

MODELOS ARIMA PARA EL ESTUDIO DE SERIES TEMPORALES DE CONTAMINACIÓN DEL AIRE

Graciela Noemí Avila¹, Julio Cesar Arroyo¹

¹Departamento de Química, Facultad de Ciencias Exactas, Universidad Nacional de Salta.

²Consejo de Investigación Universidad Nacional de Salta (CIUNSA)

Av. Bolivia 5150, Salta, CP 4400, Salta. (+54) (387) 4255354.

E-mail: graciela.avila@exa.unsa.edu.ar, julioarroyo@exa.unsa.edu.ar

Recibido 06/2025; Aceptado 06/2025

RESUMEN.- Se analizan las series temporales de concentración de óxidos de nitrógeno, NO_x, registrada mensualmente entre 2014 y 2019 en la capa atmosférica límite de la ciudad de Salta en 15 sitios de monitoreo. El objetivo es caracterizar la evolución temporal de la contaminación y generar pronósticos usando modelos ARIMA. La metodología de Box y Jenkins se implementa en tres etapas: identificación, estimación y validación. La identificación implica analizar la tendencia, heterocedasticidad, estacionalidad y estacionariedad de las series. La estimación consiste en ajustar modelos ARIMA(p,d,q)xARIMA(P,D,Q)s a las series, considerando la presencia de estacionalidad. Finalmente, la validación compara los valores predichos por el modelo con los datos experimentales. Los resultados revelan que la complejidad de los modelos ARIMA para NO_x se relaciona adecuadamente con la magnitud de la contaminación. Modelos simples como ARIMA(0,0,0)x(0,1,1)₁₂, se ajustan satisfactoriamente a las series de NO_x de sitios con baja contaminación, mientras que los sitios con mayor contaminación, más cercanos al centro, requieren modelos más complejos.

En conclusión, los modelos ARIMA son herramientas útiles para analizar y predecir la contaminación atmosférica. La elección del modelo adecuado depende de las características específicas de cada serie.

Palabras clave: contaminación del aire, óxidos de nitrógeno, modelización de series temporales

ARIMA MODELS FOR THE STUDY OF AIR POLLUTION TIME SERIES

ABSTRACT.- Time series of nitrogen oxides, NO_x, monthly concentrations between 2014 and 2019 in the boundary layer of air in the city of Salta are analyzed at 15 monitoring sites. The objective is to characterize the temporal evolution of pollution and generate forecasts using ARIMA models. The Box and Jenkins methodology is implemented in three stages: identification, estimation and validation. Identification involves analyzing the trend, heteroscedasticity, seasonality and stationarity of the series. Estimation consists of fitting ARIMA(p,d,q)xARIMA(P,D,Q)s models to the series, considering the presence of seasonality. Finally, validation compares the values predicted by the model with the experimental data. The results reveal that the complexity of the ARIMA models for NO_x is adequately related to the magnitude of pollution. Simple models such as ARIMA(0,0,0)x(0,1,1)₁₂ fit well to NO_x series of sites with low pollution, while sites with higher pollution, require more complex models.

In conclusion, ARIMA models are useful tools for analyzing and predicting air pollution. The choice of the appropriate model depends on the specific characteristics of each series.

Keywords: Air pollution, nitrogen oxides, time series modelling, ARIMA



Esta obra está bajo Creative Commons Reconocimiento-NoComercial-CompartirIgual 4.0 Internacional.
<https://creativecommons.org/licenses/by-nc-sa/4.0/>

1. INTRODUCCIÓN

La infraestructura, el microclima, la topografía y tipos de fuentes de emisión de contaminantes de un núcleo urbano pueden condicionar los niveles de concentraciones de los agentes contaminantes del aire en distintas áreas. Para el estudio de contaminación a nivel troposférico en diferentes sitios es necesario recurrir a métodos estadísticos que permiten analizar los datos de fenómenos de contaminación atmosférica. La zona de este estudio corresponde al área geográfica de la ciudad de Salta capital de la Provincia de Salta de la República Argentina. Entre enero de 2014 y diciembre de 2019 se monitorearon mensualmente, en forma pasiva, 15 sitios de muestreo como lo muestra el mapa de la Figura 1. Donde se especifican las distancias desde el centro de la ciudad de los sitios 1,2 (periferia Norte), 13 y 14 (periferia Sur)

Los sitios de muestreo más alejados del microcentro de la ciudad son: 1, 2, 8, 12 y 15. Los sitios ubicados en el centro de la ciudad de Salta son: 3, 4, 5, 6, 7, 8, 9, 10, 11 y 12.

El registro sistemático de contaminantes atmosféricos no sólo sirve para establecer la calidad del aire (Avila, et al., 2021) de un área sino también permite obtener modelos de las series de tiempo constituidas por los datos con la finalidad de caracterizar su evolución, lo que ulteriormente permitirá formular pronósticos. Una serie temporal es una secuencia ordenada de observaciones $\{y_i\} = \{y_1, y_2, \dots, y_n\}$, obtenidas a intervalos regulares de tiempo (anual, trimestral, mensual, diaria, etc.), donde el subíndice especifica el número de orden del intervalo considerado.

En la literatura se puede encontrar una gran variedad de modelos de series de tiempo de datos de contaminación ambiental entre los cuales los más sencillos son los modelos de Box y Jenkins conocidos como modelos ARIMA. Esta metodología fue empleada, por ejemplo, por Zickus para la descripción y pronóstico a corto plazo de la contaminación atmosférica en la ciudad de Vilnius (Zickus, 1999) y por Xie-Kang Wang para modelar el índice de calidad del aire de Hong Kong (Xie-Kang, et al., 2006), Tricio et al. usaron el método de promedios móviles para calcular índices



Fig. 1: Sitios de monitoreo en la ciudad de Salta que conforman cuatro zonas de diferente nivel de contaminación con NOx

de estacionalidad de datos de NO y ozono (Tricio Gómez, et al., 2009). García et al., utilizaron la metodología ARIMA para modelar concentraciones de ozono en la ciudad de México (García, A., et al., 2000).

1.1. Modelos para series de tiempo

La construcción de un modelo de una serie de tiempo sirve para explicar la estructura de la variable estudiada y prever su evolución. La variable de interés puede pertenecer a diferentes campos de la ciencia, una variable macroeconómica, meteorológica, ambiental, etc. En general se desea llegar a un modelo que permita predecir con cierta probabilidad que el valor de la variable de una fecha futura caerá dentro de ciertos límites. Este tipo de modelo se llama modelo de probabilidad o modelo estocástico. El modelo o proceso (estocástico) en realidad representa una población infinita de series de tiempo, de las cuales la que originó el modelo es solamente una realización. Los modelos estocásticos pueden ser estacionarios o no estacionarios. Los modelos estacionarios suponen que el proceso mantiene una estructura constante para la media, la varianza y la autocorrelación, sin tendencia ni fluctuaciones periódicas.

1.2. Fundamentos de la modelización ARIMA

Los modelos autorregresivos integrados de media móvil o ARIMA (Autoregressive Integrated Moving Average) introducidos por Box y Jenkins (Box, G. E. P. et al., 1976) en 1976 extienden el modelo ARMA, que solo es adecuado para series de tiempo estacionarias, también a series no estacionarias. Consiste en aplicar el modelo ARMA no directamente a los datos de la serie, sino a la serie que se obtiene por diferenciación de los datos (diferencias entre datos consecutivos) (Brockwell, P. J. et al., 1991).

El instrumento fundamental para analizar las propiedades de una serie temporal en términos de la interrelación temporal de sus observaciones es el denominado *coeficiente de autocorrelación*, que mide la correlación, es decir, el grado de asociación lineal que existe entre observaciones separadas por k periodos. Estos coeficientes proporcionan información sobre cómo están relacionadas entre sí las distintas observaciones de una serie temporal, lo que ayuda a construir el modelo apropiado para los datos.

La teoría clásica considera que una serie de tiempo está formada por cuatro componentes teóricas: de **tendencia**, **variaciones estacionales**, **variaciones cíclicas** y **componente residual**.

La **tendencia** está dada por el movimiento general a largo plazo de la serie. De acuerdo a los valores de la serie temporal se pueden usar varios métodos para estudiar su tendencia.

Las **variaciones estacionales** son oscilaciones que se producen con un período igual o inferior a un año, y que se reproducen de manera reconocible en los diferentes años. El motivo principal que induce a estudiar la componente estacional es que en la inmensa mayoría de las series temporales esta componente distorsiona su verdadero movimiento. En algunos casos es necesario eliminar estas distorsiones para captar el movimiento real de la serie.

Las **variaciones cíclicas** son oscilaciones que se producen con un período superior a un año y que se deben, principalmente, a la alternancia de etapas largas (ciclos) en las que se repite el comportamiento de la serie. Es la más difícil de detectar, pues a diferencia de la tendencia, que es un movimiento a largo pla-

zo muy general, y de las variaciones estacionales, que tienen un período fijo, estas variaciones tienen un período no fácilmente identificable que además en muchos casos es variable, incluso siendo frecuente la existencia de ciclos que se superponen, lo que hace todavía más difícil su identificación. En la práctica suele eliminarse la tendencia y las variaciones estacionales y estudiar la parte restante de la serie para poder detectar las eventuales variaciones cíclicas.

El **componente residual** en la modelización de Box y Jenkins surge de considerar que el valor observado de una serie en un momento determinado de tiempo t es una realización de una variable aleatoria y_t , definida en un momento de tiempo determinado. Por lo tanto una serie de datos es una muestra de un vector de t variables aleatorias ordenadas en el tiempo al que se denomina proceso estocástico. En ocasiones se pretende predecir el comportamiento de una variable y_t en un momento futuro t , a partir del comportamiento que la variable tuvo en un momento pasado, por ejemplo, y_{t-1} . La notación formal correspondiente es: $y_t = f(y_{t-1})$, es decir, que el valor de la variable en el momento t es función del valor tomado en el período anterior $t-1$. Dado que en el comportamiento de una variable influyen varios aspectos, se debe incluir en la relación anterior un término de error o residuo ε_t , que es una variable aleatoria con media cero y varianza constante. Como consecuencia la variable depende de su valor anterior y de una componente residual.

1.3. Procedimiento usado para identificar modelos ARIMA para las series temporales estudiadas

La metodología de los modelos de series temporales ARIMA, según el enfoque sugerido por Box y Jenkins, consta de tres fases o pasos principales: identificación, estimación y validación. A continuación se describen brevemente los pasos seguidos en este estudio.

1.3.1. Etapa de identificación de la estructura de las series

Para llegar a la fase de identificación en primer lugar se representaron gráficamente las series, su función de autocorrelación (FAC) y su función de autocorrelación parcial (FAP) con el fin de inferir la eventual estacionalidad y estacionariedad de la serie.

1.3.1.1. Tendencia

Se aplicó el Test de Daniel o de Correlación por Rangos de Spearman a las 15 series de 72 datos de concentración de NO_x para detectar la eventual existencia de **tendencia**. Los resultados están resumidos en la Tabla 1. Sólo para los Sitios 8, 11 y 15 se detectó existencia de tendencia. En los tres sitios la tendencia es negativa, lo que indica una disminución sustancial de la contaminación con NO_x .

Desde el punto de vista de la modelación estos resultados implican la necesidad de diferenciar las series en estos sitios; en los tres casos una sola diferenciación fue suficiente.

1.3.1.2. Heterocedasticidad

Se aplicó la prueba de contraste de varianzas de Levene para detectar la heterocedasticidad de las series. Los resultados figuran en la Tabla 2. Las series con heterocedasticidad, o sea en las que la varianza no es constante a lo largo de la serie (Sitios 5, 8, 11 y 12), se transformaron tomando el logaritmo y se usó la mediana como medida de tendencia central para la prueba de Levene, con lo que resultaron **homocedásticas**.

Tabla 1: Valores de los estadísticos resultantes de la aplicación del test de Daniel a las series de tiempo de 72 datos de concentración de NO_x en los 15 sitios monitoreados - Nivel de significación 0,05

Sitio	R Spearman	Percentil Z (tabla)	p-nivel	Tendencia	Sitio	R Spearman	Percentil Z (tabla)	p-nivel	Tendencia
1	0,083	0,697	0,487	No existe	9	0,072	0,610	0,543	No existe
2	0,051	0,432	0,666	No existe	10	0,065	0,545	0,587	No existe
3	0,024	0,201	0,841	No existe	11	-0,432	-4,014	0,0001	Existe
4	-0,008	-0,070	0,944	No existe	12	-0,098	-0,825	0,411	No existe
5	-0,127	-1,073	0,286	No existe	13	0,046	0,391	0,696	No existe
6	0,122	1,032	0,305	No existe	14	0,060	0,508	0,612	No existe
7	-0,060	-0,511	0,610	No existe	15	-0,272	-2,366	0,0207	Existe
8	-0,538	-5,353	0,000001	Existe					

Tabla 2: Resultados del estudio de heterocedasticidad de las series de tiempo de concentración de NO_x en los 15 sitios de muestreo aplicando la prueba de Levene al 95% de seguridad

Sitio	F crítico = 2,354		
	F observada	p-valor	Homocedástico
1	0,1	0,961	SÍ
2	0,844	0,524	SÍ
3	0,299	0,912	SÍ
4	0,363	0,872	SÍ
5	1,688	0,15	SÍ
6	1,14	0,348	SÍ
7	1,306	0,272	SÍ
8	1,87	0,112	SÍ
9	1,197	0,321	SÍ
10	1,442	0,221	SÍ
11	1,933	0,101	SÍ
12	1,933	0,101	SÍ
13	1,292	0,278	SÍ
14	0,95	0,455	SÍ
15	0,658	0,656	SÍ

1.3.1.3. Estacionalidad

Se usó la prueba no paramétrica de Kruskal-Wallis para detectar la estacionalidad de las series. La Tabla 3 resume los resultados. A pesar de que todos los sitios presentan un patrón que se repite anualmente, los resultados indican que en los Sitios 2, 7 y 8 la estacionalidad no es estadísticamente significativa.

1.3.1.4. Estacionariedad

Se usaron las pruebas Test de Dickey-Fuller (DF) y Test de Dickey-Fuller Aumentado (ADF), conocidas como “de raíces unitarias”, para detectar la estacionariedad de las series.

Estas pruebas también indican el orden de diferenciación necesario para lograr la estacionariedad. Los resultados se presentan en la Tabla 4. Solamente seis series pueden considerarse estacionarias (1, 2, 7, 11, 14, 15); los demás no tienen una media estable

Tabla 3: Resultados del estudio de Kruskal – Wallis de estacionalidad de las series de tiempo de concentración de NO_x en los 15 sitios de muestreo

Sitio	Prueba no paramétrica de estacionalidad con estabilidad 1%
1	Presente
2	Ausente
3	Presente
4	Presente
5	Presente
6	Presente
7	Ausente
8	Ausente
9	Presente
10	Presente
11	Presente
12	Presente
13	Presente
14	Presente
15	Presente

Tabla 4: Resultados del estudio de estacionariedad de las series de tiempo de 72 datos de concentración de NO_x en los 15 sitios de muestreo - H₀: la serie es no estacionaria

Sitio	t calculado	DF _{0,05} = -2.89	Conclusión
1	-5,345	H ₀ rechazada	Es estacionaria
2	-2930	H ₀ rechazada	Es estacionaria
3	-2124	H ₀ no rechazada	Tiene raíz unitaria
4	-2039	H ₀ no rechazada	Tiene raíz unitaria
5	-1488	H ₀ no rechazada	Tiene raíz unitaria
6	-2044	H ₀ no rechazada	Tiene raíz unitaria
7	-4522	H ₀ rechazada	Es estacionaria
8	-2144	H ₀ no rechazada	Tiene raíz unitaria
9	-2257	H ₀ no rechazada	Tiene raíz
10	-2284	H ₀ no rechazada	Tiene raíz
11	-5701	H ₀ rechazada	Es estacionaria
12	-1787	H ₀ no rechazada	Tiene raíz
13	-2322	H ₀ no rechazada	Tiene raíz
14	-5886	H ₀ rechazada	Es estacionaria
15	-5952	H ₀ rechazada	Es estacionaria

Tabla 5: Resultados de los estudios de caracterización de las series de tiempo de concentración de NO_x en los 15 sitios de muestreo

Sitio	Tendencia	Heterocedasticidad	Estacionalidad	Estacionariedad
1	-	-	+	+
2	-	-	-	+
3	-	-	+	-
4	-	-	+	-
5	-	-	+	-
6	-	-	+	-
7	-	-	-	+
8	+	-	-	-
9	-	-	+	-
10	-	-	+	-
11	+	-	+	+
12	-	-	+	-
13	-	-	+	-
14	-	-	+	+
15	+	-	+	+

y por lo tanto requirieron una diferenciación adecuada para su modelación por la metodología ARIMA.

En la Tabla 5 se reunieron las características estructurales de las series que surgen de los estudios anteriores.

1.5. Etapa de estimación y validación del modelo

1.5.1. Estimación del modelo

Para la estimación del modelo se usaron 66 datos dejando los último 6 para la validación y pronóstico. Observando las gráficas de la serie, su función de autocorrelación y de autocorrelación parcial se puede tener una idea del modelo que podría ajustarse a la serie, o cuáles serían los candidatos que se deben probar. (Software XLSTAT for Excel. Addinsoft., Software Gretl. Versión 1.9.9).

Cuando se concluye que la serie presenta estacionalidad significativa se ajusta un modelo ARIMA(p,d,q) donde el valor de «p» es el orden de la parte AR, «d» es el número de diferencias regulares necesarias para hacer la serie estacionaria y «q» el número de orden de la parte MA. Frecuentemente se deben probar varios modelos ARIMA(p,d,q) hasta que se encuentre el modelo óptimo. Se debe estudiar si sus términos son necesarios y si los residuos que quedan son ruido. Cuando el orden de una parte, AR ó MA, es muy alto, se puede ensayar añadir algún término de la otra parte y restarle orden a ésta.

1.5.2. Pruebas de significación

Para evaluar la significación de los parámetros del modelo elegido se usó el criterio de si el intervalo de confianza al 95% no incluye el cero, el parámetro es significativo

El análisis de los residuales se realizó mediante la aplicación de varios contrastes de hipótesis. En el de Box-Pierce, después de ajustar el modelo, el contraste confirmaba que cuando el *p*-valor del contraste era mayor que 0,05 se trataba de ruido blanco, y cuanto mayor era éste, tanto mayor era la evidencia de que los residuos realmente eran ruido blanco y por lo tanto mejor se había captado en el modelo la parte que no se debía al azar.

La observación gráfica, tras ajustar el modelo ARIMA, de las funciones FAS y FAP de los residuos, también proporcionaba una información análoga a la del contraste de aleatoriedad de Box-Pierce.

Cuando se había transformado la serie original, el modelo encontrado era el de la transformación, no el de la serie original.

1.5.3. Elección del modelo óptimo

En general no hay un único modelo correcto que explique suficientemente bien los datos. Existen diversos criterios que se pueden seguir para elegir un modelo satisfactorio.

1.5.3.1. Criterios aplicados para la selección del modelo óptimo

- Los residuos, después de ajustar el modelo, debían ser ruido blanco.
- Entre dos modelos que sólo difieran en el *p*-valor del contraste de **Box-Pierce**, se prefirió el que tenía un **valor mayor**, puesto que daba mayor garantía de que los residuos que proporcionaba el modelo eran ruido.
- Entre los que verificaban más o menos las mismas condiciones numéricas, se seleccionaba el más sencillo posible, es decir, el que necesitaba **menos parámetros**. Los parámetros que necesita un modelo ARMA(p,q) son «p+q».
- En último término, la elección del modelo final dependía de la parsimonia, un equilibrio entre simplicidad y eficacia, dado que siempre sería posible añadir parámetros indefinidamente para obtener mejoras.

1.5.3.2. Estadísticos de evaluación de bondad de ajuste

Existen numerosos cálculos que permiten evaluar la bondad del ajuste de un modelo a la serie de datos estudiados. Entre ellos se utilizaron los siguientes:

- SEC (sumatoria de errores cuadráticos) es una evaluación de la discrepancia entre cada uno de los valores generados por el modelo y el correspondiente de la serie que se modela (original o transformada

mediante logaritimación); MAPE y MAPE (Dif) es el error porcentual medio (valor absoluto) en el caso de series sin y con estacionalidad, especialmente útiles para la comparación de modelos; la varianza RB (del ruido blanco) es la suma de errores cuadráticos relativa al número total de datos, de modo que permite la comparación de modelos basados en distinto número total de datos; -2LogVerosim : entre dos modelos para una misma serie se elige aquel que tenga el menor valor de este estadístico, lo cual equivale a elegir el de mayor probabilidad de ajuste (el más veraz); FPE (función de error de predicción de Akaike); permite comparar la indeterminación de la componente autorregresiva de los modelos; AIC y SBC (criterio de información de Akaike y de Schwartz – Bayes) indican la bondad del ajuste del modelo por ser directamente proporcionales al error de predicción y son funciones de los grados de libertad asociados al modelo. Los modelos más robustos son los que pierden menor número de grados de libertad en el proceso de ajuste.

1.5.4. Validación

Una vez construido uno o más modelos para representar una serie de tiempo dada, se obtuvieron pronósticos a partir de él o ellos, para analizar y comparar la capacidad de pronóstico de cada uno de los modelos en consideración. Para ello se comparaban con los

datos experimentales los valores predichos por el modelo para el último segmento de la serie.

La discrepancia entre los valores pronosticados y las observaciones (los errores de pronóstico) permitían distinguir la existencia de sub o sobre estimación en los pronósticos mediante el cálculo de la media de los errores de pronóstico. Al disponer de datos fuera del conjunto de los utilizados para el modelado, se comparan éstos con los valores sintéticos que surgen del modelo.

2. MODELOS ARIMA DE LA EVOLUCIÓN DE LA CONTAMINACIÓN CON NO_x

Los Sitios 1, 2, 12, 13 y 14 son los más alejados del centro comercial de la ciudad, al Norte y al Sur, y todos ellos se caracterizan por ser homocedásticas sin tendencia, pero con estacionalidad, por lo que algunos aparecen como carecientes de estacionariedad (Sitios 12 y 13).

Los gráficos de correlación de estas series de un retardo $y(t)$ vs $y(t-1)$ y de doce retardos $y(t)$ vs $y(t-12)$ presentan correlación positiva, de modo que con toda seguridad el modelo debe contener componentes SMA (Simple Moving Average). Los gráficos de FAC de estas series no decaen (o solo muy lentamente)

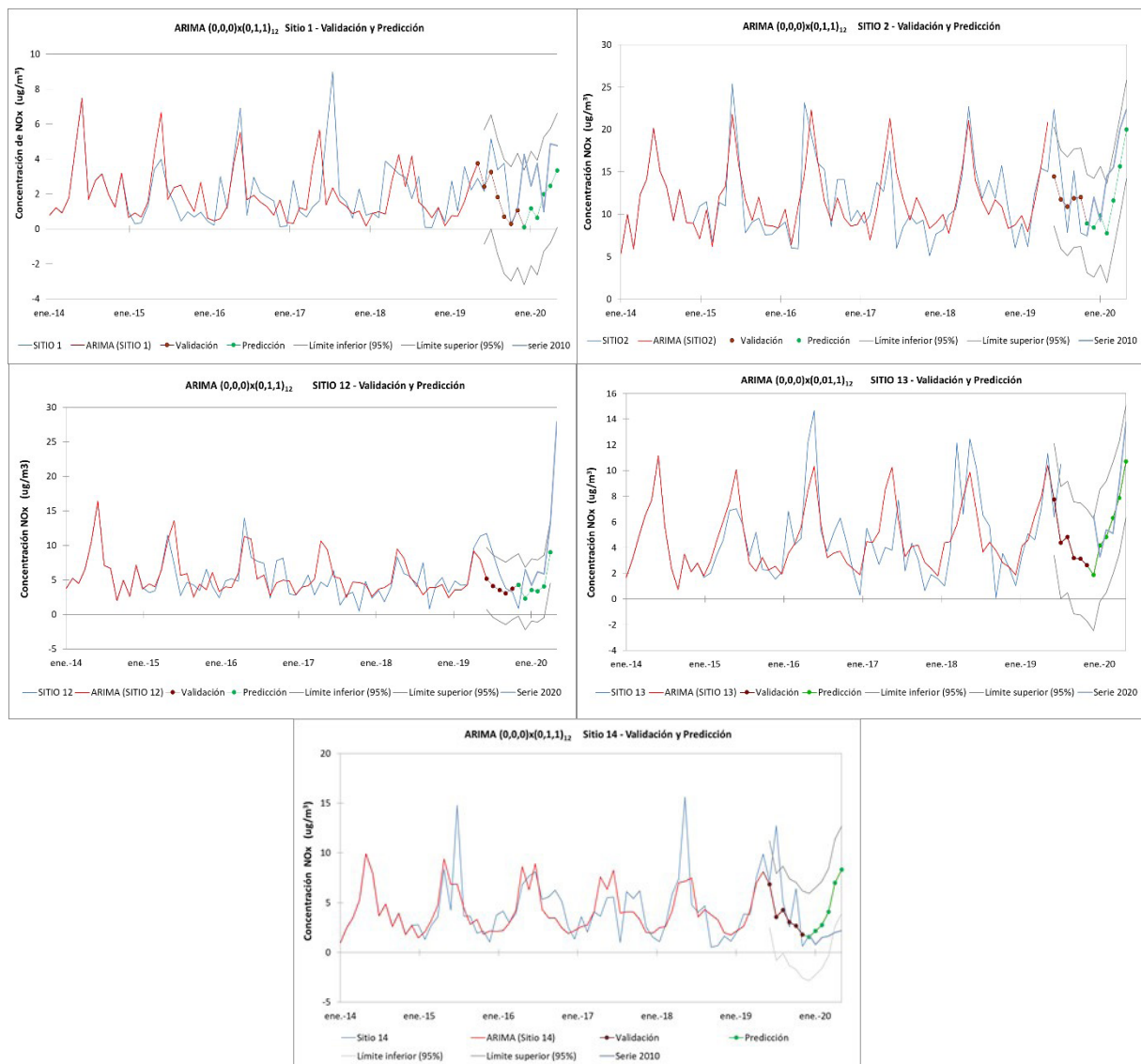


Fig. 2: Gráficos de las series temporales de NO_x y sus modelos de los Sitios de muestreo 1, 2, 12, 13 y 14.

a cero, indicando que en realidad no son estacionarias, con el primer coeficiente fuera de la banda de confianza del 95%, lo que indicaría un modelo con componente MA(1). En todos los casos los gráficos de FACP indican modelo AR(1), decayendo a cero.

La desestacionalización en todos los casos produjo FAC con todos los coeficientes dentro de la franja de confianza, indicando que el modelo no necesita componente MA. En consecuencia para todas ellas se ensayó el modelo ARIMA(0,0,0)x(0,1,1)₁₂. Estos modelos presentan coeficientes significativos, sus coeficientes de criterio se pueden considerar buenos y los residuos en todos los casos son de distribución normal y de tipo ruido blanco, lo que significa que los modelos dan cuenta correctamente del movimiento temporal de los datos. Estos modelos se basan fuertemente en el desarrollo de la serie durante el primer año, en muchos casos siendo idénticos a la serie (Figura 2). El ajuste se vuelve menos preciso para los años posteriores, la validación con los últimos seis meses de 2019 es buena, mientras que el pronóstico en general forma franjas de confianza que contienen los datos de los primeros seis meses de 2010, con lo cual se los puede considerar satisfactorios desde este punto de vista también.

En la tabla 6 se presentan los valores de los estadísticos de evaluación de bondad de ajuste de estos modelos.

En el caso del Sitio 14 el ajuste no es tan bueno como para los otros sitios; se probaron varios otros modelos, de los cuales sobresale el ARIMA(0,0,0)x(1,0,1)₁₂ cuyas características también figuran en la tabla 6. El "software" no produjo banda de confianza alguna para las etapas de validación y pronóstico, pero los datos generados son consistentes con los observados, especialmente en el período 2010. El ajuste se puede comparar con el del otro modelo en la Figura 2.

Aquellos sitios en los cuales se detectó mayor contaminación con NO_x, Sitios 3, 4, 5 y 6, presentan series de tiempo más complicadas. Prácticamente todas tienen estacionalidad marcada, algunos con tendencia e incluso heterocedasticidad, salvo el sitio más contaminado, el Sitio 7. Todas estas series se pudieron representar mediante modelos ARIMA más o menos ajustados a los datos, en general validados, pero no todos con buen pronóstico. En la Tabla 7 se reúnen los coeficientes de criterio junto con los parámetros de los modelos y los gráficos correspondientes se pueden observar en la Figura 3.

Tabla 6: Parámetros y coeficientes de criterio de las series de NO_x de los Sitios de muestreo 1, 2, 12, 13 y 14

ARIMA NO _x (0,0,0)x(0,1,1) ₁₂	Sitio 1	Sitio 2	Sitio 12	Sitio 13	Sitio 14
Observaciones	66	66	66	66	66
SEC	151	476	286	267	269
MAPE(Dif)	173	95	75	270	97
MAPE	103	18	47	100	46
Varianza RB	3	9	5	5	5
-2Log(Verosi.)	217	291	264	260	260
FPE	3	9	5	4,9	5
AIC	221	295	268	264	264
SBC	225	299	272	268	268
Iteraciones	47	32	51	76	28
Constante	-0,165	0,135	-0,231	0,161	-0,090
SMA(1)	-0,722	-1,000	-1,000	-1,000	-1,000

Tabla 7: Parámetros y coeficientes de criterio de las series de NO_x de los Sitios de muestreo 3, 4, 5 y 6.

ARIMA NO _x	Sitio 3	Sitio 4	ln Sitio 5	Sitio 6
Modelo	(1,0,2)x(1,0,0) ₁₂	(1,1,1)x(0,1,1) ₁₂	(2,1,1)x(1,1,1) ₁₂	(1,0,1)x(1,0,1) ₁₂
Observaciones	66	66	66	66
SEC	661	1752	7	163
MAPE(Dif)	246	93	108	17
MAPE	83	65	10	8,5
Varianza RB	10	30	0	2,3
-2Log(Verosi.)	348	390	86	508
FPE	15	31	0	3,3
AIC	380	398	98	562
SBC	415	406	110	623
Iteraciones	66	183	500	72

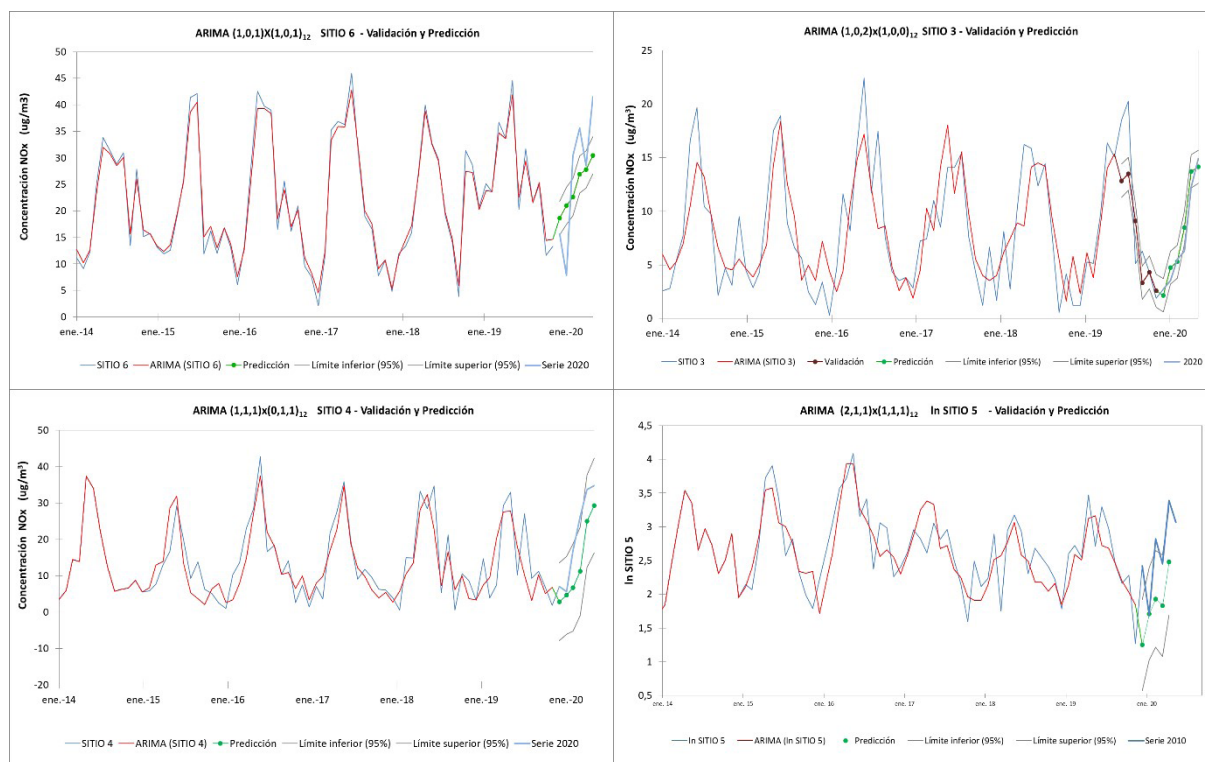


Fig. 3: Gráficos de las series temporales de NOx y sus modelos de los Sitios de muestreo 3, 4, 5 y 6.

Tabla 8: Parámetros y coeficientes de criterio de las series de NOx de los Sitios de muestreo 9, 10, 11 y 15

ARIMA NO _x	ln Sitio 9	ln Sitio 10	ln Sitio 11	ln Sitio 15
Modelo	(1,1,1)x(1,0,1) ₁₂	(3,0,1)x(3,0,2) ₁₂	(3,1,2)x(3,1,2) ₁₂	(1,1,1)x(1,0,1) ₁₂
Observaciones	66	66	66	65
SEC	11	9,18	4,34	8,5
MAPE(Dif)	103	674	198	410,3
MAPE	16	8,37	6,73	10,2
Varianza RB	0,16	0,14	0,08	0,1
-2Log(Verosi.)	85	72	49	70
FPE	0,25	0,54	0,54	0,2
AIC	95	202	71	79,9
SBC	106	344	93	90,7
Iteración	501	66	500	500

Las series de los Sitios 9, 10, 11 y 15 son tan irregulares que, a pesar de ser homocedásticas, se optó por logaritarlas previo a la modelación (Tabla 8). Estas series presentan bruscos descensos de concentración de NOx en el medio de los meses de máxima contaminación y aumentos en los meses de menor contaminación, en especial durante el año 2019.

Estos cambios repentinos pueden considerarse como “volatilidad estocástica”, característica que dificulta el modelado con ARIMA. Los sitios más contaminados, Sitios 10 y 11 necesitaron modelos muy poco parsimoniosos. En la Figura 4 se presentan sus gráficos.

Aquellas series que no poseen estructura, sino características de ruido blanco, como la del Sitio 7 no se prestan al modelado con ARIMA. En el caso del Sitio 8 la existencia de dos etapas diferentes impidió el modelado por poseer insuficientes datos cada etapa.

3. CONCLUSIÓN

La modelación con ARIMA es una herramienta poderosa para el estudio de series temporales de contaminantes de muy variadas características. El estudio previo de estas características es fundamental para lograr modelos parsimoniosos.

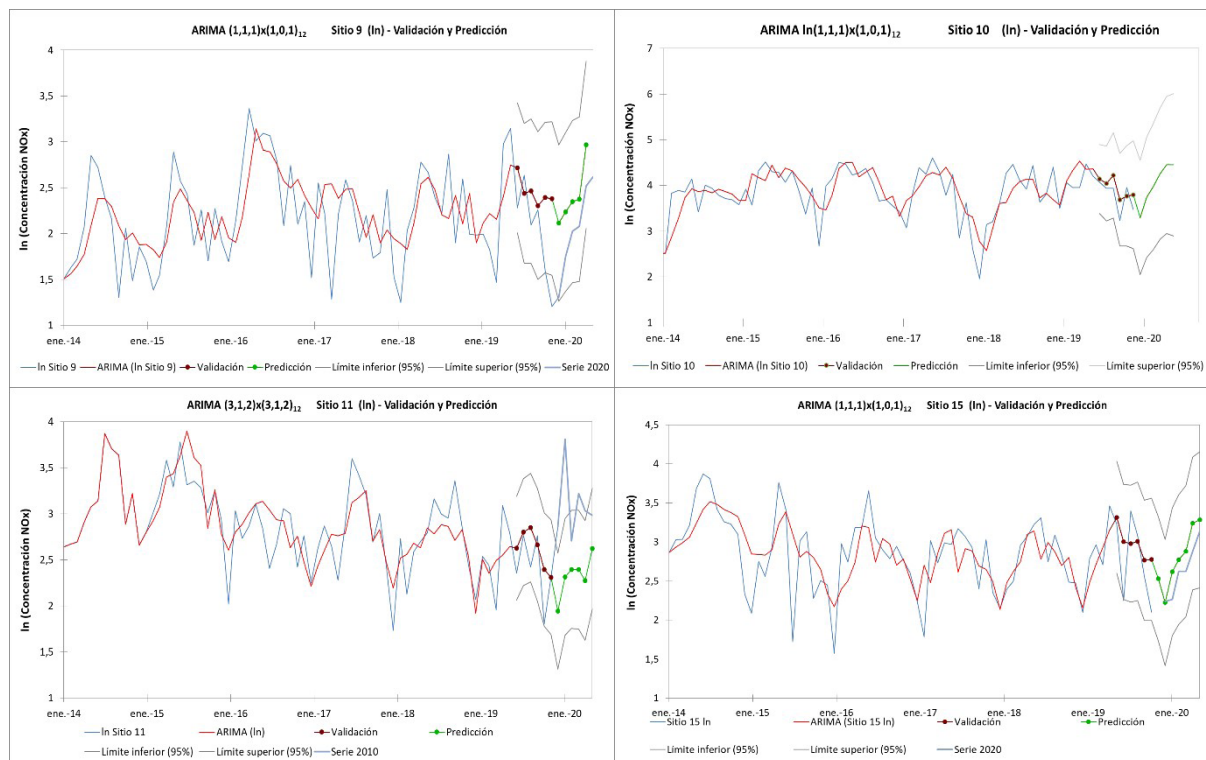


Fig. 4: Gráficos de las series temporales de NOx y sus modelos de los Sitios de muestreo 9, 10, 11 y 15.

La modelación de series temporales de contaminantes es de mucha utilidad para identificar las fuentes de la polución ya sea emisión constante o variable, lo que facilita encontrar las causales de la misma. El método ARIMA es lo suficientemente versátil para los requerimientos de este tipo de estudio. Pero no solamente es necesario reconocer la causa y la ubicación de la contaminación emitida, sino también el pronóstico, que permite evaluar la verdadera gravedad de la polución. Como este trabajo lo demuestra, el método ARIMA provee pronósticos fiables para aquellas series que son estacionarias y por lo tanto permiten ajustar modelos parsimoniosos.

REFERENCIAS

- Box, G. E. P. and Jenkins, G. (1976). *Time Series Analysis: forecasting and control*. United State: Holden Day.
- Brockwell, P. J. and Davis, R. A. (1991). *Time Series: Theory and Methods*. Colorado: Springer Verlag. G. Avila, J. Arroyo. (2019). *Mapa de contaminación de óxidos de nitrógeno en la troposfera baja de salta Capital*. Revista Científica Avances en Energías Renovables y Medio Ambiente (AVERMA). Vol. 55, pp. 01.09-01.16, 2017. ISSN 27968111. XLIII Reunión de Trabajos de la ASADES
- Secretaría del Medio Ambiente, Dirección General de Prevención y Control de la Contaminación. García, A.; Muñoz Cruz, R. y García, G. (2000). *Análisis de ozono y óxidos de nitrógeno mediante técnicas de series de tiempo y regresión lineal*. Jallapa: Autor.
- Software Gretl. Versión 1.9.9. Versión libre disponible en: http://gretl.sourceforge.net/gretl_espanol.html
- Software XLSTAT for Excel. Addinsoft.
- Tricio Gómez, V., Carneiro, G. (2009). Evolución temporal de O₃ y NO en zonas urbanas diferenciadas: Buenos Aires, Argentina-Burgos, España. AIDIS Argentina. Ingeniería Sanitaria y Ambiental, 81. 56-59.
- Xie- Kang Wang, Wei-Zhen Lu. (2006). Seasonal variation of air pollution index: Hon Kong case study. *Chemospher*, 63, 1261-1227.
- Zickus, Mindaugas. (1999). *Influence of Meteorological Parameters On The Urban Air Pollution And Its Forecast*. Internet: Ministry of Environment. Vilnius, Lithuania.