

# Análisis De Series de Tiempo Para La Estimación Del Consumo Combustible En Aeronaves ATR42 Para Aerolínea Colombiana

1<sup>st</sup> Edwin Alexander Rativa Saenz  
Fundación Universitaria Los Libertadores  
Bogotá, Colombia  
Email: earativas@libertadores.edu.co

**Resumen**—La gestión del combustible de las aerolíneas es una tarea compleja, ya que en ella se ven inmersa distintas variables involucradas que dificultan su gestión y así mismo la adquisición en términos de cantidades y dinero puede convertirse en un problema. Este artículo realiza un análisis de series de tiempo de datos históricos de consumo de combustible para el avión ATR42, que describe y predice el consumo de combustible, lo que lo convierte en una nueva herramienta útil para resolver este problema. Los datos consisten en 730 días de registros de consumo de combustible Jet A1 (en kilogramos) y tiempo de vuelo (en horas). Utilizando los modelos ARIMA, SARIMA y SARIMAX, se proponen diferentes soluciones para explicar el comportamiento de esta variable en el tiempo. Como factores de aceptación se toma el RMSE (Root Mean Square Error) y MAPE (Mean Mean Absolute Percentage Error) y se da la descripción de aplicación a los datos. Este análisis puede ser usado como punto de partida para conocer el costo del combustible para la compañía en términos de optimización, así como herramienta de control y auditoria de combustible, ya que el modelo puede ajustarse a distintas variables que pueden describir su comportamiento.

**Index Terms**—Series de Tiempo, ARIMA, SARIMA, SARIMAX, Gestión del Combustible.

## I. INTRODUCCIÓN

El Jet A1 Aviation Turbine Fuel (ATF) es el combustible empleado para el funcionamiento y despegue de las aeronaves, obviamente siendo este un elemento esencial, su adquisición es un factor tan importante para determinar la rentabilidad de las aerolíneas en términos económicos. En promedio, el combustible representa entre un 22 % y un 38 % de los gastos de una aerolínea y representa el primer o segundo gasto más importante. Por lo anterior la gestión de combustibles es muy importante, y aunque dependa de diversos factores tales como, el estado del motor, el paso de la hélice/rotor, las RPM de la hélice, la concentración de la mezcla, el porcentaje de potencia utilizada para el vuelo, el viento, etc, se deben generar soluciones aplicables. La escasez de combustible causada por una planificación inadecuada previa a los vuelos, la confirmación o no de que el combustible a bordo de la aeronave se verificó minuciosamente antes del despegue o que el consumo real no se controló de manera efectiva en comparación con lo planificado, son tareas que dificultan la operación. Este trabajo proporciona una interesante

herramienta en este esfuerzo, empleando modelos estadísticos que ocupan series de tiempo, donde gracias a datos históricos o Big Data se genere un control, contribuyendo a la economía de la empresa, la seguridad operacional, la adecuada planificación y la mejora a la calidad del servicio.

## II. OBJETIVO GENERAL

Ajustar modelos de series de tiempo para el pronóstico del consumo de combustible de las aeronaves ATR42 de una aerolínea colombiana.

## III. MATERIALES Y MÉTODOS

La base de datos corresponde a un archivo que contiene el registro día por día desde el 01 de enero del 2018 hasta el 31 de diciembre del 2019 (730 datos), de una aerolínea colombiana donde se registran todos los vuelos realizados por las aeronaves ATR42 (ya que poseen varias), estos aviones realizan varios recorridos en el día por lo cual se decide cuantificar el total de combustible y tiempo de horas de vuelo por día. Esta aeronave es especial por encima de otras usada por la empresa ya que son las más empleadas, la cantidad de pasajeros es mayor que otras (48 pasajeros). Se emplea la teoría entorno a series de tiempo empleando el modelo de predicción ARIMA, SARIMA y SARIMAX incluyendo dos variables exógenas (dummy y otra serie temporal).

La metodología Box - Jenkins aplicada en esta investigación se divide en las siguientes fases para cada método usado [2]:

- Análisis exploratorio de datos.
- Si es necesario, utilizar una transformación de Box-Cox para estabilizar la varianza.
- Si es necesario, diferenciar los datos hasta que sean estacionarios. Utilizando pruebas de raíz unitaria.
- Trace el ACF/PACF de los datos diferenciados y determine posibles modelos candidatos.
- Pruebe los modelos elegidos y utilice el AICC, RMSE y MAPE para buscar un modelo mejor [6].
- Verifique los residuos de su modelo elegido trazando el ACF de los residuos.
- ¿Los residuos parecen ruido blanco?.
- Escriba el modelo.
- Calcular pronósticos.

#### IV. ANÁLISIS EXPLORATORIO DE DATOS

Los datos suministrados por una aerolínea de transporte de pasajeros corresponden al periodo de tiempo del 01 de enero del 2018 hasta el 31 de diciembre del 2019 (con formato fecha) la cantidad total de combustible medido en Kilogramos (kg – variable cuantitativa) empleado por las aeronaves ATR42 que poseen en su flota, como también las horas de vuelo totales de estas aeronaves medido en horas (variable cuantitativa). Es una tabla con extensión de 730X4, los datos son diarios y no tienen valores faltantes, como se observa en la tabla I.

	Fecha	Combustible (kg)	Tvuelo (h)
1	1/01/2018	5298	8,39
2	2/01/2018	27157	41,74
3	3/01/2018	27327	41,02
⋮	⋮	⋮	⋮
730	31/12/2019	20155	30,28

Tabla I

Los datos son analizados de manera diaria, estableciendo ciclos de 7 días (semanal), se procede a graficar la serie de tiempo de interés “Combustible”. Se observa en la figura 1 que el uso del combustible es estacional en el tiempo, con media constante, varianza constante en la mayoría de los datos y nula tendencia de crecimiento o decrecimiento, presenta algunos picos de valores mínimos (0 kg) y valores máximos en torno a 40.000 kg, por ello se podría caracterizar la serie como un modelo estacional.

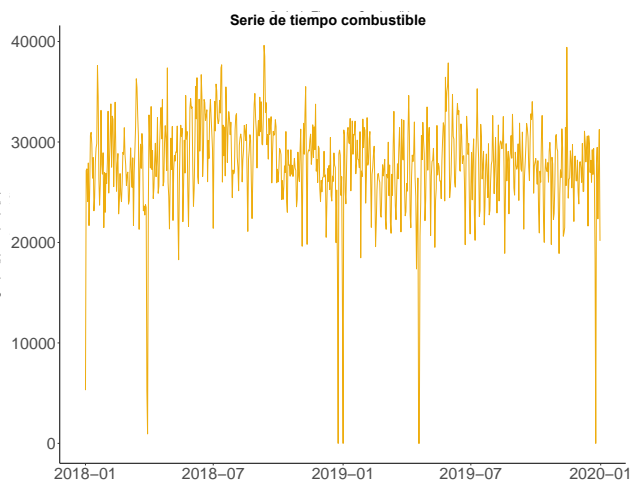


Figura 1. Serie de tiempo del consumo de combustible

En el seasonal plot de la figura 2 se muestra la tendencia del consumo de combustible para cada día de la semana, en la cual se observa que el consumo de promedio de combustible

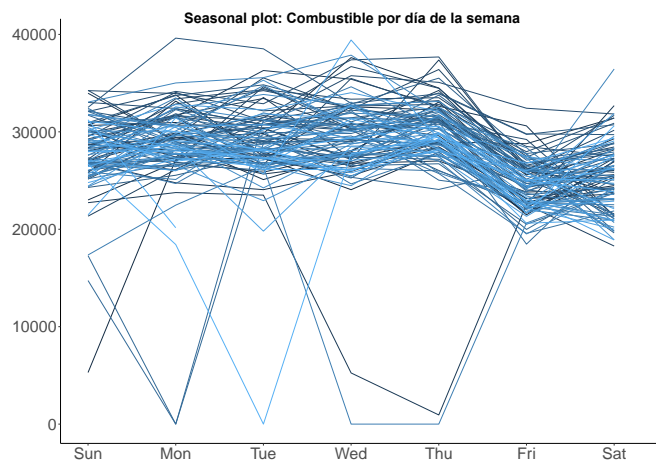


Figura 2. Combustible por días de la semana

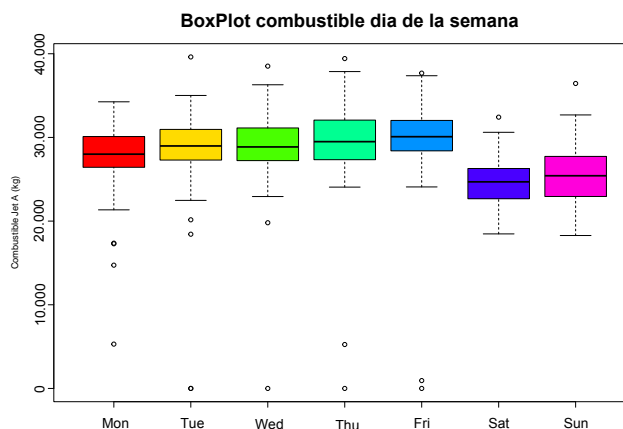


Figura 3. Boxplot combustible por días de la semana

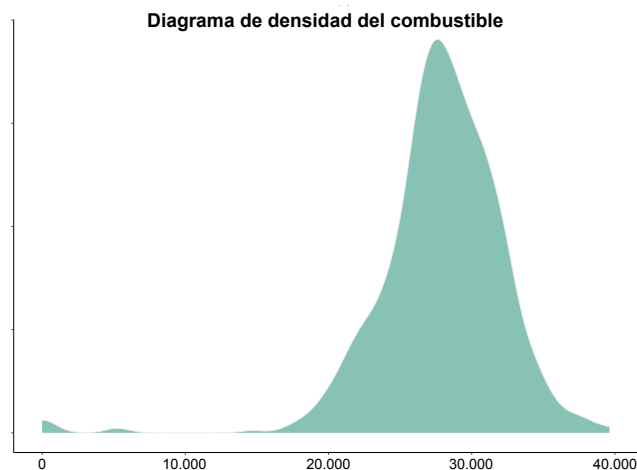


Figura 4. Diagrama de densidad del combustible

ronda en los rangos 25.000 kg a 35.000 kg de lunes a viernes, también se observa que los sábados y domingos el consumo disminuye respecto a los demás días, en algunos días feriados el valor disminuye a 0 kg (en 5 fechas concretas). El diagrama de la Figura 3 agrupa en un boxplot la información, donde se observa de forma más clara la ubicación de los cuartiles y medianas de los datos, así como los puntos de las colas más cercanos al 0. Este comportamiento de los datos simplifica el manejo de estos ya que la serie no requiere algún tipo de transformación al no presentar tendencias. Por último, se realiza un diagrama de densidades figura 4 el cual presenta un elevado grado de concentración alrededor de los valores centrales de la variable, dada su apariencia leptocúrtica.

## V. ESTACIONARIEDAD EN LOS DATOS

Se proponen 4 diferentes pruebas que permitan esclarecer y tomar la decisión si los datos requieren ser diferenciados a la hora de estimar modelos predictores, estas pruebas son:

### 1. Prueba de Dickey-Fuller aumentada:

El p - valor para la serie combustibles es de 0,005306, la estadística de prueba es menor que el nivel de significancia al 0,01, la decisión es rechazar la hipótesis nula. Debido a que los datos proporcionan evidencia de que los datos son estacionarios, la recomendación del análisis es proceder sin diferencias.

### 2. Valor de la desviación estándar:

Se realiza el cálculo de las desviaciones estándar, para los datos con 0, 1 y 2 diferencias, como se observa en la tabla II, verificando los resultados. Dicho proceso arroja que el menor valor de la SD es sin diferenciar los datos, la recomendación del análisis es proceder sin diferencias [7].

Código	Valor
sd(Combustible)	4624.653*
sd(diff(Combustible,differences=1))	4919.036
sd(diff(Combustible,differences=2))	7596.934

Tabla II

### 3. Test de diferencias paquete Forecast en R- Studio:

En la tabla III se observa los resultados del software R Studio aplicando su prueba interna de raíz unitaria para verificación de diferenciación tanto en los componentes autorregresivos y estacionales de un modelo SARIMA. Las pruebas anteriormente realizadas demuestran que, si bien la serie original es estacionaria, La última prueba realizada sugiere diferenciar 1 vez el componente ordinario y 0 el componente estacional.

### 4. Interpretación de la FAC y FACP:

La FAC muestra un decaimiento lento en los múltiplos de la estación, contemplando un posible grado de

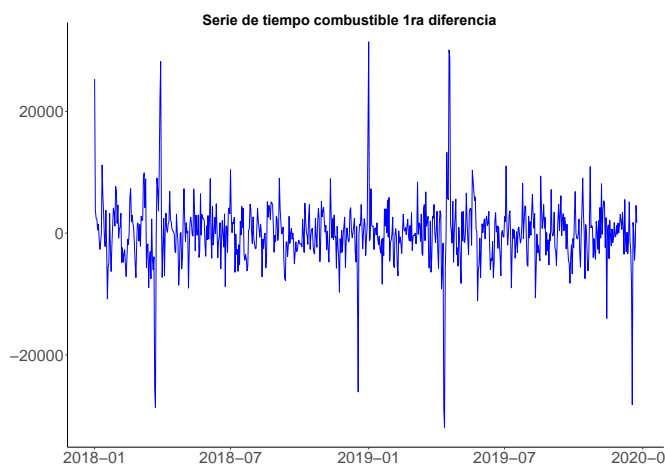


Figura 5. Serie de tiempo de combustible diferenciada 1 vez.

asociación lineal y existentes observaciones sucesivas a dichos rezagos, esto sugiere también el uso de un modelo SARIMA (sin descartar otros modelos). De la FAC y FACP resaltan rezagos  $k = 1 - 3, 6, 7, 15$ .

Método	Código	Valor	Interpretación
Número de diferencias requeridas para una serie estacionaria	ndiffs	1	R, sugiere usar 1 diferencia en el componente autorregresivo.
Número de diferencias necesarias para una serie estacionaria estacional.	nsdiffs	0	R, sugiere usar 0 diferencias en el componente estacional.

Tabla III

¿La serie requiere ser diferenciada?		
Prueba	SI	NO
Prueba de Dickey-Fuller aumentada		✓
Menor valor desviación estándar		✓
Test de diferencias en R	✓	
Interpretación de la FAC y FACP	✓	

Tabla IV

Las pruebas realizadas no son concluyentes tal como se observa en la tabla de decisión IV, se decide diferenciar la serie y realizar los correlogramas buscando interpretar y determinar el orden del mejor modelo.

Los correlogramas de la primera diferencia, describen un comportamiento estacional, con decaimiento rápido en la FAC y FACP, y una importante característica en el rezago  $k = 3, , 6, 7, 8, 15$  donde siempre fue más significativo.

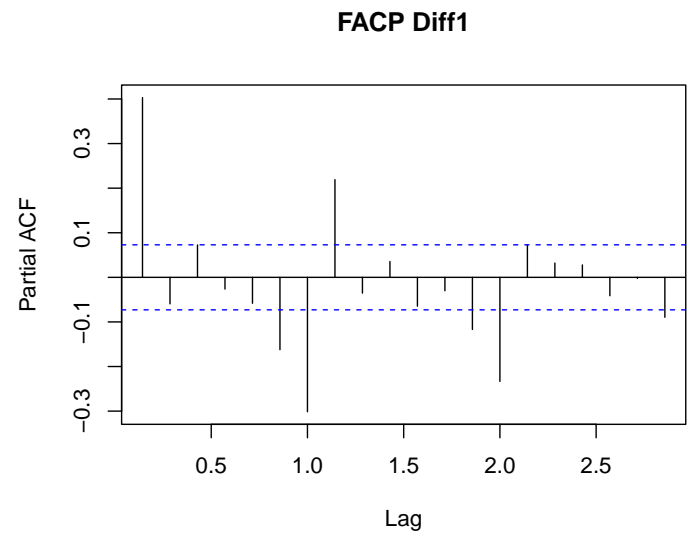
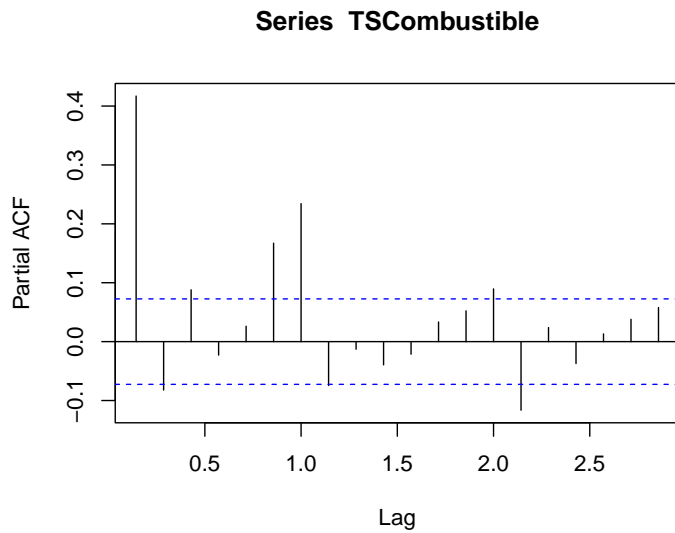
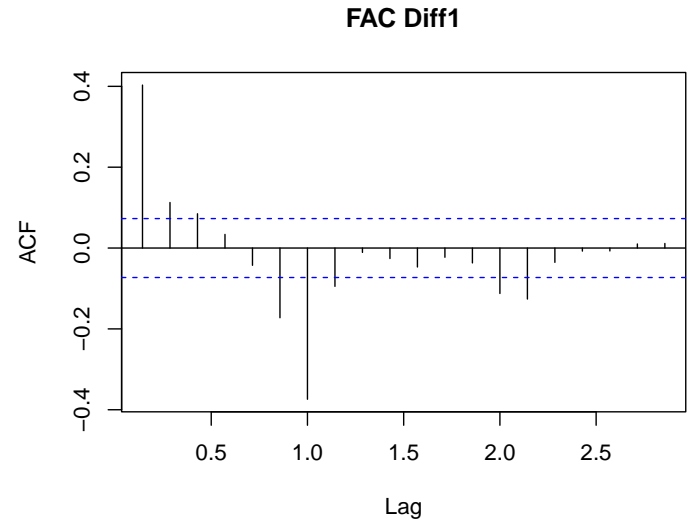
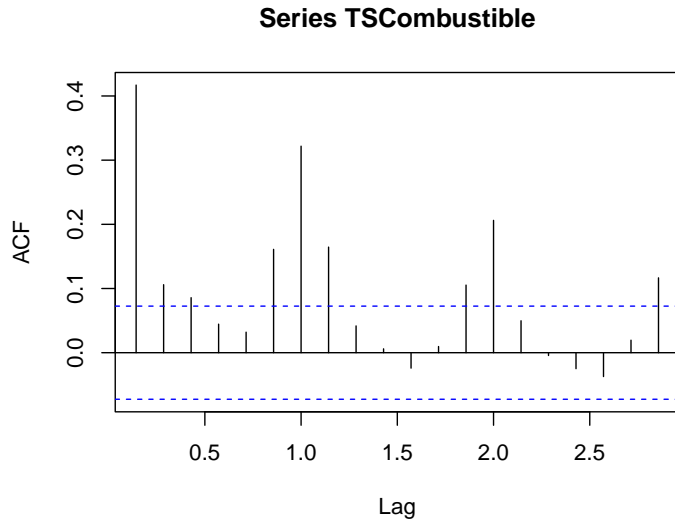


Figura 6. Función de autocorrelación y Función de autocorrelación parcial serie combustible

Figura 7. Función de autocorrelación y Función de autocorrelación parcial serie combustible 1ra diferencia

Al diferenciar la serie y graficar su comportamiento se observa en la figura 5 el comportamiento estacionario, la serie por ende ya puede ser empleada para identificar y proponer modelos.

## VI. MODELO 1: ARIMA

### VI-A. Propuestas de modelos:

El modelo ARIMA (p,d,q) es un modelo de promedia móvil autorregresivo (ARMA) integrado d veces [1]. Se construyen en análisis univariados de series de tiempo. A continuación, se presentan 3 modelos aplicando esta técnica, todos ellos programados en el software R Studio acorde a los valores de la FAC y FACP, así como el uso de un análisis exploratorio

iterando valores. Se presentan las siguientes propuestas dadas sus cualidades predictoras, tomando como referencia los menores valores del RMSE. El modelo que mejor expreso capacidad de predicción es el modelo  $ARIMA(15,1,8)$  dado que su valor de RMSE fue el menor así como el AIC (Akaike Information Criterion) como se observa en la tabla V, con 15 rezagos en la parte autorregresiva, una diferencia ordinaria y 8 parámetros ponderando a los 8 choques aleatorios inmediatamente anteriores.

### VI-B. Graficar el modelo seleccionado:

El modelo  $ARIMA(15,1,8)$  de la figura 8 se comporta de manera similar a la serie temporal original, no se ajusta

Ítem	Modelo	RMSE (kg)	AIC
1	$ARIMA(10, 1, 11)$	3731.139	14118.28
2	$ARIMA(11, 1, 13)$	3730.482	14124.39
3	$ARIMA(15, 1, 8)$	<b>3708.218</b>	<b>14117.57</b>

Tabla V

a los valores extremos donde pierde previsibilidad, en el año 2019 logra la mejor aproximación a los datos originales en comparación del 2018. Es un modelo práctico y sencillo no requiere demasiadas iteraciones y su poder de predicción no se aleja a modelos más elaborados que se presentan en este documento.

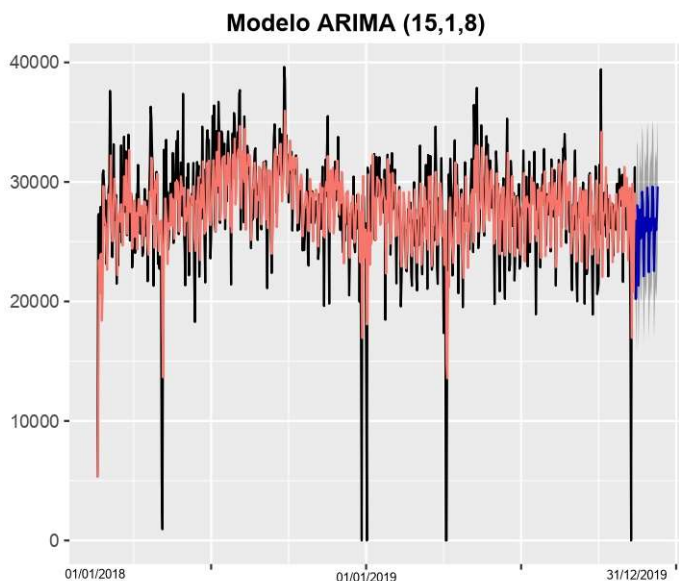


Figura 8. Modelo  $ARIMA(15, 1, 8)$

#### VI-C. Análisis de los residuales:

El análisis de los residuos para el modelo  $ARIMA(15, 1, 8)$  se evidencia que la media gira en torno a 0, además de presentar normalidad, en el FAC se no se observa rezagos fuera de la banda de confianza, los residuos no tienen correlaciones significativas por ende el modelo captura toda la información pasada y es tomada en cuenta. En la línea de tiempo se observa distintos picos que toman valores muy extremos, es por ello que el modelo no es prolijo y debe mejorarse. De la tabla VI se realizan las pruebas de normalidad de residuos donde el modelo cumple con el supuesto rechazando la hipótesis nula cumpliendo con los requerimientos del modelo. En la prueba Ljung-Box el p valor para 26 lags es inferior a 0.05 esto significa que los residuos de nuestro modelo de serie temporal no son independientes. En la figura 11 se muestra la dispersión de datos de los valores reales y la predicción, su ajuste al  $R^2$  es muy bajo siendo este 0.3599.

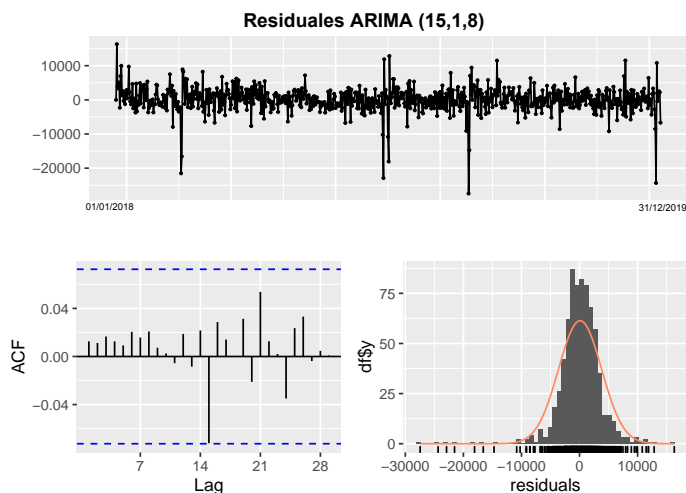


Figura 9. Residuos modelo  $ARIMA(15, 1, 8)$

	Test Residuales	p-value	Interpetación
1	Ljung-Box	0,00632@k=26	Rechaza la hipótesis nula
2	Box-Pierce test	0.9814	Rechaza la hipótesis nula
3	Shapiro-Wilk	<2.2e-16	Rechaza la hipótesis nula
4	T Student	0.5031	Rechaza la hipótesis nula

Tabla VI

#### VI-D. Predicción del modelo seleccionado:

Se realiza la predicción de los primeros 30 días consecuentes al final de los datos originales figura 10, el modelo, en azul se observa el valor con el correspondiente sombreado del error del valor, los valores tienen a comportarse de manera estacionaria y cíclica entorno a la mediana.

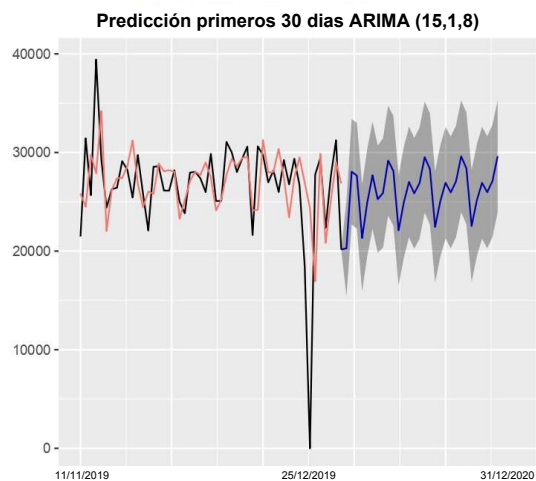


Figura 10. Predicción modelo  $ARIMA(15, 1, 8)$



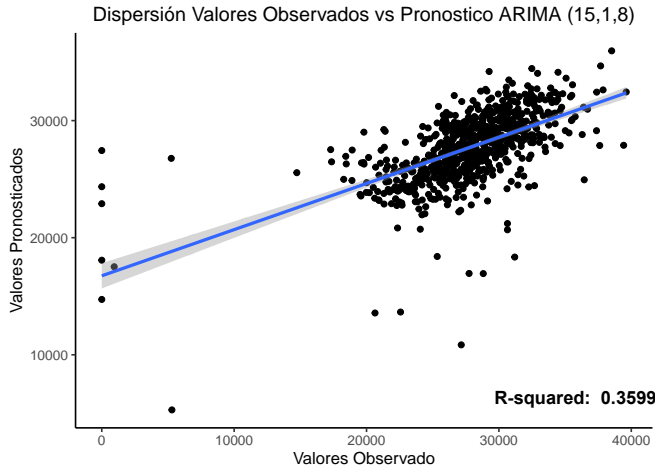


Figura 11. Dispersión  $ARIMA(15,1,8)$

#### VI-E. Escribir el modelo:

El modelo ARIMA seleccionado se escribe de la siguiente manera 1 :

$$ARIMA(15,1,8)$$

$$(1 - \alpha_1 B - \dots - \alpha_{15} B^{15})(1 - B)y_t = (1 - \phi_1 B - \dots - \phi_8 B^8)\varepsilon, \varepsilon \sim RB(0, \sigma_a^2)$$

$$\begin{aligned} \alpha_1 &= -0,473, \alpha_2 = -0,601, \alpha_3 = -0,498, \alpha_4 = -0,491 \\ \alpha_5 &= -0,520, \alpha_6 = -0,517, \alpha_7 = 0,631, \alpha_8 = 0,063 \\ \alpha_9 &= 0,210, \alpha_{10} = 0,104, \alpha_{11} = 0,097, \alpha_{12} = 0,126 \\ \alpha_{13} &= 0,124, \alpha_{14} = -0,026, \alpha_{15} = 0,015 \end{aligned}$$

$$\begin{aligned} \phi_1 &= -0,061, \phi_2 = 0,0084, \phi_3 = 0,0309, \phi_4 = 0,0061 \\ \phi_5 &= 0,0113, \phi_6 = 0,0275, \phi_7 = -0,9788, \phi_8 = 0,0607 \end{aligned} \quad (1)$$

### VII. MODELO 2: SARIMA

#### VII-A. Propuestas de modelos:

Se desarrolla la elaboración del modelo de media móvil integrada autorregresiva estacional SARIMA  $(p, d, q)(P, D, Q)_s$  que tenga la mejor capacidad de acoplarse a los datos [3]. El modelo SARIMA es una extensión del modelo ARIMA que incluye un componente estacional y una interpretación donde es probable que algo suceda a una cadencia regular y que impacte nuestra variable objetivo de manera similar cada vez que ocurra [5]. Es por ello que se decide incluir esta versión del ARIMA, dado que los datos presentan cierta estacionalidad semanal que puede ayudar a comprender los datos y por ende a modelar un mejor predictor. Se proponen los 3 modelos más adecuados gracias al análisis exploratorio de datos la FAC y FACP y la comparación del RMSE para cada uno, tal como se observa en la tabla VII.

ítem	Modelo	RMSE (kg)	AIC
1	$SARIMA(2, 0, 6)(2, 1, 3)_7$	3691.161	14104.47
2	$SARIMA(2, 0, 6)(2, 1, 5)_7$	3687.471	13990.4
3	$SARIMA(6, 1, 4)(6, 0, 3)_7$	<b>3683.658</b>	14104.47

Tabla VII

#### VII-B. Gráficar el modelo seleccionado:

Se realiza el gráfico del modelo SARIMA seleccionado  $SARIMA(6, 1, 4)(6, 0, 3)_7$ , su comportamiento es adecuado al compararlos con los datos originales, valores cercanos a la media de los datos originales, pierde potencial de predicción con los valores más extremos, aunque modela muy bien el primer valor extremo del 2018; es un modelo que toma en cuenta sus valores pasados cada 7 días. El modelo autorregresivo  $(p, d, q)$ ,  $p$  (6) del modelo toma sus valores pasados 6 periodos de tiempo atrás, el promedio móvil  $q$  (4) toma los errores de predicción retrasados 4 periodos de tiempo atrás y emplea  $d$  (1) diferencia. Para el componente estacional  $(P, D, Q)$  este modelo toma el orden autorregresivo estacional en  $P$  (6), el orden promedio móvil estacional  $Q$  (3) y no aplica orden de diferencia estacional y por último se toman  $s$  7 pasos de tiempo para un solo período estacional.

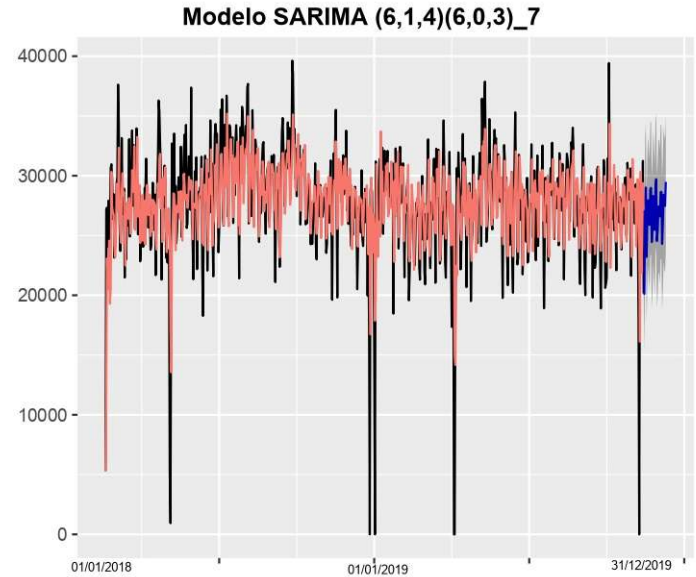


Figura 12. Modelo  $SARIMA(6, 1, 4)(6, 0, 3)_7$

#### VII-C. Análisis de los residuales:

Se encuentra ahora el análisis de los residuales para el modelo  $SARIMA(6, 1, 4)(6, 0, 3)_7$  se evidencia varianza y media constante en torno a 0, además de presentar normalidad, en el FAC se no se observa rezagos fuera de la banda de confianza, los residuales no tienen correlaciones significativas por ende el modelo captura toda la información pasada y es tomada en cuenta. En la línea de tiempo de valores residuales

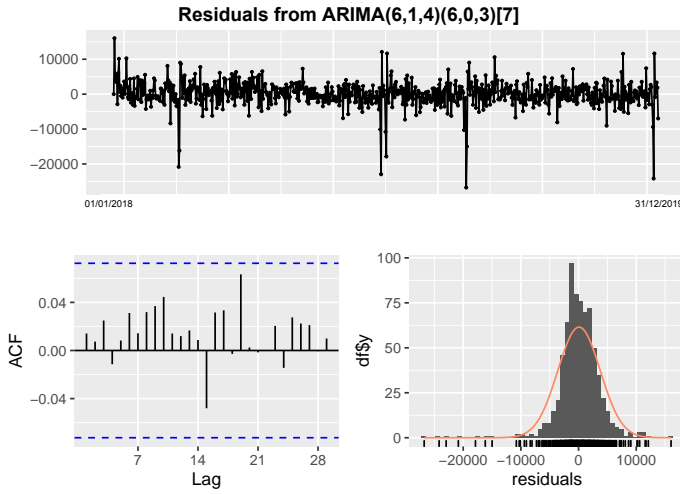


Figura 13. Residuos modelo  $SARIMA(6, 1, 4)(6, 0, 3)_7$

observados en la figura 13 se distinguen distintos picos que toman valores muy extremos, es por ello que el modelo es prolijo a pesar de ser el más sobresaliente de su categoría. De la tabla VIII se realizan las pruebas de normalidad de residuos. En la figura 14 se muestra la dispersión de datos de los valores reales y la predicción, su ajuste al  $R^2$  es muy bajo siendo este 0.3673.

	Test Residuales	p-value	Interpretación
1	Ljung-Box	0.008308	Rechaza la hipótesis nula
2	Box-Pierce test	0.957	Acepta la hipótesis nula
3	Shapiro-Wilk	<2.2e-16	Rechaza la hipótesis nula
4	T Student	0.5405	Acepta la hipótesis nula

Tabla VIII

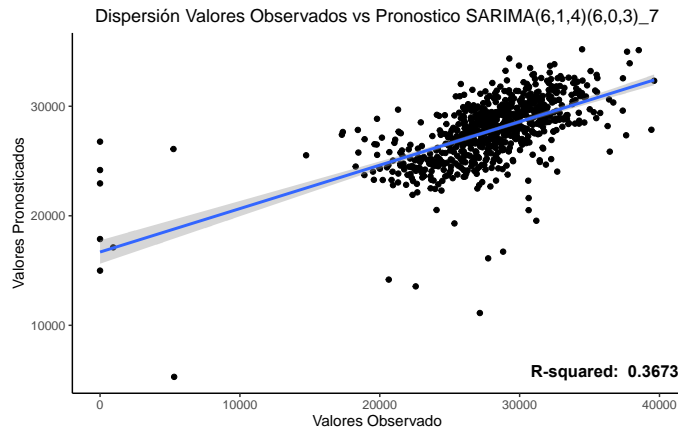


Figura 14. Dispersión  $SARIMA(6, 1, 4)(6, 0, 3)_7$

#### VII-D. Predicción del modelo seleccionado:

En la figura 15 se ve la predicción del modelo  $SARIMA(6, 1, 4)(6, 0, 3)_7$ , empleando los primeros 30 días después del dato final de la serie original, la línea azul es el valor que tomara el consumo de combustible, donde se ve que no toma una tendencia estacional como si ocurriera en el modelo ARIMA figura 10, este por el contrario toma diferentes valores pero acordes a los valores medios de su pasado, tampoco se observan tendencias que no estarían acorde a los datos.

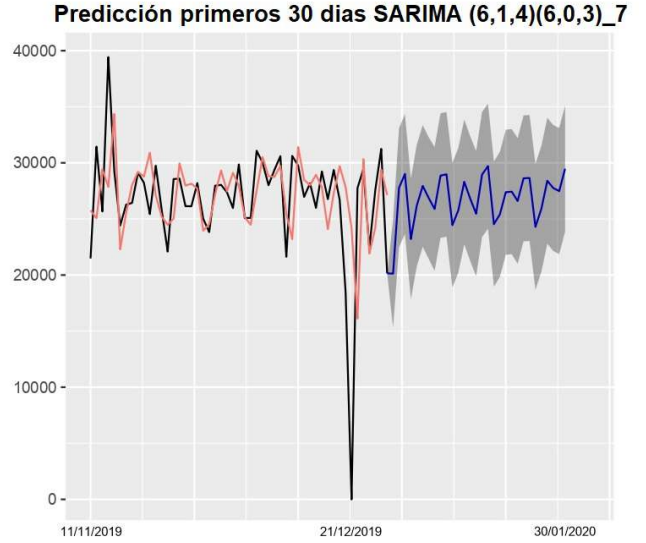


Figura 15. Predicción modelo  $SARIMA(6, 1, 4)(6, 0, 3)_7$

#### VII-E. Escribir el modelo:

El modelo  $SARIMA(6, 1, 4)(6, 0, 3)_7$  seleccionado se escribe de la siguiente manera ecuación 2:

$$SARIMA(6, 1, 4)(6, 0, 3)_7$$

$$(1 - \alpha_1 B - \dots - \alpha_6 B^6) (1 - \gamma_1 B^7 - \dots - \gamma_6 B^{42}) (1 - B) y_t = (1 - \phi_1 B - \dots - \phi_4 B^4) (1 - \theta_1 B^3 - \dots - \theta_3 B^{21}) \varepsilon, \varepsilon \sim RB(0, \sigma_a^2)$$

$$\begin{aligned} \alpha_1 &= -0,100, \alpha_2 = -0,452, \alpha_3 = -0,581, \alpha_4 = 0,4707 \\ \alpha_5 &= -0,0342, \alpha_6 = 0,1191 \\ \gamma_1 &= -0,421, \gamma_2 = 0,7131, \gamma_3 = 0,971, \gamma_4 = -0,126 \\ \gamma_5 &= -0,063, \gamma_6 = -0,077 \\ \phi_1 &= -0,434, \phi_2 = 0,070, \phi_3 = 0,313, \phi_4 = -0,949 \\ \theta_1 &= 0,053, \theta_2 = 0,031, \theta_3 = 0,0622 \end{aligned} \quad (2)$$

#### VIII. MODELO 3: SARIMAX + DUMMY

El modelo SARIMAX se diferencia del modelo SARIMA en que agrega una variable exógena o dummy que realiza la tarea explicativa de agregar información adicional al modelo,

en este caso del análisis de los datos exploratorios podemos ver que los viernes, sábados y domingos y los días festivos colombianos están asociados al consumo de combustible porque disminuye, por lo que se decidió agregar una variable que actúe como un vector de 0s y 1s, donde 1 representa esta información extra que se agrega a los datos originales, como se ve en la tabla IX.

	Fecha	Combustible (kg)	Tvuelo (h)	Festivos
1	1/01/2018	5298	8,39	1
2	2/01/2018	27157	41,74	0
3	3/01/2018	27327	41,02	0
4	4/01/2018	24061	35,56	0
5	5/01/2018	27911	42,98	0
6	6/01/2018	21680	33,38	1
7	7/01/2018	25340	39,51	1
.	.	.	.	.
.	.	.	.	.
.	.	.	.	.
730	31/12/2019	20155	30,28	0

Tabla IX

#### VIII-A. Propuestas de modelos:

Usando esta información adicional, se propusieron 3 modelos que intentaron hacer las predicciones más ajustadas a los datos históricos, usando los valores RMSE y AIC como puntos de comparación; todos los modelos se corrieron en el software R Studio. La variable “festivo” añadida, a diferencia de los modelos vistos anteriormente, mejora la predicción, reduciendo así el valor de RMSE. R sugiere el uso de modelos “SARIMAX Errors”, lo cual significa que se corre una regresión sobre la variable exógena y con los residuales se ajusta un modelo de series de tiempo. Esto viola la suposición habitual de errores independientes que se cometen en la regresión ordinaria. La consecuencia es que las estimaciones de los coeficientes y sus errores estándar serán incorrectas si se ignora la estructura de la serie temporal de los errores.

ítem	Modelo	RMSE (kg)	AIC
1	$SARIMAX(6, 1, 4)(6, 0, 3)_7$ errors	3529.476	14104.47
2	$SARIMAX(3, 0, 4)(5, 1, 3)_7$ errors	3509.913	13990.4
3	$SARIMAX(3, 0, 6)(2, 1, 3)_7$ errors	<b>3498.238</b>	14104.47

Tabla X

#### VIII-B. Gráficar el modelo seleccionado:

Se realiza el gráfico del modelo SARIMAX seleccionado  $SARIMAX(3, 0, 6)(2, 1, 3)_7$ , su comportamiento es adecuado, se comporta bien en las mediciones elocuentes a valores cercanos a la media de los datos originales ampliando el barrido de valores, gana potencial de predicción con los valores más extremos ya que la dummy logra explicarlas, pero

aun así se esperaría que ajustara mejor los valores extremos, pero no fue así. El modelo autorregresivo  $(p, d, q)$ , se describe  $p$  del modelo toma sus valores pasados 3 periodos de tiempo atrás, el promedio móvil  $q$  toma los errores de predicción retrasados 6 periodos de tiempo atrás y no emplea diferenciar la serie. Para el componente estacional  $(P, D, Q)$  este modelo toma el orden autorregresivo estacional en  $P$  (2), el orden promedio móvil estacional  $Q$  (3) y aplica orden de diferencia estacional por último se toman  $s$  7 pasos de tiempo para un solo período estacional.

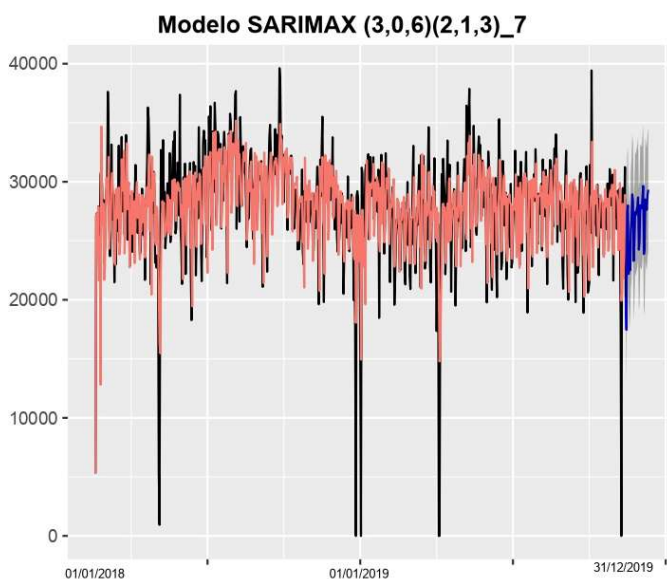


Figura 16. Modelo  $SARIMAX(3, 0, 6)(2, 1, 3)_7$

#### VIII-C. Análisis de los residuales:

Se encuentra ahora el análisis de los residuales para el modelo  $SARIMAX(3, 0, 6)(2, 1, 3)_7$  se evidencia variancia y media constante en torno a 0, además de presentar normalidad, en el FAC se no se observa rezagos fuera de la banda de confianza, los residuales no tienen correlaciones significativas por ende el modelo captura toda la información pasada y es tomada en cuenta. En la línea de tiempo de valores residuales observados en la figura 17 se distinguen distintos picos que toman valores extremos (disminuyen respecto a modelos anteriores), es por ello que el modelo es prolijo a pesar de ser el más sobresaliente de su categoría. De la tabla XI se realizan las pruebas de normalidad de residuos donde el modelo cumple con el supuesto rechazando la hipótesis nula cumpliendo con los requerimientos del modelo.

#### VIII-D. Predicción del modelo seleccionado:

En la figura 19 se ve la predicción del modelo seleccionado  $SARIMA(6, 1, 4)(6, 0, 3)_7$ , empleando los primeros 30 días después del dato final de la serie original, la línea azul es el valor que tomara el consumo de combustible, donde se ve que no toma una tendencia estacional, este toma diferentes valores acordes a los valores medios de su pasado. Para la predicción



se debe ingresar un vector de 0s y 1s acordes a los próximos 30 días calendario siendo la exógena (1, 0, 0, 1, 1, ..., 0, 0).

#### VIII-E. Escribir el modelo:

El modelo  $SARIMAX(3, 0, 6)(2, 1, 3)_7$  con la variable exogena Festivo con errores seleccionado se escribe de la siguiente manera 3 :

$$SARIMAX(3, 0, 6)(2, 1, 3)_7 + Festivo$$

$$y_t = \beta x_t + \eta_t$$

$$\eta_t = (1 - \alpha_1 B - \dots - \alpha_3 B^3) - (1 - \gamma_1 B^7 - \gamma_2 B^{14}) \eta_t$$

$$(1 - \phi_1 B - \dots - \phi_6 B^6) - (1 - \theta_1 B^3 - \dots - \theta_3 B^{21}) \varepsilon,$$

$$\varepsilon \sim RB(0, \sigma_a^2)$$

$$\beta = -5820,9195, \alpha_1 = 0,078, \alpha_2 = -0,101, \alpha_3 = 0,856$$

$$\gamma_1 = -0,184, \gamma_2 = 0,713$$

$$\phi_1 = 0,291, \phi_2 = 0,1994, \phi_3 = -0,666, \phi_4 = -0,195$$

$$\phi_5 = -0,031, \phi_6 = -0,078$$

$$\theta_1 = -0,747, \theta_2 = -0,959, \theta_3 = 0,707$$

(3)

Donde  $\beta$  es la constante que suministra la variable Festivo.

#### Predicción primeros 30 días SARIMAX (3,0,6)(2,1,3)\_7 + Festivo

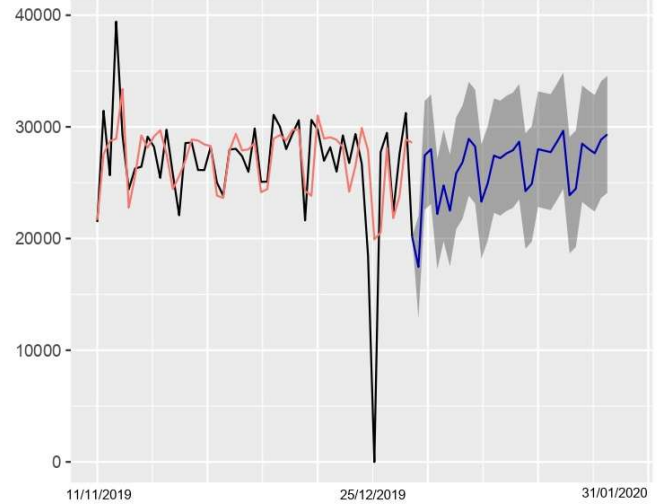


Figura 19. Predicción modelo  $SARIMAX(3, 0, 6)(2, 1, 3)_7$

#### IX. MODELO 4: SARIMAX + DUMMY + CUANTITATIVA

De los datos originales se decide también incluir la variable horas de vuelo, ya que como se observa en la figura 20 son directamente proporcionales una a la otra esto interpretado por el valor del  $R^2 = 0,9818$ , es decir entre más horas de vuelo realice las aeronaves más combustible usara y viceversa, por ello esta variable es muy explicativa y de ser tomada en cuenta logre que la predicción del modelo mejore. Es por ello que además de incluir la variable dummy "festivo" se agrega

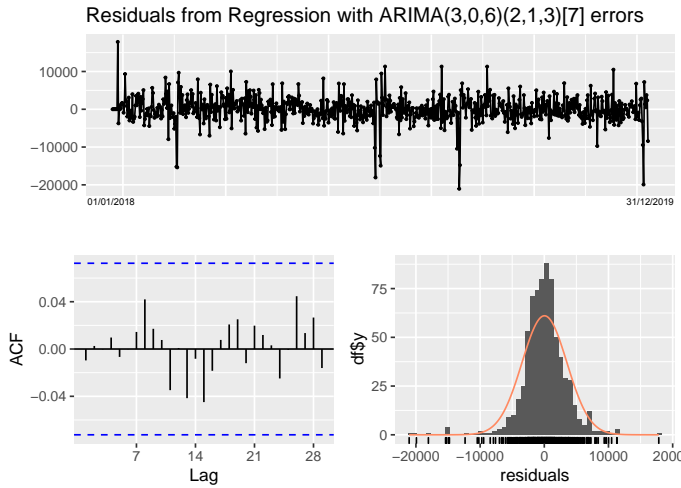


Figura 17. Residuos modelo  $SARIMAX(3, 0, 6)(2, 1, 3)_7$

Test Residuales	p-value	Interpretación
1 Ljung-Box	0.1147	Rechaza la hipótesis nula
2 Box-Pierce test	0.957	Acepta la hipótesis nula
3 Shapiro-Wilk	<2.2e-16	Rechaza la hipótesis nula
4 T Student	0.8481	Acepta la hipótesis nula

Tabla XI

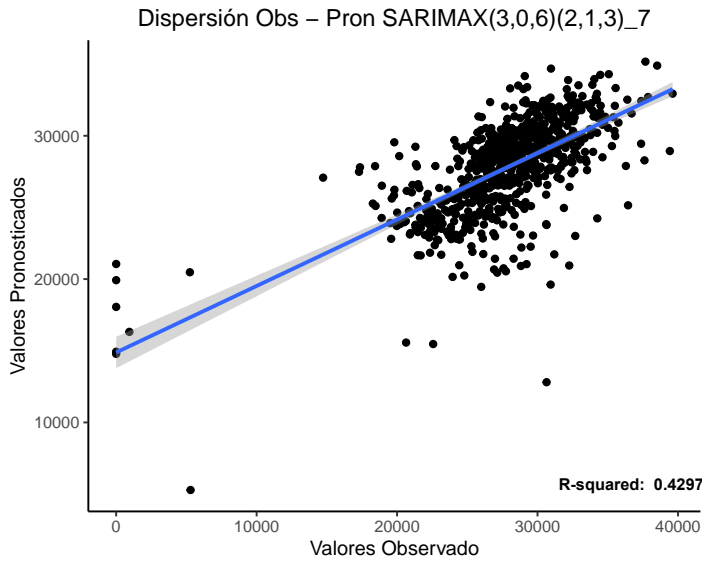


Figura 18. Dispersion  $SARIMAX(3, 0, 6)(2, 1, 3)_7$

la variable cuantitativa "horas de vuelo" dentro de un modelo SARIMAX.

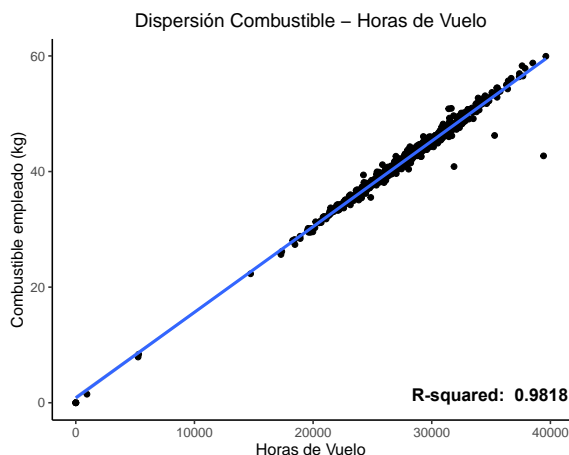


Figura 20. Correlación entre el combustible y las horas de vuelo

#### IX-A. Propuestas de modelos:

Usando esta información adicional, se propusieron 3 modelos que intentaron hacer las predicciones más ajustadas a los datos históricos, usando los valores RMSE y AIC como puntos de comparación; todos los modelos se corrieron en el software R Studio. Se sugiere el uso de un modelo "SARIMAX Errors", lo cual significa que se corre una regresión sobre la variable exógena y con los residuales se ajusta un modelo de series de tiempo. La consecuencia es que las estimaciones de los coeficientes y sus errores estándar serán incorrectas si se ignora la estructura de la serie temporal de los errores.

ítem	Modelo	RMSE (kg)	AIC
1	$SARIMAX(2, 1, 5)(2, 0, 3)_7$ errors	594.7692	11411.49
2	$SARIMAX(2, 0, 3)(4, 1, 3)_7$ errors	594.1376	11411.49
3	$SARIMAX(7, 0, 7)(3, 0, 3)_7$ errors	<b>584.5058</b>	11411.49

Tabla XII

#### IX-B. Gráficar el modelo seleccionado:

Del RMSE de la tabla XII es seleccionado el modelo  $SARIMAX(7, 0, 7)(3, 0, 3)_7$ , su comportamiento es adecuado, gana potencial de predicción con los valores más extremos ya que la dummy logra explicarlas y además las horas de vuelo justifican muy bien los picos en 0 kg y valores máximos. El modelo autorregresivo  $(p, d, q)$ , se describe  $p$  del modelo toma sus valores pasados 7 periodos de tiempo atrás, el promedio móvil  $q$  toma los errores de predicción retrasados 7 periodos de tiempo atrás y no emplea diferencias. Para el componente estacional  $(P, D, Q)$  este modelo toma el orden autorregresivo estacional en  $P$  (3), el orden promedio móvil estacional  $Q$  (3) y no aplica orden de diferencia estacional por último se toman

Modelo SARIMAX (7,0,7)(3,0,3)\_7 + Festivo + Horas Vuelo

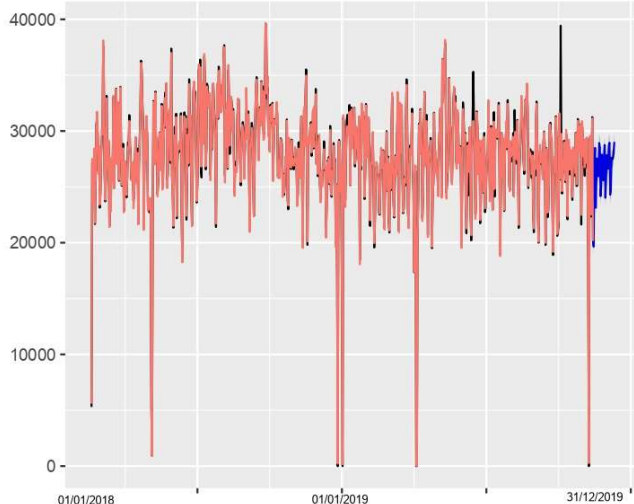


Figura 21. Modelo SARIMAX(7, 0, 7)(3, 0, 3)<sub>7</sub>

s 7 pasos de tiempo para un solo período estacional. Este modelo explica muy bien el comportamiento del combustible.

#### IX-C. Análisis de los residuales:

El análisis de los residuales para el modelo  $SARIMAX(7, 0, 7)(3, 0, 3)_7$  evidencia variancia y media constante en torno a 0, en el FAC se no se observa rezagos fuera de la banda de confianza, los residuales no tienen correlaciones significativas por ende el modelo captura toda la información pasada y es tomada en cuenta. En la línea de tiempo de valores residuales observados en la figura 22 solo se representan picos que toman valores extremos (3) pero el resto de información es muy cercana a 0, es por ello que el modelo excelente. De la tabla XIII se realizan las pruebas de normalidad de residuos donde el modelo cumple con el supuesto rechazando la hipótesis nula cumpliendo con los requerimientos del modelo.

	Test Residuales	p-value	Interpetación
1	Ljung-Box	0.09008	Acepta la hipótesis nula
2	Box-Pierce test	0.957	Acepta la hipótesis nula
3	Shapiro-Wilk	<2.2e-16	Rechaza la hipótesis nula
4	T Student	0.4501	Acepta la hipótesis nula

Tabla XIII

#### IX-D. Predicción del modelo seleccionado:

En la figura 25 se ve la predicción del modelo seleccionado  $SARIMAX(7, 0, 7)(3, 0, 3)_7$ , empleando los primeros 30 días después del dato final de la serie original, la línea azul es el valor que tomara el consumo de combustible, este toma diferentes valores acordes a los valores medios de su pasado. Para la predicción se debe ingresar un vector de 0s y 1s acordes a los próximos 30 días calendario siendo la exógena

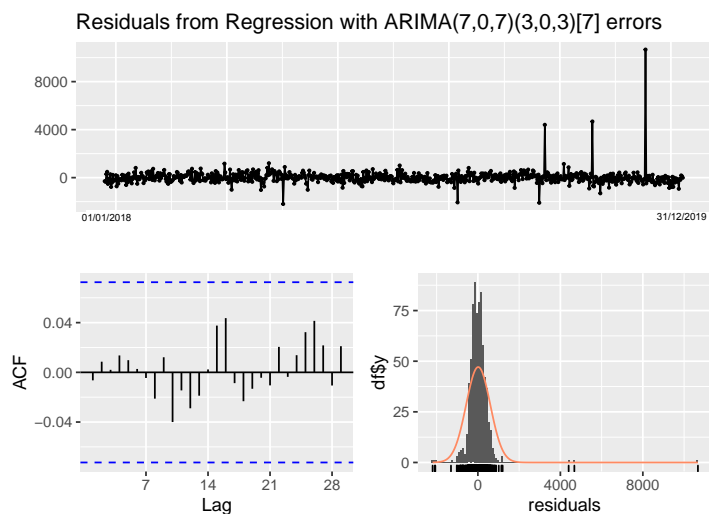


Figura 22. Residuos modelo  $SARIMAX(7, 0, 7)(3, 0, 3)_7$

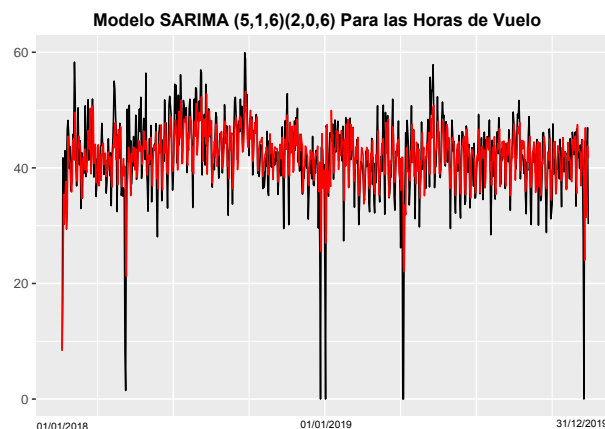


Figura 24. Modelo  $SARIMA(5, 1, 6)(2, 0, 6)$  Para las Horas de Vuelo

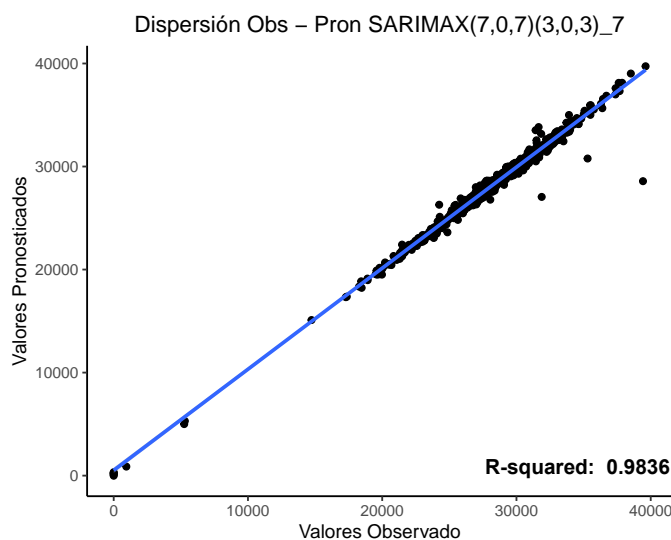


Figura 23. Dispersión  $SARIMAX(7, 0, 7)(3, 0, 3)_7$

	Fecha	Tvuelo (h)	Festivos
1	1/01/2020	29.45	1
2	2/01/2020	40.74	0
3	3/01/2020	43.02	0
4	4/01/2020	35.13	1
.	.	.	.
.	.	.	.
.	.	.	.
30	30/12/2020	42.39	0

Tabla XIV

$(1, 0, 0, 1, 1, \dots, 0, 0)$ , a parte de ello las horas de vuelo al ser una variable que cambia en el tiempo no se conoce su valor en el futuro, por ello se debe realizar un modelo que también prediga su comportamiento, en esta ocasión se realiza un modelo  $SARIMA(5, 1, 6)(2, 0, 6)$  figura 24 con valor de  $RMSE = 5.563$  h, gracias a este modelo se obtienen los valores futuros de las horas de vuelo para emplear en la predicción del  $SARIMAX(7, 0, 7)(3, 0, 3)_7$ , en la tabla XIV se muestran los valores empleados para la predicción.

#### IX-E. Escribir el modelo:

El modelo  $SARIMAX(7, 0, 7)(3, 0, 3)_7$  con la variable exógena Festivo e incluyendo las horas de vuelo con errores seleccionado se escribe de la siguiente manera 4:

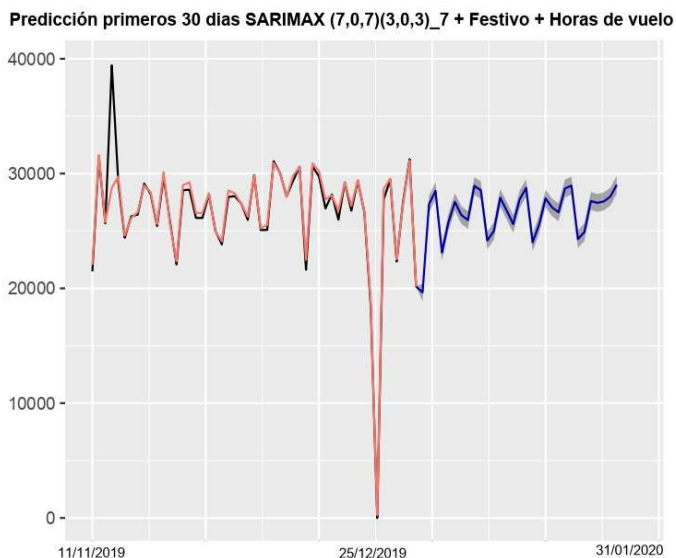


Figura 25. Predicción modelo  $SARIMAX(7, 0, 7)(3, 0, 3)_7$  + Festivos + Horas de Vuelo

$$SARIMAX(7,0,7)(3,0,3)_7 + Festivo + HorasVuelo$$

$$y_t = \beta_0 + \beta_1 x_t + \beta_2 x_t + \eta_t$$

$$\begin{aligned} \eta_t &= (1 - \alpha_1 B - \dots - \alpha_7 B^7) \\ &\quad - (1 - \gamma_1 B^7 \dots - \gamma_3 B^{21}) \eta_t = \\ &= (1 - \phi_1 B - \dots - \phi_7 B^7) - (1 - \theta_1 B^3 - \dots - \theta_3 B^{21}) \varepsilon, \\ \varepsilon &\sim RB(0, \sigma_a^2) \end{aligned}$$

$$\begin{aligned} \beta_0 &= 451,8114, \beta_1 = -277,1365, \beta_2 = 651,3673 \\ \alpha_1 &= -0,877, \alpha_2 = 0,065, \alpha_3 = 0,164, \alpha_4 = -0,339 \\ \alpha_5 &= 0,267, \alpha_6 = 0,912, \alpha_7 = 0,789, \\ \gamma_1 &= -0,271, \gamma_2 = 0,115, \gamma_3 = -0,289 \\ \phi_1 &= 0,942, \phi_2 = 0,047, \phi_3 = -0,112, \phi_4 = 0,355 \\ \phi_5 &= -0,183, \phi_6 = -0,877, \phi_7 = -0,820 \\ \theta_1 &= 0,215, \theta_2 = -0,123, \theta_3 = 0,360 \end{aligned}$$

(4)

Donde  $\beta_0$  es una constante que se interpreta como el intercepto,  $\beta_1$  la constante que suministra la variable Festivo y  $\beta_3$  es la constante suministrada por las Horas de Vuelo.

## X. RESULTADOS Y DISCUSIÓN

En la tabla XV se describen los modelos seleccionados, donde se aprecia la comparativa de los resultados además se incluye el valor Error Porcentual Absoluto Medio (MAPE) el cual es el tamaño del error en términos porcentuales (para los valores de 0 kg que corresponden a 5 fechas, se extraen de este cálculo dado que es una división por 0); cada modelo mejoro a su antecesor brindando un porcentaje menor de error en la predicción, siendo el último el más cercanos a los datos reales con un error del 1.05 %.

Modelo	MAPE	RMSE	AIC
ARIMA (15,1,8)	11.41 %	3708.218	14117.57
SARIMA(6,1,4)(6,0,3)_7	11.29 %	3683.658	14104.47
SARIMAX(3,0,6)(2,1,3)_7 errors+ Festivo	10.91 %	3498.238	14104.47
SARIMAX(7,0,7)(3,0,3)_7 errors+ Festivo + Horas de vuelo	1.05 %	584.5058	11411.49

Tabla XV

Los modelos planteados para describir el comportamiento del combustible descritos en este documento han cumplido cada uno de los pasos planteados en la metodología Box – Jenkins descrita en la metodología, a continuación, en la tabla XVI se describen algunas fortalezas y debilidades para cada uno de ellos.

Se observa que ningún modelo logro adaptarse a ciertos valores picos o atipicos de la serie de datos originales como se observa en la figura 26, (estos picos coinciden con los que se aprecian en los residuales del último modelo descrito figura 22) esta es señal que el combustible empleado no es acorde a las horas de vuelo registradas.

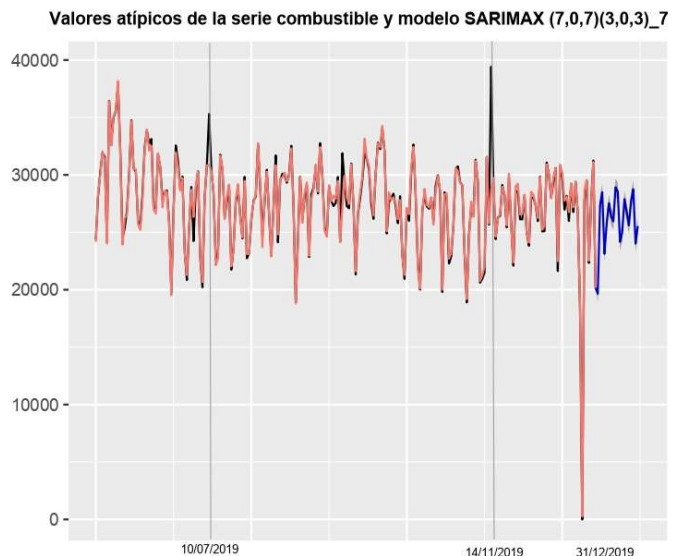


Figura 26. Valores atípicos de la serie original

Modelo	Debilidad	Fortaleza
ARIMA (15,1,8)	Posee el valor de RMSE más alto de todos los modelos	Es un modelo sencillo, no requiere mayor análisis y suministra información adecuada
ARIMA (6,1,4)(6,0,3)_7	Los residuales son muy extremos a comparación de los modelos más elaborados	Sus valores de predicción son muy similares al SARIMAX con 2 exógenas
SARIMAX (3,0,6)(2,1,3)_7 + Festivo	La reducción del RMSE con respecto al modelo ARIMA no es tan significativo	Esta exógena brinda información adicional que no se tenían de la base de datos original
SARIMAX (7,0,7)(3,0,3)_7 + Festivo + Horas de Vuelo	Requiere conocer valores futuros de variable	Tiene la capacidad de acoplarse a los datos originales por ende tiene el RMSE más bajo

Tabla XVI

## XI. CONCLUSIONES

En este estudio, proponemos modelos aplicando la teoría de series de tiempo, incluyendo modelos ARIMA, SARIMA y SARIMAX, que pueden ser utilizados de acuerdo con el análisis requerido por la empresa, cada uno de los cuales expresa un análisis diferente, es decir. se puede utilizar en diferentes situaciones de acuerdo con el requerimiento. Utilizamos el MAPE y el RMSE como índices de error para testificar cada modelo. La metodología propuesta no sólo es efectiva sino también eficiente. La información derivada de los modelos propuestos puede ser empleada para pronosticar el consumo de combustible de las aeronaves ATR 42 y con ello cuantificar el valor de este, los modelos también pueden ser usados como herramienta de control y auditoria del insumo.

## REFERENCIAS

- [1] González Casimiro, María Pilar, "Análisis de series temporales: Modelos ARIMA," Economía Aplicada III/Ekonomia Aplikatua III, UPV/EHU, 2009.
- [2] Hyndman, Rob J., and George Athanasopoulos. "Forecasting: principles and practice," OTexts, 2018.
- [3] Yu, He, et al. "A hybrid model for financial time series forecasting—integration of EWT, ARIMA with the improved ABC optimized ELM," IEEE Access 8, 2020: 84501-84518.

- 
- [4] Dritsaki, Chaido. "Oil consumption forecasting using ARIMA models: An empirical study for Greece," 670216917, 2021.
  - [5] Mariño, Maria D., et al. "Modelos de series temporales para pronóstico de la demanda eléctrica del sector de explotación de minas y canteras en Colombia." Revista EIA 18.35, 2021 : 35007-pp.
  - [6] Aladag, Cagdas Hakan, ed. "Advances in time series forecasting. Vol. 2. Bentham Science Publishers." 2017.
  - [7] Guerrero Guzmán, Víctor Manuel. "Análisis estadístico de series de tiempo económicas. No. 04; Q280, G8 2003." 2003.