

Artículo Original

**MODELO MATEMÁTICO PARA EL CONTROL DEL CONTENIDO
DE SACAROSA EN BAGAZO EN EL PROCESO
DE FABRICACIÓN DE AZÚCAR**

**MATHEMATICAL MODEL FOR CONTROL OF SUCROSE CONTENT IN
BAGASSE IN THE SUGAR MANUFACTURING PROCESS**

Jonathan Serrano Febles^{1*} <https://orcid.org/0000-0002-7434-8611>

Jesús Luis Orozco¹ <https://orcid.org/0000-0001-6484-0672>

Osney Pérez Ones² <https://orcid.org/0000-0002-0366-0317>

Julio César Peña¹ <https://orcid.org/0000-0001-7996-0167>

¹ Facultad de Ciencias Técnicas, Universidad de Matanzas, Matanzas, Cuba.

² Universidad Tecnológica de La Habana José Antonio Echeverría (CUJAE),
Facultad de Ingeniería Química, Grupo de Análisis de Procesos, La Habana, Cuba.

Recibido: Octubre 3, 2024; Revisado: Octubre 19, 2024; Aceptado: Octubre 31, 2024

RESUMEN

Introducción:

Los modelos matemáticos constituyen herramientas muy útiles para el control operacional de la etapa de extracción en el proceso de fabricación de azúcar crudo.

Objetivo:

Obtener un modelo matemático para predecir, en base a las condiciones operacionales de molienda y de imbibición que se apliquen, el contenido de sacarosa en bagazo.

Materiales y Métodos:

El modelo matemático se desarrolla mediante regresión lineal múltiple sobre la base de información experimental obtenida a nivel industrial. Se emplean diversas pruebas estadísticas para garantizar un adecuado ajuste del modelo matemático y evaluar su desempeño predictivo. El dominio de aplicación se establece a partir de análisis de “extrapolación oculta”.

Resultados y Discusión:

Este modelo matemático es válido para una razón de imbibición entre un 0,80-2,90; una



Este es un artículo de acceso abierto bajo una Licencia *Creative Commons* Atribución-No Comercial 4.0 Internacional, lo que permite copiar, distribuir, exhibir y representar la obra y hacer obras derivadas para fines no comerciales.

* Autor para la correspondencia: Jonathan Serrano, Email: jserrano@cubaron.co.cu



temperatura del agua de imbibición entre 40-80 °C y un contenido de sacarosa en el jugo primario entre 15,1-19,2 %. El modelo matemático demuestra que el contenido de sacarosa en bagazo tiene una dependencia no lineal respecto a la cantidad de imbibición aplicada y aporta nuevas evidencias estadísticas que prueban la influencia significativa de la temperatura de la imbibición sobre la extracción de sacarosa. El procedimiento que se emplea es de gran utilidad como material metodológico para el desarrollo y evaluación de modelos de regresión lineal en la investigación operacional.

Conclusiones:

Con la utilización del modelo matemático como herramienta predictiva de análisis es posible mejorar el desempeño operacional de la etapa de extracción reduciendo las pérdidas de sacarosa en bagazo.

Palabras Clave: extracción; imbibición; maceración; modelación matemática; molienda.

ABSTRACT

Introduction:

Mathematical models are very useful tools for the operational control of the extraction stage in the raw sugar manufacturing process.

Objective:

To obtain a mathematical model to predict, based on the operational conditions of milling and imbibition which are applied, the sucrose content in bagasse.

Materials and Methods:

The mathematical model is developed by means of multiple linear regression based on experimental information obtained at industrial scale. Several statistical tests are used to guarantee an adequate adjustment of the mathematical model and to evaluate its predictive performance. The application range is established using “hidden extrapolation” analysis.

Results and Discussion:

This mathematical model is valid for an imbibition ratio between 0.80-2.90; imbibition water temperature between 40-80 °C and sucrose content in primary juice between 15.1-19.2 %. The mathematical model demonstrates that the sucrose content in bagasse has a non-linear dependence on the amount of imbibition applied and provides new statistical evidence proving the significant influence of the imbibition temperature on sucrose extraction. The procedure used is of great utility as methodological material for the development and evaluation of linear regression models in operational research.

Conclusions:

Through the use of the mathematical model as a predictive analysis tool, it is possible to improve the operational performance of the extraction stage by reducing sucrose losses in bagasse.

Keywords: extraction; imbibition; mathematical modeling; maceration; milling.

1. INTRODUCCIÓN

El control operacional del proceso industrial de fabricación de azúcar de caña es extremadamente complejo debido a las numerosas etapas que lo componen. La primera de ellas es la etapa de extracción, en la cual la caña de azúcar se muele con el objetivo de extraer el jugo rico en sacarosa. De la molienda de la caña se obtiene el jugo, que se envía hacia otras etapas del proceso para obtener el azúcar crudo y el bagazo, constituyendo este último un residuo para la generación de vapor (Hugot, 2014).

En la etapa de molienda se adiciona una cantidad de agua, conocida como agua de imbibición, para lavar el bagazo y aumentar la extracción de sacarosa (Jenkins, 2013). Esta es una de las operaciones tecnológicas del proceso de fabricación más influyentes sobre la eficiencia industrial. Una cantidad de agua de imbibición insuficiente, conlleva a que el bagazo contenga mucha sacarosa, incurriéndose en serias pérdidas económicas por este concepto. De modo contrario, una cantidad excesiva supone una dilución inadecuada del jugo y representa una sobrecarga inviable a la estación de evaporación (Rein, 2012).

La regulación de la imbibición es compleja, ya que depende de las características de la caña que se esté procesando en ese momento. La cantidad de agua de imbibición adicionada normalmente se refiere a la composición de fibra de la caña, lo que da lugar al término “razón de imbibición” (Hugot, 2014). Sobre la cantidad de agua de imbibición a adicionarse influye también la temperatura del agua y la composición de sacarosa del jugo del primer molino (jugo primario). Tales elementos hacen que sea difícil y propenso a errores la toma de decisiones operacionales en la aplicación de la imbibición.

Ante esta situación se hace indispensable predecir la composición de sacarosa del bagazo según la razón de imbibición que se aplique, la temperatura del agua de imbibición y el contenido de sacarosa del jugo de caña del primer molino (jugo primario). Ello permitiría minimizar las pérdidas de sacarosa en bagazo y garantizar por este sentido un adecuado rendimiento industrial.

Este trabajo se realiza con el objetivo de obtener un modelo matemático para predecir, en base a las condiciones operacionales de molienda y de imbibición que se apliquen, el contenido de sacarosa en bagazo.

2. MATERIALES Y MÉTODOS

2.1 Concepción del modelo matemático y colección de los datos experimentales

Las variables operacionales de la imbibición consideradas para la modelación matemática fueron la cantidad y temperatura del agua añadida. La cantidad de agua de imbibición se refirió según la composición de fibra de la caña, para lo cual se empleó el término “razón de imbibición”; variable ampliamente reconocida en la literatura especializada (Rein, 2012). En la modelación matemática se incluyó como variable independiente al contenido de sacarosa del jugo primario para considerar el efecto de la riqueza en azúcar que trae la caña sobre el contenido de sacarosa del bagazo. Un mayor contenido en sacarosa en la caña, conlleva para iguales condiciones de operación, a un mayor contenido de sacarosa en bagazo.

La técnica que se empleó para el desarrollo del modelo matemático fue la regresión

lineal múltiple para correlacionar las siguientes variables a partir de la estructura mostrada en la ecuación 1.

$$polBgz = f(\lambda, T, polJP) \quad (1)$$

Donde:

polBgz: Contenido de sacarosa en bagazo (%).

λ : Razón de imbibición.

T: Temperatura de imbibición (°C).

polJP: Contenido de sacarosa en el Jugo Primario (%).

La razón de imbibición se determinó mediante la ecuación 2 (Hugot, 2014):

$$\lambda = \frac{AI}{CM \cdot fCM} \cdot 100 \quad (2)$$

Donde:

AI: Flujo de agua de imbibición (kg/h).

CM: Flujo de caña molida (kg/h).

fCM: Composición de la fibra en la caña (%).

La experimentación para la colección de los datos empleados (106 observaciones) en la modelación matemática se realizó en un central azucarero (denominado A) durante tres meses de la zafra 2020-2021. De igual manera, se recopiló información experimental en otro central azucarero (denominado B) con el objetivo de validar la capacidad predictiva del modelo matemático en un central azucarero diferente al cual se obtienen los datos para su desarrollo. Estos centrales tienen una capacidad de molida de 145 000 kg/h de caña y disponen de un tándem de cinco molinos *Hamilton*, con presiones de trabajo en los molinos de 3 100 psia (21,38 MPa) en donde se practica la imbibición compuesta a la entrada del quinto molino.

Durante los experimentos, el flujo de caña, la composición de fibra en caña y el contenido de sacarosa del jugo primario quedaron condicionados por el ritmo de molienda y las características de la caña que se procesó (variables no manipuladas durante la experimentación). La temperatura y cantidad de agua de imbibición se fijaron durante la operación de la molienda. El central azucarero garantizó a través de su tecnología que se mantuvieran constantes parámetros influyentes sobre las mediciones como: las presiones de los molinos, contenido de materia extraña y el índice de preparación de la caña.

2.2 Desarrollo de la modelación matemática. Procesamiento estadístico

El ajuste de los datos a un modelo matemático se efectuó en RStudio (versión 1.3.959).

Para la selección de un modelo entre varias propuestas se empleó el valor del error cuadrático medio (RMSE), el valor del coeficiente de significación ajustado, el criterio de información de Akaike (AIC) y el de información bayesiano (BIC) mediante el paquete “performance” (Lüdecke y col., 2021). El valor de inflación de la varianza (VIF, por sus siglas en inglés) se determinó a través del paquete “car” (Fox y Sanford, 2019). En el análisis de la capacidad explicativa del modelo seleccionado a los datos experimentales se empleó la prueba de pérdida de calidad de ajuste de Fisher

(Montgomery y col., 2002).

Se analizó el cumplimiento del supuesto de normalidad de los residuales a través de la prueba de Kolmogorov-Smirnov con la modificación de Lilliefors (Razali y Wah, 2011) mediante el paquete “nortest” (Gross y Ligges, 2015). El supuesto de homocedasticidad en los residuos (varianza constante) se realizó mediante la prueba de Breusch-Pagan (Fox y Sanford, 2019) a través del paquete “lmtest” mientras que el supuesto de independencia de los residuales se analizó mediante la prueba de Durbin-Watson (Hassan y col., 2019) con el paquete “car”.

La validación cruzada k-folds se realizó con el fin de reducir los sesgos de muestreo aleatorio y el problema de sobreajuste en el modelo matemático (Shah y col., 2021). Este es un método estadístico para la evaluación del comportamiento predictivo de modelos matemático más exhaustivo y fiable que la simple división de los datos en dos conjuntos: uno de entrenamiento y otro de validación (método tradicional) (Gkioulekas y Papageorgiou, 2019). Para este caso en la validación cruzada por “k-folds”, se estableció un número de pliegues (folds) igual a 10 (Shah y col., 2021). Se empleó como métrica para evaluar el desempeño predictivo del modelo al error cuadrático medio (RMSE) mediante el paquete DAAG (Maindonald y col., 2019) tal como se aplica en Hodeghatta y Nayak, (2016).

2.2.1 Definición del intervalo de aplicación del modelo matemático y detección de “extrapolación oculta”

Los valores máximos y mínimos de cada variable restringen el intervalo de aplicación de los modelos de regresión, pero solo en primera instancia. Ello no supone que dentro de tales límites se haya dispuesto de información suficiente para contemplar efectos entre diferentes niveles de los términos independientes y se puede dar lugar entonces a predicciones erróneas por “extrapolación oculta” (Montgomery y col., 2002).

Por ello se analizó la posible inclusión de las nuevas observaciones dentro de la región envolvente o cáscara de variables regresoras (RVH, siglas en inglés de *Regressor Variable Hull*) (Montgomery y col., 2002). Se tiene que se está en presencia de extrapolación oculta si el punto para la predicción está fuera de la RVH (Montgomery y col., 2002), tal que:

$$h_o \geq h_{m\acute{a}x} \quad (3)$$

Siendo:

$$h_o = x^{-1}(X'X)^{-1}x \quad (4)$$

Donde:

X : Matriz de los predictores que se emplea en el desarrollo del modelo de regresión (en cuya primera columna todos sus términos son 1 para la estimación del coeficiente del intercepto).

x : Punto constituido para la nueva predicción con coordenadas $1, x_1, x_2, \dots, x_k$; donde k es el número de variables independientes del modelo según el orden de estos en la matriz X .

$h_{m\acute{a}x}$: Valor máximo de los elementos de la diagonal de la matriz sombrero o como también se conoce, matriz H (del inglés “hat”).

2.2.2 Validación de la capacidad predictiva del modelo matemático con datos del central azucarero B

Se empleó como criterio para la validación del modelo matemático la prueba de bondad de ajuste de Fisher. Tienen que cumplirse para la validez de esta prueba igualmente con los supuestos de normalidad, homocedasticidad e independencia de los residuales (Montgomery y col., 2002).

3. RESULTADOS Y DISCUSIÓN

3.1 Procesamiento estadístico de los datos experimentales y desarrollo de la regresión lineal

Se disponen de 106 observaciones de las variables operacionales del central azucarero A y de 45 observaciones del central azucarero B. La Figura 1 muestra que el contenido de sacarosa en bagazo presenta una tendencia poco lineal con la razón de imbibición.

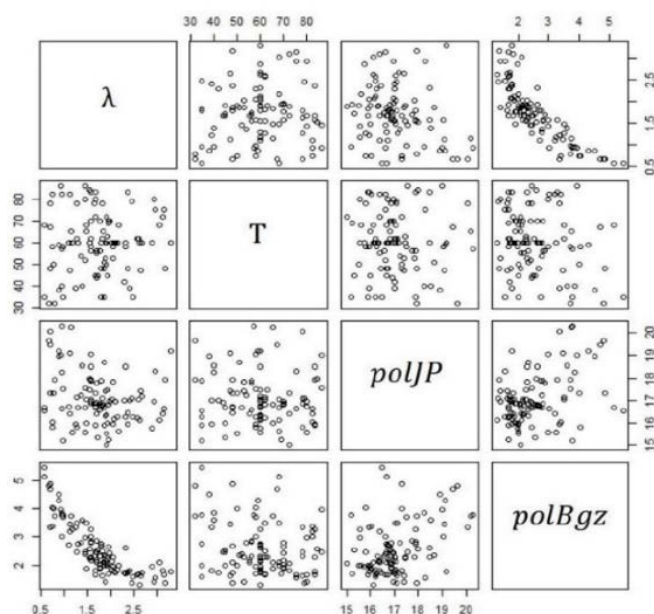


Figura 1. Disposición gráfica del contenido de sacarosa del bagazo ($polBgz$) según la razón de imbibición (λ), temperatura del agua de imbibición (T) y contenido de sacarosa del jugo primario ($polJP$)

El comportamiento mostrado en la Figura 1 es totalmente lógico, ya que el contenido de sacarosa en bagazo es dependiente e inversamente proporcional a la extracción de sacarosa en el jugo mezclado (Hugot, 2014). La extracción de sacarosa en el jugo mezclado de acuerdo a Hugot (2014) y Rein (2012) aumenta rápidamente a razones de imbibición entre 0-1, sigue aumentando aunque más lentamente entre 2-3; mientras que por encima de 3 no se verifican tasas de extracción significativas.

En el análisis de la Figura 1 no se detectan patrones definidos para la temperatura del agua de imbibición y el contenido de sacarosa en el jugo primario. Ello se debe a que los cambios en la composición de sacarosa en bagazo por estas variables pueden estar enmascarados por el efecto de la razón de imbibición; ya que esta última variable es la más influyente (Hugot, 2014). Favorablemente, no se detecta ninguna interrelación entre las variables independientes, lo que anticipa la posible no existencia de multicolinealidad entre los futuros predictores.

Dado el comportamiento poco lineal entre la razón de imbibición y el contenido de sacarosa en bagazo, se evalúan diferentes alternativas de estructuras matemáticas correspondientes a la linealización de la variable dependiente y/o de la relación de imbibición como término independiente, a través de transformaciones logarítmicas, recíprocas, inversa y polinómica de segundo orden.

En todos los casos evaluados, resultan significativos los predictores y el modelo de regresión de acuerdo al ANOVA de la regresión lineal, ya que el P-valor de cada uno de ellos es menor que el nivel de significación para una confianza del 95 %. El coeficiente de determinación ajustado para todas ellas fue mayor a 0,70; indicando que se explica más del 70 % de la variabilidad total observada. Por este motivo se selecciona la variante que presente menor valor de RMSE (criterio principal) y menor valor de AIC y BIC (criterio secundario).

Es deseable que el RMSE sea lo más cercano a cero posible, ya que permite inferir por este sentido un mejor ajuste (Gharekhani y col., 2021). Este parámetro se emplea como criterio fundamental en la selección de alternativas de modelos en numerosas investigaciones (Ayetigbo y col., 2021, Güven y Şimşir, 2020, Yang y col., 2020). El criterio de información de Akaike y el de información bayesiano permiten analizar la predicción en base a la complejidad del modelo matemático, para lo cual penalizan el incremento del número de parámetros. El mejor modelo en base a estos criterios es el que presente menor valor de AIC y de BIC (Mullor y col., 2019). De esta forma resulta como modelo matemático final (ecuación 5):

$$polBgz = EXP(0,691 - 0,683 \cdot \ln(\lambda) - 0,002 \cdot T + 0,039 \cdot polJP) \quad (5)$$

Se demuestra que no existe multicolinealidad entre los predictores, al presentarse en cada caso un bajo coeficiente de correlación. Ello se comprueba con el valor de inflación de la varianza, ya que es inferior a cinco (Montgomery y Runger, 2018). A su vez, mediante la distancia de Cook se comprueba la inexistencia de valores atípicos influyentes, dado que en todos los casos esta es inferior a 1 (Liu y Smith, 2020, Maindonald y Braun, 2010).

Como valor del estadígrafo de Fisher calculado (3,280) es menor al valor del estadígrafo de Fisher tabulado (4,406) no existen evidencias estadísticas de pérdidas de calidad de ajuste por errores de la predicción. Ello indica que no existen diferencias significativas entre los estimados del error “puro” o experimental con los del error por pérdida de ajuste. De esta forma, se prueba el correcto ajuste del modelo matemático a los datos del central A.

Montgomery y col., (2002) plantean que por lo general, no es probable que el modelo sea desde un punto de vista predictivo satisfactorio, a menos que el rango de los valores ajustados sea grande en relación con su error estándar promedio estimado. Para este caso el rango de los valores ajustados es de 1,278 y el error estándar promedio estimado es de 0,031; siendo dicho rango 41 veces mayor.

Para que el modelo matemático tenga validez, deben satisfacerse los supuestos de normalidad, homocedasticidad e independencia de los residuales que se establecen al aplicar la técnica de regresión lineal (Montgomery y col., 2002). El P-valor de la prueba de Lilliefors (0,730) es mayor que el nivel de significación, por lo que se infiere que los

residuales presentan una distribución normal (Priego y col., 2018). El modelo es homocedástico, ya que el P-valor de la prueba de Breusch-Pagan (0,141) es mayor que el nivel de significación (Hernández-Chover y col., 2019). Los errores no se encuentran correlacionados entre sí, dado que el P-valor de la prueba de Durbin-Watson (0,494) es mayor que el nivel de significación (0,05).

3.1.1 Validación cruzada por k-folds

La Figura 2 muestra los resultados de las predicciones según los pliegues (folds) en la validación cruzada. Se aprecia una adecuada correspondencia entre los valores predichos a partir del modelo ajustado a todos los datos (forma linealizada de la ecuación 5) y las predicciones obtenidas en las diferentes iteraciones del algoritmo de validación. De hecho, el error cuadrático medio derivado de la validación cruzada es de 0,151; valor que difiere en tan solo 3,36 % al que se verifica con todos los datos (0,146). Tal resultado valida la robustez del modelo matemático por este concepto y demuestra que no existen problemas de sobreajuste.

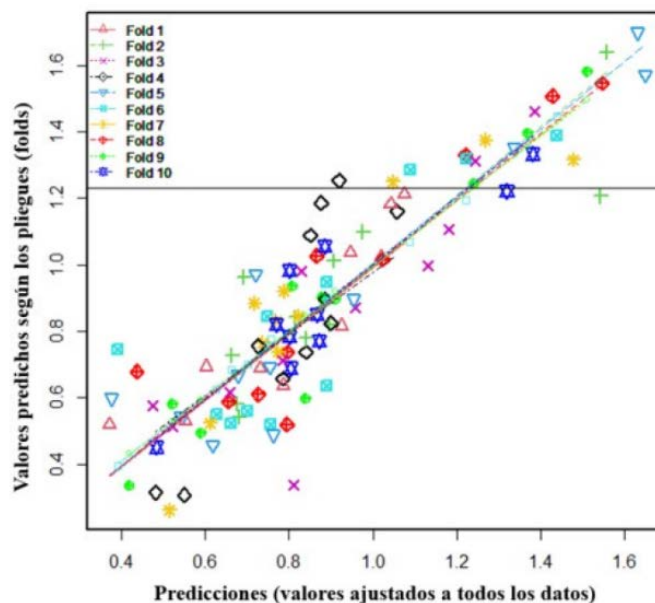


Figura 2. Predicciones de “ $\ln(\frac{polBgz}{100})$ ” (variable dependiente linealizada) mediante validación cruzada por k-folds

3.1.2 Definición de los intervalos de aplicación del modelo y análisis de extrapolación oculta

Para la determinación de los intervalos de las variables independientes en los cuales el modelo es válido, se realiza primeramente una evaluación de los intervalos definidos por los valores máximos y mínimos observados de cada una de ellas. Para ello, se establece un mallado a partir de la combinación de valores de cada variable independiente, que están igualmente espaciados y que se encuentran dentro de los intervalos anteriores. A partir estos se determina el valor de h_o en cada caso y se comparan con $h_{m\acute{a}x}$; de lo que resulta la existencia de puntos interiores que satisfacen la ecuación 3 y que por lo tanto, invalidan el intervalo de aplicación del modelo inicial por concepto de “extrapolación oculta”. Por ello, se procede a acotar más el intervalo hasta que no se logre no satisfacer dicha condición, obteniéndose finalmente que no existen

evidencias de extrapolación cuando:

$$0,80 \leq \lambda \leq 2,90;$$

$$40 \leq T \leq 80 \text{ y}$$

$$15,01 \leq polJP \leq 19,24.$$

Estos nuevos intervalos no interfieren en la aplicabilidad práctica al modelo de regresión, ya que contemplan a los valores típicos que presentan estas variables en el proceso industrial y a una desviación moderada de estos (Rein, 2012, Jenkins, 2013).

3.1.3 Validación del modelo matemático con los datos experimentales del central B

A través de la prueba de bondad de ajuste de Fisher se demuestra que el modelo es capaz de explicar satisfactoriamente los datos recopilados en este central azucarero; ya que el valor del estadígrafo de Fisher calculado (3,634) es menor al tabulado (4,469).

3.2 Valor práctico del modelo matemático

El modelo matemático desarrollado resulta de gran utilidad para la reducción de las pérdidas de sacarosa en bagazo. Mediante este, es posible determinar según la composición de sacarosa del jugo primario, los valores más adecuados de las variables operacionales de la imbibición (razón de imbibición y temperatura) en el proceso productivo para disminuir el porcentaje de sacarosa del bagazo.

Ello es de gran importancia dada las características cambiantes en el tiempo de la caña, como el flujo y su composición de fibra. Tal dinamismo hace que sea necesario con mucha frecuencia reajustar la razón de imbibición para mantener en valores adecuadamente bajos el contenido de sacarosa del bagazo; para lo cual este modelo sirve de guía. De modo contrario, al no disponerse de herramientas de análisis de este tipo, esta decisión operacional se realizaría arbitrariamente y sería más propensa a errores con graves pérdidas económicas.

Como ejemplo de su aplicación se tiene dada su predicción que para un ritmo de molienda de 150 000 kg/h, con una caña que posee un contenido de fibra del 16 % y una composición de sacarosa en el jugo de 16,5 %; debe aplicarse una razón de imbibición de 2,66 si la temperatura del agua disponible es de 40 °C para garantizar que el valor de sacarosa del bagazo sea de 1,80 % (norma operacional). Esta razón de imbibición (ecuación 2) implica adicionar 65 835 kg/h de agua (aproximadamente 66 m³/h).

A su vez, se corrobora la significación de la temperatura del agua de imbibición sobre el contenido de sacarosa del bagazo a través del ANOVA de la regresión. Algunos tecnólogos aluden que tal significación es menor y que desde un punto de vista práctico es despreciable (Batule, 2009), pese a que algunas investigaciones realizadas al respecto demuestran todo lo contrario (Kent, 2011, Serrano y col., 2023). Muchos de los criterios que se refieren acerca del efecto operacional de la temperatura del agua de imbibición parten de conocimientos empíricos y no del procesamiento matemático de datos industriales. Precisamente otro aporte de este trabajo es brindar evidencias desde un punto de vista estadístico sobre la influencia significativa de la temperatura de la imbibición sobre el contenido de sacarosa del bagazo; lo que corrobora la importancia de su control operacional para minimizar las pérdidas de sacarosa.

De hecho, para las condiciones de molienda anteriores y la razón de imbibición calculada ($\lambda=2,66$); el contenido de sacarosa en bagazo puede reducirse en un 7,7 % si

la temperatura del agua de imbibición cambia de 40 a 80 °C de acuerdo al modelo matemático. O en igual sentido, para alcanzar un mismo contenido de sacarosa en bagazo de 1,80 % se debería aplicar una razón de imbibición menor ($\lambda=2,36$) si se aplicase la imbibición a 80 °C en lugar de 40 °C. Ello implica un ahorro de agua de imbibición de 7,4 m³/h y una reducción significativa del consumo de energía térmica en el proceso.

4. CONCLUSIONES

En este artículo se presenta una metodología de modelación estadística rigurosa, de gran valor como referente bibliográfico en el desarrollo y evaluación de modelos matemáticos de regresión lineal. El modelo matemático obtenido representa una herramienta de análisis útil para un mejor control de las pérdidas de sacarosa en bagazo en el proceso de fabricación de azúcar crudo. Este modelo matemático es válido para una razón de imbibición entre un 0,80-2,90; una temperatura del agua de imbibición entre 40-80 °C y un contenido de sacarosa en el jugo primario entre 15,1-19,2 %.

REFERENCIAS

- Ayetigbo, O., Latif, S., Abass, A., & Müller, J., Dataset on influence of drying variables on properties of cassava foam produced from white- and yellow-fleshed cassava varieties., *Data in Brief*, Vol. 37, No. 107192, 2021, pp. 1-14. <https://doi.org/10.1016/j.dib.2021.107192>
- Batule, E., La molienda y difusión de la caña de azúcar., *Serie azucarera* 20, San Salvador (El Salvador), 2009, pp. 17-234.
- Fox, J., & Sanford, W., *An R Companion to Applied Regression.*, SAGE, California, 2019, pp. 582-597. <https://www.john-fox.ca/Companion>
- Gharekhani, M., Nadiri, A.A., Khatibi, R., & Sadeghfam, S., An investigation into time-variant subsidence potentials using inclusive multiple modelling strategies., *Journal of Environmental Management*, Vol. 294, 2021, 112949. <https://doi.org/10.1016/j.jenvman.2021.112949>
- Gkioulekas, I., & Papageorgiou, L.G., Piecewise regression analysis through information criteria using mathematical programming., *Expert Systems with Applications*, Vol. 121, 2019, pp. 362-372. <https://doi.org/10.1016/j.eswa.2018.12.013>
- Gross, J., & Ligges, U., Package "nortest". Paquete de RStudio (Programa informático): CRAN, 2015. <https://cran.r-project.org/web/packages/nortest/index.html>
- Güven, İ., & Şimşir, F., Demand forecasting with color parameter in retail apparel industry using artificial neural networks (ANN) and support vector machines (SVM) methods., *Computers & Industrial Engineering*, Vol. 147, No. 106678, 2020, pp. 3-9. <https://doi.org/10.1016/j.cie.2020.106678>
- Hassan, S., Aziz, H., Zakariah, Z., Noor, S., Majid, M., & Karim, N., Development of linear regression model for brick waste generation in Malaysian construction industry., *Journal of Physics: Conference Series*, Vol. 1349, No. 1, 2019, pp. 3-12. <https://iopscience.iop.org/article/10.1088/1742-6596/1349/1/012112>
- Hernández-Chover, V., Castellet-Viciano, L., & Hernández-Sancho, F., Cost analysis of
-

- the facilities deterioration in wastewater treatment plants: A dynamic approach., *Sustainable Cities and Society*, Vol. 49, 2019, 101613. <https://doi.org/10.1016/j.scs.2019.101613>
- Hodeghatta, U.R., & Nayak, U., *Business analytics using R-a practical approach.*, Springer, Berlín, 2016. <https://www.springerprofessional.de/en/business-analytics-using-r-a-practical-approach/11961462>
- Hugot, E., *Handbook of Cane Sugar Engineering.*, Elsevier, 2014, pp. 223-245. <https://doi.org/10.1016/B978-1-4832-3190-7.50028-3>
- Jenkins, G.H., *Introduction to Cane Sugar Technology.*, Elsevier, 2013, pp. 141-158. <https://doi.org/10.1016/B978-1-4832-3195-2.50016-6>
- Kent, G.A., The effect of added water temperature on milling train operation and performance., *Sugar Tech*, Vol. 13, No. 1, 2011, pp. 1-6. <https://doi.org/10.1007/s12355-011-0062-8>
- Liu, J., & Smith, S.R., A multi-level biogas model to optimise the energy balance of full-scale sewage sludge conventional and THP anaerobic digestion., *Renewable Energy*, Vol. 159, No, 2020, pp. 756-766. <https://doi.org/10.1016/j.renene.2020.06.029>
- Lüdecke, D., Makowski, D., Ben-Shachar, M., Waggoner, P., Wiernik, B.M., Arel-Bundock, V., & Jullum, M., *Package: Performance (Assessment of Regression Models Performance).*, Paquete de R (Programa informático), 2021. <https://easystats.github.io/performance>
- Maindonald, J., & Braun, J., *Data Analysis and Graphics Using R – an Example-Based Approach.*, Cambridge University Press, Cambridge (Reino Unido), 2010, pp. 514-565. <https://www.cambridge.org/core/books/data-analysis-and-graphics-using-r/E04AEC5BCEF09D2E51A63EB5A8CB0680>
- Maindonald, J., Braun, J., & DAAG, J., *Data Analysis and Graphics Data and Functions.*, Paquete de R (Programa informático), 2019. <https://cran.r-project.org/web/packages/DAAG/index.html>
- Montgomery, D., Peck, E., & Vining, G., *Introducción al análisis de regresión lineal.*, Compañía Editorial Continental, México, 2002, pp. 45-374.
- Montgomery, D., & Runger, G., *Applied Statistics and Probability for Engineers.*, Wiley, Estados Unidos, 2018, pp. 15-413. <https://www.wiley.com/en-us/Applied+Statistics+and+Probability+for+Engineers%2C+7th+Edition-p-9781119400363>
- Mullor, R., Mulero, J., & Trottini, M., A modelling approach to optimal imperfect maintenance of repairable equipment with multiple failure modes., *Computers & Industrial Engineering*, Vol. 128, No, 2019, pp. 24-31. <https://doi.org/10.1016/j.cie.2018.12.032>
- Priego, J., Palmer, R.S., Pérez-Soriano, P., Izaguirre, J., & De Anda, R., *Multi Regression Analysis of Skin Temperature Variation During Cycling Exercise.*, Springer International Publishing, VipIMAGE 2017, 2018 Cham., pp. 962-969. https://link.springer.com/chapter/10.1007/978-3-319-68195-5_105
- Razali, N.M., & Wah, Y.B., Power comparisons of Shapiro-Wilk, Kolmogorov-Smirnov, Lilliefors and Anderson-Darling tests., *Journal of statistical modeling and analytics*, Vol. 2, No. 1, 2011, pp. 21-33.
-

<https://www.nrc.gov/docs/ml1714/ml17143a100.pdf>

Rein, P., Ingeniería de la caña de azúcar., Bartens, Berlín, 2012, pp. 35-211.

Serrano, J., Orozco, J.L., García, A., Dueñas, J., León, M.L., & Herrera, Z., Influence of imbibition temperature on the extraction stage of the cane sugar production., Journal of Food Engineering, Vol. 345, No. 111414, 2023, pp. 3-14.

<https://doi.org/10.1016/j.jfoodeng.2023.111414>

Shah, M.I., Javed, M.F., & Abunama, T., Proposed formulation of surface water quality and modelling using gene expression, machine learning, and regression techniques., Environmental Science and Pollution Research, Vol. 28, No. 11, 2021, pp. 13202-13220.

<https://doi.org/10.1007/s11356-020-11490-9>

Yang, Y., He, M., & Li, L., Power consumption estimation for mask image projection stereolithography additive manufacturing using machine learning based approach., Journal of Cleaner Production, Vol. 251, No. 3, 2020, pp. 119-124.

<https://doi.org/10.1016/j.jclepro.2019.119710>

CONFLICTO DE INTERÉS

Los autores declaran que no existen conflictos de interés.

CONTRIBUCIONES DE LOS AUTORES

- Dr.C. Jonathan Serrano Febles. Investigación, análisis formal, redacción - revisión y edición.
 - Dr.C. Jesús Luis Orozco. Metodología, análisis formal, supervisión.
 - Dr.C. Osney Pérez Ones. Metodología, análisis formal, supervisión.
 - Ing. Julio César Peña. Investigación, redacción - primera redacción.
-