CSIRO PUBLISHING
Crop & Pasture Science, 2014, **65**, 300–310
http://dx.doi.org/10.1071/CP13160

Multivariate analysis of the temporal variability of sugarcane ripening in south-eastern Brazil

Nilceu P. Cardozo[A], *Paulo C. Sentelhas*[B,E], *Alan R. Panosso*[C], *and Antonio S. Ferraudo*[D]

[A]Sugarcane Research Center, CTC, Piracicaba, SP, Brazil.
[B]Department of Biosystems Engineering, ESALQ, University of São Paulo, Piracicaba, SP, Brazil.
[C]Department of Mathematics, FEIS-UNESP, Ilha Solteira, SP, Brazil.
[D]Department of Exact Sciences, FCAV-UNESP, Jaboticabal, SP, Brazil.
[E]Corresponding author. Email: pcsentel.esalq@usp.br

Abstract. Sugarcane ripening is a process controlled by cultivar characteristics and the interaction of genotypes with local climate. The objective of this study was to characterise the temporal variation of sugarcane ripening by assessing the multivariate structure contained in sugarcane quality data, and by correlating the results with local climatic conditions. Eight sugarcane cultivars were evaluated from March to October in Piracicaba, São Paulo State, Brazil. Characteristics related to the quality of raw sugarcane juice were submitted to statistical analysis by ANOVA, hierarchical and non-hierarchical (k-means) clustering methods, and principal components, in order to classify the cultivars into groups for each month of sampling. The ANOVA showed a clear difference ($P < 0.001$) among harvesting months for all sugarcane quality variables, which was reinforced by the cluster analysis for the whole dataset that selected groups according to the month of harvest. By analysing the quality variables by months, patterns of similarity among sugarcane cultivars were identified, which allowed three ripening groups to be established: early, middle and late. As the harvesting season progressed, the variations within each group, as well as among groups, were explained mainly by local soil-water availability conditions. The early ripening cultivars showed polarisable sugar (Pol) values >13% in early May, whereas these values were reached by the middle cultivars in July, and the late ones in August–September. However, the differences among groups tended to decrease through the harvest season, as expressed by the Euclidean distance, which decreased from 5.62 in March to 1.82 in September, when the water deficit reached the maximum accumulated value, totalling >130 mm. The non-hierarchical analyses (k-means) and principal components methods agreed, resulting in the identification of the same three main cultivar groups. The approach proposed for cultivar classification in this study is more complete than the usual analysis of Pol variation over time, since it allowed all of the variability contained in the sugarcane quality dataset to be analysed in an integrated way, providing a better understanding of the differences observed in the ripening of different cultivars.

Received 8 May 2013, accepted 20 February 2014, published online 14 April 2014

Introduction

Brazil is the largest sugarcane producer in the world, with a production of 697.8 million tonnes on 8.8 million hectares (Informa Economics FNP 2012). Production is concentrated mainly in the south-eastern region of the country, where the State of São Paulo is responsible for 60% of the production, over an area of 4.87 million hectares (UNICA 2012; Rudorff *et al.* 2010). This makes sugarcane the most important bioenergy crop in Brazil. Sugarcane has expanded across Brazil into regions with climates and soils that differ from the regions where it was traditionally grown, with consequential effects on its growth and ripening patterns (Marin *et al.* 2011).

Clements (1962) described sugarcane ripening as a physiological process that occurs between the phases of growth and plant death. Generally, it is said that sugarcane ripening is the process of sucrose accumulation in the stalks (Legendre 1975), which occurs from the basal internodes to the apex, until they reach a common value (Alexander 1973). The ripening point is determined primarily by the levels of sucrose, reducing sugars (RS) and stalk water content during the season. According to the Brazilian Sugarcane, Sugar and Ethanol Producers Council (CONSECANA) standards, millable cane, i.e. stalks with the technological properties and economic value suitable for industrial processing, need sucrose (polarisable sugars, Pol) levels of at least 12.257%. Early cultivars are those with Pol levels greater than this value at the beginning of the harvesting season, while middle and late cultivars have Pol levels above this threshold from the middle to the end of the season (Lavanholi 2008). Within a region, sugarcane ripening depends on a wide range of factors, some of which are cultivar, crop age, climate, soil type, month of harvest and rainfall (Lawes *et al.* 2002; Ferraro *et al.* 2009).

According to Alexander (1973), sugarcane ripening can be determined by technological parameters such as soluble solids (Brix), Pol, purity and RS. Knowledge of ripening patterns of each cultivar is essential for planning sugarcane harvesting, as

well as for breeding programs, which usually study it in a traditional univariate way, without considering the interactions between all components of the raw material when classifying cultivars as early, medium or late ripening (Cardozo 2012).

Considering the multiple genetic and environmental factors that affect sugarcane ripening, the best way to evaluate its spatial and temporal variability for a group of cultivars is by the use of multivariate statistical analysis. According to Lawes and Lawn (2005), even though such techniques are essentially empirical in nature, they can be applied to define and solve problems related to the sugarcane industry. Multivariate statistical techniques enable the analysis of a phenomenon considering all the factors that affect it in an integrated way. Therefore, it is not enough to know these factors in isolation, but in their entirety, since there are interactions among them (Hair *et al.* 2005). Multivariate statistical techniques such as cluster, canonical and principal components analyses have been used to study genetic diversity of plant species and to classify germplasm collections (Wagih *et al.* 2004; Jackson *et al.* 1995; Rahman *et al.* 2011; Ferreira *et al.* 2011; Maji and Shaibu 2012). The most commonly used technique in these studies is cluster analysis, which groups samples of individuals, such as cultivars, varieties or clones, into clusters, seeking the maximum internal homogeneity among them (Johnson and Wichern 2001). Principal components analysis has also been widely used for genotype discrimination (Beiragi *et al.* 2012). The categorisation of genotypes into groups with similar morphological and agronomic characteristics can be used for designing operational strategies to improve crop performance in terms of yield and quality (Ariyo 1993; Wagih *et al.* 2004). In Australia, Lawes *et al*. (2004) applied multivariate techniques to identify the relationships between farms that performed similarly in terms of sugarcane yield and sugar content.

Few studies on genetic similarities among cultivars have been carried out for sugarcane (Viana *et al.* 1991; Wagih *et al.* 2004), despite their potential use for parental selection (Rao *et al.* 1985) and for evaluating genotype × environment interactions (Jackson and Hogarth 1992; Jackson *et al.* 1995).

Considering the evident differences among sugarcane cultivars in terms of the ripening process, and the influence of climatic conditions on this biological process, the fact that multivariate analysis can integrate a suite of variables should provide useful information for optimising harvesting logistics and sugar production. The objectives of the present study were therefore to characterise the temporal variability of sugarcane ripening in different cultivars using multivariate analyses to identify relationships with local climatic conditions, and to group cultivars with similar ripening patterns in order to plan harvesting to optimise juice quality.

Materials and methods

Location and experimental design

The experiment was carried out in a commercial area of the Costa Pinto Sugar Mill, Raízen Co., Piracicaba (22°36′45″S, 47°37′47″W; altitude 597 m a.s.l.), São Paulo State, Brazil, in the period March 2002 (first planting)–October 2003 (final sampling). The sugarcane was planted every month from March to October 2002 and harvested from March to October 2003, every 30 days on the 15th day of the month. The experimental design was randomised blocks with eight cultivars and three repetitions. Each plot consisted of four rows of 15 m, with 1.4 m between rows. Eight sugarcane commercial cultivars, with different maturity patterns, were evaluated: SP91-1049, SP86-42, SP90-3414, SP86-155, SP87-365, SP83-2847, RB928064 and RB867515, all of which are still used and represent ~40% of the cultivated sugarcane area in the state of São Paulo.

Sugarcane samples and weather data

Sugarcane samples were collected monthly during 8 months, from March to October 2003, when the crop age was 12 months. The samples consisted of 10 stalks per cultivar, with dead and green leaves removed. The samples of each cultivar were collected randomly from the central lines of each experimental plot with three repetitions, totalling 30 stalks per cultivar. These were submitted individually for analyses to determine the levels of RS, Brix, purity (Pur), Pol, total recoverable sugar (TRS), fibre (F) and moisture (M), according to the standards of the Brazilian Sugarcane Growers Council (CONSECANA 2006).

Weather data were obtained from the closest weather station in Piracicaba (22°42′30″S, 47°30′00″W; altitude 546 m a.s.l.). The climatic conditions of the experimental period were typical for the region, as characterised by monthly weather variables (Table 1) and the regional soil water availability. Soil moisture was determined by a climatological water balance (Thornthwaite and Mather 1955) on a 10-day time scale using a soil water-holding capacity of 125 mm for a depth of 1 m, which is commonly observed in the soils of this region (Fig. 1). To run the water balance for the experimental period, the reference evapotranspiration was estimated by the Penman–Monteith equation (Allen *et al.* 1998), which uses inputs of net radiation, maximum and minimum air temperature, and relative humidity, and wind speed at a height of 2 m.

Statistical analyses

To evaluate the quality data as well as the interactions among them, analysis of variance (ANOVA) was performed for each of the sugarcane quality variables, considering the eight cultivars and the months of sampling, from March to October 2003. Differences between the overall averages among months of harvesting and among cultivars were evaluated by Tukey's test at $P = 0.05$.

To evaluate the statistical hypotheses of homoscedasticity and normality of variance, Levene ($P > 0.05$) and Shapiro–Wilk ($P > 0.05$) tests were conducted on all the quality variables, without data transformation, to ensure that all data came from a normally distributed population and to detect the presence of outliers. The multivariate structure contained in the sugarcane quality data was studied using three different multivariate statistical methods, in order to classify the cultivars into groups according to their ripening patterns (early, middle or late): (*i*) cluster analysis by the hierarchical method; (*ii*) cluster analyses by the non-hierarchical (k-means) method; and (*iii*) principal components (Sneath and Sokal 1973). These analyses were performed on the entire dataset and on each month independently.

Table 1. Weather conditions in Piracicaba, SP, Brazil, from January 2002 to December 2003
SRad, Incoming solar radiation; Tmax, maximum air temperature; Tav, average air temperature; Tmin, minimum air temperature; RH, relative humidity; U_{2m}. wind speed at 2 m

Year	Month	Rainfall (mm)	SRad ($MJ\,m^{-2}\,day^{-1}$)	Tmax	Tav (°C)	Tmin	RH (%)	U_{2m} ($m\,s^{-1}$)
2002	Jan.	275.1	17.9	29.2	23.1	19.2	91.0	1.0
	Feb.	167.8	17.7	28.5	22.7	18.8	90.5	1.1
	Mar.	259.5	20.3	31.5	24.6	19.5	85.9	0.8
	Apr.	26.7	18.5	31.2	23.7	17.7	82.5	0.9
	May	104.7	13.0	26.3	19.6	14.4	88.9	1.2
	June	0.3	13.5	27.4	19.3	12.6	82.9	1.2
	July	21.3	12.0	24.7	17.1	10.3	82.0	1.1
	Aug.	63.0	14.4	28.5	20.9	14.4	77.2	1.2
	Sept.	42.8	16.0	26.7	19.9	13.6	77.5	1.5
	Oct.	48.9	20.2	33.0	25.0	18.5	69.7	1.7
	Nov.	166.1	19.5	29.9	23.6	18.6	81.3	1.8
	Dec.	169.1	20.7	30.7	24.3	19.8	85.7	1.5
2003	Jan.	297.6	16.8	29.1	23.7	20.1	92.0	1.1
	Feb.	52.5	21.1	31.9	25.3	20.4	84.9	1.0
	Mar.	177.5	18.1	29.4	23.2	18.5	80.6	1.3
	Apr.	55.0	16.0	28.4	21.8	16.5	76.5	1.1
	May	54.2	14.0	25.3	18.1	11.8	75.0	1.1
	June	8.9	12.8	27.5	19.2	12.4	74.1	0.9
	July	16.4	12.7	26.4	18.2	10.9	65.8	1.3
	Aug.	17.6	15.1	25.8	17.7	10.6	65.9	1.4
	Sept.	12.0	17.1	28.9	21.0	13.7	63.1	1.7
	Oct.	88.6	18.2	29.6	22.3	16.2	66.4	1.8
	Nov.	138.3	19.4	29.2	22.6	17.5	72.5	1.8
	Dec.	133.1	20.6	30.2	24.0	19.3	78.2	1.5

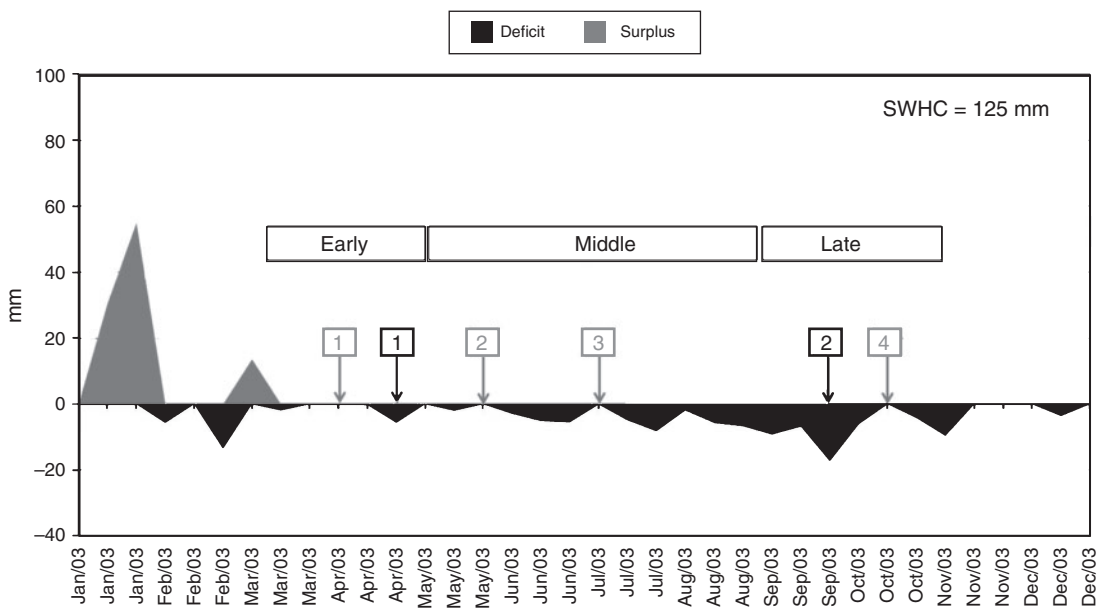

Fig. 1. Climatological water balance (water surplus and water deficit), over 10-day time periods from January to December 2003, in Piracicaba, SP, Brazil. Black arrows indicate maximum early and late season water deficit, grey arrows indicate rainfall events, and the numbers indicate the order of occurrence of such events. SWHC, Soil water-holding capacity.

The analyses were applied to the standardised values of RS, Brix, Pur, Pol, TRS, F and M, resulting in relative values with null means and unit variance (Hartingan 1975). The standardisation was done by removing the mean and dividing by the standard deviation. Non-standardisation of the variables could lead to inconsistencies in the results, since most of the distance measures are quite sensitive to different scales or magnitudes of the variables.

Clustering analysis was carried out for all standardised sugarcane quality variables by calculating the Euclidean distance among cultivars and by using Ward's algorithm for generating groups of cultivars with similar characteristics. This was done for the entire dataset and for each month independently, from March to October, using the dendrograms to identify the cultivar influence on sugarcane quality throughout the harvesting season and the possible changes in group structures over time.

Principal components analysis (PCA) was also used; this condenses the relevant information contained in p variables (eight in this study) into a smaller set of latent orthogonal variables, called principal components (eigenvectors), which are generated by linear combinations of the original variables from the eigenvalues of the covariance matrix. With the PCA, it is possible to explain in a better way the structure of the variability, which is represented by a set of variables more correlated with the maximum variability. Each pair of principal components generates a biplot, a two-dimensional representation of the original sample space. Only the principal components whose eigenvalues were greater than unity were considered (Kaiser 1958). The coefficients of the linear functions, which define the principal components, were used in the interpretation of their meaning, with the sign and the relative size of the coefficients indicating the weight to be assigned to each variable. Only coefficients with high values were considered in the interpretation, usually those ≥ 0.5. Analysis of the biplot graph was based on the spatial distribution of the quality variables and cultivars showing which variables were closer to each cultivar group.

Although the statistical techniques described above are useful for defining the composition of cultivar groups with similar characteristics, the hierarchical clustering and PCAs were not able to identify those sugarcane quality variables with greater weight in discriminating cultivar groups. Therefore, the k-means clustering analysis was performed to show which variables were more effective in discriminating cultivar groups for ripening throughout the season (Hair *et al.* 2005). This analysis computes the classification of an n-column (sugarcane quality parameters) and m-row array (varieties per month), where n is the number of variables (seven) and m is the number of observations (eight), done monthly through the harvesting season. This method minimises the variance of accessions within each group.

All statistical analyses were processed using the STATISTICA software, version 9.0 (StatSoft Inc., Tulsa, OK).

Results

Analysis of variance

The ANOVA for the entire dataset is presented in Table 2. The overall P-values were very small, suggesting it is unlikely that the differences observed were due to random sampling. In addition, the F-values, varying from 41.6 for F to 156.2 for M, were high enough to show statistically significant differences among the months. Based on that, the hypothesis that there were no differences in the means of the sugarcane quality variables among the 8 months of sampling and among cultivars was rejected.

The overall average values of each of the sugarcane quality variables throughout the months of the harvesting season, and for the eight cultivars assessed, are presented in Tables 3 and 4, respectively. The magnitude of the differences observed for the quality variables among the months of sampling can be easily identified, with a clear tendency for a decrease in RS and M and a tendency for an increase in Brix, Pur, Pol, TRS and F later in the harvesting season (Table 3). Comparison of the averages of quality variables among cultivars (Table 4) did not show any trend, varying substantially for each variable, making it difficult to group the cultivars, and emphasising the requirement for more sophisticated statistical analysis for that purpose.

Clustering analysis by the hierarchical method

The dendrogram obtained from clustering analysis by the hierarchical method for the entire dataset, including the sugarcane cultivars and quality variables for all months of sampling, is shown in Fig. 2. From this analysis, it is clear

Table 2. Analysis of variance ($P < 0.001$) for reducing sugars (RS), purity, total recoverable sugar (TRS), fibre, polarisable sugars (Pol), moisture and soluble solids (Brix) variables of sugarcane cultivars sampled from March to October 2003, in Piracicaba

Variable	*F*	*F-critical*	*P*
RS	51.47	*2.98*	4.14×10^{-20}
Purity	45.93		6.17×10^{-21}
TRS	110.63		1.98×10^{-30}
Fibre	41.60		6.18×10^{-23}
Pol	104.94		7.78×10^{-30}
Moisture	156.21		2.32×10^{-34}
Brix	147.44		1.07×10^{-33}

Table 3. Overall average values of reducing sugars (RS), purity (Pur), total recoverable sugar (TRS), fibre (F), polarisable sugars (Pol), moisture (M) and soluble solids (Brix) for all sugarcane cultivars assessed in each month of sampling, from March to October 2003, in Piracicaba

Within rows, means followed by the same letter are not significantly different by the Tukey test at $P = 0.05$

Var.	Unit	Mar.	Apr.	May	June	July	Aug.	Sept.	Oct.
RS	%	1.15a	0.92b	0.77c	0.69cd	0.67cd	0.62cd	0.61cd	0.53d
Pur	%	68.41d	76.00c	80.17b	82.82b	83.39b	84.65ab	84.98ab	87.52a
TRS	$kg\,t^{-1}$	77.96g	93.22f	110.62e	125.92d	133.35c	145.46b	152.69ab	157.79a
F	%	8.98f	9.99e	10.46d	11.06cd	11.36bc	12.07ab	12.07ab	12.71a
Pol	%	7.32f	9.19e	11.20d	12.80c	13.77c	15.09b	15.90ab	16.53a
M	%	80.17a	77.57b	75.17c	72.98d	71.56e	69.39f	68.52f	67.59f
Brix	%	11.92g	13.82f	16.04e	17.94d	19.26c	21.08b	22.11a	22.57a

Table 4. Overall average values during the harvesting season (from March to October 2003) of reducing sugars (RS), purity, total recoverable sugar (TRS), fibre, moisture, polarisable sugars (Pol) and soluble solids (Brix) for each sugarcane cultivar assessed in Piracicaba

Within columns, means followed by the same letter are not significantly different by the Tukey test at $P = 0.05$

Cultivar	RS (%)	Purity (%)	TRS (kg t^{-1})	Fibre (%)	Moisture (%)	Pol (%)	Brix (%)
RB867515	0.81a	78.99cd	118.43de	10.66b	73.92a	12.01d	17.31e
RB928064	0.75abc	80.79abc	123.22cd	10.91ab	73.19ab	12.59bcd	17.89bcd
SP83-2847	0.81a	78.52d	116.10e	11.52ab	73.21ab	11.76d	17.31e
SP86-155	0.66c	83.64a	129.92b	12.03a	71.43c	13.40b	18.85b
SP86-42	0.78ab	79.80bcd	122.00cd	10.85ab	73.30ab	12.43bcd	17.83bcd
SP87-365	0.78ab	79.94bcd	121.17cd	10.60b	73.71a	12.34cd	17.60bcd
SP90-3414	0.70bc	82.30ab	126.11bc	11.36ab	72.54bc	12.94bc	18.22bc
SP91-1049	0.66c	83.86a	138.54a	10.75b	71.71c	14.33a	19.68a

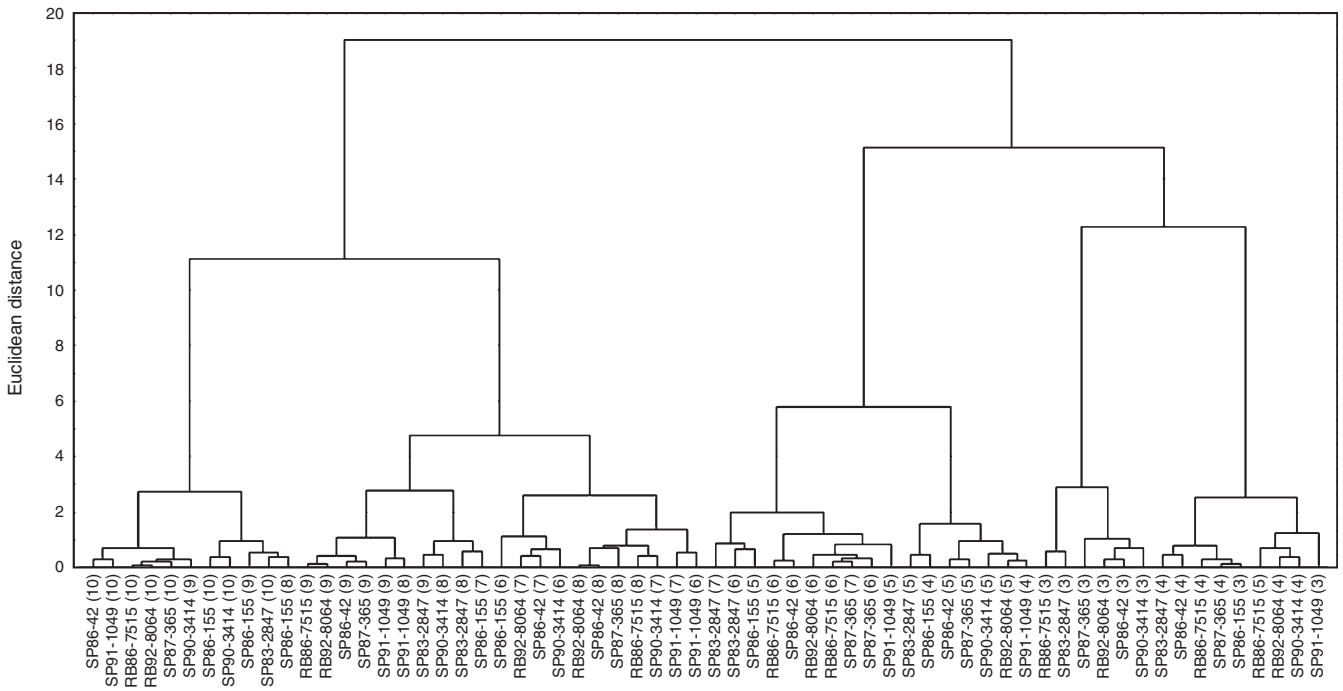

Fig. 2. Dendrogram obtained from clustering analysis using the hierarchical method for the quality variables of eight sugarcane cultivars sampled from March to October 2003, in Piracicaba. Numbers in parentheses represent the month of sampling (from 3 = March to 10 = October).

that the groups were formed by month of harvest (sampling), which shows the strong influence of weather conditions on the sugarcane quality variables. No tendency was observed for cultivar groups, which means that it is not possible to identify cultivars with similar quality characteristics by this approach. Using a Euclidian distance of 4, seven groups were identified, which basically follow the 8 months of sampling. Based on that, the clustering analysis was applied to each month independently in order to identify trends in cultivar grouping according to the sugarcane quality variables (Fig. 3).

The dendrograms obtained from the monthly clustering analysis by the hierarchical method for the sugarcane cultivar quality variables are presented in Fig. 3. In March the maximum difference between groups, expressed by the Euclidean distance, was 5.62, which is much more evident than in the other months, when the distance was <4 and reached a minimum of 1.82 in September.

Cultivars SP91-1049 and SP86-155 were in a group with the highest Euclidean distance compared with the others (5.62); this group's ripening characteristics differ most from the others and it was classified as early ripening. A second group of cultivars (SP90-3414, SP87-365, RB928064, RB867515), with a Euclidian distance of 3.06 separating it from the third group, was classified as middle ripening. Finally, a third group, classified as late ripening, was formed by cvv. SP83-2847 and SP86-42. The differences among cultivars became smaller as the harvesting season progressed and water stress increased (as can be seen in Fig. 1, the summary water balance chart for the experimental period, where positive values represent water surplus and negative values water deficit).

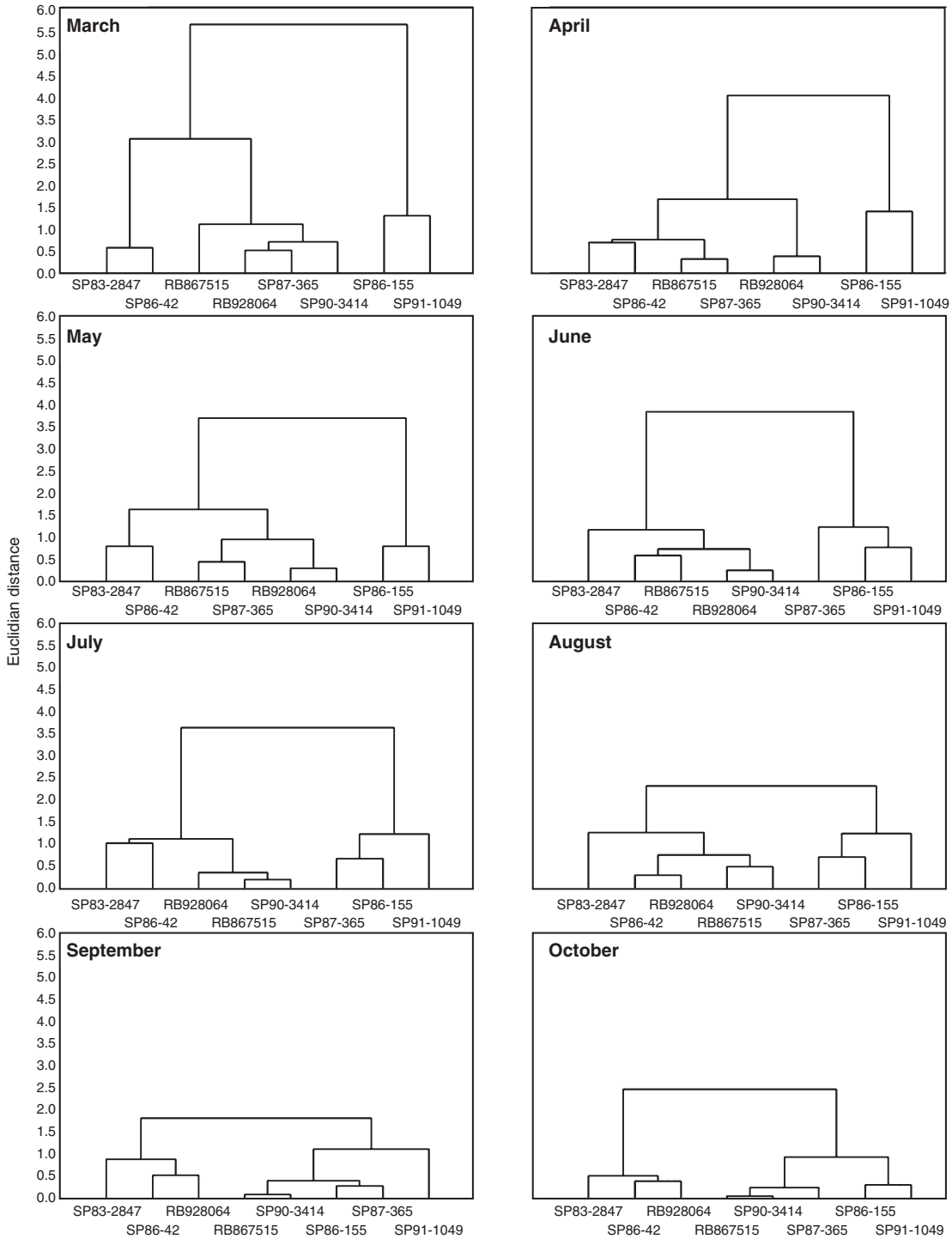

Fig. 3. Dendrograms obtained from clustering analysis using the hierarchical method for the quality variables of eight sugarcane cultivars for each month of sampling in the 2003 harvesting season, in Piracicaba.

Principal components analysis

To confirm the ripening groups (early, middle and late) obtained early in the season (April–May) by cluster analysis using the hierarchical method, the results were submitted to PCA (Fig. 4, Table 5). In this case, PC1 and PC2 showed eigenvalues >1, which explained 78.4% (eigenvalue equal to 8.6) and 15.4% (eigenvalue equal to 1.7) of the original data variability, respectively. Therefore, the total amount of information contained in the original variables that was explained by the two major components was 93.8%.

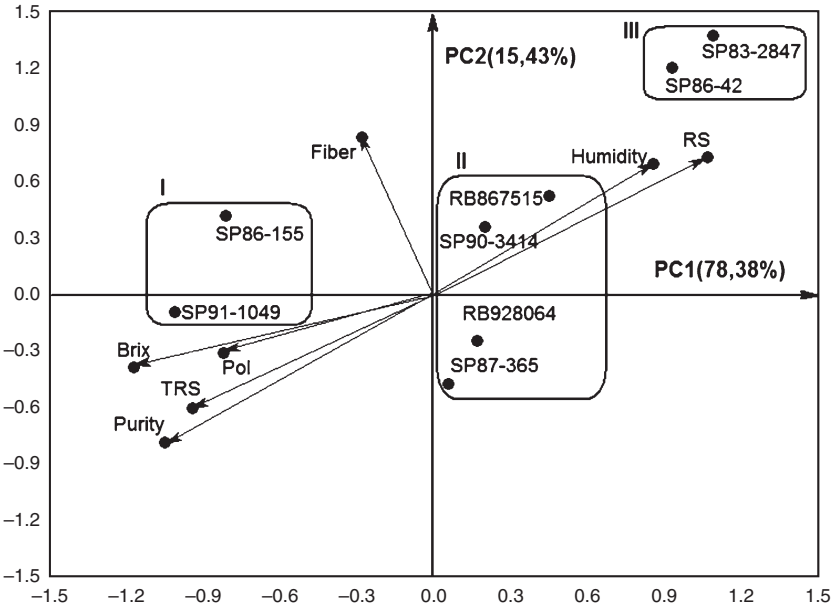

Fig. 4. Dispersion (biplot graphic) of quality variables related to eight sugarcane cultivars studied during the early season (April–May) of 2003, in Piracicaba, and their principal components (PC1 and PC2). Pol, Polarisable sugars; RS, reducing sugars; TRS, total recoverable sugar.

Table 5. Correlation between each principal component (PC1, PC2) and each quality variable of eight sugarcane cultivars harvested in the 2003 season, in Piracicaba

RS, Reducing sugars; TRS, total recoverable sugar; Pol, polarisable sugars; Brix, soluble solids

	PC1	PC2
Eigenvalues	8.6	1.7
Variance explained (%)	78.4	15.4
Correlations		
RS	–0.97	0.07
Purity	0.96	–0.12
TRS	0.99	–0.13
Fibre	0.28	0.96
Pol	0.99	–0.12
Moisture	–0.96	–0.23
Brix	0.98	–0.05

The biplot graphic (Fig. 4), which expresses the correlation between each variable and the principal components, indicates the formation of at least three distinct groups, similar to those observed with the hierarchical clustering analysis in the early season (April–May) (Fig. 3). The discriminatory power of each variable within a principal component is measured by the linear correlation between each variable and its principal component. In PC1, and by order of importance at the beginning of the harvest season, the variables that have the greatest correlation coefficients are TRS and Pol (0.99), followed by Brix (0.98), RS (–0.97), Pur (0.96) and M (–0.96) (Table 5).

Group I (early) contained cvv. SP91-1049 and SP86-155 and stayed to the left of PC1, which related to variables that indicate sugarcane ripening (Brix, Pol, TRS and Pur). Meanwhile, Group III (late) contained cvv. SP83-2847 and SP86-42 to the right of PC1, next to the variables related to vegetative growth (RS and M) rather than variables that indicate ripening. Group II (middle ripening) contained cvv. SP90-3414, SP87-365, RB867515 and RB928064 and remained in an intermediate position, but also to the right of PC1, closer to Group III (late cultivars). Unlike other technological variables, the variable F was the only one to show greater correlation with PC2. However, it was unable to influence the discrimination among groups, mainly because F is a variable strongly influenced by the genetic characteristics of each cultivar rather than by the weather conditions.

Clustering analysis by the non-hierarchical method (k-means)

Three periods of the harvesting season, early (April–May), middle (July–August) and late (October), were selected to study clustering patterns. In the early season, the cultivars were grouped following the same pattern previously obtained by the monthly dendrograms: Group 1 (early ripening), SP91-1049 and SP86-155; Group 2 (middle ripening), SP87-365, RB928064, SP90-3414 and RB867515; and Group 3 (late ripening), SP83-2847 and SP86-42 (Fig. 5).

According to the ANOVA (Table 6), almost all variables were useful for group discrimination at the beginning of the season ($P < 0.05$). The exception was F, which had no significance for discriminating groups, since all cultivars presented similar values at that time of the season.

In the middle of the season, the discrimination of groups was not so evident and there were changes in the cultivars in each group from those previously defined: Group 1 (SP91-1049); Group 2 (SP86-155 and SP87-365); and Group 3 (SP90-3414, RB928064, RB867515, SP86-42 and SP83-2847) (Fig. 6).

For all cultivars in the middle of the season, only F and M were effective variables for discriminating among the groups

(Table 7). Sugar content was not a discriminator. As observed previously, at this time of the harvest season there is no value in classifying the cultivars into early, middle and late ripening, since there are no differences among them in terms of sucrose. The observed differences in F levels, while influenced by seasonal weather, are heavily influenced by cultivar genetic characteristics (Fernandes 2011) and by the soil water supply (Cardozo and Sentelhas 2013).

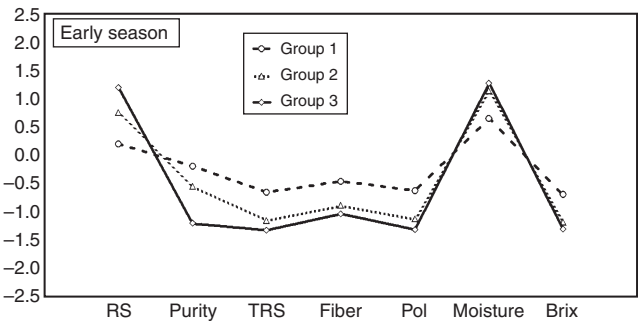

Fig. 5. Standardised averages of sugarcane quality variables during the early season, according to non-hierarchical clustering analysis k-means, for the eight sugarcane cultivars analysed in 2003, in Piracicaba. Group 1: SP91-1049 and SP86-155; Group 2: SP90-3414, RB928064, SP87-365 and RB867515; Group 3: SP83-2847 and SP86-42.

Table 6. Analysis of variance for the quality variables of sugarcane groups formed by non-hierarchical clustering analysis k-means during the early season in Piracicaba

RS, Reducing sugars; TRS, total recoverable sugar; Pol, polarisable sugars; Brix, soluble solids. $^{**}P < 0.05$

Variables	SS among groups	SS inside groups	*F*
RS	1.33	0.13	25.69**
Purity	1.51	0.14	26.49**
TRS	0.60	0.09	16.17**
Fibre	0.45	0.39	2.89
Pol	0.64	0.09	18.13**
Moisture	0.52	0.03	45.50**
Brix	0.50	0.08	16.37**

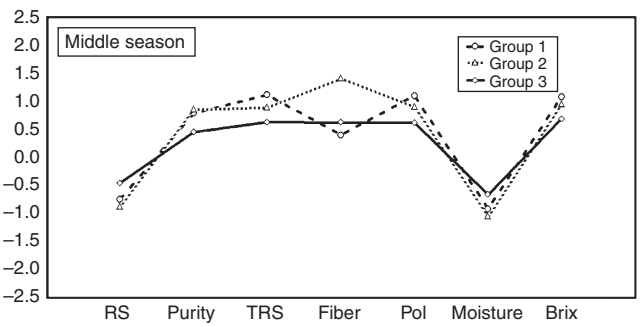

Fig. 6. Standardised averages of sugarcane quality variables during the middle season, according to non-hierarchical analysis clustering k-means, for the eight sugarcane cultivars analysed in 2003, in Piracicaba. Group 1: SP91-1049; Group 2: SP86-155 and SP87-365 Group 3: RB867515, SP90-3414, RB928064, SP83-2847 and SP86-42.

At the end of the season, the composition of the groups changed again (Fig. 7) and the relevant variables for discriminating them changed (Table 8). In this case, the groups comprised: Group 1 (SP91-1049 and SP86-155); Group 2 (SP90-3414, SP87-365 and RB867515); and Group 3 (RB928064, SP83-2847 and SP86-42). The relevant variables for discriminating them were F, M, and RS. As mentioned, the F content is more influenced by cultivar characteristics than by weather conditions. However, although RS and M are partly influenced by cultivar characteristics, they were heavily

Table 7. Analysis of variance for the quality variables of sugarcane groups formed by non-hierarchical clustering analysis k-means during the middle season, in Piracicaba

RS, Reducing sugars; TRS, total recoverable sugar; Pol, polarisable sugars; Brix, soluble solids. $^{**}P < 0.05$

Variables	SS among groups	SS inside groups	*F*
RS	0.08	0.06	3.37
Purity	0.07	0.07	2.62
TRS	0.09	0.07	3.14
Fibre	0.99	0.13	18.60**
Pol	0.09	0.07	3.06
Moisture	0.10	0.04	6.28**
Brix	0.06	0.05	3.20

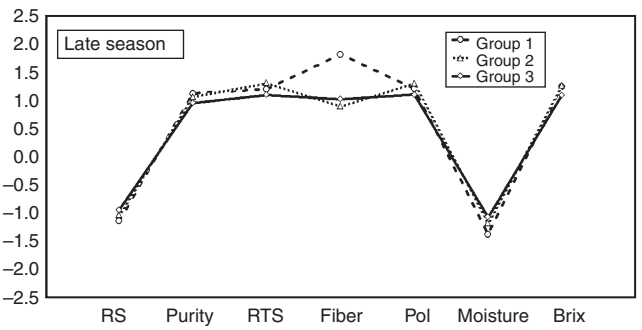

Fig. 7. Standardised averages of sugarcane quality variables during the late season, according to non-hierarchical analysis clustering k-means, for the eight sugarcane cultivars analysed in 2003, in Piracicaba. Group 1: SP86-SP91-1049 and 155; Group 2: SP90-3414, SP87-365 and RB867515; Group 3: RB928064, SP83-2847 and SP86-42.

Table 8. Analysis of variance for the quality variable of sugarcane groups formed by non-hierarchical clustering analysis k-means during the late season, in Piracicaba

RS, Reducing sugars; TRS, total recoverable sugar; Pol, polarisable sugars; Brix, soluble solids. $^{**}P < 0.05$

Variables	SS among groups	SS inside groups	*F*
RS	0.06	0.02	8.69**
Purity	0.04	0.02	5.15
TRS	0.05	0.04	3.57
Fibre	1.35	0.10	33.27**
Pol	0.05	0.04	3.54
Moisture	0.14	0.02	22.84**
Brix	0.05	0.02	5.79

influenced by climate, mainly water deficit and temperature, which affected both growth and development.

Discussion

All of the multivariate statistical methods for monthly analyses resulted in similar clusters when the same harvest period was considered. Group I (early ripening) contained cvv. SP91-1049 and SP86-155; Group II (middle ripening) contained cvv. SP90-3414, SP87-365, RB867515 and RB928064; and Group III (late ripening), cvv. SP83-2847 and SP86-42. According to Horii (2004), sugarcane cultivars in south-eastern Brazil are classified as early ripening when they show Pol values >13% in early May, as middle ripening when they reach this level in July, and as late ripening when they reach their highest sucrose levels in August–September. However, the differences among groups tended to decrease later in the harvest season, as was also observed by Lawes *et al*. (2002). From the April data (Fig. 3), there was a reduction in the Euclidean distance among the three groups, falling from 5.62 for the early group to 1.82 for the late-ripening group.

The reduction in the differences among cultivars as the harvesting season progressed was directly correlated with soil water supply, as observed in the water balance for the experimental period (Fig. 1). The more water that was available for the cane plants, the greater was the difference among cultivars. During March, the water supply was high; however, in April the cumulative water deficit intensified, until its value reached 29 mm at the end of May. Robertson and Donaldson (1998) reported a higher sucrose concentration under moderate water deficit and an increase in dry matter content up to 15%, while the average increase under wet conditions was only ~8%.

From June to July, when similarity among cultivars was expected because of the increasing water deficit, the opposite was observed. The differences among groups, expressed by Euclidean distance values, became greater, especially between early and middle/late cultivars (3.69 in May to 3.84 in June). This behaviour can be explained by the rainfall that occurred between the end of May and the beginning of July, temporarily reducing the water deficit, as shown in the water balance (indicated by the grey arrows 2 and 3 in Fig. 1). These results suggest that the high soil water availability affected the quality of sugarcane juice, increasing RS and decreasing Pol levels. According to Mamet and Galwey (1999), the more pronounced response of early sugarcane cultivars to soil water availability is due to the fact they are physiologically prepared to switch quickly between growth and ripening phases, even under favourable conditions for growing.

In July, the cumulative water deficit increased again and reached its highest value in September–October, when it totalled >130 mm (Fig. 1, black arrow 2). At this time, the differences in the quality attributes among cultivars of all groups became insignificant, although cultivars with the highest (e.g. SP91-1049) and the lowest (SP83-2847) sucrose contents retained their positions inside the groups. Based on that, the water deficit that accumulated over the months during the harvest season became the most important environmental factor for sugarcane ripening, not only by inducing sucrose accumulation, but also by promoting stalk dehydration, as also observed by Robertson and Donaldson (1998). Inman-Bamber (2004) concluded that stalk biomass accumulation is affected when the water deficit is >120 mm, while the accumulation of sucrose is affected when water deficit reaches >145 mm. Scarpari and Beauclair (2004) also reported that a water deficit >130 mm in the months preceding harvest has positive influence on sucrose accumulation. However, the right amount of water deficit for this process is not clearly defined, since it also depends on other variables, such as evapotranspiration rate, soil water-holding capacity and crop phase (Scarpari and Beauclair 2009).

Genotype factors are also important for the ripening process (Ferraro *et al.* 2009). This was reported by Stuppiello (1987), who observed that middle-ripening sugarcane cultivars rarely reach maximum sucrose content in the beginning of the harvesting season, when early cultivars exhibit their best sucrose accumulation. The same pattern was observed when early- and middle-ripening cultivars were compared with the late ones. Thus, the classification of cultivars into early-, middle- and late-ripening groups is for crop management purposes, in order to maximise sucrose production during the season.

The behaviour described above was confirmed when the k-means clustering analysis showed which variables were more effective at discriminating cultivar ripening patterns as the season progressed (Figs 5–7). The cultivars in Group 1 (SP91-1049 and SP86-155) showed an increase in M and RS over the previous month, suggesting that their precocity makes them more sensitive to climatic variations and more synchronised in vegetative growth. This was mentioned by Venkataramana *et al*. (1991) and Mamet and Galwey (1999), who considered early-ripening sugarcane cultivars more efficient physiologically since they are able to switch very quickly from vegetative growth to ripening and *vice-versa.* Higher air temperature and rainfall induce sugarcane plants to grow, due to tissue re-hydration and sucrose inversion (Alexander 1973), whereas lower air temperature and restricted soil water supply stimulate the plants to stop growth and store sucrose.

Conclusions

Clustering analyses by hierarchical and non-hierarchical methods and principal components analysis were able to discriminate sugarcane cultivars into three groups with similar ripening patterns: early (SP91-1049 and SP86-155), middle (RB928064, SP87-365, SP90-3414 and RB867515) and late (SP83-2847 and SP86-42). These methods considered all the quality variables involved in the ripening process and therefore provided a more complete and rational analysis than traditional univariate approaches. The differences between the groups formed early in the season (April–May) became smaller and essentially insignificant late in the season (August–September), but cultivars with the highest and lowest sucrose levels kept this characteristic throughout the harvest season, indicating that the genetic potential of cultivars is very important in determining the magnitude of their response, in terms of sucrose accumulation, to local climate. Rainfall events promoted inversion in the observed clustering patterns, reducing the

differences among groups, and this effect was more pronounced for early cultivars than for the others. For early cultivars, all quality attributes considered influenced their ripening, except fibre, whereas for middle and late cultivars, fibre and moisture were the attributes that most influenced their ripening. The use of multivariate methods to classify sugarcane cultivars in relation to ripening makes it possible to determine which cultivars ripen earlier and which quality variables contribute more to this process at different times of the year. This approach also helps in planning the harvest of cultivars with similar characteristics, optimising the process and the return in terms of sucrose per hectare. In addition, the techniques presented may be useful for selecting clones in the initial stages of breeding, by comparing them to standard cultivars. The method proposed by this study is more complete than the usual analysis of Pol variation over time, allowing a better understanding of sugarcane ripening in terms of the variability contained in the wider dataset.

Acknowledgements

The authors thank the Raízen Co. for the sugarcane quality data used in this study, the two anonymous reviewers for the valuable suggestions, and Prof. Terry Gillespie, from University of Guelph, Canada, for reviewing the manuscript. The second author is thankful for the fellowship provided by the Brazilian Research Council (CNPq).

References

Alexander AG (1973) 'Sugarcane physiology: a comprehensive study of the Saccharum source-to-sink system.' (Elsevier Scientific: Amsterdam)

Allen RG, Pereira LS, Raes D, Smith M (1998) 'Crop evapotranspiration: guidelines for computing crop water requirements.' Irrigation and Drainage Paper, 56. (FAO: Rome)

Ariyo OJ (1993) Genetic diversity in West African Okra (*Abelmoschus caillei* (A. Chev.) Stevels)—Multivariate analysis of morphological and agronomic characteristics. *Genetic Resources and Crop Evolution* **40**, 25–32. doi:10.1007/BF00053461

Beiragi MA, Sar BAS, Geive HS, Alhossini MN, Rahmani A, Gharibdoosti AB (2012) Application of the multivariate analysis method for some traits in maize. *African Journal of Agricultural Research* **7**, 1524–1533.

Cardozo NP (2012) Modeling sugarcane ripening as function of meteorological variables. MSc Thesis, Universidade de São Paulo, Piracicaba, SP, Brazil. Available at: www.teses.usp.br/teses/disponiveis/11/11131/tde-14032012-080359/

Cardozo NP, Sentelhas PC (2013) Climatic effects on sugarcane ripening under the influence of cultivars and crop age. *Scientia Agricola* **70**, 449–456.

Clements HF (1962) The ripening of sugar cane. *Sugar y Azúcar* **57**, 29–78.

CONSECANA (2006) 'Manual de instruções.' (Conselho dos Produtores de Cana-de-açúcar, Açúcar e Álcool do Estado de São Paulo: Piracicaba, SP, Brasil)

Fernandes AC (2011) 'Cálculos na agroindústria da cana-de-açúcar.' (STAB: Piracicaba, SP, Brasil)

Ferraro DO, Rivero DE, Ghersa CM (2009) An analysis of the factors that influence sugarcane yield in Northern Argentina using classification and regression trees. *Field Crops Research* **112**, 149–157. doi:10.1016/j.fcr.2009.02.014

Ferreira EA, Aspiazú I, Concenço G, SIlva AF, Silva AA, Galon LL, SIlva DV (2011) Evaluation and grouping of sugarcane genotypes in agreement with their physiologic characteristics types. *Revista Trópica – Ciências Agrárias e Biológicas* **5**, 30–38.

Hair JF, Black B, Babin B, Anderson RE, Tatham RL (2005) 'Multivariate data analysis.' 6th edn (Prentice Hall: Englewood Cliffs, NJ)

Hartingan JA (1975) 'Clustering algorithms.' (Wiley: New York)

Horii J (2004) A qualidade da matéria-prima, na visão agrícola. *Visão Agrícola.* **1**, 91–93.

Informa Economics FNP (2012) 'AGRIANUAL 2012: Anuário da Agricultura Brasileira.' (Informa Economics FNP: São Paulo)

Inman-Bamber NG (2004) Sugarcane water stress criteria for irrigation and drying off. *Field Crops Research* **89**, 107–122. doi:10.1016/j.fcr.2004.01.018

Jackson PA, Hogarth DM (1992) Genotype × environment interactions in sugarcane. I. Patterns of response across sites and crop-years in north Queensland. *Australian Journal of Agricultural Research* **43**, 1447–1459. doi:10.1071/AR9921447

Jackson PA, Mcrae T, Hogarth DM (1995) Selection of sugarcane families across variable environments II. Patterns of response and association with environmental factors. *Field Crops Research* **43**, 119–130. doi:10.1016/0378-4290(95)00040-W

Johnson RA, Wichern DW (2001) 'Applied multivariate statistical analysis.' 5th edn (Prentice Hall: Englewood Cliffs, NJ)

Kaiser HF (1958) The varimax criterion for analytic rotation in factor analysis. *Psychometrika* **23**, 187–200. doi:10.1007/BF02289233

Lavanholi MGDP (2008) 'Qualidade da cana-de-açúcar como material-prima para produção de açúcar e álcool.' In 'Sugarcane'. (Eds LL Dinardo-Miranda, ACM Vasconcelos, MGA Landell) (Instituto Agronômico: Campinas)

Lawes RA, Lawn RJ (2005) Applications of industry information in sugarcane production systems. *Field Crops Research* **92**, 353–363. doi:10.1016/j.fcr.2005.01.033

Lawes RA, McDonald LM, Wegener MK, Basford KE, Lawn RJ (2002) Factors affecting cane yield and commercial cane sugar in the Tully district. *Australian Journal of Experimental Agriculture* **42**, 473–480. doi:10.1071/EA01020

Lawes RA, Wegener MK, Basford KE, Lawn RJ (2004) The evaluation of the spatial and temporal stability of sugarcane farm performance based on yield and commercial cane sugar. *Australian Journal of Experimental Agriculture* **55**, 335–344. doi:10.1071/AR03169

Legendre BL (1975) Ripening of sugarcane: effects of sunlight, temperature, and rainfall. *Crop Science* **15**, 349–352.

Maji AT, Shaibu AA (2012) Application of principal component analysis for rice germplasm characterization and evaluation. *Journal of Plant Breeding and Crop Science* **4**, 87–93.

Mamet LD, Galwey NW (1999) A relationship between stalk elongation and earliness of ripening in sugarcane. *Experimental Agriculture* **35**, 283–291. doi:10.1017/S0014479799003087

Marin FR, Jones JW, Royce F, Suguitani C, Donzeli JL, Pallone Filho WJ, Nassif DSP (2011) Parameterization and evaluation of predictions of DSSAT/CANEGRO for sugarcane Brazilian production systems. *Agronomy Journal* **103**, 304–315. doi:10.2134/agronj2010.0302

Rahman MM, Hussain A, Syed MA, Ansari A, Mahmud MAA (2011) Comparison among clustering in multivariate analysis of rice using morphological traits, physiological traits and simple sequence repeat markers. *American-Eurasian Journal of Agriculture and Environmental Science* **1**, 876–882.

Rao CP, Rahman MA, Rao PN, Reddy JR (1985) Genetic divergence analysis in sugarcane. *Genetica Agraria.* **39**, 237–247.

Robertson MJ, Donaldson RA (1998) Changes in the components of cane and sucrose yield in response to drying-off of sugarcane before harvest. *Field Crops Research* **55**, 201–208. doi:10.1016/S0378-4290(97)00065-8

Rudorff BFT, Aguiar DA, Silva WF, Sugawara LM, Adami M, Moreira MA (2010) Studies on the rapid expansion of sugarcane for ethanol

production in São Paulo State (Brazil) using landsat data. *Remote Sensing* **2**, 1057–1076. doi:10.3390/rs2041057

Scarpari MS, Beauclair EGF (2004) Sugarcane maturity estimation through edaphic-climatic parameters. *Scientia Agricola* **61**, 486–491. doi:10.1590/S0103-90162004000500004

Scarpari MS, Beauclair EGF (2009) Physiological model to estimate the maturity of sugarcane. *Scientia Agricola* **66**, 622–628. doi:10.1590/S0103-90162009000500006

Sneath PH, Sokal RR (1973) 'Numerical taxonomy: The principles and practice of numerical classification.' (W.H. Freeman: San Francisco, CA)

Stuppiello J (1987) Cana-de-açúcar como matéria-prima. In 'Cana-de-açúcar cultivo e utilização'. pp. 761–804. (Fundação Cargill: Campinas, SP, Brazil)

Thornthwaite CW, Mather JR (1955) The water balance. In 'Publications in climatology VIII'. pp. 1–104. (Drexel Institute of Climatology: Centerton, NJ)

UNICA (2012) Dados e cotações – estatísticas, Produção Brasil. UNICA (Sugarcane Industry Association Brazil). Available at: www.unica.com.br/dadosCotacao/estatistica

Venkataramana S, Mohan Naidu K, Singh S (1991) Invertases and growth factors dependent sucrose accumulation in sugarcane. *Plant Science* **74**, 65–72. doi:10.1016/0168-9452(91)90256-8

Viana JMS, Cardoso AA, Cruz CDR, Regazzi AJ, Del Giudice RM (1991) Genetic divergence in sugarcane (*Saccharum* spp.) clones. *Brazilian Journal of Genetics.* **14**, 753–763.

Wagih ME, Ala A, Musa Y (2004) Evaluation of sugarcane varieties for maturity earliness and selection for efficient sugar accumulation. *Sugar Technology* **6**, 297–304. doi:10.1007/BF02942512