

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



Ajuste de un modelo ARIMA para la precipitación diaria en la zona oriental de El Salvador

**Trabajo de Graduación presentado por:
Deysi Maribel Renderos
Mario Giovanni Molina Masferrer**

**Para Optar al grado de:
Maestro(a) en Estadística**

**Asesor:
Msc. Rolando Lemus Gómez**

Noviembre, 2008

San Salvador

El Salvador

Centroamérica

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

**Trabajo de Graduación:
Ajuste de un modelo ARIMA para la precipitación
diaria en la zona oriental de El Salvador**

**Presentado por:
Deysi Maribel Renderos
Mario Giovanni Molina Masferrer**

**Msc. Rolando Lemus Gómez
Asesor**

Ciudad Universitaria, San Salvador, Noviembre de 2008

Universidad de El Salvador

Ing. Rufino Antonio Quezada Sánchez
Rector

Lic. Douglas Vladimir Alfaro Chávez
Secretario General

Facultad de Ciencias Naturales y Matemática

Dr. Rafael Gómez Escoto
Decano

Lic. Trinidad de Castro
Secretaria

Escuela de Matemática

Ing. Carlos Mauricio Canjura Linares
Director

Lic. Ernesto Américo Hidalgo
Secretario

Contenido

1.	Introducción.....	1
2.	Planteamiento del Problema.....	3
3.	Antecedentes y Justificación.....	4
4.	Objetivos.....	7
5.	Fundamento Teórico.....	8
5.1	Definiciones de Modelos Estocásticos de Series Temporales.....	10
5.2	Modelos Autorregresivos (Modelos AR).....	11
5.3	Modelos de Medias Móviles (Modelos MA).....	14
5.4	Modelos ARMA.....	18
5.5	Modelos ARIMA.....	20
5.6	Modelos ARIMA generales.....	23
6.	Metodología de Estimación del Modelo de Predicción.....	25
7.	Estimación del modelo.....	28
7.1	Identificación del modelo.....	28
7.1.1	Análisis de la serie original.....	31
7.1.2	Identificación de los posibles modelos.....	38
7.2	Estimación de los parámetros de los modelos identificados.....	40
7.3	Chequeo.....	47
7.3.1	Significación de los parámetros.....	47
7.3.2	Análisis de los residuos.....	49
7.3.3	Análisis de Sobreajuste.....	62
7.3.4	Análisis de Estabilidad.....	63
7.4	Predicción.....	65
7.4.1	El Predictor óptimo.....	65
7.4.2	Cálculo de la predicción puntual.....	66
7.4.3	Bondad de ajuste de los modelos seleccionados.....	68
7.5	Análisis de la Serie Temporal con Outliers.....	70
7.5.1	Modelos de Intervención.....	70
7.5.2	Identificación del Modelo ARIMA con intervención.....	75
8.	Conclusiones y Recomendaciones.....	83
9.	Fuentes Bibliográficas.....	85
	ANEXOS.....	87

Lista de Figuras y Gráficas

Figuras

Figura 1. Ubicación de Estaciones que serán utilizadas en el estudio. Fuente: SMN-SNET	4
Figura 2. Etapas en la construcción de un modelos de series temporales. Peña (1989)	27

Gráficas

Gráfica 1. Evolución de los datos en el tiempo	31
Gráfica 2. Diagrama de caja de la serie original	32
Gráfica 3. La serie original sin datos atípicos	33
Gráfica 4. FAS de los datos originales	33
Gráfica 5. FAP de los datos originales	34
Gráfica 6. Evolución de la serie con una diferencia	36
Gráfica 7. Función de autocorrelación simple de la lluvia diaria	39
Gráfica 8. Función de autocorrelación parcial de la lluvia diaria.....	39
Gráfica 9. Residuos de la serie de lluvia diaria, IMA(1,1).....	49
Gráfica 10. Probabilidad normal e histograma de los residuos , IMA(1,1)	50
Gráfica 11. Autocorrelación simple de los residuos del IMA(1,1)	51
Gráfica 12. Autocorrelación parcial de los residuos del IMA(1,1).....	52
Gráfica 13. Residuos de la serie de lluvia diaria, ARI (1,1).....	52
Gráfica 14. Probabilidad normal e histograma de los residuos , ARI (1,1).....	53
Gráfica 15. Autocorrelación simple de los residuos del ARI(1,1)	53
Gráfica 16. Autocorrelación parcial de los residuos del ARI(1,1).....	54
Gráfica 17. Residuos de la serie de lluvia diaria, ARI (1,1).....	54
Gráfica 18. Probabilidad normal e histograma de los residuos. ARI(2,1)	55
Gráfica 19. Autocorrelación simple de los residuos del ARI (2,1)	55
Gráfica 20. Autocorrelación parcial de los residuos del ARI (2,1).....	56
Gráfica 21. Residuos de la serie de lluvia diaria, ARIMA(4,1,4).....	56
Gráfica 22. Probabilidad normal e histograma de los residuos. ARIMA(4,1,4).....	57
Gráfica 23. Autocorrelación simple de los residuos del ARIMA (4,1,4)	57
Gráfica 24. Autocorrelación parcial de los residuos del ARIMA (4,1,4)	58
Gráfica 25. Residuos de la serie de lluvia diaria, IMA(1,2).	58
Gráfica 26. Probabilidad normal e histograma de los residuos. IMA(1,2)	59
Gráfica 27. Autocorrelación simple de los residuos del IMA (1,2).....	59
Gráfica 28. Autocorrelación parcial de los residuos del IMA (1,2).....	60
Gráfica 29. Residuos de la serie de lluvia diaria, ARI(4,1)	60
Gráfica 30. Probabilidad normal e histograma de los residuos. ARI(4,1)	61
Gráfica 31. Autocorrelación simple de los residuos del ARI (4,1)	61
Gráfica 32. Autocorrelación parcial de los residuos del ARI (4,1).....	62
Gráfica 33. Diferentes formas del Comportamiento de la variable escalón en una	72
Gráfica 34. Diferentes formas del Comportamiento de la variable impulso en una	73
Gráfica 35. Comportamiento de la variable impulso y escalón en una serie temporal.....	73
Gráfica 36. Comportamiento de los residuos del modelo IMA(1,2) con Intervención.....	78
Gráfica 37. Funciones de Autocorrelación Simple y Parcial de los residuos del IMA(1,2).....	78
Gráfica 38. Histograma y Probabilidad Normal de los residuos del IMA(1,2)	78
Gráfica 39. Comportamiento de los residuos del modelo ARI(1,1) con Intervención.....	79

Gráfica 40. Funciones de Autocorrelación Simple y Parcial de los residuos del ARI (1,1).....	80
Gráfica 41. Histograma y Probabilidad Normal de los residuos del ARI(1,1)	80

Lista de Tablas

Tabla 1. Categorización de los años según el	5
Tabla 2. Correspondencia entre los modelos AR(p) y MA(q).....	17
Tabla 3. Características de las FAS y FAP en los modelos MA(q), AR(P) y ARMA(p,q).....	31
Tabla 4. Estadísticos descriptivos de la serie original.....	32
Tabla 5. Resultados del Test de Dickey-Fuller Aumentado (ADF)	35
Tabla 6. Resultados del Test de Levene.....	36
Tabla 7. Parámetros del Modelo ARIMA(1,1,2).....	40
Tabla 8. Parámetros del Modelo ARIMA(1,1,3).....	41
Tabla 9. Parámetros del Modelo ARIMA(1,1,4).....	41
Tabla 10. Parámetros del Modelo ARIMA(2,1,3)	42
Tabla 11. Parámetros del Modelo ARIMA(2,1,4)	42
Tabla 12. Parámetros del Modelo ARIMA(4,1,3)	43
Tabla 13. Parámetros del Modelo ARIMA(5,1,3)	44
Tabla 14. Parámetros del Modelo ARIMA(5,1,4)	44
Tabla 15. Parámetros del Modelo ARIMA(6,1,3)	45
Tabla 16. Parámetros del Modelo ARIMA(6,1,4)	46
Tabla 17. Modelos identificados, reducidos y sus expresiones generales que pasaran a la etapa de chequeo.....	46
Tabla 18. Análisis de estacionariedad e invertibilidad de los modelos identificados	48
Tabla 19. Diferencias de los Parámetros del modelo IMA (1,1) y según muestras 1 y 2	63
Tabla 20. Diferencias de los Parámetros del modelo ARI (1,1), según muestras 1 y 2.....	63
Tabla 21. Diferencias de los Parámetros del modelo ARI (2,1), según muestras 1 y 2.....	64
Tabla 22. Diferencias de los Parámetros del modelo IMA (1,2), según muestras 1 y 2.....	64
Tabla 23. Diferencias de los Parámetros del modelo ARI (4,1), según muestras 1 y 2.....	64
Tabla 24. Resultados de las predicciones de \hat{y}_{46} (15 de junio)	67
Tabla 25. Resultados de las predicciones de \hat{y}_{107} (15 de agosto).....	68
Tabla 26. Resultados de las predicciones de \hat{y}_{168} (15 de octubre).....	68
Tabla 27. Resultados de los criterios DER, AIC y SBC para los modelos encontrados	69
Tabla 28. Resultados de la corrida del modelo IMA (1,2) con intervenciones.....	77
Tabla 29. Resultados de las predicciones de \hat{y}_{46} (15 de junio), sin y con intervención.....	79
Tabla 30. Resultados de las predicciones de \hat{y}_{46} (15 de junio), sin y con intervención.....	81
Tabla 31. Resultados de las predicciones de \hat{y}_{107} (15 de agosto), sin y con intervención.....	81
Tabla 32. Resultados de las predicciones de \hat{y}_{168} (15 de octubre), sin y con intervención.....	81

1. Introducción

El presente estudio se enmarca en las actividades del Servicio Meteorológico Nacional del Servicio Nacional de Estudios Territoriales (SNET)¹ del Ministerio de Medio Ambiente de El Salvador (MARN).

El Servicio Meteorológico Nacional (SMN) es la dirección encargada, dentro de SNET, del estudio y monitoreo de los fenómenos relacionados con el clima en El Salvador. Para tal efecto el SMN cuenta con diferentes herramientas de análisis, monitoreo y predicción que le permiten el estudio constante de las condiciones climáticas de nuestro país.

El estudio de las condiciones climáticas, que hoy día cobra más importancia debido a los cambios del clima a nivel global, pasa por diferentes etapas, pero una de las más importantes es la observación de las diferentes variables como precipitación, luz solar, temperatura, etc. Estas observaciones forman la base del conocimiento de las variables, que al analizarlas permiten establecer patrones de comportamiento del clima en general para la región.

Cabe mencionar, para los objetivos de este estudio, que El Salvador cuenta con dos estaciones, relacionadas al régimen de lluvia, bien definidas: la estación seca (de noviembre a abril) y la época lluviosa (de mayo a octubre), presentándose entre los meses de julio y agosto una canícula (período de ausencia de lluvia), y en estos meses la precipitación se concentra sobre todo en las horas nocturnas. Este período de canícula, trae implicaciones a diferentes sectores del país, pero el más afectado es el sector agrícola, ya que se generan diferentes niveles de sequía en todo el país, generándose el nivel de sequía más severa en la región oriental, la cual se ve afectada cada año por este fenómeno, y se agudiza todavía más en los casos cuando existe la presencia del fenómeno del Niño, en los que se han registrado a través de la historia períodos de hasta 60 días consecutivos, registrándose también las mayores pérdidas en las producciones agrícolas en los países afectados por dicho fenómeno.

En el país, se han realizado una gran cantidad de estudios sobre el comportamiento del clima, que van desde el estudio del comportamiento de series normales (períodos de 30 años), hasta el estudio del régimen de precipitación mensual. No existen estudios que permitan estimar la precipitación diaria, que en muchos casos, sería de utilidad para incorporar estos pronósticos en los diferentes análisis que los especialistas realizan con objetivos de monitoreo y alertamiento temprano sobre el fenómeno de la sequía.

¹ El Servicio Nacional de Estudios Territoriales, fue creado después de los terremotos del 2001, mediante decreto ejecutivo No. 96, Diario Oficial tomo No. 353, del 18 de octubre de 2001.

En el presente estudio se hará un análisis de series temporales sobre una serie de datos que corresponden a los meses de abril a octubre del año 2003, el cual fue un año normal de precipitación (ver tabla 1).

2. Planteamiento del Problema

El Salvador, es el único país de la región que no tiene frontera con el Mar del Caribe, sino solamente con el Océano Pacífico, por lo que la variación estacional del clima depende de los sistemas sinópticos cuasi estacionarios y de los migratorios tanto de las latitudes altas como del mismo trópico, a lo cual se le suma la interacción de los sistemas diurnos modificados por el Océano Pacífico².

El régimen lluvioso interanual es bimodal y presenta los valores máximos en Septiembre o en Junio, con una disminución importante entre julio y agosto, llamada Canícula. La ubicación de la orografía en El Salvador relativo al flujo del viento predominante provoca que la canícula sea en algunas ocasiones más intensa, produciéndose entonces sequías asociadas en la mayoría de los casos al fenómeno El Niño.

Las variaciones extremas del clima se han convertido en grave problema en especial para los pequeños productores, mayoría en Centro América, que trabajan la tierra en las extensas zonas rurales; en nuestro país la zona más afectada es la oriental (departamentos de Usulután, San Miguel, Morazán y La Unión). Tanto las sequías como las inundaciones provocan pérdidas en el sector agropecuario que traen hambruna y luego migraciones. Además, se ven afectados otros sectores como la generación de energía hidroeléctrica, la salud humana, la vida silvestre, y finalmente la socioeconomía del país.

Actualmente, el Servicio Meteorológico Nacional se encuentra trabajando en la estimación de valores de precipitación diaria, para lo cual utilizan modelos generados por el proyecto MFEWS (Mesoamerican Food Security Early Warning System – Sistema mesoamericano de Alerta Temprana para Seguridad Alimentaria), los cuales inicialmente fueron generados para grandes extensiones como África, pero que han sido ajustados a la zona centroamericana, y que para el caso de El Salvador, con una extensión territorial tan pequeña, no se consideran modelos muy adecuados para la estimación de la precipitación diaria.

En este trabajo se hará un análisis para la construcción de un modelo de predicción de la precipitación diaria para la zona oriental de El Salvador, y para lo cual se estudiará los datos de precipitaciones diarias para un período de seis meses de un año normal, específicamente para el 2003 (en términos del régimen climático, los meteorólogos consideran los años como años Niño, años Niña o años Normales)

² García y otros, Estudio Técnico Mejoramiento de la capacidad técnica para mitigar los efectos de futuros eventos de la Variabilidad Climática (El Niño), Proyecto TC-97-09-46-3, BID-CEPREDENAC, 2002.

3. Antecedentes y Justificación.

Como se menciono antes, en el país se ha desarrollado muy poco la estimación de modelos que permitan ajustar datos de diferentes variables climáticas, como la precipitación, temperatura, etc. Más bien se han estado usando modelos que se han ajustado para otras regiones y luego se hace la transferencia tecnológica hacia países como el nuestro. Con la variabilidad climática actual se ha vuelto una necesidad estar trabajando con pronósticos más adecuados a las necesidades de la población y que permitan un alertamiento temprano ante fenómenos climáticos que pueden afectar la economía nacional en sectores tales como agrícola, generación de energía, salud y pesca, entre otros.

En el caso de las estimaciones de la precipitación diaria, no se cuenta con modelos ajustados para El Salvador. Un modelo utilizado para Alerta Temprana ante sequía o inundaciones es el utilizado por el MFEWS, el cual estima valores de precipitación diaria usando datos mesoescalares, lo que representa para nuestro país valores de igual precipitación para áreas muy grandes lo que impide analizar el microclima de ciertas regiones que se ven afectadas año con año por fenómenos de sequía o inundaciones.

La serie temporal objeto de este estudio corresponde al período de mayo a octubre del año 2003, ya que en estos meses la precipitación es más estable dentro del período lluvioso (de abril a octubre, extendiéndose en algunos casos hasta noviembre). Se ha escogido el año 2003, porque fue un año normal en relación al fenómeno del ENOS (conocido por el El Niño) (Ver Tabla 1).

Los datos a procesar corresponden a los valores de precipitación diaria de 19 estaciones ubicadas en la región oriental del país (Figura 1), los cuales son recopilados por el personal del

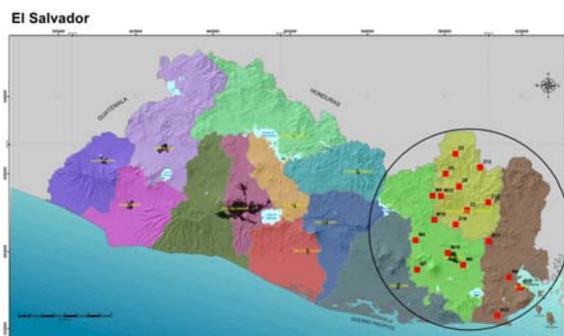


Figura 1. Ubicación de Estaciones que serán utilizadas en el estudio. Fuente: SMN-SNET

Servicio Meteorológico Nacional (SMN) del Servicio Nacional de Estudios Territoriales (SNET), como parte de sus funciones y a través de una red de estaciones de monitoreo y predicción climática, la cual está distribuida en todo el territorio nacional.

Esta red de estaciones, dependiendo del tipo, registran información de variables climáticas, entre ellas la precipitación. El

registro de la precipitación puede ser en intervalos de minutos, horas o días (depende del tipo

de estación). Dentro de la red existen estaciones que transmiten de forma automática los datos de precipitación, en otros casos los datos son leídos por un observador meteorológico y transmitidos a la sede a través de radio o teléfono y en otros casos, el observador anota la lectura de los instrumentos en una bitácora que luego es recogida por personal de la sede del SMN. En todos los casos, la información es almacenada en Bases de Datos u Hojas Electrónicas.

Para efectos de este estudio se tomarán los totales de precipitación diaria registrada para cada una de las estaciones consideradas.

Año	Fase/Intensidad ENOS
1980	Normal
1981	Normal
1982	Niño
1983	Niño
1984	Niña
1985	Niña
1986	Niño
1987	Niño
1988	Niña
1989	Niña
1990	Niño
1991	Niño
1992	Niño
1993	Niño
1994	Niño
1995	Normal
1996	Normal
1997	Niño
1998	Niña
1999	Niña
2000	Normal
2001	Normal
2002	Niño
2003	Normal

Tabla 1. Categorización de los años según el comportamiento del Niño

Este estudio sentará un precedente para ir abordando las necesidades de construcción y ajuste de modelos de predicción propios relacionados con esta y otras variables climáticas de interés para las instituciones sectoriales y población en general.

También puede ser complemento en la implementación de programas de alerta temprana para seguridad alimentaria, sobre todo en la zona oriental del país, que se ve

afectada por la recurrencia de fenómenos de sequía, y en la cual han puesto interés organismos nacionales y regionales que trabajan en el tema.

4. Objetivos

General:

Aportar al Servicio Meteorológico Nacional del Servicio Nacional de Estudios Territoriales el ajuste de un modelo de predicción de la precipitación diaria para la zona oriental de El Salvador que permita generar proyecciones útiles y hacer recomendaciones pertinentes a las instituciones encargadas de la prevención de desastres naturales ocasionados por fenómenos climatológicos.

Específicos:

- Analizar las series de datos de precipitación diaria de 19 estaciones meteorológicas y/o pluviométricas de San Miguel, Morazán y La Unión (Departamentos con más afectación por sequía severa en el país), para el período mayo-octubre del 2003.
- Ajustar un modelo de serie temporal para la predicción de precipitación pluvial diaria en la zona oriental de El Salvador para años con régimen de lluvia normal.

5. Fundamento Teórico

En el marco del estudio denominaremos **predicción** a la estimación de valores futuros de una variable en función del comportamiento pasado de ésta. La predicción de un modelo se concibe como un proceso que incluye desde la obtención de información relevante para predecir la variable, hasta la obtención de la predicción luego del análisis y procesamiento de la información.

Una **serie temporal** es el conjunto de valores de una variable correspondientes a períodos temporales sucesivos de igual duración. Dichos períodos tienen la misma amplitud y la serie tiene un carácter discreto.

Toda técnica de predicción está caracterizada por dos elementos básicos:

- La forma en que se elabora la base informativa
- El procedimiento que se sigue para procesar la información de la base

Los procedimientos a seguir para procesar la información pueden clasificarse en Métodos subjetivos y Métodos objetivos.

En este estudio se hará uso de métodos objetivos. Según Aznar³ en los métodos objetivos la base informativa está disponible al comienzo del ejercicio de predicción en forma de valores que toma la variable a lo largo de una serie de períodos. El procedimiento que se sigue para procesar la información relevante es un proceso objetivo y replicable. En este sentido, el concepto relevante es el Modelo. Un modelo estadístico es una representación abstracta de la realidad expresada en lenguaje matemático, que recoge los aspectos esenciales del fenómeno que se quiere predecir. En los modelos estadísticos es importante distinguir entre modelos paramétricos y modelos no paramétricos. En los primeros se especifica una estructura estocástica de la población a partir de la cual se supone que se han obtenido las observaciones de las variables que componen la base informativa. En los métodos no paramétricos, por el contrario, no se hace mención a dicha estructura.

Si podemos encontrar patrones de regularidad en diferentes secciones de una serie temporal, podremos también describirlas mediante modelos basados en distribuciones de probabilidad. La secuencia ordenada de variables aleatorias $X(t)$ y su distribución de probabilidad asociada,

³ Aznar, A., Trivez, Francisco, "Métodos de predicción en Economía, Tomo II, Ariel Economía, 1993.

se denomina **proceso estocástico**. Un proceso estocástico es por lo tanto el modelo matemático para una serie temporal, el cual se define de la siguiente manera:

Definición: Un proceso estocástico es lineal discreto si cada observación y_t puede expresarse de la forma:

$$y_t = \mu + u_t + \Psi_1 u_{t-1} + \Psi_2 u_{t-2} + \dots \quad (1)$$

Donde μ y los Ψ_i 's son parámetros desconocidos, y $u_t, u_{t-1}, u_{t-2}, \dots$ es una secuencia de perturbaciones aleatorias distribuidas idéntica e independientemente con media cero y varianza σ_u^2 .

En este contexto, cuando la sucesión de variables aleatorias (proceso estocástico) presenta esperanza nula, varianza constante, y covarianzas nulas para distintos valores de t , se conoce como **ruido blanco**.

Los principales inconvenientes en la construcción de un modelo para un proceso estocástico son:

- Para elaborar un modelo para la variable **Y** en estudio a partir de una única serie temporal **y** hay que suponer ciertas simplificaciones sobre la estructura probabilística del proceso Y_t del que procede **Y**, de manera que lo que no se suponga se pueda inferir a partir de una única serie temporal de **Y**. La solución a este inconveniente es el establecimiento de la hipótesis de estacionariedad, que tiene como propiedades teóricas las siguientes:
 - Estacionariedad: Cierta estado de equilibrio estadístico que caracteriza la evolución temporal de un proceso estocástico que ha generado una serie temporal.
 - Media: Valor constante en el tiempo que mide el nivel alrededor del cual evoluciona un proceso estocástico estacionario.
 - Varianza: Valor constante en el tiempo que mide la dispersión o la variabilidad de la evolución temporal de un proceso estocástico estacionario alrededor de su media.
 - Autocorrelaciones: Valores constantes en el tiempo que miden el grado de asociación lineal entre cada par de componentes de un proceso estocástico estacionario separados por diferentes intervalos temporales o retardos.

- Otro inconveniente consiste en que muchas series temporales no pueden considerarse generadas por procesos estocásticos estacionarios (es decir, son series no estacionarias) porque presentan ciertas tendencias claras en su evolución temporal, de manera que no presentan afinidad hacia algún valor constante en el tiempo, porque su variabilidad no es constante, o porque son estacionales. La solución a este inconveniente es la transformación de las series originales. Es decir, una serie temporal no estacionaria puede transformarse de forma adecuada para obtener una serie temporal de aspecto estacionario. La serie transformada puede utilizarse para elaborar un modelo estadístico que resulte compatible con la hipótesis de estacionariedad. Las transformaciones más utilizadas en la práctica para obtener series temporales estacionarias son: la **transformación logarítmica**, la **diferenciación regular** y la **diferenciación estacional**.

5.1 Definiciones de Modelos Estocásticos de Series Temporales

Cuando se trata de representar la influencia de los hechos pasados sobre el presente y en consecuencia sobre el futuro de un proceso estocástico, podemos considerar diferentes modelos alternativos.

Una alternativa consiste en colocar el valor actual del proceso como dependiente de modo lineal de valores pasados del propio proceso, más una perturbación aleatoria que evite que el modelo sea determinista:

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

A esta expresión la llamaremos “autorregresiva”, porque es un modelo de regresión del proceso sobre sí mismo.

La otra alternativa de representación de la dependencia de la variable respecto al pasado consiste en considerar el valor actual como el resultado de la combinación de una serie de factores aleatorios independientes entre sí (modelo de ruido blanco). Al igual que en el modelo autorregresivo se debe añadir una perturbación aleatoria al modelo para garantizar su carácter estocástico:

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

A esta expresión se le denomina de “medias móviles”.

Un elemento importante en el estudio de estos modelos lineales es el denominado operador de retardo L. Tal operador actúa sobre un término de un proceso estocástico reduciendo el índice de temporalidad en una unidad: $L y_t = y_{t-1}$

5.2 Modelos Autorregresivos (Modelos AR)

Un modelo autorregresivo de orden p, o abreviadamente AR(p), se define como:

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

En términos teóricos, no hay nada que nos limite el número de períodos del pasado que influyen en el valor actual del proceso, sin embargo tal planteamiento provoca problemas no justificables en términos prácticos; es decir, no es necesario tener que estimar infinitos coeficientes de regresión cuando en la práctica se observa que solo los períodos más recientes tienen influencia significativa en el valor actual del proceso.

Por tanto, la ecuación 2 puede escribirse de manera abreviada como: $\phi(L)y_t = \delta + u_t$, donde $\phi(L)$ es un operador polinomial de retardos, siendo $\phi(L) = 1 - \phi_1 L - \phi_2 L^2 - \dots - \phi_p L^p$.

Estos modelos, deben cumplir la condición de estacionariedad.

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

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

Al cumplir el supuesto 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

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

Propiedades de los modelos AR(p):

Estacionariedad: La condición necesaria y suficiente de estacionariedad en un modelo AR(p) es que las raíces de su polinomio autorregresivo $\phi(L) = 1 - \phi_1 L - \phi_2 L^2 - \dots - \phi_p L^p = 0$ estén fuera del círculo unidad del plano complejo (sean en modulo estrictamente mayores que la unidad).

Invertibilidad: Los modelos AR(p) son siempre invertibles.

Correlograma simple: El correlograma simple de un modelo AR(p) está formado por infinitos coeficientes de correlación no nulos, decreciendo según una mezcla de exponenciales y sinusoidales amortiguadas.

Correlograma parcial: El correlograma parcial de un modelo AR(p) tiene no nulos los p primeros coeficientes de correlación ρ'_k (desde $k=1$ hasta $k=p$), siendo nulos a partir de $k=p+1$.

5.3 Modelos de Medias Móviles (Modelos MA)

Un modelo de Medias Móviles de orden q denotado por (MA(q)) se define como:

$$y_t = \mu + u_t - \theta_1 u_{t-1} - \theta_2 u_{t-2} - \dots - \theta_q u_{t-q} \quad (3)$$

Donde:

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

μ : 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 .

La expresión dada en la ecuación 3 puede escribirse de manera simplificada como

$y_t = \mu + \theta(L)u_t$, donde $\theta(L)$ es un operador polinomial de retardos, definido como:

$\theta(L) = 1 - \theta_1 L - \theta_2 L^2 - \dots - \theta_q L^q$, siendo L el operador de retardo, tal que el valor de su

exponente indica el orden del retardo de la variable a la que multiplica; es decir:

$$L^s y_t = y_{t-s}, \forall s \geq 0.$$

Este modelo siempre será estacionario, ya que es un caso particular del proceso estocástico discreto general, definido en (1), con las restricciones:

$\Psi_i = 0, \forall i \geq q, \Psi_i = -\theta_i$ para $1 \leq i \leq q$ y $\Psi_0 = 1$, debe cumplirse entonces:

$$\sum_{i=0}^{\infty} \Psi_i = 1 - \theta_1 - \theta_2 - \dots - \theta_q = 1 - \sum_{i=0}^{\infty} \theta_i = k, \text{ donde } k \text{ es una constante, } q \text{ el número de}$$

parámetros de medias móviles.

$$\sum_{i=0}^{\infty} \Psi_i^2 = 1 + \theta_1^2 + \theta_2^2 + \dots + \theta_q^2 < \infty$$

$$\sum_{i=0}^{\infty} \Psi_i \Psi_{i+j} = \Psi_0 \Psi_j + \Psi_1 \Psi_{j+1} + \Psi_2 \Psi_{j+2} + \dots + \Psi_{q-j} \Psi_q = -\theta_j + \theta_1 \theta_{j+1} + \theta_2 \theta_{j+2} + \dots + \theta_{q-j} \theta_q < \infty$$

Además, un modelo de medias móviles es invertible cuando se puede escribir como un proceso autorregresivo de orden infinito, para lo cual debe cumplir que las raíces de $\theta(L) = 1 - \theta_1 L - \theta_2 L^2 - \dots - \theta_q L^q$, caigan fuera del círculo unitario, es decir si L_1, L_2, \dots, L_q , son las q raíces, estas deben ser mayores que uno en valor absoluto, cuando las raíces son reales; si las raíces son complejas el módulo debe ser mayor a la unidad.

La esperanza matemática del modelo de medias móviles, 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

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

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

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

.....

$$\phi_{jj} = \begin{vmatrix} 1 & \rho_1 & \rho_2 & \dots & \rho_{j-2} & \rho_1 \\ \rho_1 & 1 & \rho_1 & \dots & \rho_{j-3} & \rho_2 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots \\ \rho_{j-1} & \rho_{j-2} & \rho_{j-3} & \dots & \rho_1 & 1 \\ \hline 1 & \rho_1 & \rho_2 & \dots & \rho_{j-2} & \rho_{j-1} \\ \rho_1 & 1 & \rho_1 & \dots & \rho_{j-3} & \rho_{j-2} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots \\ \rho_{j-1} & \rho_{j-2} & \rho_{j-3} & \dots & \rho_1 & 1 \end{vmatrix}$$

Propiedades del modelo MA (q):

Estacionariedad: Los procesos de medias móviles de orden q son siempre estacionarios, al cumplir tres condiciones de la estacionariedad en sentido amplio.

Invertibilidad: Un proceso 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, donde

$\theta(L)=0$ se define como $\theta(L)=1 - \theta_1L - \theta_2L^2 - \dots - \theta_qL^q = 0$

Correlograma simple: El correlograma simple de un proceso MA(q) tiene los q primeros coeficientes de correlación ρ'_k desde k=1 hasta k=q no nulos, siendo el resto nulos. Es decir, 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 q del proceso.

Correlograma parcial: El correlograma parcial de un proceso MA(q) consta de infinitos coeficientes de correlación no nulos, decreciendo según una mezcla de exponenciales y sinusoidales amortiguadas. Es decir, 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).

Al observar las características de los modelos AR(p) y MA(q) se aprecia cierta simetría que se denomina dualidad, la cual es consecuencia del hecho de que un mismo proceso podemos representarlo como autorregresivo o como de medias móviles según nuestra conveniencia, dado que un AR(p) puede representarse como un MA(∞) y un MA(q) como un AR(∞). La ventaja de usar en cada caso la representación adecuada reside en tener un modelo con pocos parámetros.

En resumen, las correspondencias entre ambos modelos se muestran en la tabla siguiente:

Concepto	AR (p)	MA(q)
Expresión	$\phi(L)y_t = \delta + u_t$	$y_t = \mu + \theta(L)u_t$
Condición de estacionariedad	Raíces de $\phi(L)$ en módulo mayores que 1	Siempre es estacionario
Condición de invertibilidad	Siempre es invertible	Raíces de $\theta(L)$ en módulo mayores que 1
Correlograma simple	Infinitos valores no nulos decreciendo según mezcla de exponenciales y sinusoidales amortiguadas	Coefficientes no nulos hasta el retardo q, el resto nulos
Correlograma parcial	Coefficientes no nulos hasta el retardo p, el resto nulos	Infinitos valores no nulos decreciendo según mezcla de exponenciales y sinusoidales amortiguadas
Modelo Dual	MA(∞)	AR(∞)

Tabla 2. Correspondencia entre los modelos AR(p) y MA(q)

5.4 Modelos ARMA.

Un modelo mixto que tiene parte autorregresiva y parte de medias móviles (ARMA(p,q)) se representa como:

$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \dots + \phi_p y_{t-p} + \delta + u_t - \theta_1 u_{t-1} - \theta_2 u_{t-2} - \dots - \theta_q u_{t-q}$, y el cual, al igual que en los casos anteriores, puede expresarse simplifcadamente, usando los polinomios de retardo como: $\phi(L)y_t = \delta + \theta(L)u_t$. Dado que este modelo contiene una parte AR y una parte MA, deberá cumplir las condiciones de estacionariedad como de invertibilidad. O sea que diremos que es estacionario si su parte AR lo es y diremos que es invertible si su parte MA lo es.

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

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

Varianza

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

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

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

Propiedades del modelo ARMA(p,q)

Estacionariedad: Un proceso ARMA(p,q) es estacionario si y solo si las raíces de su polinomio autorregresivo ($\phi(L) = 0$) están todas fuera del círculo unidad del plano complejo.

Invertibilidad: Un proceso ARMA(p,q) es invertible si y solo si las raíces de su polinomio de medias móviles ($\theta(L) = 0$) están todas fuera del círculo unidad del plano complejo.

Función de Autocorrelación: 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).

Correlograma Simple: 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.

Correlograma Parcial: 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.

Los Modelos ARMA permiten una máxima flexibilidad de representación con muy pocos parámetros, y rara vez los órdenes p y q sobrepasan el valor 2.

5.5 Modelos ARIMA.

En los modelos ARMA, los cuales están constituidos por una combinación de p términos **AR** (proceso autorregresivo), y q términos **MA** (proceso de medias móviles). Donde la parte AR modela la influencia de los valores anteriores de la serie (y_{t-1} hacia atrás), y la parte MA modela la influencia del ruido en valores anteriores de la serie (u_{t-1} hacia atrás), junto con el término u_t que corresponde al ruido esperado en el mismo momento t en el que se estima el nuevo valor de la variable y_t . Una de las ventajas de estos modelos es su gran simplicidad (sumas de términos), frente a los modelos propuestos en la formulación clásica.

La letra **I** que aparece en el nombre del modelo completo **-ARIMA-**, corresponde al proceso último a realizar, una vez definido el tipo de modelo y estimados los coeficientes de éste, ya que entonces hay que restablecer las características originales de la serie de datos, que fue

transformada para inducir estacionaridad. A ese proceso inverso se denomina en general Integración y aporta esa letra que completa el nombre.

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 aplicar ciertas transformaciones para convertir dicha serie temporal en estacionaria.

Estas transformaciones se concretan en diferenciaciones sucesivas en la serie (la determinación del parámetro d).

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:

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

Abreviadamente, 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) representa un modelo estocástico general. Así, si $p = d = 0$ estamos ante el modelo ARMA(0,0,q) o bien MA(q); si $q = d = 0$ se tiene el modelo ARMA(p, 0, 0) o bien un modelo AR(p) abordados anteriormente. Por otro lado, si $d = 0$ se tiene el modelo ARMA(p, 0, q) o bien ARMA(p,q).

Otros casos posibles 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"

La mayoría de las series temporales presentan 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. También puede observarse tendencia en la varianza; es decir, la dispersión de las observaciones respecto a la media no es constante a lo largo del tiempo. Este tipo de tendencia no se elimina mediante el proceso de diferenciación, siendo lo adecuado transformar la serie aplicando logaritmos naturales.

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.

5.6 Modelos ARIMA generales

En los apartados anteriores se han desarrollado los modelos teóricos para series de tiempo no estacionales. Sin embargo, es frecuente encontrar series de tiempo cuya estructura incorpora comportamientos regulares y estacionales.

En los procesos ARIMA estacionales existe un tipo de no estacionariedad peculiar, que es la denominada estacionalidad. Los comportamientos estacionales son la repetición periódica regular de pautas de comportamiento en la serie (pueden ser mensuales, semestrales, anuales, etc). Estos comportamientos implican una falta de estacionariedad en la serie, pues significan un comportamiento sistemáticamente diferenciado en los distintos periodos del ciclo. Los procesos que presentan este tipo de comportamientos pueden ser estacionarizados mediante la aplicación del operador diferencial estacional; es decir, aplicar un número D de diferencias estacionales, que se denomina orden de diferenciación estacional y rara vez es superior a uno.

La forma general para esta clase de modelos, denotados por ARIMA(P,D,Q)_s está dado por:

$$w_t = \Phi_1 w_{t-s} + \Phi_2 w_{t-2s} + \dots + \Phi_p w_{t-ps} + \delta + u_t - \Theta_1 u_{t-s} - \dots - \Theta_q u_{t-qs}$$

dónde:

$$w_t = \Delta_s^D y_t = (1-L^s)^D y_t, \text{ lo que se puede representar como: } \Phi_p(L^s)w_t = \delta + \Theta_q(L^s)u_t;$$

$$\Theta_q(L^s) = 1 - \Theta_1 L^s - \Theta_2 L^{2s} - \dots - \Theta_q L^{qs}, \text{ es el operador de retardo para el modelo}$$

estacional de medias móviles estacional MA(Q)_s y $\Phi_p(L^s) = 1 - \Phi_1 L^s - \Phi_2 L^{2s} - \dots - \Phi_p L^{ps}$ es el operador de retardo para el modelo autorregresivo estacional AR(P)_s.

Si escribimos el modelo ARIMA(P,D,Q)_s como:

$$(1 - \Phi_1 L^s - \Phi_2 L^{2s} - \dots - \Phi_p L^{ps})(1 - L^s)^D y_t = (1 - \Theta_1 L^s - \Theta_2 L^{2s} - \dots - \Theta_q L^{qs})\varepsilon_t, \quad \text{entonces}$$

ε_t supondremos que viene generado por un proceso ARIMA(p,d,q), es decir:

$$(1 - \phi_1 L - \phi_2 L^2 - \dots - \phi_p L^p)(1 - L)^d \varepsilon_t = (1 - \theta_1 L - \theta_2 L^2 - \dots - \theta_q L^q)u_t, \text{ siendo } u_t \text{ un ruido}$$

blanco. Si sustituimos la segunda en la primera de las ecuaciones, obtendremos el modelo estacional general que denotaremos abreviadamente como ARIMA(p,d,q)x(ARIMA(P,D,Q)_s y que escribiremos como:

$$(1 - \Phi_1 L^s - \Phi_2 L^{2s} - \dots - \Phi_p L^{ps})(1 - \phi_1 L - \phi_2 L^2 - \dots - \phi_p L^p)(1 - L^s)^D (1 - L)^d y_t = (1 - \Theta_1 L^s - \Theta_2 L^{2s} - \dots - \Theta_q L^{qs})(1 - \theta_1 L - \theta_2 L^2 - \dots - \theta_q L^q)u_t$$

o, abreviadamente:

$$\Phi_p(L^s)\phi_p(L)(1 - L^s)^D (1 - L)^d y_t = \Theta_q(L^s)\theta_q(L)u_t$$

En su forma más general, esta expresión puede escribirse como:

$$\Phi_p(L^s)\phi_p(L)[(1 - L^s)^D (1 - L)^d y_t - \mu] = \Theta_q(L^s)\theta_q(L)u_t,$$

siendo μ la media de $w_t = (1 - L^s)^D (1 - L)^d y_t$.

6. Metodología de Estimación del Modelo de Predicción

A comienzo de los años 70, G.E.P. Box, profesor de Estadística de la Universidad de Wisconsin, y G.M. Jenkins, profesor de Ingeniería de Sistemas de la Universidad de Lancaster, introdujeron una pequeña revolución en el enfoque del análisis de series temporales, en sus trabajos sobre el comportamiento de la contaminación en la bahía de San Francisco, con el propósito de establecer mejores mecanismos de pronóstico y control. El libro (publicado en 1976) en el que describen su metodología, se convirtió rápidamente en un clásico, y sus procedimientos se utilizan ampliamente desde entonces en diferentes ramas de la ciencia, conociéndose como modelos ARIMA y también como **modelos Box-Jenkins**.

Para este tipo de modelos, el primer paso consiste en convertir nuestra serie de observaciones en una **serie estacionaria**, que es aquella en la que ni la media, ni la varianza, ni las autocorrelaciones dependen del tiempo. Una vez "estabilizada" la serie mediante las transformaciones adecuadas, se procede a estudiar la presencia de regularidades en la serie, para identificar un posible modelo matemático. Para ello se calcula la función de autocorrelación simple y parcial, y se compara su forma con un catálogo de patrones gráficos, que son típicos de los diferentes modelos propuestos, seleccionando el modelo que más se adecue a la forma de las funciones de autocorrelación que hemos obtenido con nuestros datos.

Una vez elegida la forma del modelo, se estiman los coeficientes del mismo, y finalmente se procede a efectuar un análisis de los residuos (diferencia entre el valor realmente observado y el valor previsto por el modelo), con el fin de comprobar si el ajuste del modelo a nuestros datos es adecuado. Si no lo fuera repetimos el proceso buscando otros modelos. Una vez determinado un modelo suficientemente válido, sobre la serie estacionaria, procedemos a deshacer la transformación inicialmente efectuada para estabilizar la serie, y ahora comprobamos si los pronósticos del modelo son adecuados con nuestros datos, volviendo a comenzar la búsqueda de otro modelo si no fuera el caso. Puede, por tanto, tratarse de un proceso iterativo de mejora del modelo.

En el modelo, cada valor tomado por la variable en un instante dado, está influido por los valores de la variable en momentos anteriores, y se expresa como una relación lineal, función de:

1. Valores recientes de la variable.
2. Ruidos en valores recientes de la variable.
3. Valores remotos de la variable.

4. Ruidos en valores remotos de la variable.

En resumen, para la búsqueda del modelo, según la metodología de Box-Jenkins, se exigen los siguientes pasos:

- **Identificación del Modelo:** La estabilidad de la serie de datos, la variación y la tendencia se obtienen a partir del cálculo de los valores de correlación simple y de la autocorrelación parcial FAS y FAP respectivamente. Este trabajo permite realizar el paso de determinación del modelo.
- **Determinación de los parámetros de estimación:** La determinación de la autocorrelación de la serie permite identificar el nivel de ruido de esta y consecuentemente determinar los parámetros de estimación.
- **Selección de un modelo ARIMA apropiado:** Durante el paso de estimación de los parámetros, el chequeo de los residuales determinará la selección del modelo ARIMA adecuado.
- **Predicción en base al modelo apropiado:** Finalmente el proceso de predicción supone la aplicación sobre la serie de datos y de sus regresores del modelo ARIMA determinado anteriormente.

Una representación gráfica de la metodología Box-Jenkins para el ajuste del modelo se muestra a continuación:

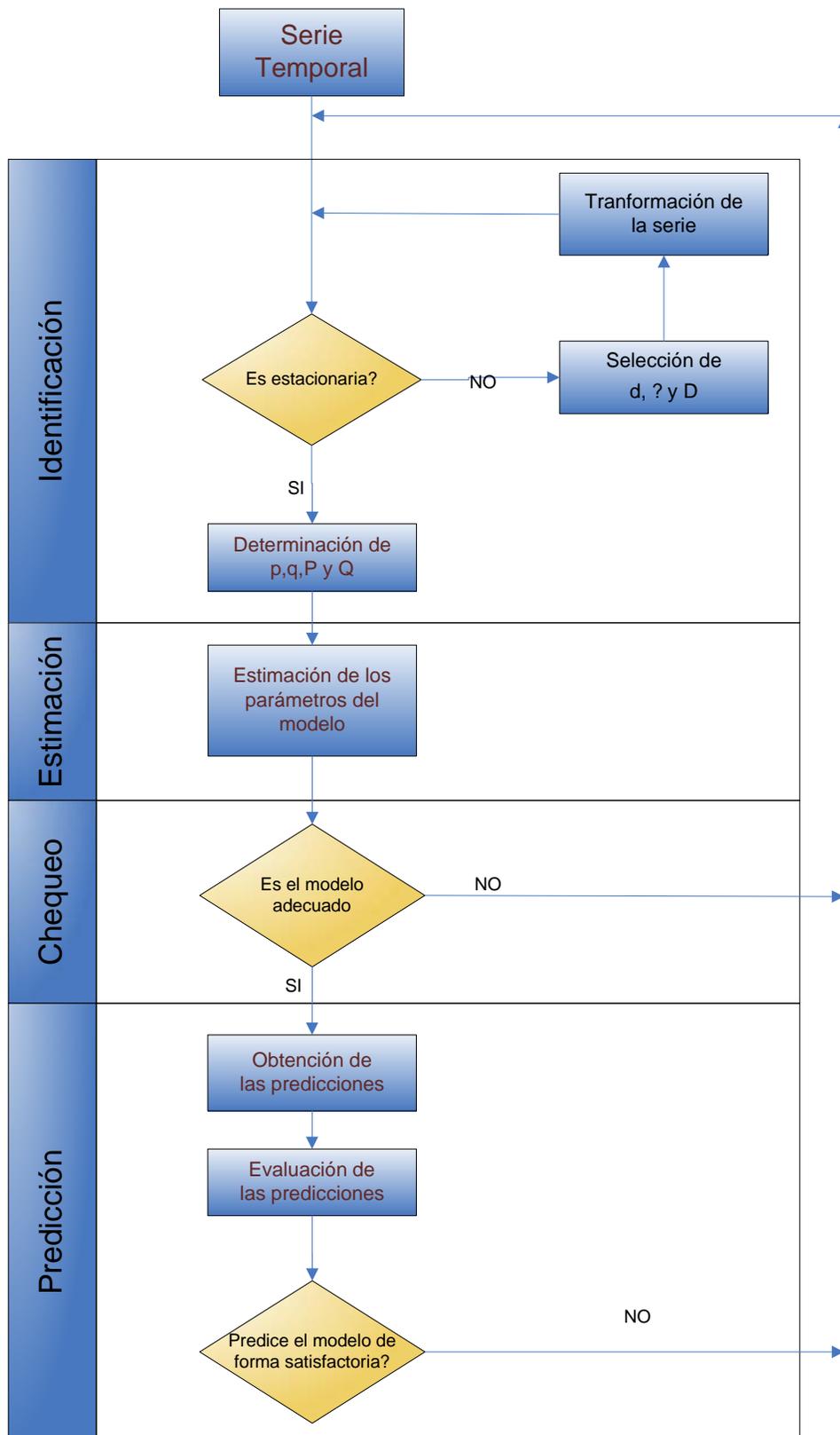


Figura 2. Etapas en la construcción de un modelos de series temporales. Peña (1989)

7. Estimación del modelo.

7.1 Identificación del modelo

Previo a la búsqueda y a la estimación del modelo, debemos analizar la serie temporal original observada y tomar en cuenta ciertas características:

- a) **Análisis de datos atípicos (outliers)**
- b) **Tendencia.** Como su nombre lo indica, es el crecimiento o decrecimiento de la serie 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.
- c) **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**.
- d) **Periodicidad,** es el intervalo de tiempo en la que se observa cada valor de la serie de los datos. Suele 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), horaria (se toma un dato cada hora), etc.

Para medir el grado de influencia de las observaciones pasadas en las futuras se utilizan las definiciones de las Funciones de Autocorrelación Simple (FAS) y Autocorrelación Parcial (FAP), mencionadas en el capítulo 5 y que a continuación se definen:

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 $Var(y_t) = Var(y_{t-j}) = \gamma_0$; 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. De manera intuitiva podemos señalar que la ergodicidad posibilita obtener estimadores consistentes de dichos parámetros. Sí 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 descrita en el capítulo 6. Identificaremos los procesos ARIMA susceptibles de generar la serie estacionaria o no, $\{y_t\}$. Para llevar a cabo este paso utilizaremos los instrumentos que caracterizan a los procesos AR, MA, ARMA y ARIMA: las funciones de autocorrelación simple y parcial cuya lectura es relevante para la búsqueda de los modelos y sus respectivos órdenes.

	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

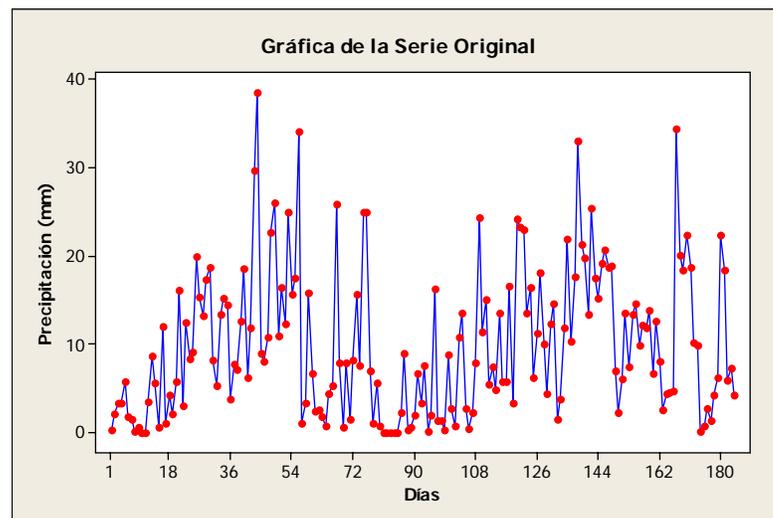
Tabla 3. Características de las FAS y FAP en los modelos MA(q), AR(P) y ARMA(p,q)

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.

7.1.1 Análisis de la serie original

Análisis de datos atípicos (outliers).

Para poder analizar la serie de datos se parte de su gráfica y sus estadísticos descriptivos, con el objetivo de determinar si existen o no datos atípicos (outliers) al comportamiento de la variable observada. La detección de los outliers es importante dado que i) permiten mejorar la comprensión de la serie al informarnos sobre el momento en que se producen y cómo actúan sobre la serie; ii) permiten mejorar la modelización y estimación de las serie; iii) mejorar nuestro análisis de intervención; iv) mejoran la calidad de las predicciones.

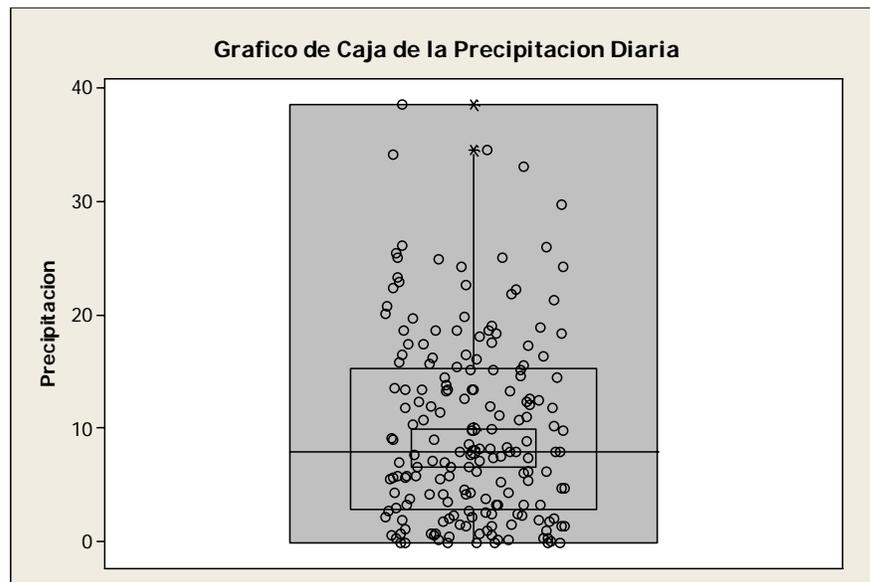


Gráfica 1. Evolución de los datos en el tiempo

Estadístico	Valor
Número de Datos (N)	184
Media	9.932
Desviación Estándar	8.238
Valor mínimo	0
Quartil 1	2.840
Mediana	7.940
Quartil 3	15.355
Valor máximo	38.440

Tabla 4. Estadísticos descriptivos de la serie original

A simple vista, en la gráfica 1 no se puede identificar los datos atípicos, pero se sospecha que existen algunos valores de precipitación muy altos, los cuales podrían ser considerados como tal. Para determinar cuáles de estos valores pueden ser considerados como atípicos, se hace uso del diagrama de cajas que está incorporado con el Minitab, y el cual se muestra a continuación:



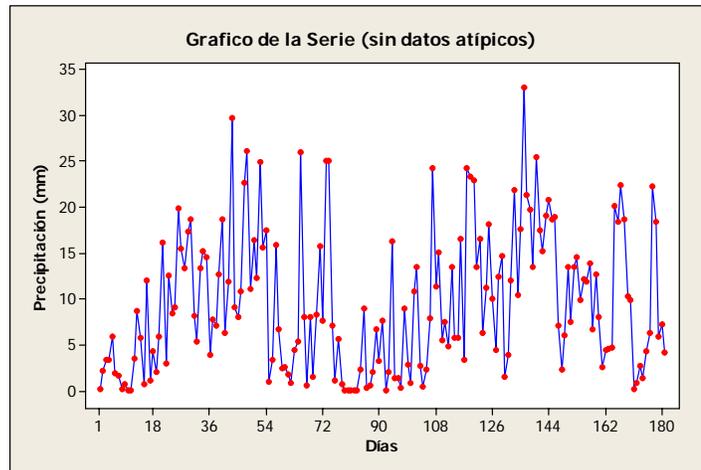
Gráfica 2. Diagrama de caja de la serie original

Al observar la figura se pueden considerar como datos atípicos aquellos que caen fuera de los bigotes del gráfico. A simple vista podemos considerar que se tienen 3 valores atípicos. Al auxiliarse del software SPSS para identificación de outlier, da como resultado los tres valores siguientes de precipitación (en mm): 38.44, 34.10 y 38.45 que corresponden a las fechas: 13 de junio, 25 de junio y 14 de octubre de 2003 respectivamente. Estos valores no serán tomados en

cuenta para las fases de identificación, chequeo y predicción, y luego serán retomados para su respectivo análisis.

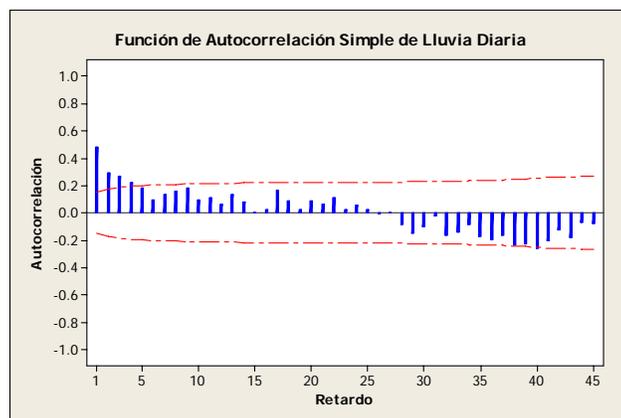
Análisis de la estacionariedad de la serie.

A continuación se muestra el gráfico de la serie, en la cual se han descartado los datos considerados como atípicos, y de los cuales se hará análisis posteriormente.

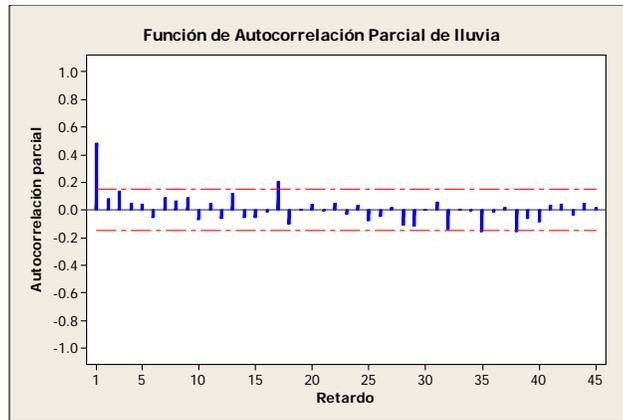


Gráfica 3. La serie original sin datos atípicos

Se observa que la serie, aparentemente, es estacionaria en varianza pero no en media. Esto puede verificarse al observar los gráficos de las autocorrelaciones simples y las autocorrelaciones parciales de la serie original. (Gráficas 4 y 5).



Gráfica 4. FAS de los datos originales



Gráfica 5. FAP de los datos originales

Para confirmar que la serie no presenta estacionariedad en media, se procede a realizar el test de raíces unitarias de Dickey-Fuller (DF)⁴.

Este test plantea las siguientes hipótesis:

Hipótesis nula Ho: existencia de raíz unitaria

Hipótesis alterna H1: no existencia de raíz unitaria.

De donde:

Si $t^* > DF$ no se rechaza H_0 , es decir existen raíces unitarias, por tanto la serie no es estacionaria.

Si $t^* < DF$ se rechaza H_0 , es decir no existen raíces unitarias, por tanto la serie es estacionaria.

Donde t^* son los valores críticos que Dickey-Fuller (1979) y MacKinnon (1994) construyeron mediante simulaciones de Montecarlo, y establecieron unas tablas especiales en las que la t es superior.

Para realizar el contraste de hipótesis, haciendo uso del programa Eviews, se obtienen los resultados que se muestran a continuación:

Null Hypothesis: LLUVIASO has a unit root
 Exogenous: None
 Lag Length: 8 (Automatic based on AIC, MAXLAG=13)

	t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic	-1.029669	0.2722
Test critical values:		
1% level	-2.578476	
5% level	-1.942688	
10% level	-1.615474	

⁴ Aznar, A., Trivez, Francisco, "Métodos de predicción en Economía, Tomo I, Capítulo 4, apartado 6.2, Ariel Economía, 1993.

*MacKinnon (1996) one-sided p-values.

Augmented Dickey-Fuller Test Equation

Dependent Variable: D(LLUVIASO)

Method: Least Squares

Date: 10/27/08 Time: 14:06

Sample (adjusted): 10 181

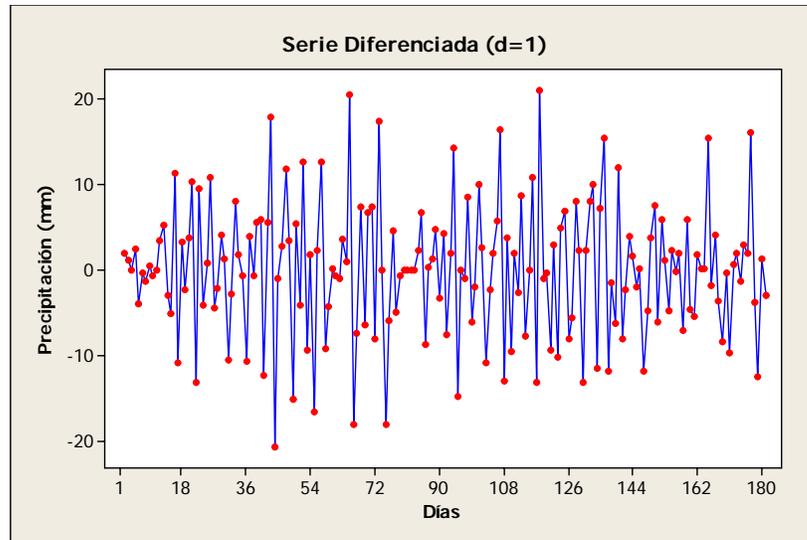
Included observations: 172 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
LLUVIASO(-1)	-0.050543	0.049087	-1.029669	0.3047
D(LLUVIASO(-1))	-0.494706	0.087883	-5.629162	0.0000
D(LLUVIASO(-2))	-0.450630	0.093979	-4.795027	0.0000
D(LLUVIASO(-3))	-0.295873	0.097298	-3.040893	0.0027
D(LLUVIASO(-4))	-0.257862	0.098085	-2.628962	0.0094
D(LLUVIASO(-5))	-0.202550	0.098094	-2.064861	0.0405
D(LLUVIASO(-6))	-0.293782	0.095802	-3.066569	0.0025
D(LLUVIASO(-7))	-0.210272	0.090000	-2.336361	0.0207
D(LLUVIASO(-8))	-0.147226	0.079529	-1.851227	0.0659
R-squared	0.274340	Mean dependent var		0.020465
Adjusted R-squared	0.238724	S.D. dependent var		7.892620
S.E. of regression	6.886399	Akaike info criterion		6.747881
Sum squared resid	7729.866	Schwarz criterion		6.912575
Log likelihood	-571.3177	Durbin-Watson stat		1.990967

Tabla 5. Resultados del Test de Dickey-Fuller Aumentado (ADF)

En los resultados puede verse que el valor del estadístico t (-1.029) es inferior a los valores críticos de la distribución DF por lo que no se puede rechazar la hipótesis nula de la existencia de una raíz unitaria y por lo tanto, la serie de lluvia diaria, no es estacionaria en media.

Para lograr la estacionariedad en media, aplicaremos una diferencia. Esto se muestra en el siguiente gráfico, donde se visualiza que la serie cumple la condición de estacionariedad en media:



Gráfica 6. Evolución de la serie con una diferencia

Para verificar que la serie es estacionaria en varianza, como lo muestra la gráfica 3, se aplicará el test de Levene que establece las siguientes hipótesis:

$H_0: \sigma_1 = \sigma_2 = \sigma_3 = \dots = \sigma_k$, es decir, las varianzas son iguales para k submuestras

$H_a: \sigma_i \neq \sigma_j$ para al menos un par (i,j)

La hipótesis nula se rechaza si la significancia del test de Levene es menor que 0.05 (valor de significación nominal), es decir, en este caso las varianzas son significativamente diferentes.

En esta aplicación, la serie se dividió en 6 submuestras, tomando un grupo por mes. La salida de SPSS se muestra en la tabla siguiente:

Test of Homogeneity of Variances

PrecDiariaSO			
Levene Statistic	df1	df2	Sig.
.318	5	175	.902

Tabla 6. Resultados del Test de Levene

En la tabla se observa que el estadístico de Levene con valor de 0.318 tiene una significancia de 0.902, por tanto no se puede rechazar la hipótesis nula y se concluye que las varianzas son iguales; por tanto se verifica que la serie es estacionaria en varianza.

Análisis de Estacionalidad

Debido a que la naturaleza de la serie que se está analizando es de carácter meteorológico, hace suponer que podría existir cierta tendencia estacional, lo cual se presenta en muchas ocasiones en este tipo de series, sin embargo debido a que se está analizando la serie de los meses con la precipitación más estable dentro del periodo lluvioso, podemos suponer que este componente podría no estar presente; sin embargo analizaremos la presencia de la estacionalidad usando la prueba de Kruskal-Wallis, a pesar que a simple vista del gráfico 6 no se observa dicha característica en la serie.

El contraste de Kruskal-Wallis sirve para establecer la presencia o ausencia del componente estacional en la serie. Concretamente se pretende contrastar la hipótesis nula:

H_0 : La serie no posee componente estacional

frente a la alternativa:

H_1 : La serie tiene componente estacional.

El estadístico que sirve de base para este contraste es el siguiente:

$$H = \frac{12}{T(T+1)} \left[\sum_{i=1}^s \frac{R_i^2}{T_i} \right] - 3(T+1), \text{ en donde:}$$

s: número de períodos estacionales dentro del período a evaluar.

T_i : número de observaciones que corresponden a la estación i -ésima, siendo $T = T_1 + T_2 + \dots + T_s$;

R_i : suma de los rangos que resultan de una ordenación de menor a mayor de los valores de la variable correspondientes a la estación i -ésima.

Bajo la hipótesis nula se obtiene que este estadístico, H , sigue una distribución χ^2 con $(s-1)$ grados de libertad. Adoptando un nivel de significación, ε , y siendo $\chi_{\varepsilon}^2(s-1)$ el punto crítico correspondiente a dicho nivel de significación, la regla de decisión es la siguiente:

- Si $H > \chi_{\varepsilon}^2(s-1)$ se rechaza la hipótesis nula. En este caso concluiremos que la serie tiene componente estacional.
- Si $H < \chi_{\varepsilon}^2(s-1)$ no se rechaza la hipótesis nula, pudiéndose concluir que la serie no tiene componente estacional.

Para realizar la prueba, se procedió de la siguiente manera:

- Se agruparon los datos en dos muestras, una con los datos de la serie de mayo a julio y la otra con los datos de agosto a octubre del 2003.
- Se realizó la corrida de la prueba Kruskal-Wallis para serie de datos diferenciada, usando el programa Minitab.

El resultado se muestra en la siguiente salida:

Kruskal-Wallis Test: PrecDiariaSODif versus Grupos_Diferenciada

Kruskal-Wallis Test on PrecDiariaSODif

Grupos_Diferenciada	N	Median	Ave Rank	Z
1	90	0.11500	92.1	0.42
2	90	0.05000	88.9	-0.42
Overall	180		90.5	

H = 0.18 DF = 1 P = 0.675

H = 0.18 DF = 1 P = 0.675 (adjusted for ties)

Las medianas de las dos muestras en el caso de la serie diferenciada fueron: 0.12 y 0.05. El valor z para la muestra 1 es 0.42 y el valor para la muestra 2 es de -0.42, siendo ambos iguales en valor absoluto, por lo que el rango medio para las dos muestras difiere en igual magnitud que el rango medio de todas las observaciones.

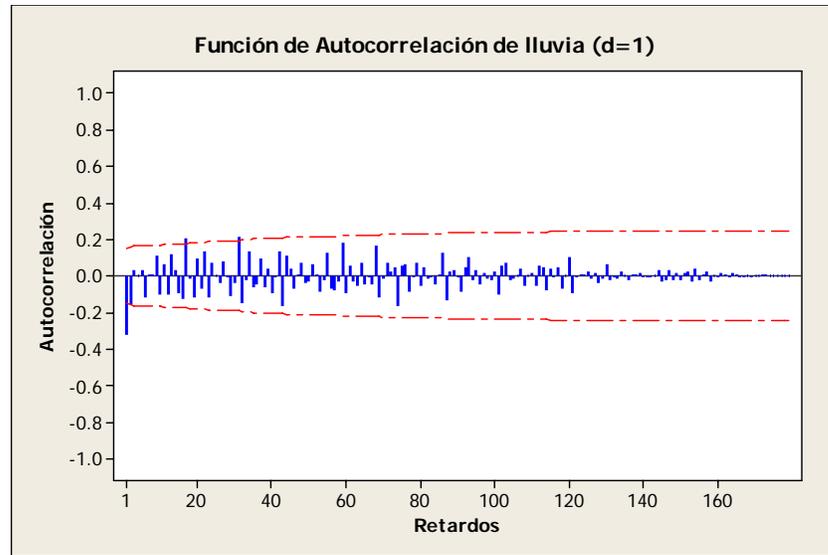
La prueba estadística (H) tiene un valor de p de 0.675, para el caso no ajustado o ajustado por empates, indicando que la hipótesis nula no puede ser rechazada.

A partir de estos resultados, podemos concluir que la serie no tiene componente estacional.

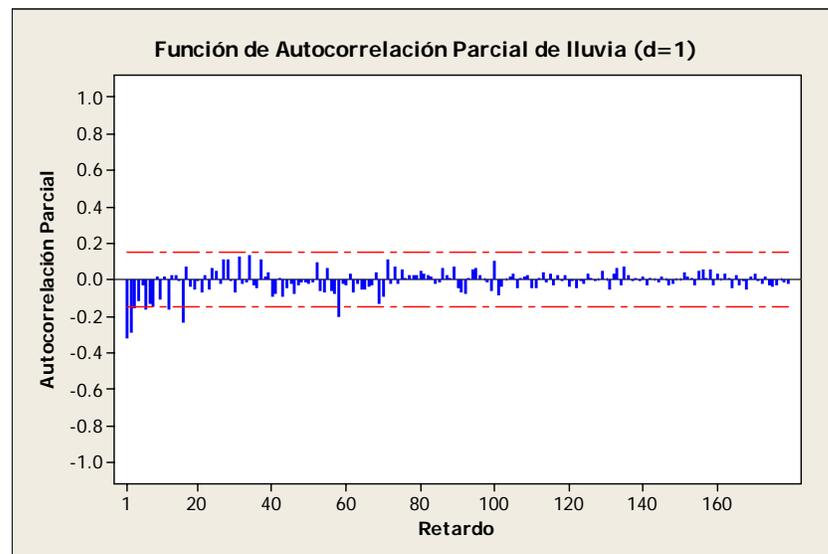
Después de haber analizado la serie, eliminando el ruido por valores atípicos, estableciendo la estacionariedad en media y en varianza y verificando que no posee componente estacional, podemos ahora pasar a la búsqueda de un modelo adecuado a la serie.

7.1.2 Identificación de los posibles modelos.

El primer paso en la identificación de un posible modelo, es elaborar los gráficos de las funciones de autocorrelación muestral simple y la función de autocorrelación muestral parcial.



Gráfica 7. Función de autocorrelación simple de la lluvia diaria



Gráfica 8. Función de autocorrelación parcial de la lluvia diaria

En la gráfica de Autocorrelación simple puede observarse que existen al menos 4 retardos significativos y además la gráfica puede observarse que existe una tendencia a anularse. Por otro lado en la gráfica de la Función de Autocorrelación parcial pueden observarse al menos 6 retardos significativos, y además en la gráfica puede observarse una tendencia lenta a anularse.

Siguiendo el análisis de las autocorrelaciones, se podría pensar que el modelo buscado es del tipo $ARIMA(p,d,q)$, el cual es expresado a través de la siguiente ecuación:

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

para la cual se tendrá que encontrar su orden. A partir, entonces, del análisis anterior los posibles modelos identificados son: ARIMA(1,1,2), ARIMA(1,1,3), ARIMA(1,1,4), ARIMA(2,1,3), ARIMA(2,1,4), ARIMA(4,1,3), ARIMA(5,1,3), ARIMA(5,1,4), ARIMA(6,1,3) y el ARIMA(6,1,4).

7.2 Estimación de los parámetros de los modelos identificados.

El siguiente paso será estimar los parámetros de los diferentes modelos identificados en el apartado anterior. Para esto haremos uso del programa SPSS.

Modelo Integrado de Medias Móviles, ARIMA (1,1,2)

La ecuación general de este modelo viene dada por:

$$w_t = \phi_1 w_{t-1} + u_t - \theta_1 u_{t-1} - \theta_2 u_{t-2}$$

La salida al correr el modelo en SPSS se muestra a continuación:

Parámetro		Estimación	SE	Prueba t	p-valor
Constante		0.033	0.087	0.378	0.706
AR	Lag 1	0.253	0.278	0.910	0.364
Diferencia		1			
MA	Lag 1	0.807	0.283	2.849	0.005
	Lag 2	0.069	0.227	0.303	0.762

Tabla 7. Parámetros del Modelo ARIMA(1,1,2)

Tomando los valores obtenidos en la tabla 5, el modelo quedaría de la siguiente manera:

$$w_t = u_t - 0.807u_{t-1}$$

Nótese que no se incluye el valor de la media (μ), dado que la constante resulta no ser significativa para un nivel de significación fijado $\varepsilon=0.05$, dado que el estadístico de prueba ($t=0.01$) es menor, en valor absoluto, que el valor en tablas ($t_{0.025, 181}=1.96$), de donde se concluye que la constante es cero y por tanto la media es cero, es decir; $\mu = \frac{\hat{\mu}}{(1 - \theta_1)} = 0$. Lo mismo ocurre

con los coeficientes para ϕ_1 y θ_2 .

El coeficiente θ_1 resulta ser significativo para la prueba t, es decir; el estadístico de prueba ($t=2.849$) es mayor que ($t_{0.025, 181}=1.96$). El mismo análisis se hará en los subsiguientes modelos

Modelo Integrado de Medias Móviles, ARIMA (1,1,3)

La ecuación general de este modelo viene dada por:

$$w_t = \phi_1 w_{t-1} + u_t - \theta_1 u_{t-1} - \theta_2 u_{t-2} - \theta_3 u_{t-3}$$

La salida al correr el modelo en SPSS se muestra a continuación:

Parámetro		Estimación	DE	Prueba t	p-valor
Constante		0.033	0.086	0.382	0.703
AR	Lag 1	-0.943	0.047	-20.260	0.000
Diferencia		1			
MA	Lag 1	-0.435	3.816	-0.114	0.909
	Lag 2	0.836	2.150	0.389	0.698
	Lag 3	0.271	1.048	0.259	0.796

Tabla 8. Parámetros del Modelo ARIMA(1,1,3)

Quedando el modelo de la siguiente manera:

$$w_t = -0.943w_{t-1} + u_t$$

Modelo Integrado de Medias Móviles, ARIMA (1,1,4)

La ecuación general de este modelo viene dada por:

$$w_t = \phi_1 w_{t-1} + u_t - \theta_1 u_{t-1} - \theta_2 u_{t-2} - \theta_3 u_{t-3} - \theta_4 u_{t-4}$$

La salida al correr el modelo en SPSS se muestra a continuación:

Parámetro		Estimación	DE	Prueba t	p-valor
Constante		0.033	0.088	0.369	0.713
AR	Lag 1	-0.912	0.051	-17.933	0.000
Diferencia		1			
MA	Lag 1	-0.393	0.464	-0.848	0.397
	Lag 2	0.759	0.289	2.626	0.009
	Lag 3	0.229	0.116	1.968	0.051
	Lag 4	0.075	0.085	0.882	0.379

Tabla 9. Parámetros del Modelo ARIMA(1,1,4)

Quedando el modelo de la siguiente manera:

$$w_t = -0.912w_{t-1} + u_t - 0.7596u_{t-2} - 0.229u_{t-3}$$

En este caso se incluye el coeficiente de θ_3 ya que es ligeramente mayor que el valor en tablas ($t_{0.025, 181}=1.96$).

Modelo Integrado de Medias Móviles, ARIMA (2,1,3)

La ecuación general de este modelo viene dada por:

$$w_t = \phi_1 w_{t-1} + \phi_2 w_{t-2} + u_t - \theta_1 u_{t-1} - \theta_2 u_{t-2} - \theta_3 u_{t-3}$$

La salida al correr el modelo en SPSS se muestra a continuación:

Parámetro		Estimación	DE	Prueba t	p-valor
Constante		0.029	0.029	0.994	0.322
AR	Lag 1	0.684	0.541	1.265	0.207
	Lag 2	0.162	0.406	0.398	0.691
Diferencia		1			
MA	Lag 1	1.260	1.252	1.006	0.316
	Lag 2	0.004	0.829	0.004	0.996
	Lag 3	-0.264	0.302	-0.872	0.384

Tabla 10. Parámetros del Modelo ARIMA(2,1,3)

Este modelo queda descartado desde este momento ya que ninguno de sus parámetros resulto significativo.

Modelo Integrado de Medias Móviles, ARIMA (2,1,4)

La ecuación general de este modelo viene dada por:

$$w_t = \phi_1 w_{t-1} + \phi_2 w_{t-2} + u_t - \theta_1 u_{t-1} - \theta_2 u_{t-2} - \theta_3 u_{t-3} - \theta_4 u_{t-4}$$

La salida al correr el modelo en SPSS se muestra a continuación:

Parámetro		Estimación	DE	Prueba t	p-valor
Constante		0.028	0.026	1.086	0.279
AR	Lag 1	-0.076	0.134	-0.569	0.570
	Lag 2	0.773	0.132	5.849	0.000
Diferencia		1			
MA	Lag 1	0.472	2.376	0.199	0.843
	Lag 2	1.094	1.240	0.882	0.379
	Lag 3	-0.466	1.361	-0.343	0.732
	Lag 4	-0.100	0.272	-0.366	0.715

Tabla 11. Parámetros del Modelo ARIMA(2,1,4)

En el cual, el modelo queda de la siguiente manera:

$$w_t = 0.773w_{t-2} + u_t$$

Los demás parámetros no resultaron significativos.

Modelo Integrado de Medias Móviles, ARIMA (4,1,3)

La ecuación general de este modelo viene dada por:

$$w_t = \phi_1 w_{t-1} + \phi_2 w_{t-2} + \phi_3 w_{t-3} + \phi_4 w_{t-4} + u_t - \theta_1 u_{t-1} - \theta_2 u_{t-2} - \theta_3 u_{t-3}$$

La salida al correr el modelo en SPSS se muestra a continuación:

Parámetro		Estimación	DE	Prueba t	p-valor
Constante		0.028	0.074	0.384	0.701
AR	Lag 1	-0.341	1.699	-0.200	0.841
	Lag 2	0.486	1.038	0.468	0.640
	Lag 3	-0.005	0.573	-0.009	0.993
	Lag 4	0.035	0.127	0.279	0.781
Diferencia		1			
MA	Lag 1	0.179	2.312	0.078	0.938
	Lag 2	0.942	1.419	0.664	0.508
	Lag 3	-0.237	1.504	-0.157	0.875

Tabla 12. Parámetros del Modelo ARIMA(4,1,3)

Este modelo también queda descartado, ya que ninguno de sus parámetros resulto significativo.

Modelo Integrado de Medias Móviles, ARIMA (5,1,3)

La ecuación general de este modelo viene dada por:

$$w_t = \phi_1 w_{t-1} + \phi_2 w_{t-2} + \phi_3 w_{t-3} + \phi_4 w_{t-4} + \phi_5 w_{t-5} + u_t - \theta_1 u_{t-1} - \theta_2 u_{t-2} - \theta_3 u_{t-3}$$

La salida al correr el modelo en SPSS se muestra a continuación:

Parámetro		Estimación	DE	Prueba t	p-valor
Constante		0.031	0.079	0.390	0.697
AR	Lag 1	-1.356	0.260	-5.204	0.000
	Lag 2	-0.161	0.228	-0.704	0.482
	Lag 3	0.320	0.193	1.660	0.099
	Lag 4	0.125	0.162	0.770	0.442
	Lag 5	0.100	0.097	1.034	0.303

Diferencia		1			
MA	Lag 1	-0.834	0.253	-3.289	0.001
	Lag 2	0.816	0.093	8.772	0.000
	Lag 3	0.723	0.207	3.497	0.001

Tabla 13. Parámetros del Modelo ARIMA(5,1,3)

En el cual, el modelo queda de la siguiente manera:

$$w_t = -1.356w_{t-1} + u_t + 0.834u_{t-1} - 0.816u_{t-2} - 0.723u_{t-3}$$

Los demás parámetros no resultaron significativos.

Modelo Integrado de Medias Móviles, ARIMA (5,1,4)

La ecuación general de este modelo viene dada por:

$$w_t = \phi_1 w_{t-1} + \phi_2 w_{t-2} + \phi_3 w_{t-3} + \phi_4 w_{t-4} + \phi_5 w_{t-5} + u_t - \theta_1 u_{t-1} - \theta_2 u_{t-2} - \theta_3 u_{t-3} - \theta_4 u_{t-4}$$

La salida al correr el modelo en SPSS se muestra a continuación:

Parámetro		Estimación	DE	Prueba t	p-valor
Constante		0.025	0.023	1.106	0.271
AR	Lag 1	-0.185	0.167	-1.108	0.269
	Lag 2	-0.196	0.139	-1.411	0.160
	Lag 3	-0.440	0.122	-3.597	0.000
	Lag 4	0.453	0.080	5.660	0.000
	Lag 5	0.139	0.092	1.516	0.131
Diferencia		1			
MA	Lag 1	0.359	0.226	1.589	0.114
	Lag 2	0.162	0.235	0.687	0.493
	Lag 3	-0.359	0.230	-1.561	0.120
	Lag 4	0.831	0.188	4.418	0.000

Tabla 14. Parámetros del Modelo ARIMA(5,1,4)

En el cual, el modelo queda de la siguiente manera:

$$w_t = -0.44w_{t-3} + 0.453w_{t-4} + u_t - 0.831u_{t-4}$$

Los demás parámetros no resultaron significativos.

Modelo Integrado de Medias Móviles, ARIMA (6,1,3)

La ecuación general de este modelo viene dada por:

$$w_t = \phi_1 w_{t-1} + \phi_2 w_{t-2} + \phi_3 w_{t-3} + \phi_4 w_{t-4} + \phi_5 w_{t-5} + \phi_6 w_{t-6} + u_t - \theta_1 u_{t-1} - \theta_2 u_{t-2} - \theta_3 u_{t-3}$$

La salida al correr el modelo en SPSS se muestra a continuación:

Parámetro		Estimación	DE	Prueba t	p-valor
Constante		0.031	0.092	0.340	0.734
AR	Lag 1	-0.146	0.664	-0.219	0.827
	Lag 2	0.499	0.399	1.251	0.213
	Lag 3	-0.101	0.226	-0.448	0.655
	Lag 4	0.016	0.103	0.160	0.873
	Lag 5	-0.037	0.095	-0.384	0.702
	Lag 6	-0.127	0.090	-1.410	0.160
Diferencia		1			
MA	Lag 1	0.382	0.667	0.573	0.567
	Lag 2	0.834	0.101	8.250	0.000
	Lag 3	-0.376	0.583	-0.645	0.520

Tabla 15. Parámetros del Modelo ARIMA(6,1,3)

En el cual, el modelo queda de la siguiente manera:

$$w_t = u_t - 0.834u_{t-2}$$

Los demás parámetros no resultaron significativos.

Modelo Integrado de Medias Móviles, ARIMA (6,1,4)

La ecuación general de este modelo viene dada por:

$$w_t = \phi_1 w_{t-1} + \phi_2 w_{t-2} + \phi_3 w_{t-3} + \phi_4 w_{t-4} + \phi_5 w_{t-5} + \phi_6 w_{t-6} + u_t - \theta_1 u_{t-1} - \theta_2 u_{t-2} - \theta_3 u_{t-3} - \theta_4 u_{t-4}$$

La salida al correr el modelo en SPSS se muestra a continuación:

Parámetro		Estimación	DE	Prueba t	p-valor
Constante		0.026	0.023	1.118	0.265
AR	Lag 1	-0.093	0.103	-0.907	0.366
	Lag 2	-0.234	0.104	-2.241	0.026
	Lag 3	-0.506	0.090	-5.619	0.000

	Lag 4	0.469	0.083	5.654	0.000
	Lag 5	0.118	0.087	1.353	0.178
	Lag 6	0.017	0.086	0.194	0.846
Diferencia		1			
MA	Lag 1	0.443	1.399	0.317	0.752
	Lag 2	0.065	1.992	0.032	0.974
	Lag 3	-0.447	1.914	-0.233	0.816
	Lag 4	0.932	1.293	0.721	0.472

Tabla 16. Parámetros del Modelo ARIMA(6,1,4)

En el cual, el modelo queda de la siguiente manera:

$$w_t = -0.234w_{t-2} - 0.506w_{t-3} + 0.469w_{t-4} + u_t$$

Los demás parámetros no resultaron significativos.

En resumen, después de revisar la significancia de los parámetros de los modelos identificados, se pueden expresar en los siguientes modelos reducidos junto con sus respectivas expresiones, excluyendo los que no tuvieron parámetros significativos:

No.	Modelo Identificado	Modelo Reducido	Expresión
1	ARIMA (1,1,2)	IMA (1,1)	$w_t = u_t - 0.807u_{t-1}$
2	ARIMA (1,1,3)	ARI (1,1)	$w_t = -0.943w_{t-1} + u_t$
3	ARIMA (1,1,4)	ARIMA (1,1,3)	$w_t = -0.912w_{t-1} + u_t - 0.759u_{t-2} - 0.229u_{t-3}$
4	ARIMA (2,1,4)	ARI (2,1)	$w_t = 0.773w_{t-2} + u_t$
5	ARIMA (5,1,3)	ARIMA (1,1,3)	$w_t = -1.356w_{t-1} + u_t + 0.834u_{t-1} - 0.816u_{t-2} - 0.723u_{t-3}$
6	ARIMA (5,1,4)	ARIMA (4,1,4)	$w_t = -0.44w_{t-3} + 0.453w_{t-4} + u_t - 0.831u_{t-4}$
7	ARIMA (6,1,3)	IMA (1,2)	$w_t = u_t - 0.834u_{t-2}$
8	ARIMA (6,1,4)	ARI (4,1)	$w_t = -0.234w_{t-2} - 0.506w_{t-3} + 0.469w_{t-4} + u_t$

Tabla 17. Modelos identificados, reducidos y sus expresiones generales que pasaran a la etapa de chequeo

7.3 Chequeo

En esta etapa veremos si alguno de los modelos identificados resulta adecuado para nuestro propósito. Esta etapa se centra en el análisis de los residuos del modelo.

Para establecer si un modelo resulta adecuado, debemos analizar:

1. Que las estimaciones de sus parámetros sean significativas, además de cumplir con las condiciones requeridas de estacionariedad e invertibilidad.
2. Que los residuos (estimaciones de los ruidos) se comporten como ruidos blancos, es decir, media cero, varianza constante y presenten incorrelación.
3. Que el modelo no omita parámetros relevantes
4. Que el modelo sea estable.

7.3.1 Significación de los parámetros.

Para realizar este análisis se realiza un contraste de significatividad individual de cada coeficiente, tomando como hipótesis nula la siguiente:

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

y como alternativa

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

Donde β_i pertenece al vector de estimadores de un modelo ARIMA, expresado como:

$\hat{\beta}' = (\hat{\delta}, \hat{\phi}_1, \dots, \hat{\phi}_p, \hat{\theta}_1, \dots, \hat{\theta}_q)$ y el cual se considera normal.

El estadístico de contraste que resulta en este caso es: $t = \frac{\hat{\beta}_i}{\hat{\sigma}_{\beta_i}}$, en el cual al fijar un cierto nivel de significancia ε , se puede concluir que el coeficiente β_i es significativo cuando el valor absoluto del estadístico $|t|$ sea superior al valor obtenido en tablas $t_{\varepsilon/2}(T-k)$. Cuando $\varepsilon=0.05$, el valor de $t_{0.025}(T-k)$, para muestras suficientemente grandes, se aproxima a 2.

La tabla 12 resume los modelos identificados con sus respectivas representaciones generales, en las cuales están expresados los coeficientes que resultaron ser significativos, luego de analizarlos bajo las condiciones señaladas en los apartados anteriores.

Se pasará ahora a revisar si los modelos identificados cumplen con la condición de estacionariedad e invertibilidad.

No.	Modelo	Estacionariedad	Invertibilidad
1	IMA (1,1)	Debido a que el modelo es un MA y una de las características de estos modelos es que siempre es estacionario	Debe cumplirse que $ \theta_1 < 1$, es decir $ 0.807 < 1$, el modelo es invertible
2	ARI (1,1)	Debe cumplirse que $ \phi_1 < 1$. En este caso $0.94 < 1$, luego el modelo es estacionario	Dado que el modelo es un AR, una de las características de estos modelos es que siempre son invertibles
3	ARIMA (1,1,3)	Se cumple que $ \phi_1 < 1$, ya que $ -0.912 < 1$, por tanto la serie es estacionaria	En este caso $ \theta_2 + \theta_3 < 1$, tenemos que $ 0.76 + 0.23 = 0.99 < 1$, el valor resultante es prácticamente 1. Además $ \theta_3 - \theta_2 < 1$ ya que $ 0.23 - 0.76 = 0.53 < 1$ y $\theta_3 = 0.23 < 1$. En este caso podemos decir que el proceso es no invertible.
4	ARI (2,1)	La serie es estacionaria dado que $ 0.77 < 1$.	Es invertible, ya que es un modelo AR
5	ARIMA (1,1,3)	El modelo no es estacionario ya que $ \phi_1 $ no es menor que 1, su valor es 1.36.	Para que sea invertible debe cumplir se que $\theta_1 + \theta_2 + \theta_3 < 1$, es decir $(-0.83 + 0.82 + 0.72) = 0.71 < 1$, podríamos decir que la serie es invertible.
6	ARIMA (4,1,4)	Para que la serie sea estacionaria debe cumplir que $ \phi_3 + \phi_4 < 1$, $ \phi_4 - \phi_3 < 1$ y $ \phi_4 < 1$, luego $ -0.44 + 0.45 = 0.01 < 1$, $ 0.45 - (-0.44) = 0.89 < 1$ y $ 0.45 < 1$, luego podemos concluir que es estacionaria	Para que sea invertible debe cumplir se que $ \theta_4 < 1$ entonces podemos concluir que es invertible ya que $ -0.83 < 1$
7	IMA (1,2)	Es estacionario ya que el modelo es un MA	Para que sea invertible debe cumplir que $ \theta_2 < 1$ y ya que $ -0.834 < 1$ se concluye que la serie es invertible
8	ARI (4,1)	Para que sea estacionario debe cumplir que $\phi_2 + \phi_3 + \phi_4 < 1$, en este caso $0.27 < 1$, podríamos concluir que el modelo es estacionario.	Es invertible puesto que es un modelo AR

Tabla 18. Análisis de estacionariedad e invertibilidad de los modelos identificados

De los modelos analizados, el numero 3 y el numero 5 en la tabla 18 no satisfacen las condiciones de estacionariedad e invertibilidad, por lo que se descartan, es decir, no se le aplicarán las siguientes etapas del chequeo.

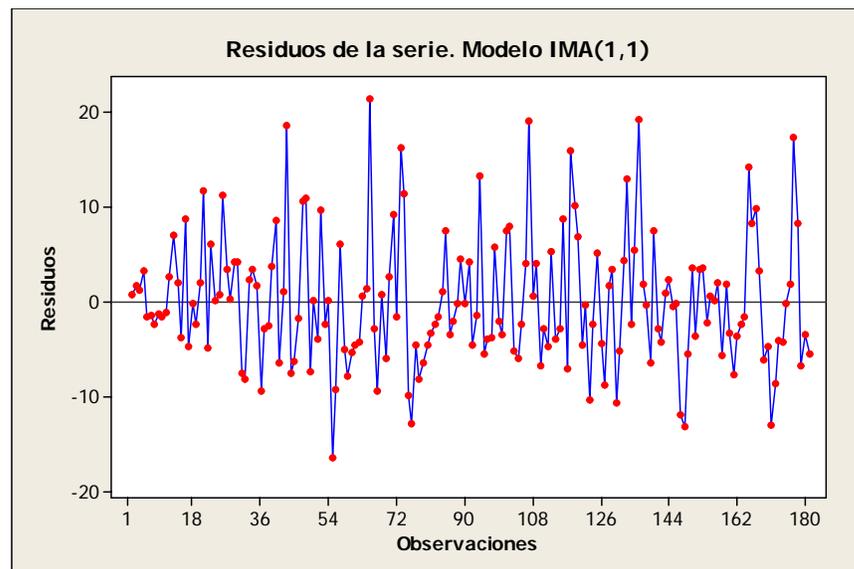
7.3.2 Análisis de los residuos.

Para llevar a cabo este análisis estudiaremos los residuos obtenidos en cada uno de los modelos identificados, los cuales deben aproximarse al comportamiento de un ruido blanco, es decir, deben presentar media cero, varianza constante y covarianza cero (incorrelación)

a) Análisis de Residuos del Modelo IMA (1,1)

¿Es la media de los residuos igual a cero?

Para analizar si la media de los residuos es cero, podemos comenzar representando gráficamente la serie de los residuos del modelo IMA (1,1). A partir de este gráfico se puede ver si los valores de los residuos giran alrededor de una media cero.



Gráfica 9. Residuos de la serie de lluvia diaria, IMA(1,1)

De esta última gráfica se puede concluir que la media es cero ($\mu = 0$), sin embargo, para tener mayor seguridad se procede a validar este hecho que parece cierto a simple vista a través de un contraste de significancia de la media de los residuos, esto se hace usando el estadístico:

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

El cual se distribuye como una t de Student con $T-1$ grados de libertad. Debido a que el número de observaciones (T) necesarias para aplicar la metodología de Box-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 significancia de la media de los residuos se adopta la siguiente regla de decisión: si $|t| < 2$ se concluye que la hipótesis de que la media de los residuos no difiere significativamente de cero es aceptada, en caso de que $|t| > 2$, se rechaza.

En la formula $t = \frac{\bar{u}_t}{\hat{\sigma}_{\bar{u}_t}}$, el valor de $\hat{\sigma}_{\bar{u}_t}$, se aproxima a través de la siguiente expresión:

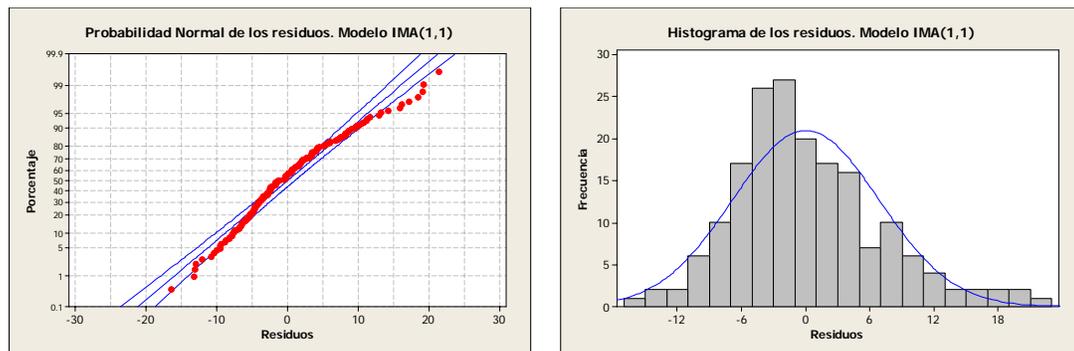
$$\hat{\sigma}_{\bar{u}_t} = \left(\frac{C_0}{T} \right)^{1/2}, \text{ donde } C_0 = \frac{\sum_{t=1}^T (u_t - \bar{u}_t)^2}{T}$$

Para el caso del modelo IMA (1,1), estos valores quedan de la siguiente manera:

$\bar{u}_t = 0.01$, $C_0 = 46.72$, luego $\hat{\sigma}_{\bar{u}_t} = 0.51$, entonces $t = \frac{0.01}{0.51} = 0.02$, el cual es un valor menor

a 2, por lo tanto concluimos que la media del modelo IMA (1,1) es cero.

Además con las graficas del histograma de los residuos y la de probabilidad normal de los residuos, se comprueba que estos se comportan de una manera aproximadamente normal.



Gráfica 10. Probabilidad normal e histograma de los residuos , IMA(1,1)

¿La Varianza de los residuos es constante?

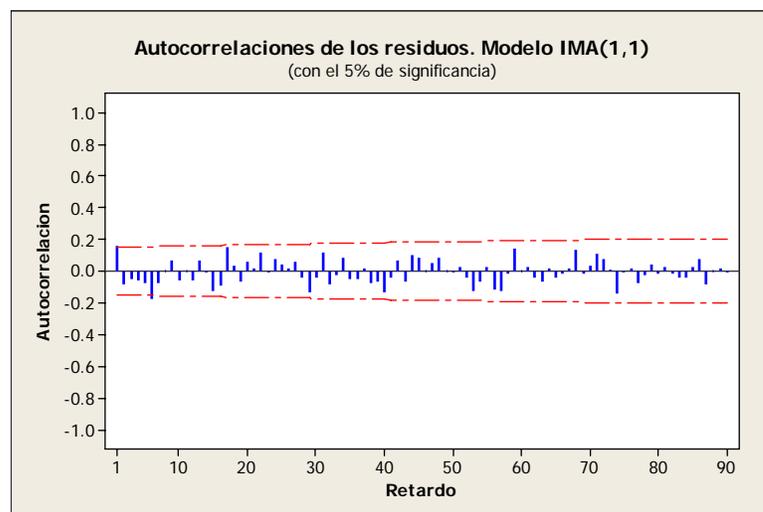
A partir del gráfico 8, se puede observar que la varianza es aproximadamente constante a lo largo del tiempo, con lo cual se cumple el requisito.

¿Se cumple la Incorrección de los residuos?

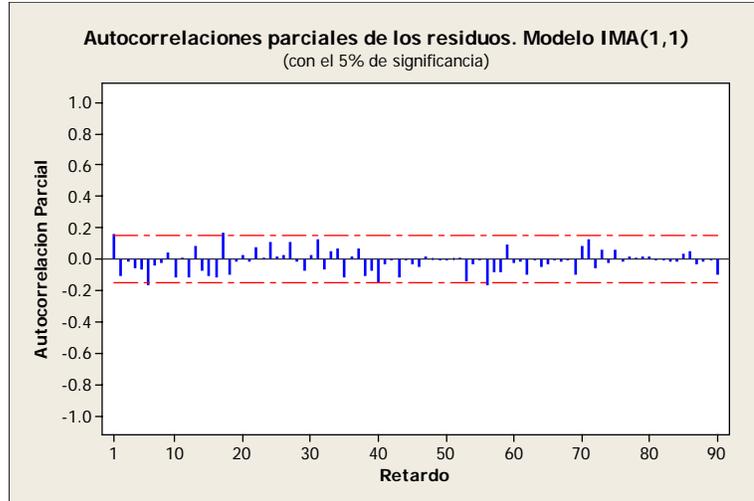
Para el establecer si no existe correlación entre los residuos obtenidos del modelo, debemos revisar el comportamiento de las autocorrelaciones simples y de las autocorrelaciones parciales.

Para las autocorrelaciones simples de los residuos, considerando un nivel de significación $\varepsilon = 0.05$ podremos aceptar la no significación de cada $\gamma_j(u_t)$ siempre que su valor se encuentre dentro del intervalo de $\pm 1.96(1/\sqrt{T})$. En el caso de las autocorrelaciones parciales de los residuos, se deberá cumplir que los coeficientes $\phi_{jj}(u_t)$ no sean significativos, es decir; en caso de que $\varepsilon = 0.05$, todos los coeficientes deben caer dentro del intervalo $\pm 1.96(1/\sqrt{T})$

Los gráficos de las autocorrelaciones muestrales y las autocorrelaciones parciales muestrales de los residuos para el modelo IMA (1,1), se muestran en los gráficos siguientes:



Gráfica 11. Autocorrelación simple de los residuos del IMA(1,1)

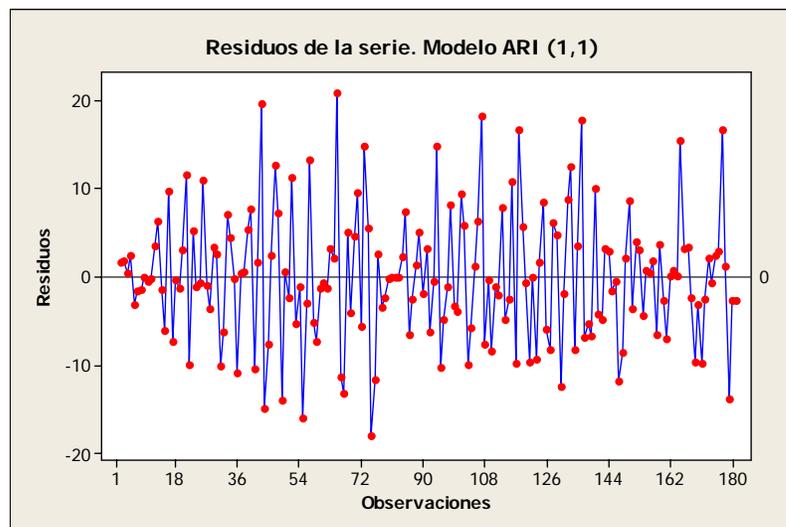


Gráfica 12. Autocorrelación parcial de los residuos del IMA(1,1)

Al observar las gráficas 9 y 10 se verifica que las autocorrelaciones simples y parciales obedecen aproximadamente al intervalo de aceptación propuesto, es decir, aproximadamente un 95% de las autocorrelaciones de los residuos se encuentran en el intervalo $\pm 1.96(1/\sqrt{T})$. Se concluye que las autocorrelaciones de los residuos son no significativas o que están muy próximas a cero para un nivel de significación $\varepsilon = 0.05$ y se acepta que no existe correlación entre los residuos del modelo.

b) Análisis de Residuos del Modelo ARI (1,1)

¿Es la media de los residuos igual a cero?



Gráfica 13. Residuos de la serie de lluvia diaria, ARI (1,1)

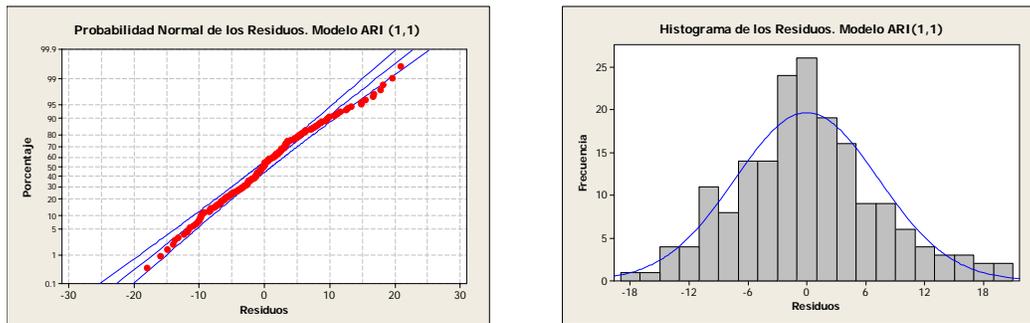
De la gráfica 11 se puede deducir que la media de los residuos es cero y al aplicar el

estadístico de prueba $t = \frac{\bar{u}_t}{\hat{\sigma}_{\bar{u}_t}}$, para el modelo, se satisface que $t < 2$,

Es decir;

$$\bar{u}_t = 0.002, C_0 = 53.24, \text{ luego } \hat{\sigma}_{\bar{u}_t} = 0.54, \text{ entonces } t = \frac{0.002}{0.54} = 0.004$$

Con lo que se verifica que la media de los residuos es cero.



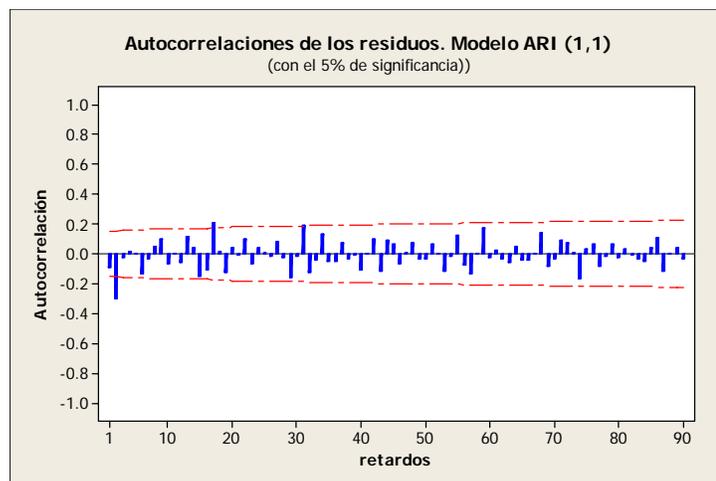
Gráfica 14. Probabilidad normal e histograma de los residuos , ARI (1,1)

Al observar la grafica 12, se comprueba que los residuos se comportan de una manera aproximadamente normal.

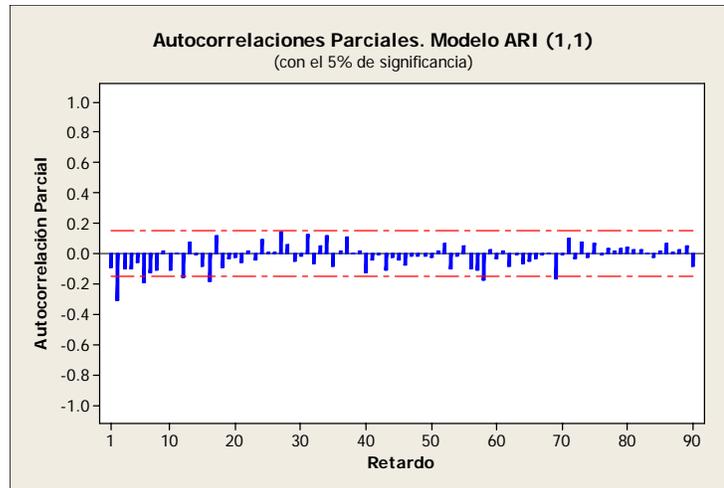
¿La Varianza de los residuos es constante?

A partir del gráfico 12, se puede observar que la varianza es aproximadamente constante a lo largo del tiempo, con lo cual se cumple el requisito.

¿Se cumple la In correlación de los residuos?



Gráfica 15. Autocorrelación simple de los residuos del ARI(1,1)



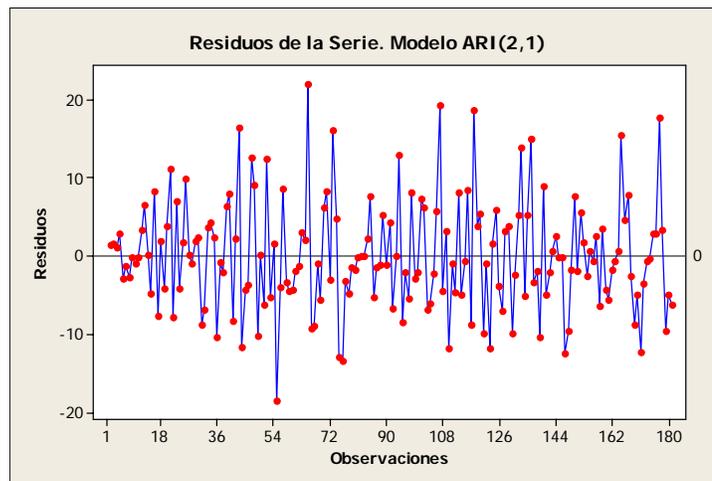
Gráfica 16. Autocorrelación parcial de los residuos del ARI(1,1)

Al observar las gráficas 13 y 14 se verifica que las autocorrelaciones simples y parciales obedecen aproximadamente al intervalo de aceptación propuesto, es decir, aproximadamente un 95% de las autocorrelaciones de los residuos se encuentran en el intervalo $\pm 1.96(1/\sqrt{T})$. Se concluye que las autocorrelaciones de los residuos son no significativas o que están muy próximas a cero para un nivel de significación $\varepsilon = 0.05$ y se acepta que no existe correlación entre los residuos del modelo.

c) Análisis de Residuos del Modelo ARI (2,1)

¿Es la media de los residuos igual a cero?

La gráfica de los residuos del modelo se muestra a continuación:



Gráfica 17. Residuos de la serie de lluvia diaria, ARI (1,1)

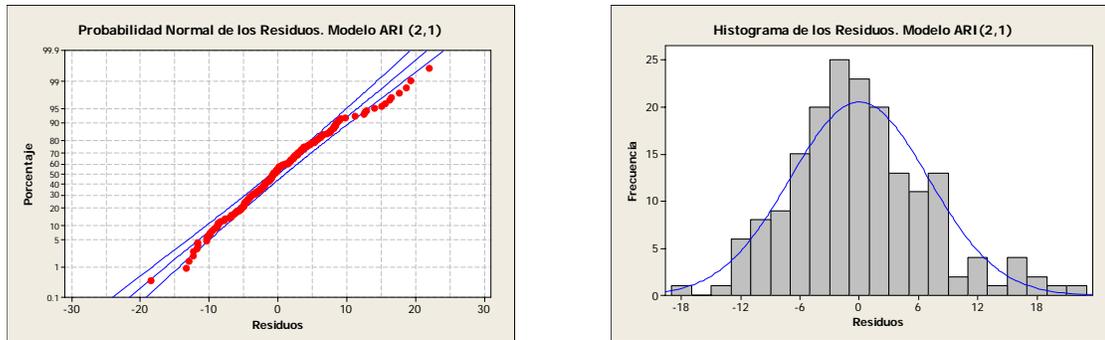
De la gráfica 15 se puede deducir que la media de los residuos es cero y al aplicar el

estadístico de prueba $t = \frac{\bar{u}_t}{\hat{\sigma}_{\bar{u}_t}}$, para el modelo, se satisface que $t < 2$,

Es decir;

$$\bar{u}_t = 0.005, C_0 = 48.69, \text{ luego } \hat{\sigma}_{\bar{u}_t} = 0.52, \text{ entonces } t = \frac{0.005}{0.52} = 0.010$$

Con lo que se verifica que la media de los residuos es cero.



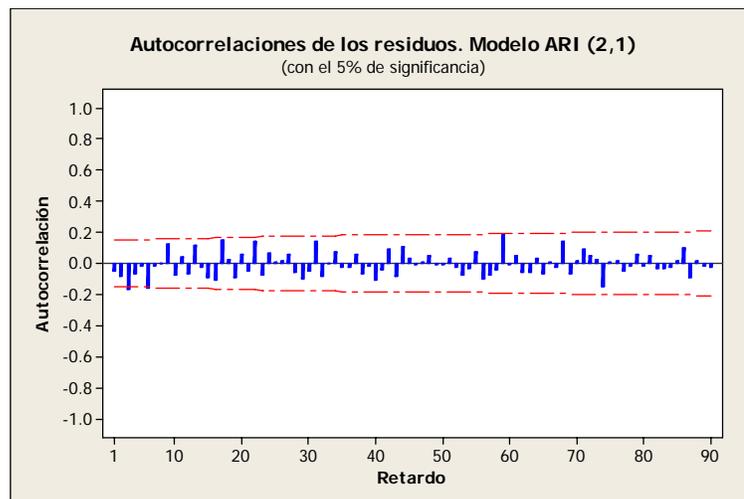
Gráfica 18. Probabilidad normal e histograma de los residuos. ARI(2,1)

Al observar la grafica 16, se comprueba que los residuos se comportan de una manera aproximadamente normal.

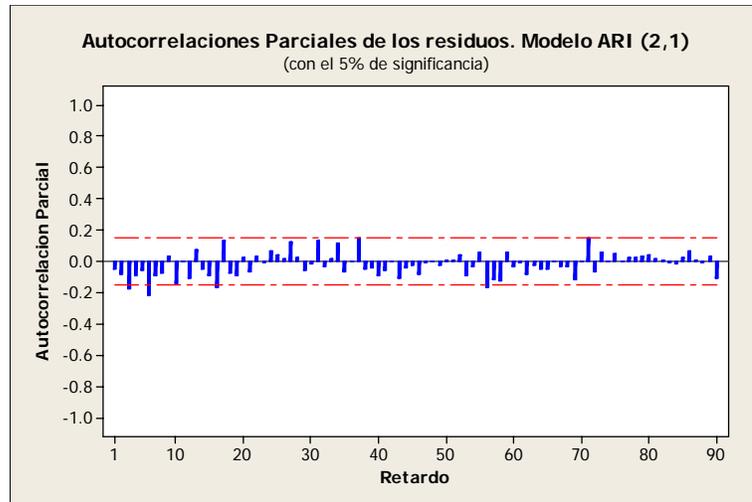
¿La varianza de los residuos es constante?

A partir del gráfico 15, se puede observar que la varianza es aproximadamente constante a lo largo del tiempo, con lo cual se cumple el requisito.

¿Se cumple la incorrelación de los residuos?



Gráfica 19. Autocorrelación simple de los residuos del ARI (2,1)



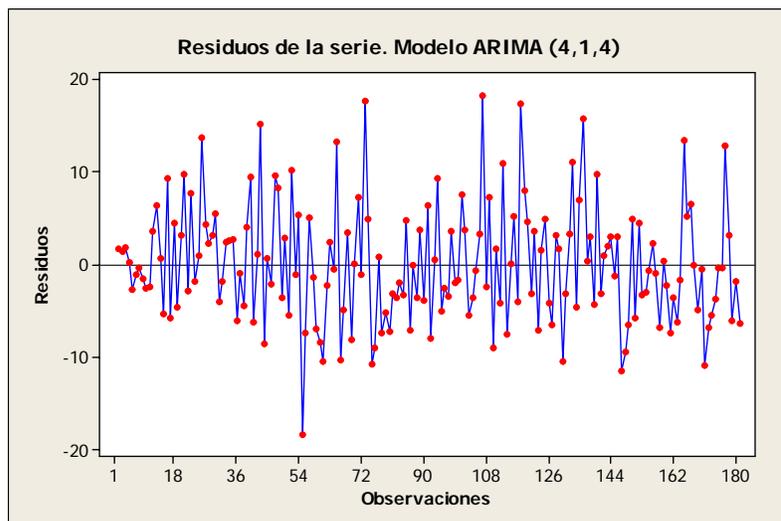
Gráfica 20. Autocorrelación parcial de los residuos del ARI (2,1)

Al observar las gráficas 17 y 18 se verifica que las autocorrelaciones simples y parciales obedecen aproximadamente al intervalo de aceptación propuesto, es decir, aproximadamente un 95% de las autocorrelaciones de los residuos se encuentran en el intervalo $\pm 1.96(1/\sqrt{T})$. Se concluye que las autocorrelaciones de los residuos son no significativas o que están muy próximas a cero para un nivel de significación $\varepsilon = 0.05$ y se acepta que no existe correlación entre los residuos del modelo.

d) Análisis de Residuos del Modelo ARIMA (4,1,4)

¿Es la media de los residuos igual a cero?

La gráfica de los residuos del modelo se muestra a continuación:



Gráfica 21. Residuos de la serie de lluvia diaria, ARIMA(4,1,4).

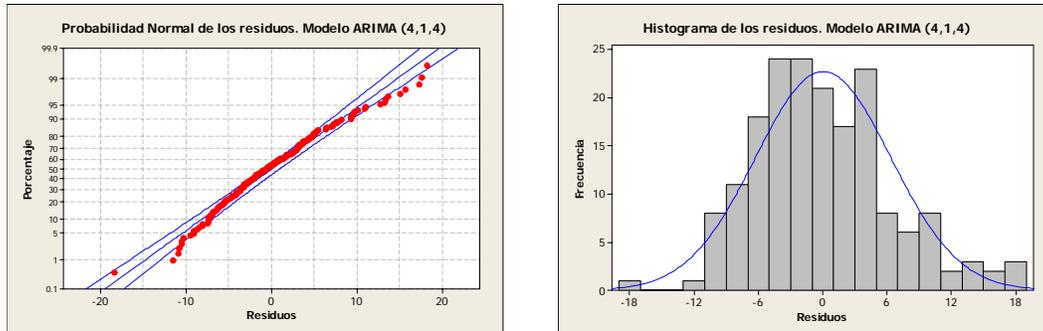
De la gráfica 19 se puede deducir que la media de los residuos es igual a cero y al aplicar el

estadístico de prueba $t = \frac{\bar{u}_t}{\hat{\sigma}_{\bar{u}_t}}$, para el modelo, satisface que $t < 2$,

Es decir;

$$\bar{u}_t = 0.04, C_0 = 39.75, \text{ luego } \hat{\sigma}_{\bar{u}_t} = 0.47, \text{ entonces } t = \frac{0.04}{0.47} = 0.085$$

Con lo que se verifica que la media de los residuos es igual a cero.



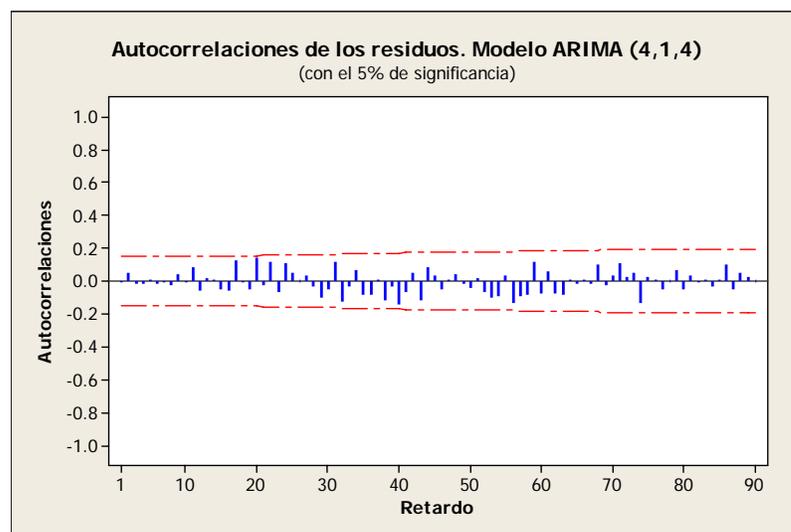
Gráfica 22. Probabilidad normal e histograma de los residuos. ARIMA(4,1,4)

Al observar la grafica 20, se comprueba que los residuos se comportan de una manera aproximadamente normal.

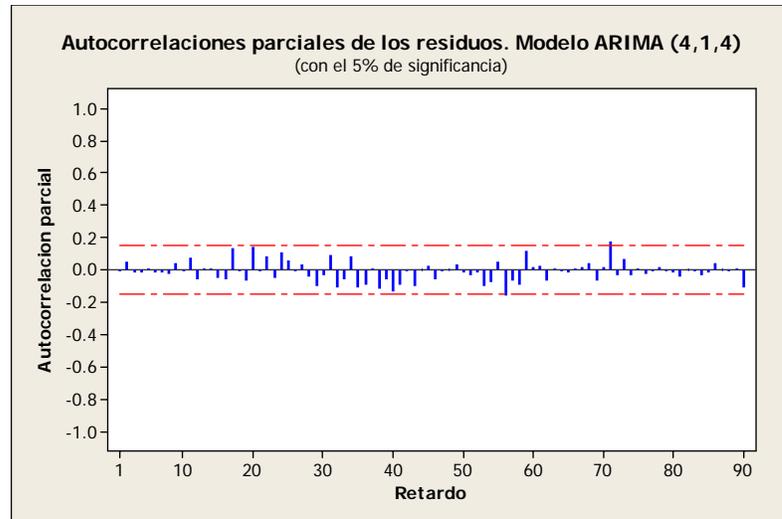
¿La varianza de los residuos es constante?

A partir del gráfico 19, se puede observar que la varianza es aproximadamente constante a lo largo del tiempo, con lo cual se cumple el requisito.

¿Se cumple la In correlación de los residuos?



Gráfica 23. Autocorrelación simple de los residuos del ARIMA (4,1,4)



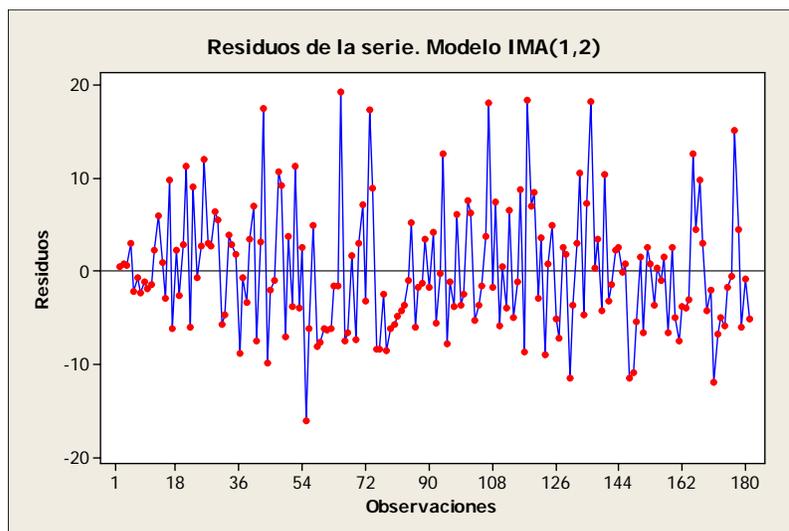
Gráfica 24. Autocorrelación parcial de los residuos del ARIMA (4,1,4)

Al observar las gráficas 21 y 22 se verifica que las autocorrelaciones simples y parciales obedecen aproximadamente al intervalo de aceptación propuesto, es decir, aproximadamente un 95% de las autocorrelaciones de los residuos se encuentran en el intervalo $\pm 1.96(1/\sqrt{T})$. Se concluye que las autocorrelaciones de los residuos son no significativas o que están muy próximas a cero para un nivel de significación $\varepsilon = 0.05$ y se acepta que no existe correlación entre los residuos del modelo.

e) Análisis de Residuos del Modelo IMA (1,2)

¿Es la media de los residuos igual a cero?

La gráfica de los residuos del modelo se muestra a continuación:



Gráfica 25. Residuos de la serie de lluvia diaria, IMA(1,2).

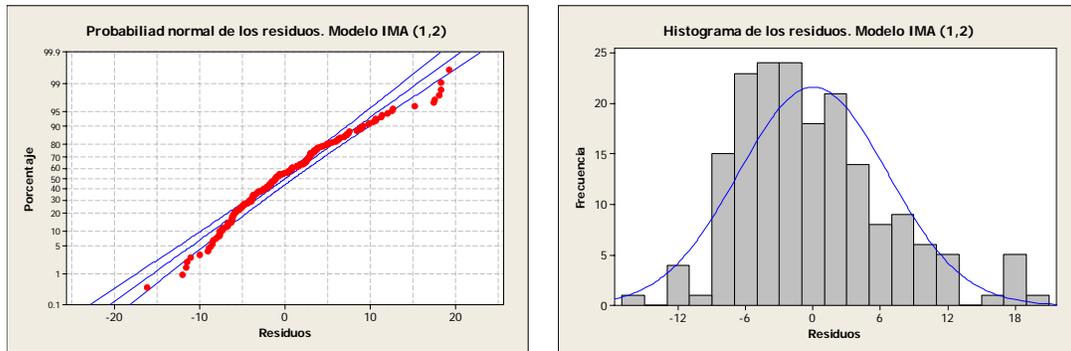
De la gráfica 23 se puede deducir que la media de los residuos es igual a cero y al aplicar el

estadístico de prueba $t = \frac{\bar{u}_t}{\hat{\sigma}_{\bar{u}_t}}$, para el modelo, satisface que $t < 2$,

Es decir;

$$\bar{u}_t = 0.016, C_0 = 43.9, \text{ luego } \hat{\sigma}_{\bar{u}_t} = 0.49, \text{ entonces } t = \frac{0.016}{0.49} = 0.032$$

Con lo que se verifica que la media de los residuos es igual a cero.



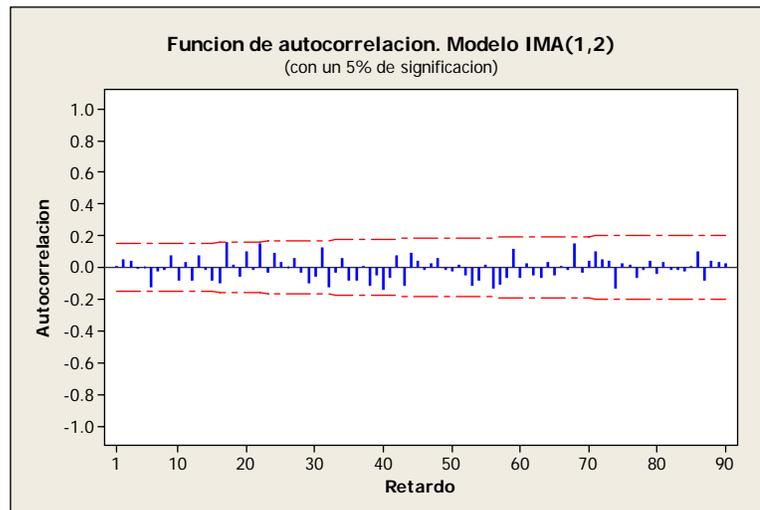
Gráfica 26. Probabilidad normal e histograma de los residuos. IMA(1,2)

Al observar la grafica 24, se comprueba que los residuos se comportan de una manera aproximadamente normal.

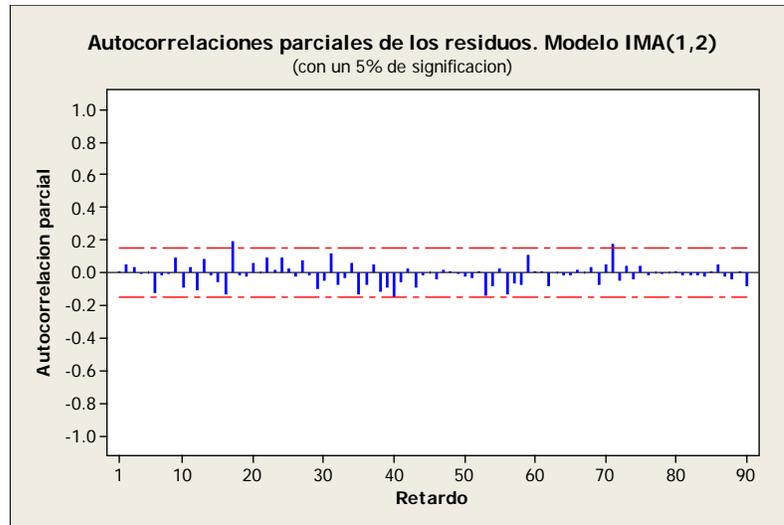
¿La varianza de los residuos es constante?

A partir del gráfico 23, se puede observar que la varianza es aproximadamente constante a lo largo del tiempo, con lo cual se cumple el requisito.

¿Se cumple la incorrelación de los residuos?



Gráfica 27. Autocorrelación simple de los residuos del IMA (1,2)



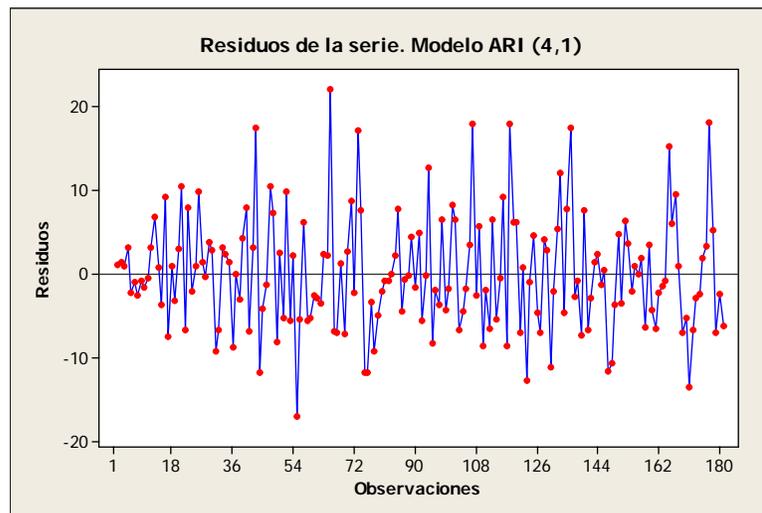
Gráfica 28. Autocorrelación parcial de los residuos del IMA (1,2)

Al observar las gráficas 25 y 26 se verifica que las autocorrelaciones simples y parciales obedecen aproximadamente al intervalo de aceptación propuesto, es decir, aproximadamente un 95% de las autocorrelaciones de los residuos se encuentran en el intervalo $\pm 1.96(1/\sqrt{T})$. Se concluye que las autocorrelaciones de los residuos son no significativas o que están muy próximas a cero para un nivel de significación $\varepsilon = 0.05$ y se acepta que no existe correlación entre los residuos del modelo.

f) Análisis de Residuos del Modelo ARI (4,1)

¿Es la media de los residuos igual a cero?

La gráfica de los residuos del modelo se muestra a continuación:



Gráfica 29. Residuos de la serie de lluvia diaria, ARI(4,1)

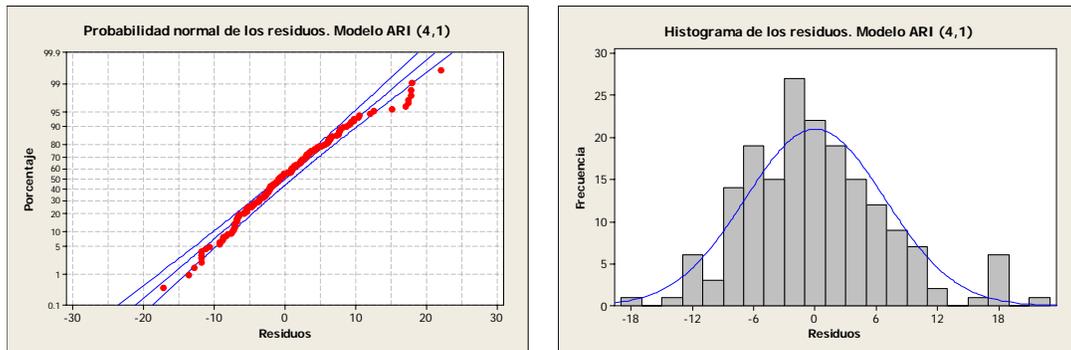
De la gráfica 27 se puede deducir que la media de los residuos es igual a cero y al aplicar el

estadístico de prueba $t = \frac{\bar{u}_t}{\hat{\sigma}_{\bar{u}_t}}$, para el modelo, satisface que $t < 2$,

Es decir;

$$\bar{u}_t = 0.008, C_0 = 46.63, \text{ luego } \hat{\sigma}_{\bar{u}_t} = 0.51, \text{ entonces } t = \frac{0.008}{0.51} = 0.016$$

Con lo que se verifica que la media de los residuos es igual a cero.



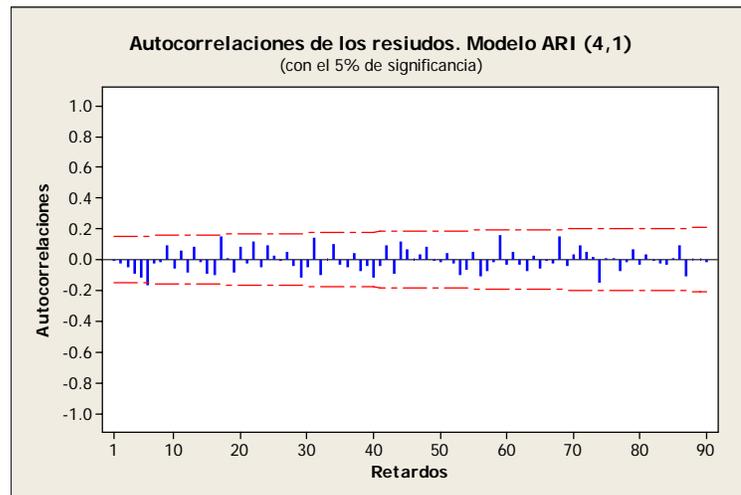
Gráfica 30. Probabilidad normal e histograma de los residuos. ARI(4,1)

Al observar la grafica 28, se comprueba que los residuos se comportan de una manera aproximadamente normal.

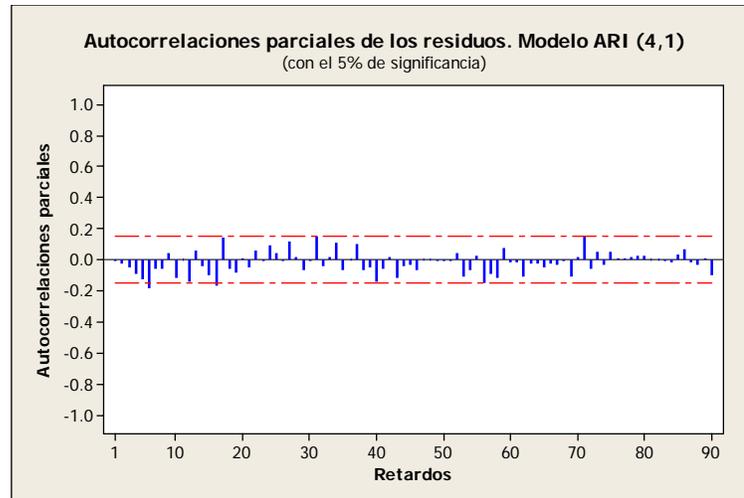
¿La varianza de los residuos es constante?

A partir del gráfico 27, se puede observar que la varianza es aproximadamente constante a lo largo del tiempo, con lo cual se cumple el requisito.

¿Se cumple la incorrelación de los residuos?



Gráfica 31. Autocorrelación simple de los residuos del ARI (4,1)



Gráfica 32. Autocorrelación parcial de los residuos del ARI (4,1)

Al observar las gráficas 29 y 30 se verifica que las autocorrelaciones simples y parciales obedecen aproximadamente al intervalo de aceptación propuesto, es decir, aproximadamente un 95% de las autocorrelaciones de los residuos se encuentran en el intervalo $\pm 1.96(1/\sqrt{T})$. Se concluye que las autocorrelaciones de los residuos son no significativas o que están muy próximas a cero para un nivel de significación $\varepsilon = 0.05$ y se acepta que no existe correlación entre los residuos del modelo.

7.3.3 Análisis de Sobreajuste

Según Box y Jenkins (1976) este proceso consiste en incluir parámetros adicionales en el modelo identificado (sobreajuste) para comprobar si éstos modelos con un mayor número de parámetros resultan apropiados. Es decir, una vez identificado el modelo ARIMA (p, d, q) que parece adecuado, el sobreajuste se aplica estimando los modelos ARIMA (p + 1, d, q) y ARIMA (p, d, q + 1) y comprobando, si los parámetros adicionales son significativos.

En este estudio, el análisis de sobreajuste fue abordado en la etapa de identificación del modelo, dando como resultado los modelos planteados en la tabla 13. Dichos modelos incluyen solamente los parámetros significativos, por tanto, en los modelos estudiados no se omiten parámetros relevantes.

Sin embargo al volver a correr los modelos, ahora en sus formas reducidas, se descartara el modelo identificado como ARIMA (4,1,4), ya que los parámetros de su parte de medias móviles no fueron significativos, lo cual lo convierte en un modelo ARI (4,1) que ya fue definido también como un modelo a ser considerado.

Hasta esta etapa, los modelos que pasarán a revisarse en su estabilidad son: IMA(1,1), ARI (1,1), ARI (2,1), IMA (1,2) y el ARI(4,1).

7.3.4 Análisis de Estabilidad

Al analizar la estabilidad del modelo, se verá la consistencia de los valores de los parámetros en todo el período muestral seleccionado, con esto se evitan problemas en la etapa de predicción.

Para analizar la estabilidad se pueden tomar submuestras, de períodos muestrales distintos. Una manera consiste en dividir la serie en dos partes, estimar sus parámetros y luego compararlos con los parámetros de los modelos estimados en las etapas previas. Diremos que el modelo es estable cuando la diferencia de los parámetros de las submuestras no sea muy grande en relación a los parámetros del modelo estimado con la muestra completa.

En la serie de lluvia diaria, se tienen observaciones correspondientes a 184 días de lluvia correspondientes a los meses de mayo a octubre del 2003, sin embargo se trabajará con 181 datos, dado que en la etapa de detección de valores atípicos fueron encontrados 3 valores, los cuales se han descartado para el análisis y posteriormente serán considerados al hacer análisis de intervención. En este caso se ha dividido las observaciones en dos partes, 90 para la primera submuestra y 91 para la segunda.

Los resultados de las estimaciones son los siguientes:

a. Modelo IMA (1,1) Estimación de parámetros de 2 submuestras.

Parámetro	Coficiente IMA (1,1)	Coficiente Muestra 1	Coficiente Muestra 2	Diferencia Muestra 1	Diferencia Muestra 2
MA 1	0.692	0.714	0.585	-0.022	0.107

Tabla 19. Diferencias de los Parámetros del modelo IMA (1,1) y según muestras 1 y 2

b. Modelo ARI (1,1) Estimación de parámetros de 2 submuestras.

Parámetro	Coficiente ARI (1,1)	Coficiente Muestra 1	Coficiente Muestra 2	Diferencia Muestra 1	Diferencia Muestra 2
AR 1	-0.319	-0.331	-0.299	-0.012	-0.02

Tabla 20. Diferencias de los Parámetros del modelo ARI (1,1), según muestras 1 y 2

c. Modelo ARI (2,1) Estimación de parámetros de 2 submuestras.

Parámetro	Coficiente	Coficiente	Coficiente	Diferencia	Diferencia
	ARI (2,1)	Muestra 1	Muestra 2	Muestra 1	Muestra 2
AR 1	-0.412	-0.450	-0.368	0.038	-0.043
AR 2	-0.289	-0.354	-0.227	0.065	-0.062

Tabla 21. Diferencias de los Parámetros del modelo ARI (2,1), según muestras 1 y 2

d. Modelo IMA (1,2) Estimación de parámetros de 2 submuestras.

Parámetro	Coficiente	Coficiente	Coficiente	Diferencia	Diferencia
	IMA (1,2)	Muestra 1	Muestra 2	Muestra 1	Muestra 2
MA 1	0.560	0.569	0.556	-0.009	0.004
MA 2	0.257	0.215	0.288	0.042	-0.031

Tabla 22. Diferencias de los Parámetros del modelo IMA (1,2), según muestras 1 y 2

e. Modelo ARI (4,1) Estimación de parámetros de 2 submuestras.

Parámetro	Coficiente	Coficiente	Coficiente	Diferencia	Diferencia
	ARI (4,1)	Muestra 1	Muestra 2	Muestra 1	Muestra 2
AR 1	-0.482	-0.554	-0.407	0.072	-0.075
AR 2	-0.403	-0.536	-0.292	0.133	-0.111
AR 3	-0.222	-0.290	-0.173	0.068	-0.049
AR 4	-0.128	-0.230	-0.013	0.102	-0.115

Tabla 23. Diferencias de los Parámetros del modelo ARI (4,1), según muestras 1 y 2

Se puede observar que las diferencias entre los parámetros del modelo de la serie completa y los parámetros de las submuestras no son significativas para ninguno de los modelos analizados, por lo que los cinco modelos: IMA (1,1), ARI (1,1), ARI (2,1), IMA (1,2) y el ARI (4,1) pueden considerarse estables y debería ser tomados en cuenta en la etapa de predicción.

Por otra parte, al relacionar el resultado de esta prueba con el análisis de sobreajuste del modelo de la etapa anterior, tomando en cuenta el **Principio de Parsimonia**, que establece que “entre los modelos de similares comportamientos debemos elegir siempre el modelo más sencillo (con menos parámetros)”, podemos concluir que los modelos candidatos serán: IMA (1,1), ARI (1,1), ARI (2,1) y el IMA (1,2), descartando el ARI (4,1)

7.4 Predicción

Según la metodología de Box-Jenkins, esta es la última etapa en el análisis de series temporales, en la cual podemos utilizar los modelos seleccionados para hacer predicciones de lluvia diaria dentro del período analizado. En este punto es bueno recordar que se está analizando la serie de datos de mayo a octubre del 2003. Entonces, a partir de los modelos seleccionados se podrá hacer predicciones para días que están dentro de estos meses y para años con régimen de lluvia normal⁵.

7.4.1 El Predictor óptimo

En general, el objetivo principal del análisis de series temporales es predecir valores futuros de $\{y_t\}$. En este sentido siempre conocemos las observaciones hasta el periodo T , y se desea obtener la predicción de y_{T+l} . 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 óptimo 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 arbitrario.

El predictor $\hat{y}_T(l)$ que cumple la desigualdad anterior es óptimo y es igual a:

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

El ECM el cual se define de la forma siguiente:

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

⁵ Ver Tabla 1, Caracterización de los años según el fenómeno del Niño.

7.4.2 Cálculo de la predicción puntual

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

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

Para obtener $\hat{y}_T(l)$, del modelo ARIMA(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:

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

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:

$$u_T = y_T - \hat{y}_{T-1} \quad (1)$$

$$u_{T-1} = y_{T-1} - \hat{y}_{T-2} \quad (1)$$

.....

$$u_{T+l-q} = y_{T+l-q} - \hat{y}_{T-l-q-1} \quad (1)$$

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 revisar los aspectos teóricos relacionados con la predicción, pasamos a realizar ciertas predicciones haciendo uso del software MINITAB.

Se sabe que la serie de datos analizada corresponde al periodo de mayo a octubre del 2003. Este año tuvo un régimen de lluvia normal, por lo que las predicciones se enmarcan en el período de mayo a octubre del siguiente año clasificado con un régimen de lluvia normal, que fue el 2005.

Las predicciones a estimar con los modelos disponibles (a manera de ejemplo) son las siguientes:

- al 15 de junio
- al 15 agosto
- al 15 de octubre

Las predicciones puntuales que aporta cada modelo, para las fechas mencionadas, se muestran en las tablas siguientes:

Primero se ha estimado la predicción de \hat{y}_{46} , es decir, la predicción del 15 de junio. Para esto calculamos: $\hat{y}_{46}(1)$, $\hat{y}_{46}(2)$ y $\hat{y}_{46}(3)$. Los valores obtenidos con el Minitab, fueron los siguientes:

	IMA (1,1)	ARI (1,1)	ARI (2,1)	IMA (1,2)
	Predicción	Predicción	Predicción	Predicción
$\hat{y}_{46}(1)$	12.0733	9.9429	9.9768	11.9466
$\hat{y}_{46}(2)$	12.1072	10.2523	9.5422	12.2420
$\hat{y}_{46}(3)$	12.1411	10.1842	10.0041	12.2755

Tabla 24. Resultados de las predicciones de \hat{y}_{46} (15 de junio)

El valor observado en campo de la serie analizada para $t=46$ es de 8.01. De la tabla 26 se observa que el valor de predicción más cercano a este dato es el del modelo ARI (2,1), seguido del valor predicho por el modelo ARI(1,1)

Los resultados para las otras fechas, se muestran a continuación:

	IMA (1,1)	ARI (1,1)	ARI (2,1)	IMA (1,2)
	Predicción	Predicción	Predicción	Predicción
$\hat{y}_{107}(1)$	4.08232	1.70505	2.29498	4.18772
$\hat{y}_{107}(2)$	4.11619	1.90634	1.67399	4.68125
$\hat{y}_{107}(3)$	4.15005	1.87468	1.95655	4.71566

Tabla 25. Resultados de las predicciones de \hat{y}_{107} (15 de agosto).

	IMA (1,1)	ARI (1,1)	ARI (2,1)	IMA (1,2)
	Predicción	Predicción	Predicción	Predicción
$\hat{y}_{168}(1)$	15.6516	24.5785	15.9888	12.8577
$\hat{y}_{168}(2)$	15.6855	23.2416	22.5960	13.0598
$\hat{y}_{168}(3)$	15.7193	23.6837	21.3173	13.0942

Tabla 26. Resultados de las predicciones de \hat{y}_{168} (15 de octubre).

Para el caso del 15 de agosto, el valor observado en campo fue de 2.26 mm de lluvia, en este caso la estimación más cercana, es de nuevo, la del modelo ARI (2,1).

En el tercer caso (15 de octubre) el valor observado es de 20.14 mm, la mejor predicción podría la del modelo ARI (1,1) o la predicción del modelo ARI (2,1).

De las predicciones calculadas, se observa que los cuatro modelos: IMA (1,1), ARI (1,1), ARI (2,1) y el IMA (1,2) aportan predicciones aceptables dada la naturaleza de la variable que se está analizando.

7.4.3 Bondad de ajuste de los modelos seleccionados

Para determinar la bondad de ajuste del modelo que mejor realiza las predicciones, se puede utilizar diversos criterios, tales como: Criterio de la Desviación Estándar Residual (DER),

el Criterio de Información de Akaike (AIC) o el Criterio Bayesiano de Schwarz (SBC), entre otros.

En este caso se usarán los tres métodos, para poder hacer una comparación de los modelos seleccionados hasta este momento y escoger el que aporte las mejores predicciones en términos de la variable lluvia diaria.

- Desviación Estándar Residual (DER): Es un criterio de selección de modelos y un valor pequeño indica una mayor adecuación del modelo, su expresión es:

$$DER = \sqrt{S/(T - r)}$$

Donde S es la suma cuadrática de los residuos, T es el número de las observaciones efectivas que se usan en el ajuste y r es el número de parámetros estimados en el modelo, incluyendo la constante.

- Criterio de Información de Akaike:
 $AIC = T \ln(S/T) + 2r$, Este criterio permite seleccionar un modelo. Se prefiere el modelo que tenga el menor valor de AIC.
- Criterio Bayesiano de Schwarz (SBC):
 $SBC = T \ln(S/T) + r \ln(T)$. Este criterio es un método para la selección de un modelo. Se prefiere al modelo que tenga el mínimo valor del estadístico. El SBC penaliza los parámetros adicionales más severamente que el AIC, conduciendo a modelos más simples.

Los resultados para los tres criterios y los cuatro modelos seleccionados hasta ahora se resumen en el siguiente cuadro:

Modelo	DER	AIC	SBC
IMA (1,1)	6.87	695.97	702.36
ARI (1,1)	7.34	719.45	725.84
ARI (2,1)	7.02	703.39	709.78
IMA (1,2)	6.66	684.59	690.98

Tabla 27. Resultados de los criterios DER, AIC y SBC para los modelos encontrados

En base a los resultados obtenidos el modelo que obtiene los valores menores es el IMA (1,2), con lo que concluimos que según esta prueba, este modelo es el que mejor se ajusta para

la precipitación diaria promedio de la zona oriental, quedando su expresión de la forma siguiente, en su modelo resumido:

$$w_t = u_t - 0.560u_{t-1} - 0.2574u_{t-2}$$

7.5 Análisis de la Serie Temporal con Outliers.

En la sección 7.1.1 Análisis de la serie temporal, se identificaron tres valores atípicos correspondientes a las fechas de 13 de junio (t=44), 25 de junio (t=56) y 14 de octubre (t=167), que serán tomados en cuenta ahora para analizar el modelo seleccionado y ver si tomando en cuenta estos valores aportan peso en las estimaciones de las predicciones, siempre recordando que el objetivo de estimar un modelo de serie temporal es hacer predicciones aceptables.

Estas observaciones identificadas como datos atípicos pudieran ser generadas por tormentas aleatorias normales a la época lluviosa, sobre todo que en el año en estudio (2003) no se presentó ninguna actividad de tormentas generadas por eventos especiales tales como Temporales⁶, tormentas tropicales, huracanes y la actividad de la Zona de convergencia intertropical (ZCIT) (Ver Anexo I).

Cuando se analizan los datos de series temporales, es común encontrar atípicos u outliers que tienen su origen en intervenciones incontroladas o inesperadas, es por eso que Box y Tiao⁷ denominaron a estos sucesos *intervenciones*.

La técnica del análisis de intervención consiste en valorar el efecto de las intervenciones en el proceso de comportamiento de la serie temporal, y para su aplicación se debe contar con la identificación de dos características de los modelos de intervención:

- a) El período de comienzo de dichos sucesos externos (intervenciones)
- b) La forma general del impacto de dichas intervenciones.

7.5.1 Modelos de Intervención

Los modelos de intervención son utilizados cuando algún suceso o evento ocurre y hace que los valores de la serie temporal en estudio aumenten (o disminuyan), a partir de la

⁶ Temporal: Sistemas ciclónicos o bajas presiones que provocan lluvias intermitentes, sin actividad eléctrica, con claros ocasionales y precipitaciones mayores a los 100mm.

⁷ Aznar, A., Trivez, Francisco, "Métodos de predicción en Economía, Tomo II, Ariel Economía, 1993.

ocurrencia del suceso. La disminución de la serie temporal puede ser de forma continua o un cambio brusco, únicamente en el instante de ocurrencia del suceso.

Son dos tipos de variables de intervención los que, determinan si el cambio de la serie es continuo ó temporal.

Una de estas variables de intervención es la **variable escalón**, la cual tiene efectos a partir del tiempo t_0 en el que ocurre, y se emplea solamente cuando los sucesos externos tienen una duración permanente.

La variable escalón se puede representar de la siguiente forma:

$$S_t^{(t_0)} = \begin{cases} 0, & t < t_0 \\ 1, & t \geq t_0 \end{cases}$$

El otro tipo de variable de intervención es la **variable impulso**, que tiene efectos solo en el tiempo t_0 de ocurrencia y se emplea cuando los sucesos externos tienen una duración temporal.

Este tipo de variable se representa de la siguiente forma:

$$P_t^{(t_0)} = \begin{cases} 1, & t = t_0 \\ 0, & t \neq t_0 \end{cases}$$

Nótese, que las variables escalón e impulso se relacionan mediante la siguiente expresión: $P_t^{t_0} = (1 - L)S_t^{t_0}$, donde L es el operador de retardo.

Luego entonces, un modelo de intervención puede representarse igualmente, bien con la variable escalón o con la variable impulso. El uso de una forma específica de estas variables es usualmente basado en la conveniencia de la interpretación en el modelo.

Algunos modelos pueden estructurarse, tomando en cuenta algunos criterios como los siguientes:

1. El modelo más simple sería aquel en que el impacto de una intervención es fijo y de magnitud desconocida. Así, dependiendo del tipo de intervención, y tomando en cuenta que el modelo se puede generalizar transcurridos b períodos, el impacto es:

$$X_t = \omega L^b S_t^{(t_0)} \text{ o } X_t = \omega L^b P_t^{(t_0)}$$

2. Si se tiene un impacto de una intervención con comienzo gradual, para los dos tipos de duración del efecto, con b períodos transcurridos después de la intervención se tiene para la variable escalón el modelo siguiente:

$$X_t = \frac{\omega L^b}{1 - \delta L} S_t^{(t_0)}$$

Y para la variable impulso se tiene:

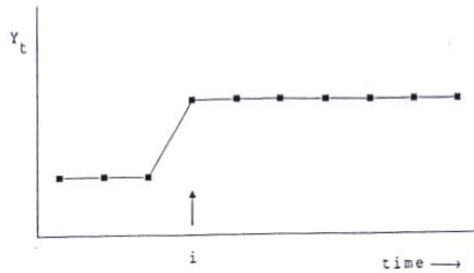
$$X_t = \frac{\omega L^b}{1 - \delta L} P_t^{(t_0)}$$

Donde $0 \leq \delta \leq 1$.

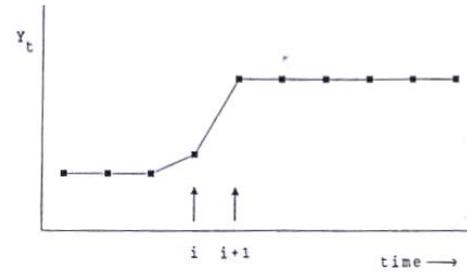
En las siguientes figuras, se muestran las gráficas para ambas variables de intervención con $b=1$ y $0 < \delta < 1$.

Para la variable escalón se tienen los siguientes casos:

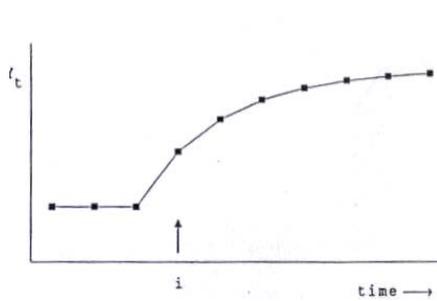
a) $X_t = \omega S_t^{t_0}$



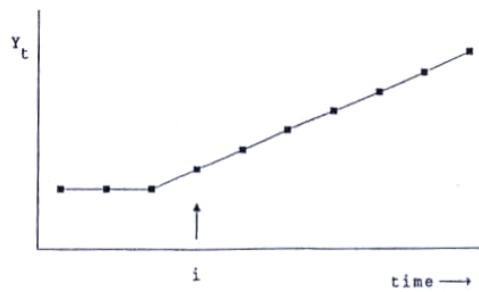
b) $X_t = \omega L S_t^{t_0}$



c) $X_t = \frac{\omega}{1 - \delta_1 L} S_t^{t_0}, \quad |\delta_1| < 1$

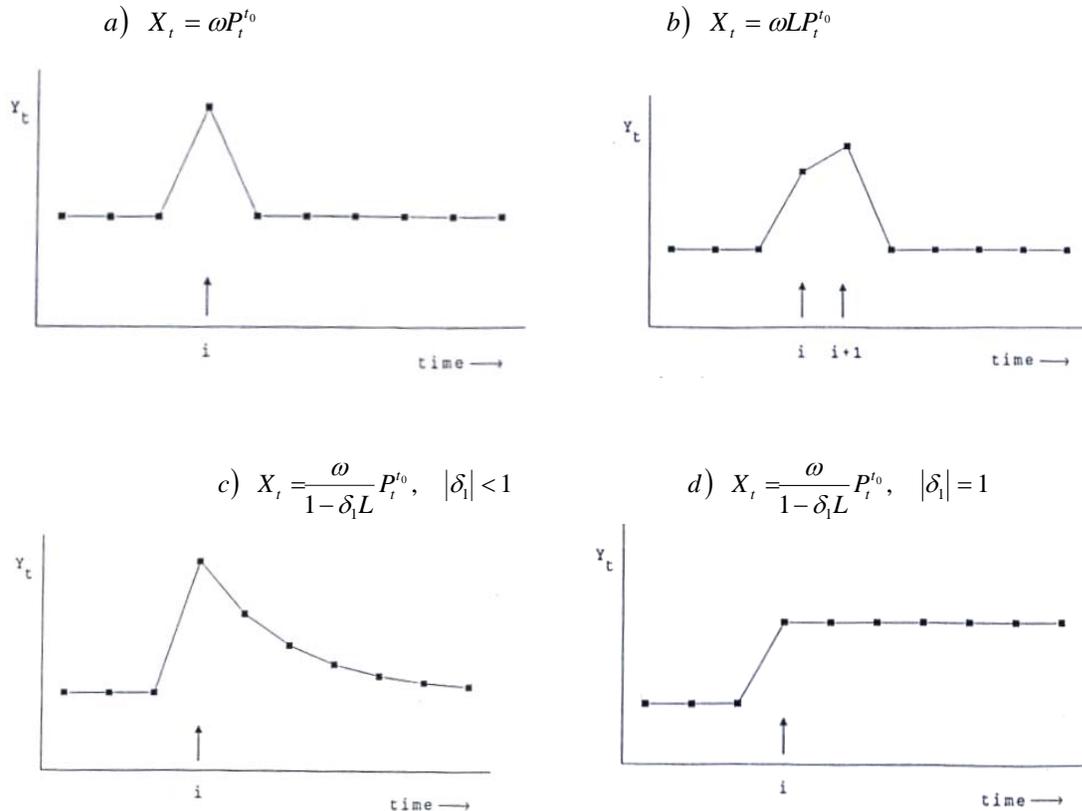


d) $X_t = \frac{\omega}{1 - \delta_1 L} S_t^{t_0}, \quad |\delta_1| = 1$



Gráfica 33. Diferentes formas del Comportamiento de la variable escalón en una serie temporal continúa

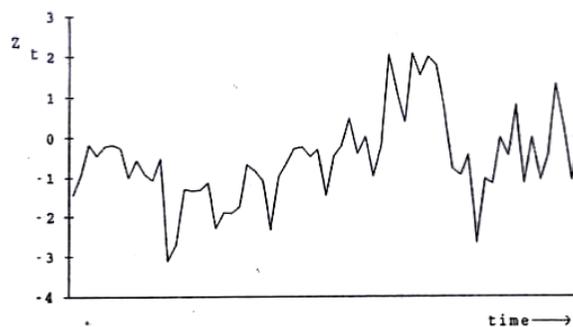
Y para la variable impulso se tienen los siguientes casos



Gráfica 34. Diferentes formas del Comportamiento de la variable impulso en una serie temporal continua

Es de hacer notar que en ocasiones suelen combinarse las variables escalón e impulso, en cuyo caso puede obtenerse el siguiente modelo de intervención, y resultará de utilidad cuando se espera que el efecto externo deje un efecto residual permanente después de una declinación gradual del impacto.

$$X_t = \left(\frac{\omega_1 L}{1 - \delta L} + \frac{\omega_2 L}{1 - L} \right) P_t^{t_0}$$



Gráfica 35. Comportamiento de la variable impulso y escalón en una serie temporal

Sin embargo, teniendo en cuenta la relación: $P_t^{t_0} = (1 - L)S_t^{t_0}$, la última expresión, puede expresarse de la forma siguiente:

$$X_t = \left(\frac{\omega_1 L}{1 - \delta L} P_t^{t_0} + \omega_2 L S_t^{t_0} \right)$$

En conclusión, se puede señalar que los modelos de intervención pueden especificarse mediante la siguiente expresión general:

$$X_t = v(L)I_t^{t_0}$$

Siendo $I_t^{t_0}$ una variable impulso o escalón.

$$v(L) = \frac{\omega(L)L^b}{\delta(L)}$$

Donde:

$$\omega(L) = \omega_0 - \omega_1 L - \omega_2 L^2 - \dots - \omega_s L^s$$

$$\delta(L) = 1 - \delta_1 L - \delta_2 L^2 - \dots - \delta_r L^r$$

En el cuadro siguiente, se recogen los valores que deberán tener b, r y s (subíndice de la expansión de cada uno de los parámetros) en la expresión anterior, para poder representar en $X_t = v(L)I_t^{t_0}$ cada uno de los modelos expuestos.

	$I_t^{t_0} = S_t^{t_0}$	$I_t^{t_0} = P_t^{t_0}$
r=0 b=0	$X_t = \omega S_t^{t_0}$	Modelo no significativo
r=0 b=1	$X_t = \omega L S_t^{t_0}$	$X_t = \omega L P_t^{t_0}$
r=1 b=0	$X_t = \frac{\omega}{1 - \delta L} S_t^{t_0}$	$X_t = \frac{\omega}{1 - \delta L} P_t^{t_0}$
r=1 b=1	$X_t = \frac{\omega L}{1 - \delta L} S_t^{t_0}$	$X_t = \frac{\omega L}{1 - \delta L} P_t^{t_0}$
r=2 b=0	Modelo no significativo	$X_t = \frac{\omega}{1 - \delta_1 L - \delta_2 L^2} P_t^{t_0}$

Obsérvese que en todos ellos el valor de s es cero porque $\omega(L) = \omega_0 = \omega$. Un modelo para k intervenciones sería:

$$X_t = \sum_{j=1}^k v_j(L) I_{jt}^{t_0^j}$$

$$\text{Donde: } v_j(L) = \frac{\omega_j(L)L^{b_j}}{\delta_j(L)}$$

7.5.2 Identificación del Modelo ARIMA con intervención.

Dada una serie temporal $\{y_t\}$, en la que algún suceso externo produce modificaciones sustanciales en el comportamiento de la misma, será preciso tener en cuenta en su modelización la inclusión de alguna intervención (o varias). Veamos ahora como identificar un modelo ARIMA con intervención (intervenciones), para ello comenzamos obteniendo el modelo ARIMA univariante, como sigue:

$$\phi(L)(1-L)^d y_t = \theta(L)u_t$$

Donde:

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

$$\theta(L) = 1 - \theta_1 L - \dots - \theta_p L^p$$

De manera que las raíces de ambos polinomios, $\phi(L)$ y $\theta(L)$, deben caer fuera del círculo unitario para cumplir la invertibilidad; otra condición que deben de cumplir estos polinomios es la estacionariedad, teniendo en cuenta que $(1-L)^d$, es el operador de diferencias consecutivas donde el exponente d indica el numero de diferencias que se realizaron para suavizar la serie y u_t un ruido blanco.

Considerando que la media de $\{y_t\}$ es igual a cero.

Obviamente, si el modelo fuese estacional multiplicativo ARIMA (p,d,q) x ARIMA (P,D,Q)s, La forma general de escribirla sería:

$$y_t = \frac{\Theta(L^s)\theta(L)}{\Phi(L^s)\phi(L)(1-L^s)^D(1-L)^d} u_t$$

Si quisiéramos incorporar ahora el modelo de intervención, la ecuación debería modificarse así:

$$y_t = v(L)I_t^{t_0} + \frac{\Theta(L^s)\theta(L)}{\Phi(L^s)\phi(L)(1-L^s)^D(1-L)^d} u_t$$

La ecuación puede generalizarse al modelo con k intervenciones, en cuyo caso escribiremos:

$$y_t = \sum_{j=1}^K v_j(L)I_{jt}^{t_0j} + N_t$$

Donde:

$$N_t = \frac{\Theta(L^S)\theta(L)}{\Phi(L^S)\phi(L)(1-L^S)^D(1-L)^d} u_t$$

Ahora, la tarea consiste en identificar el modelo recién escrito.

Como se menciono antes, se han detectado 3 valores atípicos (para t=44, 56 y 167), que podemos considerar como producto de tormentas aleatorias o convectivas que se producen en parte o todo el territorio nacional dentro de la época lluviosa y que son normales en el trópico, y las cuales no tienen una duración mayor a un día. Esto produce que los valores se vean reflejados en la serie como cambios bruscos y de duración temporal, en la cual la duración no excede a un período analizado, en este caso a un día. Por lo tanto para el análisis de intervención se considera como variables impulso estos valores atípicos.

Tomando en consideración que el modelo encontrado como adecuado en la fase de bondad de ajuste fue el IMA (1,2), y que también el modelo ARI (1,1) presentó buenas estimaciones en las predicciones, se hará el análisis de intervención con estos dos modelos.

Los modelos que serán analizados tienen las siguientes expresiones:

IMA (1,2): $w_t = u_t - 0.560u_{t-1} - 0.2574u_{t-2}$, que se puede expresar también como

$$y_t = (1 - \theta_1 L - \theta_2 L)u_t$$

ARI (1,1): $w_t = -0.319w_{t-1} + u_t$, que se puede expresar también como $(1 - \phi_1 L)y_t = u_t$

Dado que se tienen 3 intervenciones, que corresponden a los outliers identificados en la serie original, se pueden expresar los modelos con intervenciones como:

$$\text{IMA (1,2): } y_t = \frac{\omega_1}{1 - \delta_1 L} I_t^{t_0=44} + \frac{\omega_2}{1 - \delta_2 L} I_{2t}^{t_0=56} + \frac{\omega_3}{1 - \delta_3 L} I_{3t}^{t_0=167} + (1 - \theta_1 L - \theta_2 L)u_t$$

$$\text{ARI (1,1): } y_t = \frac{\omega_1}{1 - \delta_1 L} I_t^{t_0=44} + \frac{\omega_2}{1 - \delta_2 L} I_{2t}^{t_0=56} + \frac{\omega_3}{1 - \delta_3 L} I_{3t}^{t_0=167} + \frac{1}{(1 - \phi_1 L)} u_t$$

Donde:

$$I_{1t}^{t_0=44} = \begin{cases} 1 & \text{para } t = 44 \\ 0 & \text{para } t \neq 44 \end{cases} \quad I_{2t}^{t_0=56} = \begin{cases} 1 & \text{para } t = 56 \\ 0 & \text{para } t \neq 56 \end{cases}$$

$$I_{3t}^{t_0=167} = \begin{cases} 1 & \text{para } t = 167 \\ 0 & \text{para } t \neq 167 \end{cases}$$

Luego haciendo uso del software Statistica 7.0, procedemos a analizar estos modelos con intervención.

Modelo IMA (1,2)

Las salidas del análisis en Statistica están reflejadas en la siguiente tabla:

Input: Prec_Diaria (lluvia_diaria) Transformations: D(1) (Interrupted ARIMA) Model:(0,1,2) MS Residual= 42.886								
	Param.	Asympt. Std. Err.	Asympt. t(174)	p	Lower 95% Conf	Upper 95% Conf	Interv. Case No	Interv. type
Constant	0.03399	0.101812	0.333848	0.738896	-0.16696	0.23494		
q(1)	0.54653	0.075968	7.194236	0.000000	0.39660	0.69647		
q(2)	0.24853	0.077705	3.198425	0.001642	0.09517	0.40190		
Omega(1)	20.12462	6.199196	3.246328	0.001402	7.88932	32.35992	44	Abr/Temp
Delta(1)	-0.12731	0.255684	-0.497919	0.619170	-0.63195	0.37733	44	Abr/Temp
Omega(2)	22.31669	5.895517	3.785367	0.000211	10.68076	33.95263	56	Abr/Temp
Delta(2)	-0.15731	0.180213	-0.872931	0.383904	-0.51300	0.19837	56	Abr/Temp
Omega(3)	27.45775	6.231978	4.405945	0.000018	15.15775	39.75775	167	Abr/Temp
Delta(3)	0.66461	0.130447	5.094851	0.000001	0.40715	0.92207	167	Abr/Temp

Tabla 28. Resultados de la corrida del modelo IMA (1,2) con intervenciones

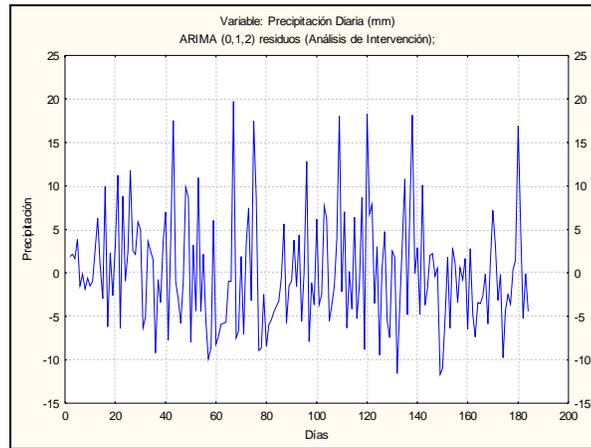
En la tabla se puede observar que los valores de la parte de medias móviles son muy parecidos a los obtenidos en el modelo sin intervención. Además se han señalado en color rojo los parámetros que resultaron significativos en la corrida del modelo.

Luego el modelo quedaría expresado de la siguiente manera:

$$y_t = 20.12I_{1t}^{t_0=44} + 22.32I_{2t}^{t_0=56} + \frac{27.46}{1-0.66L}I_{3t}^{t_0=167} + (1-0.55L-0.25L)u_t$$

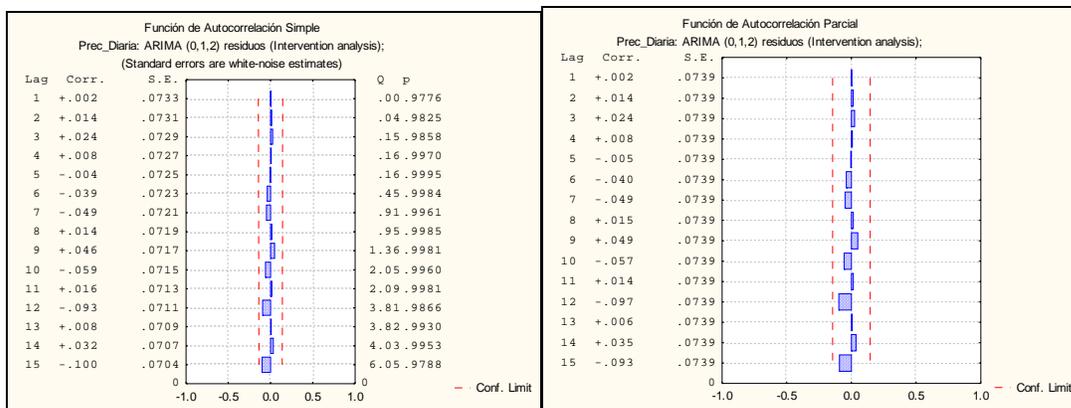
Los valores de δ_1 y δ_2 no se incluyen por haber resultado no significativos.

Residuos del modelo IMA(1,2) con intervención.



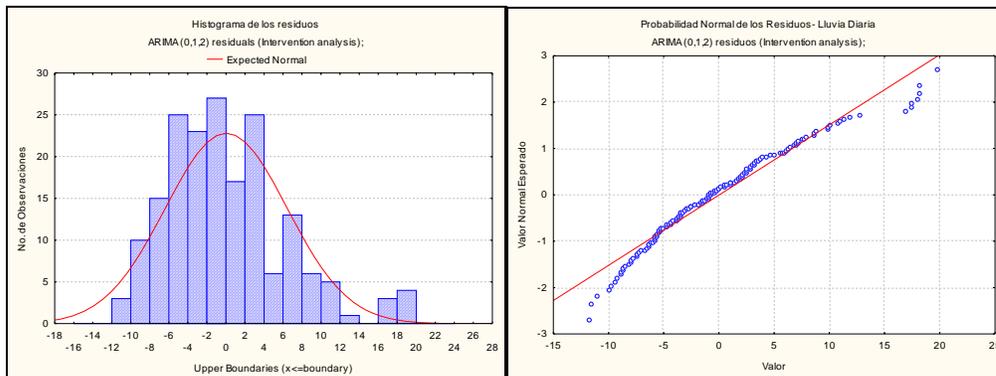
Gráfica 36. Comportamiento de los residuos del modelo IMA(1,2) con Intervención

FAS y FAP de los residuos del modelo IMA(1,2) con intervención.



Gráfica 37. Funciones de Autocorrelación Simple y Parcial de los residuos del IMA(1,2)

Histograma y Probabilidad Normal de los residuos del modelo IMA(1,2).



Gráfica 38. Histograma y Probabilidad Normal de los residuos del IMA(1,2)

Como se puede observar a partir del gráfico los residuos de este modelo con intervención, se comportan de manera constante a través del tiempo, adicionalmente en las gráficas de la FAS y la FAP vemos que más del 95% de los retardos no son significativos, cumpliéndose la incorrelación de los residuos. De la gráfica 38, histograma y probabilidad normal, podemos concluir que los residuos se comportan como un ruido blanco.

Modelo ARI (1,1)

Las salidas del programa para este modelo:

Input: Prec_Diaria (lluvia_diaria.sta) Transformations: D(1) (Interrupted ARIMA) Model:(1,1,0) MS Residual= 51.603								
	Param.	Asympt. Std. Err.	Asympt. t(174)	p	Lower 95% Conf	Upper 95% Conf	Interv. Case No	Interv. type
p(1)	-0.32236	0.074578	-4.32243	0.000026	-0.46954	-0.17518		
Omega(1)	22.12222	5.844406	3.78520	0.000210	10.58808	33.65635	44	Abr/Temp
Delta(1)	-0.18593	0.211874	-0.87755	0.381384	-0.60407	0.23221	44	Abr/Temp
Omega(2)	24.06420	5.580697	4.31204	0.000027	13.05051	35.07790	56	Abr/Temp
Delta(2)	-0.20598	0.155007	-1.32885	0.185618	-0.51189	0.09993	56	Abr/Temp
Omega(3)	27.81242	6.680667	4.16312	0.000049	14.62790	40.99695	167	Abr/Temp
Delta(3)	0.70337	0.217809	3.22930	0.001481	0.27352	1.13322	167	Abr/Temp

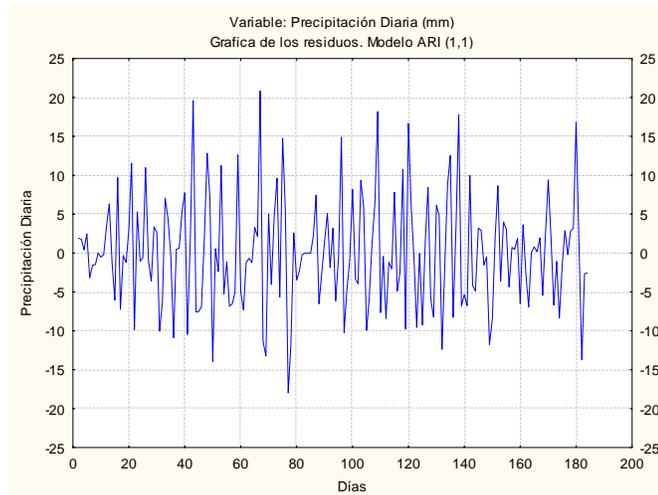
Tabla 29. Resultados de las predicciones de \hat{y}_{46} (15 de junio), sin y con intervención

A partir de estos resultados, el modelo queda de la siguiente manera:

$$y_t = 22.12I_t^{t_0=44} + 24.06I_{2t}^{t_0=56} + \frac{27.81}{1-0.7L} I_{3t}^{t_0=167} + \frac{1}{(1+0.32L)} u_t$$

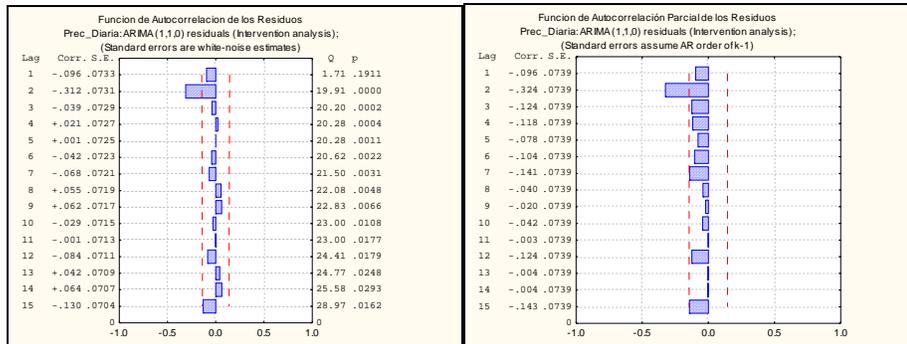
Los valores de δ_1 y δ_2 no se incluyen por haber resultado no significativos.

Residuos del modelo ARI(1,1) con intervención.



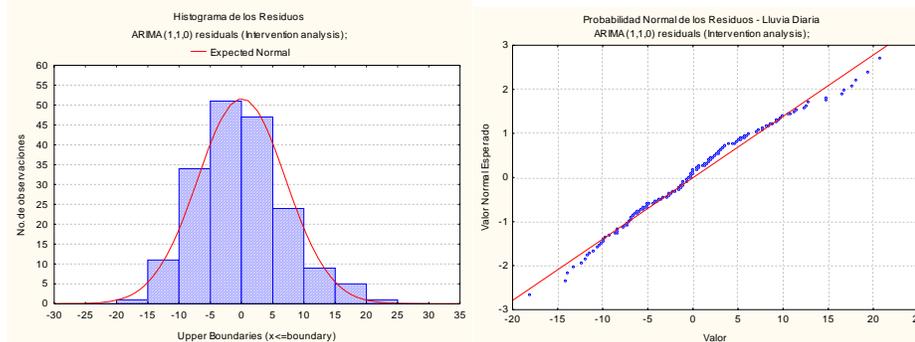
Gráfica 39. Comportamiento de los residuos del modelo ARI(1,1) con Intervención

FAS y FAP de los residuos del modelo ARI (1,1) con intervención.



Gráfica 40. Funciones de Autocorrelación Simple y Parcial de los residuos del ARI (1,1)

Histograma y Probabilidad Normal de los residuos del modelo ARI (1,1).



Gráfica 41. Histograma y Probabilidad Normal de los residuos del ARI(1,1)

Como se puede observar a partir del gráfico los residuos de este modelo con intervención, se comportan de manera constante a través del tiempo, adicionalmente de las gráficas de la FAS y FAP vemos que más del 95% de los retardos no son significativos, cumpliéndose la incorrelación de los residuos; y de la gráfica 41, histograma y probabilidad normal, podemos concluir que los residuos se comportan como un ruido blanco.

Con lo anterior se puede decir que los modelos con intervención son fuertes candidatos a modelar la serie de precipitación diaria, pero se pasará a ver las predicciones con estos nuevos modelos, tomando en cuenta las mismas predicciones que se hicieron con los modelos sin intervención.

Predicciones para el 15 de junio.

	IMA (1,2)	IMA (1,2)	ARI (1,1)	ARI (1,1)	Valor
	Sin intervención	Con intervención	Sin intervención	Con intervención	Observado
	Predicción	Predicción	Predicción	Predicción	
$\hat{y}_{46}(1)$	11.9466	13.66193	9.9429	14.91100	8.00667
$\hat{y}_{46}(2)$	12.2420	13.99445	10.2523	13.67103	10.81333
$\hat{y}_{46}(3)$	12.2755	14.04182	10.1842	13.94701	22.62000

Tabla 30. Resultados de las predicciones de \hat{y}_{46} (15 de junio), sin y con intervención

Predicciones para el 15 de agosto.

	IMA (1,2)	IMA (1,2)	ARI (1,1)	ARI (1,1)	Valor
	Sin intervención	Con intervención	Sin intervención	Con intervención	Observado
	Predicción	Predicción	Predicción	Predicción	
$\hat{y}_{107}(1)$	4.18772	3.677597	1.70505	1.101913	2.25625
$\hat{y}_{107}(2)$	4.68125	4.578587	1.90634	0.863556	7.90625
$\hat{y}_{107}(3)$	4.71566	4.578587	1.87468	0.940393	24.28125

Tabla 31. Resultados de las predicciones de \hat{y}_{107} (15 de agosto), sin y con intervención

Predicciones para el 15 de octubre.

	IMA (1,2)	IMA (1,2)	ARI (1,1)	ARI (1,1)	Valor
	Sin intervención	Con intervención	Sin intervención	Con intervención	Observado
	Predicción	Predicción	Predicción	Predicción	
$\hat{y}_{168}(1)$	12.8577	25.87835	12.8577	25.57811	20.14167
$\hat{y}_{168}(2)$	13.0598	19.75621	13.0598	19.97579	18.35000
$\hat{y}_{168}(3)$	13.0942	15.68503	13.0942	15.82965	22.37500

Tabla 32. Resultados de las predicciones de \hat{y}_{168} (15 de octubre), sin y con intervención

Al observar las predicciones se puede notar que en dos de los casos (15 de junio y 15 de agosto), el modelo ARI (1,1) sin intervención genera mejores predicciones, sin embargo en el caso del 15 de octubre las mejores predicciones son hechas por el modelo ARI (1,1) con

intervención, reflejándose además que las predicciones para esta fecha son muy cercanas entre los dos modelos $ARI(1,1)$ e $IMA(1,2)$ con intervención.

Se puede concluir que cualquiera de los dos modelos (con o sin intervención) se ajusta a la serie de datos de precipitación diaria y puede ser utilizado para realizar pronósticos de lluvia diaria para la zona oriental del país para años con régimen de lluvia normal. Atendiendo al criterio de parsimonia se sugiere el modelo $ARI(1,1)$ para generar las predicciones.

8. Conclusiones y Recomendaciones

- Todos los modelos que pasaron la fase de chequeo son buenos para hacer predicciones. En lo que se debe tener cuidado es en la interpretación de las predicciones dadas por cada modelo. Es de observar que cuando resultan predicciones que se vuelve constantes a partir del retardo “x”, ese dato nos dice que la predicción más fiable es hasta ese retardo. En el caso de la lluvia diaria, según la experiencia de los meteorólogos, no es adecuado predecir a más de 48 horas, sobre todo en los países tropicales como el nuestro.
- A partir de las pruebas de bondad de ajuste de los modelos, se pudo verificar la consistencia de la prueba en la selección del modelo, ya que el IMA (1,2) resulto ser el mejor predictor en las tres pruebas de bondad de ajuste realizadas: Criterio de la Desviación Estándar Residual (DER), el Criterio de Información de Akaike (AIC) o el Criterio Bayesiano de Schwarz (SBC).
- El análisis de intervención para tratamiento de outliers fue aplicado a los modelos: IMA(1,2) y ARI(1,1) observando que en general, no existen diferencias significativas en las predicciones generadas por el modelo con intervención, en relación a las predicciones con el modelo sin intervención. Sin embargo en la predicción de algunos valores puntuales, las pruebas de ambos modelos con intervención aportaron un mejor ajuste a los datos observados.
- En términos prácticos se ha observado que el modelo ARI(1,1) a pesar de no tener el mejor resultado en la prueba de bondad de ajuste, cumple a satisfacción todas las fases del chequeo y proporciona muy buenas predicciones antes y después del análisis de intervención para tratamiento de outliers. Por lo que, atendiendo también al criterio de parsimonia, se concluye que el ARI(1,1) es el mejor modelo para predecir la variable lluvia diaria observada.
- En este trabajo se aplicó la metodología de Box-Jenkins, abordando paso a paso cada etapa para el análisis de la serie temporal. Para ello se hizo uso de los programas estadísticos: Minitab 14, SPSS 15, TRAMO-SEATS, y el Statistica 7.0 y

se realizaron adicionalmente pruebas estadísticas para la consistencia del estudio. Sin embargo, también se pudo comprobar que el desarrollo de software estadístico ha evolucionado mucho y es posible, por ejemplo, en el caso del SPSS, a partir de la versión 15, utilizar la opción de “modelador experto”, la cual identifica el modelo o los modelos probables para el conjunto de datos, sin tener que analizar con rigor cada una de las etapas planteadas por la metodología Box-Jenkins. Esto en términos prácticos puede resultar de mucha utilidad para personas o instituciones que estén únicamente interesados en la selección de un modelo.

- Se recomienda hacer estudios similares para años no normales (niño o niña) considerando toda la estación lluviosa.
- En futuros trabajos de tesis de maestría, se sugiere el estudio de la variable precipitación diaria para El Salvador, en diferentes estados, como persistencia (días consecutivos de lluvia constante) o estimación de lluvias máximas o mínimas diarias, para los cuales existen antecedentes en otros países, en los que se han usado técnicas como Cadenas de Markov y la distribución de Pareto.

9. Fuentes Bibliográficas.

- Aznar, A.: **Métodos de Predicción en Economía**, Tomo II, Análisis de Series Temporales, Ariel Economía, 1993.
- Barrantes, Jorge A.; Vega G, Norman: **La persistencia y las cadenas de Markov aplicadas a la lluvia en San José durante el periodo 1950-1990**. Instituto Meteorológico Nacional, Ministerio del Ambiente y Energía. San José, Costa Rica, 2003
- Collantes Duarte, J.; Colmenares La Cruz, G.; Orlandoni Merli, G.; Rivas Echeverría, F.: **“Predicción con redes neuronales artificiales: comparación con las metodologías de Box y Jenkins”**. Tesis de Maestría, Universidad de Los Andes, 2001.
- García Guirola, L; Soriano de Cruz, L; Rivas Pacheco, T & Sermeño Arias, T : **Mejoramiento de la Capacidad Técnica para Mitigar los Efectos de Futuros Eventos de la Variabilidad Climática (El Niño)**, 2002.
- Maravall, Agustín.: **An Application of the TRAMO-SEATS automatic procedure; direct versus indirect adjustment**. Documentos de Trabajo No. 0524. Banco de España. 2005.
- Morata Gasca, A. y Almarza Mata, C.: **Ajuste de la Distribución de Probabilidad de Máxima Precipitación diaria en Santa Cruz de Tenerife a la Distribución generalizada de Pareto**. Instituto Nacional de Meteorología, 2002.
- Peña Sánchez, Daniel: **Estadística. Modelos y Métodos**. Volumen II, Alianza Editorial, 1989.
- Peña Sánchez, Daniel y Arnáiz Tovar, Gonzalo: **Criterios de Selección de Modelos ARIMA**. Trabajos de Estadística e Investigación Operativa. Vól. 32, No. 1, 1981, Págs. 70 a 93.

- Wei, William.: **Time Series Analysis, Univariate and Multivariate Methods**, Addison Wesley, 1990.

ANEXOS

ANEXO 1.
Eventos extremos ocurridos desde 1927. Fuente: SNET,
Servicio Meteorológico

Año	Evento	Año	Evento
1927	Ago, Sept	1968	Sept, Oct/HUR Gladys
1928	Sept	1969	Jun, Ag/HUR Camille, Sep/HUR Francelia, Oct./TT Jenny y Kara
1929	Mayo, Sept	1970	Atemp Junio, Jul, Ago/HUR Celia, Sept.
1930		1971	Atemp TT Bridget Pacifico Jun, Sept/HUR Edith+ HUR Irene=Olivia Pac, Nov/TT Laura
1931	Jun, Sept/HUR S/N	1972	Jun/HUR Agnes
1932		1973	Atemp Agos, Oct/TT Gilda
1933		1974	Jun/SN, Sept/HUR Fifi
1934	Jun/TT S/N	1975	Sept.
1935		1976	Jun
1936	Jun/TT S/N	1977	
1937		1978	Sept(Greta) Atemp(Ago-David, Sept- Frederic, Jul, Oct.
1938		1979	
1939		1980	
1940		1981	
1941		1982	Sept
1942		1983	
1943		1984	
1944		1985	Sept
1945	Oct/HUR S/N	1986	
1946		1987	
1947		1988	Ago/Gilbert -Sept-Oct/Joan
1948		1989	Sept.
1949	Sept-Oct/HUR S/N	1990	Oct.
1950	Oct/Hur ITEM	1991	
1951		1992	Sept-Oct
1952		1993	Jun., Sep (Gert)
1953	Sept	1994	
1954	Sept/TT GILDA ?, Oct	1995	Oct. (Roxanne)
1955	Sept/ HUR Hilda/HUR Janet	1996	Julio (Cesar, Kile)
1956	Sept/HUR FLOSSY	1997	
1957	Oct	1998	Oct/Mitch
1958	Junio TT ALMA	1999	Sept.
1959		2000	Sept.
1960	Oct	2001	
1961	Jul/HUR Anna, Sept/HUR Carla, Oct/ HUR Hattie, Nov/TT INGA	2002	
1962	Junio	2003	
1963	Jun., Sept., Nov Julio, Oct/HUR Isbell	2004	Mayo/Hur Adrian, Oct/HUR Stan
1964		2005	
1965	Jun, Sept/TT Debbie	2006	
1966	Jun/ZCIT, Sept	2007	Ago/Dean, Sept/Felix, Oct/BAJAS
1967	Jun, Sept/HUR Beulah		