



Universidad de San Carlos de Guatemala

Facultad de Ingeniería

Escuela de Estudios de Postgrado

Maestría en Estadística Aplicada

**USO DE SERIES TEMPORALES PARA LA DETECCIÓN DE PATRONES Y PROYECCIÓN
DE TIEMPOS PERDIDOS EN LA FABRICACIÓN DE AZÚCAR**

Ing. Angel Oseas Ajcabul Raxhón

Asesorado por el Mtro(a). Byron de Jesús López Maldonado

Guatemala, junio de 2022

UNIVERSIDAD DE SAN CARLOS DE GUATEMALA



FACULTAD DE INGENIERÍA

**USO DE SERIES TEMPORALES PARA LA DETECCIÓN DE PATRONES Y PROYECCIÓN
DE TIEMPOS PERDIDOS EN LA FABRICACIÓN DE AZÚCAR**

TRABAJO DE GRADUACIÓN

PRESENTADO A LA JUNTA DIRECTIVA DE LA
FACULTAD DE INGENIERÍA
POR

Ing. ANGEL OSEAS AJCABUL RAXHÓN

ASESORADO POR EL MTRO. BYRON DE JESÚS LÓPEZ MALDONADO

AL CONFERÍRSELE EL TÍTULO DE

MAESTRO EN ESTADÍSTICA APLICADA

GUATEMALA, JUNIO DE 2022

UNIVERSIDAD DE SAN CARLOS DE GUATEMALA
FACULTAD DE INGENIERÍA



NÓMINA DE JUNTA DIRECTIVA

DECANA	Ing. Aurelia Anabela Córdova Estrada.
VOCAL I	Ing. José Francisco Gómez Rivera.
VOCAL II	Ing. Mario Renato Escobedo Martínez.
VOCAL III	Ing. José Milton de León Bran.
VOCAL IV	Br. Kevin Vladimir Cruz Lorente.
VOCAL V	Br. Fernando José Paz González.
SECRETARIO	Ing. Hugo Humberto Rivera Pérez.

TRIBUNAL QUE PRACTICÓ EL EXAMEN PRIVADO

DECANA	Ing. Aurelia Anabela Córdova Estrada.
EXAMINADOR(A)	Ing. Edgar Darío Álvarez Cotí.
EXAMINADOR(A)	Ing. Edwin Adalberto Bracamonte Orozco.
EXAMINADOR(A)	Ing. William Eduardo Fagiani Cruz.
SECRETARIO	Ing. Hugo Humberto Rivera Pérez.

La Decana de la Facultad de Ingeniería de la Universidad de San Carlos de Guatemala, luego de conocer la aprobación por parte del Director de la Escuela de Estudios de Posgrado, al Trabajo de Graduación titulado: **USO DE SERIES TEMPORALES PARA LA DETECCIÓN DE PATRONES Y PROYECCIÓN DE TIEMPOS PERDIDOS EN LA FABRICACION DE AZÚCAR**, presentado por: **Angel Oseas Ajcabul Raxhon**, que pertenece al programa de Maestría en artes en Estadística aplicada después de haber culminado las revisiones previas bajo la responsabilidad de las instancias correspondientes, **autoriza la impresión del mismo.**

IMPRÍMASE:



Inga. Aurelia Anabela Cordova Estrada

Decana

Guatemala, junio de 2022

AACE/gaoc



Guatemala, junio de 2022

LNG.EEP.OI.425.2022

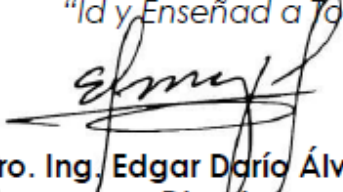
En mi calidad de Director de la Escuela de Estudios de Postgrado de la Facultad de Ingeniería de la Universidad de San Carlos de Guatemala, luego de conocer el dictamen del asesor, verificar la aprobación del Coordinador de Maestría y la aprobación del Área de Lingüística al trabajo de graduación titulado:

"USO DE SERIES TEMPORALES PARA LA DETECCIÓN DE PATRONES Y PROYECCIÓN DE TIEMPOS PERDIDOS EN LA FABRICACIÓN DE AZÚCAR"

presentado por **Angel Oseas Ajcabul Raxhón** correspondiente al programa de **Maestría en artes en Estadística aplicada**; apruebo y autorizo el mismo.

Atentamente,

"Id y Enseñad a Todos"


Mtro. Ing. Edgar Darío Álvarez Cotí
Director
Escuela de Estudios de Postgrado
Facultad de Ingeniería





Guatemala 27 de febrero 2022.

M.A. Edgar Darío Álvarez Cotí
Director
Escuela de Estudios de Postgrado
Presente

M.A. Ingeniero Álvarez Cotí:

Por este medio informo que he revisado y aprobado el Informe Final del trabajo de graduación titulado **"USO DE SERIES TEMPORALES PARA LA DETECCIÓN DE PATRONES Y PROYECCIÓN DE TIEMPOS PERDIDOS EN LA FABRICACIÓN DE AZÚCAR"** del estudiante **Angel Oseas Ajcabul Raxhón** quien se identifica con número de carné 201212888 del programa de Maestría en Estadística Aplicada.

Con base en la evaluación realizada hago constar que he evaluado la calidad, validez, pertinencia y coherencia de los resultados obtenidos en el trabajo presentado y según lo establecido en el *Normativo de Tesis y Trabajos de Graduación aprobado por Junta Directiva de la Facultad de Ingeniería Punto Sexto inciso 6.10 del Acta 04-2014 de sesión celebrada el 04 de febrero de 2014*. Por lo cual el trabajo evaluado cuenta con mi aprobación.

Agradeciendo su atención y deseándole éxitos en sus actividades profesionales me suscribo.

Atentamente,


MSc. Ing. Edwin Adalberto Bracamonte Orozco
Coordinador
Maestría en Estadística Aplicada
Escuela de Estudios de Postgrado

Guatemala 10 de julio de 2021

Mtro. Edgar Darío Álvarez Cotí

Director de la Escuela de Estudios de Posgrado, FIUSAC

Presente

Estimado Maestro Álvarez Cotí:

Reciba un cordial saludo.

Por medio de la presente hago de su conocimiento que Angel Oseas Ajcabul Raxhón, estudiante de la Maestría en Estadística Aplicada, quien se identifica con el carné 2012-12888, me ha presentado su trabajo de graduación titulado: "USO DE SERIES TEMPORALES PARA LA DETECCIÓN DE PATRONES Y PROYECCIÓN DE TIEMPOS PERDIDOS EN LA FABRICACIÓN DE AZÚCAR", el cual realizó bajo mi asesoría, brindada Ad-Honorem.

Luego de revisar el documento que contiene el informe del trabajo de investigación, manifiesto que le doy mi aprobación y considero que puede continuar las gestiones correspondientes.

Sin otro particular, me suscribo a sus respetables órdenes.

Atentamente,



Byron de Jesús López Maldonado

Maestro en Energía y Ambiente

Byron de Jesús López Maldonado
Ingeniero Químico
Colegiado 1088

HONORABLE TRIBUNAL EXAMINADOR

En cumplimiento con los preceptos que establece la ley de la Universidad de San Carlos de Guatemala, presento a su consideración mi trabajo de graduación titulado:

USO DE SERIES TEMPORALES PARA LA DETECCIÓN DE PATRONES Y PROYECCIÓN DE TIEMPOS PERDIDOS EN LA FABRICACIÓN DE AZÚCAR

Tema que me fuera aprobado por la Dirección de la Escuela de Estudios de Postgrado, con fecha 19 de agosto de 2020.

A handwritten signature in black ink, featuring a large, stylized 'A' and 'R' that are interconnected, with a horizontal line crossing through them.

Ing. Angel Oseas Ajcabul Raxhón

ACTO QUE DEDICO A:

Dios

Por darme la vida y guiarme en cada decisión que doy.

Mis padres

Diego Ajcabul y Elena Raxhón por apoyarme en todo momento.

Mis hermanas

Sandy y Wendy Ajcabul por darme el aliento necesario para terminar cada meta que me propongo.

AGRADECIMIENTOS A:

Universidad de San Carlos de Guatemala	Por ser mi lugar de estudios y una excelente fuente de conocimiento.
Facultad de Ingeniería	Por la formación que me ha brindado desde la licenciatura hasta la maestría.
Ing. Byron López	Por compartir sus conocimientos y brindarme apoyo al momento de asesorar el presente trabajo.

ÍNDICE GENERAL

ÍNDICE DE ILUSTRACIONES.....	V
LISTA DE SÍMBOLOS	VII
GLOSARIO	IX
RESUMEN DEL MARCO METODOLÓGICO	XI
PLANTEAMIENTO DEL PROBLEMA.....	XIII
OBJETIVOS.....	XV
RESUMEN DEL MARCO METODOLÓGICO	XVII
INTRODUCCIÓN	XXIII
1. MARCO REFERENCIAL.....	1
1.1. Estudios Previos.....	1
2. MARCO TEÓRICO.....	5
2.1. Datos, información y conocimiento	5
2.2. Calidad de los datos	5
2.3. Distribuciones muestrales.....	6
2.4. Distribución normal.....	6
2.5. Series de tiempo.....	7
2.6. Procesos estocásticos estacionarios	7
2.7. Procesos estocásticos integrados	8
2.8. Pruebas de estacionariedad	8
2.9. Prueba de raíz unitaria (Dickey Fuller)	9
2.10. Pronósticos de una serie de tiempo	10
2.11. Métodos de suavizado exponencial.....	11
2.12. Modelos ARIMA.....	11

2.13.	Principio 80-20	12
2.14.	Proceso de producción del azúcar	13
2.15.	Etapa 1: preparación de la caña de azúcar	13
2.16.	Etapa 2: molienda de la caña de azúcar	14
2.17.	Etapa 3: clarificación	14
2.18.	Etapa 4: evaporación	15
2.19.	Etapa 5: cristalización	16
2.20.	Etapa 6: centrifugación.....	16
3.	PRESENTACIÓN DE RESULTADOS.....	17
3.1.	Objetivo 1: categorizar los problemas que causan tiempos perdidos en la operación, en función de la descripción.....	17
3.2.	Objetivo 2: clasificar los problemas que generan la mayor cantidad de tiempo perdido en la operación mediante la regla 80/20	18
3.3.	Objetivo 3: describir el comportamiento de los problemas que causan la mayor cantidad de tiempo perdido mediante el uso de series de tiempo	23
3.4.	Objetivo 4: pronóstico del tiempo perdido para la zafra 2020-2021, usando las técnicas de medias móviles y suavización exponencial	29
3.5.	Objetivo general: describir los problemas que causan la mayor cantidad de tiempo perdido con el uso de series de tiempo, generando un pronóstico para el período 2021	31
4.	DISCUSIÓN DE RESULTADOS.....	33
4.1.	Análisis interno	33
4.2.	Análisis externo	34
4.3.	Categorización de los tiempos perdidos.....	34

4.4.	Clasificación de los problemas más importantes.....	35
4.5.	Análisis descriptivo de los tiempos perdidos con series temporales.....	36
4.6.	Predicción de los tiempos perdidos	38
CONCLUSIONES		39
RECOMENDACIONES.....		41
REFERENCIAS		43
APÉNDICES		49
ANEXOS		53

ÍNDICE DE ILUSTRACIONES

FIGURAS

1.	Diagrama de Pareto para la duración de los tiempos perdidos.....	19
2.	Tiempos perdidos según su frecuencia.....	21
3.	Duración de los tiempos perdidos	23
4.	Correlograma de la duración de los tiempos perdidos	24
5.	Descomposición de los tiempos perdidos (duración)	25
6.	Frecuencia en los tiempos perdidos.....	26
7.	Correlograma para la frecuencia de los tiempos perdidos	27
8.	Descomposición de la frecuencia de los tiempos perdidos	28

TABLAS

I.	Variables de estudio	XVIII
II.	Clasificación de problemas.....	18
III.	Pronósticos con el método de Holt Winters.....	29
IV.	Pronósticos con el método ARIMA.....	30
V.	Criterios de selección	31

LISTA DE SÍMBOLOS

Símbolo	Significado
ρ	Coefficiente de correlación parcial
Y_k	Covarianza del retraso k
$H1$	Hipótesis alterna
$H0$	Hipótesis nula
$E(Y_t)$	Media constante para una serie de tiempo
μ	Media de los residuos asociada al ruido blanco
α	Nivel de significancia del contraste de hipótesis (5 %)
qq	Quintales
k	Retraso de una serie de tiempo
(Y_t)	Varianza de una serie de tiempo

GLOSARIO

Correlograma	Gráfica que muestra los efectos de diferentes retrasos sobre el valor presente en una serie de tiempo.
Hipótesis nula	Suposición que se desea contrastar y validar.
Proceso determinista	Proceso que se puede predecir con exactitud usando modelos matemáticos.
Proceso estocástico	Conjunto de variables aleatorias ordenadas en el tiempo que no se pueden predecir con exactitud.
Regresión espuria	Regresión engañosa que proporciona pruebas estadísticas de relación lineal entre variables, pero en realidad no tienen efectos causales.
Residuos	Diferencia entre los valores observados y los valores predichos con algún modelo.
Ruido blanco	Serie de tiempo cuya media es igual a cero, tiene varianza constante y no está seriamente correlacionado.
Serie de tiempo	Conjunto de datos en orden cronológico e intervalo constante.

RESUMEN DEL MARCO METODOLÓGICO

La presente investigación tiene como propósito mostrar que el uso adecuado de los registros relacionados a los tiempos perdidos en una fábrica de producción de azúcar, puede ser una herramienta útil para tomar decisiones al momento de elaborar planes de acción que mitiguen los paros en la producción.

El objetivo general de esta investigación consiste en describir los problemas que causan la mayor cantidad de tiempo perdido en la producción de azúcar analizando series de tiempo y generar un pronóstico para el período de zafra 2020-2021 que pueda ser utilizado en la toma de acciones preventivas por parte de los miembros de la organización.

Los resultados se elaboraron mediante la categorización de los paros de molienda, identificando con el principio de Pareto, las causas que mayor influencia tuvieron en los tiempos perdidos. También se utilizó la descomposición clásica de las series temporales en cuanto a la duración y frecuencia de los paros de molienda, con el fin de identificar patrones en su ocurrencia, posteriormente se aplicaron los métodos de Holt-Winters y ARIMA para pronosticar los tiempos perdidos para los primeros 15 días de la zafra 2020-2021.

Con todo lo anterior, se identificó que los problemas que más influyeron en la continuidad de la producción de la fábrica analizada son los relacionados a la calidad de la materia prima y a los fallos en los equipos, que pueden acumular hasta 4,600 minutos en paros de molienda para la zafra 2020-2021. Se concluyó que el enfoque adecuado para los planes de mitigación debe tomar en cuenta la

materia prima y el mantenimiento de los equipos y se recomienda seguir categorizando los problemas registrados para facilitar próximos análisis.

PLANTEAMIENTO DEL PROBLEMA

- Contexto general

El azúcar es un producto de consumo masivo a nivel mundial y el proceso de fabricación involucra una constante evolución en el proceso, siendo parte de la mejora continua. En los últimos años el precio del azúcar ha fluctuado de manera negativa, impactando en los costos de producción y generando un panorama muy difícil para muchas fábricas en Guatemala. Bajo este contexto, se llevó a cabo el estudio de tiempos perdidos en un ingenio de Guatemala, ubicado en la costa sur, pues los tiempos perdidos influyen directamente en el costo final por cada quintal de azúcar producido y pueden ser una oportunidad de mejora para aumentar la eficiencia del proceso.

- Descripción del problema

En las últimas tres zafras se produjeron paros de molienda que afectaron la eficiencia de la producción de azúcar, los cuales no fueron analizados estadísticamente para encontrar la causa raíz de los problemas y cuya oportunidad de mejora es evidente. Tampoco se hicieron proyecciones para las zafras siguientes, cuya importancia es alta para atacar los factores que influyen en los paros de molienda y con ello, aumentar la eficiencia del proceso.

- Formulación del problema

El trabajo de investigación buscó responder a las siguientes preguntas:

- Pregunta general
 - ¿Cuál es el comportamiento de los problemas que causan la mayor cantidad de tiempo perdido en la producción de azúcar?
- Preguntas auxiliares
 - ¿Cuáles son los problemas que causan tiempos perdidos en la producción de azúcar?
 - ¿Cuáles son los problemas que generan mayor tiempo perdido en la operación?
 - ¿Cuál es el comportamiento de las principales causas de paros de molienda en función del tiempo?
 - ¿Cuánto tiempo perdido se registrará en la zafra 2020-2021?
- Delimitación del problema

El estudio de los tiempos perdidos en la fabricación de azúcar se llevó a cabo con registros que se extrajeron de las bases de datos de las bitácoras de turno del personal en operación, desde la zafra 2017-2018 hasta la zafra 2019-2020. La información que se utilizó corresponde a un ingenio cuya capacidad de molienda es de 19,500 toneladas de caña y la capacidad de producción ronda los 45,000 quintales diarios.

OBJETIVOS

General

Describir los problemas que causan la mayor cantidad de tiempo perdido en la producción de azúcar, con el uso de series de tiempo, generando un pronóstico para el período 2020-2021 para que se tomen acciones preventivas por parte de los miembros de la organización.

Específicos

- Categorizar los problemas que causan tiempos perdidos en la operación, en función de la descripción registrada para agruparlos por causas comunes.
- Clasificar los problemas que generan la mayor cantidad de tiempo perdido en la operación mediante la regla 80/20 para establecer una prioridad en el análisis de la información.
- Describir el comportamiento de los problemas que causan la mayor cantidad de tiempo perdido mediante el uso de series de tiempo para identificar patrones en la ocurrencia de estos.
- Realizar un pronóstico del tiempo perdido para la zafra 2020-2021, usando las técnicas de medias móviles y suavización exponencial, para tomar acciones preventivas en el proceso de producción.

RESUMEN DEL MARCO METODOLÓGICO

- Características del estudio

La presente investigación tiene un enfoque cuantitativo con un diseño no experimental de tipo longitudinal, pues se tienen los registros de los tiempos perdidos que afectan la producción de azúcar en un ingenio de Guatemala, desde el año 2017 hasta el año 2020, con los cuales se harán estimaciones de frecuencia y duración para un período posterior. El alcance inicial de la investigación es descriptivo porque propone un método para analizar los registros de los tiempos perdidos, permitiendo identificar las causas de estos, un problema poco estudiado hasta el momento en la industria azucarera, finalizando con un alcance explicativo, pues se busca determinar la ocurrencia en función del tiempo y hacer proyecciones para el año 2021.

Entre los objetivos de la investigación se buscó identificar las causas principales de los tiempos perdidos en las últimas 3 zafas (2017-2020), así como identificar los patrones de ocurrencia de los paros de molienda y con ello, realizar un pronóstico para la zafa 2020-2021. Los resultados de esta investigación serán de utilidad los ingenieros encargados de velar por la producción en un ingenio azucarero.

- Unidades de análisis

La población de estudio estuvo conformada por los registros ingresados en las bases de datos con la descripción de cada paro de molienda y el responsable

del proceso a partir del año 2017, dicha información se encuentra dividida en subpoblaciones relacionadas con el área en la cual se originó el tiempo perdido, la hora de inicio y la hora final del paro. No se realizaron muestreos, sino que se utilizó la información almacenada en las bitácoras registradas por el personal de operación.

- Variables

Las variables del estudio se describen a continuación:

Tabla I. **Variables de estudio**

Variable	Definición teórica	Definición operativa
Causa del paro	Descripción del motivo por el cual se detuvo la molienda de caña en la fabricación de azúcar.	1=Tanques llenos 2=Piedras detectadas 3=Falta de caña 4=Fuga 5=Centrifuga fuera de línea 6=Lleno de meladura 7=Recibidores llenos de masa
Área operativa	Parte del proceso en donde ocurrió el incidente	1= Patio de caña 2= Molinos 3= Calderas 4= Clarificación 5= Evaporación 6= Cristalización 7= Centrifugas 8=Refinería 9= Envasado
Duración	Tiempo que tardó el paro de molienda en minutos	Número real
Zafra	Período de producción de azúcar.	1= 2017/2018 2= 2018/2019 3= 2019/2020

Fuente: elaboración propia, realizado con Microsoft Excel 2019.

- Fases del estudio

Las siguientes fases fueron las que se llevaron a cabo en el estudio para lograr los objetivos planteados:

- Revisión de la literatura

En esta etapa se llevó a cabo la recopilación de todas las técnicas, herramientas y estudios necesarios para usarla de referencia en esta investigación.

- Gestión de la información

Etapa en la cual se llevó a cabo la extracción de los datos con los permisos necesarios, se ubicó el servidor y el nombre del fichero en donde los operarios guardaron la información. También se prepararon y limpiaron los datos para poder ser analizados en la etapa posterior con un formato estándar.

- Análisis de la información

Etapa en la cual se aplicaron las diversas técnicas estadísticas de series de tiempo y se obtuvieron los resultados necesarios para responder a los objetivos planteados.

- Interpretación de la información

En esta fase se les dio sentido a los resultados obtenidos, explicándolos desde el punto de vista estadístico y técnico.

- Elaboración del informe final

Etapa en la cual se plasmó de forma escrita y formal todo el trabajo de investigación, el cual fue planteado con anterioridad como un protocolo de investigación.

- Presentación del trabajo terminado

Período en el cual se mostró todo el trabajo de investigación a los revisores y asesor para ser aprobado.

- Realización de correcciones

Etapa en la cual se corrigieron errores de formato y forma para obtener finalmente la aprobación por parte de la escuela de posgrado.

Técnicas de análisis de la información

- Regla del 80/20

Consistió en ordenar los paros de molienda o tiempos perdidos de forma descendente tomando en cuenta la duración y la frecuencia, seleccionando únicamente el 20 % de los problemas que causaron el 80 % del tiempo y frecuencia totales.

- Prueba de hipótesis de raíz unitaria

Esta prueba se utilizó para identificar la existencia o no de estacionariedad en cada serie de tiempo analizada, tomando en cuenta las siguientes hipótesis:

$$H_0: \delta=0$$

$$H_1: \delta<0$$

Donde se estableció la existencia de raíz unitaria con el estadístico de Dickey-Fuller (DF) para la hipótesis nula (serie de tiempo no estacionaria) y fue contrastada en la hipótesis alterna con el estadístico DF (serie de tiempo estacionaria).

- Descomposición de la serie temporal

Proceso de análisis de series de tiempo clásico, en el cual se separaron las diferentes componentes (tendencia, estacionalidad y componente aleatoria) para cada serie de tiempo y fue tomado como referencia el modelo aditivo.

- Algoritmo Auto Arima

Este algoritmo se implementó para cada serie de tiempo, el cual busca cada coeficiente de un modelo SARIMAX, el modelo más complejo. Se tomó como parámetro de optimización el criterio de pérdida de información de Akaike (AIC).

- Software estadístico

Para el análisis completo de la información, se utilizaron las siguientes herramientas: Software R v.3.8.1, Python 3.8.3, Minitab 19 y Microsoft Excel 2016.

INTRODUCCIÓN

La presente investigación se basa en la sistematización al momento de analizar los tiempos perdidos en una fábrica de producción de azúcar, pues toma en cuenta la recolección y estandarización de los datos, así como también el análisis y proyección para períodos posteriores. La sistematización propuesta puede replicarse para diferentes períodos de datos históricos y adaptarse a otras fábricas de producción.

Los tiempos perdidos en las fábricas de producción de azúcar, impactan directamente en el costo final del producto y la calidad, en muchos casos, estos problemas pueden ser mitigados si se realizan planes preventivos que tomen en consideración la causa por la cual ocurren. Actualmente, no se cuentan con muchos estudios sobre los tiempos perdidos aplicados a la producción de azúcar debido a que cada fábrica tiene configuraciones distintas en las líneas de producción, ajustes específicos acordes a la materia prima y otros factores que hacen del proceso de producción, un sistema complejo y diferente entre cada fábrica.

Ante la falta de estudios y metodologías para realizar un análisis de tiempos perdidos en la producción de azúcar, se propone una estandarización en el registro de estos y a partir de ellos, realizar proyecciones para los próximos períodos utilizando series temporales. La solución mostrada en esta investigación permite identificar los problemas que mayor impacto tienen en el costo y la calidad del producto final, siendo parte fundamental para tomar decisiones al momento de hacer los planes de producción.

La metodología utilizada en esta investigación comienza con la categorización y agrupación de los paros de molienda, logrando clasificar a los más importantes mediante el principio de Pareto (para la priorización en la mitigación de estos). También se presenta un método para identificar patrones en la ocurrencia de los tiempos perdidos por medio de la descomposición de la serie temporal que los representa, con el fin de visualizar comportamientos estacionales, cíclicos y aleatorios en la frecuencia y duración. Por último, se muestran dos modelos para la proyección de los paros de molienda de la zafra 2020-2021, el primero es el método de Holt-Winters y el segundo es el método ARIMA, seleccionando el mejor con el criterio de información AIC.

Los resultados obtenidos indican que la materia prima y las fallas en los equipos en la operación influyeron considerablemente en los tiempos perdidos de las zafas 2017-2018, 2018-2019 y 2019-2020, mostrando la importancia de la estandarización y categorización de estos. Con la descomposición de la serie temporal de los tiempos perdidos, se logró identificar que los cambios de turnos semanales aportan un incremento de la duración y frecuencia de los paros de molienda, pues cambian los criterios de operación al no tener una directriz en el traslado de turnos. Por último, las proyecciones indican que para los primeros 15 días de la zafra 2020-2021 los paros de molienda sumarán más de cuatro mil minutos, aportando una métrica a reducir previo a empezar la producción. Los beneficios de esta investigación se resumen en una propuesta de análisis de los resultados de los tiempos perdidos, enfocada en aportar información útil para tomar decisiones al momento de optimizar los procesos de producción y cuyos pronósticos pueden utilizarse como métrica para evaluar los resultados de los planes de mitigación elaborados.

El estudio es factible para implementarlo previo al inicio de un período productivo, pues proporciona información que puede ser útil para reducir los

tiempos perdidos. También se puede utilizar la metodología utilizada en esta investigación para otras fábricas de producción similares, teniendo en consideración, el correcto análisis de los resultados obtenidos, adaptados al entorno al cual se desee aplicar.

En el capítulo 1 de la investigación, se muestran estudios previos que aportan metodologías, resultados y conceptos para analizar la información obtenida, relacionadas con los tiempos perdidos en fábricas de producción, el uso correcto de las series temporales para describir un fenómeno, hacer predicciones y otros aportes que sirven de referencia para alcanzar los objetivos propuestos.

El capítulo 2 muestra los fundamentos teóricos en los cuales, se sustenta el análisis, desarrollo y presentación de los resultados, desde el punto de vista estadístico y desde la perspectiva de operación en la fábrica típica de un ingenio azucarero.

El capítulo 3 muestra los resultados obtenidos a partir de los datos recolectados, aplicando los conocimientos y técnicas mostradas en el capítulo 1 y 2, con la finalidad de categorizar y clasificar los paros de molienda, encontrar patrones en su ocurrencia a lo largo de las zafas 2017-2018, 2018-2019 y 2019-2020 para posteriormente, hacer predicciones para los primeros 15 días de producción de la zafra 2020-2021.

El capítulo 4 muestra la interpretación de los resultados obtenidos, validándolos internamente (muestras representativas y escalabilidad en otras investigaciones), externamente (comparación entre las referencias de estudios previos) y analizando cada uno de los resultados en función de los objetivos planteados.

1. MARCO REFERENCIAL

1.1. Estudios Previos

Durante los procesos de mejora continua en las fábricas de producción de azúcar, ha sido importante seleccionar los datos adecuados para generar información útil y que impacten significativamente a los costos de producción para poder hacer estudios factibles. Los tiempos perdidos son una fuente primaria de información que involucra un impacto evidente en los costos de producción, motivo por el cual, es necesario su análisis con determinada frecuencia, pues es una forma de diagnóstico del proceso actual.

Rein (2012), determinó la importancia de tener sistemas de medición para evaluar el desempeño de una fábrica, haciendo comparaciones entre las mismas en diferentes países como Sudáfrica, Brasil y Estados Unidos. Por otro lado, Alfonso, Consuegra y Ribas (2016) encontraron algunas variables que afectaron el rendimiento en las fábricas de producción de azúcar en Cuba, una de las cuales es el tiempo perdido y utilizaron un análisis de componentes principales y regresión múltiple. Esta investigación aportó diversas técnicas estadísticas utilizadas para las variables del proceso de producción similares a las del Ingenio analizado en este trabajo.

En un estudio realizado por Pérez, Navarro y Garrido (2014) sobre tiempos perdidos en la fabricación de azúcar, concluyeron que se redujo la producción de azúcar en un 23 % debido a un incremento del 20 % en los paros de producción en la fábrica analizada y resaltaron la importancia de un análisis estadístico frecuente para evitar una reducción importante en la productividad. El estudio

anterior aportó información sobre el enfoque a tomar al analizar los tiempos perdidos en la producción de azúcar.

Garcés y Castrillón (2007) presentaron una metodología de análisis de información para reducir los tiempos muertos en un sistema de producción, el cual comenzó con la etapa de recolección de los datos, estandarización de la descripción de los fallos, la aplicación de diferentes técnicas de minería de datos, la selección del modelo estadístico adecuado y conclusión de los resultados obtenidos. Esta metodología se utilizó en gran parte durante el desarrollo de este trabajo.

Por otro lado, Reyes y García (2005) mostraron una metodología para extraer información útil de las bases de datos con un proceso conocido como KDD (*Knowledge Discovery Database*). Este proceso se utilizó para definir el enfoque de la investigación en función de los cuatro grupos posibles: agrupación, clasificación, asociación y predicción. Se seleccionó el enfoque de predicción para cumplir los objetivos del informe.

Ali, Badar y Fatima (2015) utilizaron los modelos autorregresivos integrados con medias móviles para hacer proyecciones en la producción de azúcar y algodón. Esta investigación influyó en el presente informe al seleccionar los criterios AIC y BIC, como parámetros del mejor modelo para hacer predicciones. Santana (2006) utilizó las técnicas de suavizado exponencial, modelos SARIMA y redes neuronales aplicados a la inflación colombiana. De esta investigación se utilizó el tratamiento de los datos con el suavizado exponencial y las técnicas de comparación de modelos.

También Contreras, Zúñiga, Martínez y Sánchez (2016) utilizaron modelos de suavizado exponencial y medias móviles aplicados al almacenamiento de

productos perecederos para organizar distintas cámaras frigoríficas disponibles. El análisis de series de tiempo no se aplicó a todas las cámaras frigoríficas por la cantidad, pero se utilizó un diagrama de Pareto para seleccionar las que mayor impacto tuvieron. Esta técnica de selección de los datos representativos utilizó Koch (2013) como la regla de 80/20 para variables que impactan en el resultado de cualquier proceso. Estas investigaciones influyeron en el método utilizado para identificar los problemas recurrentes que mayor impacto tuvieron en la producción del ingenio bajo estudio.

Wongkoon, Pollar y Jaroensutasinee (2009) aplicaron los modelos ARIMA y SARIMA para la predicción de casos de dengue hemorrágico en Tailandia, aplicando el coeficiente de correlación y coeficiente de correlación parcial para cada modelo obtenido. Esta investigación aportó una correcta forma de abordar los correlogramas, con interpretación puntual y un ejemplo sobre la influencia lineal de valores pasados sobre el valor presente en una serie de tiempo estacionaria y no estacionaria. Ruiz, Hernández y Zulueta (2010) utilizaron un modelo ARIMA para pronosticar la producción de azúcar en una fábrica en Cuba, dándole valor agregado al análisis para planificar la etapa de producción siguiente, de este estudio se utilizó el enfoque de planificación con la predicción de series temporales.

Por último, Delgadillo, Ruíz, Ramírez, Leos, Salas y Valdez (2016) elaboraron un modelo estadístico para el pronóstico del rendimiento de granos básicos en México y concluyeron que las series temporales son útiles para dar soporte en la toma de decisiones, pero no se puede predecir un valor exacto dada la característica estocástica de las series de tiempo. El aporte de esta investigación fue la búsqueda y selección del mejor modelo que minimice los errores con un buen ajuste en los datos reales.

2. MARCO TEÓRICO

2.1. Datos, información y conocimiento

Es importante conocer que, en la jerarquía del conocimiento, el dato es unidad mínima de almacenamiento (una fecha, un dígito, un conjunto de letras), la información como un conjunto de datos formados de modo que pueda transmitir un mensaje y el conocimiento que implica una combinación de experiencias, información relevante y contextualizada sobre ciertos datos obtenidos (Cairo ,2011).

2.2. Calidad de los datos

Para cuantificar la calidad de los datos con que se trabaja, se debe tomar en cuenta las siguientes características UNIR (2019):

- **Completitud:** es la cantidad de datos disponibles respecto al total que se tiene.
- **Credibilidad:** representa la confianza y veracidad de los datos que provienen de una fuente específica.
- **Precisión:** es la cantidad de datos correctos en un conjunto finito de información.
- **Consistencia:** hace referencia a la coherencia de los datos respecto a su valor.

- Interpretabilidad: indica el grado en que los datos pueden tener sentido al momento de ser analizados por una persona.

2.3. Distribuciones muestrales

Al momento de analizar información, es importante conocer cómo se comportan los datos en función de la frecuencia de aparición de cada uno de sus valores correspondientes, es en donde se introduce el concepto de distribuciones muestrales. Se dice entonces que una distribución muestral está asociada a una distribución de probabilidad que se obtiene de forma empírica con muestras tomadas de una población (Celis y Labrada, 2014).

2.4. Distribución normal

Es una distribución de probabilidad específica para variables continuas, utilizada en diversos campos como economía, ciencias exactas, medicina y en aspectos de la vida cotidiana (Milton, 2007). Las características más importantes de una distribución normal son:

- Es simétrica: sus valores se centran en la esperanza matemática de la población, la media.
- Medidas de tendencia central: en la distribución normal, la media toma el mismo valor que la mediana y la moda.
- Valor de probabilidad: el área bajo la curva suma 1, debido a que es un valor de probabilidad.

- Regla empírica: la cantidad de datos que se encuentran en un rango de $\pm$ una desviación estándar respecto al valor central corresponde a un 68 % del área total. Cuando el rango aumenta a dos desviaciones estándar, la cantidad de datos bajo la curva es de un 95 % y cuando el rango se extiende a tres desviaciones estándar, la cantidad de datos dentro de este intervalo es del 99.7 % (Gutiérrez y Salazar, 2013).

2.5. Series de tiempo

Una serie de tiempo es un conjunto de valores ordenados cronológicamente en un intervalo de tiempo constante y cuyo elemento de análisis que se suele utilizar es un modelo que permita reproducir el comportamiento de una variable de interés (Caridad y Ocerín 2012). Cuando únicamente se analiza una serie temporal en función de sus valores pasados, se denomina univariante. Por el contrario, si se analizan varias series temporales a la vez con relación entre ellas, se denominan multivariantes (González, 2009).

2.6. Procesos estocásticos estacionarios

Se define un proceso estocástico estacionario cuando su media y su varianza son constantes en el tiempo, si el valor de la covarianza entre dos períodos depende de la distancia o rezago entre estos dos períodos, y no del tiempo en el cual se calculó la covarianza (Gujarati y Porter, 2009).

$$E(Y_t) = \mu \quad (\text{Ec.1})$$

$$E(Y_t) = E(Y_t - \mu)^2 \quad (\text{Ec.2})$$

$$Cov(Y_t, Y_{t+k}) = E(Y_t - \mu)(Y_{t+k} - \mu) \quad (\text{Ec.3})$$

Donde:

$E(Y_t)$ = es la media constante de la serie de tiempo

Y_k = es la covarianza en el rezago k y Y_{t+k}

(Y_t) = es la varianza de la serie de tiempo

En otras palabras, una serie de tiempo no estacionaria tendrá una media que varía con el tiempo o una varianza que cambia con el tiempo, o ambas.

2.7. Procesos estocásticos integrados

Cuando una serie no es estacionaria, un procedimiento común es calcular las diferencias entre los valores de un período anterior y el actual, esto ayuda a volver estacionaria la serie en estudio. Generalmente cuando la serie es estacionaria sin necesidad de integrarla, es de orden 0. Cuando se calcula las primeras diferencias de la serie, se le conoce como serie integrada de orden 1 y así sucesivamente (Gujarati y Porter, 2009).

2.8. Pruebas de estacionariedad

Existen diversas pruebas para ver la estacionariedad de una serie temporal, las siguientes opciones son las más utilizadas a nivel general y las que se aplicaron en este trabajo de investigación:

- Análisis gráfico: una gráfica de líneas suele representar correctamente si una serie temporal es estacionaria, esto debido a que intuitivamente se aprecian los cambios y tendencias de los valores sobre un valor específico cuando se grafican en función del tiempo.

- Función de autocorrelación FAC: puede representar la dependencia lineal de los diferentes retrasos sobre el valor actual, determinado por un nivel de significancia y puede ser visualizado en un correlograma.

$$\rho = \frac{Y_k}{Y_0} \text{ (Ec.4)}$$

Donde:

ρ = representa el coeficiente de correlación parcial.

Y_k = es la covarianza en el retraso k.

Y_0 = es la varianza del conjunto de datos.

2.9. Prueba de raíz unitaria (Dickey Fuller)

La prueba de raíz unitaria es útil cuando se desea determinar si una serie de tiempo es estacionaria.

Prueba de raíz unitaria: dado un proceso estocástico que se inicia con:

$$Y_t = \rho Y_{t-1} + \mu_t \text{ (Ec.5)}$$

Donde:

Y_t = representa el valor actual de una serie de tiempo

ρ = es el coeficiente de influencia del retraso en el tiempo t-1

Y_{t-1} = es el valor de la serie en el retraso t-1

μ_t = es un término de error de ruido blanco

Debido a que no se puede probar la hipótesis de que $\rho = 1$ por medio de mínimos cuadrados ordinarios, la ecuación 5 se puede modificar para el cálculo de las primeras diferencias de la siguiente manera:

$$\Delta Y_t = \delta Y_{t-1} + \mu_t \text{ (Ec.6)}$$

Donde:

ΔY_t = es la diferencia del valor actual y su primer retraso

δ = es el valor del estadístico de prueba calculado como $(\rho - 1)$

La prueba de hipótesis de la raíz unitaria consiste en probar que el valor del estadístico de prueba $\delta=0$ (hipótesis nula) o que $\delta < 0$ (hipótesis alterna). Entonces, si $\delta=0$, implica que $\rho = 1$ y se concluye que la serie de tiempo no es estacionaria. Si no se acepta la hipótesis nula, se concluye que la serie es estacionaria. Una de las pruebas más populares de raíz unitaria es la de Dickey-Fuller, con el supuesto de que el término de ruido blanco está correlacionado con en número de retrasos.

2.10. Pronósticos de una serie de tiempo

Existen 5 enfoques económicos aplicados a las series de tiempo para la realización de pronósticos: modelos de suavizamiento exponencial, modelos de regresión uniecuacionales, modelos de regresión con ecuaciones simultáneas, modelos autorregresivos integrados en promedios móviles (ARIMA) y modelos de vectores autorregresivos (VAR).

Los enfoques más utilizados para el pronóstico de series de tiempo son el suavizado exponencial y los modelos ARIMA, proporcionando una solución complementaria a diversos problemas.

Por un lado, los modelos de suavizado exponencial se basan en una descripción de la tendencia y la estacionalidad de los datos, los modelos ARIMA describen las autocorrelaciones existentes en los datos analizados (Hyndman y George, 2021).

2.11. Métodos de suavizado exponencial

Son métodos que se utilizan generalmente para ajustar datos históricos de una serie temporal, utilizados generalmente en áreas en donde no es necesario un proceso sumamente riguroso. Dentro de los métodos de suavizado exponencial, los más importantes son: suavizado exponencial simple, método de Holt y el método de Holt-Winters, este último toma en cuenta la tendencia y estacionalidad de la serie analizada (Gujarati y Porter, 2010).

Los pronósticos generados con los métodos de suavizado exponencial son promedios ponderados de observaciones pasadas, con los pesos que se reducen exponencialmente a medida que las observaciones se alejan del dato más reciente. Estos métodos generan pronósticos confiables de forma rápida y cuya aplicación es amplia para diferentes series de tiempo, siendo atractivos para su utilización en la industria (Hyndman y George, 2021).

2.12. Modelos ARIMA

Estos métodos son muy conocidos con el nombre de Box-Jenckins o como modelos ARIMA, cuya base es la construcción de un modelo que tome en cuenta las propiedades estocásticas de los datos analizados, explicados por sus valores pasados, sin un fundamento económico (Shumway y Stoffer, 2011). Los modelos ARIMA pueden ser explicar una serie de tiempo o muchas series de tiempo (multivariantes).

En un modelo autorregresivo se pronostica la variable de interés utilizando una combinación lineal de valores pasados de la variable. El término auto regresión indica que se trata de una regresión de la variable contra sí mismo, el cual se puede escribir de la siguiente manera:

$$Y_t = C + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \dots + \phi_p y_{t-p} + \varepsilon \text{ (Ec.7)}$$

Donde:

Y_t = es la serie temporal analizada

ε = es el ruido blanco

c = es una constante para la serie analizada

ϕ_p = es el coeficiente del retraso p

De igual forma, para un modelo de medias móviles tiene una estructura similar a un modelo autorregresivo, con unas pequeñas diferencias:

$$Y_t = C + \phi_1 e_{t-1} + e_t \text{ (Ec.8)}$$

Donde:

Y_t = es la serie temporal analizada

e_t = representa a los residuos en el período t

c = es una constante para la serie analizada

ϕ = es el coeficiente del residuo en el período t

2.13. Principio 80-20

Este principio fue propuesto por el economista italiano Wilfrido Pareto, cuya utilidad inicial fue la descripción de la riqueza de diferentes poblaciones, pero posteriormente se aplicó para entender diferentes fenómenos estudiados en ciencia, política o cualquier otra disciplina. El principio de Pareto, como se le conoce también, se basa en que un reducido número de sucesos tiene un efecto mayoritario en la mayor parte de aplicaciones (Koch, 2013). Lo anterior se puede definir como: el 20 % de los eventos estudiados tienen un efecto del 80 % de la variable de respuesta de interés.

2.14. Proceso de producción del azúcar

Describir todo el proceso de producción de azúcar resulta muy complejo y extenso, pero al tomar en cuenta únicamente los procesos en los cuales existan problemas que logren detener la línea de producción, se puede explicar a través de etapas clave:

- Preparación de la caña
- Molienda
- Clarificación
- Evaporación
- Cristalización
- Centrifugación

2.15. Etapa 1: preparación de la caña de azúcar

La preparación de la caña consiste en reducir el tamaño de estas utilizando equipos de corte, posterior a ello, se busca exponer las fibras que contienen la sacarosa utilizando desfibradoras (equipos con martillos rotativos que golpean la caña a gran velocidad).

La preparación ayuda a una alimentación constante y homogénea, permite también cuantificar el material entrante mediante el uso de básculas colocadas en las bandas transportadoras. Con el fin de lograr una adecuada preparación de

la caña, se puede colocar en diferentes configuraciones las precuchillas, picadoras y la desfibradora (Melgar et al., 2014).

2.16. Etapa 2: molienda de la caña de azúcar

Partiendo de que la caña está compuesta por fibra, agua y sólidos disueltos (dentro de los cuales se encuentra la sacarosa), el objetivo principal de la molienda es extraer el jugo (agua y sólidos disueltos) de la fibra. Para lograr lo anterior, se hace pasar la caña a través de los molinos distribuidos en serie. Los molinos son rodillos de gran tamaño diseñados para extraer tanto jugo como sea posible, luego el jugo es recolectado en bandejas que lo almacenan en tanques previo al proceso de clarificación y el bagazo es utilizado como biomasa para las calderas, que se encargan de producir el vapor que utiliza la fábrica y en la generación de energía eléctrica (Rein, 2012).

2.17. Etapa 3: clarificación

Cuando el jugo es extraído de la caña en los molinos, suele arrastrar muchos elementos como tierra, hojas, arena, bagacillo y otras formas propias de la caña. La clarificación es entonces, la remoción de todas esas impurezas en el jugo de caña, con el objetivo de alcanzar la calidad mínima necesaria para la producción de azúcar.

La clarificación se lleva a cabo mediante el uso de calentadores de placas, calentadores de concha y tubos, tanques de remoción de lodo (conocidos como clarificadores) y el uso de químicos como la cal, floculantes, decolorantes, ácido fosfórico y azufre que cumplen una función específica (Melgar et al., 2014). Los subprocesos que se pueden encontrar en la clarificación de jugo son los siguientes:

- Sulfitación
- Alcalización
- Calentamiento
- Decantación de sólidos en suspensión
- Desalojo de cachaza
- Separación de jugo en la cachaza

El proceso de clarificación es de suma importancia para el resto de las etapas en la producción de azúcar, pues es fundamental cumplir con la calidad mínima para poder asegurar la producción de azúcar de calidad.

2.18. Etapa 4: evaporación

Etapa en la cual se concentra el jugo clarificado, llegando a obtener un 68 % de contenido de sólidos disueltos y por lo tanto, es el proceso en el cual se utiliza la mayor parte de la energía térmica obtenida del vapor de calderas y es crucial para la eficiencia energética de la fábrica. El proceso comienza con la recepción del jugo claro, el cual es precalentado a 110 °C para que se facilite la separación del agua en el jugo en los evaporadores (equipos que se encargan de concentrar el jugo de caña utilizando vapor y vacío) y termina con la entrega del jugo concentrado, conocido como meladura, la cual puede clarificarse para la producción de azúcar con bajo color o bien ir directo a los cristalizadores para la producción de azúcar con alto color. Actualmente la evaporación se basa en los principios de Rillieux, el cual enfoca el uso del vapor con un arreglo de los

evaporadores de múltiple efecto y el arreglo adecuado que permita liberar continuamente los gases incondensables (Chen y Chou, 1993).

2.19. Etapa 5: cristalización

Proceso en el cual se produce la formación de los cristales de sacarosa y se controla su desarrollo para obtener las características adecuadas para el almacenamiento. Los cristales de sacarosa se desarrollan principalmente porque las moléculas de sacarosa en la miel y en disolución, logran concentrarse en una red cristalina, de tal manera que crecen y por lo tanto, se reduce la pureza de la miel, lo cual se le conoce como agotamiento (Melgar et al., 2014). Los equipos responsables de dicho proceso son los tachos o cristalizadores, son equipos muy parecidos a los evaporadores pero son alimentados por meladura y semillas de sacarosa en una solución de alcohol, que facilita el desarrollo y formación de los cristales

2.20. Etapa 6: centrifugación

En esta etapa se separa a los cristales de azúcar de la masa cocida (cristales más miel) por medio de centrifugación, debido al alto contenido de sólidos disueltos y a la consistencia de la masa. Los equipos encargados de dicho proceso son las centrífugas, las cuales pueden funcionar por lotes o de forma continua, siendo de mucha importancia la temperatura de la masa a centrifugar, para asegurar el correcto funcionamiento de estas. Luego del proceso de centrifugación, se procede a transportar el grano hacia las bodegas, en donde son almacenados en sacos de diferentes presentaciones (Rein, 2012).

3. PRESENTACIÓN DE RESULTADOS

De acuerdo con los objetivos propuestos, se presentan los siguientes resultados:

3.1. Objetivo 1: categorizar los problemas que causan tiempos perdidos en la operación, en función de la descripción

Se categorizaron los tiempos perdidos por causas comunes en la descripción, los cuales involucran seis grupos distintos según el problema asociado: la materia prima, fallas en los equipos, fallas en la operación, mantenimientos programados y otros que incluyen partes del proceso dependientes de la logística y transporte.

También se agregaron las descripciones específicas para cada tiempo perdido, con el fin de explicar detalladamente los problemas que ocurren con mayor frecuencia.

En la siguiente tabla se muestran las categorías y las descripciones específicas para los tiempos perdidos, ordenados por frecuencia y duración, tanto en valores absolutos como en valores porcentuales.

Tabla II. Categorización de problemas

Descripción general	Descripción específica	Frecuencia	Frecuencia [%]	Tiempo perdido [min]	Tiempo perdido [%]
Materia prima	Tanques llenos de meladura	714	25%	15,150	17%
Otros	Problemas con piedras	497	17%	5,889	7%
Materia prima	Tanques llenos de miel A	264	9%	7,805	9%
Materia prima	Bagazo atorado	208	7%	2,816	3%
Materia prima	Tanque lleno de jugo claro	165	6%	2,105	2%
Fallas en equipos	Problemas eléctricos	107	4%	2,399	3%
Fallas en equipos	Cadenas descarriladas en patio	105	4%	4,108	5%
Materia prima	Bajo brix de meladura	103	4%	1,763	2%
Fallas en equipos	Ajustes mecánicos en molinos	98	3%	1,466	2%
Fallas en equipos	Problemas mecánicos en centrífugas	70	2%	2,574	3%
Materia prima	Caña atorada	70	2%	3,604	4%
Mantenimiento	Mantenimiento programado	62	2%	8,928	10%
Mantenimiento	Limpieza del tachó continuo	58	2%	8,650	10%
Abastecimiento	Falta de caña	52	2%	9,885	11%
Materia prima	Baja presión de vapor de escape	51	2%	1,287	1%
Fallas en equipos	Fallas en bombas	45	2%	1,824	2%
Fallas en equipos	Problemas con el virador	39	1%	1,933	2%
Fallas en equipos	Problemas mecánicos en molinos	37	1%	1,133	1%
Fallas en equipos	Problemas mecánicos en patios	34	1%	862	1%
Fallas en equipos	Falla en sistemas de comunicación	33	1%	636	1%
Fallas en equipos	Ajustes mecánicos en patios	26	1%	966	1%
Fallas de operación	Baja presión de vapor vegetal II	25	1%	451	1%
Falla de operación	Problemas mecánicos en tachos	14	0.5%	418	0.5%
Materia prima	Tanque de jugo alcalizado lleno	9	0.3%	100	0.1%
Materia prima	Lleno de masa A	6	0.2%	114	0.1%
Fallas de operación	Pérdida de vacío	6	0.2%	275	0.3%
Total		2,898	100%	87,141	100%

Fuente: elaboración propia, realizado con Microsoft Excel.

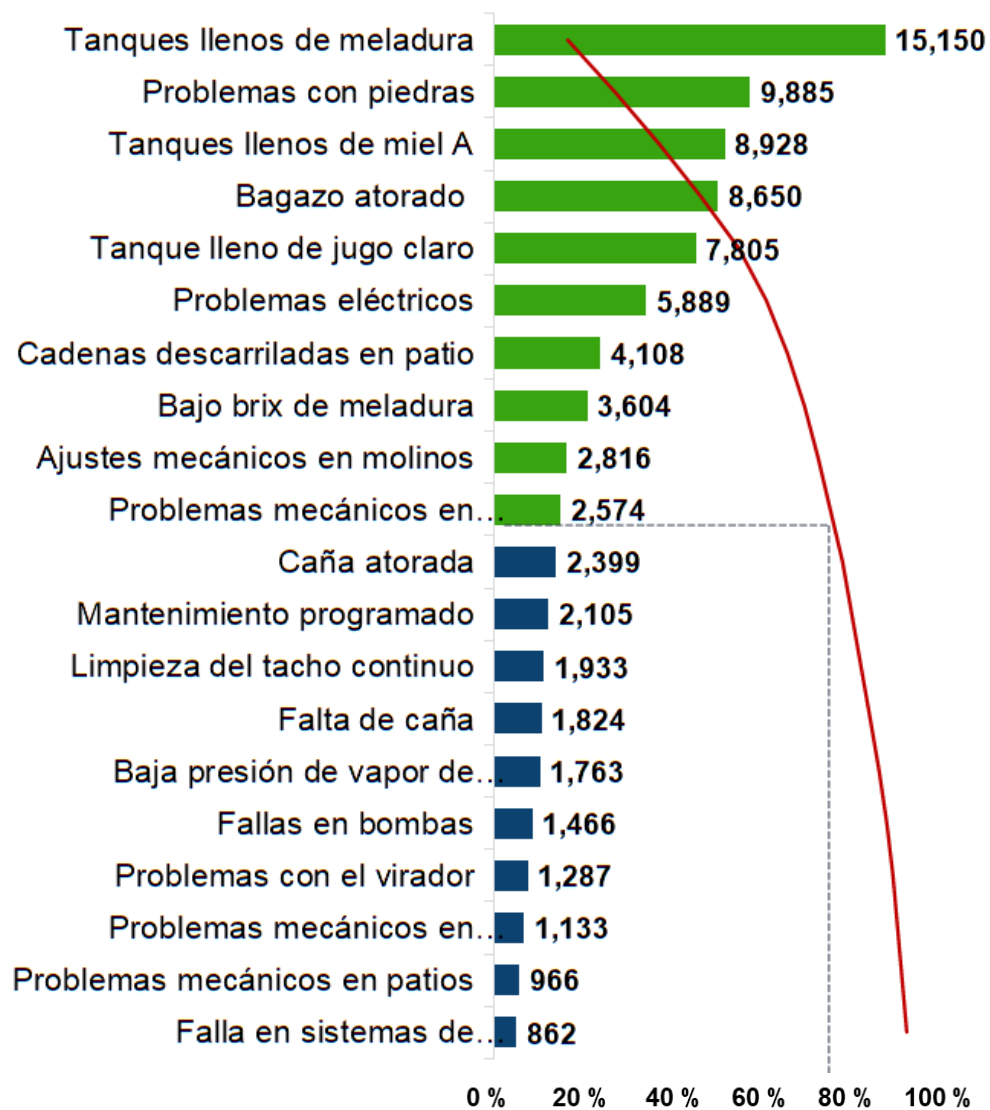
3.2. Objetivo 2: clasificar los problemas que generan la mayor cantidad de tiempo perdido en la operación mediante la regla 80/20

Luego de categorizar los problemas que causaron tiempos perdidos según la descripción con la cual se registró en operación, fue necesario clasificarlos en función del impacto acumulado que provocaron durante las zafas 2017-2018, 2018-2019 y 2019-2020.

La clasificación se basa en la regla del 80-20 y se muestra utilizando un diagrama de Pareto, con ello se logró clasificar al 20 % de los problemas que tuvieron un impacto del 80 % en la producción de azúcar, tomando en cuenta la

frecuencia (asociada a la continuidad del proceso a corto plazo) y la duración de cada uno (asociada a la continuidad del proceso a mediano y largo plazo).

Figura 1. **Diagrama de Pareto para la duración de los tiempos**

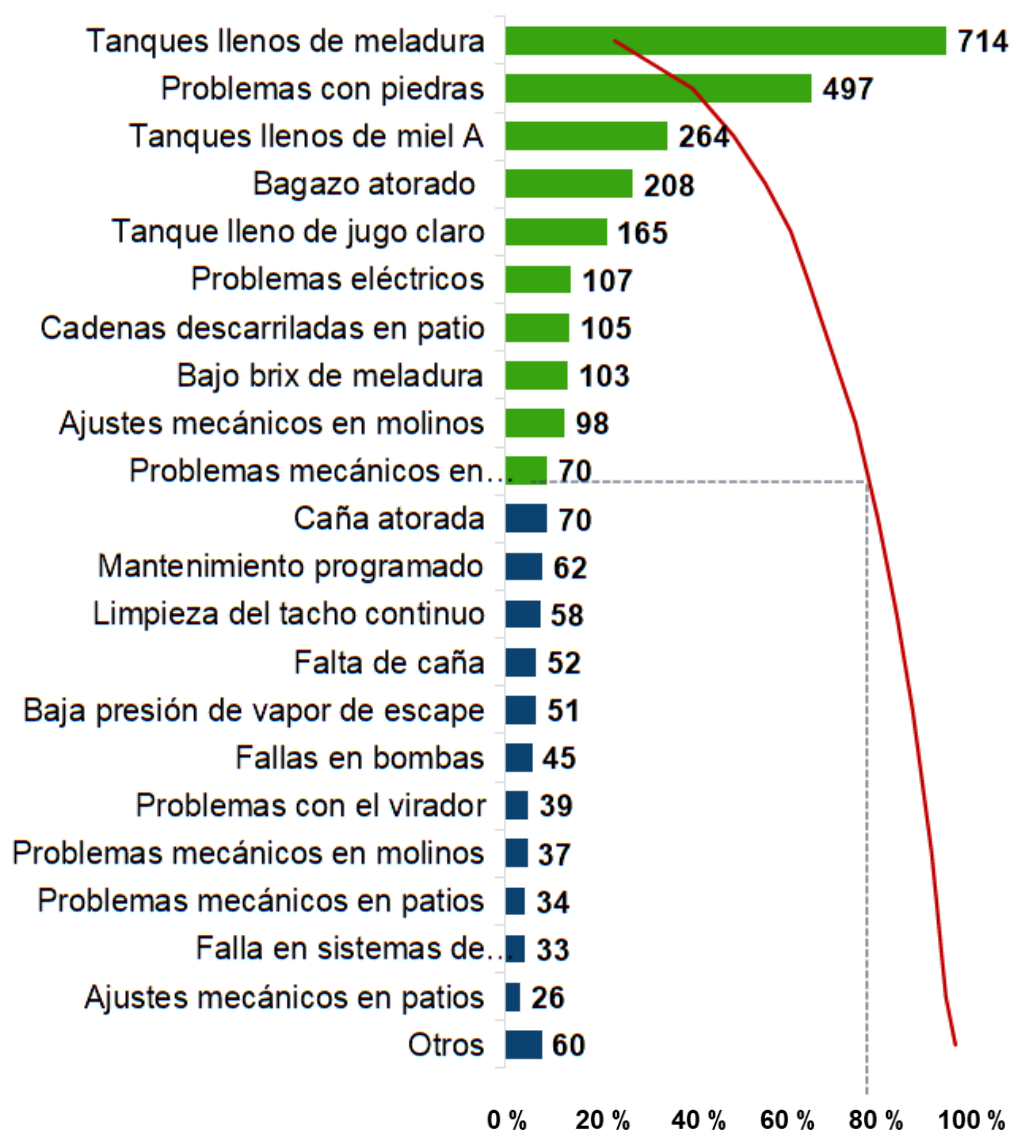


Fuente: elaboración propia, realizado con Microsoft Excel 2019.

En la figura 1 se muestran los tiempos perdidos ordenados de mayor a menor según la duración total registrada, la línea roja representa la duración acumulada en porcentaje, la línea discontinua gris representa el punto de corte para identificar al 20 % de los problemas que impactaron en el 80 % de la duración total de los tiempos perdidos, mostrados también con las barras de color verde y el resto de los problemas con las barras de color azul. Los problemas clasificados por el gran impacto que tuvieron en la operación según su duración son los siguientes:

- Tanques llenos de meladura
- Falta de caña
- Mantenimiento programado
- Limpieza del tacho continuo
- Tanques llenos de miel A
- Problemas con piedras
- Cadenas descarriladas
- Caña atorada
- Bagazo atorado
- Problemas en centrífugas

Figura 2. **Tiempos perdidos según su frecuencia**



Fuente: elaboración propia, realizado con Microsoft Excel 2019.

En la figura 2 se muestran los tiempos perdidos ordenados en función de su frecuencia de mayor a menor, la línea roja representa el porcentaje acumulado de la frecuencia y la línea discontinua indica el punto de corte entre el 20 % de

los problemas que causaron un impacto del 80 % en la operación (barras de color verde) y el resto de los problemas (barras de color azul).

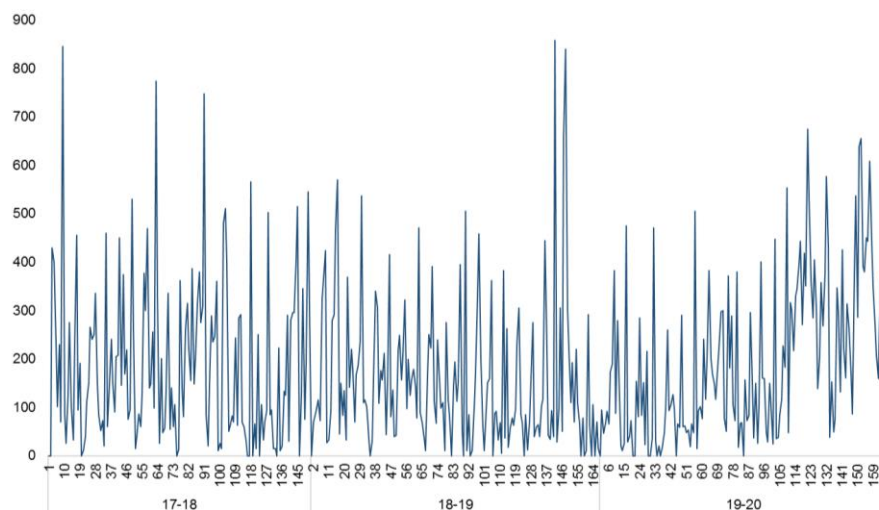
La clasificación de los problemas con mayor impacto en la continuidad del proceso según su frecuencia son los siguientes:

- Tanques llenos de meladura
- Problemas con piedras
- Tanques llenos de miel A
- Bagazo atorado
- Tanque lleno de jugo claro
- Problemas eléctricos
- Cadenas descarriladas en el patio
- Bajo brix de meladura
- Ajustes mecánicos en molinos
- Problemas mecánicos en centrífugas

3.3. Objetivo 3: describir el comportamiento de los problemas que causan la mayor cantidad de tiempo perdido mediante el uso de series de tiempo

Teniendo los problemas agrupados por causas comunes, fue necesario describir el comportamiento de los paros de molienda a través del tiempo. Dichos paros de molienda se graficaron utilizando series temporales, tomando en cuenta la duración y la frecuencia de estos. También utilizó un correlograma con los primeros 40 retrasos para determinar el efecto de los días de producción sobre los paros de molienda y se utilizó la descomposición clásica con un modelo aditivo para separar las componentes de tendencia, estacionalidad y residual.

Figura 3. Duración de los tiempos perdidos

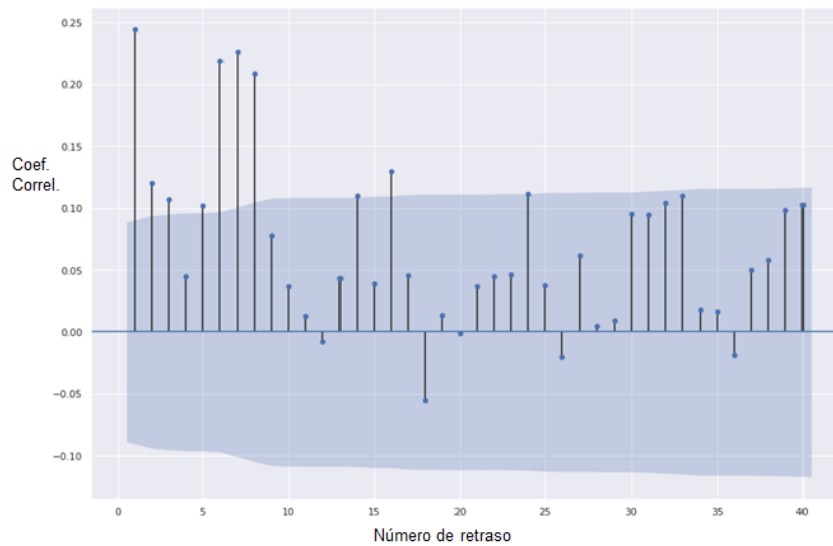


Fuente: elaboración propia, realizado con Microsoft Excel 2019.

La figura 3 muestra la duración de los tiempos perdidos registrados (mostrados en minutos) de las zafras 2017-2018, 2018-2019 y 2019-2020,

tomando en cuenta únicamente el período productivo (de noviembre a mayo de cada año).

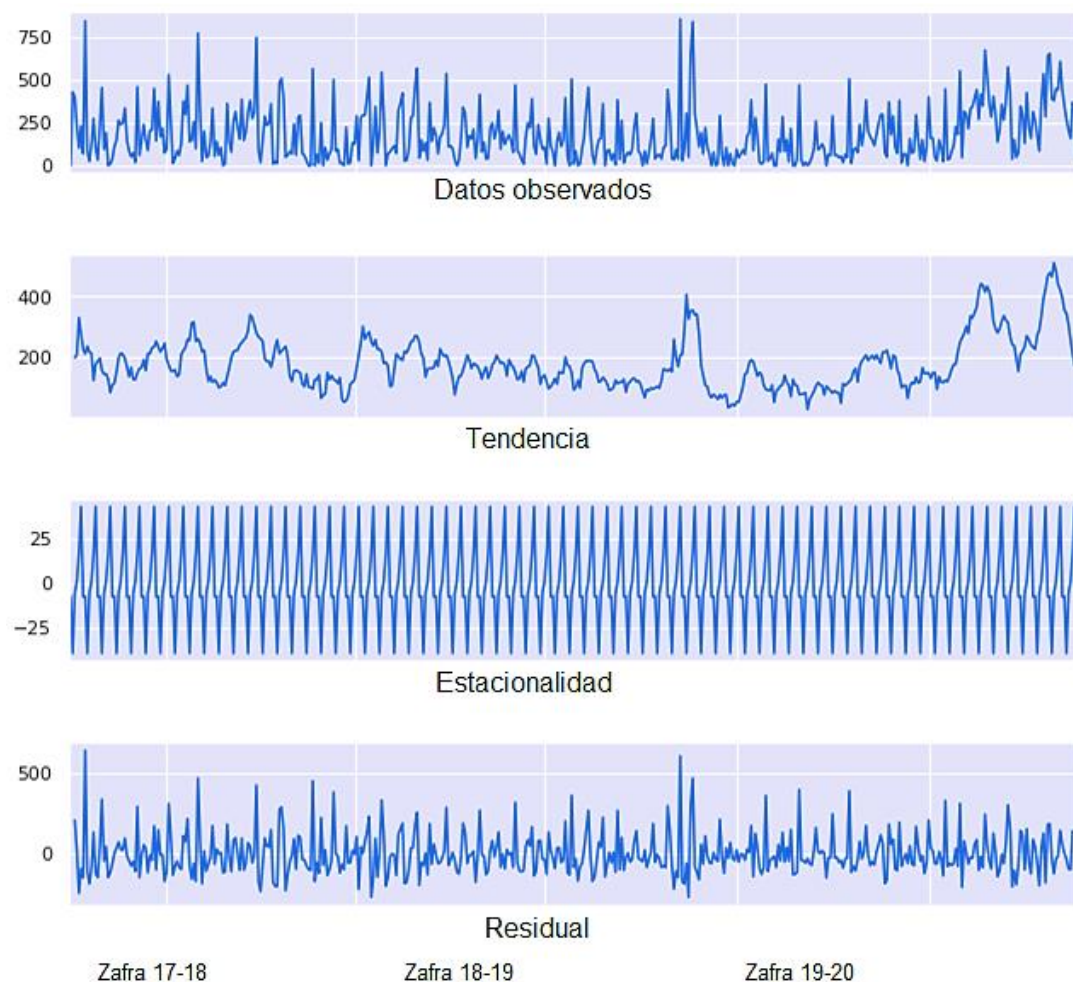
Figura 4. **Correlograma de la duración de los tiempos perdidos**



Fuente: elaboración propia, realizado con R 4.1.2

La figura 4 muestra el correlograma de la duración de los paros de molienda para los primeros 40 retrasos, esta gráfica indica la influencia de la operación en días pasados sobre el valor actual del tiempo perdido. El área más oscura, es el nivel de significancia del 5 % y está asociado a un efecto despreciable. En el eje Y, se muestra el coeficiente de correlación de cada retraso (entre el valor actual y el valor del período anterior).

Figura 5. **Descomposición de los tiempos perdidos (duración)**

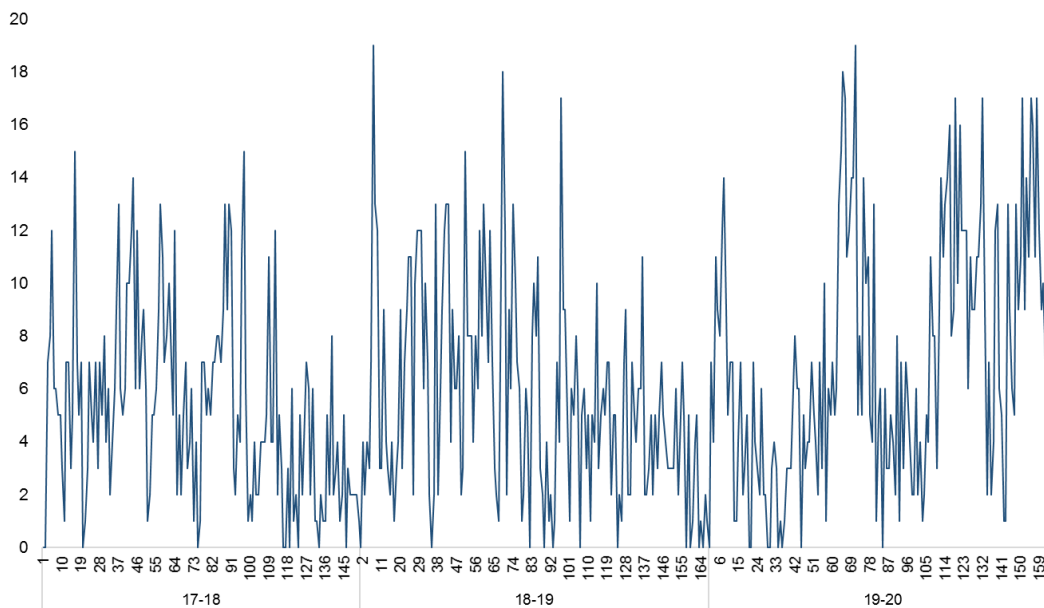


Fuente: elaboración propia, realizado con R 4.1.2.

La figura 5 muestra la descomposición de la serie temporal de los tiempos perdidos en cuanto a su duración, utilizando un modelo aditivo. La serie original está en el primer cuadro, la tendencia se observa en el segundo cuadro, la estacionalidad en el tercero y por último, la componente aleatoria o residual. Al igual que en la figura 3, la serie original y sus componentes se basan

exclusivamente en el período productivo de un ingenio (de noviembre a abril de cada año).

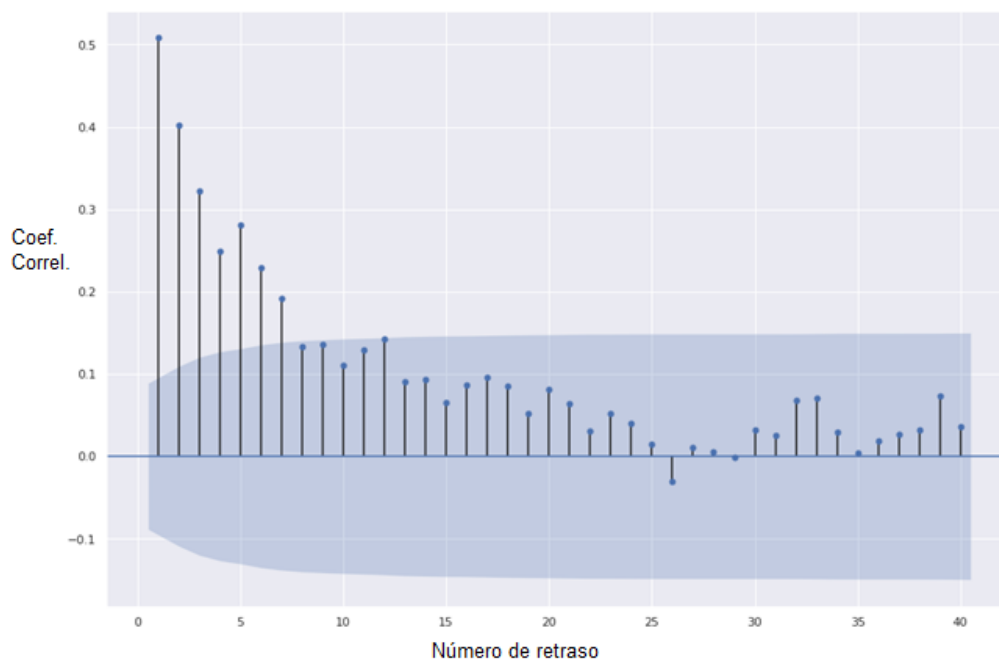
Figura 6. Frecuencia en los tiempos perdidos



Fuente: elaboración propia, realizado con Microsoft Excel 2019.

La figura 6 muestra la frecuencia de los tiempos perdidos registrados en función del tiempo, medidos por día productivo para la zafra 2017-2018, 2018-2019 y 2019-2020.

Figura 7. **Correlograma para la frecuencia de los tiempos perdidos**

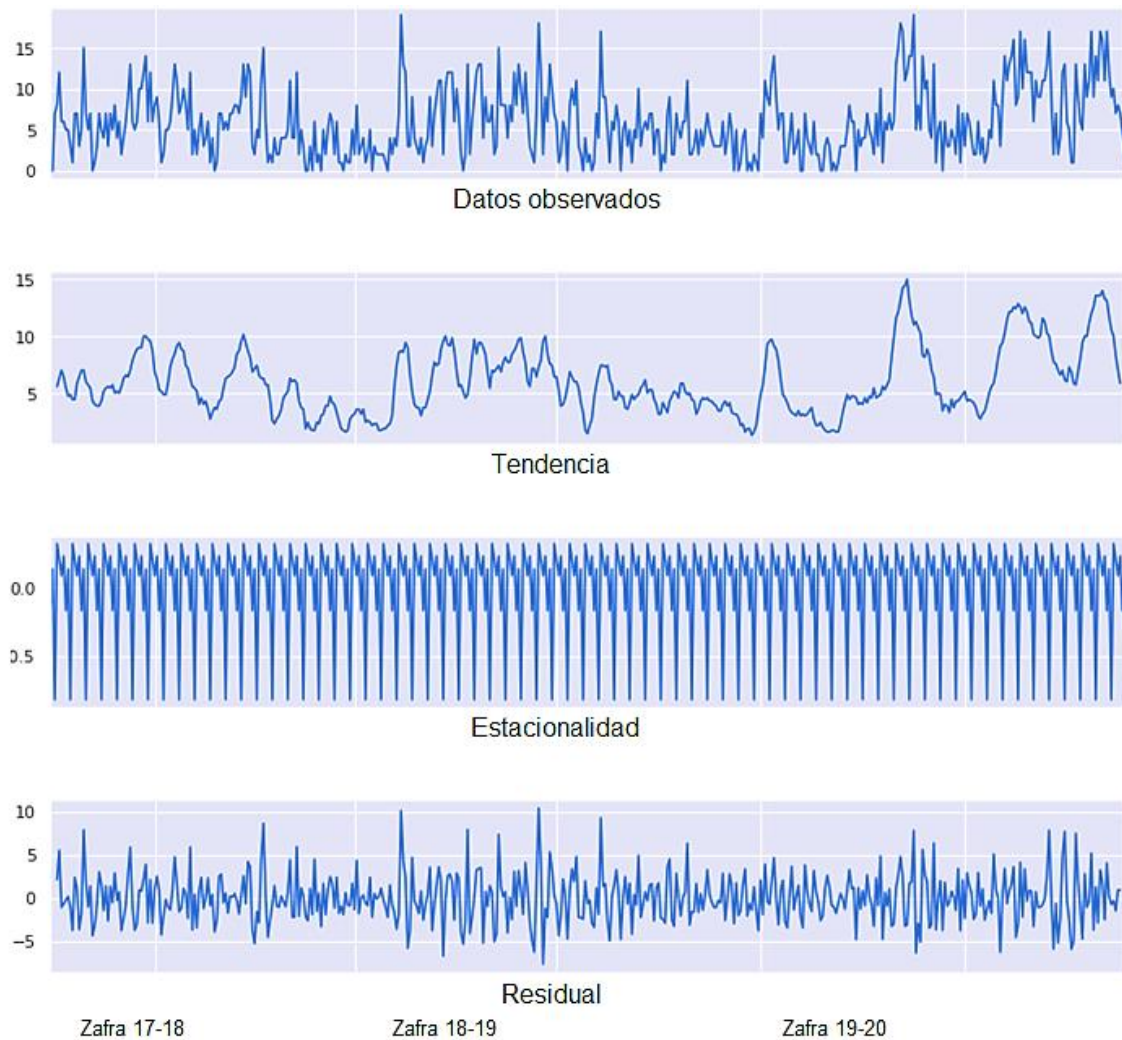


Fuente: elaboración propia, realizado con R 4.1.2

La figura 7 es el correlograma de frecuencia de los paros de molienda, que muestra la influencia de los primeros 40 retrasos (eje x) sobre el valor actual de la serie. El área oscura, es el nivel de significancia del 5 % y está asociada a la falta de significancia estadística de los retrasos sobre el valor actual, medidos con el coeficiente de correlación en el eje y.

Al igual que las series temporales de la figura 3 y figura 6, los correlogramas toman en cuenta todos los valores de tiempos perdidos para las zafras 17-18, 18-19 y 19-20.

Figura 8. **Descomposición de la frecuencia de los tiempos perdidos**



Fuente: elaboración propia, realizado con R 4.1.2.

La figura 8 muestra la descomposición de la serie temporal mostrada en la figura 6 con un modelo aditivo. En el primer cuadro está la serie original, en el segundo se observa la tendencia, en el tercero la estacionalidad y en el último cuadro, la componente residual o aleatoria. En el eje x se muestra el período del tiempo perdido y en el eje y, el recuento de los paros de molienda registrados.

3.4. Objetivo 4: pronóstico del tiempo perdido para la zafra 2020-2021, usando las técnicas de medias móviles y suavización exponencial

Entendiendo el comportamiento de los tiempos perdidos de las zafra 17-18, 18-19 y 19-20, se realizaron los pronósticos de estos para la duración (minutos) y la frecuencia (recuento de eventos) de los primeros 15 días de operación de la zafra 2020-2021, utilizando las técnicas de Holt Winters y ARIMA. Para la selección del mejor modelo, se tomó en cuenta los criterios de pérdida de información AIC y BIC.

Tabla III. Pronósticos con el método de Holt Winters

Zafra	Día	Tiempo [min]	Frecuencia
2020-2021	1	272	5
2020-2021	2	285	4
2020-2021	3	354	4
2020-2021	4	334	4
2020-2021	5	332	4
2020-2021	6	302	4
2020-2021	7	307	3
2020-2021	8	269	5
2020-2021	9	282	4
2020-2021	10	350	4
2020-2021	11	331	4
2020-2021	12	329	4
2020-2021	13	299	4
2020-2021	14	304	3
2020-2021	15	266	5

Fuente: elaboración propia, realizado con Microsoft Excel 2019.

La tabla III muestra las proyecciones realizadas con el método de Holt Winters, tomando en cuenta la información generada desde el año 2017 hasta el 2019.

Tabla IIIV. Pronósticos con el método ARIMA

Zafra	Día	Tiempo [min]	Frecuencia
2020-2021	1	248	4
2020-2021	2	285	5
2020-2021	3	291	6
2020-2021	4	292	7
2020-2021	5	292	8
2020-2021	6	292	9
2020-2021	7	292	10
2020-2021	8	292	11
2020-2021	9	292	11
2020-2021	10	292	11
2020-2021	11	292	12
2020-2021	12	292	12
2020-2021	13	292	12
2020-2021	14	292	12
2020-2021	15	292	12

Fuente: elaboración propia, realizado con Microsoft Excel 2019.

La tabla IV muestra los pronósticos obtenidos usando un modelo ARIMA (3,1,1) para la frecuencia y para la duración, un modelo ARIMA (1,1,2).

Con el fin de seleccionar el mejor modelo para los pronósticos, se utilizaron los criterios de pérdida de información de Akaike y Bayesiano, mostrados en la tabla V. Es importante tomar en cuenta el valor del AIC y BIC, siendo los valores más bajos, los que representan un mejor modelo.

Tabla IV. Criterios de selección

Tipo	Modelo	AIC	BIC
Duración	HW	4,997	5,043
Duración	ARIMA (3,1,1)	6,278	6,291
Frecuencia	HW	1,274	1,320
Frecuencia	ARIMA (1,1,2)	2,615	2,631

Fuente: elaboración propia, realizado con Microsoft Excel 2019.

3.5. Objetivo general: describir los problemas que causas la mayor cantidad de tiempo perdido con el uso de series de tiempo, generando un pronóstico para el período 2021

Los resultados muestran que los problemas que afectan más a la operación son los relacionados con la materia prima y los equipos, motivo por el cual, se sugiere que sean prioridad al momento de elaborar planes de mejora en los procesos de producción, pudiendo asegurar una reducción considerable de los mismos para la zafra 2020-2021. Los problemas que más influyen en la operación son los siguientes:

- Tanques llenos de meladura
- Problemas con piedras
- Tanques llenos de miel A
- Bagazo atorado
- Tanque lleno de jugo claro

- Problemas eléctricos
- Cadenas descarriladas en el patio
- Bajo brix de meladura
- Ajustes mecánicos en molino
- Problemas mecánicos en centrífugas
- Falta de caña
- Mantenimiento programado
- Limpieza del tachó continuo
- Cadenas descarriladas
- Caña atorada
- Problemas en centrífugas

Con todo lo anterior, se espera para los primeros 15 días de operación de la zafra 2020-2021 un total de 61 paros de molienda con una duración total de 4,626 minutos según el modelo de Holt-Winters (modelo con mejor ajuste).

4. DISCUSIÓN DE RESULTADOS

La presente investigación busca identificar problemas recurrentes y de gran impacto para una fábrica de producción de azúcar, apoyándose en la caracterización, clasificación y el uso de modelos estadísticos de series temporales para explicar el comportamiento de estos y generar pronósticos. Con todo lo anterior, se muestra una forma de analizar los datos de tiempos perdidos para ser una herramienta potencial al momento de tomar decisiones y generar planes de acción para mitigar dichos problemas encontrados.

4.1. Análisis interno

El método empleado para el desarrollo de los resultados se basa en el ordenamiento y estandarización de los registros, categorizando y clasificando los problemas que afectaron a la fábrica analizada, explicando la mayor parte de los paros de molienda que ocurrieron en las zafras 2017-2018, 2018-2019 y 2019-2020 y a partir de ello, generar pronósticos para el inicio del período siguiente, en este caso, la zafra 2020-2021. El método utilizado se considera el más adecuado para datos que se registran de forma abierta por parte de los operadores, pues cada uno de ellos registra cada paro en la línea de producción con sus propias palabras y esto aporta un grado de complejidad en los análisis más detallados a realizar.

Esta investigación muestra una limitación en cuanto a la recolección y análisis de los datos. La limitación encontrada está relacionada con los registros abiertos utilizados, pues al ser almacenados por diferentes personas durante la operación, se encontraron registros sin descripción alguna, los cuales fueron

eliminados por no poderse clasificar (estos registros representan el 1 % del total). A pesar de la primera limitación, se considera que la cantidad de datos es suficiente para garantizar la calidad de los resultados obtenidos.

4.2. Análisis externo

Tal como se abordó en la investigación sobre un diseño inteligente para reducir y eliminar los tiempos muertos en producción (Garcés y Castrillón, 2007), es importante emplear una metodología de recolección, ordenamiento y estandarización de los tiempos perdidos para poder clasificarlos y analizarlos correctamente, siendo un paso previo a la selección del modelo estadístico a utilizar y la generación de pronósticos, escalable a cualquier fábrica de producción que desee analizar sus tiempos muertos. Sin embargo, los resultados de esta investigación solo son aplicables a la fábrica que proporcionó los datos, ya que cuenta con una capacidad de molienda específica, con equipos colocados en configuraciones que pueden variar entre fábricas con capacidad de molienda similar, el plan de producción anual y la calidad de la materia prima e insumos que manejan. A pesar de esto, el método para abordar el análisis de los tiempos perdidos si puede ser aplicado a cualquier fábrica de producción, en diferentes contextos en cuanto al material a producir y la naturaleza de la operación.

4.3. Categorización de los tiempos perdidos

Uno de los pasos fundamentales previo a aplicar cualquier modelo estadístico es la categorización de los problemas registrados. En su investigación (Garcés y Castrillón, 2007) condensaron los tiempos muertos en 5 grupos según la naturaleza del acontecimiento registrado, que a diferencia de la presente investigación, se categorizaron los tiempos perdidos en 6 grupos importantes,

debido a la complejidad del proceso de producción y a los diferentes departamentos que aportan el abastecimiento de insumos y materia prima.

Los grupos en los cuales se categorizaron los tiempos perdidos se muestran en la tabla II, ordenados por una descripción específica, duración y frecuencia de cada uno. Los 6 grupos principales son:

- Materia prima
- Fallas en los equipos
- Mantenimientos
- Abastecimiento de materia prima
- Fallas en la operación
- Otros (el resto de los problemas que son menos frecuentes)

A su vez, cada categoría contiene una descripción más detallada del problema, para que en función de su duración y frecuencia, puedan ser abordados al momento de evaluar un plan de acción para su mitigación.

4.4. Clasificación de los problemas más importantes

De acuerdo con el principio 80-20 (Koch, 2013), la mayor parte de los efectos provienen de unas cuantas causas, a lo cual, fue necesario identificar las más importantes para la fábrica analizada en esta investigación. Se logró identificar las que mayor impacto tuvieron en el período analizado en función de

la frecuencia y la duración de estas, las cuales se muestran en la figura 1 y figura 2 respectivamente.

Se encontraron 10 problemas que provocaron el 80 % del tiempo total registrado, de los cuales, 4 pertenecen a la categoría de materia prima, 2 a fallas en equipos, 2 a mantenimientos programados, 1 a abastecimiento y 1 a la categoría de otros problemas menores. Para el total de paros registrados, el 80 % fue provocado por 10 problemas, dentro de los cuales, 10 están asociados a la materia prima, 4 a fallas en los equipos y 1 a problemas menores.

La clasificación anterior indica que la mayor parte de los problemas en cuanto a frecuencia y duración, se deben a la materia prima procesada. Según (Pérez, Navarro y Garrido, 2014), los tiempos perdidos relacionados con la materia prima provocan una reducción en las ganancias de hasta el 20 %, mostrando una oportunidad importante para profundizar y cuantificar las pérdidas obtenidas. Algo interesante obtenido de la clasificación es que los problemas con mayor tiempo acumulado pueden ser por diversas causas (materia prima, abastecimiento, fallas en equipos y problemas menores), pero los problemas cuya frecuencia acumulada es alta, están asociados únicamente a la materia prima, fallas en equipos y en menor proporción a los problemas menores. Se puede decir de manera general, que un plan de mitigación debe estar enfocado a la materia prima y a las fallas en los equipos.

4.5. Análisis descriptivo de los tiempos perdidos con series temporales

Según (Peña, 2010), las series temporales pueden perseguir dos fines, la descripción de un fenómeno y la predicción. Para los tiempos perdidos analizados por su frecuencia y duración, se utilizaron las series temporales con

el objetivo de encontrar patrones claros y posteriormente predecirlos para los primeros días de la zafra 2020-2021. Los resultados de describir los tiempos perdidos se muestran en la figura 3 y figura 6, mostrando un comportamiento con alta variabilidad, sin un patrón claro, tanto en la frecuencia como en la duración de estos.

También se utilizaron herramientas más detalladas para describir el comportamiento de los paros de molienda, una de ellas fue graficar un correlograma con 40 retrasos, figura 4 y figura 7. Ambas figuras muestran que los paros de molienda pueden estar influenciados por las condiciones operativas de hasta 7 días atrás, pues su valor deja de ser significativo en el octavo retraso. Tomando en cuenta esta información y la clasificación obtenida con la regla 80-20, se puede confirmar que la materia prima y las fallas en los equipos afectan significativamente las condiciones de operación por 7 días.

Por último, (Caridad, 2012) indica que la descomposición de una serie temporal ayuda a describir y reproducir la evolución de un evento estudiado, con una pauta de regularidad en los datos. Fue con la descomposición de la duración y frecuencia de los paros de molienda, que se logró identificar un patrón estacional, dichas descomposiciones se muestran en la figura 5 y figura 8. Se observa un patrón repetitivo en la componente estacional con un período de 7 días, lo que corresponde a los cambios de turno. Se asocian los cambios de turno a un incremento en los tiempos perdidos, influenciados por las condiciones propias que pueden manejar los operadores y que muestran un aspecto de mejora y estandarización, ya que con ello, se reduciría dicho patrón estacional. No se encontró ningún otro comportamiento en la tendencia y en la componente aleatoria-cíclica.

4.6. Predicción de los tiempos perdidos

Se utilizaron dos modelos para hacer las predicciones de los paros de molienda para los primeros 15 días de la zafra 2020-2021, los resultados se muestran en la tabla III utilizando el método de Holt Winters y en la tabla IV, utilizando el método ARIMA. El modelo obtenido con el mejor ajuste utilizando el criterio de pérdida de información AIC, es el de Holt Winters, el cual indica un total de 61 paros de molienda y un tiempo acumulado de 4,616 minutos (tiempo medio de 76 minutos). Por otro lado, el modelo obtenido con el método ARIMA, estima un total de 142 paros de molienda, un tiempo acumulado de 4,328 minutos y un tiempo medio de 30 minutos. Tal y como lo presenta (Contreras, Zúñiga, Martínez y Sánchez, 2016), la técnica de suavizado exponencial triple o Holt Winters, es sencilla pero con un buen ajuste para hacer pronósticos, en los resultados de esta investigación, fue el mejor modelo obtenido.

Con todos los resultados mostrados anteriormente, se logró categorizar a los problemas que afectan la continuidad de la producción de azúcar, identificando los que tienen un efecto significativo (materia prima y fallas en los equipos), que posteriormente se utilizaron para hallar patrones repetitivos en su ocurrencia (p.ej. cambios de turnos y el efecto de los 7 días) y a partir de ello, pronosticar una cantidad total de fallas esperadas y su respectiva duración, lo cual puede ser de vital importancia al momento de planificar acciones de mitigación y mejora del proceso productivo.

CONCLUSIONES

1. Los problemas que afectan la continuidad del proceso de producción de azúcar se categorizan en 6 grupos, siendo estos los problemas relacionados con la materia prima, fallas en los equipos, mantenimiento de equipos, abastecimiento de materia prima, fallas en la operación y otros problemas menores de diferente índole.
2. Clasificando los tiempos perdidos en función de la regla 80-20, se identifica el 20 % de los problemas que aportan el 80 % del tiempo total registrado y el 80 % de la frecuencia total, los cuales son: tanques llenos de meladura, miel A y jugo claro, bajo brix de meladura, falta de caña, caña y bagazo atorado, mantenimientos, limpieza de tacho continuo, cadenas descarriladas, ajustes y problemas mecánicos en centrífugas, molinos y problemas con piedras.
3. Se observa un comportamiento altamente aleatorio en los paros de molienda con ciclos de siete días correspondientes a los cambios en el plan de producción y cambios en los turnos del personal operativo. Se observa también un incremento de los paros de molienda de menor duración y mayor frecuencia debido a los cambios en la calidad de la materia prima.
4. Se infiere con un modelo de suavizado exponencial que los primeros 15 días de la zafra 2020-2021 tendrán 61 paros de molienda con un tiempo total de 4,616 minutos (76 minutos por paro). Por otro lado, se infiere con

un modelo ARIMA que los primeros 15 días de la zafra 2020-2021 tendrán 142 paros de molienda con un tiempo total de 4,328 (30 minutos por paro).

5. Se determina que los paros de molienda para los primeros 15 días de la zafra 2020-2021 están relacionados con la calidad de la materia prima (tanques llenos de miel, meladura o jugo), mantenimiento de los equipos (problemas mecánicos y eléctricos (cadenas y centrífugas), problemas operativos (bagazo atorado) y otros ajenos a la operación (piedras en el proceso).

RECOMENDACIONES

1. Categorizar los nuevos problemas registrados, con el fin de estandarizar el almacenamiento de la información que posteriormente facilitará el análisis y la toma de decisiones.
2. Priorizar los problemas clasificados con la regla del 80-20, pues son los que más impacto tendrán en los resultados al momento de reducir el tiempo perdido total.
3. Analizar el comportamiento que tienen los paros de molienda en función del tiempo, desde el punto de vista de asignación de turnos que ocurren cada 7 días, porque tienen un incremento detectado con la descomposición de las series de tiempo.
4. Comparar el pronóstico para los primeros 15 días de la zafra 2020-2021 con los datos reales en operación, para evaluar el ajuste del modelo obtenido en esta investigación.

REFERENCIAS

1. Ali, S., Nouman, B. y Fatima, H. (2015). Forecasting production and yield of sugarcane and cotton crops of Pakistan for 2013-2030. *Sarhad Journal of Agriculture*, 31(1), 1-9.
2. Anderson, D., Sweeney, D. y Williams, T. (2008). *Estadística para la administración y economía*. México, Ciudad de México: Cengage Learning Editores, S.A.
3. Asociación de azucareros de Guatemala. (20 de julio de 2020). *Productividad de toneladas de azúcar por hectárea*. Recuperado de <https://www.azucar.com.gt/azucar-de-guatemala-en-el-mundo/>.
4. Cairo, A. (2011). *El arte funcional*. Madrid, España: Fareso S.A.
5. Caridad, J. (2012). *Econometría: Modelos econométricos y series temporales*. Barcelona, España: Editorial Reverté, S.A.
6. Celis, A. y Labrada, V. (2014). *Bioestadística*. México: Editorial El Manual Moderno.
7. Chatfield, C. (2000). *Time-Series Forecasting*. Florida, Estados Unidos: Editorial CRC Press LLC.

8. Chen, J. y Chou, C. (1993). *Cane Sugar Handbook, a manual for cane sugar manufacturers and their chemist*. Estados Unidos: John Wiley & Sons, Inc.
9. Contreras, J., Zúñiga, C., Martínez, J. y Sánchez, D. (2016). Análisis de series de tiempo en el pronóstico de la demanda de almacenamiento de productos perecederos. *Estudios Gerenciales*, 32(141), 387-396.
10. Delgadillo-Ruiz, O., Ramírez-Moreno, P., Leos-Rodríguez, J., Salas González, J. y Valdez-Cepeda, R. (2016). Pronósticos y series de tiempo de rendimientos de granos básicos en México. *Acta Universitaria*, 26(3), 23-32.
11. Dettling, M. (2015). *Applied Time Series Analysis*. Zúrich, Suiza: Zurich University of Applied Sciences. Recuperado de <https://ethz.ch/de.html>.
12. Garcés, D. y Castrillón, O. (2017). Diseño de una técnica inteligente para identificar y reducir los tiempos muertos en un sistema de producción. *Información tecnológica*, 28(3), 157-170.
13. Gutiérrez, H. y Salazar, R. (2013). *Control estadístico de la calidad y Seis Sigma*. Ciudad de México, México: McGraw Hill.
14. Han, J., Kamber M., Pei J. (2012). *Data Mining, Concepts and Techniques*. Estados Unidos: Morgan Kaufman Editor.
15. Hyndman, R.J., & Athanasopoulos, G. (2021). *Forecasting: principles and practice*. Tercera edición. OTexts: Melbourne, Australia.

16. Koch, R. (2013). *The 80/20 Principle and other power laws of nature*. Boston, Estados Unidos: Nicholas Brealey Publishing.
17. Melgar, M., Meneses, A., Orozco, H., Pérez, O. y Espinoza, R. (2014). *El cultivo de la caña de azúcar en Guatemala*. Guatemala: Librerías Artemis Edinter, S.A.
18. Milton, S. (2007). *Estadística para biología y ciencias de la salud*. Madrid, España: Editorial McGraw-Hill Interamericana.
19. Newbold, P., Carlson, W. y Thorne, B. (2008). *Estadística para la administración y la economía*. Madrid, España: Editorial Pearson.
20. Otero, J. (1993). *Econometría, series temporales y predicción*. Madrid, España: Editorial AC.
21. Peña, D. (2010). *Análisis de series temporales*. Madrid, España: Alianza Editorial.
22. Pérez, I., Navarro, H. y Garrido, N. (2014). Influencia del tiempo perdido industrial sobre la economía de los ingenios azucareros. Sobre los derivados de la caña de azúcar, 48(2), 56-59.
23. Pérez M. (2015). *Business Intelligence: Técnicas, herramientas y aplicaciones*. España: Madrid. Editorial: RC Libros.
24. Rein, P. (2012). *Ingeniería de la caña de azúcar*. Alemania, Berlín: Editorial Bartens.

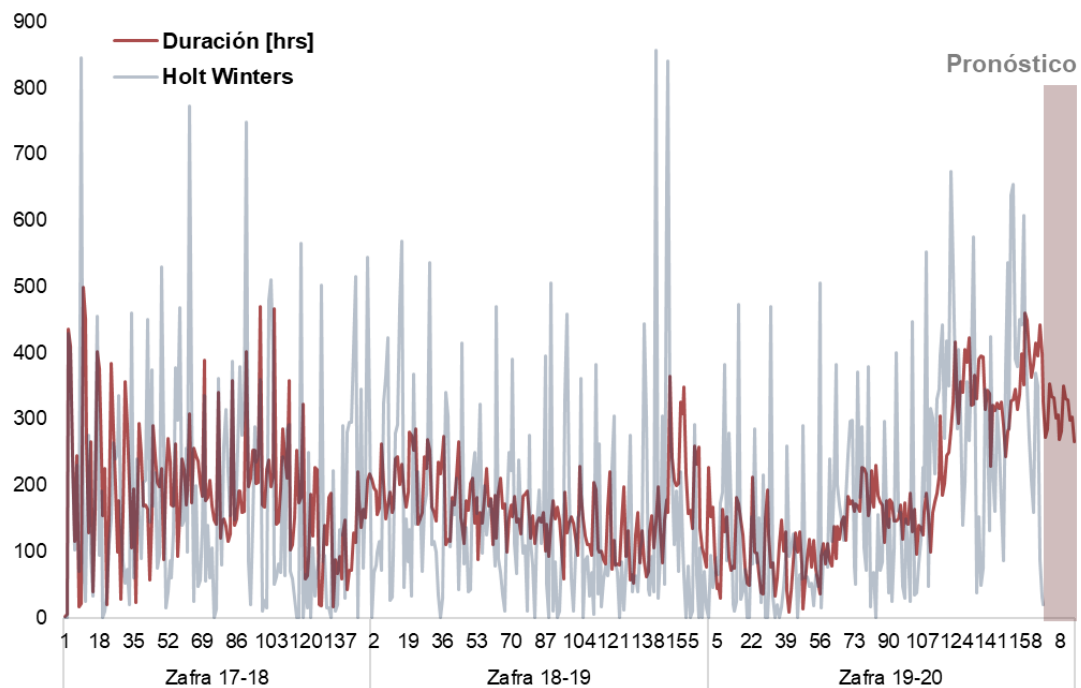
25. Reyes, J. y García, F. (2005). *Descubrimiento de patrones de desempeño académico con árboles de decisión en las competencias genéricas de la formación profesional (tesis de doctorado)*. Bogotá, Colombia: Ediciones Universidad Cooperativa de Colombia. Recuperado de: <https://ediciones.ucc.edu.co/>.
26. Ribas, M., Consuegra, R. y Alfonso, M. (2015). Análisis de los factores que más inciden sobre el rendimiento industrial azucarero. *Centro azúcar*, 43(1), 51-60.
27. Ruiz, J., Hernández, G. y Zulueta, R. (2010). Análisis de series de tiempo en el pronóstico de la producción de la caña de azúcar. *Terra Latinoamericana* 29(1), 103-109.
28. Rius, F., Barón, J., Parras, L. y Sánchez, E. (2005). *Bioestadística: Métodos y aplicaciones*. Málaga, España: Editorial Paraninfo.
29. Santana, J. (2006). Predicción de series temporales con redes neuronales: una aplicación a la inflación colombiana. *Revista colombiana de estadística*, 29(1), 77-92.
30. Siegel, S. y Castellán, N. (1995). *Estadística no paramétrica aplicada a ciencias de la conducta*. Ciudad de México, México: Editorial Trillas.
31. Shumway, R. y Stoffer, D. (2011). *Time Series Analysis and Its Applications*. Nueva York: Estados Unidos: Editorial Springer.
32. Tsay, R. (2002). *Analysis of Financial Time Series*. Estados Unidos: Editorial John Wiley & Sons, Inc.

33. Universidad Internacional de La Rioja. (2019). *Métodos de captura de la información: Tema 1*. España: Autor.
34. Walpole, R., Myers, R., Myers, S. y Ye, K. (2012). *Probabilidad y estadística para ingeniería y ciencias*. México, Ciudad de México: Pearson Educación de México, S.A. de C.V.
35. Wayne, D. (2012). *Bioestadística: Base para el análisis de las ciencias de la salud*. México: Editorial Limusa S.A.
36. Webster, A. (2000). *Estadística aplicada a los negocios*. Bogotá, Colombia: McGraw-Hill Interamericana, S.A.
37. Wongkoon, S., Pollar, M., Jaroensutasinee, M. y Jaroensutasinee, K. (2016). Predicting DHF Incidence in Northern Thailand using Time Series Analysis Technique. *International Journal of Biological and Medical Sciences*, 4(3), 117-123.

APÉNDICES

Para visualizar el ajuste del modelo obtenido mediante Holt Winters, la siguiente gráfica muestra la serie temporal de la duración de los paros de molienda, indicando el ajuste del modelo y su respectivo pronóstico para los primeros 15 días de zafra del período 20-21.

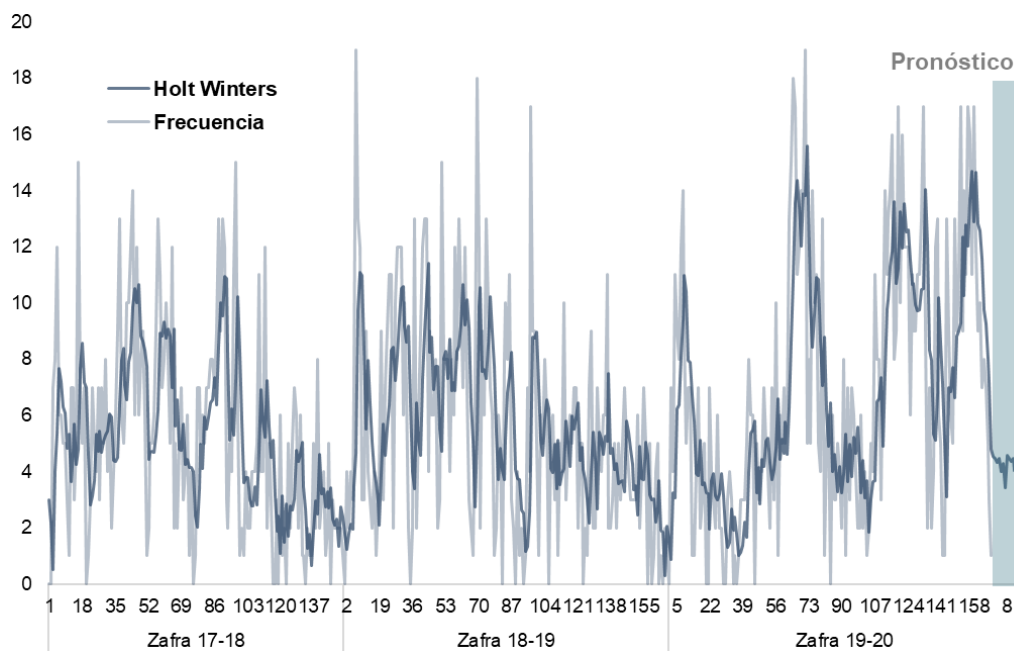
Apéndice 1. **Predicción de la duración de los tiempos perdidos (Holt-Winters)**



Fuente: elaboración propia, realizado con Microsoft Excel 2019.

También la siguiente gráfica muestra el ajuste del modelo con sus respectivos pronósticos para la frecuencia de cada paro de molienda.

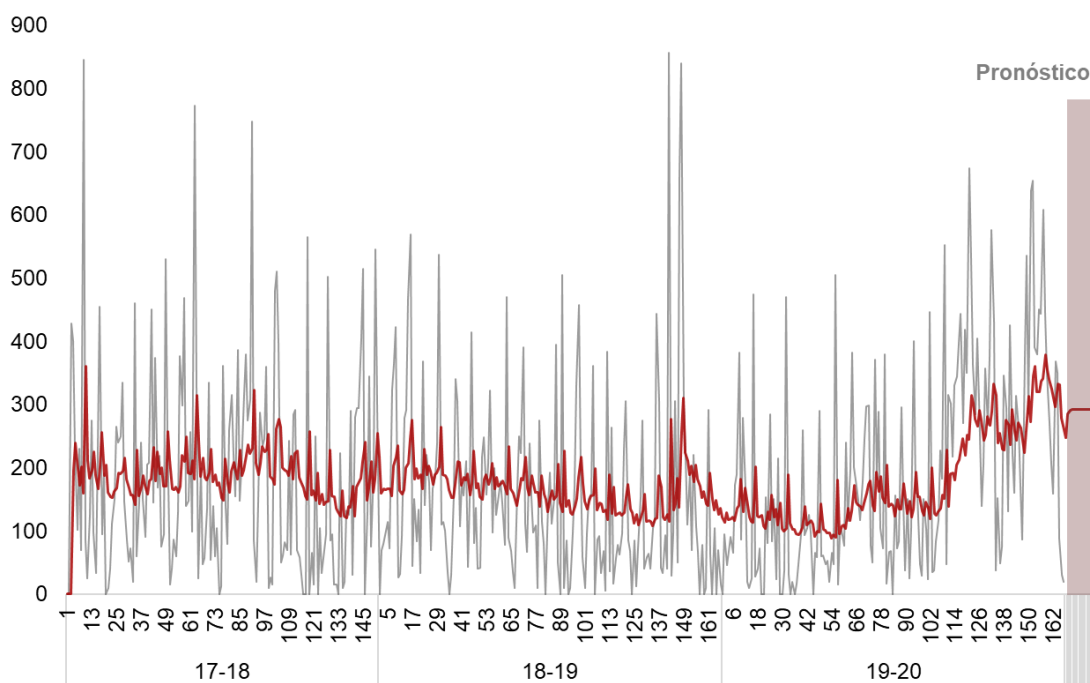
Apéndice 2. Predicción de la frecuencia de los tiempos perdidos (Holt-Winters)



Fuente: elaboración propia, realizado con Microsoft Excel 2019.

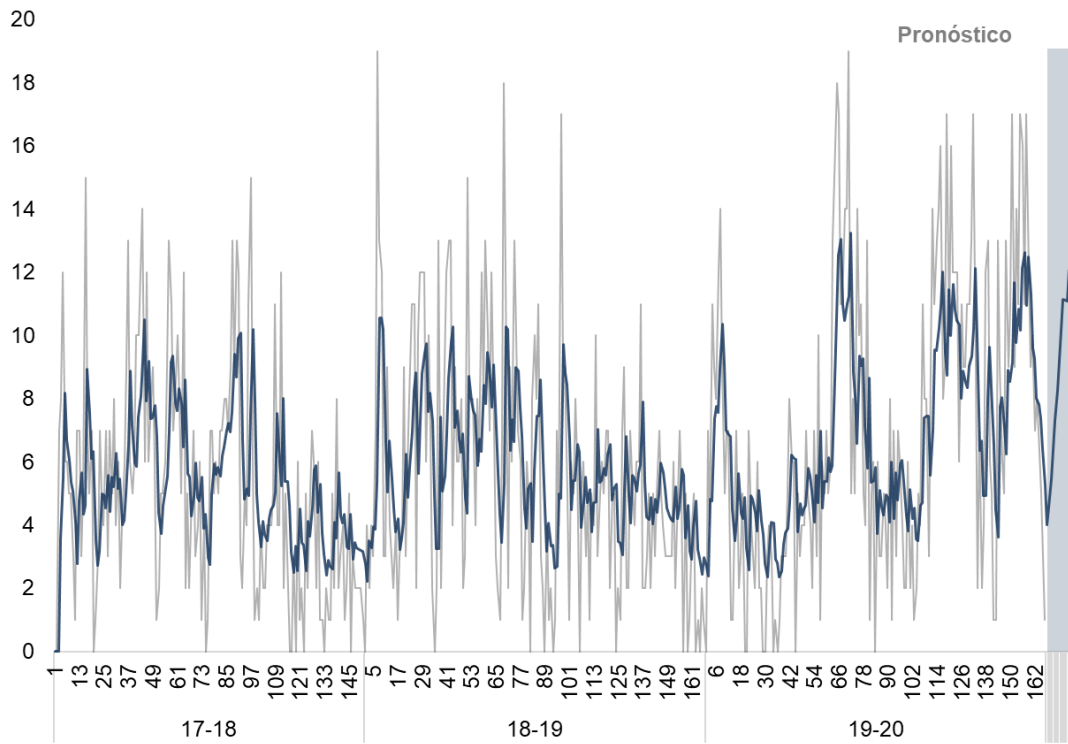
Por último, se muestran las gráficas y los ajustes del modelo obtenido mediante el método ARIMA (3,1,1) para la duración y para la frecuencia (1,1,2).

Apéndice 3. Ajuste ARIMA (3,1,1) para la duración



Fuente: elaboración propia, realizado con Microsoft Excel 2019.

Apéndice 4. Ajuste ARIMA (1,1,2) para la frecuencia

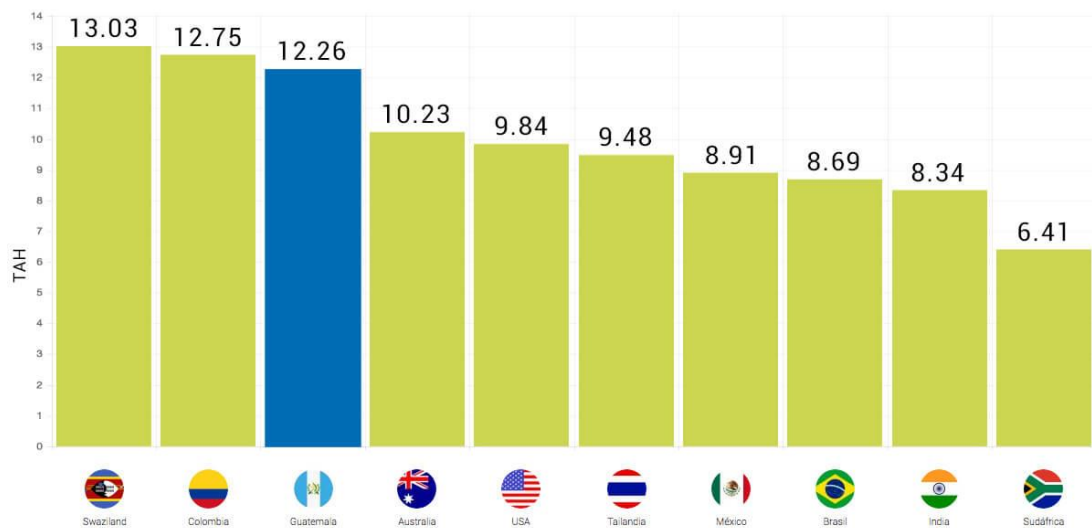


Fuente: elaboración propia, realizado con Microsoft Excel 2019.

ANEXOS

Guatemala es uno de los principales países productores de azúcar a nivel mundial, en la siguiente gráfica se muestra que Guatemala es el tercer lugar en función de las toneladas por hectárea producidos, la siguiente gráfica indica los países con mayor producción por área.

Anexo 1. Productividad a nivel mundial en la producción de azúcar



Fuente: Asociación de Azucareros de Guatemala (2021). *Productividad de toneladas de azúcar por hectárea*. Consultado el 02 de julio de 2021. Recuperado de: <https://www.azucar.com.gt/azucar-de-guatemala-en-el-mundo>.