

# Modelos SARIMA para pronóstico de pasajeros en vuelos nacionales e internacionales en Colombia

10.53358/ideas.v3i2.629

Luis Garzón<sup>1</sup>, Julián Arellana<sup>2</sup>, Víctor Cantillo<sup>3</sup>

<sup>1</sup> Universidad del Norte, Barranquilla, Colombia

<sup>1</sup>lagarzon@uninorte.edu.co, <sup>2</sup>jarellana@uninorte.edu.co, <sup>3</sup>vcantillo@uninorte.edu.co

## RESUMEN

El sector del transporte aéreo de Colombia se posiciona como el tercer país con mayor incremento de operaciones en las terminales aéreas de América Latina. Para alcanzar un crecimiento sostenible del sector, es necesario contar con metodologías de planificación basadas en estimaciones sobre la previsión de las cantidades de pasajeros. En este contexto de planificación operativa, este trabajo presenta varios modelos de pronóstico predictivo de largo plazo con base a las operaciones en vuelos nacionales e internacionales registradas por Aerocivil desde el año 2005. Se aplicó un enfoque de Modelado de Orden Diferenciado Estacional, Autoregresivo y de Medias Móviles (SARIMA). Se presenta una metodología secuencial para la selección de cinco modelos sobre un total de 256 calculados. El modelo  $(0,1,0) \times (0,1,0)_{12}$  resultó ser el más apropiado con una estimación del RMSEA del 3.1%. Los modelos presentan altos rendimientos de estimación a pesar de los efectos causados por la última pandemia COVID-19.

**Palabras Clave:** SARIMA, pronóstico, pasajeros.

**Abstract.** The Colombian aviation sector is positioned as the country with the third largest increase in operations at air terminals in Latin America. Sustainable expansion of this sector demands planning methodology based on estimates of forecast passengers' flows. In this context in operational planning, this paper presents several long-term predictive forecasting models based on domestic and international flight traffic operations reported by Aerocivil since 2005. A Seasonal Autoregressive Integrated Moving Average (SARIMA) model was conducted. A sequential methodology was performed to select five models from 256 estimated. The  $(0,1,0) \times (0,1,0)_{12}$  model showed the best fit in terms of RMSEA 3.1%. The models demonstrate high performance in estimation despite the effects of COVID-19.

**Keywords:** SARIMA, forecasting, passengers.

## Introducción

La variación del tráfico de pasajeros del sector aeroportuario en la región de América Latina y el Caribe está marcada por la evolución del ciclo económico mundial, con una estrecha relación económica entre la demanda de vuelos nacionales e internacionales con respecto a las tasas de crecimiento del PIB. Según registros del Banco Mundial [1], durante los últimos años en la región de América Latina y el Caribe se registraron aproximadamente 338 millones de pasajeros, 4.6 millones de toneladas métricas de carga y 6.4 millones de operaciones aéreas.

El sector del transporte aéreo de Colombia se posiciona en el mercado de América Latina y el Caribe (ALC) como el tercer país con mayor incremento de operaciones en las terminales aéreas. Este posicionamiento responde a la dinamización del sector del transporte aéreo y a la suscripción de acuerdos internacionales para la liberación del espacio aéreo comercial. Este avance ha permitido la diversificación y aumento de la oferta de vuelos con nuevos operadores aéreos. El establecimiento de nuevas empresas en el país se explica en gran medida por las difíciles condiciones orográficas que dificultan los desplazamientos de personas y carga por vía terrestre.

La entrada a la libre competencia de nuevos operadores aéreos, permitió concretar varios acuerdos bilaterales, así como de la decisión de crear acuerdos de cielos abiertos con varios países [2]. Según los registros de la oferta y la demanda de la cantidad de pasajeros existe un incremento significativo desde los años 90 con 8.6 millones de pasajeros para vuelos nacionales y 2.2 millones en vuelos internacionales hasta el 2016 mostrados en la Figura 1>.

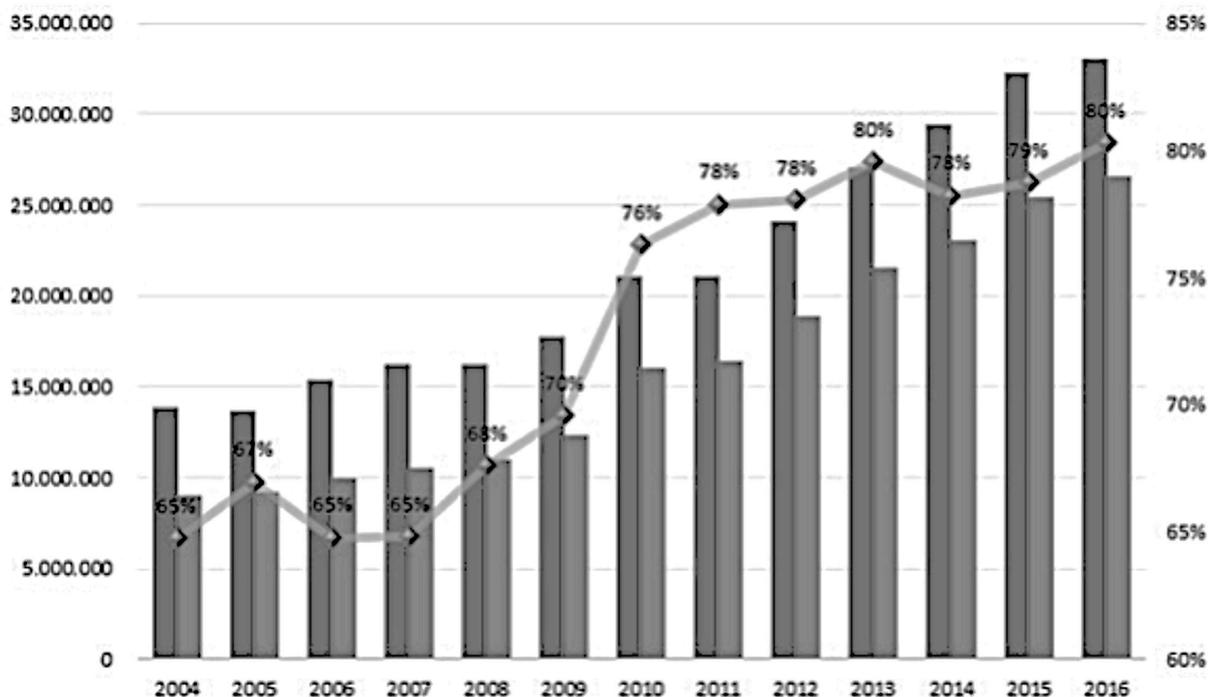


Fig. 1. Oferta y demanda de pasajeros en vuelos nacionales

La Figura 2 muestra los incrementos del número de pasajeros en los puntos más alejados de las grandes urbes, razón por la cual, los tomadores de decisión demandan de una estructuración y planificación mejor organizada con respecto a la previsión de las operaciones en los nuevos y existentes terminales aeroportuarios [3]. El gobierno nacional de Colombia a través del Departamento Nacional de Planeación [4] estableció un programa para la mejora en la capacidad de la infraestructura para 17 aeropuertos, los cuales representan una operación regular de pasajeros en las capitales de departamento y puntos estratégicos identificados por Aerocivil. En otros 32 aeropuertos se implementará un programa para mantener las condiciones básicas de operación y vigilancia del espacio aéreo que se establecen en el Plan de Navegación Aérea de Aerocivil [5].

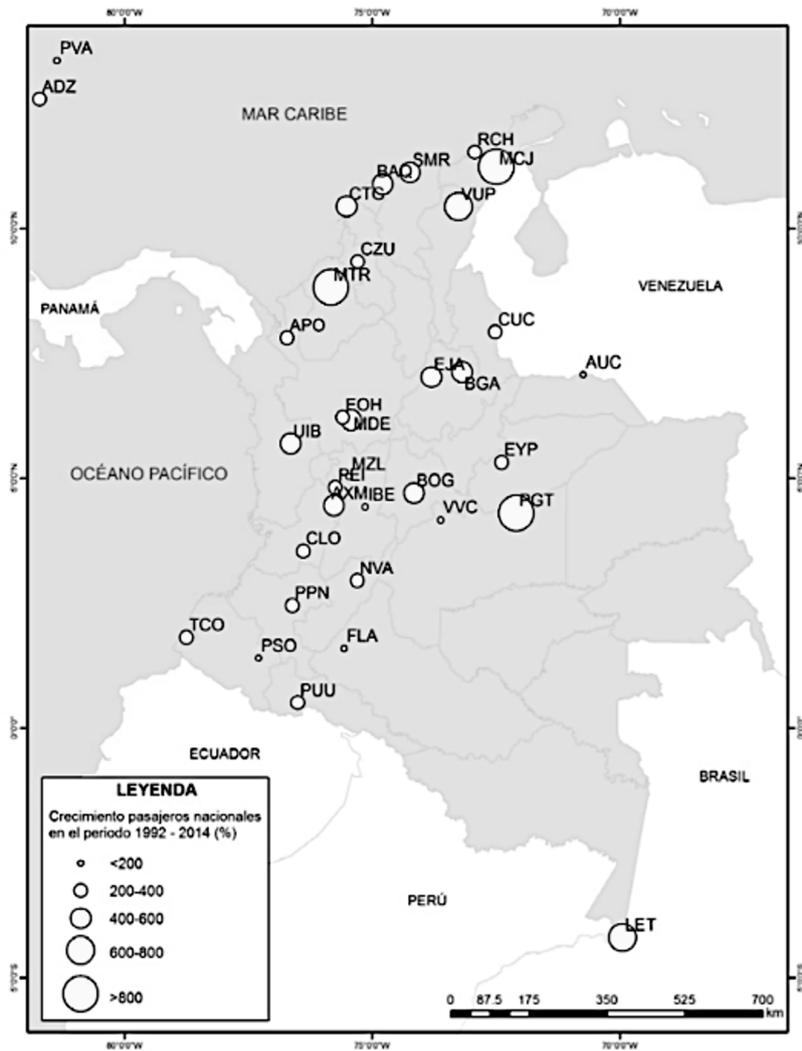


Fig. 2. Crecimiento de pasajeros en aeropuertos de Colombia

Para cumplir con dicho programa es necesario lograr una correcta estimación y previsión de la demanda de pasajeros, lo cual permitirá estructurar de manera sostenible la planificación de inversión en el sector aeroportuario tanto en infraestructura como en operaciones. Además, los responsables en la toma de decisiones necesitan planificar de acuerdo a las previsiones y condiciones actuales y pasadas de la demanda del flujo de pasajeros. Para esto, se han propuesto varios enfoques cuantitativos que permiten predecir la demanda futura mediante un análisis estadístico con datos existentes del mercado.

El análisis de regresión aplicado sobre la especificación de una función es el método más utilizado, y permite observar la causalidad entre las variables observadas. Este enfoque tiene como objetivo encontrar a la variable independiente que exprese de mejor manera a la variable dependiente, y de esta manera, predecir los valores en el futuro. Sin embargo, el análisis de series temporales supone un trabajo más amplio para la obtención de valores del futuro como una función de la visión del pasado. En el sector de la aviación, este análisis es posible realizarlo ya que existe una bitácora de información bastante amplia en el tiempo que permiten estimar pronósticos con altos niveles de rendimiento.



Los modelos estocásticos y dinámicos de series de tiempo se presentan como la herramienta adecuada para el cálculo de pronósticos, permitiendo la interpretación de los datos de tipo univariantes y multivariantes [6]. Estos modelos han sido mejorados integrando herramientas computacionales que disponen de una mayor capacidad de procesamiento, teniendo como resultado series de tiempo con mejores especificaciones y más eficientes [7].

Existen varias herramientas y modelos para la realización de pronósticos a partir de los datos disponibles de una serie de tiempo. Uno de ellos se refiere al método de suavizamiento exponencial propuesto a fines de la década de 1950 por Holt [8] y Durban y Koopman [9]. Otras especificaciones permiten mejorar el ajuste de los datos [10]. Sin embargo, el modelo clásico ARIMA propuesto por Box y Jenkins [11] y posteriormente los modelos de media móvil integrada autorregresivo estacional multiplicativo SARIMA han permitido capturar de mejor manera las características de orden integrado de diferenciación relacionadas con la estacionariedad y estacionalidad requeridas para la interpretación de resultados sobre el pronóstico de pasajeros del transporte aéreo [12].

El análisis de series temporales se divide en análisis en el dominio del tiempo y análisis en el dominio de la frecuencia. El análisis en el dominio del tiempo supone que la dependencia o relatividad entre los datos adjuntos retrocede con los datos anteriores. Basándose en este supuesto, descompone los datos de las series temporales en factores deterministas, como la tendencia, y factores estadísticos, como el ruido. Los enfoques típicos son autoregresivos (AR), medias móviles (MA), autoregresivos y medias móviles (ARMA), ARIMA y SARIMA [13]; [14]; [15]; [16]. Por otro lado, el análisis del dominio de la frecuencia convierte los datos de las series temporales normales con una media y una varianza estables en un compuesto lineal de funciones periódicas fuertes.

Independientemente del método de predicción utilizado, los resultados de las estimaciones siempre resultarán diferentes. Sin embargo, la precisión de la predicción viene determinada por la magnitud del error. La métrica de la precisión calculada sobre el error incluye la comparación de varios criterios estadísticos como el error porcentual medio absoluto (MAPE), el error cuadrático medio (RMSEA), y la desviación media absoluta (MAD).

En este contexto de planificación operativa, esta investigación tiene como objetivo estimar modelos de pronóstico para la cantidad de pasajeros mensual para vuelos nacionales e internacionales en Colombia. Los pronósticos contendrán la especificación del nivel de incertidumbre y confianza de acuerdo con los períodos de cálculos especificados, se tratará de identificar posibles factores externos asociadas con altas variaciones del número de vuelos realizados en función del pronóstico previsto, por ejemplo, el relacionado a la disminución del número de vuelos debido a situaciones de amenaza mundial para la salud. Al final, se determinará el modelo que mejor represente y caracterice a las dos bases de datos en función de la estimación y comparación de diferentes criterios estadísticos como la raíz del error cuadrático medio.

La estructura general de este artículo está compuesta por cuatro capítulos, incluida esta sección introductoria que incluye una revisión literaria sobre el uso de modelos de pronósticos basados en series de tiempo. En el segundo capítulo se menciona la metodología para el diseño y especificación secuencial de modelos SARIMA. En el tercer capítulo se discuten y analizan los resultados. La última muestra las conclusiones y se brinda una visión para seguir investigando con la aplicación de este tipo de modelos de pronóstico con información de las cantidades de pasajeros.

### Metodología

El marco metodológico como se muestra en la Figura 3 está compuesto en una primera etapa por la identificación de varias propiedades empíricas y estadísticas que permiten analizar la base de datos y evaluar las condiciones de estacionariedad, estacionalidad, el orden de estimadores de auto regresión y medias móviles. Luego se procede con la estimación del modelo y la determinación de varios estimadores como el criterio estadístico AKAIKE (AIC) o Bayesiano (BIC) y Máxima Verosimilitud (ML). Con los modelos definidos se diagnostican varias condiciones de las series de tiempo relacionadas con las propiedades de normalidad y correlación de los datos, así como la comprobación de estacionariedad. Finalmente, si el modelo cumple con todos los criterios de bondad de ajuste estadístico se realiza una categorización de los modelos en función de las estimaciones del error. Una vez definido el mejor modelo se procede a calcular los pronósticos de las cantidades de pasajeros.

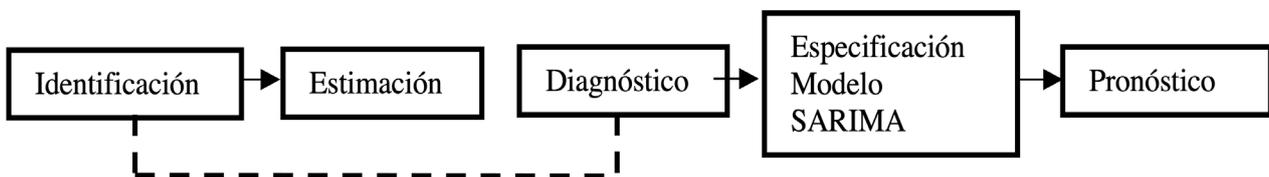


Fig. 3. Marco metodológico

La base de datos se encuentra disponible en los boletines de oferta y demanda de transporte de pasajeros recopilados de forma mensual en el sitio web oficial de la Administración de Aerocivil [17] y del Ministerio del Transporte [18]. El registro corresponde a 196 meses sobre el flujo de pasajeros en vuelos nacionales e internacionales reportados en territorio Colombiano, correspondiente al período entre enero del 2005 hasta abril del 2021, según el detalle mostrado en la Tabla 1.

Tabla 1. Promedio mensual de pasajeros movilizados (x103)

Año	2005	2006	2007	2008	2009	2010	2011	2012	2013
Nacionales	759	821	861	891	1,005	1,336	1,368	1,570	1,792
Internacionales	351	406	406	462	502	573	637	721	823

Año	2014	2015	2016	2017	2018	2019	2020	2021
Nacionales	1,918	2,120	2,210	2,138	2,221	2,500	852	1,455
Internacionales	904	997	1,077	1,142	1,249	1,295	357	393

El detalle del número de pasajeros a bordo en vuelos domésticos es mayor en comparación a la cantidad de pasajeros con destinos internacionales y su comportamiento ha sido creciente en el tiempo. Sin embargo, se puede observar una fuerte caída provocada por el surgimiento de la actual pandemia COVID-19 cuando el gobierno Colombiano decretó el cese parcial y total de las operaciones regulares ofrecidas por las diferentes empresas aéreas. Por esta razón, se puede observar en la Figura 4 una reducida cantidad de vuelos realizados en condiciones de operación no regular durante este período.

La serie de tiempo es construida a partir de la función general:  $X_t=U_t+Y_t$ , donde  $U_t$  representa a la tendencia y  $Y_t$  el proceso estacionario. La primera parte de la metodología se relaciona con la identificación de los procedimientos empíricos, a través del análisis de las propiedades gráficas; además, estas son contrastadas con el uso de criterios estadísticos bajo el uso de ciertas consideraciones sobre el comportamiento de los datos de normalidad e independencia.

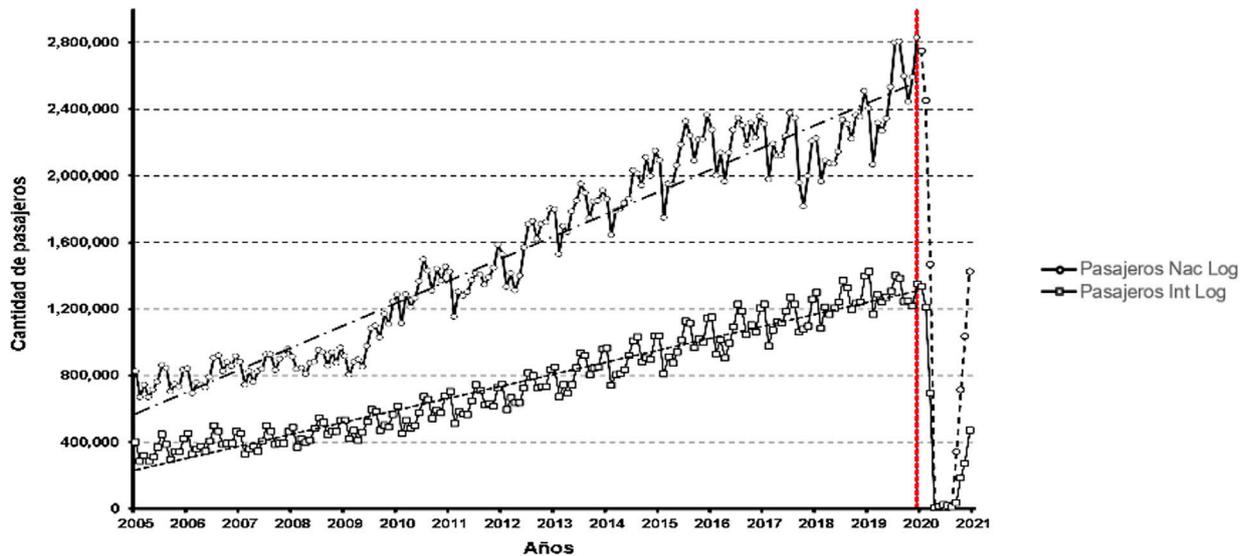


Fig. 4. Cantidad de pasajeros en vuelos nacionales e internacionales

La serie de datos presenta un comportamiento no estacionario y no lineal con tendencia; cualquier cambio brusco aparente en el comportamiento de la serie de tiempo puede sugerir una condición de estacionariedad. Esta característica indica la consistencia de los datos en función de estimadores estadísticos como la media, la varianza, la asimetría y la curtosis, de forma independiente del tiempo transcurrido. La condición de estacionariedad cumple dos condiciones, la primera que la función de distribución acumulativa no cambie con el paso del tiempo y la segunda referida a la covarianza después del tiempo  $h$  sea la misma que la anterior.

$$F_{t_1, t_2, \dots, t_n}(x_1, \dots, x_n) = F_{t_1+h, t_2+h, \dots, t_n+h}(x_1, \dots, x_n) \quad (1)$$

$$Cov(x_t, x_s) = Cov(x_{t+h}, x_{s+h}) \quad (2)$$

Como un primer tratamiento de los datos, se procede a realizar una transformación Box-Cox [11]. En este caso, la intervención es realizada para conseguir una transformación de tipo logarítmica mostrada en la Figura 5. Este cambio permite disminuir la variabilidad de la desviación estándar con respecto a la media calculada.

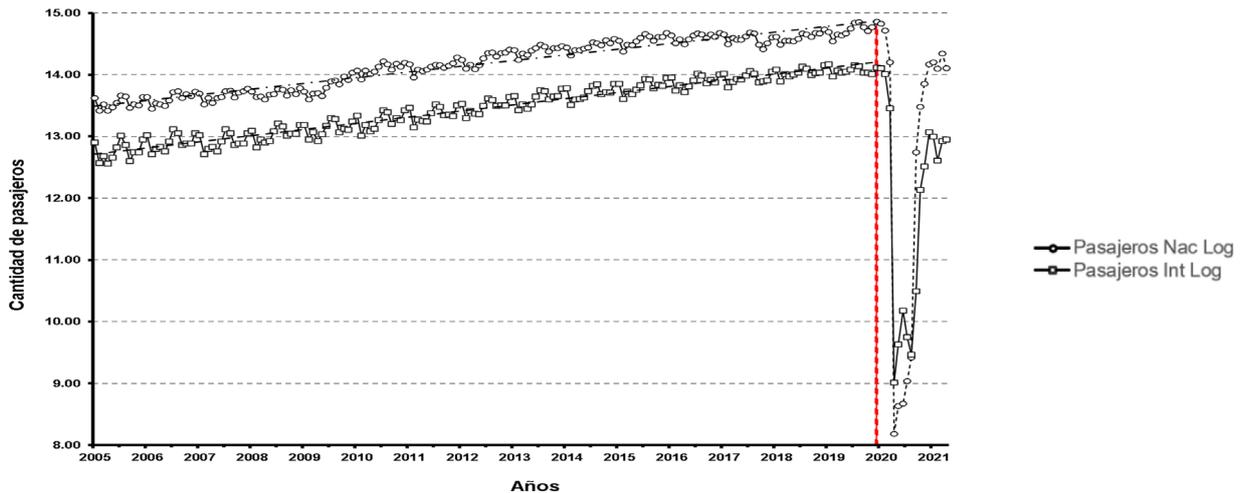


Fig. 5. Serie de datos con transformación logarítmica.

Para una segunda parte del desarrollo secuencial del marco metodológico sobre la estimación de los parámetros del modelo, se identifican de forma secuencial los efectos sobre sus componentes de orden autorregresivo, (AR)  $p$ , orden de media móvil, (MA)  $q$  y de orden de diferenciación integrado,  $d$ . La característica de no estacionariedad es tratada aplicando una primera diferenciación de orden  $i$  por medio de la técnica correspondiente al modelo de media móvil integrada autorregresiva SARIMA. Un gráfico temporal de los datos facilita identificar si es necesario diferenciar una vez,  $d = 1$ , o recurrir a una segunda diferenciación  $D=1$  para eliminar la condición de estacionalidad y establecer un modelo SARIMA.

Los modelos de pronóstico de largo plazo SARIMA denominados de forma general como  $ARIMA(p, d, q) \times (P, D, Q)_s$ , se caracterizan por disminuir efectos de tipo estacionales [13], y están constituidos por la siguiente expresión:

$$\phi_P(B^s)\phi(B)\nabla_s^D\nabla^d x_t = \theta_Q(B^s)\theta(B)w_t \tag{3}$$

Donde los componentes autorregresivos ordinarios (AR) y con media móvil (MA) están representados por polinomios  $\phi(B)$  y  $\theta(B)$  de orden  $p$  y  $q$ , respectivamente, y los componentes autorregresivos estacionales y de media móvil por  $\phi_P(B^s)$  y  $\theta_Q(B^s)$  de órdenes  $P$  y  $Q$ , obteniéndose los componentes de diferencia ordinaria y estacional por  $\nabla^d = (1-B)^d$  y  $\nabla_s^D = (1-B^s)^D$ . El último término  $w_t$  se refiere al efecto de ruido blanco gaussiano.

Para las series de datos de vuelos se realizó una primera diferencia y las características de no estacionalidad aún persisten, y esto es comprobado por medio de la estimación de las raíces unitarias de los estimadores Dickey Fuller [19]. Las raíces son determinadas a partir de la función  $Y_t$  del proceso estacionario  $\Delta y = \alpha + \beta t + \gamma y_{(t-1)} + \epsilon_t$ , donde  $\alpha$  es una constante,  $\beta t$  es la tendencia y  $\gamma y_{(t-1)}$  son las raíces de la función para su estimación. Al realizarse una segunda diferencia correspondiente a la definición de un modelo SARIMA, se puede observar en la parte inferior de la Figura 6 que los residuos de la serie de datos se comportan de forma estacionaria y estacional.



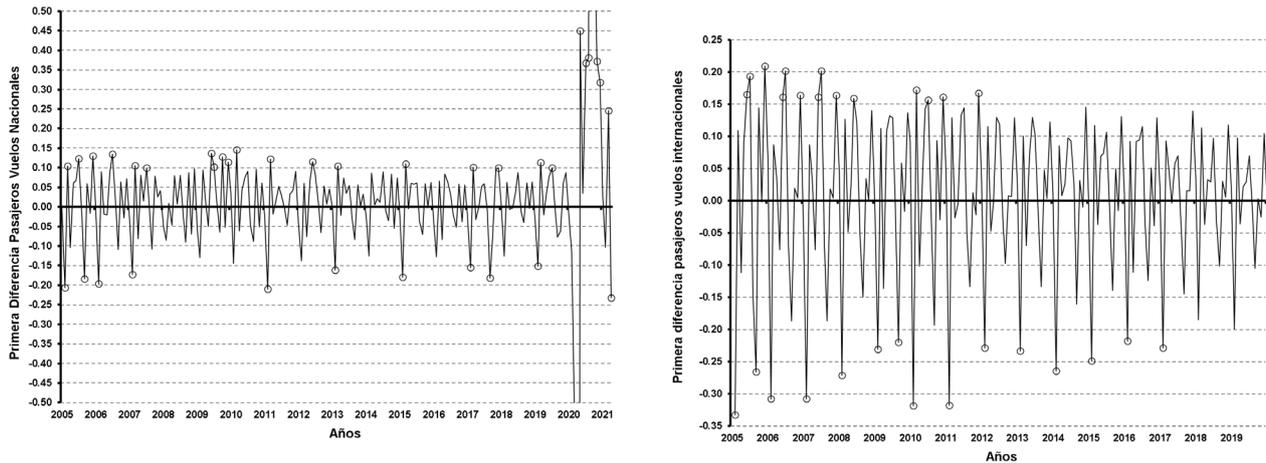
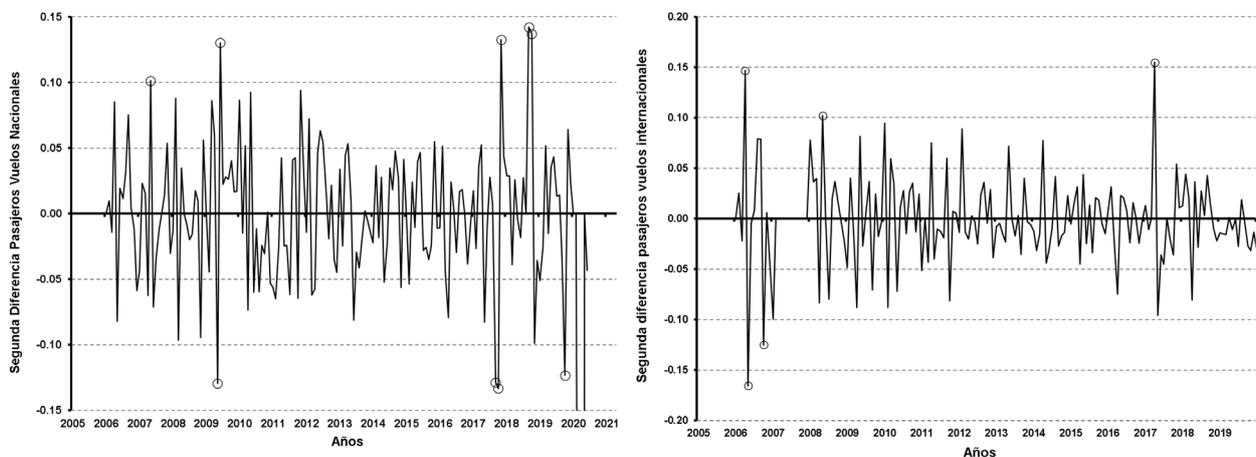


Fig. 6. Residuos con orden de diferenciación estacionario ( $d=1$ ) y estacional ( $D=1$ ).



Luego de la primera y segunda diferencia se puede observar el comportamiento de los residuos, y se procede a evaluar las funciones de autocorrelación simple (ACF) y autocorrelación parcial (PACF). El uso de las gráficas de estas dos funciones presentado en la Figura 7 permite identificar los picos de los residuos y posteriormente ajustarlos seleccionando el componente autorregresivo (AR) o de media móvil (MA) que mejor corresponda. Los picos que aparecen en estas funciones se eliminan ajustando mediante la selección más adecuada de dichos componentes de acuerdo con las propiedades de análisis sugeridas en la literatura [20].

Para una tercera parte metodológica relacionada con el diagnóstico de las series de datos, dado que se tratan de estimaciones, no siempre es evidente identificar el comportamiento de la ACF o la PACF, por esta razón, se comprueban los criterios de normalidad en función del modelo seleccionado por medio de la prueba de Anderson-Darling (AD) o Kolmogorov-Smirnov [21]; [19]. Además, para su categorización se calculan los estimadores de máxima verosimilitud (ML), el criterio de Akaike (AIC) y el criterio de información Bayesiana (BIC).

La selección del modelo más apropiado para un conjunto de datos no es una tarea sencilla, dado que existen muchos tipos de combinaciones entre los operadores, los cuales producen diferentes conjuntos de aproximaciones estacionarias, estacionales, autorregresivos y de medias móviles. Pueden presentarse similitudes entre los estimadores de los modelos SARIMA, sin embargo, es necesario proceder con esta fase de ajuste del modelo e ir seleccionando de manera preliminar valores de  $p$ ,  $d$  y  $q$ , y  $P$ ,  $D$  y  $Q$ .

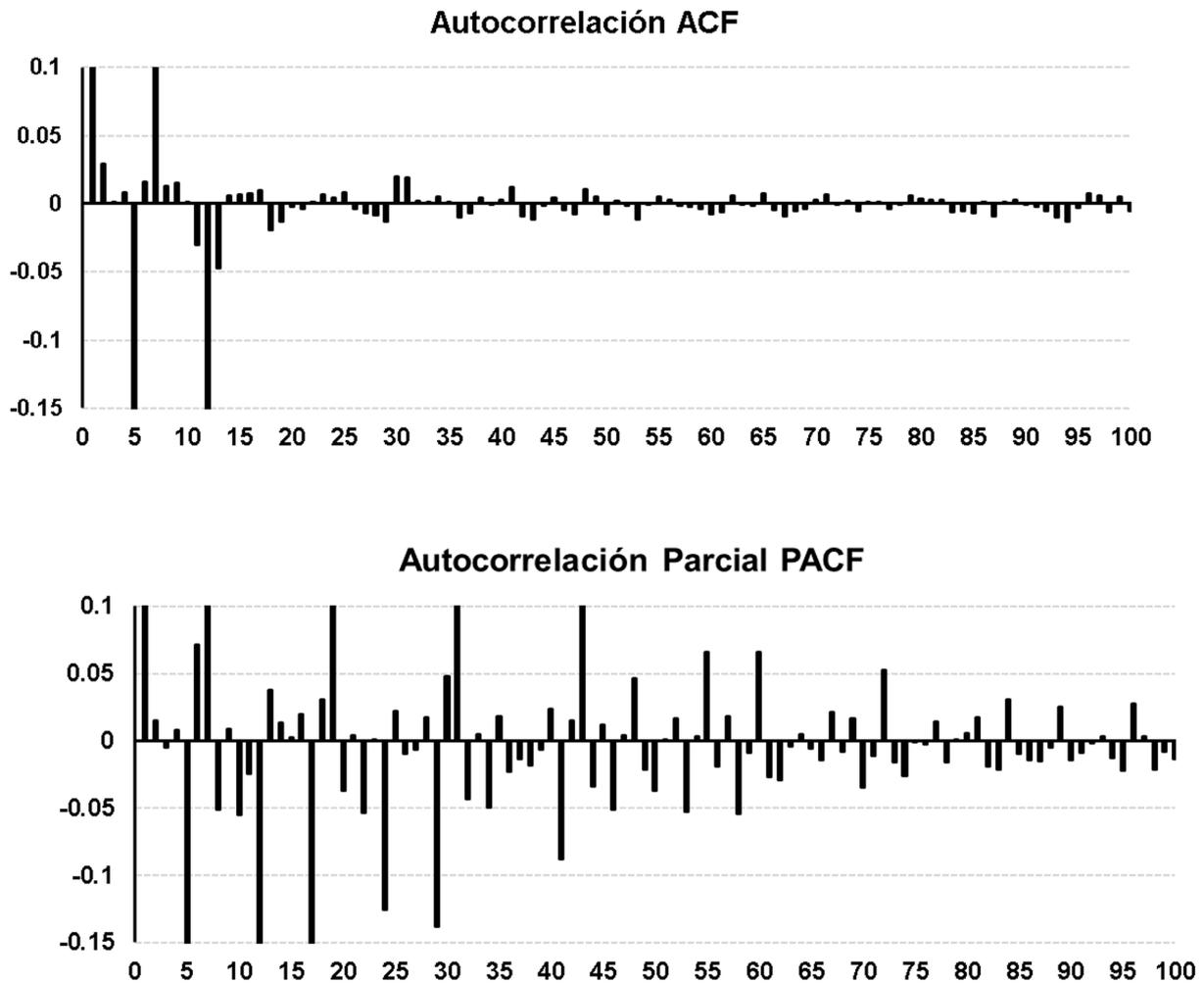


Fig. 7. Función de autocorrelación (superior) (b) Función Autocorrelación parcial (inferior),

Para la última parte relacionada con el análisis de pronósticos sobre las series de datos, se evaluaron distintas especificaciones, partiendo de los modelos básicos SARIMA  $(p,1,q) \times (P,1,Q)_2$ , de esta forma, se determina el modelo que mejor represente las características de ajuste sobre las estimaciones obtenidas. Los valores pronosticados fueron comparados con los valores reales usando el criterio estadístico de la raíz del error cuadrático medio. Los métodos representativos para estimar los parámetros de cada término también incluyen el método de estimación de máxima verosimilitud.

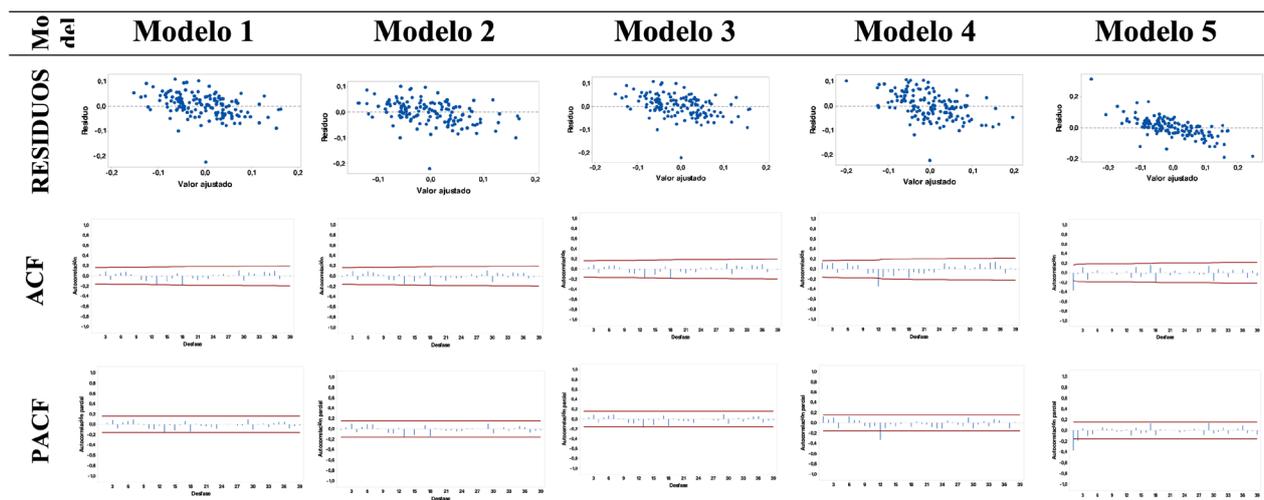
### Resultados y discusión

Las series de tiempo de la cantidad de pasajeros en vuelos nacionales e internacionales estuvieron conformadas por un total de 196 observaciones cada una. Durante los últimos meses se observa un comportamiento atípico de los datos, debido a la aparición de la pandemia actual del COVID-19. La disminución registrada modifica el comportamiento creciente del número de pasajeros y esto conlleva a un comportamiento atípico de los datos para los últimos meses, razón por la cual, los pronósticos realizados disminuyen la potencia de la estimación y los errores calculados se hacen más grandes. Por esta razón, se propone estimar cuatro modelos diferentes, dos modelos que contiene datos de la cantidad de pasajeros en vuelos nacionales, en condiciones antes y después de la pandemia; y bajo la misma consideración para los datos de la cantidad de pasajeros de vuelos internacionales.



Los resultados se muestran mediante la selección y análisis de 5 especificaciones que presentan las mejores características de un total de 256 posibles modelos para cada serie de datos. Como parte del diagnóstico de los modelos, según la metodología Box & Jenkins [13] se examina la aleatoriedad de los residuos y el criterio de independencia de los datos por medio de las funciones de autocorrelación simple y autocorrelación parcial. Por esta razón, se realiza una comparación de los mejores modelos estimados, los cuales son observados de forma gráfica y contrastados con estimadores estadísticos mostrados en la Tabla 2.

Tabla 2. Comparación de modelos SARIMA



El modelo de pronóstico para la cantidad de pasajeros en vuelos nacionales sin la consideración de la disminución de vuelos provocados por la pandemia COVID-19 correspondió a la especificación SARIMA  $(0,1,1) \times (0,1,0)_{12}$  y en vuelos internacionales SARIMA  $(0,1,0)(0,1,0)_{12}$ , para los cuales se estimó un error cuadrático medio de 0,036 para vuelos nacionales y 0,031 para vuelos internacionales. Los modelos son expresados algebraicamente de la siguiente forma y se muestran los valores estimados de los coeficientes de autoregresión y medias móviles:

Modelo para pasajeros en vuelos nacionales:

$$(1 - B^{12})\log(1 - B)x_t = (1 + \theta_1 B) w_t \tag{4}$$

$$\theta_1 = 6.47,$$

Modelo SARIMA para pasajeros en vuelos internacionales:

$$(1 - B^{12})\log(1 - B)x_t = w_t \tag{5}$$

Se presentan el resto de las especificaciones y estimaciones que permitieron categorizar los modelos principalmente en función de la estimación de la raíz del error cuadrático medio, el cual fue calculado por medio de la comparación entre los valores pronosticados y reales.

Tabla 3. Comparación de modelos SARIMA previo a la época de pandemia COVID-19

Vuelos Nacionales				
Modelo	RMSEA	AIC	BIC	ML
$(0,1,1)(0,1,0)_{12}$	0,036	-515.19	-508.95	259.59
$(0,1,0)(0,1,0)_{12}$	0,040	-514.6	-511.48	258,30
$(10,1,1)(0,1,0)_{12}$	0.041	-512.79	-469.88	260.39
$(6,1,2)(0,1,0)_{12}$	0.042	-508.19	-480.13	263.1
$(9,1,1)(0,1,1)_{12}$	0,047	-589.17	-551.75	306.58

Vuelos Internacionales				
Modelo	RMSEA	AIC	BIC	ML
$(0,1,0)(0,1,0)_{12}$	0.031	-574.53	-571.41	288.26
$(11,1,0)(0,1,0)_{12}$	0.037	-587.83	-550.41	305.91
$(8,1,1)(0,1,0)_{12}$	0.038	-589.81	-558.63	304.91
$(9,1,0)(0,1,0)_{12}$	0.038	-590.68	-559.49	305.34
$(4,1,1)(0,1,0)_{12}$	0.038	-593.07	-574.31	302.53

Cuando se considera el efecto de la disminución parcial y total de la cantidad de pasajeros durante la época de la pandemia, en la Tabla 4 se estiman los modelos de acuerdo a los coeficientes que presentan el mejor ajuste sobre la serie de tiempo, y en donde se evidencia la significancia estadística del efecto de la disminución de la cantidad de pasajeros. Para vuelos nacionales se especifica un modelo SARIMA  $(1,1,5)(1,1,0)_{12}$  y para vuelos internacionales SARIMA  $(6,1,1)(0,1,1)_{12}$  con un error medio calculado de 0.041 y 0.071 respectivamente. Como se puede apreciar, el efecto de la disminución brusca de las cantidades de pasajeros provoca que las estimaciones sean calculadas con un mayor error. De igual forma, se presenta algebraicamente los modelos con las estimaciones de los respectivos parámetros.

Modelo SARIMA para pasajeros en vuelos nacionales:

Modelo SARIMA para pasajeros en vuelos internacionales:

$$(1 - B^{12})\log(1 - B)x_t = (1 - \phi B)(1 - \theta B^{12})(1 + \theta_1 B + \theta_2 B^2 + \theta_3 B^3 + \theta_4 B^4 + \theta_5 B^5) w_t \tag{6}$$

$$\theta_1 = 0.36, \theta_2 = 0.36, \theta_3 = -0.9478, \theta_4 = -0.947 \text{ y } \theta_5 = 1.1716$$

$$(1 - B^{12})\log(1 - B)x_t = (1 - \phi_1 B - \phi_2 B^2 - \phi_3 B^3 - \phi_4 B^4 - \phi_5 B^5 - \phi_6 B^6) (1 - \theta B^{12}) (1 + \theta_1 B)(1 + \theta B)w_t \tag{7}$$

$$\phi_1 = -1.16, \phi_2 = -1.16, \phi_3 = 1.16, \phi_4 = 1.16, \phi_5 = -1.34, \phi_6 = 1.34, \theta_1 = -6.3, \theta = 1.07$$



El efecto de la pandemia COVID-19 se muestra con un efecto más significativo para los vuelos internacionales, ya que dentro del territorio Colombiano se efectuaron vuelos domésticos en operaciones no regulares, mientras que los vuelos con frecuencias internacionales cesaron totalmente por el decreto nacional emitido para las restricciones de movilidad y cierre total de fronteras.

**Tabla 4. Estimación de las series de tiempo con efecto de la pandemia**

Vuelos nacionales			
Parámetro	estimate	std	z-value
cte	0.006	0.012	0.447
phi_1	0.090	0.081	1.118
Phi_1	-0.840	0.053	-15.72
theta_5	-0.453	0.074	-6.102
Efecto COVID	-0.473	0.212	-2.234

Vuelos Internacionales			
Parámetro	estimate	std	z-value
cte	0.006	0.006	0.991
phi_6	-0.169	0.081	-2.076
theta_1	0.159	0.104	1.527
Theta_1	-0.932	0.112	-8.352
Efecto COVID	-0.670	0.183	-3.656

De igual forma, en función de la estimación de los parámetros del resto de modelos, se efectúa una categorización de las diferentes estimaciones obtenidas posterior a la etapa de diagnóstico de modelos, los cuales se adecuaron a criterios de normalidad e independencia.

**Tabla 5. Comparación de modelos SARIMA**

Vuelos Nacionales				
Modelo	RMSEA	AIC	BIC	ML
(1,1,5)(1,1,0) <sub>12</sub>	0.041	262.38	289.07	-125.19
(0,1,8)(1,1,0) <sub>12</sub>	0.043	293.14	299.23	-141.57
(11,1,1)(0,1,1) <sub>12</sub>	0.042	289.48	313.89	-138.74
(12,1,0)(0,1,1) <sub>12</sub>	0.043	284.24	313.88	-137.12
(1,1,13)(1,1,0) <sub>12</sub>	0.042	290.10	322.5	-139.05

Vuelos Internacionales				
Modelo	RMSEA	AIC	BIC	ML
$(6,1,1)(0,1,1)_{12}$	0.071	163.87	198.71	-75.93
$(1,1,6)(1,1,0)_{12}$	0.071	164.05	186.43	-76.02
$(5,1,1)(0,1,1)_{12}$	0.072	167.78	186.43	-77.89
$(13,1,0)(0,1,1)_{12}$	0.074	168.36	193.82	-79.18
$(0,1,13)(1,1,0)_{12}$	0.074	170.12	211.74	-80.06

La Figura 8 muestra los pronósticos realizados sobre la demanda de vuelos nacionales e internacionales, para los cuales no se consideraron el período de la pandemia COVID-19. Las estimaciones de los errores cuadráticos se muestran errores cercanos al 3% con un intervalo al 95% de confianza.

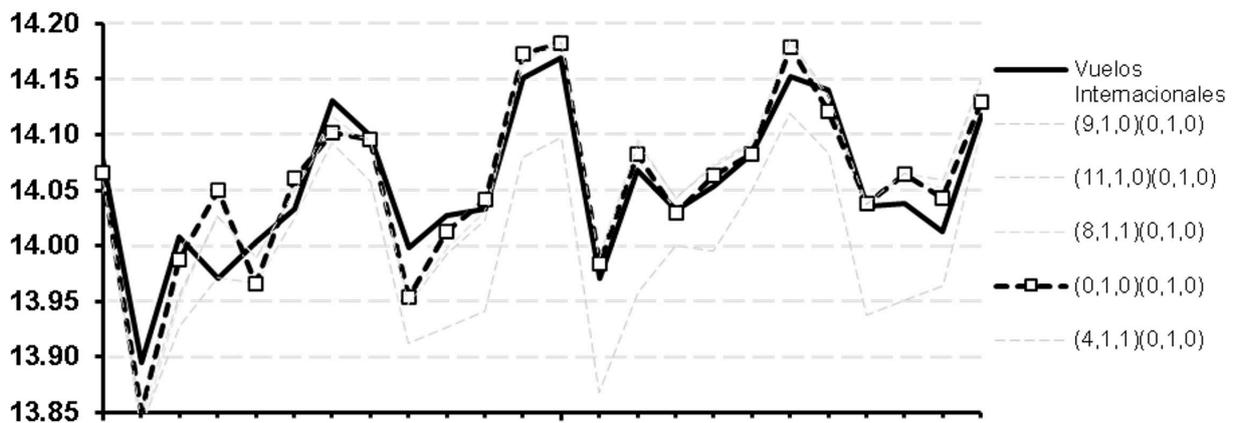
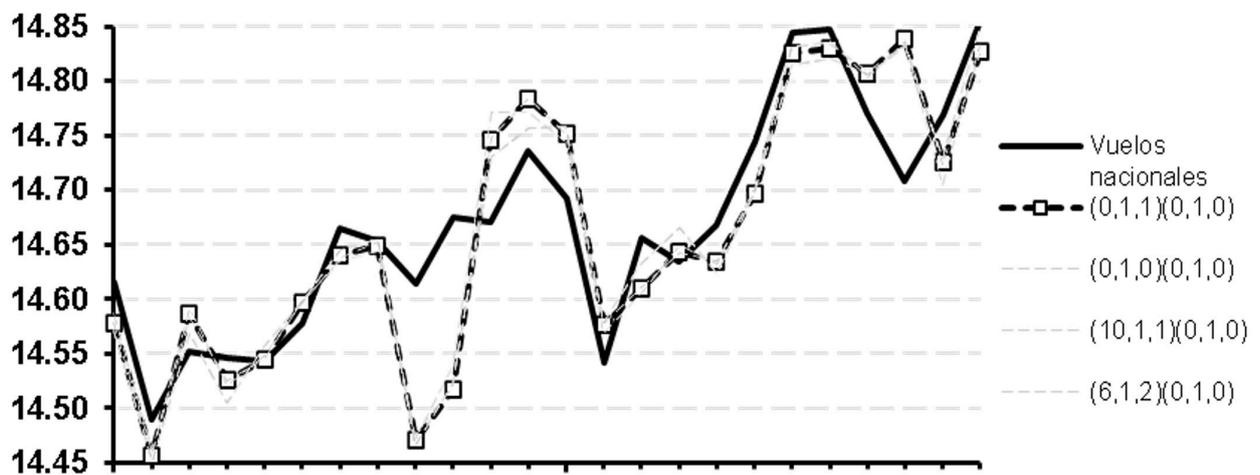


Fig. 8. (a) Pronósticos vuelos nacionales (superior) e internacionales (inferior) sin efecto COVID



La Figura 9 presenta las disminuciones de las cantidades del número de pasajeros por el efecto de la pandemia actual, los valores pronosticados aumentan considerablemente los errores estimados por sobre al 7%. Este efecto es causado debido a las grandes variaciones sobre los datos, y aunque los modelos SARIMA tienen la capacidad de aislar este tipo de efectos sobre los pronósticos efectuados a largo plazo, estos no pudieron ser disminuidos en su totalidad.

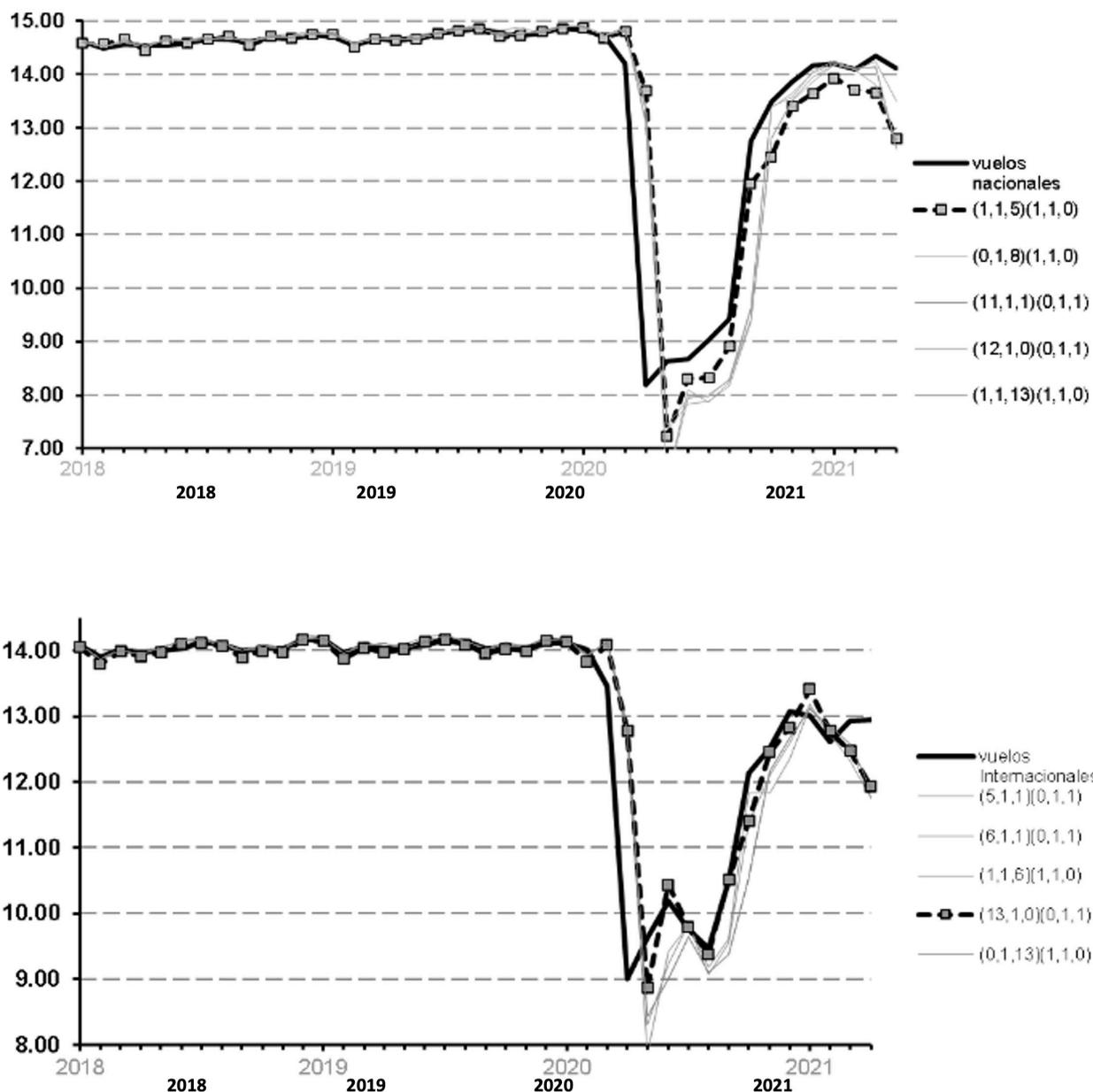


Fig. 9. (a) Pronósticos vuelos nacionales (superior) (b) Pronóstico vuelos internacionales (inferior)

Los modelos de pronóstico predictivo determinado para las series de datos de vuelos nacionales e internacionales presentan un alto rendimiento de estimación. Los valores calculados de la serie de datos, sin la consideración del efecto de la pandemia COVID-19 son comparados con los valores reales del año 2019, previo al período de estado de excepción de movilidad y cierre de fronteras. Los pronósticos realizados sobre la serie de tiempo considerando el efecto COVID-19 son comparados con los últimos valores reportados para los meses del año 2021.

**Tabla 6. Comparación de la demanda de pasajeros reales y pronosticados**

Mes	Vuelos Año 2019				Vuelos Año 2021			
	Nacionales		Internacionales		Nacionales		Internacionales	
	Reales	Pronos-tico	Reales	Pronos-tico	Reales	Pronos-tico	Reales	Pronos-tico
Ene	2,404,914	2,477,485	1,424,646	1,439,779	2,404,914	1,537,441	1,424,646	534,829
Feb	2,067,546	2,080,638	1,167,343	1,178,791	2,067,546	1,299,866	1,167,343	331,309
Mar	2,316,572	2,254,553	1,286,799	1,315,859	2,316,572	990,157	1,286,799	268,566
Abr	2,268,004	2,241,230	1,241,209	1,239,229	2,268,004	348,326	1,241,209	128,199
May	2,343,948	2,329,503	1,269,031	1,276,969				
Jun	2,532,460	2,475,356	1,306,488	1,302,766				
Jul	2,799,156	2,691,735	1,400,969	1,439,779				
Ago	2,806,062	2,727,950	1,384,113	1,355,933				
Sep	2,598,355	2,574,857	1,246,050	1,251,683				
Oct	2,443,470	2,713,894	1,249,684	1,289,803				
Nov	2,594,237	2,504,476	1,218,211	1,251,683				
Dic	2,830,620	2,769,696	1,351,795	1,369,560				

El error cuadrático medio (RMSE) se presentó como una métrica adecuada para expresar la precisión entre los valores pronosticados y los valores reales, lo que muestra los altos rendimientos y capacidad de pronóstico de los modelos SARIMA.

### Conclusiones y trabajos futuros

En este estudio, se estimaron varias especificaciones de modelo de predicción a mediano y largo plazo de orden diferenciado estacional, autoregresivo y de medias móviles (SARIMA). Los modelos utilizados para la predicción de las series de tiempo del número de pasajeros en vuelos nacionales e internacionales, muestran un gran rendimiento alcanzado con errores estimados alrededor del 3%. Se ajustaron modelos utilizando los datos registrados del número de pasajeros para el período durante la aparición de la pandemia del COVID-19, y a pesar de la amplia variación presente en los datos, se pudo obtener un ajuste adecuado sobre parámetros por medio de la definición de un modelo estimado con un error del 7%.

La metodología aplicada secuencial para el análisis de las series temporales bajo el procedimiento de los modelos de predicción SARIMA facilitan comprender la estructura de los datos por medio de la aplicación de criterios empíricos y estadísticos, los cuales permiten realizar una categorización de las especificaciones definidas.

Los resultados obtenidos muestran que los modelos SARIMA son una herramienta adecuada para la estimación de pronósticos de mediano a largo plazo y válidos para la toma de decisiones, incluso cuando se presentan períodos específicos con grandes variaciones sobre la serie.

En futuras propuestas se debería comparar los valores entre los pronósticos alcanzados y los valores registrados para el período del año 2021, además comparar el rendimiento y potencia de predicción que podría alcanzarse con nuevas herramientas basadas en redes neuronales y computacionales.



## Referencias

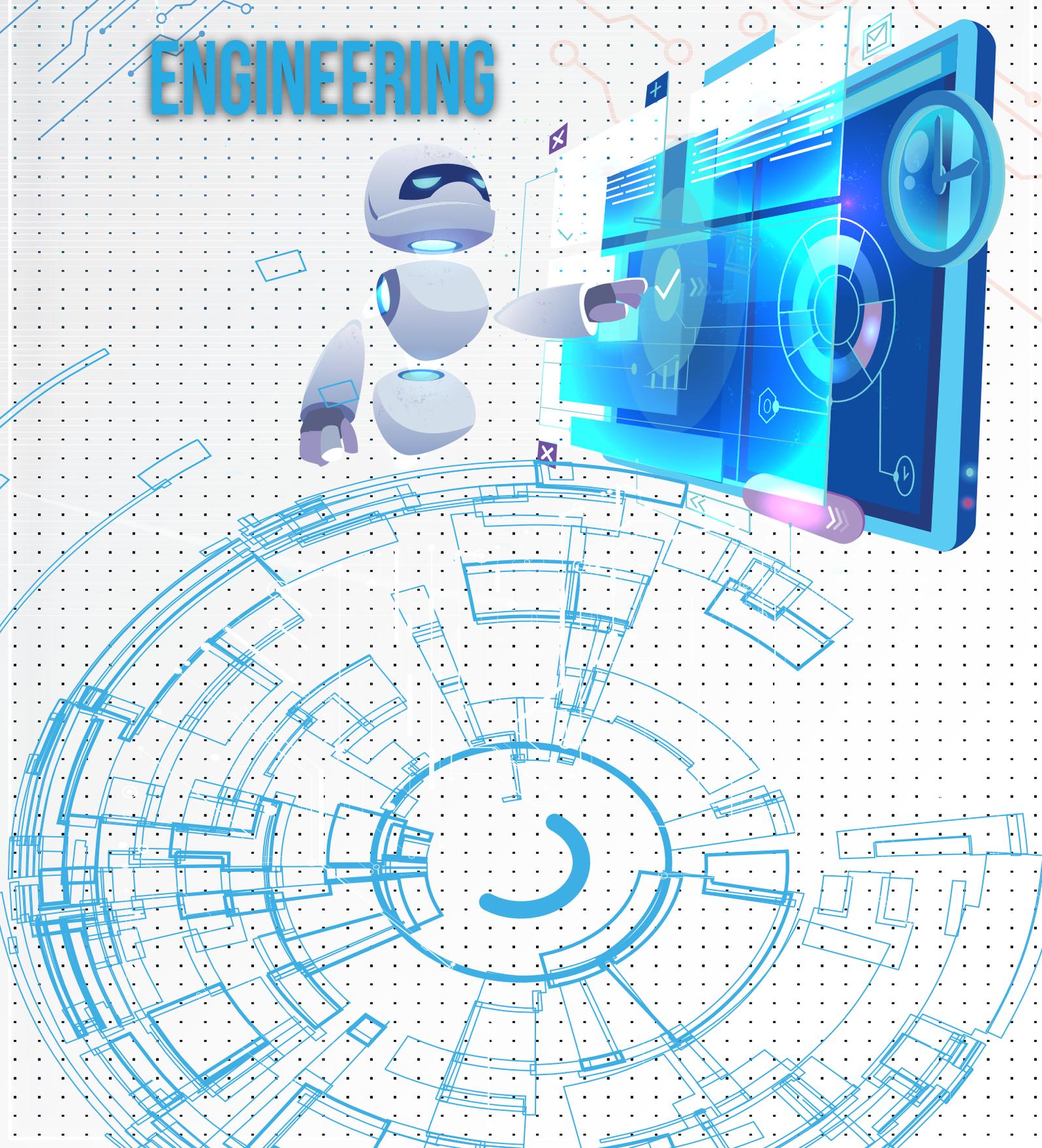
- Banco Mundial. Airport Economics in Latin America and the Caribbean. Serebrisky T. (Eds): The World Bank. Washington DC. (2012).
- Olariaga OD.; Carvajal AF.: Efectos de la liberalización en la geografía del transporte aéreo en Colombia. Transporte y Territorio Journal, pp.344–64 (2016)
- Keeling DJ.: Transport research challenges in Latin America. Journal of Transport and Geography, pp.103–104. (2012)
- Departamento Nacional de Planeación.: Plan Nacional de Desarrollo. (2014)
- Aerocivil.: Boletín estadístico de Oferta y Demanda Pasajeros Operación regular. (2006)
- Chatfield C.: The Analysis of Time Series: An Introduction. Chapman & Hall. London. (2003)
- Wang WC.; Chau KW.; Cheng CT.; Qiu L.: A comparison of performance of several artificial intelligence methods for forecasting monthly discharge time series. Journal of Hydrology. pp. 294–306.
- Holt CC.: Forecasting seasonals and trends by exponentially weighted moving averages. International Journal of Forecasting. pp.5–10. (2004).
- Durban J.; Koopman SJ.; Time series Analysis by state space methods.: Oxford University Press; Oxford. (2012)
- Winters PR.; Forecasting Sales by Exponentially Weighted Moving Averages. Management Science pp. 324–342. (1960)
- Box G.; Jenkins G.; Gregry R.; Ljung G.: Times Series Analysis: Forecasting and Control. Holden Bay. San Francisco, (2015)
- Grubb H, Mason A. Long lead-time forecasting of UK air passengers by Holt-Winters methods with damped trend. International Journal of Forecasting. pp. 71–82. (2001).
- Box G.; Jenkins G.; Reinsel G.; Ljung G.: Time Series Analysis. Wiley. (2016).
- Wold H.: A Study in Analysis of Stationary Time Series. Journal of the royal statistical society. pp.295–298. (1939)
- Slutzky E.: The Summation of Random Causes as the Source of Cyclic Processes. Econometrica; pp.105-146. (1937)
- Yule GU.: Why do we Sometimes get Nonsense-Correlations between Time-Series?. A Study in Sampling and the Nature of Time-Series. Journal of the Royal Statistical Society, pp . 1-63. (1926)
- Aerocivil.: La Aviación en Cifras Edición 2020. Bogotá-Colombia: (2020)
- Ministerio de Transporte.: Transporte en Cifras 2014. (2014)

Phillips PCB.: Time series regression with a unit root and infinite-variance errors. *Economic Theory*. pp 44–62. (1990)

Enders W.: *Applied Econometric Time Series*. Four. New York: (2014)

Harris RID.: Testing for unit roots using the augmented Dickey-Fuller test. Some issues relating to the size, power and the lag structure of the test. *Economic Letters*. pp. 38:381. (1992)

# INFORMATION AND ELECTRONIC ENGINEERING





JOURNAL

IDEAS