The Journal of Engineering and Exact Sciences – JCEC, Vol. 05 N. 03 (2019)
journal homepage: https://periodicos.ufv.br/ojs/jcec
doi: 10.18540/jcecvl5iss3pp0228-0237
OPEN ACCESS – ISSN: 2527-1075

EVALUATION OF CANE SUGAR PRODUCTION USING MULTIVARIATE STATISTICAL METHODS

AVALIAÇÃO DA PRODUÇÃO DE AÇÚCAR DE CANA UTILIZANDO MÉTODOS ESTATÍSTICOS MULTIVARIADOS

B. J. C. de CASTRO[1] and A. BERNARDO[1]

[1] Federal University of São Carlos, Graduate Program in Chemical Engineering, São Carlos, São Paulo, Brazil, abernardo@ufscar.br

ARTICLE INFO

Article history:
Received 2018-11-09
Accepted 2019-06-27
Available online 2019-06-30

palavras-chave
Indústria de açúcar
Monitoramento de processos
Qualidade de açúcar
Estatística multivariada

keywords
Sugar industry
Process monitoring
Sugar quality
Multivariate statistics

ABSTRACT

In sugarcane industries, process monitoring has the main purpose of maximizing sugar and ethanol production, meeting the quality parameters demanded by customers. The aim of this work was to identify industrial process variables that presented the greatest impacts on the quantity and quality of the produced sugar, by applying principal component analysis (PCA) and partial least squares regression (PLS) to the process data of a sugar and ethanol industry. The PCA correlation matrix highlighted the correlation between the presence of alcoholic flocs in sugar and the concentrations of starch and dextran in it. Both PCA and PLS showed that the color of the sugar was highly correlated to its moisture content. The first three principal components accounted for 40.92% of the total data variability.

RESUMO

Nas indústrias de cana-de-açúcar, o monitoramento de processos tem como principal objetivo maximizar as produções de açúcar e etanol, atendendo aos parâmetros de qualidade exigidos pelos clientes. O objetivo deste trabalho foi identificar as variáveis do processo industrial que apresentaram os maiores impactos sobre a quantidade e a qualidade do açúcar produzido, aplicando a análise por componentes principais (PCA) e a regressão por mínimos quadrados parciais (PLS) aos dados de processo de uma usina produtora de açúcar e etanol. A matriz de correlações do PCA destacou a correlação entre a presença de flocos alcoólicos no açúcar e sua concentração de amido e dextrana. Ambas análises mostraram que a cor do açúcar esteve altamente correlacionada à sua umidade. Os três primeiros componentes principais responderam, juntos, por 40,92% da variabilidade total dos dados.

1. INTRODUCTION

Process monitoring is becoming increasingly necessary in modern industries in order to meet higher demands in terms of production, cost reduction, and product quality, according to the requirements of the consumer market and sustainability standards. The most widely applied traditional method of process monitoring is the model-based method, which uses conservation laws and physical/chemical relations. As long as the model is reliable, this method tends to provide more accurate and decisive results, compared to other methods. However, for modern and complicated industrial processes, constructing models that accurately characterize them becomes extremely hard, time-consuming, expensive, and even impossible (Ge et al., 2013; Kano and Nakagawa, 2008; Sun et al., 2001; Yin et al., 2015).

For complex processes, or processes for which a first-principle model is unknown, knowledge-based methods can be used. These methods use the accumulated knowledge of expert plant operators and, therefore, tend to be more intuitive. However, many years of effort and experience are necessary to create a reliable knowledge base. On the other hand, data-based methods do not require any process modeling or operator knowledge. These techniques have been applied in various industrial sectors in recent years, with the purpose of extracting as much information as possible, based on large amounts of available data, in order to model, monitor, and control processes (Ge et al., 2013; Qin, 2012; Sun et al., 2001; Yin et al., 2015).

Among data-based methods, multivariate statistical techniques have become popular in process monitoring and fault diagnosis, particularly principal component analysis (PCA) and partial least squares regression (PLS), due mainly to their simplicity, low computational effort, and ability to handle large numbers of highly correlated variables (Johnson and Wichern, 2007; Sun et al., 2001; Yin et al., 2014).

PCA is a multivariate statistical method that enables reduction of the dimensionality of large amounts of correlated data, preserving the significant variability information extracted from process measurements. It has been successfully applied to image analysis, pattern recognition, data compression, time series prediction, and process monitoring (Chen et al., 2015; Jiang et al., 2013; Kano and Nakagawa, 2008; Ku et al., 1995; Ramburan et al., 2011; Rodushkin et al., 2011; Santchurn et al., 2012; Yin et al., 2014). PLS is another multivariate statistical method widely used in model construction, process monitoring and control, and fault detection. This technique enables the prediction of key performance indicators of a process by the application of partial least squares regression to uncorrelated variables generated from measurements (Kano and Nakagawa, 2008; Qin, 2012; Roy and Roy, 2008; Yin et al., 2014).

In the sugarcane industry, monitoring of agricultural and industrial processes aims at increasing agroindustrial efficiency, by increasing the sugar and ethanol production per hectare of cultivated sugarcane, and at ensuring the maintenance of quality parameters demanded by the consumer market (Fernandes, 2011). Therefore, the aim of this work was to use these methods to analyze the industrial process data of a sugar and ethanol industry located in São Paulo State, collected during the 2015/2016 harvest season. This was expected to assist in understanding the relations among the variables and in identifying those that had the greatest impacts on the quantity and quality of the sugar produced. The equations of the PCA and PLS methods are provided in Appendix A.

2. METHODS

Table 1 presents 50 process variables that were monitored on 101 production days, between 1 June and 31 October 2015. Days with an absence of information for at least one of the 50 variables were excluded from the analyses.

Figure 1 shows a flow diagram of the sugar crystallization in the two-boiling system of the industry studied, indicating the syrup, massecuite, and molasses streams listed in Table 1. Light molasses could also be recycled to the B vacuum pans, depending on the goals for each period of the season.

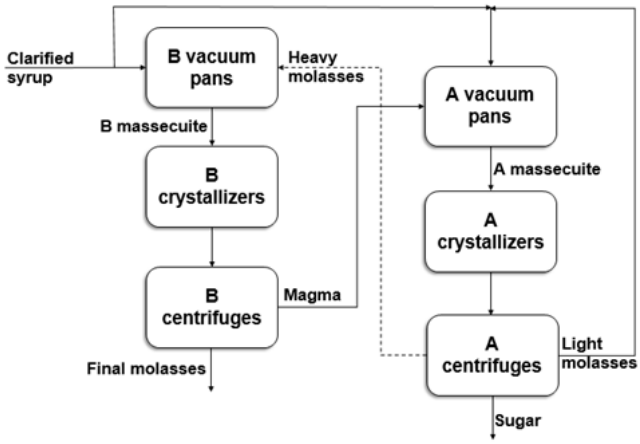

Figure 1 - Flow diagram of sugar crystallization (two-boiling system).

The data were normalized to zero mean and unit variance, followed by construction of a normalized data matrix that was used in the PCA and PLS procedures. Both sets of analyses were performed using Minitab software. Cross-validation was used to determine the numbers of components to be retained in the PLS models, calculating the predictive ability of potential models, excluding one observation at a time (leave-one-out method).

By identifying the highest coefficients, c_{ij}, in the PCA correlation matrix, the most highly correlated variables among the 50 original variables were identified and analyzed. The highest values for coefficients $c_{i,34}$ and $c_{i,36}$ enabled identification of the variables most highly correlated with sugar production and sugar color, respectively. A coefficient matrix was obtained by applying PLS regression with the sugar production variable as the key performance indicator and all the other parameters as input variables. The variables presenting the highest coefficients were compared with the variables identified by the PCA as being highly correlated with sugar production. A similar comparison procedure was applied between the highest PLS coefficients, with sugar color as the key performance indicator, and the variables shown by PCA to be highly correlated with sugar color.

The percentages of total variability contained in each linear combination were calculated from the eigenvalues resulting from decomposition of the PCA correlation matrix. The first three principal components were analyzed according to the original variables with the highest loadings in them.

Table 1 – Process variables considered in the PCA and PLS procedures.

	Variable	Unit	Reference for chemical analysis
1	Processed cane	t/day	-
2	Concentration of TRS (total reducing sugars) in cane	kg/t	ICUMSA, 2015
3	Cane cutting time (average time from cutting to processing)	h	-
4	Concentration of mineral impurities in cane	kg/t	UDOP, 2014 a
5	Concentration of vegetal impurities in cane	kg/t	UDOP, 2014 b
6	Dextran concentration in cane	mg/brix	Chen and Chou, 1993
7	Time use	%	-
8	Vapor pressure	psi	-
9	Water pressure in vacuum pan condensers	psi	-
10	Brix[a] of clarified syrup	%	ICUMSA, 2015
11	Pol[b] of clarified syrup	%	ICUMSA, 2015
12	Purity[c] of clarified syrup	%	-
13	pH of clarified syrup	-	ICUMSA, 2015
14	ICUMSA[d] color of clarified syrup	IU	ICUMSA, 2015
15	Clarified syrup turbidity	IU	ICUMSA, 2015
16	Brix[a] of heavy molasses	%	ICUMSA, 2015
17	Pol[b] of heavy molasses	%	ICUMSA, 2015
18	Purity[c] of heavy molasses	%	-
19	Brix[a] of light molasses	%	ICUMSA, 2015
20	Pol[b] of light molasses	%	ICUMSA, 2015
21	Purity[c] of light molasses	%	-
22	Brix[a] of A massecuite	%	ICUMSA, 2015
23	Pol[b] of A massecuite	%	ICUMSA, 2015
24	Purity[c] of A massecuite	%	-
25	Brix[a] of B massecuite	%	ICUMSA, 2015
26	Pol[b] of B massecuite	%	ICUMSA, 2015
27	Purity[c] of B massecuite	%	-
28	Brix[a] of magma	%	ICUMSA, 2015
29	Pol[b] of magma	%	ICUMSA, 2015
30	Purity[c] of magma	%	-
31	Brix[a] of final molasses	%	ICUMSA, 2015
32	Pol[b] of final molasses	%	ICUMSA, 2015
33	Purity[c] of final molasses	%	-
34	Sugar production	sacks (50 kg)/day	-
35	Sugar polarization	°Z	ICUMSA, 2015
36	ICUMSA[d] color of sugar	IU	ICUMSA, 2015
37	Sugar moisture	%	ICUMSA, 2015
38	Ash concentration in sugar	%	ICUMSA, 2015
39	Sulfite concentration in sugar	mg/kg	ICUMSA, 2015
40	Insoluble solids in sugar	n. (1 to 10)	ICUMSA, 2015
41	Concentration of black points in sugar	n./100 g	Lopes and Borges, 2004
42	Concentration of magnetizable particles in sugar	mg/kg	Lopes and Borges, 2004
43	pH of sugar	-	ICUMSA, 2015
44	Sugar turbidity	NTU	ICUMSA, 2015
45	Dextran concentration in sugar	mg/kg	ICUMSA, 2015
46	Starch concentration in sugar	mg/kg	Lopes and Borges, 2004
47	Alcoholic flocs in sugar	-	Lopes and Borges, 2004
48	Sugar filterability	min	Lopes and Borges, 2004
49	Particle mean size of sugar crystals	mm	ICUMSA, 2015
50	Coefficient of variation of sugar crystals	%	ICUMSA, 2015

[a]Brix: mass percentage of dissolved solids.
[b]Pol: mass percentage of dissolved sucrose.
[c]Purity: mass percentage of sucrose in dissolved solids (purity = pol/brix).
[d]ICUMSA: International Commission for Uniform Methods of Sugar Analysis.

3. RESULTS AND DISCUSSION

In this section, the results obtained in the PCA and PLS procedures are presented and discussed. First, the correlation matrix is analyzed, then the PCA and PLS results for sugar production and sugar color are compared. Finally, the first three principal components are presented.

3.1 Analysis of the correlation matrix

Table 2 shows the variables that presented the highest correlations in the correlation matrix. High correlations between brix and pol, or between pol and purity, for the same material, occurred for the clarified syrup ($c_{10,11} = 0.9909$), heavy molasses ($c_{16,17} = 0.8834$), light molasses ($c_{19,20} = 0.8752$), B massecuite ($c_{26,27} = 0.8513$), magma ($c_{28,29} = 0.8100$), and final molasses ($c_{32,33} = 0.8070$). Brix is the mass percentage of dissolved solids, while pol is the polarimetric reading of the solution, considered as its sucrose mass percentage. Purity is the relation between pol and brix, hence representing the mass percentage of sucrose in the dissolved solids. The high values of c_{ij} in these cases indicated that the purities of these materials did not vary significantly. Therefore, pol was intimately related to brix during the entire period.

Time use was calculated taking into account the hours when the mills were broken or at standstill, and indicated the percentage of time during which cane was effectively processed on a particular day. Hence, the value reflected the degree of stability of the plant on that day. Processed cane showed a high correlation with time use ($c_{1,7} = 0.9591$), which was not an obvious relation. The mills had the possibility of rotation velocity adjustments, so high time use did not necessarily indicate a high quantity of processed cane. For instance, on a day when the mills had operated for 24 hours, but at low velocity, the amount of processed cane would be low, while time use would continue to be 100%. However, the high correlation obtained for the plant studied showed that the amount of processed cane was related to the effective hours of milling, irrespective of velocity adjustments.

The dextran concentration in the sugar presented high correlations with alcoholic flocs ($c_{45,47} = 0.8591$) and the starch concentration ($c_{45,46} = 0.8071$) in the sugar. Dextran is a polysaccharide resulting from the action on sucrose of microorganisms of the genus *Leuconostoc*. Its presence in sugar increases the viscosity of sugar solutions and consequently negatively affects products that use this sugar as an ingredient, such as candies and cereal bars (Oliveira et al., 2007).

On the other hand, starch is a polysaccharide produced by the sugarcane itself, and is found mainly in the leaves, internodes, and tops. Starch is partially removed during juice treatment and sugar manufacture, but part of it remains in the process and ends up in the sugar crystals. The greatest impact of a high starch concentration is associated with the difficulty of sugar solution filtration in food industries where sugar is dissolved and filtered (Oliveira et al., 2007).

Alcoholic flocs are polysaccharide precipitates that appear in hydroalcoholic solutions produced using sugar with the presence of these compounds as an ingredient. Although they do not affect product quality and can be easily re-dissolved under slight agitation, alcoholic flocs may cause the impression of a deteriorated or contaminated product (Oliveira et al., 2007). After the dextran concentration in the sugar, the second highest correlation of alcoholic flocs was with the starch concentration in the sugar ($c_{46,47} = 0.7310$).

The results indicated that dextran and starch in sugar were responsible for the emergence of alcoholic flocs in solutions. Lemos et al. (2013) reached a similar conclusion in assessment of the influence of different concentrations of dextran and starch in sugar on the formation of alcoholic flocs in solutions. The tests, which were performed using acidic beverages and hydroalcoholic solutions with volumetric concentrations between 55 and 89%, showed positive correlations between dextran and starch in sugar and floc formation. Merheb et al. (2016) studied the effects of dextran and starch on the appearance of alcoholic flocs in sugar crystals produced using crystallization by cooling or evaporation. In the tests performed using evaporation crystallization, the appearance of alcoholic flocs was significantly influenced by the presence of dextran, starch, and the combined action of the two contaminants (Merheb et al., 2016).

Table 2 – Ten highest correlations between variables *i* and *j* of the correlation matrix.

c_{ij} value	Variable *i*	Variable *j*
0.9909	10 – Brix of clarified syrup	11 – Pol of clarified syrup
0.9591	1 – Processed cane	7 – Time use
0.8834	16 – Brix of heavy molasses	17 – Pol of heavy molasses
0.8752	19 – Brix of light molasses	20 – Pol of light molasses
0.8591	45 – Dextran concentration in sugar	47 – Alcoholic flocs in sugar
0.8513	26 – Pol of B massecuite	27 – Purity of B massecuite
0.8101	36 – ICUMSA color of sugar	37 – Sugar moisture
0.8100	28 – Brix of magma	29 – Pol of magma
0.8071	45 – Dextran concentration in sugar	46 – Starch concentration in sugar
0.8070	32 – Pol of final molasses	33 – Purity of final molasses

The increase in mechanized harvesting of raw cane has led to vegetal components with high starch contents entering the industrial process. Furthermore, the greater area of raw cane exposed to microorganism contamination has also resulted in increased dextran formation during the period between cutting and processing (Merheb et al., 2016). Oliveira et al. (2002) concluded that longer periods of time between cane cutting and processing, as well as high temperatures and high moisture levels, contribute to increased contamination and consequently to higher dextran production.

Table 3 presents the correlations between the concentrations of starch and dextran in the sugar and the variables related to sugarcane quality. Both starch and dextran showed negative correlations with the concentration of TRS in the cane, which represents sugarcane purity. Positive correlations with cutting time confirmed the conclusion reached by Oliveira et al. (2002). Concentrations of mineral and vegetal impurities and starch in the cane also showed positive correlations with dextran and starch in the sugar.

The ICUMSA color of the sugar showed a high correlation with sugar moisture content ($c_{36,37} = 0.8101$). Color is the main parameter for sugar classification, and high values of this variable reflect visible yellowing of the crystals. Sugar with high color negatively impacts the quality of products whose desirable attributes involve color or transparency, such as beverages, candies, and ice creams (Oliveira et al., 2007). In the industry studied here, the main product was type 2 white crystals, with color below 150 IU. The sugar moisture content after passing through the dryers is recommended to be less than 0.1%, in order to avoid problems during sugar conditioning, such as crystal agglomeration and hardening, loss of brightness, sucrose inversion, and yellowing (Sarantópoulos et al., 2002).

High correlation between these two important sugar quality parameters indicated that high color was related to liquid occlusion and inclusion. Occlusion is the presence of a molasses film on the crystal surface, which also increases the sugar moisture content, and is usually a consequence of centrifugation failures. On the other hand, inclusion is the presence of molasses inside the crystals, which mainly occurs during crystal growth. Molasses inclusion is associated with abrupt changes in supersaturation conditions during crystallization, and is favored by high growth rates and large crystals (Mullin, 2001). The presence of compounds such as starch and dextran, which can modify the crystalline habit by inhibition of one of the faces, also favors liquid inclusion (Schlumbach et al., 2017). Molasses, which is rich in contaminants and colored compounds, decreases sugar purity and can also contribute to increasing its color, as discussed in detail in Section 3.2, where the PCA and PLS results for sugar color are presented.

3.2 PCA and PLS results for sugar production and sugar color

Table 4 shows the variables that presented the highest correlations with sugar production, according to the PCA correlation matrix, together with the variables that presented the highest coefficients in PLS, using sugar production as the key performance indicator.

Figure 2 illustrates the coefficients resulting from PLS applied to sugar production. Following application of the cross-validation method for prediction of the key performance indicator, the model with two principal components was selected, with adjusted R^2 equal to 0.7968 and predicted R^2 equal to 0.7090.

In both analyses, sugar production was highly correlated with the variables related to the quantity of sucrose available for sugar manufacturing: processed cane, concentration of TRS in the cane, and purity of the clarified syrup. In both analyses, time use also presented high correlations with sugar production. As described in Section 3.1, this variable reflects the degree of stability of the industrial plant. Days with high time use had high sugar production, because in addition to the material available for processing, sugar production is also highly dependent on process stability.

Table 3 – Correlations between dextran and starch in sugar and sugarcane characteristics.

	45 – Dextran concentration in sugar	46 – Starch concentration in sugar
2 – Concentration of TRS in cane	-0.2760	-0.4056
3 – Cane cutting time	0.3677	0.3202
4 – Concentration of mineral impurities in cane	0.4550	0.3641
5 – Concentration of vegetal impurities in cane	0.2369	0.1701
6 – Dextran concentration in cane	0.2283	0.2641

Table 4 – Variables highly correlated with sugar production, according to the PCA correlation matrix and PLS regression.

PCA correlation matrix		PLS regression	
$c_{i,34}$ value	Variable i	$m_{i,1}$ value	Variable i
0.6944	7 – Time use	0.1954	1 – Processed cane
0.6771	1 – Processed cane	0.1909	7 – Time use
0.6129	2 – Concentration of TRS in cane	0.1127	12 – Purity of clarified syrup
-0.5607	36 – ICUMSA color of sugar	-0.1083	5 – Concentration of vegetal impurities in cane
0.5413	12 – Purity of clarified syrup	0.1071	2 – Concentration of TRS in cane

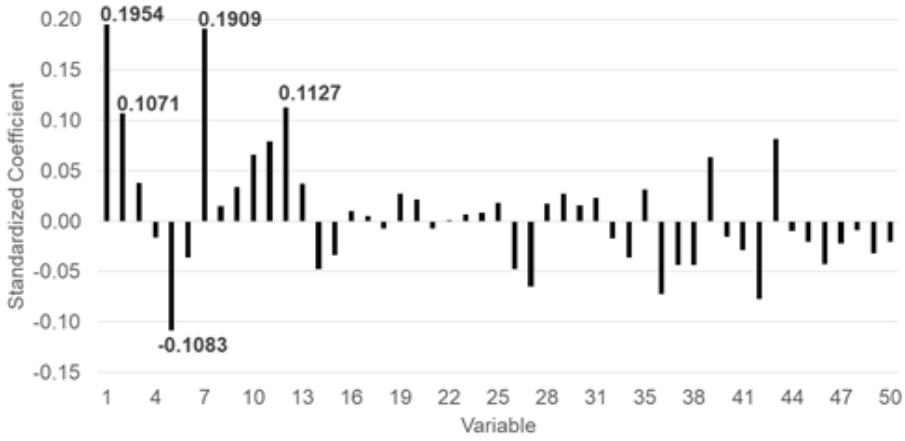

Figure 2 - PLS coefficients with sugar production as the key performance indicator.

In the PCA, the correlation with sugar color was negative, indicating that high sugar production was highly correlated to low sugar color (better quality sugar). This was because during the period studied, the days with high sugar production were also days with high plant stability, when quality control of the sugar was easier. On days when the plant was unstable, in addition to producing less sugar, it was more difficult to maintain sugar quality. This reinforces the need for a good quality of maintenance, in order to ensure high time use and the associated process stability.

The PLS results also highlighted the coefficient for the concentration of vegetal impurities in the cane, consisting of leaves, tops, and roots, which are parts of the plant with high starch contents. In the juice treatment, more specifically during its heating, the starch concentration is reduced because high temperatures cause its flocculation. However, raw cane milling, which has increased with mechanized harvesting, has led to significant increases of vegetal impurity concentrations, making it more difficult to totally remove starch during the juice treatment. Part of this impurity remains in the process and ends up delaying the crystallization, hence explaining the high and negative value of this coefficient in the regression (Merheb, 2014).

Table 5 lists the variables showing the highest correlations with sugar color in the PCA correlation matrix, together with those that presented the highest coefficients in the PLS regression, using the ICUMSA color of the sugar as the key performance indicator.

In this case, the model with three principal components was selected for PLS regression, with adjusted R^2 equal to 0.8854 and predicted R^2 equal to 0.7816. Figure 3 shows the coefficients resulting from the model.

Sugar moisture content presented the highest correlation with sugar color, as well as the highest coefficient in the PLS regression. The relation between these two variables, discussed previously in Section 3.1, is related to the molasses present in the crystals due to the phenomena of occlusion and inclusion. Molasses is rich in contaminants and, as Table 5 shows, the presence of these contaminants in the crystals was responsible for increased sugar color.

The ash concentration in sugar is the content of inorganic compounds, such as potassium, sodium, calcium, and iron in the form of chlorides, sulfates, carbonates, silicates, and sulfites, and its determination is based on electrical conductivity. These saline substances originate from the sugarcane juice itself and are incorporated during the industrial process, especially the juice treatment (Merheb, 2014). Schlumbach et al. (2017) studied the contribution of molasses inclusion to increased sugar color by determination of the transfer of ash from the syrup to the crystals, adopting the assumption that the presence of ash in the crystals resulted from inclusion.

Determination of the black point concentration in sugar, which is another quality parameter, is performed by counting the particles of color contrasting with that of the sugar crystals, such as rust, iron dusts, bagasse, caramel compounds, and soot (Lopes and Borges, 2004). Black points can be visually perceived in beverages and transparent liquids, or perceived by the tongue in candies and chocolates (Oliveira et al., 2007).

Table 5 – Variables highly correlated with the ICUMSA color of sugar, according to the PCA correlation matrix and the PLS regression.

PCA correlation matrix		PLS regression	
$c_{i,36}$ value	Variable i	$m_{i,36}$ value	Variable i
0.8101	37 – Sugar moisture	0.2139	37 – Sugar moisture
0.7926	38 – Ash concentration in sugar	0.2029	38 – Ash concentration in sugar
0.7724	41 – Concentration of black points in sugar	-0.1500	35 – Sugar polarization
-0.7702	35 – Sugar polarization	0.1277	44 – Sugar turbidity
0.7414	44 – Sugar turbidity	0.1272	41 – Concentration of black points in sugar

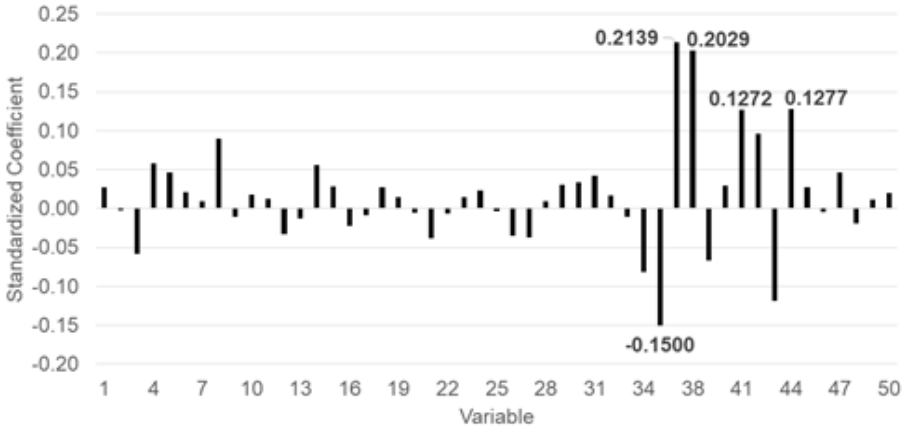

Figure 3 - PLS coefficients using the ICUMSA color of sugar as the key performance indicator.

Sugar polarization is the sucrose percentage in sugar, and it presented a negative correlation with the color. In other words, the lower the polarization, the higher the sugar color value. Sugar with low polarization has high concentrations of contaminants such as dextran, starch, and ash, which are associated with lower product quality (Oliveira et al., 2007).

Lastly, sugar turbidity reflects the obstruction of light passing through a sugar solution, due to the presence of particles in suspension that cause dispersion of the transmitted light. Turbidity is often an undesirable characteristic in soft drinks and alcoholic beverages, and is directly related to the presence of dextran and starch in sugar, as shown in the evaporation crystallization experiments performed by Merheb et al. (2016). In the correlation matrix, sugar color also presented high positive correlations with the concentrations of these two contaminants in the sugar crystals: 0.5557 for starch and 0.5843 for dextran.

3.3 Principal components of process variability

Table 6 shows the eigenvectors corresponding to the first three principal components obtained from decomposition of the correlation matrix, the eigenvalues referring to each one, and the loadings of each original variable in them. The three highest loadings for each principal component are highlighted in bold italic. The variance proportion for each principal component was calculated by dividing the eigenvalue corresponding to it by the sum of all the eigenvalues. Figure 4 shows the eigenvalues for each component. The first three principal components together accounted for 40.92% of the total data variability.

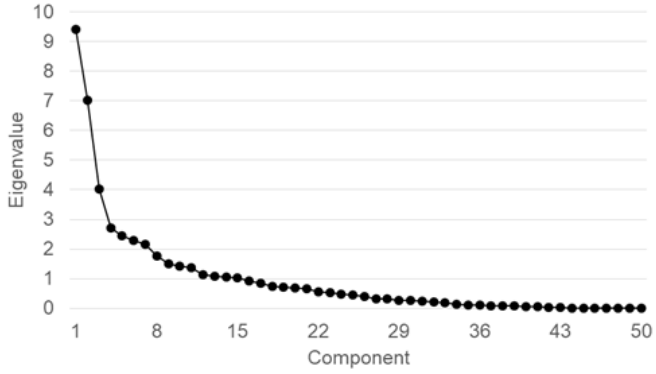

Figure 4 - Eigenvalues for each principal component.

In the linear combination of the 50 original variables that composed the first principal component, PC_1, the variables with the highest loadings were ICUMSA color of the sugar, ash concentration in the sugar, and concentration of black points in the sugar, with loadings of 0.2757, 0.2583, and 0.2557, respectively. In PC_1, responsible for 18.82% of the total data variability, the major factors were three important variables related to sugar quality, highlighting the substantial variation of sugar quality in the period studied, resulting from the high instability of the process. This instability was associated with mechanical problems in the mills and with disarms of the electric power generator. These problems occurred frequently during the whole period and consequently caused high instability in the process flow rates and the vapor pressure.

The second principal component, PC_2, responsible for 14.06% of the total variability, highlighted the variables pol of the clarified syrup, brix of the clarified syrup, and pol of the final molasses, with loadings of 0.2830, 0.2825, and 0.2448, respectively. These variables are related to the sugar concentration in the sugar manufacturing streams: the clarified syrup, which is the input of crystallization, and the final molasses, which is the output of this process. PC_2 suggested that greater control of the concentrations of the sugar manufacturing process streams would play a key role in reducing plant variability.

The original variables with the highest loadings in the third principal component, PC_3, were the concentration of vegetal impurities in the cane, processed cane, and time use, with loadings of 0.3126, -0.2936, and -0.2796, respectively. PC_3, responsible for 8.04% of the total variability, highlighted the sugarcane characteristics and the industry stability indicator.

4. CONCLUSIONS

The variables that presented the highest correlations with sugar production were time use, processed cane, concentration of TRS in the cane, ICUMSA color of the sugar, and purity of the clarified syrup. In the PLS regression, the variables with the highest coefficients, using sugar production as the key performance indicator, were processed cane, time use, purity of the clarified syrup, concentration of vegetal impurities, and TRS in the cane. Both PCA and PLS showed that high sugar production occurred on days with high stability, high cane milling, and high sugarcane purity, which also resulted in better sugar quality.

Table 6 – Eigenvectors, eigenvalues, and individual and cumulative variance percentages for the first three principal components.

	Variable	PC_1	PC_2	PC_3
1	Processed cane	-0.0940	0.0476	***-0.2936***
2	Concentration of TRS in cane	-0.2055	0.1502	-0.1581
3	Cane cutting time	0.0836	0.0976	-0.1594
4	Concentration of mineral impurities in cane	0.1388	0.1293	0.0749
5	Concentration of vegetal impurities in cane	0.0803	0.0570	***0.3126***
6	Dextran concentration in cane	0.1247	-0.0201	0.1394
7	Time use	-0.1143	0.0534	***-0.2796***
8	Vapor pressure	-0.0505	0.1827	-0.0185
9	Water pressure in vacuum pan condensers	-0.0335	0.1204	-0.0096
10	Brix of clarified syrup	-0.0496	***0.2825***	-0.0354
11	Pol of clarified syrup	-0.0694	***0.2830***	-0.0466
12	Purity of clarified syrup	-0.1574	0.0991	-0.0949
13	pH of clarified syrup	0.0391	-0.0672	-0.1402
14	ICUMSA color of clarified syrup	0.1760	-0.0562	0.0347
15	Clarified syrup turbidity	0.1657	0.0386	0.0694
16	Brix of heavy molasses	-0.0670	0.1442	-0.0345
17	Pol of heavy molasses	-0.0696	0.2437	0.0228
18	Purity of heavy molasses	-0.0172	0.2357	0.1147
19	Brix of light molasses	0.0257	0.1211	-0.1091
20	Pol of light molasses	-0.0102	0.2046	-0.0185
21	Purity of light molasses	-0.0679	0.2015	0.1659
22	Brix of A massecuite	-0.0494	0.0616	-0.0893
23	Pol of A massecuite	-0.1217	0.1707	0.1188
24	Purity of A massecuite	-0.1081	0.1586	0.2160
25	Brix of B massecuite	-0.0558	0.0756	-0.1552
26	Pol of B massecuite	-0.0003	0.2113	0.1061
27	Purity of B massecuite	0.0331	0.1955	0.2142
28	Brix of magma	0.0301	0.0610	-0.2593
29	Pol of magma	-0.0180	0.0774	-0.2282
30	Purity of magma	-0.0781	0.0292	0.0550
31	Brix of final molasses	-0.0603	0.0855	-0.1637
32	Pol of final molasses	-0.1037	***0.2448***	0.0720
33	Purity of final molasses	-0.0835	0.2354	0.2028
34	Sugar production	-0.2240	0.0476	-0.2441
35	Sugar polarization	-0.2538	-0.1177	0.0615
36	ICUMSA color of sugar	***0.2757***	0.0730	-0.0030
37	Sugar moisture	0.2394	0.0519	-0.0416
38	Ash concentration in sugar	***0.2583***	-0.0354	-0.0814
39	Sulfite concentration in sugar	0.0643	0.0116	-0.2544
40	Insoluble solids in sugar	0.2215	0.1176	-0.0895
41	Concentration of black points in sugar	***0.2557***	0.1321	-0.1053
42	Concentration of magnetizable particles in sugar	0.1817	0.0403	-0.0238
43	pH of sugar	-0.0280	0.1661	-0.1882
44	Sugar turbidity	0.2524	0.1162	-0.1271
45	Dextran concentration in sugar	0.2228	0.1641	-0.0169
46	Starch concentration in sugar	0.2387	0.1081	-0.0289
47	Alcoholic flocs in sugar	0.1874	0.1726	0.0396
48	Sugar filterability	0.1823	-0.0255	-0.0741
49	Particle mean size of sugar	0.0178	-0.1735	0.0536
50	Coefficient of variation of sugar	0.0955	-0.0648	-0.0712
	Eigenvalues	**9.41**	**7.03**	**4.02**
	Individual percentage (%)	**18.82**	**14.06**	**8.04**
	Cumulative percentage (%)	**18.82**	**32.88**	**40.92**

When analyzing sugar color, both PCA and PLS highlighted variables associated with sugar quality: sugar moisture, ash and black point concentrations, sugar polarization, and turbidity. The analyses indicated that sugar with high color was related to high concentrations of non-sucrose compounds such as water, dextran, starch, inorganic salts, soot, bagasse, rust, and caramel compounds, which were present in the crystals due to the occlusion and inclusion of molasses.

The first three PCA principal components together accounted for 40.92% of the total data variability. PC_1, accounting for 18.82%, was mainly composed of variables related to sugar quality: color, ash concentration, and black point concentration. In the case of PC_2, responsible for 14.06%, the major factors were variables associated with the sugar concentrations in the crystallization streams: brix and pol of the clarified syrup, and pol of the final molasses. PC_3, responsible for 8.04% of the variation, highlighted variables related to sugarcane characteristics and process stability: concentration of vegetal impurities in cane, processed cane, and time use.

APPENDIX A

PCA – Considering N measurements of the industrial process, each one with m variables, after scaling data to zero mean and unit variance, the normalized data matrix, X, is defined by Equation 1. Data normalizing is required when the measured variables differ in magnitude, in order to avoid an improper influence of one variable or another in the analysis.

$$X = \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1m} \\ x_{21} & x_{22} & \cdots & x_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ x_{N1} & x_{N2} & \ldots & x_{Nm} \end{pmatrix} \in \mathcal{R}^{N\times m} \quad (1)$$

The correlation matrix, ϕ, is given by Equation 2, and has the form shown in Equation 3.

$$\Phi = \frac{1}{N-1} X^t X \quad (2)$$

$$\Phi = \begin{pmatrix} c_{11} & c_{12} & \cdots & c_{1m} \\ c_{21} & c_{22} & \cdots & c_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ c_{m1} & c_{m2} & \ldots & c_{mm} \end{pmatrix} \in \mathcal{R}^{m\times m} \quad (3)$$

Significant variability information is extracted from the data by performing the singular value decomposition of Φ, as shown in Equation 4, where $Z \in \mathcal{R}^{m\times m}$ is an orthogonal matrix and $\Lambda \in \mathcal{R}^{m\times m}$ is a diagonal matrix. The columns of Z are the eigenvectors and the elements from the principal diagonal of Λ are the eigenvalues $\lambda_1, \ldots, \lambda_m$ of Φ.

$$\frac{1}{N-1} X^t X = Z \Lambda Z^t \quad (4)$$

According to the magnitude of the eigenvalues, Z and Λ can be divided as shown in Equations 5 and 6.

$$Z = [Z_{pc} \quad Z_{res}] \quad (5)$$

$$\Lambda = \begin{bmatrix} \Lambda_{pc} & 0 \\ 0 & \Lambda_{res} \end{bmatrix} \quad (6)$$

Let β denote the number of principal components. Then, $Z_{pc} \in \mathcal{R}^{m\times\beta}$, called principal subspace, contains the eigenvectors corresponding to the first β large eigenvalues in Λ, while $Z_{res} \in \mathcal{R}^{m\times(m-\beta)}$, called residual subspace, contains the eigenvectors corresponding to the last $(m-\beta)$ small eigenvalues in Λ (Yin et al., 2014).

PLS – Let U be the normalized data matrix, containing N measurements, each one with l variables, and Y be the key performance indicator matrix, with N measurements, each one with m variables, as shown in Equations 7 and 8.

$$U = \begin{pmatrix} u_{11} & u_{12} & \cdots & u_{1l} \\ u_{21} & u_{22} & \cdots & u_{2l} \\ \vdots & \vdots & \ddots & \vdots \\ u_{N1} & u_{N2} & \ldots & u_{Nl} \end{pmatrix} \in \mathcal{R}^{N\times l} \quad (7)$$

$$Y = \begin{pmatrix} y_{11} & y_{12} & \cdots & y_{1m} \\ y_{21} & y_{22} & \cdots & y_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ y_{N1} & y_{N2} & \ldots & y_{Nm} \end{pmatrix} \in \mathcal{R}^{N\times m} \quad (8)$$

The latent variables, t, are defined by Equation 9, where γ is the number of latent variables.

$$T = (t_1 \quad t_2 \quad \cdots \quad t_\gamma) \in \mathcal{R}^{N\times\gamma} \quad (9)$$

Based on the projection of U and Y onto latent variables, the correlation between U and Y can be determined by Equations 10 and 11.

$$U = TP^t + \tilde{U} = \hat{U} + \tilde{U} \quad (10)$$

$$Y = TQ^t + E_y = UM + E_y \quad (11)$$

where $P \in \mathcal{R}^{l\times\gamma}$ and $Q \in \mathcal{R}^{m\times\gamma}$ are the loading matrices of U and Y, respectively. $\hat{U}$ is highly correlated to Y. $\tilde{U}$ and E_y are residual subspaces and assumed to be uncorrelated to U e Y, respectively. According to the correlation between U and Y, given by Equations 10 and 11, matrices T and $M \in \mathcal{R}^{l\times m}$, called coefficient matrix, can be determined by Equations 12 and 13, respectively.

$$T = UR \quad (12)$$

$$M = RQ^t \quad (13)$$

where $P^tR = R^tP = I_{\gamma\times\gamma}$ and $R \in \mathcal{R}^{l\times\gamma}$. R is called loading matrix. Thus, the coefficient matrix, M, can be used to predict the key performance indicators from process measurements (Yin et al., 2014).

ACKNOWLEDGMENTS

The authors obtained financial support from CNPq (*Conselho Nacional de Desenvolvimento Científico e Tecnológico*), grant number 132045/2017-6 and from CAPES (*Coordenação de Aperfeiçoamento de Pessoal de Nível Superior*), finance code 001.

REFERENCES

CHEN, J. C. P.; CHOU, C. C. Cane sugar handbook: A manual for cane sugar manufacturers and their chemists. 12th ed., New York: John Wiley & Sons, 1993.

CHEN, R.; KANG, S.; HAO, X.; LI, F.; DU, T.; QIU, R.; CHEN, J. Variations in tomato yield and quality in relation to soil properties and evapotranspiration under greenhouse condition. **Sci. Hortic**, v. 197, n. 3, p. 318–328, 2015.

FERNANDES, A. C. Cálculos na agroindústria da cana-de-açúcar. 3rd ed., Piracicaba: STAB, 2011.

GE, Z.; SONG, Z.; GAO, F. Review of recent research on data-based process monitoring. **Ind. Eng. Chem. Res**, v. 52, n. 10, p. 3543–3562, 2013.

ICUMSA. ICUMSA Methods Book (2015). England: ICUMSA, 2015.

JIANG, Q.; YAN, X.; ZHAO, W. Fault detection and diagnosis in chemical processes using sensitive principal component analysis. **Ind. Eng. Chem. Res**, v. 52, n. 4, p. 1635–1644, 2013.

JOHNSON, R. A.; WICHERN, D. W. Applied multivariate statistical analysis. 6th ed., Upper Saddle River: Pearson Prentice Hall, 2007.

KANO, M.; NAKAGAWA, Y. Data-based process monitoring, process control, and quality improvement: Recent developments and applications in steel industry. **Comput. Chem. Eng**, v. 32, n. 1–2, p. 12–24, 2008.

KU, W.; STORER, R. H.; GEORGAKIS, C. Disturbance detection and isolation by dynamic principal component analysis. **Chemom. Intell. Lab. Syst**, v. 30, n. 1, p. 179–196, 1995.

LEMOS, L. R.; NOGUEIRA, A.; WOSIACKI, G.; LACERDA, L. G.; DEMIATE, I. M. The influence of different amounts of dextran and starch in crystallized sugar in the formation of floc in acidic carbonated solutions and alcoholic solutions. **Sugar Tech**, v. 15, n. 1, p. 65–70, 2013.

LOPES, C. H.; BORGES, M. T. M. R. Manual de análise de açúcar: açúcar VHP, VVHP, demerara, cristal, refinado e açúcar líquido. Araras: Sucral, 2004.

MERHEB, G. A. Influência da contaminação combinada de dextrana e amido na cristalização do açúcar. 2014. 300 p. Tese (Doutorado em Engenharia Química) - Universidade Federal de São Carlos, São Carlos, SP, 2014.

MERHEB, G. A.; OLIVEIRA, N. de; GIULIETTI, M.; BERNARDO, A. Combined effect of starch and dextran in sucrose crystallization. **Sugar Ind**, v. 141, p. 697–704, 2016.

MULLIN, J. W. Crystallization. 4th ed., Woburn: Butterworth Heinemann, 2001.

OLIVEIRA, D. T.; ESQUIAVETO, M. M. M.; SILVA JÚNIOR, J. F. Impacto dos itens da especificação do açúcar na indústria alimentícia. **Ciênc. Tecnol. Aliment**, v. 27, p. 99–102, 2007.

OLIVEIRA, A. S.; RINALDI, D. A.; TAMANINI, C.; VOLL, C. E.; HAULY, M. C. O. Fatores que interferem na produção de dextrana por microrganismos contaminantes da cana-de-açúcar. **Semin., Ciênc. Exatas Tecnol.**, v. 23, n. 1, p. 99–104, 2002.

QIN, S. J. Survey on data-driven industrial process monitoring and diagnosis. **Annu. Rev. Control**, v. 36, n. 2, p. 220–234, 2012.

RAMBURAN, S.; ZHOU, M.; LABUSCHAGNE, M. Interpretation of genotype×environment interactions of sugarcane: Identifying significant environmental factors. **Field Crops Res**, v. 124, n. 3, p. 392–399, 2011.

RODUSHKIN, I.; BAXTER, D. C.; ENGSTRÖM, E.; HOOGEWERFF, J.; HORN, P.; PAPESCH, W.; WATLING, J.; LATKOCZY, C.; VAN DER PEIJL, G.; BERENDS-MONTERO, S.; EHLERINGER, J.; ZDANOWICZ, V. Elemental and isotopic characterization of cane and beet sugars. **J. Food Compos. Anal**, v. 24, n. 1, p. 70–78, 2011.

ROY, P. P.; ROY, K. On some aspects of variable selection for partial least squares regression models. **QSAR Comb. Sci**, v. 27, n. 3, p. 302–313, 2008.

SANTCHURN, D.; RAMDOYAL, K.; BADALOO, M. G. H.; LABUSCHAGNE, M. From sugar industry to cane industry: investigations on multivariate data analysis techniques in the identification of different high biomass sugarcane varieties. **Euphytica**, v. 185, n. 3, p. 543–558, 2012.

SARANTÓPOULOS, C. I. G. L.; OLIVEIRA, L. M.; CANAVESI, É. Requisitos de conservação de alimentos em embalagens flexíveis. 2nd ed., Campinas: CETEA/ITAL, 2002.

SCHLUMBACH, K.; PAUTOV, A.; FLÖTER, E. Crystallization and analysis of beet and cane sugar blends. **J. Food Eng**, v. 196, p. 159–169, 2017.

SUN, X.; CHEN, T.; MARQUEZ, H. J. Detecting leaks and sensor biases by recursive identification with forgetting factors. In: CONFERENCE ON DECISION AND CONTROL, 2001, Edmonton. Anais... Edmonton: IEEE, 2001. p. 3716-3721.

UDOP. Determinação das impurezas minerais em carregamentos de cana-de-açúcar pelo método da incineração em forno mufla. UDOP, 2014a. Avaiable at: <http://www.udop.com.br/download/legislacao/bioenergia/institucional_site_juridico/impurezas_minerais_cana_objetivo_equipamentos_procedimentos.pdf>. Accessed on September 19, 2017.

UDOP. Determinação das impurezas vegetais e totais em carregamentos de cana-de-açúcar pelo método de limpeza manual e a seco. UDOP, 2014b. Avaiable at: <http://www.udop.com.br/download/legislacao/bioenergia/institucional_site_juridico/impurezas_vegetais_totais_objetivo_equipamentos_procedimentos.pdf>. Accessed on September 19, 2017.

YIN, S.; DING, S. X.; XIE, X.; LUO, H. A review on basic data-driven approaches for industrial process monitoring. **IEEE Trans. Ind. Electron**, v. 61, n. 11, p. 6418–6428, 2014.

YIN, S.; LI, X.; GAO, H.; KAYNAK, O. Data-based techniques focused on modern industry: an overview. **IEEE Trans. Ind. Electron**, v. 62, n. 1, p. 657–667, 2015.