



03043
2
2y

**UNIVERSIDAD NACIONAL AUTONOMA
DE MEXICO**

UNIDAD ACADÉMICA DEL CICLO PROFESIONAL
Y POSGRADO

INSTITUTO DE INVESTIGACIONES EN MATEMÁTICAS
APLICADAS Y EN SISTEMAS

“APLICACION DE MODELOS ARIMA AL OZONO EN
LA ZONA SUROESTE DE LA CIUDAD DE MEXICO”

T E S I S

QUE PARA OBTENER EL DIPLOMA DE
ESPECIALIZACION EN ESTADISTICA
A P L I C A D A
P R E S E N T A :
MONICA ESTRADA HERNANDEZ

DIRECTOR: M EN C. JOSE SALVADOR ZAMORA MUNOZ

JUNIO 1998

26034

**TESIS CON
FALLA DE ORIGEN**



Universidad Nacional
Autónoma de México



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

Cuando se trata de agradecer, siempre hay una larga lista que mencionar, y que sin lugar a dudas, nunca puede estar completa. Con temor a equivocarme y omitir a muchas personas, quiero agradecer de una manera muy especial todo el tiempo y la paciencia que mi querido asesor Salvador Zamora invirtió en mí, en verdad mil gracias.

De igual manera, a Guillermina Eslava, Leticia Gracia, Martín Romero y Patricia Romero por su interés en lograr un mejor trabajo.

Y a todas esas personas que me ayudaron en tantas actividades que este trabajo demandó: a mi muy amado Temo, a mamá, Angie, Archie, Verona, el niño Jaime y a todos aquellos que me apoyaron y me dieron aliento. Gracias.

Finalmente, más que un agradecimiento es una dedicatoria, a mi pequeño hermano para que pegué el brinco que le hace falta.

CONTENIDO

INTRODUCCION

RESUMEN

1. GENERALIDADES	1
1.1 MARCO TEORICO	1
1.2 ANTECEDENTES	2
1.3 AREA DE ESTUDIO	4
1.4 PLANTEAMIENTO DEL PROBLEMA	5
2. SERIES DE TIEMPO	6
2.1 INTRODUCCION	6
2.1.1 FUENTES DE VARIACION	6
2.1.2 ELEMENTOS DE UNA SERIE DE TIEMPO	7
2.2 MODELOS PARA SERIES DE TIEMPO	8
2.2.1 RUIDO BLANCO	8
2.2.2 MODELO DE MEDIAS MOVILES (MA)	9
2.2.3 MODELO AUTORREGRESIVO (AR)	10
2.2.4 MODELOS AUTOREGRESIVOS Y DE MEDIAS MOVILES (ARMA)	11
2.2.5 MODELOS ARIMA	12
2.3 MODELOS PARA SERIES PERIODICAS	13
2.4 PREDICCIÓN CON ARIMA	15
3. CONSTRUCCION DE MODELOS	17
3.1 IDENTIFICACION	18
3.1.1 ESTABILIZACION DE LA VARIANZA	18
3.1.2 ESTABILIZACION DEL NIVEL	20
3.1.3 APLICACION DE FAC Y FACP	21
3.2 ESTIMACION	29
3.3 VERIFICACION	31
4. ANALISIS ESTADISTICO DEL OZONO EN LA ZONA SUROESTE DE LA CIUDAD DE MÉXICO	35
4.1 IDENTIFICACION	37
4.2 ESTIMACION	43
4.3 VERIFICACION	44
4.4 PRONOSTICOS	47
CONCLUSIONES	51
BIBLIOGRAFIA	52

INTRODUCCION

La Ciudad de México desde su fundación ha sido núcleo de los acontecimientos más importantes del país. Desde un principio, la vida política, económica y cultural se concentró en ella, lo cual trajo como consecuencia un crecimiento acelerado y este fenómeno se fue incrementando con los años.

Actualmente, en esta gran urbe se encuentran gran variedad de empresas y fábricas que generan diversos productos y servicios, cuenta con grandes centros recreativos y los más importantes en cuanto a exhibición, difusión, esparcimiento, cultura y educación.

Como resultado de la alta concentración de la actividad económica, se observa un crecimiento muy acelerado de la población, provocando la expansión de la urbe y, con ello, demanda de vivienda y servicios, colocando a la Ciudad de México como el centro de mayor concentración demográfica en el país. Este desarrollo ha generado grandes problemas de contaminación, principalmente de tipo atmosférico, ubicándola entre las ciudades más contaminadas del mundo.

El Area Metropolitana cuenta con una gran variedad de fuentes de contaminación atmosférica: industrias químicas, del cemento, fábricas, fundidoras, baños públicos, incineradores industriales y domésticos, vehículos automotores de combustión interna, aviones, etc. El Departamento del Distrito Federal reporta que estas fuentes emitieron en 1990 aproximadamente 11,000 toneladas de contaminantes al día.

La contaminación generada por la industria y el transporte puede adquirir grandes dimensiones por la ubicación geográfica de las ciudades; la Ciudad de México se encuentra rodeada por complejos montañosos, que en conjunto con su clima dificultan la dispersión de los contaminantes. Además está estrechamente relacionada con la cantidad y calidad de combustibles consumidos, tecnología de combustión y control de emisiones.

Existen algunas sustancias que son utilizadas para evaluar la calidad del aire: partículas suspendidas totales, plomo, monóxido de carbono, óxidos de azufre, óxidos de nitrógeno, hidrocarburos, ozono y otros oxidantes fotoquímicos. A excepción del ozono, todas estas sustancias son conocidas como contaminantes primarios por ser emitidas de manera directa por combustión, procesos industriales, erosión, incendios y algunos otros procesos naturales. El ozono es un contaminante secundario (no es descargado directamente a la atmósfera), pues surge como resultado de reacciones fotoquímicas entre contaminantes primarios, básicamente óxidos de nitrógeno (NOx) e hidrocarburos.

El ozono es un contaminante que permanece mucho tiempo en la atmósfera, es altamente tóxico y muy reactivo. Esta sustancia deteriora materiales y es dañina para la salud humana, cultivos, animales y plantas de ornato. Un adecuado conocimiento de su comportamiento y la posibilidad de predecir adecuadamente sus valores futuros permitirían la implementación de estrategias de manejo y advertencia al público para el ozono, particularmente en áreas densamente pobladas como lo es la Ciudad de México.

RESUMEN

La Ciudad de México es la urbe más importante y grande del país, como consecuencia de su crecimiento desmedido se han desarrollado serios problemas ambientales, principalmente de tipo atmosférico, que la han convertido en la ciudad más contaminada del territorio nacional.

En cuanto a contaminación ambiental, el ozono representa el problema más fuerte para la ciudad, debido a que por lo general registra valores más altos de la norma, es tóxico, muy reactivo y es el que tiene mayor permanencia en la atmósfera. Por ello sería de gran ayuda el poder describir y predecir este fenómeno, para que pudiera ser utilizado como herramienta en el establecimiento de medidas de control.

La manera de registrar este contaminante (secuencia de valores a lo largo del tiempo) permite suponer que la serie refleja sus variaciones en el tiempo, así pues, se aplicaron modelos de series de tiempo ARIMA para ajustar la serie. La zona que se analiza es la región suroeste de la Ciudad de México, en donde, generalmente se presentan los niveles más altos de ozono, siendo considerada como un área crítica.

En la primera sección del trabajo se exponen las generalidades para una mejor visión del problema, describiendo el fenómeno, su relación con la Zona Metropolitana y el por qué se empleará esta técnica. En el siguiente capítulo, se plantean las bases de la técnica de series de tiempo con modelos ARIMA, y se describe brevemente cada uno de los modelos. La tercera parte, explica cuál es el procedimiento para identificar un modelo, estimar sus parámetros y verificar los supuestos.

En la última sección se realiza el ajuste de dos modelos, un AR(2) y uno ARMA $(2,0,0)(0,0,1)_6$. Con ambos modelos propuestos se hicieron predicciones y sus correspondientes intervalos de confianza. Los pronósticos pueden ser una ayuda para comprender el comportamiento de la serie y, quizá, para su control, pero deben de tomarse con cuidado, pues la naturaleza de este fenómeno involucra muchos factores, cuya alteración puede llevar a un comportamiento atípico del ozono.

1 GENERALIDADES

1.1 MARCO TEORICO

La contaminación por oxidantes fotoquímicos es una mezcla de diversos compuestos gaseosos y aerosoles. Típicamente el ozono representa a estos oxidantes, debido a que se encuentra en mayor concentración en la atmósfera y que es de fácil medición; aunque también existen otros oxidantes importantes como el peroxiacetilo (PAN) y el peróxido de hidrógeno (H_2O_2).

Los oxidantes fotoquímicos (entre ellos el ozono) se conocen como contaminantes secundarios pues no son emitidos directamente a la atmósfera, más bien, son formados en el aire como resultado de la acción de la energía ultravioleta proveniente del sol sobre las emisiones de los óxidos de nitrógeno (conocidos como NO_x) y de compuesto orgánicos reactivos. Además, interviene el mezclado atmosférico, las concentraciones y la relación de los precursores y la reactividad de los mismos.

Existe una estrecha dependencia entre la cantidad de oxidantes formados y la hora del día, las condiciones meteorológicas y la naturaleza de las fuentes emisoras de precursores.

Los compuestos conocidos como NO_x tienen su origen en los procesos de combustión a alta temperatura, son producidos por la combustión de energéticos fósiles (gasolina, combustóleo, diesel) en vehículos e industrias. Los compuestos orgánicos reactivos son principalmente hidrocarburos, provenientes de solventes, evaporación de combustibles (gasolina) y combustión incompleta de energéticos que contienen carbono e hidrógeno.

Los óxidos de nitrógeno son 7, pero sólo el óxido nítrico (NO) y el bióxido de nitrógeno (NO_2) se consideran contaminantes importantes; aunque hay otras cuatro formas que también se ven involucradas en las reacciones que generan los oxidantes fotoquímicos. Los compuestos NO_x se requieren para la formación de ozono, una vez que son consumidos la formación de ozono se detiene.

El ozono forma parte de los cinco contaminantes criterio (sustancias empleadas para evaluar la calidad del aire) que se manejan en nuestro país, dichas sustancias son: monóxido de carbono (CO), bióxido de nitrógeno (NO_2), bióxido de azufre (SO_2), partículas suspendidas totales (PST) y ozono (O_3).

Para conocer el comportamiento de éstos y otros contaminantes, en 1984 inició operaciones la Red Automática de Monitoreo Atmosférico (RAMA), la cual cuenta con 32 estaciones de monitoreo (21 ubicadas en el Distrito Federal y 11 en el Estado de México). La información es procesada cada hora durante todo el día, el reporte se hace en forma de IMECA (Índice Metropolitano de la Calidad del Aire). Este índice transforma a una escala numérica los niveles de contaminación de cada uno de los 5

contaminantes criterio del país y el algoritmo que es empleado por la RAMA para su cálculo se muestra en la Tabla 1.

UNIDAD	INTERVALO	ECUACION
ppm	[0, 0.11] ppm	909.090909 x (concentración)
	(0.11, 0.60] ppm	816.32653 x (concentración) + 10.20409

Tabla 1. Algoritmo para calcular IMECA de ozono.

Los valores que toma se ubican en una escala de 0 a 500, siendo un valor de 100 satisfactorio y el de 500 representa una muy mala calidad del aire, para dicho nivel existen evidencias de daños significativos a la salud.

1.2 ANTECEDENTES

Desde la fundación de Tenochtitlan, la Ciudad de México ha sido un centro de gran importancia en el país, su crecimiento ha variado a lo largo de las diversas épocas que ha atravesado. Es en el presente siglo cuando el proceso de crecimiento la convirtió en una de las metrópolis más grandes del mundo, fenómeno que se inicia a partir de la década de los treinta. El crecimiento de la población entre 1940 y 1950 fue de 73.5%, este incremento en la tasa demográfica tuvo su origen en el desarrollo económico, el centralismo y la migración hacia la ciudad por el enorme mercado de trabajo, y es justo en esta época, cuando se inicia la conurbación con el Estado de México.

Para 1970 el 14.5% de la población total del país vivía en la capital y en ella se concentraban el 35% de la actividad industrial y el 70% de los servicios. Es la industria la que ofreció empleos y fortaleza económica, motivando así un incremento demográfico. En los ochentas, la tasa de crecimiento general anual fue de 1.86% y en la presente década la zona metropolitana abarca 17 municipios del Estado de México.

El gran crecimiento que presenta la ciudad, tanto física como demográficamente, en conjunto con la concentración industrial y el incremento en los vehículos han traído como consecuencias varios desequilibrios, de los cuales, el deterioro ambiental se considera un problema fuerte. Cuando inicia el crecimiento del sector industrial la contaminación del aire también comienza a manifestarse; sin embargo, el crecimiento de la urbe trajo consigo el incremento en la principal causa de la contaminación urbana: el parque vehicular.

La contribución del automóvil, que es un vehículo de combustión interna, a la contaminación atmosférica es alta. De 1950 a 1990 su número en la Ciudad de México

aumentó en más de 24 veces, a tal grado, que sólo en un par de décadas logró alcanzar el primer lugar como fuente contaminante. Dicha posición la obtuvo en la década de los ochentas. Estudios realizados por dependencias públicas indican que entre 1982 y 1990 los automóviles produjeron entre el 75% y 85% del total de las emisiones contaminantes. Se observa un crecimiento directamente proporcional entre el número de vehículos y combustible consumido con el de contaminantes emitidos.

La industria y el transporte de combustión interna, son precisamente las mayores fuentes de contribución de precursores del ozono. Actualmente el ozono representa un problema de gran magnitud para esta ciudad, pues es considerado como el contaminante más significativo en la urbe, el más permanente en la atmósfera y el que por lo general se encuentra por encima de la norma. Este compuesto muestra una tendencia notoriamente en aumento, superando los 300 IMECA, siendo este nivel de advertencia, pues es considerado como muy insalubre para la salud humana.

Los efectos del ozono a la salud son considerables, es un gas de elevada toxicidad y de acción rápida, y existe evidencia de que sus efectos están relacionados con la exposición acumulativa diaria. Los problemas de salud que puede originar se basan en que es un poderoso irritante de las vías respiratorias y que penetra fácilmente a los pulmones, por ello produce irritación de ojos y garganta, dolor de cabeza, tos moderada, disminución en función pulmonar, cambios estructurales irreversibles en los pulmones y, con las exposiciones a largo plazo, puede acelerar los procesos normales de envejecimiento o aumentar el riesgo de cáncer. Además, daña materiales (caucho, fibras textiles, colorantes) y vegetales (daño en tejidos superficiales e intermedios, reducción del crecimiento, floración y cosechas).

En la Ciudad de México, la contaminación por oxidantes fotoquímicos quizá represente el problema más fuerte que tiene que enfrentar en este rubro. Aunque el ciclo fotoquímico del ozono en la ciudad fue detectado a finales de los cincuentas, fue hasta 1986 cuando el problema del ozono se magnificó; a partir de entonces se aprecia una clara tendencia al incremento. Es alarmante que en 1988 la norma (establecida en 0.11 partes por millón, ppm) haya sido violada 330 días y 3 años más tarde 345 días.

Generalizando, para cualquier zona de la ciudad, entre 1988 y 1991 se registró un 85% de días con excedencia (con máximos de más de 300 IMECA), y para 1991 se alcanzó un 95%.

Esto puede deberse a la combinación que se da entre las fuentes contaminantes y las características ambientales de la ciudad. En las atmósferas urbanas la formación del ozono se encuentra en relación con compuestos orgánicos reactivos y de tipo NO_x. En la Zona Metropolitana, se han favorecido las emisiones de estos contaminantes, las principales fuentes son los vehículos de combustión interna, la industria del vidrio, química y celulosa, y de los servicios los que más contribuyen son los baños públicos, tintorerías y hoteles.

Aunado a esto, las características de la Ciudad de México agravan el problema. Por su altitud, la deficiencia de oxígeno no permiten un proceso de combustión completa y eficiente, generándose más óxidos de nitrógeno; y su latitud se traduce en elevada radiación que favorece la formación de contaminantes fotoquímicos. Las temperaturas bajas que se presentan debido a su clima provocan inversiones térmicas.

Finalmente, con esta visión global, se puede comprender lo importante que es conocer el comportamiento histórico y las tendencias del ozono (y otros contaminantes), principalmente para las áreas críticas, como la zona suroeste de la ciudad; con ello, se puede conocer con cierta precisión si aumenta o disminuye, lo cual permite establecer mejores medidas de manejo, implementando estrategias tecnológicas, legales y sociales, y de ser necesario, alertar a la población en caso crítico. Todavía en los primeros años de la década de los noventas, esta información no existía, sin embargo, en la actualidad el problema ha adquirido gran importancia, por lo que se está generando la información necesaria para describir, predecir y establecer vínculos de los contaminantes.

1.3 AREA DE ESTUDIO

La Ciudad de México se encuentra en el Valle de México, cuya extensión es de aproximadamente 9,600 km², tiene forma semejante a una elipse y una altitud promedio de 2,240 msnmm (metros sobre el nivel medio del mar). El relieve que presenta esta región es ligero y sólo presenta algunas elevaciones de altura media (como la Sierra de Guadalupe y Santa Catarina).

El Valle de México está delimitado al norte por las sierras de Tezontlalpan, Tepetzotlán y Pachuca; al sur por las del Ajusco y de Chichinautzin; la Sierra Nevada al oriente y la Sierra de las Cruces, el Monte Alto y el Monte Bajo al poniente. Está completamente rodeado por montañas y gracias a esto, la ciudad se ve protegida de la acción directa de los huracanes. Los principales ríos y arroyos que posee bajan de las sierras por el poniente de la cuenca: ríos Magdalena, La Piedad, Los Remedios y Cuautitlán.

El clima se cataloga como subtropical de altura, templado, semiseco con lluvias de mayo a octubre con época de secas el resto del año, sin estación invernal bien definida y con temperatura media anual de 15° C.

1.4 PLANTEAMIENTO DEL PROBLEMA

Los registros, proporcionados por el Instituto Nacional de Ecología (INE), se utilizarán en el presente trabajo para desarrollar un análisis de serie de tiempo de los promedios semanales de los registros máximos diarios de ozono, reportados como IMECA. La evaluación será con el propósito de conocer su comportamiento y realizar predicciones. Esta información podría ser empleada como auxiliar en el desarrollo de diversas estrategias de control de contaminación atmosférica.

Los datos de ozono se pueden considerar como históricos, lo cual implica que son colecciones de fenómenos naturales que pueden ser observados sólo una vez y por lo tanto no ocurrirá de nuevo. Debido a la naturaleza del ozono, que es una variable que se registra en el orden en que ocurrió (es una secuencia de valores), el grupo de observaciones que se obtiene de su registro es una serie de tiempo, pues concuerda totalmente con la definición: sucesiones cronológicas (Guerrero, 1991), conjunto de observaciones tomadas a intervalos de tiempo específicos y generalmente equidistantes (Kazmier, 1967).

Con este trabajo se desea saber si el promedio semanal de la concentración máxima diaria de ozono que se presenta en la región suroeste de la Ciudad de México (que se caracteriza por presentar los niveles más altos de ozono) sigue un comportamiento reconocible a través del tiempo y por ello predecible, y un análisis de series de tiempo permite identificar los componentes que tienen influencia en cada uno de los valores de la serie.

La forma en que el ozono se mide permite reconocerlo como una variable aleatoria continua, pues los valores que puede tomar son todos aquellos que no sean negativos (la variable se mide en partes por millón y posteriormente se transforma en IMECA).

2 SERIES DE TIEMPO

2.1 INTRODUCCION

Se define una serie de tiempo como un conjunto de magnitudes pertenecientes a diferentes periodos de tiempo, de cierta variable o conjunto de variables (Chou, 1990), cuyos valores se presentan en el orden en que ocurrieron. Se pueden presentar en diversos fenómenos: económicos, físicos (meteorológicos), demográficos, etc.

Los objetivos de realizar un análisis de este tipo son la descripción, explicación, predicción de valores futuros de la serie y control del mecanismo generador de las observaciones. La descripción es el primer paso en el análisis y generalmente es una gráfica de la cual se obtienen medidas descriptivas de las principales propiedades de la serie, además en ella se mostrará la presencia de tendencia, datos aberrantes y discontinuidades.

El análisis de series de tiempo toma en cuenta el hecho de que observaciones sucesivas por lo general no son independientes y que el análisis debe de considerar el orden de las observaciones (Chatfield, 1989). Así pues, cuando las observaciones están correlacionadas, se pueden hacer predicciones de valores futuros a partir de observaciones anteriores. Cuando una serie puede describirse completamente sin margen de error y por lo tanto predecir valores futuros con toda exactitud, se denomina determinista; por el contrario, si la serie sólo es influenciada parcialmente por valores anteriores y existe cierta incertidumbre en la predicción, se le llama serie estocástica, no determinista o aleatoria. En toda serie de tiempo existe un elemento aleatorio y uno no aleatorio. Para este último caso se considera que existe una distribución de probabilidad de la variable, la cual está condicionada por el conocimiento de los valores ya ocurridos.

2.1.1 FUENTES DE VARIACIÓN

La variabilidad en el tiempo que presenta una serie puede clasificarse en:

1. Efecto estacional: variación periódica con amplitud constante, por ejemplo un ciclo anual o semanal.
2. Cambios cíclicos: variación de una serie pero con amplitud variable.
3. Tendencia: es un cambio suave del nivel de la serie a lo largo de un periodo de tiempo.
4. Otras fluctuaciones: una vez que no hay variación cíclica ni tendencia, sólo permanece la variación no determinista causada por situaciones no controlables que afectan los valores de la serie y que bien puede ser o no de origen aleatorio.

2.1.2 ELEMENTOS DE UNA SERIE DE TIEMPO

Una manera sencilla de describir un proceso estocástico es por medio de la media, varianza y covarianza, que permiten determinar la estacionariedad y la autocorrelación de una serie, que se verán más adelante. Estos momentos del proceso se definen como

Función media

$$\mu(t) = E[X(t)]$$

Función varianza

$$\sigma^2(t) = Var[X(t)]$$

Función autocovarianza

$$\gamma(t_1, t_2) = E\{[X(t_1) - \mu(t_1)][X(t_2) - \mu(t_2)]\}$$

La autocovarianza es la covarianza de observaciones de la misma variable, cuando el proceso es estacionario, también puede expresarse como:

$$\gamma(\tau) = Cov[X(t), X(t + \tau)]$$

Se dice que una serie es estrictamente estacionaria cuando la función de densidad para un conjunto arbitrario de variables ($Z_t, Z_{t+1}, \dots, Z_{t+m}$) es invariante respecto al desplazamiento en el tiempo, es decir, cuando la distribución es desplazada k lugares no se ve afectada, por ello la distribución es la misma para toda t . La estacionariedad estricta es un supuesto fuerte y difícil de verificar, por lo que se utiliza la de segundo orden que es un supuesto menos rígido (Diggle, 1990). La estacionariedad de segundo orden se presenta si los momentos de primer y segundo orden no dependen del tiempo. Es común suponer que la distribución asociada con las series de tiempo es la normal, por ello es suficiente conocer la media y la función de autocovarianza para caracterizar completamente una serie estacionaria (Guerrero, 1991).

La estacionariedad estricta implica estacionariedad de segundo orden, pero no a la inversa, esta implicación sólo se da cuando se tiene un proceso normal (variables con distribución conjunta normal multivariada).

La función de autocorrelación (FAC) se obtiene de estandarizar la función de autocovarianza

$$\rho(\tau) = \frac{\gamma(\tau)}{\gamma(0)}$$

esta función mide la correlación entre $X(t)$ y $X(t + \tau)$, es decir, entre observaciones a diferentes tiempos.

La gráfica de los coeficientes obtenidos con la función de autocorrelación (correlograma) da indicio sobre cuál modelo puede ser aplicado. El correlograma indica si la serie es completamente aleatoria, si existe correlación a "corto plazo", no estacionariedad y fluctuaciones estacionales.

También se cuenta con la función de autocorrelación parcial (FACP) que es un instrumento para identificar los procesos de tipo AR que serán explicados más adelante. Esta función cuantifica la correlación que existe entre Z_t y Z_{t+k} después de descontar las variables $Z_{t+1}, Z_{t+2}, \dots, Z_{t+k-1}$. Queda definida por la siguiente expresión (Tapia, 1995)

$$P_k = \frac{\text{Cov}\left[(Z_t - \hat{Z}_t), (Z_{t+k} - \hat{Z}_{t+k})\right]}{\sqrt{\text{Var}(Z_t - \hat{Z}_t)} \sqrt{\text{Var}(Z_{t+k} - \hat{Z}_{t+k})}}$$

en donde

$$\hat{Z}_{t+k} = E\{Z_{t+k} | Z_{t+1} \dots Z_{t+k-1}\}$$

es un estimador de la dependencia lineal de Z_{t+k} sobre $Z_{t+k-1}, Z_{t+k-2}, \dots, Z_{t+1}$.

2.2 MODELOS PARA SERIES DE TIEMPO

2.2.1 RUIDO BLANCO

A la sucesión de valores que se generan como resultado de la ocurrencia aleatoria e independiente de una variable se le denomina ruido blanco, que es un proceso puramente aleatorio, estacionario, con distribución fija, media (generalmente considerada igual a cero) y varianza constantes (Wei, 1990), cuya función de autocorrelación es

$$\rho(k) = \begin{cases} 1 & k = 0 \\ 0 & k \neq 0 \end{cases}$$

y la de autocorrelación parcial

$$\phi_{kk} = \begin{cases} 1 & k = 0 \\ 0 & k \neq 0 \end{cases}$$

El ruido blanco consiste de una secuencia de variables aleatorias no correlacionadas, y se utiliza para construir otras secuencias como se verá a continuación.

2.2.2 MODELO DE MEDIAS MOVILES (MA)

Dado un proceso de ruido blanco $\{a_t\}$ con media $\mu = 0$ y varianza σ_a^2 para toda t , se dice que $\{Z_t\}$ es un proceso de medias móviles de orden q , si se puede representar como

$$Z_t = \mu + a_t + \theta_1 a_{t-1} + \dots + \theta_q a_{t-q}$$

En donde $\theta_1, \theta_2, \dots, \theta_q$ son los parámetros de promedios móviles, indican la intensidad de efecto del evento. Un proceso de medias móviles es simplemente un filtro lineal (transformación de una secuencia de variables aleatorias a otras) aplicado a Z_t .

Todo proceso MA es estacionario y debido a que ni la media, ni la varianza, ni la covarianza dependen del tiempo, el proceso es estacionario de segundo orden.

Media

$$E(\tilde{Z}_t) = 0$$

Varianza

$$Var(\tilde{Z}_t) = \sigma_a^2 \sum_{i=0}^q \theta_i^2$$

donde

$$\tilde{Z} = Z_t - \mu$$

y representa a la desviación de Z_t con respecto a la media. La nueva serie obtenida por las desviaciones de la media, se comporta igual que la original, sólo que ahora la media de la serie será igual a cero, en lugar de μ .

La autocorrelación queda expresada como

$$\rho_k \left\{ \begin{array}{ll} 1 & k = 0 \\ \frac{\sum_{i=0}^{q-k} \theta_i \theta_{i+k}}{\sum_{i=0}^{q-k} \theta_i^2} & k = 1, 2, \dots, q \\ 0 & k > q \end{array} \right.$$

Un proceso de este tipo sólo se ve influenciado por las q últimas observaciones, por ejemplo, un proceso MA(1) tendrá una influencia sobre sí limitada a un sólo periodo.

Todo proceso MA cumple con ser estacionario, pero se debe de comprobar su invertibilidad, lo que implica que también puede escribirse como un proceso autorregresivo (descrito a continuación), asegurando así que existe sólo un proceso MA para una función de autocorrelación. La condición de invertibilidad se cumple si las raíces de la ecuación

$$\theta(x) = 1 - \theta_1 x - \theta_2 x^2 - \dots - \theta_p x^p = 0$$

están fuera del círculo unitario.

2.2.3 MODELO AUTORREGRESIVO (AR)

Es similar a un modelo de regresión múltiple, pero en este caso las variables explicativas son los valores de la misma variable que se han observado en periodos anteriores (de aquí el prefijo "auto") y ponderados con coeficientes autorregresivos ϕ_1, \dots, ϕ_p . La forma general de este modelo es

$$Z_t = \mu + \phi_1 Z_{t-1} + \dots + \phi_p Z_{t-p} + a_t$$

en donde a_t es un proceso de ruido blanco, el primer término del segundo miembro es una constante y el número de variables explicativas están determinadas por las primeras p autocorrelaciones.

Para un proceso AR(p) la media, la varianza y la función de autocorrelación están dadas respectivamente por

$$E(Z) = \mu$$

$$\text{Var}(\tilde{Z}) = \phi_1 \gamma_1 + \dots + \phi_p \gamma_p + \sigma_a^2$$

$$\rho_k = \phi_1 \rho_{k-1} + \dots + \phi_p \rho_{k-p} \quad \text{para } k \geq p$$

Esta última es conocida como ecuación Yule-Walker. Las tres expresiones son útiles para determinar las propiedades de un modelo, dados los parámetros del modelo, y para estimar los parámetros de un modelo dada una serie de observaciones.

El modelo autorregresivo más sencillo es el de orden uno y se conoce como serie o proceso Markov, en el cual se requiere que $|\phi| < 1$ para cumplir con la condición de estacionariedad, lo que expresado en términos de autocorrelaciones es equivalente a requerir $|\rho_1| < 1$.

En un caso general de un proceso AR de orden p , la estacionariedad se cumple sí y sólo si las raíces de la ecuación

$$\phi(x) = 1 - \phi_1 x - \phi_2 x^2 - \dots - \phi_p x^p = 0$$

se encuentran fuera del círculo unitario. Esto es similar a la manera de comprobar las condiciones de invertibilidad en un proceso MA.

Todo proceso AR es invertible, lo que implica que está determinado de manera única por su FAC, lo cual no ocurre con los procesos no invertibles.

2.2.4 MODELOS AUTOREGRESIVOS Y DE MEDIAS MÓVILES (ARMA)

Son una combinación de los modelos AR y MA, originándose así los modelos autorregresivos y de medias móviles. Se obtendrá un proceso de este tipo de orden (p, q) de la mezcla de un proceso AR con p términos y uno MA con q términos. Dado por

$$\begin{aligned} Z_t = & \phi_1 Z_{t-1} + \phi_2 Z_{t-2} + \dots + \phi_p Z_{t-p} + a_t \\ & - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \dots - \theta_q a_{t-q} \end{aligned}$$

Para encontrar las autocorrelaciones sólo se mencionará que es necesario resolver un sistema simultáneo de ecuaciones para $\gamma_1, \dots, \gamma_p$ (previamente definidas como las autocovarianzas). Para que se cumplan las condiciones de que el proceso es estacionario e invertible, es necesario que las raíces de las ecuaciones $\phi(x) = 0$ y $\theta(x) = 0$, respectivamente, estén fuera del círculo unitario.

Una serie de tiempo estacionaria puede describirse por un modelo ARMA con menos parámetros que aquellos requeridos por un proceso AR o MA por sí solos, lo que le confiere una ventaja en cuanto a parsimonia.

2.2.5 MODELOS ARIMA

Reciben este nombre los modelos autorregresivos e integrados de medias móviles y se pueden considerar como una generalización de los modelos ARMA, en ellos se emplean diferencias para lograr un proceso estacionario en cuanto al nivel de la serie.

$$W_t = \nabla^d \tilde{Z}_t = (1 - B)^d \tilde{Z}_t \quad \text{para toda } t$$

donde ∇^d indica la diferencia de orden d de la serie, el operador diferencia se define por

$$\nabla Z_t = Z_t - Z_{t-1}$$

y el operador de retraso por

$$BZ_t = Z_{t-1}$$

la relación entre ambos operadores está dada por

$$\nabla = 1 - B$$

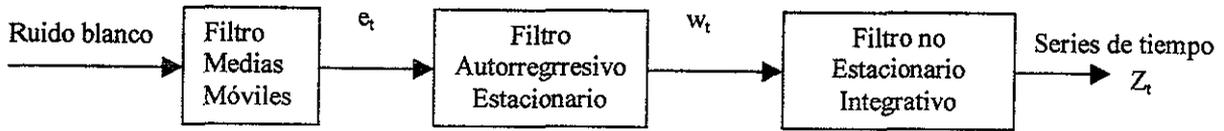
Cabe mencionar que los polinomios de retraso también permiten expresar de una manera concisa y simple los modelos MA, AR y ARMA.

Se aplican d diferencias con el propósito de volver estacionario el comportamiento de la serie, así la nueva serie será estacionaria con respecto a su nivel, lo cual se logra después de cierto número de retrasos. Para esta nueva serie se puede obtener un modelo ARMA, que es equivalente a considerar uno ARIMA

$$\phi(B)\nabla^d \tilde{Z}_t = \theta(B)a_t$$

donde a_t es un proceso de ruido blanco.

Un modelo ARIMA (p,d,q) consta de un polinomio de retraso de grado p ($\phi(B)$, autorregresivo), una diferencia de orden d (exponente en el operador diferencia) y un polinomio de retraso de grado q ($\theta(B)$ de promedios móviles). Este tipo de proceso puede ser generado a partir de un ruido blanco como lo ilustra Box y Jenkins (1970).



Debido a que un proceso ARIMA (p,d,q) es la generalización de un ARMA (p,q) , para cuestiones de identificación la función de autocorrelación y autocorrelación parcial son las mismas para ambos.

Para que este modelo cumpla con ser estacionario e invertible, es necesario que las raíces de $\phi(x) = 0$ y $\theta(x) = 0$ se encuentren fuera del círculo unitario.

2.3 MODELOS PARA SERIES PERIODICAS

Los modelos de tipo ARIMA pueden ser empleados para ajustar series periódicas, empleando el mismo procedimiento iterativo que es usado para cuando los procesos no presentan periodicidad.

Este análisis se basa en que los datos separados por un periodo s de tiempo serán similares: Z_t será similar a $Z_{t+i(s)}$, con $i = 1, 2, 3, \dots$; y se espera que estas observaciones estén correlacionadas.

Para eliminar la estacionalidad en este tipo de series, se tienen que realizar diferencias periódicas con una longitud de s (intervalo del periodo); también aquí es conveniente cuidar de no sobrediferenciar, pues en una serie estacional se perderán todas las observaciones correspondientes al primer periodo.

El valor asignado a s , está relacionado con la frecuencia en que los datos fueron registrados (anuales, mensuales, semanales), y el intervalo periódico se puede determinar visualmente en función de las observaciones similares presentadas a lo largo del tiempo (cada 7, 14, etc. datos se presentan registros similares).

El uso y aplicación de las FAC y FACP teóricas y estimadas es el mismo que el planteado para los modelos no estacionales, sirviendo de guía para escoger un modelo tentativo. Sólo que ahora es obtenido a retrasos múltiples de s ($s, 2s, 3s, \dots$) y, de existir

correlación, los coeficientes de estas funciones serán diferentes de cero a retrasos múltiplos de s . Una vez que se escoge un modelo se estiman sus parámetros AR y MA.

En muchas series se observa que además de presentarse un patrón periódico también hay uno no estacional que debe ser removido, esto generalmente se logra con diferencias. Así pues, ambos elementos deberán ser considerados en el modelo multiplicativo presentado a continuación:

$$\phi_p(B)\Phi_p(B^s)\nabla^d\nabla_s^D\tilde{Z}_t = \Theta_Q(B^s)\theta_q(B)a_t$$

donde

- $\phi_p(B)$ = operador AR no estacional
- $\theta_q(B)$ = operador MA no estacional
- $\Phi_p(B^s)$ = operador AR estacional
- $\Theta_Q(B^s)$ = operador MA estacional
- ∇^d = operador diferencia no estacional
- ∇_s^D = operador diferencia estacional

La ecuación arriba expuesta representa el modelo ARIMA sugerido por Box y Jenkins para representar series con elementos estacionales y no estacionales, en el cual los elementos AR se multiplican uno por el otro al igual que los MA. En un caso así, el problema se torna un poco complejo pues las funciones FAC y FACP reflejan ambos elementos, sin embargo, Guerrero (1991) presenta las formas de las autocorrelaciones más comunes para este tipo de procesos.

Pero de no existir la tendencia (elemento no estacional), el modelo se reduce a:

$$\Phi_p(B^s)\nabla_s^D\tilde{Z}_t = \Theta_Q(B^s)a_t$$

Aquí también es necesario realizar una verificación de los supuestos. El modelo tiene que satisfacer que a_t sea ruido blanco, lo cual se observa por medio de la estimación de los residuales \hat{a}_t del modelo. De igual manera deben ser revisadas las condiciones de estacionariedad e invertibilidad, que para un modelo multiplicativo es relativamente simple, pues las condiciones se aplican por separado a los coeficientes estacionales y no estacionales, por ejemplo, las condiciones de estacionariedad se verifican si las raíces de

$$\phi_p(B)\Phi_p(B^s) = 0$$

caen fuera del círculo unitario, o bien, si las raíces de las ecuaciones

$$\phi_p(B) = 0$$

$$\Phi_p(B^s) = 0$$

caen fuera del círculo unitario; una situación similar se presenta con la invertibilidad.

2.4 PREDICCIÓN CON ARIMA

Se despeja Z_t del modelo ARIMA, generalmente queda igualada a un polinomio que puede depender tanto de observaciones anteriores de la serie (Z_{t-k} , $k=1,2,\dots$), como de choques aleatorios (a_{t-k} , $k=1,2,\dots$). Para obtener la estimación puntual del pronóstico, se sustituyen los valores observados de la serie que se involucran en la ecuación, así como los residuos, que sustituyen a los choques aleatorios.

La dependencia que existe de la predicción con las observaciones anteriores, se puede expresar como

$$\hat{Z}_t(k) = E(Z_{t+k} | k_t)$$

donde k_t es la información contenida en la serie de observaciones (Z_t, Z_{t-1}, \dots); $Z_t(k)$ representa Z_{t+k} , y k es el tiempo al cual se va a predecir.

Los valores anteriores de a_t son reemplazados por sus correspondientes estimaciones (\hat{a}_t), pero cuando se requiere un valor posterior al correspondiente a t , es reemplazado por su valor esperado, que es cero.

Para determinar un intervalo de confianza de un pronóstico dado, hay que tener presente que cualquier modelo ARIMA estacionario se puede escribir en forma de choques aleatorios (un término AR puede reemplazarse por una serie infinita de términos MA), y esa propiedad permite elaborar un intervalo de confianza, que se obtiene por medio de la siguiente expresión:

$$\hat{Z}_n(l) \pm Z_{\alpha/2} \left(\sum_{j=0}^{l-1} \psi_j^2 \right)^{1/2} \sigma_a$$

donde ψ proviene de

$$\phi(B)Z_t = \theta(B)a_t$$

$$Z_t = \psi(B)a_t$$

$$Z_{\alpha/2} = \text{cuantil de la normal estándar}$$

donde

$$\psi(B) = \frac{\theta(B)}{\phi(B)} = \sum_{j=0}^{\infty} \psi_j B_j$$

y

$$\psi_0 = 1$$

3 CONSTRUCCION DE MODELOS

Existen cuatro etapas propuestas por Box y Jenkins para encontrar un buen modelo que describa cómo se encuentran relacionadas entre sí las observaciones de una serie de tiempo, y que básicamente representa el mecanismo generador de la serie de tiempo que se analiza:

1. Identificación: se determinan los valores de p , d , q que especifican el modelo ARIMA (recordando que los modelos AR, MA y ARMA son casos especiales de los modelos ARIMA) adecuado para los datos que se están analizando.
2. Estimación: estimar los parámetros del modelo (coeficientes de polinomios autorregresivos y de promedios móviles que sean necesarios en el modelo) que representen a la serie.
3. Verificación: corroborar el cumplimiento de los supuestos y de que el ajuste del modelo es adecuado.
4. Uso del modelo: aplicar el modelo obtenido para pronosticar, controlar, simular o explicar el fenómeno.

Estas etapas se representan gráficamente en la Figura 2 tomada de Guerrero (1991).

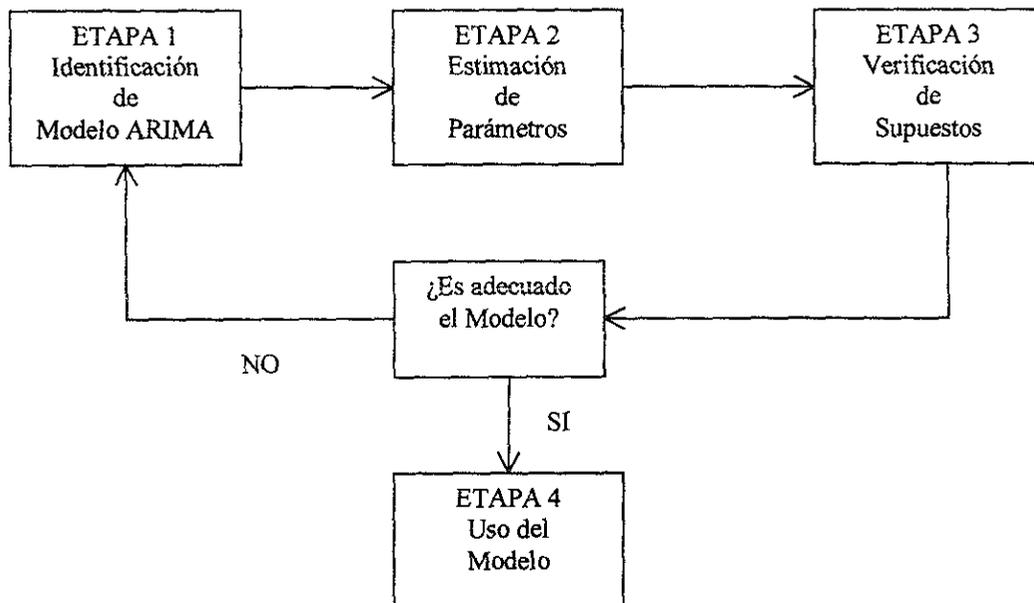


Figura 1

Este proceso iterativo es importante pues proporciona elementos para reformular el modelo hasta encontrar uno satisfactorio.

3.1 IDENTIFICACION

Es de suma utilidad que, como primer paso, se realice una gráfica de los valores observados contra los tiempos de observación. En ella se busca, básicamente, evidencia de no estacionariedad, que puede manifestarse por algún tipo de tendencia, oscilación, periodicidad y estacionalidad; estas fuentes de variación representan un cambio de la media y/o varianza de la serie en el tiempo. Cabe mencionar que la media y la varianza muestrales de una serie de tiempo se calculan como cualquier otra media o varianza.

Debido a que los modelos ARIMA únicamente se aplican a series estacionarias es importante lograr este requisito antes de continuar con el análisis, y muchas de las series no estacionarias pueden ser transformadas en estacionarias con el empleo de operaciones sencillas.

3.1.1 ESTABILIZACION DE LA VARIANZA

Si la varianza fuera no estacionaria (cambia con el nivel de las mediciones), se debe lograr que ésta sea constante antes de empezar a modelar. La varianza se puede corregir aplicando transformaciones como logaritmo natural, raíz cuadrada de los datos originales o empleando la transformación de Box-Cox (Box y Cox, 1964, Pankratz, 1983 y Guerrero, 1991). En las series que presentan ambos tipos de no estacionariedad (varianza y nivel) primero se obtiene una varianza constante y luego se ajusta la media.

Para determinar si la varianza es constante, graficar la serie a lo largo del tiempo es de gran ayuda. En ella se observa el comportamiento de la varianza (constante, creciente, decreciente, no monótona). Guerrero (1991) presenta un método para verificar el resultado de este análisis visual. En dicha técnica se requiere conocer el cociente

$$\frac{\sigma_t}{\mu_t^{1-\lambda}}$$

donde λ representa la potencia de la transformación estabilizadora de Box-Cox. Dicho cociente será constante si la varianza también lo es.

El procedimiento inicia con la división de los datos en H grupos con R elementos contiguos, sin importar que algunos datos queden fuera:

$$R = \frac{(N - n)}{H}$$

con N = número total de observaciones de la serie
n = número de observaciones fuera de los grupos.

De cada grupo, se calculan la media, varianza y desviación estándar. Esta información sirve para determinar el valor de λ de la transformación Box-Cox que estabiliza la varianza de Z_i :

$$T(Z_i) = \begin{cases} Z_i^\lambda & \lambda \neq 0 \\ \log(Z_i) & \lambda = 0 \end{cases}$$

ahora, para obtener la relación

$$\frac{S_h}{\bar{Z}_h^{1-\lambda}}$$

que representa una aproximación a

$$\frac{\sigma_i}{\mu_i^{1-\lambda}}$$

donde $h=1, \dots, H$. Es necesario escoger, sin que exista alguna regla, "q" valores de λ de tal manera que cumplan con $-1 \leq \lambda \leq 1$.

Con los cocientes obtenidos se calcula el coeficiente de variación de la siguiente manera

$$CV(\lambda_i) = \frac{\sigma_{\lambda_i}}{M(\lambda_i)}$$

donde

$$M(\lambda_i) = \frac{\sum_{h=1}^H \left(\frac{S_h}{\bar{Z}_h^{1-\lambda}} \right)}{H}$$

$$\sigma_{\lambda_i} = \sqrt{\frac{\sum_{h=1}^H \left[\left(\frac{S_h}{\bar{Z}_h^{1-\lambda}} \right) - M(\lambda) \right]^2}{(H-1)}}$$

y el valor de la potencia (λ) que se empleará para estabilizar la varianza será aquella dada por el coeficiente de variación más pequeño para $i=1,2,3,\dots,q$. Cuando $\lambda=1$, la serie no requiere de ser transformada.

3.1.2 ESTABILIZACION DEL NIVEL

El tipo más común de no estacionariedad es cuando la media (nivel de la serie) cambia a través del tiempo. En este caso, la diferenciación de los datos es utilizada con frecuencia para inducir a la estacionariedad. Hay que cuidar el no sobrediferenciar la serie pues se incrementa la varianza de la misma y se pierden observaciones.

El gráfico de la serie puede dar evidencia de la presencia de tendencia en el proceso. Sin embargo, la inspección visual sólo es una guía y el cambio en el nivel debe ser evaluado por otros medios.

Cuando existe no estacionariedad debida a la media, ésta se refleja en el correlograma. En la gráfica de la FAC, se observa un decaimiento lento de las autocorrelaciones a cero (Figura 7.b), lo cual indica la necesidad de estabilizar el nivel y, al mismo tiempo, sirve como herramienta para determinar el grado de diferenciación requerido.

Anderson (citado por Guerrero) sugiere otra manera de lograr lo anterior. Se calcula la desviación estandar de varias diferencias (de bajo orden) y se comparan entre sí. La varianza decrecerá hasta que la serie se torne estacionaria, pero se presentará un incremento en la misma al sobrediferenciar.

También se puede utilizar un análisis de regresión simple, aplicado a las observaciones contra el tiempo, para probar la presencia de tendencia en una serie (Kendall y Ord, 1990). En él se evaluará si la pendiente es significativa, y de ser así, será necesario remover la tendencia por medio de diferencias.

Los datos que tienen comportamiento estacional (generalmente se presentan en economía y negocios) requieren de un tipo particular de diferenciación denominado diferenciación estacional (definida con anterioridad); también hay series con patrones estacionales y no estacionales, para ellas será necesario un modelo multiplicativo o un aditivo estacional (consultar Box y Jenkins, 1970 y Guerrero, 1991).

Siempre hay que recordar que si una serie fue transformada para ser modelada se puede aplicar la transformación inversa para obtener la serie no estacionaria original.

Resumiendo, si los datos parecen ser estacionarios no se requiere de aplicar transformación alguna (en un proceso ARIMA si no hay diferenciación equivale a tener $d = 0$, es decir, un orden de cero). Si aparentemente no hay estacionariedad, se procede a

transformar la serie (generalmente con una diferenciación) hasta lograr que en el gráfico de observaciones contra tiempo luzca como estacionario.

3.1.3 APLICACION DE FAC Y FACP

Con el objetivo de medir la correlación entre las observaciones de una serie de tiempo se emplean la función de autocorrelación y la función de autocorrelación parcial; ambas funcionan como una guía para escoger cuál modelo ARIMA se ajustará. Cada modelo ARIMA tiene una FAC y FACP teórica y la idea es comparar éstas con las estimadas. En la Tabla 2 se mencionan las características generales más importantes de las funciones teóricas:

MODELO	FAC TEORICA	FACP TEORICA
AR	Decae suavemente hacia cero.	Es a cero después de pocas (p) autocorrelaciones parciales.
MA	Es a cero después de pocas (q) autocorrelaciones.	Decae suavemente a cero.
ARMA	Decae suavemente hacia cero.	Decae suavemente hacia cero.

Tabla 2. Comportamiento teórico de las FAC y FACP.

Para ejemplificar gráficamente el comportamiento de las funciones FAC y FACP, se efectuó una simulación para cada uno de los modelos ARIMA. Las figuras que ilustran los procesos son: el gráfico de la serie (a), y sus correspondientes funciones de autocorrelación (b) y autocorrelación parcial (c) muestrales.

- 1) MA (1): proceso $Z_t = 0.6a_{t-1} + a_t$ (Figura 4)
- 2) AR (2): proceso $Z_t = 0.6 Z_{t-1} + 0.3 Z_{t-2} + a_t$ (Figura 5)
- 3) ARMA (1, 1): proceso $Z_t = -0.8 a_{t-1} - 0.7 Z_{t-1} + a_t$ (Figura 6)
- 4) ARIMA (1, 1, 1): proceso $W_t = 0.5 W_{t-1} - 0.7 a_{t-1} + a_t$ (Figura 7)

En la práctica los coeficientes de estas funciones se obtienen a partir de los coeficientes de autocovarianza muestral, los cuales están dados para un retraso k por

$$C_k = \frac{\sum_{t=1}^{N-k} (x_t - \bar{x})(x_{t+k} - \bar{x})}{N}$$

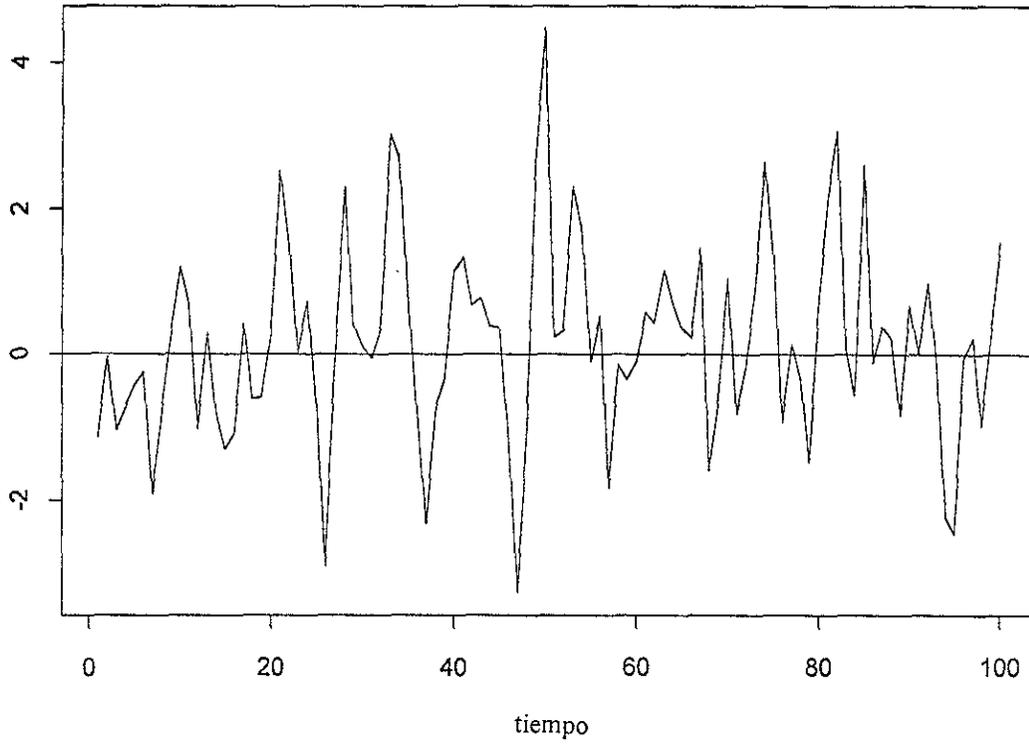


Figura 2.a: Realización del proceso $Z_t = 0.6a_{t-1} + a_t$

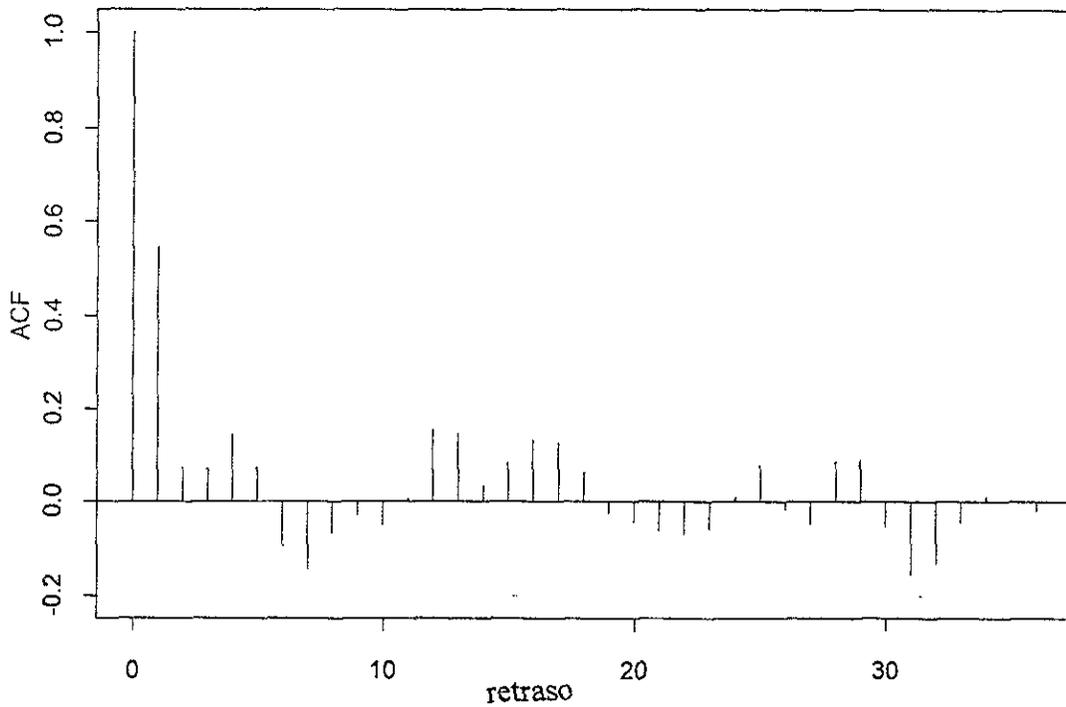


Figura 2.b: Función de autocorrelación muestral del proceso $Z_t = 0.6a_{t-1} + a_t$

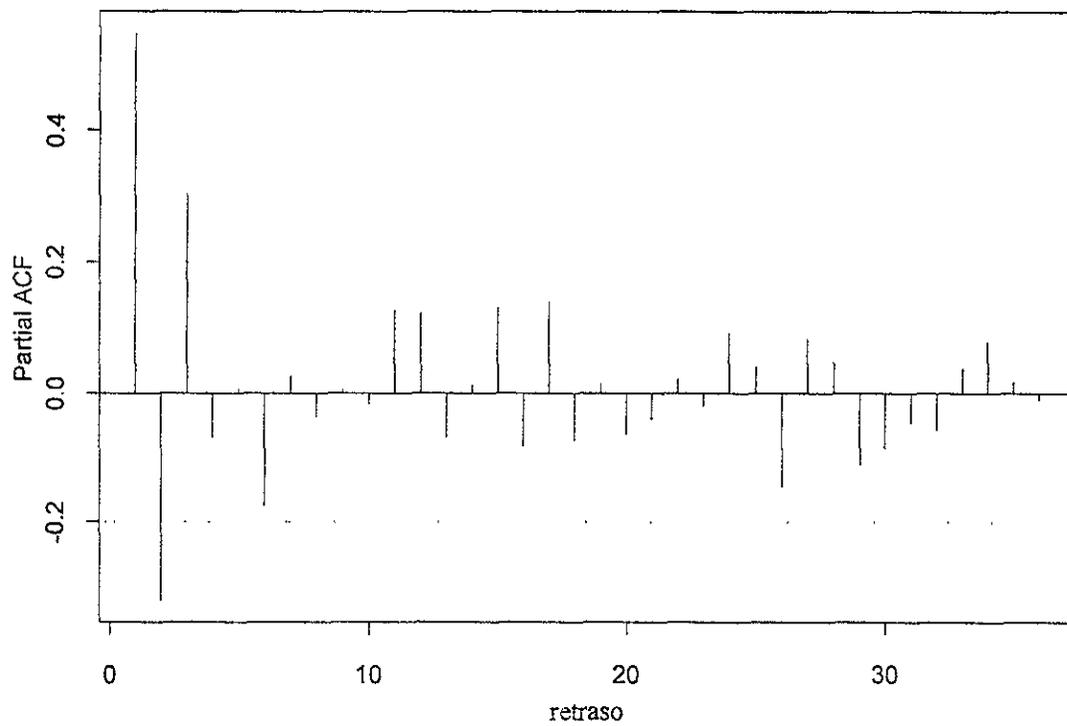


Figura 2.c: Función de autocorrelación parcial muestral del proceso $Z_t = 0.6a_{t-1} + a_t$

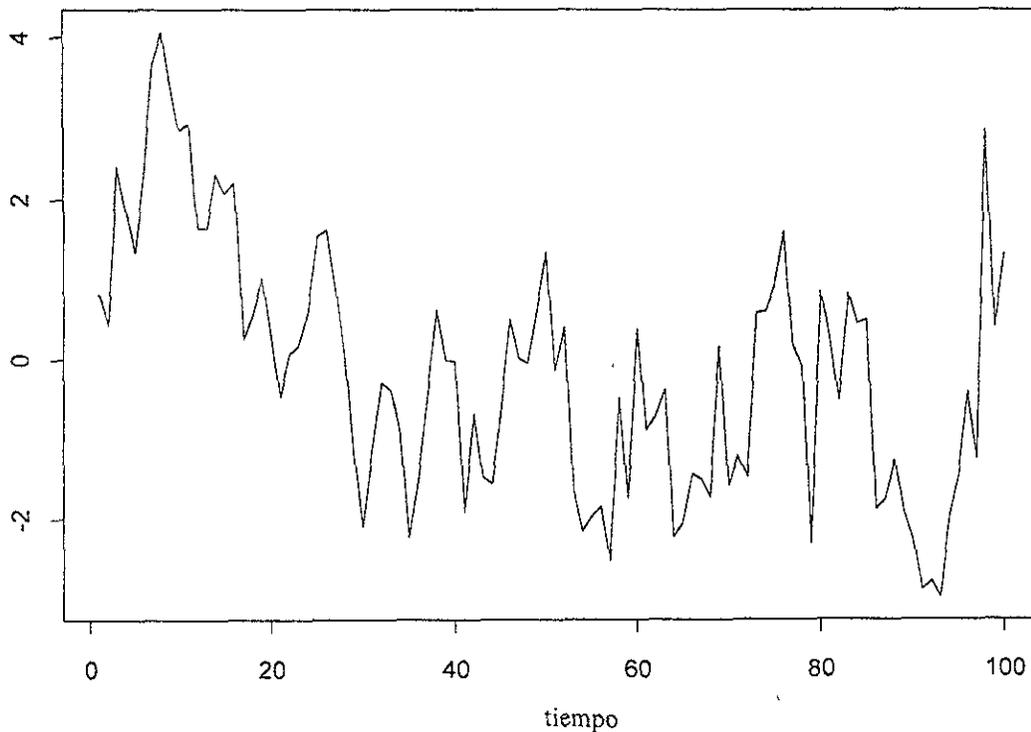


Figura 3.a: Realización del proceso $Z_t = 0.6 Z_{t-1} + 0.3 Z_{t-2} + a_t$

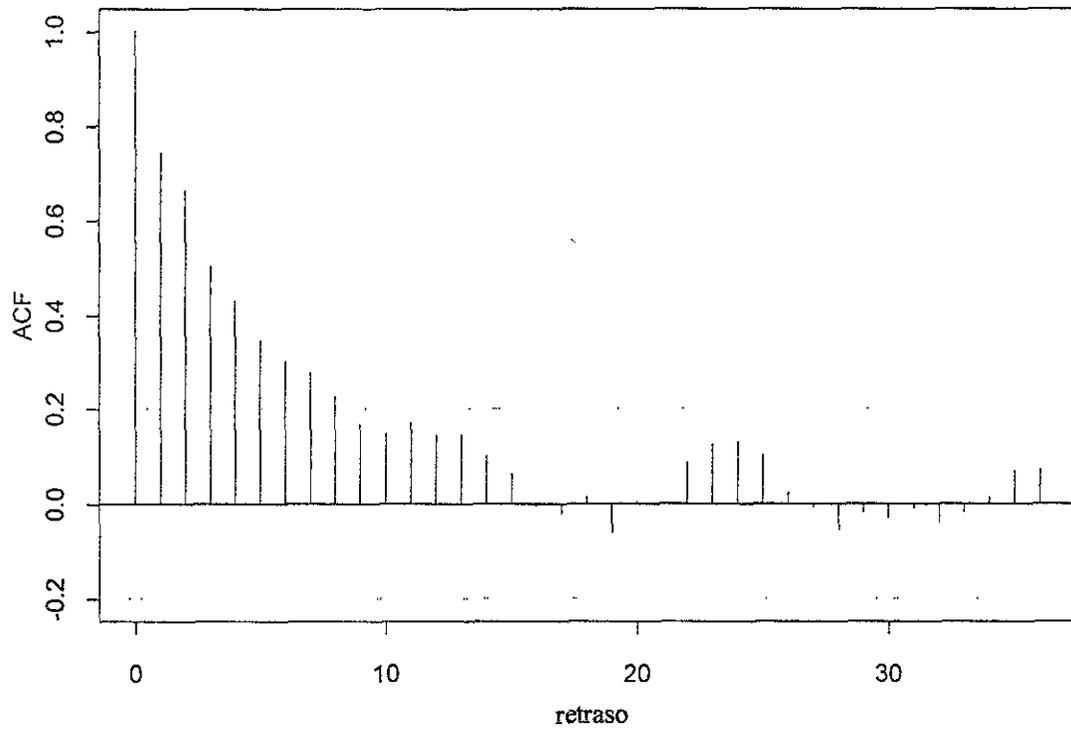


Figura 3.b: Función de autocorrelación muestral del proceso $Z_t = 0.6 Z_{t-1} + 0.3 Z_{t-2} + a_t$

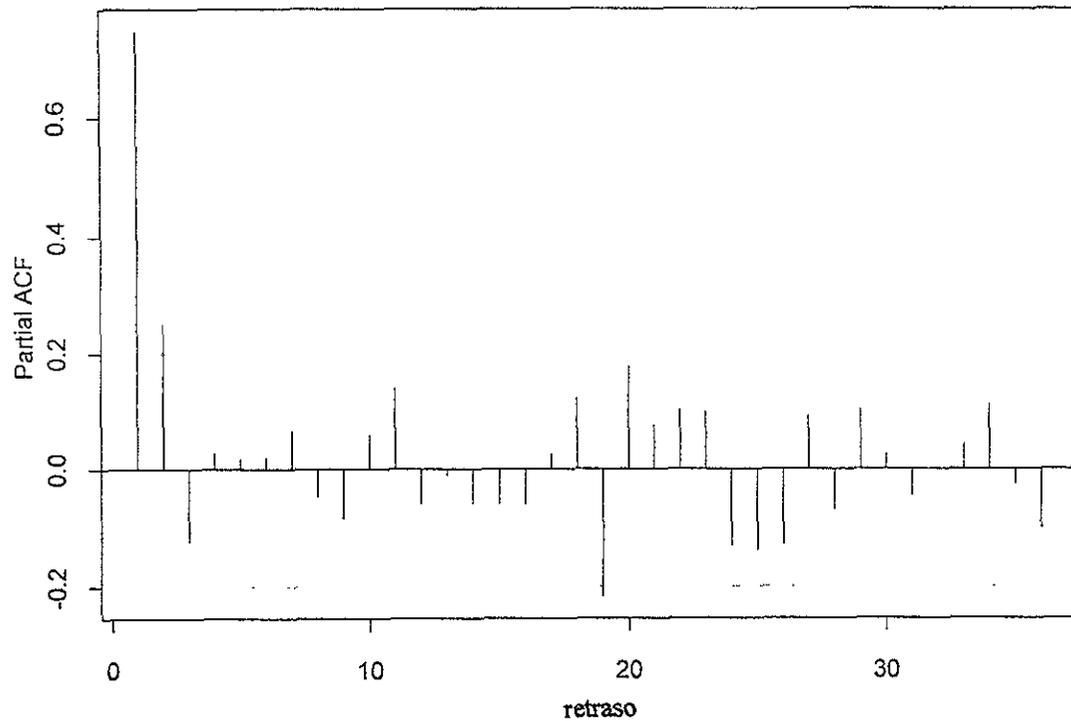


Figura 3.c: Función de autocorrelación parcial muestral del proceso $Z_t = 0.6 Z_{t-1} + 0.3 Z_{t-2} + a_t$

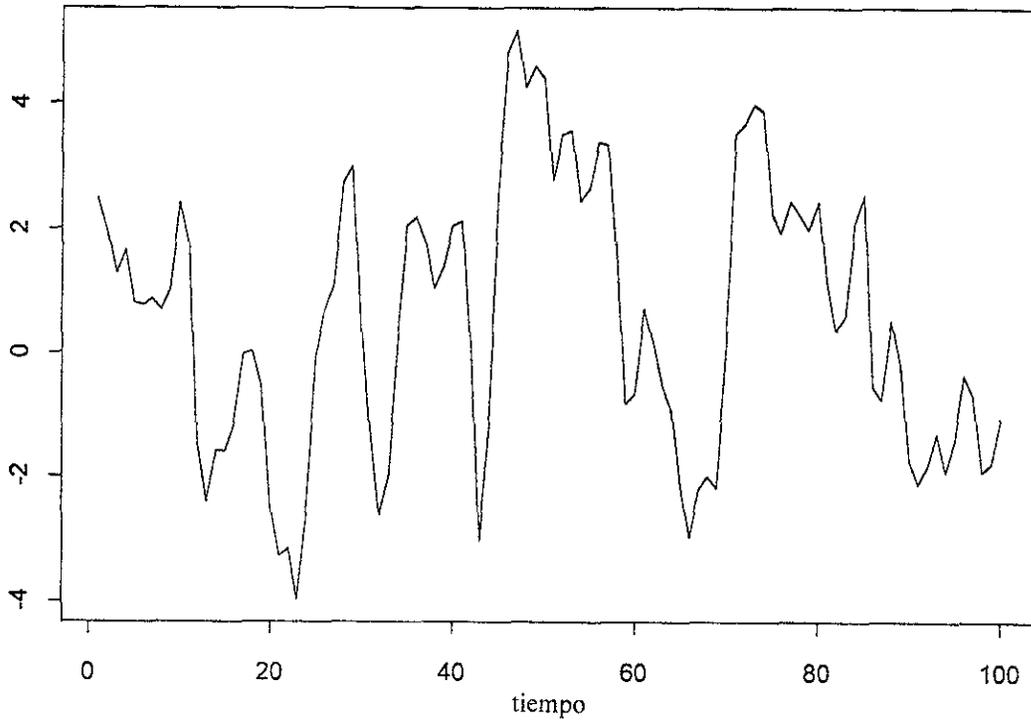


Figura 4.a: Realización del proceso $Z_t = -0.8 a_{t-1} - 0.7 Z_{t-1} + a_t$

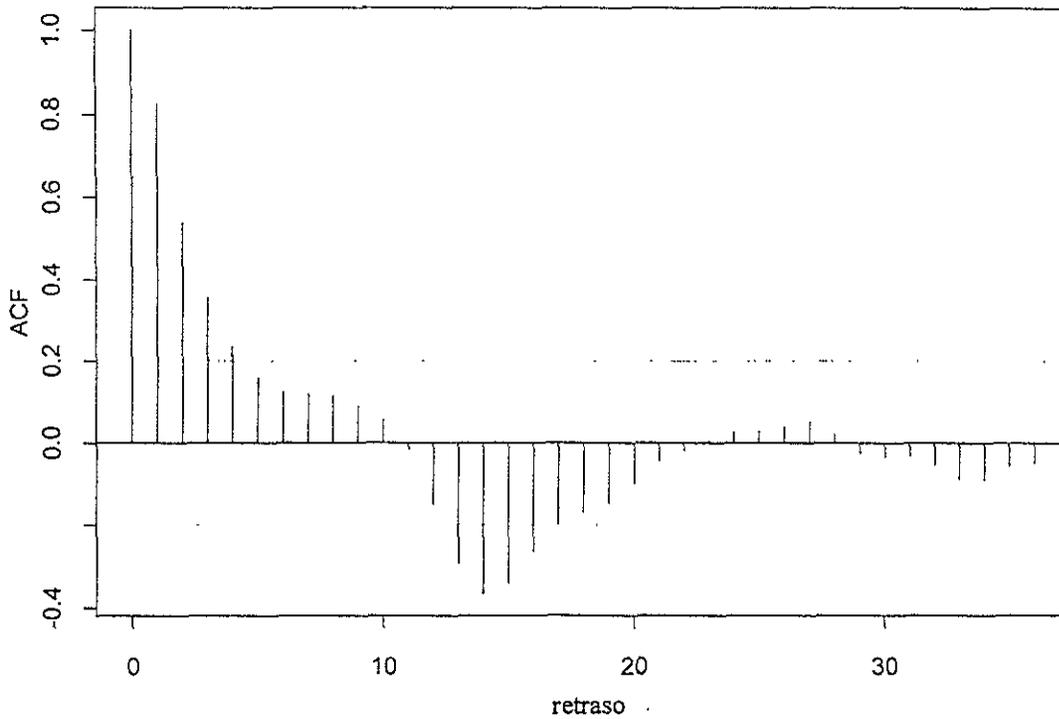


Figura 4.b: Función de autocorrelación muestral del proceso $Z_t = -0.8 a_{t-1} - 0.7 Z_{t-1} + a_t$

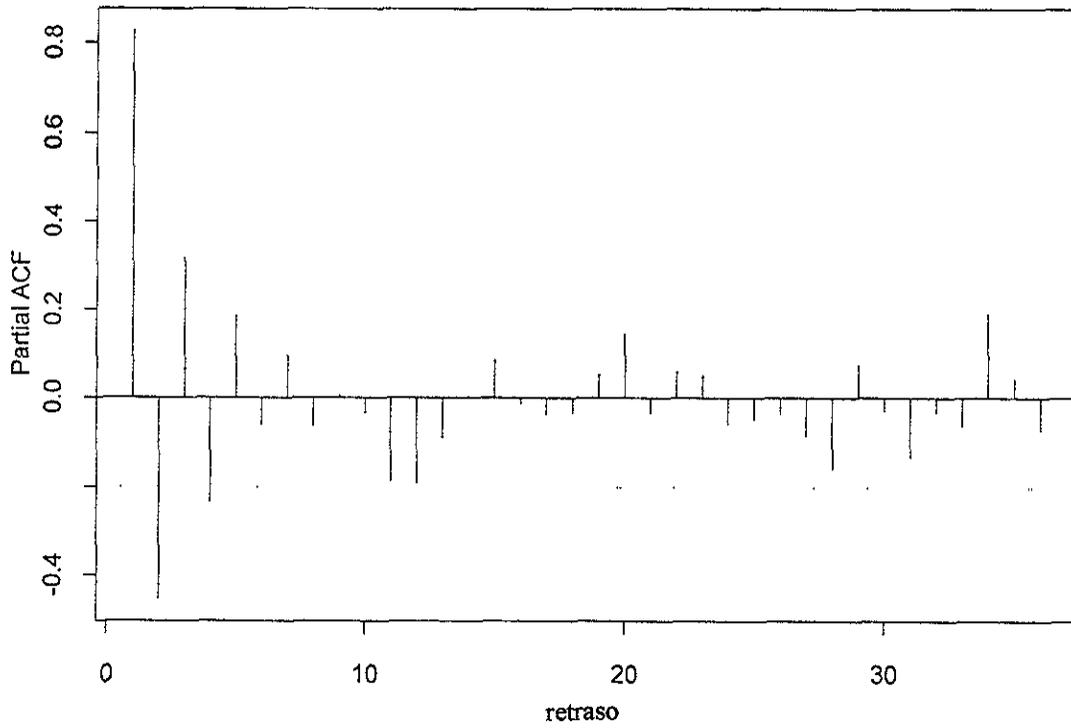


Figura 4.c Función de autocorrelación parcial muestral del proceso
 $Z_t = -0.8 a_{t-1} - 0.7 Z_{t-1} + a_t$

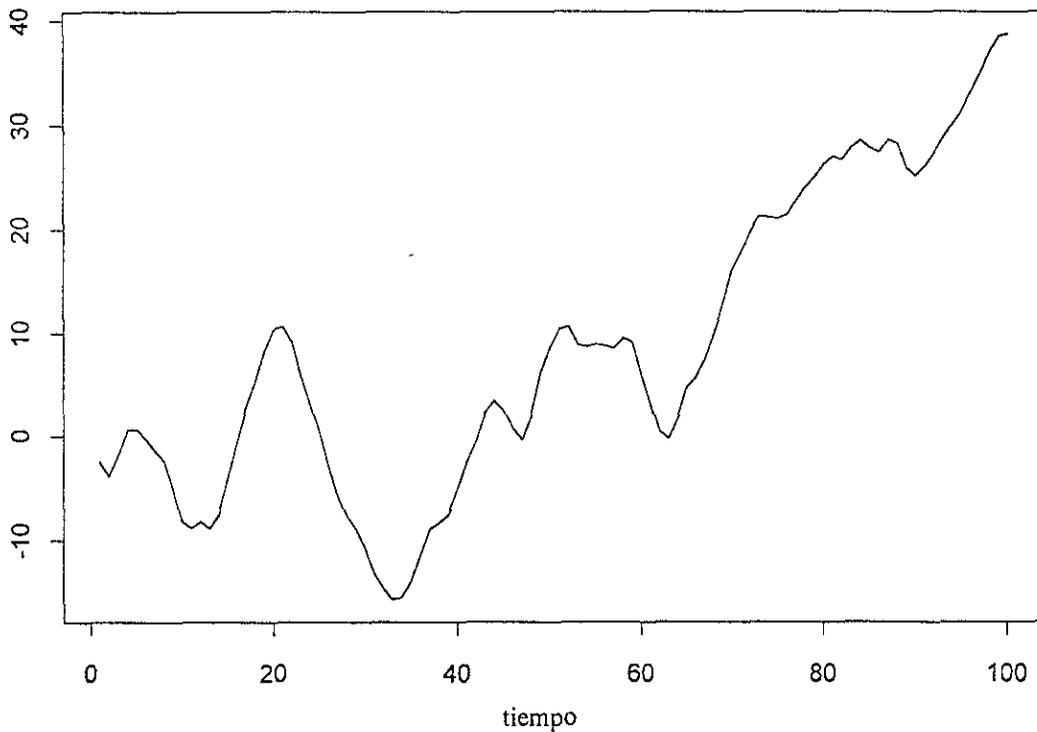


Figura 5.a: Realización del proceso $W_t = 0.5 W_{t-1} - 0.7 a_{t-1} + a_t$

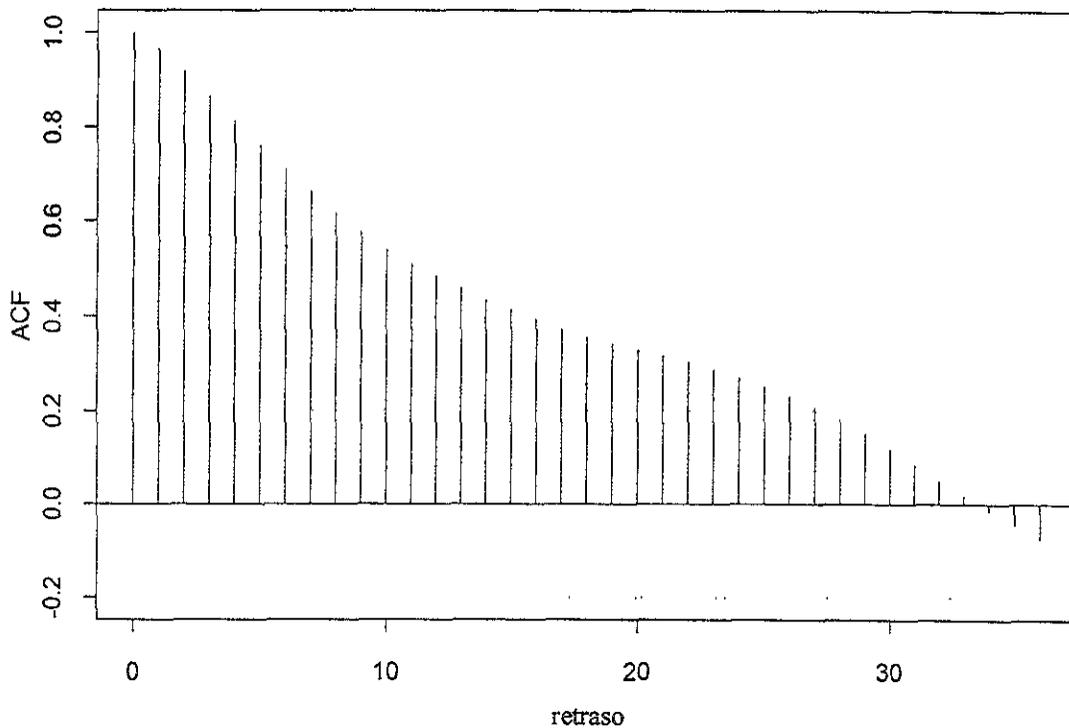


Figura 5.b: Función de autocorrelación muestral del proceso $W_t = 0.5 W_{t-1} - 0.7 a_{t-1} + a_t$, antes de ser diferenciada.

siendo x_t el valor observado y \bar{x} la media muestral. Partiendo de esta expresión se tiene que un estimador para $\rho(k)$ es

$$r_k = \frac{C_k}{C_0}$$

Una vez calculados los coeficientes de autocorrelación se pueden obtener las estimaciones para la autocorrelación parcial:

$$\hat{\phi}_{11} = r_1$$

$$\hat{\phi}_{kk} = \frac{r_k - \sum_{j=1}^{k-1} \hat{\phi}_{k-1,j} r_{k-j}}{1 - \sum_{j=1}^{k-1} \hat{\phi}_{k-1,j} r_j} \quad (k = 2, 3 \dots)$$

donde

$$\hat{\phi}_{kj} = \hat{\phi}_{k-1,j} - \hat{\phi}_{kk} \hat{\phi}_{k-1,k-j} \quad (k = 3, 4, \dots; j = 1, 2, \dots, k - 1)$$

La gráfica de los coeficientes de autocorrelación de la muestra se conoce como correlograma, en él se puede observar el comportamiento de los coeficientes de autocorrelación muestral, si muestran una caída abrupta a cero se puede pensar en un proceso MA (q) como un posible modelo para los datos, donde q es el número de coeficientes diferentes de cero y p es igual a cero (Figura 4). En contraste, el orden de un proceso AR (p) no es posible detectarlo con el correlograma, por ello se utiliza el correlograma parcial para su identificación.

Para construir este último gráfico se emplean los coeficientes de autocorrelación parcial muestrales y su interpretación es similar al correlograma, si en el gráfico existe una caída rápida hacia cero el modelo propuesto será uno de tipo AR (p) con un valor de q igual a cero. Así pues, el número de autocorrelaciones parciales que sean diferentes de cero indicarán el orden del modelo AR (Figura 5).

Si ninguno de los decaimientos en ambas gráficas se observa a un periodo lo suficientemente pequeño (generalmente 2 ó 3) será mejor considerar un proceso mixto con p y q coeficientes diferentes a cero, iniciando naturalmente con el orden más pequeño (Figura 6).

Para decidir si las autocorrelaciones son cero a partir de un cierto retraso q (en un modelo MA(q)), se comparan los coeficientes con sus desviaciones estándar. Cuando se tiene $N \geq 50$ la distribución de la autocorrelación r_k (coeficiente estimado) puede aproximarse a una distribución normal con media cero, por ello los valores ± 1.96 (error estándar) son los límites para determinar las observaciones que son significativamente distintas de cero con un nivel de significancia del 5%. En el caso de un proceso AR(p) las autocorrelaciones parciales se consideran diferentes de cero cuando están fuera del intervalo establecido por

$$\pm 2(\text{error estándar}) = \pm \frac{2}{\sqrt{N-d}} \quad \text{para } k > p$$

pero si el proceso fuera un MA(q), las autocorrelaciones serían significativas si exceden el intervalo delimitado por

$$\pm 2 \sqrt{\frac{1}{N-d} \left(1 + 2 \sum_{j=1}^q r_j^2 \right)} \quad \text{para } k > q$$

Diggle (1990) presenta un diagrama de flujo que ayuda a visualizar de manera general el proceso de identificación, se muestra en la Figura 8.

3.2 ESTIMACION

El método de máxima verosimilitud, para los procesos ARIMA(p,d,q) de cualquier orden, permite la estimación de los parámetros

$$\begin{aligned}\phi &= (\phi_1, \dots, \phi_p), \theta_0, \\ \theta &= (\theta_1, \dots, \theta_q) \text{ y } \sigma_a^2.\end{aligned}$$

Se recordará , de manera general, que el método de estimación por máxima verosimilitud selecciona como estimador del parámetro a aquel valor que tiene la propiedad de maximizar el valor de la probabilidad de la muestra aleatoria observada, es decir, encuentra el valor del parámetro que maximiza la función de verosimilitud; además suele generar estimadores insesgados de mínima varianza.

Para utilizar este método se incorpora al modelo una media diferente de cero, se asume que el proceso de ruido blanco tiene una distribución normal, con media cero y varianza σ_a^2 , de esto se deriva que Z_t tiene una distribución normal mutivariada.

Para un modelo ARIMA la verosimilitud para los parámetros ϕ, μ (que es θ_0), θ y σ_a^2 , se expresa como

$$L(Z_1, \dots, Z_{t-1}, Z_t | \mu, \phi, \theta, \sigma_a^2) = (2\pi\sigma_a^2)^{-N/2} \exp\left[-\frac{1}{2\sigma_a^2} \sum_{t=1}^N a_t^2\right]$$

Por analogía con la regresión lineal clásica, los parámetros pueden estimarse minimizando la suma de cuadrados de los residuos (parte de las observaciones que no es explicada por el modelo). El método de mínimos cuadrados es un concepto más fácil de entender, sin embargo, la estimación por máxima verosimilitud se prefiere porque es más eficiente, aunque en la mayoría de los casos la variación entre ambas estimaciones es mínima.

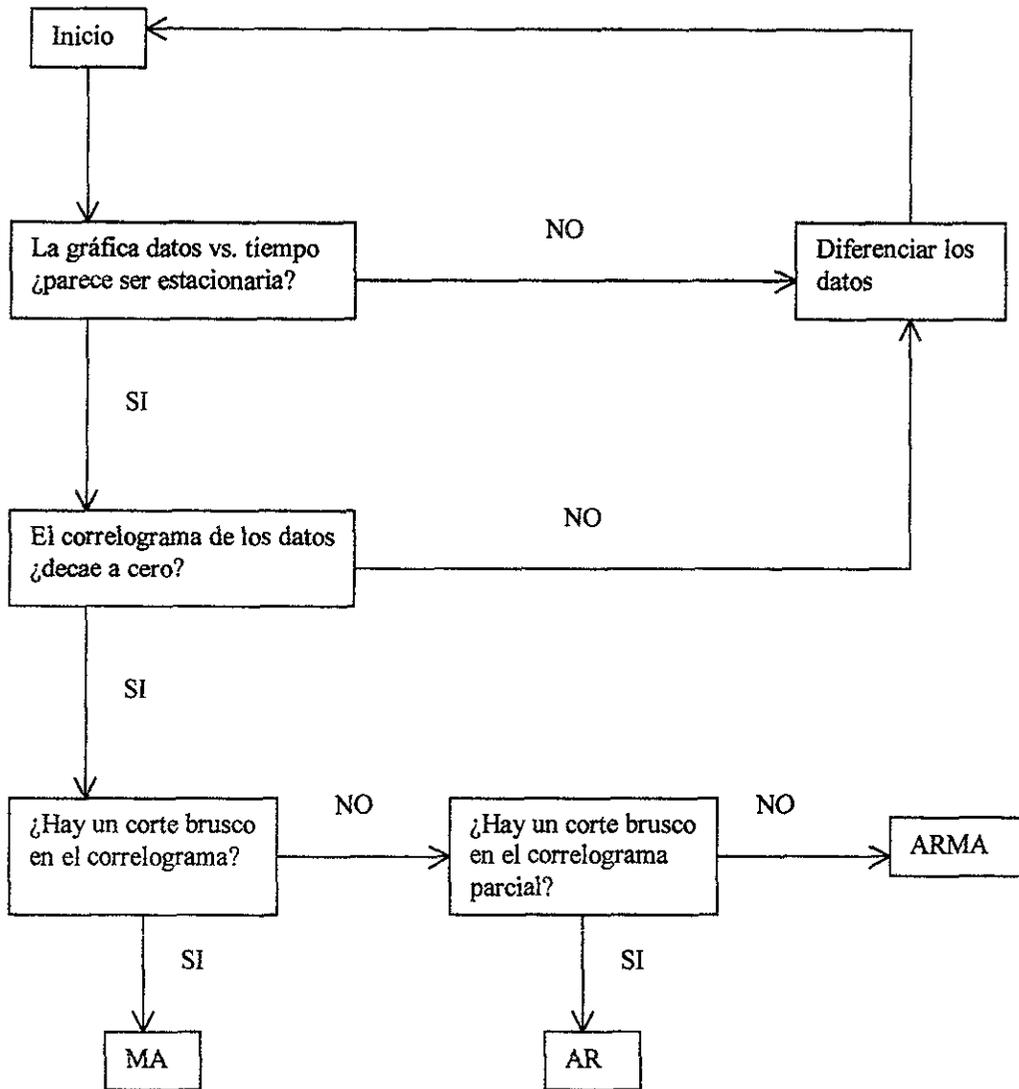


Figura 6

3.3 VERIFICACION

El objetivo de esta etapa es comprobar que no existan violaciones de consideración a los supuestos que fundamentan al modelo, de tal manera que exista la certeza de que el modelo describe razonablemente a los datos. Como un modelo es la representación de un proceso, es importante elegir aquel que parezca más adecuado (de acuerdo a los datos disponibles).

Este análisis generalmente se basa en los residuales, los cuales se definen como la diferencia entre el valor observado y el ajustado. Partiendo de que los errores aleatorios $\{a_t\}$ deberían ser una realización de un proceso de ruido blanco y, además, para muestras grandes, son casi iguales a los residuos \hat{a}_t ; entonces, para la verificación de un modelo se revisa que las características de $\{a_t\}$ las cumpla \hat{a}_t .

I. $\{a_t\}$ tiene distribución normal para toda t .

Para corroborar este supuesto se puede utilizar un histograma de residuos y revisar la simetría, o bien, elaborar una gráfica en papel normal. También se pueden graficar los residuos contra el tiempo y, si se cumple este supuesto, el 95% de las observaciones deben estar dentro del intervalo

$$(- 1.96 \sigma_a, 1.96 \sigma_a)$$

En este mismo paso se puede revisar que no existan observaciones aberrantes. Se consideran como tales a aquellas que se encuentran fuera del intervalo (de acuerdo a Guerrero, 1991)

$$(- 3 \sigma_a, 3 \sigma_a)$$

II. $\{a_t\}$, como todo ruido blanco, debe tener media cero y varianza constante.

Para verificar que la media es igual a cero, es necesario conocer a la media aritmética y a la desviación estándar muestral de los residuales, que se calculan mediante las siguientes expresiones

$$m(\hat{a}) = \sum_{t=d+p+1}^N \frac{\hat{a}_t}{(N - d - p - q)}$$

$$\sigma_a = \sqrt{\sum_{t=d+p+1}^N \frac{[\hat{a}_t - m(\hat{a})]^2}{(N - d - p)}}$$

que se utilizan para determinar si la media de los residuales es significativamente distinta de cero, esto se presenta si se cumple la siguiente desigualdad

$$\left| \frac{\sqrt{N - d - p} \bar{m}(\hat{a})}{\sigma_a} \right| \geq 1.96$$

De ocurrir lo anterior, será necesario un parámetro más.

La varianza constante se puede comprobar de manera visual en una gráfica de residuos contra el tiempo, en donde se observa si existe la presencia de un patrón y de ser así, se requerirá de una transformación.

III. Las variables aleatorias $\{a_t\}$ son no correlacionadas.

Esto implica autocorrelación cero, para poder determinarlo se construye una FAC empleando a los residuales y se estima el error estándar de $r_k(\hat{a})$

$$\sqrt{V \hat{a} r [r_k(\hat{a})]} = \frac{1}{\sqrt{N - d - p}}$$

y si se cumple que

$$|r_k(\hat{a})| \geq \frac{1.96}{\sqrt{N - d - p}}$$

con

$$r_k(\hat{a}) = \frac{\sum_{t=t'}^{N-k} \hat{a}_t \hat{a}_{t+k}}{\sum_{t=t'}^N \hat{a}_t^2}$$

entonces se tiene que la autocorrelación de orden k es significativamente diferente de cero al 95%. También se puede emplear el estadístico Q' (encontrado por Box y Ljung en 1978)

$$Q' = \frac{(N - d - p)(N - d - p + 2) \sum_{k=1}^k r_k^2(\hat{a})}{N - d - p - k}$$

el cual, si $k > 20$, tiene aproximadamente una distribución χ^2_{k-p-q} , así pues, si $Q' > \chi^2_{k-p-q}$ el modelo no es adecuado y será necesario revisar los residuales para encontrar algún indicio del cambio que será necesario realizar.

Además, es conveniente verificar que todos los parámetros son necesarios, por ello para cada parámetro se puede construir un intervalo de confianza al 95%

$$\left(\hat{\theta} - 1.96 \sqrt{\text{Var}(\hat{\theta})}, \hat{\theta} + 1.96 \sqrt{\text{Var}(\hat{\theta})} \right)$$

si el cero se encuentra dentro del intervalo es posible que el parámetro no sea necesario.

También hay que cuidar que los coeficientes estimados ϕ y θ no estén altamente correlacionados unos con otros, si lo están se tiende a producir inestabilidad en el modelo. El cuidado de estos aspectos ayudan a mantener el principio de parsimonia, el cual establece que el modelo debe ajustar los datos adecuadamente sin el empleo de parámetros innecesarios, es decir, que el modelo involucre el menor número de parámetros. Finalmente, para que el modelo sea admisible hay que verificar las condiciones de estacionariedad y/o invertibilidad.

Además, si se considera que la raíz del error cuadrado medio (RMSE) estima a σ_a , ésta resulta ser de ayuda al elegir un modelo pues indica cuál tiene la menor varianza en el error.

Otra herramienta útil para seleccionar el modelo es el criterio de Akaike (AIC), el cual penaliza a la log verosimilitud condicional del modelo (condicionada a las primeras m observaciones de la serie) por la introducción de nuevos parámetros que realmente no aporten significativamente al modelo (Brockwell, 1987). Se define como:

$$AIC = -2 \log L(x_{m+1}, \dots, x_n | x_1, \dots, x_m) + 2r$$

donde r es el número de parámetros estimados (número de coeficientes). El mejor modelo será aquel que presente el índice AIC de menor valor. Para evaluar si la diferencia entre dos valores de este índice es significativa, dicha diferencia se compara con el de una $\chi^2_{\alpha, m}$, donde m representa el incremento en el número de parámetros de un modelo a otro.

4 ANALISIS ESTADISTICO DEL OZONO EN LA ZONA SUROESTE DE LA CIUDAD DE MEXICO

La información que se presenta corresponde a la zona suroeste de la Ciudad de México, aquí se encuentra una estación de monitoreo en el Pedregal de San Angel, dentro del Colegio Alexander Bain. Esta región de la ciudad es considerada como una de las más críticas, pues con frecuencia presenta los índices más altos de contaminación por ozono, por esta razón, se considera importante su análisis. La serie está formada por los promedios semanales de los registros diarios máximos de ozono, reportados como IMECA. La Tabla 3 muestra los valores que constituyen la serie y la Figura 7 su gráfica.

El valor de la media muestral es de 153.363 puntos IMECA, con una varianza muestral de 1,043.0541. La inspección visual de la gráfica no muestra indicios de varianza no homogénea ni de comportamiento periódico, aunque se puede sospechar de una ligera tendencia, pero es necesario confirmar estos hechos.

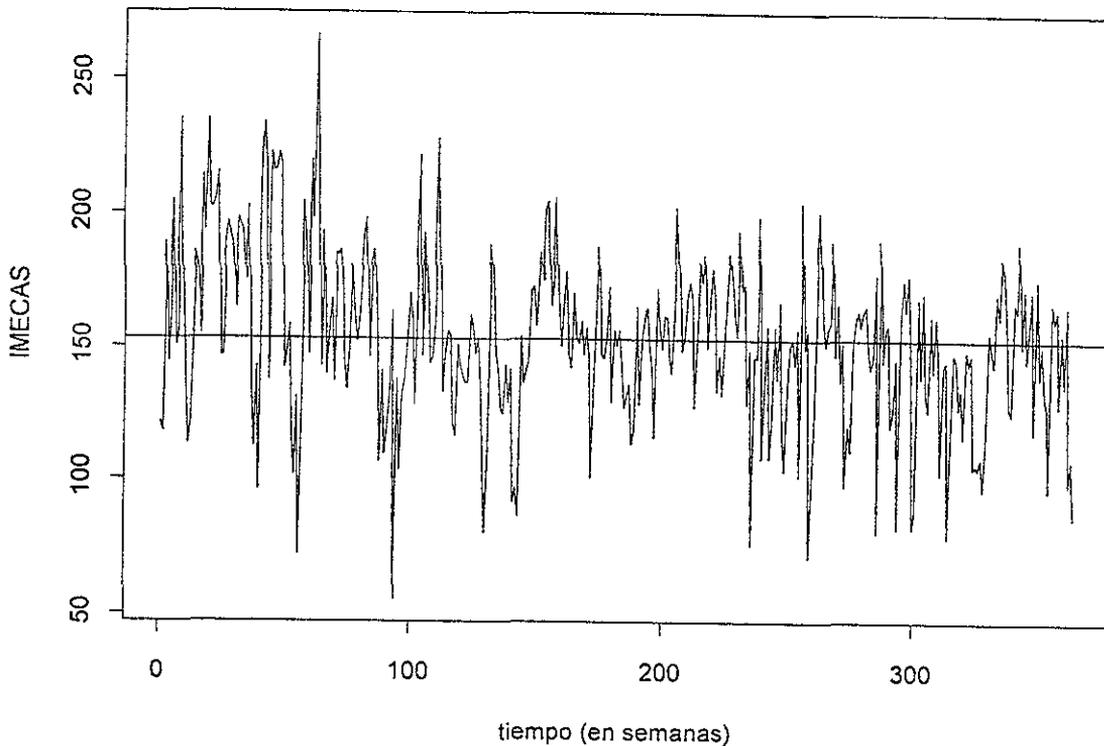


Figura 7. Ozono de la zona suroeste de la Cd. de México

120.857	234	198.429	162.286	170.571	202.714	167.857	145.143	113.571
117.571	217.286	146.286	157.143	154.286	177	118.857	157.714	156.143
189	137.429	181.857	147.571	152	149.286	103.571	158.857	149.286
144.429	222.571	186.571	153.571	160.143	154.429	133.857	120	143.571
165.143	216.429	158.857	136.571	147.571	169.429	151.286	125.571	170.714
205	216.286	106.571	109.429	157.714	174.857	153	145.857	162.143
150.714	222.714	140.714	80.2857	145.714	169.286	144	82.2857	184.143
156.286	218.143	109.429	109.857	101	127.429	157.286	161.571	178.714
235.286	141.857	117.857	188.429	136.857	154.429	101.571	175.714	159.143
168.571	148.571	130.857	180.857	171	182.714	204.714	164.714	127.286
135.571	158.714	163.714	148.429	187.857	175.286	150.286	177.429	125
113.143	116.857	55.4286	141.143	147.571	185.143	156.571	145.286	167.429
121.714	101.857	138	126.143	146.286	150.429	71.2857	82.1429	164.571
159.857	130.714	103.714	124.429	154.714	166.857	99.4286	87.7143	189.714
185.714	72.1429	131.571	143.143	173	180.143	125.714	169	151.286
181.286	128.857	138.143	128.857	129.143	168.714	181.857	139	172.571
154.857	204.571	154	141.571	156.714	133.714	201.143	170.857	145.571
214.143	192.286	170.143	91.7143	151	147.143	161.429	133.143	158.571
193.714	147.571	161.143	96.8571	156.857	132.143	151.143	125.857	171.714
235	219.857	128.143	86.5714	144.286	154.714	157.857	162.429	117.714
202.857	198.286	178.143	154.429	127.143	171.714	160	141.286	176.143
202.286	266.714	221.857	136.571	131.571	185.714	190.143	161.714	139.286
205.714	207.714	146.857	141.143	135.857	175.714	147.286	143.143	151.286
215.429	142.714	192.857	144.857	113.286	162	166.714	102.429	132.571
146.571	193.286	169.571	171.571	118.571	154.571	137.429	143.286	127
147.571	139.857	144.143	173.143	165.714	193.857	152.143	145.143	96.2857
188.714	159.714	146.429	158.571	128.429	172.143	97.8571	78.7143	167.286
197	168.143	163	175	155	174.143	119.857	111.286	160.857
191.857	137	228.143	185.857	162	128.571	110.714	148	164.857
188.857	185.571	193.286	175.571	165.143	149.143	152.714	145.429	128
165.143	184.714	133	201.714	150	75.8571	161.429	127	155.714
198.571	186.429	149.143	205	140.857	146.714	164.143	133.857	135.857
196.286	145.143	156.429	166.143	115.857	146.286	158.571	116.143	166.429
193.429	133.857	154.571	176.429	172.429	199.143	163.857	149.286	98.2857
175.571	148.714	121	206.143	157.429	108	166	144.286	107.429
202.857	181	116.286	167.857	150.143	134.571	150	147.714	86.2857
139	162.143	151	150.857	162.286	158.429	142.429	104.714	
112	152.286	142.857	169.429	161.714	108.143	145.857	105.571	
142.429	161.714	139	178.571	146.143	120.429	177.714	104.429	
96.2857	177.857	136.286	147.857	140.286	158.286	80.4286	108	
221.429	191.143	136.714	142.429	166.143	138.571	190.571	96.7143	

Tabla 3. Promedio semanal de los registros máximos diarios de ozono (reportados como IMECA), desde enero de 1991 hasta diciembre de 1997.

4.1 IDENTIFICACION

Cómo primer paso es necesario establecer que el proceso sea estacionario, para ello se verifica que la varianza sea constante y que el nivel no cambia en el tiempo.

Con el propósito de determinar si la serie requiere de algún tipo de transformación para estabilizar la varianza, los datos se dividieron en 7 grupos ($H=7$) con 50 elementos cada uno ($R=50$) quedando fuera de los cálculos 14 datos ($N=364$ y $n=14$). La media y desviación estándar correspondientes se muestran en la Tabla 4a. Con esta información y escogiendo 10 valores para λ ($q = 10$), se calculó el coeficiente de variación (Tablas 4b y 4c), cuyo valor más pequeño se obtuvo para un $\lambda = -0.2$. Considerando que este método sólo da estimaciones aproximadas de dicho parámetro, se optó por tomar $\lambda = 0$ cuya transformación correspondiente es

$$T(Z_t) = \ln(Z_t)$$

Una vez realizada la transformación para estabilizar la varianza, se volvió a aplicar el método anteriormente descrito, bajo las mismas condiciones de H , R y λ . Los resultados se muestran en las Tablas 5a, 5b y 5c, en donde se puede apreciar que los coeficientes de variación son muy similares, considerándose estabilizada la varianza.

Ya que se logró la estabilización de la varianza, es necesario evaluar si la media del proceso no depende del tiempo (Figura 8). En la gráfica de la FAC muestral (de la serie ya transformada), se aprecia una caída que puede ser considerada como suave pero sin llegar a ser lenta (Fig. 9), comportamiento que sugiere que no es necesario diferenciar.

Para confirmar los resultados de la herramienta anterior, se procedió a aplicar el criterio de la presencia de sobrediferenciación si hay un incremento de la varianza. Para ello, se realizaron dos diferencias de $\ln Z_t$ ($\nabla \ln Z_t$ y $\nabla^2 \ln Z_t$) y se calcularon sus correspondientes varianzas. Al compararlas, se observó que incrementaban conforme aumentaba el grado de diferenciación (Tabla 6). Este hecho indica que no se requiere de transformar para estabilizar la media.

Grupo	Media	Desviación Estándar
1	178.0886	37.3334
2	155.5629	37.8820
3	146.9000	29.9012
4	155.6514	22.5340
5	154.3714	25.8723
6	146.9343	30.7561
7	141.4086	26.4233

Tabla 4a. Medidas de tendencia central de cada grupo.

λ	-1	-0.8	-0.6	-0.4	-0.2	0
Grupo 1	0.0011771	0.0033186	0.0093558	0.0263759	0.0743592	0.209634
Grupo 2	0.0015654	0.0042954	0.0117865	0.0323419	0.0887455	0.2435159
Grupo 3	0.0013856	0.0037588	0.0101966	0.0276604	0.0750348	0.2035481
Grupo 4	0.0009301	0.0025525	0.0070047	0.0192231	0.0527538	0.144772
Grupo 5	0.0010857	0.0029745	0.0081494	0.0223275	0.0611722	0.1675975
Grupo 6	0.0014246	0.0038646	0.0104842	0.0284419	0.0771583	0.2093184
Grupo 7	0.0013214	0.0035574	0.0095769	0.0257822	0.069409	0.1868577
Media	0.00127	0.0034745	0.0095077	0.0260218	0.0712333	0.1950348
σ	0.0002181	0.0005828	0.0015683	0.0042532	0.0116336	0.0321094
C.V.	0.1716979	0.1677454	0.1649553	0.1634468	0.1633175	0.1646341

Tabla 4b. Cálculo del coeficiente de variación.

λ	0.2	0.4	0.6	0.8	1
Grupo 1	0.591002	1.6661578	4.6972464	13.242517	37.333419
Grupo 2	0.6682032	1.8335371	5.0311919	13.805498	37.882031
Grupo 3	0.5521682	1.4978757	4.0633115	11.022611	29.901215
Grupo 4	0.3972968	1.0902991	2.9921013	8.2112054	22.533961
Grupo 5	0.4591779	1.2580399	3.4467342	9.4432432	25.872271
Grupo 6	0.5678478	1.540482	4.1790854	11.337202	30.756046
Grupo 7	0.5030443	1.3542578	3.6458307	9.8150303	26.423284
Media	0.5341057	1.4629499	4.0079288	10.982472	30.100318
σ	0.0894239	0.2511739	0.7109323	2.0255365	5.8021424
C.V.	0.1674273	0.17169	0.1773815	0.1844336	0.1927602

Tabla 4c. Cálculo del coeficiente de variación (continuación).

Grupo	Media	Desviación Estándar
1	5.1585	0.226235
2	5.0139	0.272575
3	4.9686	0.211301
4	5.0370	0.148661
5	5.0237	0.185825
6	4.9643	0.240374
7	4.9329	0.201025

Tabla 5a. Medidas de tendencia central de cada grupo.

λ	-1	-0.8	-0.6	-0.4	-0.2	0
Grupo 1	0.008502	0.011804	0.016388	0.022752	0.031589	0.226235
Grupo 2	0.010842	0.014968	0.020663	0.028525	0.039379	0.272575
Grupo 3	0.008559	0.011794	0.016253	0.022396	0.030861	0.211301
Grupo 4	0.005859	0.008096	0.011187	0.015458	0.021359	0.148661
Grupo 5	0.007363	0.010169	0.014043	0.019394	0.026784	0.185825
Grupo 6	0.009754	0.013438	0.018514	0.025508	0.035144	0.240374
Grupo 7	0.008261	0.011368	0.015642	0.021523	0.029616	0.201025
Media	0.008449	0.011662	0.016099	0.022222	0.030676	0.212285
σ	0.0016	0.002204	0.003037	0.004185	0.005768	0.039779
C.V.	0.189412	0.189002	0.188637	0.188316	0.188042	0.187387

Tabla 5b. Cálculo del coeficiente de variación.

λ	0.2	0.4	0.6	0.8	1
Grupo 1	0.060889	0.084537	0.117368	0.16295	0.226235
Grupo 2	0.075048	0.103604	0.143025	0.197446	0.272575
Grupo 3	0.058602	0.080753	0.111277	0.153339	0.211301
Grupo 4	0.040781	0.05635	0.077862	0.107587	0.148661
Grupo 5	0.051084	0.070549	0.09743	0.134555	0.185825
Grupo 6	0.066711	0.091912	0.126632	0.174468	0.240374
Grupo 7	0.056075	0.07716	0.106172	0.146093	0.201025
Media	0.058456	0.080695	0.111395	0.153777	0.212285
σ	0.010968	0.01513	0.020877	0.028814	0.039779
C.V.	0.187632	0.187498	0.187412	0.187375	0.187387

Tabla 5c. Cálculo del coeficiente de variación (continuación).

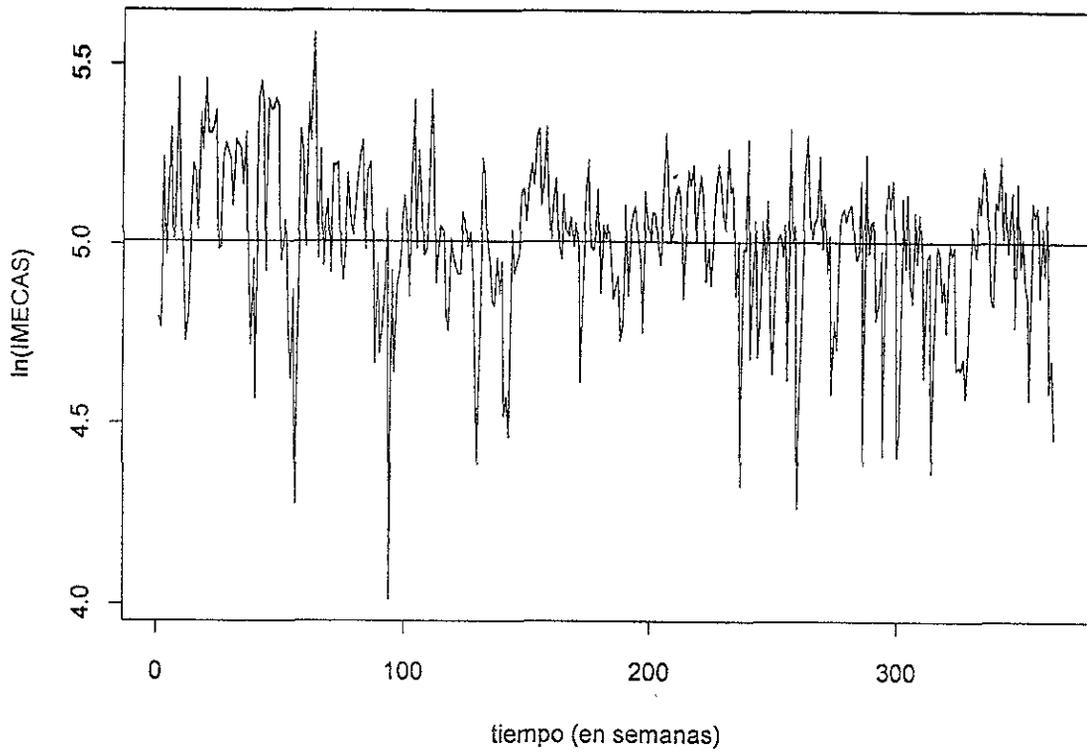


Figura 8. Serie de ozono con varianza estabilizada

Parámetro	$\ln Z_t$	Primera Diferencia	Segunda Diferencia
Media	5.00886	-0.00093	-0.00053
Varianza	0.05087	0.06237	0.17455

Tabla 6. Medidas de tendencia central para las diferencias.

Ahora se tiene un proceso que es considerado como estacionario: con varianza constante y sin cambios en el nivel a lo largo del tiempo. Para proceder a asociarle un modelo ARIMA, primero se hace el cálculo de μ (5.00886), que es utilizada para obtener

$$\tilde{Z}_t = Z_t - \mu$$

y con esta nueva variable, se realiza el ajuste del modelo.

Para ello se examina la FAC y la FACP muestrales. En el gráfico de la primera función, se observa, como ya se había mencionado con anterioridad, una caída ligeramente lenta hacia cero (Figura 9); por el contrario, el correlograma parcial muestra una caída a cero después del segundo retraso, aunque se presentan correlaciones que nuevamente rebasan las bandas de confianza a el retraso 7 (Figura 10). Debido a que en el correlograma parcial los dos primeros retrasos salen de las bandas de confianza, se propone un modelo AR de orden 2.

Así pues, el modelo propuesto para modelar la serie transformada es un AR(2):

$$\tilde{Z}_t = \mu + \phi_1 \tilde{Z}_{t-1} + \phi_2 \tilde{Z}_{t-2} + a_t$$

si se escribe esta expresión en función de Z_t queda como

$$Z_t = (1 - \phi_1 - \phi_2)\mu + \phi_1 Z_{t-1} + \phi_2 Z_{t-2} + a_t$$

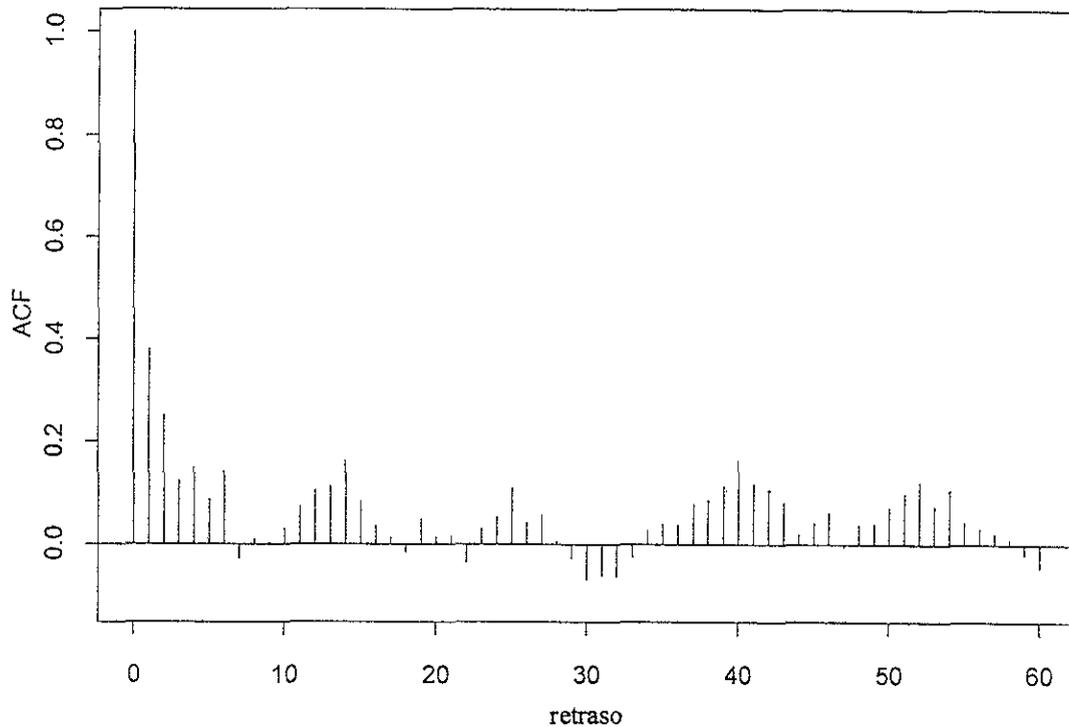


Figura 9. Autocorrelación muestral del ozono.

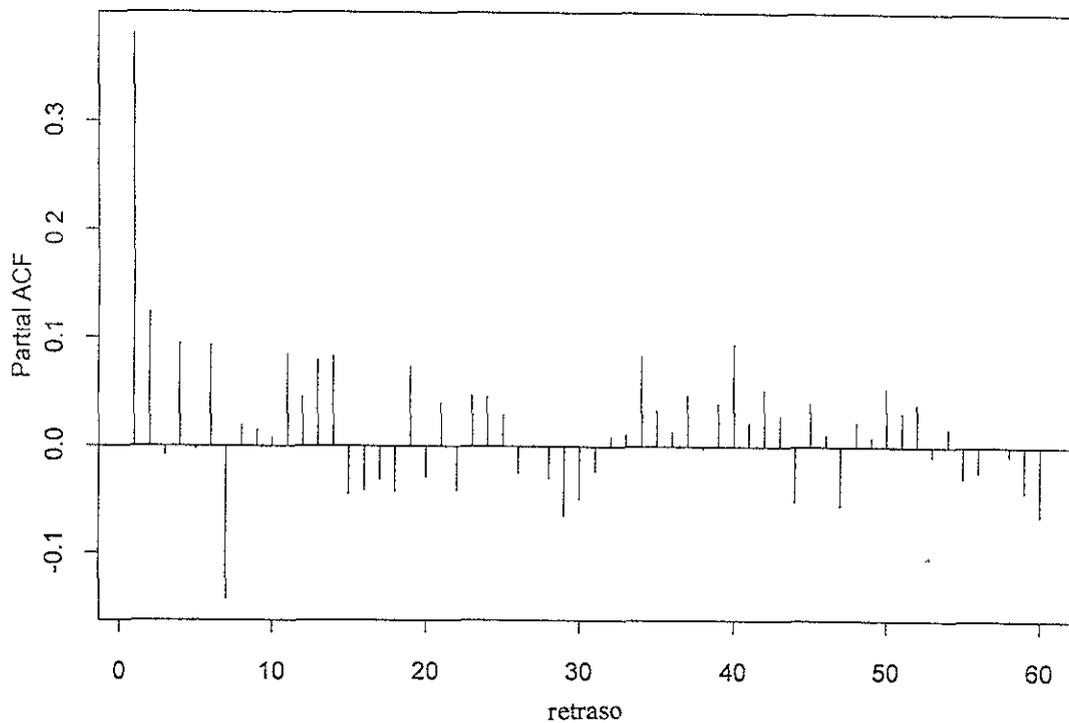


Figura 10. Autocorrelación parcial muestral del ozono.

4.2 ESTIMACION

Los parámetros estimados para el modelo se muestran en la Tabla 7 con sus correspondientes valores de significancia, las cuales indican que en ambos coeficientes se rechaza la hipótesis nula (igualdad a cero).

Coficiente	Estimación	Varianza	Normalización (valor absoluto)	Valores p
ϕ_1	0.3353	0.0027	6.4324	$6.28 \cdot 10^{-11}$
ϕ_2	0.1283	0.0027	2.4392	0.0074

Tabla 7

El modelo estimado que se obtiene es:

$$\hat{Z}_t = (1 - 0.3353 - 0.1283)\hat{\mu}_t + 0.3353\hat{Z}_{t-1} + 0.1283\hat{Z}_{t-2} + \hat{a}_t$$

4.3 VERIFICACION

El comportamiento de los residuos estandarizados permite vislumbrar si el modelo es apropiado. La gráfica de los residuos estandarizados contra el tiempo (Figura 11) muestra un comportamiento de ruido blanco, pues no hay evidencia de varianza no constante ni de presencia de tendencia o ciclos, que en este caso implica una media igual a cero y se corrobora al calcular la media aritmética.

Además, la mayoría de los valores se ubican entre la franja delimitada por las bandas en ± 1.96 ; existen 20 valores que rebasan las 2 desviaciones estándar y representan el 5.52%, pero sólo 6 de ellos (1.66%) llegan más allá de las 3 desviaciones estándar y corresponden a las observaciones 36, 94, 240, 259 y 286. Siguiendo el criterio de Guerrero (1991), estas observaciones se pueden considerar como aberrantes, por lo que sería conveniente sondear las posibles causas de tales datos, aunque estas pueden ser muy amplias debido a la naturaleza del fenómeno (toma de información, transformación a IMECA, días feriados, inversión térmica, época de nortes, etc.).

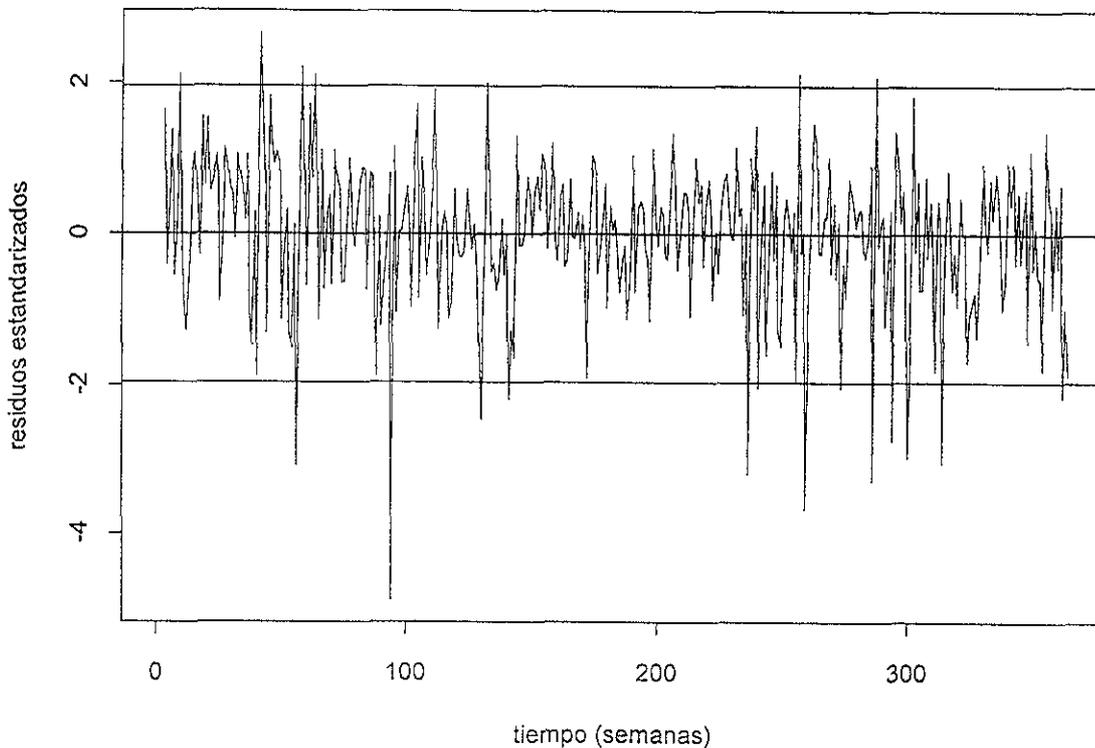


Figura 11. Residuos estandarizados del modelo de ozono

En la última gráfica (Figura 11) no es posible evaluar la correlación existente, para ello se revisa la FAC de los residuos (Figura 12). En él, se aprecia que las autocorrelaciones caen dentro de las bandas de confianza (característica de un ruido blanco) a excepción de las correspondientes a los retrasos 6 y 14, así que se ajustó un modelo multiplicativo para incorporarlas y analizar su significancia, será expuesto más adelante.

El estadístico Q' (Ljung y Box) permite evaluar si los residuos son mutuamente independientes. El valor calculado de Q' (para $K=27$) es de 27.15, y es inferior al de $\chi^2_{0.05, 20}$, cuyo valor es 31.42. Este resultado indica que existe independencia y se tiene un proceso de ruido blanco. Debido a que K es un parámetro arbitrario, se probó el valor de Q' con diferentes valores para K , y se encontró que en todos los casos la conclusión era la misma: se aceptaba la independencia de los residuos.

Para determinar la normalidad de los residuos, se empleó una gráfica en papel normal (Figura 12), en ella se aprecia que existe parecido entre la distribución de los residuos y la recta, las desviaciones presentes no se consideran graves. Además, hay que

recordar que únicamente el 5.52% de las observaciones se encontraban más allá de las 2 desviaciones estándar, así que aproximadamente el 95% de los datos están dentro de ellas, y esto es típico de una distribución normal.

Por último se verificó la invertibilidad del modelo encontrando las raíces de la ecuación:

$$1 + 0.3353 x + 0.1283 x^2 = 0$$

encontrándose que todas las raíces se hallaban fuera del círculo unitario, lo que determina que es un proceso invertible y que tiene una FAC única.

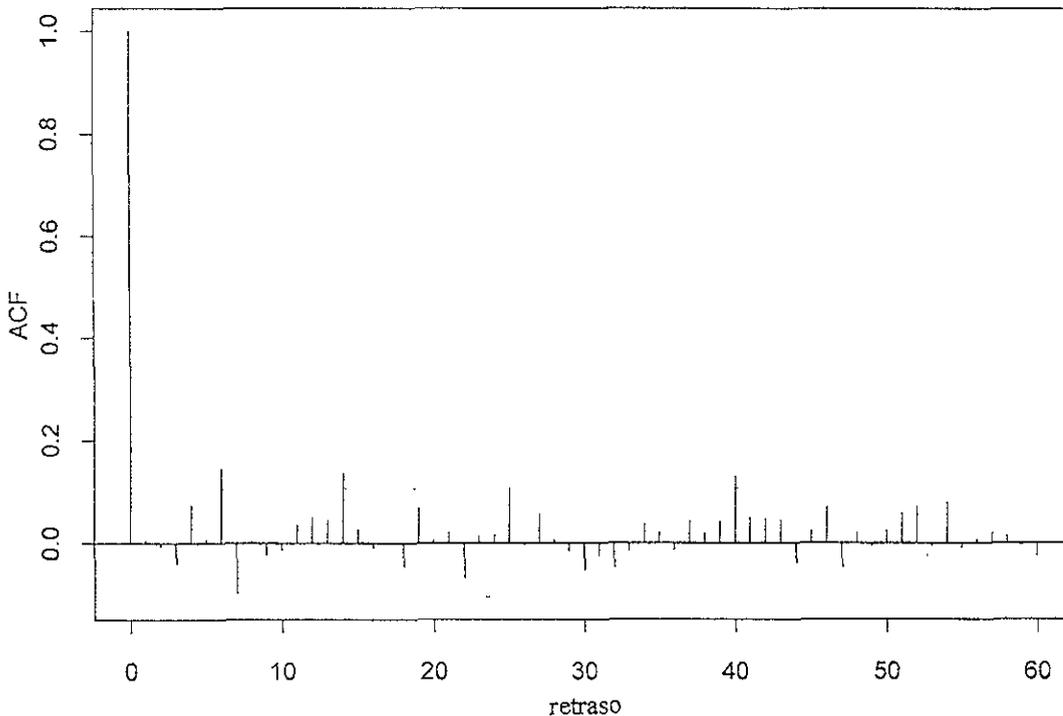


Figura 12. Autocorrelación parcial de los residuos del modelo de ozono.

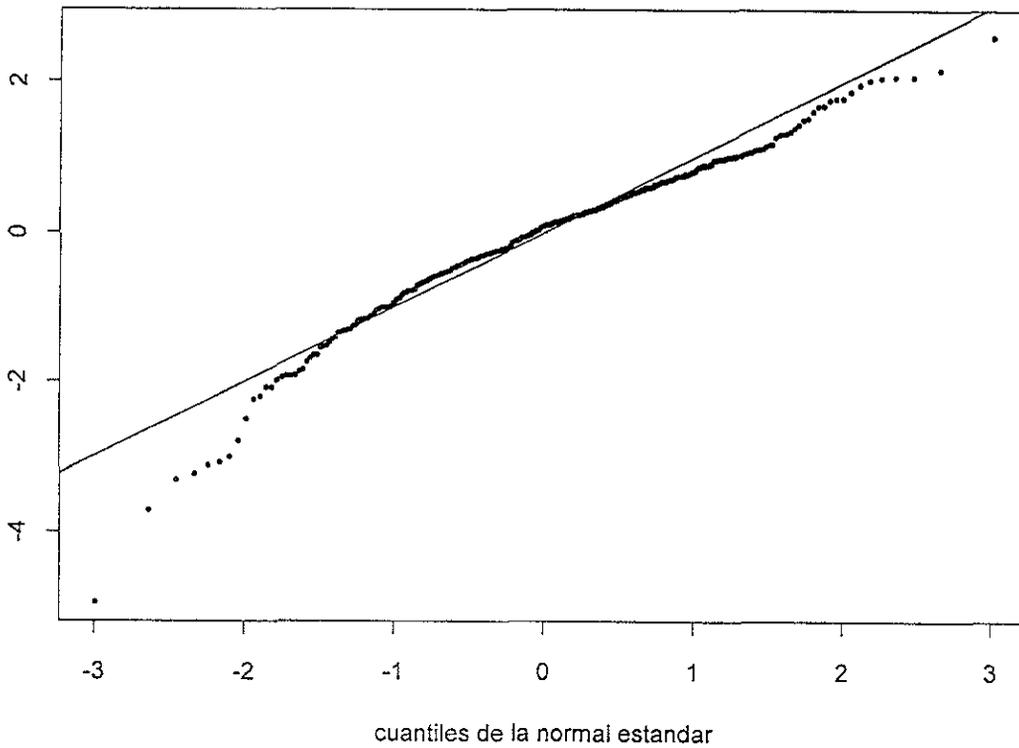


Figura 13. Gráfica en papel normal de los residuos del modelo de ozono.

El modelo multiplicativo que se analizó incorpora el retraso 6 empleando una componente de media móviles de periodo 6, obteniéndose un modelo

$$\text{ARIMA}(2,0,0)(0,0,1)_6$$

que expresado algebraicamente es equivalente a tener:

$$(1 - \phi_1 B - \phi_2 B^2) \tilde{Z}_t = (1 - \Theta B^6) \alpha_t$$

Se realizó la estimación de los parámetros y se obtuvo que el parámetro de medias móviles, justamente el que incorpora el retraso 6, es significativo (Tabla 8).

Coefficiente	Estimación	Varianza	Normalización (valor absoluto)	Valores p
ϕ_1	0.3493	0.00273	6.6858	0.0000
ϕ_2	0.1092	0.00273	2.0871	0.0184
Θ_1	-0.1368	0.00272	2.6245	0.0043

Tabla 8

El modelo estimado que se obtiene es

$$\hat{Z}_t = (1 - \phi_1 - \phi_2)\hat{\mu} + \phi_1\hat{Z}_{t-1} + \phi_2\hat{Z}_{t-2} + \Theta_1 a_{t-6} + a_t$$

Del mismo modo que para el modelo AR(2) (que será denominado como modelo 1), se verificaron los supuestos para el modelo 2 (ARIMA (2,0,0)(0,0,1)₆). Para facilitar una comparación entre ambos modelos, los resultados de la verificación se muestran resumidos en la Tabla 9 y en las Figuras 14, 15 y 16 se observa el comportamiento de los residuales del modelo 2.

	Modelo 1	Modelo 2
Parámetros estimados	$\phi_1 = 0.3353$ $\phi_2 = 0.1283$	$\phi_1 = 0.3493$ $\phi_2 = 0.1092$ $\Theta_1 = -0.1368$
Varianza del modelo	0.04265	0.04179
No. residuales > 1.96	20	21
No. residuales > 3	6	5
AIC	-110.69	-115.93
Q'	27.15 (k=27)	20.20 (k=23)

Tabla 9

Al verificar la invertibilidad del segundo modelo, se encontró que las raíces de las ecuaciones

$$1+0.3493x + 0.1092x^2 = 0$$

$$1-0.1368x = 0$$

caían fuera del círculo unitario.

En la Tabla 9 se aprecia que el modelo con el índice más pequeño es el del modelo ARIMA(2, 0, 0)(0, 0, 1)₆, y la diferencia entre éste y el AR(2) es de 5.2452. Si se compara este valor con una $\chi^2_{0.05,1} = 3.84$, se concluye que existe diferencia significativa entre ambos valores del índice AIC. Por esta razón, se considera al modelo multiplicativo más adecuado vía AIC, ya que incorpora la correlación de retraso 6. Sin embargo, las demás características de ambos modelos son muy semejantes.

Para fines descriptivos del proceso, el modelo AR(2) permite una interpretación más sencilla del fenómeno analizado, pues relaciona las 2 últimas semanas con un tiempo dado, y ésto se puede vincular con el ozono, ya que una de sus características es que tiene una gran permanencia.

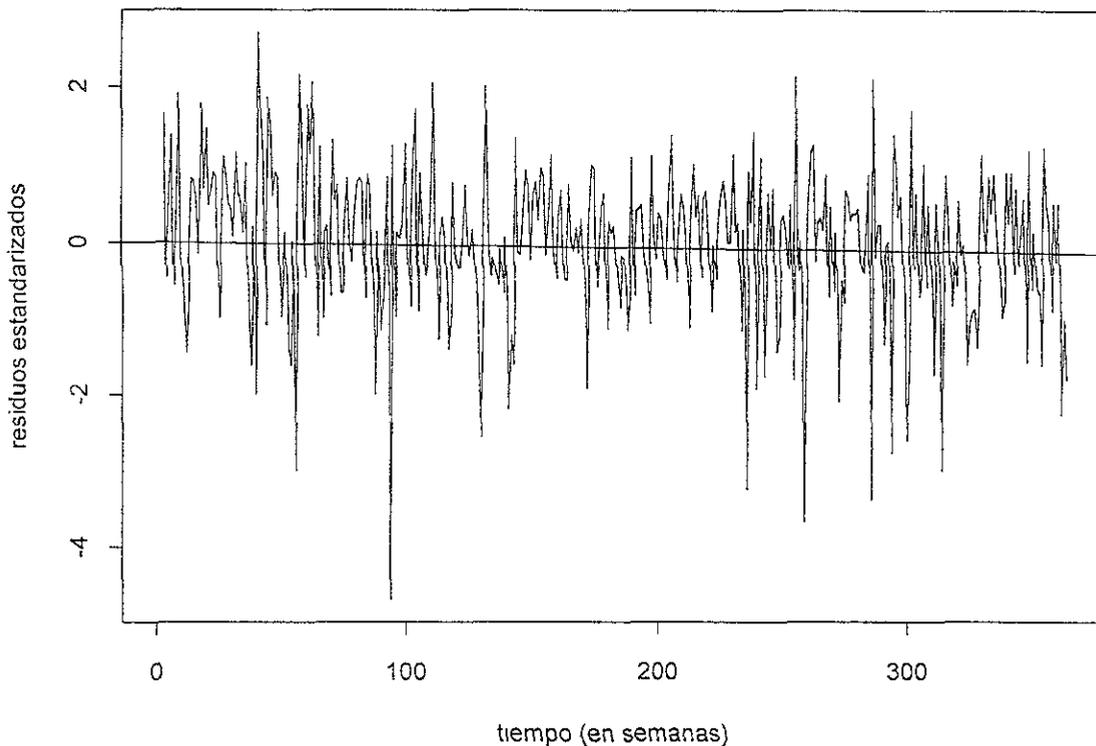


Figura 14. Residuos estandarizados del modelo multiplicativo.

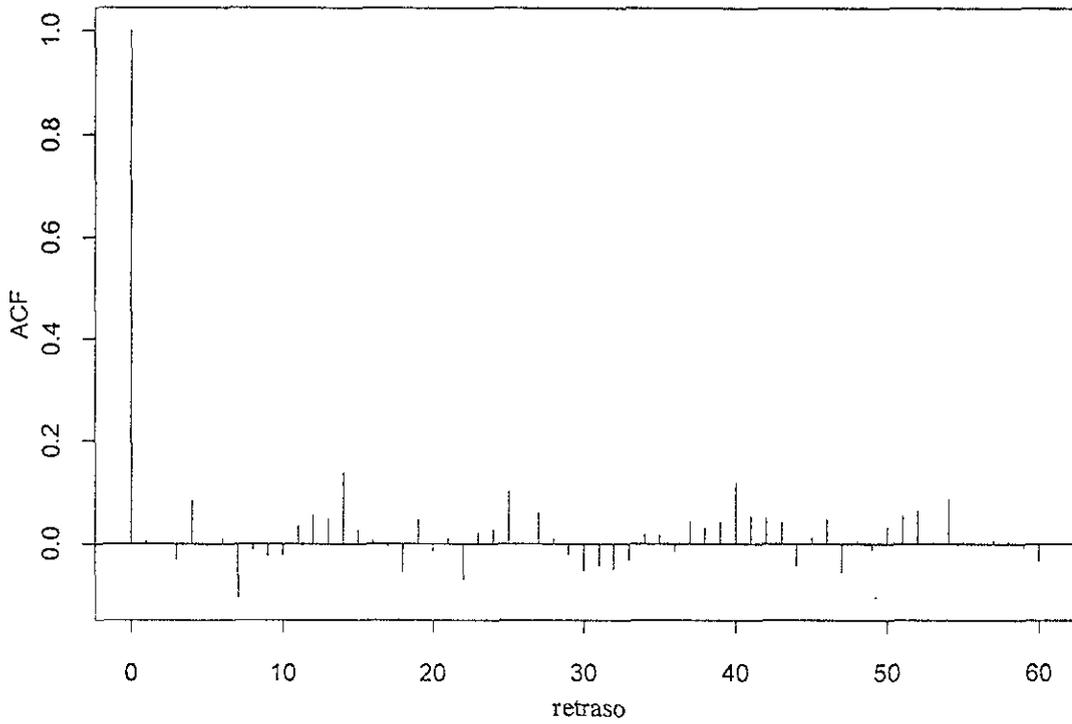


Figura 15. Autocorrelación de los residuos del modelo multiplicativo.

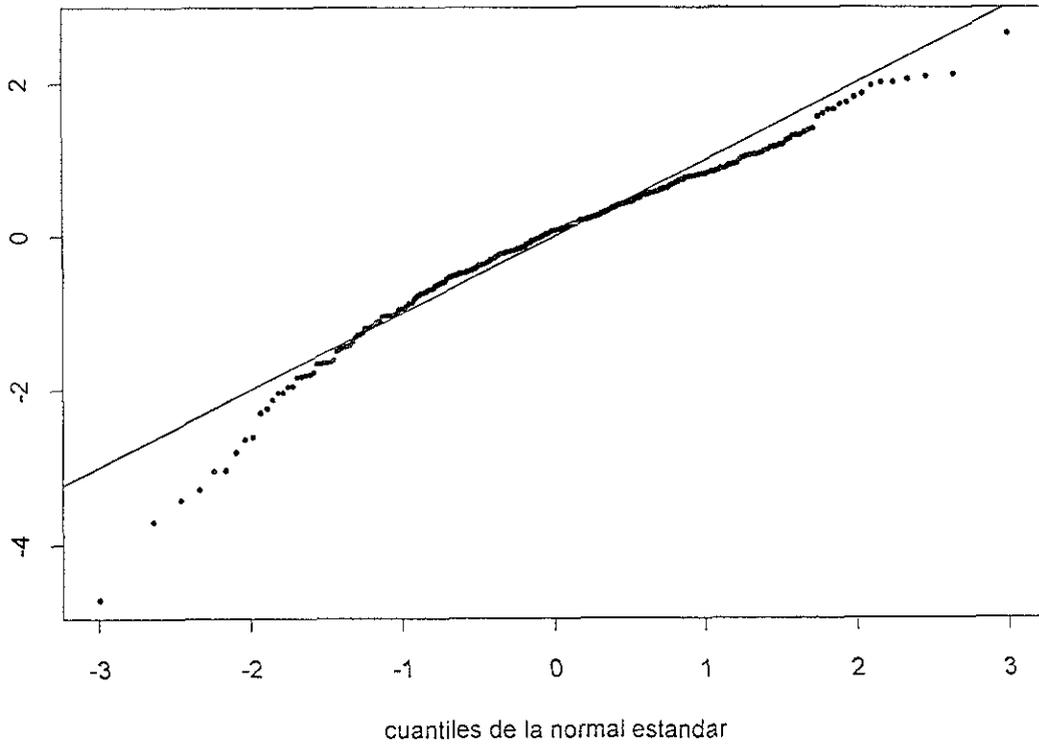


Figura 16. Gráfica en papel normal de los residuos del modelo multiplicativo.

4.4 PRONOSTICOS

Se realizaron predicciones para ambos modelos para las siguientes 2 semanas a partir la última de 1997; la primera semana pronosticada está constituida por los cuatro últimos días de 1997 y los 3 primeros de 1998, y de aquí en adelante se continúan formando las semanas para el presente año. Los pronósticos calculados y sus intervalos de confianza se reportan en la Tabla 10.

Modelo	Pronóstico	Límite inferior	Límite superior	Valor real
1	119.279	79.537	178.797	136.43
	129.269	84.348	198.111	122.14
2	120.79	80.908	180.322	136.43
	129.926	86.254	195.708	122.14

Tabla 10

Para ilustrar la manera en que se calcularon los pronósticos, a continuación se muestra el procedimiento que se siguió para obtener los del modelo 1. Para ello, se utilizan el modelo propuesto y los valores de Z_t .

El modelo general para pronósticos con el modelo AR(2) queda determinado por

$$\hat{Z}_{t+l} = (1 - 0.3353 - 0.1283)\mu_{t+l} + 0.3353\hat{Z}_{t+l-1} + 0.1283\hat{Z}_{t+l-2} + \hat{a}_{t+l}$$

Para calcular el primer valor pronosticado ($l = 1$) se tiene que:

$$\mu_{t+1} = 5.008859$$

$$Z_t = 4.4577$$

$$Z_{t-1} = 4.6768$$

$$\hat{a}_{t+1} = 0$$

Sustituyendo los valores indicados Z_{t+1} (registro número 365) adquiere un valor de 4.7814522, exponenciando para obtener IMECAS se convierte en 119.2774. Las diferencias entre este valor y el reportado en la tabla radica básicamente en el manejo de los decimales, pues los mostrados en la tabla fueron calculados en computadora.

Para el segundo pronóstico ($l = 2$), se emplean los valores:

$$\begin{aligned}\mu_{t+2} &= 5.008859 \\ Z_{t+1} &= 4.7814 \\ Z_t &= 4.4577 \\ \hat{a}_{t+2} &= 0\end{aligned}$$

ahora, se obtiene un valor de 4.8619 para Z_{366} , y de aquí se obtiene 129.27 IMECA de ozono. En esta cifra, también se aprecian ligeras diferencias por redondeo.

Los intervalos (a un $\alpha = 0.05$) fueron obtenidos con la fórmula

$$\hat{Z}_n(l) \pm Z_{\alpha/2} \left(\sum_{j=0}^{l-1} \Psi_j^2 \right)^{1/2} \sigma_a$$

Para el primer pronóstico el intervalo se calcula como:

$$\begin{aligned}\hat{Z}(1) &= 4.781 \\ \psi_0^2 &= 1 \\ \sigma_a &= 0.2065228 \\ 4.781 \pm 1.96(1)(0.2065228) \\ 4.781 \pm 0.40478\end{aligned}$$

exponenciando para tener los valores en IMECA

$$(79.537, 178.797)$$

El intervalo de confianza para la segunda predicción involucra los valores de

$$\begin{aligned}\hat{Z}(2) &= 4.862 \\ \psi_1^2 &= 0.11242609\end{aligned}$$

realizando los cálculos necesario se tiene que

$$\begin{aligned}4.862 \pm 1.96(1.0547)(0.2065228) \\ 4.862 \pm 0.426933\end{aligned}$$

y una vez transformado a IMECA el intervalo queda como

$$(86.254, 195.708)$$

CONCLUSIONES

En el caso del ozono, las diversas estrategias que se han planteado no muestran efectos inmediatos, más bien, su reflejo se llega a apreciar incluso varios años después de haber sido implantadas (como es el caso del convertidor catalítico y las gasolinas sin plomo que han salido al mercado). Quizá, la sutil tendencia decreciente que presenta la serie de ozono se debe a este tipo de estrategias que han estado funcionando desde hace varios años.

El ozono de la región suroeste en la ciudad de México, fue ajustado por dos modelos ARIMA: un AR(2) y un ARIMA (2,0,0) (0,0,1)₆. En el modelo multiplicativo se empleó sólo el operador autorregresivo de la componente no estacionaria y el operador de medias móviles para la componente estacional. A este modelo se llegó después de estabilizar la varianza (aplicando logaritmo natural) y de haber ajustado primero un modelo AR(2), en el cual surge la necesidad de incorporar un nuevo parámetro. El modelo cumplió adecuadamente los supuestos, encontrándose que $\{\hat{a}_t\}$ se mostró como un ruido blanco con distribución normal, y al cubrirse la condición de invertibilidad, se aseguró la existencia de únicamente un FAC para este proceso.

Cabe mencionar que también se ajustaron otros modelos que incorporaban estos retrasos de diversas maneras, pero durante la verificación de los mismos surgían problemas como raíces unitarias, índice de AIC muy grande y parámetros no significativos. Esto lleva a pensar que es necesario evaluar todas las posibilidades de ajuste (con base en las FAC y FACP) y que con el análisis de los diversos modelos, poco a poco se irán descartando aquellos que no resulten adecuados.

Los pronósticos obtenidos con ambos modelos son muy similares, de aquí que, para fines de predicción, podría ser empleado cualquiera de ellos, aunque se espera que el modelo multiplicativo sea mejor. Pero para describir el fenómeno, el modelo AR(2) resulta mejor, pues si se considera que el ozono es uno de los contaminantes con mayor permanencia en la atmósfera, se puede comprender que existe una influencia de dos semanas anteriores sobre un tiempo dado.

Los pronósticos obtenidos por medio del análisis de series de tiempo pueden funcionar como herramientas para conocer el comportamiento de este proceso, lo que pudiera permitir un mejor planteamiento de medidas de control. Además, las predicciones pueden ser valiosas para evaluar el efecto que tienen dichas medidas sobre el ozono, básicamente, evaluando a largo plazo si se logran abatir las altas concentraciones.

Al considerar los pronósticos, estos deben ser tomados con la reserva que amerita el caso, pues en la formación de este contaminante, intervienen muchos factores tanto naturales como antropogénicos, y cualquier alteración que sufran éstos, alejándose de lo cotidiano, se verá reflejado en un comportamiento atípico del ozono. Probablemente esta sea la razón de la variabilidad que presenta la serie.

A pesar de lo anterior, se puede decir que el ozono tiene una tendencia a decrecer, sin embargo, hay que recordar que el análisis muestra que es muy pequeña, y si se desea disminuir aún más este contaminante, es necesario proponer otras vías de solución, pues se pudo ver que los niveles que por lo general se alcanzan en esta zona permanecen por arriba de los 100 IMECAS cotidianamente.

BIBLIOGRAFIA

- Box, G.E.P. and Cox, D.R. (1964). *An Analysis of Transformation*. Journal of the Royal Statistical Society B-26, 211.
- Box, G.E.P. and G.M. Jenkins. (1970). *Time Series Analysis Forecasting and Control*. San Francisco: Holden-Day.
- Brockwell, P.J. y R.A. Davis. (1987). *Time Series: Theory and Methods*. Segunda edición. EUA: Springer-Verlag.
- Chatfield, C. (1984). *The Analysis of Time Series: An Introduction*. Londres: Chapman and Hall.
- Chou, Y. (1990). *Análisis Estadístico*. México: McGraw-Hill.
- Diggle, P.J. 1990. *Time Series: A Biostatistical Introduction*. Oxford: Clarendon Press-Oxford.
- Gamboa De Buen, J. (1994). *Ciudad de México: una visión*. México: Fondo de Cultura Económica.
- Godish, T. (1986). *Air Quality*. EUA: Lewis Publishers.
- Guerrero, V.M. (1991). *Análisis Estadístico de Series de Tiempo Económicas*. México: UAM.
- Kazmier, L. (1967). *Statistical Analysis for Business and Economics*. EUA: McGraw Hill.
- Kendall, S.M. and Ord, J.K. (1990). *Time Series*. Londres: Edward Arnold.
- Pankratz, A. (1983). *Forecasting with Univariate Box-Jenkins Models. Concepts and Cases*. EUA: John Wiley & Sons.
- Quadri, G. y L.R. Sánchez. (1992). *La Ciudad de México y la Contaminación Atmosférica*. México: Limusa.
- Restrepo, I. (1992). *La Contaminación Atmosférica en México, sus Causas y Efectos en la Salud*. México: Comisión Nacional de Derechos Humanos.
- Tapia, C. (1995). *Pronósticos para Series de Tiempo con Modelos ARIMA*. Tesis de Licenciatura. UNAM.
- Wei, W.W.S. (1990). *Time Series Analysis. Univariate and Multivariate Methods*. EUA: Addison-Wesley Publishing Co.