Improving the efficiency of the estimation of agricultural and factory yields in Cuba

Alberto N González¹, Maira Ferrer¹, Gilda Jiménez¹ and Eduardo Casanova²

¹Sugarcane Research Institute, Havana, Cujae, Boyeros, CP 19390 La Habana, Cuba; alberto.gonzalez@inica.azcuba.cu
²Sugarcane Byproducts Research Institute, Via Blanca No. 804, La Habana, Cuba

Abstract The Cuban sugarcane needs efficient methods to estimate yields, both from an agriculture and industry perspective. We present a Cuban case study related to this subject. The study used yield data from 2001 to 2013 from 55 mill areas. Four existing models were evaluated and three mathematical models were proposed to predict sugar and potential yields (RPC). The statistical analysis explained more than 88% of the total variance of the data. The results obtained through stepwise regression analysis were validated for the mills using the 2014 sugar harvest and enabled indicators such as Pol % cane, fibre % cane, brix and Pol of the primary juice and factory losses to be used to predict both factory yield and RPC. The yield average of the 2013 harvest differed by as little as 0.9 t/ha from the actual yield obtained. One of the models (Model 3) was considered to best represent the planting and harvesting system used in Cuba. The relatively high regression coefficients indicate that the models can be used by technical staff and the managers of the mills for predictive purposes.

Key words Sugarcane modeling, yield agriculture, factory efficiency

INTRODUCTION

The development of modern informatics technologies linked to large data bases (DBs) held within research institutions or associated with agricultural production allows various possibilities for modeling yields to estimate, and predict and improve all the facets of harvest operations/organization on-farm and in relation to mills. Modern and efficient yield estimation tools could include:

- Empirical models supported by field information linked to crop behavior associated with environmental and management stimuli (Galvez et al. 2004);
- Mechanistic models supported by the plant’s simplified behavior, which integrates the plant with water-soil-climate data (Galvez et al. 2004);
- Remote and other sensing technologies linked to precision agriculture (PA) applications (Galvez and Lopez 2007).

The most common crop model applications (Wallach et al. 2006) include:

- prediction,
- determining optimal management options,
- large spatial-scale applications,
- characterizing plant varieties and plant-breeding techniques, and
- other applications such as educational tools.

Here, we present our first experiences of a sugarcane modeling project with the main objective of improving the estimation of yields, including in both the agriculture area and the mills.

MATERIAL AND METHODS

Evaluation of agricultural models

Our study included yield data from 50 mills and their associated mill areas in different provinces of Cuba. Four mathematical models were evaluated for estimating agricultural fields (Gonzalez 1995; Ferrer et al. 2014). The models were based on statistical multiple linear regression equations:

$$\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots$$ (Sokal and Rolf 1981).
The following models provided estimations of total yields for four different planting and harvest scenarios:

1. Model 1: Total yield = 19.9618 + 0.3956 * yield of the remaining previous spring-planted cane
2. Model 2: Total yield = -8.34 + 0.2357 * yield of the remaining previous spring-planted cane + 0.9629 * yield of ratoons
3. Model 3: Total yield = -8.2939 + 0.2332 * yield of the remaining previous spring-planted cane + 0.9572 * yield of ratoons + 0.0053 * yield of autumn-planted cane.
4. Model 4: Total yield = -0.4341 + 0.2248 * yield of the remaining previous spring-planted cane + 0.8756 * yield of ratoons - 0.0046 * autumn-planted cane + 0.1063 * yield of spring-planted cane.

The value of $\beta$ that precedes the yields of each crop provides a weight on the overall yield.

In all of the above cases the total number of data points exceeded 1,850. Statistical analysis was performed using the software Statistica 6.1.

Complementing this work, we performed monomodal modeling of the growth cycles of sugarcane to estimate evapotranspiration of sugarcane. This was done by modeling sugarcane yield using the relationship obtained from degree days, evapotranspiration and agricultural yields of sugarcane. A linear correlation between evapotranspiration and degree days accumulated by monomodal growth cycles was obtained (Ferrer et al. 2014).

**Evaluation of mill models**

The determination of recoverable sugar is central to sugarcane quality evaluation. Wood et al. (1972) developed a formula whereby recoverable sugar was expressed as a function of Pol less the factory losses, the non-sugar values and the sugarcane fiber content (Equation 1):

$$ERS \%\text{cane} = S - 0.485\left( N + \frac{F}{80} \right)$$

(Equation 1)

where ERS is the expected recoverable sugar, S is cane Pol, N is the non-sugar content entering the mill, and F is the cane fiber. This formula has been used extensively by others (Buchanan 1975; Moor 2002; Murray 2002; Le Gal et al. 2005; Wynne et al. 2009).

Four mathematical models were proposed to predict the sugar and potential yields (RPC = Spanish acronym for Potential Cane Yield) following González (1995) who performed a stepwise regression using industrial yield as a dependent variable, and a range of industrial and weather variables as independent variables. The progressive introduction of variables into the regression equation established equations showing that industrial yields were determined predominantly by Pol % in cane and % of total losses. This was confirmed using multiple factorial analyses.

The aim was to evaluate three mathematical models that would establish the dynamics among quality indicators of sugar cane to predict its sugar yield and potential yield (RPC) using chronological series of data collected during 2001-2013 period.

Model A

$$y = 0.6972 + 0.9434 \times PCI$$

PCI= % of cane Pol entering the mill.

Model B

$$y = 2.2528 + 0.8646 \times PCI - 0.1252 \times PPT$$

PPT= % total industrial losses.

Model C

$$y = 3.0582 + 0.8630 \times PCI - 0.2232 \times PPT + 0.0026 \times PPT^2$$

The results obtained through regression analysis using Statistica 6.1 were validated using the 2014 data.
RESULTS AND DISCUSSION

Agricultural models

Application of the four agricultural models indicated that Model 1 produced a regression coefficient ($R^2$) of 0.58 (Table 1). However, the $R^2$ values increased to 0.90, 0.89 and 0.90 for Models 2, 3 and 4, respectively. This indicated the importance of including ratoons yields in the relationship. This particularly suits the current structure of sugarcane harvesting in Cuba, by considering the harvest of spring-planted cane, ratoons and autumn-planted cane. Using the remaining previous spring-planted cane was not a good indicator of yield in the subsequent year, as indicated by a lower $R^2$ value (0.58). This may be because the majority of Cuban sugarcane varieties are selected for harvesting between 12 and 15 months. The remaining previous spring-planted cane was harvested after more than 15 months.

Table 1. Summary of the regression analyses of the four agricultural models considered in this study.

<table>
<thead>
<tr>
<th>Model</th>
<th>Linear relationships from regression analyses</th>
<th>$R^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>$Y = 1.18x - 13.15$</td>
<td>0.58</td>
</tr>
<tr>
<td>2</td>
<td>$Y = 0.81x + 7.52$</td>
<td>0.90</td>
</tr>
<tr>
<td>3</td>
<td>$Y = 8.16x + 7.68$</td>
<td>0.89</td>
</tr>
<tr>
<td>4</td>
<td>$Y = 0.81x + 2.06$</td>
<td>0.88</td>
</tr>
</tbody>
</table>

When Model 3 was applied to the 2013 harvest, the predicted yield was within 0.9 t/ha of the actual average yield. This model was found to better represent the planting and harvesting system used in Cuba. A possible improvement to Model 3 is the substitution of the yield of the remaining previous spring-planted cane and the ratoons with a relationship involving the total rainfall between May 15 and October 20 of each year, and utilizing the ratoon yield of the previous harvest (data not presented here). This modification could have some merit as more than 80% of the Cuban sugarcane area is rainfed.

Table 2 shows the results of using a monomodal model to estimate the yield potential of cane based on evapotranspiration. The results showed large differences between the potential yields and the actual yields, with potential yields often more than twice the actual yield. Although agricultural production is complex and could be affected by many variables, it is interesting to note that the actual average yield for the agricultural enterprise identified as ’Jesus Rabi’ was much closer to the estimated yield potential. The ‘Jesus Rabi’ enterprise is recognized locally for utilizing improved management practices.

Table 2. Potential yield determined from a monomodal model compared to actual sugarcane yields in three seasons (2009-2011).

<table>
<thead>
<tr>
<th>Agricultural enterprise</th>
<th>Potential yield (Estimated) (t cane/ha)</th>
<th>Actual yield (2009-11) (t cane/ha)</th>
<th>Number of harvests</th>
</tr>
</thead>
<tbody>
<tr>
<td>30 de Noviembre</td>
<td>172.8</td>
<td>79.5</td>
<td>6</td>
</tr>
<tr>
<td>Jesús Rabí</td>
<td>173.6</td>
<td>119.0</td>
<td>15</td>
</tr>
<tr>
<td>Primero de Enero</td>
<td>160.2</td>
<td>85.7</td>
<td>25</td>
</tr>
<tr>
<td>Brasil</td>
<td>178.1</td>
<td>104.2</td>
<td>2</td>
</tr>
<tr>
<td>Amancio Rodriguez</td>
<td>205.7</td>
<td>98.6</td>
<td>5</td>
</tr>
<tr>
<td>Argeo Martinez</td>
<td>198.6</td>
<td>65.5</td>
<td>4</td>
</tr>
<tr>
<td>Urbano Noris</td>
<td>195.6</td>
<td>87.0</td>
<td>2</td>
</tr>
</tbody>
</table>

Mill models

Estimated sugar yields (t sugar/ha) using Model A were plotted against the actual yields for sugarcane variety RB96 during the period 2001 to 2013 (Fig. 1). The regression equation resulted in an $R^2$ value of 0.86. Model B resulted in a close fit between the estimated and actual yields, except in 2002 when the model underestimated the yield (Fig. 2). This is probably due to industry inefficiency in that year that included lost time at the mill due to unexpected interruptions. Similar results were obtained for Model C (Fig. 3).
The three models were validated using the 2014 data. A summary of the regression analyses is shown in Table 3. All three models appeared to perform well with $R^2$ values of 0.87, 0.99 and 0.99 for Models A, B and C, respectively.

**Fig. 1.** Estimated sugar yields using Model A plotted against actual yields for sugarcane variety RB96.

**Fig. 2.** Estimated sugar yields using Model B against actual yields for sugarcane variety RB96.

**Fig. 3.** Estimated sugar yields using Model C against actual yields for sugarcane variety RB96.
Table 3. Validation of the models using the 2014 harvest data.

<table>
<thead>
<tr>
<th>Validation data</th>
<th>R²</th>
<th>Model 1</th>
<th>Model 2</th>
<th>Model 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Harvests 2001-2013</td>
<td>0.86</td>
<td>0.72</td>
<td>0.74</td>
<td></td>
</tr>
<tr>
<td>Harvest 2014 until April</td>
<td>0.87</td>
<td>0.99</td>
<td>0.99</td>
<td></td>
</tr>
</tbody>
</table>

CONCLUSIONS

1. Four mathematical models were evaluated for predicting total annual sugarcane yield. Models 2, 3 and 4 provided a ‘good fit’. The exception was Model 1 that was based on the remaining previous spring-planted cane. Model 3 was considered to best represent the planting and harvesting system used in Cuba.
2. The agricultural models supported the importance of including ratoon yields in the predictive capacity and the need for careful management of ratoons.
3. Three mathematical models were validated for predicting sugar yields.
4. The relatively high regression coefficients indicate that the models can be used by technical staff and the managers of the mills for predictive purposes.

Our results support further research with these models to refine the relationships.

ACKNOWLEDGEMENTS

We acknowledge the revision and integration of the agricultural and mill areas of the paper by Professor Guillermo Galvez. The collaboration of Mr. Pedro P. Acosta for supplying some of the agricultural data is also recognized.

REFERENCES

Vitores X. 2006. Especificación de un Modelo Econometrístico Espacial Bayesian para el Pronóstico de la Producción de un Cultivo de Caña de Azúcar. Ingenio San Carlos, Ecuador.
Une amélioration dans l’efficience de l’estimation des rendements aux champs et à l’usine à Cuba

Résumé. L’industrie cannière cubaine a besoin de méthodes efficaces pour estimer les rendements, tant du point de vue agricole et industriel. Une étude de cas liée à ce sujet est présentée. L’étude a utilisé les données de rendement de 2001 à 2013 provenant de 55 périmètres sucriers. Quatre outils de modélisation furent évalués et trois modèles mathématiques furent proposés pour prévoir le taux de sucre et les rendements potentiels (RPC). L’analyse statistique a révélé un taux de plus de 88% dans la variabilité des données. Les résultats obtenus grâce à l’analyse de régression par étapes ont été validés pour les usines en utilisant les données de la récolte sucrière de 2014 et les paramètres, tels que Pol% canne, fibre% canne, Brix et Pol du jus de première pression et des pertes à l’usine, ont pu être utilisés pour prévoir à la fois le rendement de l’usine et le RPC. Pour la récolte de 2013, la différence dans la moyenne entre le rendement modélisé et effectivement obtenu était de l’ordre de 0,9 t/ha. Parmi les modèles testés, le Modèle 3 est considéré comme le plus représentatif des systèmes de plantation et de récolte en cours à Cuba. Les coefficients de régression relativement élevés indiquent que les modèles peuvent être utilisés par le personnel technique et les gestionnaires des usines pour faire des prédictions.

Mots-clés: Modélisation de la culture de la cane à sucre, rendement agricole, efficience de l’usine

Mejorando la eficiencia de la estimacion de los rendimientos agrícolas e industriales en Cuba

Resumen. La agroindustria cubana de la caña de azúcar necesita de métodos eficientes para estimar los rendimientos tanto en la agricultura como en la industria. Se presenta un caso de estudio cubano sobre este tema. El estudio usó los rendimientos del periodo comprendido entre los años 2001 a 2013 en 55 ingenios y sus áreas agrícolas. Para la estimación de los rendimientos agrícolas se validaron cuatro modelos y en las fábricas se validaron 3 modelos matemáticos para predecir el azúcar y el potencial de sus rendimientos (RPC). El modelo de más perspectivas se ajustó a la actual estructura de cepas de los rendimientos agrícolas de Cuba. El análisis estadístico realizado explico más del 88 % de la varianza total de los datos. Los resultados obtenidos utilizando un análisis de regresión paso a paso fueron validados en los ingenios en la zafra del 2014 y permitió una estrategia de trabajo basada en los indicadores como Pol % en caña, % de fibra en caña, Brix, el Pol de los jugos primarios y las pérdidas de fábrica y permitió predecir el rendimiento del ingenio y el RPC. El rendimiento agrícola promedio de la zafra del 2013 difirió en 0,89 t/ha con el rendimiento real obtenido y con el modelo que se ajustó mejor a la estructura de cepas de Cuba, además la cepa de retoños fue la que mayor peso en las ecuaciones utilizadas. El modelo más ajustado para los ingenios ($R^2$ entre 0.88 y 0.99) brindó una herramienta eficiente para el staff técnico de los ingenios.

Palabras clave: Caña de azúcar, modelos, rendimiento agrícola, azúcar recuperable