation (BS) and organic matter (OM). The soil samples were sent to the soil analysis laboratory of the Department of Soil Science of the Federal University of Lavras for proper analysis of these properties.

The six coffee plant-related variables studied were yield (YIELD), plant height (HEIGHT), crown diameter (CROD), fruit maturation index (FMI), degree of fruit maturation (DFM) and leafing (LEAF).

The coffee YIELD (L plant^{-1}) was obtained by manual harvesting on cloths around the four plants at the sampling point, and the volume harvested from each plant after shaking was measured in a container graduated in litres. After this measurement, the mean YIELD of these four plants was obtained and used as the YIELD for the sampling point.

At each sampling point, after the YIELD measurements, the fruits collected from the four plants composing the point were placed in the same container, where they were homogenized to obtain a 0.5-L sample of fruits (Carvalho et al., 2003; Silva, 2008). Using this sample, the number of fruits at each maturation stage (dry, raisin, cherry and green) was counted and transformed to a percentage for use in equation 1 described by Alves et al. (2009) to calculate the FMI.

$$FMI = \quad \%cherry, \%raisin, \%dry \tag{1}$$

The same fruits used to obtain the FMI were also used to determine the DFM. According to Silva et al. (2010a), the literature does not provide information about the maturation stage parameters or aims and only reports the indices related to fruit colour and days after flowering. Thus, these authors stipulated a scale of scores ranging from 1 to 4, where a value of 1.0 is given for the green maturation stage, 2.0 is given for the cherry maturation stage, 3.8 is given for the raisin maturation stage and 4.0 is given for the dry maturation stage. Thus, the authors created an index called the DFM, and it is presented in Eq. 2.

$$DFM = \frac{4(\%\, dry) + 3.8(\%\, raisin) + 2(\%\, cherry) + 1(\%\, gree}{(\%\, dry + \%raisin + \%cherry + \%\, gree} \tag{2}$$

In the four plants that represented the sampling point, the HEIGHT and CROD were measured using a ruler graduated in millimetres. HEIGHT was measured from the soil surface to the top of the plant, and the CROD represented the measurement of the longest branch. After this measurement, the mean height and CROD of each sampling point was obtained in metres.

The visual scale proposed by Boldini (2001) was used to evaluate the LEAF, and the classes range from 0 to 20%, from 21 to 40%, from 41 to 60%, from 61 to 80% and from 81 to 100%.

Souza et al. (2008) reinforced that the multivariate analysis technique has an advantage relative to the univariate analysis methods in that it evaluates the level to which each characteristic studied explains the total variance among the evaluated treatments. Thus, less discriminating characteristics could be discarded since they are already correlated with other redundant variables or present invariance.

The methodology proposed by Jolliffe (1972) was used to select the number of principal components (PCs), where the p number of significant PCs can be defined by the number of PCs required to explain a percentage of the total data variance (cumulative) greater than 90% or by a value of the associated eigenvalue (λ) greater than 0.7. The eigenvalue can be obtained by Eq. 3.

$$det([T]_A - \lambda I_n) = 0 \tag{3}$$

where I_n is the identity matrix of order n and is called the characteristic equation. From this equation, the polynomial $[T]_A - \lambda I_n$ is obtained, which is the polynomial characteristic of T, and its roots in R are the eigenvalues of the linear operator T.

In addition, the classical B4 method for discarding variables proposed by Jolliffe (1972), which is based on the preservation of most of the variation in the data, was adopted for discarding the variables. The B4 method involves the use of the first p PCs selected, and the variable with the greatest absolute value of eigenvectors corresponding to the first PC is selected. The next variable to be selected will be the highest absolute value of the eigenvectors corresponding to the second PC, which continues until the selected p PC. Unselected variables will be discarded. In this way, the number of variables selected is equal to the number p of PCs.

All analyses were performed using the R (R Development Core Team, 2018) statistical platform.

RESULTS AND DISCUSSION

The PCA generated 20 PCs, and the first 7 PCs explained 88.98% of the total variance of the data under study and presented a variance greater than 0.7 (eigenvalue greater than 0.7); thus, they were considered significant PCs (Table 1). Moreover, 45.10%of the data variation can be explained by the first PC. Silva et al. (2010b) studied 22 soil variables and observed that the first PC explained 42.00% of the data variation.

Following the B4 method for discarding variables proposed by Jolliffe (1972) the values with the highest absolute value of each of the significant PCs (PC1, PC2, PC3, PC4, PC5, PC6 and PC7) can be selected: BS, OM, FMI, HEIGHT, YIELD, LEAF and P, respectively. These variables correspond to the values highlighted in bold in Table 2. The other variables can be discarded.

The variables to be discarded showed significant simple linear correlations with the other variables (Table 3) and thus were redundant because such correlations present a statistical relationship that involves dependence between variables.

Table 1. Principal components (PCs), eigenvalues, percentage of variance explained by the PCs (% VPC) and cumulative percentage of variance explained by the PCs of the soil and plant attributes of the studied coffee plants

	Eigenvalue	%VPC	Cumulative %VPC
PC1	**9.02**	45.10	45.10
PC2	**2.28**	11.40	56.50
PC3	**1.96**	9.81	66.31
PC4	**1.67**	8.35	74.66
PC5	**1.13**	5.63	80.28
PC6	**0.92**	4.60	84.90
PC7	**0.82**	4.09	88.98
PC8	0.58	2.90	91.88
PC9	0.53	2.64	94.52
PC10	0.36	1.79	96.32
PC11	0.32	1.62	97.93
PC12	0.17	0.84	98.77
PC13	0.10	0.51	99.29
PC14	0.08	0.38	99.67
PC15	0.05	0.22	99.90
PC16	0.01	0.05	99.95
PC17	0.01	0.05	100.00
PC18	0.00	0.00	100.00
PC19	0.00	0.00	100.00
PC20	0.00	0.00	100.00

Table 2. Coefficients of the 20 principal components (eigenvectors)

	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9	PC10	PC11	PC12	PC13	PC14	PC15	PC16	PC17	PC18	PC19	PC20
YIELD	0.111	-0.178	-0.074	0.004	**-0.479**	-0.556	0.404	-0.219	-0.430	0.064	0.072	0.000	0.010	0.075	0.034	-0.006	0.010	0.000	0.000	0.000
HEIGHT	-0.012	-0.135	0.098	**0.624**	-0.261	0.081	0.053	-0.201	0.406	-0.315	0.382	0.221	-0.006	0.064	-0.021	0.010	-0.009	0.000	0.000	0.000
CROD	0.033	-0.330	-0.056	0.413	-0.305	0.255	-0.019	**0.625**	-0.176	0.233	-0.258	-0.111	0.049	-0.017	0.017	-0.004	0.011	-0.001	0.000	0.000
FMI	-0.011	0.290	**-0.599**	0.105	-0.108	-0.116	0.051	0.095	0.101	-0.132	-0.053	0.018	-0.299	-0.616	-0.097	0.009	0.014	-0.001	0.000	0.000
DFM	-0.005	0.319	-0.594	0.087	-0.049	-0.052	-0.118	0.057	0.057	-0.046	-0.080	0.038	0.229	**0.669**	0.051	0.008	-0.008	0.001	0.000	0.000
LEAF	0.024	0.182	0.297	0.265	0.285	**-0.645**	-0.148	0.414	0.001	-0.301	-0.047	-0.146	-0.011	0.038	-0.018	0.029	-0.005	0.000	0.000	0.000
Ph	0.322	-0.046	0.019	-0.026	-0.016	0.084	-0.043	-0.033	-0.061	-0.058	-0.019	-0.113	-0.352	0.229	**-0.793**	0.236	0.035	-0.002	0.001	0.000
P	-0.042	0.308	0.090	0.200	0.272	0.290	**0.776**	0.023	-0.153	-0.180	-0.142	-0.116	0.038	0.063	-0.008	0.008	0.000	0.000	0.000	0.000
K	0.256	-0.078	0.022	0.208	-0.049	-0.019	-0.145	-0.467	0.151	-0.121	**-0.626**	-0.408	0.104	-0.078	0.122	-0.013	-0.084	0.084	0.036	0.000
Ca	0.320	-0.067	-0.021	-0.044	0.116	-0.034	0.103	0.102	0.083	0.063	-0.049	0.333	-0.152	0.043	0.167	0.128	-0.244	**0.712**	0.308	0.000
Mg	0.290	-0.073	-0.169	0.004	0.155	-0.021	0.069	0.048	0.087	0.139	0.437	-0.410	**0.599**	-0.212	-0.167	0.078	0.003	0.153	0.066	0.000
Al	-0.310	-0.109	-0.064	0.093	0.109	-0.092	0.061	-0.080	0.145	0.229	0.045	-0.275	-0.247	0.111	0.052	0.040	0.585	0.323	0.015	-0.417
H + Al	-0.302	-0.211	-0.088	0.047	0.151	-0.107	0.051	-0.059	0.029	0.018	-0.157	0.149	0.160	-0.030	-0.265	-0.125	-0.002	-0.216	**0.780**	0.000
SB	0.325	-0.072	-0.044	-0.015	0.114	-0.032	0.078	0.045	0.093	0.061	-0.022	0.156	-0.008	-0.009	0.116	0.109	-0.201	-0.393	-0.016	**-0.779**
CECp	0.264	-0.217	-0.129	0.058	0.286	-0.136	0.185	0.004	0.285	0.305	0.004	0.015	-0.233	0.085	0.239	0.216	0.186	-0.367	-0.014	0.468
CECa	-0.229	-0.351	-0.155	0.059	0.292	-0.176	0.125	-0.057	0.102	0.066	-0.242	0.317	0.226	-0.050	-0.309	-0.093	-0.130	0.150	-0.539	0.000
BS	**0.328**	-0.002	-0.009	-0.017	0.051	-0.007	0.055	0.058	0.097	0.049	0.056	-0.020	-0.142	0.073	-0.066	**-0.910**	0.101	0.006	-0.008	0.000
$N_{(Al)}$	-0.314	-0.045	-0.057	0.069	0.043	-0.054	0.050	-0.032	0.099	0.241	0.186	**-0.414**	-0.314	0.123	0.011	-0.086	**-0.698**	0.001	-0.001	0.000
OM	0.029	**-0.443**	-0.292	0.036	0.337	0.140	-0.146	-0.059	-0.446	**-0.495**	0.181	-0.106	-0.163	0.035	0.203	-0.028	0.009	0.002	0.000	0.000
Prem	0.075	0.281	0.033	0.490	0.259	0.023	-0.258	-0.287	**-0.448**	0.446	0.093	0.197	-0.019	-0.090	-0.032	-0.027	-0.003	0.000	0.000	0.000

Table 3. Simple correlation coefficient between the studied variables

	YIELD	HEIGHT	CROD	FMI	DFM	LEAF	Ph	P	K	Ca	Mg	Al	H + Al	SB	CECp	CECa	BS	N(Al)	MO	Prem
YIELD	1.00																			
HEIGHT	0.09	1.00																		
CROD	0.17	0.45	1.00																	
FMI	0.05	-0.05	-0.06	1.00																
DFM	-0.05	-0.11	-0.09	0.90	1.00															
LEAF	-0.02	0.11	-0.10	-0.12	-0.13	1.00														
Ph	0.31	-0.04	0.13	-0.10	-0.07	0.00	1.00													
P	-0.19	0.07	-0.13	0.10	0.05	0.11	-0.16	1.00												
K	0.28	0.23	0.16	-0.07	-0.04	0.06	0.76	-0.16	1.00											
Ca	0.31	-0.09	0.10	-0.05	-0.05	0.07	0.92	-0.11	0.68	1.00										
Mg	0.28	-0.05	0.10	0.09	0.11	-0.01	0.81	-0.13	0.59	0.85	1.00									
Al	-0.26	0.13	-0.02	0.04	0.01	-0.06	-0.90	0.08	-0.65	-0.87	-0.72	1.00								
H + Al	-0.22	0.08	0.02	0.00	-0.03	-0.09	-0.87	0.02	-0.62	-0.81	-0.71	0.94	1.00							
SB	0.32	-0.06	0.11	-0.03	-0.03	0.06	0.93	-0.12	0.73	0.99	0.89	-0.87	-0.81	1.00						
CECp	0.29	0.02	0.16	-0.02	-0.03	0.04	0.75	-0.13	0.63	0.88	0.84	-0.55	-0.52	0.89	1.00					
CECa	-0.11	0.07	0.10	-0.03	-0.07	-0.10	-0.67	-0.05	-0.43	-0.53	-0.46	0.80	0.93	-0.54	-0.18	1.00				
BS	0.30	-0.05	0.08	-0.02	-0.01	0.08	0.95	-0.10	0.72	0.97	0.88	-0.90	-0.89	0.98	0.83	-0.66	1.00			
N(Al)	-0.28	0.10	-0.04	0.07	0.04	-0.08	-0.91	0.10	-0.71	-0.90	-0.73	0.98	0.89	-0.90	-0.63	0.71	-0.91	1.00		
OM	0.05	0.01	0.28	-0.01	0.00	-0.25	0.15	-0.25	0.11	0.15	0.27	0.03	0.20	0.17	0.32	0.40	0.08	-0.04	1.00	
Prem	-0.13	0.24	0.03	0.12	0.21	0.35	0.18	0.25	0.30	0.13	0.17	-0.18	-0.27	0.16	0.10	-0.29	0.19	-0.18	-0.08	1.00

The parameter BS is an important index of soil acidity for establishing adequate limestone doses for crops and management strategies for agricultural production (Fageria, 2011). The concept of BS is related to the supply of bases (Ca, Mg, K) at optimal levels for the development of plants (McLean, 1977). Thus, the choice of BS can be explained since this variable indicates the percentage of cation exchange sites that are occupied by bases, that is, the percentage of negative charges at pH 7,0 occupied by Ca^{2+}, K^{+}, (Na^{+}) compared to the sites occupied by H^{+} and Al^{3+}.

PC1 included BS and presented correlation values with the items included in this index (Table 2). A positive correlation was observed with Ca, K and CEC at pH 7,0 (CECp), and a negative correlation was observed with Al and with H + Al, which justifies the choice of V since these parameters are already included in the composition of the index.

PC2 included OM as the variable to be chosen. According to Mielniczuk et al. (2003), OM is one of the main parameters in the evaluation of soil quality and had a strong effect on the physical, chemical and biological characteristics of the soil.

Xavier et al. (2006) stated that soil OM is one of the main sources of energy and nutrients to the system, and it is capable of maintaining the soil YIELD in general. Silva & Resck (1997) reported that the benefits generated by OM include improvements in the physical conditions of the soil and the energy supply for microbial growth, which Paes et al. (1996) indicated leads to increased nutrient cycling and soil CEC. Table 3 shows that the t and T are directly correlated with the OM. These two variables are also correlated with PC2.

It should be noted that the behaviour of PC1 and PC2 in this study was similar to that observed by Silva et al. (2010b).

According to Novais & Smyth (1999), P is one of the most limiting elements for crops grown under tropical soil conditions due to their soil-sink characteristics. In the coffee crop, Guimarães et al. (2002) indicated that the P requirement is small in the adult stage of the crop relative to the nitrogen and K requirements. In the young stage, the P requirement is greater, which is also observed for the other nutrients. P affects the development of the root system and formation of the xylem of the plant, and it is also very important in grain formation. Therefore, the study of this variable is very important.

The plant characteristics selected by the PCs were HEIGHT, FMI, YIELD and LEAF. DFM was excluded because its information was already included in the FMI, which made it redundant.

The study of YIELD in precision coffee farming is of great importance because these data can be used to infer soil parameters that may be detrimental to the good productive performance of the plant, and they are also important for the adequate planning and management of the harvesting stage and for determining whether to implement manual, semi-mechanized or mechanized strategies.

According to Silva et al. (2006), the maturation index can be used to define the harvest period of a given plot. A plot that exhibits plants with 20 to 25% of green fruits (FMI of 75 to 80%) is considered to be at the beginning of the harvest; a plot that exhibits between 10 and 15% of green fruit is considered to be at the middle of the harvest (FMI of 85 to 90%); and a plot that exhibits less than 5% of green fruit is considered to be at the end of harvest (FMI of 95%).

Boldini (2001) developed a grading scale for classifying the LEAF a coffee plant: 0 to 20% LEAF is assigned a score of one, 21 to 40% is assigned a score of two, 41 and

60% is assigned a score of three, 61 to 80% is assigned a score of four; and from 81 to 100% is assigned a score of five. This variable is analysed to assess the effect of pest organisms in the crop LEAF (Silva et al., 2013).

PC4 indicated that HEIGHT is a variable to be evaluated. This parameter is an important growth trait of the plant and indicates its development. This characteristic is closely related to the crop management conditions. When analysing PC4 (Table 2), CROD also stood out as one of the characteristics strongly related to this PC, and it is also an indicator of plant development. Therefore, because HEIGHT is a sufficient indicator of plant development, CROD was redundant and discarded.

The study of fruit detachment force, both for green and cherry fruits, can be an important indicator for selecting selective mechanized harvesting. Silva et al. (2010b) reported that a greater difference in the detachment force between green and cherry fruits corresponds to improved selective mechanized harvesting of the coffee fruits. In addition, this variable can be used to indicate when to start this type of harvest (Silva, 2008).

When studied together, these four coffee plant-related variables provide the foundation for the study of crop development based on the type of management implemented and allow for the identification of abnormal development. These variables can also be used to define the most adequate crop management activities.

Therefore, the selection of these soil and plant variables can be valuable for the coffee grower. The use of precision agriculture can generate a lot of information that may not always be useful to producers who may be still being spent on time and money. So, focusing your time on variables that can really give accurate information to farm management becomes essential. This paper proves that was possible reduce the number of variables that can be useful for the coffee producers.

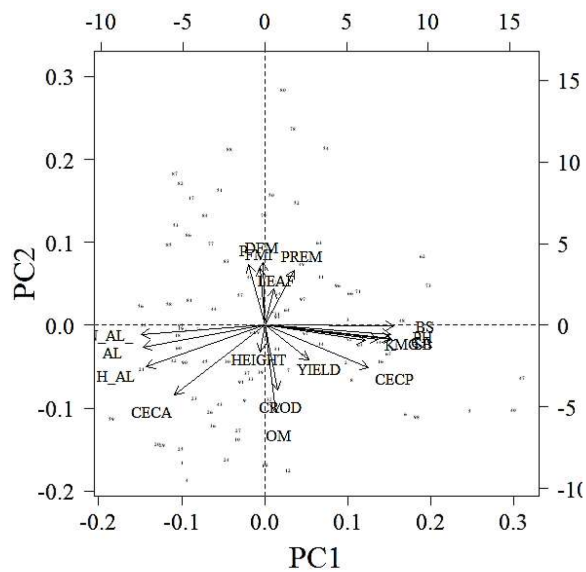

Figure 2. Biplot of the first two principal components.

The PCA analyses can also give us a good information about the relationship between variables. So, in order to study the relationship of the chosen variables with the yield it was performed a biplot chart of PC1 and PC2 (Fig. 2). It is possible to observe in the Fig. 2 that exists a direct relationship between BS and YIELD, OM and YIELD and HEIGHT and YIELD. Otherwise, there are an inverse relationship between FMI and YIELD, LEAF and YIELD and P and YIELD. So, the right management of this variables will be important to increase the coffee yield.

CONCLUSIONS

The use of the PCA is important in the field of precision coffee farming because it can identify soil and plant variables that can be discarded to remove repetitive and difficult-to-measure information. Thus, the variables selected for this study were SB, OM, FMI, HEIGHT, YIELD, LEAF and P. It was possible as well to use the PCA to study the relationship among the coffee yield among the other six selected variables. So, with the results present in this study the coffee farmer could be more focused in some variables, so reduce time consuming to analyse all of the many data that precision agriculture can generate.

ACKNOWLEDGEMENTS. The authors thank the National Council for Scientific and Technological Development (CNPq), the Coordination for the Improvement of Higher Education Personnel (CAPES), the Foundation for Research of the State of Minas Gerais (FAPEMIG), the Federal University of Lavras (UFLA), the Federal Rural University of Rio de Janeiro (UFRRJ) the Postgraduate Program in Agricultural and Environmental Engineering (PGEAAmb).

REFERENCES

Almeida, A.P.S. 2011. Impact of the ICMS Credit on the Cost of Production in Coffee Production: a Study of the Main Producing Regions of Arabica Coffee in Brazil. 2011. Available: < https://repositorio.ufu.br/handle/123456789/11948> Accessed 7.12.2018 (in Portuguese).

Alves, M.C., Silva, F.M., Ferreira, T.A. & Silva, F.C. 2009. Neuro-fuzzy operational performance of a coffee harvester machine. *Journal of convergence information technology* **4**, n.2, 52–59.

Arruda, N.P., Hovell, A.M., Rezende, C.M., Freitas, S.P., Couri, S. & Bizzo, H.R. 2011. Discrimination between maturation stages and types of post-harvest processing of arabica coffees by solid phase microextraction and principal component analysis. *Química Nova*, São Paulo **34**(5), 819–824 (in Portuguese).

Boldini, J.M. 2001. *Epidemiology of rust and cercosporiosis in irrigated and fertigated coffee.* 67 pp. Dissertation (Masters in Phytopathology), Federal University of Larvas, Lavras, 2001 (in Portuguese).

Brazilian Agricultural Research Corporation. National Soil Research Center. *Brazilian classification soil system (*Empresa Brasileira de Pesquisa Agropecuária. Centro Nacional de Pesquisas de Solos. Sistema Brasileiro de classificação do Solo - in Portuguese). 2 ed. Rio de Janeiro, 2006, 306 pp.

Carvalho, G.R., Mendes, A.N.G., Carvalho, L.F. & Bartholo, G.F. 2003. Ethephon efficiency in the uniformization and anticipation of fruit maturation (Coffea arabica L.) and beverage quality. *Science and Agrotechnology* **27**(1), 98–106 (in Portuguese).

Carvalho, L.G.D., Sediyama, G.C., Cecon, P.R. & Alves, H.M. 2004. Regression model for the coffee produdivity forecast in the State of Minas Gerais. *Brazilian Journal of Agricultural and Environmental Engineering* **8**, 204–211 (in Portuguese).

Fageria, N.K. 2001. Response of upland rice, dry bean, corn and soybean to base saturation in cerrado soil. *Brazilian Journal of Agricultural and Environmental Engineering* **5**(3), 416–424. (in Portuguese).

Ferraz, G.A.S, Silva, F.M., Alves, M.C., Bueno, R.L. & da Costa, P.A.N. 2012a. Geostatistical analysis of fruit yield and detachment force in coffee. *Precision Agriculture* **13**(1), 76–89.

Ferraz, G.A.S., Silva, F.M.D., Carvalho, L.C., Alves, M.D.C. & Franco, B.C. 2012b. Spatial and temporal variability of phosphorus, potassium and of the yield of a coffee field. *Engenharia Agrícola* **32**(1), 140–150 (in Portuguese).

Freitas, Z.M.T.S.D., Oliveira, F.J.D., Carvalho, S.P.D., Santos, V.F.D. & Santos, J.P.D.O. 2007. Evaluation of quantitative traits related to the vegetative growth among cultivars of low size Arabica coffee. *Bragantia*, Campinas, **66**(2), 267–275 (in Portuguese).

Guimarães, R.J., Mendes, A.N.G., Souza, C.A.S. 2002. Coffee Nutrition: Extraction of Nutrients, Liming and Gessam in the Planting, Training and Production phases. In: *Coffee*. Lavras: UFLA / FAEPE, pp. 194–234 (in Portuguese).

Jolliffe, I.T. & Cadima, J.2016. Principal component analysis: a review and recent developments. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, **374**(2065), 1–16.

Jolliffe, I.T. 1972. Discarding variables in a principal component analysis. I. Artificial data. *Applied Statistics* **21**, 160–173.

Leite, C.D.S., Corrêa, G.D.S.S., Barbosa, L.T., Melo, A.L.P.D., Yamaki, M., Silva, M.D.A. & Torres, R.D.A. 2009. Evaluation of performance and carcass characteristics of cutting quails by principal components analysis. *Brazilian Journal of Veterinary Medicine and Animal Science* **61**(2), 498–503 (in Portuguese).

Matiello, J.B., Santinato, R., Garcia, A.W.R., Almeida, S.R. & Fernandes, D.R. 2010. *Coffe crop in Brazil*: *recommendation manual*. Rio de janeiro e Varginha: Fundação Procafé (in Portuguese).

McLean, E.O. 1977. Contrasting concepts in soil test interpretation: Sufficiency levels of available nutrients versus basic cation saturation ratios. In: Stelly, M. (Ed.) *Soil testing: Correlating and interpreting the analytical results*. Madison: American. Society of Agronomy, pp. 39–54.

Mielniczuk, J., Bayer, C., Vezzani, F.M., Lovato, T., Fernandes, F.F. & Debarba, L. 2003. Soil and crop management and their relationship with carbon and soil nitrogen stocks. In: Curi, N.; Marques, J.J., Guilherme, L.R.G., Lima, J.M., Lopes, A.S., Alvarez, V.H., eds. *Topics in soil science*. Viçosa, MG, Brazilian Society of Soil Science, **3**, pp. 209–248 (in Portuguese).

Morais, J.T.G. 2011. *Principal components analysis integrated to artificial neural networks for prediction of organic matter*. 67 pp. Dissertation (Master in Industrial Engineering) - Federal University of Bahia (in Portuguese).

Novais, R.F. & Smyth, T.J. Source-drain relationship of phosphorus in soil. In*: Phosphorus in soil and plant in tropical conditions*. UFV, 1999. 2–6 (in Portuguese).

Olive, D.J. 2017. Principal Component Analysis. In: Robust Multivariate Analysis. Springer, Cham, pp. 189–217.

Paes, J.M.V., Andreola, F., Brito, C.H. & Loudes, E.G. 1996. Decomposition of coffee straw in three types of soil and its influence on CTC and pH. *Ceres* **43**, 337–392.

R DEVELOPMENT CORE TEAM. R: a language and environment for statistical computing. Vienna: R Foundation for Statistical Computing, 2018. Available <http://www.R-project.org/>. Accessed:13.8. 2018.

Sá Junior, A., Carvalho, L.G., Silva, F.F. & Alves, M.C. 2012. Application of the Köppen classification for climatic zoning in the state of Minas Gerais, Brazil. *Theoretical and Applied Climatology* **108**, 1–7.

Silva, A.F., Lima, J.S.D.S., Souza, G.S. & Oliveira, R.B. 2010b. Spatial variability of chemical attributes of a humic red-yellow latosol cultivated with coffee. *Revista Brasileira de Ciências do Solo* **34**, 15–32 (in Portuguese).

Silva, F.C.D., Silva, F.M.D., Alves, M.D.C., Barros, M.M.D. & Sales, R.D.S. 2010a. Behavior of the detachment force of fruits of coffee trees during the harvest period. *Science and Agrotechnology* **34**(2), 468–474 (in Portuguese).

Silva, F.C. *Effect of the Detachment force and maturation of coffee fruits on mechanized harvesting*. 2008. 106 pp. Dissertation (Master in Agricultural Engineering) - Federal University of Lavras (in Portuguese).

Silva, F.M.D., Oliveira, E., Guimarães, R.J., Figueiredo, C.A.P.D. & Silva, F.C. 2006. Operational and economical performance of the coffee harvest with the use of the lateral harvester. *Coffee Science* **1**(2), 119–125 (in Portuguese).

Silva, F.M., Ferraz, G.A.S., Alves, M.C. 2013. Study case on precision coffee cultivation. In: Silva, F. M., Alves, M. C. (Ed). *Precision coffee makers*. Lavras: Editora UFLA, pp. 179–208. (in Portuguese).

Silva, J.E. & Resck, D.V.S. 1997. Soil organic matter. In: Vargas, M.A.T. & Hungary, M., ed. C*errado soils Biology*. Planaltina: Embrapa-CPAC, 467–524 (in Portuguese).

Silva, S.A. & Lima, J.S.S. 2012. Evaluation of the variability of nutritional status and coffee productivity through principal component analysis and geostatistics. *Ceres* **59**(2), 271–277 (in Portuguese).

Souza, J.M.L., Negreiros, J.R.S., Álvares, V.S., Leite, F.M.N, Souza, M.L., Reis, F.R & Felisberto, F.Á. V. 2008. Physicochemical variability of cassava flour. *Food Science and Technology* **28**(4), 907–912 (in Portuguese).

Teixeira, A.L., Gonçalves, F.M.A., Rezende, J.C., Rocha, R.B., Pereira, A.A. 2013. Principal components analysis on morphological characters of arabica coffee in juvenile stage. *Coffee Science* **8**(2), 205–210 (in Portuguese).

Xavier, F.A.D.S., Maia, S.M.F., Oliveira, T.S.D. & Mendonça, E.D.S. 2006. Microbial biomass and light organic matter in soils under organic and conventional agricultural systems in Ibiapaba - CE. *Brazilian Journal of Soil Science* **30**, 247–258 (in Portuguese).