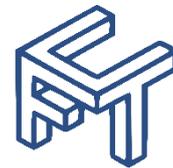


**Universidad de Matanzas**

**Facultad de Ciencias Técnicas**

**Ingeniería Informática**



**Título:** Modelos matemáticos en la etapa de extracción para el proceso de fabricación de azúcar crudo.

Trabajo de Diploma en opción al título de Ingeniero Informático

**Autor:** Héctor Elias Cruz

**Tutor:** Msc. Jonathan Serrano Febles.

Ing. Eduardo J. Berrio Turiño.

Noviembre, 2022

El futuro de nuestra Patria  
tiene que ser necesariamente  
un futuro de hombres de  
ciencias.

Fidel Castro Ruz

## **Dedicatoria**

El presente trabajo investigativo lo dedico principalmente a Esther Beatriz Norda Castro, por darme fuerzas para continuar en este proceso de obtener uno de mis anhelos y apoyarme incondicionalmente.

A mis padres, por su amor, trabajo y sacrificio en todos estos años, gracias a ustedes he logrado llegar hasta aquí.

A mi familia y amigos por el apoyo moral, que me brindaron a lo largo de esta etapa.

A todas las personas que contribuyeron a que este trabajo se realice con éxito, en especial a mis tutores que compartieron sus conocimientos conmigo.

## **Agradecimientos:**

A mis tutores: Jonathan Serrano Febles y Eduardo J. Berrio Turiño. Sin ustedes y sus virtudes, su paciencia y constancia este trabajo no lo hubiese logrado tan fácil. Sus consejos fueron siempre útiles cuando no salían de mi pensamiento las ideas para escribir lo que hoy he logrado. Ustedes formaron parte importante de esta historia con sus aportes profesionales que lo caracterizan. Muchas gracias por sus múltiples palabras de aliento, cuando más las necesite; por estar allí cuando mis horas de trabajo se hacían confusas. Gracias por sus orientaciones.

A todos los docentes de la Universidad de Matanzas, que de una forma u otra incidieron sobre mi formación. Sus palabras fueron sabias, sus conocimientos rigurosos y precisos, a ustedes les debo mis conocimientos. Donde quiera que vaya, los llevaré conmigo en mi transitar profesional. Su semilla de conocimientos germinó en el alma y el espíritu. Gracias por su paciencia, por compartir sus conocimientos de manera profesional e invaluable, por su dedicación perseverancia y tolerancia.

A Esther Beatriz Norda Castro, quien siempre estuvo a mi lado en los días y noches más difíciles durante mis horas de estudio. Orgulloso de haberte elegido mi compañera de vida y que estés a mi lado en este momento tan importante.

A mis padres y mi familia que siempre han sido mis mejores guías de vida. Hoy cuando concluyo mis estudios, les dedico a ustedes este logro, como una meta más conquistada. Gracias por ser quienes son y por creer en mí.

A mis amigos y compañeros de viaje, hoy culminan esta maravillosa aventura y no puedo dejar de recordar cuantas tardes y horas de trabajo nos juntamos a lo largo de nuestra formación. Hoy nos toca cerrar un capítulo maravilloso en esta historia de vida y no puedo dejar de agradecerles por su apoyo y constancia. Gracias por estar siempre.

## Declaración de autoría

Yo, Héctor Elías Cruz, declaro que soy el único autor de este trabajo y autorizo a la Universidad de Matanzas especialmente a la Facultad de Ciencias Técnicas, a que hagan el uso que estimen pertinente de él.



---

Firma del autor

Héctor Elías Cruz



---

Firma del tutor

Ing. Eduardo J. Berrio Turiño



---

Firma del tutor

Msc. Jonathan Serrano Febles

## **Opinión del tutor del Trabajo de Diploma.**

### **DATOS PERSONALES DEL TUTOR**

**Nombre y apellidos:** Eduardo Javier Berrio Turiño.

**Centro de trabajo:** Universidad de Matanzas.

**Organismo a que pertenece:** Ministerio de Educación Superior – MES.

**Cargo que ocupa:** Profesor e Investigador.

**Especialidad de la que es graduado:** Ingeniero Informática. Universidad de Matanzas, 2018.

**Categoría docente o investigativa:** Asistente.

### **DATOS DE LA TESIS Y EL DIPLOMANTE**

**Nombre y apellidos:** Héctor Elias Cruz.

**Centro de estudio:** Universidad de Matanzas sede “Camilo Cienfuegos”.

**Título de la Tesis:** Modelos matemáticos en la etapa de extracción para el proceso de fabricación de azúcar crudo.

### **OPINION SOBRE EL TRABAJO**

La tesis presentada posee gran actualidad, pues intenta resolver un problema real, y de gran importancia. El tutor de este trabajo de diploma considera que, durante su ejecución, el estudiante mostró las cualidades que a continuación se detallan:

Independencia y capacidad de investigación. Fueron jornadas adentrándose en temas complejos y nuevos que profundizan en gran medida los recibidos durante la carrera, logró captar con rapidez y profesionalidad el conocimiento necesario para enfrentar el problema planteado. Fue consecuente con los aspectos tanto metodológicos como de la investigación científica propiamente. Esto le permitió una feliz culminación del método desarrollado, de la documentación y de las pruebas realizadas.

En el trabajo se aprecia rigor, manifestado desde el tratamiento de los conceptos estudiados y referenciados en la bibliografía, hasta las conclusiones, lo que ha contribuido a la correcta solución de los problemas encontrados.

Una buena planificación de tiempo y recursos, dieron una gran calidad al trabajo obtenido. También fueron horas de revisión, discusión y consenso en las que demostró notables cualidades para la

investigación. El trabajo que hoy presenta y que sintetiza un periodo de aprendizaje no solo académico.

Como resultado se derivó en la obtención de un producto al nivel de las exigencias y expectativas. Por todo lo anteriormente señalado, considero que la estudiante Héctor Elías Cruz reúne los requisitos para el título de Ingeniero Informático y espero le sea otorgada la mejor calificación de este Tribunal.



---

Ing. Eduardo J. Berrio Turiño

Dpto. Informática

Universidad de Matanzas

Noviembre/2022

**Resumen:**

La presente investigación corresponde al desarrollo de modelos predictivos para su uso en la toma de decisiones operacionales en la etapa de extracción del Central Azucarero Industrial (CAI): "Jesús Rabí" de la provincia de Matanzas. Se realizaron observaciones a las variables de interés en un período muestreado de la zafra 2020-2021. Se obtiene por regresión lineal múltiple un modelo matemático para predecir el contenido de humedad a partir de la cantidad de imbibición según el peso de la fibra en caña ( $\lambda$ ) y la temperatura del agua de imbibición; así como otros dos que permiten estimar el brix del bagazo y del jugo mezclado, también bajo la consideración de estas variables independientes y del brix del jugo primario. Estos presentan una correcta explicación de la realidad observada denotado principalmente por los resultados de la prueba de bondad de ajuste de Fisher y describen adecuadamente la realidad simulada en un caso de estudio. De este último análisis se establece que no resulta conveniente desde un punto de vista extractivo aplicar niveles de  $\lambda$  superiores a 2,4 ni se justifica el empleo de elevadas temperaturas. La modelación matemática es totalmente aplicable para valores de  $\lambda$  de 0,80 a 2,90; de temperatura del agua de imbibición entre 40 y 86 °C y de contenido de sólidos disueltos del jugo primario entre 17,6 a 23,0 °Bx. Los gráficos correspondientes a la representación de los modelos predictivos facilitan el empleo práctico de la modelación matemática para la toma de decisiones operacionales.

Palabras claves: modelos matemáticos, modelos predictivos, informatización del CAI Jesús Rabí.

**Abstract:**

This research corresponds to the development of predictive models for use in making operational decisions in the extraction stage of the Industrial Sugar Mill (CAI): "Jesús Rabí" from the province of Matanzas. Observations were made to the variables of interest in a sampled period of the 2020-2021 harvest. A mathematical model is obtained by multiple linear regression to predict the moisture content from the amount of imbibition according to the weight of the cane fiber ( $\lambda$ ) and the temperature of the imbibition water; as well as two others that allow estimating the brix of the bagasse and the mixed juice, also considering these independent variables and the brix of the primary juice. These present a correct explanation of the observed reality, denoted mainly by the results of Fisher's goodness-of-fit test, and adequately describe the simulated reality in a case study. From this last analysis it is established that it is not convenient from an extractive point of view to apply levels of  $\lambda$  higher than 2.4 nor is the use of high temperatures justified. The mathematical modeling is fully applicable for values of  $\lambda$  from 0.80 to 2.90; imbibition water temperature between 40 and 86 oC and dissolved solids content of the primary juice between 17.6 and 23.0 oBx. The graphs corresponding to the representation of the predictive models facilitate the practical use of mathematical modeling for operational decision making.

Keywords: mathematical models, predictive models, computerization of the CAI Jesús Rabí.

## Índice General:

### Contents

Introducción: .....	14
Capítulo 1: Marco Teórico - Referencial.....	18
<b>1.1. Descripción de la organización donde se trabajó.....</b>	<b>18</b>
<b>1.2. Generalidades del Proceso de Producción de azúcar crudo.....</b>	<b>18</b>
<b>1.3. Etapa de extracción del proceso de producción de azúcar crudo.....</b>	<b>20</b>
<b>1.4. Normalización de la etapa de extracción de azúcar crudo.....</b>	<b>21</b>
<b>1.5. Modelación Matemática de la etapa de extracción en un tren de molinos.....</b>	<b>22</b>
<b>1.5.1 Modelos fenomenológicos aplicados al proceso de extracción.....</b>	<b>23</b>
<b>1.5.2 Modelos de caja negra aplicados al proceso de extracción .....</b>	<b>25</b>
<b>1.6. Minería de datos.....</b>	<b>26</b>
<b>1.6.1 Técnicas de minería de datos.....</b>	<b>30</b>
<b>1.7. Análisis Predictivos .....</b>	<b>30</b>
<b>1.8. Tendencias Tecnológicas.....</b>	<b>31</b>
<b>1.8.1. Lenguaje de Programación .....</b>	<b>31</b>
<b>1.8.2. Extensiones y paquetes.....</b>	<b>32</b>
<b>1.8.3. Entorno de Desarrollo Integrado (IDE).....</b>	<b>32</b>
<b>Conclusiones parciales del capítulo .....</b>	<b>33</b>
Capítulo 2: Descripción del procedimiento propuesto. ....	34
<b>2.1 Caracterización de la etapa de extracción en el CAI Jesús Rabí.....</b>	<b>34</b>
<b>2.2 Modelación del comportamiento de la etapa de extracción del proceso de producción de azúcar crudo .....</b>	<b>35</b>
<b>2.2.1 Consideraciones generales .....</b>	<b>35</b>
<b>2.2.2 Desarrollo de la experimentación.....</b>	<b>36</b>
<b>2.2.3 Procesamiento estadístico y ajuste matemático de los resultados experimentales</b>	<b>37</b>
<b>2.2.4 Definición del intervalo de aplicación de los modelos de regresión.....</b>	<b>42</b>
<b>2.2.5 Análisis de sensibilidad .....</b>	<b>44</b>
<b>2.3 Análisis de la etapa de extracción a partir de los modelos de regresión.....</b>	<b>44</b>
<b>2.4 Desarrollo de gráficos para el manejo operacional de la imbibición a partir de los modelos de regresión .....</b>	<b>46</b>
Capítulo 3: Análisis de Resultados.....	48
<b>3.1 Resultados experimentales .....</b>	<b>48</b>

<b>3.2 Análisis de resultados de la modelación estadística</b> .....	51
<b>3.2.1 Aplicación de la técnica de regresión lineal múltiple para el ajuste de los datos experimentales.</b> .....	51
<b>3.2.2 Análisis del cumplimiento de los supuestos de la regresión lineal múltiple.</b> .....	60
<b>3.2.3 Análisis de la calidad de ajuste de los modelos de regresión</b> .....	68
<b>3.2.4 Definición del intervalo de aplicación de los modelos de regresión</b> .....	69
<b>3.2.5 Análisis de sensibilidad</b> .....	71
<b>3.3 Análisis de la etapa de extracción</b> .....	74
<b>3.4 Análisis de los gráficos para el manejo operacional de la imbibición a partir de los modelos de regresión</b> .....	78
<b>Conclusiones parciales del capítulo:</b> .....	82
Conclusiones Generales.....	83
Recomendaciones: .....	84

#### Índice de Figuras:

Figura 1 Región efectiva de predicción .....	43
Figura 2 : Interrelación entre los valores de las variables observadas. ....	51
Figura 3: Interrelación entre las variables observadas .....	54
Figura 4: Interrelación entre las variables observadas .....	57
Figura 5: Disposición de los residuales en un histograma de frecuencias y en un diagrama “qq” de los residuales: A- Para el brix en el jugo mezclado B- Para la humedad del bagazo C- Para el brix del bagazo. ....	62
Figura 6: Disposición de los residuales según el valor ajustado: A- Para el brix en el jugo mezclado; B- Para la humedad del bagazo y C- Para el brix del bagazo.....	64
Figura 7: Disposición de los valores observados y predichos de la variable dependiente: A- Para el brix en el jugo mezclado; B- Para la humedad del bagazo y C- Para el brix del bagazo. ....	65
Figura 8: Distancia de Cook de las observaciones: A- Brix en el jugo mezclado; B- Humedad del bagazo y C- Brix del bagazo.....	67
Figura 9: Análisis de sensibilidad para el contenido de sólidos solubles del jugo mezclado ....	71
Figura 10: Análisis de sensibilidad para el contenido de humedad del bagazo.....	72

Figura 11: Análisis de sensibilidad para el contenido de sólidos disueltos del bagazo .....	73
Figura 12: Comportamiento de la extracción de sólidos disueltos: A- Respecto a la cantidad de agua de imbibición referida como $\lambda$ y B- Respecto a la temperatura del agua de imbibición. ..	74
Figura 13: Comportamiento según la cantidad de agua de imbibición del flujo de jugo mezclado y de bagazo (gráfico A), de la composición de fibra en bagazo (B) y de la humedad y el contenido de sólidos disueltos del bagazo (C).....	76
Figura 14: Comportamiento del contenido de sólidos disueltos de la caña: A- Respecto a la cantidad de agua de imbibición; B- Respecto a la temperatura del agua de imbibición. ....	78
Figura 15: Gráfico para la estimación de la humedad del bagazo. ....	79
Figura 16: Gráficos para la determinación del brix del jugo mezclado: A- Gráfico de imbibición; B- Gráfico de estimación. ....	80
Figura 17: Manejo de la imbibición para estimar la composición de sólidos del residuo leñoso. A- Estimación del número de imbibición; B- Estimación de la composición del contenido de sólidos disueltos del residuo leñoso. ....	81



## Introducción:

La sociedad del nuevo milenio se encuentra inmersa en un proceso de profundas transformaciones, derivadas fundamentalmente de la globalización de la economía, las crecientes exigencias a las empresas desde el punto de vista de su eficiencia, la evolución de los mercados y la competencia por estos. Múltiples cambios en las relaciones comerciales, las transformaciones tecnológicas y los adelantos científico-técnicos propician una significativa revolución en la organización de las actividades productivas y de servicios, que alcanzan niveles mundiales (Lao *et al.*, 2017).

La tecnología ayuda a las empresas a ganar en eficiencia y a mejorar buena parte de sus procesos internos. Las que la incorporan aumentan su productividad, porque pueden controlar mejor sus procesos y reducir el tiempo en la ejecución de estos, incrementando su competitividad en el mercado (Lao *et al.*, 2017).

Una de las principales empresas de nuestro país es AZCUBA, perteneciente a la industria azucarera. Las participaciones del azúcar en las exportaciones cubanas representaban el 75 % en la década de los ochenta y diez años después se redujo aproximadamente a la mitad (Salgado *et al.*, 2020). Por lo que es de interés de la empresa aumentar su eficiencia, mediante el uso de las tecnologías.

La modelación del proceso de fabricación del azúcar crudo se realiza mediante diferentes técnicas, para comprender y mejorar el rendimiento de las operaciones unitarias (Thaval y Kent, 2013). Los modelos matemáticos sólo son útiles para describir el proceso en estudio si son adecuadamente completos y precisos en el rango requerido de variables involucradas (Mann *et al.*, 2015).

La modelación de la etapa de extracción se realiza con dos objetivos principales: analizar el rendimiento operacional del proceso o predecir el resultado de la modificación de los parámetros en el tren de molienda. La modelación matemática de un tren de molienda puede realizarse de dos formas: mediante modelos de caja negra o modelos fenomenológicos, siendo la segunda variante la más extendida (Kent, 1997).

Los modelos fenomenológicos de la etapa de extracción son herramientas muy robustas e indudablemente poderosas, para la predicción del rendimiento del tren de molienda. Estos tienen una mayor connotación cuando es necesario conocer el funcionamiento de las unidades de molienda. Sin embargo, estos suelen requerir información muy específica del tándem en particular, que necesita una amplia experimentación para ser determinada (Kent, 1997).

Además, algunas variables no son bien conocidas por el personal técnico de las fábricas y no son totalmente extrapolables de un sistema a otro. La solución matemática de las ecuaciones del modelo puede ser un impedimento para tomar decisiones operacionales inmediatas si no se utiliza un software.

En cambio, los modelos de caja negra para la fase de extracción no tienen este nivel de descripción y no predicen el funcionamiento interno de las unidades de molienda. Sin embargo, son más prácticos. Estos modelos son potencialmente útiles en el control del proceso para tomar decisiones operacionales rápidas y precisas. Los modelos de caja negra también son muy atractivos para parte del personal técnico de la fábrica, ya que son más manipulables y menos complejos matemáticamente (Kent, 1997).

Por lo general, los modelos de caja negra requieren información mucho más conocida y fácil de determinar que los modelos fenomenológicos. La elección entre uno u otro depende esencialmente de la finalidad de su uso, del nivel de detalle y de la precisión de la predicción requerida, así como de la información del proceso disponible (Kent, 1997).

La empresa AZCUBA no cuenta con estos modelos matemáticos predictivos para la etapa de extracción, muchas veces se le dificulta el control tecnológico en tiempo real y la toma de decisiones operacionales.

Por tales motivos se tiene el **problema científico** siguiente: ¿Cómo desarrollar un software que contribuya al proceso de obtención de los modelos matemáticos en la etapa de extracción de azúcar crudo?

Para darle solución a lo anteriormente planteado, se formula la **hipótesis** siguiente:

Mediante el empleo de herramientas matemáticas que faciliten la toma de decisiones operacionales, es posible mejorar el comportamiento de la etapa de extracción.

Se delimita como **objetivo general** de la investigación:

Desarrollar un software con herramientas matemáticas para la toma de decisiones operacionales en la etapa de extracción del proceso de fabricación de azúcar de caña.

Para esto se tienen como **objetivos específicos**:

1.- Realizar la modelación matemática de la etapa de extracción de azúcar crudo.

2.- Establecer la validación de los modelos matemáticos de la etapa de extracción desde un punto de vista matemático y tecnológico.

3.- Establecer correspondencias gráficas entre las variables predictivas y dependientes que faciliten el empleo práctico de los modelos matemáticos.

**Campo de acción:** Proceso de obtención de modelos matemáticos predictivos en la etapa de extracción de fabricación de azúcar de caña en el CAI Jesús Rabí.

**Objeto de estudio:** Proceso de obtención de modelos matemáticos en la etapa de extracción para la fabricación del azúcar crudo.

### **Métodos de investigación y descripción empleados:**

Histórico y lógico: para realizar un análisis de las soluciones similares y las tendencias en la obtención de modelos predictivos.

Hipotético-deductivo: Para formular una hipótesis y después, a partir de inferencias lógicas deductivas, arribar a conclusiones particulares.

Modelación: Para representar gráficamente con el objetivo de explicar con mayor exactitud la realidad.

### **Métodos empíricos**

Método de la observación científica: Para obtener una percepción directa del objeto de investigación.

Consulta bibliográfica: para consultar y analizar las fuentes de información relacionadas con la obtención de modelos predictivos en la etapa de extracción.

Generalización: para sistematizar en cada capítulo los aspectos más relevantes y arribar a conclusiones más objetivas y explícitas.

Capítulo 1. Fundamentación teórica: En este capítulo se aborda el marco teórico de la investigación. Se plantean conceptos y criterios de diversos autores, sobre el proceso de producción de azúcar crudo, con detenimiento en la etapa de extracción. Se estudia la repercusión de los principales parámetros de operación de esta, así como la influencia del agua de imbibición.

Capítulo 2. Descripción del procedimiento propuesto: La investigación se efectúa en el CAI "Jesús Rabí" de la provincia de Matanzas en un período de la zafra 2020-2021. En este capítulo se reflejan

los materiales y la metodología a seguir; se compone de dos partes fundamentales: La modelación matemática en la etapa de extracción del proceso de producción de azúcar crudo y el análisis del comportamiento operacional de esta.

Capítulo 3. Análisis de Resultados: Se plantean los resultados de la modelación estadística a partir de los obtenidos en la experimentación y de la predicción del comportamiento a través de tales modelos. Se muestran diferentes herramientas gráficas para un mejor manejo práctico de los modelos matemáticos.

## **Capítulo 1: Marco Teórico - Referencial.**

### **Introducción**

En este capítulo se recoge información científica sobre el proceso de producción de azúcar crudo, con especial detenimiento en la etapa de extracción. Se analiza la repercusión de los principales parámetros de operación de esta, así como la influencia del agua de imbibición; lo que constituye el basamento teórico de la presente investigación. Se enfatizan las recomendaciones de diferentes autores sobre las condiciones a las que se debe aplicar el agua de imbibición y se investiga la repercusión de determinadas variables operacionales sobre el proceso.

### **1.1. Descripción de la organización donde se trabajó**

El Grupo Azucarero AZCUBA, es un sistema articulado de empresas y entidades con un crecimiento sostenido en la producción de azúcar y derivados de la caña, con alto valor agregado y competitivo para el mercado interno y externo, respaldada por centrales azucareros. La presente investigación se llevó a cabo en el CAI Jesús Rabí.

### **1.2. Generalidades del Proceso de Producción de azúcar crudo**

La producción de azúcar involucra una serie de procesos físicos, tales como: extracción de jugo, clarificación, evaporación, cristalización y secado. Además de estos principales procesos, un central azucarero típico tiene sistemas de generación de vapor y de potencia, sistemas de enfriamiento de agua, entre otros (Dogbe *et al.*, 2018).

- Etapa de manipulación y preparación de la caña

El proceso de producción de azúcar crudo dentro de la fábrica comienza con el pesaje de las unidades que transportan la caña desde los frentes de corte. En esta se determina la calidad de la materia prima. Luego, se siguen secciones donde mediante equipos de preparación (niveladoras, cuchillas, desfibradoras, entre otras) se acondiciona la materia prima para una mejor extracción (Hugot, 2014).

- Etapa de extracción o molienda

En esta se separa el jugo que contiene sacarosa del resto de la caña, en la molienda, más de un 95 % del azúcar pasa a formar parte del jugo mezclado (Balla *et al.*, 2018).

- Etapa de purificación

Se elimina la mayor parte de las impurezas presentes en el jugo, para lo cual se realizan operaciones de separación, así como algunas auxiliares de calentamiento (Dogbe *et al.*, 2018).

En la estación de purificación se le adiciona lechada de cal al jugo con el objetivo de neutralizar, o llevar a una condición cercana a esta el pH del jugo, para de esta manera evitar la inversión de la sacarosa durante la evaporación (y calentamiento) (Dogbe *et al.*, 2018).

En el calentamiento del jugo se eleva la temperatura para favorecer la reacción de la cal con los fosfatos presentes en el mismo y formar flóculos de fosfato de calcio insoluble, que contribuyen sustancialmente al proceso de clarificación (Casanova *et al.*, 2019).

Antes de la clarificación, se procede a la evaporación instantánea del jugo en el tanque "flash", lo que estabiliza la temperatura y elimina las burbujas de aire, de lo contrario podrían ocasionar revolturas en el clarificador. En la clarificación, el jugo atraviesa el equipo a muy baja velocidad superficial, lo que hace que los sólidos sedimenten y que salga un jugo limpio y claro (Mohammed, 2018).

Los sedimentos, denominados lodos, pueden recircularse hacia los difusores o alimentarse hacia los filtros rotatorios al vacío (Mkwananzi *et al.*, 2019); los que operan a velocidades entre 4 y 6 rpm. De estos equipos, que consumen agua para el lavado de la torta, se obtiene el jugo filtrado (que se recircula al proceso) y cachaza (Kumar *et al.*, 2020).

La cantidad de cachaza oscila entre 4 y 5 % del total de la caña, aunque localmente, las variaciones pueden ser considerables tanto inferiores al 3 % como superiores al 7 % (Caballero *et al.*, 2019).

- Etapa de concentración

La unidad de evaporación es la responsable de la concentración de jugo clarificado hasta 65 °Bx. Ello supone una demanda considerable de energía, tanto por el elevado contenido energético del calor latente de vaporización como por las cantidades de agua a evaporarse (Mkwananzi *et al.*, 2019).

El principal equipo involucrado es el evaporador a múltiple efecto, el que emplea para la evaporación en un vaso (Chantasiriwan, 2016). La presión de trabajo de cada vaso tiene que ser inferior a la del anterior; el primer vaso trabaja a la presión del vapor de escape, el segundo a presión atmosférica y los siguientes al vacío creciente. De hecho, el evaporador de múltiples efectos multiplica el uso de la unidad de vapor de escape, tanto como vasos haya, a costa del consumo de una pequeña cantidad de energía eléctrica para crear vacío donde corresponda (Casanova *et al.*, 2019).

- Etapa de cristalización

Según Gozá y Ones (2005), el objetivo principal de esta es lograr que el máximo de sacarosa. En esta etapa se concentra la solución azucarada hasta la condición de saturación a vacío en los tachos; además de emplear materiales como magma o semilla y masas cocidas (Mkwananzi *et al.*, 2019).

- Etapa de centrifugación

El objetivo de la misma es separar los cristales presentes en las masas cocidas comerciales de la miel que les rodea mediante la fuerza centrífuga (Gozá y Ones, 2005).

- Etapa de generación de vapor

Genera el vapor necesario para los requerimientos energéticos del proceso. Desde que la cantidad de bagazo requerida para la producción de vapor es mucho menor que la cantidad que se produce durante la molienda, la mayoría de las fábricas de azúcar instalan generadores de vapor para producir potencia (Chantasiriwan, 2016).

Generalmente, en los centrales tradicionales se producen entre 250 y 280 kg de bagazo por cada 1 000 kg de caña molida, lo que permite generar entre 500 y 600 kg de vapor por cada 1 000 kg de esta, o sea, 2 kg de vapor por cada kg de bagazo (Bhutani *et al.*, 2020).

Los centrales azucareros demandan gran cantidad de energía, especialmente en forma de calor (Assawamartbunlue *et al.*, 2018). Dogbe (2020) plantea que, para producir azúcar crudo, normalmente se consumen entre 350 y 600 kg de vapor y de 25 a 32 kWh por cada 1 000 kg de caña molida.

### **1.3. Etapa de extracción del proceso de producción de azúcar crudo**

Comienza con la preparación de la caña, en la misma se busca romper y desgarrar los tallos de esta, mediante diferentes máquinas con el fin de prepararla para la molienda (Morales *et al.*, 2018).

Un mayor rompimiento de las células que contienen azúcar significa una mejor extracción de jugo en los molinos y que el bagazo presente menor cantidad de humedad (Villatoro, 2019). Kusoncum y colaboradores (2020) plantean que cerca de un 90 % de las células de jugo son abiertas a través de la preparación de la caña, extrayendo entre 60 y 75 % de la sacarosa utilizando un primer molino.

La etapa de extracción es para extraer la máxima cantidad de jugo presente en la caña y al mismo tiempo producir el bagazo con la humedad conveniente para producir vapor (Velásquez *et al.*, 2019).

El tándem está compuesto por cinco o seis molinos, de tres o cuatro mazas generalmente (Casanova *et al.*, 2015). Estos tienen sus mazas puestas horizontalmente en distribución triangular; la mayoría son de 3 rodillos estriados y algunos pocos son de 5 (Velásquez *et al.*, 2019).

El primer molino normalmente se concibe para alcanzar la máxima extracción, porque esta es esencial para un buen desempeño del tren. El molino final es también importante, dado que es la "puerta de salida" y a través de él, se pretende minimizar las pérdidas en bagazo y la humedad (Rein, 2012).

En estos la caña se comprime mediante elevadas presiones entre pares de mazas o rodillos consecutivos (Rein, 2012). Las presiones máximas se aplican por lo general en el primer y último molino, los intermedios pueden operar con presiones iguales o inferiores a estos, en dependencia de las condiciones del tándem así como de la preparación de la caña (Herrera *et al.*, 1996).

Para un ingenio, el bagazo sobrante no debe ser cero, porque significa que hay que usar un segundo combustible, implicando un costo adicional. En condiciones normales es necesario disponer de cierta cantidad de bagazo sobrante, para operar durante paradas de la molienda debido a interrupciones operativas o por falta de caña. Con solo 5 % de bagazo sobrante se puede operar durante paradas de hasta 2 ó 3 horas; por lo que se debe garantizar un mínimo sobrante (Jiménez *et al.*, 2017).

#### **1.4. Normalización de la etapa de extracción de azúcar crudo**

En la industria azucarera cubana existe un documento normativo que refiere los valores por los cual se rige la misma y son dictaminados por AZCUBA, el cual es el Manual de Operaciones del 2013.

Una de las variables fundamentales que se controla en el proceso de extracción es el agua de imbibición añadida. AZCUBA (2013) concibe que la cantidad óptima de la misma a aplicar debe ser de 1,8 a 2,0 veces el contenido de fibra en la caña; dado por un compromiso entre la extracción de sacarosa y los requerimientos de evaporación.

Otra variable medida es el porcentaje de fibra en caña, según Casanova y colaboradores en el 2015 esta debe estar entre 12,0 y 13,0 %. Sin embargo, Gálves y Otero (2019) señalan valores entre 12 y 15 % y AZCUBA (2013) de 9 a 15 %. Claro está, ello se encuentra condicionado por las características propias de cada caña y de la especie botánica.

El contenido de sacarosa de diferentes corrientes también es un valor normado por diversos autores. Según ICIDCA (2012) este debe estar entre 13,5 a 14,5 %, mientras que la norma establecida por AZCUBA (2013) para esta variable es entre 10 y 16 %.

El contenido de sacarosa en el bagazo constituye una gran pérdida para el proceso; este depende principalmente de la eficiencia de la molienda, de la humedad del bagazo y en menor medida de la riqueza en azúcar de la caña (Birru, 2016). Para Casanova y Alonso (2006) este debe estar entre 1,5 y 2,0 % mientras que Castro y colaboradores (2019) plantean que debe ser inferior a 1,7 %.

En AZCUBA (2013) se plantea valores de sacarosa en el jugo mezclado de 9 a 17 %, lo que representa un intervalo mucho más amplio que el anterior. En molinos bien ajustados y con una imbibición eficiente deben esperarse valores de esta variable entre 3 y 3,5 % y aún inferiores, lo que minimiza las pérdidas en bagazo (Honing, 1987).

También usualmente se controlan otros parámetros como el contenido de sólidos disueltos del jugo primario y el del último jugo extraído. El primero se debe encontrar de 15 a 25 °Bx y el último debe ser menor a 10 °Bx (AZCUBA, 2013).

Otro valor normado según AZCUBA (2013) es el contenido de sólidos disueltos del jugo mezclado, el que debe oscilar entre 13 y 15 °Bx; valores con los que coinciden con los planteados por Cortés y colaboradores en el 2019.

También es importante en el control del área, la caída de pureza entre el jugo primario y el último jugo extraído, la que no debe ser inferior a 15 o 16, puntos mientras que la diferencia de pureza entre el jugo primario y mezclado debe ser inferior a 1 punto (AZCUBA, 2013).

Del agua de imbibición cabe destacar la temperatura a la cual es añadida. En AZCUBA (2013) se plantea que debe estar entre 60 y 80 °C, mientras que Poel y colaboradores (1998) proponen valores entre 75 y 80 °C.

Otra variable normada es la humedad del bagazo. Ordóñez (2004) da un rango de 49 a 51 % y Carrazana (1987) lo hace de 46 a 52 %. El porcentaje de fibra en bagazo también se controla; Casanova y Alonso (2006) refieren valores de 47 a 49 %, mientras que (Carrazana (1987)) lo hace para un 43 y 47 %; estos últimos un poco más bajo. Para Sahu (2018) el contenido de sólidos disueltos del bagazo oscila entre un 2 y un 5 %. La presión de los molinos debe oscilar entre 3 000 a 3600 psi (20,68 a 24,82 MPa) (Castro *et al.*, 2019).

### **1.5. Modelación Matemática de la etapa de extracción en un tren de molinos**

La modelación de procesos en las fábricas de azúcar crudo se ha realizado mediante diferentes técnicas a lo largo del tiempo con el fin de comprender y mejorar el rendimiento de las operaciones

unitarias (Thaval, 2012). Para que un modelo matemático sea útil, debe ser lo suficientemente completo y preciso para representar el sistema en el rango de variables a estudiar (Mann *et al.*, 2015).

En la etapa de extracción, la influencia del ajuste de los molinos, la calidad de la caña, la imbibición, entre otros son extremadamente difíciles de cuantificar. Un modelo de extracción puede ayudar a atenuar este problema y permite un mejor entendimiento y mejora de este proceso (Wienese, 1990).

Según Kent (2010) los modelos de la etapa de extracción se realizan con dos propósitos fundamentales: para el análisis del comportamiento de la etapa de extracción de un tren de molino y para predecir el resultando del cambio de parámetros en este o en la planta moledora.

Los modelos matemáticos pueden ser utilizados también como punto de partida para la estimación de parámetros como la cantidad de bagazo, que tal como señala Kent (2010), no se mide directamente.

Según Wienese (1990) la extracción depende de la operación de los molinos, de las características del bagazo, entre otros. Los parámetros operacionales típicos en el área son: el número de molinos del tándem, las características de la caña a procesar, la presión hidráulica de la compresión, la velocidad de los molinos, la cantidad de imbibición, la aspereza de las mazas de los molinos y la abertura de estos. De modo análogo, las características del bagazo que afectan la extracción son: la preparación de la caña, la pol, el porcentaje de fibra y el de cenizas.

Sin embargo, la consideración de estas variables en la modelación es compleja y a veces ineficiente. De hecho, tal como lo plantea Wienese (1990), varios esfuerzos se han realizado para cuantificar el efecto de esos parámetros individuales con pequeño o ningún éxito.

La composición de la caña y el nivel de imbibición influyen en el comportamiento de la fábrica. El conocer el efecto de las mismas es importante por dos razones: para compararlo con parámetros similares y para predecir el cambio de estos parámetros (Wienese, 1990).

### **1.5.1 Modelos fenomenológicos aplicados al proceso de extracción**

Pedroso (2018a) desarrolla modelos basados en balances de masa en cada molino con la consideración de parámetros como: eficiencia de la separación de fibra, eficacia de la imbibición y el coeficiente de este; los que son en última instancia de naturaleza empírica.

En esta investigación se ajusta a un modelo polinómico de cuarto grado el comportamiento logarítmico de la extracción de brix con la cantidad de agua de imbibición según la fibra. De esta manera al

correlacionar la extracción de sacarosa con la de brix (polinomio de segundo grado), se define la influencia de la imbibición sobre la recuperación de pol.

Russell (1965) desarrolla la modelación de un tren de molido a través de balances de masa para lo cual le resulta necesario considerar el factor de reabsorción y el coeficiente de imbibición.

Edwards (1995) presenta un modelo de extracción para lo cual reemplaza el coeficiente de imbibición por dos parámetros: el factor de trituración y la eficiencia de la mezcla. Este primero contempla el efecto del proceso de apertura de las celdas de jugo mediante un molino, mientras que el segundo se centra en la mezcla de imbibición con bagazo. Tiene la desventaja que requiere del conocimiento de los grados brix en celdas abiertas para cada molino, información que rara vez se conoce. Thaval y Kent (2012) coinciden con ello y señalan que este modelo se ha adoptado ampliamente.

Bajo el principio de balances de masa, Kent (2010) presenta un modelo para estimar la producción de bagazo a partir de la consideración de insolubles. Igualmente contempla términos no medidos en la industria frecuentemente, a tal punto que el propio autor no lo considera adecuado para su implementación en la fábrica.

En Thaval y Kent (2012) se desarrolla un modelo para definir el comportamiento de una unidad de extracción. Este modelo toma en consideración la fibra suspendida en el jugo y recalcula parámetros como el factor de reabsorción, el coeficiente de imbibición y la eficiencia de separación para lo cual se usa, a juicio de los autores, definiciones más completas que en modelos anteriores.

Según Thaval y Kent (2012) el factor de reabsorción determina las cantidades relativas de jugo total en el bagazo de entrega y el jugo exprimido, mientras que el coeficiente de imbibición determina la división de brix y humedad en el bagazo de entrega y el jugo exprimido; el modelo que se propone se deriva directamente del modelo MILSIM de Russell, el que según este artículo es el primer y único modelo de extracción con una amplia aplicación en Australia.

El modelo que se presenta en Thaval y Kent (2012) se implementó exitosamente en el paquete profesional "SysCAd".

Chantasiriwan (2013) fundamenta su modelación de la etapa con balances de masa para lo cual le resulta necesario emplear variables como eficiencia en la separación de fibra y coeficientes de mezclado y de obstrucción de fibra. Ofrece un gráfico en el que se aprecia el comportamiento logarítmico de la extracción con respecto al nivel de imbibición.

Hugot (2014) presenta varios modelos sustentados en balances de masa en el área según el sistema de imbibición, el agua de imbibición según la fibra en caña, a partir de la consideración de coeficientes alfa y beta. El autor construye curvas características teniendo en cuenta el sistema de imbibición y el coeficiente beta que facilita indiscutiblemente su empleo.

De esta manera se aprecia que estos modelos fenomenológicos no son apropiados para un control operacional del área desde un punto de vista práctico, debido a que demandan información que comúnmente no se maneja en la industria. Además, generalmente implican un sistema de ecuaciones, derivadas de balances de masa que le contrarresta "operabilidad" a la modelación; cuya resolución puede llegar a ser compleja para los operadores. Estos modelos tienen la ventaja de una detallada descripción del proceso de extracción de cada molino, pero dejan de ser muy "atractivos" cuando el interés de la predicción se centra solamente en las salidas de la etapa de extracción.

### **1.5.2 Modelos de caja negra aplicados al proceso de extracción**

Para Pedroso (2018a) los molinos pueden considerarse como una caja negra en la que entra caña y agua de imbibición, y sale jugo mezclado y bagazo. Plantea además que la perspectiva de caja negra es muy útil, pero no provee suficiente información para entender que sucede en cada molino, por lo que se desarrollan modelos que describen el comportamiento de cada uno de ellos.

Para Rein (1975) las variables que rigen el comportamiento de un tren de molinos se pueden agrupar básicamente en tres categorías:

- Variables fijas tales como: el número y tipo de molinos, número de mazas, entre otras.
- Aquellas variables de control manipulables por los operadores como son: la velocidad de las mazas y su ajuste, la cantidad de imbibición aplicada, entre otras.
- Variables con comportamiento aleatorio como es el asociado con las características de la caña molida, que varía constantemente de un corte a otro.

De todas las categorías presentadas anteriormente, la más interesante resulta la tercera, pues no existe control sobre ella y su análisis sobre el comportamiento de la molida es muy difícil. Para mitigar ese efecto y mejorar el control del área se pueden desarrollar modelos de caja negra y emplear técnicas como la regresión lineal múltiple. En tal caso es importante la adecuada selección de la colección de valores y tratar que la muestra sea suficiente y confiable (Rein, 1975).

Entre los trabajos más significativos al respecto en la literatura especializada se encuentra el de Rein (1975). En este se muestra un estudio, en el que se desarrolla un modelo de caja negra por regresión

lineal múltiple en la etapa de extracción. Considera como variables independientes la pérdida en molienda, la razón de extracción, la extracción reducida en base a Deer y en base a Mittal; así como la extracción de pol, el porcentaje de pol en el bagazo, la pérdida absoluta en el jugo según la fibra en base a la pol y en base al brix.

Rein (1975) establece como término independiente a la calidad de la caña, la que puede ser representada a través de la composición de la caña. Señala que esta puede ser expresada en términos de concentración de sacarosa en el jugo de la caña y la fracción del jugo con respecto a la fibra en caña. Muy importante resulta el hecho que plantea que el brix de la caña no se emplea en la modelación porque la composición del jugo presenta una alta correlación con este parámetro y que su empleo no aporta una ganancia significativa con relación a la otra variable.

En este estudio se prueba que el flujo de bagazo producido no es significativo sobre las variables respuestas. Es importante destacar que la respuesta depende del tipo de molinos, de las características particulares del tándem, entre otras cuestiones.

Munro (1963) desarrolla un estudio en dos molinos experimentales en el que determina la influencia de la cantidad de agua de imbibición. Para ello emplea una muestra de caña de 12,7 % de fibra previamente preparada. De ello se evidencia un comportamiento lineal de la extracción de brix con respecto a la cantidad del agua de imbibición según la fibra en el segundo molino.

En Lionnet (1981) se desarrollan modelos empíricos que relacionan linealmente la extracción de pol con la fracción de la pureza de la caña entre la del jugo mezclado. Los modelos presentan una adecuada calidad de ajuste con respecto a la colección de datos industriales que se emplean para su obtención y son validados experimentalmente mediante una prensa a nivel de laboratorio.

Es importante que se consideren como términos independientes las características de entrada vinculadas con la caña ya que influyen de manera importante sobre la composición del jugo mezclado y el bagazo, como es el caso del brix o pol del jugo primario. A su vez, ello le otorga una mayor robustez a la modelación de caja negra en sí.

## **1.6. Minería de datos**

La minería de datos es un campo de la estadística y las ciencias de la computación referido al proceso de detectar la información procesable de los conjuntos grandes de datos. El término es un concepto de moda, y es frecuentemente mal utilizado para referirse a cualquier forma de datos a gran escala o

procesamiento de la información. En el uso de la palabra, el término clave es el descubrimiento, comúnmente se define como "la detección de algo nuevo", para esto utiliza el análisis matemático para deducir los patrones y tendencias que existen. Normalmente, estos patrones no se pueden detectar mediante la exploración tradicional porque las relaciones son demasiado complejas o porque hay demasiados datos. El objetivo general del proceso consiste en extraer información de un conjunto de datos y transformarla en conocimiento. Para entender su significado es imprescindible conocer los términos relacionados en esta:

- Datos: son cualquier hecho, número o texto que puede ser procesado por una computadora. Hoy en día, las organizaciones están acumulando grandes cantidades, y cada vez mayores, en diferentes formatos y diferentes bases de datos.
- Información: los patrones, asociaciones, o relaciones entre todos estos datos pueden proporcionar información. Por ejemplo, el análisis del punto de venta de datos de transacciones puede dar información sobre qué productos se venden y cuándo.
- Conocimiento: la información puede ser convertida en conocimiento acerca de los patrones históricos y las tendencias futuras. Por ejemplo, la información resumida sobre las ventas de supermercados minoristas puede ser analizada a la luz de los esfuerzos de promoción para facilitar el conocimiento del comportamiento de compra del consumidor. Por lo tanto, un fabricante o distribuidor puede determinar qué elementos son los más susceptibles a los esfuerzos de promoción.

De forma general, los datos son la materia prima bruta. En el momento que el usuario les atribuye algún significado especial pasan a convertirse en información. Cuando los especialistas elaboran o encuentran un modelo, haciendo que la interpretación que surge entre la información y ese modelo represente un valor agregado, entonces nos referimos al conocimiento. (Tecnologías Información, n.d.)

Suscita cierta polémica el definir las fronteras existentes entre la minería de datos y las disciplinas análogas como la estadística y la informática. Pero a pesar de existir muchas similitudes, en la minería de datos se encuentran una serie de problemas y métodos específicos que la hacen distinta de otras disciplinas.

El hecho es que, en la práctica la totalidad de los modelos y algoritmos de uso general en minería de datos como redes neuronales, árboles de regresión y clasificación, modelos logísticos y análisis de componentes principales gozan de una tradición relativamente larga en otros campos.

Ciertamente, la minería de datos bebe de la estadística, de la que toma las siguientes técnicas:

- Análisis de varianza, mediante el cual se evalúa la existencia de diferencias significativas entre las medias de una o más variables continuas en poblaciones distintas.
- Regresión: define la relación entre una o más variables y un conjunto de variables predictoras de las primeras.
- Análisis de agrupamiento o clustering: permite la clasificación de una población de individuos caracterizados por múltiples atributos (binarios, cualitativos o cuantitativos) en un número determinado de grupos, con base en las semejanzas o diferencias de los individuos.
- Análisis discriminante: permite la clasificación de individuos en grupos que previamente se han establecido, permite encontrar la regla de clasificación de los elementos de estos grupos, y por tanto una mejor identificación de cuáles son las variables que definan la pertenencia al grupo.
- Series de tiempo: permite el estudio de la evolución de una variable a través del tiempo para poder realizar predicciones, a partir de ese conocimiento y bajo el supuesto de que no van a producirse cambios estructurales.

De la informática toma las siguientes técnicas:

- Inteligencia Artificial: Mediante un sistema informático que simula un sistema inteligente, se procede al análisis de los datos disponibles. Entre los sistemas de Inteligencia Artificial se encuadrarían los Sistemas Expertos y las Redes Neuronales.
- Sistemas Expertos: Son sistemas que han sido creados a partir de reglas prácticas extraídas del conocimiento de expertos. Principalmente a base de inferencias o de causa-efecto.
- Sistemas Inteligentes: Son similares a los sistemas expertos, pero con mayor ventaja ante nuevas situaciones desconocidas para el experto.
- Redes neuronales: Genéricamente, son métodos de proceso numérico en paralelo, en el que las variables interactúan mediante transformaciones lineales o no lineales, hasta obtener unas salidas. Estas salidas se contrastan con los que tenían que haber salido, basándose en unos datos de prueba, dando lugar a un proceso de retroalimentación mediante el cual la red se reconfigura, hasta obtener un modelo adecuado.

La tarea real de minería de datos es el análisis automático o semiautomático de grandes cantidades de datos para extraer patrones interesantes hasta ahora desconocidos. Estos patrones pueden entonces

ser vistos como una especie de resumen de los datos de entrada, y pueden ser utilizados en el análisis adicional o, por ejemplo, en el aprendizaje automático y análisis predictivo.

Un proceso típico de minería de datos consta de los siguientes pasos generales:

- 1- Selección del conjunto de datos, tanto en lo que se refiere a las variables objetivo (aquellas que se quiere predecir, calcular o inferir), como a las variables independientes (las que sirven para hacer el cálculo o proceso), como posiblemente al muestreo de los registros disponibles.
- 2- Análisis de las propiedades de los datos, en especial los histogramas, diagramas de dispersión, presencia de valores atípicos y ausencia de datos (valores nulos).
- 3- Transformación del conjunto de datos de entrada, se realizará de diversas formas en función del análisis previo, con el objetivo de prepararlo para aplicar la técnica de minería de datos que mejor se adapte a los datos y al problema, a este paso también se le conoce como preprocesamiento de los datos.
- 4- Seleccionar y aplicar la técnica de minería de datos, se construye el modelo predictivo o de clasificación
- 5- Extracción de conocimiento, mediante una técnica de minería de datos, se obtiene un modelo de conocimiento, que representa patrones de comportamiento observados en los valores de las variables del problema o relaciones de asociación entre dichas variables. También pueden usarse varias técnicas a la vez para generar distintos modelos, aunque generalmente cada técnica obliga a un pre procesado diferente de los datos.
- 6- Interpretación y evaluación de datos, una vez obtenido el modelo, se debe proceder a su validación comprobando que las conclusiones que arroja son válidas y suficientemente satisfactorias. En el caso de haber obtenido varios modelos mediante el uso de distintas técnicas, se deben comparar los modelos en busca de aquel que se ajuste mejor al problema.

Si el modelo final no superara esta evaluación el proceso se podría repetir desde el principio o, si el experto lo considera oportuno, a partir de cualquiera de los pasos anteriores. Esta retroalimentación se podrá repetir cuantas veces se considere necesario hasta obtener un modelo válido. Una vez validado el modelo, si resulta ser aceptable (proporciona salidas adecuadas y/o con márgenes de error admisibles) éste ya está listo para su explotación. Los modelos obtenidos por técnicas de minería de datos se aplican incorporándolos en los sistemas de análisis de información de las organizaciones.

### **1.6.1 Técnicas de minería de datos**

Como ya se ha comentado, las técnicas de la minería de datos provienen de la inteligencia artificial y de la estadística, dichas técnicas, no son más que algoritmos, más o menos sofisticados que se aplican sobre un conjunto de datos para obtener unos resultados.

Las técnicas más representativas son:

- **Redes neuronales.** Son un paradigma de aprendizaje y procesamiento automático inspirado en la forma en que funciona el sistema nervioso de los animales. Se trata de un sistema de interconexión de neuronas en una red que colabora para producir un estímulo de salida.
- **Regresión lineal.** Es la más utilizada para formar relaciones entre datos. Rápida y eficaz pero insuficiente en espacios multidimensionales donde puedan relacionarse más de 2 variables.
- **Árboles de decisión.** Es un modelo de predicción utilizado en el ámbito de la inteligencia artificial y el análisis predictivo, dada una base de datos se construyen estos diagramas de construcciones lógicas, muy similares a los sistemas de predicción basados en reglas, que sirven para representar y categorizar una serie de condiciones que suceden de forma sucesiva, para la resolución de un problema.
- **Reglas de asociación.** Se utilizan para descubrir hechos que ocurren en común dentro de un determinado conjunto de datos.
- **Series temporales.** Son utilizadas para el análisis de la relación causal entre diversas variables que cambian con el tiempo y se influyen entre sí.

En resumen, la minería de datos se presenta como una tecnología emergente, con varias ventajas: por un lado, resulta un buen punto de encuentro entre los investigadores y las personas de negocios; por otro, ahorra grandes cantidades de dinero a una empresa y abre nuevas oportunidades de negocios. Además, no hay duda de que trabajar con esta tecnología implica cuidar un sinnúmero de detalles debido a que el producto final involucra "toma de decisiones".

### **1.7. Análisis Predictivos**

El análisis predictivo es un área de la minería del dato que pretende extraer conocimiento que le permita predecir tendencias y patrones de comportamiento. A menudo una circunstancia desconocida de interés se va a producir en el futuro, pero el análisis predictivo se puede aplicar igualmente a lo desconocido tanto en el pasado, el presente o el futuro. Por ejemplo, identificar sospechosos después de haberse producido un crimen o un fraude con tarjeta de crédito. Lo fundamental del análisis

predictivo está en identificar relaciones entre las variables explicativas y las variables predictivas del pasado de forma que se pueda escalar a lo que está por ocurrir. Es importante advertir, en cualquier caso, que la fiabilidad y usabilidad de los resultados dependerán mucho del nivel de análisis del dato y la calidad de las hipótesis.

Este es el proceso de utilizar el análisis de datos para realizar predicciones basadas en estos, se hace uso de los datos junto con técnicas analíticas, estadísticas y de aprendizaje automático a fin de crear un modelo predictivo para predecir eventos futuros. Agrupa una variedad de técnicas estadísticas de modelización, aprendizaje automático y minería de datos que analiza los datos actuales e históricos reales para hacer predicciones acerca del futuro o acontecimientos no conocidos.

El efecto funcional que pretenden estas iniciativas técnicas es que el análisis predictivo provea una puntuación (probabilidad) para cada sujeto (cliente, empleado, paciente, producto, vehículo, componente, máquina y otra unidad en la organización) con el objeto de determinar, informar o influir procesos en la organización en el que participen un gran número de sujetos, tal y como ocurre en marketing, evaluación de riesgo de crédito, detección de fraudes, fabricación, salud y operaciones gubernamentales como el orden público.

### **1.8. Tendencias Tecnológicas**

Es necesario para el desarrollo de un producto informático que satisfaga una necesidad existente, el empleo de diferentes herramientas y tecnologías con demostrada idoneidad. Ya que, según el problema a resolver, es de suma importancia una correcta elección de las mismas.

#### **1.8.1. Lenguaje de Programación**

##### **Lenguaje R:**

Creado en 1993, en la universidad de Auckland. Viene derivado de otros dos lenguajes, que son S y Scheme. Sus creadores son Ross Ihaka y Robert Gentleman. Es un lenguaje con licencia GNU, es decir, es libre, gratuito y abierto. En resumen, lo puede usar cualquiera y no es propiedad de nadie. R funciona con paquetes gratuitos, como las librerías en otros lenguajes, y puedes descargar y usar esos paquetes. Algunas de sus características principales como lenguaje son:

- Posibilidad de crear gráficos, basado en LaTeX.
- Gran cantidad de herramientas estadísticas:

- modelos lineales y no lineales.
  - tests estadísticos.
  - algoritmos de clasificación y agrupamiento.
- Posibilidad de crear tus propias funciones, además de objetos al ser su programación POO (orientada a objetos).
  - Integración con distintas bases de datos.
  - Puede tener un uso matemático.

### **1.8.2. Extensiones y paquetes**

R forma parte de un proyecto colaborativo y abierto. Sus usuarios pueden publicar paquetes que extienden su configuración básica. Existe un repositorio oficial de paquetes cuyo número superó en otoño de 2009 la cifra de los 2000.

Dado el enorme número de nuevos paquetes, estos se han organizado en vistas (o temas), que permiten agruparlos según su naturaleza y función. Por ejemplo, hay grupos de paquetes relacionados con estadística bayesiana, econometría, series temporales, etc.

Para facilitar el desarrollo de nuevos paquetes, se ha puesto a servicio de la comunidad una forja de desarrollo que facilita las tareas relativas a dicho proceso.

### **1.8.3. Entorno de Desarrollo Integrado (IDE)**

#### **RStudio:**

Es un entorno de desarrollo integrado (IDE) para el lenguaje de programación R, dedicado a la computación estadística y gráficos. Incluye una consola, editor de sintaxis que apoya la ejecución de código, así como herramientas para el trazado, la depuración y la gestión del espacio de trabajo.

RStudio está disponible para Windows, Mac y Linux o para navegadores conectados a RStudio Server o RStudio Server Pro (Debian / Ubuntu, RedHat / CentOS, y SUSE Linux). RStudio tiene la misión de proporcionar el entorno informático estadístico R. Permite un análisis y desarrollo para que cualquiera pueda analizar los datos con R. entre sus principales características se encuentran: (RStudio, n.d.)

- IDE construido exclusivo para R

- El resaltado de sintaxis, auto completado de código y sangría inteligente.
- Ejecutar código R directamente desde el editor de código fuente.
- Salto rápido a las funciones definidas.
- Potente autoría y depuración.
- Depurador interactivo para diagnosticar y corregir los errores rápidamente.
- Herramientas de desarrollo extensas.
- Autoría con Sweave y R Markdown.

Aunque se cuenta con dos excelentes entornos de desarrollo se decidió optar por RStudio por ser un IDE exclusivo para R. Este logra un entorno más agradable para trabajar y cuenta con extensas herramientas para el desarrollo de aplicaciones específicamente en este lenguaje. Además, Visual Studio en su versión gratuita cuenta con opciones de desarrollo muy básicas, aprovechadas principalmente por programadores principiantes, quedándose corto en muchas funcionalidades respecto a RStudio.

### **Conclusiones parciales del capítulo**

1. El correcto ajuste de la imbibición es muy importante para minimizar las pérdidas de sacarosa en bagazo y lograr una adecuada economía de vapor.
2. No se encuentra bien definido en la literatura la influencia de la cantidad y temperatura del agua de imbibición sobre el comportamiento de las variables operacionales del área.
3. No existen modelos matemáticos lo suficientemente prácticos y robustos que puedan ser empleados para el control operacional de la etapa de extracción en la literatura especializada.
4. La modelación empírica es más adecuada que la fenomenológica en el desarrollo de herramientas matemáticas para un mejor control y la toma de desiciones operacionales con relación a la imbibición.
5. La minería de datos es una tecnología emergente que facilita la toma de decisiones.
6. El principal beneficio del análisis predictivo que permite a las empresas aprender de su experiencia pasadas a partir de sus datos y tomar decisiones.

## Capítulo 2: Descripción del procedimiento propuesto.

En este capítulo se abordan los materiales y la metodología a seguir en la presente investigación. Metodológicamente se compone de dos partes fundamentales. La primera concierne a la modelación matemática de la etapa de extracción del proceso de producción de azúcar crudo y la segunda al análisis del comportamiento operacional de esta, a partir de una realidad simulada que se sustenta en los modelos que se obtienen. Esta investigación se efectúa en el CAI “Jesús Rabí” de la provincia de Matanzas en un período de la zafra 2020-2021.

### 2.1 Caracterización de la etapa de extracción en el CAI Jesús Rabí.

La caña se recibe en el basculador, donde se prepara a través de niveladores y martillos, para facilitar el proceso de extracción. En esta etapa existe un colador rotatorio, cinco conductores de arrastre, tres bombas de maceración, dos de jugo diluido y una de agua de imbibición.

El tándem de molinos está integrado por cinco unidades de cuatro mazas, con un diámetro medio de 40 pulg (1,02 m) y 7 pies (2,13 m) de longitud. Para la zafra correspondiente presentan un ajuste (comúnmente conocido por su denominación en inglés, “setting”) para procesar 4 000 000 kg de caña por día; que define una molienda horaria nominal ininterrumpida de 166 670 kg/h.

El primer molino presenta una velocidad de rotación de 705 rpm mientras que los restantes de 590 rpm. El primer molino tiene función desfibradora dado el rayado especial de sus mazas. Presentan tolva tipo *Donelly* y están accionados por motores eléctricos de rotor bobinado (un motor por unidad de molida). La presión de trabajo de todos estos equipos es de 21,38 MPa (3100 psi).

En los molinos, el primero extrae el jugo de mayor concentración de sacarosa (jugo primario) que pasa directamente al colador de jugo, mientras que el bagazo que resulta se envía por los restantes molinos para continuar la extracción de sacarosa. Para una mayor recuperación de esta, se aplica imbibición compuesta triple con el condensado de los calentadores de jugo en forma de aspersores. De esta manera, se obtiene a la salida del segundo molino el jugo diluido, el cual se une con el jugo primario en las canaletas de los molinos para constituir el jugo mezclado.

El último molino entrega el bagazo a un conductor de banda que lo eleva al rastrillo repartidor de los generadores de vapor. Todo el bagazo obtenido se separa del bagacillo y pasa hacia los generadores de vapor correspondientes. Tanto el área de báscula y preparación como propiamente la de extracción se encuentran parcialmente automatizadas.

## 2.2 Modelación del comportamiento de la etapa de extracción del proceso de producción de azúcar crudo

### 2.2.1 Consideraciones generales

La modelación del comportamiento de la etapa extracción del proceso de producción de azúcar crudo, según la cantidad y temperatura del agua de imbibición, se efectúa en el CAI: Jesús Rabí de la provincia de Matanzas (Cuba). Para ello resulta de interés predecir, dado su alta significación en el proceso algunas variables, esta investigación se enfoca en las siguientes:

1. Contenido de sólidos disueltos del Jugo Mezclado ( $^{\circ}\text{Bx}$ )
2. Composición de humedad del residuo leñoso (%)
3. Contenido de sólidos disueltos del residuo leñoso ( $^{\circ}\text{Bx}$ )

Se emplea como técnica de modelación matemática la modelación estadística por regresión lineal múltiple. Se consideran como términos independientes en la modelación, a la temperatura del agua de imbibición, la cantidad de agua de imbibición, la cantidad de caña molida y la fibra de esta. Estas últimas tres se tienen en cuenta a través del peso del agua de imbibición con respecto al de la fibra en caña ( $\lambda$ ). Este término, además de ser ampliamente conocido, es adimensional y permite generalizar el grado de imbibición para otros centrales azucareros con diferente capacidad de molienda.

Ambas variables se pueden manipular de acuerdo con la molida, a la cantidad de fibra en la caña y a los intereses que se persigan propiamente con la observación experimental. El agua de imbibición se varía de forma tal que se verifiquen aproximadamente relaciones de  $\lambda$  entre 0,5 a 3; intervalo que incluye a los valores típicos de operación y a una desviación prudencial de estos (Rein, 2012).

De forma similar, la temperatura del agua de imbibición se analiza entre 30 y 80  $^{\circ}\text{C}$ , siendo este último el máximo valor operacionalmente permitido (AZCUBA, 2013).

Para la estimación del contenido de sólidos disueltos, tanto en el bagazo como en el jugo mezclado, debe considerarse también a la riqueza en la caña de este componente. Ello se debe a que, para un mismo grado de imbibición, mientras mayor sea la composición de este en la caña, mayor será su contenido en las corrientes de salida del tándem. Sin embargo, el contenido de sólidos disueltos en la caña es un término que se conocen luego de concluida la molida, a partir de la determinación de la composición correspondiente en el residuo leñoso y el jugo mezclado. Por lo que se establece para este propósito como términos independientes a la composición de sólidos disueltos en el jugo primario.

De este modo se tienen las siguientes relaciones básicas para la regresión:

$$BxJM = f(\lambda, T, BxJP) \quad (2.1)$$

$$HumdRL = f(\lambda, T) \quad (2.2)$$

$$BxRL = f(\lambda, T, BxJP) \quad (2.3)$$

Donde:

*BxJM*: Contenido de sólidos disueltos del jugo mezclado ( $^{\circ}BxJM$ )

*BxJP* : Contenido de sólidos disueltos del jugo primario ( $^{\circ}BxJP$ )

$\lambda$ : Cantidad de agua de imbibición según el peso de la fibra en caña (kg/kg)

*T*: Temperatura del agua de imbibición ( $^{\circ}C$ )

*HumdRL*: Contenido de humedad del residuo leñoso (%)

*BxRL*: Contenido de sólidos disueltos del residuo leñoso (%)

### **2.2.2 Desarrollo de la experimentación**

La colección de datos se efectúa por observaciones del proceso. Se persigue lograr recolectar la mayor cantidad de datos posibles de forma pasiva y completar las observaciones necesarias para apreciar desviaciones de los valores normales de operación a partir de la manipulación de las variables operacionales de la imbibición. Deben contarse con réplicas, al menos en algunas observaciones, para la estimación del error puro en el análisis de la bondad de ajuste de los modelos. La cantidad de datos experimentales tiene que ser suficiente, de acuerdo con el número de variables predictoras del modelo.

Es importante lograr que los parámetros restantes se mantengan controlados a los valores comunes de operación, de forma tal que estos no interfieran sobre las mediciones. Entre ellos se encuentran las presiones de trabajo de los molinos, el índice de preparación de la caña, el porcentaje de materia extraña, la velocidad de rotación de las mazas, entre otras.

El flujo de agua de imbibición se regula de forma manual mediante una válvula en la línea de dosificación. Se tiene instalado en el campo un flujómetro KF700 (China) que se encuentra conectado a un sistema de control automático.

Se manipula la temperatura del agua de imbibición mediante la válvula que regula el suministro de vapor y de agua fresca en el colector de atemperamiento correspondiente. Este último se localiza entre el tanque de almacenamiento de agua de imbibición (condensados de los calentadores) y la bomba que la envía a la entrada del quinto molino. La medición de temperatura se efectúa mediante un termómetro bimetálico (CSHERUN, China) instalado en el área y que está conectado a una pizarra de control automático.

Se toman las muestras de jugo y del residuo leñoso para su análisis a partir de treinta minutos después de haberse realizado alguna modificación en los parámetros de la imbibición.

La determinación del porcentaje de sólidos disueltos se determina según el procedimiento analítico número ocho de MINAZ (2006). Se emplea un refractómetro digital PDR-35 (China).

Para determinar el porcentaje de fibra se utiliza una prensa hidráulica Engehidro (Modelo PHE-45) a través del procedimiento analítico número dos de MINAZ (2006). La determinación de humedad en el residuo leñoso se realiza de acuerdo al procedimiento analítico número doce de MINAZ (2006).

### **2.2.3 Procesamiento estadístico y ajuste matemático de los resultados experimentales**

El ajuste matemático de los resultados experimentales se realiza mediante el software RStudio (versión 1.3.959). Esencialmente el análisis estadístico para el desarrollo matemático de la modelación empírica se divide en tres etapas:

- Etapa 1: Ajuste matemático de los modelos de regresión múltiple
  - Etapa 2: Comprobación del cumplimiento de los supuestos de la regresión múltiple
  - Etapa 3: Análisis de la calidad de ajuste de los modelos matemáticos
- ✓ Etapa 1: Ajuste matemático de los modelos de regresión múltiple

El ajuste matemático de los modelos de regresión lineal se efectúa mediante el comando "lm", para lo cual debe especificarse los datos con los cuales se efectúa el análisis y la estructura matemática del modelo. Para la propuesta de la estructura matemática se analiza de forma gráfica la disposición de la respuesta según los términos independientes mediante el comando "pairs".

En caso de apreciarse un comportamiento no lineal, deben evaluarse otras estructuras de naturaleza no lineal previamente linealizadas. Ello puede realizarse a partir de transformaciones tanto en términos independientes como en la variable dependiente (Montgomery *et al.*, 2002).

Debe garantizarse la mayor simplicidad estructural (principio de parsimonia) y que todos sus términos sean significativos, esto último se determina a partir del p-valor de la prueba de significación de los estimados. Para que sea significativo este debe ser menor que 0,05 para una confianza del 95 %. De lo contrario, debe considerarse la exclusión del regresor correspondiente de la estructura matemática del modelo.

De igual manera se analiza el valor del estadígrafo de Fisher en la prueba de significación del modelo para determinar si hay una relación lineal significativa entre la variable dependiente y las independientes (Montgomery y Runger, 2018). En caso de que el p-valor de esta sea mayor al nivel de significación, implica no rechazar la hipótesis nula y se infiere que la regresión no es significativa. Ello supone la consideración de otras modificaciones en la estructura del modelo o en última instancia, de otra técnica de modelación para la obtención de este.

Para la selección de un modelo entre varias propuestas se tienen como criterios el del valor del error cuadrático medio (RMSE, por sus siglas en inglés), el valor del coeficiente de significación ajustado, el criterio de información de Akaike (AIC) y el de información bayesiano (BIC).

Para la determinación de cada indicador se emplea el paquete *performance* de Lüdecke y colaboradores (2021). Se emplea para ello el comando: “compare\_performance”.

✓ Etapa 2: Comprobación del cumplimiento de los supuestos de la regresión múltiple

En la metodológica se analiza el cumplimiento de los supuestos de la regresión múltiple planteados por Montgomery y colaboradores en el 2002.

El supuesto de relación lineal ente los predictores numéricos y la variable dependiente se realiza mediante la prueba de significación de Fisher que se describe en la etapa 1. A su vez se determina la media de los residuos, la que debe ser de 0 o aproximadamente este valor.

Se analiza el cumplimiento del supuesto de normalidad a través de la prueba de Kolmogorov-Smirnov, la cual es más recomendable que la de Shapiro-Wilk cuando el tamaño de la muestra es mayor a 50 (Mishra *et al.*, 2019). Sin embargo, es más apropiado emplearla con la modificación de Lilliefors (Garbo *et al.*, 2021). Esta se emplea a través del paquete *nortest* de Gross y Ligges (2015).

Para ello se consideran los residuales estudentizados de acuerdo a lo recomendado por Montgomery y colaboradores en el 2002, cuyo escalamiento se realiza a través del paquete *MASS* de Ripley y colaboradores (2021).

Con igual propósito como análisis complementarios se analiza el valor de los coeficientes de Kurtosis y de asimetría de los residuales (Ballard *et al.*, 2021). Estos se determinan a través del paquete “parameters” de Lüdecke y colaboradores (2021).

Como herramientas gráficas se emplean los gráficos de distribución de residuales según los cuartiles “qq” (Ciaburro, 2018) y el histograma de frecuencias correspondiente.

El supuesto de homocedasticidad en los residuos (varianza constante) se realiza mediante la prueba de Breusch-Pagan (Fox y Sanford, 2019). Esta se efectúa mediante el paquete “*lmtest*” de Hothorn y colaboradores (2020).

Se tiene como herramienta gráfica en el análisis a la disposición de los residuales de acuerdo con los valores ajustados, los que deben distribuirse de forma aleatoria en torno a cero, manteniendo aproximadamente la misma variabilidad a lo largo del eje de las coordenadas.

La no existencia de multicolinealidad entre los predictores se analiza a partir de la matriz de correlación entre los predictores, en la que debe existir entre ellos un débil coeficiente de correlación (bastante inferior a 1). Por igual motivo a través del paquete “car” (Fox *et al.*, 2021) se determina el valor de inflación de la varianza (VIF, por sus siglas en inglés).

Se analiza que los errores no se encuentren correlacionados entre sí mediante la prueba de Durbin-Watson (Fakoya, 2020), que se realiza a través del paquete *car*. El p-valor de la prueba debe ser mayor que el nivel de significación para poder inferir que no exista auto-correlación entre los residuales. Este análisis se complementa a través de la correcta disposición sobre la diagonal de 45° de los valores observados y predicho de la variable dependiente.

Se analiza también la no existencia de valores muy influyentes mediante la distancia de Cook (Fox y Sanford, 2019).

Para la construcción de algunos gráficos auxiliares en algunos análisis se emplea el paquete “ggplot2” de Wickham *et al.* (2021).

✓ Etapa 3: Análisis de la calidad de ajuste de los modelos matemáticos

Se emplea la prueba de pérdida de calidad de ajuste de Fisher, la que requiere de contar con observaciones replicadas. Se destaca que deben ser réplicas verdaderas y no mediciones repetidas de la variable dependiente.

Se emplea para la estimación del error “puro” réplicas que se obtienen en el mismo contexto en el que se recolectan los datos para el desarrollo de la regresión; por lo que se realiza sobre un “estimador interno de la varianza del error experimental”, variante más consistente según Guerra y Sevilla (1986).

El valor de estadígrafo de prueba para la falta de ajuste según Montgomery y colaboradores (2002) se determina como:

$$F_o = \frac{SS_{LOF}/(m - 2)}{SS_{PE}/(n - m)} \quad (2.4)$$

Donde:

$F_o$ : Valor del estadígrafo de Fisher calculado

$SS_{LOF}$ : Suma de cuadrados para la falta de ajuste

$SS_{PE}$ : Suma de cuadrados total para el error puro

$n$ : Cantidad de observaciones

$m$ : Niveles de los términos independientes

La suma de cuadrados para la falta de ajuste se calcula como:

$$SS_{LOF} = SS_E - SS_{PE} \quad (2.5)$$

Donde:

$SS_E$ : Suma de cuadrados del error

$SS_{PE}$ : Suma de cuadrados total para el error puro

Si se satisface el criterio de varianza constante,  $SS_{PE}$  es una medida del error puro (independiente) del modelo porque solo se usa la variabilidad de las observaciones para cada nivel de los términos independientes para calcularla (Montgomery et al., 2002).

La suma de cuadrados del error se determina como:

$$SS_E = \sum_i^n (y_i - \hat{y}_i)^2 \quad (2.6)$$

Donde:

$y_i$ : Valor observado de la variable independiente

$\hat{y}_i$ : Valor predicho de la variable independiente

La suma de cuadrados total para el error puro se determina como:

$$SS_{PE} = \sum_{i=1}^m \sum_{u=1}^{n_i} (y_{iu} - \bar{y}_i)^2 \quad (2.7)$$

Donde:

$\bar{y}_i$ : Media de los valores observados de la variable dependiente para un nivel de los términos independientes

$y_{iu}$ : Valor observado de la variable dependiente para un nivel de los términos independientes

El valor de estadígrafo de prueba para la falta de ajuste tabulado se determina mediante la tabla de distribución correspondiente como:

Donde:

$$F_{\alpha, m-2, n-m} \quad (2.8)$$

$\alpha$ : Nivel de significación, en este caso 0,05 para una confianza del 95 %.

Si el valor del estadígrafo de Fisher calculado es menor que el tabulado, se infiere que la pérdida de ajuste no es significativa y que no hay evidencias de que el modelo no sea adecuado en este sentido (Guerra y Sevilla, 1986). Esta prueba se implementa en el *software* RStudio.

Montgomery y colaboradores (2002) señalan otra medida relativamente sencilla de la eficiencia potencial de predicción, que se determina a través de la comparación entre el recorrido de los valores ajustados con su error estándar promedio.

El rango de los valores ajustados se determina por la expresión siguiente:

$$\hat{y}_i = \hat{y}_{mxm} - \hat{y}_{mnm} \quad (2.9)$$

Donde:

$\hat{y}_i$ : Rango

$\hat{y}_{mxm}$ : Valor ajustado máximo

$\hat{y}_{mnm}$ : Valor ajustado mínimo

Mientras que el error estándar promedio se determina como:

$$e_{sp} = \sqrt{p \cdot \hat{\sigma}^2 / n} \quad (2.10)$$

Donde:

$e_{sp}$ : Error estándar promedio

$p$ : Cantidad de parámetros del modelo

$\hat{\sigma}^2$ : Estimado de la varianza del error independiente del modelo (en este caso,  $SS_{PE}$ )

$n$ : Cantidad de observaciones

#### 2.2.4 Definición del intervalo de aplicación de los modelos de regresión

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 disponga de información suficiente para contemplar efectos entre diferentes niveles de los términos independientes y puede dar lugar a predicciones desacertadas por "extrapolación oculta" (Montgomery *et al.*, 2002), tal como se muestra en la figura siguiente:

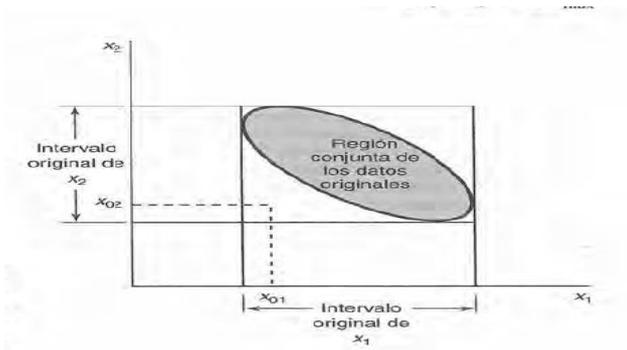


Figura 1 Región efectiva de predicción

Existe un método presentado por Montgomery y colaboradores (2002), que aunque no garantiza totalmente la definición de la región de predicción efectiva, permite analizar 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*). Se tiene que se está en presencia de extrapolación si el punto para la predicción con nuevas observaciones está fuera de la RVH.

Los elementos diagonales ( $h_{ii}$ ) de la matriz “sombrero” de la regresión ( $H$ ) son útiles para detectar la extrapolación oculta. Los valores de esta dependen tanto de la distancia euclidiana del punto al centroide, como de la densidad de los puntos en la RVH. En general, el punto que tiene mayor valor de  $h_{ii}$  (valor máximo) está en la frontera de la región RVH del espacio ( $X$ ) donde la densidad de las observaciones es relativamente baja (Montgomery *et al.*, 2002).

En base a estos elementos se define por el autor la siguiente restricción para no incurrir en extrapolación oculta (por este sentido):

$$x^{-1}(X'X)^{-1}x \leq h_{m\acute{a}x} \quad (2.11)$$

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.

La matriz sombrero matemáticamente se define por la expresión siguiente (Montgomery et al., 2002):

$$H = X^{-1}(X'X)^{-1}X \quad (2.12)$$

Esta prueba también aparece en Carmona (2003). Se implementa en el *software* Rstudio en donde se recurre al comando “hatvalues” para obtener los elementos diagonales de la matriz “sombrero” de la regresión.

A partir de la implementación de la prueba se realizan análisis para las combinaciones de las variables independientes para acotar el dominio de aplicación de cada uno de los modelos de regresión a través del *software* Matlab con los resultados que se obtienen de la prueba implementada en Rstudio).

Para ello se emplea como límites de los valores para el análisis los valores máximos y mínimos observados de los términos independientes del modelo y se evalúan los puntos comprendidos entre estos valores que surgen del mallado de los valores incluidos en tal intervalo.

En el caso de la regresión para el modelo de brix del jugo mezclado y del bagazo, estos comparten exactamente el mismo procedimiento, pues poseen iguales variables independientes con igual número de observaciones y valores experimentales de estas; solo se diferencian entre sí por la variable dependiente observada.

### **2.2.5 Análisis de sensibilidad**

Se realizan análisis de sensibilidad de la respuesta ante la variación de forma individual de cada uno de los predictores contemplados en el modelo. Para ello se define un punto central por una relación de agua de imbibición de 2 según el peso de la fibra, 60 °C de temperatura de agua de imbibición y un jugo primario con 19 °Bx. Se evalúa la variación de la predicción ante desviaciones de 5, 10 y 15 % de los términos independientes. Este análisis se efectúa mediante el *software*.

### **2.3 Análisis de la etapa de extracción a partir de los modelos de regresión**

Se realizan diferentes análisis que permitan evaluar el comportamiento de la etapa para diferentes condiciones de imbibición. Mediante la predicción de los modelos de regresión y de balances de masa se determinan parámetros operacionales que constituyen una realidad simulada de la etapa de extracción. A su vez, como parte derivada de este análisis se evalúa desde un punto de vista operacional la predicción de los modelos de regresión; etapa de gran importancia para la validación de estos. Tales análisis se implementan en el *software* Rstudio.

Es importante aclarar que en la presente investigación solo se analiza la extracción de sólidos disueltos (dado los modelos que se obtienen).

Se definen como parámetros fijos una molida de 150 000 kg/h y una composición de fibra en caña del 17 %. Se establecen como términos independientes a la cantidad del agua de imbibición (referida como  $\lambda$ ), a la temperatura del agua de imbibición, al brix y la pol del jugo primario. Como principal variable operacional en el área se tiene a la extracción de brix, la que se calcula como:

$$E_{Bx} = \frac{Bx_{JM} \cdot JM}{Bx_{CM} \cdot CM} \cdot 100 \% \quad (2.13)$$

Donde:

$E_{Bx}$ : Extracción de brix (%)

$JM$ : Flujo de jugo mezclado (kg/h)

$Bx_{CM}$ : Porcentaje de brix de la caña (%)

$CM$ : Flujo de caña molida (kg/h)

El porcentaje de brix del jugo mezclado se determina mediante el modelo empírico correspondiente a partir del conocimiento de la relación del agua de imbibición referida según el peso de la fibra en caña, de la temperatura del agua de imbibición y el brix del jugo primario.

El flujo del jugo mezclado se calcula mediante un balance de masa general y uno por composición de fibra, tal que:

$$JM = CM + AI - RL \quad (2.14)$$

Donde:

$AI$ : Flujo de agua de imbibición (m<sup>3</sup>/h)

$RL$ : Flujo de residuo leñoso (kg/h)

A su vez, el flujo de bagazo se determina como:

$$RL = \frac{f_{CM} \cdot CM}{f_{RL}} \quad (2.15)$$

Donde:

$f_{CM}$ : Porcentaje de fibra de la caña molida (%)

$f_{RL}$ : Porcentaje de fibra del bagazo (%)

El  $f_{RL}$  se determina mediante la expresión siguiente:

$$f_{RL} = 100 - Humd - Bx_{RL} \quad (2.16)$$

Donde:

$Humd$ : Humedad del residuo leñoso (%)

$Bx_{RL}$ : Brix del residuo leñoso (%)

Los términos de humedad del bagazo y el brix del residuo leñoso se determinan a partir de los modelos empíricos obtenidos en esta investigación.

El término de brix de la caña molida se determina a partir de un balance de masa, tal que:

$$Bx_{CM} = \frac{Bx_{JM} \cdot JM + Bx_{RL} \cdot RL}{CM} \quad (2.17)$$

## 2.4 Desarrollo de gráficos para el manejo operacional de la imbibición a partir de los modelos de regresión

Se establecen diferentes gráficos para viabilizar el empleo de los modelos matemáticos de regresión que se obtienen. Ello permite que puedan utilizarse directamente en el campo operacional y que las herramientas matemáticas estén al alcance de todo el personal técnico que la requiera.

Para el caso de la humedad del residuo leñoso, como depende solamente de la relación de imbibición  $\lambda$  y de la temperatura del agua de imbibición, su representación gráfica se realiza de forma muy simple en el plano "Humedad del bagazo- $\lambda$ " según funciones específicas correspondientes a diferentes valores de temperatura.

Sin embargo, para los modelos de brix del jugo mezclado y del bagazo la representación es más compleja, dado que dependen de tres términos independientes. Con el objetivo de obtener gráficos de fácil manejo e interpretación, la representación de estos modelos se realiza mediante dos gráficos dependientes entre sí.

Por ello se establece el término "número de imbibición" que se define como el valor positivo de la suma de los términos correspondientes a  $\lambda$  y a la temperatura del agua de imbibición; siendo específico para cada variable dependiente, necesitando un gráfico para su determinación dada las condiciones de imbibición y de otro que, con el conocimiento y con curvas específicas para cada brix del jugo primario permitan la estimación del brix del jugo mezclado o del bagazo.

A su vez, dado que la variable  $\lambda$  no es una información directa de proceso, se establecen dos gráficos para su rápida determinación en caso de que se requiera.

Los gráficos se construyen en Rstudio, el que permite el desarrollo de herramientas de este tipo de buena calidad y visualmente atractivas.

## Capítulo 3: Análisis de Resultados

En este capítulo se desarrolla el análisis de los resultados de la investigación. Se abordan los resultados de la modelación estadística a partir de los resultados de la experimentación en el área de extracción del CAI: "Jesús Rabí" y de la predicción del comportamiento de la etapa a través de tales modelos. Se presentan diferentes herramientas gráficas para un mejor manejo práctico de los modelos matemáticos.

### 3.1 Resultados experimentales

Se cuenta con 107 observaciones de las variables de procesos de interés para el desarrollo de la experimentación. Dentro de ellas se disponen con cuatro condiciones de molienda replicadas, lo que permite la estimación interna (que es más consistente) de la varianza del error experimental para la prueba de bondad de ajuste de Fisher (Guerra y Sevilla, 1986).

Se considera al tamaño de la muestra aceptable, ya que presenta una relación de número de observaciones por cantidad de predictor superior a lo referido como criterio general en la literatura especializada. Meyers y colaboradores (2016) señalan que la cantidad de observaciones debe satisfacer en 20 veces al número de predictores mientras Harrell (2015) acota que puede ser desde 10 veces hasta este valor, como aproximación a la mínima cantidad requerida. En este caso para el modelo de la humedad del bagazo se tiene una relación de 53,5 mientras que para los restantes de 35,7.

En la tabla 1 se muestran los resultados de la caracterización estadística preliminar de la información experimental:

Tabla 1: Resultados de la caracterización estadística de las variables operacionales observadas.

Parámetro	Variables experimentales					
	$\lambda$	Temp (°C)	BxJP (°Bx)	BxJM (°Bx)	BxRL (%)	Humd (%)
Valor máximo	3,29	86,00	23,5	20,70	9,24	51,23

Valor mínimo	0,57	32,00	17,60	12,15	2,24	47,54
Rango	2,72	54,00	5,90	8,55	7,00	3,69
Media aritmética	1,73	60,51	19,93	15,58	4,98	49,38
Mediana	1,74	60,00	19,60	15,21	4,74	49,40
Coeficiente de variación (%)	36,86	22,91	6,58	11,67	32,43	1,50

El conocimiento de estos parámetros es de gran importancia para la evaluación y crítica de la colección de datos (realidad observada) sobre la cual se efectúa la modelación matemática.

La cantidad de agua de imbibición, referida por la relación " $\lambda$ ", presenta una dispersión de 36,86 %. Es deseable que la dispersión sea baja o moderadamente baja, en conjunto con una equitativa distribución de las observaciones a lo largo del rango, lo que se analiza según la simetría del diagrama de cajas y bigotes. Ello contribuye a que se disponga de una uniforme y suficiente información de la realidad experimental en todo el dominio de definición de los modelos. De modo contrario existirían intervalos mucho más descritos que otros y el ajuste matemático podría estar encaminado hacia minimizar los errores de la predicción principalmente en las zonas más pobladas sin que presente suficiente significación al respecto las otras menos representadas.

El valor referido indica una variación alta si se compara con el límite generalmente empleado (12 %) y se considera lógica de acuerdo con el comportamiento de esta variable en la práctica. Este parámetro operacional es de por sí mismo muy variable, ya que depende por definición de tres variables igualmente cambiantes y a su vez, su valor normalmente se modifica de acuerdo con el comportamiento operacional de otras. Se cambiaron intencionalmente el valor de agua de imbibición de algunas observaciones para obtener cierta información distante de las condiciones normales de operación en la construcción de los modelos.

La media y la mediana presentan valores muy semejantes, por lo que la muestra está centrada y en este caso, en valores de la zona normal de operación. Al respecto se aprecia una región muy poblada

en torno a 1,90 kg/kg y una buena representación de esta en sus proximidades, por lo que se infiere una adecuada representación de la realidad sobre las condiciones a las cuales están encaminados mayormente la predicción de los modelos en su uso práctico con relación a  $\lambda$ . Un 50 % de los datos se localiza entre 1,2 y 2.

Otra de las razones por la que se atribuye este comportamiento es que desde un punto de vista práctico en la realización de la experimentación a nivel industrial (para la coleccionar los datos del proceso) es menos riesgoso para el buen desenvolvimiento del proceso por cuestiones de capacidad, la disminución moderada del flujo de agua de imbibición que su aumento.

Se detecta un punto atípico ( $\lambda = 3,29$  kg/kg) ya que presenta más de 1,5 veces el rango intercuartil (ancho de caja); sin que sea considerado extremo (no presenta 3,0 veces el rango intercuartil). No se considera su exclusión siempre que no sea muy influente (y que por lo tanto represente una interferencia sería en el ajuste matemático); para esta cuestión, se tiene implementada una prueba en el análisis de regresión.

La temperatura del agua de imbibición tiene una dispersión menor y mejor distribuida a lo largo del rango de acuerdo con la simetría del diagrama de cajas y bigotes y a la densidad de la nube de observaciones. En esta variable, el valor central es próximo a la media de 60,51°C, destacable dado que la temperatura del agua de imbibición usualmente se recomienda en torno a este valor (Ordóñez, 2004).

El comportamiento del brix del jugo primario se encuentra poco disperso y mejor representado en la zona en torno a 20 °Bx respectivamente. Se aprecian puntos atípicos, aunque no extremos. Los puntos atípicos se deben precisamente al procesamiento de cañas con elevado brix puntualmente en esas observaciones y no constituyen un comportamiento habitual de este parámetro en la práctica ingenieril.

Esta no es una variable manipulable en la experimentación y depende de la variedad botánica de la caña, el entorno del cultivo y la planificación del corte. Se considera aceptable las observaciones de este parámetro para el desarrollo de los modelos.

Con relación a las variables respuesta, se evidencia que al menos un 50 % de los datos se encuentran dispersos en la zona de cumplimiento con las normas operacionales; sin que ello implique déficit de información en la vecindad de esta. Se acepta tal comportamiento dado que el principal propósito de los modelos de regresión es permitir el reajuste de los parámetros de operación para el cumplimiento con las mismas.

El rango delimitado por los bigotes de los gráficos correspondientes (no tiene en cuenta los valores atípicos y extremos) acota mejor el intervalo a recomendar para la aplicación de los modelos matemáticos, pero solo en primera instancia. Este se puede restringir más según resultados del análisis de la regresión.

### 3.2 Análisis de resultados de la modelación estadística

En este epígrafe se analizan los resultados de la modelación por regresión lineal múltiple para predecir la composición de brix del jugo mezclado, la humedad del bagazo y el brix del bagazo.

#### 3.2.1 Aplicación de la técnica de regresión lineal múltiple para el ajuste de los datos experimentales.

❖ Modelo para predecir el brix del jugo mezclado

La disposición del brix del jugo mezclado con relación a las variables independientes que se muestra en la figura siguiente:

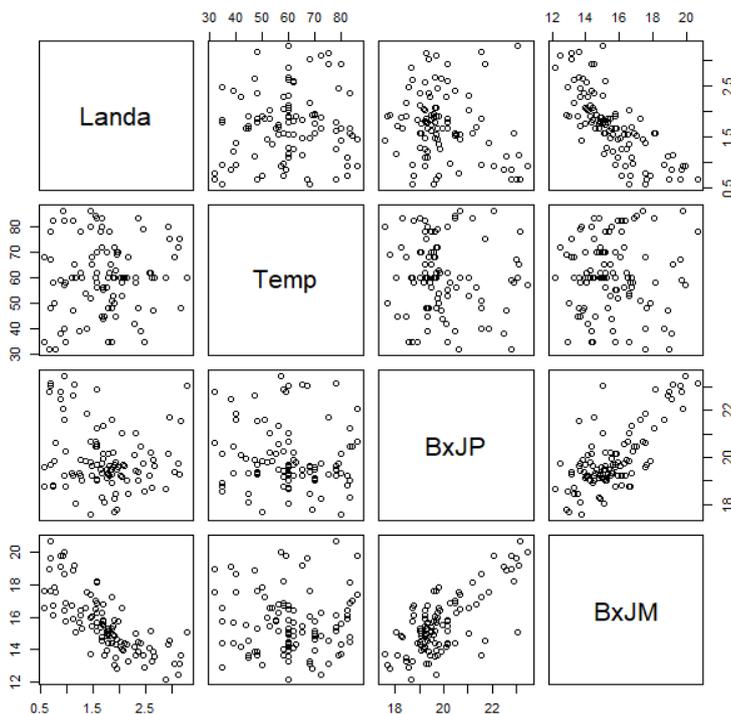


Figura 2 : Interrelación entre los valores de las variables observadas.

Se evidencia una clara dependencia entre el brix del jugo mezclado con la cantidad de agua de imbibición, así con el brix del jugo primario. No se aprecia un patrón bien definido con la temperatura del agua de imbibición y es lógico inferir que tal vez se encuentre enmascarado por los niveles de  $\lambda$ .

El comportamiento con relación a  $\lambda$  y el brix del jugo primario aparenta ser lineal y justifica la propuesta de la estructura siguiente para el desarrollo de la regresión:

$$BxJM = \beta_0 + \beta_1 \cdot BxJP + \beta_2 \cdot \lambda + \beta_3 \cdot T + \beta_4 \cdot \lambda \cdot T \quad (3.1)$$

Donde:

$\beta_0, \beta_1, \beta_2, \beta_3, \beta_4$ : Coeficientes de regresión

Las significaciones de los regresores para este caso se muestran en la tabla siguiente:

Tabla 2: Resultados de la regresión según la expresión 3.1

Coeficiente de:	Estimado	p-valor
Intercepto	1,375	0,104
T	0,006	0,431
$\lambda$	-2,183	1,09e-11
BxJP	0,857	Menor que 2e-16
$\lambda \cdot T$	0,005	0,263
p-valor	Menor que 2,2e-16	
R <sup>2</sup> ajustado	0,9456	

Para una confianza del 95 % solamente son significativos en el modelo los coeficientes de  $\lambda$  y el brix del jugo primario. Por ello se procede a la exclusión de la interacción y a efectuar la regresión para la nueva variante, de lo que resulta:

Tabla 3: Resultados de la regresión

Coeficiente de:	Estimado	P-valor
Intercepto	0,876	0,222
T	0,014	9,84e-06***
$\lambda$	-1,872	Menor que 2e-16***
BxJP	0,858	Menor que 2e-16***
P-valor	Menor que 2,2e-16	
R <sup>2</sup> ajustado	0,9455	

Se evidencia que en este caso la consideración solo de la temperatura como término independiente (sin interacción) resulta que esta variable sea significativa para una confianza del 95 %. Tal comportamiento se verifica de igual modo para la cantidad de agua de imbibición y el brix del jugo primario; de hecho, tienen atribuidos el símbolo “\*\*\*” que establece el *software* para una significación mayor (de acuerdo con su escala).

El p-valor de la regresión es mucho menor que 0,05 lo que indica una relación lineal muy significativa entre los predictores y la variable respuesta, lo que avala en este sentido el empleo de la regresión múltiple para el desarrollo del modelo. El coeficiente de determinación ajustado, a pesar de no ser una medida de calidad de ajuste (Montgomery *et al.*, 2002), presenta un valor muy aceptable e indica que cerca de un 95 % de la variabilidad observada del jugo mezclado se describe a través del modelo. Se

tiene también que la exclusión de la interacción no afecta el valor del coeficiente de determinación ajustado.

De esta manera se obtiene el modelo predictor del contenido de sólidos disueltos en el jugo mezclado siguiente:

$$BxJM = 0,858 \cdot BxJP - 1,872 \cdot \lambda + 0,014 \cdot T + 0,876 \quad (3.2)$$

❖ Modelo para predecir el contenido de humedad del bagazo

La disposición del contenido de humedad con relación a las variables independientes que se muestra en la figura siguiente:

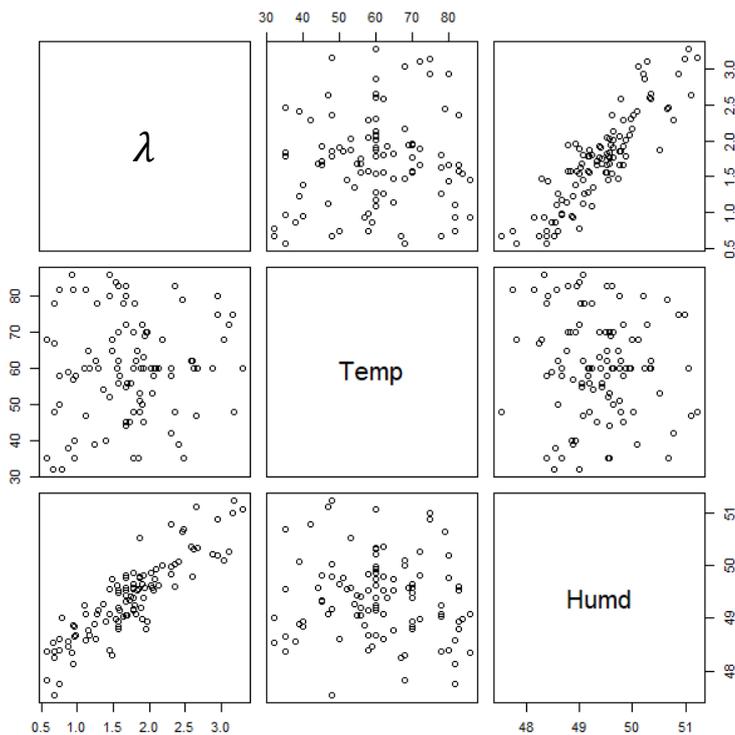


Figura 3: Interrelación entre las variables observadas

En la figura anterior se aprecia un probable comportamiento lineal de humedad del bagazo con la relación del agua de imbibición con relación al peso de la fibra en caña. Por tal motivo, se ajustan los datos a un modelo de regresión de la manera siguiente:

$$Humd = \beta_0 + \beta_1 \cdot \lambda + \beta_2 \cdot T + \beta_3 \cdot \lambda \cdot T \quad (3.3)$$

En donde resulta que:

Tabla 4: Resultados del análisis de regresión

Coefficiente de:	Estimado	p-valor
Intercepto	47,894	Menor que 2e-16***
T	-0,005	0,388
$\lambda$	1,214	3,190e-07***
$\lambda \cdot T$	-0,003	0,431
p-valor	Menor que 2,2e-16	
R <sup>2</sup> ajustado	0,8007	

Para una confianza del 95 % solamente es significativo en el modelo el coeficiente de  $\lambda$  . Por tal motivo se procede a la exclusión de la interacción y se ejecuta la regresión de esta manera, de lo que resulta:

Tabla 5: Resultados de la regresión

Coefficiente de:	Estimado	p-valor
Intercepto	48,156	Menor que 2e-16***
T	-0,010	8,82e-05***
$\lambda$	1,043	Menor que 2e-16***

p-valor	Menor que 2,2e-16
R <sup>2</sup> ajustado	0,8014

La exclusión de la interacción entre la temperatura y la relación de agua de imbibición y, por lo tanto, la asignación de todo el efecto a la temperatura como término independiente hace que este sea estadísticamente muy significativo (“\*\*\*\*”) para una confianza del 95 %. La demostración a través de este análisis de varianza, en base a la realidad observada, que la temperatura del agua de imbibición es influyente de modo significativo sobre esta variable constituye un aporte de la investigación.

El p-valor de la regresión es muy inferior a 0,05 lo que se debe a una relación lineal muy significativa entre los predictores y la variable respuesta y justifica por este sentido al empleo de la regresión lineal múltiple como técnica para el ajuste de los valores observados.

El coeficiente de determinación ajustado indica que se explica a través del modelo obtenido un 80 % de la variabilidad observada. De acuerdo con la variabilidad propia que presenta tal parámetro en la práctica ingenieril y a la exactitud con que se determina analíticamente su valor se considera adecuado. Se aprecia además que la exclusión de la interacción no afecta prácticamente el valor del coeficiente de determinación ajustado.

De esta manera se tiene el modelo predictor de la humedad del bagazo siguiente:

$$\text{Humd} = 48,156 + 1,043 \cdot \lambda - 0,010 \cdot T \quad (3.4)$$

❖ Modelo para predecir el contenido de sólidos solubles del bagazo

La disposición del brix del bagazo con relación a las variables independientes que se representa en la figura siguiente:

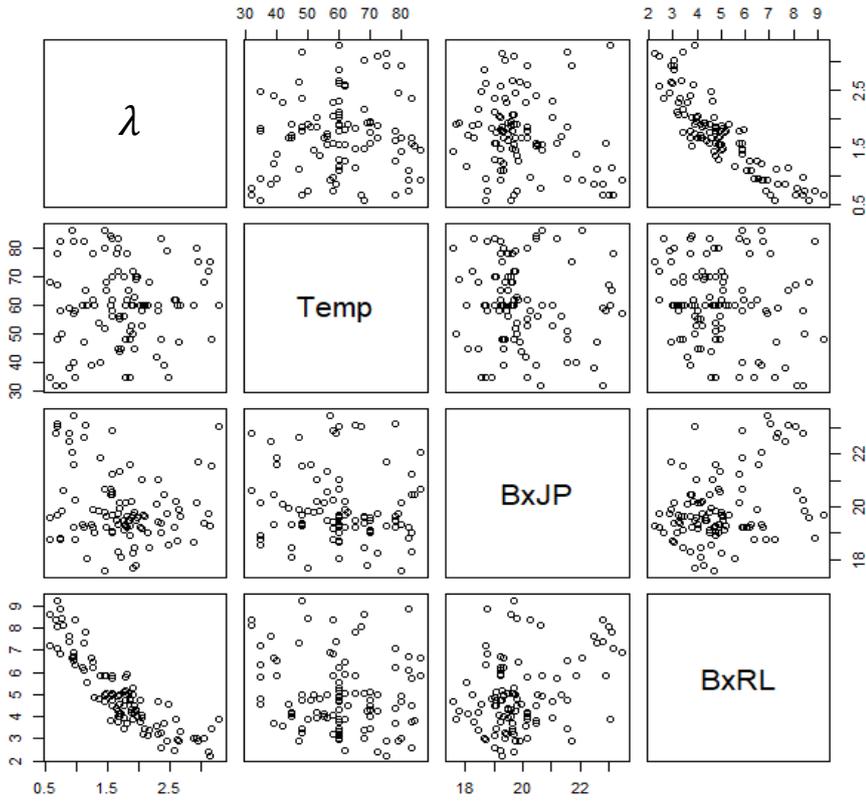


Figura 4: Interrelación entre las variables observadas

En la figura anterior se nota cierto comportamiento no lineal entre  $\lambda$  y la temperatura del agua de imbibición. Por esta razón se evalúan siete posibles estructuras matemáticas para la regresión, definidas a partir de tal disposición gráfica:

$$\text{BxRL} = \beta_0 + \beta_1 \cdot \ln(\lambda) + \beta_2 \cdot T + \beta_3 \cdot \text{BxJP} \quad (3.5)$$

$$\ln(\text{BxRL}) = \beta_0 + \beta_1 \cdot \lambda + \beta_2 \cdot T + \beta_3 \cdot \text{BxJP} \quad (3.6)$$

$$\ln(\text{BxRL}) = \beta_0 + \beta_1 \cdot \ln(\lambda) + \beta_2 \cdot T + \beta_3 \cdot \text{BxJP} \quad (3.7)$$

$$\text{BxRL} = \beta_0 + \beta_1 \cdot \lambda^2 + \beta_2 \cdot \lambda + \beta_3 \cdot T + \beta_4 \cdot \text{BxJP} \quad (3.8)$$

$$\text{BxRL} = \beta_0 + \beta_1 \cdot \sqrt{\lambda} + \beta_2 \cdot T + \beta_3 \cdot \text{BxJP} \quad (3.9)$$

$$\text{BxRL} = \beta_0 + \beta_1 \cdot \frac{1}{\lambda} + \beta_2 \cdot T + \beta_3 \cdot \text{BxJP} \quad (3.10)$$

$$\text{BxRL} = \beta_0 + \beta_1 \cdot \lambda + \beta_2 \cdot T + \beta_3 \cdot \text{BxJP} \quad (3.11)$$

Con relación a la estructura 3.8 se limita el orden del solamente a dos para evitar problemas de sobreajuste, por el principio de simplicidad estructural y por las recomendaciones de Montgomery y colaboradores en el 2002.

Para la selección de la mejor variante se analiza en primer lugar el valor del error cuadrático medio y luego como criterios complementarios el de información de Akaike (AIC) y el bayesiano (BIC). Estos resultados se muestran en la tabla 6.

Tabla 6: Indicadores para la selección del modelo de regresión

Estructura	AIC	BIC	RMSE
3.6	-117,947	-104,629	0,132
3.7	-114,089	-100,771	0,135
3.8	205,968	221,949	0,604
3.5	208,073	221,391	0,616
3.9	217,486	230,803	0,644
3.10	236,595	249,912	0,705
3.11	238,578	251,895	0,711

El RMSE mientras más cercano a cero sea es mejor y permite inferir por este sentido un mejor ajuste (Gharekhani *et al.*, 2021).

El criterio de información de Akaike y el de información bayesiano permiten analizar la predicción en base a la complejidad del modelo matemático (principio de parsimonia), para lo cual penalizan el incremento del número de parámetros del mismo. El mejor modelo en base a estos criterios es el que presente menor valor de AIC y de BIC (Mullor *et al.*, 2019).

Se tiene entonces marcada superioridad del modelo realizado bajo la estructura 3.6 con relación a las estructuras restantes, excepto para el caso de la 3.7, cuya diferencia no es notable. Para este último

caso, la expresión 3.7 presenta un BIC superior pero no significativo. Por tales criterios se selecciona el modelo resultante de la estructura 3.6.

Tal como se realiza para el caso del brix del jugo mezclado y la humedad del bagazo, se evalúa la significación de la interacción entre la cantidad de agua de imbibición y la temperatura, cuyos resultados se muestran en la tabla siguiente:

Tabla 7: Resultados de la regresión para la estructura 3.6 bajo la consideración del efecto de la temperatura con  $\lambda$

Coefficiente de:	Estimado	p-valor
Intercepto	1,404	6.82e-07 ***
T	0,012	0.598380
$\lambda$	-0,290	0.001670 **
BxJP	0,041	0.000132 ***
$\lambda \cdot T$	-0,002	0.125162
p-valor	Menor que 2,2e-16	
R <sup>2</sup> ajustado	0,824	

Del desarrollo de la regresión con la inclusión de este efecto resulta que tal término y el de la temperatura no son estadísticamente significativos en la predicción de la variable dependiente. Por ello se establecen los resultados de la regresión por la estructura 3.6 como modelo final para la predicción del brix del bagazo, que se muestran en la tabla siguiente:

Tabla 8: Resultados de la regresión para la estructura 3.6 sin la consideración del efecto de la temperatura con  $\lambda$

Coefficiente de:	Estimado	p-valor
------------------	----------	---------

Intercepto	1,620	1,310e-07 ***
$\Lambda$	-0,425	Menor que 2e-16 ***
BxJP	0,040	0,000 ***
Temp	-0,002	0,0255*
p-valor	Menor que 2,2e-16	
R <sup>2</sup> ajustado	0,822	

De esta manera se tiene el modelo predictor del brix del bagazo siguiente:

$$BxRL = \text{EXP}(1,620 - 0,425 \cdot \lambda - 0,002 \cdot T + 0,040 \cdot BxJP) \quad (3.12)$$

### 3.2.2 Análisis del cumplimiento de los supuestos de la regresión lineal múltiple.

Resulta de gran importancia analizar el cumplimiento de las suposiciones que se siguen al emplear este método de ajuste; un incumplimiento con algunos de ellos conduce inmediatamente a la invalidación del modelo en cuestión.

#### ❖ Supuesto de normalidad de los residuales

En la tabla siguiente se muestran los resultados de la prueba de normalidad de Kolmogorov-Smirnov con modificación de Lilliefors y de los coeficientes de Kurtosis y asimetría:

Tabla 9: Resultados de las pruebas de normalidad

Modelo para la predicción de:	p-valor de la prueba de Lilliefors	Coficiente de Kurtosis	Coficiente de asimetría
Brix del jugo mezclado	0,272	-0,249	0,011

Humedad del bagazo	0,266	0,274	0,122
Brix del bagazo	0,177	-0,331	0,249

El valor de probabilidad de la prueba de Lilliefors en cada caso es mayor que el nivel de significación (0,05), por lo que se infiere de acuerdo a la misma que los residuales presentan una distribución normal (Priego y Palmer, 2018). Ello se corresponde con los valores de los coeficientes de Kurtosis y de asimetría dado que se encuentran entre -2 y 2 (Sejati *et al.*, 2019). En la figura siguiente se muestran la distribución de los residuales en un histograma de frecuencias y según los cuartiles (gráfico “qq”):

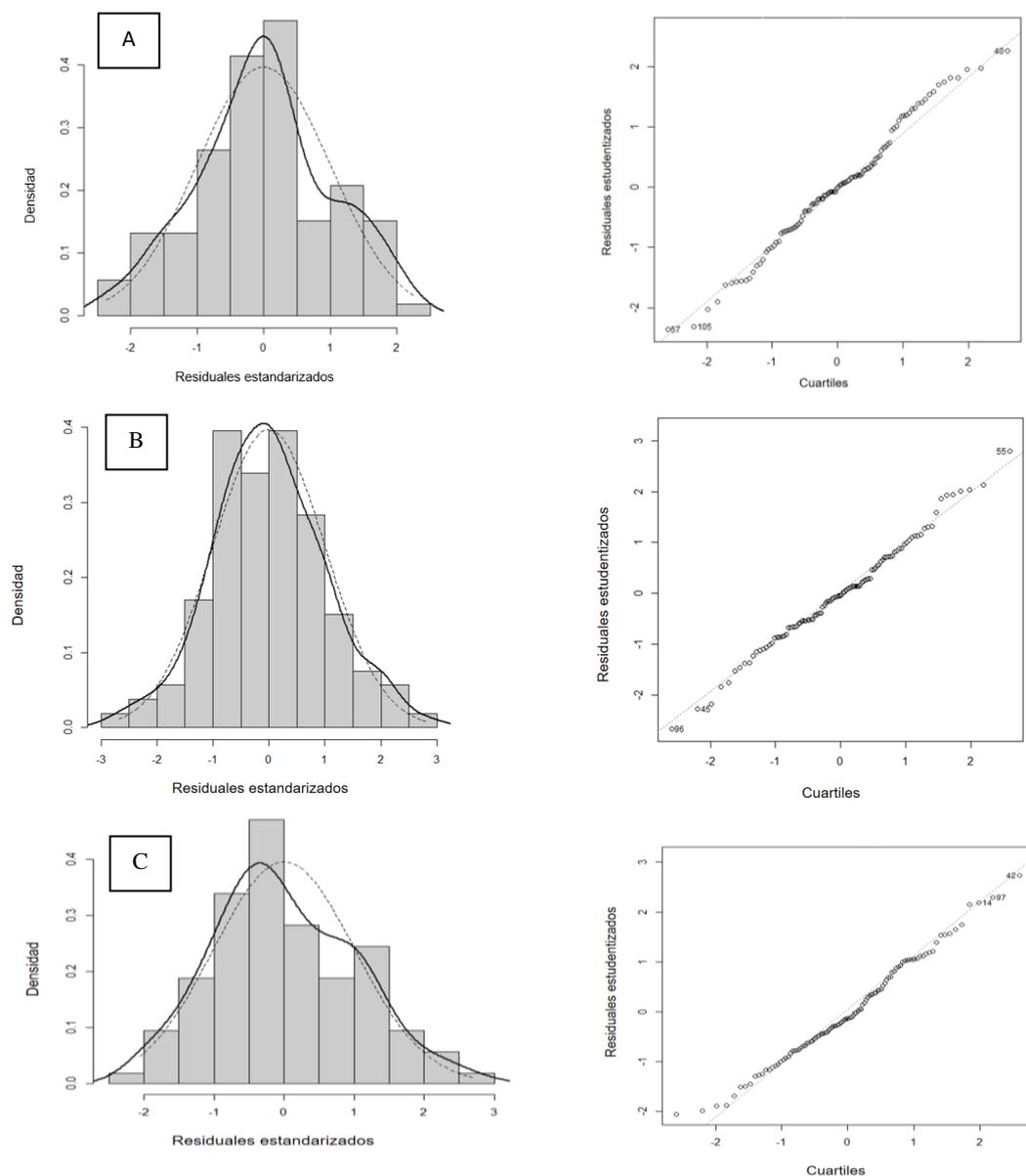


Figura 5: Disposición de los residuales en un histograma de frecuencias y en un diagrama “qq” de los residuales: A- Para el brix en el jugo mezclado B- Para la humedad del bagazo C- Para el brix del bagazo.

Se aprecia que en todos los casos la distribución de densidad descrita por el histograma de frecuencias de los residuales se asemeja bastante a la correspondiente a la distribución normal (líneas discontinuas en la figura). A su vez se disponen favorablemente sobre la línea de distribución normal en el gráfico “qq”, en donde no se aprecia ningún patrón en específico que indique un comportamiento atípico. Por ello, antes los resultados presentados se establece el cumplimiento con el supuesto de normalidad.

❖ Supuesto de homocedasticidad en los residuales

En la tabla siguiente se muestran los resultados de la prueba de Breusch-Pagan para la prueba de homocedasticidad:

Tabla 10: Resultados de la prueba de Breusch-Pagan

Modelo para la predicción de:	p-valor de la prueba
Brix del jugo mezclado	0,260
Humedad del bagazo	0,472
Brix del bagazo	0,193

En base a esta prueba los residuales que se definen a partir de la explicación de los modelos son homocedásticos, ya que el valor de probabilidad de la misma para cada caso es mayor que el nivel de significación (Hernández *et al.*, 2019).

En la figura siguiente se muestra el comportamiento de los residuales según el valor ajustado de la variable predicha:

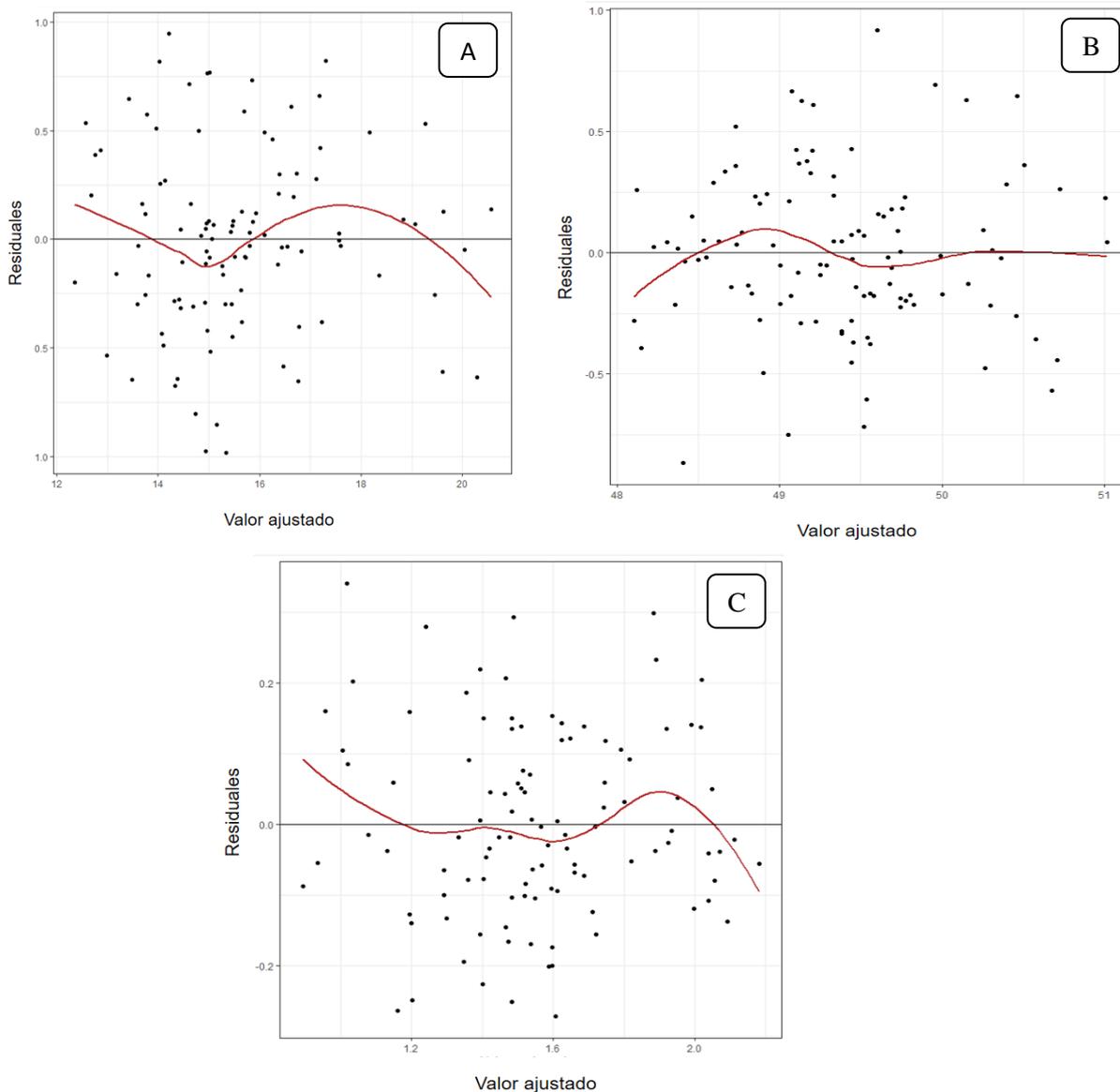


Figura 6: Disposición de los residuales según el valor ajustado: A- Para el brix en el jugo mezclado; B- Para la humedad del bagazo y C- Para el brix del bagazo.

Se aprecia que los residuales se encuentran dispersos de forma relativamente uniforme a lo largo de los valores predichos de las variables dependientes. No se observa algún patrón específico, por ejemplo: forma cónica acentuada o mayor dispersión en los extremos que indique que la variabilidad es dependiente del valor ajustado y por ende heterocedástica.

❖ Supuesto de independencia de los residuales

En la tabla siguiente se muestran los resultados de la prueba de Durbin-Watson:

Tabla 11: Resultados de la prueba de Durbin-Watson

Modelo para la predicción de:	p-valor de la prueba
Brix del jugo mezclado	0,665
Humedad del bagazo	0,902
Brix del bagazo	0,088

En todos los casos el valor de probabilidad de la prueba es mayor que 0,05 por lo que se infiere a partir de este criterio que los errores no se encuentran correlacionados entre sí.

Ello es de gran importancia, no solo para avalar la aplicabilidad de la técnica de modelación sino para evaluar la calidad de la experimentación por este sentido. Permite inferir que todas las variables que influyen sobre la respuesta de forma significativa se mantuvieron controladas en las observaciones o fueron consideradas en la modelación matemática.

Este análisis puede corroborarse de forma gráfica mediante la disposición de los valores observados y predichos de la variable independiente mediante la figura siguiente:

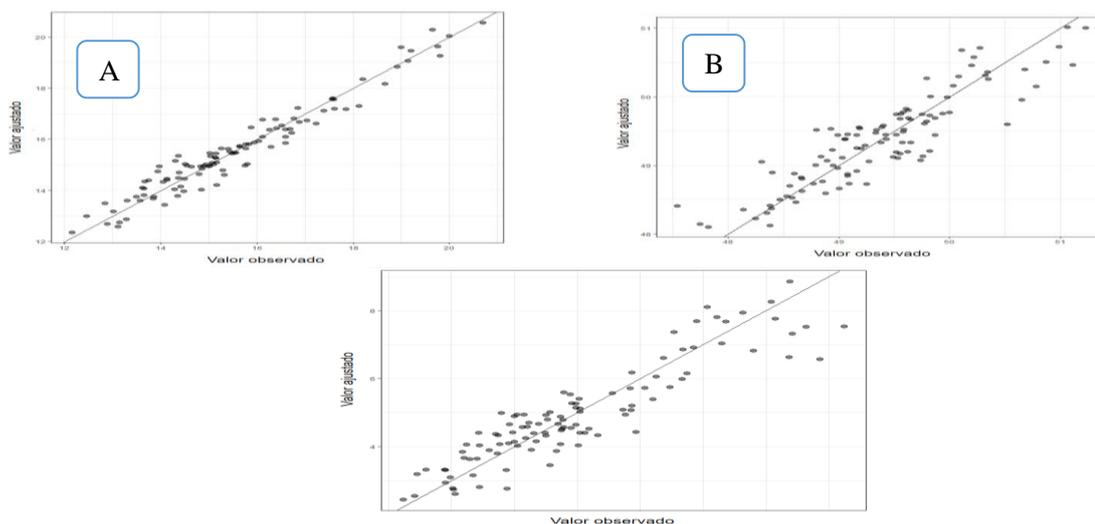


Figura 7: Disposición de los valores observados y predichos de la variable dependiente: A- Para el brix en el jugo mezclado; B- Para la humedad del bagazo y C- Para el brix del bagazo.

No se aprecia ningún patrón que no sea la tendencia hacia la disposición sobre la diagonal de 45°C, lo que ratifica el cumplimiento del supuesto de independencia. A su vez, este comportamiento denota por este sentido un adecuado ajuste de la realidad experimental.

❖ Otras consideraciones

El supuesto de linealidad se comprueba a través de la significación de la regresión mediante el estadígrafo de Fisher, anteriormente analizada en el subepígrafe 3.2.1. A su vez, la media de los residuales la que debe ser cero o aproximadamente a este valor, cuyos resultados se muestran en la tabla siguiente:

Tabla 12: Media de los residuales:

Modelo para la predicción de:	Media de los residuales
Brix del jugo mezclado	1,46e-17
Humedad del bagazo	3,32e-18
Brix del bagazo	2,48e-18

Los valores puntuales de cada uno de ellos se muestran en la figura 6 y denotan la tendencia entorno a cero.

A través de la matriz de correlación entre los predictores se aprecia que no existe una relación fuerte de dependencia entre ellos ya que el coeficiente de correlación se encuentra bastante distante de uno. Otro análisis que se efectúa para probar la no existencia de multicolinealidad es mediante el valor de inflación de la varianza (VIF) (Montgomery y Runger, 2018). Los resultados del VIF para cada modelo se muestran en la tabla siguiente:

Tabla 13: Valores de inflación de la varianza:

Modelo para la predicción de:	VIF		
	Temperatura	$\lambda$	BxJP

Brix del jugo mezclado	1,02	1,04	1,04
Humedad del bagazo	1,01	1,01	-
Brix del bagazo	1,02	1,04	1,04

En este caso como es menor que cinco se infiere por este criterio la no multicolinealidad entre predictores (Montgomery y Runger, 2018). De esta manera se prueba que estos son independientes entre sí.

Otro aspecto importante que debe tenerse en consideración es la no existencia de puntos atípicos influyentes, que pudieran describir una desviación muy alejada de la realidad y que tener mucho peso para el ajuste del modelo

En la figura siguiente se muestran la distancia de Cook de las observaciones para cada caso:

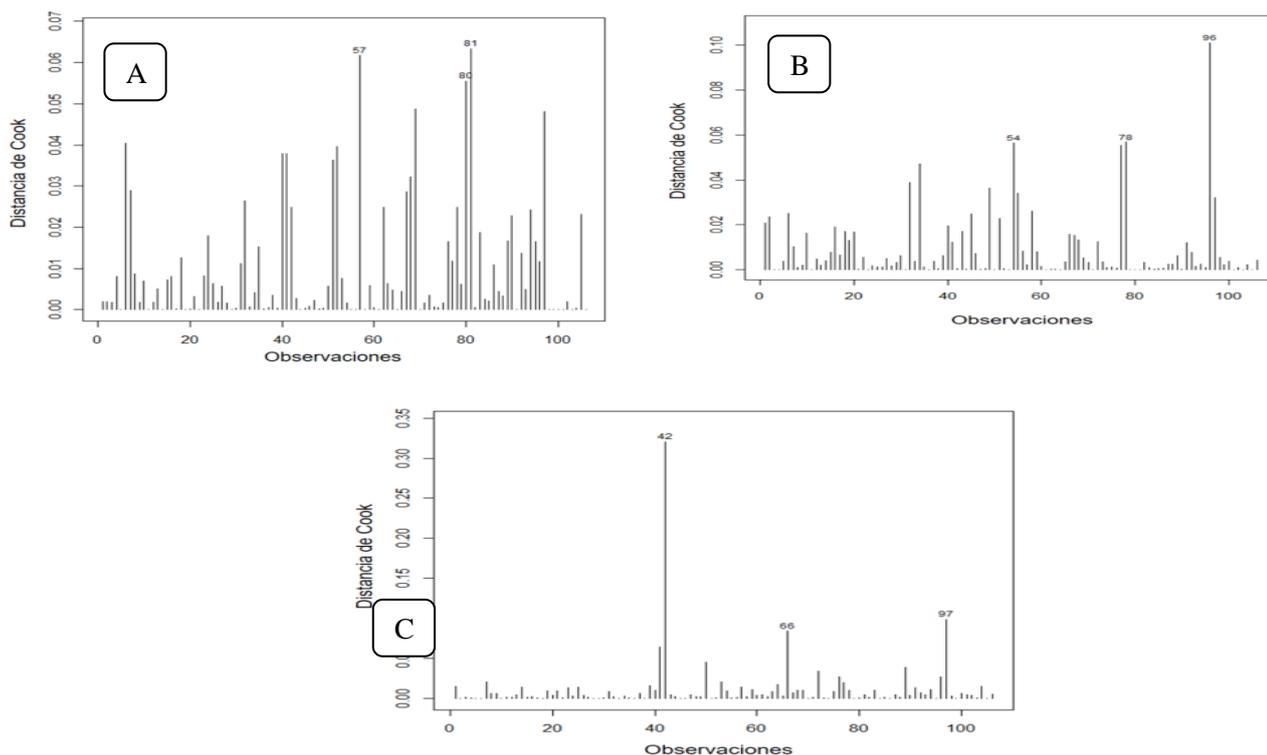


Figura 8: Distancia de Cook de las observaciones: A- Brix en el jugo mezclado; B- Humedad del bagazo y C- Brix del bagazo.

Se aprecia en la figura anterior la existencia de determinados puntos atípicos. Sin embargo, estos no son influyentes ya que la distancia de Cook en todos los casos no es superior a 1 (Liu y Smith, 2020) y se justifica por este sentido su inclusión.

### 3.2.3 Análisis de la calidad de ajuste de los modelos de regresión

Se pueden realizar inferencias con relación a la calidad de ajuste de los modelos matemáticos por análisis anteriores como la prueba de significación de la regresión de Fisher; la disposición de los valores observados y predichos sobre la diagonal de 45°; la tendencia en torno a 0 de los residuales y los valores de RMSE. Sin embargo, estos no son propiamente una prueba “formal” de ajuste, por ello se analiza la prueba de bondad de ajuste de Fisher, cuyos resultados se muestran en la tabla siguiente:

Tabla 14: Resultados de la prueba de bondad de ajuste de Fisher:

Modelo para la predicción de:	F-calculado (Fo)	F-tabulado (Ft)	Ft/Fo
Brix del jugo mezclado	0,965	4,405	4,565
Humedad del bagazo	0,711	4,405	6,200
Brix del bagazo	0,270	4,405	16,327

Para la aplicación de esta prueba es necesario comprobar determinados supuestos que coinciden con los supuestos comprobados para el ajuste de los datos por regresión lineal (Montgomery *et al.*, 2002).

Como el valor del estadígrafo de Fisher calculado es menor que el tabulado se infiere que no hay evidencias desde un punto de vista estadístico que denoten pérdida de calidad de ajuste por los errores de la predicción (Guerra y Sevilla, 1986). Ello significa que no existen diferencias significativas entre los estimados del error “puro” o experimental con los del error por pérdida de ajuste.

Sin embargo, eso no garantiza que el modelo sea satisfactorio como ecuación de predicción. A menos que la variación de los valores predichos sea grande en relación con el error aleatorio, el modelo no se

estima con la precisión suficiente para producir pronósticos satisfactorios. O sea, el modelo puede estar ajustado para los errores solamente (Montgomery *et al.*, 2002).

Según este mismo autor, existen varios trabajos analíticos para desarrollar criterios para juzgar la adecuación del modelo de regresión desde un punto de vista predictivo. Destaca una investigación realizada que refiere que el estadígrafo de Fisher calculado debe ser cuando menos cuatro o cinco veces mayor que el valor crítico de la tabla para que el modelo de regresión sea útil como predictor, esto es, si la dispersión de los valores pronosticados debe ser grande con respecto al ruido. En este caso se aprecia en la tabla anterior que se cumple también con esta relación.

Montgomery y colaboradores en el 2002 plantean que en general no es probable que el modelo sea un predictor satisfactorio a menos que el rango de los valores ajustados sea grande en relación con su error estándar promedio estimado; comportamiento que se verifica en la tabla siguiente:

Tabla 15: Relación del rango de los valores predichos con el error promedio estándar:

Modelo para la predicción de:	Rango	Error estándar promedio	Relación (%)
Brix del jugo mezclado	8,213	0,161	5 103,093
Humedad del bagazo	2,915	0,119	2 452,410
Brix del bagazo	1,288	0,091	1 415,455

De acuerdo con los resultados anteriores se aceptan los modelos de regresión desarrollados desde un punto de vista estadístico como adecuados para describir la realidad experimental.

### 3.2.4 Definición del intervalo de aplicación de los modelos de regresión

Para la definición del intervalo de aplicación se establece en primer lugar un dominio definido por los valores máximos y mínimos de cada término independiente, que se muestran en la tabla 1. Sin embargo, se debe evaluar si existe suficiente información dentro del interior de la región definida por estos valores para explicar todo este espacio y no cometer “extrapolación oculta” al predecir con

valores de dicha región interior, tal como se describe en el capítulo 2. En la tabla siguiente se muestra el valor máximo de los elementos de la diagonal de la matriz sombrero para cada modelo:

Tabla 16: valor máximo de los elementos de la diagonal de la matriz sombrero

Modelo para la predicción de:	$h_{m\acute{a}x}$
Brix del jugo mezclado	0,1463314
Humedad del bagazo	0,07062596
Brix del bagazo	0,1463314

En el caso del modelo del brix del jugo mezclado y del brix del bagazo si el intervalo de aplicación de  $\lambda$  se acota con respecto a  $\lambda$  hasta 3,0 (en lugar del valor máximo observado de 3,29) y el brix del jugo primario hasta 23,0 °Bx (en vez del valor máximo observado de 23,5 °Bx) se cumple que se pueden aplicar en todo en el intervalo resultante sin que se detecte extrapolación. Este mismo análisis se realiza para el modelo de la humedad del bagazo, en que se sustituye al valor mínimo de temperatura de 32 °C por 40 °C y los valores extremos de  $\lambda$  de 0,57 y 3,29 por 0,80 y 2,90 respectivamente.

Tales acotaciones no comprometen la aplicabilidad de los modelos predictivos ya que su nuevo dominio de definición acota a los valores típicos y/o normados de operación, así como a una desviación moderada de estos. En la tabla siguiente se definen el intervalo de aplicación recomendado para cada modelo de forma individual y el intervalo que permite la aplicación de todos en conjunto:

Tabla 17: Intervalo recomendado de aplicación para la modelación:

Modelo para la predicción de:	Temperatura (°C)	$\lambda$	BxJP (°Bx)
Brix del jugo mezclado y brix del residuo leñoso	32-86	0,57-3,00	17,6-23,0
Humedad del bagazo	40-86	0,80-2,90	-

Todos	40-86	0,80-2,90	17,6-23,0
-------	-------	-----------	-----------

### 3.2.5 Análisis de sensibilidad

Se efectúan análisis de sensibilidad para evaluar la influencia que presenta cada variable independiente en la predicción. Parte de este comportamiento se anticipa por el signo matemático del coeficiente estimado de los términos predictores y su significación en cada caso, recogidos en las tablas 3, 5 y 8 según corresponda.

En la siguiente figura se muestra el gráfico de sensibilidad para el contenido de sólidos solubles del jugo mezclado:

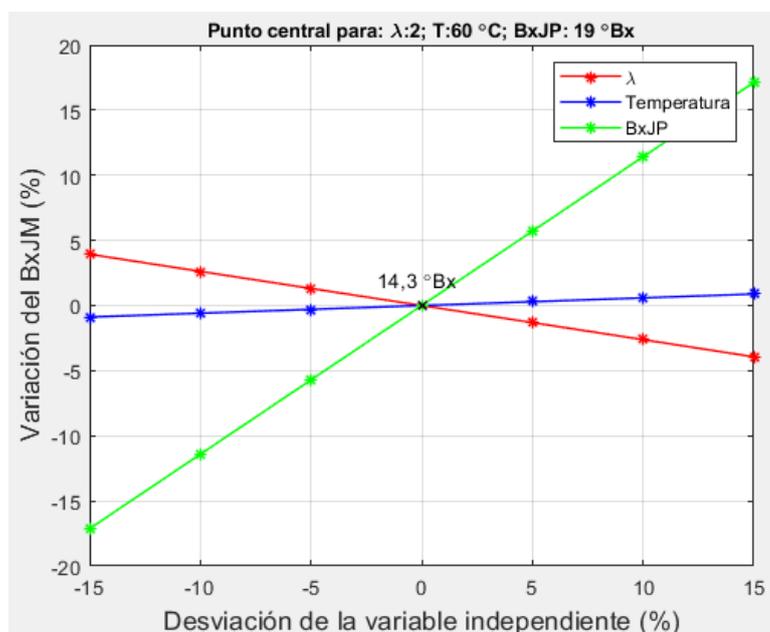


Figura 9: Análisis de sensibilidad para el contenido de sólidos solubles del jugo mezclado

En este gráfico se aprecia que un aumento de la cantidad de agua provoca una disminución del contenido de sólidos disueltos del jugo mezclado, lo que denota una influencia negativa. De igual modo, aunque en menor magnitud se tiene el efecto de la temperatura del agua de imbibición. Caso contrario sucede con el contenido de sólidos disueltos del jugo primario, que tiene la mayor influencia y de modo positivo.

Ello se debe precisamente a que el jugo mezclado está constituido por jugo diluido y jugo primario. El jugo primario posee mucho más flujo y contenido de brix, lo que implica que por el efecto del mezclado este tenga más significación en la variable dependiente que el diluido, el que en última instancia es efecto de la imbibición.

De hecho, un decremento del 15 % de la cantidad de agua de imbibición del punto central hace que se exceda justamente el límite superior normado del brix del jugo mezclado (15 °Bx); mientras que tal efecto se alcanza para un incremento del 5 % del valor del jugo primario. De ahí la importancia de reajustar consecuentemente la cantidad de agua de imbibición según cambien las condiciones de entrada al proceso (en este caso el brix del jugo primario).

En la figura siguiente se muestra el gráfico de sensibilidad para el contenido de humedad del bagazo:

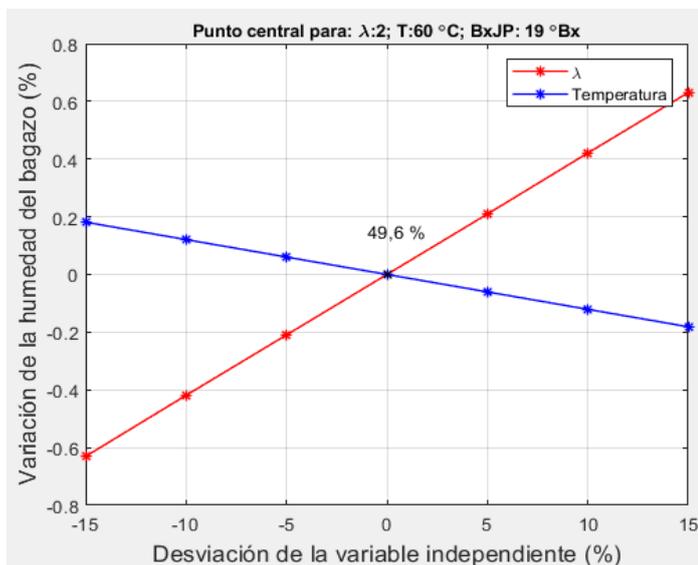


Figura 10: Análisis de sensibilidad para el contenido de humedad del bagazo

En este gráfico se observa que un aumento de la temperatura del agua de imbibición provoca un decremento de la humedad del bagazo (efecto negativo). Ello está en correspondencia con lo referido por Pérez y colaboradores en el 2009, que plantea que a mayor temperatura del agua de imbibición el bagazo sale más caliente y tiende a tener menor humedad por efecto de la autoevaporación del agua contenida.

La cantidad de agua de imbibición influye mucho más que la temperatura y de modo positivo. La consideración de que, por la estructura fibrosa del bagazo, este siempre retenga “aproximadamente” la misma cantidad de líquido para igual presión, solo es aplicable en términos de cantidad de jugo

extraído, cuyas variaciones son despreciables; pero en términos de humedad ello no se aplica (tal como lo denota la significación de esta variable). Castro y colaboradores en el 2019, reconocen el efecto del incremento de la cantidad de agua de imbibición sobre el contenido de humedad del bagazo.

En la figura siguiente se muestra el gráfico de sensibilidad para el contenido de brix del bagazo:

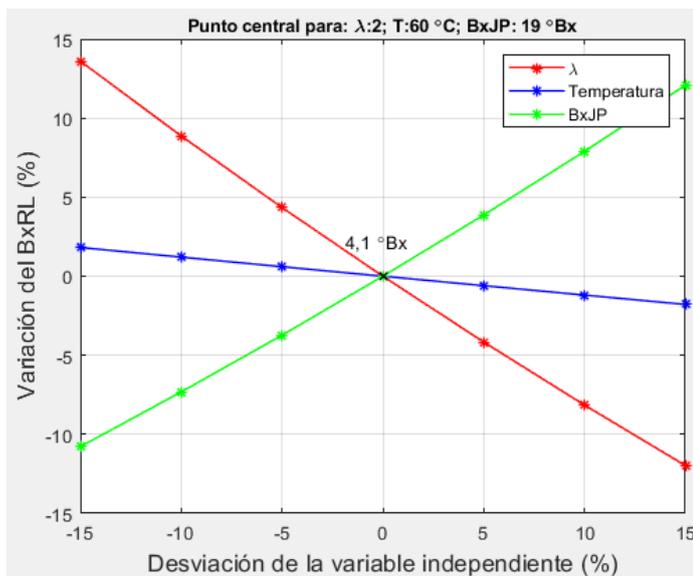


Figura 11: Análisis de sensibilidad para el contenido de sólidos disueltos del bagazo

En este gráfico se aprecia que un aumento de la temperatura implica una disminución del contenido de sólidos disueltos del residuo leñoso; lo que denota un efecto negativo, aunque con menor significación que los restantes términos. Ello concuerda con lo planteado por Hugot (2014), quien plantea que al aumentar la temperatura la eficacia de la imbibición es ligeramente superior.

El brix del jugo primario influye en buena medida y de forma positiva. Este comportamiento es totalmente lógico ya que, para un mismo nivel y eficiencia de la imbibición, a mayor contenido de sólidos disueltos del jugo de caña, mayor es la riqueza de estos en el jugo residual ocluido en la estructura fibrosa del bagazo.

Con un efecto negativo y preponderante se manifiesta la cantidad de agua de imbibición referida según el peso de la fibra. Ello se debe a que, para un mismo brix del jugo primario, a mayor cantidad de agua de imbibición mayor es la dilución del jugo remanente en el bagazo y, por lo tanto, menor es su composición de brix.

Estos análisis también validan la veracidad de la predicción de las variables dependientes que realizan los modelos matemáticos de regresión obtenidos.

### 3.3 Análisis de la etapa de extracción

A partir de los modelos de regresión que se obtienen y de diferentes balances de masa, tal como se describe en el capítulo 2, se establece una realidad simulada correspondiente a la etapa de extracción. En la siguiente figura se muestra el comportamiento de la extracción de sólidos disueltos según la cantidad y temperatura de imbibición para el caso base:

- ✓ Análisis de la etapa de extracción según la cantidad y temperatura del agua de imbibición

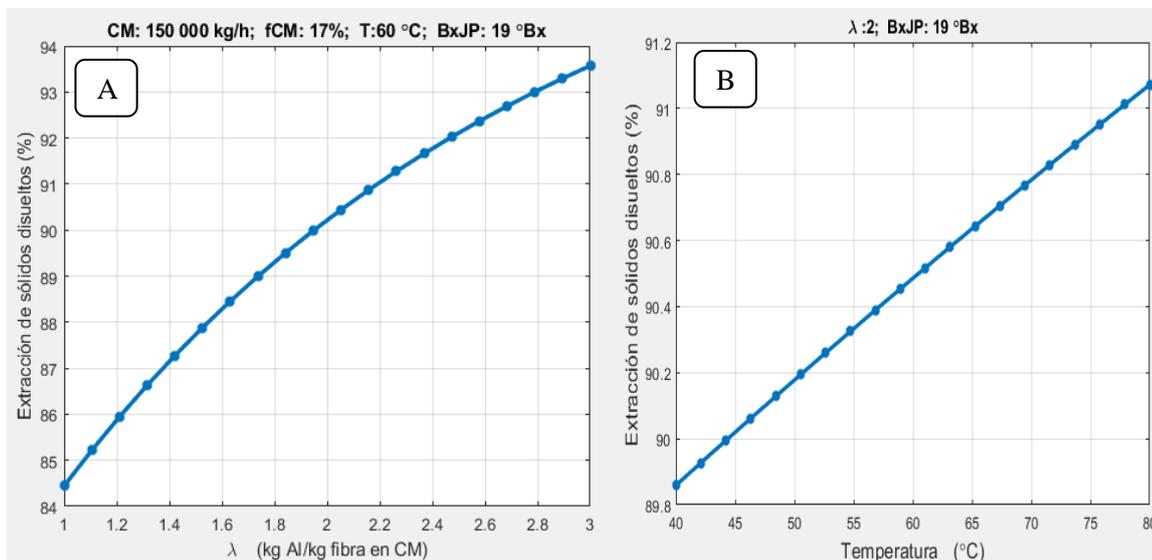


Figura 12: Comportamiento de la extracción de sólidos disueltos: A- Respecto a la cantidad de agua de imbibición referida como  $\lambda$  y B- Respecto a la temperatura del agua de imbibición.

En el primero gráfico (A) se observa un comportamiento casi lineal de la extracción de sólidos disueltos para una relación de  $\lambda$  de 1 a 2, a partir del cual la extracción comienza a presentar cierta desviación a dicha tendencia; lo que denota una disminución gradual de la tasa de extracción para igual incremento en los niveles de  $\lambda$ . En las proximidades de una relación de imbibición de 2,4 se aprecia una razón de la extracción menor que para 2 ante un incremento en la cantidad de agua adicionada.

Lo mencionado anteriormente coincide con lo referido por Hugot (2014), que plantea que a valores de  $\lambda$  entre 1 y 2, la extracción de sólidos disueltos aumenta bastante rápido a media que se incrementa el flujo de agua de imbibición; mientras que por encima de este, la ganancia en la extracción comienza a disminuir hasta que por encima de 3 no resulta ventajoso aplicar más cantidad de agua. Este

comportamiento el autor lo infiere a partir de la modelación fenomenológica que establece para el proceso de extracción. Por ello se recomienda como valores de operación para esta variable 2 o cercanos a este (Cortés *et al.*, 2019).

Con el comportamiento de la figura 12 (A) se coincide adecuadamente, tanto en valores numéricos de extracción como en tendencia con el ofrecido por Wienese (1990) como parte de la modelación fenomenológica que realiza al respecto; esta investigación es referenciada por Rein (2012). Presenta similitud de forma semejante con el de Pedroso (2018b), aunque con valores de extracción un tanto menores y un comportamiento menos acentuado de la linealidad.

El gráfico B de la figura 12 denota una influencia mucho menor de la temperatura sobre la extracción, en donde para todo el dominio de análisis la diferencia es tan solo del 2 %. Ello no implica que no se deba establecer un control estricto sobre esa variable, de hecho, en los resultados del análisis de regresión se aprecia su influencia sobre importantes variables operacionales; sino que con relación a la extracción de brix no resulta tan significativo su variación. No obstante, un 2 % de pérdidas en la extracción por este concepto pueden representar una pérdida económica relativamente razonable y fácilmente evitable.

El comportamiento de la extracción de sólidos solubles con relación a la temperatura coincide con lo referido por Christodoulou (2003). Este plantea que un incremento de 75 a 80 °C en la temperatura ocasiona un incremento en la eficiencia de la extracción en un 0,2 % del azúcar en caña. Bocha (2019) plantea que a valores superiores a 70°C la ganancia en la extracción es en un 0,4 %.

Estos elementos avalan la realidad simulada y, por lo tanto, a los modelos de regresión en este sentido. Le otorga mayor connotación al respecto, el hecho de que la extracción de brix no es una variable que se predice directamente por un modelo, sino que necesita en última instancia de la respuesta en conjunto de todos ellos.

- ✓ Análisis del flujo de jugo mezclado y del flujo, composición de fibra, humedad y el contenido de sólidos disueltos del bagazo con respecto a la cantidad de agua de imbibición.

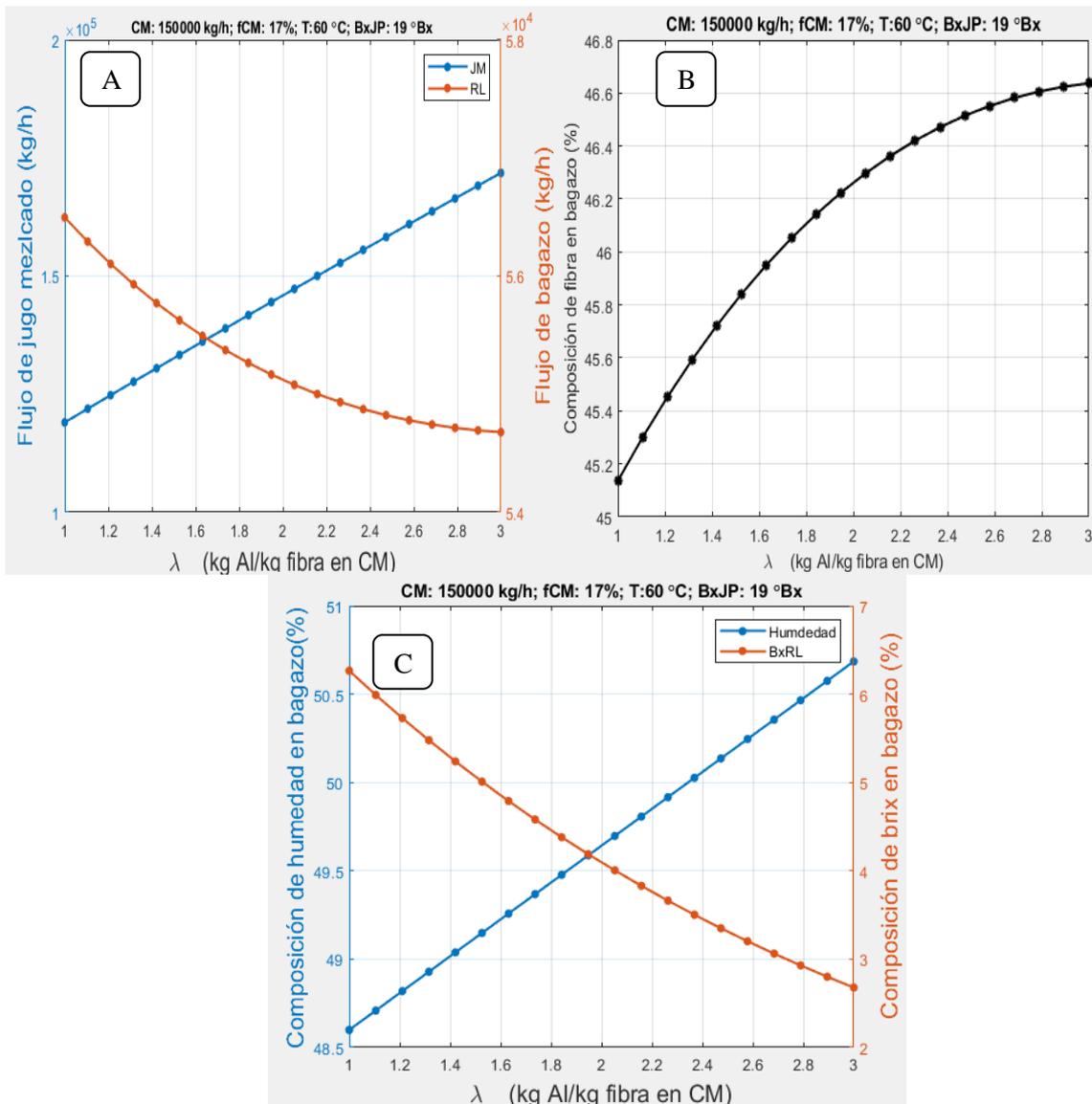


Figura 13: Comportamiento según la cantidad de agua de imbibición del flujo de jugo mezclado y de bagazo (gráfico A), de la composición de fibra en bagazo (B) y de la humedad y el contenido de sólidos disueltos del bagazo (C).

La figura 13 (A) muestra el comportamiento del flujo de jugo mezclado y del flujo de residuo leñoso con respecto a la cantidad de agua de imbibición. El primero presenta un comportamiento lineal ascendente a medida que aumenta la cantidad de agua de imbibición debido a que el residuo leñoso retiene “aproximadamente” la misma cantidad de líquido y la cantidad de agua añadida es devuelta aproximadamente en su totalidad como jugo diluido. Este comportamiento es totalmente lógico y coincide con el referido por Wienese (1994).

El flujo de residuo leñoso (gráfico A) presenta un comportamiento decreciente que se desvía de la tendencia lineal a medida que aumenta la cantidad de agua de imbibición con un rango de variación bastante pequeño. Ello se debe a que el flujo de residuo leñoso solamente depende, de forma inversamente proporcional, de la composición de fibra en residuo leñoso (el flujo de caña molida y su composición de fibra se mantienen constantes en el caso base). Tal como se evidencia en el gráfico B de la figura, existe una total correspondencia de esta variable según la cantidad de agua de imbibición con la manifestada por la composición de fibra del bagazo.

La composición de fibra en bagazo se justifica por el efecto que presenta la humedad y el brix (del bagazo) ante cambios en la cantidad de agua de imbibición (gráfico C). Resulta que al aumentar  $\lambda$ , se incrementa la humedad del bagazo, pero disminuye en mayor magnitud el brix, lo que origina dicha tendencia.

El comportamiento de la humedad del residuo leñoso concuerda con lo referido por González y colaboradores en el 2010 quien plantea que la humedad del residuo leñoso se incrementa ligeramente con el aumento de la cantidad de agua de imbibición. Esta variación de la humedad del residuo leñoso, aunque es menor de un 3 %, es significativa para el intervalo de análisis de la investigación e implica operacionalmente cambios sustanciales en el agarre de las mazas de los molinos y en el índice de la generación de vapor.

Se coincide con el comportamiento lineal evidenciado en el trabajo de Wienese (1995) para iguales niveles de agua de imbibición que el mostrado en el gráfico C bajo similares valores de humedad.

La tendencia descendente y ligeramente desviada del comportamiento lineal de la composición de brix del bagazo frente a la cantidad de agua de imbibición, se asemeja adecuadamente con la presentada por Wienese (1994). Se coincide también en similitud de valores de caída de esta variable para iguales niveles de agua de imbibición que los considerados en el caso de estudio.

En la figura siguiente se muestra el comportamiento del brix de la caña molida con relación a la cantidad de agua y temperatura de imbibición:

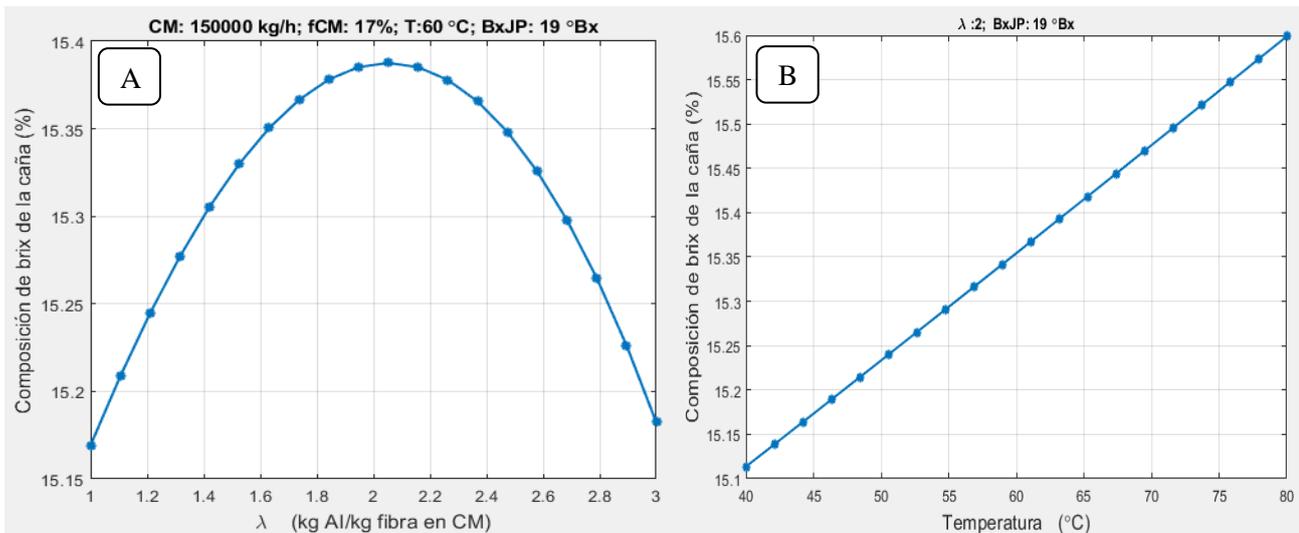


Figura 14: Comportamiento del contenido de sólidos disueltos de la caña: A- Respecto a la cantidad de agua de imbibición; B- Respecto a la temperatura del agua de imbibición.

Es de esperar que el brix de la caña sea independiente de las condiciones de imbibición, ya que solamente depende de las características intrínsecas de la caña. En este caso se aprecia que no es exactamente constante; sin embargo, la variación es totalmente despreciable, de menos de un 0,5 % en todo el intervalo de análisis para el mayor de los casos. El comportamiento de esta variable (que no es directamente estimada) se debe a pequeñas incongruencias entre la explicación en conjunto de los modelos involucrados; hecho que dada su insignificancia no compromete la calidad de la descripción y avala a la modelación en este sentido.

### 3.4 Análisis de los gráficos para el manejo operacional de la imbibición a partir de los modelos de regresión

Se construyen gráficos a partir de los modelos predictivos desarrollados para facilitar su empleo. Ello viabiliza la toma de decisiones operacionales rápidas y eficaces en el reajuste de las condiciones operacionales de la imbibición según cambien las condiciones de entrada.

Para el caso de la determinación del contenido de humedad del bagazo, solamente se requiere de un gráfico (figura 15) dado que se cuentan con tres variables en el modelo.

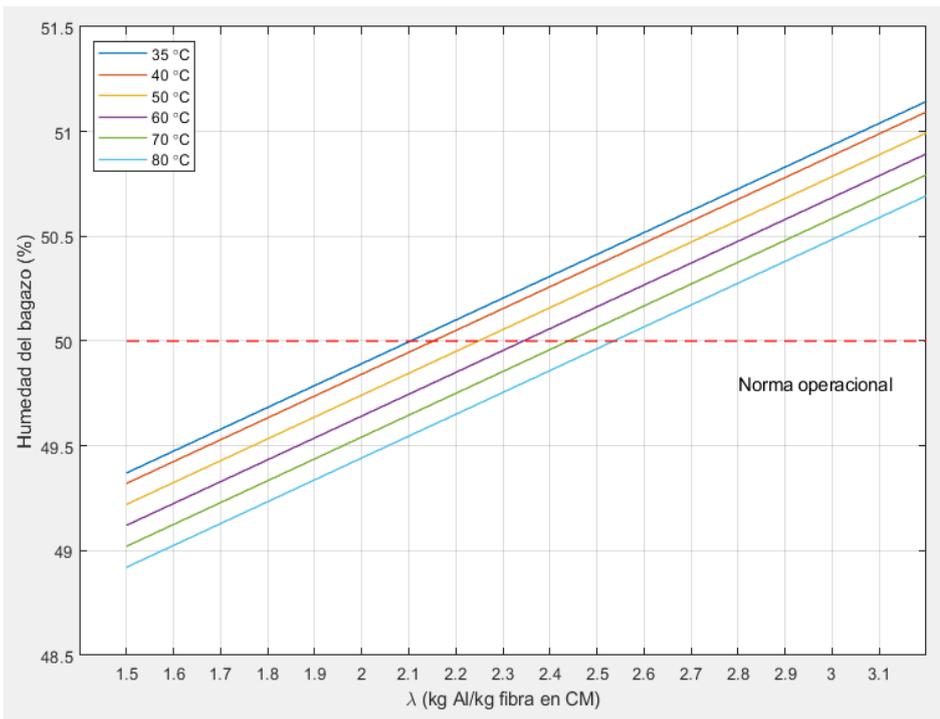


Figura 15: Gráfico para la estimación de la humedad del bagazo.

En este caso, corresponde al valor de las ordenadas del intercepto de la relación de imbibición  $\lambda$  (eje de las coordenadas) con la curva característica para la temperatura del agua de imbibición de operación. Se tiene representada una línea discontinua que delimita el límite de cumplimiento con el valor normado.

Para el caso de la determinación del jugo mezclado se requiere de dos gráficos: uno que relaciona a las variables de operación de la imbibición con el término definido como “número de imbibición” (gráfico de imbibición) y otro que establece la correspondencia entre esta última variable con el parámetro de interés (gráfico de estimación). La figura 16 representan a estos gráficos según corresponda:

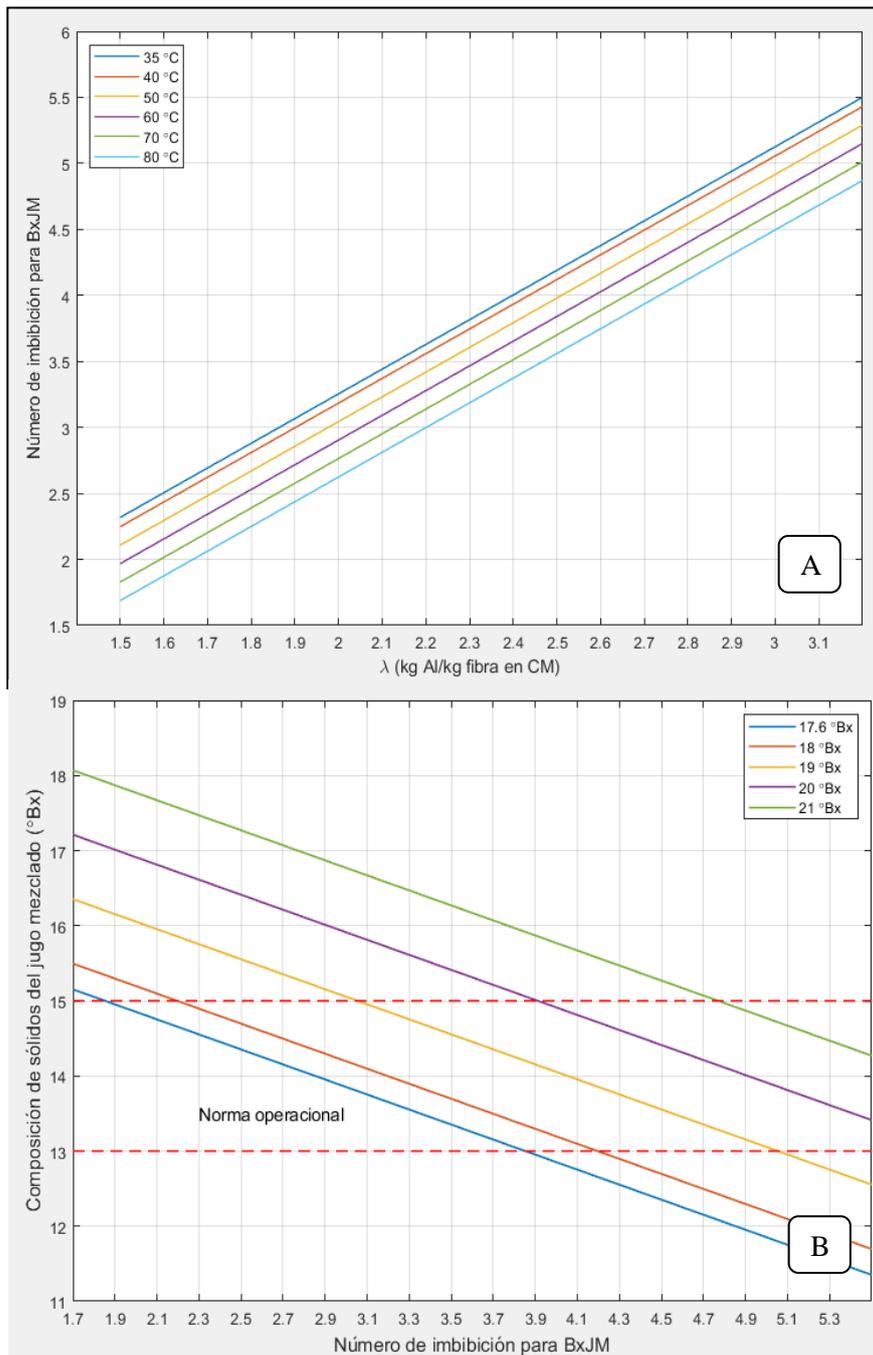


Figura 16: Gráficos para la determinación del brix del jugo mezclado: A- Gráfico de imbibición; B- Gráfico de estimación.

En este caso a través del gráfico A se determina el número de imbibición mediante la ordenada del intercepto del valor de  $\lambda$  con la temperatura de imbibición. Luego con el número de imbibición y el brix del jugo primario (curvas específicas) se establece el valor del brix del jugo mezclado. Igualmente se señala la zona de cumplimiento con lo normado tecnológicamente.

La figura siguiente muestra los gráficos necesarios para la determinación del brix del bagazo, para lo cual se sigue un procedimiento análogo al descrito en el caso anterior:

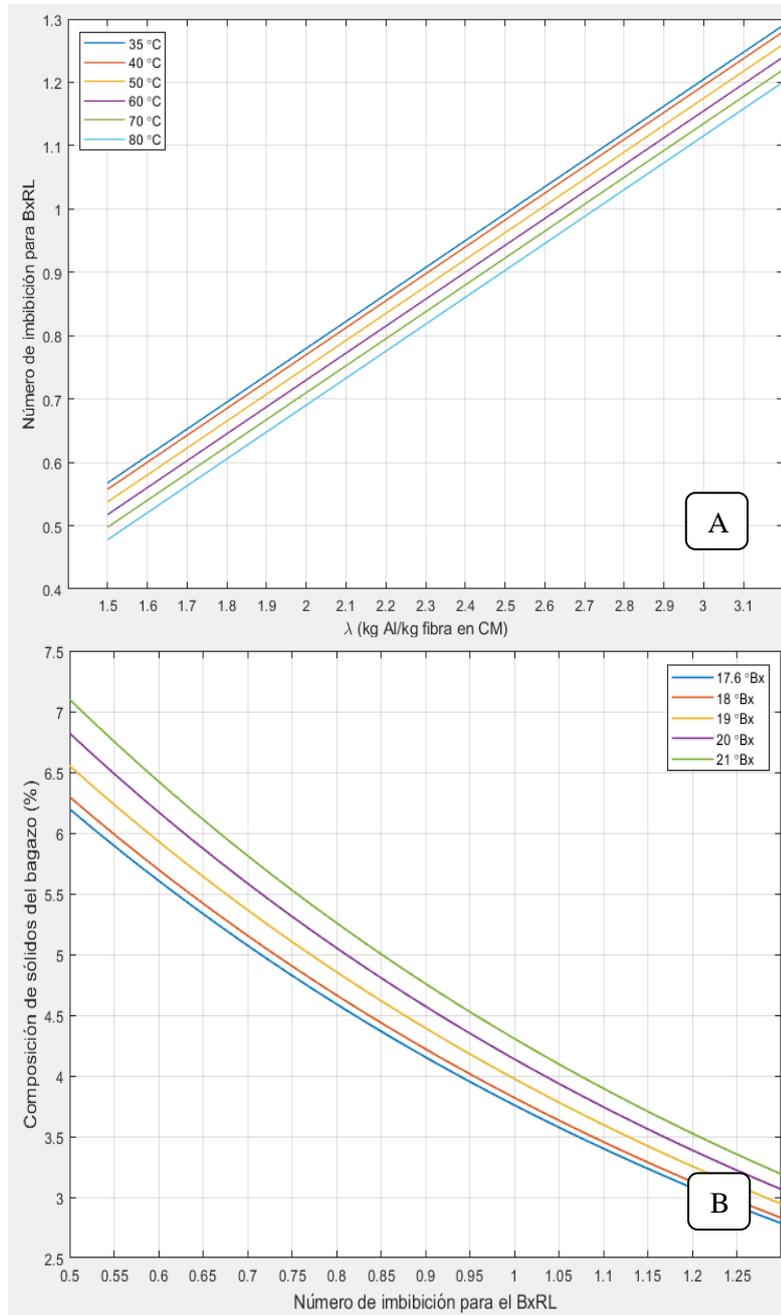


Figura 17: Manejo de la imbibición para estimar la composición de sólidos del residuo leñoso. A- Estimación del número de imbibición; B- Estimación de la composición del contenido de sólidos disueltos del residuo leñoso.

El empleo de estos gráficos facilita la determinación de algunas de las variables operacionales de interés en condiciones de molida inestable, cuando no existe uniformidad en el colchón de bagazo o en

otras condiciones donde resulta difícil e impreciso estimar el flujo de caña molida y por consiguiente la relación  $\lambda$ . Para ello si se conoce el valor del brix del jugo mezclado y primario, cuya determinación es muy fácil en el laboratorio y hasta inclusive en el campo a través de un refractómetro manual, se puede determinar a través del gráfico de estimación específico (figura 16-B) el valor del “número de imbibición”. Luego mediante este valor y el de la temperatura del agua de imbibición se define el valor de  $\lambda$  en el “gráfico de imbibición”. Finalmente, con el valor de este parámetro determinado pueden estimarse las variables operacionales restantes.

### **Conclusiones parciales del capítulo:**

1. No existen evidencias desde un punto de vista estadístico de pérdida de calidad de ajuste por los errores experimentales en los modelos de regresión ajustados por regresión lineal múltiple.
2. El modelo predictivo para el brix del jugo mezclado y del brix del bagazo son aplicables para un valor de temperatura entre 32 y 86 °C, una relación de agua de imbibición según el peso de fibra entre 0,57 y 3,00 y un brix del jugo primario entre 17,6 y 23 °Bx. En cambio, el modelo predictor de la humedad del bagazo es válido a una temperatura entre 40 y 86 °C y una relación de imbibición de 0,80 a 2,90.
3. No resulta conveniente desde un punto de vista extractivo aplicar niveles de agua de imbibición según el peso de la fibra en caña superiores a 2,4.
4. La temperatura del agua de imbibición es un factor que influye de forma significativa sobre la humedad del bagazo y sobre el contenido de sólidos disueltos del jugo mezclado y bagazo; y presenta una influencia menor sobre la extracción de azúcares en la etapa.
5. Para predecir la composición de sólidos disueltos del jugo mezclado y bagazo a partir de la modelación matemática se necesita de dos gráficos interrelacionados entre sí mediante un término definido como “número de imbibición”. En cambio, para la estimación de la humedad del bagazo solamente se requiere de un gráfico que relaciona directamente a esta variable con  $\lambda$  y la temperatura del agua de imbibición.

## Conclusiones Generales

A partir de los modelos matemáticos obtenidos en el software R Studio se obtiene que:

1. El comportamiento tecnológico de la etapa de extracción mejora a través de la toma de decisiones operacionales adecuadas mediante el empleo de los modelos predictivos; por lo que se corrobora la hipótesis planteada.
2. Los modelos predictivos (expresiones 3.2; 3.4 y 3.12) presentan una adecuada calidad de ajuste denotado principalmente por los resultados de la prueba de bondad de ajuste de Fisher y son totalmente válidos para valores de  $\lambda$  de 0,80 a 2,90; de temperatura del agua de imbibición entre 40 y 86 °C y de contenido de sólidos disueltos del jugo primario entre 17,6 a 23,0 °Bx.
3. La realidad simulada mediante los modelos predictivos presenta una descripción adecuada y avala a la calidad de la modelación matemática en este sentido.
4. Los gráficos correspondientes a la representación de los modelos predictivos facilitan el empleo práctico de la modelación matemática para la toma de decisiones operacionales.

## **Recomendaciones:**

- ✓ Aplicar las herramientas de análisis matemáticas y gráficas en la operación de la etapa de extracción del CAI: “Jesús Rabí” de la provincia de Matanzas.
- ✓ Evaluar la potencialidad del uso de estas herramientas en otros centrales con características semejantes.

## Referencias Bibliográficas:

- Abraham, K., Brykczynski, H., Rudolph-Floter, E., and Floter, E. (2021). Targeted dextranase application for problem mitigation during sucrose crystallisation. *International Sugar Journal: World Sugar Yearbook 2021*.
- Anand, M. (2012). Chapter 12 Sugar and By-Products from Sugarcane: Recent Developments in Manufacturing Techniques. En G. Hunsigi (Ed.), *Production of Sugarcane: Theory and Practice (Vol. 21)*: Springer.
- Assawamartbunlue, K., Kunrapeegayson, N., and Limwattana, P. (2018). Specific energy consumption of sugar cane mills in Thailand. *Chemical Engineering Transactions*, 70, 625-630.
- AZCUBA. (2013). *Planta moledora El Consultor*. La Habana.
- Balla, W. H., Rabah, A. A., and Abdallah, B. K. (2018). Pinch Analysis of Sugarcane Refinery Water Integration. *Sugar Tech*, 20(2), 122-134. doi: 10.1007/s12355-017-0535-5
- Ballard, J., Borden, L., and Perkins, D. F. (2021). Program quality components related to youth civic engagement. *Children and Youth Services Review*, 126, 106022. doi: <https://doi.org/10.1016/j.childyouth.2021.106022>
- Bell, S. R., y Acosta, Y. L. (2017). Acciones de Producción más Limpia para implementar en la industria azucarera cubana. *ICIDCA. Sobre los Derivados de la Caña de Azúcar*, 51, 60-66.
- Bhutani, R., Tharani, K., Sudha, K., and Tomar, Y. (2020). Design of a cogeneration plant for sugar industries using renewable energy resources. *Journal of Statistics and Management Systems*, 23(1), 181-190. doi: 10.1080/09720510.2020.172163
- Birru, E. (2016). *Sugar cane industry overview and energy efficiency considerations*.
- Bocha, M. (2019). *Performance Evaluation of Sugar Cane Milling Plant and Optimization of Process Parameters: A Case of Wonji-Shoa Sugar Factory*. (Máster en Ciencias en Ingeniería Química), Universidad de Addis Ababa, Addis Ababa (Etiopía).
- Caballero, A., Fonte, A., y González, G. (2019). La cachaza y sus usos. En E. ICIDCA (Ed.), *Resultados de los institutos cubanos de investigación, desarrollo e innovación en las tecnologías sobre azúcar y derivados*. La Habana (Cuba): ICIDCA.

- Carrazana, L. (1987). Análisis Agro Industrial Azucarero Teoría y práctica: Editorial Pueblo y Educación.
- Casanova, E., Almazán, O., Cordovés, M., y Lodos, J. (2015). Tecnología azucarera. Ciudad de La Habana: ICIDCA.
- Casanova, E., Almazán, O., y Lodos, J. (2019). Capítulo 4: Tecnología Azucarera. En E. ICIDCA (Ed.), Resultados de los Institutos Cubanos de Investigación, Desarrollo e Innovación en las Tecnologías sobre Azúcar y Derivados (pp. 57-91). Ciudad de La Habana (Cuba).
- Casanova, E., y Alonso, J. (2006). Métodos de cálculos para el control azucarero. La Habana (Cuba): Instituto Cubano de Investigaciones Azucareras. MINAZ.
- Castro, B., Silva, A., y Santos, P. (2019). Aplicação de controle fuzzy na combustão de bagaço de cana-de-açúcar em caldeiras. II Encontro de Desenvolvimento de Processos Agroindustriais Uniube-UFTM-IFTM.
- Chantasiriwan, S. (2013). Numerical model of sugar mills with compound Imbibition. Presentado en: Proceedings of the 35th Conference of the Australian Society of Sugar Cane Technologists, Townsville, Queensland, Australia.
- Chantasiriwan, S. (2016). Optimum imbibition for cogeneration in sugar factories. Applied Thermal Engineering, 103, 1031-1038.
- Christodoulou, P. (2003). Sugars and Carbohydrates. En C. Tzia and G. Liadakis (Eds.), Extraction Optimization in Food Engineering. Estados Unidos: Marcel Dekker.
- Ciaburro, G. (2018). Regression Analysis with R: Design and develop statistical nodes to identify unique relationships within data at scale. Livery Place (Reino Unido): Packt.
- Corrales, J. M. (2017). Propuesta para el ahorro de energía en tándem de molinos cañeros sin afectar la eficiencia del proceso de extracción de la sacarosa de la caña de azúcar. (Tesis Doctoral), Universidad Central "Marta Abreu" de Las Villas. Facultad de Química y Farmacia, Villa Clara (Cuba).
- Cortés Falcón, M., Cornes Izquierdo, Y., Alomá Vicente, I., y González Suárez, E. (2019). Evaluación del sistema energético en el central azucarero Quintín BAndera. Centro Azúcar, 46(3), 66-78.

- Cortés, M. G., Vicente, I. A., Pedraja, R. E., y Suárez, E. G. (2015). Tecnologías y sistemas para el manejo de agua y condensados en la producción de azúcar. *Revista Centro Azúcar*, 42(1).
- Dogbe, E. S. (2020). Exergy, techno-economic and exergoeconomic analyses for improving energy efficiency of a typical sugar mill and designing integrated biorefineries.
- Dogbe, E. S., Mandegari, M. A., and Görgens, J. F. (2018). Exergetic diagnosis and performance analysis of a typical sugar mill based on Aspen Plus® simulation of the process. *Energy*, 145, 614-625. doi: <https://doi.org/10.1016/j.energy.2017.12.134>
- Edwards, B. (1995). Extraction performance of milling trains—imbibition processes and calculations. Presentado en: Proc. Aust. Soc. Sugar Cane Technol.
- Eggleston, G., Legendre, B., & Godshall, M. (2017). *Sugar and Other Sweeteners* (pp. 933-978).
- Fakoya, M. B. 2020. Investment in hazardous solid waste reduction and financial performance of selected companies listed in the Johannesburg Stock Exchange Socially Responsible Investment Index. *Sustainable Production and Consumption* 23: 21-29.
- Fox, J. and Sanford, W. 2019. *An R Companion to Applied Regression*. California: SAGE.
- Fox, J., Weisberg, S., Price, B., Adler, D., Bates, D., Baud-Bovy, G., Zeileis, A. (2021). Package: "car"(Companion to Applied Regression) (Version 3.0-11).
- Gálves, L., y Otero, I. (2019). Capítulo 2: La disponibilidad de la materia prima caña y de sus coproductos. En E. ICIDCA (Ed.), *Resultados de los Institutos Cubanos de Investigación, Desarrollo e Innovación en las Tecnologías sobre Azúcar y Derivados* (pp. 23-35). Ciudad de La Habana (Cuba).
- Gharekhani, M., Nadiri, A. A., Khatibi, R., and Sadeghfam, S. (2021). An investigation into time-variant subsidence potentials using inclusive multiple modelling strategies. *Journal of Environmental Management*, 294, 112949. doi: <https://doi.org/10.1016/j.jenvman.2021.112949>
- Gil, J. M. (1998). Estudio del efecto de los surfactantes en el agua de imbibición de los molinos sobre el porcentaje de Pol y de Humedad del bagazo. (Tesis de Maestría), Universidad Central de Las Villas.
- Ginste, J., and Partanen, S. (2020). Feasibility analysis of upgrading the cogeneration unit of George Washington sugar mill in Cuba.

- Gomaa, R. (2019). Effect of Imbibition Rate on Milling Quality of some Sugar Cane Varieties.
- González, F., Castellanos, J. A., y Puertas, J. F. (2010). Método para determinar la cantidad de agua de imbibición a utilizar en la industria de azúcar de caña. *Ingeniería Mecánica*, 13(1), 41-48.
- Gozá, O., y Ones, O. P. (2005). Notas sobre la producción de azúcar crudo. ISPJAE. La Habana.
- Gross, J. and Ligges, U. 2015. Package 'nortest'. 1.0-4: CRAN
- Guerra, J. and Sevilla, É. 1986. Introducción al análisis estadístico para procesos. La Habana: ISPJAE.
- Hamill, T. M. (1972). Estados Unidos Patent No.: U. S. P. Office.
- Harrell, F. (2015). *Regression Modeling Strategies: With Applications to Linear Models, Logistic and Ordinal Regression, and Survival Analysis*. Nueva York: Springer.
- Hernández-Chover, V., Castellet-Viciano, L., and Hernández-Sancho, F. (2019). Cost analysis of the facilities deterioration in wastewater treatment plants: A dynamic approach. *Sustainable Cities and Society*, 49, 101613. doi: <https://doi.org/10.1016/j.scs.2019.101613>
- Herrera, E., Zuazaga, J., y Díaz, A. (1996). *Planta Moledora: Operación en zafra*.
- Honing, P. (1987). *Principios de Tecnología Azucarera: Propiedades de los azúcares y no-azúcares. La purificación de los jugos*. La Habana, Cuba: Edición Revolucionaria.
- Hothorn, T., Zeileis, A., Farebrother, R. W., Cummins, C., Millo, G. and Mitchell, D. 2020. Package: "lmtest" (Testing Linear Regression Models). 0.9-38:
- Hugot, E. (2014). *Handbook of Cane Sugar Engineering*: Elsevier.
- ICIDCA. (2012). *Índice de Capacidades*. La Habana (Cuba): Instituto Cubano de Investigaciones de los Derivados de la Caña de Azúcar.
- Jiménez, R., Monteagudo, J. P., y Lorenzo, J. (2017). Potencialidades de mejoras energéticas en el proceso tecnológico en los centrales azucareros. *Revista de Ingeniería Energética*, 38(2), 88-96.
- Kent, G. (2010). Estimating bagasse production. Presentado en: *Proceedings of the 2010 Conference of the Australian Society of Sugar Cane Technologists held at Bundaberg,, Queensland, Australia*.

- Kent, G. A. 1997. Modelling the extraction processes of milling trains. *Proceedings of the Australian Society of Sugar Cane Technologists* 19: 315-321.
- Kumar, S., Yadav, A., and Kumar, A. (2020). Design modification to enhance the performance of rotary drum vacuum filter. En STAI (Ed.), e-Proceedings of 78th Annual Convention of STAI (pp. 124-132). Nueva Delhi (India): STAI.
- Kusoncum, C., Sethanan, K., Hartl, R. F., and Jamrus, T. (2020). Modified differential evolution and heuristic algorithms for dump tippler machine allocation in a typical sugar mill in Thailand. *Operational Research*. doi: 10.1007/s12351-020-00597-z
- Lao León, Y. O., Vega de la Cruz, L. O., Marrero Delgado, F., y Pérez-Pravia, M. C. (2017). Procedimiento para modelar recursos restrictivos en el sistema logístico de empresas comercializadoras. *Ingeniería Industrial*, 38(1), 43-55.
- Lionnet, G. (1981). The effect of the level of extraction on mixed juice purity. *EXTRACTION*, 96, 97.
- Liu, J., and Smith, S. R. (2020). A multi-level biogas model to optimise the energy balance of full-scale sewage sludge conventional and THP anaerobic digestion. *Renewable Energy*, 159, 756-766. doi: <https://doi.org/10.1016/j.renene.2020.06.029>
- Londhe, M., Kale, K., Dixit, K., y Raj, S. (2020). Case study - Impact of high imbibition % cane (50 % +) on sugar recovery and steam consumption. En STAI (Ed.), e-Proceedings of 78th Annual Convention of STAI (pp. 155-164). Nueva Delhi (India): STAI.
- Lüdecke, D., Makowski, D., Ben-Shachar, M., Waggoner, P., Wiernik, B. M., Arel-Bundock, V. and Jullum, M. 2021. Package: Performance (Assessment of Regression Models Performance). 0.7.3:
- Macedo, I. d. C. (2005). A energia da cana-de-açúcar. Doze estudos sobre a agroindústria da cana-de-açúcar no Brasil ea sua sustentabilidade. *União da Agroindústria Canavieira do Estado de São Paulo (UNICA)*, 35.
- Mann, A., Thaval, O., Broadfoot, R. and Mcfeaters, J. 2015. Improving factory performance using an integrated sugar factory model. *Proceedings of the Australian Society of Sugar Cane Technologists* 37: 305-315.
- Marthino De Oliveira, T. (2019). Análise da eficiência energética e extração de sacarose no setor sucroenergético.
- MINAZ. (2006). Manual de métodos analíticos para azúcar crudo (MACU): Instituto cubano de investigaciones azucareras. División de aseguramiento y calidad.

- Meyers, L., Gamst, G., and Guarino, A. (2016). *Applied Multivariate Research: Design and Interpretation*: Sage.
- Mishra, P., Pandey, C., Singh, U., Gupta, A., Sahu, C. and Keshri, A. 2019. Descriptive Statistics and Normality Tests for Statistical Data. *Annals of Cardiac Anaesthesia*.
- Mkwanzani, T., Mandegari, M., and Görgens, J. F. (2019). Disturbance modelling through steady-state value deviations: The determination of suitable energy indicators and parameters for energy consumption monitoring in a typical sugar mill. *Energy*, 176, 211-223. doi: <https://doi.org/10.1016/j.energy.2019.03.191>
- Mohammed, Y. (2018). *Optimization of Steam Generation and Consumption: A Case study Elguneid Sugar Factory, Gezira State, Sudan*. (Master of Science), Universidad de Gezira, Sudán.
- Montgomery, D., Peck, E. y Vining, G. 2002. *Introducción al análisis de regresión lineal*. México: Compañía Editorial Continental.
- Montgomery, D., and Runger, G. (2018). *Applied Statistics and Probability for Engineers* (7ma ed.). Estados Unidos: Wiley.
- Morales, Y. L., Kafarov, V., Ruiz, F., y Castillo, E. F. (2018). Modelación de los procesos de producción de bioetanol de primera y segunda generación a partir de la caña de azúcar. Etapas; preparación, molienda y clarificación. *Umbral Científico*, 16.
- Mullor, R., Mulero, J., and Trottini, M. (2019). A modelling approach to optimal imperfect maintenance of repairable equipment with multiple failure modes. *Computers and Industrial Engineering*, 128, 24-31. doi: <https://doi.org/10.1016/j.cie.2018.12.032>
- Munro, B. M. (1963). Further Imbibition Experiments. Presentado en: *Proceedings of the Queensland Society of Sugar Cane Technologists*.
- Ordóñez, S. C. (2004). *Modelo de diagnóstico para la evaluación de la productividad de un proceso agroindustrial azucarero*. (Trabajo de graduación de Ingeniera Química), Universidad de San Carlos de Guatemala, Guatemala.
- Pedroso, I. (2018a). Mass balance of a sugarmill with compound imbibition—a numerical model. *Sugar Journal*, 80 (11), 20-27.
- Pedroso, I. (2018b). Mass balance of a sugarmill with compound imbibition—a numerical model. *Sugar Journal*, 80 (11), 20-27.

- Pérez de Alejo, H., Perez, A., Corrales, Y., Pérez, O., Curbelo, T., González, Y., Marrero, R. (2009). El análisis de procesos y el empleo adecuado de la energía en la producción de azúcar crudo y electricidad en ingenios.
- Priego, Q., and Palmer, R. (2018). Multi Regression Analysis of Skin Temperature Variation During Cycling Exercise. *Lecture Notes in Computational Vision and Biomechanics*, 27, 962-969. doi: 10.1007/978-3-319-68195-5\_105
- Poel, P. W. V. d., Schiweck, H., and Schwartz, T. (1998). Sugar Technology beet and cane manufacture. Denver, Estados Unidos: Publicado con el apoyo de la Fundación para el Desarrollo de la Azúcar de Remolacha.
- Rein, P. (1975). A statistical analysis of the effect of cane quality on extraction performance. Presentado en: Proc S Afr Sug Technol Ass.
- Rein, P. (2012). Ingeniería de la caña de azúcar: Bartens.
- Ripley, B., Bill, V., Bates, D., Hornik, K., Gebhardt, A., and Firth, D. (2021). Package: MASS (Functions and datasets to support Venables and Ripley, "Modern Applied Statistics with S" (4th edition, 2002).) (Version 7.3-54): CRAN.
- Russell, G. E. (1965). The use of analogue computation to predict the performance of a milling train. Presentado en: Proceedings of the Queensland Society of Sugar Cane Technologists.
- Sahu, O. (2018). Assessment of sugarcane industry: Suitability for production, consumption, and utilization. *Annals of Agrarian Science*, 16(4), 389-395. doi: <https://doi.org/10.1016/j.aasci.2018.08.001>
- Salgado, A. K. G., Leyva, C. A. G., Ceballos, M. d. I. C. B., Alfonso, J. C. Q., y Morales, J. C. P. (2020). Análisis de la eficiencia operacional y energética del proceso de producción de azúcar crudo en la UEB Central Azucarero Industrial Jesús Rabí. Encuentro con la Química.
- Sejati, A., Kasmianti, S., and Ikhsan, F. (2019). The relationship between learning process interactions and student's learning outcomes in environmental sustainability matter geography-social science education subject. Presentado en: IOP Conference Series: Earth and Environmental Science.

- Thaval, O. (2012). Modelling the flow of cane constituents through the milling process of a raw sugar factory. Queensland University of Technology.
- Thaval, O., & Kent, G. (2012). Modelling milling train performance. Presentado en: Proceedings of the 37th Conference of the Australian Society of Sugar Cane Technologists, Bundaberg, Queensland, Australia.
- Thaval, O. and Kent, G. A. 2013. Advanced computer simulation of the milling process. International Society of Sugar Cane Technologists 28: 1595-1607.
- Villatoro, M. A. S. (2019). Evaluación de mazas perforadas, en el agotamiento del bagazo de caña de azúcar en el proceso de extracción de jugo, en ingenio PAlo Gordo, S.A., Universidad de San Carlos de Guatemala, Guatemala.
- Wickham , H., Chang, W., Henry , L., Lin, T., Takahashi, K., Wilke, C., Dunnington, D. (2021). Package "ggplot2" (Create Elegant Data Visualisations Using the Grammar of Graphics) (Version 3.3.5).
- Wienese, A. (1990). Mill settings and extraction. Presentado en: Proceedings of The South African Sugar Technologists' Association.
- Wienese, A. (1994). Imbibition optimisation at Mount Edgecombe. Proceedings of The South African Sugar Technologists' Association, 6.
- Wienese, A. (1995). The effect of imbibition and cane quality on the front end mass balance. Presentado en: Proc S Afr Sug Technol Ass.