

**UNIVERSIDAD DE EL SALVADOR
FACULTAD DE CIENCIAS NATURALES Y MATEMÁTICA
ESCUELA DE MATEMÁTICA**



**AJUSTE DE UN MODELO ARIMA A LA
EMANACIÓN DE GAS CO₂
DEL VOLCÁN DE SAN VICENTE**

TRABAJO DE INVESTIGACIÓN PRESENTADO POR:

PEDRO ARMANDO RAMOS ALBERTO

COORDINADOR DE MAESTRÍA Y ASESOR:

DR. JOSÉ NERYS FUNES TORRES

PARA OPTAR AL TÍTULO DE

MAESTRÍA EN ESTADÍSTICA

CIUDAD UNIVERSITARIA, SAN SALVADOR, DICIEMBRE 2006

**UNIVERSIDAD DE EL SALVADOR
FACULTAD DE CIENCIAS NATURALES Y MATEMÁTICA
ESCUELA DE MATEMÁTICA**



AJUSTE DE UN MODELO ARIMA A LA EMANACIÓN DE GAS CO₂ DEL VOLCÁN DE SAN VICENTE

TRABAJO DE INVESTIGACIÓN PRESENTADO POR:

PEDRO ARMANDO RAMOS ALBERTO

PARA OPTAR AL TÍTULO DE

MAESTRÍA EN ESTADÍSTICA

COORDINADOR DE MAESTRÍA Y ASESOR:

DR. JOSÉ NERYS FUNES TORRES _____

UNIVERSIDAD DE EL SALVADOR

AUTORIDADES UNIVERSITARIAS

Dra. María Isabel Rodríguez
Rectora

Ing. Joaquín Orlando Machuca
Vice-rector Académico

Dra. Carmen Rodríguez de Rivas
Vice-rectora Administrativa

Licda. Alicia Margarita Rivas de Recinos
Secretaria General

Lic. Pedro Rosalío Escobar
Fiscal

FACULTAD DE CIENCIAS NATURALES Y MATEMÁTICA

M.Sc. José Héctor Elías Díaz
Decano

Magíster Francisco Antonio Chicas
Vice-decano

ESCUELA DE MATEMÁTICA

Lic. Mauricio Hernán Lovo
Director

DEDICATORIA

A DIOS TODO PODEROSO

Que es quien me da la fortaleza espiritual para concluir todas mis metas trazadas en mi vida.

A MI ESPOSA IRMA

Que con gran sacrificio, paciencia y amor, me animo en los momentos más difíciles a salir adelante y culminar mi trabajo.

A MIS HIJOS

Alexis Armando y William Ernesto por el apoyo, cariño y ánimos que me brindaron en todos los momentos de mi estudio.

A MIS FAMILIARES

En especial a mi Tía Julia y primas por estar pendientes de la finalización de mi trabajo.

AGRADECIMIENTOS

A DIOS TODO PODEROSO

Por permitirme alcanzar una meta más en mi vida.

A MIS COMPAÑEROS Y AMIGOS

Por todas las experiencias compartidas.

A MI ASESOR

Dr. José Nerys Funes Torres por su paciencia, confianza y profesionalismo con que me oriento en el trabajo.

A LA UNIVERSIDAD DE EL SALVADOR

Por darme la oportunidad de formarme profesionalmente.

UNIVERSIDAD DE EL SALVADOR
FACULTAD DE CIENCIAS NATURALES Y MATEMÁTICA
ESCUELA DE MATEMÁTICA



**AJUSTE DE UN MODELO ARIMA A LA
EMANACIÓN DE GAS CO₂
DEL VOLCÁN DE SAN VICENTE**

TRABAJO DE INVESTIGACIÓN PRESENTADO POR:

PEDRO ARMANDO RAMOS ALBERTO

COORDINADOR DE MAESTRÍA Y ASESOR:

DR. JOSÉ NERYS FUNES TORRES

PARA OPTAR AL TÍTULO DE

MAESTRÍA EN ESTADÍSTICA

CIUDAD UNIVERSITARIA, SAN SALVADOR, DICIEMBRE 2006

PRESENTACIÓN

El presente trabajo explora la utilización de los modelos de series temporales para predecir el comportamiento de la emanación del gas dióxido de carbono del volcán de San Vicente (2003). Para este análisis se recurre a la metodología de Box-Jenkins que consiste en las fases de identificación (seleccionar todos los posibles modelos que haya generado la serie), estimación de los parámetros (cuantificar los parámetros de los modelos), revisión de los modelos (la evaluación de los modelos previamente identificados y estimados; chequeo) y finalmente la predicción (predecir los valores futuros del modelo).

Para lograr este objetivo, el trabajo se ha estructurado en tres capítulos; en el primer capítulo se aborda la descripción, la depuración y el análisis de los datos de las emanaciones de CO₂. En el segundo capítulo se hace una síntesis de los diferentes modelos ARIMA y sus características importantes de cada modelo. Seguidamente el capítulo tres trata sobre el ajuste del modelo, el análisis de los residuos y la presentación de algunas predicciones del modelo ajustado. Finalmente, se presentan las conclusiones del modelo y se comentan las perspectivas del trabajo relacionados con esta investigación.

Por otra parte, quiero agradecer grandemente al Lic. Francisco Barahona por haberme proporcionado los datos y brindado las explicaciones de la obtención de las mediciones de la emanación del gas CO₂ para realizar este estudio, ya que sin su apoyo no hubiese sido posible la realización de este trabajo de investigación.

ÍNDICE

CAPÍTULO 1. DESCRIPCIÓN Y DEPURACIÓN DE LOS DATOS EMANACIÓN DE DIÓXIDO DE CARBONO, CO₂.	1
1.1 INTRODUCCIÓN.....	1
1.2 RECOLECCIÓN Y DEPURACIÓN DE LOS DATOS.....	4
CAPÍTULO 2. MODELOS ARIMA.....	9
2.1 INTRODUCCIÓN.....	9
2.2 MODELOS DE MEDIAS MÓVILES (MA).....	10
2.3 MODELOS AUTORREGRESIVOS (AR).....	19
2.4 MODELOS AUTORREGRESIVOS DE MEDIAS MÓVILES (ARMA).....	31
2.5 MODELOS ARIMA (p, d, q).....	38
CAPÍTULO 3. AJUSTE DEL MODELO.....	42
3.1 ELEMENTOS A CONSIDERAR EN LA BÚSQUEDA DEL MODELO.....	42
3.2 DESCRIPCIÓN DE LA METODOLOGÍA DE BOX-JENKINS.....	45
3.3 APLICACIÓN DE LA METODOLOGÍA DE BOX-JENKINS.....	47
3.3.1 IDENTIFICACIÓN DEL MODELO.....	48
3.3.2 ESTIMACIÓN DEL MODELO.....	52
3.3.3 CHEQUEO.....	56
3.3.4 PREDICCIÓN.....	75
3.3.5 CONCLUSIONES.....	82
BIBLIOGRAFÍA.....	83

CAPÍTULO 1

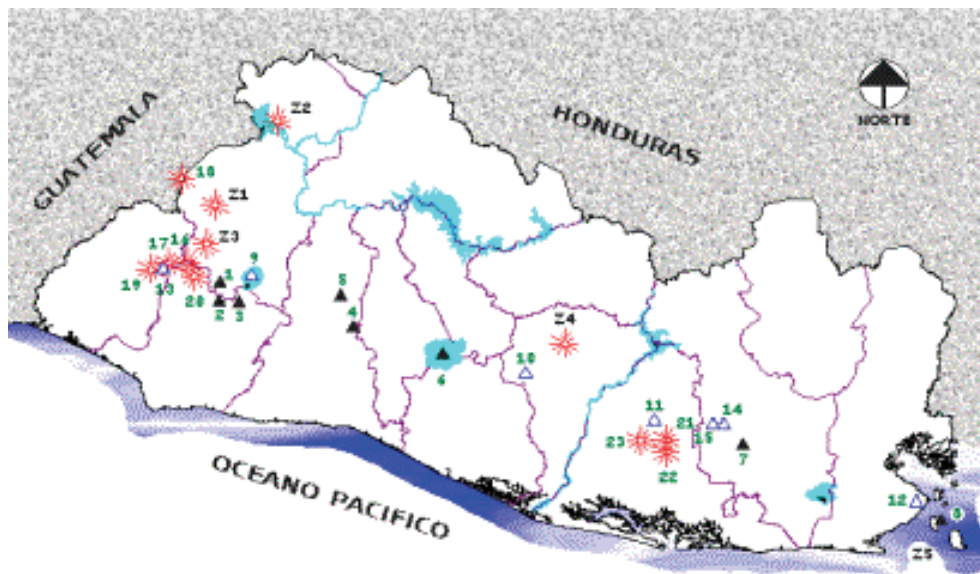
DESCRIPCIÓN Y DEPURACIÓN DE LOS DATOS EMANACIÓN DE CO₂

1.1 INTRODUCCIÓN

El Salvador, se ubica en el cinturón de fuego del Pacífico, es un país con volcanes muy activos. Tiene un 90% de su territorio conformado por materiales volcánicos.

Sobre la base de las investigaciones geológicas y a los reconocimientos de campo desarrollados recientemente en la Cordillera Volcánica, se establecieron criterios específicos para la clasificación de volcanes activos. Se determinó que en El Salvador existen 23 volcanes y cinco de ellos con antecedentes sísmicos que agrupan estructuras volcánicas y lagos cratericos.

Figura 1.1 Ubicación de los volcanes activos de El Salvador



A 8 volcanes se les conoce historia eruptiva: Santa Ana, Izalco (Sonsonate), Boquerón (San Salvador), San Marcelino (Sonsonate), El Playón (San Salvador), Ilopango (San Salvador), Chaparrastique (San Miguel) y Conchaguita (Golfo de Fonseca); 7 tienen antecedentes de actividad sísmica: Caldera de Coatepeque (Santa Ana), San Vicente, Tecapa (Usulután), Conchagua (La Unión), Cuyanausu (Ahuachapán), Laguna Seca El

Pacayal (San Miguel), Limbo (San Miguel) y dos volcanes presentan fumarolas: el volcán de Izalco y el Chaparrastique.

La identificación y estudio de los volcanes activos en El Salvador se reinició en el año 1985, con instituciones internacionales y la participación de la Universidad de El Salvador. Se han elaborado conjuntamente, mapas preliminares en las que se muestran las zonas de peligro volcánico, en especial aquellos con características eruptivas (emanación de gases, cenizas, rocas, etc.), tales como el volcán de San Miguel, Izalco, San Vicente, San Salvador y Santa Ana.

Es importante señalar, que para hablar de reactivación o crisis volcánica es necesario observar o medir los cambios de los parámetros, como la emanación de gases, temperatura del aire, actividad sísmica, cambios del microclima, etc. Situación que se pudo verificar en los meses de septiembre y octubre del año 2005 en el volcán de Santa Ana.

Para observar los cambios en los parámetros volcánicos se desarrolla la vigilancia volcánica la cual consiste en realizar mediciones sistemáticas y periódicas de diferentes parámetros físicos y químicos de los volcanes activos.

Los objetivos de la vigilancia volcánica son detectar cualquier anomalía o cambio que pudiera indicar intranquilidad volcánica, con la finalidad de aportar los conocimientos técnicos necesarios para alertar a la población y a las instituciones en caso de reactivación de la actividad volcánica.

La vigilancia volcánica se lleva a cabo por diferentes sistemas de monitoreo que permiten la recolección de datos, dentro de ellos tenemos:

- Observación visual por parte de personal local y vulcanólogos especializados.
- Observación visual remota con cámaras automáticas (Webcam, vídeo).
- Medición de temperatura en fumarolas, fisuras, pozos y cuerpos de agua.

- Análisis químicos de gases, aguas y cenizas.
- Utilización de sensores remotos (fotografías aéreas, imágenes de satélite, imágenes de radar, etc.).
- Vigilancia sísmica (que se realiza con estaciones telemétricas).

Así, la vigilancia de los volcanes, cuenta actualmente con una red de estaciones telemétricas y sismógrafos portátiles. Se han instalado en diferentes lugares de los volcanes para el control de su actividad. Entre ellos tenemos; el volcán de San Vicente, el volcán de San Salvador, volcán de Santa Ana, en los alrededores de la caldera de Ilopango y el Volcán de San Miguel.

Actualmente, se cuenta con instituciones que participan en el monitoreo de los diferentes volcanes: Servicio Nacional de Estudios Territoriales (SNET), Universidad de El Salvador (UES), el grupo técnico de ITER-Canarias y observadores locales del lugar.

Los tipos de monitoreo o vigilancia que se obtiene de estas instituciones, se clasifican en el Monitoreo geoquímico (gases) y el sismológico. El monitoreo geoquímico es el encargado de recopilar la información de las emanaciones de los gases emitidos por cada uno de los volcanes por períodos de tiempo, y el monitoreo sismológico se refiere a la recopilación de datos que registra los movimientos sísmicos, tales como, características del sismo, lugar, profundidad, magnitud, zonas afectadas, etc. Estas medidas son obtenidas por un sensor que registra el movimiento del sismo que es captada por el equipo de telemetría, que posteriormente es transportada a un sistema central de procesamiento.

Los datos se obtienen a través de las visitas periódicas que realizan los encargados del monitoreo volcánico, dentro de éstas medidas tenemos: la temperatura en fumarolas, emanación de gases, abertura de grietas y los cambios morfológicos en el cono y cráter del volcán.

1.2 RECOLECCIÓN DE LOS DATOS

La información de las mediciones de la emanación de gas Dióxido de Carbono, CO_2 , se recolectan a través de mediciones sistemáticas de diferentes parámetros físicos y químicos que se observan en los volcanes, con el fin de establecer la línea base de su comportamiento e identificar cambios que puedan indicar una posible reactivación volcánica. La información registrada y recolectada son datos de las mediciones químicas obtenidas de los volcanes de San Vicente, de San Salvador, de Santa Ana, en los alrededores de la caldera de Ilopango y el Volcán de San Miguel.

Los compuestos químicos medidos, son gases de origen magmático, entre los cuales tenemos:

- Dióxido de Carbono, CO_2 . Es un gas incoloro, inodoro y con un ligero sabor ácido.

El dióxido de carbono se produce por diversos procesos entre los cuales tenemos:

- a) Combustión u oxidación de materiales que contienen carbono, como el carbón, la madera, el aceite o algunos alimentos;
 - b) Por la fermentación de azúcares.
 - c) Por la descomposición de los carbonatos bajo la acción del calor o los ácidos.
- Sulfuro de Hidrógeno, H_2S es un gas venenoso e incoloro, con olor a huevo podrido. Este se combina con hidrógeno y con elementos metálicos por calentamiento, formando sulfuros.
 - Otros registros obtenidos son: medidas de la velocidad del viento, humedad, temperatura de la tierra, temperatura del viento, etc.

Para iniciar el trabajo de investigación consideramos necesario conocer algunas características del volcán en estudio, para ello mostramos la Tabla 1.1, en la que se

presenta la información que se refiere a la altura, localización y el tipo de roca del que esta formado el Volcán de San Vicente.

Tabla 1.1. CARACTERÍSTICAS DEL VOLCÁN DE SAN VICENTE

Nombre del volcán	Altura sobre el nivel del mar (m)	Tipo de volcán	Localización	Tipo de roca predominante
San Vicente	2,173	Estrato volcán	Departamento de San Vicente	Andesita Piroplástico

Por el momento, el volcán de San Vicente no presenta peligro para la población con relación a una modificación de los niveles de actividad volcánica. Sin embargo, el SNET, la Universidad de El Salvador y otras instituciones, continúan monitoreando la actividad gaseosa y sísmica.

Por otra parte, las instituciones de investigación vulcanológica y la Universidad de El Salvador con la cooperación de unidades extranjeras, trabajan para la puesta a punto de la instrumentación destinada al monitoreo geoquímico en los volcanes de San Miguel, Izalco, Santa Ana y San Vicente, la cual podrá aportar información valiosa para la comprensión del comportamiento de los diferentes volcanes bajo vigilancia.

La información que se obtiene es de diferentes parámetros, pero, para éste análisis se hace uso de datos recolectados de mediciones de la emanación del gas Dióxido de Carbono, CO₂, obtenidos de los alrededores del volcán de San Vicente, a través del proceso de monitoreo llevado a cabo por un grupo de vulcanólogos de la Universidad de El Salvador y otras instituciones.

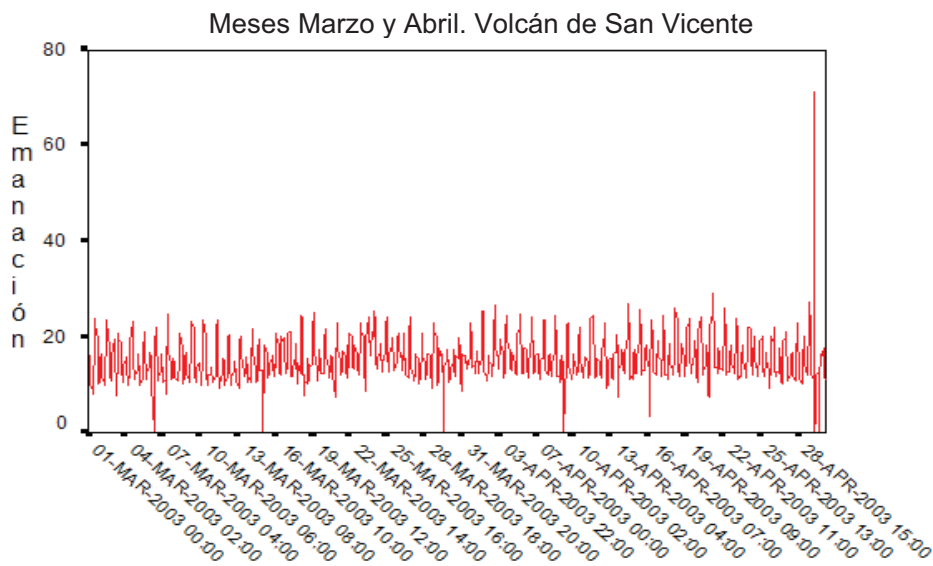
Para iniciar el estudio de los datos de la emanación de gases de Dióxido de Carbono, CO₂, se contó al inicio con los datos del año 2003, recogidos en los meses de enero, febrero, marzo, abril, mayo, junio y julio. Se realizó una revisión y análisis de estos datos y nos llevó a determinar que la serie *no* presentaba las condiciones que se requieren para ser modeladas a través de la teoría ARIMA (teoría que se presentará más adelante) por lo que a continuación se detallan algunas conclusiones de este análisis:

- a) Ausencia de datos en días, semanas y horas de los meses de Enero, Febrero, Mayo, Junio y Julio.
- b) Datos atípicos: datos con signos negativos, valores muy cercanos a cero, otros datos que al observar el comportamiento de la serie presentaba alteraciones que no coinciden con la secuencia de la serie, estos elementos dificultan la interpretación y modelación.
- c) Repetición de datos a determinadas horas, días o semanas.

Cada una de estas fallas se deben a muchos factores en las que mencionamos: fallas en el aparato telemétrico, el agotamiento de la batería que da soporte el sistema, falta de mantenimiento adecuado del sistema, protección al sistema, vigilancia y otros que se deben al error humano y al estado del tiempo.

Se continuó con la revisión, lectura y el análisis de la base de datos del año 2003 y se encontró que los meses de Marzo y Abril, meses de la estación de verano, no presenta irregularidades en los datos, es decir, la secuencia no presenta cortes o discontinuidades. Por lo que se **determinó** que es la información para la realización del presente trabajo de investigación. Se presenta a continuación en la Figura 1.2 el gráfico de los datos con los detalles expuestos.

Figura 1.2 Gráfico de las mediciones de la Emanación de gas CO₂ (ppm/sec.)

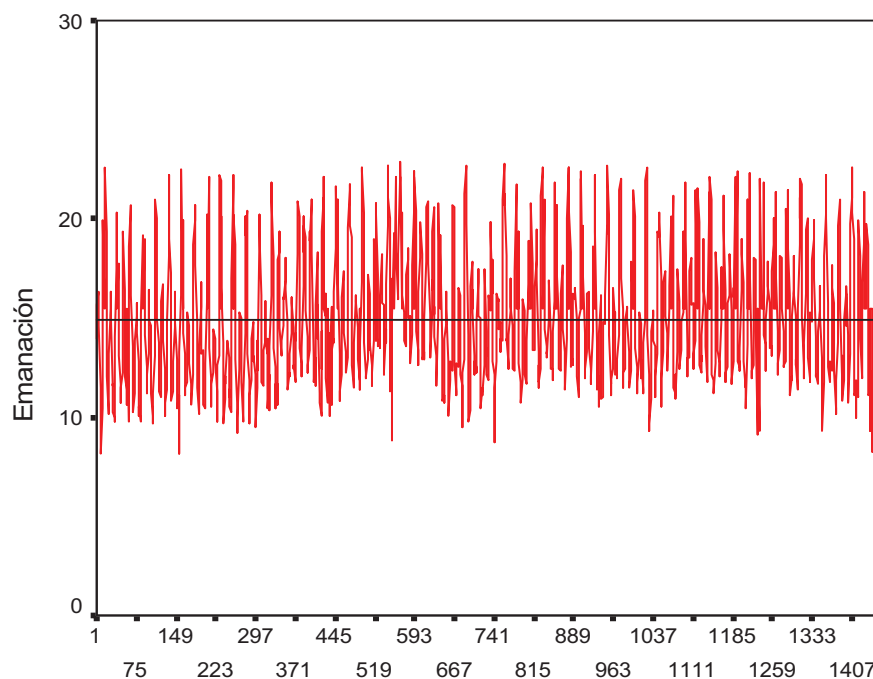


Al observar el gráfico 1.2, en el comportamiento de la serie se detecta la presencia de valores extremos, esto nos indica que es necesario la aplicación de una técnica para tratar estos “*datos atípicos*”.

Para darle tratamiento a los datos atípicos se utilizó el siguiente procedimiento: Se le consulto a los responsables del monitoreo volcánico a fin de constatar la veracidad de dichos datos, sin embargo se nos dijo “que una vez recolectados se encargarían de revisar la información y depurar los datos.....”, por lo tanto, se decidió que los valores extremos serán sustituidos por la media muestral.

A continuación se muestra la gráfica de la serie con los datos “atípicos” sustituidos por la media muestral. También se omitió las fechas para obtener una mejor apreciación del comportamiento de los datos, una mejor ubicación entorno a la media, expresándolo en número de horas, no afectando así el comportamiento y la dispersión de la información.

Figura 1.3. Serie de datos de los Meses de Marzo- Abril
(Valores atípicos que fueron reemplazados por la media muestral)



La figura 1.3 nos muestra tal como se planteo, la acumulación de la información más aceptable, alrededor de su media, las mediciones se estiman entre 10 ppm/sec y 22 ppm/sec, además, se observa que la variabilidad se mantiene constante en los períodos de tiempo, el comportamiento de la serie se despliega con claridad y facilita la observación de sus características, algunos punto altos se compensan con algunos que están debajo de la media, por lo que iniciaremos la investigación con los datos presentados en la figura 1.3 y el marco teórico del que se dispone para buscar el modelo que más se ajuste a la serie.

La metodología a utilizar en esta investigación consiste en la aplicación de diferentes etapas para encontrar y estimar los parámetros del modelo ARIMA capaz de generar la serie, las etapas son las siguientes: identificación, estimación, chequeo y predicción.

Las etapas mencionadas son las sugeridas en la metodología Box-Jenkins. En el desarrollo de éste trabajo se detallará el análisis para cada una, hasta llegar al modelo ARIMA que supera cada una de las fases de la metodología Box-Jenkins.

CAPÍTULO 2

MODELOS ARIMA

2.1 INTRODUCCIÓN

Box y Jenkins (1978) desarrollaron un recurso metodológico destinado a identificar, estimar y diagnosticar modelos dinámicos de series temporales en los que la variable tiempo juega un papel fundamental. Una parte importante de esta metodología está pensada para liberar al investigador de la tarea tediosa de efectuarlo utilizando los "procedimientos tradicionales", intentando por cada modelo establecer cálculos que hasta cierto punto se vuelven aburridos, siempre en la búsqueda del modelo y es posible que durante el procedimiento se pierdan los propios datos (serie temporal) de la variable a estudiar y se nos omitan las características de la estructura probabilística subyacente. En parte, los procedimientos que vamos a analizar se contraponen a la "forma tradicional" de identificar y especificar un modelo apoyándonos en los recursos informáticos y de las teorías de autores que hacen referencias a estos fenómenos analizados.

El estudio de series temporales tiene por objeto analizar la evolución de una variable a través del tiempo. La diferencia esencial entre las series temporales y los análisis no temporales (Estadística descriptiva, Diseño de experimentos o Regresión) es que en los análisis previos no importaba el orden en que estaban tomadas las observaciones y éste se podía variar sin problemas. En series temporales el orden es muy importante y variarlo supone cambiar la información contenida en la serie.

Determinar las características más notables asociadas a un conjunto de series temporales es generalmente considerado como un paso crucial en el área del análisis estadístico. Ésta es una de las razones por las que es deseable formular un modelo para una serie temporal.

Sin embargo, para que tales modelos sean útiles deben ser consistentes con las propiedades estocásticas de los datos y ofrecer información relevante. En este sentido, los modelos estructurales de series temporales constituyen un enfoque adecuado que,

basado explícitamente en las propiedades estocásticas de los datos, permite presentar sus rasgos más sobresalientes.

El objetivo de este trabajo es establecer la relación entre los modelos teóricos y modelos prácticos extraídos de nuestra realidad, como también la utilidad en diferentes áreas de las ciencias, además, proponer un modelo base para el estudio del comportamiento de los volcanes a través de las emanaciones de gas CO₂ en diferentes volcanes de nuestro país. También proporcionar un aporte importante para nuestra comunidad y de beneficio para su aplicación en otras áreas de las ciencias.

A continuación definimos en forma breve los modelos ARIMA extraídos de la literatura clásica de series temporales. Se presenta la ecuación en diferencias del modelo y las características más importantes.

2.2 MODELOS DE MEDIAS MÓVILES (MA)

Se abordan en este apartado los modelos de medias móviles, MA. Estos modelos se obtienen a partir de observaciones y que juegan un papel importante en el problema de la inferencia.

Nuestro problema es identificar un modelo apropiado para los datos observados por lo tanto es necesario conocer cuales son las características de los modelos MA que más se ajustan a la serie. Este es uno de los procesos más utilizados ya que es un procedimiento que detecta y elimina la tendencia de una serie basado en la aplicación de filtros, es decir, filtros para suavizar las irregularidades y fluctuaciones (variaciones cíclicas y estacionarias) de una serie temporal con el fin de obtener la línea de tendencia.

La serie así obtenida es una versión suavizada de la original. Este filtro no es más que una función matemática que aplicada a los valores de la serie produce una nueva serie con características determinadas. Uno de estos filtros es el llamado procesos de medias móviles. Pero para encontrar tal función matemática de un fenómeno no se pueden modelar exclusivamente con modelos autorregresivos, es necesario introducir una familia de modelos para poder representarlo.

A continuación se presentan los modelos de medias móviles de orden 1 y 2 con sus respectivas ecuaciones en diferencias y las principales características de cada uno de ellos, finalmente se generaliza al modelo de medias móviles de orden q.

2.2.1. Modelo de Medias Móviles de Orden uno, MA (1)

Su ecuación en diferencias viene dada por:

$$y_t = \mu + u_t - \theta_1 u_{t-1}$$

o bien:

$$y_t = \mu + \theta(L)u_t$$

de donde:

$$\theta(L) = 1 - \theta_1 L, \quad Lu_t = u_{t-1}$$

El modelo de medias móviles, MA, siempre es estacionario.

Para conseguir su invertibilidad es necesario que cumpla que la raíz de la ecuación

$$\theta(L) = 1 - \theta_1 L = 0$$

caiga fuera del círculo unitario (esto es, sea en valor absoluto mayor que la unidad) lo cual implica que $|\theta_1| < 1$.

Una vez cumplida esta condición, puede escribirse el Modelo MA(1) como un modelo AR de orden infinito, de la forma siguiente

$$y_t = \pi_1 y_{t-1} + \pi_2 y_{t-2} + \pi_3 y_{t-3} + \dots + \delta + u_t$$

Donde

$$\delta = \frac{\mu}{1 - \theta_1}$$

de ahí que se dice que es invertible.

La esperanza matemática y la varianza del modelo MA (1) vienen dadas por

$$E(y_t) = \mu$$
$$\gamma_0 = \sigma_u^2 (1 + \theta_1^2)$$

La función de autocovarianza por

$$\gamma_j = \begin{cases} -\theta_1 \sigma_u^2 & \text{para } j = 1 \\ 0 & \text{para } j > 1 \end{cases}$$

La función de autocorrelación

$$\rho_j = \begin{cases} -\frac{\theta_1}{1 + \theta_1^2} & \text{para } j = 1 \\ 0 & \text{para } j > 1 \end{cases}$$

La función de autocorrelación parcial

$$\phi_{jj} = \frac{-\theta_1^j (1 - \theta_1^2)}{1 - \theta_1^{2(j+1)}} \quad \text{para } j \geq 1$$

Las Características más relevantes del modelo MA (1) son las siguientes:

1- El modelo MA (1) siempre es estacionario.

2- Para que sea invertible debe cumplirse que $|\theta_1| < 1$.

3- La función de autocorrelación se anula para retardos superiores a 1 (el orden del modelo). El modelo MA(1) tiene memoria de un solo período; esto es, cualquier valor de y_t está correlacionado con y_{t-1} e y_{t+1} pero no lo está con otro valor de la serie, por lo que podemos decir que el modelo “olvida” lo que ocurrió en períodos previos por lo que los sucesos ocurridos en dichos períodos no tiene efecto alguno sobre lo que ocurre en el momento presente. En general, “la limitación de la memoria” de los

modelos de medias móviles es importante porque sugiere, que un modelo de este tipo proporciona información para la predicción de un número limitado de períodos futuros; en concreto, tantos como el orden del modelo.

4- El correlograma, representación gráfica de la función de autocorrelación, tendrá un solo pico, distinto de cero, el correspondiente al primer retardo, siendo del mismo signo la posición que ocupa en el gráfico del correspondiente retardo.

5- La función de autocorrelación parcial no se anula. Su representación gráfica presentará un comportamiento amortiguado hacia cero con todos los valores negativos, caso de que $\theta_1 > 0$, o alternado el signo, comenzando con positivo, si $\theta_1 < 0$.

2.2.2 Modelo de medias Móviles de Orden Dos, MA (2):

Su ecuación en diferencias viene dada por:

$$y_t = \mu + u_t - \theta_1 u_{t-1} - \theta_2 u_{t-2}$$

O bien

$$y_t = \mu + \theta(L)u_t$$

Donde

$$\theta(L) = 1 - \theta_1 L - \theta_2 L^2$$

Como cualquier modelo de medias móviles, el modelo MA (2) será estacionario, mientras que para ser invertible deberá cumplirse que las raíces de la ecuación

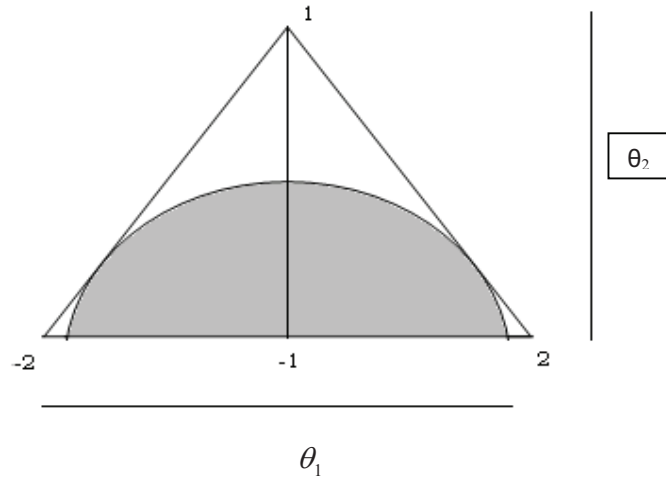
$$\theta(L) = 1 - \theta_1 L - \theta_2 L^2 = 0$$

caigan fuera del círculo unitario.

Al resolver dicha ecuación se tiene

$$L_1 = \frac{-\theta_1 + \sqrt{\theta_1^2 + 4\theta_2}}{2\theta_2} \quad \text{y} \quad L_2 = \frac{-\theta_1 - \sqrt{\theta_1^2 + 4\theta_2}}{2\theta_2}$$

Figura 2.1 Zona de invertibilidad del modelo MA (2)



Estas raíces podrán ser reales o complejas dependiendo de si $\theta_1^2 + 4\theta_2$ es negativo o positivo. Así pues, $\theta_1^2 + 4\theta_2 \geq 0$ (caso de raíces reales), diremos que se cumple la condición de invertibilidad si $|L_i| > 1$ para $i = 1, 2$. Por el contrario, en el caso de raíces complejas, esto es, cuando $\theta_1^2 + 4\theta_2 < 0$ (y, por lo tanto, $\theta_2 < 0$), las raíces serán iguales a una expresión de la forma $a \pm bi$, donde i es el número imaginario.

En este caso, la condición de invertibilidad se cumplirá cuando el módulo, definido como $\sqrt{a^2 + b^2}$ sea mayor que la unidad. En cualquiera de los dos casos considerados, raíces reales o raíces complejas, en el que las raíces caigan fuera del círculo unitario implica las siguientes restricciones de los valores de los parámetros θ_1 y θ_2 :

$$\begin{aligned} \theta_1 + \theta_2 &< 1 \\ \theta_1 - \theta_2 &< 1 \\ |\theta_2| &< 1 \end{aligned}$$

A partir de estas restricciones, sin más que reemplazar en las tres desigualdades el signo $<$ por el $=$, se obtienen las tres líneas rectas que definen la zona de invertibilidad, la cual esta constituida por todos los puntos del interior del triángulo representado en la figura 2.1. Obsérvese en el caso de las raíces complejas debe cumplirse que $\theta_2 < 0$, de donde resulta la zona sombreada de la figura 2.1 mencionada.

De manera análoga se obtendría considerando las raíces complejas y, por tanto, tomando

$$\theta_1^2 + 4\theta_2 < 0.$$

La esperanza matemática y la varianza del modelo MA (2) viene dada por

$$E(y_t) = \mu$$

$$\gamma_0 = \sigma_u (1 + \theta_1^2 + \theta_2^2)$$

La función de autocovarianzas

$$\gamma_j = \begin{cases} (-\theta_1 + \theta_1\theta_2)\sigma_u^2 & \text{para } j = 1 \\ -\theta_2\sigma_u^2 & \text{para } j = 2 \\ 0 & \text{para } j > 2 \end{cases}$$

La función de autocorrelación

$$\rho_j = \begin{cases} \frac{-\theta_1 + \theta_1\theta_2}{1 + \theta_1^2 + \theta_2^2} & \text{para } j = 1 \\ \frac{-\theta_2}{1 + \theta_1^2 + \theta_2^2} & \text{para } j = 2 \\ 0 & \text{para } j > 2 \end{cases}$$

Función de autocorrelación parcial

$$\begin{aligned}\phi_{11} &= \rho_1 \\ \phi_{22} &= \frac{\rho_2 - \rho_1^2}{1 - \rho_1^2} \\ \phi_{33} &= \frac{\rho_3 - \rho_1\rho_2(2 - \rho_2)}{1 - \rho_2^2 - 2\rho_1^2(1 - \rho_2)} \\ &\dots\dots\end{aligned}$$

Las características más importantes del modelo MA (2) son las siguientes:

1- El modelo MA (2) siempre es estacionario

2- Es invertible si se cumple

$$\begin{aligned}\theta_1 + \theta_2 &< 1 \\ \theta_1 - \theta_2 &< 1 \\ |\theta_2| &< 1\end{aligned}$$

3- La función de autocorrelación se anula para retardos superiores a dos (el orden del modelo). Diremos, por ello, que el modelo MA (2) tiene memoria de solo dos períodos; esto es, cualquier valor y_t está correlacionado con $y_{t-1}, y_{t-2}, y_{t+1}, y_{t+2}$ pero no lo está con ningún otro valor de la serie.

4- El correlograma, representación gráfica de la función de autocorrelación, tendrá tan solo dos picos distintos de cero, los correspondientes a los dos primeros retardos. Respecto al signo de estas correlaciones, se puede concluir que a partir de ρ_j autocorrelaciones el primer retardo tendrá signo inverso al del coeficiente de θ_1 , mientras que el del segundo retardo tendrá signo inverso al de θ_2 .

5- La función de autocorrelación parcial no se anula. Su representación gráfica presentará un comportamiento amortiguado hacia cero, caso de que las raíces $\theta(L) = 1 - \theta_1 L - \theta_2 L^2 = 0$ sean reales o un comportamiento sinusoidal, también tendiente a cero, en el caso de que las raíces sean complejas.

2.2.3 Modelos de Medias Móviles de orden q, MA (q)

Su ecuación en diferencias viene dada por:

$$y_t = \mu + u_t - \theta_1 u_{t-1} - \theta_2 u_{t-2} - \dots - \theta_q u_{t-q}$$

Donde

y_t : Variable en estudio, la cual ha sido observada durante cierto periodo de tiempo

μ : Representa la esperanza matemática del modelo

θ_i : Parámetros o coeficientes desconocidos del modelo de Medias Móviles de orden q

u_t : Variables aleatorias distribuidas idéntica e independientemente con media cero y varianza σ^2 .

Este modelo siempre es estacionario y para que sea invertible debe cumplir que las raíces

$$\theta(L) = 1 - \theta_1 L - \theta_2 L^2 - \dots - \theta_q L^q = 0$$

caigan fuera del círculo unitario. Esto es, denominando mediante L_1, L_2, \dots, L_q las q raíces, estas deben ser mayores que la unidad en valor absoluto, si dichas raíces son reales, o bien el módulo debe ser superior a la unidad si las raíces son complejas.

En los modelos particulares MA (1) y MA (2) se han obtenido las restricciones paramétricas implicadas por esta condición de invertibilidad. Una condición necesaria, aunque no suficiente, para que un modelo de medias móviles de cualquier orden (por ejemplo, q) sea invertible, es la siguiente:

$$\theta_1 + \theta_2 + \theta_3 + \dots + \theta_q < 1$$

La esperanza matemática del modelo medias móvil, MA (q) viene dada por:

$$E(y_t) = \mu$$

La varianza del modelo MA (q)

$$\gamma_0 = \sigma_u^2 (1 + \theta_1^2 + \theta_2^2 + \dots + \theta_q^2)$$

La función de autocovarianzas

$$\gamma_j = \begin{cases} (-\theta_j + \theta_1\theta_{j+1} + \dots + \theta_{q-j}\theta_q)\sigma_u^2 & \text{para } j=1, 2, 3, \dots, q \\ 0 & \text{para } j > q \end{cases}$$

Y la función de autocorrelación

$$\rho_j = \begin{cases} \frac{-\theta_j + \theta_1\theta_{j+1} + \dots + \theta_{q-j}\theta_q}{1 + \theta_1^2 + \theta_2^2 + \dots + \theta_q^2} & \text{para } j=1, 2, 3, \dots, q \\ 0 & \text{para } j > q \end{cases}$$

Finalmente, La función de autocorrelación parcial

$$\phi_{11} = \rho_1$$

$$\phi_{22} = \frac{\rho_2 - \rho_1^2}{1 - \rho_1^2}$$

$$\phi_{33} = \frac{\rho_1^3 - \rho_1\rho_2(2 - \rho_2) + \rho_3(1 - \rho_1^2)}{1 - \rho_2^2 - 2\rho_1^2(1 - \rho_2)}$$

.....

$$\phi_{jj} = \begin{array}{c} \left| \begin{array}{cccccc} 1 & \rho_1 & \rho_2 & \dots & \rho_{j-2} & \rho_1 \\ \rho_1 & 1 & \rho_1 & \dots & \rho_{j-3} & \rho_2 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots \\ \rho_{j-1} & \rho_{j-2} & \rho_{j-3} & \dots & \rho_1 & 1 \end{array} \right| \\ \left| \begin{array}{cccccc} 1 & \rho_1 & \rho_2 & \dots & \rho_{j-2} & \rho_{j-1} \\ \rho_1 & 1 & \rho_1 & \dots & \rho_{j-3} & \rho_{j-2} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots \\ \rho_{j-1} & \rho_{j-2} & \rho_{j-3} & \dots & \rho_1 & 1 \end{array} \right| \end{array}$$

Características más relevantes del modelo MA (q) son las siguientes:

- a) Un modelo de medias móviles siempre es estacionario
- b) Un modelo de medias móviles siempre es invertible, esto es, puede escribirse como un modelo AR(∞) siempre y cuando las raíces de la ecuación $\theta(L)=0$ se ubiquen fuera del círculo unidad, y la ecuación $\theta(L)=0$ se define así $\theta(L)=1 - \theta_1L - \theta_2L^2 - \dots - \theta_qL^q = 0$
- c) Las autocovarianzas se anulan cuando el desfase temporal considerado es superior al orden del modelo. Consecuentemente, diremos que la memoria de un proceso de medias móviles es igual al orden del proceso, en general el orden del modelo se denota por q .
- d) Las autocorrelaciones parciales no se anulan, presentando un comportamiento amortiguado hacia cero (caso de raíces reales) o sinusoidal también tendiente a cero (caso de que las raíces sean complejas).

2.3 MODELOS AUTORREGRESIVOS (AR)

Definimos un modelo como autorregresivo, si la variable endógena de un período t es explicada por las observaciones de ella misma correspondientes a períodos anteriores

añadiéndose, como en los modelos estructurales, un término de error. En el caso de procesos estacionarios con distribución normal, la teoría estadística de los procesos estocásticos dice que, bajo determinadas condiciones previas, toda y_t puede expresarse como una combinación lineal de sus valores pasados (parte sistemática) más un término de error (innovación). Los modelos autorregresivos se abrevian con la palabra **AR** tras la que se indica el orden del modelo: AR(1), AR(2),....etc. El orden del modelo expresa el número de observaciones retrasadas de la serie temporal analizada que intervienen en la ecuación.

A continuación se presentan los modelos autorregresivos de orden 1 y 2 con sus respectivas ecuaciones en diferencias y las principales características de cada uno de ellos, finalmente se generaliza con el modelo autorregresivo de orden p.

2.3.1 Modelos Autorregresivos de orden uno. AR (1)

Un modelo AR(1) viene definido por

$$y_t = \phi_1 y_{t-1} + \delta + u_t$$

Abreviadamente se puede escribir

$$\phi(L)y_t = \delta + u_t$$

siendo:

$$\phi(L) = 1 - \phi_1 L$$

La condición de estacionariedad del modelo AR (1) implica que

$$|\phi_1| < 1$$

Supuesto el cumplimiento de la condición anterior, podrá escribirse la esperanza matemática, igual a una constante μ , definida de la forma siguiente:

$$E(y_t) = \mu = \frac{\delta}{1 - \phi_1}$$

Una forma alternativa de escribir un Modelo AR (1) estacionario es en forma de desviaciones con respecto a la media:

$$\tilde{y}_{t-j} = \phi_1 \tilde{y}_{t-1} + u_t$$

donde:

$$\tilde{y}_{t-j} = y_{t-j} - \mu = y_{t-j} - \frac{\delta}{1 - \phi_1} \quad \text{para } j = 0 \text{ y } 1$$

o bien:

$$\phi(L)\tilde{y}_t = u_t$$

estando definido

$$\phi(L) \text{ como } \phi(L) = 1 - \phi_1 L.$$

La varianza viene dada por

$$\gamma_0 = \frac{\sigma_u^2}{1 - \phi_1^2}$$

La función de autocovarianza

$$\gamma_j = \phi_1^j \frac{\sigma_u^2}{1 - \phi_1^2} \quad \text{para } j \geq 1$$

La función de autocorrelación

$$\rho_j = \phi_1^j \quad \text{para } j \geq 1$$

La función de autocorrelación parcial

$$\phi_{jj} = \begin{cases} \rho_1 = \phi_1 & \text{para } j = 1 \\ 0 & \text{para } j > 1 \end{cases}$$

Las características más relevantes del modelo AR(1) son las siguientes:

1- El modelo AR (1) es siempre invertible

2- El modelo AR (1) es estacionario siempre que se cumpla que $|\phi_1| < 1$.

3- La función de autocorrelación no se anula, si bien las correlaciones son menos intensas a medida que aumentan los desfases temporales. Diremos, por ello, que el modelo AR(1) tiene memoria infinita; esto es, cualquier valor de y_t esta correlacionado con cualquier valor pasado o futuro de la serie, si bien la correlación es tanto más pequeña cuanto más distantes en el tiempo sean las observaciones.

4- El correlograma, la representación gráfica de la función de autocorrelación, tendrá un comportamiento amortiguado hacia cero con todos los valores positivos, caso de que $\phi_1 > 0$, o bien, alternando el signo, comenzando con negativo, si $\phi_1 < 0$.

5- La función de autocorrelación parcial se anula para retardos superiores a uno (el orden del modelo), coincidiendo, como es norma, este coeficiente de autocorrelación parcial (de primer orden) no nulo con el coeficiente de autocorrelación de primer orden. Su representación gráfica tendrá un solo pico distinto de cero, el correspondiente al primer retardo, coincidiendo el signo del mismo con el del parámetro ϕ_1 .

2.3.2 Modelos Autorregresivos de orden dos. AR (2)

Un modelo AR(2) viene definido por:

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \delta + u_t$$

Este modelo también puede escribirse como:

$$\phi(L)y_t = \delta + u_t$$

El modelo AR (2) será estacionario si las raíces de la ecuación

$$\phi(L) = 1 - \phi_1 L - \phi_2 L^2 = 0$$

caen fuera del círculo unitario. Recuérdese que esto quiere decir que si las raíces son reales deben ser en valor absoluto mayores que uno; mientras que si son complejas, el módulo debe ser mayor que la unidad.

Por analogía, con la condición de invertibilidad del modelo MA (2), resulta inmediato deducir que las restricciones paramétricas implicadas por la estacionariedad de un modelo AR (2) son las siguientes

$$\begin{aligned}\phi_1 + \phi_2 &< 1 \\ \phi_2 - \phi_1 &< 1 \\ |\phi_2| &< 1\end{aligned}$$

pudiéndose representar la zona de estacionariedad con los puntos interiores del triángulo representados en la figura 2.2, en la que la zona sombreada corresponde al caso en que las raíces son complejas.

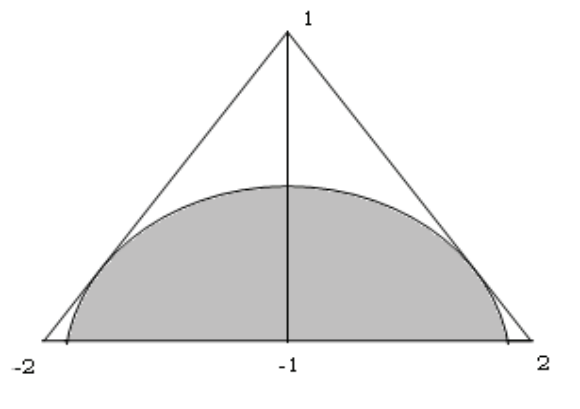


Figura 2.2 Zona de estacionariedad del modelo AR(2)

Supuesto que el modelo AR (2) es estacionario podemos escribir la esperanza matemática como:

$$E(y_t) = \mu = \frac{\delta}{1 - \phi_1 - \phi_2}$$

Bajo la condición de estacionariedad, el modelo AR (2) puede escribirse en forma de desviaciones respecto a la media:

$$\tilde{y}_t = \phi_1 \tilde{y}_{t-1} + \phi_2 \tilde{y}_{t-2} + u_t$$

donde:

$$\tilde{y}_{t-j} = y_{t-j} - \mu = y_{t-j} - \frac{\delta}{1 - \phi_1 - \phi_2} \quad \text{para } j = 0, 1 \text{ y } 2$$

o bien

$$\phi(L)\tilde{y}_t = u_t$$

La varianza del modelo AR (2) es igual a

$$\gamma_0 = \phi_1 \gamma_1 + \phi_2 \gamma_2 + \sigma_u^2$$

La función de autocovarianza

$$\gamma_j = \phi_1 \gamma_{j-1} + \phi_2 \gamma_{j-2} \quad \text{para } j \geq 1$$

A partir de la varianza del modelo AR(2) y de los valores de la función de autocovarianzas para $j = 1$ y 2 se obtiene el sistema de ecuaciones

$$\begin{aligned} \gamma_0 &= \phi_1 \gamma_1 + \phi_2 \gamma_2 + \sigma_u^2 \\ \gamma_1 &= \phi_1 \gamma_0 + \phi_2 \gamma_1 \\ \gamma_2 &= \phi_1 \gamma_1 + \phi_2 \gamma_0 \end{aligned}$$

Desde el cual pueden calcularse γ_0, γ_1 y γ_2 dados los valores de los parámetros ϕ_1, ϕ_2 y σ_u^2 .

Las autocovarianzas de orden superior al del modelo se obtienen según la función de autocovarianza, de forma recursiva; es decir

$$\begin{aligned}\gamma_3 &= \phi_1\gamma_2 + \phi_2\gamma_1 \\ \gamma_4 &= \phi_1\gamma_3 + \phi_2\gamma_2 \\ \gamma_5 &= \phi_1\gamma_4 + \phi_2\gamma_3 \\ &\dots\dots\dots\end{aligned}$$

La función de autocorrelación. Se obtiene dividiendo la función de autocovarianzas por γ_0 , y se obtiene la expresión general de la función de autocorrelación del modelo AR(2)

$$\rho_j = \phi_1\rho_{j-1} + \phi_2\rho_{j-2} \quad \text{para } j \geq 1$$

Las ecuaciones de Yule Walter se deducen directamente de la función de autocorrelación sustituyendo $j = 1$ y 2 y teniendo en cuenta la propiedad de simetría $\rho_j = \rho_{-j}$, así como del hecho que $\rho_0 = 1$:

$$\begin{aligned}\rho_1 &= \phi_1 + \phi_2\rho_1 \\ \rho_2 &= \phi_1\rho_1 + \phi_2\end{aligned}$$

obteniéndose las correspondientes autocorrelaciones de primer y segundo orden, en función de ϕ_1 y ϕ_2 :

$$\rho_1 = \frac{\phi_1}{1 - \phi_2} \quad \text{y} \quad \rho_2 = \frac{\phi_1^2 + \phi_2(1 - \phi_2)}{1 - \phi_2}$$

Función de autocorrelación parcial

$$\phi_{jj} = \begin{cases} \rho_1 = \frac{\phi_1}{1 - \phi_2} & \text{para } j = 1 \\ \frac{\rho_2 - \rho_1^2}{1 - \rho_1^2} & \text{para } j = 2 \\ 0 & \text{para } j > 2 \end{cases}$$

Las características más relevantes del modelo son las siguientes:

- 1- El modelo AR (2) es siempre invertible
- 2- El modelo AR (2) es estacionario si cumple

$$\begin{aligned} \phi_1 + \phi_2 &< 1 \\ \phi_2 - \phi_1 &< 1 \\ |\phi_2| &< 1 \end{aligned}$$

- 3- La función de autocorrelación no se anula. Diremos que su memoria es infinita.
- 4- La forma del correlograma, dependerá de cómo son las raíces de la ecuación

$$\phi(L) = 1 - \phi_1 L - \phi_2 L^2 = 0$$

Si ésta son reales, tendrá un comportamiento amortiguado a cero, siendo todos los valores positivos, o alternando el signo, aunque cumpliéndose siempre que el signo de la primera autocorrelación coincide con el signo de ϕ_1 . Si las raíces son complejas para lo cual necesariamente debe cumplirse que $\phi_2 < 0$, el correlograma tendrá un comportamiento sinusoidal pero también con tendencia a cero.

5- La función de autocorrelación parcial se anula para retardos superiores a dos (orden del modelo). Por consiguiente, su representación gráfica presentará dos autocorrelaciones distintos de cero, correspondientes a los dos primeros retardos. Respecto al signo de las dos autocorrelaciones, primero y segundo, coincidirán, respectivamente, con los signos de ϕ_1 y ϕ_2 .

2.3.3 Modelos Autorregresivos de orden p. AR (p)

Generalizando los modelos anteriores, tenemos que su ecuación en diferencias viene dada por:

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \dots + \phi_p y_{t-p} + \delta + u_t$$

Este modelo puede escribirse abreviadamente

$$\phi(L)y_t = \delta + u_t$$

Donde

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

Diremos que este modelo siempre será invertible y que para que sea estacionario deberá cumplirse

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

caigan fuera del círculo unitario. Es decir, denominando mediante L_1, L_2, \dots, L_p las p raíces, éstas deben ser mayores que la unidad en valor absoluto, si dichas raíces son reales, o bien el módulo debe ser superior a la unidad si las raíces son complejas.

En los modelos particulares AR (1) y AR (2) se han obtenido las restricciones paramétricas implicadas por esta condición de estacionariedad las cuales son análogas a las condiciones de invertibilidad de los correspondientes modelos de medias

móviles. Una condición necesaria, aunque no suficiente, para que un modelo autorregresivo de cualquier orden sea estacionario es la siguiente

$$\phi_1 + \phi_2 + \dots + \phi_p < 1$$

Supuesto el cumplimiento de estacionariedad, la esperanza matemática viene dada por

$$E(y_t) = \mu = \frac{\delta}{1 - \phi_1 - \phi_2 - \dots - \phi_p}$$

Su varianza

$$\gamma_0 = \phi_1 \gamma_1 + \phi_2 \gamma_2 + \dots + \phi_p \gamma_p + \sigma_u^2$$

Función de autocovarianzas

$$\gamma_j = \phi_1 \gamma_{j-1} + \phi_2 \gamma_{j-2} + \dots + \phi_p \gamma_{j-p} \quad \text{para } j > 0$$

Considerando γ_0 y γ_j para $j = 1, 2, 3, \dots, p$ se obtiene el sistema de $p + 1$ ecuaciones:

$$\begin{aligned} \gamma_0 &= \phi_1 \gamma_1 + \phi_2 \gamma_2 + \dots + \phi_p \gamma_p + \sigma_u^2 \\ \gamma_1 &= \phi_1 \gamma_0 + \phi_2 \gamma_1 + \dots + \phi_p \gamma_{p-1} \\ \gamma_2 &= \phi_1 \gamma_1 + \phi_2 \gamma_0 + \dots + \phi_p \gamma_{p-2} \\ &\dots\dots\dots \\ \gamma_p &= \phi_1 \gamma_{p-1} + \phi_2 \gamma_{p-2} + \dots + \phi_p \gamma_0 \end{aligned}$$

Si se conocen los valores de los parámetros $\phi_1, \phi_2, \dots, \phi_p$ y σ_u^2 , pueden calcularse las varianzas y autocovarianzas de orden no superior al del proceso. A partir de dicho

orden, es decir, para $j > p$, las autocovarianzas se obtendrán de forma recursiva según las ecuaciones siguientes:

$$\begin{aligned}\gamma_{p+1} &= \phi_1 \gamma_p + \phi_2 \gamma_{p-1} + \dots + \phi_p \gamma_1 \\ \gamma_{p+2} &= \phi_1 \gamma_{p+1} + \phi_2 \gamma_p + \dots + \phi_p \gamma_2 \\ \gamma_{p+3} &= \phi_1 \gamma_{p+2} + \phi_2 \gamma_{p+1} + \dots + \phi_p \gamma_3\end{aligned}$$

Función de autocorrelación. Dividiendo por la función de autocovarianzas, γ_0 , se tiene

$$\rho_j = \phi_1 \rho_{j-1} + \phi_2 \rho_{j-2} + \dots + \phi_p \rho_{j-p} \quad \text{para } j > 0$$

Las ecuaciones de Yule-Walker, que se obtienen directamente de la función de autocorrelación, serán

$$\begin{aligned}\rho_1 &= \phi_1 + \phi_2 \rho_1 + \dots + \phi_p \rho_{p-1} \\ \rho_2 &= \phi_1 \rho_1 + \phi_2 + \dots + \phi_p \rho_{p-2} \\ &\dots\dots\dots \\ \rho_p &= \phi_1 \rho_{p-1} + \phi_2 \rho_{p-2} + \dots + \phi_p\end{aligned}$$

Resolviendo dicho sistema para $\phi_1, \phi_2, \dots, \phi_p$ y expresado en forma de matrices se tiene

$$\begin{pmatrix} \phi_1 \\ \phi_2 \\ \dots \\ \phi_p \end{pmatrix} = \begin{pmatrix} 1 & \rho_1 & \dots & \rho_{p-1} \\ \rho_1 & 1 & \dots & \rho_{p-2} \\ \dots & \dots & \dots & \dots \\ \rho_{p-1} & \rho_{p-2} & \dots & 1 \end{pmatrix}^{-1} \begin{pmatrix} \rho_1 \\ \rho_2 \\ \dots \\ \rho_p \end{pmatrix}$$

Función de Autocorrelaciones Parciales

$$\phi_{jj} = \begin{cases} \rho_1 & \text{para } j = 1 \\ \frac{\rho_2 - \rho_1^2}{1 - \rho_1^2} = \frac{\phi_2 [(1 - \phi_2)^2 - \phi_1^2]}{(1 - \phi_2)^2 - \phi_1^2} & \text{para } j = 2 \\ \dots\dots\dots & \\ \left| \begin{array}{ccccc} 1 & \rho_1 & \dots & \rho_{p-2} & \rho_1 \\ \rho_1 & 1 & \dots & \rho_{p-3} & \rho_2 \\ \dots & \dots & \dots & \dots & \dots \\ \rho_{p-1} & \rho_{p-2} & \dots & \rho_1 & \rho_p \end{array} \right| & \text{para } j = p \\ \left| \begin{array}{ccccc} 1 & \rho_1 & \dots & \rho_{p-2} & \rho_{p-1} \\ \rho_1 & 1 & \dots & \rho_{p-3} & \rho_{p-2} \\ \dots & \dots & \dots & \dots & \dots \\ \rho_{p-1} & \rho_{p-2} & \dots & \rho_1 & 1 \end{array} \right| & \\ 0 & \text{para } j > p \end{cases}$$

Características:

- a) Un modelo autorregresivo es siempre invertible.
- b) Un modelo AR estacionario debe satisfacer la condición de que las raíces de la ecuación $\phi(L) = 1 - \phi_1 L - \phi_2 L^2 - \dots - \phi_p L^p = 0$ caigan fuera del círculo unidad.
- c) En el modelo AR, a diferencia de lo que sucede con el MA, las autocovarianzas, y consecuentemente las autocorrelaciones no lleguen a anularse (la memoria del proceso es por lo tanto, infinita), observándose que, una vez superado el orden del modelo, las diferentes autocovarianzas –y autocorrelaciones – pueden obtenerse de forma recursiva.
- d) Las ecuaciones de Yule-Walker de un AR(p) están constituidas por p ecuaciones lineales que permiten obtener los valores de los parámetros

$(\phi_1, \phi_2, \dots, \phi_p)$ dadas las autocorrelaciones $(\rho_1, \rho_2, \dots, \rho_p)$ o bien la estimación de dichos parámetros $(\hat{\phi}_1, \hat{\phi}_2, \dots, \hat{\phi}_p)$ a partir de las autocorrelaciones muestrales (r_1, r_2, \dots, r_p)

- e) La forma del correlograma, representación gráfica de la función de autocorrelación, dependerá de cómo sean las raíces de la ecuación. Si éstas son reales, tendrá un comportamiento amortiguado hacia cero, siendo todos los valores positivos, o alternando el signo, aunque cumpliéndose siempre que el signo del primer pico coincide con el valor de ϕ_1 . Si las raíces son complejas, para lo cual necesariamente debe cumplirse que $\phi_j < 0$, el correlograma tendrá un comportamiento sinusoidal, pero también tendiente a cero.
- f) La función de autocorrelación parcial se anula para retardos superiores al orden. Por consiguiente, su representación gráfica presentará el número de picos de acuerdo al orden distintos de cero. Con respecto a los signos de estos picos coincidirán, respectivamente, con los signos de los valores de ϕ .

2.4 MODELOS AUTORREGRESIVOS Y DE MEDIAS MÓVILES (ARMA)

Estos modelos pueden representarse en forma asociada con modelos de medias móviles o autorregresiva. A estos tipos de modelos se denominan modelos mixtos autorregresivo- medias móviles de orden p, q, abreviadamente denotado por ARMA (p, q).

2.4.1 Modelos mixtos Autorregresivos- medias móviles de orden uno, uno. ARMA (1,1)

Su ecuación en diferencias viene dada por:

$$y_t = \phi_1 y_{t-1} + \delta + u_t - \theta_1 u_{t-1}$$

o bien, utilizando los operadores polinomiales de retardos se tiene

$$\phi(L)y_t = \delta + \theta(L)u_t$$

siendo

$$\phi(L) = 1 - \phi_1 L$$

$$\theta(L) = 1 - \theta_1 L$$

El modelo ARMA (1,1) será invertible cuando la ecuación de la raíz de la ecuación $1 - \theta_1 L = 0$ sea, en valor absoluto, mayor que uno; esto es cuando:

$$|L| = \frac{1}{|\theta_1|} > 1$$

de donde se concluye que la condición de invertibilidad en este caso implica que:

$$|\theta_1| < 1$$

Por otro lado, el modelo ARMA (1,1) será estacionario cuando la raíz cuadrada de la ecuación $1 - \phi_1 L = 0$ sea, en valor absoluto, mayor que uno; esto es, cuando:

$$|L| = \frac{1}{|\phi_1|} > 1$$

lo que implica que:

$$|\phi_1| < 1$$

Cumplida la condición de estacionariedad obtenemos la esperanza matemática, la cual viene dada por:

$$E(y_t) = \mu = \frac{\delta}{1 - \phi_1}$$

Una forma alternativa de escribir un modelo ARMA(1,1) estacionario, es en forma de desviaciones con respecto a la media:

$$\tilde{y}_t = \phi_1 \tilde{y}_{t-1} + u_t - \theta_1 u_{t-1}$$

donde:

$$\tilde{y}_{t-j} = y_{t-j} - \mu = y_{t-j} - \frac{\delta}{1 - \phi_1} \quad \text{para } j = 0 \text{ y } 1$$

o bien

$$\phi(L)\tilde{y}_t = \theta(L)u_t$$

La varianza del modelo viene dada por

$$\gamma_0 = \frac{\sigma_u^2(1 + \theta_1^2 - 2\phi_1\theta_1)}{1 - \phi_1^2}$$

La función de autocovarianzas viene dado por

$$\gamma_j = \begin{cases} \phi_1 \gamma_0 - \theta_1 \sigma_u^2 = \frac{\sigma_u^2(1 - \phi_1\theta_1)(\phi_1 - \theta_1)}{1 - \phi_1^2} & \text{para } j = 1 \\ \phi_1 \gamma_{j-1} & \text{para } j > 1 \end{cases}$$

La función de autocorrelación

$$\rho_j = \begin{cases} \frac{(1 - \phi_1\theta_1)(\phi_1 - \theta_1)}{1 - \theta_1^2 - 2\phi_1\theta_1} & \text{para } j = 1 \\ \phi_1 \rho_{j-1} & \text{para } j > 1 \end{cases}$$

La función de autocorrelación parcial

$$\begin{aligned}\phi_{11} &= \rho_1 \\ \phi_{22} &= \frac{\rho_2 - \rho_1^2}{1 - \rho_1^2} \\ \phi_{33} &= \frac{\rho_1^3 - \rho_1\rho_2(2 - \rho_2) + \rho_3(1 - \rho_1^2)}{1 - \rho_2^2 - 2\rho_1^2(1 - \rho_2)} \\ &\dots\end{aligned}$$

Las características del modelo ARMA (1,1) son las siguientes

1- El modelo ARMA (1,1) será estacionario e invertible siempre y cuando se cumpla que $|\phi_1| < 1$ y $|\theta_1| < 1$. Adicionalmente, debe satisfacerse que las raíces de $1 - \phi_1L = 0$ y $1 - \theta_1L = 0$ no se cancelen, esto es, debe cumplirse que $\phi_1 \neq \theta_1$.

2- La función de autocorrelación no se anula, si bien tiende a cero a medida que aumentan los desfases temporales considerados. Consecuentemente la memoria de este modelo es infinita. La presencia del término de medias móviles en el modelo ARMA (1,1) afecta únicamente a la determinación de ρ_1 , estando el resto de las autocorrelaciones afectadas únicamente por la parte autorregresiva del modelo. Concretamente, el análisis de la función de autocorrelación muestra que una vez superado el orden de la parte de las medias móviles, las autocorrelaciones del modelo ARMA (1,1) se comportan como las de un modelo AR(1).

3- El correlograma presenta un pico para el retardo 1, en el que influye tanto la parte autorregresiva como las de medias móviles del modelo y, a continuación, su forma coincide con la de los modelos AR(1), esto es, comportamiento amortiguado hacia cero sin llegar a anularse.

4- La función de autocorrelación parcial no se anula, como consecuencia de que el modelo ARMA (1,1) contiene al modelo MA(1) como caso especial. La forma de la representación gráfica de la función de autocorrelación parcial es complicada, si bien

la característica que cabe resaltar es su comportamiento amortiguado hacia cero sin llegar a anularse.

2.4.2 Modelos mixtos Autorregresivos- medias móviles de orden p, q. ARMA (p, q)

Su ecuación en diferencias viene dada por:

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \dots + \phi_p y_{t-p} + \delta + u_t - \theta_1 u_{t-1} - \theta_2 u_{t-2} - \dots - \theta_q u_{t-q}$$

Este modelo puede escribirse utilizando los operadores polinomiales de retardos

$$\phi(L)y_t = \delta + \theta(L)u_t$$

Este modelo es estacionario e invertible si tanto las raíces de $\phi(L) = 1 - \phi_1 L - \dots - \phi_p L^p = 0$ como las de $\theta(L) = 1 - \theta_1 L - \dots - \theta_q L^q$ caigan fuera del círculo unidad.

Una condición necesaria, aunque no suficiente, para que un modelo mixto autorregresivo- medias móviles de cualquier orden sea estacionario e invertible es la siguiente:

$$\begin{aligned} \theta_1 + \theta_2 + \dots + \theta_q &< 1 \\ \phi_1 + \phi_2 + \dots + \phi_p &< 1 \end{aligned}$$

Adicionalmente a la estacionariedad e invertibilidad supondremos que las p raíces de $\phi(L) = 0$ y las q raíces de $\theta(L) = 0$ no se cancelan, esto es, no hay raíces comunes. Bajo estas condiciones, la esperanza matemática y la varianza del modelo ARMA(p,q) se definen así:

Esperanza matemática

$$\mu = \frac{\delta}{1 - \phi_1 - \phi_2 - \dots - \phi_p}$$

Prueba.

Sea

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \dots + \phi_p y_{t-p} + \delta + u_t$$

Si el proceso es estacionario

$$E(y_t) = \phi_1 E(y_{t-1}) + \phi_2 E(y_{t-2}) + \dots + \phi_p E(y_{t-p}) + E(\delta) \\ + E(u_t) - \theta_1 E(u_{t-1}) - \theta_2 E(u_{t-2}) - \dots - \theta_q E(u_{t-q})$$

Todos los términos de la parte de medias móviles, su esperanza para éstos es cero, ya que se trata de un ruido blanco (media constante y varianza constante) por lo que se obtiene la expresión

$$E(y_t) = \phi_1 E(y_{t-1}) + \phi_2 E(y_{t-2}) + \dots + \phi_p E(y_{t-p}) + E(\delta)$$

De donde:

$$\mu = \phi_1 \mu + \phi_2 \mu + \dots + \phi_p \mu + \delta$$

Despejando δ se tiene:

$$\mu(1 - \phi_1 - \phi_2 - \dots - \phi_p) = \delta$$

Así tenemos que:

$$\mu = \frac{\delta}{(1 - \phi_1 - \phi_2 - \dots - \phi_p)}$$

Varianza

$$\gamma_0 = \frac{\sigma_u^2 (1 + \theta_1^2 + \dots + \theta_q^2 - 2\phi_1\theta_1 - 2\phi_2\theta_2 - \dots - 2\phi_h\theta_h)}{1 - \phi_1^2 - \phi_2^2 - \dots - \phi_p^2}$$

Función de autocovarianzas

$$\gamma_j = \begin{cases} \phi_1\gamma_0 + \phi_2\gamma_1 + \dots + \phi_p\gamma_{p-1} - \theta_1\sigma_u^2 - \theta_2E(\tilde{y}_{t-1}u_{t-2}) - \dots - \theta_qE(\tilde{y}_{t-1}u_{t-q}) & \text{para } j=1 \\ \phi_1\gamma_1 + \phi_2\gamma_0 + \dots + \phi_p\gamma_{p-2} - \theta_2\sigma_u^2 - \theta_3E(\tilde{y}_{t-1}u_{t-3}) - \dots - \theta_qE(\tilde{y}_{t-2}u_{t-q}) & \text{para } j=2 \\ \dots\dots\dots & \\ \phi_1\gamma_{q-1} + \phi_2\gamma_{q-2} + \dots + \phi_p\gamma_{p-q} - \theta_q\sigma_u^2 & \text{para } j=q \\ \phi_1\gamma_{j-1} + \phi_2\gamma_{j-2} + \dots + \phi_p\gamma_{p-q} & \text{para } j>q \end{cases}$$

Función de Autocorrelaciones

$$\rho_j = \begin{cases} \phi_1 + \phi_2\rho_1 + \dots + \phi_p\rho_{p-1} + \frac{1}{\gamma_0}[-\theta_1\sigma_u^2 - \theta_2E(\tilde{y}_{t-2}u_{t-2}) - \dots - \theta_qE(\tilde{y}_{t-1}u_{t-q})] & \text{para } j=1 \\ \phi_1\rho_1 + \phi_2 + \dots + \phi_p\rho_{p-2} + \frac{1}{\gamma_0}[-\theta_2\sigma_u^2 - \theta_3E(\tilde{y}_{t-2}u_{t-3}) - \dots - \theta_qE(\tilde{y}_{t-2}u_{t-q})] & \text{para } j=2 \\ \dots\dots\dots & \\ \phi_1\rho_{q-1} + \phi_2\rho_{p-2} + \dots + \phi_p\rho_{p-q} + \frac{1}{\gamma_0}\theta_q\sigma_u^2 & \text{para } j=q \\ \phi_1\rho_{j-1} + \phi_2\rho_{j-2} + \dots + \phi_p\rho_{j-q} & \text{para } j>q \end{cases}$$

Características:

- a) El modelo ARMA será estacionario e invertible siempre y cuando se cumpla que las raíces de $\phi(L) = 0$ y de $\theta(L) = 0$ se ubican afuera del círculo de la unidad. Adicionalmente, supondremos que no hay raíces comunes.
- b) La función de autocorrelación no se anula, si bien tiende a cero a medida que aumentan los desfases temporales considerados. Consecuentemente, la memoria de este modelo es infinita. La presencia del término de medias móviles en el modelo ARMA(p,q) afecta la determinación de las q primeras autocorrelaciones ($\rho_1, \rho_2, \dots, \rho_q$) estando el resto de las autocorrelaciones afectadas únicamente por la parte autorregresivo. Concretamente, el análisis de la función de autocorrelación muestra que una vez superado el orden de la parte de

medias móviles, las autocorrelaciones del modelo ARMA(p,q) se comportan como las de un modelo AR(p).

- c) El correlograma, la representación gráfica de autocorrelación, presentará q autocorrelaciones, hasta el retardo q , en los que influye tanto la parte autorregresiva como el de las medias móviles del modelo y, a continuación, su forma coincide con los del modelo AR(p), es decir, un comportamiento amortiguado hacia cero sin llegar a anularse, si las raíces son reales, o comportamiento sinusoidal, pero también amortiguado hacia cero, si las raíces son complejas.
- d) La función de autocorrelación parcial no se anula, como consecuencia de que el modelo ARMA(p,q) contiene el modelo MA(q) como un caso especial. La representación gráfica de la función de autocorrelación parcial tendrá también un comportamiento amortiguado hacia cero.

2.5 MODELOS ARIMA(p , d, q)

Dada una serie y_t y supóngase que no cumple con los requisitos de estacionariedad en media y en varianza. Para obtener un modelo ARIMA será necesario la aplicación de ciertas transformaciones para convertir dicha serie temporal en estacionaria, esto nos posibilitará continuar con la aplicación de la metodología de Box-Jenkins.

Estas transformaciones se concretan en diferenciaciones sucesivas en la serie (la determinación del parámetro d), diferenciaciones estacionales en la serie (la determinación de D), y aplicaciones de la transformada de Box-Cox (determinación de λ). Las dos primeras se aplican para conseguir estacionariedad en media y la última para lograr estacionariedad en varianza.

Si la serie temporal no presenta componente estacional (aunque ésta puede ser mensual, trimestral, anual, etc.), la serie esta desestacionalizada previamente o bien porque carece de un patrón de comportamiento estacional.

Ahora bien, si diferenciamos una vez y_t , se obtiene un modelo transformado de la serie

$$w_t = y_t - y_{t-1}$$

donde:

$$w_t = \Delta y_t = y_t - y_{t-1} = (1 - L)y_t$$

Al diferenciar la serie temporal se pasa de y_t al proceso w_t . Ahora considérese el caso inverso de obtener y_t a partir del proceso w_t . Por sustituciones sucesivas se obtiene que

$$y_t = w_t + y_{t-1} = w_t + w_{t-1} + y_{t-2} = w_t + w_{t-1} + w_{t-2} + y_{t-3} = \dots = w_t + w_{t-1} + w_{t-2} + w_{t-3} \dots$$

obteniéndose así

$$y_t = w_t + w_{t-1} + w_{t-2} + w_{t-3} + \dots$$

Como puede notarse, que el proceso y_t se obtiene sumando, o lo que es lo mismo “integrando” el proceso w_t .

A este proceso integrado, y_t se le denomina proceso autorregresivo-medias móviles integrado o más abreviadamente ARIMA(p,d,q), si tomando diferencias de orden d se obtiene un proceso estacionario w_t del tipo ARMA(p,q).

La forma general del modelo ARIMA(p,d,q) es

$$w_t = \phi_1 w_{t-1} + \dots + \phi_p w_{t-p} + u_t - \theta_1 u_{t-1} - \dots - \theta_q u_{t-q}$$

donde:

$$w_{t-j} = \Delta^d y_{t-j} = (1-L)^d y_{t-j} \quad \text{para } j = 0, 1, 2, 3, \dots, p$$

Más brevemente utilizando los operadores de polinomios de retardos definidos, en los apartados anteriores podemos escribir:

$$\phi(L)w_t = \theta(L)u_t$$

donde:

$$w_t = \Delta^d y_t = (1 - L)^d y_t$$

El modelo ARIMA(p,d,q) puede considerarse como un modelo estocástico general que englobe a los modelos analizados. Así, si $p = d = 0$ estamos ante el modelo ARMA(0,0,q) o bien MA(q) abordados en la sección 2.2; si $q = d = 0$ se tiene el modelo ARMA(p, 0, 0) o bien un modelo AR(p) abordado en la sección 2.3. Finalmente, si $d = 0$ se tiene el modelo ARMA(p, 0, q) o bien ARMA(p,q) abordado en la sección 2.4

Otros casos de interés son:

a) $q = 0 \rightarrow ARIMA(p,d,0)$ o $ARI(p,d)$

b) $p = 0 \rightarrow ARIMA(0,d,q)$ o $IMA(d,q)$

c) $p = q = 0; d = 1 \rightarrow ARIMA(0,d,0)$ o "paseo aleatorio"

d) $p = d = q = 0$ (y $\delta = 0$) $\rightarrow ARIMA(0,0,0)$ o "ruido blanco"

Al analizar la evolución de la mayoría de las series temporales suele obtenerse que éstas presentan una tendencia creciente o decreciente. La eliminación de esta tendencia (no estacionaria en media) de la serie se puede conseguir mediante las diferenciaciones implícitas en los modelos ARIMA. En ocasiones se observa también, que existe una tendencia en la varianza, es decir, la dispersión de las observaciones no es constante a lo largo del tiempo, la cual no se elimina mediante estas diferenciaciones. Cuando se presenta esta situación la transformación adecuada puede consistir en tomar logaritmos neperianos.

Esta posibilidad de transformar la serie puede concretarse de la forma más general mediante la transformada de Box-Cox. De esta forma, el modelo ARIMA más general puede enunciarse como:

$$\phi(L)\Delta^d y_t^{(\lambda)} = \delta + \theta(L)u_t$$

O bien

$$\phi(L)(1-L)^d (y_t^{(\lambda)} - \mu) = \theta(L)u_t$$

donde μ es la media de $y_t^{(\lambda)}$, siendo:

$$y_t^{(\lambda)} = \begin{cases} \frac{y_t^{(\lambda)} - 1}{\lambda} & \text{para } \lambda \neq 0 \\ \ln y_t & \text{para } \lambda = 0 \end{cases}$$

Restringiéndonos a los casos $\lambda = 1$ y $\lambda = 0$, tendremos respectivamente la serie sin transformar (la serie original) o la serie transformada aplicando logaritmos neperianos.

CAPÍTULO 3 AJUSTE DEL MODELO.

3.1. ELEMENTOS A CONSIDERAR EN LA BÚSQUEDA DEL MODELO.

El obtener el posible modelo estocástico capaz de generar los datos de la serie observada es lo que preocupa a todo investigador, por lo que partiremos revisando los elementos importantes a considerar en la búsqueda del modelo. Para comenzar el estudio de una serie se hace un gráfico que muestre la evolución de los datos a través del tiempo y se observa la presencia o ausencias de las siguientes características

- **LA TENDENCIA.** La tendencia implica que la serie tiende a crecer o a decrecer a largo plazo. Cuando una serie permanece más o menos constante, oscilando en torno a un valor, decimos que la serie no tiene tendencia, es decir, la serie es estacionaria en media.

- **VARIABILIDAD.** Decimos que una serie es **HOMOCEDÁSTICA**, si su variabilidad se mantiene constante a lo largo del tiempo, o bien se dice que la serie es estacionaria en varianza. Cuando la variabilidad de la serie aumenta o disminuye a lo largo del tiempo, decimos que la serie es **HETEROCEDÁSTICA**.

- **PERIODICIDAD** es el intervalo de tiempo en la que se observa cada valor de la serie de los datos. La periodicidad puede ser: Anual (Se toma un dato cada año), Mensual (Se toma un dato cada mes), Semanal (Se toma un dato cada semana), Diaria (Se toma un dato cada día), por Horas (se toma un dato cada hora), etc.

Hasta ahora hemos descrito el aspecto de la serie. Sin embargo, cuando se quiere analizar la serie es necesario identificar la estructura que la genera, es decir, cómo influyen las observaciones del pasado en las observaciones del futuro. Para identificar esta dependencia utilizamos dos definiciones importantes, la Función de Autocorrelación Simple (FAS) y la Función de Autocorrelación Parcial (FAP).

A continuación se definen y se comenta en forma breve cada uno de estos elementos:

Función de autocorrelación simple (FAS)

La autocorrelación de orden j (ρ_j) es la correlación separada j períodos de la misma serie temporal. Esto es:

$$\rho_j = \text{Corr}(y_t, y_{t-j}) = \frac{\text{Cov}(y_t, y_{t-j})}{\sqrt{\text{Var}(y_t)}\sqrt{\text{Var}(y_{t-j})}} = \frac{\gamma_j}{\gamma_0}$$

Obsérvese que se ha supuesto que $\text{Var}(y_t) = \text{Var}(y_{t-j}) = \gamma_0$; al conjunto de autocorrelaciones obtenidas para distintos valores de j se le denomina función de autocorrelación. Es fácil verificar las siguientes propiedades de la función de autocorrelación:

- a) $\rho_0 = 1$
- b) $|\rho_j| \leq 1$
- c) $\rho_j = \rho_{-j}$ (simetría)

Adicionalmente a estas tres propiedades, supondremos también que se satisface la siguiente:

$$d) \lim_{j \rightarrow \infty} \rho_j = 0$$

El cumplimiento de la propiedad d) es una condición necesaria para poder establecer que el proceso estocástico en cuestión, además de ser estacionario es ergódico. Precisaremos trabajar con procesos estocásticos estacionarios y ergódicos para poder efectuar la inferencia del proceso estocástico que ha podido generar la serie temporal en estudio. Para ello debemos estimar los parámetros que configuran la función de autocovarianza y autocorrelación tal como se expondrá en apartados posteriores. Pues bien, de manera intuitiva podemos señalar que la ergodicidad posibilita obtener estimadores consistentes

de dichos parámetros. Si no se satisface la propiedad d), esto es, el valor de ρ_j tuviera valores elevados para ordenes j altos, significaría que al aumentar el tamaño de la muestra se añadiría poca información nueva como consecuencia que en dicho caso deberá calcularse un mayor número de autocovarianzas (o autocorrelaciones) para caracterizar adecuadamente el proceso.

Función de Autocorrelación Parcial (FAP).

El concepto de coeficiente de autocorrelación parcial entre y_t y y_{t-j} es similar al de coeficiente de autocorrelación entre dichos valores de la serie temporal, salvo porque se mide dicha correlación ajustada por el efecto de los retardos intermedios. Es decir, al obtener la autocorrelación parcial entre y_t e y_{t-j} se elimina la influencia que, sobre ambas, tienen $y_{t-1}, y_{t-2}, \dots, y_{t-j+1}$.

Analíticamente, se define el coeficiente de autocorrelación parcial de orden j (Π_j) como la correlación condicional siguiente:

$$\Pi_j = \text{corr}(y_t, y_{t-j} / y_{t-1}, y_{t-2}, \dots, y_{t-j+1})$$

Considerando sin pérdida de generalidad, que el proceso estacionario $\{y_t\}$ tiene media nula, el coeficiente de autocorrelación parcial (Π_j) será igual al coeficiente de autocorrelación entre $(y_t - \hat{y}_t)$ e $(y_{t-j} - \hat{y}_{t-j})$ siendo \hat{y}_t e \hat{y}_{t-j} , respectivamente las mejores predicciones lineales, en el sentido de mínimo error cuadrático medio, de y_t y de y_{t-j} como funciones lineales de $y_{t-1}, y_{t-2}, \dots, y_{t-j+1}$. Esto es:

$$\Pi_j = \frac{\text{Cov}[(y_t - \hat{y}_t), (y_{t-j} - \hat{y}_{t-j})]}{\sqrt{\text{Var}(y_t - \hat{y}_t)} \sqrt{\text{Var}(y_{t-j} - \hat{y}_{t-j})}}$$

Después de realizar una serie de operaciones algebraicas se tiene

$$\Pi_j = \frac{\rho_j - \alpha_1 \rho_{j-1} - \alpha_2 \rho_{j-2} - \dots - \alpha_{j-1} \rho_1}{1 - \alpha_1 \rho_1 - \alpha_2 \rho_2 - \dots - \alpha_{j-1} \rho_{j-1}}$$

Una vez comentadas en forma breve las características a considerar en la búsqueda del modelo aplicaremos la metodología de Box-Jenkins. Identificaremos los procesos ARIMA susceptible de generar la serie estacionaria o no, $\{y_t\}$. Para llevar a cabo este paso utilizaremos los instrumentos importantes que caracterizan a los procesos AR, MA, ARMA y ARIMA: las funciones de autocorrelaciones simples y parciales cuyas características de su lectura son las más relevantes para la búsqueda de los modelos y sus respectivos órdenes.

Tabla 3.1 Características Teóricas de la FAS y la FAP de los Procesos Estacionarios

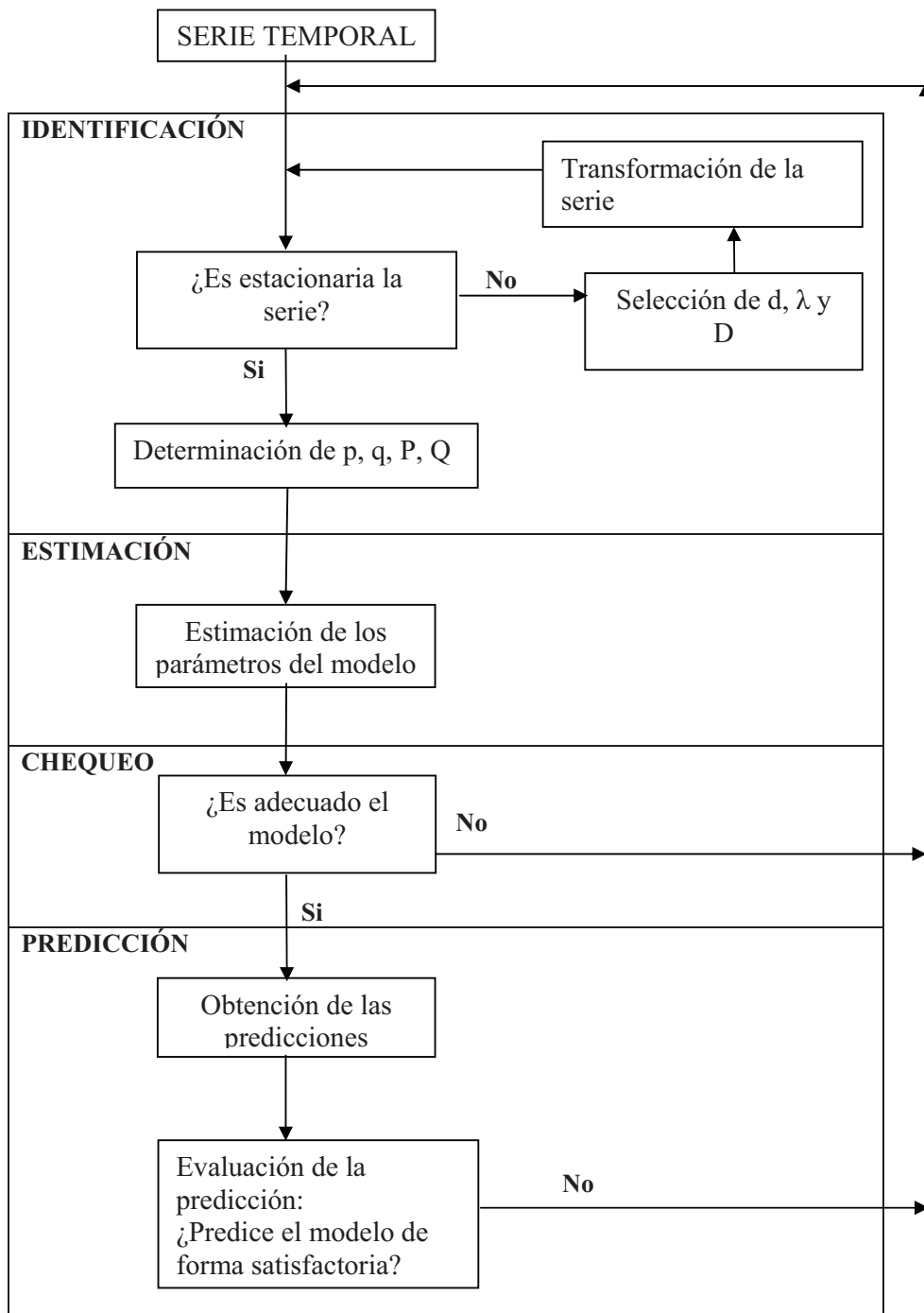
	FAS	FAP
MA(q)	Se anula para retardos superiores a q	Decrecimiento rápido sin llegar a anularse
AR(p)	Decrecimiento rápido sin llegar a anularse	Se anula para retardos superiores a p
ARMA(p, q)	Decrecimiento rápido sin llegar a anularse	Decrecimiento rápido sin llegar a anularse

Aunque los valores de las autocorrelaciones simples (FAS) y parciales (FAP), serán desde luego desconocidos podrán estimarse a partir de la serie temporal observada, $\{y_t\}$. Se trata de comparar el comportamiento de los estimadores de las FAS y FAP con el comportamiento teórico que tiene cada uno de los diferentes modelos.

3.2 DESCRIPCIÓN DE LA METODOLOGÍA DE BOX - JENKINS.

Se presenta el esquema a seguir para la aplicación de la metodología de Box-Jenkins, mencionando las etapas y en forma breve se comentan en que consisten cada uno.

ETAPAS DE LA METODOLOGÍA DE BOX – JENKINS



A continuación se hace referencia en forma breve a cada una de las etapas presentadas en el esquema anterior:

a) **Identificación:** En esta etapa es habitual de que existan más de un modelo que haya podido generar la serie. En estos casos se trata de seleccionar todos los posibles modelos candidatos, discriminando en etapas posteriores, algunos de ellos.

b) **Estimación:** En esta etapa se trata de cuantificar los parámetros de los modelos seleccionados. Los dos problemas fundamentales a los que se enfrentan la estimación de los modelos ARIMA son el de los valores iniciales (de los parámetros, de la serie y de los ruidos) y la falta de linealidad.

c) **Validación o Chequeo:** En esta etapa se procede a realizar la evaluación de la adecuación de los modelos previamente identificados y estimados. Esta etapa se centra en analizar si los residuos del modelo, tienen un comportamiento similar a las perturbaciones del mismo, esto es sí puede afirmarse que son semejantes a un ruido blanco.

También, se trata de comprobar la calidad de las estimaciones, así como el cumplimiento de las estimaciones de los parámetros, de las condiciones de estacionariedad e invertibilidad que deben satisfacer los parámetros de estos modelos.

d) **Predicción:** Una vez los modelos superan la etapa del chequeo se llega al fin último de la metodología, la obtención de las predicciones (puntuales y/o por intervalo) de valores futuros de la serie temporal.

3.3 APLICACIÓN DE LA METODOLOGÍA DE BOX-JENKINS.

Para la aplicación de esta metodología utilizaremos los datos de la emanación de gas dióxido de carbono, CO₂ obtenidos por los vulcanólogos en el volcán de San Vicente. Datos que fueron recolectados durante los meses de marzo y abril del año 2003.

Es importante señalar, que en la captura de la información se obtuvieron otras mediciones, tales como la temperatura de la tierra, humedad, temperatura del aire, actividad sísmica, cambios del microclima, etc. Sin embargo, en este estudio se ajustará un modelo ARIMA para la información de la emanación de gas CO₂. Con esta información se pretende establecer una alternativa para detectar cualquier anomalía o

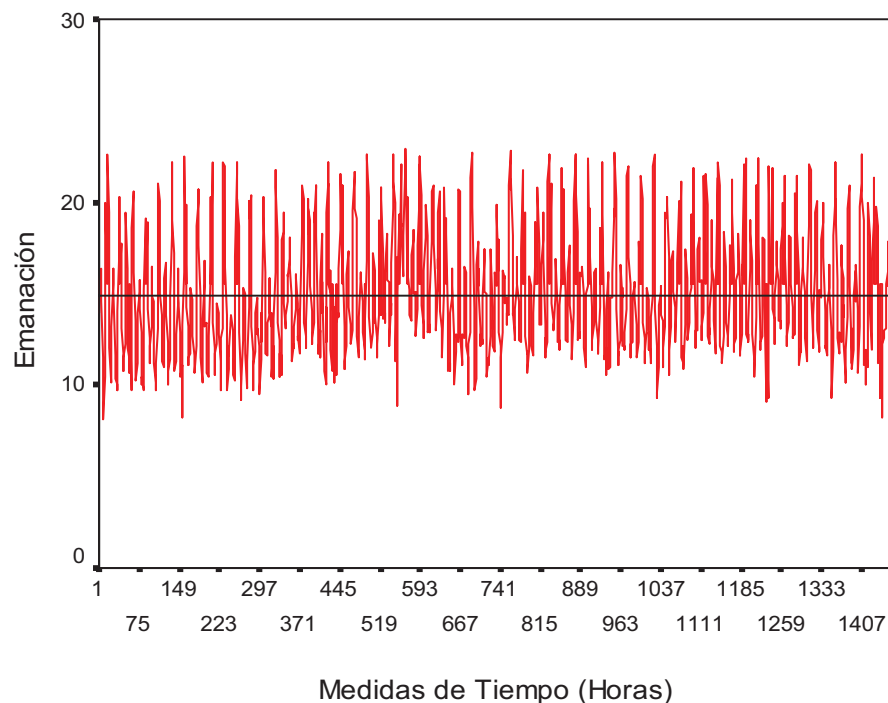
cambio que pudiera indicar intranquilidad volcánica a través del tiempo y de aportar los conocimientos técnicos necesarios para estudiar otros volcanes de nuestra región con la intención de alertar a las instituciones y a la población en caso de reactivación de la actividad volcánica.

3.3.1 IDENTIFICACIÓN DEL MODELO

Esta fase como se ha comentado, se tratará de identificar los posibles modelos capaces de generar la serie, se pretende en esta fase catalogar todos los posibles modelos candidatos que posteriormente se irán descartando en la revisión de cada una de las siguientes etapas propuestas por la metodología de Box-Jenkins.

Para comenzar nuestro estudio de la serie emanación del gas CO_2 del volcán de San Vicente elaboramos el gráfico que muestre la evolución de los datos a través del tiempo y se analizará si se observa la presencia o ausencias de las siguientes características: tendencia, homocedasticidad y estacionalidad.

Figura 3.1 Gráfico de la Serie Emanación del gas CO_2
Volcán de San Vicente



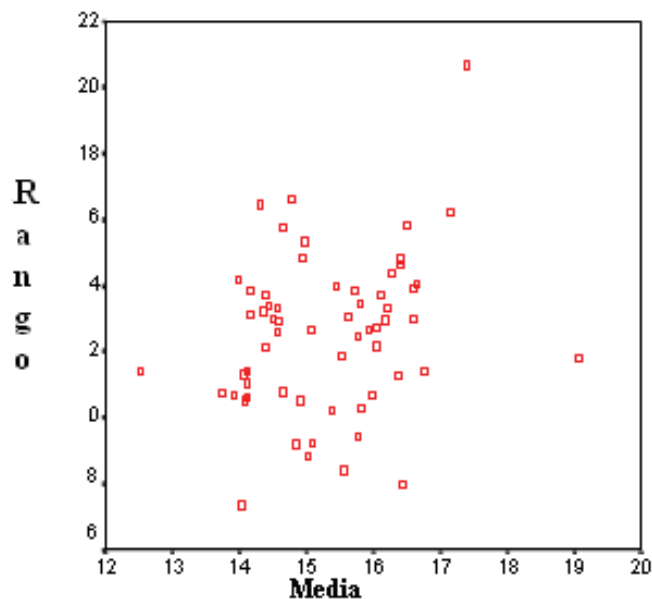
De la figura 3.1 se pueden extraer la información siguiente:

a) Los valores de la serie giran en torno a su media a lo largo del tiempo, por lo que se puede conjeturar que la serie es estacionaria en media.

b) No se observa el comportamiento de datos que se manifiesten en forma periódica a lo largo de la serie, es decir, no se observa componente estacional. Esto se debe en gran medida por la longitud del intervalo de tiempo en que fueron obtenidos los datos (por horas y de forma continua), por lo que podemos concluir que en la serie emanación del gas CO₂ no se contempla la componente estacional a lo largo de la serie.

c) Se observa que la variabilidad de la serie permanece aproximadamente constante a lo largo del tiempo, por lo que cabe señalar que la serie es estacionaria en varianza. Con el fin de verificar esta conjetura se presenta a continuación el diagrama rango media.

Figura 3.2 Diagrama rango media de la serie Emanación de gas CO₂



La conclusión que puede extraerse de la figura 3.2 es que el cúmulo de puntos no están alineados en torno a una recta con una pendiente positiva, no se observa un comportamiento definido, por lo tanto, la serie es estacionaria en varianza.

Después de hacer el análisis de los gráficos anteriores se concluye, *que la serie de las mediciones de CO₂ cumple los siguientes aspectos:*

a) Los valores de la serie giran en torno a una constante de manera continua, por lo cual parece plausible afirmar que esta serie es estacionaria en media.

b) No se observa componente estacional, comportamiento de una medición que se manifieste de manera periódica.

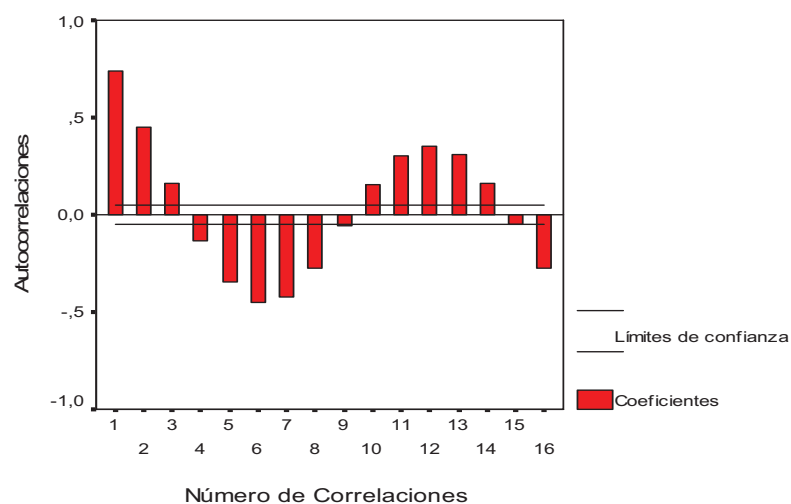
c) La variabilidad de la serie permanece aproximadamente constante a lo largo del tiempo, por lo cual cabe señalar que la serie es estacionaria en varianza.

Por lo anterior, no parece adecuado realizar ninguna transformación en la serie.

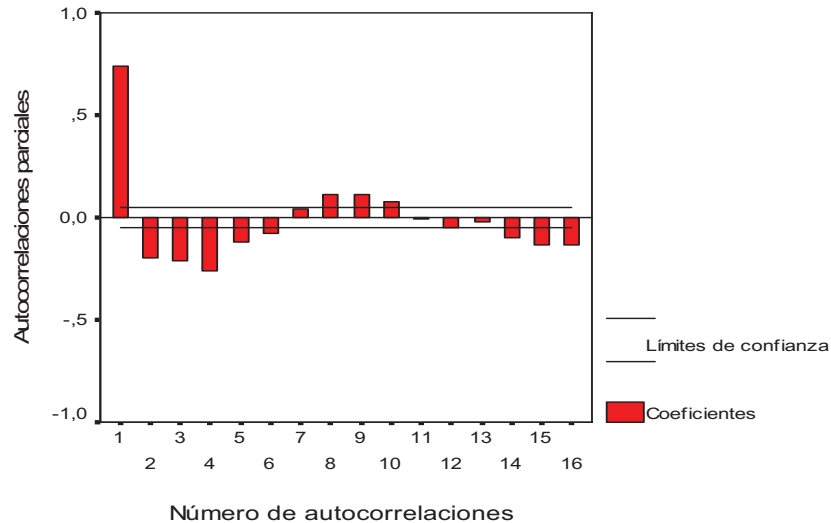
Una vez determinada la estacionariedad en media y en varianza, los instrumentos que utilizaremos para identificar el modelo son la función de autocorrelación muestral simple y la función de autocorrelación muestral parcial. A continuación se presentan los gráficos:

Figura 3.3 Autocorrelaciones de la Serie Emanación del gas CO₂ Volcán de San Vicente

a) Autocorrelaciones Muestrales de la Serie Emanación CO₂



b) Autocorrelaciones Parciales de la Serie Emanación CO₂



Revisamos el comportamiento de las autocorrelaciones:

En las Autocorrelaciones Muestrales Simples (Fig. 3.3.a):

Se observa que las dos primeras autocorrelaciones que son altamente significativas, diferentes de cero, por lo que estamos ante la presencia de un modelo de medias móviles de orden 2, MA (2).

También, puede observarse que las autocorrelaciones 3, 4, 9 y 10 no son estadísticamente significativas, por lo que pueden considerarse igual a cero.

También, podríamos conjeturar la existencia de otros modelos, como el modelo MA de orden 5 (MA(5)), 6 (MA (6)), 7 (MA (7)), 11(MA(11)), 12(MA(12)) y 13(MA(13)) esto se debe por el número de autocorrelaciones que son las más significativas.

Las autocorrelaciones decrecen de manera lenta hacia cero por lo que podríamos conjeturar la existencia de un modelo autorregresivo.

En las autocorrelaciones Parciales, (Fig. 3.3.b)

Se observa que la primera, la segunda, tercera y cuarta autocorrelación son estadísticamente significativas; luego el resto de las autocorrelaciones pueden considerarse iguales a cero. Por lo tanto, el gráfico 3.3.b cumple con las características de un modelo autorregresivo de orden 4 (AR(4)). Sin embargo, la función de

autocorrelación simple no muestra una tendencia hacia cero como para poder afirmar que únicamente se tiene un modelo AR(4), por tanto esto nos indica que dicho modelo esta acompañado de un modelo de medias móviles.

A partir del análisis de las autocorrelaciones, es evidente que el modelo buscado es del tipo ARMA (p, q), cuya ecuación es de la forma

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \dots + \phi_p y_{t-p} + \delta + u_t - \theta_1 u_{t-1} - \theta_2 u_{t-2} - \dots - \theta_q u_{t-q}$$

y habrá que encontrar el orden del modelo. Los posibles modelos obtenidos de la observación de las autocorrelaciones simples y parciales son los siguientes: ARMA(1,6), ARMA(1,7), ARMA(2,2), ARMA(4,1) y ARMA(4,2).

3.3.2 ESTIMACIÓN DEL MODELO

Una vez identificados los posibles modelos ARMA(1,6), ARMA(1,7), ARMA(2,2), ARMA(4,1) y ARMA(4,2) que han podido generar la serie temporal, se tratará de estimar los parámetros de los mismos.

A continuación se presenta la ecuación en diferencias en forma general, los parámetros estimados, los coeficientes del modelo y sus respectivos P-valores.

Modelo mixto autorregresivo - medias móviles; ARMA (1,6)

La ecuación general de este modelo viene dada por

$$y_t = \phi_1 y_{t-1} + \delta + u_t - \theta_1 u_{t-1} - \theta_2 u_{t-2} - \theta_3 u_{t-3} - \theta_4 u_{t-4} - \theta_5 u_{t-5} - \theta_6 u_{t-6}$$

Tabla 3.2 PARÁMETROS DEL MODELO ARMA (1, 6)

Variables del Modelo				
Modelos	Coeficientes	Desviación Estándar de los coeficientes	Prueba T-RATIO	P- valor
AR1	0.754252	0.07386190	10.21165	0.00000000
MA1	-0.006017	0.07517444	-0.,08004	0.93621701
MA2	0.037105	0.05730637	0.64748	0.51742470
MA3	0.013549	0.04210016	0.32183	0.74763042
MA4	0.135033	0.03497780	3.86053	0.00011810
MA5	0.139271	0.02664105	5.22770	0.00000020
MA6	0.188201	0.02616688	7.19232	0.00000000
CONSTANTE	14.939890	0.10528042	141.90568	0,00000000

El coeficiente que aparece en la última fila y segunda columna de la tabla 3.2 representa la media muestral de los datos, es decir, un estimador de μ es $\hat{\mu} = 14.933$ pero la media para el modelo ARMA (1,6) está dada por:

$$\hat{\mu} = \phi_1 \mu + \hat{\delta}$$

Por lo tanto, un estimador de $\hat{\delta}$ es:

$$\hat{\delta} = \mu(1 - \phi_1)$$

Para el modelo ARMA (1,6), el estimador¹, $\hat{\delta}$ la constante del modelo está dado por

$$\hat{\delta} = 14.93(1 - 0.7543) = 3.67$$

De la tabla 3.2 extraemos los parámetros estimados del modelo y juntamente con su estimador, $\hat{\delta}$ formulamos la ecuación en diferencias del modelo ARMA (1,6), la cual está dada por:

$$y_t = 0.75y_{t-1} + 3.67 + u_t + 0.006_1u_{t-1} - 0.037u_{t-2} - 0.013u_{t-3} - 0.13u_{t-4} - 0.14u_{t-5} - 0.19u_{t-6}$$

Modelo mixto autorregresivo - medias móviles; ARMA (1,7)

La ecuación general de este modelo viene dada por

$$y_t = \phi_1 y_{t-1} + \delta + u_t - \theta_1 u_{t-1} - \theta_2 u_{t-2} - \theta_3 u_{t-3} - \theta_4 u_{t-4} - \theta_5 u_{t-5} - \theta_6 u_{t-6} - \theta_7 u_{t-7}$$

Tabla 3.3 PARÁMETROS DEL MODELO ARMA (1,7)

Variables del Modelo				
Modelos	Coefficientes	Desviación Estándar de los coeficientes	Prueba T-RATIO	P- valor
AR1	0,658356	0,07679431	8,57298	0,00000000
MA1	-0,105478	0,07809880	-1,35058	0,17704105
MA2	-0,050233	0,05907892	-0,85027	0,39531351
MA3	-0,063301	0,04457779	-1,42000	0,15582145
MA4	0,092055	0,03720792	2,47406	0,01347214
MA5	0,119519	0,02822906	4,23392	0,00002441
MA6	0,173291	0,02567563	6,74924	0,00000000
MA7	0,159811	0,02830772	5,64549	0,00000002
CONSTANTE	14,940300	0,10225599	146,10684	0,00000000

¹ La constante del estimador $\hat{\delta}$ se obtiene al calcular la media en el modelo ARMA(p,q). Capítulo 2 Apartado 2.4

El estimador de $\hat{\delta}$ para el modelo ARMA (1,7) es:

$$\hat{\delta} = 14.94(1 - 0.65) = 5.2$$

De la tabla 3.3 extraemos los parámetros estimados del modelo y juntamente con su estimador, $\hat{\delta}$ formulamos la ecuación en diferencias del modelo ARMA (1,7), la cual está dada por:

$$y_t = 0.65y_{t-1} + 5.2 + u_t + 0.10u_{t-1} + 0.05u_{t-2} + 0.06u_{t-3} - 0.09u_{t-4} - 0.12u_{t-5} - 0.17u_{t-6} - 0.16u_{t-7}$$

Modelo mixto autorregresivo - medias móviles; ARMA (2,2)

La ecuación general de este modelo viene dada por

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \delta + u_t - \theta_1 u_{t-1} - \theta_2 u_{t-2}$$

Tabla 3.4 PARÁMETROS DEL MODELO ARMA (2, 2)

Variables del Modelo				
Modelo	Coefficientes	Desviación Estándar de los Coeficientes	Prueba (T-RATIO)	P- valor
AR1	1,611472	0,02331744	69,11015	0,0000000
AR2	-0,831889	0,01944203	-42,78819	0,0000000
MA1	0,940350	0,03457706	27,19576	0,0000000
MA2	-0,330054	0,03191618	-10,34127	0,0000000
CONSTANTE	14,938,669	0,08987654	166,21322	0,0000000

El estimador de $\hat{\delta}$ para el modelo ARMA (2,2) está dado por:

$$\hat{\delta} = \mu(1 - \phi_1 - \phi_2)$$

estimamos la constante del modelo:

$$\hat{\delta} = 14.93(1 - 1.611 + 0.8318) = 3.29$$

De la tabla 3.4 extraemos los parámetros estimados del modelo y juntamente con su estimador, $\hat{\delta}$ formulamos la ecuación en diferencias del modelo ARMA (2,2), la cual está dada por:

$$y_t = 1.61y_{t-1} - 0.83y_{t-2} + 3.29 + u_t - 0.94u_{t-1} + 0.33\theta_2u_{t-2}$$

Modelo mixto autorregresivo - medias móviles; ARMA (4,1)

La ecuación general de este modelo viene dada por

$$y_t = \phi_1y_{t-1} + \phi_2y_{t-2} + \phi_3y_{t-3} + \phi_4y_{t-4} + \delta + u_t - \theta_1u_{t-1}$$

Tabla 3. 5 PARÁMETROS DEL MODELO ARMA (4, 1)

Variables del Modelo				
Modelo	Coefficientes	Desviación Estándar de los Coeficientes	Prueba (T-RATIO)	P- valor
AR1	1,096194	0,07971235	13,75187	0,00000000
AR2	- 0,277235	0,07295701	-3,79997	0,00015067
AR3	0,023081	0,03890005	0,59335	0,55303888
AR4	-0,207313	0,03457249	-5,99647	0,00000000
MA1	0,341800	0,08066205	4,23743	0,00002403
CONSTANTE	14,939024	0,09145904	163,34114	0,00000000

El estimador de $\hat{\delta}$ para el modelo ARMA (4,1) está dado por:

$$\hat{\delta} = \mu(1 - \phi_1 - \phi_2 - \phi_3 - \phi_4)$$

estimamos la constante del modelo:

$$\hat{\delta} = 14.93(1 - 1.096 + 0.277 - 0.023 + 0.20) = 5.34$$

De la tabla 3.5 extraemos los parámetros estimados del modelo y juntamente con su estimador, $\hat{\delta}$ formulamos la ecuación en diferencias del modelo ARMA (4,1), la cual está dada por:

$$y_t = 1.096y_{t-1} - 0.277y_{t-2} + 0.023y_{t-3} - 0.20y_{t-4} + 5.34 + u_t - 0.34u_{t-1}$$

Comentamos que el modelo ARMA(4,2) no se presenta la estimación de los parámetros ya que al introducir el orden del modelo en el ordenador, con ayuda del software SPSS nos presenta como resultado que el modelo no cumple las condiciones para dicha estimación. Por lo que descartamos este modelo en la investigación.

Es importante señalar que, las ecuaciones de los modelos ARMA(1,6), ARMA(1,7), ARMA(2,2), ARMA(4,1) se elaboraron tomando de cada una de las tablas de los modelos, la columna de los coeficientes, sin considerar los demás elementos, pero que, posteriormente en la fase del chequeo se analizarán las características principales de los coeficientes (Condición de significatividad de los coeficientes), para definir la ecuación que más se ajusta a la serie utilizando la columna de los P-valores.

En conclusión, se estimaron los parámetros para los modelos ARMA (1,6), ARMA (1,7), ARMA (2,2) y el modelo ARMA (4,1), el cálculo de su constante y la respectiva ecuación en diferencias. Ahora procederemos al análisis de la siguiente fase: chequeo.

3.3.3 CHEQUEO

Estimados los parámetros, se procede en esta etapa a la evaluación de la adecuación de los modelos previamente identificados y estimados. Esta etapa se centra en analizar los residuos del modelo (u_t), esto es, si puede afirmarse que son semejantes a un ruido blanco y verificar que los estimadores de los parámetros son estadísticamente diferentes de cero; las fases que se analizarán en la etapa de chequeo son las siguientes:

- 1) Las estimaciones de los parámetros de los posibles modelos sean significativas, condición de significatividad de los coeficientes, además debe cumplirse las condiciones de estacionariedad e invertibilidad.
- 2) Que los residuos (“estimaciones” de los ruidos) se comporten como ruidos blancos, esto es, tenga media cero, varianza constante, y que presente incorrelación.
- 3) El modelo sea estable. Diremos que un modelo es estable, cuando al dividir la serie en muestras de igual tamaño (por ejemplo, tomando la serie en dos mitades) y al estimar sus respectivos parámetros no muestren diferencias significativas con los parámetros del modelo en estudio. Esta situación ha sido verificada en la serie que se viene estudiando en este trabajo.
- 4) Que el modelo no omita parámetros relevantes (o sobreajuste).
Este proceso consiste en incluir parámetros adicionales en el modelo identificado, con el

fin de comprobar si éstos modelos menos parsimoniosos (falta de otro criterio, se prefieren modelos con el menor número de parámetros) resultan apropiados.

En general, una vez identificado el modelo ARIMA (p, d, q) que parece adecuado, el sobreajuste que se aplica es, estimando los modelos ARIMA (p + 1, d, q) y ARIMA (p, d, q + 1) y comprobando, si los parámetros adicionales son significativos. Obsérvese que se ha sugerido considerar modelos alternativos en los que se incrementa la parte autorregresiva o la de medias móviles. Por lo tanto, el numeral 4 ha sido incluido en la etapa de identificación del modelo.

A continuación verificaremos el comportamiento de estas condiciones enunciadas en los numerales 1 y 2, para cada uno de los modelos estimados.

En cuanto al **primer requisito** que se plantea, la condición de significatividad de los coeficientes² efectuamos el siguiente proceso. Para analizar si los coeficientes estimados son estadísticamente significativos, se debe efectuar el contraste de hipótesis respecto a los coeficientes del modelo. En general, se efectuará el contraste individual de cada coeficiente, adoptando como hipótesis nula la siguiente:

$$H_0: \beta_i = 0, \quad i = 1, 2, 3, \dots, k$$

y como alternativa

$$H_1: \beta_i \neq 0, \quad i = 1, 2, 3, \dots, k$$

El estadístico que se utiliza para el contraste es:

$$t = \frac{\hat{\beta}_i}{\hat{\sigma}_{\hat{\beta}_i}}$$

y fijando un nivel de significación ε , se concluye que el coeficiente es significativo cuando el valor absoluto del estadístico $|t|$ sea superior al valor de tabla $t_{\frac{\varepsilon}{2}}(T - k)$.

Una vez se ha descrito brevemente el contraste de significatividad, procedemos a verificarlo para cada uno de los modelos seleccionados.

² La condición de significatividad de los coeficientes se plantea en forma general y para los modelos posteriores solo se enunciará el análisis ya que el proceso es similar.

3.3.3.1 ANÁLISIS DE LA ETAPA DE CHEQUEO PARA EL MODELO ARMA (1, 6)

Efectuando el contraste para la parte autorregresiva del modelo ARMA (1,6) se tiene que:

$H_0 : \phi_1 = 0$ (El coeficiente de la parte autorregresiva de orden 1 no es estadísticamente significativo)

$H_1 : \phi_1 \neq 0$ (El coeficiente de la parte autorregresiva es estadísticamente significativo)

Ahora bien, al observar en la última columna de la tabla 3.2 se encuentra el P-valor de cada uno de los parámetros estimados los cuales pueden ser utilizados para realizar los contrastes de hipótesis que se requieren en esta etapa de chequeo.

Comparamos el P-valor de la parte autorregresiva ($\hat{\phi}_1 = 0.75$) con el nivel de significación fijado; $\varepsilon = 0.00001$. Concluimos, que el coeficiente es significativo, es decir, el P-valor es menor que ε , por lo que se acepta la hipótesis alternativa de que el parámetro del modelo es diferente de cero.

Aplicamos nuevamente, el contraste para la parte de medias móviles:

$$H_0 : \theta_i = 0, i = 1, 2, 3, 4, 5 \text{ y } 6$$

$$H_1 : \theta_i$$

Comparamos el P-valor de cada uno de los parámetros de la parte de medias móviles (Ver Tabla 3.2) con el nivel de significación fijado, $\varepsilon = 0.00001$ y podemos concluir que los coeficientes estimados $\theta_1 = -0.006$, $\theta_2 = 0.037$, $\theta_3 = 0.013$ y $\theta_4 = 0.13$, **no** son estadísticamente significativos, debido a que, sus P-valores son mayores que ε , por lo que se acepta la hipótesis nula, de que los parámetros del modelo en el análisis son iguales a cero.

Revisamos los P-valores correspondientes a los estimadores $\hat{\theta}_5 = 0.14$ y $\hat{\theta}_6 = 0.19$,

concluimos que son estadísticamente significativos, sus P-valores son menores que ε .

Una vez analizado la significatividad de los parámetros del modelo ARMA(1,6), su ecuación en diferencias y su respectiva constante estimada $\hat{\delta}$ viene dada por:

$$y_t = 0.75y_{t-1} + 3.67 + u_t - 0.14u_{t-5} - 0.19u_{t-6}$$

Revisamos la condición de estacionariedad e invertibilidad del modelo ARMA (1,6)

$|\phi_1| < 1 \Rightarrow \{ |0.75| < 1$. Por tanto, la serie es estacionaria.

$\theta_5 + \theta_6 < 1 \Rightarrow \{ 0.12 + 0.17 = 0.29 < 1$, satisface la condición de invertibilidad.

Verificamos el **segundo requisito** para el modelo ARMA (1,6), si los residuos (“estimaciones” de los ruidos) se comportan como ruidos blancos, esto es que tenga

a) media cero,

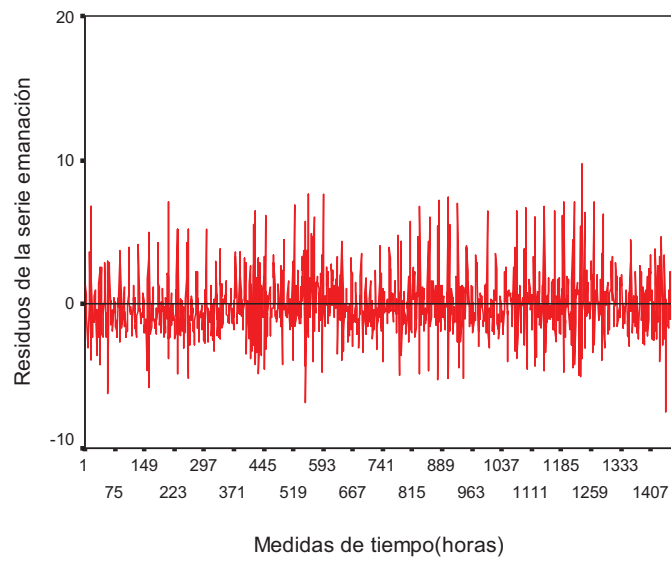
b) Varianza Constante

c) que presente incorrelación y

d) distribución normal.

a) Para analizar si la media de los residuos es cero, podemos comenzar representando gráficamente la serie de los residuos del modelo ARMA (1, 6). Esta representación resulta interesante, dado que del examen de la misma podremos observar si la serie de los residuos se comporta como un ruido blanco. En concreto, si los valores de los residuos giran alrededor de cero, que será la evidencia respecto del cumplimiento de esta condición.

Figura 3.4. Gráfico de los residuos de la serie emanación gas CO₂,
Modelo ARMA (1,6)



De la figura 3.4, se observa que los valores de cada uno de los residuos giran alrededor de cero, por lo que puede conjeturarse de que $\mu = 0$.

a) Para comprobar que $\mu_{u_t} = 0$, efectuaremos el contraste de significatividad de la media de los residuos³, utilizando el siguiente estadístico:

$$t = \frac{\bar{u}_t}{\hat{\sigma}_{\bar{u}_t}}$$

Donde \bar{u}_t representa la media de los residuos, definido por:

$$\bar{u}_t = \frac{\sum_{t=1}^T u_t}{T}$$

y aproximando $\hat{\sigma}_{\bar{u}_t}$ (Wei, 1990) mediante la expresión

$$\hat{\sigma}_{\bar{u}_t} = \left(\frac{C_0}{T} \right)^{1/2}$$

³ El contraste de significatividad de la media de los residuos se plantea en forma general y para los modelos posteriores solo se enunciará el análisis ya que el proceso es similar.

donde C_0 es la varianza muestral, que viene dada por

$$C_0 = \frac{\sum_{t=1}^T (u_t - \bar{u}_t)^2}{T}$$

El estadístico t , $t = \frac{\bar{u}_t}{\hat{\sigma}_{\bar{u}_t}}$, se distribuye bajo la hipótesis nula como una t de Student con $T - 1$ grados de libertad. Ahora bien, dado que el número de observaciones (T) necesarias para aplicar la metodología de Box y Jenkins debe ser elevado (como mínimo se requiere que $T \cong 50$), y dado que la distribución t de Student tiende a una distribución normal a medida que aumenta el tamaño de la serie. Así pues, para aceptar o rechazar el contraste de significatividad de la media de los residuos adoptaremos la siguiente regla de decisión: si $|t| < 2$ concluiremos que podemos aceptar la hipótesis que la media de los residuos no difiere significativamente de cero, en caso de que $|t| > 2$, lo rechazaremos.

Retomando el modelo ARMA (1,6) para verificar que $\mu_{u_t} = 0$, se tiene los siguientes resultados

$$\begin{aligned} \bar{u}_t &= -0.00028 & C_0 &= \frac{5851,21899}{1463} = 4 \\ \hat{\sigma}_{\bar{u}_t} &= \left(\frac{4}{1463} \right)^{1/2} = 0.052 & t &= \frac{\bar{u}_t}{\hat{\sigma}_{\bar{u}_t}} = \frac{-0.00028}{0.052} = -0.0054 \end{aligned}$$

Obsérvese que $|t| < 2$, por lo que podemos aceptar que la media de los residuos no difiere significativamente de cero. Por lo tanto, la media de los residuos del modelo ARMA (1,6) es cero.

b) Para verificar que la varianza es constante se puede utilizar el gráfico de los residuos; observaremos la dispersión de los residuos y si éstos se mantienen constante a lo largo del tiempo.

De la Figura 3.4, podemos concluir que se observa poca dispersión y que se mantiene aproximadamente constante a lo largo del tiempo; por lo que se acepta la homocedasticidad del modelo ARMA (1,6).

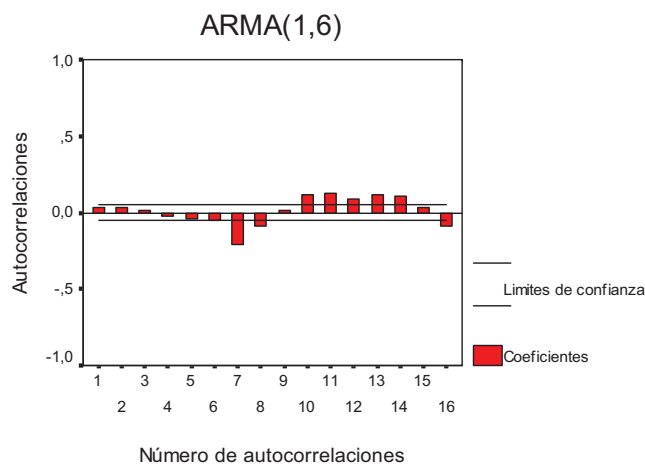
c) Revisamos el comportamiento de las autocorrelaciones de los residuos, para observar si presenta incorrelación, es decir, si las autocorrelaciones muestrales de los residuos no son significativos o sea tienen valores que son muy pequeños; considerando un nivel de significación de $\varepsilon = 0.005$, que podremos aceptar la no significatividad de cada $\gamma_j(u_t)$, siempre que el valor este dentro del intervalo de $\pm Z_{\varepsilon/2}(1/\sqrt{T})$.

También, obtendremos las autocorrelaciones parciales muestrales de los residuos $\phi_{jj}(u_t)$, debiéndose cumplir también que estos coeficientes no sean significativos, es decir, que se encuentren (al menos el 95% de ellos, caso de que $\varepsilon = 0.005$) dentro del intervalo $\pm 1.96(1/\sqrt{T})$.

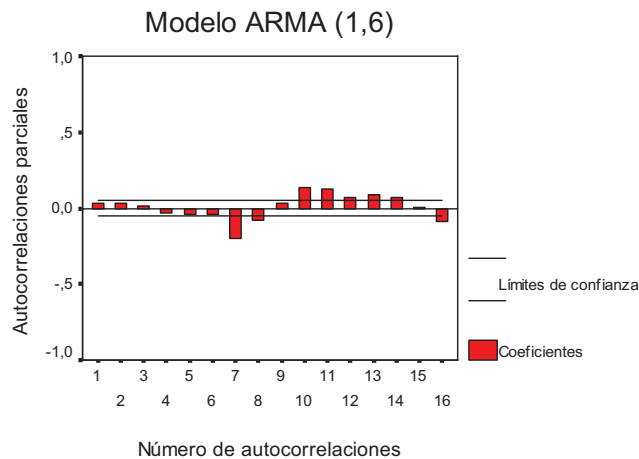
Los gráficos correspondientes a las autocorrelaciones muestrales y autocorrelaciones parciales muestrales de los residuos del modelo ARMA (1,6) se presentan a continuación.

Figura 3.5. Autocorrelaciones de los residuos, ARMA (1, 6)

a) Autocorrelaciones muestrales de los residuos



b) Autocorrelaciones parciales de los residuos



Del gráfico de las autocorrelaciones muestrales simples y parciales (Figura 3.5.a, y Figura 3.5.b), se observa que la autocorrelación siete (en la parte negativa de ambos gráficos) es significativa, no se ajusta a la banda de significación que se propone de acuerdo al total de observaciones de la serie, es decir, $\pm 1.96 \left(\frac{1}{\sqrt{1463}} \right) = \pm 0.05$.

También, en las demás autocorrelaciones se observa que algunas de ellas no se encuentran dentro de la banda de significación. Por no satisfacerse esta condición descartamos el modelo ARMA (1,6).

3.3.3.2 ANÁLISIS DE LA ETAPA DE CHEQUEO PARA EL MODELO ARMA (1, 7)

Revisamos el *primer requisito* la condición de Significatividad de los coeficientes.

Efectuamos el contraste para la parte autorregresiva del modelo ARMA (1,7):

$$H_0 : \phi_1 = 0$$

$$H_1 : \phi_1 \neq 0$$

Ahora bien, al observar en la última columna de la tabla 3.3, nuevamente, revisamos el P-valor de cada uno de los parámetros estimados para realizar los contrastes de hipótesis que se requieren en esta etapa de chequeo.

Comparamos el P-valor de la parte autorregresiva ($\hat{\phi}_1 = 0.65$) con el nivel de significación $\varepsilon = 0.00001$, concluimos así, que el coeficiente es significativo, es decir, el P-valor es menor a ε , por lo que se acepta la hipótesis alternativa de que el parámetro del modelo es diferente de cero.

Aplicamos nuevamente, el contraste para la parte de medias móviles:

$$H_0 : \theta_i = 0, \quad i = 1, 2, 3, 4, 5, 6 \text{ y } 7$$

$$H_1 : \theta_i \neq 0$$

Nuevamente, comparamos el P-valor de cada uno de los parámetros de la parte de medias móviles (Ver Tabla 3.3) con el nivel de significación fijado, $\varepsilon = 0.00001$ y concluimos que los coeficientes estimados $\hat{\theta}_1 = -0.10$, $\hat{\theta}_2 = -0.05$, $\hat{\theta}_3 = -0.063$ y $\hat{\theta}_4 = 0.009$, **no** son estadísticamente significativos, debido a que, sus P-valores son mayores a ε , por lo que se acepta la hipótesis nula de que los parámetros del modelo en el análisis son iguales a cero.

Revisamos los P-valores correspondientes a los estimadores $\hat{\theta}_5 = 0.12$, $\hat{\theta}_6 = 0.17$ y $\hat{\theta}_7 = 0.16$, concluimos que son estadísticamente significativos sus P-valores son menores a ε .

Analizado la significatividad de los parámetros del modelo ARMA (1,7), su ecuación en diferencias y su respectiva constante estimada $\hat{\delta}$ viene dada por:

$$y_t = 0.65y_{t-1} + 5.2 + u_t - 0.12u_{t-5} - 0.17u_{t-6} - 0.16u_{t-7}$$

Continuamos con la revisión de la condición de estacionariedad e invertibilidad del modelo ARMA (1,7).

$$|\phi_1| < 1 \Rightarrow \{ |0.65| < 1. \text{ Por tanto, la serie es estacionaria.}$$

$$\theta_5 + \theta_6 + \theta_7 < 1 \Rightarrow \{ 0.12 + 0.17 + 0.16 = 0.48 < 1, \text{ satisface la condición de}$$

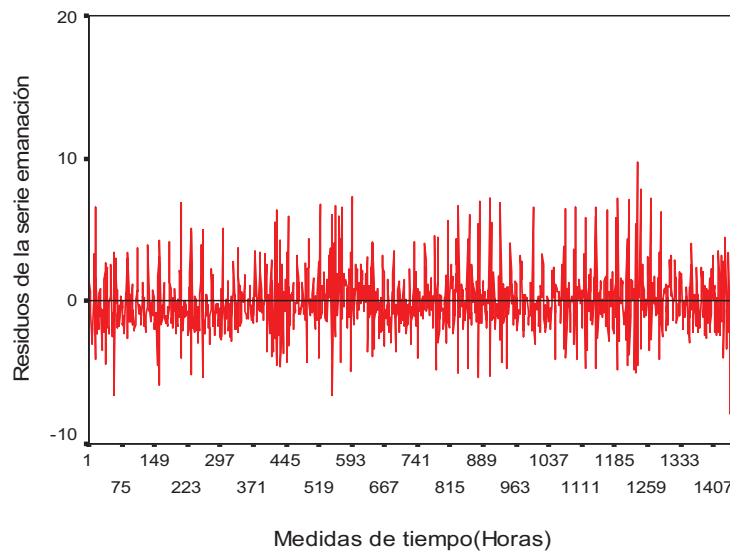
invertibilidad.

Verificamos el *segundo requisito* para el modelo ARMA (1,7), si los residuos (“estimaciones”) se comportan como ruidos blancos, esto es que tenga:

a) Media cero.

De manera similar al modelo anterior, representamos gráficamente la serie de los residuos del modelo ARMA (1,7). Analizaremos, si la serie de los residuos se comporta como un ruido blanco.

Figura 3.6. Gráfico de los residuos de la serie emanación gas CO₂,
Modelo ARMA (1,7)



Se observa de la Figura 3.6 que los valores de cada uno de los residuos giran alrededor de cero, por lo que puede conjeturarse de que $\mu=0$.

Para comprobar que $\mu_{u_t} = 0$ efectuaremos, nuevamente, el contraste de significatividad de la media de los residuos tal como se aplicó en el modelo anterior.

Retomando el modelo ARMA (1,7) para verificar que $\mu_{u_t} = 0$, se tienen los siguientes resultados

$$\bar{u}_t = -0,00031834 \quad Co = \frac{5693,84296}{1463} = 3.89$$

$$\hat{\sigma}_{\bar{u}_t} = \left(\frac{3.89}{1463}\right)^{1/2} = 0.051 \quad t = \frac{-0.00031834}{0.051} = -0.0062$$

Obsérvese que $|t| < 2$, por lo que podemos aceptar el contraste de que la media no difiere significativamente de cero. Por lo tanto, la media de los residuos del modelo ARMA (1,7) es cero.

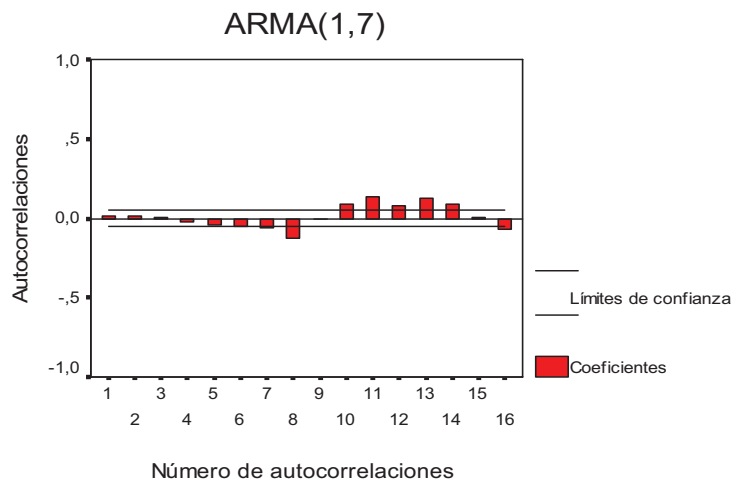
b) Para verificar que la varianza es constante se puede utilizar, de igual forma que el modelo anterior, examinando el gráfico de los residuos. De la Figura 3.6, podemos concluir que se observa poca dispersión y que se mantiene aproximadamente constante a lo largo del tiempo; por lo que se acepta la homocedasticidad del modelo ARMA (1,7).

c) Revisamos el comportamiento de las autocorrelaciones de los residuos, para observar si presenta incorrelación, es decir, si las autocorrelaciones muestrales y parciales de los residuos no son significativos o sea tiene valores que son muy pequeños; considerando un nivel de significación de $\varepsilon = 0.005$, que podremos aceptar la no significatividad siempre que el valor este dentro del intervalo de $\pm Z_{\varepsilon/2}(1/\sqrt{T})$.

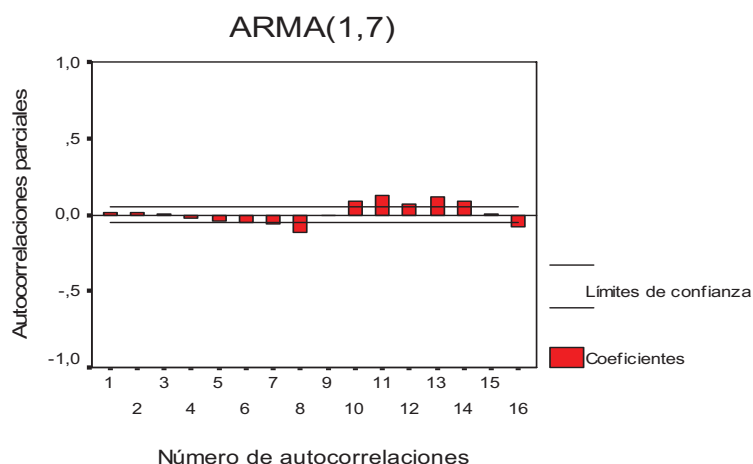
Se presentan a continuación los gráficos correspondientes a las autocorrelaciones muestrales y autocorrelaciones parciales de los residuos del modelo ARMA (1,7):

Figura 3.7. Autocorrelaciones de los residuos, ARMA (1, 7)

a) Autocorrelaciones muestrales de los residuos



b) Autocorrelaciones parciales de los residuos



Del gráfico de las autocorrelaciones muestrales simples y parciales (Figura 3.7.a, y Figura 3.7.b), se observa que la autocorrelaciones en ambos gráficos, algunas son significativas (por ejemplo, la ocho, once, trece), no se ajustan a la banda de significación que se propone, $\pm 1.96 \left(\frac{1}{\sqrt{1463}} \right) = \pm 0.05$. También, en las demás autocorrelaciones se observa que algunas de ellas no se encuentran dentro de la banda de significación. Por lo tanto, por no satisfacerse esta condición descartamos el modelo ARMA (1,7).

3.3.3.3 ANÁLISIS DE LA ETAPA DE CHEQUEO PARA EL MODELO ARMA (2, 2)

Revisamos los requisitos que debe satisfacer un modelo ARIMA para afirmar sí el modelo ARMA (2, 2) es el adecuado.

Aplicamos el análisis para el **primer requisito**, la significatividad de los coeficientes del modelo ARMA (2,2).

Efectuamos el contraste para la parte autorregresiva del modelo ARMA (2,2):

$$H_0 : \phi_i = 0, \quad i = 1, 2.$$

$$H_i : \phi_i \neq 0$$

Al observar en la última columna de la tabla 3.4, nuevamente, revisamos el P-valor de

cada uno de los parámetros estimados del modelo ARMA (2,2), que utilizaremos para realizar los contrastes de hipótesis que se requieren en la etapa de chequeo.

Comparamos el P-valor de la parte autorregresiva ($\hat{\phi}_1 = 1.61$ y $\hat{\phi}_2 = -0.83$) con el nivel de significación $\varepsilon = 0.00001$, concluimos así, que los coeficientes son significativos, es decir, el P-valor es menor a ε , por lo que se acepta la hipótesis alternativa de que el parámetro del modelo es diferente de cero.

Aplicamos nuevamente, el contraste para la parte de medias móviles:

$$H_0 : \theta_i = 0, \quad i = 1, 2$$

$$H_1 : \theta_i \neq 0$$

Comparamos el P-valor de cada uno de los parámetros de la parte de medias móviles (Ver Tabla 3.4) con el nivel de significación fijado, $\varepsilon = 0.00001$ y concluimos que los coeficientes estimados $\theta_1 = 0.94$, $\theta_2 = -0.33$ son estadísticamente significativos.

Una vez analizado la significatividad de los coeficientes del modelo ARMA (2,2), su ecuación en diferencias y su respectiva constante estimada $\hat{\delta}$ viene dada por:

$$y_t = 1.61y_{t-1} - 0.83y_{t-2} + 3.29 + u_t - 0.94u_{t-1} + 0.33u_{t-2}$$

Revisamos la condición de estacionariedad e invertibilidad para el modelo ARMA (2,2)

Tabla 3.6 Coeficientes del modelo ARMA (2,2)

Coeficientes de la parte autorregresivo	Coeficientes de la parte de medias móviles
$\phi_1 = 1.61$	$\theta_1 = 0.94$
$\phi_2 = -0.83$	$\theta_2 = -0.33$

Tenemos que:

$$\left. \begin{array}{l} \phi_1 + \phi_2 < 1 \\ \phi_2 - \phi_1 < 1 \\ |\phi_2| < 1 \end{array} \right\} \Rightarrow \left\{ \begin{array}{l} 1.61 + (-0.83) = 0.78 < 1 \\ -0.83 - 1.61 = -2.44 < 1 \\ |-0.83| = 0.83 < 1 \end{array} \right.$$

Por tanto, el modelo ARMA (2,2) satisface la estacionariedad.

Además, para que sea invertible debe cumplirse que

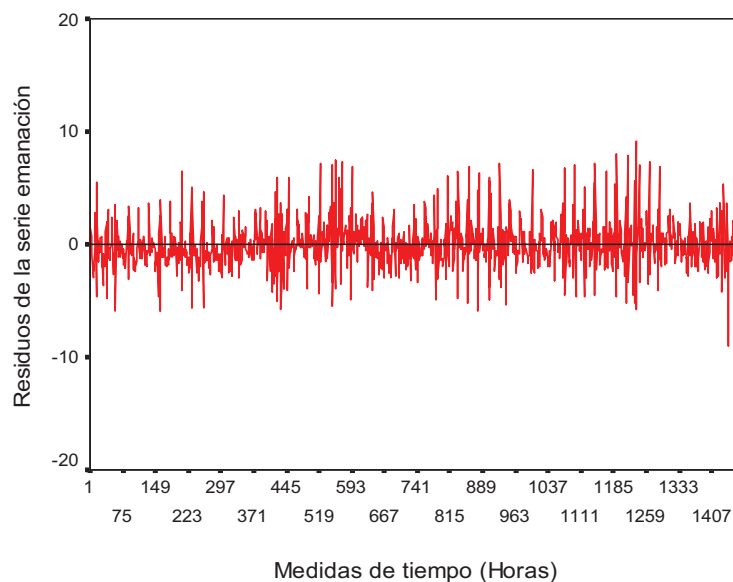
$$\left. \begin{array}{l} \theta_1 + \theta_2 < 1 \\ \theta_2 - \theta_1 < 1 \\ |\theta_2| < 1 \end{array} \right\} \Rightarrow \left\{ \begin{array}{l} 0.94 + (-0.33) = 0.61 < 1 \\ -0.33 - 0.94 = -1.27 < 1 \\ |0.94| = 0.94 < 1 \end{array} \right.$$

Por tanto, satisface la condición de invertibilidad.

Verificamos el *segundo requisito* para el modelo ARMA (2,2), si los residuos (“estimaciones” de los ruidos) se comportan como ruidos blancos, esto es que tenga media cero, varianza constante, incorrelación y una distribución normal.

a) Para analizar si la media de los residuos es cero, representamos el gráfico de la serie de los residuos del modelo ARMA(2, 2). Nuevamente, esta representación es interesante, dado que del análisis de la misma se observará si la serie de los residuos se comporta como un ruido blanco.

Figura 3.8. Gráfico de los residuos de la serie emanación gas CO₂,
Modelo ARMA (2,2)



De la figura 3.8, se observa que los valores de cada uno de los residuos giran alrededor de cero, por lo que puede conjeturarse de que se trata de un ruido blanco, $\mu=0$.

Comprobaremos que $\mu_{u_t} = 0$, efectuando el contraste de significatividad de la media de los residuos, tal como se aplico en el modelo anterior.

Verificamos dicho contraste para el modelo ARMA(2,2), en la que se tiene los resultados siguientes:

$$\bar{u}_t = 0.00071773, \quad \sum_{t=1}^{1463} (u_t - 0.00071773)^2 = 5505.97527$$

$$C_0 = \frac{5505.97527}{1463} = 3.7635, \quad \hat{\sigma}_{\bar{u}_t} = \left(\frac{C_0}{T} \right)^{1/2} = \left(\frac{3.7635}{1463} \right)^{1/2} = 0.05072$$

$$t = \frac{\bar{u}_t}{\hat{\sigma}_{\bar{u}_t}} = \frac{0.00071773}{0.05072} = 0.01415$$

Obsérvese que $|t| < 2$, por lo que podemos aceptar que la media de los residuos no difiere significativamente de cero. Por lo tanto, la media de los residuos del modelo ARMA (2,2) es cero.

b) Para verificar que la varianza es constante, de igual forma que en los modelos anteriores, examinamos el gráfico de los residuos.

De la Figura 3.8, podemos concluir que se observa poca dispersión y que se mantiene constante a lo largo del tiempo; por lo que se acepta la homocedasticidad del modelo ARMA (2,2).

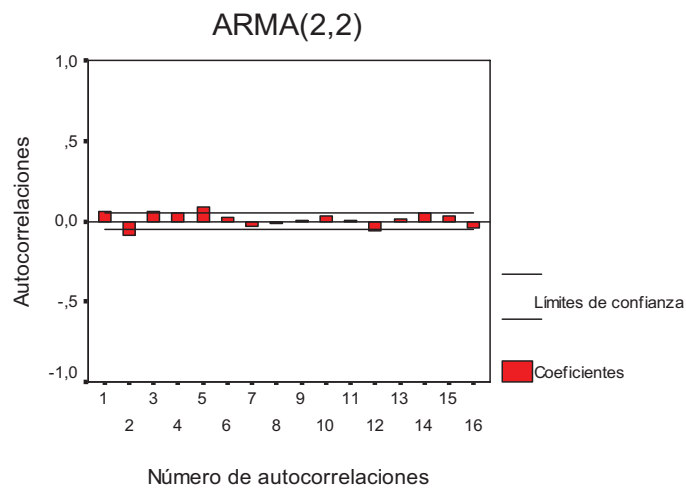
c) Revisamos el comportamiento de las autocorrelaciones de los residuos, para observar si presenta incorrelación, es decir, si los residuos de los coeficientes de las autocorrelaciones muestrales no son significativos o sea tiene valores que son muy pequeños, considerando un nivel de significación de $\varepsilon = 0.005$, que podremos aceptar la no significatividad de cada $\gamma_j(u_t)$, siempre que el valor este dentro del intervalo de $\pm 1.96(1/\sqrt{T})$.

También, obtendremos las autocorrelaciones parciales muestrales de los residuos $\phi_{jj}(u_t)$, debiéndose cumplir también que estos coeficientes no sean significativos, es decir, que se encuentren (al menos el 95% de ellos, caso de que $\varepsilon = 0.005$) dentro del intervalo $\pm 1.96(1/\sqrt{T})$

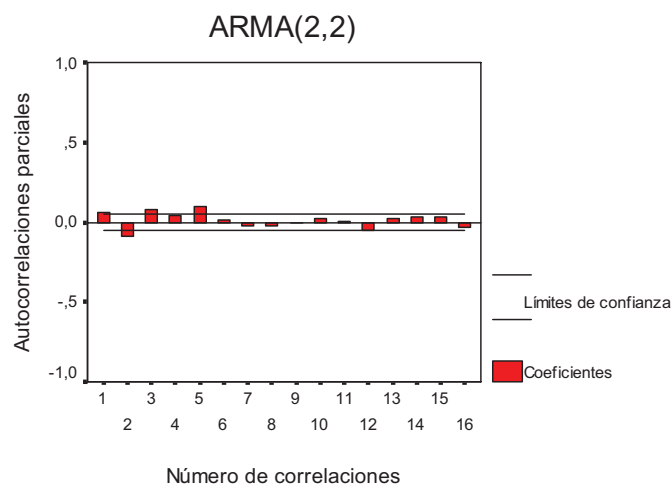
Los gráficos correspondientes a las autocorrelaciones muestrales y autocorrelaciones parciales muestrales de los residuos del modelo ARMA (2,2) se presentan a continuación

Figura 3.9. Autocorrelaciones de los residuos, ARMA (2,2)

a) Autocorrelaciones muestrales de los residuos



b) Autocorrelaciones parciales de los residuos

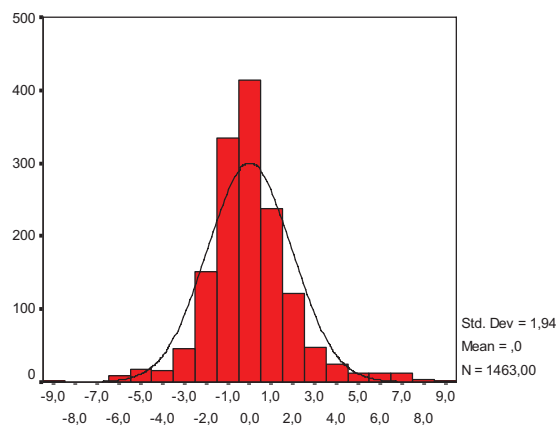


Se evidencia que las autocorrelaciones muestrales y parciales (ver Figura 3.9.a y Figura 3.9.b) obedecen aproximadamente al intervalo de aceptación propuesto; $\pm 1.96(1/\sqrt{T})$.

Por lo que se acepta que las autocorrelaciones de los residuos no son significativas, es decir, son muy próximos a cero considerando un nivel de significación de $\varepsilon = 0.005$.

d) Procedemos a observar si el comportamiento de los residuos del modelo ARMA (2,2) se distribuyen normalmente. Se presenta a continuación el histograma de frecuencia de los residuos:

Figura 3.10 Histograma de los residuos, serie Emanación de Gas CO₂



Histograma y Curva Normal de los residuos , Modelo ARMA(2, 2)

En el histograma de la Figura 3.10 se puede observar:

- Que efectivamente, la concentración de la mayor parte de los residuos (errores) se acumula en la parte central, y se observa que la media de los residuos es cero, $\mu = 0$.
- La distribución es aproximadamente simétrica.

La distribución de los residuos parece seguir aproximadamente el modelo de probabilidad normal, pero los resultados del análisis deben ser interpretados con cautela.

También, podemos comprobar que los residuos se distribuyen como una normal por medio del coeficiente de asimetría y curtosis, que se define cada uno de ellos de la forma siguiente:

$$g_1 = \frac{m_3}{m^{3/2}} = \frac{\frac{\sum_{t=1}^T (\hat{u}_t - \bar{\hat{u}})^3}{T}}{\left(\frac{\sum_{t=1}^T (\hat{u}_t - \bar{\hat{u}})^2}{T} \right)^{3/2}} ; \quad g_2 = \frac{m_4}{m^2} - 3 = \frac{\frac{\sum_{t=1}^T (u_t - \bar{u}_t)^4}{T}}{\left(\frac{\sum_{t=1}^T (u_t - \bar{u}_t)^2}{T} \right)^2} - 3$$

A continuación se presentan en la tabla 3.7 esta información:

Tabla 3.7 Estadísticos Descriptivos de los errores⁴. Modelo ARMA (2,2)

Media	Desviación Estándar	Varianza	Asimetría
7,176795E-04	1,9406331	3,766	0,625

Se observa de la tabla 3.7 que el valor de la asimetría es un valor cerca de cero por lo que da lugar a decir que los residuos se comportan de manera normal.

f) Finalmente, revisamos el *tercer requisito*. El análisis de estabilidad del modelo, el cuál consiste en dividir la serie, y_t , de la emanación de gas CO₂ en dos muestras distintas (o bien puede dividirse en más partes iguales distintas) y estimar los parámetros, compararse con los calculados en los modelos identificados de la serie y así concluir, que el modelo es estable cuando sus estimaciones no sean muy diferentes a los estimados en la totalidad de la serie.

A continuación se presenta la estimación de los parámetros de las dos muestras, y luego se efectúa su comparación.

Tabla 3.8 MUESTRA 1 .PARÁMETROS DEL MODELO ARMA (2, 2)

Variables del Modelo				
Modelo	Coefficientes	Desviación Estándar de los Coeficientes	Prueba (T-RATIO)	P-valor
AR1	1,586411	0,04695640	33,78477	0,000000
AR2	0,790933	0,03858529	-20,49831	0,000000
MA1	0,901964	0,06748023	13,36634	0,000000
MA2	-0,221774	0,06211926	-3,570130	0,003943
CONSTANTE	15,191318	0,11030268	137,72393	0,000000

⁴ Se utilizo el Software SPSS para la estimación de los estadísticos

Tabla 3.9 MUESTRA 2 .PARÁMETROS DEL MODELO ARMA (2, 2)

Variables del Modelo				
Modelo	Coefficientes	Desviación Estándar de los Coeficientes	Prueba (T-RATIO)	P-valor
AR1	1,624110	0,02660252	61,05097	0,000000
AR2	-0,871122	0,02380193	-36,59879	0,000000
MA1	1,035121	0,03884129	26,65001	0,000000
MA2	-0,550152	0,03522935	-15,61629	0,000000
CONSTANTE	14,684119	0,14192227	103,46592	0,000000

Comparando la tabla 3.8 y la tabla 3.9 puede observarse que los parámetros de cada uno de los periodos no son muy diferentes comparados con los parámetros de la serie, ARMA (2, 2), en la tabla 3.3, por lo que se concluye que dicho modelo es estable.

Calculamos sus diferencias de los parámetros del modelo ARMA(2,2) con las dos muestras estudiadas obteniendo los siguientes resultados:

Tabla 3.10 Diferencias de la serie emanación CO₂ con las muestras obtenidas.

MODELO	Coefficientes Modelo ARMA(2,2)	Coefficientes muestra 1	Coefficientes muestra 2	Diferencias muestra 1	Diferencias muestra 2
AR1	1,611472	1,586411	1,62411	0,025061	-0,012638
AR2	-0,83189	0,790933	-0,871122	-1,622822	0,039233
MA1	0,94035	0,901964	1,035121	0,038386	-0,094771
MA2	-0,33005	-0,221774	-0,550152	-0,108280	0,220098

Puede observarse en la tabla 3.10 en las columnas de las diferencias, que al compararse los coeficientes las diferencias son pequeñas, por lo que diremos que el modelo ARMA(2,2) es el que más se aproxima a la serie emanación CO₂, por lo que esta situación nos lleva a continuar su estudio en la fase de predicción.

3.3.3.4 ANÁLISIS DE LA ETAPA DE CHEQUEO PARA EL MODELO ARMA (4,1)

Revisamos el *primer requisito* la condición de Significatividad de los coeficientes.

Efectuamos el contraste para la parte autorregresiva del modelo ARMA (4,1):

$$H_0 : \phi_i = 0 , i = 1, 2, 3, 4$$

$$H_1 : \phi_i \neq 0$$

Ahora bien, al observar en la última columna de la tabla 3.5, nuevamente, revisamos el P-valor de cada uno de los parámetros estimados para realizar los contrastes de hipótesis requeridas en la etapa de chequeo.

Comparamos el P-valor de la parte autorregresiva ($\hat{\phi}_1 = 1.096$, $\hat{\phi}_2 = -0.27$, $\hat{\phi}_3 = 0.02$ y $\hat{\phi}_4 = -0.20$) con el nivel de significación $\varepsilon = 0.0001$, concluimos así, que el coeficiente estimado, $\hat{\phi}_3$ **no** es estadísticamente significativo, debido a que su P-valor es mayor a ε , por lo que se acepta la hipótesis nula. Por lo tanto, el modelo ARMA(4,2) se descarta en esta fase.

En conclusión, el modelo ARMA (2,2) cuya ecuación en diferencias es:

$$y_t = 1.61y_{t-1} - 0.83y_{t-2} + 3.29 + u_t - 0.94u_{t-1} + 0.33u_{t-2}$$

supera la fase del chequeo, por lo que este modelo puede ser utilizado para obtener las predicciones.

3.3.4 PREDICCIÓN

Esta es, la última fase y la primordial del análisis de la serie temporal Emanación del gas CO₂ del volcán de San Vicente, según el esquema seguido y sugerido por Box- Jenkins. En este apartado estamos interesados en predecir los valores futuros de la serie en estudio utilizando el modelo ARMA(2,2), sujeto al menor error posible.

3.3.4.1 El predictor óptimo

En general, el objetivo principal del análisis de series temporales es predecir valores futuros de $\{y_t\}$.

En este estudio se conocen las observaciones hasta el periodo T, se desea obtener la predicción de y_{T+1} . Entonces, al conjunto de observaciones disponibles hasta T lo

designaremos por $I_T = (y_0, y_1, \dots, y_T, y_{T-2}, \dots, y_{T-1}, \dots)$, se denota a $\hat{y}_T(l)$ el predictor óptimo para y_{T+l} , que utiliza toda la información disponible hasta T de I_T .

Así, el predictor óptimo⁵ será aquel que minimice el error cuadrático medio (ECM) de predicción, que se define como:

$$E[(y_{T+l} - \hat{y}_T(l))^2 / I_T]$$

Consecuentemente, $\hat{y}_T(l)$ será óptimo si cumple

$$E[(y_{T+l} - \hat{y}_T(l))^2 / I_T] \leq E[(y_{T+l} - \hat{y}_T^*(l))^2 / I_T]$$

donde $\hat{y}_T^*(l)$ es cualquier predictor óptimo.

El predictor $\hat{y}_T(l)$ que cumple la desigualdad anterior, está definido por:

$$\hat{y}_T(l) = E(y_{T+l} / I_T)$$

El ECM el cual se define de la forma siguiente:

$$ECM = \frac{1}{T-1} \sum_{t=1}^{T-1} e_t^2(l) = \frac{1}{T-1} \sum_{t=1}^{T-1} (y_{T+1} - \hat{y}_T(l))^2$$

En nuestro estudio no fue necesario calcular el error cuadrático medio ya que solo un modelo fue el que superó las fases de la metodología de Box-Jenkins, el modelo ARMA(2,2).

⁵ Pág. 258. Antonio Aznar, Javier Trivez. Métodos de predicción en economía (1).

3.3.4.2 Cálculo de la predicción puntual

El punto de partida del modelo ARIMA (p, q) podría expresarse de la forma siguiente:

$$y_t = \varphi_1 y_{t-1} + \varphi_2 y_{t-2} + \dots + \varphi_{p+d} y_{t-p-d} + \delta + u_t - \theta_1 u_{t-1} - \theta_2 u_{t-2} - \dots - \theta_l u_{t-q}$$

Para obtener $\hat{y}_T(l)$, del modelo ARMA(p,q) debemos considerar los supuestos siguientes:

- Los parámetros son conocidos.
- Las perturbaciones (ruidos) se conocen para los periodos muestrales, pero tiene el carácter de ruidos blancos para periodos posteriores a los incluidos en la base informativa, es decir

$$E(u_{T+j} / I_T) \begin{cases} u_{T+j} & \text{para } j \leq 0 \\ E(u_{T+j}) = 0 & \text{para } j > 0 \end{cases}$$

Teniendo en cuenta estos supuestos, $\hat{y}_T(l)$ será igual a:

$$\hat{y}_T(l) = \varphi_1 \hat{y}_T(l-1) + \varphi_2 \hat{y}_T(l-2) + \dots + \varphi_l \hat{y}_{l-1}(1) + \varphi_l y_T + \varphi_{l+1} y_{T-1} + \dots + \varphi_{p+d} y_{T+l-p-d} + \delta - \theta_l u_T - \theta_{l+1} u_{T-1} - \dots - \theta_q u_{T+l-q}$$

Con esta ecuación se puede obtener $\hat{y}_T(l)$ las predicciones desde la base informativa disponible (T para nuestro caso), para un periodo, dos periodos, ..., (h - 1) periodos hacia delante.

Los valores para calcular los $u_T, u_{T-1}, u_{T-2}, \dots, u_{T+l-q}$, se obtienen a partir de las perturbaciones referidas a periodos de tiempos incluidos en la bases de datos de la sucesión de los errores de predicción un período hacia adelante, es decir:

$$u_{T+j} = y_{T+j} - \hat{y}_{T+j-l}(1) \quad \text{para } j \leq 0$$

Así, para calcular $\hat{y}_T(l)$, podemos escribir:

$$\begin{aligned} u_T &= y_T - \hat{y}_{T-1}(1) \\ u_{T-1} &= y_{T-1} - \hat{y}_{T-2}(1) \\ &\dots\dots \\ u_{T+l-q} &= y_{T+l-q} - \hat{y}_{T-l-q-1}(1) \end{aligned}$$

Para calcular cada uno de estos valores, debe haberse calculado previamente, las predicciones siguientes:

$$\hat{y}_{T+l-q-1}(1), \dots, \hat{y}_{T-2}(1), \hat{y}_{T-1}(1), \hat{y}_T(1), \hat{y}_T(2), \dots, \hat{y}_T(l-1)$$

Después de describir brevemente la forma de obtener las predicciones para períodos futuros podemos obtener la predicción puntual para $\hat{y}_{1464}, \hat{y}_{1465}, \hat{y}_{1466}, \hat{y}_{1467}$ con el modelo seleccionado; ARMA (2,2).

Iniciamos el proceso calculando las predicciones puntuales propuestas

$$\begin{aligned} \hat{y}_{1463}(1) &= E(y_{1464} / I_{1463}) \\ &= E[(3.229 + 1.61y_{1463} - 0.83y_{1462} + u_{1464} - 0.94u_{1463} + 0.33u_{1462}) / I_{1463}] \end{aligned}$$

$$\hat{y}_{1463}(1) = 3.29 + 1.61y_{1463} - 0.83y_{1462} - 0.94[y_{1463} - \hat{y}_{1462}(1)] + 0.33[y_{1462} - \hat{y}_{1461}(1)]$$

A continuación en la tabla 3.11 se presenta las predicciones de los valores $\dots, \hat{y}_{1459}, \hat{y}_{1460}, \hat{y}_{1461}, \hat{y}_{1462}, \dots$

Tabla 3.11 Serie Emanación CO₂ (y_t) Serie Ajustada, (\hat{y}_t)

Tiempo de Medida (Horas) t	Serie Emanación es CO ₂ y_t	Predicciones
...
1455	15,473	16,46336
1456	17,890	15,74803
1457	16,311	16,90908
1458	16,626	15,96433
1459	13,606	15,69653
1460	11,294	13,57165
1461	10,907	11,62582
1462	11,500	11,39791
1463	11,422	12,41800
...

Ahora se obtendrán las predicciones para los valores $\dots, \hat{y}_{1463}, \hat{y}_{1464}, \hat{y}_{1465}, \hat{y}_{1466}, \dots$

Sustituyendo los valores para $\hat{y}_{1464}(1)$, utilizando la tabla y la ecuación, tenemos:

$$\begin{aligned}\hat{y}_{1463}(1) &= 3.29 + 1.61(11.42) - 0.83(11.50) - 0.94[11.42 - 12.418] + 0.33[11.50 - 11.39791] \\ &= 13.103\end{aligned}$$

Calculamos el valor de la serie $\hat{y}_{1463}(2)$ y se obtiene:

$$\begin{aligned}\hat{y}_{1463}(2) &= E[y_{1465} / I_{1463}] \\ &= E[(3.29 + 1.61y_{1464} - 0.83y_{1463} + u_{1465} - 0.94u_{1464} + 0.33u_{1463}) / I_{1463}] \\ &= 3.29 + 1.61\hat{y}_{1463}(1) - 0.83y_{1463} + 0.33u_{1463} \\ &= 3.29 + 1.61(13.1030) - 0.83(11.422) + 0.33(11.42 - 12.418) \\ &= 14.57\end{aligned}$$

La predicción puntual de $\hat{y}_{1463}(3)$ es:

$$\begin{aligned}
\hat{y}_{1463}(3) &= E[y_{1466} / I_{1463}] \\
&= E[(3.29 + 1.61y_{1465} - 0.83y_{1464} + u_{1466} - 0.94u_{1465} + 0.33u_{1464}) / I_{1463}] \\
&= 3.29 + 1.61 \hat{y}_{1465}(2) - 0.83 \hat{y}_{1464}(1) \\
&= 3.29 + 1.61(14.57) - 0.83(13.103) \\
&= 15.87
\end{aligned}$$

La predicción puntual del valor $\hat{y}_{1463}(4)$:

$$\begin{aligned}
\hat{y}_{1463}(4) &= E[y_{1467} / I_{1463}] \\
&= E[(3.29 + 1.61y_{1466} - 0.83y_{1465} + u_{1467} - 0.94u_{1466} + 0.33u_{1465}) / I_{1463}] \\
&= 3.29 + 1.61 \hat{y}_{1463}(3) - 0.83 \hat{y}_{1463}(2) \\
&= 3.29 + 1.61(15.87) - 0.83(14.57) \\
&= 16.75
\end{aligned}$$

A continuación se presentan en la tabla 3.12 las predicciones puntuales para \hat{y}_{1464} , \hat{y}_{1465} , \hat{y}_{1466} , \hat{y}_{1467} del modelo ARMA (2,2)

Tabla 3.12. Predicciones para el modelo ARMA (2,2)

Tiempo de Medida (Horas) t	Serie Emanaciones CO ₂ y_t	Predicciones
...
1461	10,907	11,62582
1462	11,500	11,39791
1463	11,422	12,41800
1464		13,10253
1465		14,57654
1466		15,88256
1467		16,76098
...

Por lo tanto, revisada la serie emanación CO₂, utilizando las fases sugeridas por Box-Jenkins: se estudio la estacionariedad en media, la componente estacional (no se contempla) y su varianza, se concluyó que dicha serie cumple con estas condiciones; lo que permitió estudiar las autocorrelaciones e identificar los posibles modelos

obteniéndose los modelos ARMA(1,6), ARMA(1,7), ARMA(2,2), ARMA(4,1) y ARMA(4,2). Se estiman sus respectivos parámetros de cada modelo y posteriormente se realiza el análisis de los coeficientes a partir de la revisión de la significatividad de éstos, definimos luego, la ecuación de cada modelo. Superada esta fase, abordamos el chequeo de los modelos estimados; se realizó el análisis de los residuos encontrando que los modelos ARMA(1,6), ARMA(1,7), ARMA(4,1) y ARMA(4,2) no superan dicho análisis. Así, el único modelo que supera todas las fases es el modelo ARMA(2,2), lo que permite concluir que es el modelo que mejor se ajusta a la serie emanación CO₂. Finalmente, se procedió a efectuar el cálculo de las predicciones puntuales para los valores $\dots, \hat{y}_{1463}, \hat{y}_{1464}, \hat{y}_{1465}, \hat{y}_{1466}, \dots$ obteniéndose los resultados mostrados en tabla 3.12.

3.3.5 Conclusiones y futuras líneas de investigación

Después de exponer el comportamiento de la serie emanación de gases dióxido de carbono del Volcán de San Vicente (de los meses de marzo y abril del año 2003) y haberse aplicado la metodología de Box-Jenkins, en donde se evaluó cada uno de los modelos encontrados y se comprobó la calidad de las estimaciones se obtuvo como resultado de este proceso, que el modelo ARMA (2,2) es el que más se ajustó a nuestra serie en estudio, en el sentido que cumplió con las fases siguientes:

- a) Identificación
- b) Estimación
- c) Chequeo
- d) Predicción

Es importante señalar que el proceso para la obtención del modelo ARMA (2,2) provee al analista de Ciencias Físicas una herramienta efectiva para la investigación referido al monitoreo volcánico que se efectúa diariamente en nuestro país. Además, que puedan utilizar esta investigación para estudiar las emanaciones del dióxido de carbono que actualmente emiten otros volcanes.

El análisis de esta serie se realizó con el fin de estudiar y establecer una metodología para modelar series que se utilicen como línea base de comportamientos de fenómenos, en nuestro caso la emanación de gas dióxido de carbono del volcán de San Vicente o bien de otros fenómenos observables.

También, los resultados de esta investigación pueden ser utilizados como una técnica, por instituciones afines al monitoreo volcánico, para que tomen las precauciones en caso de una modificación de los niveles de actividad volcánica y estar atentos a cualquier tipo de manifestación que pueda detectarse en el transcurso del tiempo.

Este resultado sirva de motivación y guía a los estudiantes para la realización de trabajos de investigación en las diferentes áreas de las ciencias y además, trabajar en problemas que implique una serie temporal de fenómenos que nos proporcione información de manera continua y prolongada.

El trabajo demuestra el carácter interdisciplinario que en la actualidad adquiere el trabajo científico.

BIBLIOGRAFÍA

- Arrondo, Vicente Manzano (1997) Inferencia Estadística. Aplicaciones con SPSS / PC+. GRUPO EDITOR ALFA OMEGA, S.A. de C.V. México.
- Aznar, Antonio y Trivez, Francisco Javier (1993). Métodos de Predicción en Economía I. Editorial Ariel , S.A. Barcelona. 1ª. Edición.
- Aznar, Antonio y Trivez, Francisco Javier (1993). Métodos de Predicción en Economía II. Editorial Ariel Economía. Barcelona. 1ª. Edición.
- Kazmier, Leonard J. (1999). Estadística Aplicada a la Administración a la Economía. Editorial Mc Graw Hill, México. 3ª Edición
- Montgomery, Douglas C. , Peck, Elizabeth A., Vining, Geoffrey (2002). Introducción al análisis de Regresión Lineal. Compañía editorial Continental, , México. 3ª Edición en Inglés.
- Otero, José María (1993) Economía. Series Temporales. Editorial AC.
- Pardo, Antonio. , Ruiz Miguel Ángel. (2002) SPSS 11 Guía para el análisis de datos. Editorial Mc Graw Hill, México. Primera edición.
- SPSS, for Windows release 10.0.7 (2000) Standard Version.
- Tobías Garcés, Aurelio Análisis, Predicción y Regresión de series temporales epidemiológicas Modelos Box-Jenkins con SPSS y Stata. Internet Explorer. <http://perso.wanadoo.es/bledatobias/series.html>,<http://www.geocities.com/bledatobias/series.html>.
- Wei, William (1994). Time Series Analysis Univariate and Multivariate Methods. Editorial Addison-Wesley Publishing Company Inc.