Sugar Tech (Nov-Dec 2020) 22(6):954–966
https://doi.org/10.1007/s12355-020-00854-7

RESEARCH ARTICLE

A Multivariate Approach to Determine the Economic Profitability of Sugarcane Production Under Diverse Climatic Conditions in Brazil

Nilceu Piffer Cardozo[1] · Ricardo de Oliveira Bordonal[2] · Alan Rodrigo Panosso[3] · Carlos Alexandre Costa Crusciol[1]

Received: 27 January 2020 / Accepted: 22 June 2020 / Published online: 2 July 2020
© Society for Sugar Research & Promotion 2020

Abstract The prominent position of sugarcane as a source of renewable and sustainable energy resulted in the expansion of its production into regions under limiting climatic conditions, thus affecting patterns related to growth, ripening, and profitability. This study provides an assessment of the factors that compose the economic return of sugarcane production using a multivariate approach. Monthly data, including total recoverable sugars, price, productivity (for an average of five mechanized harvesting), and rainfall during seven harvest seasons (2011/2012–2017/2018), were used to perform the multivariate statistical analyses considering the climatic conditions of four regions in the state of São Paulo, southeastern Brazil (Araçatuba, Assis, Ribeirão Preto, and Piracicaba). The chosen techniques were hierarchical and non-hierarchical (*k*-means) cluster analysis and principal component analysis. The data indicated the existence of three groups of months that exhibited different performances: Groups I, II, and III with intermediate, high, and low gross economic returns, respectively. Although group organization presented regional variations, July, August, September, and eventually October (Group II) generally exhibited the best gross economic returns (R$5065.1 ha^{-1}). December and November (Group III) exhibited the lowest economic returns (R$4731.5 ha^{-1}), and April, May, and June (Group I) exhibited intermediate returns that were close to the annual average (R$4829.6 ha^{-1}). Given the territorial extent of Brazil and the significant variations in environmental conditions, the adaptation of sugarcane cultivation and harvesting strategies to the characteristics of each producing region is fundamental for the rational and sustainable exploitation of the crop in the country.

Keywords *Saccharum* spp. · Biomass production · Industrial quality · Cluster analysis · Strategic planning

Introduction

An increasing population has aggravated the stress on land and other natural resources to meet the growing demands for food, fiber, fodder, and fuel. In this regard, the sustainability of human action and the growing need for energy and food are the main challenges that humanity faces in the coming decades (Tammisola 2010; Lal 2013). Renewable energy sources, such as biofuels, are promising means to promote sustainable development and minimize human impact on the planet (Cardozo et al. 2016). Concerns over energy shortage, greenhouse gas (GHG) mitigation policies, and the pursuit of new sources of farm income can explain why energy policymakers consider biofuels to be the main alternative to fossil fuels in many countries (Bordonal et al. 2015).

As a leading global producer of agricultural commodities, Brazil is often considered a key player in the

✉ Nilceu Piffer Cardozo
nilceu.cardozo@terra.com.br

1 FCA/UNESP – Department of Crop Science and Plant Breeding, College of Agricultural Sciences/São Paulo State University, Av. Universitária 3780, Botucatu, SP 18610-307, Brazil

2 LNBR/CNPEM – Brazilian Biorenewables National Laboratory/Brazilian Center for Research in Energy and Materials, Rua Giuseppe Máximo Scolfaro 10000, Polo II de Alta Tecnologia, Campinas, SP 13083-100, Brazil

3 FCAV/UNESP – Department of Engineering and Exact Sciences, College of Agricultural and Veterinarian Sciences/ São Paulo State University, Via de Acesso Prof. Paulo Donato Castellane s/n, Jaboticabal, SP, Brazil

production of food and bioenergy (Chagas et al. 2016). Brazilian ethanol is a renewable source of energy derived from sugarcane, which has been recognized as a very important bioenergetic crop because of its rapid growth, high productivity, and greater energy balance (i.e., energy output in a liter of ethanol over the fossil fuel energy input required to produce it) (Bordonal et al. 2018; Jaiswal et al. 2017). Recent studies indicate that sugarcane-based ethanol is the most competitive fuel in terms of energy use and net carbon balance (Souza et al. 2017) and is the only fully developed and stabilized alternative (Goldemberg 2007).

After the USA, with 2018 production of 60.8 billion liters (56%) of global ethanol, Brazil is the world's second-largest producer with 29.9 billion liters (28%) of ethanol (RFA 2019). Through a pioneering and successful program to encourage the production and use of ethanol as an alternative to fossil fuel, Brazil has become the largest producer of sugarcane. Its production has increased by 160% during the past decades, from 257 million tons (Mt) in 2000/2001 to approximately 590 Mt in 2019/2020—achieved from a cultivated area of 9.0 million hectares mostly in the south–central (90%) region of the country (UNICA 2020; CONAB 2020).

In view of the growing world demand for liquid biofuels, the Brazilian ethanol industry has expanded its production into areas of the Brazilian Cerrado, which is currently responsible for 70% of Brazil's agricultural production and 43% of sugarcane produced in Brazil (Gomes et al., 2019). However, expanding a sugarcane plantation into new agricultural frontiers has drawn attention, especially for the Cerrado regions that exhibit limiting agroclimatic patterns for crop development and growth (Cardozo et al. 2016; Marin et al. 2016). When combined with interannual climate variability that affects the growth and ripening patterns of the crop, this expansion directly affects the production and quality of sugarcane (Vianna and Sentelhas 2016; Scarpare et al. 2015). The use of Cerrado areas, for instance, shows the fragility of sugarcane expansion in Brazil, which was based on projects that did not consider the effects of climate on crop production, thus providing insufficient information to enable managers to make the best investment and management decisions (Fachinelli and Pereira 2015). These regions are characterized by poor soil quality (Scarpare et al. 2016) and reduced potential for water storage and rainfall distribution (Walter et al. 2014), where highly intense dry periods may occur (Dias and Sentelhas 2017). Faced with this situation, the cultivation of sugarcane in the country goes through a new phase in which efficiency and sustainability are fundamental priorities.

The sugarcane sector has been going through a deep crisis in recent years after crop expansion to regions under limited environmental resources. Among other aspects contributing to this crisis include changes in the crop cultivation system, such as the intensification of mechanization, labor laws pressure, increased costs, and reduced remuneration. The consequence of these problems has been to extend the harvest window to months that are agronomically limiting as a means of reducing the industry's fixed costs. Whereas the sugarcane harvest was carried out between May and October/November before the expansion, harvests today are initiated in March/April and extend until December. Therefore, the crushing capacity in the intermediate months of harvest has been reduced by more than 10%, resulting in an imbalance in the production systems with adverse effects on sugarcane yield and raw material quality (Cardozo 2017).

The tools that can assist with such decision making include multivariate statistical analyses that simultaneously scrutinize variables, such as the economic return of the sugarcane crop, the features of crop production and quality, and the climatic variables involved in the growth and ripening of sugarcane. Multivariate statistics has emerged as an important tool for providing greater amounts of information than that generated using univariate methods (Saed-Moucheshi et al. 2013) such that multivariate statistical techniques allow a phenomenon to be analyzed by considering all of the relevant factors in an integrated manner. Understanding the factors individually is inadequate because of the interactions among themselves; therefore, they must be considered jointly (Hair et al. 2005; Hartigan 1975). Given the multiple genetic and environmental factors involved in the ripening of sugarcane, Cardozo et al. (2014) reported that the evaluation of the temporal variability of the ripening of a group of cultivars is best achieved using multivariate statistical techniques. Lawes and Lawn (2005) claim that, although multivariate techniques are empirical in nature, they can be applied to define and solve problems related to the sugarcane industry.

In this context, de Castro and Bernardo (2019) applied principal component analysis and partial least squares regression to identify the variables—among 50 related to the industrial process—that presented the greatest impact on the quality and quantity of the produced sugar in São Paulo, Brazil. Their result indicated that high sugar production occurred in conditions of a high industry stability indicator, high cane milling, and high sugarcane purity, also resulting in better sugar quality. In addition, multivariate analysis techniques are applied to understanding the influence of environmental factors on the genotype × environment interactions, helping implement more efficient plant material identification and evaluation procedures (Ramburan et al. 2011; Santchurn et al. 2012). This study aimed to evaluate the factors that compose the economic return of sugarcane production on the basis of climatic conditions prevailing in diverse locations in the state of São Paulo in southeastern Brazil. The specific objective is to provide sustainable ways to further increase producers' profitability.

Materials and Methods

This section presents the locations and characterization of the areas considered in the study and the variables considered to perform the statistical analysis (MANOVA, cluster, and principal component analyses).

Location and Characterization of Experimental Sites and Data Set

Four traditional sugarcane-producing regions in the state of São Paulo with long histories of crop production were considered (Fig. 1): Araçatuba (21°12′32″ S–50°25′58″ W; 390 m altitude), Assis (22°39′42″ S–50°24′44″ W; 560 m altitude), Ribeirão Preto (21°14′05″ S–48°17′09″ W; 615 m altitude), and Piracicaba (22°42′30″ S–47°30′00″ W; 546 m altitude). According to Rolim et al. (2007), the climates of these regions are classified as follows: tropical savanna (Aw) in Araçatuba; humid tropical (Cwa) with a transition to subtropical (Cfa) in Assis; tropical savanna (Aw) with a transition to humid tropical (Cwa) in Ribeirão Preto; and subtropical (Cfa) in Piracicaba. The classification of the soils that are most representative of each region and the normal climatic conditions (1984–2013) are presented in Table 1.

The data on sugarcane yield (tons of cane per hectare—TCH) and total recoverable sugars (TRS, kg per ton of cane) were used considering the monthly average values obtained from sugarcane mills (20 per location) in each region for seven harvest seasons (2011/2012–2017/2018). Typically, sugarcane in southeastern Brazil is cropped in a 5-year cycle of yearly harvests; thus, TCH comprises mean values of sugarcane yield obtained in five crop harvests to reduce possible uncertainties associated with the number of harvests between the assessed regions.

TRS (kg t^{-1}) represents the monthly average values obtained during each crop season. The TRS monthly prices (R\$ kg TRS^{-1}) correspond to the historical values determined by the Council of Sugarcane, Sugar and Ethanol Producers of the State of São Paulo (Conselho dos Produtores de Cana-de-Açúcar, Açúcar e Álcool do Estado de São Paulo—CONSECANA). Economic profitability (R\$ ha^{-1}) was obtained by multiplying the sugarcane yield per hectare (TCH) by its value in reais (obtained by multiplying the cane TRS and TRS price values). Production costs were not considered when calculating profitability.

Rainfall data corresponding to each region were obtained at nearby weather stations between 2010 and 2017 (Table 2). Accumulated rainfall values (mm) that represent the period of 120 days (4 months) before each month of harvest were considered because of their effects on the values of TCH and TRS, as described in Cardozo et al. (2015). The data set consists of 252 observations.

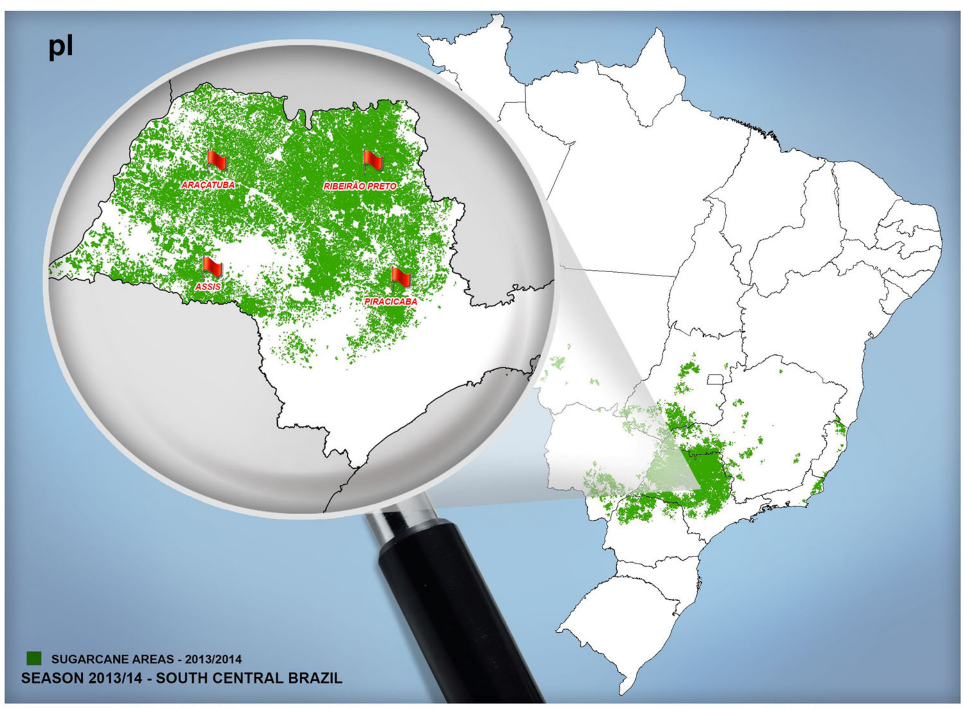

Fig. 1 Areas of sugarcane in Brazil. Red flags indicate the areas considered in this study Source: CANASAT-INPE (2014)

Table 1 Average climatic conditions (1988–2017) of the regions analyzed in this study

Region	Soil classification	Rainfall mm $year^{-1}$	Tmax (°C)	Tmin (°C)	Tmed (°C)
Araçatuba	Dystrophic Oxisol medium texture	1267.7	30.5	17.0	23.8
Assis	Yellow–red Ultisol	1441.5	28.7	15.4	22.1
Ribeirão Preto	Dystroferric Oxisol	1456.7	29.1	16.9	23.2
Piracicaba	Yellow–red Ultisol	1328.1	28.2	14.8	21.6

Table 2 Accumulated annual rainfall (2010–2017) of the regions evaluated in the state of Sao Paulo, southeastern Brazil

Rainfall (mm $year^{-1}$)				
Year	Araçatuba	Assis	Ribeirão Preto	Piracicaba
2010	1207.0	1452.7	1272.6	1378.4
2011	1206.1	1277.5	1615.0	1575.0
2012	1024.4	1217.2	1178.0	1244.8
2013	1046.5	1347.1	1453.0	1268.1
2014	1324.3	1461.0	815.6	1015.7
2015	1578.0	2227.1	1711.3	1648.3
2016	1354.5	1893.6	1604.5	1571.9
2017	1627.3	1646.6	1208.2	1249.7
Mean	1296.0	1565.4	1357.3	1369.0
SD[a]	222.2	343.8	296.9	215.9
CV[b] (%)	17.1	22.0	21.9	15.8

[a]Standard deviation

[b]Coefficient of variation (%)

Multivariate Analysis of Variance (MANOVA)

To assess the statistical hypothesis of the homoskedasticity and normality of the variance, Levene ($p > 0.05$) and Shapiro–Wilk ($p > 0.05$) tests were carried out for all variables before processing the data. These tests were used to confirm that all of the data represent a normally distributed population and to detect the presence of outliers. A multivariate analysis of variance (MANOVA) was performed to evaluate interactions among the production variables (months and regions). The purpose of MANOVA is to test whether the vectors of means (production variables) for the groups of factors' months and regions are sampled from the same sampling distribution. In other words, MANOVA is the extended version of the univariate analysis of variance and is a statistical method that examines the effect of two or more independent variables on two or more dependent variables (Ateş et al. 2019). The ith observation in the jth group on the mth variable was modeled additively in the same way as in ANOVA (Stahle and Wold 1990—Eq. 1)

$$x_{ijm} = \mu_m + \alpha_{jm} + e_{ijm} \quad (1)$$

where μ_m is the grand mean of the mth variable, α_{jm} is the effect of the jth treatment on the mth variable, and e_{ijm} is the error term assumed to have a multinormal distribution. The expected value is 0 for each variable (vector 0), and the dispersion around 0 is determined by the covariance matrix $\boldsymbol{\Sigma}$.

MANOVA has several advantages over ANOVA. First, by measuring several dependent variables in a single experiment, the chance of discovering the factor that is truly important is better. Second, the use of MANOVA can protect against Type I errors (the rejection of a true null hypothesis) that might occur if multiple ANOVAs were conducted independently. Multivariate hypothesis tests of the mean vector equality are determined based on four criteria (Hair et al. 2005): Wilks' lambda, Pillai's trace, Hotelling–Lawley trace, and Roy's greatest root. The Wilks' criterion (a test based on the likelihood ratio) is the most often used criterion for decision making, whereas the remaining criteria serve to complement it. In the present study, only two criteria were used: the Wilks and Pillai (a generalization of Roy's criterion) criteria (Gasperik 2010; Górecki and Smaga 2017). In addition, the main proposal for the use of MANOVA in this study was to explore how independent variables influence some response patterns in the production variables.

Multivariate Data Analyses

Subsequent multivariate analyses were performed using standardized values of TCH, TRS, TRS price, and rainfall, resulting in a null mean and unit variance (Hartigan 1975) by subtracting the mean and dividing it by the standard deviation (z-score):

$$Z_s(x_i) = \frac{Z(x_i) - \bar{Z}(x)}{\sigma(x)} \quad (2)$$

where $Z_s(x_i)$ is the variable $Z(x)$ value at the i observation after standardization; $\bar{Z}(x)$ is the $Z(x)$ value mean; and $\sigma(x)$ is the standard deviation of the variable $Z(x)$.

To characterize the multivariate structure contained in the data, three multivariate statistical methods were applied to sort group hits: hierarchical cluster analysis, non-hierarchical cluster analysis (k-means), and principal component analysis. Thus, the harvest months can be sorted according to the gross economic profitability obtained for each. Non-standardization could lead to inconsistencies in the solutions of the two techniques because most distance measurements are highly sensitive to different scales or to the magnitudes of the variables.

Cluster Analyses

In the context of multivariate analysis, these clustering techniques are classified as interdependence techniques in which no variable is defined as independent or dependent because the process involves the simultaneous analysis of all variables considered together. Cluster analysis (Sneath and Sokal 1973) was performed by calculating the Euclidean distance between months, for the four variables, and using Ward's algorithm to obtain clusters of similar months. The results of the analyses are presented in graphical form (a dendrogram), which aided in the identification of clusters of months. The clustering of months into groups was also obtained based on the k-means method (Hair et al. 2005), which belongs to the class of non-hierarchical and unsupervised clustering methods. This method minimizes the variance in hits within each group.

Principal Component Analysis

The multivariate structure of the initial dataset was analyzed using principal component analysis (PCA). The main idea of PCA is to reduce the dimensionality of a data set consisting of many variables that are correlated with each other, either strongly or weakly, while retaining the variation present in the dataset to the maximum extent possible. The first principal component extracted from the covariance matrix is a linear combination of the original variables and accounts for as much of the variation in the samples as possible. The second component is the second linear function of the original variables and accounts for the majority of the remaining variability. The remaining components are similarly defined. Principal component analysis summarizes the relevant information contained in p variables (p = 5, in the present study) in a smaller set of orthogonal latent variables termed principal components (eigenvectors), which are generated by linear combinations of the original variables from the eigenvalues of the covariance matrix. Each pair of principal components (PCs) generates a two-dimensional representation of the original sample space termed a biplot. The structure of the variables can be explained in this plot by directing beams of variables in the maximum variability regions. The first two principal components, PC1 and PC2, were considered, and their eigenvalues were greater than unity (Kaiser 1958). The coefficients of the linear functions that define the PCs were used to interpret the meaning, to enable the sign and relative size of the coefficients to be used as an indication of the weight assigned to each variable. Only coefficients greater than or equal to |0.50| were considered for analysis. SAS 9.3 software was used for the MANOVA, and the remaining analyses were run using STATISTICA version 9.0 (STATSOFT 2010).

Results and Discussion

Multivariate Analysis of Variance

MANOVA was used to investigate whether the treatment mean vectors were the same and, if not, the mean component that differed significantly. The results of the MANOVA for the entire dataset are shown in Table 3. The test of the hypothesis of no treatment effects was performed considering the extent of the sums of squares, treatment products, and residue on the basis of the generalized variance. Significant differences in the TCH and profitability variables were detected for the harvest months in all studied variables and locations (Table 3).

The source of the total variation was partitioned into the causes of the variation because of regions, months, and experimental error. For the two criteria used, the test statistics were calculated and converted to F statistics. The results of the test of the hypothesis regarding the equality of the mean vectors are shown in Table 4. According to the two criteria used (Wilks' and Pillai's criteria), the hypothesis of equality of the vector treatment effects

Table 3 Results of the multivariate analysis of variance for the variables TCH (t ha^{-1}), TRS (kg.t^{-1}), TRS price (R$ t^{-1}), rainfall (mm day^{-1}), and profitability (R$ ha^{-1}) in the studied months and regions

MS	TCH	TRS	TRS price	Rainfall	Profitability
Month	6.916**	146.318**	2.636**	91.755**	7666.933**
Regions	34.543**	0.374^{ns}	0.0001^{ns}	3.27^{ns}	15.293**
CV	9.39	3.55	17.07	25.02	17.41
SD	7.21	4.62	0.089	104.65	905.12
Mean	76.892	130.1502	0.524	418.165	5198.71

MS mean square, *CV* coefficient of variation, *SD* standard deviation

**significant at 1%, ns = nonsignificant according to the Snedecor F test. (1 BRL = 0.17 USD)

Table 4 Multivariate hypothesis tests for equality of the effects in the analyzed months and regions

Criteria	Value	*F*	Num DF	Den DF	Pr > *F*
Months					
Wilks' lambda	0.063	23.096	40	89.972	< .0001
Pillai's trace	1.491	12.904	40	120	< .0001
Regions					
Wilks' lambda	0.542	11.164	12	55.852	< .0001
Pillai's trace	0.489	9.572	12	69	< .0001

F = *F* approaches; Pr > *F* = *p* value associated with the F statistic; Num DF = numerator degrees of freedom; Den DF = denominator degrees of freedom

($p < 0.01$) was rejected, demonstrating that all mean components differed significantly. Cluster analyses (hierarchical and non-hierarchical) and principal component analysis were used to reduce the size and/or discard variables that were not relevant for describing the variability of the dataset.

Hierarchical Cluster Analysis

The dendrograms obtained by cluster analysis are presented in Fig. 2. For the set of variables considered, it was possible to divide the groups each time a significant variation was obtained for the values of the Euclidean distance between the hits. Three groups were identified: Group I (months of intermediate economic return); Group II (months of higher economic return); and Group III (months of lower economic return). Table 5 presents the Euclidean distances among the groups. Although the organization of the groups showed some regional differences, the months of July, August, September, and eventually October (Group II) generally provided the best economic return (R$5065.1 ha^{-1}) in all regions analyzed (mean Euclidean distance of 2.670 between regions). December and November presented the lowest return (R$4731.5 ha^{-1}) in Araçatuba, Ribeirão Preto, Piracicaba, and Assis and were grouped in Group III (mean Euclidean distance of 7.254). April, May, and June presented an intermediate position (Group I) and exhibited returns that were close to the annual average (R$4829.6 ha^{-1}) and exhibited a mean Euclidean distance of 4.738.

Non-hierarchical Cluster Analysis

Considering the existence of the three groups (based on the dendrograms), the clustering method based on *k*-means was applied to confirm the classification based on the analysis of hierarchical clusters (Fig. 3 and Table 6). The data confirmed the previously obtained groupings and indicated the variables that were significant for such sorting. In Araçatuba (Fig. 3a), TRS and rainfall were the decisive variables in the group formation ($p < 0.05$), suggesting that these variables are the most important factors for the creation of revenue in this region (Table 6). In contrast, in Assis (Fig. 3b), TCH, rainfall, and TRS price were the variables that were most responsible for the differences between the groups (Table 6). In this case, TRS did not significantly show the months that represented the best economic return in the region.

In Ribeirão Preto (Fig. 3c), all variables were significant ($p < 0.05$) for defining groups (Table 6). Thus, it can be

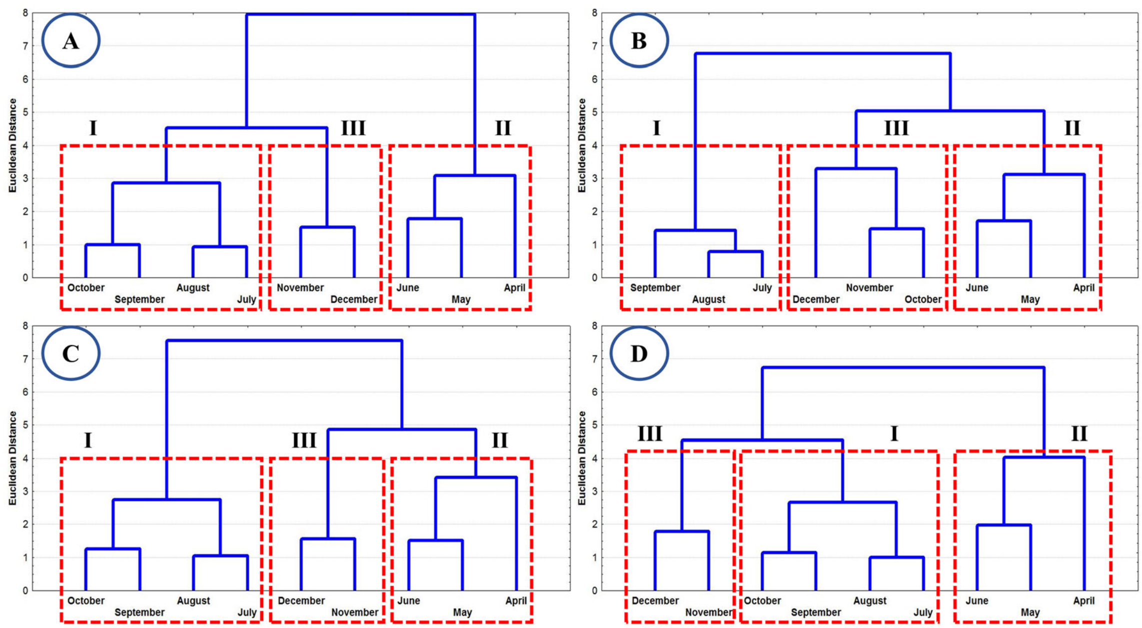

Fig. 2 Dendrogram resulting from the hierarchical cluster analysis showing the formation of groups according to TCH, TRS, TRS price, rainfall, and economic profitability in each region under study (**a** Araçatuba; **b** Assis; **c** Ribeirão Preto; **d** Piracicaba)

Table 5 Euclidean distances of each group of months clustered according to the economic return obtained from sugarcane production in various regions of the Sao Paulo state, southeastern Brazil

Region	Euclidean distance		
	Group I	Group II	Group III
Araçatuba	2.858	4.514	7.955
Assis	2.421	5.043	6.774
Ribeirão Preto	2.736	4.862	7.560
Piracicaba	2.663	4.532	6.728
Mean	2.670	4.738	7.254

inferred that the rainfall in this region is distributed in such a way that the balance between productivity gains (TCH) and TRS varied according to the months considered. At the beginning of the harvesting season (April and May), when TRS values were low, TCH values were higher, and this group of months differed from the others. At midharvest, TCH values were lower because of stalk dehydration, and then, the high TRS allowed group discrimination. The TRS price played an important role at the end of the harvesting season, when both TCH and TRS were low, especially in association with the degradation of raw material. Such degradation was caused by either rainfall (causing an inversion of sucrose and a decrease in TRS) or the prolonged dry season (causing excessive dehydration and decreases in TCH and TRS). Finally, in Piracicaba (Fig. 3d), the grouping of months based on the similarity of economic returns was mainly based on TRS values and rainfall ($p < 0.05$) (Table 6), suggesting that TRS in this region was the main driver of the variations in economic returns according to time of year.

Principal Component Analysis (PCA)

The distribution of sugarcane's gross economic profitability during the harvest months and the variables that control (TCH, TRS, TRS price) or influence (rainfall) this distribution in each region are shown in Figs. 4, 5, 6, and 7. Table 7 shows the correlation values measured for each variable and their relation to the PCs, the eigenvalues, and the percentage of variability retained in each component. The PCA enabled a single distribution of hits (PC1 $\times$ PC2) because only two eigenvalues were greater than unity in all cases. For the Araçatuba region (Fig. 4), the largest eigenvalue (the component that best explained the original variability) was 3.30, and the second-largest eigenvalue (the component that best explained the data variability after excluding those of the first component) was 1.46, explaining approximately 95.29% [66.02% (PC1) + 29.27% (PC2)] of the total information of original variables. In Assis (Fig. 5), the two largest eigenvalues were 3.22 and 1.33, which explained 91.08% [64.46% (PC1) + 26.61% (PC2)] of the total data variance. In the Ribeirão Preto region, the largest eigenvalues were 2.86 and 1.56, with 88.61% [57.37% (PC1) + 31.25% (PC2)] of the data variance being explained by the two principal components (Fig. 6). Finally, in Piracicaba (Fig. 7), the largest eigenvalues were 2.81 and 1.39, accounting for

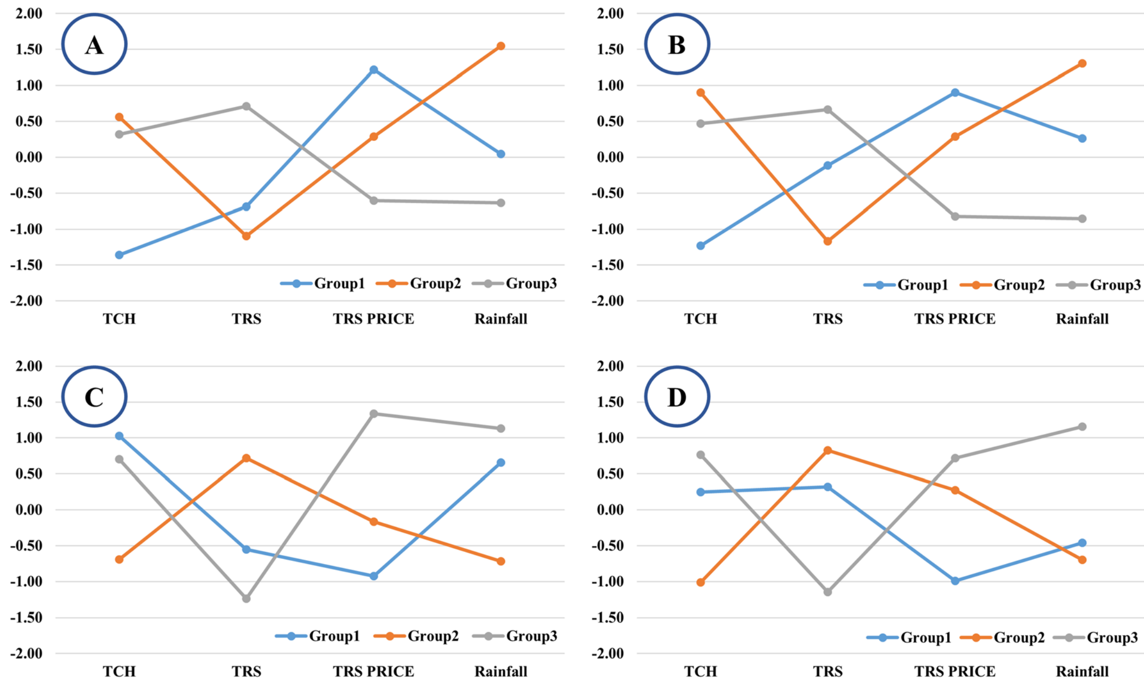

Fig. 3 Standard means of the variables determinant of the economic profitability of sugarcane (R$ ha^{-1}) for each group of months according to *k*-means non-hierarchical cluster analysis for the regions of Araçatuba (**a**), Assis (**b**), Ribeirão Preto (**c**), and Piracicaba (**d**)

Table 6 Analysis of variance for each variable of the groups formed by *k*-means non-hierarchical cluster analyses in various regions of the state of São Paulo, southeastern Brazil

Variable	SQ between groups	DF	SQ within groups	DF	*F*	Prob.
Araçatuba						
TCH	4.8166	2	3.1834	6	4.5390	0.0630
TRS	5.8756	2	2.1244	6	8.2975	0.0187
TRS price	4.9704	2	3.0296	6	4.9217	0.0543
Rainfall	6.8442	2	1.1558	6	17.7643	0.0030
Assis						
TCH	7.0773	2	0.9227	6	23.0107	0.0015
TRS	4.5367	2	3.4633	6	3.9298	0.0811
TRS price	5.3078	2	2.6922	6	5.9147	0.0381
Rainfall	6.5156	2	1.4844	6	13.1679	0.0064
Ribeirão Preto						
TCH	5.4914	2	2.5086	6	6.5673	0.0308
TRS	6.2666	2	1.7334	6	10.8458	0.0102
TRS price	5.4278	2	2.5722	6	6.3304	0.0332
Rainfall	5.9956	2	2.0044	6	8.9736	0.0157
Piracicaba						
TCH	5.0103	2	2.9897	6	5.0276	0.0522
TRS	6.3038	2	1.6962	6	11.1493	0.0095
TRS price	4.7067	2	3.2933	6	4.2875	0.0698
Rainfall	6.1022	2	1.8978	6	9.6464	0.0133

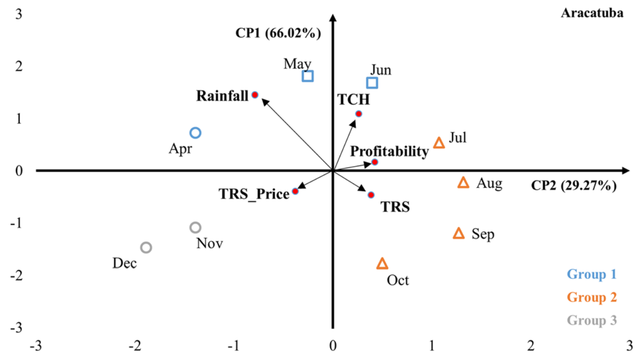

Fig. 4 Dispersion (biplot graph) of sugarcane profitability (R$ ha^{-1}) over the harvest months in Araçatuba region

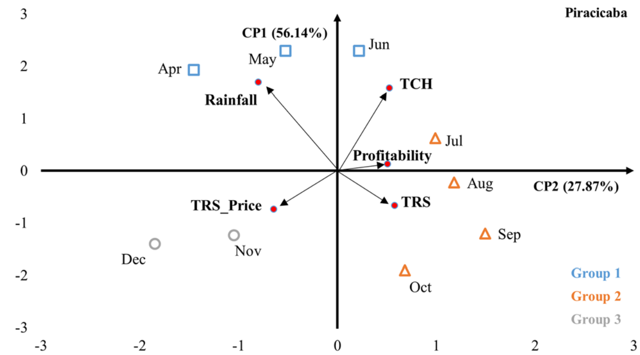

Fig. 7 Dispersion (biplot graph) of sugarcane profitability (R$ ha^{-1}) over the harvest months in Piracicaba region

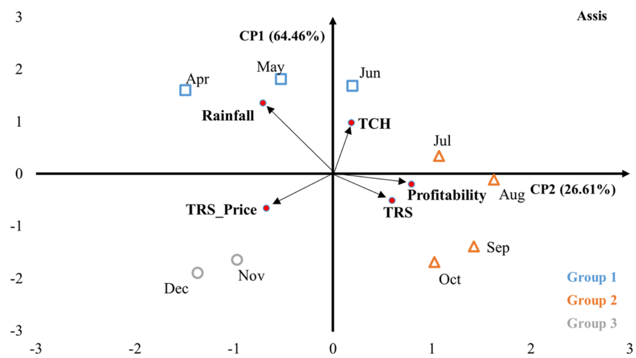

Fig. 5 Dispersion (biplot graph) of sugarcane profitability (R$ ha^{-1}) over the harvest months in Assis region

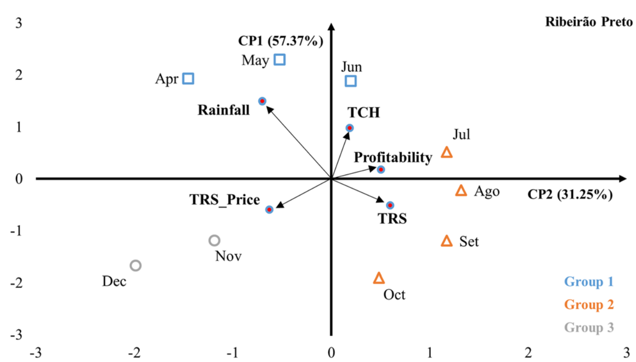

Fig. 6 Dispersion (biplot graph) of sugarcane profitability (R$ ha^{-1}) over the harvest months in Ribeirão Preto region

84.01% [56.14% (PC1) + 27.87% (PC2)] of the total information of original variables.

The graphical representation and the correlation of the variables in the PCs enabled the characterization of variables that defined the formation of groups in each region. In all cases, the ranking of months according to the first two PCs confirmed the classification into three groups based on the hierarchical and non-hierarchical cluster analyses (Figs. 2 and 3). Regardless of region, the TRS price enabled the discrimination of months in Group III (lower profitability) on the basis of their location to the left of PC1 (negative correlations, ranging from − 0.83 to − 0.89). TRS (0.87–0.99) enabled the discrimination of Group II (greater profitability), and rainfall enabled the discrimination of Group I. For the second PCs, only TCH, which exhibited positive correlations (0.81–0.90), enabled the discrimination of months that were in the upper part of the biplot graph (Group II), thus indicating the months in which these hits had a greater TCH than the other groups (Fig. 4). This finding supports the conclusion that Group II (months of greater economic profitability) is characterized by sugarcane with higher sugar content (i.e., higher TRS values) and intermediate TCH values, whereas Group III is characterized by higher TRS prices and Group I by early harvest, which is strongly influenced by greater rainfall (0.72–0.98) occurring in previous months, low sugar content, and higher TCH values. This information is highly important because it helps characterize the variables that most influence profitability at each moment of harvest, the variations depending on the region, and the related impacts when considering changes in the harvest seasons. Because the worst results (e.g., financial or agronomic) are obtained at the beginning and end of the harvest, the expansion of the crushing capacity over these months can be considered one of the factors that have hindered the sector's performance in recent harvest seasons (Cardozo 2017).

Regions Economic Returns

All multivariate statistical methods resulted in similar clustering patterns, in which the sugarcane harvest months in four of the main sugarcane regions were divided into three groups of economic profitability. Group II (months of high economic return) included July, August, September, and October. The economic return was highest in these

Table 7 Correlation between each major component and variables related to economic return in various regions of the state of Sao Paulo, southeastern Brazil

Variables	Araçatuba		Assis		Ribeirão Preto		Piracicaba	
	PC1	PC2	PC1	PC2	PC1	PC2	PC1	PC2
Eigenvalues	3.30	1.46	3.22	1.33	2.86	1.56	2.81	1.39
Explained variation (%)	66.02	29.27	64.46	26.61	57.37	31.25	56.14	27.87
	Correlations							
TCH	0.587	**0.805**	0.409	**0.907**	0.523	**0.836**	0.397	**0.822**
TRS	**0.874**	− 0.465	**0.877**	− 0.471	**0.990**	− 0.090	**0.981**	− 0.128
TRS price	**− 0.874**	− 0.393	**− 0.835**	− 0.421	**− 0.845**	− 0.456	**− 0.894**	− 0.319
Rainfall	**− 0.717**	0.664	**− 0.911**	0.330	**− 0.977**	0.135	**− 0.976**	0.133

months, which was mainly influenced by higher TRS. According to Cardozo and Sentelhas (2013), sugarcane maturation is highly dependent on weather conditions, and water availability is the main factor of this process under Brazilian conditions. Robertson and Donaldson (1998) reported that high sucrose concentration was obtained under moderate water deficits, with a 15% increase compared with 8% under wet conditions. According to Cardozo et al. (2014), in most sugarcane-producing regions of Brazil, a water deficit (related to lower rainfall index) usually begins in May and reaches its maximum extent in September. Thus, water deficit is directly related to sugarcane sucrose concentration because the latter also usually reaches a maximum between August and October. Inman-Bamber (2004) demonstrated that the accumulation of stalk biomass is affected by water deficits greater than 120 mm, whereas sucrose accumulation is affected when the water deficit reaches greater than 145 mm. Scarpari and Beauclair (2004) reported that water deficits greater than 130 mm in the months preceding the harvest positively affect sucrose accumulation. However, the optimal amount of water deficit for this process is not well defined because sucrose accumulation also depends on other variables, such as evapotranspiration rate, soil water retention capacity, and harvesting phase (Robertson et al. 1999b; White et al. 2019; Scarpari and Beauclair 2009).

Group I (months of intermediate economic return) included April, May, and June. In these months, TCH had the strongest influence on economic returns, which was expected because productivity is directly related to climatic variables, including rainfall (water availability), air temperature, and solar radiation (Keating et al. 1999). In all regions evaluated, such requirements are fully met during the spring and summer months (November to March). During this period, sugarcane tissues are highly hydrated, and the contents of reducing sugars (glucose and fructose) are significantly higher than those of sucrose (Cardozo et al. 2014). The moisture and reducing sugars in stalks are used to grow and expand plant tissues (Inman-Bamber 2004; Robertson et al. 1999a), thus resulting in higher biomass values during the harvest.

Group III (months of low economic return) included November and December—periods mainly influenced by the TRS price. With the end of the harvest season and the expected decreases in the supply of sugar and alcohol, the TRS price increased again and became the main variable influencing economic returns. At that time, TCH values were low because of the loss of moisture in the stalks induced by the water deficit (Cardozo and Sentelhas 2013). Additionally, TRS values begin to decrease with the outset of rainfall, increases in mineral and vegetable impurities (related to the harvesting procedure), and the inversion of sucrose (Cardozo et al. 2014).

Although the economic returns were very similar during the harvest months, some differences existed across the assessed regions. In Araçatuba, for instance, TRS and rainfall were the critical variables in the formation of groups ($p < 0.05$), suggesting that these variables are the most important factors in the creation of the region's economic returns (Table 4). The concentration of rainfall in the summer months, which is typical of this region, followed by intense water deficit in the autumn and winter, promoted the intense accumulation of sucrose during midharvest (June to August). However, under extreme water deficit conditions, sugar recovering tends to decrease because of excessive loss of moisture in the sugarcane and the phenomena of "pithing" (Silva and Caputo 2012). Such phenomena prejudice the sucrose extraction process because it makes the "catch" of the mill difficult and requires extra water during the soaking process (e.g., "sponge effect"). The addition of extra water in the industrial process is undesirable because it reduces industrial efficiency by requiring a greater expenditure of energy during broth concentration (Cardozo and Sentelhas 2013).

In Assis, TCH, rainfall, and TRS price were the variables that differentiated the groups ($p < 0.05$). In this case, TRS was not a significant factor for indicating months of better economic returns in the region. This non-significance might occur because the sugarcane ripening in this region is traditionally regarded as problematic because of the low water deficits accumulated during the crop season. Because rainfall is better distributed throughout the year, this region exhibits a smaller variation in TRS, whereas TCH varies widely. In Ribeirão Preto, all variables were significant ($p < 0.05$) for the defining groups, suggesting that this region's rainfall is distributed such that a better balance exists between cane yield and sugar content (TRS) according to the months considered. At the beginning of a harvest, when TRS values were low, TCH values were high, discriminating this group from the other months (April and May). At midharvest, when TCH values were lower (because of the dehydration of the stems), high TRS values allowed group discrimination. The TRS price played an important role at the end of a harvest, when TCH and TRS values were lower because of the degradation of raw materials from either the onset of rains (inversion of sucrose and decreases in TRS) or prolongation of the dry season.

In the Piracicaba region, groups of months of similar economic returns were discriminated based using TRS and rainfall ($p < 0.05$). Thus, TRS can be inferred to have the strongest influence on the variation in sugarcane economic returns depending on the time of year. Local climate conditions, with lower water deficits than the regions of Ribeirão Preto and Araçatuba, but more intense than in Assis, put the region in the "intermediate" position regarding water availability. Thus, the change in TRS, although not as intense as those observed in Araçatuba and Ribeirão Preto, was important for its economic return and more significant than that of Assis.

Conclusion

The economic profitability of sugarcane crop varied widely depending on the region and time of year analyzed. This variation is related to how the crop interacts with the climate throughout the year. The multivariate analysis described herein enabled patterns of similarity between economic profitability obtained in each month of harvest across the assessed regions, where the midharvest months (generally from July to September, Group II) presented the highest values. This information enables decision makers to determine why certain regions have better returns over longer periods during the year, thus implying differentiated harvest windows. Such information is especially important after the period of rapid expansion of sugarcane plantations in the country, which was marked by many problems with characterizing regional productive potential and more appropriate ways to manage them. This discussion is especially important when Brazil seeks to retake the importance of the sugarcane sector as a producer of cleaner energy and ethanol. However, the adequacy of the production strategy (particularly the harvesting periods) is essential for optimizing the project's results, including more efficient exploitation of available resources and the economic return of sugarcane production, all of which are crucial factors for greater sustainability.

The different environmental characteristics of each region require an adaptation of the production system depending on the harvest period in which the results presented a significant difference. For instance, the extension of the harvest season without any compensatory alteration in the crop production system has undermined the sugarcane sector. Before the large expansion in sugarcane plantations prior to 2008, the harvest used to be carried out during months of greater profitability (Group II). Despite growing industrial performance, the increased harvest in the months for Groups I and III led to impacts on sugarcane yield, raw material quality, and consequently in the economic returns to the agricultural sector. These aspects have not been considered in the expansion of the sugarcane sector, in which a large number of mills are currently unable to remain economically viable.

Acknowledgments The authors would like to thank the Continued Educational Program in Economy and Agro-business Management (Programa de Educação Continuada em Economia e Gestão de Empresas—PECEGE) for granting a scholarship for the specialization in Management and Investment in the Sugar and Alcohol Industry.

Author Contributions NPC developed research concepts with inputs from ROB, ARP, and CACC, who supported the study design. ARP supported data analysis. All authors contributed to manuscript writing and reviewed the manuscript.

Compliance with Ethical Standards

Conflict of interest The authors declare that they have no conflict of interest.

Ethical Statement None.

Informed Consent None.

References

Ateş, C., Ö. Kaymaz, H.E. Kale, and M.A. Tekindal. 2019. Comparison of test statistics of nonnormal and unbalanced samples for multivariate analysis of variance in terms of Type-I error rates. *Computational and Mathematical Methods in Medicine* 2019: 2173638.

Bordonal, R.O., J.L.N. Carvalho, R. Lal, E.B. Figueiredo, B.G. Oliveira, and N. La Scala. 2018. Sustainability of sugarcane production in Brazil: A review. *Agronomy for Sustainable Development* 38(2): 13.

Bordonal, R.O., R. Lal, D.A. Aguiar, E.B. Figueiredo, L.I. Perillo, M. Adami, B.F.T. Rudorff, and N. La Scala. 2015. Greenhouse gas balance from cultivation and direct land use change of recently

established sugarcane (*Saccharum officinarum*) plantation in south-central Brazil. *Renewable and Sustainable Energy Reviews* 52: 547–556.
CANASAT-INPE. 2014. Sugarcane Crop Monitoring in Brazil by Earth Observing Satellite Images. National Institute for Space Research (INPE). http://www.dsr.inpe.br/laf/canasat/en/. Accessed 22 Sept 2019.
Cardozo, N.P., R.O. Bordonal, and N. La Scala. 2016. Greenhouse gas emission estimate in sugarcane irrigation in Brazil: Is it possible to reduce it, and still increase crop yield? *Journal of Cleaner Production* 112: 3988–3997.
Cardozo, N.P. 2017. Flowering of sugarcane: Genotypic and climatic effects, losses and control strategies. Available at: https://repositorio.unesp.br/handle/11449/152444. Accessed 01 May 2020.
Cardozo, N.P., and P.C. Sentelhas. 2013. Climatic effects on sugarcane ripening under the influence of cultivars and crop age. *Scientia Agricola* 70: 449–456.
Cardozo, N.P., P.C. Sentelhas, A.R. Panosso, and A.S. Ferraudo. 2014. Multivariate analysis of the temporal variability of sugarcane ripening in south-eastern Brazil. *Crop and Pasture Science* 65(3): 300–310.
Cardozo, N.P., P.C. Sentelhas, A.R. Panosso, A.L. Palhares, and B.Y. Ide. 2015. Modeling sugarcane ripening as a function of accumulated rainfall in Southern Brazil. *International Journal of Biometeorology* 59(12): 1913–1925.
Chagas, M.F., R.O. Bordonal, O. Cavalett, J.L.N. Carvalho, A. Bonomi, and N. La Scala. 2016. Environmental and economic impacts of different sugarcane production systems in the ethanol biorefinery. *Biofuels Bioproducts & Biorefining-Biofpr* 10(1): 89–106.
CONAB. 2020 Companhia Nacional de Abastecimento. Acompanhamento da safra brasileira de cana-de-açúcar. Quarto Levantamento-Safra 2019/20, 58p. http://www.conab.gov.br. Accessed 01 May 20.
de Castro, B.J.C., and A. Bernardo. 2019. Evaluation of cane sugar production using multivariate statistical methods. *Journal of Engineering and Exact Sciences* 5: 228–237. https://doi.org/10.18540/jcecvl5iss3.
Dias, H.B., and P.C. Sentelhas. 2017. Evaluation of three sugarcane simulation models and their ensemble for yield estimation in commercially managed fields. *Field Crops Research* 213: 174–185.
Fachinelli, N.P., and A.O. Pereira. 2015. Impacts of sugarcane ethanol production in the Paranaiba basin water resources. *Biomass and Bioenergy* 83: 8–16.
Gasperik, K.W. 2010. *MANOVA: Type I error rate analysis*. San Luis Obispo, CA: The Faculty of the Statistics Department California Polytechnic State University, A Senior Project.
Goldemberg, J. 2007. Ethanol for a sustainable energy future. *Science* 315(5813): 808–810.
Gomes, L., S.J.C. Simões, E.L. Dalla Nora, E.R. de Sousa-Neto, M.C. Forti, and J.P.H.B. Ometto. 2019. Agricultural expansion in the Brazilian Cerrado: Increased soil and nutrient losses and decreased agricultural productivity. *Land* 8: 12.
Górecki, T., and L. Smaga. 2017. Multivariate analysis of variance for functional data. *Journal of Applied Statistics* 44(12): 2172–2189.
Hair, J.F., R.E. Anderson, R.L. Tatham, and W. Black. 2005. *Multivariate data analysis*. Richmond: Prentice Hall.
Hartigan, J.A. 1975. *Clustering algorithms*. New York: Wiley.
Inman-Bamber, N.G. 2004. Sugarcane water stress criteria for irrigation and drying off. *Field Crops Research* 89(1): 107–122.
Jaiswal, D., A.P. Souza, S. Larsen, D.S. LeBauer, F.E. Miguez, G. Sparovek, G. Bollero, M.S. Buckeridge, and S.P. Long. 2017. Brazilian sugarcane ethanol as an expandable green alternative to crude oil use. *Nature Climate Change* 7: 788–792.
Kaiser, H.F. 1958. The varimax criterion for analytic rotation in factor-analysis. *Psychometrika* 23(3): 187–200.
Keating, B.A., M.J. Robertson, R.C. Muchow, and N.I. Huth. 1999. Modelling sugarcane production systems I. Development and performance of the sugarcane module. *Field Crops Research* 61(3): 253–271.
Lal, R. 2013. Food security in a changing climate. *Ecohydrology & Hydrobiology* 13(1): 8–21.
Lawes, R.A., and R.J. Lawn. 2005. Applications of industry information in sugarcane production systems. *Field Crops Research* 92(2–3): 353–363.
Marin, F.R., G.B. Martha, K.G. Cassman, and P. Grassini. 2016. Prospects for increasing sugarcane and bioethanol production on existing crop area in Brazil. *BioScience* 66(4): 307–316.
Ramburan, S., M. Zhou, and M. Labuschagne. 2011. Interpretation of genotype × environment interactions of sugarcane: Identifying significant environmental factors. *Field Crops Research* 124: 392–399. https://doi.org/10.1016/j.fcr.2011.07.008.
Renewable Fuels Association (RFA). 2019. Markets and Statistics. Available at: https://ethanolrfa.org/statistics/.
Robertson, M.J., and R.A. Donaldson. 1998. Changes in the components of cane and sucrose yield in response to drying-off of sugarcane before harvest. *Field Crops Research* 55(3): 201–208.
Robertson, M.J., N.G. Inman-Bamber, R.C. Muchow, and A.W. Wood. 1999a. Physiology and productivity of sugarcane with early and mid-season water deficit. *Field Crops Research* 64(3): 211–227.
Robertson, M.J., R.C. Muchow, R.A. Donaldson, N.G. Inman-Bamber, and A.W. Wood. 1999b. Estimating the risk associated with drying-off strategies for irrigated sugarcane before harvest. *Australian Journal of Agricultural Research* 50(1): 65–77.
Rolim, G.S., M.B.P. Camargo, D.G. Lania, and J.F.L. Moraes. 2007. Classificação climática de Köppen e de Thornthwaite e sua aplicabilidade na determinação de zonas agroclimáticas para o estado de São Paulo. *Bragantia* 66(4): 711–720.
Saed-Moucheshi, A., E. Fasihfar, H. Hasheminasab, A. Rahmani, and A. Ahmadi. 2013. A review on applied multivariate statistical techniques in agriculture and plant science. *International Journal of Agronomy and Plant Production* 4: 127–141.
Santchurn, D., K. Ramdoyal, M.G.H. Badaloo, and M. Labuschagne. 2012. From sugar industry to cane industry: Investigations on multivariate data analysis techniques in the identification of different high biomass sugarcane varieties. *Euphytica* 185: 543–558. https://doi.org/10.1007/s10681-012-0682-4.
Scarpare, F.V., T.A.D. Hernandes, S.T. Ruiz-Correa, M.C.A. Picoli, B.R. Scanlon, M.F. Chagas, D.G. Duft, and T.D. Cardoso. 2016. Sugarcane land use and water resources assessment in the expansion area in Brazil. *Journal of Cleaner Production* 133: 1318–1327.
Scarpare, F.V., M.R.L.V. Leal, and R.L. Victoria. 2015. The challenges of sugarcane ethanol in Brazil: Past, present and future. In *The challenges of sugarcane ethanol in Brazil: Past, present and future*, ed. J.F. Dallemand, J.A. Hilbert, and F. Monforti, 91–104. Luxembourg: Publications Office of the European Union.
Scarpari, M.S., and E.G.F. Beauclair. 2004. Sugarcane maturity estimation through edaphic-climatic parameters. *Scientia Agricola* 61: 486–491.
Scarpari, M.S., and E.G.F. Beauclair. 2009. Physiological model to estimate the maturity of sugarcane. *Scientia Agricola* 66: 622–628.
Silva, M.A., and M.M. Caputo. 2012. Ripening and the use of ripeners for better sugarcane management. In *Crop management—cases and tools for higher yield and sustainability*, ed. F.R. Marin, 2–24. Rijeka: InTech.

Sneath, P.H., and R.R. Sokal. 1973. *Numerical taxonomy: The principles and practice of numerical classification*. San Francisco: W.H. Freeman.

Souza, G.M., M.V.R. Ballester, C.H. de Brito Cruz, H. Chum, B. Dale, V.H. Dale, E.C.M. Fernandes, et al. 2017. The role of bioenergy in a climate-changing world. *Environmental Development* 23: 57–64.

Stahle, L., and S. Wold. 1990. Multivariate analysis of variance (MANOVA). *Chemometrics and Intelligent Laboratory Systems* 9(2): 127–141.

STATSOFT, Inc. 2010. STATISTICA (data analysis software system), version 7.

Tammisola, J. 2010. Towards much more efficient biofuel crops—can sugarcane pave the way? *GM Crops* 1(4): 181–198.

UNICA. 2020. União da Indústria de Cana-de-açúcar. UnicaData, Área cultivada com cana-de-açúcar, Mapeamento de área Centro-sul. http://www.unicadata.com.br. Accessed 10 May 2020.

Vianna, M.D., and P.C. Sentelhas. 2016. Performance of DSSAT CSM-CANEGRO under operational conditions and its use in determining the 'Saving Irrigation' impact on Sugarcane Crop. *Sugar Tech* 18(1): 75–86.

Walter, A., M.V. Galdos, F.V. Scarpare, M.R.L.V. Leal, J.E.A. Seabra, M.P. da Cunha, M.C.A. Picoli, and C.O.F. Oliveira. 2014. Brazilian sugarcane ethanol: Developments so far and challenges for the future. *Wiley Interdisciplinary Reviews: Energy and Environment* 3(1): 70–92.

White, P.M., C.L. Webbe, R.P. Viator, and G. Alta. 2019. sugarcane biomass, dry matter, and sucrose availability and variability when grown on a bioenergy feedstock production cycle. *Bioenergy Research* 12(1): 55–67.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.