



Benemérita Universidad Autónoma de Puebla
Facultad de Ciencias Físico Matemáticas
Posgrado en Ciencias Matemáticas

Análisis de Variables Climatológicas del Estado de Tlaxcala
Utilizando Series de Tiempo y Redes Neuronales

Tesis presentada al

Posgrado en Ciencias Matemáticas

Como requisito para la obtención del grado de

Maestro en Ciencias Matemáticas

por

Silvia Herrera Cortés

Bajo la dirección de

Dr. Bulmaro Juárez Hernández

Puebla, Puebla. 2015

Agradecimientos

Agradezco a todas aquellas personas que de alguna manera contribuyeron a mi formación personal y profesional en este trayecto de mi vida, en especial:

A Dr. José Enrique Arrazola y Dr. Juan Alberto Escamilla Reyna por su paciencia y comprensión durante el curso propedeúutico.

A quienes formaron parte de mi formación académica: Dr. Jorge Bustamante González, Dr. Miguel Antonio Jiménez Pozo, Dr. Jacobo Oliveros Oliveros, Dr. David Hernández Villa, Dr. Arnoldo Bezanilla López y Dr. Bulmaro Juárez Hernández, por los conocimientos transmitidos.

A mi director de tesis, Dr. Bulmaro Juárez Hernández, por sus consejos, sugerencias, amabilidad y comprensión.

A Dr. Francisco Tajonar Sanabria, Dra. Hortencia Josefina Reyes Cervantes, Dr. Víctor Hugo Vázquez Guevara y Dra. Galdys Linares Fleites, por sus acertadas comentarios para mejorar el proyecto.

A Dr. Tomás Morales Acoltzi, Ing. Rogelio Bernal Morales, Dr. Saturnino Orozco Flores, M.E. F. Ma. de Lourdes Blanca Nieto Pérez por todo su apoyo.

A la Benemérita Universidad Autónoma de Puebla por acogerme durante estos dos años.

Finalmente, agradezco a CONACyT ya que sin su apoyo no hubiera sido posible realizar mis estudios.

Índice de Figuras	IV
Índice de Tablas	VII
Introducción	1
1. Variables atmosféricas	3
1.1. Estaciones meteorológicas	4
1.2. Temperatura	5
1.3. Precipitación	6
2. Series de tiempo	8
2.1. Operadores y polinomios de retraso	9
2.2. Ecuaciones en diferencia	10
2.2.1. Ecuaciones en diferencia de primer orden	10
2.2.2. Ecuaciones en diferencia de p -ésimo orden	12
2.2.3. Solución general de una ecuación en diferencia de p -ésimo orden con eigenvalores distintos	15
2.2.4. Solución de una ecuación en diferencia de segundo orden con distintos eigenvalores	19
2.2.5. Solución de una ecuación en diferencia de p -ésimo orden con eigenvalores repetidos	20
2.2.6. Ecuaciones en diferencia utilizando operadores de retraso	22
2.3. Función de autocovarianza y función de autocorrelación	26
2.4. Propiedades de la media muestral y la función de autocorrelación	30
2.5. Ejemplos de procesos estacionarios no estacionales	33
2.5.1. Procesos de medias móviles (MA)	33

2.5.2. Modelos autorregresivos (AR)	37
2.5.3. Procesos autorregresivos de medias móviles de orden (p, q) ($ARMA(p, q)$)	39
2.5.4. Identificación tentativa de modelos Box-Jenkins no estacionales	40
2.6. Invertibilidad	40
2.7. Procesos Lineales	42
2.8. Pronóstico de series de tiempo estacionarias	44
2.9. Estimación por máxima verosimilitud	49
2.9.1. Función de verosimilitud para un proceso gaussiano $AR(1)$. .	49
2.10. Función de verosimilitud para un proceso gaussiano $AR(p)$	54
2.11. Función de verosimilitud para un proceso gaussiano $MA(1)$	56
2.12. Función de verosimilitud para un proceso gaussiano $MA(q)$	57
2.13. Función de verosimilitud para un proceso $ARMA(p, q)$	57
2.14. Ejemplos de procesos estacionales	58
2.14.1. Procesos de promedios móviles estacionales de orden Q	59
2.14.2. Procesos autorregresivos estacionales de orden P	59
3. Redes Neuronales	61
3.1. Definiciones básicas de redes neuronales	62
3.2. Redes neuronales artificiales	63
3.3. Tipos de función de activación	64
3.4. Redes neuronales vistas como grafos dirigidos	66
3.5. Arquitectura de una red neuronal	67
3.5.1. Red de alimentación directa de una sola capa	67
3.5.2. Red de alimentación directa multicapa	68
3.5.3. Redes recurrentes	68
3.6. Representación del conocimiento de una red neuronal	69
3.7. Procesos de aprendizaje de una red neuronal	71
3.7.1. Aprendizaje error-corrección	71
3.8. Aprendizaje basado en memoria	72
3.8.1. Aprendizaje Hebbiano	73
3.8.2. Aprendizaje competitivo	74
3.9. Perceptrones de una sola capa	75
3.10. Perceptrón multicapa	78
3.11. Algoritmo Back-propagation	81
3.12. Algoritmo de Levenberg-Maquardt	89
4. Caso de estudio	93
4.1. Análisis de los datos utilizando la metodología de Box-Jenkins	94
4.1.1. Estación de Atlangatepec, Tlaxcala	99
4.2. Análisis de los datos utilizando redes neuronales	116
4.2.1. Estación de Atlangatepec, Tlaxcala	117

5. Análisis de Resultados	120
Conclusión	125
Bibliografía	126
Apéndice 1	128
Apéndice 2	129
Apéndice 3	148

Índice de figuras

1.1. Componentes del sistema climático	4
1.2. Elementos de una estación meteorológica automatizada	5
1.3. Escalas termométricas	6
1.4. Pluviómetro	7
2.1. Análisis de un multiplicador dinámico para una ecuación en diferencia de primer orden	12
2.2. Resumen de los multiplicadores dinámicos de una ecuación en diferencia de segundo orden	21
2.3. Diagrama de flujo	60
3.1. Esquema de una red neuronal biológica	63
3.2. Esquema de una red neuronal artificial	63
3.3. Transformación afín producida por la presencia del sesgo	64
3.4. Funciones de activación: a) Función umbral, b) Función lineal a trozos, c) Función logística	65
3.5. Reglas básicas para la construcción de grafos dirigidos	66
3.6. Gráfico de flujo de señal de una neurona	67
3.7. Representación gráfica de una red de una sola capa	68
3.8. Representación gráfica de una red multicapa	69
3.9. Red neuronal recurrente sin bucles de autoalimentación y sin neuronas ocultas	70
3.10. Representación gráfica del aprendizaje de error-corrección	72
3.11. Representación gráfica del aprendizaje basado en memoria	73
3.12. Representación gráfica de un perceptrón	76
3.13. Regla de decisión para un perceptrón con dos entradas x_1, x_2	77
3.14. Representación gráfica del perceptrón equivalente a la Figura 3.12	77

3.15. Regla de clasificación. a) Clases linealmente separables, b) Clases no linealmente separables	77
3.16. Red completamente conectada	79
3.17. Representación gráfica de una neurona destacando los detalles de la neurona de salida j	82
3.18. Representación gráfica de una neurona destacando los detalles de la neurona k conectada a la neurona oculta j	84
4.1. Ubicación geográfica de las 5 estaciones meteorológicas estudiadas	94
4.2. Diagramas de caja para la temperatura máxima, mínima y precipitación de Atlangatepec, Huamatla y Tlaxco	95
4.3. Diagramas de caja para la temperatura máxima, mínima y precipitación de Españita y El Carmen Tequexquitla	96
4.4. Series de tiempo y funciones de autocorrelación y autocorrelación parcial para las estaciones meteorológicas de Atlangatepec y Huamantla	97
4.5. Series de tiempo y funciones de autocorrelación y autocorrelación parcial para las estaciones meteorológicas de Tlaxco, Españita	98
4.6. Series de tiempo y funciones de autocorrelación y autocorrelación parcial para la estación meteorológica El Carmen Tequexquitla	99
4.7. Serie de tiempo transformada, función de autocorrelación y función de autocorrelación parcial para los datos de temperatura mínima de Atlangatepec, Tlaxcala	100
4.8. Serie de tiempo, función de autocorrelación y función de autocorrelación parcial para los datos de temperatura mínima de Atlangatepec, Tlaxcala, con una diferencia en el nivel no estacional y una diferencia en el nivel estacional	101
4.9. Resultados de la prueba de Ljung-Box, para los datos de la serie de tiempo de la temperatura mínima de Atlangatepec, Tlaxcala	104
4.10. Serie de tiempo transformada, función de autocorrelación y función de autocorrelación parcial para los datos de temperatura máxima de Atlangatepec, Tlaxcala	105
4.11. Serie de tiempo, función de autocorrelación y función de autocorrelación parcial para los datos de temperatura máxima de Atlangatepec, Tlaxcala, con una diferencia en el nivel no estacional y una diferencia en el nivel estacional	106
4.12. Resultados de la prueba de Ljung-Box, para los datos de la serie de tiempo de la temperatura máxima de Atlangatepec, Tlaxcala	108
4.13. Serie de tiempo transformada, función de autocorrelación y función de autocorrelación parcial para los datos de la precipitación de Atlangatepec, Tlaxcala	110

4.14. Serie de tiempo, función de autocorrelación y función de autocorrelación parcial para los datos de precipitación de Atlangatepec, Tlaxcala, con una diferencia en el nivel no estacional y una diferencia en el nivel estacional	111
4.15. Resultados de la prueba de Ljung-Box, para los datos de la serie de tiempo de la precipitación de Atlangatepec, Tlaxcala	113

Índice de tablas

2.1. Identificación de los modelos por medio del comportamiento de las funciones f_{am} y f_{amp}	40
4.1. Periodos de información para 5 estaciones del estado de Tlaxcala	94
4.2. Estimación de los parámetros para el modelo (4.4)	102
4.3. Valor observado y pronosticado para la temperatura mínima de Atlangatepec, Tlaxcala	103
4.4. Modelos que fueron propuestos para la temperatura mínima de Atlangatepec, Tlaxcala	103
4.5. Estimación de los parámetros para el modelo (4.9)	107
4.6. Valor observado y pronosticado para la temperatura máxima de Atlangatepec, Tlaxcala	108
4.7. Modelos que fueron propuestos para la temperatura máxima de Atlangatepec, Tlaxcala	109
4.8. Estimación de los parámetros para el modelo 4.14	110
4.9. Valor observado y pronosticado para la precipitación de Atlangatepec, Tlaxcala	112
4.10. Modelos que fueron propuesto para la precipitación de Atlangatepec, Tlaxcala	112
4.11. Mejor modelo y suma del cuadrado de los errores de predicción	115
4.12. Valor observado y valor pronosticado en cada una de las estaciones para las variables estudiadas, utilizando series de tiempo	116
4.13. Valor observado y pronosticado para la temperatura mínima de Atlangatepec, Tlaxcala, utilizando redes neuronales	117
4.14. Valor observado y pronosticado para la temperatura máxima de Atlangatepec, Tlaxcala, utilizando redes neuronales	118
4.15. Valor observado y pronosticado para la precipitación de Atlangatepec, Tlaxcala, utilizando redes neuronales	118

4.16. Valor Observado y valor pronosticado en cada una de las estaciones para las variables estudiadas, utilizando redes neuronales 119

Introducción

Los datos que se obtienen de variables climáticas juegan un papel importante para la caracterización del clima de alguna región, para el análisis de algunas variables como lo es el caso de la precipitación es necesario tener un registro de datos de por lo menos 30 años de observación, a este periodo de datos observados se le conoce como periodo largo de observación [1]. Estudiar la tendencia de estos datos es esencial, por ejemplo; una pequeña variación en la temperatura puede suponer grandes cambios en situaciones como la disponibilidad de agua para consumo o la distribución de especies [8].

El pronóstico de variables climatológicas como es la precipitación, la temperatura, el inicio de la temporada de heladas, etc, juegan un papel importante en las actividades de los productores agrícolas y también permiten al ser humano tomar precauciones sobre posibles desastres naturales ocasionados por lluvias intensas, sequías extremas o inicio temprano de la temporada de heladas.

Para los productores agrícolas, el pronóstico de lluvias es importante porque les permite tomar decisiones respecto del tipo de producto que deben sembrar para obtener algún beneficio que a su vez repercute en el desarrollo económico del país.

El estudio del comportamiento de las variables climatológicas cuyo registro numérico de sus características son longitudinales, se puede relacionar con el estudio de las series de tiempo cuyo objetivo es el de predecir o pronosticar datos. Una serie de tiempo se entiende como el registro metódico de la medición u observación numérica, efectuada a intervalos de tiempos fijos acerca de ciertas características de ciertas unidades de observación [11].

En este trabajo de investigación se realiza un estudio sobre la temperatura máxima,

mínima y precipitación de 5 estaciones meteorológicas del estado de Tlaxcala por medio de series de tiempo y redes neuronales.

Para el análisis de series de tiempo se aplica la metodología de Box y Jenkins con la finalidad de elegir el modelo que mejor se ajuste a los datos y realizar pronósticos. Para el análisis de los datos con redes neuronales se aplican perceptrones multicapa para finalizar con una comparación entre los pronósticos obtenidos por ambas metodologías aplicadas.

En este trabajo se tiene como objetivo general, realizar un análisis de bases de datos de estaciones meteorológicas no automáticas del estado de Tlaxcala (ver disco anexo) para explicar el comportamiento de algunos fenómenos meteorológicos relacionados con éstos, por medio de la metodología de Box y Jenkins para series de tiempo, además de realizar una comparación de los pronósticos obtenidos con series de tiempo y redes neuronales.

Por lo anterior el siguiente documento se divide en 5 capítulos. En el capítulo uno, se presenta de manera breve el concepto de meteorología, estación meteorológica, estación meteorológica automática y no automática así como los conceptos de temperatura y precipitación. En el capítulo dos, se presenta la teoría sobre series de tiempo, definiendo: proceso estacionario, procesos autorregresivos, procesos de medias móviles y autorregresivos de promedios móviles. En el capítulo tres se presenta la teoría de redes neuronales. En el capítulo cuatro se muestra el tratamiento de los datos con series de tiempo y redes neuronales y finalmente, en el capítulo cinco, se presenta un análisis de los resultados obtenidos y conclusiones así como el trabajo futuro a realizar. También se presenta un anexo con las líneas de código utilizadas para el análisis con series de tiempo y redes neuronales.

CAPÍTULO 1

Variables atmosféricas

El clima se suele definir en sentido restringido como el estado promedio del tiempo y, más rigurosamente, como una descripción estadística del tiempo atmosférico en términos de los valores medios y de la variabilidad de las magnitudes correspondientes durante periodos que pueden abarcar desde meses hasta miles o millones de años [13]. Las magnitudes son casi siempre variables de superficie (por ejemplo, temperatura, precipitación o viento). En un sentido más amplio, el clima es el estado del sistema climático en términos tanto clásicos como estadísticos. El periodo de promedio habitual para su caracterización es de 30 años, según la definición de la Organización Meteorológica Mundial (OMM).

Se puede considerar al clima como el estado del sistema climático el cual comprende:

- La **atmósfera** es la envoltura del gas que rodea la Tierra.
- La **hidrósfera** es la parte del sistema climático que contiene agua líquida en la superficie de la Tierra y subterráneas (por ejemplo, océanos, ríos, lagos ...).
- La **criósfera** contiene agua en su estado congelado (por ejemplo, glaciares, nieve, hielo, ...).
- La **litósfera** es el componente sólido de nuestro planeta, con un grosor de algo más de 100 km, está conformada por rocas y metales sólidos en la superficie.
- La **biósfera** contiene todos los organismos y los ecosistemas que viven sobre la tierra y en los océanos [17].

Se considera a la meteorología como la ciencia que estudia la atmósfera y los fenómenos como la lluvia, viento, humedad relativa, etc., que en ella acontecen. Se

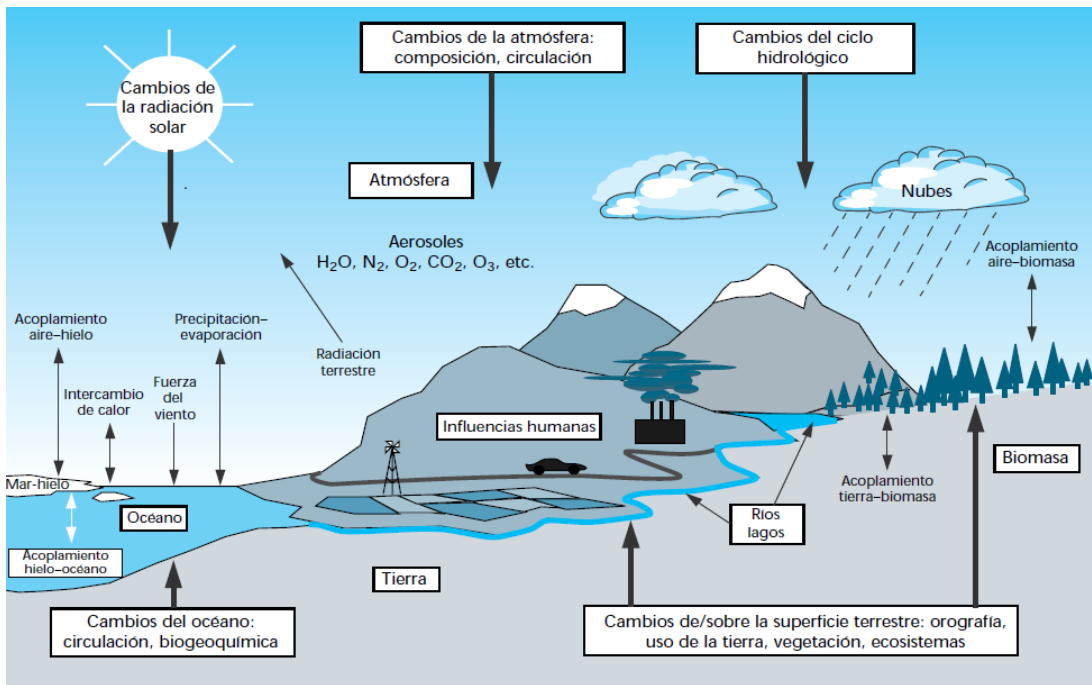


Figura 1.1: Componentes del sistema climático

conoce como tiempo atmosférico al estado de la atmósfera en un momento dado y se encuentra caracterizado por la temperatura, presión del aire, humedad y viento [18].

1.1. Estaciones meteorológicas

De acuerdo a la guía práctica de hidrometeorología de la Organización Mundial de Meteorología (OMM) [17], una estación meteorológica se utiliza para medir variables climatológicas como temperatura, precipitación, velocidad del viento, radiación solar, humedad relativa, etc.

Una **estación meteorológica** se define como el espacio físico donde están instalados los instrumentos para la medición de diversos elementos meteorológicos.

Una **estación meteorológica automática** está conformada por un grupo de sensores que registran y transmiten información meteorológica de forma automática del sitio donde está estratégicamente colocada, su función principal es la recopilación y monitoreo de algunas variables meteorológicas para generar archivos del promedio de cada 30 minutos de todas las variables.

Los sensores que conforman una estación meteorológica automática se utilizan para registrar información numérica de variables como: velocidad del viento, presión atmosférica, temperatura y humedad relativa, radiación solar y precipitación (Figura 1.2).



Figura 1.2: Elementos de una estación meteorológica automatizada

En una **estación meteorológica no automática**, sólo se registra información numérica sobre precipitación y temperatura.

A continuación, se presenta el concepto de temperatura y precipitación, variables meteorológicas que serán analizadas en este trabajo de investigación.

1.2. Temperatura

La **temperatura** es la condición que determina la dirección del flujo neto de calor entre dos cuerpos [17]. Esta magnitud nos permite expresar el grado de calentamiento o enfriamiento de los cuerpos. La temperatura termodinámica (T) expresada en grados Kelvin es la temperatura básica. En meteorología se utiliza casi siempre la temperatura (t) expresada en grados Celsius definida por:

$$t = T - 273.16 \cdot$$

En la escala Celsius se asigna el valor 0 ($0^{\circ}C$) a la temperatura de congelación del agua y el valor 100 ($100^{\circ}C$) a la temperatura de ebullición del agua a nivel del mar.

El intervalo entre estas dos temperaturas se divide en 100 partes iguales, cada una de las cuales corresponde a un grado. En la escala Kelvin se asigna el 0 a aquella temperatura a la cual las partículas carecen de energía cinética (temperatura más baja posible) la cual equivale a -273°C de la escala Celsius (Figura 1.3).

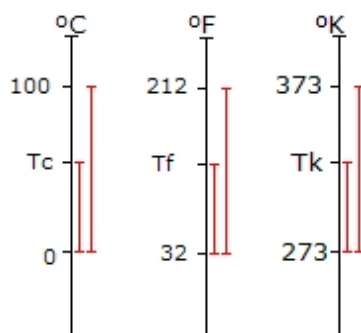


Figura 1.3: Escalas termométricas

De acuerdo al Servicio Meteorológico Nacional (SMN) [9], se definen los siguientes conceptos:

Temperatura Ambiente. Es la temperatura del aire registrada en el instante de la lectura.

Temperatura Máxima. Es la mayor temperatura registrada en un día, y que se presenta entre las 14 : 00 y las 16 : 00 horas.

Temperatura Mínima. Es la menor temperatura registrada en un día, y se puede observar entre las 6 : 00 y las 8 : 00 horas.

1.3. Precipitación

Se denomina **precipitación** al agua de la atmósfera que cae en forma líquida, sólida, o líquida y sólida desde las nubes hasta la superficie de la tierra. También se denomina precipitación a la caída del agua en estado líquido y/o sólido que alcanza la superficie, proveniente de las nubes que están formadas de pequeñas gotas de agua y cristales de hielo que se han formado sobre los núcleos de condensación y núcleos de congelación. El objetivo de medir la precipitación es obtener tanta información como sea posible, acerca de la cantidad y distribución, en el tiempo y el espacio de

ésta, generalmente, la precipitación se mide en un recipiente circular de 12.5 a 30 cm de diámetro expuesto a una altura que varía de 45 a 100 cm con respecto a la superficie del suelo.



Figura 1.4: **Pluviómetro**

CAPÍTULO 2

Series de tiempo

Una serie de tiempo es un conjunto de valores numéricos correspondientes a la realización de un proceso estocástico, $\{X_t(\omega), t \in T\}$, los cuales se obtienen en periodos regulares a través del tiempo y su comportamiento se debe a componentes como tendencia, estacionalidad y estacionaridad. Su principal objetivo, es el pronóstico de datos futuros de la variable aleatoria para lo cual es importante contar con un modelo adecuado que describa los datos generados por el fenómeno en estudio.

Definición 2.1 *Un proceso estocástico es una familia de variables aleatorias $\{X_t, t \in T\}$ definida sobre un espacio de probabilidad (Ω, \mathcal{F}, P) y asociada a un conjunto de índices T de números reales de forma tal que a cada elemento del conjunto le corresponda una y sólo una variable aleatoria.*

En el análisis de series de tiempo y en este documento el conjunto de índices T es un subconjunto de \mathbb{R} , algunas veces $\{0, 1, 2, \dots\}$, $\{1, 2, 3, \dots\}$, $[0, \infty)$, o $(-\infty, \infty)$ aunque también puede ocurrir que $T \not\subseteq \mathbb{R}$ [3].

Si T es un intervalo de números reales abierto o cerrado, se dice que el proceso estocástico es continuo, si T , es un conjunto finito o infinito numerable, entonces se dice que el proceso estocástico es discreto.

Definición 2.2 *Las funciones sobre T , tales que $t \rightarrow X_t(\omega)$, para cada $\omega \in \mathbb{R}$ se conocen como realizaciones del proceso $\{X_t, t \in T\}$. Si el conjunto de estados T , es discreto, entonces, al conjunto de realizaciones se le denomina **serie de tiempo**.*

Se sabe que el comportamiento de una variable aleatoria X puede caracterizarse por medio de su función de densidad $f(x)$, similarmente dos variables aleatorias X_1, X_2

quedan descritas por su función de densidad conjunta $f(x_1, x_2)$.

En prácticamente todo el análisis estadístico, es costumbre suponer que las observaciones que se tienen provienen de variables aleatorias independientes de forma tal que sólo con el conocimiento de las funciones de densidad individuales, es posible obtener fácilmente la función de densidad conjunta. En el caso de las series de tiempo se supone que existe toda una estructura de correlación entre las observaciones; por consiguiente no es posible obtener la función de densidad conjunta de manera directa y se debe utilizar alguna otra forma para caracterizar a las variables aleatorias que intervienen. Con este objetivo se presenta a continuación algunos operadores y polinomios que se utilizan frecuentemente en el análisis de series de tiempo.

2.1. Operadores y polinomios de retraso

Definición 2.3 *El operador de retraso denotado por B se define mediante la relación*

$$BX_t = X_{t-1}, \text{ para toda } t \cdot$$

Si se aplica de manera sucesiva se obtiene que

$$B^k X_t = X_{t-k}, \text{ para } k = 0, 1, \dots \text{ y toda } t \cdot$$

Definición 2.4 *El operador diferencia denotado por ∇ se utiliza para expresar relaciones del tipo $Y_t = X_t - X_{t-1}$, así*

$$\nabla X_t = X_t - X_{t-1}, \text{ para toda } t \text{ y}$$

$$Y_t = \nabla X_t \cdot$$

Se puede comprobar que $\nabla = 1 - B$, es decir

$$\nabla X_t = (1 - B)X_t,$$

de esta manera se define a ∇^k como

$$\nabla^k X_t = \sum_{j=0}^k \binom{j}{k} (-1)^j X_{t-j} = (1 - B)^k X_t \cdot$$

En el análisis de series de tiempo, también se utilizan operadores de retraso en forma de polinomios, así el polinomio

$$X_t - g_1 X_{t-1} - g_2 X_{t-2} - \dots - g_k X_{t-k} = X_t - \sum_{j=1}^k g_j X_{t-j} ,$$

es un polinomio de retraso que puede expresarse como $G(B)X_t$ en donde

$$G(B)X_t = \left(1 - g_1 B - g_2 B^2 - \dots - g_k B^k\right) X_t = 1 - \sum_{j=1}^k g_j B^j X_{t-j} ,$$

y los coeficientes g_i , $i = 1, 2, \dots, k$ son constantes que ponderan la importancia de los retrasos con los cuales están asociados, además k puede tomar valores $1, 2, \dots$. También es posible trabajar con polinomios de retraso racionales, los cuales pueden expresarse como cocientes de polinomios de retraso, así, si a_j y c_j son constantes, $G(B)$ será un polinomio racional si

$$G(B) = \frac{A(B)}{C(B)} \text{ con } A(B) = 1 - \sum_{j=1}^k a_j B^j \text{ y } C(B) = 1 - \sum_{j=1}^m c_j B^j ,$$

considerando que se debe restringir los coeficientes del g_j del polinomio $G(B)$ [11].

2.2. Ecuaciones en diferencia

2.2.1. Ecuaciones en diferencia de primer orden

Una ecuación en diferencia es una expresión que relaciona una variable y_t con sus valores previos.

Definición 2.5 *Una ecuación en diferencia de primer orden es una ecuación de la forma*

$$y_t = \phi y_{t-1} + \omega_t . \quad (2.1)$$

La cual se puede escribir como

$$(1 - \phi B)y_t = \omega_t .$$

Para cada dato se tiene una ecuación que relaciona el valor de y para el tiempo t con su valor anterior y_{t-1} y el valor actual de ω , es decir:

Tiempo	Ecuación
0	$y_0 = \phi y_{-1} + \omega_0$
1	$y_1 = \phi y_0 + \omega_1$
.	.
.	.
.	.
t	$y_t = \phi y_{t-1} + \omega_t$

Supóngase que se conoce el valor de y para $t = -1$ y el valor de ω para $t = 0, 1, 2, \dots$, entonces es posible encontrar el valor de y para cualquier tiempo t , en efecto; si se conoce el valor de y para $t = -1$ y el valor de ω para $t = 0$ se tiene

$$y_0 = \phi y_{-1} + \omega_0 \cdot$$

Ahora, dado y_0 , si se conoce ω_1 entonces

$$y_1 = \phi y_0 + \omega_1 = \phi(\phi y_{-1} + \omega_0) + \omega_1 = \phi^2 y_{-1} + \phi \omega_0 + \omega_1 \cdot$$

Continuando de manera recursiva se tiene

$$y_t = \phi^{t+1} y_{-1} + \sum_{i=0}^t \phi^{t-i} \omega_i \cdot \quad (2.2)$$

Al procedimiento anterior se le conoce como la **solución de la ecuación en diferencia por sustitución recursiva**.

Multiplicadores dinámicos

La expresión (2.2) expresa a y_t como una función lineal del valor inicial y_{-1} y los valores históricos de ω lo que hace fácil calcular el efecto de ω_0 sobre y_t , el cual está dado por:

$$\frac{\partial y_t}{\partial \omega_0} = \phi^t \cdot$$

Si la ecuación en diferencia empieza en el tiempo t , tomando y_{t-1} conocido, al realizar cálculos similares a los anteriores, y_{t+j} se puede expresar como una función de y_{t-1} y $\omega_t, \omega_{t+1}, \dots, \omega_{t+j}$ como

$$y_{t+j} = \phi^{j+1} y_{t-1} + \sum_{i=0}^j \phi^{j-i} \omega_{t+i} \cdot \quad (2.3)$$

y el efecto de ω_t sobre y_{t+j} está dado por

$$\frac{\partial y_{t+j}}{\partial \omega_t} = \phi^j .$$

Observación 2.1 *Diferentes valores de ϕ en (2.1), pueden producir una variedad de respuestas dinámicas de y en ω . En efecto,*

1. Si $0 < \phi < 1$, el multiplicador $\frac{\partial y_{t+j}}{\partial \omega_t} = \phi^j \rightarrow 0$, cuando $j \rightarrow \infty$.
2. Si $-1 < \phi < 0$, el multiplicador $\frac{\partial y_{t+j}}{\partial \omega_t} = \phi^j$ decrece a 0 alternándose en signo, cuando $j \rightarrow \infty$.
3. Si $\phi > 1$, el multiplicador $\frac{\partial y_{t+j}}{\partial \omega_t} = \phi^j$ crece exponencialmente cuando $j \rightarrow \infty$.
4. Si $\phi < -1$, el sistema (2.1) exhibe oscilaciones explosivas cuando $j \rightarrow \infty$.

Los casos anteriores se pueden visualizar en la Figura 2.1.

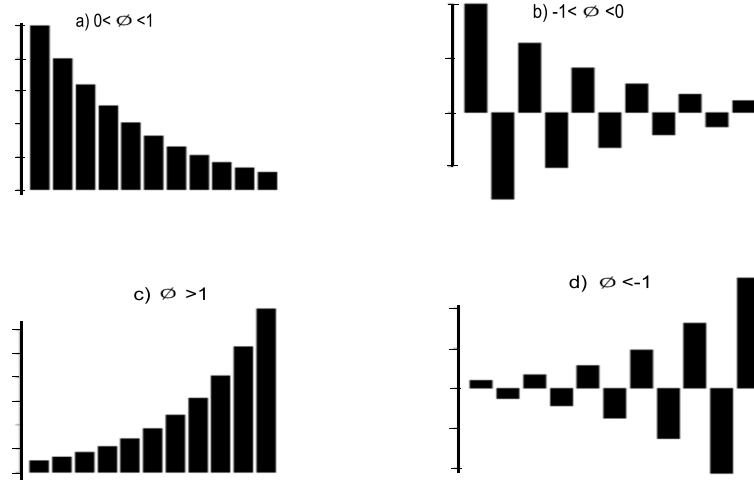


Figura 2.1: Análisis de un multiplicador dinámico para una ecuación en diferencia de primer orden

2.2.2. Ecuaciones en diferencia de p -ésimo orden

Una ecuación en diferencia de orden p , es una expresión de la forma

$$y_t = \sum_{i=1}^p \phi_i y_{t-i} + \omega_t , \quad (2.4)$$

la cual se puede expresar como

$$\left(1 - \sum_{i=1}^p \phi_i B^i\right) y_t = \omega_t \cdot$$

Es conveniente reescribir a (2.4) como una ecuación de primer orden, para ésto, se considera

$$\boldsymbol{\xi}_t = (y_t, y_{t-1}, \dots, y_{t-p+1})' ,$$

la matriz \mathbf{F} de tamaño $p \times p$

$$\mathbf{F} = \begin{bmatrix} \phi_1 & \phi_2 & \dots & \phi_{p-1} & \phi_p \\ 1 & 0 & \dots & 0 & 0 \\ \vdots & \vdots & \dots & \vdots & \vdots \\ 0 & 0 & \dots & 1 & 0 \end{bmatrix} , \quad (2.5)$$

y el vector \mathbf{v}_t por

$$\mathbf{v}_t = (\omega_t, 0, \dots, 0)' ,$$

para expresar la ecuación en diferencia de p -ésimo orden como

$$\boldsymbol{\xi}_t = \mathbf{F}\boldsymbol{\xi}_{t-1} + \mathbf{v}_t \cdot \quad (2.6)$$

Donde el primer término del sistema (2.6) es la expresión (2.4) y el resto sólo es $y_{t-i} = y_{t-i}$, $i = 1, 2, \dots, p-1$, así, (2.6) es una representación matricial del sistema (2.4).

De manera similar a las ecuaciones en diferencia de primer orden, supóngase que se conoce el vector $\boldsymbol{\xi}$ para $t = -1$ y el vector \mathbf{v} para $t = 0$, entonces

$$\boldsymbol{\xi}_0 = \mathbf{F}\boldsymbol{\xi}_{-1} + \mathbf{v}_0 \cdot$$

El valor de $\boldsymbol{\xi}$ para $t = 1$ es

$$\boldsymbol{\xi}_1 = \mathbf{F}\boldsymbol{\xi}_0 + \mathbf{v}_1 = \mathbf{F}(\mathbf{F}\boldsymbol{\xi}_{-1} + \mathbf{v}_0) + \mathbf{v}_1 = \mathbf{F}^2\boldsymbol{\xi}_{-1} + \mathbf{F}\mathbf{v}_0 + \mathbf{v}_1 \cdot$$

Siguiendo de manera recursiva se tiene que:

$$\boldsymbol{\xi}_t = \mathbf{F}^{t+1}\boldsymbol{\xi}_{-1} + \sum_{i=0}^t \mathbf{F}^{t-i}\mathbf{v}_i ,$$

que en notación matricial se puede expresar como

$$\begin{bmatrix} y_t \\ y_{t-1} \\ \vdots \\ y_{t-p+1} \end{bmatrix} = \mathbf{F}^{t+1} \begin{bmatrix} y_{-1} \\ y_{-2} \\ \vdots \\ y_{-p} \end{bmatrix} + \mathbf{F}^t \begin{bmatrix} \omega_0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} + \mathbf{F}^{t-1} \begin{bmatrix} \omega_1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} + \dots + \mathbf{F} \begin{bmatrix} \omega_{t-1} \\ 0 \\ \vdots \\ 0 \end{bmatrix} + \begin{bmatrix} \omega_t \\ 0 \\ \vdots \\ 0 \end{bmatrix} . \quad (2.7)$$

Si se considera la primer expresión de (2.7), la cual caracteriza el valor de y_t y se denota por $f_{11}^{(t)}$ como el primer elemento de \mathbf{F}^t , $f_{12}^{(t)}$ el elemento (1, 2) de \mathbf{F}^t y así sucesivamente, se tiene que:

$$y_t = f_{11}^{(t+1)} y_{-1} + f_{12}^{(t+1)} y_{-2} + \dots + f_{1p}^{(t+1)} y_{-p} + f_{11}^{(t)} \omega_0 + f_{11}^{(t-1)} \omega_1 + \dots + f_{11} \omega_{t-1} + \omega_t .$$

La cual describe el valor de y en el tiempo t como una función lineal con p valores iniciales de y , $y_{-1}, y_{-2}, \dots, y_{-p}$ y los valores históricos de la variable de entrada $\omega_0, \omega_1, \dots, \omega_t$.

La generalización de (2.3) es

$$\boldsymbol{\xi}_{t+j} = \mathbf{F}^{j+1} \boldsymbol{\xi}_{t-1} + \mathbf{F}^j \mathbf{v}_t + \mathbf{F}^{j-1} \mathbf{v}_{t-1} + \dots + \mathbf{F} \mathbf{v}_{t+j-1} + \mathbf{v}_{t+j} . \quad (2.8)$$

Luego para una ecuación en diferencia de p -ésimo orden, el multiplicador dinámico está dado por

$$\frac{\partial y_{t+j}}{\partial \omega_t} = f_{11}^{(j)} .$$

El elemento (1, 1) de \mathbf{F}^j se puede obtener fácilmente en términos de los eigenvalores de la matriz \mathbf{F} .

Lema 2.1 *Los eigenvalores de la matriz \mathbf{F} definida en (2.5) son los valores de λ que satisfacen*

$$\lambda^p - \phi_1 \lambda^{p-1} - \phi_2 \lambda^{p-2} - \dots - \phi_{p-1} \lambda - \phi_p = 0 . \quad (2.9)$$

2.2.3. Solución general de una ecuación en diferencia de p -ésimo orden con eigenvalores distintos

Por el teorema de la descomposición espectral, si los eigenvalores de \mathbf{F} definida en (2.5) son distintos, existe una matriz invertible T de tamaño $p \times p$ tal que

$$\mathbf{F} = \mathbf{T}\mathbf{\Lambda}\mathbf{T}^{-1} . \quad (2.10)$$

Donde la matriz $\mathbf{\Lambda}$ de tamaño $p \times p$ es una matriz diagonal formada por los eigenvalores de \mathbf{F} ,

$$\mathbf{\Lambda} = \begin{bmatrix} \lambda_1 & 0 & 0 & \cdots & 0 & 0 \\ 0 & \lambda_2 & 0 & \cdots & 0 & 0 \\ 0 & 0 & \cdots & 0 & 0 & \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & \lambda_{p-1} & 0 \\ 0 & 0 & 0 & \cdots & 0 & \lambda_p \end{bmatrix} .$$

La expresión (2.10) permite caracterizar de una manera más sencilla el elemento $(1, 1)$ de \mathbf{F}^j , pues

$$\mathbf{F}^j = \underbrace{\mathbf{T}\mathbf{\Lambda}\mathbf{T}^{-1} \times \mathbf{T}\mathbf{\Lambda}\mathbf{T}^{-1} \times \cdots \times \mathbf{T}\mathbf{\Lambda}\mathbf{T}^{-1}}_{j\text{-veces}} = \mathbf{T}\mathbf{\Lambda}^j\mathbf{T}^{-1} . \quad (2.11)$$

La estructura diagonal de $\mathbf{\Lambda}$ implica que $\mathbf{\Lambda}^j$ también es una matriz diagonal cuyos elementos son los eigenvalores de \mathbf{F} elevados a la potencia j .

Sea t_{ij} el elemento del i -ésimo renglón y la j -ésima columna de T y t^{ij} el elemento del i -ésimo renglón y la j -ésima columna de T^{-1} , la ecuación (2.11) se puede escribir de manera explícita como

$$\begin{aligned} \mathbf{F}^j &= \begin{bmatrix} t_{11} & t_{12} & \cdots & t_{1p} \\ t_{21} & t_{22} & \cdots & t_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ t_{p1} & t_{p2} & \cdots & t_{pp} \end{bmatrix} \begin{bmatrix} \lambda_1^j & 0 & \cdots & 0 \\ 0 & \lambda_2^j & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_p^j \end{bmatrix} \begin{bmatrix} t^{11} & t^{12} & \cdots & t^{1p} \\ t^{21} & t^{22} & \cdots & t^{2p} \\ \vdots & \vdots & \ddots & \vdots \\ t^{p1} & t^{p2} & \cdots & t^{pp} \end{bmatrix} \\ &= \begin{bmatrix} t_{11}\lambda_1^j & t_{12}\lambda_2^j & \cdots & t_{1p}\lambda_p^j \\ t_{21}\lambda_1^j & t_{22}\lambda_2^j & \cdots & t_{2p}\lambda_p^j \\ \vdots & \vdots & \ddots & \vdots \\ t_{p1}\lambda_1^j & t_{p2}\lambda_2^j & \cdots & t_{pp}\lambda_p^j \end{bmatrix} \begin{bmatrix} t^{11} & t^{12} & \cdots & t^{1p} \\ t^{21} & t^{22} & \cdots & t^{2p} \\ \vdots & \vdots & \ddots & \vdots \\ t^{p1} & t^{p2} & \cdots & t^{pp} \end{bmatrix} , \end{aligned}$$

donde el elemento $(1, 1)$ de \mathbf{F}^j está dado por

$$f_{11}^{(j)} = \sum_{i=1}^p [t_{1i} t^{i1}] \lambda_i^j = \sum_{i=1}^p c_i \lambda_i^j, \quad (2.12)$$

donde $c_i = t_{1i} t^{i1}$, para $i = 1, 2, \dots, p$.

Por lo tanto,

$$\frac{\partial y_{t+j}}{\partial \omega_t} = f_{11}^{(j)} = \sum_{i=1}^p c_i \lambda_i^j.$$

Observación 2.2 Note que $\sum_{i=1}^p c_i = t_{11} t^{11} + t_{22} t^{22} + \dots + t_{1p} t^{p1}$ es el elemento $(1, 1)$ de TT^{-1} , la cual es la matriz identidad de tamaño $p \times p$, lo que implica que $\sum_{i=1}^p c_i = 1$.

Proposición 2.1 Si los eigenvalores λ_i , $i = 1, 2, \dots, p$ de la matriz \mathbf{F} definida en (2.5) son distintos, entonces

$$c_i = \frac{\lambda_i^{p-1}}{\prod_{k=1}^p (\lambda_i - \lambda_k)}. \quad (2.13)$$

Demostración. Supongamos que λ_i , $i = 1, 2, \dots, p$ son eigenvalores distintos de \mathbf{F} , la matriz \mathbf{T} de la descomposición $\mathbf{F} = \mathbf{T}\mathbf{\Lambda}\mathbf{T}^{-1}$ se puede construir con los eigenvectores de \mathbf{F} . Sea \mathbf{t}_i un p -vector de la forma

$$\mathbf{t}_i = \begin{bmatrix} \lambda_i^{p-1} \\ \lambda_i^{p-2} \\ \vdots \\ \lambda_i \\ 1 \end{bmatrix}, \quad (2.14)$$

donde λ_i denota el i -ésimo eigenvalor de \mathbf{F} . Obsérvese que

$$\begin{aligned} \mathbf{F}\mathbf{t}_i &= \begin{bmatrix} \phi_1 & \phi_2 & \phi_3 & \cdots & \phi_{p-1} & \phi_p \\ 1 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 1 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \cdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 1 & 0 \end{bmatrix} \begin{bmatrix} \lambda_i^{p-1} \\ \lambda_i^{p-2} \\ \vdots \\ \lambda_i \\ 1 \end{bmatrix} \\ &= \begin{bmatrix} \sum_{k=1}^p \phi_k \lambda_i^{p-k} \\ \lambda_i^{p-1} \\ \vdots \\ \lambda_i \end{bmatrix} . \end{aligned}$$

Como λ_i es un eigenvalor de \mathbf{F} , satisface (2.9), así $\lambda_i^p = \sum_{k=1}^p \phi_k \lambda_i^{p-k}$ por lo que

$$\mathbf{F}\mathbf{t}_i = \begin{bmatrix} \lambda_i^p \\ \lambda_i^{p-1} \\ \vdots \\ \lambda_i \end{bmatrix} = \lambda_i \mathbf{t}_i ,$$

es decir; \mathbf{t}_i es un eigenvector de \mathbf{F} asociado con el eigenvalor λ_i .

Se puede calcular la matriz \mathbf{T} combinando los eigenvectores $(\mathbf{t}_1, \dots, \mathbf{t}_p)$ en una matriz de tamaño $p \times p$,

$$\mathbf{T} = [\mathbf{t}_1, \dots, \mathbf{t}_p] . \quad (2.15)$$

Para calcular los valores de c_i , $i = 1, 2, \dots, p$, se sabe

$$\mathbf{T}\mathbf{T}^{-1} = \mathbf{I}_p , \quad (2.16)$$

donde \mathbf{T} está dada por (2.14) y (2.15).

Escribiendo la primer columna del sistema de ecuaciones (2.16), se tiene que

$$\begin{bmatrix} \lambda_1^{p-1} & \lambda_2^{p-1} & \cdots & \lambda_p^{p-1} \\ \lambda_1^{p-2} & \lambda_2^{p-2} & \cdots & \lambda_p^{p-2} \\ \vdots & \vdots & \cdots & \vdots \\ \lambda_1 & \lambda_2 & \cdots & \lambda_p \\ 1 & 1 & \cdots & 1 \end{bmatrix} \begin{bmatrix} t^{11} \\ t^{21} \\ \vdots \\ t^{p-1,1} \\ t^{p1} \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \\ 0 \end{bmatrix} . \quad (2.17)$$

El cual es un sistema de ecuaciones de tamaño $p \times p$ con variables desconocidas (t_{11}, \dots, t_{p1}) . Siempre que los eigenvalores λ_i sean distintos, la solución de (2.17) está dada por

$$\begin{aligned} t^{11} &= \frac{1}{(\lambda_1 - \lambda_2)(\lambda_1 - \lambda_3) \cdots (\lambda_1 - \lambda_p)} \quad , \\ t^{21} &= \frac{1}{(\lambda_2 - \lambda_1)(\lambda_2 - \lambda_3) \cdots (\lambda_2 - \lambda_p)} \quad , \\ &\quad \vdots \\ t^{p1} &= \frac{1}{(\lambda_p - \lambda_1)(\lambda_p - \lambda_2) \cdots (\lambda_p - \lambda_{p-1})} \quad , \end{aligned}$$

luego,

$$c_i = t_{1i} t^{i1} = \frac{\lambda_i^{p-1}}{\prod_{k=1}^p (\lambda_i - \lambda_k)} \quad .$$

■

Observación 2.3 Para ecuaciones en diferencia de orden $p > 1$, (2.12) permite una variedad de multiplicadores dinámicos más complicados. Por ejemplo, para una ecuación en diferencia de segundo orden $y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \omega_t$,

$$F = \begin{bmatrix} \phi_1 & \phi_2 \\ 1 & 0 \end{bmatrix} \quad ,$$

cuyos eigenvalores se encuentran resolviendo

$$\lambda^2 - \phi_1 \lambda - \phi_2 = 0 \quad , \quad (2.18)$$

es decir,

$$\lambda = \frac{\phi_1 \pm \sqrt{\phi_1^2 + 4\phi_2}}{2} \quad .$$

1. Supóngase que todos los eigenvalores de F son distintos, es decir, $\phi_1^2 + 4\phi_2 > 0$. Si los eigenvalores λ_1 y λ_2 son ambos menores que la unidad en valor absoluto, se dice que el sistema es estable y sus multiplicadores dinámicos son representados como un promedio ponderado de exponenciales decrecientes o de exponenciales decrecientes oscilando en signo.
2. Si λ_1 y $\lambda_2 \in \mathbb{R}$ con $|\lambda_1| > 1$ ó $|\lambda_2| > 1$, se dice que el sistema es explosivo. Si λ_1 denota el eigenvalor más grande en valor absoluto, el multiplicador dinámico es eventualmente dominado por una función exponencial de aquel eigenvalor

$$\lim_{j \rightarrow \infty} \frac{\partial y_{t+j}}{\partial \omega_t} = c_1 \quad .$$

3. Otra posibilidad interesante ocurre cuando λ_1 y λ_2 son complejos, es decir, si $\phi_1^2 + 4\phi_2 < 0$ entonces λ_1 y λ_2 tienen la representación

$$\lambda_1 = a + ib, \quad \lambda_2 = a - ib,$$

donde $a = \frac{\phi_1}{2}$, $b = \frac{\sqrt{-\phi_1^2 - 4\phi_2}}{2}$. Utilizando coordenadas polares

$$\lambda_1 = R[\cos\theta + i\text{sen}\theta] , \quad (2.19)$$

donde $R = \sqrt{a^2 + b^2}$, $\cos\theta = \frac{a}{R}$ y $\text{sen}\theta = \frac{b}{R}$.

Por el teorema de Moivre, λ_1 en (2.19) se puede escribir como $R[e^{i\theta}]$, así

$$\lambda_1^j = R^j[e^{i\theta j}] = R^j[\cos(\theta j) + i\text{sen}(\theta j)] ,$$

de manera análoga,

$$\lambda_2^j = R^j[e^{-i\theta j}] = R^j[\cos(\theta j) - i\text{sen}(\theta j)] ,$$

y el multiplicador dinámico se expresa por

$$\frac{\partial y_{t+j}}{\partial \omega_t} = 2\alpha R^j \cos(\theta j) - 2\beta R^j \text{sen}(\theta j) .$$

En el siguiente apartado se presenta un análisis de los eigenvalores de una ecuación en diferencia de segundo orden.

2.2.4. Solución de una ecuación en diferencia de segundo orden con distintos eigenvalores

Para una ecuación en diferencia de segundo orden los eigenvalores λ_1 , λ_2 en (2.18) son complejos siempre que

$$\phi_1^2 + 4\phi_2 < 0 ,$$

o bien, siempre que (ϕ_1, ϕ_2) pertenezca a la parábola

$$\phi_1^2 = -4\phi_2 .$$

Si los eigenvalores son reales, sean λ_1 y λ_2 , el eigenvalor aritméticamente más grande y el eigenvalor aritméticamente más pequeño, respectivamente, λ_1 será mayor a la unidad siempre que

$$\frac{\phi_1 + \sqrt{\phi_1^2 + 4\phi_2}}{2} > 1 ,$$

o bien

$$\sqrt{\phi_1^2 + 4\phi_2} > 2 - \phi_1 .$$

Suponiendo que λ_1 es real, el lado izquierdo de la expresión anterior es un número positivo y la desigualdad se cumple para cualquier $\phi_1 > 2$, por otro lado, si $\phi_1 < 2$, se puede elevar al cuadrado ambos lados de la expresión para concluir que λ_1 excede la unidad siempre que $\phi_2 > 1 - \phi_1$.

Así en el plano real, $\lambda_1 > 1$ siempre que $\phi_1 > 2$ o las coordenadas (ϕ_1, ϕ_2) estén por arriba de la recta $\phi_2 = 1 - \phi_1$.

Similarmente λ_2 será menor que -1 siempre que

$$\sqrt{\phi_1^2 + 4\phi_2} > 2 + \phi_1 ,$$

de nuevo la desigualdad anterior se cumple si $\phi_1 < -2$, si $\phi_1 > -2$, se puede elevar al cuadrado ambos lados de la desigualdad para obtener

$$\phi_2 > 1 + \phi_1 ,$$

en el plano real, $\lambda_1 < -1$ si $\phi_1 < -2$ o las coordenadas (ϕ_1, ϕ_2) están por abajo de la recta $\phi_2 = 1 + \phi_1$.

El sistema se dice estable siempre que (ϕ_1, ϕ_2) se encuentre dentro de la región triangular de la Figura 2.3.

2.2.5. Solución de una ecuación en diferencia de p -ésimo orden con eigenvalores repetidos

En el caso de una ecuación en diferencia de orden p en la que la matriz \mathbf{F} tiene eigenvalores repetidos y $s < p$ eigenvectores linealmente independientes, la expresión (2.10) se generaliza utilizando la descomposición de Jordan, es decir, se considera \mathbf{F} de la forma

$$\mathbf{F} = \mathbf{M}\mathbf{J}\mathbf{M}^{-1} , \quad (2.20)$$

donde \mathbf{M} es una matriz de tamaño $p \times p$ y \mathbf{J} toma la forma

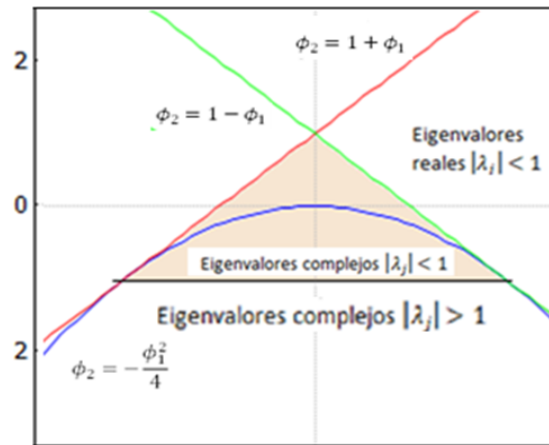


Figura 2.2: Resumen de los multiplicadores dinámicos de una ecuación en diferencia de segundo orden

$$\mathbf{J} = \begin{bmatrix} \mathbf{J}_1 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{J}_2 & \cdots & \mathbf{0} \\ \vdots & \vdots & \cdots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{J}_s \end{bmatrix},$$

y

$$\mathbf{J}_i = \begin{bmatrix} \lambda_i & 1 & 0 & \cdots & 0 \\ 0 & \lambda_i & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & 0 & \cdots & \lambda_i \end{bmatrix},$$

para λ_i un eigenvalor de \mathbf{F} . Si (2.10) se reemplaza por (2.20), entonces (2.11) se generaliza a

$$\mathbf{F}^j = \mathbf{M} \mathbf{J}^j \mathbf{M}^{-1}, \quad (2.21)$$

donde

$$\mathbf{J}^j = \begin{bmatrix} \mathbf{J}_1^j & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{J}_2^j & \cdots & \mathbf{0} \\ \vdots & \vdots & \cdots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{J}_s^j \end{bmatrix},$$

además de (2.2.5), si \mathbf{J}_i es de dimensión $n_i \times n_i$ entonces

$$\mathbf{J}_i^j = \begin{bmatrix} \lambda_i^j & \binom{j}{1} \lambda_i^{j-1} & \binom{j}{2} \lambda_i^{j-2} & \dots & \binom{j}{n_i-1} \lambda_i^{j-n_i+1} \\ 0 & \lambda_i^j & \binom{j}{1} \lambda_i^{j-1} & \dots & \binom{j}{n_i-1} \lambda_i^{j-n_i+2} \\ \vdots & \vdots & \vdots & \dots & \vdots \\ 0 & 0 & 0 & \dots & \lambda_i^j \end{bmatrix},$$

donde

$$\binom{j}{n} = \begin{cases} \frac{j(j-1)(j-2)\dots(j-n+1)}{n!}, & \text{si } j \geq n, \\ 0, & \text{en otro caso.} \end{cases}$$

2.2.6. Ecuaciones en diferencia utilizando operadores de retraso

Utilizando operadores de retraso la ecuación en diferencia de orden p , (2.4) se puede expresar como

$$(1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p) y_t = \omega_t,$$

y

$$(1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p) = (1 - \lambda_1 B)(1 - \lambda_2 B) \dots (1 - \lambda_p B). \quad (2.22)$$

Esto es lo mismo que encontrar los valores de $\lambda_1, \lambda_2, \dots, \lambda_p$ tales que los siguientes polinomios son los mismos para todo z :

$$(1 - \phi_1 z - \phi_2 z^2 - \dots - \phi_p z^p) = (1 - \lambda_1 z)(1 - \lambda_2 z) \dots (1 - \lambda_p z),$$

al multiplicar ambos lados de la expresión anterior por z^{-p} y definiendo $\lambda = z^{-1}$ se tiene

$$\lambda^p - \phi_1 \lambda^{p-1} - \dots - \phi_p = (\lambda - \lambda_1)(\lambda - \lambda_2) \dots (\lambda - \lambda_p). \quad (2.23)$$

Ajustando $\lambda = \lambda_i$, $i = 1, 2, \dots, p$; el lado derecho de la expresión de (2.23) es igual a cero. Así los valores $\lambda_1, \lambda_2, \dots, \lambda_p$ deben ser los valores numéricos que ajustan el lado izquierdo de la expresión (2.23) a cero, es decir,

$$\lambda^p - \phi_1 \lambda^{p-1} - \dots - \phi_p = 0. \quad (2.24)$$

La expresión (2.24) es idéntica a la expresión dada en el Lema 2.1 la cual caracteriza los eigenvalores $\lambda_1, \dots, \lambda_p$, de la matriz \mathbf{F} de (2.5).

Lema 2.2 *La factorización de un polinomio de operadores de retraso*

$$(1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p) = (1 - \lambda_1 B)(1 - \lambda_2 B) \cdot \dots \cdot (1 - \lambda_p B) ,$$

requiere de los mismos cálculos utilizados para encontrar los eigenvalores de la matriz \mathbf{F} definida en (2.5). Los eigenvalores $\lambda_1, \lambda_2, \dots, \lambda_p$, de \mathbf{F} son los mismos que los parámetros $\lambda_1, \lambda_2, \dots, \lambda_p$ en (2.22) y están dados por las soluciones de la ecuación (2.24).

La ecuación en diferencias (2.4) es estable si los eigenvalores están dentro del círculo unitario, o equivalentemente si las raíces de

$$1 - \phi_1 z - \dots - \phi_p z^p = 0 ,$$

se encuentran fuera del círculo unitario.

Suponiendo que los eigenvalores se encuentran dentro del círculo unitario y al considerar sucesiones acotadas, los inversos $(1 - \lambda_1 B)^{-1}, (1 - \lambda_2 B)^{-1}, \dots, (1 - \lambda_p B)^{-1}$ existen, permitiendo que la ecuación en diferencia

$$(1 - \lambda_1 B) \cdot \dots \cdot (1 - \lambda_p B) y_t = \omega_t , \quad (2.25)$$

se pueda escribir como

$$y_t = (1 - \lambda_1 B)^{-1} (1 - \lambda_2 B)^{-1} \cdot \dots \cdot (1 - \lambda_p B)^{-1} \omega_t . \quad (2.26)$$

Siempre que los eigenvalores $\lambda_1, \lambda_2, \dots, \lambda_p$ sean todos distintos, el polinomio asociado con el operador del lado derecho de (2.26) se pueden expandir con fracciones parciales:

$$\frac{1}{(1 - \lambda_1 z)(1 - \lambda_2 z) \cdot \dots \cdot (1 - \lambda_p z)} = \frac{c_1}{(1 - \lambda_1 z) + \dots + \frac{c_p}{(1 - \lambda_p z)}} ,$$

así,

$$1 = c_1(1 - \lambda_2 \lambda_1^{-1})(1 - \lambda_3 \lambda_1^{-1}) \cdot \dots \cdot (1 - \lambda_p \lambda_1^{-1}) ,$$

o equivalentemente

$$c_1 = \frac{\lambda_1^{p-1}}{(\lambda_1 - \lambda_2)(\lambda_1 - \lambda_3) \cdot \dots \cdot (\lambda_1 - \lambda_p)} . \quad (2.27)$$

Para $z = \lambda_2^{-1}, \dots, \lambda_p^{-1}$, se tiene

$$c_2 = \frac{\lambda_2^{p-1}}{(\lambda_2 - \lambda_1)(\lambda_2 - \lambda_3) \cdots (\lambda_2 - \lambda_p)},$$

$$\vdots$$

$$c_p = \frac{\lambda_p^{p-1}}{(\lambda_p - \lambda_1)(\lambda_p - \lambda_2) \cdots (\lambda_p - \lambda_{p-1})} .$$

Obsérvese que de nuevo estas expresiones son idénticas a la expresión (2.13). Para concluir, (2.26) se puede escribir como

$$y_t = \frac{c_1}{1 - \lambda_1 B} \omega_t + \frac{c_2}{1 - \lambda_2 B} \omega_t + \dots + \frac{c_p}{1 - \lambda_p B} \omega_t$$

$$= c_1(1 + \lambda_1 B + \lambda_1^2 B^2 + \dots) \omega_t + c_2(1 + \lambda_2 B + \lambda_2^2 B^2 + \dots) \omega_t$$

$$+ \dots + c_p(1 + \lambda_p B + \lambda_p^2 B^2 + \dots) \omega_t ,$$

ó

$$y_t = [c_1 + \dots + c_p] \omega_t + [c_1 \lambda_1 + c_2 \lambda_2 + \dots + c_p \lambda_p] \omega_{t-1}$$

$$+ [c_1 \lambda_1^2 + c_2 \lambda_2^2 + \dots + c_p \lambda_p^2] \omega_{t-2} +$$

$$+ [c_1 \lambda_1^3 + c_2 \lambda_2^3 + \dots + c_p \lambda_p^3] \omega_{t-3} + \dots . \quad (2.28)$$

De nuevo el multiplicador dinámico está dado por

$$\frac{\partial y_{t+j}}{\partial \omega_t} = [c_1 \lambda_1^j + c_2 \lambda_2^j + \dots + c_p \lambda_p^j] .$$

Una forma conveniente para determinar el efecto de ω sobre el valor presente de y , es a través del uso de la representación con operadores de retraso, en efecto, escribiendo (2.2.6) como

$$y_t = \psi_0 \omega_t + \psi_1 \omega_{t-1} + \psi_2 \omega_{t-2} + \psi_3 \omega_{t-3} + \dots , \quad (2.29)$$

donde

$$\psi_j = [c_1 \lambda_1^j + c_2 \lambda_2^j + \dots + c_p \lambda_p^j] .$$

Al reescribir (2.29) en notación de operadores de retraso se tiene

$$y_t = \psi(B) \omega_t , \quad (2.30)$$

donde $\psi(B)$ denota un polinomio de orden infinito en el operador de retraso

$$\psi(B) = \psi_0 + \psi_1 B + \psi_2 B^2 + \psi_3 B^3 + \dots .$$

El efecto de ω_t sobre el valor presente de y está dado por

$$\begin{aligned} \frac{\partial \sum_{j=0}^{\infty} \beta^j y_{t+j}}{\omega_t} &= \sum_{j=0}^{\infty} \beta^j \frac{\partial y_{t+j}}{\partial \omega_t} \\ &= \sum_{j=0}^{\infty} \beta^j \psi_j . \end{aligned} \quad (2.31)$$

Pensando en $\psi(z)$ como un polinomio real en z ,

$$\psi(z) = \psi_0 + \psi_1 z + \psi_2 z^2 + \psi_3 z^3 + \dots ,$$

la expresión (2.31) es simplemente el polinomio $\psi(z)$ evaluado en $z = \beta$;

$$\frac{\partial \sum_{j=0}^{\infty} \beta^j y_{t+j}}{\partial \omega_t} = \psi(\beta) = \psi_0 + \psi_1 \beta + \psi_2 \beta^2 + \psi_3 \beta^3 + \dots . \quad (2.32)$$

Comparando (2.30) con (2.25), se hace evidente que

$$\psi(B) = [(1 - \lambda_1 B)(1 - \lambda_2 B) \dots (1 - \lambda_p B)]^{-1} ,$$

de (2.22) lo anterior significa que

$$\psi(B) = [1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p]^{-1} ,$$

por lo que se concluye que

$$\psi(z) = [1 - \phi_1 z - \phi_2 z^2 - \dots - \phi_p z^p]^{-1} . \quad (2.33)$$

Para cualquier valor de z , en particular, para $z = \beta$

$$\psi(\beta) = [1 - \phi_1 \beta - \phi_2 \beta^2 - \dots - \phi_p \beta^p]^{-1} ,$$

finalmente, se sustituye (2.33) en (2.32), para obtener

$$\frac{\partial \sum_{j=0}^{\infty} \beta^j y_{t+j}}{\partial \omega_t} = \frac{1}{1 - \phi_1 \beta - \phi_2 \beta^2 - \dots - \phi_p \beta^p} .$$

2.3. Función de autocovarianza y función de autocorrelación

Definición 2.6 Si $\{X_t\}$ es un proceso tal que $E(X_t^2) < \infty$ para cada $t \in T$, entonces la función de autocovarianza $\gamma_X(\cdot, \cdot)$ de $\{X_t\}$ se define por

$$\gamma_X(r, s) = \text{Cov}(X_r, X_s) = E[(X_r - E[X_r])(X_s - E[X_s])], \quad r, s \in T \cdot$$

Definición 2.7 La serie de tiempo $\{X_t\}$ con conjunto de índices $T = \{0, \pm 1, \pm 2, \dots\}$ es débilmente estacionaria si

1. $E|X_t^2| < \infty$.
2. $E[X_t] = m$, para todo $t \in \mathbb{Z}$.
3. $\gamma_X(r, s) = \gamma_X(r + t, s + t)$, para todo $r, s, t \in \mathbb{Z}$.

Observación 2.4 Si $\{X_t\}$ es estacionaria, entonces $\gamma_X(r, s) = \gamma_X(r - s, 0)$ para todo $r, s \in \mathbb{Z}$. En efecto, por ser $\{X_t\}$ estacionaria, se cumple la Condición 3 de la Definición 2.7, luego tomando $t = -s$, se tiene $\gamma_X(r, s) = \gamma_X(r - s, 0)$. Por lo tanto, es conveniente redefinir la función de autocovarianza de un proceso como función de una sola variable.

Definición 2.8 Si $\{X_t\}$ es una serie de tiempo débilmente estacionaria, se define

$$\gamma_X(h) = \gamma_X(h, 0) = \text{Cov}(X_{t+h}, X_t), \quad \forall h \in \mathbb{Z} \cdot,$$

como la función de autocovarianza de $\{X_t\}$ y a

$$\rho_X(h) = \frac{\gamma_X(h)}{\gamma_X(0)},$$

como la función de autocorrelación de $\{X_t\}$.

Observación 2.5 Si $\{X_t\}$ es una serie de tiempo débilmente estacionaria, entonces la función de autocovarianza es independiente de t , y está dada por $\text{Cov}(X_{t+h}, X_t)$.

La función $\gamma_X(\cdot)$ tiene las siguientes propiedades:

1. $\gamma_X(0) \geq 0$.

2. $|\gamma_X(h)| \leq \gamma_X(0)$.
3. $\gamma_X(h) = \gamma_X(-h)$.

Demostración.

1. $\gamma_X(0) = E[(X_t - E[X_t])^2] = \text{Var}(X_t) \geq 0$.
2. $|\frac{\gamma_X(h)}{\gamma_X(0)}| \leq 1 \iff |\gamma_X(h)| \leq \gamma_X(0)$, si $\gamma_X(0) > 0$.
3. $\gamma_X(h) = \text{Cov}(X_{t+h}, X_t) = \text{Cov}(X_t, X_{t+h}) = \gamma_X(-h)$.

■

Definición 2.9 Una función κ con valores reales definida sobre los enteros, es definida no negativa si

$$\sum_{i,j=1}^n a_i \kappa(i-j) a_j \geq 0 \cdot$$

Para todo entero positivo n y vectores $\mathbf{a} = (a_1, a_2, \dots, a_n)'$.

Teorema 2.1 Una función con valores reales definida sobre los enteros es la función de autocovarianza de una serie de tiempo estacionaria si y sólo si es par y definida no negativa.

Demostración. \Rightarrow) Sea $\gamma_X(\cdot)$ una función tal que $\gamma_X(\cdot)$ es la función de autocovarianza de una serie de tiempo $\{X_t\}$, por propiedades de la función, $\gamma_X(\cdot)$ es par. Sea $\mathbf{a} = (a_1, \dots, a_n)'$ cualquier vector con componentes reales, $\mathbf{X}_n = (X_1, \dots, X_n)'$ y $\Gamma_n = (\sigma_{ij}), i, j = 1, 2, \dots, n$, la matriz de covarianza de \mathbf{X}_n .

$$\text{Var}(\mathbf{a}'\mathbf{X}_n) = \mathbf{a}'\Gamma_n\mathbf{a} = \sum_{i,j=1}^n a_i \gamma(i-j) a_j \geq 0 \cdot,$$

de lo anterior, $\gamma_X(\cdot)$ es definida no negativa.

\Leftarrow) Sea $\kappa : \mathbb{Z} \rightarrow \mathbb{R}$ una función par y definida no negativa. Se debe demostrar que existe un proceso estacionario con $\kappa(\cdot)$ como su función de autocovarianza, para esto se usará el Teorema de Kolmogorov ([3] p.11). Para cada entero n y cada $\mathbf{t} = (t_1, \dots, t_n) \in \mathbb{Z}^n$ tal que $t_1 < t_2 < \dots < t_n$. Sea $F_{\mathbf{t}}$ la función de distribución sobre \mathbb{R}^n con función característica $\Phi_{\mathbf{t}}(\mathbf{u}) = e^{\mathbf{u}'\mathbf{K}\mathbf{u}/2}$ donde $\mathbf{u}' = (u_1, \dots, u_n) \in \mathbb{R}^n$ y $\mathbf{K} = [\kappa(t_i - t_j)]_{i,j=1}^n$, como κ es definida no negativa, la matriz \mathbf{K} también es definida no negativa y consecuentemente $\phi_{\mathbf{t}}$ es la función característica de una

distribución multinormal con n variables con media cero y matriz de covarianza \mathbf{K} , luego utilizando la notación del Teorema de Kolmogorov,

$$\lim_{\mathbf{u}_t \rightarrow \mathbf{0}} \phi_t(\mathbf{u}) = \phi_{t(i)}(u(i))$$

es decir, las funciones F_t son consistentes, luego por el Teorema de la página 11 de [3], existe una serie de tiempo $\{X_t\}$ con funciones de distribución F_t y funciones características $\phi_t, t \in \mathcal{F}$. En particular la distribución conjunta de X_i y X_j es la distribución normal bivariada con media $\mathbf{0}$ y matriz de covarianza

$$\begin{bmatrix} \kappa(0) & \kappa(i-j) \\ \kappa(i-j) & \kappa(0) \end{bmatrix},$$

lo cual demuestra que $Cov(X_i, X_j) = \kappa(i-j)$. ■

Definición 2.10 *La serie de tiempo $\{X_t\}$ se dice estacionaria estricta si las distribuciones conjuntas de $(X_{t_1}, \dots, X_{t_k})'$ y $(X_{t_1+h}, \dots, X_{t_k+h})'$ son las mismas para todos los enteros positivos k y para todo $t_1, \dots, t_k, h \in \mathbb{Z}$.*

Propiedades de series de tiempo estrictamente estacionarias

1. Las variables aleatorias $\{X_t\}$ son idénticamente distribuidas.
2. $(X_t, X_{t+h})' = (X_1, X_{1+h})'$ en distribución, $\forall t \in \mathbb{Z}, h \in \mathbb{Z}$.
3. Si $E(X_t^2) < \infty$, $\{X_t\}$ es débilmente estacionaria.
4. Una sucesión independiente e idénticamente distribuida (*iid*) es estacionaria estricta.

Demostración. Sea $\{X_t\}$ una serie estacionaria estricta.

1. Por la definición de estacionaridad estricta, las variables aleatorias $\{X_t\}$ son idénticamente distribuidas.
2. Por la Propiedad 1, $(X_t, X_{t+h})' = (X_{t+h'}, X_{t+h+h'})'$ en distribución, en particular si $h' = 1-t$, entonces, $(X_t, X_{t+h})' = (X_{t+1-t}, X_{t+h+1-t})' = (X_1, X_{1+h})'$.
3. Supongamos que $E(X_t^2) < \infty, \forall t \in \mathbb{Z}$, como $\{X_t\}$ es un conjunto de variables aleatorias independientes e idénticamente distribuidas, $E(X_t)$ es independiente de t y por la Propiedad 2, $Cov(X_t, X_{t+h}) = Cov(X_1, X_{1+h})$ también es independiente de t , de donde se tiene que $\{X_t\}$ es débilmente estacionaria.

4. Sea $\{X_t\}$ una sucesión *iid* de variables aleatorias con distribución común F , entonces la distribución conjunta de $(X_{t_1}, \dots, X_{t_k})'$ y $(X_{t_1+h}, \dots, X_{t_k+h})'$ claramente son las mismas, independientemente de h , esto es, la sucesión es estacionaria estricta.

■

Observación 2.6 *Estacionaridad débil no implica estacionariedad estricta. En efecto, sea $\{Z_t\} \sim iid N(0, 1)$ y defínase*

$$X_t = \begin{cases} Z_t, & \text{si } t \text{ es par,} \\ \frac{Z_{t-1}^2 - 1}{\sqrt{2}}, & \text{si } t \text{ es impar.} \end{cases}$$

$$E[X_t] = \begin{cases} E[Z_t] = 0, & \text{si } t \text{ es par,} \\ E\left[\frac{Z_{t-1}^2 - 1}{\sqrt{2}}\right] = 0, & \text{si } t \text{ es impar.} \end{cases}$$

$$Var[X_t] = \begin{cases} E[Z_t^2] = 1, & \text{si } t \text{ es par,} \\ E\left[\left(\frac{Z_{t-1}^2 - 1}{\sqrt{2}}\right)^2\right] = 1, & \text{si } t \text{ es impar.} \end{cases}$$

$$\rho(h) = \begin{cases} E[Z_{t+h}Z_t] = 0, & \text{si } t \text{ y } h \text{ son pares,} \\ E[Z_{t+h}Z_t] = 0, & \text{si } t \text{ es impar y } h \text{ es par,} \\ E\left[Z_{t+h}\left(\frac{Z_{t-1}^2 - 1}{\sqrt{2}}\right)\right] = 0, & \text{si } t \text{ es impar y } h \text{ es impar.} \end{cases}$$

Lo que demuestra que $\{X_t\}$ es una sucesión de variables aleatorias con media 0 y varianza 1 que es débilmente estacionaria pero que no es estrictamente estacionaria puesto que $\{X_t\}$ no es idénticamente distribuida para todo t .

Observación 2.7

- i. A una serie de tiempo $\{X_t\}$ que es no correlacionada, con media 0 y varianza σ^2 se le conoce como una serie de ruido blanco y se denota como $Z_t \sim WN(0, \sigma^2)$.

- ii. Una manera simple de construir una serie de tiempo estrictamente estacionaria $\{X_t\}$ es filtrar una sucesión *iid* de variables aleatorias, para esto, sea $\{Z_t\}$ una sucesión *iid* que por la propiedad (4) es estrictamente estacionaria y defina

$$X_t = g(Z_t, Z_{t-1}, \dots, Z_{t-q}) .$$

Para alguna función con valores reales $g(\cdot, \dots, \cdot)$. Entonces $\{X_t\}$ es estacionaria, como $(Z_{t+h}, \dots, Z_{t+h-q})' = (Z_t, \dots, Z_{t-q})'$ en distribución, para todos los enteros h , se sigue de la ecuación (2.3) que $\{X_t\}$ es q -**dependiente**, esto es, X_s y X_t son independientes siempre que $|t - s| > q$. De igual manera, se dice que una sucesión es q -correlacionada si $\gamma(h) = 0$ siempre que $|h| > q$.

2.4. Propiedades de la media muestral y la función de autocorrelación

El primer momento muestral de un proceso estacionario es la media muestral

$$\bar{X}_n = n^{-1}(X_1 + \dots + X_n) ,$$

el cual es un estimador insesgado de la media μ del proceso. El error cuadrado medio de \bar{X}_n es

$$\begin{aligned} E(\bar{X}_n - \mu)^2 &= Var(\bar{X}_n) \\ &= n^{-2} \sum_{i=1}^n \sum_{j=1}^n Cov(X_i, X_j) \\ &= n^{-2} \sum_{i-j=1}^n (n - |i - j|) \gamma(i - j) \\ &= n^{-1} \sum_{h=-n}^n \left(1 - \frac{|h|}{n}\right) \gamma(h) , \end{aligned}$$

Ahora, si $\gamma(h) \rightarrow 0$ cuando $h \rightarrow \infty$, $E(\bar{X}_n - \mu)^2$ converge a 0 y si

$$\sum_{h=-\infty}^{\infty} |\gamma(h)| < \infty ,$$

entonces $\lim_{n \rightarrow \infty} Var(\bar{X}_n) = \sum_{|h| < \infty} \gamma(h)$.

Estimación de $\gamma(\cdot)$ y $\rho(\cdot)$

Se define la función de autocovarianza muestral y autocorrelación muestral como

$$\hat{\gamma}(h) = n^{-1} \sum_{t=1}^{n-|h|} (X_{t+|h|} - \bar{X}_n) (X_t - \bar{X}_n)$$

y

$$\hat{\rho}(h) = \frac{\hat{\gamma}(h)}{\hat{\gamma}(0)}$$

respectivamente, en donde n es el tamaño de la muestra. Estos dos estimadores son sesgados. La función de autocovarianza muestral tiene la propiedad de que para cada $k \geq 1$, la matriz de covarianza muestral es de dimensión k y es definida no negativa [2].

$$\hat{\Gamma}_k = \begin{bmatrix} \hat{\gamma}(0) & \hat{\gamma}(1) & \cdots & \hat{\gamma}(k-1) \\ \hat{\gamma}(1) & \hat{\gamma}(0) & \cdots & \hat{\gamma}(k-2) \\ \vdots & \vdots & \cdots & \vdots \\ \hat{\gamma}(k-1) & \hat{\gamma}(k-2) & \cdots & \hat{\gamma}(0) \end{bmatrix},$$

De manera intuitiva, una serie de tiempo es estacionaria si las propiedades estadísticas como la media y la varianza de la serie de tiempo son esencialmente constantes a través del tiempo. Si se tienen n valores observados y_1, y_2, \dots, y_n de una serie de tiempo se puede usar una gráfica de estos valores en función del tiempo para determinar si la serie de tiempo es estacionaria. Si los n valores fluctúan con variación constante respecto de una media constante μ , es razonable pensar que la serie de tiempo es estacionaria, en caso contrario, se puede pensar que la serie de tiempo no es estacionaria.

Si la serie de tiempo resulta ser no estacionaria, algunas veces se puede transformar la serie de tiempo no estacionaria, en valores de una serie de tiempo estacionaria mediante las primeras o segundas diferencias.

Las **primeras diferencias** de los valores de una serie de tiempo y_1, y_2, \dots, y_n son

$$y_t^* = y_t - y_{t-1}, t = 2, 3, \dots, n.$$

En ocasiones al obtener las primeras diferencias se transforman los valores de la serie de tiempo no estacionaria en valores de una serie de tiempo estacionaria, pero en otras ocasiones esto no ocurre por lo que se recomienda generar las segundas diferencias de los valores originales de la serie de tiempo.

Las **segundas diferencias** de los valores de una serie de tiempo y_1, y_2, \dots, y_n son

$$\begin{aligned} y_t^* &= (y_t - y_{t-1}) - (y_{t-1} - y_{t-2}) \\ &= y_t - 2y_{t-1} + y_{t-2} . \end{aligned}$$

Los modelos de pronóstico de Box-Jenkins se identifican en forma tentativa examinando el comportamiento de la función de autocorrelación muestral (FAM) y la función de autocorrelación muestral parcial (FAMP) para los valores de una serie de tiempo estacionaria $y_b^*, y_{b+1}^*, \dots, y_n^*$, donde y_i^* , $i = b, b+1, \dots, n$ pueden ser los valores originales de la serie de tiempo o los valores transformados.

Definición 2.11 *Considérese la serie de trabajo, $y_b^*, y_{b+1}^*, \dots, y_n^*$. Se define la función de autocorrelación muestral (FAM) en el desfase k como*

$$r_k = \frac{\sum_{t=b}^{n-k} (y_t^* - \bar{y}^*)(y_{t+k}^* - \bar{y}^*)}{\sum_{t=b}^n (y_t^* - \bar{y}^*)^2} ,$$

donde,

$$\bar{y}^* = \frac{\sum_{t=b}^n y_t^*}{n - b + 1} .$$

Definición 2.12 *Dada la serie de trabajo, $y_b^*, y_{b+1}^*, \dots, y_n^*$, se define la función de autocorrelación muestral parcial (FAMP) en el desfase k como*

$$r_{kk} = \begin{cases} r_1, & \text{si } k = 1, \\ \frac{r_k - \sum_{j=1}^{k-1} r_{k-1,j} r_{k-j}}{1 - \sum_{j=1}^{k-1} r_{k-1,j} r_j}, & \text{si } k = 2, 3, \dots, \end{cases}$$

donde $r_{kj} = r_{k-1,j} r_{k-1,k-j}$, para $j = 1, 2, \dots, k-1$.

Con el objeto de aplicar la metodología de Box-Jenkins, se puede demostrar que para datos no estacionales,

1. Si la FAM de los valores $y_b^*, y_{b+1}^*, \dots, y_n^*$ de la serie de tiempo se corta claramente con rapidez o si se corta rápidamente, entonces se debe considerar que los valores de la serie de tiempo son estacionarios.
2. Si la FAMP de los valores de la serie de tiempo $y_b^*, y_{b+1}^*, \dots, y_n^*$, se cortan con lentitud extrema, entonces se deben considerar que los valores de la serie de tiempo son no estacionarios.

2.5. Ejemplos de procesos estacionarios no estacionales

Un proceso es ergódico para la media siempre que la función de autocovarianza $\gamma_j \rightarrow 0$ suficientemente rápido conforme j es suficientemente grande.

2.5.1. Procesos de medias móviles (MA)

Procesos de Medias Móviles de Primer orden (MA(1))

Definición 2.13 Sea $\{Z_t\} \sim iid WN(0, \sigma^2)$, el proceso

$$X_t = \mu + Z_t + \theta Z_{t-1}, \text{ con } \mu, \theta, \in \mathbb{R}.$$

Es una serie de tiempo de medias móviles de primer orden.

Es fácil verificar que $E(X_t) = \mu$, $Var(X_t) = \sigma^2(1 + \theta^2)$ y

$$\begin{aligned} Cov(X_t, X_{t-1}) &= E[(X_t - \mu)(X_{t-1} - \mu)] \\ &= E[(Z_t + \theta Z_{t-1})(Z_{t-1} + \theta Z_{t-2})] \\ &= E[Z_t Z_{t-1} + \theta Z_{t-1}^2 + \theta Z_t Z_{t-2} + \theta^2 Z_{t-1} Z_{t-2}] \\ &= \theta \sigma^2, \end{aligned}$$

además

$$\begin{aligned} Cov(X_t, X_{t-j}) &= E[(X_t - \mu)(X_{t-j} - \mu)] \\ &= E[(Z_t + \theta Z_{t-1})(Z_{t-j} + \theta Z_{t-j-1})] \\ &= 0, \text{ para } j > 1. \end{aligned}$$

De lo anterior

$$\gamma(h) = \begin{cases} \sigma^2(1 + \theta^2), & \text{si } h = 0, \\ \theta \sigma^2, & \text{si } h = 1, \\ 0, & \text{si } h > 1. \end{cases}$$

Procesos de Medias Móviles de orden q ($MA(q)$)

Definición 2.14 $\{X_t\}$ es un proceso de medias móviles de orden q si

$$X_t = \mu + Z_t + \theta_1 Z_{t-1} + \dots + \theta_q Z_{t-q} \cdot$$

donde $\{Z_t\} \sim iid WN(0, \sigma^2)$ y $\theta_1, \dots, \theta_q$ son constantes.

De manera similar a un proceso $MA(1)$, $E(X_t) = \mu$ y

$$\begin{aligned} Var(X_t) &= E[(X_t - \mu)]^2 \\ &= E(Z_t + \theta_1 Z_{t-1} + \theta_2 Z_{t-2} + \dots + \theta_q Z_{t-q})^2 \\ &= \left(1 + \sum_{j=1}^q \theta_j^2\right) \sigma^2, \text{ Para } j = 1, 2, \dots, q, \end{aligned}$$

además

$$\begin{aligned} Cov(X_t, X_{t-j}) &= E[(Z_t + \theta_1 Z_{t-1} + \dots + \theta_q Z_{t-q})(Z_{t-j} + \theta_1 Z_{t-j-1} + \dots + \theta_q Z_{t-j-q})] \\ &= E(\theta_j Z_{t-j}^2 + \theta_{j+1} \theta_1 Z_{t-j-1}^2 + \theta_{j+2} \theta_2 Z_{t-j-2}^2 + \dots + \theta_q \theta_{q-j} Z_{t-q}^2) \cdot \end{aligned}$$

Así,

$$\gamma(h) = \begin{cases} 0, & \text{si } h = 0, \\ [\theta_h \theta_{h+1} \theta_1 + \theta_{h+2} \theta_2 + \dots + \theta_q \theta_{h-q}] \sigma^2, & \text{si } j = 1, 2, \dots, q, \\ 0, & \text{si } h > q. \end{cases}$$

Procesos de medias móviles de orden infinito ($MA(\infty)$).

El proceso $MA(q)$ se puede escribir como

$$X_t = \mu + \sum_{j=0}^q \theta_j Z_{t-j},$$

con $\theta_0 = 1$. Al considerar el proceso que resulta cuando $q \rightarrow \infty$,

$$X_t = \mu + \sum_{j=0}^{\infty} \psi_j Z_{t-j} , \quad (2.34)$$

el cual describe un proceso $MA(\infty)$.

Observación 2.8 *La sucesión infinita descrita en (2.34) genera un proceso de covarianza estacionaria siempre que*

$$\sum_{j=0}^{\infty} \psi_j^2 < \infty . \quad (2.35)$$

Para esto, se demostrará que sumabilidad absoluta de coeficientes de promedios de medias móviles implica sumabilidad cuadrática.

Supóngase que $\{\psi_j\}$ es absolutamente sumable, entonces existe $N < \infty$ tal que $|\psi_j| < 1$ para todo $j \geq N$, lo que implica $\psi_j^2 < |\psi_j|$ para todo $j \geq N$. Entonces

$$\sum_{j=0}^{\infty} \psi_j^2 = \sum_{j=0}^{N-1} \psi_j^2 + \sum_{j=N}^{\infty} \psi_j^2 < \sum_{j=0}^{N-1} \psi_j^2 + \sum_{j=N}^{\infty} |\psi_j| .$$

Pero $\sum_{j=0}^{\infty} \psi_j^2$ es finita, porque N es finito, y $\sum_{j=0}^{\infty} |\psi_j|$ es finita, pues $\{\psi_j\}$ es absolutamente sumable. Así, $\sum_{j=0}^{\infty} \psi_j^2 < \infty$, luego sumabilidad absoluta implica sumabilidad cuadrática.

Ahora se demostrará que sumabilidad cuadrática de coeficientes de medias móviles implica que un proceso $MA(\infty)$ en la representación (2.34) genera una variable aleatoria convergente en media cuadrática. Para esto hay que recordar el significado de convergencia de una suma determinística tal como $\sum_{j=0}^{\infty} a_j$ donde $\{a_j\}$ es sólo una sucesión de números.

Un criterio para determinar cuando $\sum_{j=0}^T a_j$ converge a algún número finito conforme $T \rightarrow \infty$ es el Criterio de Cauchy, el cual afirma que $\sum_{j=0}^{\infty} a_j$ converge si y sólo sí, para cualquier $\epsilon > 0$, existe un entero N tal que para cualquier entero $M > N$,

$$\left| \sum_{j=0}^M a_j - \sum_{j=0}^N a_j \right| < \epsilon . \quad (2.36)$$

Para un proceso estocástico como (2.36), el interés de la comparación es cuando $\sum_{j=0}^T \psi_j Z_{t-j}$ converge en media cuadrática a alguna variable aleatoria Y_t conforme $T \rightarrow \infty$. En este caso, el criterio de Cauchy asegura que $\sum_{j=0}^{\infty} \psi_j Z_{t-j}$ converge si y sólo sí, para cualquier $\epsilon > 0$, existe un entero N tal que para cualquier entero $M > N$

$$E \left[\sum_{j=0}^M \psi_j Z_{t-j} - \sum_{j=0}^N \psi_j Z_{t-j} \right]^2 < \epsilon . \quad (2.37)$$

Ahora el lado izquierdo de la expresión (2.37) es simplemente

$$\begin{aligned} E [\psi_M Z_{t-M} + \psi_{M-1} Z_{t-M+1} + \dots + \psi_{N+1} Z_{t-N-1}]^2 &= (\psi_M^2 + \psi_{M-1}^2 + \dots + \psi_{N+1}^2) \sigma^2 \\ &= \left[\sum_{j=0}^M \psi_j^2 - \sum_{j=0}^N \psi_j^2 \right] \sigma^2 . \end{aligned}$$

Ahora, si $\sum_{j=0}^{\infty} \psi_j^2$ converge como se requería por (2.35), entonces aplicando el criterio de Cauchy, el lado derecho de la expresión anterior, éste se puede hacer tan pequeño como se quiera eligiendo un entero N adecuado. Así, la serie infinita en (2.34) converge en media cuadrática con lo que se demuestra que (2.37) se cumple.

Finalmente, se demostrará que la sumabilidad absoluta de coeficientes de medias móviles implica que el proceso es **ergódico** para la media. Para esto, se expresa

$$\gamma_j = \sigma^2 \sum_{k=0}^{\infty} \psi_{j+k} \psi_k .$$

Entonces

$$|\gamma_j| = \sigma^2 \left| \sum_{k=0}^{\infty} \psi_{j+k} \psi_k \right| .$$

Luego por propiedades de valor absoluto,

$$|\gamma_j| \leq \sigma^2 \sum_{k=0}^{\infty} |\psi_{j+k} \psi_k| ,$$

y

$$\begin{aligned} \sum_{j=0}^{\infty} |\gamma_j| &\leq \sum_{j=0}^{\infty} \sum_{k=0}^{\infty} |\psi_{j+k} \psi_k| = \\ &= \sigma^2 \sum_{j=0}^{\infty} \sum_{k=0}^{\infty} |\psi_{j+k}| |\psi_k| \\ &= \sigma^2 \sum_{k=0}^{\infty} |\psi_k| \sum_{k=0}^{\infty} |\psi_{j+k}| . \end{aligned}$$

Pero existe un $M < \infty$ tal que $\sum_{j=0}^{\infty} |\psi_j| < M$, y por lo tanto, $\sum_{j=0}^{\infty} |\psi_{j+k}| < M$ para $k = 0, 1, 2, \dots$, lo que significa que

$$\sum_{j=0}^{\infty} |\gamma_j| < \sigma^2 \sum_{k=0}^{\infty} |\psi_k \cdot M| < \sigma^2 M^2 < \infty .$$

Un proceso $MA(\infty)$ con coeficientes absolutamente sumables tiene autocovarianzas absolutamente sumables, por lo tanto, un proceso $MA(\infty)$ que satisface $\sum_{j=0}^{\infty} |\psi_j| < \infty$ es ergódico para la media. ■

2.5.2. Modelos autorregresivos (AR)

Procesos Autorregresivos de primer orden (AR(1))

Un proceso autorregresivo de primer orden $AR(1)$, es un proceso estacionario de la forma

$$X_t = c + \phi X_{t-1} + Z_t , \quad (2.38)$$

donde $\{Z_t\} \sim iid WN(0, \sigma^2)$.

Se observa que (2.38) toma la forma de una ecuación en diferencia en donde $\omega_t = c + Z_t$ y se sabe del análisis realizado en la sección 2.2 que si $|\phi| \geq 1$, el multiplicador dinámico $\frac{\partial x_{t+j}}{\partial \omega_t}$ crece exponencialmente por lo que no existe un proceso de covarianza estacionario $\{X_t\}$ con varianza finita que satisfaga (2.38). Si $|\phi| < 1$ existe un proceso estacionario $\{X_t\}$ que satisface (2.38) y su solución estable de este proceso está dada por (2.2), es decir

$$\begin{aligned} X_t &= (c + Z_t) + \phi(c + Z_{t-1}) + \phi^2(c + Z_{t-2}) + \dots \\ &= \left[\frac{c}{1 - \phi} \right] + Z_t + \phi Z_{t-1} + \phi^2 Z_{t-2} + \dots . \end{aligned}$$

Lo anterior se puede visualizar como un proceso $MA(\infty)$ como en (2.34) con ψ_j dada por ϕ^j . Cuando $|\phi| < 1$, la condición

$$\sum_{j=0}^{\infty} |\psi_j| = \sum_{j=0}^{\infty} |\phi^j| = \frac{1}{1 - \phi} ,$$

se cumple.

Calculando la esperanza de (2.38), se observa

$$E[X_t] = \frac{c}{1 - \phi} , \quad (2.39)$$

así, $\mu = \frac{c}{1 - \phi}$.

La varianza de un proceso $AR(1)$ está dada por

$$\begin{aligned} \gamma(0) &= E[Z_t + \phi Z_{t-1} + \dots]^2 \\ &= (1 + \phi^2 + \phi^4 + \phi^6 + \dots)\sigma^2 \\ &= \frac{\sigma^2}{1 - \phi^2} , \end{aligned} \quad (2.40)$$

la j -ésima autocovarianza es

$$\begin{aligned} \gamma(j) &= E(X_t - \mu)(X_{t-j} - \mu) \\ &= E[Z_t + \phi Z_{t-1} + \dots + \phi^j Z_{t-j} + \phi^{j+1} Z_{t-j-1} + \phi^{j+2} Z_{t-j-2} + \dots] \\ &\quad \times [Z_{t-j} + \phi Z_{t-j-1} + \phi^2 Z_{t-j-2} + \dots] \\ &= (\phi^j + \phi^{j+2} + \phi^{j+4} + \dots)\sigma^2 \\ &= \phi^j [1 + \phi^2 + \phi^4 + \dots]\sigma^2 \\ &= \left[\frac{\phi^j}{1 - \phi^2} \right] \sigma^2 , \end{aligned} \quad (2.41)$$

de (2.40) y (2.41), se obtiene

$$\rho_j = \phi^j .$$

El cual sigue un patrón decreciente. Obsérvese que la función de autocorrelación (2.5.2) para un proceso estacionario $AR(1)$ coincide con el multiplicador dinámico (2.2.1), luego el efecto por el incremento de una unidad de Z_t sobre X_{t+j} es igual a la correlación entre X_t y X_{t+j} . Así, un valor positivo de ϕ implica una correlación positiva entre X_t y X_{t+j} y un valor negativo de ϕ implica una ecuación en diferencia de primer orden negativa pero autocorrelación de segundo orden positivo (Ver Figura 2.1 incisos (a) y (b)).

Procesos autorregresivos de orden q ($AR(q)$)

Definición 2.15 *Un proceso autorregresivo de orden q denotado como $AR(q)$, satisface*

$$X_t = c + \phi_1 X_{t-1} + \phi_2 X_{t-2} + \dots + \phi_q X_{t-q} + Z_t .$$

Es posible apreciar que un proceso autorregresivo de orden q es en realidad una ecuación en diferencia más un elemento aleatorio (en este caso un ruido blanco).

Un proceso autorregresivo puede ser estacionario o no, dependiendo de los valores que tomen las raíces de la ecuación característica

$$1 - \phi_1\lambda - \phi_2\lambda^2 - \dots - \phi_q\lambda^q = 0, \quad (2.42)$$

la cual dicta el comportamiento del proceso autorregresivo. Al recordar el caso general de una ecuación en diferencia (ver subsección 2.2.6), se sabe que $\phi(B) = 1 - \phi_1B - \phi_2B^2 - \dots - \phi_qB^q$ se puede escribir como

$$\phi(B) = (1 - \lambda_1B)(1 - \lambda_2B) \dots (1 - \lambda_qB) \cdot$$

De tal manera que el proceso autorregresivo definido por $\phi(B)$ será estacionario siempre y cuando $|\lambda_i| < 1$ para $i = 1, 2, \dots, q$, es decir, si y sólo si las raíces de (2.42) se encuentran fuera del círculo unitario en el plano no complejo.

2.5.3. Procesos autorregresivos de medias móviles de orden (p, q) ($ARMA(p, q)$)

Definición 2.16 *Un proceso $ARMA(p, q)$ es un proceso de la forma*

$$(1 - \phi_1B - \phi_2B^2 - \dots - \phi_pB^p)X_t = c + (1 + \theta_1B + \theta_2B^2 + \dots + \theta_qB^q)Z_t, \quad (2.43)$$

siempre que las raíces de

$$1 - \phi_1z - \phi_2z^2 - \dots - \phi_pz^p = 0,$$

estén fuera del círculo unitario. Tomando ambos lados de la expresión (2.43), pueden dividirse por $(1 - \phi_1B - \phi_2B^2 - \dots - \phi_pB^p)$ para obtener

$$X_t = \mu + \psi(B)Z_t,$$

donde

$$\psi(B) = \frac{1 + \theta_1B + \theta_2B^2 + \dots + \theta_qB^q}{1 - \phi_1B - \phi_2B^2 - \dots - \phi_pB^p},$$

$$\mu = \frac{c}{1 - \phi_1 - \dots - \phi_p}.$$

Así, la estacionariedad de un proceso *ARMA* depende enteramente de los parámetros autorregresivos $(\phi_1, \phi_2, \dots, \phi_p)$ y no de los parámetros de los promedios móviles $(\theta_1, \theta_2, \dots, \theta_q)$.

2.5.4. Identificación tentativa de modelos Box-Jenkins no estacionales

Dada una serie de tiempo no estacional y estacionaria, es posible identificar un proceso estacionario que se ajuste a los datos a través del comportamiento de la FAM y la FAMP, lo cual se resume en la Tabla 2.1.

Modelo	Comportamiento	
	fam	famp
MA(q)	Se trunca después del desfaseamiento q .	Se extingue.
AR(p)	Se extingue.	Se trunca después del desfaseamiento p .
ARMA(p,q)	Se extingue	Se extingue.

Tabla 2.1: Identificación de los modelos por medio del comportamiento de las funciones **fam** y **famp**

2.6. Invertibilidad

Los procesos autorregresivos estacionarios también pueden representarse como procesos de promedios móviles y algunos procesos de promedios móviles como ocurre con los procesos *MA(1)* pueden representarse de manera autorregresiva siempre que $|\theta| < 1$.

En general cuando un proceso de promedios móviles puede presentarse como un proceso autorregresivo, se dice que dicho proceso es invertible, lo cual significa que tiene como posible representación a

$$\pi(B)X_t = Z_t ,$$

donde, $\pi(B) = 1 - \pi_1 B - \pi_2 B^2 - \dots$, es un polinomio de retraso que cumple $\pi(x) =$

$1 - \sum_{i=1}^{\infty} \pi_i x^i$ converge dentro o sobre el círculo unitario $|x| \leq 1$.

Considérese un proceso $AR(1)$ estacionario de la forma (2.38), si $|\phi| < 1$ entonces $(1 - \phi B)X_t = Z_t$ se transforma en

$$\begin{aligned} X_t &= (1 - \phi(B))^{-1} Z_t = \left(\sum_{j=0}^{\infty} (\phi B)^j \right) Z_t \\ &= Z_t + \phi_1 Z_{t-1} + \phi^2 Z_{t-1} + \dots \end{aligned}$$

De manera general un proceso estacionario $AR(q)$ se puede representar como una suma ponderada de choques aleatorios, esto es

$$X_t = Z_t - \psi_1 Z_{t-1} - \psi_2 Z_{t-2} - \dots - , \quad (2.44)$$

con $\sum_{j=0}^{\infty} |\psi_j| < \infty$. Los coeficientes ψ_i , $i = 1, 2, \dots$, pueden obtenerse a partir del hecho de que un proceso estacionario $AR(q)$ se debe poder expresar como

$$\phi(B)X_t = Z_t \quad \text{y} \quad X_t = \psi(B)Z_t ,$$

es decir, $1 = \phi(B)\psi(B)$, donde $\phi(B) = 1 - \phi_1 B - \phi_2 B - \dots - \phi_q B^q$.

Luego,

$$\begin{aligned} 1 &= (1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_q B^q) (1 - \psi_1 B - \psi_2 B^2 - \dots - \psi_q B^q) \\ &= 1 - (\phi_1 + \psi_1)B - (\psi_2 - \phi_1 \psi_1 + \phi_2)B^2 - \dots \\ &\quad - (\psi_q - \phi_1 \psi_{q-1} - \phi_2 \psi_{q-2} - \dots - \phi_q)B^q - \dots \end{aligned} \quad (2.45)$$

Para que la expresión (2.45) se cumpla se requiere

$$\begin{aligned} \psi_1 &= -\phi_1 \\ \psi_2 &= \phi_1 \psi_1 - \phi_2 \\ &\vdots \\ \psi_q &= \phi_1 \psi_{q-1} + \phi_2 \psi_{q-2} + \dots + \phi_{p-1} \psi_1 - \phi_q \end{aligned}$$

La expresión (2.44) no será útil si todos los coeficientes ψ_1, ψ_2, \dots , son distintos de cero, sin embargo, en ocasiones dichos coeficientes son distintos de cero hasta un cierto

ψ_j , $j \in \mathbb{N}$ a partir del cual todos los ψ_j son cero, esto es $\psi_1 \neq 0, \psi_2 \neq 0, \dots, \psi_j \neq 0$, $\psi_{j+1} = \psi_{j+2} = \dots = 0$, [11].

De lo anterior se puede concluir que un proceso $MA(1)$ tiene una representación en términos de un proceso autorregresivo siempre que $|\theta| < 1$. En efecto, considérese el proceso $MA(1)$

$$X_t - \mu = (1 + \theta B)Z_t . \quad (2.46)$$

Siempre que $|\theta| < 1$, ambos lados de la expresión (2.46) se pueden multiplicar por $(1 + \theta B)^{-1}$ para obtener

$$(1 - \theta_1 B - \theta_2 B^2 - \theta_3 B^3 - \dots)(X_t - \mu) = Z_t ,$$

la cual se puede considerar como un proceso $AR(\infty)$.

En general si un proceso de promedio móviles admite una representación en términos de un proceso autorregresivo, se dice que dicho proceso es invertible, esto es, que admite una representación de la forma

$$\pi(B)X_t = Z_t ,$$

donde $\pi(B) = 1 - \pi_1 B - \pi_2 B^2 - \dots$ es un polinomio de retraso el cual debe cumplir que $\pi(x) = 1 - \sum_{j=1}^{\infty} \pi_j x^j$ converge dentro o sobre el círculo unitario.

Si un proceso $MA(q)$ es invertible, admite la representación

$$\pi(B)X_t = 1 - \pi_1 X_{t-1} - \pi_2 X_{t-2} - \dots ,$$

de donde los valores de π_j , $j \in \mathbb{N}$ pueden obtenerse de la representación $\pi(B)\theta(B) = 1$ y siguiéndose un proceso similar como en (2.45).

La importancia de la invertibilidad de un proceso radica en que todo proceso invertible está determinado de manera única por su FAM lo cual no ocurre con procesos que no sean invertibles [11].

2.7. Procesos Lineales

Definición 2.17 Una serie de tiempo $\{X_t\}$ es un proceso lineal si se puede representar como

$$X_t = \sum_{j=-\infty}^{\infty} \psi_j Z_{t-j} , \quad (2.47)$$

para todo t , donde $\{Z_t\} \sim iidWN(0, \sigma^2)$ y ψ_j es una sucesión de constantes con $\sum_{j=-\infty}^{\infty} \psi_j < \infty$.

Observación 2.9 *i. La condición $\sum_{j=-\infty}^{\infty} |\psi_j| < \infty$ asegura que la suma en la ecuación (2.47) converge y como $E|Z_t| < \sigma$ y*

$$E|X_t| \leq \sum_{j=-\infty}^{\infty} (|\psi_j| E|Z_{t-j}|) \leq \left(\sum_{j=-\infty}^{\infty} |\psi_j| \right) \sigma < \infty .$$

También asegura que $\sum_{j=-\infty}^{\infty} \psi_j^2 < \infty$ y por lo tanto, la serie de tiempo (2.47) converge en media cuadrática.

ii. En términos del operador de retraso B , la expresión (2.47) se puede reescribir como

$$X_t = \sum_{j=-\infty}^{\infty} \psi_j B^j Z_t .$$

*iii. De la observación anterior, si $\psi_j = 0 \ \forall j < 0$, el proceso lineal es llamado **proceso lineal de medias móviles o $MA(\infty)$** .*

Proposición 2.2 *Sea $\{Y_t\}$ una serie de tiempo estacionaria con media cero y función de covarianza γ_Y . Si $\sum_{j=-\infty}^{\infty} |\psi_j| < \infty$, entonces la serie de tiempo*

$$X_t = \sum_{j=-\infty}^{\infty} \psi_j Y_{t-j} ,$$

es estacionaria con media 0 y función de autocovarianza

$$\gamma_X(h) = \sum_{j=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} \psi_j \psi_k \gamma_Y(h+k-j) .$$

En el caso en que $\{X_t\}$ es un proceso lineal,

$$\gamma_X(h) = \sum_{-\infty}^{\infty} \psi_j \psi_{j+h} \sigma^2 .$$

Demostración. Como $E(|Y_t|) < \sqrt{\gamma_0}$, entonces $\sum_{j=-\infty}^{\infty} |\psi_j| E|Y_t| < \infty$, lo que implica que $\sum_{j=-\infty}^{\infty} \psi_j Y_t$ es convergente.

Como $E(Y_t) = 0$ se tiene que

$$E(X_t) = E \left(\sum_{j=-\infty}^{\infty} \psi_j Y_{t-j} \right) = \sum_{j=-\infty}^{\infty} \psi_j E(Y_{t-j}) = 0 ,$$

$$\begin{aligned}
E(X_{t+h}X_t) &= E\left(\sum_{j=-\infty}^{\infty} \psi_j Y_{t+h-j}\right)\left(\sum_{k=-\infty}^{\infty} \psi_k Y_{t-k}\right) \\
&= \sum_{j=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} \psi_j \psi_k E(Y_{t+h-j}Y_{t-k}) \\
&= \sum_{j=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} \psi_j \psi_k \gamma(h-j+k) .
\end{aligned}$$

Si $\{Y_t\}$ es una sucesión de ruido blanco $\{Z_t\}$, entonces

$$\begin{aligned}
X_t &= \sum_{j=-\infty}^{\infty} \psi_j Z_{t-j}, \text{ y} \\
\gamma_X(h) &= \sum_{j=-\infty}^{\infty} \psi_j \psi_{j+h} \sigma^2 .
\end{aligned}$$

■

2.8. Pronóstico de series de tiempo estacionarias

En las siguientes dos secciones se presentan dos formas de estimar parámetros de los modelos de series de tiempo mediante mínimos cuadrados y por máxima verosimilitud, aunque en la aplicación se trabaja sólo con los estimadores obtenidos por máxima verosimilitud por ser ésta herramienta la que se encuentra integrada en R .

Se considera el problema de predecir los valores X_{n+h} , $h > 0$ de una serie de tiempo estacionaria con media μ conocida y función de autocovarianza en términos de los valores $\{X_n, \dots, X_1\}$, el objetivo es encontrar la combinación lineal de $1, X_n, \dots, X_1$ que pronostique X_{n+h} con el menor error cuadrado medio posible. El mejor predictor lineal en términos de $1, X_n, \dots, X_1$ se denota por $P_n X_{n+h}$ el cual tiene la forma

$$P_n X_{n+h} = a_0 + a_1 X_n + \dots + a_n X_1 . \quad (2.48)$$

Por lo que resta determinar los coeficientes a_i , $i = 0, 1, \dots, n$, que minimicen

$$S(a_0, \dots, a_n) = E \left\{ \left(X_{n+h} - a_0 - \sum_{i=1}^n a_i X_{n+1-i} \right)^2 \right\} . \quad (2.49)$$

Como S es una función cuadrática de a_i , $i = 0, 1, \dots, n$, que es acotada inferiormente por 0, es claro que hay por lo menos un valor de (a_0, \dots, a_n) que minimiza S y que satisface las ecuaciones

$$\frac{\partial S((a_0, \dots, a_n))}{\partial a_j} = 0, \quad j = 0, 1, \dots, n \cdot$$

Realizando los cálculos correspondientes se obtiene que

$$E \left[X_{n+h} - a_0 - \sum_{i=1}^n a_i X_{n+1-i} \right] = 0 \cdot \quad (2.50)$$

$$E \left[\left(X_{n+h} - a_0 - \sum_{i=1}^n a_i X_{n+1-i} \right) X_{n+1-j} \right] = 0, \quad j = 1, 2, \dots, n \cdot \quad (2.51)$$

Las cuales pueden ser escritas como

$$a_0 = \mu \left(1 - \sum_{i=1}^n a_i \right), \quad (2.52)$$

$$\Gamma_n \mathbf{a}_n = \boldsymbol{\gamma}_n(h), \quad (2.53)$$

donde

$$\mathbf{a}_n = (a_1, \dots, a_n)', \quad \Gamma_n = [\gamma(i-j)]_{i,j=1}^n \text{ y,}$$

$$\boldsymbol{\gamma}_n(h) = (\gamma(h), \gamma(h+1), \dots, \gamma(h+n-1))' \cdot$$

Por lo tanto,

$$P_n X_{n+h} = \mu + \sum_{i=1}^n a_i (X_{n+1-i} - \mu),$$

donde \mathbf{a}_n satisface (2.53) .

Propiedades de $P_n X_{n+h}$

1. $P_n X_{n+h} = \mu + \sum_{i=1}^n a_i (X_{n+1-i} - \mu)$ donde $\mathbf{a}_n = (a_1, \dots, a_n)'$ satisface (2.53).
2. $E(X_{n+h} - P_n X_{n+h})^2 = \gamma(0) - \mathbf{a}_n' \boldsymbol{\gamma}_n(h)$, donde $\boldsymbol{\gamma}_n(h) = (\gamma(h), \gamma(h+1) + \dots + \gamma(h+n-1))'$
3. $E(X_{n+h} - P_n X_{n+h}) = 0$.

$$4. E[(X_{n+h} - P_n X_{n+h})X_j] = 0, \quad j = 1, \dots, n.$$

Demostración. Primero se demostrarán la expresiones (2.52) y (2.53) bajo las suposiciones correspondientes en (2.50), por lo que se tiene

$$\begin{aligned} E[X_{n+h}] - a_0 - \sum_{i=1}^n E[X_{n+1-i}] &= 0 \\ \mu - a_0 - \mu \sum_{i=1}^n a_i &= 0 \\ a_0 &= \mu - \mu \sum_{i=1}^n a_i. \end{aligned}$$

Por otro lado,

$$E[X_{n+h}X_{n+1-j}] - a_0E[X_{n+1-j}] - \sum_{i=1}^n a_iE[X_{n+1-i}X_{n+1-j}] = 0, \text{ para } j = 1, 2, \dots, n. \quad (2.54)$$

Además,

$$\text{Cov}(X_{n+1-j}, X_{n+1-j}) = E[X_{n+1-j}X_{n+1-j}] - \mu^2,$$

luego para cada j en (2.54) se tiene

$$\begin{aligned} a_1\text{Cov}(X_n, X_n) + \dots + a_n\text{Cov}(X_n, X_1) + \mu^2 \sum_{i=1}^n a_i - \mu^2 + a_0\mu &= \text{Cov}(X_{n+h}, X_n) \\ a_1\text{Cov}(X_{n-1}, X_n) + \dots + a_n\text{Cov}(X_{n-1}, X_1) + \mu^2 \sum_{i=1}^n a_i - \mu^2 + a_0\mu &= \text{Cov}(X_{n+h}, X_{n-1}) \\ \vdots & \\ a_1\text{Cov}(X_1, X_n) + \dots + a_n\text{Cov}(X_1, X_1) + \mu^2 \sum_{i=1}^n a_i - \mu^2 + a_0\mu &= \text{Cov}(X_{n+h}, X_1). \end{aligned} \quad (2.55)$$

Reduciendo términos en (2.55), se tiene

$$\begin{bmatrix} Cov(X_n, X_n) & Cov(X_n, X_{n-1}) & \cdots & Cov(X_n, X_1) \\ Cov(X_{n-1}, X_n) & Cov(X_{n-1}, X_{n-1}) & \cdots & Cov(X_{n-1}, X_1) \\ \vdots & \vdots & \cdots & \vdots \\ Cov(X_1, X_n) & Cov(X_1, X_{n-1}) & \cdots & Cov(X_1, X_1) \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix} = \begin{bmatrix} Cov(X_{n+h}, X_n) \\ Cov(X_{n+h}, X_{n-1}) \\ \vdots \\ Cov(X_{n+h}, X_1) \end{bmatrix} .$$

Considerando la siguiente notación

$$\Gamma_n = \begin{bmatrix} Cov(X_n, X_n) & Cov(X_n, X_{n-1}) & \cdots & Cov(X_n, X_1) \\ Cov(X_{n-1}, X_n) & Cov(X_{n-1}, X_{n-1}) & \cdots & Cov(X_{n-1}, X_1) \\ \vdots & \vdots & \cdots & \vdots \\ Cov(X_1, X_n) & Cov(X_1, X_{n-1}) & \cdots & Cov(X_1, X_1) \end{bmatrix} .$$

$$\mathbf{a}_n = (a_1, a_2, \dots, a_n)' \quad y,$$

$$\boldsymbol{\gamma}_n(h) = (\gamma(h), \gamma(h+1), \dots, \gamma(h+n-1))' .$$

Se tiene que

$$\Gamma_n \mathbf{a}_n = \boldsymbol{\gamma}_n(h) . \quad (2.56)$$

1. Como el mejor predictor lineal $P_n X_{n+h}$, en términos de $1, X_n, \dots, X_1$ tiene la forma $P_n X_{n+h} = a_0 + a_1 X_n + \dots + a_n X_1$, donde los a_i , $i = 1, 2, \dots, n$, están determinados por (2.52) y (2.53), luego

$$\begin{aligned} P_n X_{n+h} &= \mu - \mu \sum_{i=1}^n a_i + \sum_{i=1}^n a_i X_{n+1-i} \\ &= \mu + \sum_{i=1}^n a_i (X_{n+1-i} - \mu) . \end{aligned}$$

2.

$$X_{n+h} - P_n X_{n+h} = X_{n+h} - a_0 - a_1 X_n - \dots - a_n X_1$$

pero $a_0 = \mu - \sum_{i=1}^n a_i \mu$, así,

$$\begin{aligned} (X_{n+h} - P_n X_{n+h})^2 &= \left\{ (X_{n+h} - \mu) + \left(\sum_{i=1}^n a_i \mu - \sum_{j=1}^n a_j X_{n+1-j} \right) \right\}^2 \\ &= (X_{n+h} - \mu)^2 + 2(X_{n+h} - \mu) \left(\sum_{i=1}^n a_i \mu - \sum_{j=1}^n a_j X_{n+1-j} \right) + \left(\sum_{i=1}^n a_i \mu - \sum_{i=1}^n a_i X_{n+1-i} \right)^2 \\ &= X_{n+h}^2 - 2X_{n+h}\mu + \mu^2 + 2 \sum_{i=1}^n \mu X_{n+h} - 2 \sum_{j=1}^n a_j X_{n+1-j} X_{n+h} - 2 \sum_{i=1}^n a_i \mu^2 \\ &\quad + 2 \sum_{j=1}^n a_j \mu X_{n+1-j} + \sum_{i=1}^n \sum_{j=1}^n a_i a_j \mu^2 - 2 \sum_{i=1}^n \sum_{j=1}^n a_i a_j \mu X_{n+1-i} \\ &\quad + \sum_{i=1}^n \sum_{j=1}^n a_i a_j X_{n+1-i} X_{n+1-j}. \end{aligned}$$

Luego,

$$\begin{aligned} E[X_{n+h} - P_n X_{n+h}]^2 &= Cov(X_{n+h}, X_{n+h}) - 2 \sum_{j=1}^n a_j Cov(X_{n+1-j}, X_{n+h}) \\ &\quad + \sum_{i=1}^n \sum_{j=1}^n a_i Cov(X_{n+1-i}, X_{n+1-j}) a_j = \gamma(0) - \mathbf{a}'_n \boldsymbol{\gamma}_n(h) . \end{aligned}$$

3. Por la expresión (2.50) se cumple $E[X_{n+h} - a_0 - \sum_{i=1}^n a_i X_{n+1-i}] = 0$, pero $P_n X_{n+h} = a_0 - \sum_{i=1}^n a_i X_{n+1-i}$, así

$$E[X_{n+h} - P_n X_{n+h}] = E \left[X_{n+h} - a_0 - \sum_{i=1}^n a_i X_{n+1-i} \right] = 0 .$$

4. Por la expresión (2.50) se cumple $E[(X_{n+h} - a_0 - \sum_{i=1}^n a_i X_{n+1-i}) X_{n+1-j}] = 0$, para $j = 1, 2, \dots, n$, además,

$$P_n X_{n+h} = a_0 - \sum_{i=1}^n a_i X_{n+1-i} ,$$

luego

$$E[(X_{n+h} - P_n X_{n+h}) X_{n+1-j}] = E\left[\left(X_{n+h} - a_0 - \sum_{i=1}^n a_i X_{n+1-i}\right) X_{n+1-j}\right] = 0,$$

para $j = 1, 2, \dots, n$.

■

2.9. Estimación por máxima verosimilitud

Considérese un proceso $ARMA(p, q)$, $Y_t = c + \phi_1 Y_{t-1} + \dots + \phi_p Y_{t-p} + Z_t + \theta_1 Z_{t-1} + \theta_2 Z_{t-2} + \dots + \theta_q Z_{t-q}$, con $\{Z_t\} \sim iid WN(0, \sigma^2)$. En esta sección se presenta cómo estimar los valores $(c, \phi_1, \phi_2, \dots, \phi_p, \theta_1, \theta_2, \dots, \theta_q, \sigma^2)$ a partir de datos observados de la variable aleatoria Y , utilizando estimación por máxima verosimilitud, para esto se asume que $\{Z_t\} \sim iid N(0, \sigma^2)$ y se considera a $\boldsymbol{\theta} = (c, \phi_1, \phi_2, \dots, \phi_p, \theta_1, \theta_2, \dots, \theta_q, \sigma^2)$ como el vector de parámetros poblacionales.

2.9.1. Función de verosimilitud para un proceso gaussiano $AR(1)$

Considérese el proceso gaussiano $AR(1)$ dado por $Y_t = c + \phi Y_{t-1} + Z_t$, con $Z_t \sim iid N(0, \sigma^2)$. Se sabe de (2.39) y (2.40) que $E(Y_t) = \frac{c}{1-\phi}$ y $E(Y_t - \mu)^2 = \frac{\sigma^2}{1-\phi^2}$.

Como $\{Z_t\}$ es gaussiano, Y_1 también lo es, por lo tanto, la densidad de la distribución de donde proviene la primera observación toma la forma

$$f_{Y_1}(y_1; \boldsymbol{\theta}) = \frac{1}{\sqrt{2\pi\sigma^2/(1-\phi^2)}} \exp\left[\frac{-(y_1 - [c/(1-\phi)^2])^2}{2\sigma^2/(1-\phi^2)}\right] , \quad (2.57)$$

ahora, $Y_2 = c + \phi Y_1 + Z_2$, entonces $(Y_2 | Y_1 = y_1) \sim N(c + \phi y_1, \sigma^2)$, es decir,

$$f_{Y_2|Y_1}(y_2|y_1; \boldsymbol{\theta}) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left[\frac{-(y_2 - c - \phi y_1)^2}{2\sigma^2} \right] ,$$

así,

$$f_{Y_2, Y_1}(y_2, y_1; \boldsymbol{\theta}) = f_{Y_1}(y_1; \boldsymbol{\theta}) f_{Y_2|Y_1}(y_2|y_1; \boldsymbol{\theta}) .$$

Similarmente, $Y_3 = c + \phi Y_2 + Z_3$ y $(Y_3|Y_2 = y_2, Y_1 = y_1) \sim N(c + \phi y_2, \sigma^2)$, por tanto

$$f_{Y_3, Y_2, Y_1}(y_3, y_2, y_1; \boldsymbol{\theta}) = f_{Y_3|Y_2, Y_1}(y_3|y_2, y_1; \boldsymbol{\theta}) f_{Y_2, Y_1}(y_2, y_1; \boldsymbol{\theta}) .$$

En general, $(Y_t|Y_{t-1}, Y_{t-2}, \dots, Y_1) \sim N(c + \phi y_{t-1}, \sigma^2)$, esto es

$$\begin{aligned} f_{Y_t|Y_{t-1}, Y_{t-2}, \dots, Y_1}(y_t|y_{t-1}, y_{t-2}, \dots, y_1; \boldsymbol{\theta}) &= f_{Y_t|Y_{t-1}}(y_t|y_{t-1}; \boldsymbol{\theta}) \\ &= \frac{1}{\sqrt{2\pi\sigma^2}} \left[\frac{-(y_t - c - \phi y_{t-1})^2}{2\sigma^2} \right] , \end{aligned} \quad (2.58)$$

así, la función de densidad conjunta de las primeras t observaciones está dada por

$$f_{Y_t, Y_{t-1}, \dots, Y_1}(y_t|y_{t-1}, \dots, y_1) = f_{Y_t|Y_{t-1}}(y_t|y_{t-1}; \boldsymbol{\theta}) f_{Y_{t-1}, Y_{t-2}, \dots, Y_1}(y_{t-1}, y_{t-2}, \dots, y_1; \boldsymbol{\theta}),$$

de forma que la función de verosimilitud de la muestra completa es

$$f_{Y_T, Y_{T-1}, \dots, Y_1}(y_T|y_{T-1}, \dots, y_1) = f_{Y_1}(y_1; \boldsymbol{\theta}) \prod_{t=2}^T f_{Y_t|Y_{t-1}}(y_t|y_{t-1}; \boldsymbol{\theta}) .$$

La función logaritmo de verosimilitud se determina por

$$\mathcal{L}(\boldsymbol{\theta}) = \log f_{Y_1}(y_1; \boldsymbol{\theta}) + \sum_{t=2}^T \log (f_{Y_t|Y_{t-1}}(y_t|y_{t-1}; \boldsymbol{\theta})) . \quad (2.59)$$

Sustituyendo (2.57) y (2.58) en (2.59), la función logaritmo de verosimilitud para una muestra de tamaño T de un proceso gaussiano $AR(1)$ se expresa por

$$\begin{aligned} \mathcal{L}(\boldsymbol{\theta}) = & -\frac{1}{2}\log(2\pi) - \frac{1}{2}\log\left[\frac{\sigma^2}{1-\phi^2}\right] - \frac{\{y_1 - [c/(1-\phi)]\}^2}{2\sigma^2/(1-\phi^2)} - \left[\frac{T-1}{2}\right]\log(2\pi) \\ & - \left[\frac{T-1}{2}\right]\log\sigma^2 - \sum_{t=2}^T \left[\frac{(y_t - c - \phi y_{t-1})^2}{2\sigma^2}\right] . \end{aligned} \quad (2.60)$$

Forma matricial de la función de verosimilitud

Sea $\mathbf{y} = (y_1, y_2, \dots, y_T)'$ una muestra de tamaño T de un proceso gaussiano $AR(1)$. Este vector puede verse como una realización simple de una distribución gaussiana T -dimensional con vector de medias

$$\boldsymbol{\mu} = E(\mathbf{Y}) = [E(Y_1), E(Y_2), \dots, E(Y_T)]' = (\mu, \mu, \dots, \mu)' ,$$

donde $\mu = \frac{c}{1-\phi}$ y matriz de covarianzas

$$\boldsymbol{\Omega} = E[(\mathbf{Y} - \boldsymbol{\mu})(\mathbf{Y} - \boldsymbol{\mu})'] = [E[(Y_i - \mu)(Y_j - \mu)]] , \text{ para } i, j = 1, 2, \dots, T .$$

De la Sección 2.5.2, se sabe que $E[(Y_t - \mu)(Y_{t-j} - \mu)] = \frac{\sigma^2\phi^j}{1-\phi^2}$, por lo que

$$\boldsymbol{\Omega} = \sigma^2\mathbf{V} ,$$

donde

$$\mathbf{V} = \frac{1}{1-\phi^2} \begin{bmatrix} 1 & \phi & \phi^2 & \dots & \phi^{T-1} \\ \phi & 1 & \phi & \dots & \phi^{T-2} \\ \vdots & \vdots & \vdots & \dots & \vdots \\ \phi^{T-1} & \phi^{T-2} & \phi^{T-3} & \dots & 1 \end{bmatrix} .$$

Visualizando la muestra observada \mathbf{y} como una realización de la distribución normal multivariada $N(\boldsymbol{\mu}, \boldsymbol{\Omega})$, la función de verosimilitud se puede expresar como

$$f_{\mathbf{Y}}(\mathbf{y}; \boldsymbol{\theta}) = (2\pi)^{-T/2} |\boldsymbol{\Omega}|^{-1/2} \exp\left[-\frac{1}{2}(\mathbf{y} - \boldsymbol{\mu})' \boldsymbol{\Omega}^{-1} (\mathbf{y} - \boldsymbol{\mu})\right] ,$$

con función logaritmo de verosimilitud

$$\mathcal{L}(\boldsymbol{\theta}) = -\frac{T}{2}\log(2\pi) + \frac{1}{2}\log|\boldsymbol{\Omega}^{-1}| - \frac{1}{2}(\mathbf{y} - \boldsymbol{\mu})' \boldsymbol{\Omega}^{-1} (\mathbf{y} - \boldsymbol{\mu}) . \quad (2.61)$$

Resta demostrar que (2.61) y (2.60) representan la misma función de (y_1, y_2, \dots, y_T) , para esto, se define

$$\mathbf{L}_{T \times T} = \begin{bmatrix} \sqrt{1-\phi^2} & 0 & 0 & \cdots & 0 & 0 \\ -\phi & 1 & 0 & \cdots & 0 & 0 \\ \vdots & \cdots & \vdots & \cdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & -\phi & 1 \end{bmatrix} .$$

Se puede demostrar que

$$\mathbf{L}'\mathbf{L} = \mathbf{V}^{-1} ,$$

por lo que

$$\mathbf{\Omega}^{-1} = \sigma^{-2}\mathbf{L}'\mathbf{L} . \quad (2.62)$$

Sustituyendo (2.62) en (2.61), se tiene que

$$\mathcal{L}(\boldsymbol{\theta}) = -\frac{T}{2}\log(2\pi) + \frac{1}{2}\log|\sigma^{-2}\mathbf{L}'\mathbf{L}| - \frac{1}{2}(\mathbf{y} - \boldsymbol{\mu})'\sigma^{-2}\mathbf{L}'\mathbf{L}(\mathbf{y} - \boldsymbol{\mu}) . \quad (2.63)$$

Se define el vector

$$\begin{aligned} \tilde{\mathbf{y}} &\equiv \mathbf{L}(\mathbf{y} - \boldsymbol{\mu}) \\ &= \begin{bmatrix} \sqrt{1-\phi^2} & 0 & 0 & \cdots & 0 & 0 \\ -\phi & 1 & 0 & \cdots & 0 & 0 \\ \vdots & \cdots & \vdots & \cdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & -\phi & 1 \end{bmatrix} \begin{bmatrix} y_1 - \mu \\ y_2 - \mu \\ \vdots \\ y_T - \mu \end{bmatrix} \\ &= \begin{bmatrix} \sqrt{1-\phi^2}(y_1 - \mu) \\ (y_2 - \mu) - \phi(y_1 - \mu) \\ (y_3 - \mu) - \phi(y_2 - \mu) \\ \vdots \\ (y_T - \mu) - \phi(y_{T-1} - \mu) \end{bmatrix} . \end{aligned} \quad (2.64)$$

El último término de (2.63) se puede escribir como

$$\begin{aligned} \frac{1}{2}(\mathbf{y} - \boldsymbol{\mu})'\sigma^{-2}\mathbf{L}'\mathbf{L}(\mathbf{y} - \boldsymbol{\mu}) &= \left(\frac{1}{2\sigma^2}\right) \tilde{\mathbf{y}}'\tilde{\mathbf{y}} \\ &= \left[\frac{1}{2\sigma^2}\right] (1 - \phi^2) \left[y_1 - \frac{c}{1-\phi}\right]^2 \\ &\quad + \frac{1}{2\sigma^2} \sum_{t=2}^T (y_t - c - \phi y_{t-1})^2 , \end{aligned} \quad (2.65)$$

el término medio de (2.63) es

$$\begin{aligned} \frac{1}{2} \log |\sigma^{-2} \mathbf{L}' \mathbf{L}| &= \frac{1}{2} \log \{ \sigma^{-2T} |\mathbf{L}' \mathbf{L}| \} \\ &= -\frac{T}{2} \log(\sigma^2) + \log(1 - \phi^2) . \end{aligned} \quad (2.66)$$

Considerando que $\mu = c/(1 - \phi)$ en (2.64) y sustituyendo (2.64), (2.65) y (2.66) en (2.63), se verifica la igualdad de (2.61) y (2.60).

La estimación por máxima verosimilitud $\hat{\theta}$ es el valor para el cual (2.60) se maximiza. En principio, esto requiere diferenciar (2.60) e igualar a cero el resultado. En la práctica cuando ésto se logra, el resultado es un sistema no lineal de ecuaciones en θ y (y_1, y_2, \dots, y_T) por lo que no hay una solución simple para θ en términos de (y_1, y_2, \dots, y_T) por lo que en ocasiones se requiere de iteración o procedimientos numéricos.

Una alternativa a la maximización numérica de la función de verosimilitud es considerar el valor de y_1 como determinístico y maximizar la función logaritmo de verosimilitud condicionada sobre la primera observación, así

$$f_{Y_T, Y_{T-1}, \dots, Y_2 | Y_1}(y_T, y_{T-1}, \dots, y_2 | y_1; \theta) = \prod_{t=2}^T f_{Y_t | Y_{t-1}}(y_t | y_{t-1}; \theta) .$$

Por lo que ahora el objetivo es maximizar

$$\begin{aligned} \log f_{Y_T, Y_{T-1}, \dots, Y_2 | Y_1}(y_T, y_{T-1}, \dots, y_2 | y_1; \theta) &= -\frac{T-1}{2} \log(2\pi) - \frac{T-1}{2} \log(\sigma^2) \\ &\quad - \sum_{t=2}^T \left[\frac{(y_t - c - \phi y_{t-1})^2}{2\sigma^2} \right] . \end{aligned} \quad (2.67)$$

Maximizar la expresión anterior con respecto a c y ϕ es equivalente a minimizar

$$\sum_{t=2}^T \left[\frac{(y_t - c - \phi y_{t-1})^2}{2\sigma^2} \right] . \quad (2.68)$$

Al realizar el proceso de minimización de (2.68), se tiene

$$\begin{bmatrix} \hat{c} \\ \hat{\phi} \end{bmatrix} = \begin{bmatrix} T-1 & \sum_{t=2}^T y_{t-1} \\ \sum_{t=2}^T y_{t-1} & \sum_{t=2}^T y_{t-1}^2 \end{bmatrix}^{-1} \begin{bmatrix} \sum_{t=2}^T y_t \\ \sum_{t=2}^T y_{t-1} y_t \end{bmatrix} ,$$

derivando (2.67) respecto de σ e igualando a cero, se tiene

$$\widehat{\sigma}^2 = \sum_{t=2}^T \left[\frac{(y_t - \widehat{c} - \widehat{\phi}y_{t-1})^2}{T-1} \right] .$$

Observación 2.10 *El EMV condicional es el promedio de los residuales al cuadrado del estimador MCO de regresión.*

2.10. Función de verosimilitud para un proceso gaussiano $AR(p)$

Consideremos el proceso autorregresivo $AR(p)$, $Y_t = c + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \dots + \phi_p Y_{t-p} + Z_t$, con $\{Z_t\} \sim iid N(0, \sigma^2)$. Supóngase que se tiene una muestra de tamaño T para un proceso de $AR(p)$ y que las primeras p observaciones de la muestra $\{y_1, y_2, \dots, y_p\}$ son colectadas en un vector \mathbf{y}_p de tamaño $p \times 1$ el cual se puede ver como una realización de una variable gaussiana de dimensión p .

Es fácil demostrar que la media de este vector es el vector de tamaño $p \times 1$, $\boldsymbol{\mu}_p$, en el que cada entrada de éste es de la forma

$$\frac{c}{1 - \phi_1 - \phi_2 - \dots - \phi_p} .$$

Sea $\sigma^2 \mathbf{V}_p$ la matriz de varianzas y covarianzas de (Y_1, Y_2, \dots, Y_p) , dada por

$$\sigma^2 \mathbf{V}_p = \begin{bmatrix} \gamma_0 & \gamma_1 & \gamma_2 & \cdots & \gamma_{p-1} \\ \gamma_1 & \gamma_0 & \gamma_1 & \cdots & \gamma_{p-2} \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ \gamma_{p-1} & \gamma_{p-2} & \gamma_{p-3} & \cdots & \gamma_0 \end{bmatrix} ,$$

donde γ_j es la j -ésima autocovarianza para un proceso $AR(p)$. La densidad de las primeras p observaciones es entonces $N \sim (\boldsymbol{\mu}_p, \sigma^2 \mathbf{V}_p)$. Para el resto de las observaciones en la muestra $(y_{p+1}, y_{p+2}, \dots, y_T)$, condicionando las primeras $t-1$ observaciones entonces, la t -ésima observación es gaussiana con media $c + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \dots + \phi_p y_{t-p}$ y varianza σ^2 . Únicamente las p observaciones más recientes influyen en esta distribución, por lo tanto, para $t > p$,

$$\begin{aligned} f_{Y_t|Y_{t-1}, Y_{t-2}, \dots, Y_1}(y_t|y_{t-1}, y_{t-2}, \dots, y_1; \boldsymbol{\theta}) &= f_{Y_t|Y_{t-1}, Y_{t-2}, \dots, Y_{t-p}}(y_t|y_{t-1}, \dots, y_{t-p}; \boldsymbol{\theta}) \\ &= \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left[\frac{-(y_t - c - \phi_1 y_{t-1} - \phi_2 y_{t-2} - \dots - \phi_p y_{t-p})^2}{2\sigma^2} \right] . \end{aligned}$$

La función de verosimilitud para la muestra completa es

$$f_{Y_T, Y_{T-1}, Y_{T-2}, \dots, Y_1}(y_T, y_{T-1}, \dots, y_1; \boldsymbol{\theta}) = f_{Y_p, Y_{p-1}, \dots, Y_1}(y_p, y_{p-1}, \dots, y_1; \boldsymbol{\theta}) \\ \times \prod_{t=p+1}^T f_{Y_t|Y_{t-1}, Y_{t-2}, \dots, Y_{t-p}}(y_t|y_{t-1}, y_{t-2}, \dots, y_{t-p}; \boldsymbol{\theta}) ,$$

con función logaritmo de verosimilitud

$$\mathcal{L}(\boldsymbol{\theta}) = -\frac{T}{2}\log(2\pi) - \frac{T}{2}\log(\sigma^2) + \frac{1}{2}\log|\mathbf{V}_p^{-1}| \\ - \frac{1}{2\sigma^2}(\mathbf{y}_p - \boldsymbol{\mu}_p)' \mathbf{V}_p^{-1}(\mathbf{y}_p - \boldsymbol{\mu}_p) \\ - \sum_{t=p+1}^T \frac{(y_t - c - \phi_1 y_{t-1} - \phi_2 y_{t-2} - \dots - \phi_p y_{t-p})^2}{2\sigma^2} . \quad (2.69)$$

La expresión (2.68) requiere de \mathbf{V}_p^{-1} . Se denota por v^{ij} al elemento del renglón i y la columna j de \mathbf{V}_p^{-1} , y

$$v^{ij}(p) = \left[\sum_{k=0}^{i-1} \phi_k \phi_{k+j-i} - \sum_{k=p+1-j}^{p+i-j} \phi_k \phi_{k+j-i} \right] ,$$

para $1 \leq i \leq j \leq p$ y con $\phi_0 = -1$ [6].

Para maximizar la función logaritmo de verosimilitud exacta para un proceso $AR(p)$, la expresión (2.10) debe resolverse a través de métodos numéricos, mientras que la maximización de la función de verosimilitud de la función logaritmo de verosimilitud condicionada sobre las primeras p observaciones asume la forma simple

$$\log f_{Y_T, Y_{T-1}, \dots, Y_{p+1}|Y_p, Y_{p-1}, \dots, Y_1}(y_T, y_{T-1}, \dots, y_{p+1}|y_p, y_{p-1}, \dots, y_1, \boldsymbol{\theta}) \\ = -\frac{T-p}{2}\log(2\pi) - \frac{T-p}{2}\log(\sigma^2) \\ - \sum_{t=p+1}^T \frac{(y_t - c - \phi_1 y_{t-1} - \phi_2 y_{t-2} - \dots - \phi_p y_{t-p})^2}{2\sigma^2} . \quad (2.70)$$

Los valores de c, ϕ_1, \dots, ϕ_p que maximizan (2.70) son aquellos que minimizan

$$\sum_{t=p+1}^T \frac{(y_t - c - \phi_1 y_{t-1} - \phi_2 y_{t-2} - \dots - \phi_p y_{t-p})^2}{2\sigma^2} .$$

Así, la estimación de los parámetros de la función logaritmo de verosimilitud condicional (2.70) se puede obtener mediante una regresión lineal de y_t sobre una constante y sus p valores de retraso.

Se puede demostrar que la estimación de σ^2 de la función logaritmo de verosimilitud condicional (2.70) es

$$\hat{\sigma}^2 = \frac{1}{T-p} \sum_{t=p+1}^T (y_t - \hat{c} - \hat{\phi}_1 y_{t-1} - \hat{\phi}_2 y_{t-2} - \dots - \hat{\phi}_p y_{t-p})^2 .$$

2.11. Función de verosimilitud para un proceso gaussiano $MA(1)$

Considérese un proceso gaussiano $MA(1)$, $Y_t = \mu + Z_t + \theta Z_{t-1}$, con $\{Z_t\} \sim iid N(0, \sigma^2)$. Sea $\boldsymbol{\theta} = (\mu, \theta, \sigma^2)$ que denota a los parámetros poblacionales a estimar. Si los valores de Z_{t-1} fueran conocidos, entonces

$$f_{Y_t|Z_{t-1}}(y_t|Z_{t-1}, \boldsymbol{\theta}) = \frac{1}{\sqrt{2\phi\sigma^2}} \exp \left[\frac{-(y_t - \mu - \theta Z_{t-1})^2}{2\sigma^2} \right] ,$$

Supóngase que $Z_0 = 0$, entonces $(Y_1|Z_0 = 0) \sim N(\mu, \sigma^2)$. Dada la observación y_1 , $y_1 = Z_1 - \mu$ y de nuevo,

$$f_{Y_2|Y_1, Z_0=0}(y_2|y_1, Z_0 = 0, \boldsymbol{\theta}) = \frac{1}{\sqrt{2\phi\sigma^2}} \exp \left[\frac{-(y_2 - \mu - \theta Z_1)^2}{2\sigma^2} \right] ,$$

y como Z_1 es conocido, entonces Z_2 se puede estimar como

$$Z_2 = y_2 - \mu - \theta Z_1 .$$

Procediendo de manera similar, partiendo de la suposición $Z_0 = 0$, se puede calcular la sucesión completa $\{Z_1, Z_2, \dots, Z_T\}$ a partir de los valores $\{y_1, y_2, \dots, y_T\}$ por iteración sobre

$$Z_t = y_t - \mu - \theta Z_{t-1}, \text{ para } t = 0, 1, \dots, T .$$

Así la función de densidad condicional de la t -ésima observación está dada por

$$f_{Y_t|Y_{t-1}, Y_{t-2}, \dots, Y_1, Z_0=0}(y_t|y_{t-1}, \dots, y_1, Z_0 = 0; \boldsymbol{\theta}) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left[-\frac{Z_t^2}{2\sigma^2} \right] .$$

De forma que, la función de verosimilitud estaría dada por el producto de las funciones anteriores de densidad individuales, esto es,

$$f_{Y_t, Y_{t-1}, Y_{t-2}, \dots, Y_1 | Z_0=0}(y_t, y_{t-1}, \dots, y_1 | Z_0 = 0; \boldsymbol{\theta}) = f_{Y_1 | Z_0=0}(y_1 | Z_0 = 0; \boldsymbol{\theta}) \prod_{t=2}^T f_{Y_t | Y_{t-1}, Y_{t-2}, \dots, Y_1, Z_0=0}(y_t | y_{t-1}, \dots, y_1, Z_0 = 0; \boldsymbol{\theta}) .$$

Con logaritmo de verosimilitud condicional

$$\mathcal{L}(\boldsymbol{\theta}) = -\frac{T}{2} \log(2\pi) - \frac{T}{2} \log(\sigma^2) - \sum_{t=1}^T \frac{Z_t^2}{2\sigma^2} . \quad (2.71)$$

2.12. Función de verosimilitud para un proceso gaussiano $MA(q)$

Considérese un proceso gaussino $MA(q)$, $Y_t = \mu + Z_t + \theta_1 Z_{t-1} + \dots + \theta_q Z_{t-q}$ con $\{Z_t\} \sim iid N(0, \sigma^2)$. Sea $\boldsymbol{\theta} = (\mu, \theta_1, \dots, \theta_q, \sigma^2)$ que denota a los parámetros poblacionales a estimar.

Una forma de aproximar la función de verosimilitud condicional, es suponer que $Z_0 = Z_{-1} = \dots = Z_{-q+1} = 0$ y considérese el vector $\boldsymbol{\varepsilon}_0 = (Z_0, Z_{-1}, \dots, Z_{-q+1})'$. Realizando un proceso similar al cálculo de la función de máxima verosimilitud de un proceso gaussiano $MA(1)$, se puede demostrar que

$$\mathcal{L}(\boldsymbol{\theta}) = -\frac{T}{2} \log(2\pi) - \frac{T}{2} \log(\sigma^2) - \sum_{t=1}^T \frac{Z_t^2}{2\sigma^2} . \quad (2.72)$$

Observación 2.11 *La expresión (2.72), es útil siempre que todos los valores de z para los cuales $1 + \theta_1 z + \dots + \theta_q z^q = 0$ estén fuera del círculo unitario.*

2.13. Función de verosimilitud para un proceso $ARMA(p, q)$

Un proceso gaussiano $ARMA(p, q)$, toma la forma $Y_t = c + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \dots + \phi_p Y_{t-p} + Z_t + \theta_1 Z_{t-1} + \theta_2 Z_{t-2} + \dots + \theta_q Z_{t-q}$, donde $Z_t \sim iid N(0, \sigma^2)$. El objetivo

es estimar el vector de parámetros $\boldsymbol{\theta} = (c, \phi_1, \phi_2, \dots, \phi_q, \theta_1, \theta_2, \dots, \theta_q)'$.

Una aproximación común a la función de verosimilitud de un proceso $ARMA(p, q)$ es condicionar y' s y Z' s. Tomando los valores iniciales para $\mathbf{y}_0 = (y_0, y_1, \dots, y_{p+1})'$ y $\mathbf{z}_0 = (Z_0, Z_{-1}, \dots, Z_{-q+1})'$ como valores dados, la sucesión $\{Z_1, Z_2, \dots, Z_T\}$ se puede calcular de $\{y_1, y_2, \dots, y_T\}$ por iteración sobre

$$\begin{aligned} Z_t &= y_t - c - \phi_1 y_{t-1} - \phi_2 y_{t-2} - \dots - \phi_p y_{t-p} \\ &\quad - \theta_1 Z_{t-1} - \theta_2 Z_{t-2} - \dots - \theta_q Z_{t-q}, \quad t = 1, 2, \dots, T. \end{aligned}$$

La función logaritmo de verosimilitud condicional se puede expresar como

$$\begin{aligned} \mathcal{L}(\boldsymbol{\theta}) &= \log f_{Y_T, Y_{T-1}, \dots, Y_1 | \mathbf{y}_0, \mathbf{z}_0}(y_T, y_{T-1}, \dots, y_1 | \mathbf{y}_0, \mathbf{z}_0, \boldsymbol{\theta}) \\ &= -\frac{T}{2} \log(2\pi) - \frac{T}{2} \sigma^2 - \sum_{t=1}^T \frac{Z_t^2}{2\sigma^2}. \end{aligned}$$

En [6], se recomienda elegir $Z_p = Z_{p-1} = \dots = Z_{p-q+1} = 0$ y considerar los p valores actuales de y , esto es y_1, y_2, \dots, y_p , así la función de verosimilitud calculada es

$$\begin{aligned} &\log f(y_T, y_{T-1}, \dots, y_{p+1} | y_p, y_{p-1}, \dots, y_1, Z_p = 0, Z_{p-1} = 0, \dots, Z_{p-q+1} = 0) \\ &= -\frac{T-p}{2} \log(2\pi) - \frac{T-p}{2} \log(\sigma^2) - \sum_{t=p+1}^T \frac{Z_t^2}{2\sigma^2}. \end{aligned}$$

Como en el caso de los procesos de medias móviles, estas aproximaciones deben ser utilizada siempre y cuando los valores de Z satisfagan que las soluciones de $1 + \theta_1 Z + \theta_2 Z^2 + \dots + \theta_q Z^q = 0$ y estén fuera del círculo unitario.

2.14. Ejemplos de procesos estacionales

Dada una serie estacional no estacionaria con n observaciones y_1, y_2, \dots, y_n , se presentan a continuación tres transformaciones que se pueden realizar para transformar la serie de tiempo en una serie de tiempo estacional y estacionaria.

Primeros valores diferenciados regulares. La transformación

$$y_t^* = y_t - y_{t-1},$$

produce los primeros valores diferenciados regulares, se utiliza la palabra regular para distinguir estas diferencias de los valores obtenidos por primera diferencia estacional.

Primeros valores diferenciados estacionales. Se denota por L el número de estaciones en un año ($L = 12$ para datos mensuales, $L = 52$ para datos semanales, etc.). La transformación

$$y_t^* = y_t - y_{t-L} ,$$

genera los primeros valores diferenciados estacionales.

Los primeros valores diferenciados regulares y los primeros valores diferenciados estacionales se generan con la transformación

$$\begin{aligned} y_t^* &= (y_t - y_{t-1}) - (y_{t-L} - y_{t-L-1}) \\ &= y_t - y_{t-1} - y_{t-L} + y_{t-L-1} . \end{aligned}$$

Para determinar si una transformación particular es adecuada, se examina el comportamiento de la FAM y FAMP de los valores $y_b^*, y_{b+1}^*, \dots, y_n^*$ que produce la transformación en el nivel no estacional y en el nivel estacional.

2.14.1. Procesos de promedios móviles estacionales de orden Q

El modelo

$$X_t = \delta + Z_t - \theta_{1,L}Z_{t-L} - \theta_{2,L}Z_{t-2L} - \dots - \theta_{Q,L}Z_{t-QL} ,$$

se denomina modelo de promedio móvil estacional de orden Q .

2.14.2. Procesos autorregresivos estacionales de orden P

El modelo

$$X_t = \delta + Z_t + \phi_{1,L}X_{t-L} - \phi_{2,L}X_{t-2L} + \dots + \phi_{P,L}X_{t-PL} ,$$

se llama modelo autorregresivo estacional de orden P .

Debido a que la FAM y la FAMP de los valores $y_b^*, y_{b+1}^*, \dots, y_n^*$ estacionarios estacionales de la serie de tiempo manifiestan con frecuencia un comportamiento tanto en el nivel no estacional como en el nivel estacional, se recomienda el siguiente procedimiento de tres pasos para identificar, en forma tentativa, un modelo que describa estos valores.

A continuación se presenta el procedimiento para ajustar modelos *SARIMA*.

1. Analizar de la FAM y FAMP en el nivel no estacional para identificar en forma tentativa, un modelo no estacional que describa los valores de la serie de tiempo.
2. Utilizar el comportamiento de la FAM y FAMP en el nivel estacional para identificar en forma tentativa, un modelo estacional que describa los valores de la serie de tiempo.
3. Combinar los modelos obtenidos en los dos pasos anteriores para llegar a un modelo global identificado tentativamente.

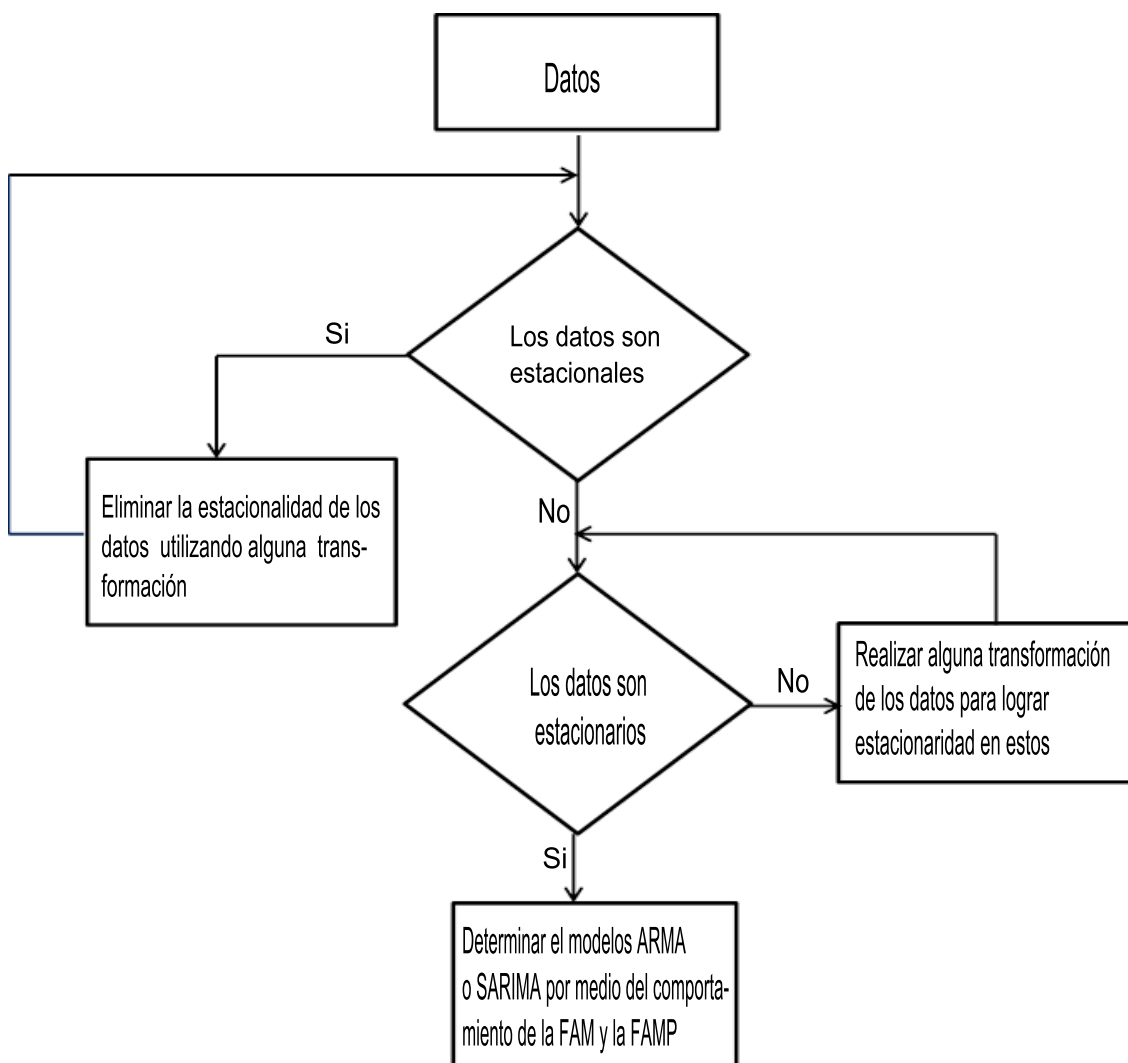


Figura 2.3: Diagrama de flujo

CAPÍTULO 3

Redes Neuronales

Las redes neuronales artificiales (RNA) se basan en el comportamiento y función del cerebro humano, en particular del sistema nervioso, el cual está compuesto por redes de neuronas biológicas que poseen bajas capacidades de procesamiento, sin embargo, toda su capacidad cognitiva se sustenta en la conectividad de éstas.

La unidad de una red neuronal artificial es un procesador elemental llamado neurona que posee la capacidad limitada de calcular, en general, una suma ponderada de sus entradas y luego le aplica una función de activación para obtener una señal que será transmitida a la próxima neurona. Estas neuronas artificiales se agrupan en capas o niveles y poseen un alto grado de conectividad entre ellas, conectividad que es ponderada por los pesos. A través de un algoritmo de aprendizaje supervisado o no supervisado, las RNA ajustan su arquitectura y parámetros de manera que puedan minimizar alguna función de error que indique el grado de ajuste a los datos y la capacidad de generalización de las RNA.

Antecedentes de las redes neuronales

Los primeros teóricos que concibieron los fundamentos de la computación neuronal fueron Warren McCulloch, un neurofisiólogo, y Walter Pitts, un matemático, quienes, en 1943, lanzaron una teoría acerca de la forma de trabajar de las neuronas (Un Cálculo Lógico de la Inminente Idea de la Actividad Nerviosa). Ellos modelaron una red neuronal simple mediante circuitos eléctricos [7].

En 1949 Donald Hebb, fue el primero en explicar los procesos del aprendizaje (que

es el elemento básico de la inteligencia humana) desde un punto de vista psicológico, desarrollando una regla de cómo el aprendizaje ocurría. En 1957, Frank Rosenblatt, comenzó el desarrollo del Perceptron. Esta es la red neuronal más antigua; utilizándose hoy en día para aplicación como identificador de patrones [7].

Para 1969, Marvin Minsky y Seymour Papert probaron (matemáticamente) que el perceptrón no era capaz de resolver problemas relativamente fáciles, tales como el aprendizaje de una función no lineal lo que llevó a demostrar que el perceptrón era muy débil, dado que las funciones no-lineales son extensamente empleadas en computación y en los problemas del mundo real. Para 1985, comienza el renacimiento de las redes neuronales con la publicación del libro: Computación neuronal de decisiones en problemas de optimización por John Hopfield, y en 1986, David Rumelhart y G. Hinton, redescubrieron el algoritmo de aprendizaje de propagación hacia atrás (back-propagation)[7].

A partir de 1986, el panorama fue alentador con respecto a las investigaciones y el desarrollo de las redes neuronales. En la actualidad, son numerosos los trabajos que se realizan y publican entorno a las redes neuronales pues son un método para resolver problemas, de forma individual o combinadas con otros métodos, para aquellas tareas de clasificación, identificación, diagnóstico, optimización o predicción en las que el balance datos/conocimiento se inclina hacia los datos y donde, adicionalmente, puede haber la necesidad de aprendizaje en tiempo de ejecución y de cierta tolerancia a fallos. En estos casos las RNA se adaptan dinámicamente reajustando constantemente los pesos de sus interconexiones [7].

3.1. Definiciones básicas de redes neuronales

Desde el punto de vista biológico, las redes neuronales son el principal elemento del Sistema Nervioso y están compuestas por un gran número de elementos llamados neuronas. Una neurona es una célula compuesta por un cuerpo o soma, un número de extensiones llamadas dendritas, que sirven de entradas, y una larga extensión llamada axón que actúa como salida. La sinapsis conecta el axón de una neurona a las dendritas de las otras neuronas. Las neuronas están dispuestas en capas. En general, las neuronas de una capa reciben entradas desde otra capa y envían sus salidas a neuronas de una tercera. Dependiendo de la aplicación también es posible que las neuronas de una capa reciban entradas y provean salidas a neuronas de la misma capa (Figura 3.1).

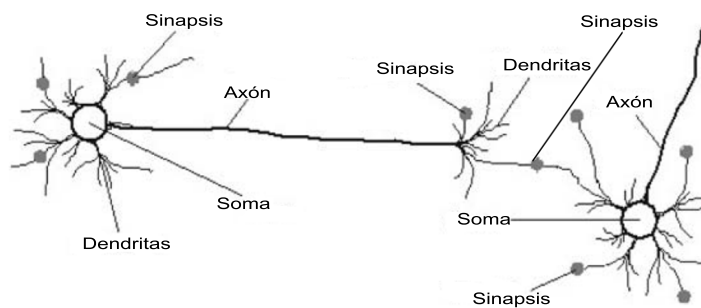


Figura 3.1: Esquema de una red neuronal biológica

3.2. Redes neuronales artificiales

Una **neurona** es una unidad de procesamiento de información para el funcionamiento de una red neuronal. En un modelo de redes neuronales se pueden identificar tres elementos básicos

1. Un conjunto de sinapsis o enlaces de conexión, cada uno de los cuales es caracterizado por un peso. Específicamente una señal x_j en la entrada de sinapsis j conectada a la neurona k es multiplicada por el peso sináptico ω_{kj} .
2. Un sumador para agregar las señales de entrada ponderado por las respectivas sinapsis de la neurona (las operaciones descritas aquí constituyen una combinación lineal).
3. Una función de activación para limitar la amplitud de la salida de una neurona (Figura 3.2).

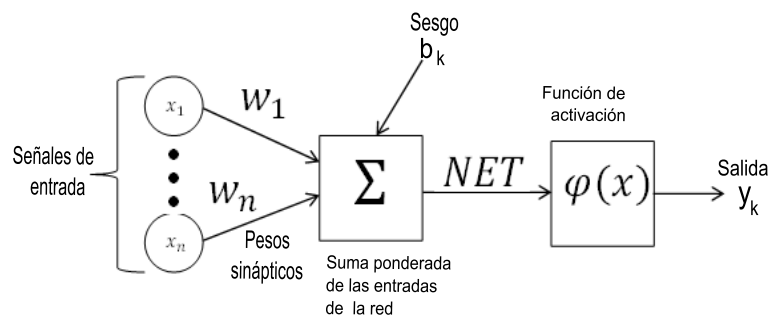


Figura 3.2: Esquema de una red neuronal artificial

En términos matemáticos se puede describir una neurona k mediante el siguiente par de elementos

$$u_k = \sum_{j=1}^m \omega_{kj} x_j , \quad (3.1)$$

$$y_k = \varphi(u_k + b_k) , \quad (3.2)$$

donde x_1, x_2, \dots, x_m son las señales de entrada, $\omega_{k1}, \omega_{k2}, \dots, \omega_{km}$ son los pesos sinápticos de la neurona k , u_k es la salida de la combinación lineal debido a las señales de entrada, b_k es el sesgo, $\varphi(\cdot)$ es la función de activación y y_k es la salida de activación de la neurona. El uso de la ganancia b_k tiene el efecto de aplicar una transformación afín a la salida u_k de la combinación lineal del modelo mostrado en la Figura 3.2, el cual es llamado entrada neta de la red y está dada por

$$net_k = u_k + b_k . \quad (3.3)$$

En particular, dependiendo si el sesgo b_k es positivo o negativo, la relación entre el campo inducido local o el potencial de activación y_k de la neurona k y de la combinación lineal de la salida u_k es modificado en la forma que se ilustra en la Figura 3.3.

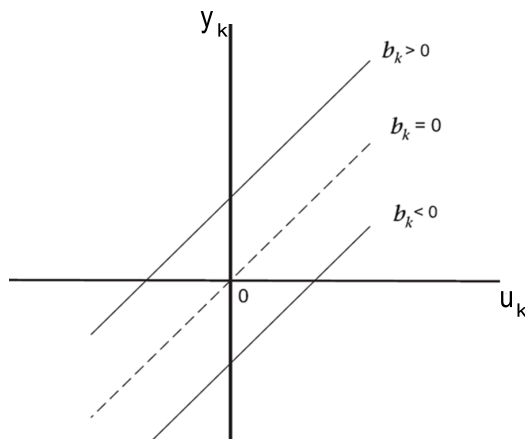


Figura 3.3: Transformación afín producida por la presencia del sesgo

3.3. Tipos de función de activación

La función de activación, denotada por $\varphi(net)$, define la salida de una neurona en términos del campo local inducido net . A continuación se presentan tres tipos de funciones de activación básicas.

Función umbral. Este tipo de función está definida por

$$\varphi(net) = \begin{cases} 1, & \text{si } net \geq 0, \\ 0, & \text{si } net < 0. \end{cases} \quad (3.4)$$

Función lineal a trozos,

$$\varphi(net) = \begin{cases} 0, & \text{si } net \leq -\frac{1}{2}, \\ net + \frac{1}{2}, & \text{si } -\frac{1}{2} < net < \frac{1}{2}, \\ 1, & \text{si } net \geq \frac{1}{2}. \end{cases} \quad (3.5)$$

Función sigmoïdal. Este tipo de función es la más usada en la construcción de redes neuronales artificiales. Un ejemplo de la función sigmoïdal es la función logística la cual se define como

$$\varphi(net) = \frac{1}{1 + \exp(-a \cdot net)}, \quad (3.6)$$

donde a es el parámetro de la pendiente de la función sigmoïdal.

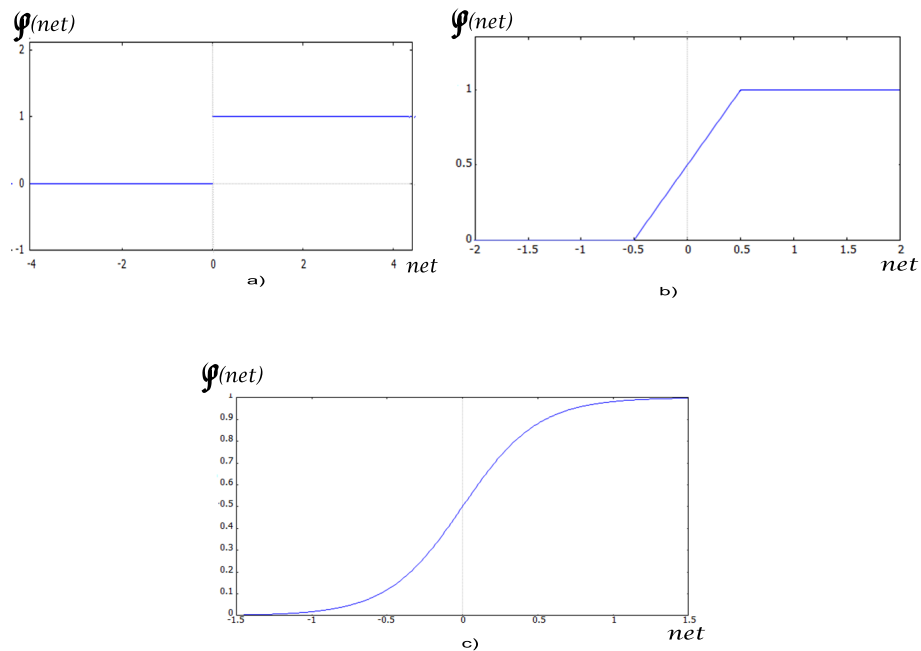


Figura 3.4: Funciones de activación: a) Función umbral, b) Función lineal a trozos, c) Función logística

3.4. Redes neuronales vistas como grafos dirigidos

Un grafo dirigido es una red de enlaces dirigidos que son interconectados en ciertos puntos llamados nodos. Un nodo j se encuentra asociado a otro nodo con señal x_j . Un enlace dirigido se origina en el nodo j y termina en el nodo k ; tiene una función de transferencia asociada que especifica la forma en que la señal y_k en el nodo k depende de la señal x_j en el nodo j . Las señales de flujo de la red en todas las partes del grafo están indicadas por tres reglas:

1. Una señal fluye a lo largo de un enlace sólo en la dirección definida por una flecha en el enlace (Figura 3.5 incisos (a) y (b)).
2. La señal de un nodo es igual a la suma algebraica de todas las señales que entran en el nodo pertinente a través de los enlaces entrantes (Figura 3.5 inciso (c)).
3. La señal de un nodo es transmitida a cada enlace de salida de aquel nodo con la función de activación totalmente independiente de las funciones de activación de los enlaces salientes (Figura 3.5 inciso (d)).

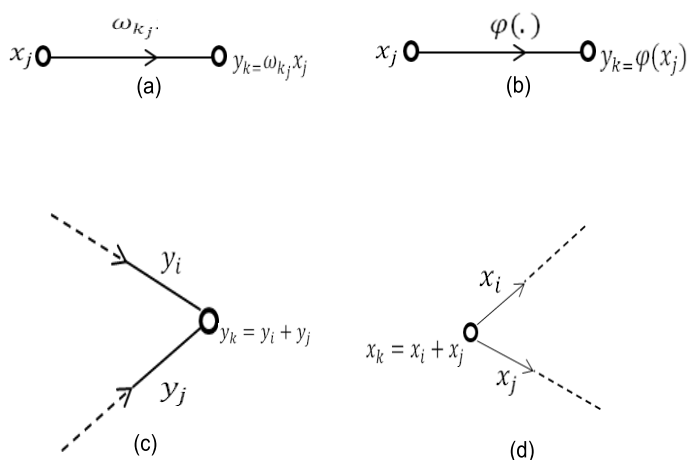


Figura 3.5: Reglas básicas para la construcción de grafos dirigidos

En base a las reglas anteriores y a la Figura 3.6, se puede definir matemáticamente a una red neuronal como sigue.

Definición 3.1 Una red neuronal es un grafo con nodos dirigidos con interconexiones sinápticas, enlaces de activación y se encuentra caracterizada por las siguientes cuatro propiedades

1. Cada neurona es representada por un conjunto de enlaces sinápticos lineales, un sesgo aplicado externamente y posiblemente un enlace de activación no lineal. El sesgo es representado por un enlace sináptico conectado a una entrada fija en $+1$.
2. Los enlaces sinápticos de una neurona pondera sus respectivas señales de entrada.
3. La suma ponderada de las señales de entrada define el campo local inducido de la neurona en cuestión.
4. El enlace de activación transforma el campo local de una presentación de la neurona para producir una salida.

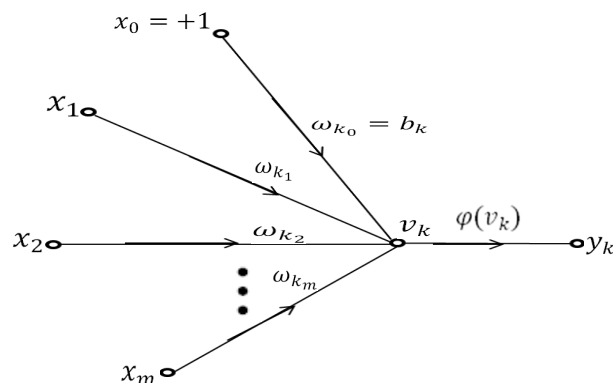


Figura 3.6: Gráfico de flujo de señal de una neurona

3.5. Arquitectura de una red neuronal

La forma en que las neuronas de una red neuronal son estructuradas está íntimamente ligada con el algoritmo de aprendizaje utilizado para entrenar la red. En general, se pueden identificar tres tipos de arquitecturas de redes neuronales.

3.5.1. Red de alimentación directa de una sola capa

En una red neuronal en capas las neuronas están organizadas en forma de capas. En la forma más simple de una red en capas, tenemos una capa de entrada de nodos de origen que se proyecta sobre una capa de salida de las neuronas (nodos de cálculo), pero no viceversa (Figura 3.7).

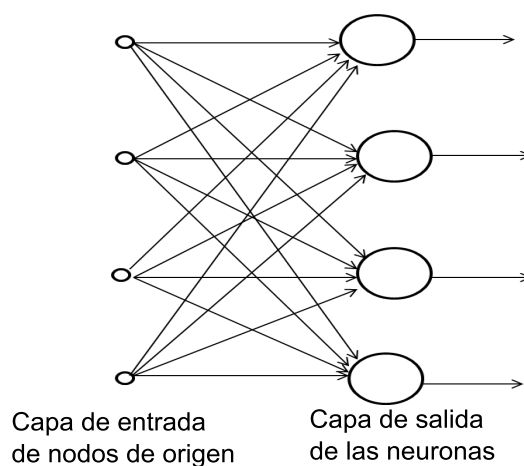


Figura 3.7: Representación gráfica de una red de una sola capa

3.5.2. Red de alimentación directa multicapa

Este tipo de redes neuronales se distinguen por la presencia de una o más capas ocultas cuyos nodos de cálculo se conocen como neuronas ocultas o unidades ocultas. La función de estas neuronas ocultas es intervenir entre la entrada externa y la salida de la red de alguna forma útil.

La señal de entrada en la capa de entrada de la red proporciona respectivamente patrones de activación las cuales constituyen la señal de entrada aplicada a la neurona en la segunda capa (es decir, la primera capa oculta). La señal de salida de la segunda capa es usada como entrada en la tercer capa y así sucesivamente para las demás capas de la red. Las neuronas en cada capa de la red tienen como entrada la señal procedente de la capa inmediata anterior. Luego, el conjunto de señales de salida final en la capa oculta de la red lo constituyen las respuestas totales de activación, las cuales fueron proporcionadas por la señal de entrada en la primera capa.

En la Figura 3.8 se muestra el diseño de una red multicapa para el caso de una capa oculta. Para abreviar la red de la Figura 3.8, el lector se puede referir a ésta como una $10 - 4 - 2$ red porque tiene 10 nodos de origen, 4 neuronas ocultas y 2 neuronas de salida.

3.5.3. Redes recurrentes

Este tipo de red se distingue de las redes con conexión hacia adelante porque existe al menos un bucle o ciclo de retroalimentación. Por ejemplo, una red recurrente puede consistir de una red de una sola capa cuya señal de salida es señal de entrada para

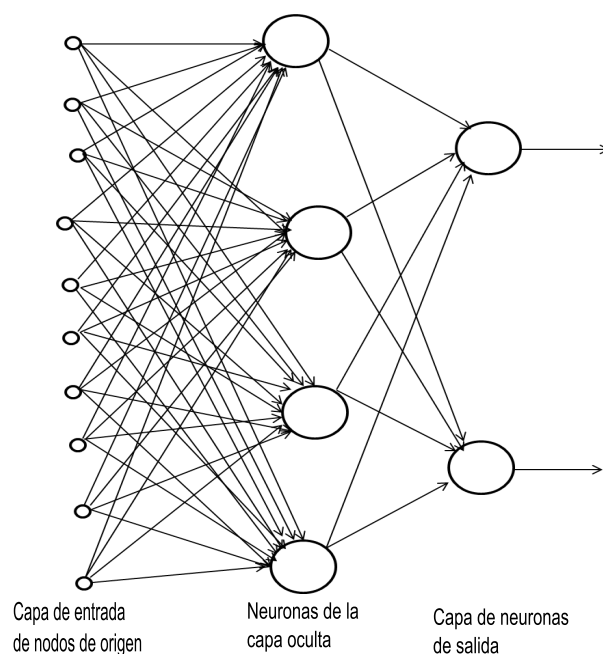


Figura 3.8: Representación gráfica de una red multicapa

las demás neuronas. De manera general este tipo de red es aquella cuya señal de entrada es también de salida para la red (ver Figura 3.9).

3.6. Representación del conocimiento de una red neuronal

Definición 3.2 *Se entiende por conocimiento de una red a la información histórica o modelos usados por una persona o máquina para interpretar, predecir y responder apropiadamente al mundo exterior.*

La representación del conocimiento de una red neuronal es complicada, sin embargo, existen cuatro reglas generales:

1. Entradas similares de clases iguales deben producir representaciones semejantes dentro de la red neuronal y por lo tanto, deben representarse como pertenecientes de la misma categoría. Dos medidas con las que se puede determinar la similitud entre entradas son:

a) **Distancia euclidiana.** Sea \mathbf{x}_i un vector de tamaño $m \times 1$

$$\mathbf{x}_i = [x_{i1}, x_{i2}, \dots, x_{im}]'$$

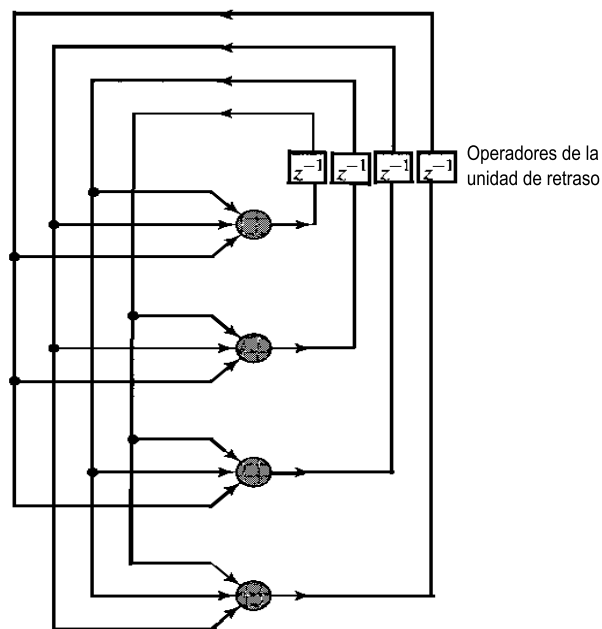


Figura 3.9: Red neuronal recurrente sin bucles de autoalimentación y sin neuronas ocultas

La distancia euclidiana entre dos vectores de tamaño $m \times 1$, $\mathbf{x}_i, \mathbf{x}_j$ se define por

$$d(\mathbf{x}_i, \mathbf{x}_j) = \|\mathbf{x}_i - \mathbf{x}_j\| = \left[\sum_{k=1}^m (x_{ik} - x_{jk})^2 \right]^{1/2}, \quad (3.7)$$

donde x_{ik} y x_{jk} son los elementos k -ésimos de los vectores \mathbf{x}_i y \mathbf{x}_j respectivamente. Así mientras más pequeña sea la distancia euclidiana entre estos elementos mayor será la similitud entre éstos.

b) **Producto interior.** Dados dos vectores de tamaño $m \times 1$, $\mathbf{x}_i, \mathbf{x}_j$, se define su producto interior como

$$(\mathbf{x}_i, \mathbf{x}_j) = \sum_{k=1}^m x_{ik} x_{jk}. \quad (3.8)$$

Así, mientras más grande sea el producto interior de estos vectores mayor será la similitud entre \mathbf{x}_i y \mathbf{x}_j .

2. Artículos que sean clasificados como clases separadas deben tener diferentes representaciones en la red.

3. Si alguna característica es importante, entonces debe haber un gran número de neuronas implicadas en la representación de esa característica en la red neuronal.
4. Estipular la información previa y las variaciones, deben ser incorporadas en el diseño, para no tener que incorporarlas en su proceso de aprendizaje.

3.7. Procesos de aprendizaje de una red neuronal

Dentro del contexto de redes neuronales, se puede definir el aprendizaje como sigue.

Definición 3.3 *El aprendizaje es un proceso por el cual los parámetros libres de una red neuronal son adaptados a través de un proceso de simulación.*

El tipo de aprendizaje está determinado por la manera en que los cambios de los parámetros libres tienen lugar. El conjunto de reglas bien definidas para la solución de un problema de aprendizaje es llamado **algoritmo de aprendizaje**.

A continuación se describen de manera general algunos procesos de aprendizaje para redes neuronales.

3.7.1. Aprendizaje error-corrección

Considérese el caso de una neurona k constituida por el único nodo computacional en la capa de salida de una red como en la Figura 3.10. La neurona k es impulsada por un vector $\mathbf{x}(n)$ producido por una o más capas de neuronas ocultas las cuales son a su vez impulsadas por un vector de entrada aplicado a los nodos de origen de la red neuronal. El argumento n denota tiempo discreto, es decir; el tiempo de paso de un proceso iterativo que participa en el ajuste de los pesos sinápticos de la neurona k .

La señal de salida de una neurona $y_k(n)$ se compara con una respuesta deseada o salida destino $d_k(n)$. Una señal de error $e_k(n) = d_n(k) - y_n(k)$, acciona un mecanismo de control, su propósito es la aplicación de una secuencia de ajustes correctivos los cuales están diseñados para hacer que la señal de salida $y_k(n)$ se acerque a la respuesta deseada $d_k(n)$ paso a paso. Este objetivo se consigue minimizando la función de costo

$$\mathcal{E}(n) = \frac{1}{2}e_k(n) \quad , \quad (3.9)$$

donde $\mathcal{E}(n)$ es un valor instantáneo del error de energía. El ajuste de los pesos sinápticos de la neurona k , realizado paso a paso, se continúa hasta que el sistema alcanza un estado de equilibrio. En particular la minimización de la función $\mathcal{E}(n)$ lleva a aplicar una regla de aprendizaje conocida como Regla Delta [12].

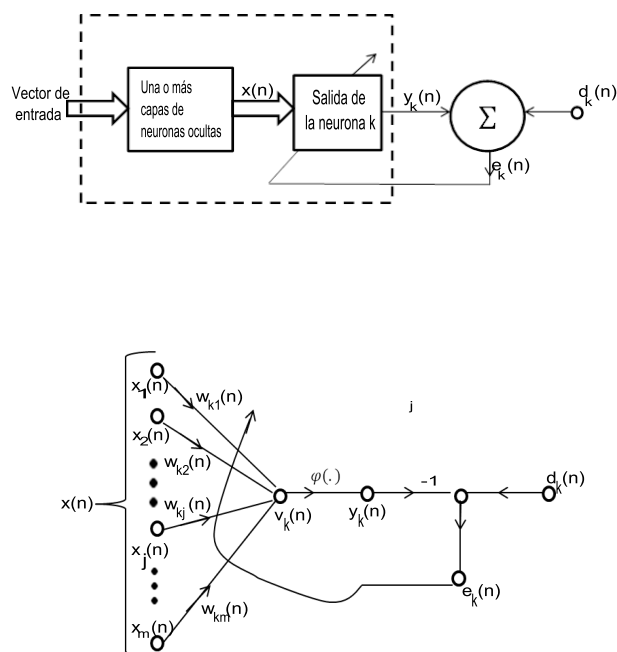


Figura 3.10: Representación gráfica del aprendizaje de error-corrección

3.8. Aprendizaje basado en memoria

En el aprendizaje basado en memoria, las experiencias pasadas se almacenan de forma explícita en un gran insumo de entrada-salida correctamente clasificado, por ejemplo: $\{(\mathbf{x}_i, d_i)\}_{i=1}^N$ donde \mathbf{x}_i denota un vector de entrada y d_i la correspondiente respuesta deseada. Sin pérdida de generalidad, se restringen las respuestas deseadas a un escalar.

Cuando se requiere la clasificación de un nuevo vector \mathbf{x}_{test} , el algoritmo responde retrocediendo y analizando el tratamiento de los datos en una vecindad local de \mathbf{x}_{test} (Figura 3.11). Todos los algoritmos basados en memoria, involucran dos cosas:

1. Criterio utilizado para definir la vecindad local del vector \mathbf{x}_{test} .
2. Una regla de aprendizaje utilizada para entrenar ejemplos en la vecindad local de \mathbf{x}_{test} .

Los algoritmos difieren unos de otros en la forma en que éstos dos ingredientes se definen. Una forma simple pero eficaz del aprendizaje basado en memoria se conoce como la regla del vecino más cercano. Se define una vecindad local como el ejemplo de entrenamiento que se encuentra en la vecindad inmediata del vector \mathbf{x}_{test} .

En particular se dice que el vector

$$\mathbf{x}'_N \in \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N\} , \quad (3.10)$$

es la vecindad más cercana de \mathbf{x}_{test} si

$$\min_i d(\mathbf{x}_i, \mathbf{x}_{test}) = d(\mathbf{x}'_N, \mathbf{x}_{test}) , \quad (3.11)$$

donde $d(\mathbf{x}_i, \mathbf{x}_{test})$ es la distancia euclideana entre los vectores \mathbf{x}_i y \mathbf{x}_{test} .

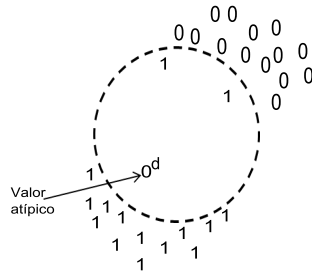


Figura 3.11: Representación gráfica del aprendizaje basado en memoria

3.8.1. Aprendizaje Hebbiano

Esta regla de aprendizaje es la más antigua y la más utilizada de todas las reglas de aprendizaje y está basada en las siguientes dos partes:

1. Si dos neuronas en cada lado de una sinapsis (conexión) se activan simultáneamente, entonces la fuerza de sinapsis se incrementa selectivamente.
2. Si dos neuronas en cada lado de una sinapsis se activan de forma asincrónica, entonces esa sinapsis se debilita o se elimina selectivamente. Una sinapsis, es llamada una sinapsis Hebbiana siempre que ésta dependa del tiempo, con un mecanismo local, interactivo y fuerte para aumentar la eficiencia sináptica como una función de la correlación entre las actividades presinápticas y postsinápticas.

En términos matemáticos consideremos un peso sináptico ω_{jk} de una neurona k con señales presinápticas y postsinápticas denotada por x_j y y_k respectivamente. El ajuste aplicado al peso sináptico ω_{jk} en el instante n se expresa en forma general como

$$\Delta\omega_{kj}(n) = F(x_j(n), y_k(n)), \quad (3.12)$$

donde $F(\cdot, \cdot)$ es una función de la señal presináptica y postsináptica. Las señales $x_j(n)$ y $y_k(n)$ se tratan algunas veces sin dimensiones.

La expresión (3.12), admite muchas formas las cuales se clasifican como Hebbianas (Ver [12]).

3.8.2. Aprendizaje competitivo

En este tipo de aprendizaje, las neuronas de salida de una red neuronal compiten entre ellas para llegar a estar activas. Mientras que en una red neuronal basada en el aprendizaje Hebbiano muchas neuronas de salida pueden estar activas simultáneamente, en el aprendizaje competitivo sólo una neurona de salida está activa en cualquier instante. Esta característica hace que el aprendizaje competitivo sea muy adecuado para descubrir características estadísticamente sobresalientes que se pueden utilizar para clasificar un conjunto de patrones de entrada.

Elementos básicos de la regla del aprendizaje competitivo:

1. Un conjunto de neuronas que sean todas iguales a excepción de algunos pesos sinápticos distribuidos al azar, y respondan de manera diferente a un determinado conjunto de patrones de entrada.
2. Un límite impuesto a la fuerza de cada neurona.
3. Un mecanismo que permita a las neuronas competir por el derecho a responder a un subconjunto de entradas dado, de tal forma que únicamente una neurona de salida o solamente una neurona por grupo, está activa en cada instante.

Para una neurona k activa, su campo local es inducido por v_k y éste, debe ser el más grande que cualquier otra neurona, para un patrón de entrada \mathbf{x} especificado. A la señal de salida y_k de la neurona ganadora k se le asigna el valor 1. A las señales de salida de todas las neuronas que perdieron la competencia le es asignado el valor de cero. Así se tiene:

$$y_k = \begin{cases} 1, & \text{si } v_k > v_j \quad \forall j, j \neq k, \\ 0, & \text{en otro caso.} \end{cases}$$

Aquí se entiende como campo local inducido a un conjunto de acciones combinadas del avance y retroalimentación de todas las entradas a la neurona k .

Sea ω_{kj} el peso sináptico de la conexión del nodo j a la neurona k . Supóngase que cada neurona tiene asignada una cantidad fija de pesos sinápticos, el cuál es distribuido entre sus nodos de entrada como

$$\sum_j \omega_{kj} = 1, \text{ para todo } k \cdot \quad (3.13)$$

Así una neurona aprende por el intercambio de pesos sinápticos de sus nodos de entrada inactivos a sus nodos de entrada activos. Si una neurona no responde a un patrón de entrada en particular no hay aprendizaje en esa neurona.

Si una neurona en particular gana la competencia, cada nodo de entrada de esa neurona renuncia a una parte de sus pesos sinápticos, para ser distribuido en partes iguales entre todos los nodos de entrada activos. De acuerdo a la regla de aprendizaje competitivo, el cambio $\Delta\omega_{kj}$ aplicado al peso sináptico ω_{kj} es definido por

$$\Delta\omega_{kj} = \begin{cases} \eta(x_j - \omega_{kj}) , & \text{si la neurona } k \text{ gana la competencia} , \\ 0 , & \text{si la neurona } k \text{ pierde la competencia} . \end{cases} \quad (3.14)$$

Donde η es el parámetro de tasa de aprendizaje. Esta regla de aprendizaje tiene el efecto global de mover el vector de peso sináptico ω_k de la neurona ganadora k hacia el patrón de entrada \mathbf{x} .

3.9. Perceptrones de una sola capa

El perceptrón es la forma más simple de una red neuronal la cual es utilizada para la clasificación de patrones que se dicen linealmente separables. De manera general, el perceptrón consiste de una neurona con pesos sinápticos, un sesgo y una función umbral. En esta sección, se describe el modelo de perceptrón y el teorema de convergencia del perceptrón.

El perceptrón de una sola capa, produce una salida igual a $+1$ si la entrada de la función umbral es positiva y -1 si es negativa. Considérese la señal de flujo de la Figura 3.12, se denota por $\omega_1, \omega_2, \dots, \omega_m$ a los pesos sinápticos del perceptrón, x_1, x_2, \dots, x_m denotan las entradas aplicadas al perceptrón y b el sesgo aplicado. Del modelo, se observa que la entrada de la función umbral también conocida como campo local inducido de la neurona es

$$v = \sum_{i=1}^m \omega_i x_i + b \cdot \quad (3.15)$$

El objetivo del perceptrón es clasificar correctamente el conjunto x_1, x_2, \dots, x_m en dos clases \mathcal{C}_1 o \mathcal{C}_2 , la regla de decisión para esta clasificación es asignar los puntos representados por x_1, x_2, \dots, x_m a las clases \mathcal{C}_1 si la salida del perceptrón es $+1$ y

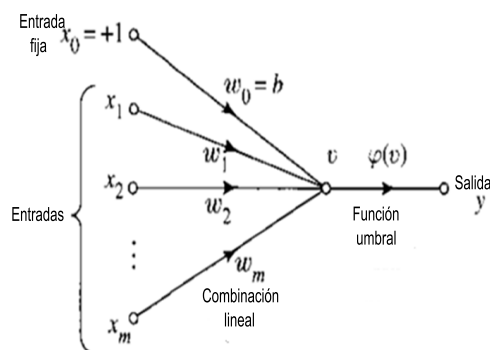


Figura 3.12: Representación gráfica de un perceptrón

a la clase \mathcal{C}_2 si la salida del perceptrón es -1 . En la forma más simple, el perceptrón tiene dos regiones de decisión separadas por el hiperplano definido por

$$\sum_{i=1}^m \omega_i x_i + b = 0 \quad (3.16)$$

La Figura 3.13, ilustra una región de decisión para el caso de dos variables de entrada x_1, x_2 , para las cuales la frontera de decisión toma la forma de una recta.

Para el caso más general del perceptrón, los pesos sinápticos $\omega_1, \omega_2, \dots, \omega_m$ del perceptrón se pueden adaptar en una base iteración por iteración, por medio de la regla de aprendizaje de error-corrección, conocida como **Algoritmo de convergencia del perceptrón**.

Para derivar el algoritmo de aprendizaje de error-corrección para el perceptrón, se considera trabajar con una modificación del perceptrón de la Figura 3.12, la cual se muestra en la Figura 3.14.

En esta modificación, el sesgo es tratado como un peso sináptico estipulado por una entrada fija igual a $+1$, así, se puede definir el vector de entrada de tamaño $((m+1) \times 1)$, $\mathbf{x}_*(n) = (1, x_1(n), x_2(n), \dots, x_m(n))'$, donde n denota la iteración en la aplicación del algoritmo, de igual forma se define el vector de tamaño $((m+1) \times 1)$ de pesos sinápticos $\boldsymbol{\omega}_*(n) = (b, \omega_1(n), \omega_2(n), \dots, \omega_m(n))'$.

Para el perceptrón, las dos clases deben ser linealmente separables, esto significa que los patrones deben ser suficientemente separados unos de otros para asegurar que la región de decisión consista de un hiperplano, para el caso de dos dimensiones, lo anterior se muestra en la Figura 3.15.

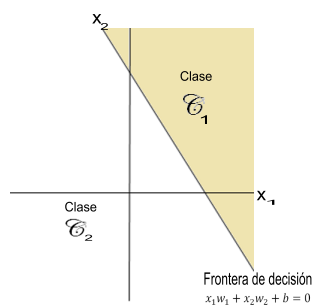


Figura 3.13: Regla de decisión para un perceptrón con dos entradas x_1, x_2

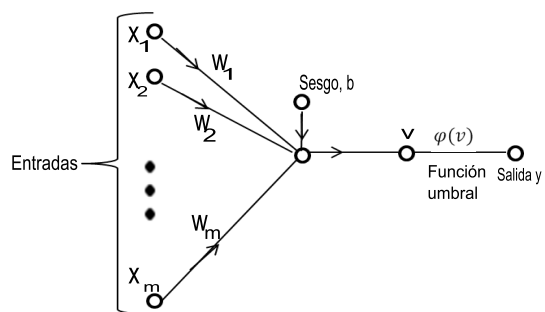


Figura 3.14: Representación gráfica del perceptrón equivalente a la Figura 3.12

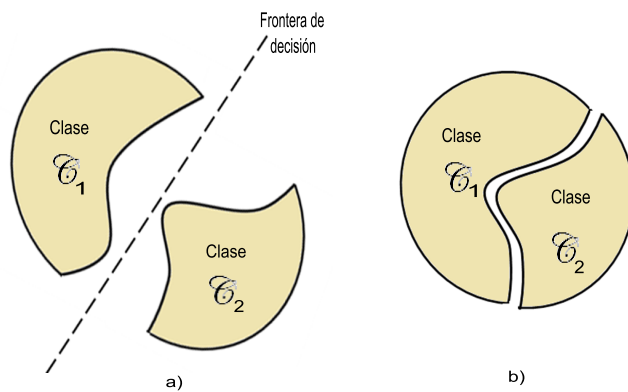


Figura 3.15: Regla de clasificación. a) Clases linealmente separables, b) Clases no linealmente separables

Supóngase que las variables de entrada del perceptrón provienen de dos clases lineal-

mente separables. Sea \mathcal{H}_1 el subconjunto de vectores $\mathbf{x}_1(1), \mathbf{x}_2(2), \dots$, que pertenecen a la clase \mathcal{C}_1 , y \mathcal{H}_2 , el subconjunto de vectores $\mathbf{x}_1(1), \mathbf{x}_2(2), \dots$, que pertenecen a la clase \mathcal{C}_2 .

Dados los subconjuntos de vectores \mathcal{H}_1 y \mathcal{H}_2 , para clasificar el proceso de tratamiento, se debe ajustar del vector de peso $\boldsymbol{\omega}$ de tal forma que las clases \mathcal{C}_1 , y \mathcal{C}_2 sean linealmente separables, esto es, existe un vector de peso $\boldsymbol{\omega}$ tal que

$$\begin{aligned} \boldsymbol{\omega}_*^t \mathbf{x} &> 0; \text{ para cada vector de entrada que pertenece a la clase } \mathcal{C}_1. \\ \boldsymbol{\omega}_*^t \mathbf{x} &\leq 0; \text{ para cada vector de entrada que pertenece a la clase } \mathcal{C}_2. \end{aligned} \quad (3.17)$$

Dados los subconjuntos de vectores de tratamiento \mathcal{H}_1 y \mathcal{H}_2 , el problema elemental del perceptrón es encontrar un vector de peso $\boldsymbol{\omega}$ tal que las dos desigualdades en (3.17) se satisfagan.

El algoritmo para adaptar el vector de peso se puede formular como sigue:

1. Si el n -ésimo miembro del conjunto de tratamiento, $\mathbf{x}(n)$, es correctamente clasificado por el vector de peso $\boldsymbol{\omega}(n)$ calculado en la n -ésima iteración del algoritmo, no se realiza ninguna corrección al vector de peso del perceptrón, esto es:

$$\begin{aligned} \boldsymbol{\omega}_*(n+1) &= \boldsymbol{\omega}_*(n) \text{ si } \boldsymbol{\omega}_*^t \mathbf{x}(n) > 0 \text{ y } \mathbf{x}(n) \in \mathcal{C}_1, \\ \boldsymbol{\omega}_*(n+1) &= \boldsymbol{\omega}_*(n) \text{ si } \boldsymbol{\omega}_*^t \mathbf{x}(n) \leq 0 \text{ y } \mathbf{x}(n) \in \mathcal{C}_2. \end{aligned} \quad (3.18)$$

2. En otro caso, el vector de peso del perceptrón se actualiza de acuerdo con la regla:

$$\begin{aligned} \boldsymbol{\omega}_*(n+1) &= \boldsymbol{\omega}_*(n) - \eta(n)\mathbf{x}(n) \text{ si } \boldsymbol{\omega}_*^t \mathbf{x}(n) > 0 \text{ y } \mathbf{x}(n) \in \mathcal{C}_2, \\ \boldsymbol{\omega}_*(n+1) &= \boldsymbol{\omega}_*(n) + \eta(n)\mathbf{x}(n) \text{ si } \boldsymbol{\omega}_*^t \mathbf{x}(n) \leq 0 \text{ y } \mathbf{x}(n) \in \mathcal{C}_1, \end{aligned} \quad (3.19)$$

donde la tasa del parámetro de aprendizaje $\eta(n)$ controla el ajuste aplicado al vector de peso en la iteración n .

3.10. Perceptrón multicapa

Una red multicapa de alimentación hacia adelante consiste de un conjunto de unidades sensoriales o nodos de origen que constituyen la capa de entrada, una o más capas ocultas de nodos de cálculo y una capa de salida de nodos de cálculo. La señal

de entrada se propaga a través de la red, en dirección hacia adelante capa por capa. Para resolver problemas con esta herramienta se utiliza el *algoritmo back-propagation* el cual está basado en la regla de aprendizaje de error corrección.

Un perceptrón multicapa se distingue por tres características:

1. El modelo de cada neurona en la red incluye una función de activación no lineal diferenciable en cualquier punto.
2. La red contiene una o más capas ocultas de neuronas que no son parte de la entrada o salida de la red. Estas neuronas ocultas permiten a la red aprender tareas complejas por extracción progresiva de hechos significativos de los patrones de entrada (vectores).
3. La red muestra un alto grado de conectividad determinada por la sinapsis de la red. Un cambio en la conectividad de la red requiere un cambio en las conexiones sinápticas o los pesos de la población.

Se considera una red completamente conectada siempre que una neurona en cualquier capa de la red esté conectada a todos los nodos o neuronas de la capa anterior, como se muestra en la Figura 3.16.

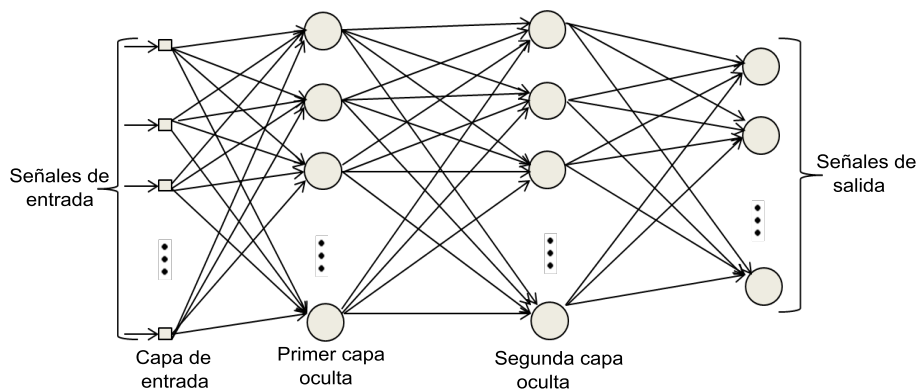


Figura 3.16: **Red completamente conectada**

De la Figura 3.16, se pueden identificar dos tipos de señales:

1. *Señal de entrada*. Una señal de entrada comienza en el extremo de la entrada de la red, se propaga hacia adelante neurona por neurona a través de la red y emerge en el extremo de la salida de la red como una señal de salida.
2. *Señal de error*. Se origina en la salida de una neurona de la red y se propaga hacia atrás capa por capa a través de la red.

Las neuronas de salida o nodos de cálculo constituyen las capas de salida de la red. El resto de las neuronas (también nodos de cálculo), constituyen capas ocultas de la red. Así, las unidades ocultas no son parte de la salida o entrada de la red (de aquí su designación como oculta). Cada neurona oculta o de salida de un perceptrón multicapa está designado a desarrollar dos cálculos:

1. El cálculo de la señal de entrada apareciendo en la salida de una neurona, el cuál se expresa como una función continua no lineal de la señal de entrada y los pesos sinápticos asociados con esa neurona.
2. El cálculo de una estimación de un vector gradiente el cual es requerido para el paso hacia atrás a través de la red.

Para describir el algoritmo *back-propagation*, se hace uso de la siguiente notación:

- Los índices i , j y k , referirán diferentes neuronas en la red, con señales de propagación a través de la red de izquierda a derecha. Así la neurona j , se encuentra en una capa a la derecha de la neurona i y a la izquierda de la neurona k , cuando la neurona j es una unidad oculta.
- En la iteración n , el n -ésimo patrón de tratamiento se presenta a la red.
- El símbolo $\mathcal{E}(n)$ se refiere a la suma instantánea del cuadrado de los errores en la iteración n . El promedio de $\mathcal{E}(n)$ sobre todos los valores de n , es el promedio del cuadrado de los errores y se denota por \mathcal{E}_{av} .
- El símbolo $e_j(n)$ se refiere a la señal de error en la salida de la neurona j en la iteración n .
- El símbolo $d_j(n)$ se refiere a la respuesta deseada de la neurona j y se usa para calcular e_j .
- El símbolo $y_j(n)$ se refiere a la señal de entrada que aparece en la salida de la neurona j en la iteración n .
- El símbolo $\omega_{ji}(n)$ denota la conexión del peso sináptico a la salida de la neurona i a la entrada de la neurona j en la iteración n . La corrección aplicada a este peso en la iteración n se denota por $\Delta\omega_{ji}(n)$.
- El campo local inducido (suma ponderada de los pesos sinápticos de todas las entradas sinápticas más un sesgo) de la neurona j en la iteración n , se denota por $v_j(n)$; este campo constituye la señal aplicada a la función de activación asociada con la neurona j .
- La función de activación que describe la relación funcional de entrada-salida de la no linealidad asociada con la neurona j se denota por $\phi_j(\cdot)$.

- El sesgo aplicado a la neurona j se denota por b_j . Este efecto está representado por un peso sináptico $\omega_{j0} = b_j$ conectado a una entrada fija igual a $+1$.
- El i -ésimo elemento del vector de entrada se denota por $x_i(n)$.
- El k -ésimo elemento del vector completo de salida se denota por $o_k(n)$.
- El parámetro de la tasa de aprendizaje es denotado por η .
- El símbolo m_l denota el tamaño en la capa l del perceptrón multicapa: $l = 0, 1, \dots, L$. Donde L es la profundidad de la red. Así, m_0 denota el tamaño de la capa de entrada, m_1 denota el tamaño de la primer capa oculta y m_L denota el tamaño de la capa de salida.

3.11. Algoritmo Back-propagation

La señal de error en la salida de la neurona j en la iteración n está definida por

$$e_j(n) = d_j(n) - y_j(n), \text{ donde la neurona } j \text{ es de salida.} \quad (3.20)$$

Se define el error energético para la neurona j como $\frac{1}{2}e_j^2(n)$. Respectivamente, $\mathcal{E}(n)$ denota el error energético total y se define por

$$\mathcal{E}(n) = \frac{1}{2} \sum_{j \in C} e_j^2(n) \cdot \quad (3.21)$$

Donde C es el conjunto que incluye a todas las neuronas en la capa de salida de la red. Sea N el total de patrones contenidos en el conjunto de tratamiento. El promedio de los cuadrados de los errores energéticos se obtiene como

$$\mathcal{E}_{av} = \frac{1}{N} \sum_{n=1}^N \mathcal{E}(n) \cdot \quad (3.22)$$

Observación 3.1 *Obsérvese que $\mathcal{E}(n)$ y \mathcal{E}_{av} , dependen de los parámetros libres de la red.*

Para un conjunto de tratamientos dado, \mathcal{E}_{av} representa la función de costo como una medida del rendimiento de aprendizaje, donde el objetivo del proceso de aprendizaje es ajustar los parámetros libres de la red para minimizar \mathcal{E}_{av} .

Considérese la Figura 3.17, la cuál describe la neurona j siendo alimentada por un conjunto de señales de entrada producidas por una capa de neuronas a su izquierda. El campo local inducido $v_j(n)$ producido en la entrada de la función de activación asociada con la neurona j es

$$v_j(n) = \sum_{i=0}^m \omega_{ji}(n) y_i(n) \cdot \quad (3.23)$$

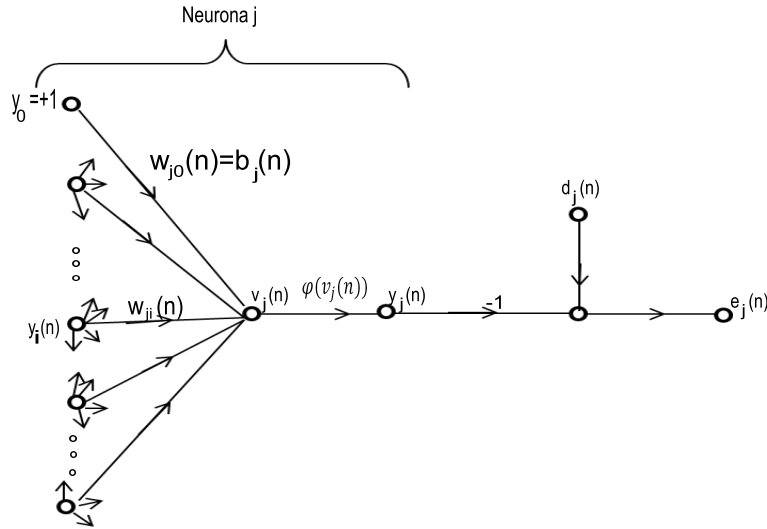


Figura 3.17: Representación gráfica de una neurona destacando los detalles de la neurona de salida j

Donde m es el total de entradas (excluyendo el sesgo) aplicadas a la neurona j . El peso sináptico ω_{j0} es igual al sesgo b_j aplicado a la neurona j . La señal de entrada $y_j(n)$ que aparece en la salida de la neurona j en la iteración n es

$$y_j(n) = \varphi(v_j(n)) \cdot \quad (3.24)$$

El algoritmo *back-propagation*, aplica corrección de $\Delta\omega_{ji}(n)$ al peso sináptico $\omega_{ji}(n)$, el cual es proporcional a $\frac{\partial \mathcal{E}(n)}{\partial \omega_{ji}(n)}$, es decir,

$$\begin{aligned} \frac{\partial \mathcal{E}(n)}{\partial \omega_{ji}(n)} &= \frac{\partial \mathcal{E}(n)}{\partial e_j(n)} \frac{\partial e_j(n)}{\partial y_j(n)} \frac{\partial y_j(n)}{\partial v_j(n)} \frac{\partial v_j(n)}{\partial \omega_{ji}(n)} \\ &= -e_j(n) \frac{\partial \varphi_j(v_j(n))}{\partial y_j(n)} y_i(n) \cdot \end{aligned} \quad (3.25)$$

Así, la corrección $\Delta\omega_{ji}(n)$ aplicado a ω_{ji} se define por la regla delta

$$\Delta\omega_{ji}(n) = -\eta \frac{\partial \mathcal{E}(n)}{\partial \omega_{ji}(n)} \cdot \quad (3.26)$$

donde η es el parámetro de la tasa de aprendizaje del algoritmo *back-propagation*.

La expresión (3.26) se puede representar como

$$\Delta\omega_{ji}(n) = \delta_j(n)y_i(n) \ , \quad (3.27)$$

donde $\delta_j(n)$ se define como

$$\begin{aligned} \delta_j(n) &= \frac{\partial \mathcal{E}(n)}{\partial v_j(n)} \\ &= \frac{\partial \mathcal{E}(n)}{\partial e_j(n)} \frac{\partial e_j(n)}{\partial y_j(n)} \frac{\partial y_j(n)}{\partial v_j(n)} \\ &= e_j(n) \frac{\partial \varphi_j(v_j(n))}{\partial v_j(n)} \ . \end{aligned} \quad (3.28)$$

De las ecuaciones (3.27) y (3.28) el factor clave para el cálculo del peso ajustado $\Delta\omega_{ji}(n)$ es la señal de error e_j en la salida de la neurona j . En este caso se pueden identificar los siguientes casos:

1. **La neurona j es un nodo de salida.** Cuando la neurona j está localizada en una capa de salida de la red, se suple con la respuesta deseada de esta neurona. Se puede utilizar la ecuación (3.20) para calcular la señal de error $e_j(n)$ asociada con esta neurona.
2. **La neurona j es un nodo oculto.** En este caso, la señal de error puede determinarse de manera recursiva en términos de las señales de error de todas las neuronas para las cuales la neurona oculta está conectada de manera directa. Para describir esto con más detalle, considérese la Figura 3.18 la cual describe una neurona j como un nodo oculto de la red.

De acuerdo a la ecuación (3.28), se debe redefinir $\delta_j(n)$ para la neurona oculta j como

$$\begin{aligned} \delta_j(n) &= - \frac{\partial \mathcal{E}(n)}{\partial y_j(n)} \frac{\partial y_j(n)}{\partial v_j(n)} \\ &= \frac{\partial \mathcal{E}(n)}{\partial y_j(n)} \varphi'_j(v_j(n)) \ . \end{aligned} \quad (3.29)$$

Para calcular la derivada parcial $\frac{\partial \mathcal{E}(n)}{\partial y_j(n)}$, se procede como sigue:

de la Figura 3.18 se observa

$$\frac{\partial \mathcal{E}(n)}{\partial y_j(n)} = \sum_k e_k(n) \frac{\partial e_k(n)}{\partial y_j(n)} \ , \quad (3.30)$$

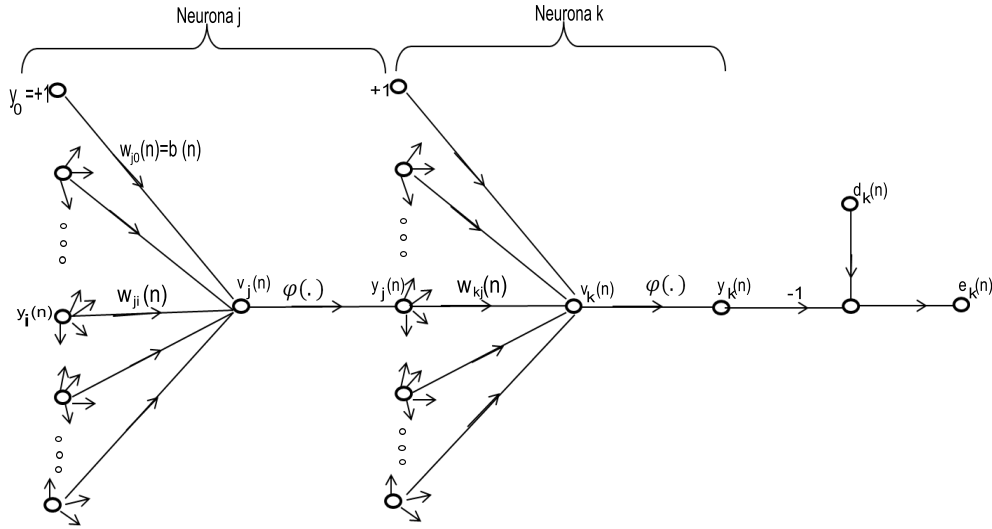


Figura 3.18: Representación gráfica de una neurona destacando los detalles de la neurona k conectada a la neurona oculta j

luego por regla de la cadena

$$\frac{\partial \mathcal{E}(n)}{\partial y_j(n)} = \sum_k e_k(n) \frac{\partial e_k(n)}{\partial v_k(n)} \frac{\partial v_k(n)}{\partial y_j(n)} , \quad (3.31)$$

de la Figura 3.18 se observa que

$$e_k(n) = d_k(n) - y_k(n) = d_k(n) - \varphi'_k(v_k(n)), \quad \text{con } k \text{ un nodo de salida.} \quad (3.32)$$

Por lo tanto,

$$\frac{\partial e_k(n)}{\partial v_k(n)} = -\varphi'(v_k(n)) , \quad (3.33)$$

con

$$v_k(n) = \sum_{j=0}^m \omega_{kj}(n) y_j(n) , \quad (3.34)$$

ya m el total de entradas (excluyendo el sesgo) aplicadas a la neurona k .

Observación 3.2 De nuevo, el peso sináptico $\omega_{k0}(n) = b_k(n)$.

Diferenciando la ecuación (3.34) se tiene

$$\frac{\partial v_k(n)}{\partial y_j(n)} = \omega_{kj}(n) , \quad (3.35)$$

al utilizar las ecuaciones (3.33) y (3.35) en la ecuación (3.31) se tiene

$$\begin{aligned} \frac{\partial \mathcal{E}(n)}{\partial y_j(n)} &= - \sum_k e_k(n) \varphi'_k(v_k(n)) \omega_{kj}(n) \\ &= - \sum_k \delta_k(n) \omega_{kj}(n) . \end{aligned} \quad (3.36)$$

Al utilizar la ecuación (3.36) en la ecuación (3.28) se tiene la expresión (3.37), la cual se conoce como la *fórmula back-propagation* para $\delta_j(n)$.

$$\delta_j(n) = \varphi'_j(v_j(n)) \sum_k \delta_k(n) \omega_{kj}(n), \quad \text{con } j \text{ una neurona oculta.} \quad (3.37)$$

Observación 3.3 Observe que $\varphi'_j(v_j(n))$ depende solamente de la función de activación asociada con la neurona oculta j . El resto de los factores involucrados en la ecuación (3.37), depende de los términos $\delta_k(n)$ (el cual requiere del conocimiento de las señales de error $e_k(n)$ de todas las neuronas que viven en la capa inmediata a la derecha de la neurona oculta j y que están directamente conectados con ella) y de $\omega_{kj}(n)$.

Dentro del algoritmo *back-propagation* se distinguen dos pasos de cálculo:

1. **Paso hacia adelante.** Aquí los pesos sinápticos y las señales de entrada de la red se calculan neurona por neurona. La señal de entrada que aparece en la salida de la neurona j se calcula como

$$y_j = \varphi(v_j(n)) , \quad (3.38)$$

donde $v_j(n)$ es el campo local inducido de la neurona j definida por

$$v_j(n) = \sum_{i=0}^m \omega_{ji}(n) y_i(n) , \quad (3.39)$$

con m el total de entradas (excluyendo el sesgo) aplicadas a la neurona j , $\omega_{ji}(n)$ los pesos sinápticos que conectan la neurona i con la neurona j y $y_j(n)$ la señal

que aparece en la salida de la neurona i .

Si j es la primer capa oculta de la red, $m = m_0$ y el índice i se refiere a la i -ésima entrada de la red. Si ésto ocurre, entonces

$$y_i(n) = x_i(n) , \quad (3.40)$$

donde $x_i(n)$ es el i -ésimo elemento del vector de entrada. Por otro lado, si la neurona j es una capa de salida de la red, $m = m_L$ y el índice j se refiere a la j -ésima salida terminal de la red, por lo tanto, se escribe

$$y_j(n) = o_j(n) , \quad (3.41)$$

donde $o_j(n)$ es el j -ésimo elemento del vector salida, esta salida se comparará con la respuesta deseada $d_j(n)$ para obtener el error $e_j(n)$ en la j -ésima neurona de salida.

2. **Paso hacia atrás.** Este paso empieza en la capa de salida haciendo pasar la señal de error hacia la izquierda a través de la red capa por capa y recursivamente calculando el valor de δ para cada neurona. Este proceso recursivo, permite a los pesos sinápticos de la red someterse a cambios de acuerdo con la regla delta de la ecuación (3.42).

$$\begin{pmatrix} \text{Corrección} \\ \text{de peso} \\ \Delta\omega_{ji}(n) \end{pmatrix} = \begin{pmatrix} \text{Parámetro} \\ \text{de la tasa} \\ \text{de aprendizaje} \\ \eta \end{pmatrix} \begin{pmatrix} \text{Gradiente} \\ \text{local} \\ \delta_j(n) \end{pmatrix} \begin{pmatrix} \text{Señal de} \\ \text{entrada de la} \\ \text{neurona } j \\ y_i(n) \end{pmatrix} \quad (3.42)$$

Para una neurona localizada en la capa de salida, δ es igual a la señal de error de esta neurona multiplicada por la primer derivada de la función no lineal. Se utiliza la ecuación (3.42), para calcular los cambios a los pesos de todas las conexiones de alimentación en la capa de salida. Dados los δ para las neuronas de la capa de salida, se utiliza la ecuación (3.41) para calcular los δ 's de todas las neuronas en la penúltima capa y por lo tanto, los cambios de los pesos de todas las conexiones de alimentación dentro de ésta. El cálculo recursivo se continúa capa por capa a través de la propagación de los cambios de todos los pesos sinápticos en la red.

El cálculo de δ para cada neurona de un perceptrón multicapa, requiere el conocimiento de la derivada de la función de activación $\varphi(\cdot)$. A continuación se presentan

dos funciones de activación no lineales continuamente diferenciables utilizadas frecuentemente en los perceptrones multicapa [12].

1. **Función logística.** En su forma más general se define como

$$y_j(n) = \varphi_j(v_j(n)) = \frac{1}{1 + \exp(-av_j(n))}, \quad a > 0, \quad y - \infty < v_j(n) < \infty, \quad (3.43)$$

donde $v_j(n)$ es el campo local inducido de la neurona j . Es sencillo demostrar que $0 < y_j(n) \leq 1$ y

$$\begin{aligned} \frac{\partial \varphi_j(v_j(n))}{\partial v_j(n)} &= \frac{a \exp(-av_j(n))}{[1 + \exp(-av_j(n))]^2} \\ &= ay_j(n)[1 - y_j(n)]. \end{aligned} \quad (3.44)$$

Para cada neurona j localizada en la capa de salida, $y_j(n) = o_j(n)$, por lo que se puede expresar a $\delta_j(n)$ como

$$\begin{aligned} \delta_j(n) &= e_j(n) \varphi'(v_j(n)) \\ &= a[d_j(n) - o_j(n)]o_j(n)[1 - o_j(n)], \quad \text{con } j \text{ un nodo de salida.} \end{aligned} \quad (3.45)$$

Donde $o_j(n)$ es la señal de entrada en la neurona de salida j y $d_j(n)$ es la respuesta deseada para j . Para una neurona oculta arbitraria, $\delta_j(n)$ se expresa como

$$\begin{aligned} \delta_j(n) &= \varphi'(v_j(n)) \sum_k \delta_k(n) \omega_{kj}(n) \\ &= ay_j(n)[1 - y_j(n)] \sum_k \delta_k(n) \omega_{kj}(n), \quad \text{para } j \text{ una neurona oculta.} \end{aligned} \quad (3.46)$$

2. **Función tangente hiperbólica.** Esta función es de la forma

$$y_j(n) = \varphi_j(n) = \operatorname{atanh}(bv_j(n)), \quad (3.47)$$

con a y b constantes positivas. Es sencillo verificar que

$$\varphi'(v_j(n)) = \frac{b}{a} [a - y_j(n)][a + y_j(n)]. \quad (3.48)$$

Luego, para una neurona j localizada en la capa de salida se puede expresar a $\delta_j(n)$ como

$$\begin{aligned}\delta_j(n) &= e_j(n)\varphi'(v_j(n)) \\ &= \frac{b}{a}[d_j(n) - o_j(n)][a - o_j(n)][a + o_j(n)] \cdot\end{aligned}\quad (3.49)$$

Para una neurona j en una capa oculta, se tiene que

$$\begin{aligned}\delta_j(n) &= \varphi'(v_j(n)) \sum_k \delta_k(n)\omega_{kj}(n) \\ &= \frac{b}{a}[a - y_j(n)][a + y_j(n)] \sum_k \delta_k(n)\omega_{kj}(n), \quad j \text{ es una neurona oculta.}\end{aligned}\quad (3.50)$$

A continuación se presenta el algoritmo *back-propagation* de manera sintetizada.

-
1. Inicializar los pesos y umbrales con valores pequeños.
 2. Presentar las entradas $X_p = \{x_1, x_2, \dots, x_n\}$ y las salidas $T_p = t_0, t_1, \dots, t_{m-1}$, donde n es el número de nodos de entrada y m el número de nodos de salida. Elegir $w_0 = -\theta$ y $x_0 = 1$.
 3. Calcular la salida actual
Para cada capa, calcular:
$$y_{pj} = f \left[\sum_{i=0}^{n-1} \omega_i x_i \right]$$
y pasar y_{pj} como entrada a la siguiente capa.
 4. Adaptar los pesos.
Empezar con la capa de salida y trabajar hacia atrás
$$\omega_{ij}(t+1) = \omega_{ij}(t) + \eta \delta_{pj} o_{pj}$$
donde $\omega_{ij}(t)$ representa los pesos del nodo i al nodo j en el tiempo t , η es la tasa de aprendizaje y δ_{pj} es un término de error del patrón p en el nodo j .
Para unidades de salida
$$\delta_{pj} = k o_{pj} (1 - o_{pj}) (t_{pj} - o_{pj})$$
Para unidades ocultas
$$\delta_{pj} = k o_{pj} (1 - o_{pj}) \sum_k \delta_{pk} \omega_{jk}$$
donde la suma es sobre los k nodos de la capa de arriba del nodo j
-

3.12. Algoritmo de Levenberg-Maquardt

El algoritmo de Levenberg-Marquardt es una alternativa popular para el método de Gauss-Newton para encontrar el mínimo de una función $f(x)$ que es una suma de cuadrados no lineales, es decir, $f(x) = \frac{1}{2} \sum_{j=1}^n [f(x_j)]^2$. El algoritmo tiene un excelente desempeño en el entrenamiento de redes neuronales donde el rendimiento de la red está determinado por el error cuadrático medio [12].

El método de Newton para optimizar el rendimiento $e(x)$ es

$$\mathbf{x}_{k+1} = \mathbf{x}_k - \mathbf{A}^{-1} \mathbf{g}_k , \quad (3.51)$$

donde $\mathbf{A} \equiv \nabla^2 \mathbf{e}(\mathbf{x})|_{\mathbf{x}=\mathbf{x}_k}$ y $\mathbf{g}_k \equiv \nabla \mathbf{e}(\mathbf{x})|_{\mathbf{x}=\mathbf{x}_k}$.

Suponiendo que $\mathbf{e}(\mathbf{x})$ es una suma de funciones cuadráticas

$$\mathbf{e}(\mathbf{x}) = \sum_{i=1}^n v_i^2 = \mathbf{v}(\mathbf{x})' \mathbf{v}(\mathbf{x}) , \quad (3.52)$$

luego el j -ésimo elemento del gradiente se expresa como

$$\nabla [\mathbf{e}(\mathbf{x})_j] = \frac{\partial \mathbf{e}(\mathbf{x})}{\partial x_j} = 2 \sum_{i=1}^n v_i(\mathbf{x}) \frac{\partial v_i(\mathbf{x})}{\partial x_j} , \quad (3.53)$$

por lo que el gradiente puede ser escrito como

$$\nabla \mathbf{e}(\mathbf{x}) = 2 \mathbf{J}'(\mathbf{x}) \mathbf{v}(\mathbf{x}) , \quad (3.54)$$

con

$$\mathbf{J}(\mathbf{x}) = \begin{bmatrix} \frac{\partial v_1(x)}{\partial x_1} & \frac{\partial v_1(x)}{\partial x_2} & \dots & \frac{\partial v_1(x)}{\partial x_n} \\ \frac{\partial v_2(x)}{\partial x_1} & \frac{\partial v_2(x)}{\partial x_2} & \dots & \frac{\partial v_2(x)}{\partial x_n} \\ \vdots & \vdots & \dots & \vdots \\ \frac{\partial v_n(x)}{\partial x_1} & \frac{\partial v_n(x)}{\partial x_2} & \dots & \frac{\partial v_n(x)}{\partial x_n} \end{bmatrix} , \quad (3.55)$$

la matriz Jacobiana.

Ajustando el método de Newton, se obtiene el algoritmo de Levenberg-Maquardt, para ésto se toma

$$\mathbf{x}_{k+1} = \mathbf{x}_k - [\mathbf{J}'(\mathbf{x}_k) \mathbf{J}(\mathbf{x}_k) + \mu_k \mathbf{I}]^{-1} \mathbf{J}'(\mathbf{x}_k) \mathbf{v}(\mathbf{x}_k) , \quad (3.56)$$

o bien,

$$\Delta \mathbf{x}_k = - [\mathbf{J}'(\mathbf{x}_k)\mathbf{J}(\mathbf{x}_k) + \mu_k \mathbf{I}]^{-1} \mathbf{J}'(\mathbf{x}_k)\mathbf{v}(\mathbf{x}_k) . \quad (3.57)$$

La nueva constante μ_k determina la tendencia del algoritmo, cuando μ_k se incrementa, este algoritmo se aproxima al algoritmo de pasos descendientes para factor de aprendizaje pequeño; cuando μ_k disminuye, este algoritmo se convierte en el método de Gauss - Newton.

El algoritmo comienza con un valor pequeño para μ_k (por lo general 0.01), si no se alcanza el valor para $\mathbf{e}(x)$ entonces el paso se repite con μ_k multiplicado por un factor $\vartheta > 1$. Si se ha escogido un valor pequeño de paso en la dirección de paso descendiente, $\mathbf{e}(x)$ debería decrecer. Si un paso produce un pequeño valor para $\mathbf{e}(x)$, entonces el algoritmo tiende al método de Gauss - Newton, el cual garantiza una rápida convergencia.

Lo principal del algoritmo de Levenberg-Marquardt es el cálculo de la matriz Jacobiana, pero en el algoritmo *back-propagation* se calculan las derivadas de los errores al cuadrado con respecto a los pesos y ganancias de la red, por lo que para el cálculo de la matriz Jacobiana se realiza el cálculo de las derivadas de los errores en lugar de las derivadas de los errores al cuadrado.

De lo anterior, sea el vector de error \mathbf{v}' definido como

$$\mathbf{v} = (v_1, v_2, \dots, v_N) = (e_{1,1}, e_{2,1}, \dots, e_{S^M,1}, e_{1,2}, \dots, e_{s^M,Q}) , \quad (3.58)$$

y, el vector de parámetros

$$\begin{aligned} \mathbf{x}' &= (x_1, x_2, \dots, x_n) \\ &= (\omega_{1,1}^1, \omega_{1,2}^1, \dots, \omega_{S^1,R}^1, b^1, \dots, b_{s^1}^1, \omega_{1,1}^2, \omega_{1,2}^2, \dots, b_{S^M}^M) , \end{aligned} \quad (3.59)$$

donde $N = Q \times S^M$ y $n = S^1(R+1) + S^2R(S^1+1) + \dots + S^M(S^{M-1}+1)$.

Sustituyendo en (3.60), la matriz Jacobiana para el entrenamiento de un perceptrón multicapa queda expresado por

$$\mathbf{J}(\mathbf{x}) = \begin{bmatrix} \frac{\partial e_{1,1}}{\partial \omega_{1,1}^1} & \frac{\partial e_{1,1}}{\partial \omega_{1,2}^1} & \dots & \frac{\partial e_{1,1}}{\partial \omega_{S^1,R}^1} & \frac{\partial e_{1,1}}{\partial b_1^1} & \dots \\ \frac{\partial e_{2,1}}{\partial \omega_{1,1}^1} & \frac{\partial e_{2,1}}{\partial \omega_{1,2}^1} & \dots & \frac{\partial e_{2,1}}{\partial \omega_{S^1,R}^1} & \frac{\partial e_{2,1}}{\partial b_1^1} & \dots \\ \vdots & \vdots & \dots & \vdots & \vdots & \dots \\ \frac{\partial e_{S^M,1}}{\partial \omega_{1,1}^1} & \frac{\partial e_{S^M,1}}{\partial \omega_{1,2}^1} & \dots & \frac{\partial e_{S^M,1}}{\partial \omega_{S^1,R}^1} & \frac{\partial e_{2,1}}{\partial b_1^1} & \dots \\ \frac{\partial e_{1,2}}{\partial \omega_{1,1}^1} & \frac{\partial e_{1,2}}{\partial \omega_{1,2}^1} & \dots & \frac{\partial e_{1,2}}{\partial \omega_{S^1,R}^1} & \frac{\partial e_{1,2}}{\partial b_1^1} & \dots \\ \vdots & \vdots & \dots & \vdots & \vdots & \dots \end{bmatrix}, \quad (3.60)$$

cada elemento del Jacobiano se puede calcular como

$$[\mathbf{J}]_{h,l} = \frac{\partial v_h}{\partial x_l} = \frac{e_{k,q}}{\partial x_l}, \quad (3.61)$$

luego haciendo uso de la regla de la cadena se tiene

$$\frac{\partial \hat{e}}{\partial \omega_{i,j}^m} = \frac{\partial \hat{e}}{\partial n_{n_i}^m} \times \frac{\partial n_{n_i}^m}{\partial \omega_{i,j}^m}, \quad (3.62)$$

donde $s_i^m = \frac{\partial \hat{e}}{\partial \omega_{i,j}^m}$ corresponde a la sensibilidad de la red.

Se define la sensibilidad del algoritmo de Levenberg-Maquardt como

$$\hat{s}_{i,h}^m \equiv \frac{\partial v_h}{\partial n_{i,q}^m}, \quad (3.63)$$

donde $h = (q-1)S^m + k$, por lo tanto, los elementos de Jacobiano se pueden calcular como

$$[\mathbf{J}]_{h,i} = \frac{\partial e_{k,q}}{\partial \omega_{i,j}^m} = \frac{\partial e_{k,q}}{\partial n_{i,q}^m} \times \frac{\partial n_{i,q}^m}{\partial \omega_{i,j}^m} = \hat{s}_{i,h}^m. \quad (3.64)$$

Para la sensibilidad del algoritmo de la última capa se tiene

$$\hat{s}_{i,h}^M = \frac{\partial v_h}{\partial n_{i,q}^M} = \frac{\partial e_{k,q}}{\partial n_{i,q}^M}$$

$$= \begin{cases} -f^M(n_{i,q}^M) , & \text{para } i = k , \\ 0 , & \text{para } i \neq k . \end{cases} \quad (3.65)$$

$$(3.66)$$

Para cada nueva entrada presentada en la red, los vectores de sensibilidad son propagados hacia atrás, ésto se debe a que se ha calculado cada error en forma individual, en lugar de derivar la suma al cuadrado de los errores. Para cada entrada aplicada a la red habrá S^M errores, uno por cada elemento de salida de la red y por cada error se generará una fila de la matriz Jacobiana.

El algoritmo de Levenberg-Maquardt puede resumirse en los siguientes pasos:

1. Se presentan todas las entradas a la red, se calculan las correspondientes salidas y cada uno de los errores según $\mathbf{e}_q = \mathbf{t}_q - \mathbf{a}_q^M$, después se calcula la suma de los errores cuadrados para cada entrada $\mathbf{e}(\mathbf{x})$.
2. Calcular las sensibilidades individuales y la matriz sensibilidad total y con éstas, se calculan los elementos de la matriz Jacobiana.
3. Obtener Δx_k .
4. Se recalcula la suma de los errores cuadrados usando $x_k + \Delta x_k$.
Si esta nueva suma es más pequeña que el valor calculado en el paso 1 entonces se divide μ por ϑ , se calcula $x_{k+1} = x_k + \Delta x_k$ y regresar al paso 1.
Si la suma no se reduce entonces se multiplica μ por ϑ y se regresa al paso 3.

CAPÍTULO 4

Caso de estudio

El Estado de Tlaxcala se localiza geográficamente en la región centro-oriental de la República Mexicana entre los $97^{\circ}37'07''$ y los $98^{\circ}42'51''$ de longitud oeste y los $19^{\circ}05'43''$ y los $19^{\circ}44'07''$ de latitud norte situado en las tierras altas del eje neovolcánico, sobre la meseta de Anáhuac, colinda al norte con los estados de Hidalgo y Puebla; al este y sur con el estado de Puebla; al oeste con los estados de Puebla, México e Hidalgo. Su extensión territorial es de 4060.93 km^2 , lo que representa el 0.2% del territorio nacional, siendo el estado de la Federación con menor superficie. Su altitud media es de 2230 m sobre el nivel del mar, con clima templado-subhúmedo, semifrío- subhúmedo y frío.

En este trabajo de investigación se cuenta con información de 5 estaciones meteorológicas no automáticas del estado de Tlaxcala, las cuales se encuentran a cargo del Sistema Meteorológico Nacional (SMN) a través de la Comisión Nacional del Agua (CONAGUA) y sólo registran datos numéricos sobre temperatura máxima, temperatura mínima y precipitación.

Para el ajuste de modelos por medio de la metodología de Box y Jenkins, se dispone de información diaria de temperatura máxima, temperatura mínima y precipitación de 5 estaciones meteorológicas ubicadas en los municipios de Atlangatepec, Huamantla, Tlaxco, Españita y el Carmen Teuexquitla (Figura4.1).

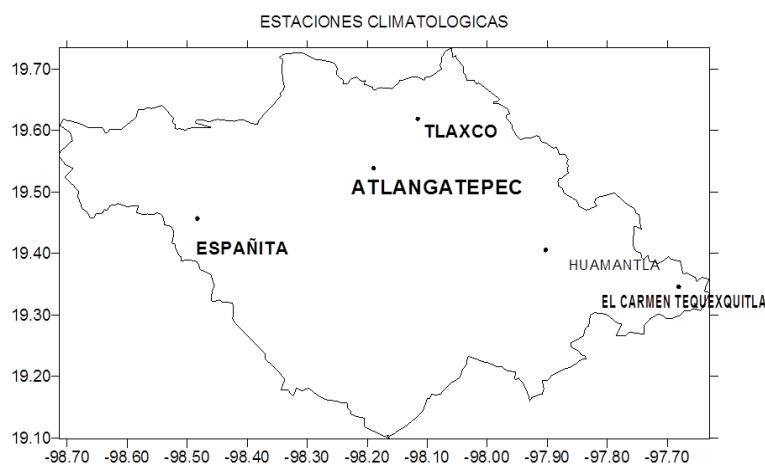


Figura 4.1: Ubicación geográfica de las 5 estaciones meteorológicas estudiadas

Estación	Periodo
Atlangatepec	1/1/1961 al 31/12/2012
Huamantla	1/1/1990 al 31/12/2012
Tlaxco	1/1/1989 al 31/12/2012
Españita	4/3/1989 al 31/12/2012
El Carmen Tequexquitla	1/1/1992 al 31/12/2012

Tabla 4.1: Periodos de información para 5 estaciones del estado de Tlaxcala

4.1. Análisis de los datos utilizando la metodología de Box-Jenkins

En la Tabla 4.1 se presentan los periodos de información obtenidos para cada una de las estaciones. Los datos analizados presentaron problemas de calidad en relación a datos faltantes, éstos fueron completados con el promedio diario de 10 años de registro del mismo mes y del mismo día, cabe mencionar que el procedimiento para la estimación de datos faltantes es heurístico y que existen otros procedimientos óptimos y/o software para la estimación de datos faltantes (ver [10], [14]). Para un mejor tratamiento de la información, una vez estimados los datos faltantes se decidió trabajar con el promedio semanal y con la finalidad de realizar pronóstico de datos, se omitieron los últimos 6 datos de las nuevas bases de datos.

Para identificar ciclos estacionales y obtener información sobre observaciones extremas, se graficaron los diagramas de caja de las variables estudiadas para cada una de las estaciones meteorológicas (Figura 4.2). De la Figura 4.2, se observa que en

todas las estaciones estudiadas, la temperatura máxima tiende a tener un mayor aumento durante los meses de marzo a junio. La temperatura mínima tiende a bajar más durante los meses de noviembre a febrero salvo la temperatura mínima para Españita, Tlaxcala, en la cual la temperatura mínima tiene un comportamiento muy similar en los meses de febrero a noviembre y se tiene una disminución de la temperatura bastante notoria para los meses de diciembre y enero. Los meses con mayor precipitación, comprenden de junio a septiembre y en municipios como Huamantla y Tlaxco también se registra precipitación hasta el mes de octubre.

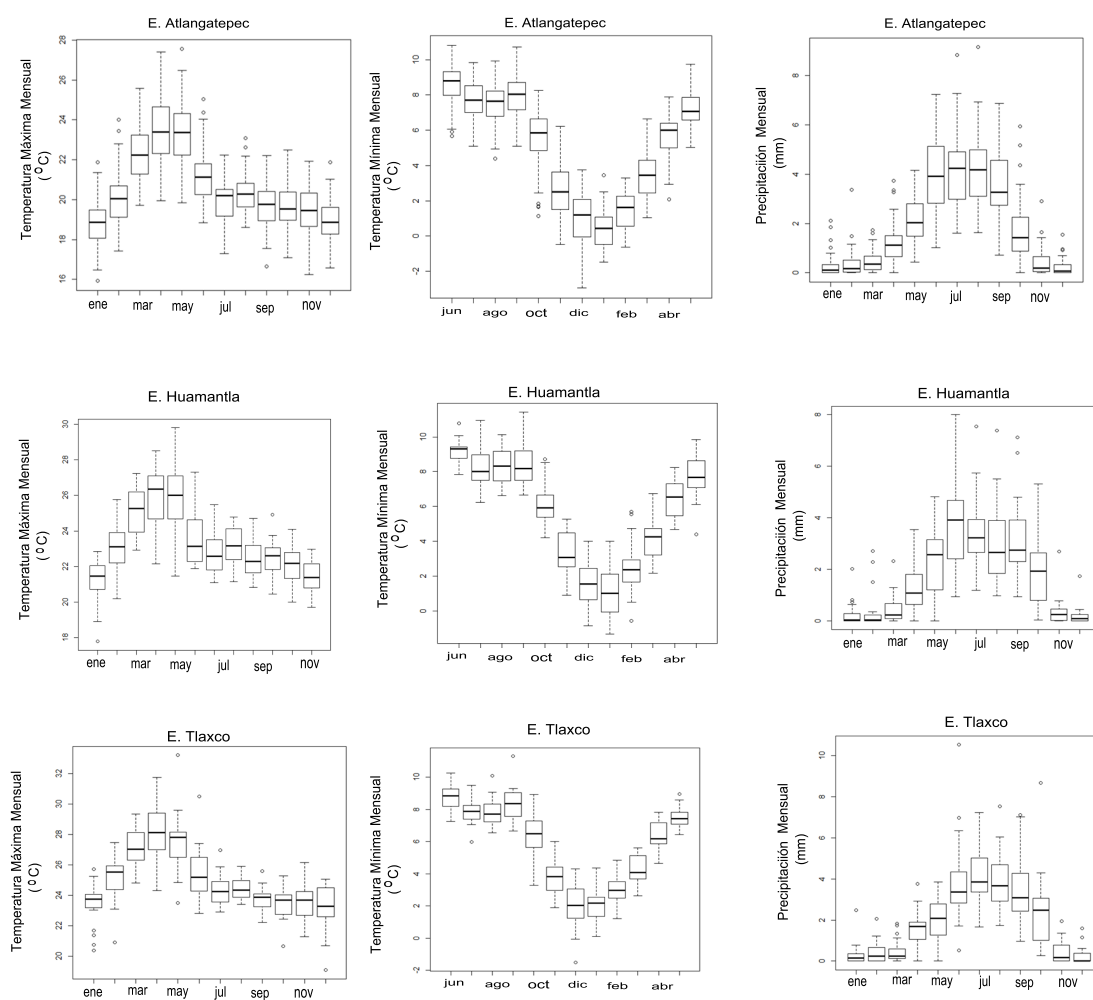


Figura 4.2: Diagramas de caja para la temperatura máxima, mínima y precipitación de Atlangatepec, Huamatla y Tlaxco

Para observar el comportamiento de los datos se graficaron las series de tiempo co-

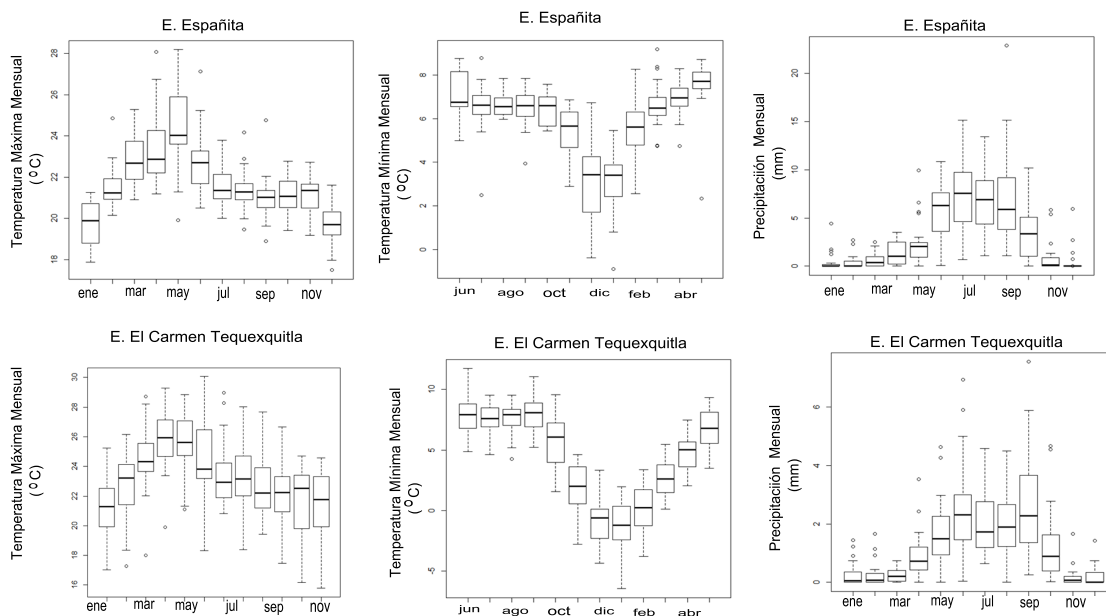


Figura 4.3: Diagramas de caja para la temperatura máxima, mínima y precipitación de España y El Carmen Tequexquitla

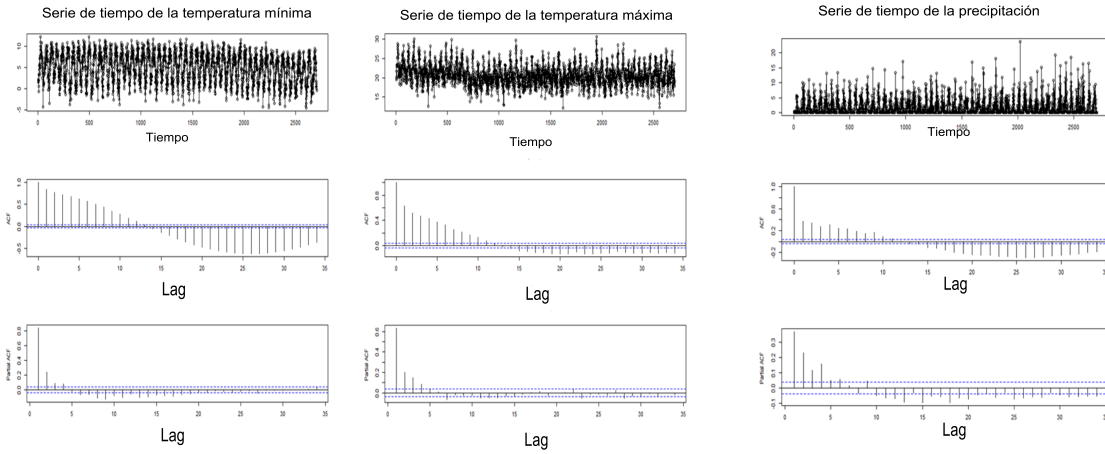
respondientes a la temperatura máxima, temperatura mínima y precipitación, además de graficar su función de autocorrelación y su función de autocorrelación parcial con la finalidad de identificar estacionaridad de las series de datos (Figuras 4.4 y 4.5).

De las Figuras 4.4, 4.5 y 4.6, se tienen las siguientes observaciones:

Para la temperatura mínima de las estaciones de Atlangatepec, Huamantla, Tlaxco, España y El Carmen Tequexquitla, la función de autocorrelación correspondiente decrece lentamente en forma de senos y cosenos y, su función de autocorrelación parcial se trunca rápidamente, por esta razón no se puede concluir la estacionaridad de las series de tiempo y se debe realizar una transformación de los datos para transformar cada una de las series de esta variable climatológica, en series de tiempo estacionarias. Para la temperatura máxima de las estaciones de Atlangatepec, Huamantla, Tlaxco y España la función de autocorrelación correspondiente decrece lentamente en forma de senos y cosenos y, su función de autocorrelación parcial se truncan rápidamente, por esta razón no se puede concluir la estacionaridad de las series de tiempo y se debe realizar una transformación de los datos para transformar cada una de las series, en series de tiempo estacionarias.

Para la serie de temperatura máxima de la estación de El Carmen Tequexquitla, la función de autocorrelación, decrece lentamente en forma exponencial y su función de

E. Atlangatepec, Tlaxcala



E. Huamantla, Tlaxcala

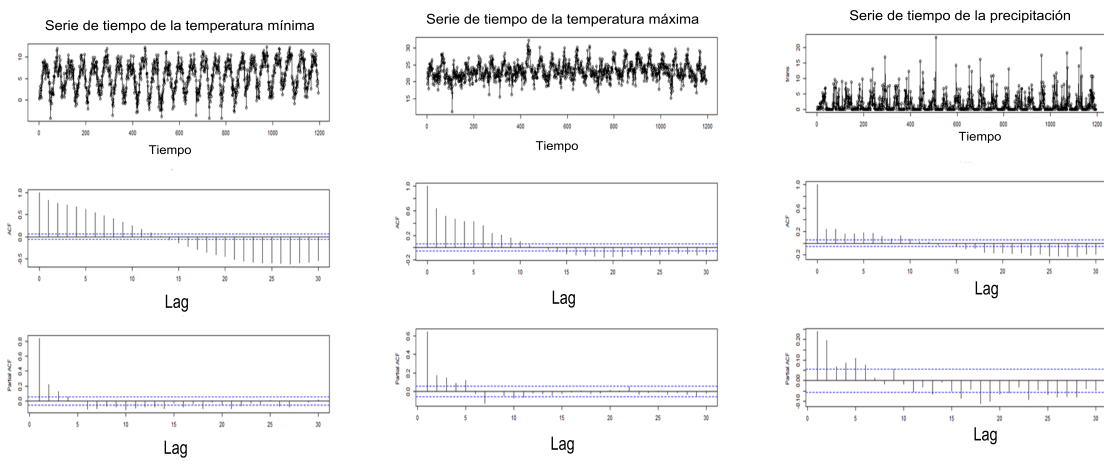
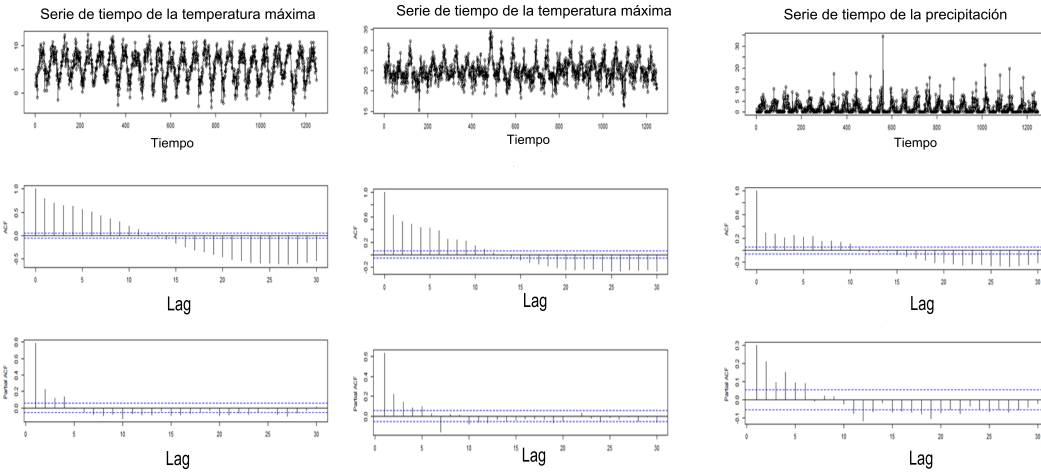


Figura 4.4: Series de tiempo y funciones de autocorrelación y autocorrelación parcial para las estaciones meteorológicas de Atlangatepec y Huamantla

E. Tlaxco, Tlaxcala



E. España, Tlaxcala

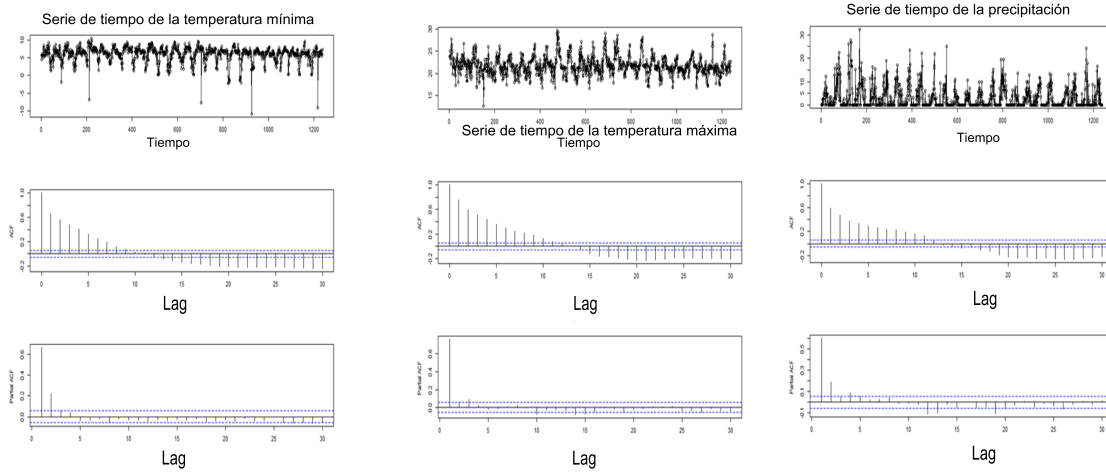


Figura 4.5: Series de tiempo y funciones de autocorrelación y autocorrelación parcial para las estaciones meteorológicas de Tlaxco, España

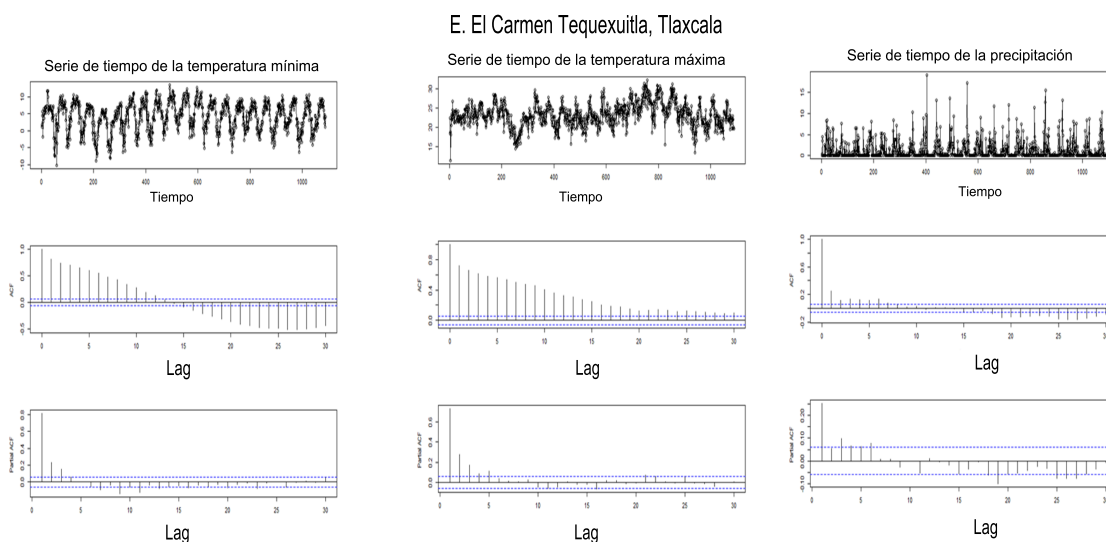


Figura 4.6: Series de tiempo y funciones de autocorrelación y autocorrelación parcial para la estación meteorológica El Carmen Tequexuitla

autocorrelación parcial se trunca de manera rápida por lo que tampoco es posible concluir la estacionaridad de la serie y se debe realizar una transformación de los datos para garantizar la estacionaridad de la serie.

Para la serie de datos sobre precipitación, la función de autocorrelación decrece de manera lenta para las estaciones de Atlangatepec, Huamantla, Tlaxco, Españita y El Carmen Tequexuitla y, la función de autocorrelación parcial decrece de manera lenta para las estaciones de Atlangatepec, Huamantla, Tlaxco y El Carmen Tequexuitla y sólo para la serie de datos de la estación de Españita, Tlaxcala, la función de autocorrelación parcial se trunca de manera rápida. Dado el comportamiento de las funciones de autocorrelación y autocorrelación parcial para la serie de datos de las variables estudiadas no se concluye la estacionaridad de los datos.

En las siguientes secciones se trabaja con más detalle con cada una de las series de tiempo mostrando la transformación de los datos realizada, el modelo propuesto que se ajusta a los datos y el pronóstico realizado. Para la información de cada una de las estaciones se probaron distintos modelos de series de tiempo y sólo se muestra el mejor modelo elegido bajo el Criterio de Información de Akaike [15].

4.1.1. Estación de Atlangatepec, Tlaxcala

El municipio de Atlangatepec, se ubica en el Altiplano central mexicano a 2500 *m* sobre el nivel del mar, se sitúa en un eje de coordenadas geográficas entre los $19^{\circ}31'51''$

latitud norte y $98^{\circ} 12' 30''$ longitud oeste. En el municipio prevalece el clima templado subhúmedo con lluvias en verano.

Temperatura mínima

Para la serie de datos de temperatura mínima de la estación ubicada en Atlangatepec, Tlaxcala se realizó la primer diferencia de los datos obteniéndose una nueva serie cuyo comportamiento se muestra en la Figura 4.7.

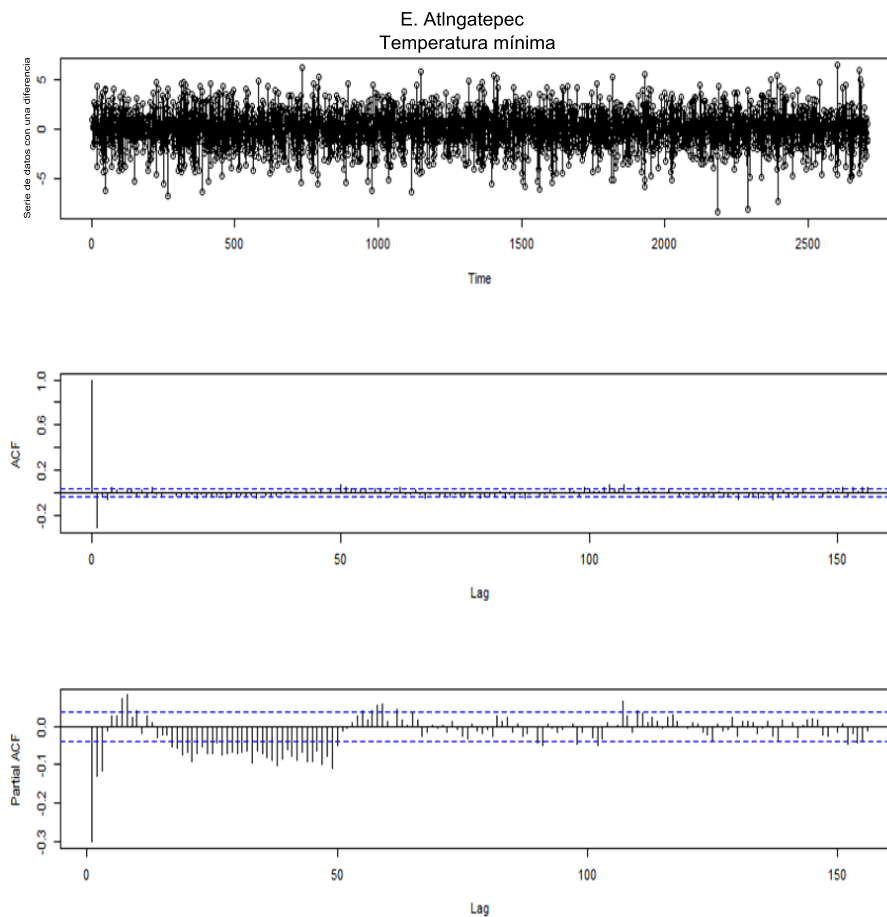


Figura 4.7: Serie de tiempo transformada, función de autocorrelación y función de autocorrelación parcial para los datos de temperatura mínima de Atlangatepec, Tlaxcala

De la Figura 4.7, se observa que la función de autocorrelación para la serie de datos transformada se extingue de manera rápida y su función de autocorrelación parcial se extingue en forma exponencial por lo que se propone un modelo no estacional

$ARIMA(1, 1, 2)$.

Para identificar un modelo estacional en la serie mostrada en la Figura 4.7, se realizó una diferencia de un periodo ($L = 52$), se graficó la serie que resulta de la transformación, además de la función de autocorrelación y la función de autocorrelación parcial de forma que del análisis del comportamiento de estas dos últimas funciones se determina o identifica un posible modelo estacionario para la parte estacional, ver Figura 4.8.

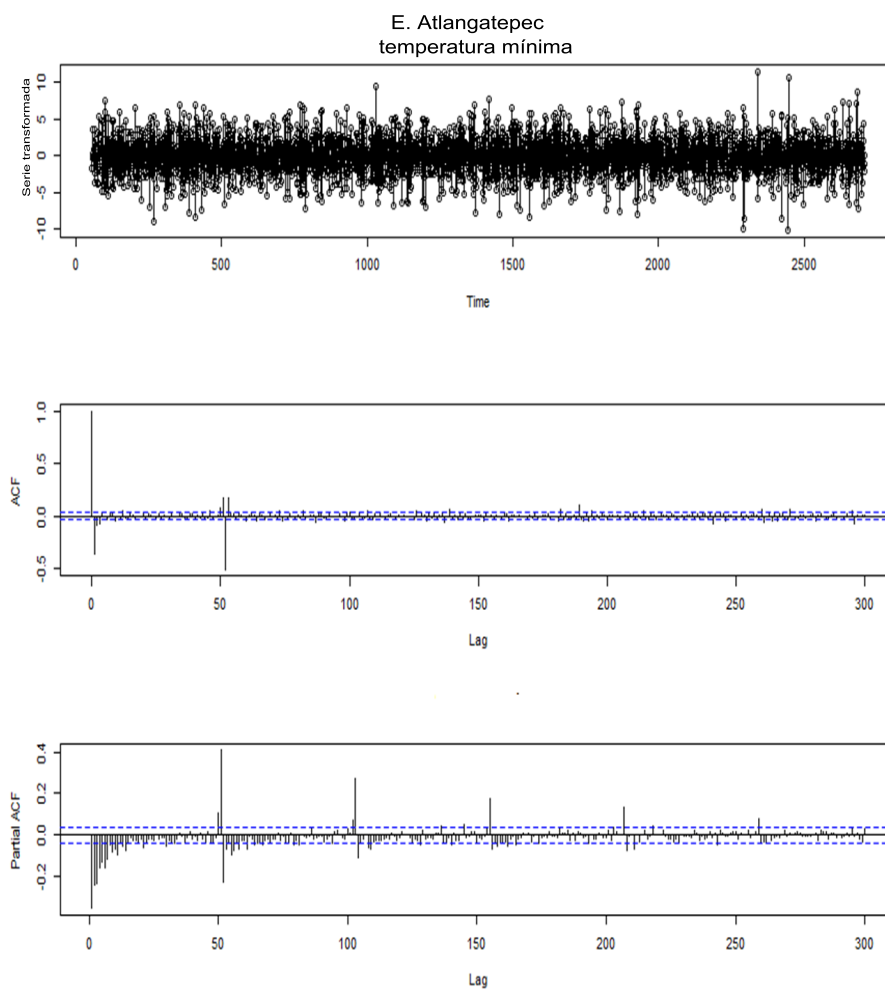


Figura 4.8: Serie de tiempo, función de autocorrelación y función de autocorrelación parcial para los datos de temperatura mínima de Atlangatepec, Tlaxcala, con una diferencia en el nivel no estacional y una diferencia en el nivel estacional

De la Figura 4.8, se observa que la función de autocorrelación se extingue de manera rápida y su función de autocorrelación parcial se extingue en forma exponencial para el modelo estacional, por lo tanto, el modelo estacionario que integra el comportamiento estacional se propone como $SARIMA(1, 1, 2)(1, 1, 3)$.

Al combinar los modelos propuestos para modelar la temperatura mínima de Atlangatepec, Tlaxcala se tiene

$$y_t^* = y_t - y_{t-1} - y_{t-52} + y_{t-53} \quad , \quad (4.1)$$

donde y_t representa una observación de la temperatura mínima en el tiempo t . Así, para el nivel no estacional se propone el modelo

$$y_t^* = \phi_1 y_{t-1}^* + z_t + \theta_1 z_{t-1} + \theta_2 z_{t-2} \quad . \quad (4.2)$$

Para el nivel estacional se propone

$$y_t^* = \phi_{1,52} y_{t-1}^* - z_t - \theta_{1,52} z_{t-1} - \theta_{2,52} z_{t-2} - \theta_{3,52} z_{t-3} \quad . \quad (4.3)$$

Finalmente, el modelo integrado es

$$\begin{aligned} y_t^* &= \phi_1 y_{t-1}^* + z_t + \theta_1 z_{t-1} + \theta_2 z_{t-2} + \phi_{1,52} y_{t-52}^* \\ &\quad - \theta_{1,52} z_t - \theta_{2,52} z_{t-104} - \theta_{3,52} z_{t-156} + \phi_1 \phi_{1,52} y_{t-53}^* \\ &\quad - \theta_1 \theta_{1,52} z_{t-53} - \theta_1 \theta_{2,52} z_{t-105} - \theta_1 \theta_{3,52} z_{t-157} \\ &\quad - \theta_2 \theta_{1,52} z_{t-54} - \theta_2 \theta_{2,52} z_{t-106} - \theta_2 \theta_{3,52} z_{t-108} . \end{aligned} \quad (4.4)$$

Para estimar los parámetros por máxima verosimilitud en el modelo (4.4) se utilizó el software R . Para verificar qué modelo SARIMA se ajustó mejor a los datos fue necesario observar el Criterio de Información de Akaike y los p -valores de la estadística de Ljung-Box [15].

	ar1	ma1	ma2	sar1	sma1	sma2	sma3	intercepto
	0.8001	-1.5978	0.6583	0.8001	-0.6835	-0.2049	-0.1116	-0.0001
e. s	0.2081	0.0815	0.0573	0.2082	0.0392	0.0249	0.0309	0.0001

σ^2 estimado: 4.66, logaritmo de verosimilitud = -5813.59, AIC = 11645.19.

Tabla 4.2: Estimación de los parámetros para el modelo (4.4)

Al sustituir los valores estimados de los parámetros (Tabla 4.2), el modelo propuesto para pronosticar la temperatura mínima de Atlangatepec, Tlaxcala queda expresado de la siguiente manera:

$$\begin{aligned}
y_t^* = & 0.8001y_{t-1}^* + z_t - 1.5978z_{t-1} + 0.6583z_{t-2} + 0.8001y_{t-52}^* \\
& + 0.6835z_{t-52} + 0.2049z_{t-104} + 0.1116z_{t-156} + 0.64016y_{t-53}^* + 1.092z_{t-53} \\
& + 0.3273z_{t-105} - 0.1783z_{t-157} - 0.4499z_{t-54} - 0.1348z_{t-106} - 0.0734z_{t-108}.
\end{aligned}
\tag{4.5}$$

En la Figura 4.9 se muestran los resultados de la prueba de Ljung-Box aplicados a los datos y, en la Tabla 4.3 se presenta el valor observado y los pronósticos del modelo estacional $SARIMA(1, 1, 2)(1, 1, 3)$ para los datos de la temperatura mínima para Atlangatepec, Tlaxcala.

Valor observado	Valor pronosticado
0.57143	1.7215190
-0.42857	0.6916233
-0.71429	-2.9139462
-2.6429	-1.3504657
0.71429	-0.9344342
0	-0.8066220

Tabla 4.3: Valor observado y pronosticado para la temperatura mínima de Atlangatepec, Tlaxcala

La Tabla 4.4 muestra algunos modelos probados para los datos.

Modelo	AIC
$SARIMA(1,1,3)(0,1,3)$	11645.55
$SARIMA(1,1,3)(1,1,3)$	11646.6
$SARIMA(1,1,2)(1,1,3)$	11645.19

Tabla 4.4: Modelos que fueron propuestos para la temperatura mínima de Atlangatepec, Tlaxcala

Temperatura máxima

Para la serie de datos de temperatura máxima de la estación ubicada en Atlangatepec, Tlaxcala se obtuvo la primer diferencia, obteniéndose una nueva serie cuyo comportamiento se muestra en la Figura 4.10.

De la Figura 4.10, se observa que la función de autocorrelación para la serie de datos transformada se extingue de manera rápida y su función de autocorrelación parcial

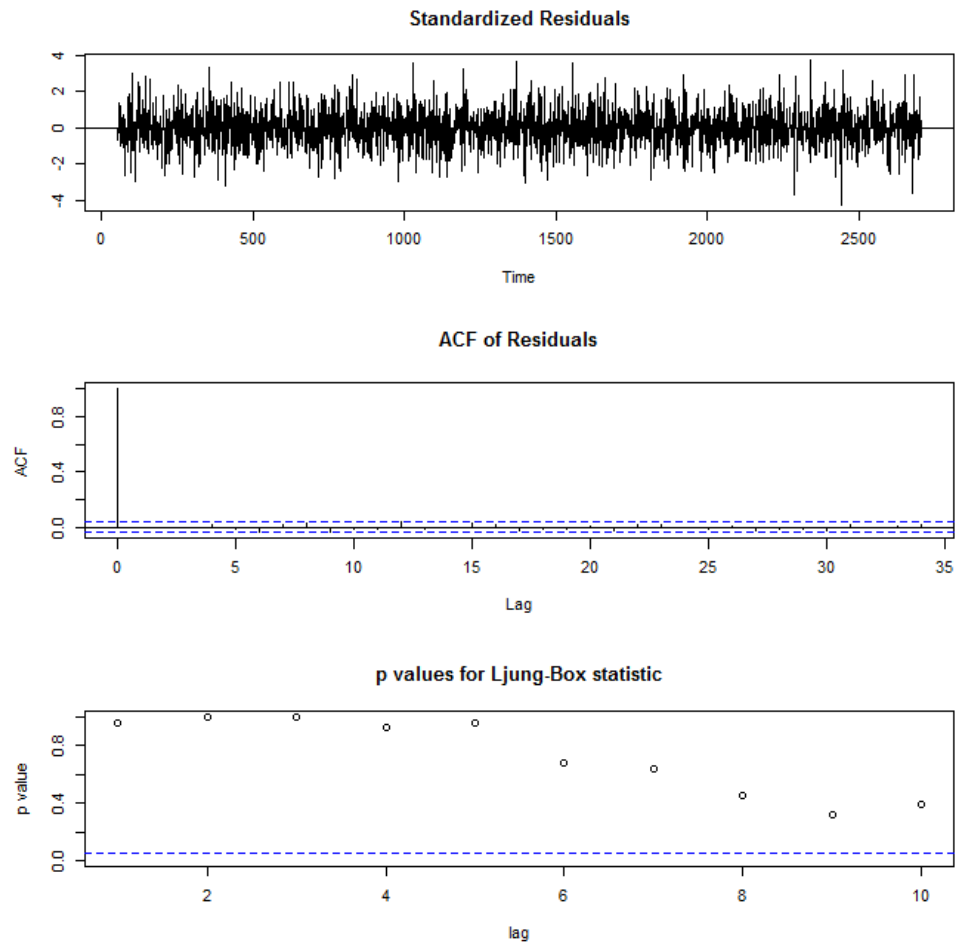


Figura 4.9: Resultados de la prueba de Ljung-Box, para los datos de la serie de tiempo de la temperatura mínima de Atlangatepec, Tlaxcala

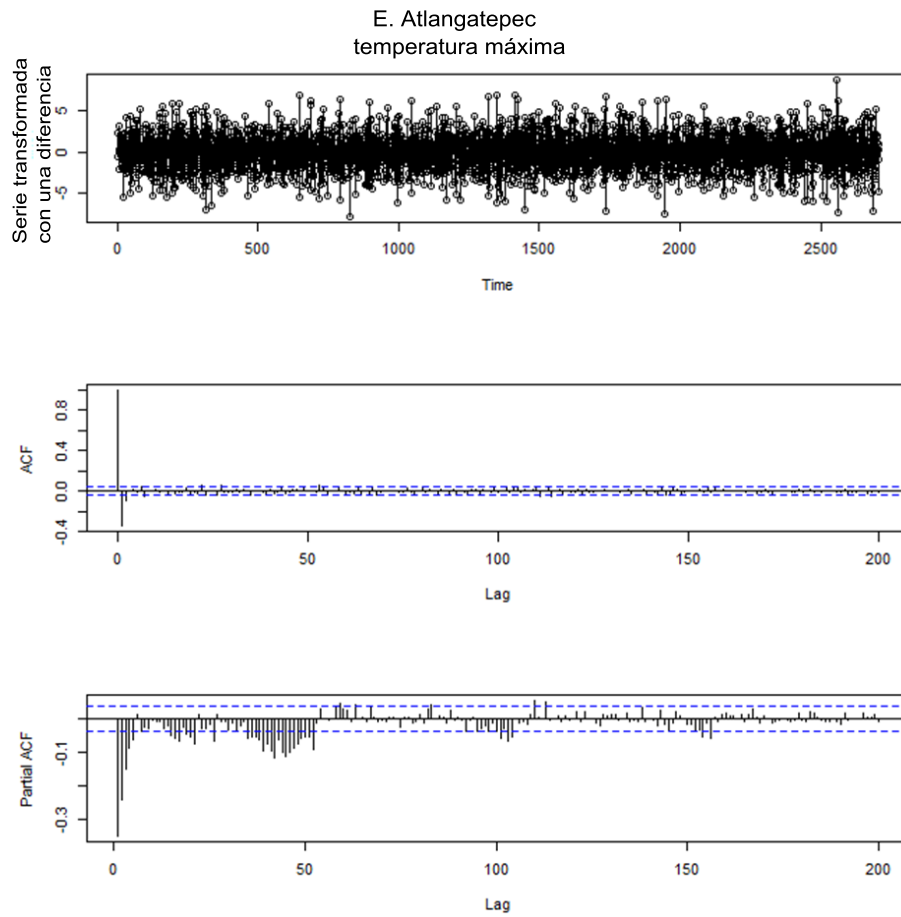


Figura 4.10: Serie de tiempo transformada, función de autocorrelación y función de autocorrelación parcial para los datos de temperatura máxima de Atlangatepec, Tlaxcala

también se extingue rápidamente en forma sinusoidal por lo que para el nivel no estacional se propone un modelo $ARIMA(1, 1, 3)$.

Para identificar un modelo estacional, se realizó una diferencia de un periodo ($L = 52$) y, se graficó la función de autocorrelación y la función de autocorrelación parcial de la nueva serie (Figura 4.11).

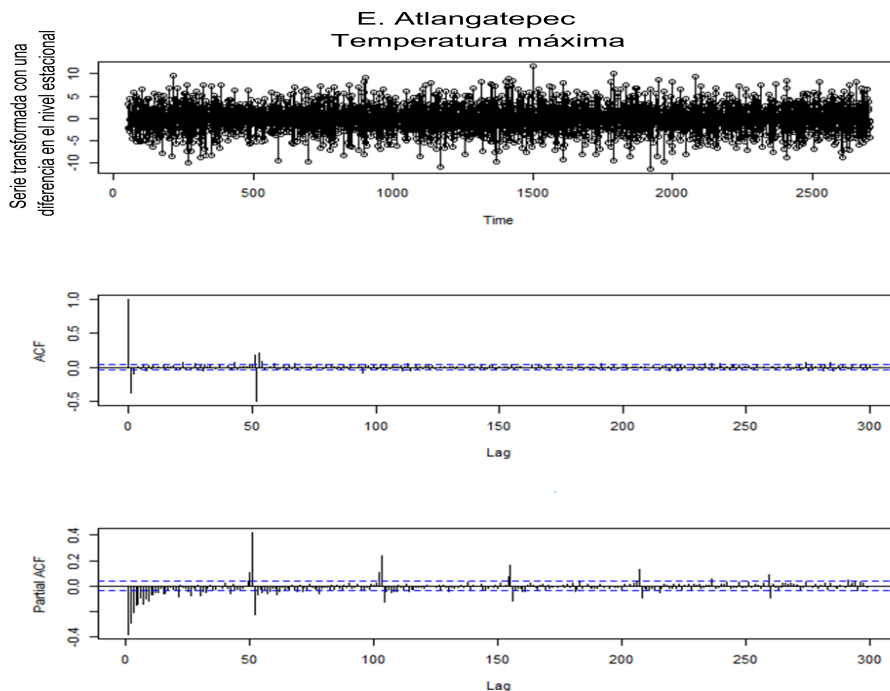


Figura 4.11: Serie de tiempo, función de autocorrelación y función de autocorrelación parcial para los datos de temperatura máxima de Atlangatepec, Tlaxcala, con una diferencia en el nivel no estacional y una diferencia en el nivel estacional

De la Figura 4.11, se observa que la función de autocorrelación se extingue de manera rápida y su función de autocorrelación parcial se extingue en forma exponencial, por lo tanto se propone un modelo estacional $ARIMA(0, 1, 3)$.

Al combinar los modelos propuestos para modelar la temperatura máxima de Atlangatepec, Tlaxcala se tiene

$$y_t^* = y_t - y_{t-1} - y_{t-52} + y_{t-53} , \quad (4.6)$$

con y_t , una observación de la temperatura máxima en el tiempo t . Así, para el nivel no estacional se propone el modelo

$$y_t^* = \phi_1 y_{t-1}^* + z_t + \theta_1 z_{t-1} + \theta_2 z_{t-2} + \theta_3 z_{t-3} \cdot \quad (4.7)$$

Para el nivel estacional se propone

$$y_t^* = z_t - \theta_{1,52} z_{t-52} - \theta_{2,52} z_{t-104} - \theta_{3,52} z_{t-156} \cdot \quad (4.8)$$

Finalmente el modelo integrado propuesto es

$$\begin{aligned} y_t^* = & \phi_1 y_{t-1}^* + z_t + \theta_1 z_{t-1} + \theta_2 z_{t-2} + \theta_3 z_{t-3} - \theta_{1,52} z_{t-52} - \theta_{2,52} z_{t-104} \\ & - \theta_{3,52} z_{t-156} - \theta_1 \theta_{1,52} z_{t-53} - \theta_1 \theta_{2,52} z_{t-105} - \theta_1 \theta_{3,52} z_{t-157} - \theta_2 \theta_{1,52} z_{t-54} \\ & - \theta_2 \theta_{2,52} z_{t-106} - \theta_2 \theta_{3,52} z_{t-108} - \theta_3 \theta_{1,52} z_{t-55} - \theta_3 \theta_{2,52} z_{t-107} - \theta_3 \theta_{3,52} z_{t-159} \cdot \end{aligned} \quad (4.9)$$

Para estimar los parámetros en el modelo (4.9) se utilizó el software *R*. Para verificar qué modelo *SARIMA* se ajustó mejor a los datos fue necesario observar el Criterio de información de Akaike y los p -valores de la estadística de Ljung-Box.

	ar1	ma1	ma2	ma3	sma1	sma2	sma3	intercepto
	0.9105	-0.8679	0.0639	-0.0372	-0.7616	-0.2450	0.0066	0.0001
e.s	0.0342	0.4734	0.4053	0.0613	0.4589	0.3542	0.1247	0.0001

σ^2 estimado: 6.252, logaritmo de verosimilitud = -201 : 44, $AIC = 12420.89$.

Tabla 4.5: Estimación de los parámetros para el modelo (4.9)

Al sustituir los valores estimados de los parámetros (Tabla 4.5), el modelo propuesto para pronosticar la temperatura máxima de Atlangatepec, Tlaxcala queda expresado de la siguiente manera:

$$\begin{aligned} y_t^* = & 0.9105 y_{t-1}^* + z_t - 0.8679 z_{t-1} + 0.0639 z_{t-2} - 0.0372 z_{t-3} + 0.7616 z_{t-52} \\ & + 0.2450 z_{t-104} - 0.00066 z_{t-156} + 0.6609 z_{t-53} + 0.2126 z_{t-105} - 0.0057 z_{t-157} \\ & - 0.0486 z_{t-54} - 0.0156 z_{t-106} + 0.0004 z_{t-108} + 0.0283 z_{t-55} + 0.0091 z_{t-107} \\ & - 0.0002 z_{t-159}. \end{aligned} \quad (4.10)$$

En la Figura 4.12 se muestran los resultados de la prueba de Ljung-Box aplicados a los datos y en la Tabla 4.6 se presentan el valor observado y los pronósticos del modelo estacional *SARIMA*(1, 1, 3)(0, 1, 3) para los datos de la temperatura máxima para Atlangatepec, Tlaxcala.

La Tabla 4.7 muestra algunos modelos propuestos para los datos.

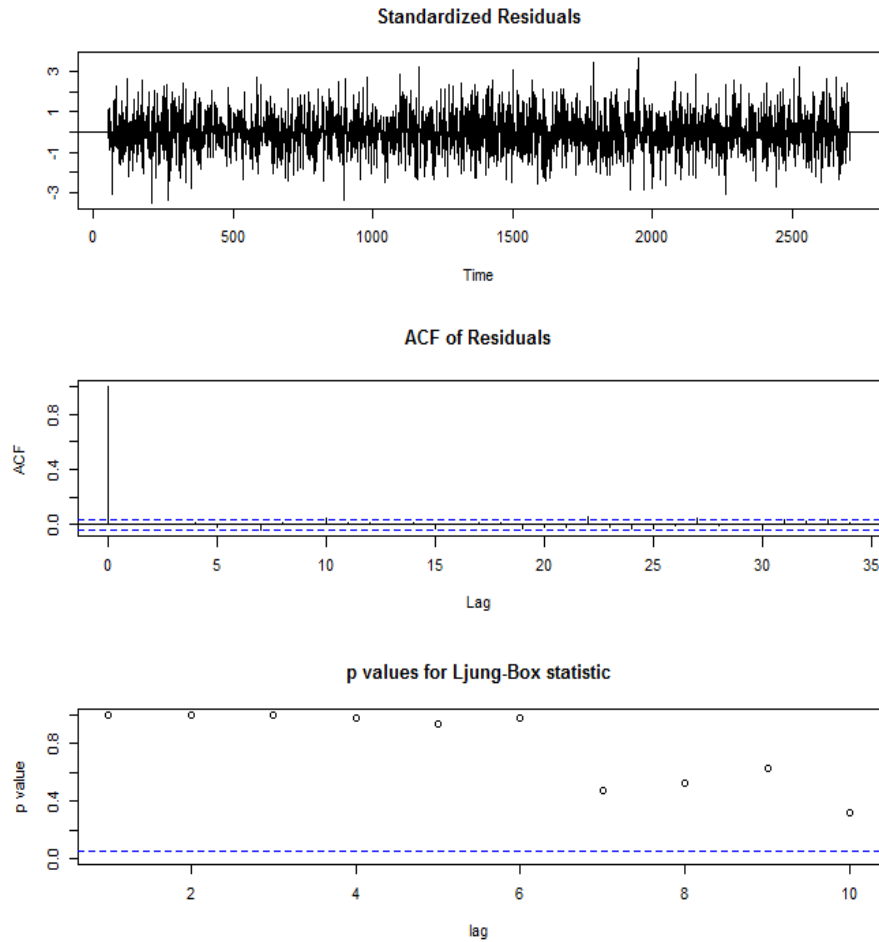


Figura 4.12: Resultados de la prueba de Ljung-Box, para los datos de la serie de tiempo de la temperatura máxima de Atlangatepec, Tlaxcala

Valor observado	Valor pronosticado
18.143	20.02774
19.143	23.14753
21.143	18.61019
19	23.02930
21.571	21.16136
21.714	19.72227

Tabla 4.6: Valor observado y pronosticado para la temperatura máxima de Atlangatepec, Tlaxcala

Modelo	AIC
SARIMA(0,1,2)(0,1,3)	12435.9
SARIMA(1,1,3)(0,1,3)	12420.89
SARIMA(1,1,3)(1,1,3)	12422.76

Tabla 4.7: Modelos que fueron propuestos para la temperatura máxima de Atlangatepec, Tlaxcala

Precipitación

Para la serie de datos de precipitación de la estación ubicada en Atlangatepec, Tlaxcala se obtuvo la primer diferencia, obteniéndose una nueva serie cuyo comportamiento se muestra en la Figura 4.13.

De la Figura 4.13, se observa que la función de autocorrelación para la serie de datos transformada se extingue de manera rápida y su función de autocorrelación parcial se extingue en forma exponencial, por lo tanto, se propone un modelo no estacional $ARIMA(1, 1, 1)$.

Para identificar un modelo estacional, se realizó una diferencia de un periodo ($L = 52$) y, se graficó la serie, la función de autocorrelación y la función de autocorrelación parcial para tratar de identificar un modelo (Figura 4.14).

Al combinar los modelos propuestos para modelar precipitación de Atlangatepec, Tlaxcala se tiene

$$y_t^* = y_t - y_{t-1} - y_{t-52} + y_{t-53} \cdot \quad (4.11)$$

Donde y_t una observación de la precipitación en el tiempo t . Así, para el nivel no estacional se propone el modelo

$$y_t^* = \phi_1 y_{t-1}^* + z_t + \theta_1 z_{t-1} \cdot \quad (4.12)$$

Para el nivel estacional se propone

$$y_t^* = z_t - \theta_{1,52} z_{t-52} - \theta_{2,52} z_{t-104} - \theta_{3,52} z_{t-156} \cdot \quad (4.13)$$

Finalmente, el modelo propuesto es

$$\begin{aligned} y_t^* = & \phi_1 y_{t-1}^* + z_t + \theta_1 z_{t-1} - \theta_{1,52} z_{t-52} - \theta_{2,52} z_{t-104} - \theta_{3,52} z_{t-156} \\ & - \theta_1 \theta_{1,52} z_{t-53} - \theta_1 \theta_{2,52} z_{t-105} - \theta_1 \theta_{3,52} z_{t-157} \cdot \end{aligned} \quad (4.14)$$

Para estimar los parámetros en el modelo (4.14) se utilizó el software R . Para verificar qué modelo $ARIMA$ se ajustó mejor a los datos fue necesario observar el Criterio de información de Akaike y los p -valores de la estadística de Ljung-Box.

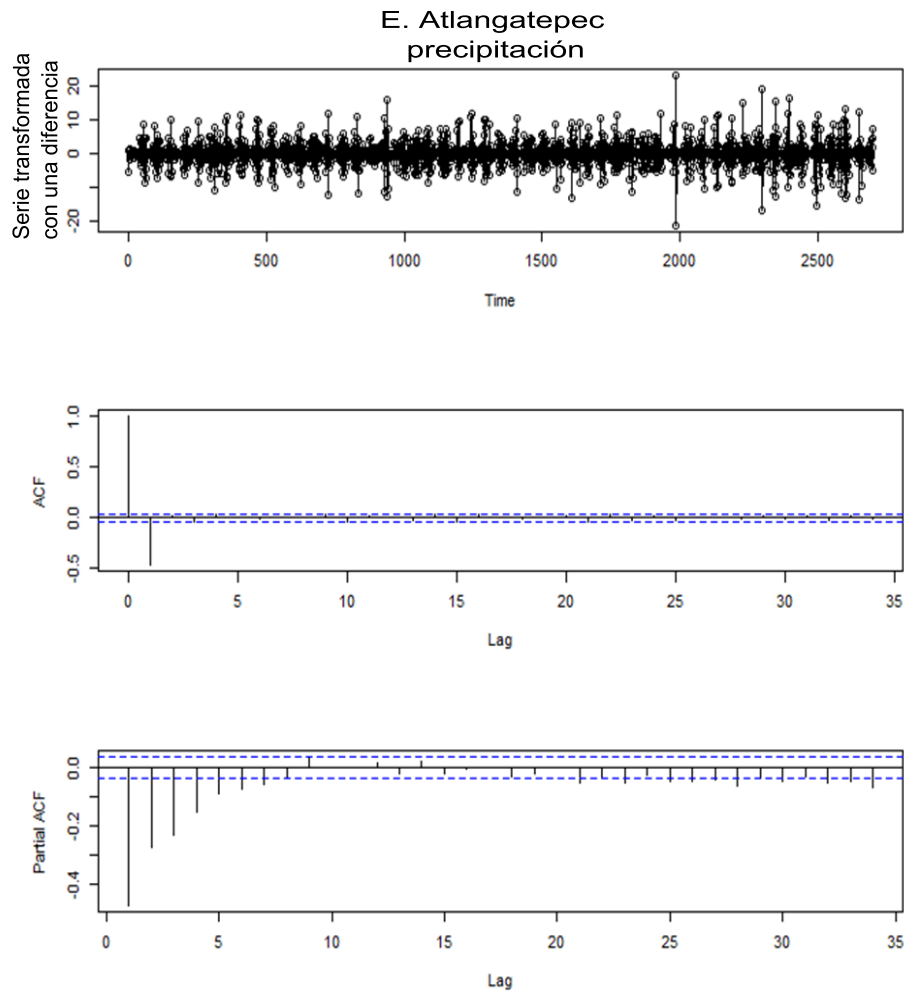


Figura 4.13: Serie de tiempo transformada, función de autocorrelación y función de autocorrelación parcial para los datos de la precipitación de Atlangatepec, Tlaxcala

	ar1	ma1	sma1	sma2	sma3	intercepto
	-0.4692	-0.4343	0.1818	-0.2706	-0.1718	0.0000
e.s	0.1377	0.1261	0.1270	0.1003	0.0250	0.0137

σ^2 estimado: 6.275, logaritmo de verosimilitud = $-6324 : 85$, $AIC = 12663.7$.

Tabla 4.8: Estimación de los parámetros para el modelo 4.14

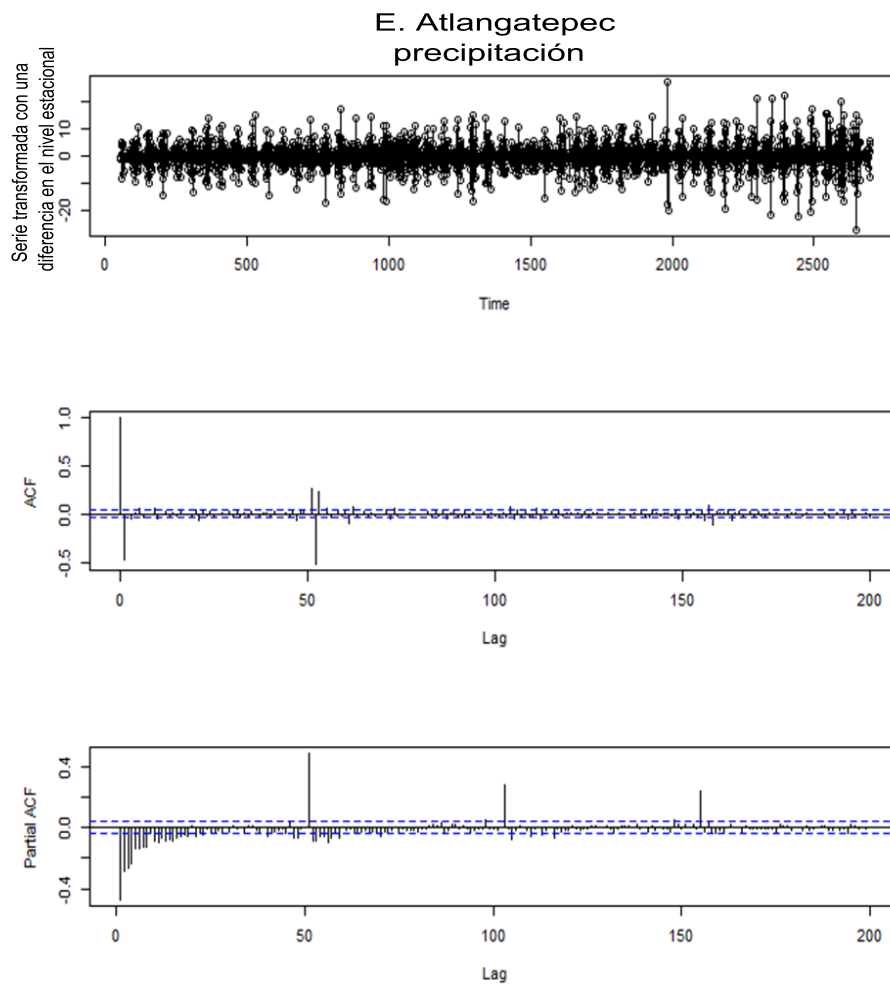


Figura 4.14: Serie de tiempo, función de autocorrelación y función de autocorrelación parcial para los datos de precipitación de Atlangatepec, Tlaxcala, con una diferencia en el nivel no estacional y una diferencia en el nivel estacional

Al sustituir los valores estimados de los parámetros (Tabla 4.8), el modelo propuesto para pronosticar la precipitación de Atlangatepec, Tlaxcala queda expresado de la siguiente manera:

$$y_t^* = -0.4692y_{t-1}^* + z_t - 0.4343z_{t-1} - 0.1818z_{t-52} - 0.2706z_{t-104} - 0.1718z_{t-156} - 0.1886z_{t-53} - 0.1175z_{t-105} - 0.0746z_{t-157} \quad (4.15)$$

En la Figura 4.15 se muestran los resultados de la prueba de Ljung-Box aplicados a los datos y en la Tabla 4.9 se presentan el valor observado y los pronósticos del modelo estacional $SARIMA(1, 1, 1)(0, 1, 3)$ para los datos de la precipitación en Atlangatepec, Tlaxcala. La Tabla 4.10 muestra algunos modelos propuestos para los datos.

Valor observado	Valor pronosticado
3.1429	4.120890
4.8857	4.332434
7.5714	5.504828
7.5714	7.248522
2.5714	5.692316
4.3571	5.264610

Tabla 4.9: Valor observado y pronosticado para la precipitación de Atlangatepec, Tlaxcala

Modelo	AIC
SARIMA(0,1,1)(0,1,3)	12666.12
SARIMA(0,1,2)(0,1,3)	12666.08
SARIMA(1,1,2)(0,1,3)	12665.67
SARIMA(1,1,1)(0,1,3)	12663.7
SARIMA(1,1,1)(1,1,3)	12665.46
SARIMA(1,1,2)(1,1,3)	12667.43

Tabla 4.10: Modelos que fueron propuesto para la precipitación de Atlangatepec, Tlaxcala

Huamantla se sitúa en un eje de coordenadas geográficas entre los $19^{\circ}18'41''$ latitud norte y $97^{\circ}55'24''$ longitud oeste. Su clima se considera semiseco templado, con régimen de lluvias en los meses de mayo, junio, agosto y septiembre. Los meses más calurosos son marzo, abril y mayo.

El municipio de Tlaxco se encuentra ubicado en el Altiplano central mexicano a 2600 m sobre el nivel del mar, se sitúa en un eje de coordenadas geográficas entre

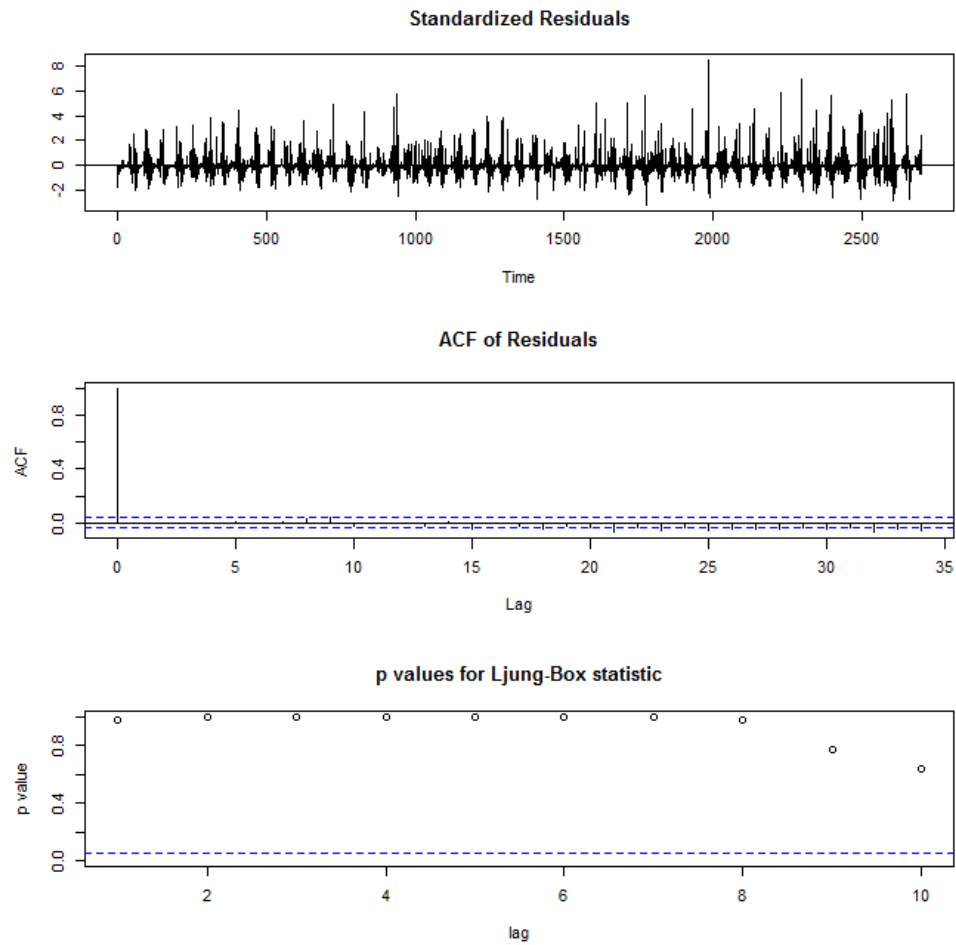


Figura 4.15: Resultados de la prueba de Ljung-Box, para los datos de la serie de tiempo de la precipitación de Atlangatepec, Tlaxcala

los $19^{\circ}36'50''$ latitud norte y $98^{\circ}7'7''$ longitud oeste. En el municipio, se considera un clima templado subhúmedo, con régimen de lluvias en los meses de junio a septiembre. Los meses más calurosos son de marzo a mayo.

De acuerdo al Instituto Nacional de Estadística Geografía e Informática, la posición geográfica de Españita, Tlaxcala es $19^{\circ}27'41''$ latitud norte y $98^{\circ}25'23''$ longitud oeste. El clima del municipio es templado frío, con régimen de lluvias en los meses de julio a septiembre.

El Carmen Tequexquitla, se sitúa en un eje de coordenadas geográficas entre los $19^{\circ}19'$ latitud norte y $97^{\circ}39'$ longitud oeste, su clima se considera templado subhúmedo con lluvias en verano.

Para la base de datos de las otras cuatro estaciones, se siguió la misma metodología aplicada a los datos de la estación de Atlangatepec, Tlaxcala. En la Tabla 4.11 se presenta el mejor modelo encontrado para cada variable analizada y en la Tabla 4.12 se presenta el valor observado y el valor pronosticado.

Estación	Variable	Mejor modelo	SCE
Huamatla	temperatura mínima	$y_t^* = z_t - 0.8002z_{t-1} + 0.3165z_{t-2}$ $- 0.1794z_{t-52} + 0.4049z_{t-104} + 0.3926z_{t-156}$ $+ 0.1983z_{t-208} + 0.1435z_{t-53} - 0.3224z_{t-105}$ $- 0.3141z_{t-157} - 0.1586z_{t-209} - 0.0567z_{t-54}$ $+ 0.1281z_{t-106} + 0.1242z_{t-158} + 0.0627z_{t-210}$	4.41
	temperatura máxima	$y_t^* = 0.7900y_{t-1}^* + z_t - 0.7057z_{t-1}$ $+ 0.00737z_{t-2} - 0.3681z_{t-3} + 0.7853z_{t-52}$ $+ 0.2344z_{t-104} + 0.3309z_{t-156} - 0.5541z_{t-53}$ $- 0.1654z_{t-105} + 0.2351z_{t-157} - 0.0290z_{t-54}$ $+ 0.0172z_{t-106} - 0.0243z_{t-158}$ $- 0.2890z_{t-55} - 0.0862z_{t-107} + 0.1218z_{t-159}$	13.48
	precipitación	$y_t^* = z_t + 0.2745z_{t-1} - 0.1983z_{t-2}$ $+ 0 - 7386z_{t-52} - 0.0343z_{t-104} + 0.02957z_{t-156}$ $- 0.2027z_{t-53} + 0.0094z_{t-105} - 0.0811z_{t-157}$ $- 0.1464z_{t-54} + 0.068z_{t-106} - 0.05863z_{t-158}$	22.18
Tlaxco	temperatura mínima	$y_t^* = z_t - 0.8724z_{t-1} + 0.3778z_{t-2} - 0.1738z_{t-52}$ $+ 0.4536z_{t-104} + 0.5300z_{t-156} + 0.1901z_{t-208}$ $+ 0.1516z_{t-53} - 0.3957z_{t-105} - 0.4623z_{t-157}$ $- 0.1658z_{t-209} - 0.0656z_{t-54} + 0.1713z_{t-106}$ $+ 0.2003z_{t-158} + 0.0718z_{t-210}$	35.34
	temperatura máxima	$y_t^* = -0.4894y_{t-1}^* - 0.9379y_{t-2}^* + z_t$ $- 0.6698z_{t-1} - 0.1654z_{t-2} - 0.00097z_{t-3}$ $- 0.4859z_{t-52} - z_{t-104} + 0.3073z_{t-53}$	19.10

Estación	Variable	Mejor Modelo	SCE
		$+0.6698z_{t-105} - 0.0759z_{t-54}$ $+0.1654z_{t-106} + 0.0047z_{t-55} + 0.0097z_{t-107}$	
	precipitación	$y_t^* = z_t - 0.4217z_{t-1} + 0.5616z_{t-52}$ $+0.2025z_{t-104} + 0.2359z_{t-156} - 0.2368z_{t-53}$ $-0.0853z_{t-105} - 0.0994z_{t-157}$	94.77
Españita	temperatura mínima	$y_t^* = 0.5038y_{t-1}^* + 0.1833y_{t-2}^* + 0.0661y_{t-3}^*$ $+z_t - z_{t-1}$	3.66
	temperatura máxima	$y_t^* = 0.7667y_{t-1}^* + 0.1082y_{t-2}^* + z_t$ $-0.8017z_{t-1} + 0.3985z_{t-52} + 0.3915z_{t-104}$ $+0.21z_{t-156} - 0.3194z_{t-53} - 0.0319z_{t-105}$ $-0.1683z_{t-157}$	25.43
	precipitación	$y_t^* = z_t + 0.3307z_{t-1} + 0.9671z_{t-52}$ $-0.1553z_{t-104} + 0.1882z_{t-156}$ $+0.3198z_{t-53} - 0.0513z_{t-105} + 0.0622z_{t-157}$	374.77
El Carmen Tequexquitla	temperatura mínima	$y_t^* = z_t - 0.5823z_{t-1} - 0.3064z_{t-2}$ $-0.4198y_{t-52}^* - 0.3348z_{t-52} + 0.1949z_{t-53}$ $+0.1025z_{t-54}$	17.45
	temperatura máxima	$y_t^* = z_t + 0.0760z_{t-1} + 0.7610z_{t-52}$ $+0.0628z_{t-104} - 0.0578z_{t-53} + 0.0047z_{t-105}$	16.64
	precipitación	$y_t^* = z_t - 0.2627z_{t-1} + 0.910z_{t-52}$ $+0.309z_{t-104} - 0.1815z_{t-53} - 0.0811z_{t-105}$	180

Tabla 4.11: Mejor modelo y suma del cuadrado de los errores de predicción

Estación	Temperatura mínima		Temperatura máxima		Precipitación	
	V. O.	V. P.	V. O.	V. P.	V. O.	V. P.
Huamantla	4.4286	4.4282879	20.429	22.29531	6.1143	1.7113017
	1.7143	0.7434613	20.429	21.03791	0.65714	1.900622
	2.4286	2.4808470	21.286	21.40124	10.771	3.836428
	1.7143	2.5952023	21.429	18.63371	1.3143	1.479240
	4.2857	2.6581274	22.429	21.63324	1.3	10.822152
	3	2.8002909	22.286	21.20478	2.4571	0.7364535
Tlaxco	4.7857	4.9959468	20.571	23.87766	1.3429	2.6882866
	2.7143	5.3840089	20.714	22.82184	3.2	1.7364943
	2.3571	-0.6164397	22.429	22.16891	4.6857	1.7934649
	1.2143	0.9131620	21.714	23.29916	7.6714	5.4076355
	4.7857	0.4493721	23.857	23.77909	0.15714	8.8789060

Estación	Temperatura mínima		Temperatura máxima		Precipitación	
	V. O.	V. P.	V. O.	V. P.	V. O.	V. P.
Tlaxco	2.5	1.840886	23	24.06419	1.4714	0.3502466
	7.2857	5.786217	22.714	20.95363	16.457	14.628920
	6.1429	5.764986	21	20.67117	4.8571	10.816500
Españita	6.1429	5.736557	21.714	18.53215	9.4286	11.010197
	5.4286	5.716324	20	17.66128	8.7143	24.829147
	5.4286	5.699410	20.286	17.74430	10.286	18.830864
	4.7143	5.685203	19.714	19.30077	8.1429	7.299247
	4.1429	3.3485148	19.571	21.72158	5.8571	1.2317796
	0.71429	1.1953804	20	18.65420	0	2.0623586
El Carmen T.	-0.637	-3.5046050	20.429	22.36348	7.6429	3.1339280
	-2.199	-0.3549704	20.429	20.64627	5.4286	0.4912674
	0.857	-1.3504528	23.143	20.78707	0	2.6342668
	-0.471	-0.7751810	20.714	21.64186	0.28571	10.42004

Tabla 4.12: Valor observado y valor pronosticado en cada una de las estaciones para las variables estudiadas, utilizando series de tiempo

4.2. Análisis de los datos utilizando redes neuronales

Las redes neuronales fueron programadas en el software matlab utilizando una mejora del algoritmo de back-propagation conocido como el algoritmo de Levenverg-Marquard (descrito en el capítulo anterior). En la programación, se tienen como datos de entrada el año y la semana y como datos de salida cada una de las variables estudiadas. Debido a que el contradominio de las funciones de activación utilizadas es $[-1, 1]$, los valores de la capa de salida de la redes utilizadas se dividieron entre el valor más grande de todo el conjunto de datos de cada una de las variables estudiadas, para lograr que el conjunto de datos de salida estuviera en el rango de valores adecuado.

En todas las redes fue necesario utilizar dos capas ocultas. Se observó que cada vez que compilaban las redes en Matlab, éste daba un pronóstico distinto aunque cercano a pronósticos realizados con anterioridad, por lo que se compiló 500 veces cada una de las redes y se promediaron los pronósticos obtenidos, dejando este promedio como el pronóstico final de la red. A continuación se muestran los pronósticos encontrados para cada una de las estaciones mediante redes neuronales.

4.2.1. Estación de Atlangatepec, Tlaxcala

Temperatura mínima

Para la estación de la temperatura mínima de la estación se eligió una red 2, 14, 7, 1, por ser ésta con la que se obtuvieron los mejores pronósticos. Para esta red, se tuvo como datos de entrada el año y la semana, dos capas ocultas con 14 y 7 neuronas, respectivamente y como salida la temperatura mínima, las funciones de activación utilizadas en cada capa oculta fueron tangentes sigmoidales.

La Tabla 4.13 muestra el valor observado y pronosticado de la temperatura mínima de la estación.

Valor observado	Valor pronosticado
0.57143	0.77192
-0.42857	1.002
-0.71429	1.2486
-2.6429	1.5325
0.71429	1.8677
0	2.2549

Tabla 4.13: Valor observado y pronosticado para la temperatura mínima de Atlangatepec, Tlaxcala, utilizando redes neuronales

Temperatura máxima

Para la estación de la temperatura máxima de la estación se eligió una red 2, 18, 16, 1, en donde las dos entradas fueron el año y la semana, dos capas ocultas con 18 y 16 neuronas, respectivamente y como salida la temperatura máxima, las funciones de activación utilizadas en cada capa oculta fueron tangentes sigmoidales.

La Tabla 4.14 muestra el valor observado y pronosticado de la temperatura máxima de la estación.

Precipitación

Para la estación de la precipitación de la estación se eligió una red 2, 1, 1, 1, en donde las dos entradas fueron el año y la semana, dos capas ocultas con 2 neuronas cada una y como salida la precipitación, las funciones de activación utilizadas en cada capa oculta fueron tangentes sigmoidales.

La Tabla 4.15 muestra el valor observado y pronosticado de la temperatura máxima de la estación.

Valor observado	Valor pronosticado
18.143	19.776
19.143	20.32
21.143	20.93
19	21.37
21.571	21.739
21.714	22.063

Tabla 4.14: Valor observado y pronosticado para la temperatura máxima de Atlangatepec, Tlaxcala, utilizando redes neuronales

Valor observado	Valor pronosticado
3.1429	2.5374
4.8857	2.4158
7.5714	2.2518
7.5714	2.0472
2.5714	1.8156
4.3571	1.5992

Tabla 4.15: Valor observado y pronosticado para la precipitación de Atlangatepec, Tlaxcala, utilizando redes neuronales

En la Tabla 4.16 se presenta el pronóstico obtenido en cada una de las estaciones restantes. La red neuronal utilizada para el análisis de los datos de cada estación se puede encontrar en el Apéndice 3.

Estación	Temperatura mínima		Temperatura máxima		Precipitación	
	V. O.	V. P.	V. O.	V. P.	V. O.	V. P.
Huamantla	4.4286	4.4205	20.429	21.8	6.1143	2.3645
	1.7143	4.6838	20.429	22.298	0.65714	2.1625
	2.4286	4.8475	21.286	22.939	10.771	1.9047
	1.7143	4.9899	21.429	23.586	1.3143	1.5969
	4.2857	5.1512	22.429	24.103	1.3	1.3332
	3	5.3716	22.286	24.51	2.4571	1.1615
Tlaxco	4.7857	3.4203	20.571	24.616	1.3429	3.0405
	2.7143	3.6087	20.714	24.885	3.2	2.7825
	2.3571	3.7432	22.429	25.252	4.6857	2.4468
	1.2143	3.8656	21.714	25.611	7.6714	2.0621
	4.7857	4.018	23.857	25.922	0.15714	1.669
	2.5	4.2537	23	26.186	1.4714	1.336

Estación	Temperatura mínima		Temperatura máxima		Precipitación	
	V. O.	V. P.	V. O.	V. P.	V. O.	V. P.
Españita	7.2857	5.5988	22.714	20.541	16.457	3.4117
	6.1429	5.8541	21	20.628	4.8571	3.4115
	6.1429	6.0356	21.714	20.714	9.4286	3.412
	5.4286	6.1434	20	20.794	8.7143	3.4147
	5.4286	6.2087	20.286	20.863	10.286	3.4186
	4.7143	6.256	19.714	20.91	8.1429	3.4228
El Carmen T.	4.1429	0.93839	19.571	21.498	5.8571	1.4516
	0.71429	1.1121	20	21.899	0	1.449
	-0.637	1.1936	20.429	22.357	7.6429	1.4483
	-2.199	1.2574	20.429	22.786	5.4286	1.4475
	0.857	1.3649	23.143	23.19	0	1.447
	-0.471	1.5835	20.714	23.535	0.28571	1.4466

Tabla 4.16: Valor Observado y valor pronosticado en cada una de las estaciones para las variables estudiadas, utilizando redes neuronales

CAPÍTULO 5

Análisis de Resultados

Para aplicar la metodología de Box-Jenkins, fué necesario eliminar la estacionalidad de los datos mediante segundas diferencias. Con la nueva serie obtenida, se propuso un modelo que se ajustara a los datos.

Una herramienta útil para determinar estacionalidad en las series de tiempo fueron los gráficos boxplot de cada variable analizada, de donde además, se observaron e identificaron datos atípicos los cuales se describen en forma más precisa a continuación.

Estación de Atlangatepec Tlaxcala. Los datos atípicos observados para la temperatura mínima son: en el mes de enero $3.4516^{\circ}C$ en el año 1964, abril de 2012 con una temperatura mínima de $2.06667^{\circ}C$, el mes de junio con $5.6667^{\circ}C$ y $5.9^{\circ}C$ en los años 2006 y 2008 respectivamente, el mes de agosto con $4.3871^{\circ}C$ en 2009 y octubre con $1.663^{\circ}C$, $1.129^{\circ}C$ y $1.6774^{\circ}C$ para los años 2008, 2010 y 2012, respectivamente. Para la temperatura máxima de esta estación, se observaron los siguientes datos: $15.919^{\circ}C$ y 21.871 en el mes de enero para los años 1981 y 1963, respectivamente. De $24^{\circ}C$ y $23.448^{\circ}C$ para el mes de febrero en los años 1962 y 1964, respectivamente. De $27.548^{\circ}C$ en el mes de mayo de 1998, $24.367^{\circ}C$ y $25.033^{\circ}C$ en junio para los años 1969 y 1998, respectivamente. De $23.081^{\circ}C$ y $22.613^{\circ}C$ para agosto de 1962 y 1964, respectivamente. De $16.65^{\circ}C$ para septiembre de 1984 y $21.871^{\circ}C$ para diciembre de 1970. Para la precipitación se registraron los siguientes datos: 1.33 mm , 1.029 mm , 2.1097 mm y 1.8548 mm para el mes de enero en los años 1967, 1980, 1992 y 2010, respectivamente. De 1.4786 mm y 3.3714 mm para febrero durante los años 2007 y 2010, respectivamente. De 1.6 mm y 1.7387 mm para marzo de los años 1978 y 1997, respectivamente. De 3.74 mm y 3.3567 mm para abril de 1968 y

2010 respectivamente, 8.8387 *mm* para julio de 2010, 9.1516 *mm* en agosto de 1995, 4.3613 *mm*, 5.9355 *mm* y 5.1645 *mm* para octubre de 1992, 1999 y 2005 respectivamente, 2.91 *mm* y 1.65 *mm* para noviembre de 1980 y 1997, respectivamente, finalmente para el mes de diciembre se registraron 1.5548 *mm* y 0.96774 *mm* en diciembre de 1995 y 2009.

Estación de Huamantla Tlaxcala. Para la temperatura mínima se observaron los siguientes datos atípicos: $-0.55357^{\circ}C$, $5.5517^{\circ}C$ y $5.6897^{\circ}C$ para febrero de 1998, 2008 y 2012, respectivamente De $4.4032^{\circ}C$ en mayo de 2006, de 10.783 para junio de 2003 y $8.7097^{\circ}C$, en octubre del 2009. Para la temperatura máxima los datos atípicos observados fueron: $17.774^{\circ}C$ en enero de 1992 y $24.919^{\circ}C$ en octubre de 2006. Para la precipitación los datos atípicos observados fueron: de 2.0323 *mm* y 0.82581 *mm* para enero de 1992 y 2010, respectivamente, en febrero los datos atípicos fueron de 2.2857 *mm*, 2.7107 *mm* y 1.5103 *mm* para 2007, 2010 y 2012, respectivamente, en marzo se registró una precipitación de 2.3258 *mm* en 1997, 7.5452 *mm* en julio del 2010, 7.3935 *mm* en agosto de 1995, 7.1233*mm* y 6.5167 *mm* en agosto de 1998 y 2009 respectivamente. En noviembre de 1992, se tuvo una precipitación de 2.7 *mm* y de 1.7516 *mm* en diciembre de 1995.

Estación de Tlaxco Tlaxcala. Para la temperatura mínima: $7.8167^{\circ}C$ en abril de 2003, $10.097^{\circ}C$ y $6.5484^{\circ}C$ en agosto de 1995 y 1997. De $11.3^{\circ}C$ en septiembre de 1998, de $-1.529^{\circ}C$ para diciembre de 2010. Para la temperatura máxima, los datos atípicos fueron: $20.742^{\circ}C$, $21.387^{\circ}C$, $25.726^{\circ}C$ y $20.387^{\circ}C$ para enero de 1992, 2003, 2007 y 2010. De $20.929^{\circ}C$ para febrero del 2010, $23.516^{\circ}C$ y $33.194^{\circ}C$ para mayo de 1992 y 1998. De $30.483^{\circ}C$ en junio de 1998, de $26.952^{\circ}C$ en julio de 1998, de $25.583^{\circ}C$ en septiembre de 1996, de $20.677^{\circ}C$ en octubre de 1999, y $19.089^{\circ}C$ en diciembre de 2009. Para la precipitación, se observaron los siguientes datos atípicos: 2.4774 *mm* en enero de 1992, 2.0586 *mm* en febrero del 2012, 1.8194 *mm*, 1.7355 *mm* y 1.3419 *mm* en marzo de 1997, 2004 y 2012, 3.7667 *mm* en abril de 1991, 0.52667 *mm*, 6.97 *mm* y 10.54 *mm* en junio de 1998, 2003 y 2008, 7.529 *mm* en agosto de 1995, 7.1233 *mm* en septiembre de 1998, 8.671 *mm* en octubre de 1991, 1.9533 *mm* en noviembre del 2002 y 1.6 *mm* y $1.1581^{\circ}C$ para diciembre de 1995 y 1996.

Estación de Españita Tlaxcala. Los datos atípicos observados para la temperatura mínima fueron: $-0.90323^{\circ}C$ en enero de 2005, $9.1742^{\circ}C$, $8.3871^{\circ}C$ y $4.7419^{\circ}C$ en marzo de 1993, 1996 y 2001 respectivamente, $4.7533^{\circ}C$ en abril de 1993, $8.7742^{\circ}C$ y $2.4968^{\circ}C$ en julio de 1998 y 2012, $3.9367^{\circ}C$ en septiembre de 2002. Para la temperatura máxima, los datos atípicos observados fueron: $24.857^{\circ}C$ en febrero del 2003, $28.083^{\circ}C$ en abril de 1998, $19.923^{\circ}C$ en mayo del 2009, $27.133^{\circ}C$ en junio del 2005, $19.452^{\circ}C$, $22.903^{\circ}C$ y $24.181^{\circ}C$ en agosto de 1995, 2002 y 2009, $18.9^{\circ}C$ $24.772^{\circ}C$ en septiembre de 1991 y 2009, $17.484^{\circ}C$ en diciembre de 2010. Para la precipitación,

los datos atípicos fueron 4.4452 mm, 1.7484 mm y 1.5484 mm en enero de 1992, 1995 y 2010, 2.6786 mm, 2.3103 mm en febrero de 2010 y 2012, 2.4903 mm en marzo de 1997, 5.6548 mm, 9.9387 mm y 6.5774 mm en mayo de 1990, 1992 y 2004, 22.847 mm en septiembre de 1991, 5.3667°C, 1.3233°C y 5.811 mm en noviembre de 1992, 1998 y 2009 y 0.761290323 mm, 1.4097 mm, 2.7032 mm, 5.9387 mm en diciembre de 1989, 1995, 1997 y 2009.

Estación de El Carmen Tequexquitla Tlaxcala. Los datos atípicos observados para la temperatura mínima fueron 4.2581°C en agosto de 1996. Para la temperatura máxima los datos atípicos observados fueron: 17.286°C en febrero de 1997, 18°C y 28.71°C en marzo de 1997 y 2007, 19.9°C en abril de 1997, 21.097°C en mayo de 1997, 28.968°C y 28.29°C en julio de 2006 y 2007. Para la precipitación, los datos atípicos fueron: 1.6774 mm y 1.4516 mm en enero de 1992 y 1995, 0.92857 mm 1.1429 mm y 1.6607 mm en febrero de 1993, 2007 y 2010, 0.74194 mm en marzo de 1997, 3.5333 mm, 2.43 mm en abril de 1997 y 2010, 4.7419 mm y 4.629 mm en mayo de 1992 y 2006, 6.9333 mm en junio de 2008, 7.55 mm en septiembre de 2009, 4.5484 mm 4.6613 mm en octubre de 1999 y 2005, 2.4 mm y 1.6667 mm en noviembre de 1992 y 2006, 1.4194 mm en diciembre de 1995.

Para encontrar una explicación a los datos atípicos se revisó la página de *Climate Prediction Center* [4], y se encontraron las siguientes caracterizaciones:

El año 1962 fue catalogado como un año neutral, 1963 fue año neutral hasta el mes de mayo donde se empezó a registrar un evento de Niño y que duró hasta febrero de 1964 y al cual procedió un evento de Niña. El año 1968 inició como un año neutro hasta el mes de Julio, donde se presentó un evento de Niño y que continuó así hasta febrero de 1970 y enseguida hubo un evento de año neutro para continuar un evento de Niña a partir del mes de julio.

Los años 1980 y 1981 se clasificaron como años neutros, el año 1989, inició como año de Niña hasta el mes de junio para continuar después como un año normal, los años siguientes fueron años normales hasta mayo de 1991 que continuó con un evento de Niño hasta julio de 1992 después de este mes, 1992, 1993 y los primeros 7 meses de 1994 se consideraron años neutros. A partir de Agosto de 1994 y hasta el primer cuatrimestre de 1995 se consideró evento de Niño, después de este periodo, le prosiguió un evento de año neutro para finalizar con un evento de Niña hasta marzo de 1996. A partir de marzo de 1996, le continuó un evento de año neutro hasta mayo de 1997. A partir de mayo de 1997, se empezó a registrar un evento de Niño que duró hasta mayo de 1998, después de mayo, se presentó un evento de Niña el cual duró hasta abril del 2001.

El año 2002 empezó como neutro hasta abril donde inició un evento de Niño el cual

duró hasta enero del 2003, después de enero del 2003, el año se consideró año neutro hasta Julio del 2004 donde se presentó un evento de Niño el cual terminó en febrero del 2005. Aunque 2005 inició con un evento de Niño, después se presentó un evento neutro y a partir de octubre inició un evento de Niña el cual duró hasta abril del 2006, después de esto, 2006 tuvo un periodo corto de año neutro y a partir de junio se presentó un año de Niño el cual se extendió hasta el primer trimestre del 2007.

El año 2008 inició y terminó como un año de Niña, 2009 inició como año de Niña, sin embargo, a partir de junio se presentó un evento de Niño el cual duró hasta abril del 2010. A partir de junio del 2010 y hasta la fecha, los años se han considerado como años Neutros con poca presencia de eventos de Niña y sin ninguna presencia de eventos de Niño.

De acuerdo a la descripción anterior, es posible explicar el comportamiento de los datos atípicos observados en los gráficos boxplot de la Figura 4.2. En el caso de los datos atípicos sobre temperatura máxima, la mayoría de éstos, ocurren en meses donde se registró un evento de Niño. Los datos atípicos sobre temperatura mínima y precipitación ocurrieron en años con eventos de Niña y años neutros.

En general, los pronósticos realizados con series de tiempo a través de la metodología Box-Jenkins para la temperatura mínima y la temperatura máxima de las 5 estaciones meteorológicas estudiadas, se obtuvieron valores cercanos a los valores observados, esto, visto a través del cálculo del error absoluto, en este sentido, los errores menos acertados se obtuvieron para la temperatura máxima de Atlangatepec y la temperatura mínima de Tlaxco. Para el caso de la precipitación en la mayoría de las estaciones estudiadas no se obtuvieron pronósticos cercanos a los valores observados salvo la estación de Atlangatepec donde se tiene buen pronóstico.

Para aplicar la metodología de redes neuronales fue necesario normalizar los datos de salida (en este caso, temperatura mínima, máxima y precipitación) antes de programar la red neuronal para cada base de datos ya que las salidas de redes neuronales están dentro del intervalo $[-1, 1]$. Para la mayoría de las variables estudiadas de cada estación, fue necesario una red con distinto número de capas y distinto número de neuronas dentro de cada capa.

Del análisis de la temperatura mínima y máxima de las 5 estaciones a través de redes neuronales, el peor pronóstico se obtuvo para la temperatura máxima de Tlaxco y el mejor pronóstico fue para la temperatura máxima de Atlangatepec y la temperatura mínima de Españita. Los pronósticos para la precipitación no estuvieron cercanos a los valores observados para la mayoría de los datos de las estaciones estudiadas.

Al comparar las dos metodologías estudiadas se observa que en estaciones como Tlaxco para la temperatura mínima, el pronóstico con redes neuronales es mejor que el

pronóstico con series de tiempo y el pronóstico realizado con series de tiempo para la precipitación de Atlangatepec, temperatura mínima de Huamantla Tlaxcala es mejor que el pronóstico obtenido con redes neuronales. En general, para todas las estaciones estudiadas, el pronóstico obtenido a través de series de tiempo es mejor en algunos datos observados pero en otros, el pronóstico es mejor con el análisis por redes neuronales, de forma que no es posible dar una recomendación general para alguno de los métodos aplicados en este trabajo.

Conclusiones

En el análisis de los datos, los gráficos boxplot fueron de utilidad para determinar estacionalidad y observar datos atípicos, además de esta herramienta estadística, fue necesario revisar la página de Climate Prediction Center en la que se encuentra la caracterización de los años (años de Niña, años de Niño, o años neutrales) la cual, fue de ayuda para explicar el comportamiento de las series de datos analizadas y la presencia de los datos atípicos.

De los pronósticos realizados para la temperatura mínima y máxima en las estaciones estudiadas, en general, se obtuvieron pronósticos aceptables con el análisis de series de tiempo y redes neuronales. Al comparar las dos metodologías aplicadas se observa que en algunos casos los pronósticos realizados con series de tiempo fueron mejores que los pronósticos obtenidos con redes neuronales. En el caso de la precipitación, aunque en general no se obtuvieron resultados aceptables con ambas metodologías, los resultados obtenidos con redes neuronales se vieron superados con el estudio de series de tiempo. Una posible explicación a este hecho, puede ser que para la precipitación, la red neuronal utilizada aprendió el cero, observación con mayor presencia en los datos observados.

Es importante un análisis de valores extremos que permita realizar pronóstico de estos, así, como aplicar otras metodologías como la de los Wavelets para realizar pronóstico en general y comparar los resultados con las metodologías aplicadas.

Bibliografía

- [1] Ayuga T. E. et al. Cambio Climático Regional: Modelos Estadísticos de predicción ARIMA de precipitaciones en dos estaciones españolas representativas de dos grupos con diferentes características climáticas. Asociación Española de Climatología. Serie No.6. 2008.
- [2] Brockwell J. P., Davis A. R. Introduction to Time series and Forecasting. Springer. Second Ediditon. ISBN: 978-0-387-95351-9. 2002.
- [3] Brockwell J. P., Davis A. R. Time series: Theory and Methods. Springer- Verlag. ISBN: 0-387-96406-1. 1987.
- [4] Climate Prediction Center. www.cpc.ncep.noaa.gov/products/precip/CWlink/MJO/enso.shtml.
- [5] Comisión Nacional del Agua. Manual Teórico Práctico del Observador Meteorológico de Superficie. www.conagua.gob.mx. 2010.
- [6] Douglas.H.J. Time Series Analysis. Princeton University Press. ISBN: 0-691-04289-6. 1994.
- [7] Du K.L., M.N.S. Swamy., Neuronal Networks in a softcomputing framework. Spriner. 2006.
- [8] Espejo G. F., et al. Cambio Climático Regional: Tendencias recientes en las series de temperatura del Pirineo central y occidental. Asociación Española de Climatología. Serie No. 6. 2008.
- [9] Glosarios CONAGUA. <http://smn.cna.gob.mx/glosario/glost.html>. 1987.
- [10] Guenni L., Degryze E., Alvarado K. Trend and Seasonality Assessment for Monthly Precipitation in Venezuela. Revista Colombiana de Estadística. **31**(1). p.p.: 41 – 65. 2008.

- [11] Guerrero V. Modelos estadísticos para series de tiempo univariadas. V Coloquio del departamento de matemáticas. Centro de Investigación y Estudios Avanzados del IPN. 1987.
- [12] Haykin S. Neural networks. A comprehensive foundation. Prentice Hall. Second edition. ISBN:0-13-908385-5. 1999.
- [13] Intergovernmental Panel on Climate Change. www.ipcc.ch.
- [14] J. Alfaro. E., Soley J.F. Descripción de dos métodos de rellenado de datos ausentes en series de tiempo meteorológicas. Revista de matemática: Teoría y Aplicaciones. **16**(1). p.p: 60 – 75.2009.
- [15] Kochler A.B. et.al. Pronósticos, Series de tiempo y Regresión. Un enfoque Aplicado. (Traducción) Cenage Learning. Cuarta Edición. 2007.
- [16] OMM. Guide to Meteorological Instruments and Methods of Observation, WMO-No. 8. Seventh Edition. ISBN:978-92-63-10008-5. 2008.
- [17] World Meteorological Organization. www.wmo.
- [18] Zuñiga L.I., Crespo A. E. Meteorología y Climatología. Universidad Nacional de Educación a Distancia. Segunda reimpresión. 2012.

Definición 5.1 La función de verosimilitud de n variables aleatorias X_1, X_2, \dots, X_n se define como la función de densidad de probabilidad conjunta de las n variables aleatorias, esto es, $L(\theta; x_1, x_2, \dots, x_n) = f_{X_1, X_2, \dots, X_n}(x_1, x_2, \dots, x_n; \theta)$ la cual se considera una función de θ .

Definición 5.2 Sea X_1, X_2, \dots, X_n una muestra aleatoria de una población con función de densidad $f(x; \theta)$ el estimador de máxima verosimilitud $\hat{\Theta} = X_1, X_2, \dots, X_n$ de θ , es la variable aleatoria tal que para cada realización (x_1, x_2, \dots, x_n) de (X_1, X_2, \dots, X_n) el valor correspondiente $\hat{\theta}$ de $\hat{\Theta}$ maximiza a $L(\theta, x_1, x_2, \dots, x_n)$.

Propiedades de los estimadores de máxima verosimilitud

Teorema 5.1 Sea $\hat{\Theta}$ el estimador de máxima verosimilitud de θ en la densidad $f(x; \theta)$. Si $\tau(\cdot)$ es una función con inversa, entonces el estimador de máxima verosimilitud de $\tau(\theta)$ es $\widehat{\tau(\Theta)} = \tau(\hat{\Theta})$.

Teorema 5.2 Los estimadores de máxima verosimilitud $(\hat{\Theta}_1, \hat{\Theta}_2, \dots, \hat{\Theta}_k)$ de los parámetros $(\theta_1, \theta_2, \dots, \theta_k)$ basados en una muestra aleatoria de tamaño k de $f(\underline{x}; \theta_1, \theta_2, \dots, \theta_k)$ son para k grande distribuidos aproximadamente como una normal de dimensión k con vector de medias $\theta_1, \theta_2, \dots, \theta_k$ y matriz de varianzas-covarianzas $\frac{V}{n}$ en donde $V = R^{-1}$ y $R = r_{ij}$, tal que

$$r_{ij} = -E \left\{ \frac{\partial^2}{\partial \theta_i \partial \theta_j} \ln [f(\underline{X}; \theta_1, \theta_2, \dots, \theta_k)] \right\} .$$

En este apartado se presentan las líneas de código en *R* que fueron compiladas para el análisis de las series de tiempo de cada una de las estaciones con la metodología de Box-Jenkins.

```
#####  
#Atlanga TMin#  
#####  
  
d<-read.csv("C:\\Users\\Silvia\\Documents\\MAESTRIA  
\\TRATAMIENTO DE LOS DATOS\\TRATAMIENTO  
SERIES DE TIEMPO\\DATOS ESTIMADOS\\atlangatminsem.csv")  
attach(d)  
serie<-ts(d[,3])  
trans1<-diff(serie,1)  
trans2<-diff(trans1,52)  
par(mfrow=c(3,1))  
plot(trans1, type='o')  
acf(trans1,lag.max=200)  
pacf(trans1,lag.max=200)  
plot(trans2, type='o')  
acf(trans2,lag.max=300)  
pacf(trans2,lag.max=300)  
fit<-arima(trans2,order=c(1,0,2),seasonal=list(order=c(1,0,3)))  
fit  
tsdiag(fit)  
prediccion<-predict(fit,n.ahead=6)$pred  
#prediccion  
x.completada<-c(trans2,prediccion)
```

```

#x.completada
xinv1<-diffinv(x.completada,lag=52,xi=trans1[1:52])
#xinv1
xinv2<-diffinv(xinv1,lag=1,xi=serie[1])
x.reconstruida<-ts(xinv2,start=1961,frequency=52)
x.reconstruida
x.reconstruida[2699:2704]

#####
#Atlanga TMAX#
#####

d<-read.csv("C:\\Users\\Silvia\\Documents\\MAESTRIA
\\TRATAMIENTO DE LOS DATOS\\TRATAMIENTO SERIES DE
TIEMPO\\DATOS ESTIMADOS\\atlangatmaxsem.csv")
attach(d)
serie<-ts(d[,3])
trans1<-diff(serie,1)
trans2<-diff(trans1,52)
par(mfrow=c(3,1))
plot(serie, type='o')
acf(serie)
pacf(serie)
plot(trans1, type='o')
acf(trans1, lag.max=200)
pacf(trans1, lag.max=200)
plot(trans2, type='o')
acf(trans2, lag.max=300)
pacf(trans2, lag.max=300)
fit<-arima(trans2,order=c(1,0,3),seasonal=list(order=c(0,0,3)))
fit
tsdiag(fit)
prediccion<-predict(fit,n.ahead=6)$pred
#prediccion
x.completada<-c(trans2,prediccion)
#x.completada
xinv1<-diffinv(x.completada,lag=52,xi=trans1[1:52])
#xinv1
xinv2<-diffinv(xinv1,lag=1,xi=serie[1])
x.reconstruida<-ts(xinv2,start=1961,frequency=52)
x.reconstruida
x.reconstruida[2699:2704]

```

```

#####
#Atlanga PCP#
#####

d<-read.csv("C:\\Users\\Silvia\\Documents\\MAESTRIA\\
TRATAMIENTO DE LOS DATOS\\TRATAMIENTO SERIES
DE TIEMPO\\DATOS ESTIMADOS\\atlangapcpsem1.csv")
attach(d)
trans1<-ts(d[,3])
trans2<-diff(trans1,1)
trans3<-diff(trans2,52)
par(mfrow=c(3,1))
plot(trans1, type='o')
acf(trans1)
pacf(trans1)
plot(trans2, type='o')
acf(trans2)
pacf(trans2)
plot(trans3, type='o')
acf(trans3,lag.max=200)
pacf(trans3,lag.max=200)

fit<-arima(trans2,order=c(0,0,1),seasonal=list(order=c(0,0,3)))
fit
tsdiag(fit)
prediccion<-predict(fit,n.ahead=6)$pred
#prediccion
x.completada<-c(trans3,prediccion)
#x.completada
xinv1<-diffinv(x.completada,lag=52,xi=trans2[1:52])
#xinv1
xinv2<-diffinv(xinv1,lag=1,xi=trans1[1])
x.reconstruida<-ts(xinv2,start=1961,frequency=52)
#x.reconstruida
x.reconstruida[2699:2704]

#####
#huamantla TMin#
#####

```

```

d<-read.csv("C:\\Users\\Silvia\\Documents\\MAESTRIA
\\TRATAMIENTO DE LOS DATOS\\TRATAMIENTO CON SERIES
DE TIEMPO\\DATOS ESTIMADOS\\huamantlatminsem.csv")
attach(d)
serie<-ts(d[,3])
par(mfrow=c(3,1))
plot(serie, type='o')
acf(serie)
pacf(serie)
trans<-log(ts(d[,3])+5)
trans1<-diff(serie,1)
trans2<-diff(trans1,52)
par(mfrow=c(3,1))
plot(trans1, type='o')
acf(trans1)
pacf(trans1)
par(mfrow=c(3,1))
plot(trans2, type='o')
acf(trans2,lag.max=300)
pacf(trans2,lag.max=300)
fit<-arima(trans2,order=c(0,0,2),seasonal=list(order=c(0,0,4)))
fit
tsdiag(fit)
prediccion<-predict(fit,n.ahead=6)$pred
#prediccion
x.completada<-c(trans2,prediccion)
#x.completada
xinv1<-diffinv(x.completada,lag=52,xi=trans1[1:52])
#xinv1
xinv2<-diffinv(xinv1,lag=1,xi=serie[1])
x.reconstruida<-ts(xinv2,start=1961,frequency=52)
x.reconstruida
x.reconstruida[1192:1197]

#####
#huamantla TMAX#
#####

d<-read.csv("C:\\Users\\Silvia\\Documents\\MAESTRIA
\\TRATAMIENTO DE LOS DATOS\\DATOS ESTIMADOS\\huamantlatmaxsem1.csv")
attach(d)
trans<-ts(d[,3])

```

```

trans1<-diff(trans,1)
trans2<-diff(trans1,52)
par(mfrow=c(3,1))
plot(trans1, type='o')
acf(trans1,lag.max=200)
pacf(trans1,lag.max=200)
plot(trans2, type='o')
acf(trans2,lag.max=200)
pacf(trans2,lag.max=200)
fit<-arima(trans2,order=c(1,0,3),seasonal=list(order=c(0,0,3)))
fit
tsdiag(fit)
prediccion<-predict(fit,n.ahead=6)$pred
#prediccion
x.completada<-c(trans2,prediccion)
#x.completada
xinv1<-diffinv(x.completada,lag=52,xi=trans1[1:52])
#xinv1
xinv2<-diffinv(xinv1,lag=1,xi=trans[1])
x.reconstruida<-ts(xinv2,start=1990,frequency=52)
x.reconstruida
x.reconstruida[1192:1197]

#####
#huamantla PCP#
#####

d<-read.csv("C:\\Users\\Silvia\\Documents\\MAESTRIA
\\TRATAMIENTO DE LOS DATOS\\TRATAMIENTO SERIES
DE TIEMPO\\DATOS ESTIMADOS\\huamantlapcpsem1.csv")
attach(d)
trans1<-ts(d[,3])
trans2<-diff(trans1,1)
trans3<-diff(trans2,52)
par(mfrow=c(3,1))
plot(trans1, type='o')
acf(trans1,lag.max=200)
pacf(trans1,lag.max=200)
par(mfrow=c(3,1))
plot(trans2, type='o')
acf(trans2,lag.max=100)

```

```

pacf(trans2,lag.max=100)
par(mfrow=c(3,1))
plot(trans3, type='o')
acf(trans3,lag.max=200)
pacf(trans3,lag.max=200)
fit<-arima(trans3,order=c(0,0,2),seasonal=list(order=c(0,0,3)))
fit
tsdiag(fit)
prediccion<-predict(fit,n.ahead=6)$pred
#prediccion
x.completada<-c(trans3,prediccion)
#x.completada
xinv1<-diffinv(x.completada,lag=52,xi=trans2[1:52])
#xinv1
xinv2<-diffinv(xinv1,lag=1,xi=trans1[1])
x.reconstruida<-ts(xinv2,start=1990,frequency=52)
x.reconstruida
x.reconstruida[1197:1197]

#####
#tlaxco tmin#
#####

d<-read.csv("C:\\Users\\Silvia\\Documents\\MAESTRIA
\\TRATAMIENTO DE LOS DATOS\\DATOS ESTIMADOS\\tlaxcotminsem.csv")
attach(d)
serie<-ts(d[,3])
par(mfrow=c(3,1))
plot(serie, type='o')
acf(serie)
pacf(serie)
t1<-diff(serie,1)
t2<-diff(t1,52)
par(mfrow=c(3,1))
plot(t1, type='o')
acf(t1,lag.max=150)
pacf(t1,lag.max=150)
plot(t2, type='o')
acf(t2,lag.max=300)
pacf(t2,lag.max=300)
fit<-arima(t2,order=c(0,0,2),seasonal=list(order=c(0,0,4)))
fit

```

```

tsdiag(fit)
prediccion<-predict(fit,n.ahead=6)$pred
#prediccion
x.completada<-c(t2,prediccion)
#x.completada
xinv1<-diffinv(x.completada,lag=52,xi=t1[1:52])
#xinv1
xinv2<-diffinv(xinv1,lag=1,xi=serie[1])
x.reconstruida<-ts(xinv2,start=1989,frequency=52)
x.reconstruida
x.reconstruida[1243:1248]

#####
#tlaxco TMAX#
#####

d<-read.csv("C:\\Users\\Silvia\\Documents\\MAESTRIA
\\TRATAMIENTO DE LOS DATOS\\DATOS ESTIMADOS\\tlaxcotmaxsem.csv")
attach(d)
trans<-ts(d[,3])
trans1<-diff(trans,1)
trans2<-diff(trans1,52)
par(mfrow=c(3,1))
plot(trans1, type='o')
acf(trans1,lag.max=100)
pacf(trans1,lag.max=100)
plot(trans2, type='o')
acf(trans2,lag.max=200)
pacf(trans2,lag.max=200)
fit<-arima(trans2,order=c(2,0,3),seasonal=list(order=c(0,0,2)))
fit
tsdiag(fit)
prediccion<-predict(fit,n.ahead=6)$pred
#prediccion
x.completada<-c(trans2,prediccion)
#x.completada
xinv1<-diffinv(x.completada,lag=52,xi=trans1[1:52])
#xinv1
xinv2<-diffinv(xinv1,lag=1,xi=trans[1])
x.reconstruida<-ts(xinv2,start=1989,frequency=52)
x.reconstruida
x.reconstruida[1243:1248]

```

```

#####
#tlaxco PCP#
#####

d<-read.csv("C:\\Users\\Silvia\\Documents\\MAESTRIA
\\TRATAMIENTO DE LOS DATOS\\TRATAMIENTO SERIES
DE TIEMPO\\DATOS ESTIMADOS\\tlaxcopcpsem2.csv")
attach(d)
trans1<-ts(d[,3])
trans2<-diff(trans1,1)
trans3<-diff(trans2,52)

par(mfrow=c(3,1))
plot(trans2, type='o')
acf(trans2,lag.max=100)
pacf(trans2,lag.max=100)

plot(trans3, type='o')
acf(trans3,lag.max=150)
pacf(trans3,lag.max=150)

fit<-arima(trans3,order=c(0,0,2),seasonal=list(order=c(0,0,3)))
fit
tsdiag(fit)
prediccion<-predict(fit,n.ahead=6)$pred
#prediccion

x.completada<-c(trans3,prediccion)
#x.completada
xinv1<-diffinv(x.completada,lag=52,xi=trans2[1:52])
#xinv1
xinv2<-diffinv(xinv1,lag=1,xi=trans1[1])
x.reconstruida<-ts(xinv2,start=1989,frequency=52)
#x.reconstruida
x.reconstruida[1243:1248]

#####

```

```

#españita tmin#
#####
d<-read.csv("C:\\Users\\Silvia\\Documents\\MAESTRIA
\\TRATAMIENTO DE LOS DATOS\\DATOS ESTIMADOS\\espanitatminsem.csv")
attach(d)
t<-ts(d[,3])
t1<-diff(t,1)
t2<-diff(t1,52)
par(mfrow=c(3,1))
plot(t, type='o')
acf(t,lag.max=200)
pacf(t,lag.max=200)
par(mfrow=c(3,1))
plot(t1, type='o')
acf(t1,lag.max=200)
pacf(t1,lag.max=200)
plot(t2, type='o')
acf(t2,lag.max=200)
pacf(t2,lag.max=200)
fit<-arima(t1,order=c(3,0,1))#,seasonal=list(order=c(0,0,1)))
fit
tsdiag(fit)
prediccion<-predict(fit,n.ahead=6)$pred
#prediccion
x.completada<-c(t1,prediccion)
xinv1<-diffinv(x.completada,lag=1,xi=t[1])
x.reconstruida<-ts(xinv1,start=1989,frequency=52)
x.reconstruida
x.reconstruida[1192:1197]

#####
#españita TMAX#
#####

d<-read.csv("C:\\Users\\Silvia\\Documents\\MAESTRIA
\\TRATAMIENTO DE LOS DATOS\\DATOS ESTIMADOS\\espanitatmaxsem.csv")
attach(d)
trans<-ts(d[,3])
trans1<-diff(trans,1)
trans2<-diff(trans1,52)
par(mfrow=c(3,1))
plot(trans1, type='o')

```

```

acf(trans1,lag.max=200)
pacf(trans1,lag.max=200)
plot(trans2, type='o')
acf(trans2,lag.max=200)
pacf(trans2,lag.max=200)
fit<-arima(trans2,order=c(2,0,1),seasonal=list(order=c(0,0,3)))
fit
tsdiag(fit)
prediccion<-predict(fit,n.ahead=6)$pred
#prediccion
x.completada<-c(trans2,prediccion)
#x.completada
xinv1<-diffinv(x.completada,lag=52,xi=trans1[1:52])
#xinv1
xinv2<-diffinv(xinv1,lag=1,xi=trans[1])
x.reconstruida<-ts(xinv2,start=1989,frequency=52)
x.reconstruida
x.reconstruida[1192:1197]

#####
#españita PCP#
#####

d<-read.csv("C:\\Users\\Silvia\\Documents\\MAESTRIA
\\TRATAMIENTO DE LOS DATOS\\TRATAMIENTO SERIES
DE TIEMPO\\DATOS ESTIMADOS\\espanitapcpsem2.csv")
attach(d)
trans1<-ts(d[,3])
trans2<-diff(trans1,1)
trans3<-diff(trans2,52)
par(mfrow=c(3,1))
plot(trans1, type='o')
acf(trans1,lag.max=100)
pacf(trans1,lag.max=100)
par(mfrow=c(3,1))
plot(trans2, type='o')
acf(trans2,lag.max=100)
pacf(trans2,lag.max=100)
par(mfrow=c(3,1))
plot(trans3, type='o')
acf(trans3,lag.max=180)
pacf(trans3,lag.max=180)

```

```

fit<-arima(trans3,order=c(2,0,2),seasonal=list(order=c(0,0,3)))
fit
tsdiag(fit)
prediccion<-predict(fit,n.ahead=6)$pred
#prediccion
x.completada<-c(trans3,prediccion)
#x.completada
xinv1<-diffinv(x.completada,lag=52,xi=trans2[1:52])
#xinv1
xinv2<-diffinv(xinv1,lag=1,xi=trans1[1])
x.reconstruida<-ts(xinv2,start=1989,frequency=52)
#x.reconstruida
x.reconstruida[1192:1197]

#####
#carmen tmin#
#####

d<-read.csv("C:\\Users\\Silvia\\Documents\\MAESTRIA
\\TRATAMIENTO DE LOS DATOS\\DATOS ESTIMADOS\\carmenminsem.csv")
attach(d)
trans<-ts(d[,3])
trans1<-diff(trans,1)
trans2<-diff(trans1,52)
par(mfrow=c(3,1))
plot(trans1, type='o')
acf(trans1,lag.max=100)
pacf(trans1,lag.max=100)
plot(trans2, type='o')
acf(trans2,lag.mx=100)
pacf(trans2,lag.max=100)
fit<-arima(trans2,order=c(0,0,2),seasonal=list(order=c(1,0,1)))
fit
tsdiag(fit)
prediccion<-predict(fit,n.ahead=6)$pred
#prediccion
x.completada<-c(trans2,prediccion)
xinv1<-diffinv(x.completada,lag=52,xi=trans1[1:52])
xinv2<-diffinv(xinv1,lag=1,xi=trans[1])
x.reconstruida<-ts(xinv2,start=1992,frequency=52)
#x.reconstruida
x.reconstruida[1087:1092]

```

```
#####
#carmen TMAX#
#####

d<-read.csv("C:\\Users\\Silvia\\Documents\\MAESTRIA
\\TRATAMIENTO DE LOS DATOS\\DATOS ESTIMADOS\\carmentmaxsem.csv")
attach(d)
trans<-ts(d[,3])
trans1<-diff(trans,1)
trans2<-diff(trans1,52)
par(mfrow=c(3,1))
plot(trans1, type='o')
acf(trans1,lag.max=100)
pacf(trans1,lag.max=100)
par(mfrow=c(3,1))
plot(trans2, type='o')
acf(trans2,lag.max=100)
pacf(trans2,lag.max=100)
fit<-arima(trans2,order=c(0,0,1),seasonal=list(order=c(0,0,2)))
fit
tsdiag(fit)
prediccion<-predict(fit,n.ahead=6)$pred
#prediccion
x.completada<-c(trans2,prediccion)
#x.completada
xinv1<-diffinv(x.completada,lag=52,xi=trans1[1:52])
#xinv1
xinv2<-diffinv(xinv1,lag=1,xi=trans[1])
x.reconstruida<-ts(xinv2,start=1992,frequency=52)
x.reconstruida
x.reconstruida[1087:1092]

#####
#carmen PCP#
#####

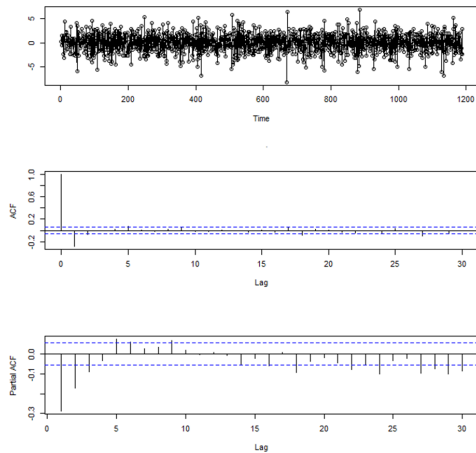
d<-read.csv("C:\\Users\\Silvia\\Documents\\MAESTRIA
\\TRATAMIENTO DE LOS DATOS\\TRATAMIENTO SERIES
DE TIEMPO\\DATOS ESTIMADOS\\carmenpcpsem2.csv")
attach(d)
trans1<-ts(d[,3])
```

```

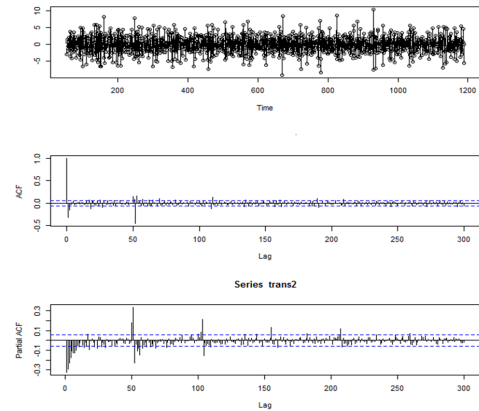
trans2<-diff(trans1,1)
trans3<-diff(trans2,52)
par(mfrow=c(3,1))
plot(trans2, type='o')
acf(trans2,lag.max=200)
pacf(trans2,lag.max=200)
plot(trans3, type='o')
acf(trans3,lag.max=200)
pacf(trans3,lag.max=200)
fit<-arima(trans3,order=c(0,0,1),seasonal=list(order=c(0,0,3)))
fit
tsdiag(fit)
prediccion<-predict(fit,n.ahead=6)$pred
#prediccion
x.completada<-c(trans3,prediccion)
#x.completada
xinv1<-diffinv(x.completada,lag=52,xi=trans2[1:52])
#xinv1
xinv2<-diffinv(xinv1,lag=1,xi=trans1[1])
x.reconstruida<-ts(xinv2,start=1992,frequency=52)
x.reconstruida
x.reconstruida[1087:1092]

```

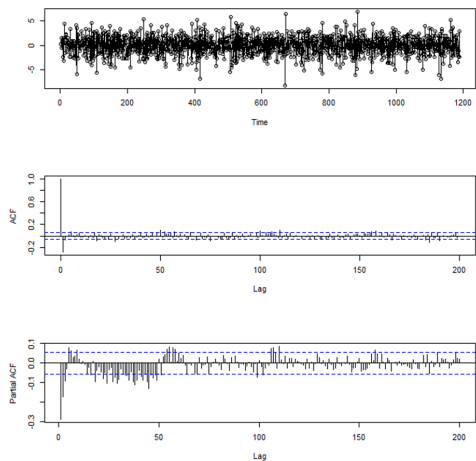
A continuación se presentan las series de tiempo transformadas de las estaciones de Huamantla, Tlaxco, Españita y El Carmen Tequexquitla, las cuales no fueron mostradas las secciones correspondientes en el capítulo 4.



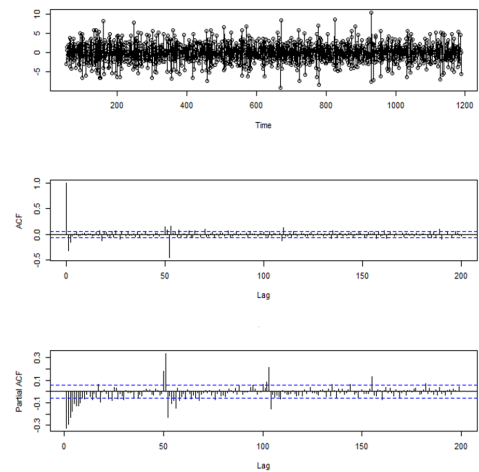
(a) Serie de tiempo con una diferencia en el nivel no estacional para la temperatura mínima de Huamantla Tlaxcala junto con sus funciones ACF y PACF.



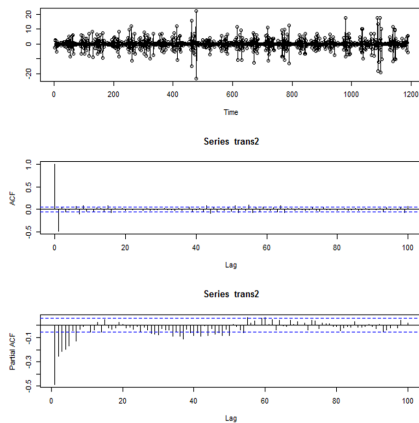
(b) Serie de tiempo con una diferencia en el nivel no estacional y una diferencia en el nivel estacional para la temperatura mínima de Huamantla Tlaxcala junto con sus funciones ACF y PACF.



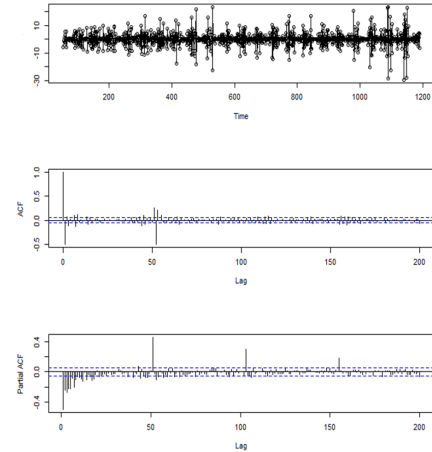
(c) Serie de tiempo con una diferencia en el nivel no estacional para la temperatura máxima de Huamantla Tlaxcala junto con sus funciones ACF y PACF.



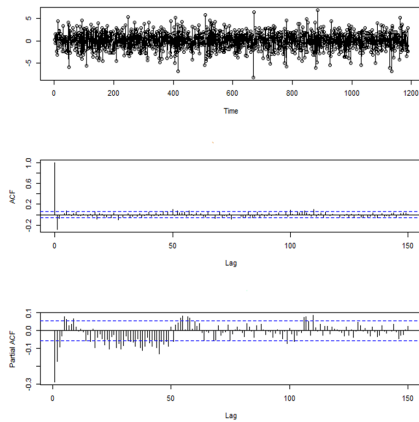
(d) Serie de tiempo con una diferencia en el nivel no estacional y una diferencia en el nivel estacional para la temperatura máxima de Huamantla Tlaxcala junto con sus funciones ACF y PACF.



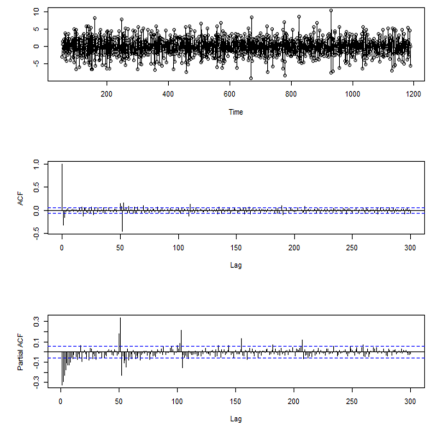
(e) Serie de tiempo con una diferencia en el nivel no estacional para la precipitación de Huamantla Tlaxcala junto con sus funciones ACF y PACF.



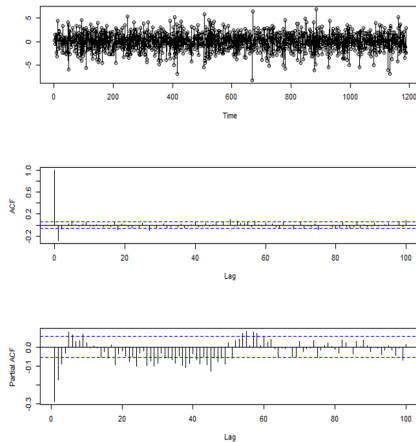
(f) Serie de tiempo con una diferencia en el nivel no estacional y una diferencia en el nivel estacional para la precipitación de Huamantla Tlaxcala junto con sus funciones ACF y PACF.



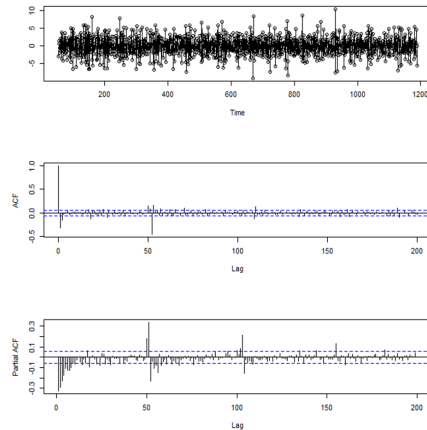
(g) Serie de tiempo con una diferencia en el nivel no estacional para la temperatura mínima de Tlaxco Tlaxcala junto con sus funciones ACF y PACF.



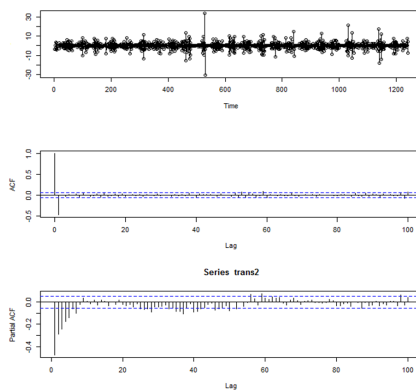
(h) Serie de tiempo con una diferencia en el nivel no estacional y una diferencia en el nivel estacional para la temperatura mínima de Tlaxco Tlaxcala junto con sus funciones ACF y PACF.



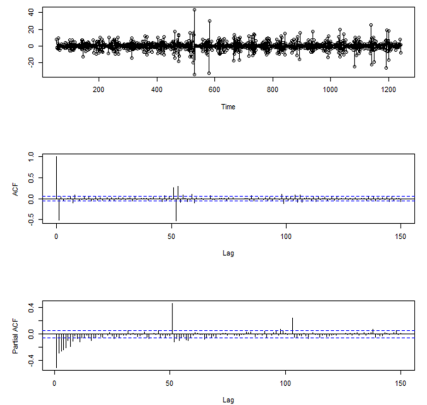
(i) Serie de tiempo con una diferencia en el nivel no estacional para la temperatura máxima de Tlaxco Tlaxcala junto con sus funciones ACF y PACF.



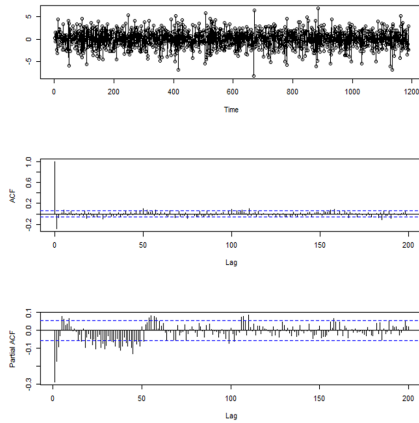
(j) Serie de tiempo con una diferencia en el nivel no estacional y una diferencia en el nivel estacional para la temperatura máxima de Tlaxco Tlaxcala junto con sus funciones ACF y PACF.



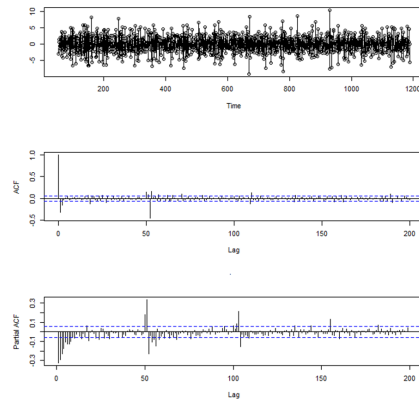
(k) Serie de tiempo con una diferencia en el nivel no estacional para la precipitación de Tlaxco Tlaxcala junto con sus funciones ACF y PACF.



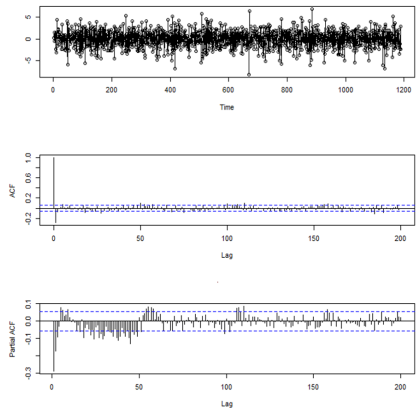
(l) Serie de tiempo con una diferencia en el nivel no estacional y una diferencia en el nivel estacional para la precipitación de Tlaxco Tlaxcala junto con sus funciones ACF y PACF.



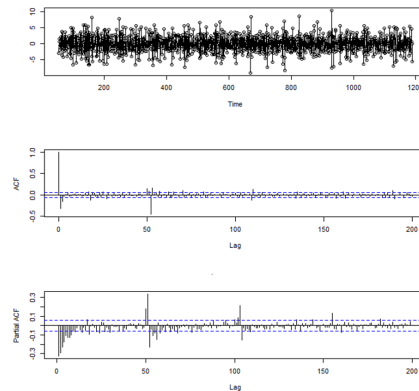
(m) Serie de tiempo con una diferencia en el nivel no estacional para la temperatura mínima de España Tlaxcala junto con sus funciones ACF y PACF.



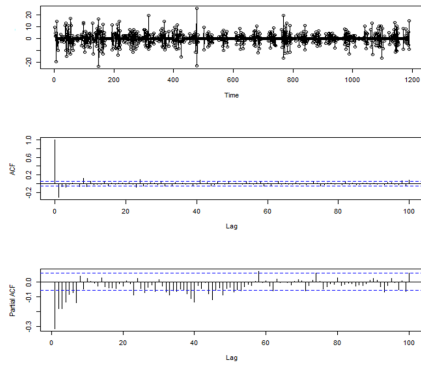
(n) Serie de tiempo con una diferencia en el nivel no estacional y una diferencia en el nivel estacional para la temperatura mínima de España Tlaxcala junto con sus funciones ACF y PACF.



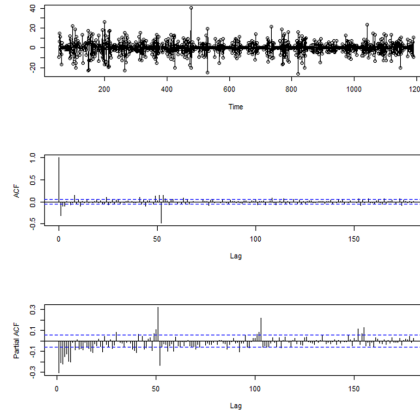
(ñ) Serie de tiempo con una diferencia en el nivel no estacional para la temperatura máxima de España Tlaxcala junto con sus funciones ACF y PACF.



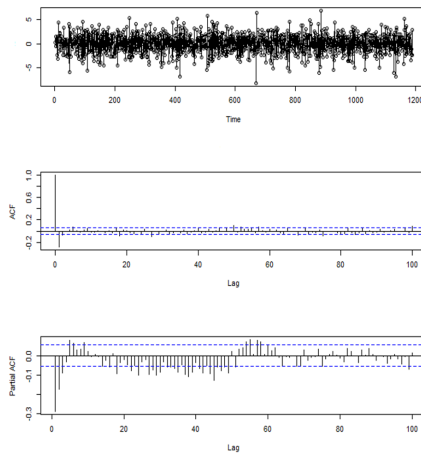
(o) Serie de tiempo con una diferencia en el nivel no estacional y una diferencia en el nivel estacional para la temperatura máxima de España Tlaxcala junto con sus funciones ACF y PACF.



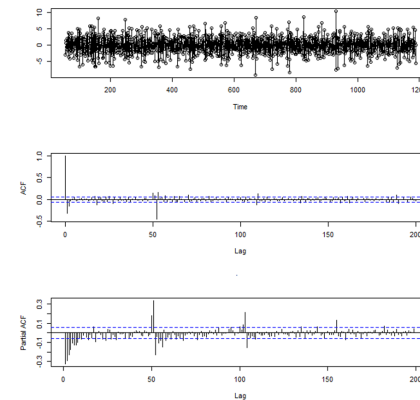
(p) Serie de tiempo con una diferencia en el nivel no estacional para la precipitación de Españita Tlaxcala junto con sus funciones ACF y PACF.



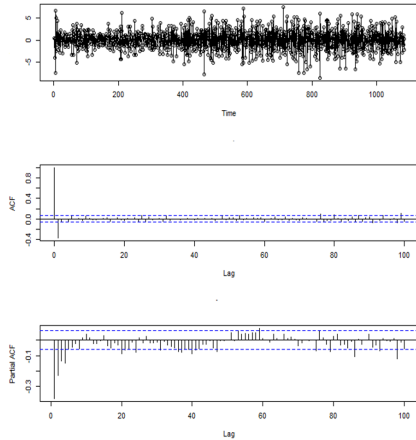
(q) Serie de tiempo con una diferencia en el nivel no estacional y una diferencia en el nivel estacional para la precipitación de Españita Tlaxcala junto con sus funciones ACF y PACF.



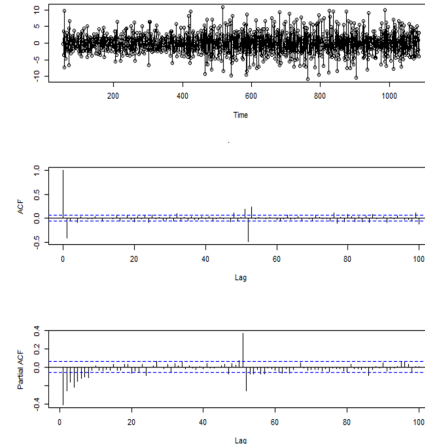
(r) Serie de tiempo con una diferencia en el nivel no estacional para la temperatura mínima de El Carmen Tequexquitla Tlaxcala junto con sus funciones ACF y PACF.



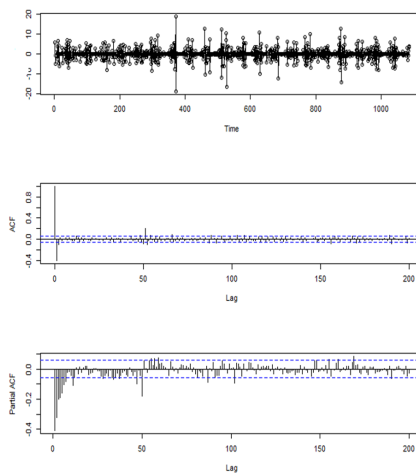
(s) Serie de tiempo con una diferencia en el nivel no estacional y una diferencia en el nivel estacional para la temperatura mínima de El Carmen Tequexquitla Tlaxcala junto con sus funciones ACF y PACF.



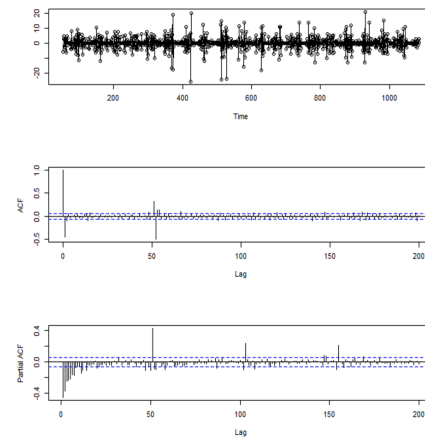
(t) Serie de tiempo con una diferencia en el nivel no estacional para la temperatura máxima de El Carmen Tequexquitla Tlaxcala junto con sus funciones ACF y PACF.



(u) Serie de tiempo con una diferencia en el nivel no estacional y una diferencia en el nivel estacional para la temperatura máxima de El Carmen Tequexquitla Tlaxcala junto con sus funciones ACF y PACF.



(v) Serie de tiempo con una diferencia en el nivel no estacional para la precipitación de El Carmen Tequexquitla Tlaxcala junto con sus funciones ACF y PACF.



(w) Serie de tiempo con una diferencia en el nivel no estacional y una diferencia en el nivel estacional para la precipitación de El Carmen Tequexquitla Tlaxcala junto con sus funciones ACF y PACF.

Apéndice 3

En este apartado se presentan la programación de cada una de las redes neuronales utilizadas para hacer el pronóstico de temperatura mínima, temperatura máxima y precipitación para las estaciones de Atlangatepec, Huamantla, Tlaxco, Españita y El Carmen Tequexquitla.

```
data=load('c:\atlangatminsem.txt');
P=(data(1:2608,1:2))';
T=(data(1:2608,3))';
minimo=min(data(1:2704,3));
t2=T-minimo;
maximo=max(data(1:2704,3));
T1=t2/maximo;
matriz=zeros(6,500);
for k=1:500
    %estructura de la red
    net=newff(P,T1,[14,7,1],{'logsig','logsig','purelin'},'trainlm');
    net=init(net);
    net.trainparam.goal=1e-5;
    net.trainparam.epochs=1500;
    net.trainparam.lr=0.02;
    [net,tr]=train(net,P,T1);

    estimacion=[2012 2012 2012 2012 2012 2012 2012 2012 2012
    2012 2012 2012; 41 42 43 44 45 46 47 48 49 50 51 52];

    datos=[4.2857 3.1429 0.42857 0 1.1429 1.7143 0.57143
    -0.42857 -0.71429 -2.6429 0.7142 0];
    pronostico=[];
```

```

pronostico=cat(2,pronostico,sim(net,estimacion(:,6)));
for i=7:12
X=sim(net,prueba(:,i));
pronostico=cat(2,pronostico,X);
end
pronostico=(pronostico*maximo)+minimo;
matriz(:,k)=pronostico;
k=k+1;
end
mf=zeros(6,1);
for k=1:6
suma=0;
  for i=1:500
    suma=suma+matriz(k,i);
    i=i+1;
  end
promedio=suma/500;
mf(k,1)=promedio;
k=k+1;
end
disp(mf)

data=load('c:\atlangatmaxsem.txt');
P=(data(1:2698,1:2))';
T=(data(1:2698,3))';
maximo=max(data(1:2704,3));
T1=T/maximo;
matriz=zeros(6,500);
for k=1:500

  net=newff(P,T1,[18,16,1],{'tansig','tansig','purelin'},'trainlm');
  net=init(net);
  net.trainparam.goal=1e-3;
  net.trainparam.epochs=1000;
  net.trainparam.lr=0.2;
  [net,tr]=train(net,P,T1);
  estimacion=[2012 2012 2012 2012 2012 2012 2012 2012
2012 2012 2012 2012; 41 42 43 44 45 46 47 48 49 50 51 52];
  datos=[19.514 19.857 21.571 23.429 18.714 17.857
18.143 19.143 21.143 19 21.57 21.714];
  pronostico=[];
  pronostico=cat(2,pronostico,sim(net,estimacion(:,6)));
for i=7:12

```

```

X=sim(net,estimacion(:,i));
pronostico=cat(2,pronostico,X);
end
pronostico=pronostico*maximo;
matriz(:,k)=pronostico;
k=k+1;
disp(k)
end
mf=zeros(6,1);
for k=1:6
    suma=0;
    for i=1:500
        suma=suma+matriz(k,i);
        i=i+1;
    end
    promedio=suma/500;
    mf(k,1)=promedio;
    k=k+1;
end
disp(mf)

data=load('c:\atlangapcpsem1.txt');
P=(data(1:2177,1:2))';
T=(data(1:2177,3))';
maximo=max(data(1:2183,3));
T1=T/maximo;
matriz=zeros(6,500);
for k=1:500
    net=newff(P,T1,[2,1,1],{'tansig','tansig','purelin'},'trainlm');
    net=init(net);
    net.trainparam.goal=1e-3;
    net.trainparam.epochs=500;
    net.trainparam.lr=0.2;
    [net,tr]=train(net,P,T1);
estimacion=[2012 2012 2012 2012 2012 2012
2012 2012 2012 2012 2012 2012; 41 42 43 44 45 46 47 48 49 50 51 52];
datos=[0.1 0 4.5714 6.2143 1.3571 8.5714
3.1429 4.8857 7.5714 7.5714 2.5714 4.3571];

pronostico=[];
pronostico=cat(2,pronostico,sim(net,estimacion(:,6)));
for i=7:12
X=sim(net,estimacion(:,i));

```

```

pronostico=cat(2,pronostico,X);
end
pronosico=pronostico*maximo;
matriz(:,k)=pronostico;
disp(k)
k=k+1;
end
mf=zeros(6,1);
for k=1:6
    suma=0;
    for i=1:500
        suma=suma+matriz(k,i);
        i=i+1;
    end
    promedio=suma/500;
    mf(k,1)=promedio;
    k=k+1;
end
disp(mf)

data=load('c:\huamantlatminsem.txt');
P=(data(1:1190,1:2))';
T=(data(1:1190,3))';
maximo=max(data(1:1196,3));
T1=T/maximo;
matriz=zeros(6,500);
for k=1:500
    net=newff(P,T1,[20,10,1],{'tansig','tansig','purelin'},'trainlm');
    net=init(net);
    net.trainparam.goal=1e-3;
    net.trainparam.epochs=500;
    net.trainparam.lr=0.2;
    [net,tr]=train(net,P,T1);
    estimacion=[2012 2012 2012 2012 2012 2012
2012 2012 2012 2012 2012; 41 42 43 44 45 46 47 48 49 50 51 52];
    datos=[6.8571 6.5714 4.4286 3.2857 6.1429 3.8571
4.4286 1.7143 2.4286 1.7143 4.2857 3];
pronostico=[];
pronostico=cat(2,pronostico,sim(net,estimacion(:,6)));
for i=7:12
    X=sim(net,estimacion(:,i));
    pronostico=cat(2,pronostico,X);
end

```

```

end
pronostico=pronostico*maximo;
matriz(:,k)=pronostico;
k=k+1;
%disp(k)
end
mf=zeros(6,1);
for k=1:6
    suma=0;
    for i=1:500
        suma=suma+matriz(k,i);
        i=i+1;
    end
    promedio=suma/500;
    mf(k,1)=promedio;
    k=k+1;
end

data=load('c:\huamantlatmaxsem.txt');
P=(data(1:1190,1:2))';
T=(data(1:1190,3))';
maximo=max(data(1:1196,3));
T1=T/maximo;
matriz=zeros(6,500);
for k=1:500
net=newff(P,T1,[20,10,1],{'tansig','tansig','purelin'},'trainlm');
net=init(net);
net.trainparam.goal=1e-3;
net.trainparam.epochs=500;
net.trainparam.lr=0.2;
[net,tr]=train(net,P,T1);
estimacion=[2012 2012 2012 2012 2012 2012 2012 2012 2012
2012 2012 2012 2012; 41 42 43 44 45 46 47 48 49 50 51 52];
datos=[21.286 22.143 24.429 24 20.714 19.571 20.429
20.429 21.286 21.429 22.429 22.286];
pronostico=[];
pronostico=cat(2,pronostico,sim(net,estimacion(:,6)));
for i=7:12
X=sim(net,estimacion(:,i));
pronostico=cat(2,predicho1,X);
end
pronostico=pronostico*maximo;
matriz(:,k)=pronostico;

```

```

%disp(k);
k=k+1;
end
mf=zeros(6,1);
for k=1:6
    suma=0;
    for i=1:500
        suma=suma+matriz(k,i);
        i=i+1;
    end
    promedio=suma/500;
    mf(k,1)=promedio;
    k=k+1;
end
disp(mf)

data=load('c:\huamantlapcpsem1.txt');
P=(data(1:1190,1:2))';
T=(data(1:1190,3))';
maximo=max(data(1:1196,3));
T1=T/maximo;
matriz=zeros(6,500);
for k=1:500
net=newff(P,T1,[2,2,1],{'tansig','tansig','purelin'},'trainlm');
net=init(net);
net.trainparam.goal=1e-3;
net.trainparam.epochs=1000;
net.trainparam.lr=0.2;
[net,tr]=train(net,P,T1);
estimacion=[2012 2012 2012 2012 2012 2012 2012
2012 2012 2012 2012 2012; 41 42 43 44 45 46 47 48 49 50 51 52];
datos=[0.1 0 4.5714 6.2143 1.3571 8.5714
3.1429 4.8857 7.5714 7.5714 2.5714 4.3571];
pronostico=[];
pronostico=cat(2,pronostico,sim(net,estimacion(:,6)));
for i=7:12
X=sim(net,estimacion(:,i));
pronostico=cat(2,pronostico,X);
end
pronostico=pronostico*maximo;
matriz(:,k)=pronostico;
disp(k);
k=k+1;

```

```

end
mf=zeros(6,1);
for k=1:6
    suma=0;
    for i=1:500
        suma=suma+matriz(k,i);
        i=i+1;
    end
    promedio=suma/500;
    mf(k,1)=promedio;
    k=k+1;
end
disp(mf)

data=load('c:\tlaxcotminsem.txt');
P=(data(1:1242,1:2))';
T=(data(1:1242,3))';
maximo=max(data(1:1248,3));
T1=T/maximo;
matriz=zeros(6,500);
for k=1:500
    net=newff(P,T1,[18,16,1],{'tansig','tansig','purelin'},'trainlm');
    net=init(net);
    net.trainparam.goal=1e-3;
    net.trainparam.epochs=1000;
    net.trainparam.lr=0.2;
    [net,tr]=train(net,P,T1);
estimacion=[2012 2012 2012 2012 2012 2012 2012 2012 2012 2012
2012 2012 2012; 41 42 43 44 45 46 47 48 49 50 51 52];
datos=[5.5714 6.2143 4.2857 7.3714 5.2143 4.7143
4.7857 2.7143 2.3571 1.2143 4.7857 2.5];
pronostico=[];
pronostico=cat(2,pronostico,sim(net,estimacion(:,6)));
for i=7:12
X=sim(net,estimacion(:,i));
pronostico=cat(2,pronostico,X);
end
pronostico=pronostico*maximo;
matriz(:,k)=pronostico;
% disp(k)
k=k+1;
end
mf=zeros(6,1);

```

```

for k=1:6
    suma=0;
    for i=1:500
        suma=suma+matriz(k,i);
        i=i+1;
    end
    promedio=suma/500;
    mf(k,1)=promedio;
k=k+1;
end
disp(mf)

clear
clc
data=load('c:\tlaxcotmaxsem.txt');
P=(data(1:1242,1:2))';
T=(data(1:1242,3))';
maximo=max(data(1:1248,3));
T1=T/maximo;
matriz=zeros(6,500);
for k=1:500
    net=newff(P,T1,[10,5,1],{'tansig','tansig','tansig','purelin'},'trainlm');
    net=init(net);
    net.trainparam.goal=1e-5;
    net.trainparam.epochs=1000;
    net.trainparam.lr=0.2;
    [net,tr]=train(net,P,T1);
estimacion=[2012 2012 2012 2012 2012 2012 2012 2012 2012 2012
2012 2012 2012; 41 42 43 44 45 46 47 48 49 50 51 52];
datos=[22.714 24.429 25.214 24.671 21.5 21.714 20.571
20.714 22.429 21.714 23.857 23];
pronostico=[];
pronostico=cat(2,pronostico,sim(net,estimacion(:,6)));
for i=7:12
X=sim(net,estimacion(:,i));
pronostico=cat(2,pronostico,X);
end
pronostico=pronostico*maximo;
matriz(:,k)=pronostico;
disp(k)
k=k+1;
end
mf=zeros(6,1);

```

```

for k=1:6
    suma=0;
    for i=1:500
        suma=suma+matriz(k,i);
        i=i+1;
    end
    promedio=suma/500;
    mf(k,1)=promedio;
    k=k+1;
end
disp(mf)

data=load('c:\tlaxcopcpsemi.txt');
P=(data(1:1242,1:2))';
T=(data(1:1242,3))';
maximo=max(data(1:1248,3));
T1=T/maximo;
matriz=zeros(6,500)
for k=1:500
    net=newff(P,T1,[2,1],{'tansig','purelin'},'trainlm');
    net=init(net);
    net.trainparam.goal=1e-4;
    net.trainparam.epochs=1000;
    net.trainparam.lr=0.1;
    [net,tr]=train(net,P,T1);
estimacion=[2012 2012 2012 2012 2012 2012 2012 2012
2012 2012 2012 2012; 41 42 43 44 45 46 47 48 49 50 51 52];
datos=[0 0 0.42857 8.0286 0.14286 4.1714 1.3429
3.2 4.6857 7.6714 0.15714 1.4714];
pronostico=[];
pronostico=cat(2,pronostico,sim(net,estimacion(:,6)));
for i=7:12
    X=sim(net,estimacion(:,i));
    pronostico=cat(2,pronostico,X);
end
pronostico=pronostico*maximo;
matriz(:,k)=pronostico;
disp(k);
k=k+1;
end
mf=zeros(6,1);
for k=1:6

```

```

        suma=0;
        for i=1:500
            suma=suma+matriz(k,i);
            i=i+1;
        end
        promedio=suma/500;
        mf(k,1)=promedio;
        k=k+1;
    end
    disp(mf)

data=load('c:\espanitatminsem.txt');
P=(data(1:1190,1:2))';
T=(data(1:1190,3))';
maximo=max(data(1:1196,3));
T1=T/maximo;
matriz=zeros(6,500);
for k=1:500
    net=newff(P,T1,[10,4,1],{'tansig','tansig','purelin'},'trainlm');
    net=init(net);
    net.trainparam.goal=1e-4;
    net.trainparam.epochs=1000;
    net.trainparam.lr=0.1;
    [net,tr]=train(net,P,T1);
    estimacion=[2012 2012 2012 2012 2012 2012 2012 2012 2012 2012
    2012 2012 2012; 41 42 43 44 45 46 47 48 49 50 51 52];
    datos=[6.1429 6.4286 6.8571 5.8571 5.5714 6 7.2857
    6.1429 6.1429 5.4286 5.4286 4.7143];
    pronostico=[];
    pronostico=cat(2,pronostico,sim(net,estimacion(:,6)));
    for i=7:12
        X=sim(net,estimacion(:,i));
        pronostico=cat(2,pronostico,X);
    end
    pronostico=pronostico*maximo;
    matriz(:,k)=pronostico;
    disp(k)
    k=k+1;
end
mf=zeros(6,1);
for k=1:6
    suma=0;
    for i=1:500

```

```

        suma=suma+matriz(k,i);
        i=i+1;
    end
    promedio=suma/500;
    mf(k,1)=promedio;
    k=k+1;
end
disp(mf)

data=load('c:\espanitatmaxsem.txt');
P=(data(1:1190,1:2))';
T=(data(1:1190,3))';
maximo=max(data(1:1196,3));
T1=T/maximo;
matriz=zeros(6,500);
for k=1:500
net=newff(P,T1,[17,13,1],{'tansig','tansig','purelin'},'trainlm');
net=init(net);
net.trainparam.goal=1e-4;
net.trainparam.epochs=1000;
net.trainparam.lr=0.1;
[net,tr]=train(net,P,T1);
estimacion=[2012 2012 2012 2012 2012 2012 2012 2012
2012 2012 2012 2012; 41 42 43 44 45 46 47 48 49 50 51 52];
datos=[21.143 21.714 22.714 20.286 20.571
21.714 22.714 21 21.714 20 20.286 19.714];
pronostico=[];
pronostico=cat(2,pronostico,sim(net,estimacion(:,6)));
for i=7:12
X=sim(net,estimacion(:,i));
pronostico=cat(2,pronostico,X);
end
pronostico=pronostico*maximo;
matriz(:,k)=pronostico;
disp(k)
k=k+1;
end
mf=zeros(6,1);
for k=1:6
    suma=0;
    for i=1:500
        suma=suma+matriz(k,i);
        i=i+1;
    end
end

```

```

        end
        promedio=suma/500;
        mf(k,1)=promedio;
        k=k+1;
    end
    disp(mf)

    data=load('c:\espanitapcpsem.txt');
    P=(data(1:1190,1:2))';
    T=(data(1:1190,3))';

    maximo=max(data(1:1196,3));
    T1=T/maximo;
    matriz=zeros(6,500);
    for k=1:500
        net=newff(P,T1,[1,1,1],{'tansig','tansig','purelin'},'trainlm');
        net=init(net);
        net.trainparam.goal=1e-4;
        net.trainparam.epochs=1000;
        net.trainparam.lr=0.1;
        [net,tr]=train(net,P,T1);
        estimacion=[2012 2012 2012 2012 2012 2012 2012 2012 2012 2012
        2012 2012; 41 42 43 44 45 46 47 48 49 50 51 52];

        datos=[0 0 5.4286 0 14.714 5.6586 16.457 4.8571 9.4286 8.7143 10.286 8.1429 ];
        pronostico=[];
        pronostico=cat(2,pronostico,sim(net,estimacion(:,6)));
        for i=7:12
            X=sim(net,estimacion(:,i));
            pronostico=cat(2,pronostico,X);
        end
        pronostico=pronostico*maximo;
        matriz(:,k)=pronostico;
        disp(k)
        k=k+1;
    end
    mf=zeros(6,1);
    for k=1:6
        suma=0;
        for i=1:500
            suma=suma+matriz(k,i);
            i=i+1;
        end
    end

```

```

        promedio=suma/500;
        mf(k,1)=promedio;
        k=k+1;
end
disp(mf)

data=load('c:\carmentminsem.txt');
P=(data(1:1086,1:2))';
T=(data(1:1086,3))';
    %T1=T;
maximo=max(data(1:1086,3));
minimo=min(data(1:1086,3));
t2=T-minimo;
T1=t2/maximo;
matriz=zeros(6,500);
for k=1:500
net=newff(P,T1,[18,9,1],{'tansig','tansig','purelin'},'trainlm');
net=init(net);
net.trainparam.goal=1e-3;
net.trainparam.epochs=1000;
net.trainparam.lr=0.02;
    [net,tr]=train(net,P,T1);
estimacion=[2012 2012 2012 2012 2012 2012 2012 2012 2012 2012
2012 2012 2012; 41 42 43 44 45 46 47 48 49 50 51 52];

datos=[6 4.1429 3.5714 2.1429 4.7143 1.1429 4.1429
0.71429 -0.637 -2.199 0.857 -0.471];

pronostico=[];
pronostico=cat(2,pronostico,sim(net,estimacion(:,6)));
for i=7:12
X=sim(net,estimacion(:,i));
pronostico=cat(2,pronostico,X);
end
pronostico=(pronostico*maximo)+minimo;
matriz(:,k)=pronostico;
disp(k)
k=k+1;
end
for k=1:6
    suma=0;
    for i=1:500
        suma=suma+matriz(k,i);

```

```

        i=i+1;
    end
    promedio=suma/500;
    mf(k,1)=promedio;
    k=k+1;
end
disp(mf)

data=load('c:\carmentmaxsem.txt');
P=(data(1:1086,1:2))';
T=(data(1:1086,3))';
maximo=max(data(1:1086,3));
T1=T/maximo;
matriz=zeros(6,500);
for k=1:500
net=newff(P,T1,[9,3,1],{'tansig','tansig','purelin'},'trainlm');
net=init(net);
net.trainparam.goal=1e-3;
net.trainparam.epochs=1000;
net.trainparam.lr=0.02;
[net,tr]=train(net,P,T1);
estimacion=[2012 2012 2012 2012 2012 2012 2012 2012
 2012 2012 2012 2012; 41 42 43 44 45 46 47 48 49 50 51 52];
datos=[21.286 21.857 23 22 19.714 20 19.571
20 20.429 20.429 23.143 20.714];
pronostico=[];
pronostico=cat(2,pronostico,sim(net,estimacion(:,6)));
for i=7:12
X=sim(net,estimacion(:,i));
pronostico=cat(2,pronostico,X);
end
pronostico=pronostico*maximo;
matriz(:,k)=pronostico;
disp(k)
k=k+1;
end
mf=zeros(6,1);
for k=1:6
    suma=0;
    for i=1:500
        suma=suma+matriz(k,i);
        i=i+1;
    end
end

```

```

        end
        promedio=suma/500;
        mf(k,1)=promedio;
        k=k+1;
    end
    disp(mf)

data=load('c:\carmenpcpsem.txt');
P=(data(1:1086,1:2))';
T=(data(1:1086,3))';
maximo=max(data(1:1086,3));
T1=T/maximo;
matriz=zeros(6,500);
for j=1:500
    net=newff(P,T1,[1,1,1],{'tansig','tansig','purelin'},'trainlm');
    net=init(net);
    net.trainparam.goal=1e-3;
    net.trainparam.epochs=1000;
    net.trainparam.lr=0.02;
    [net,tr]=train(net,P,T1);
    estimacion=[2012 2012 2012 2012 2012 2012 2012 2012
    2012 2012 2012 2012; 41 42 43 44 45 46 47 48 49 50 51 52];
    datos=[0 0 1.5714 1 4.95 8.5714 5.8571 0 7.6429 5.4286 0 0.28571];

    pronostico=[];
    pronostico=cat(2,pronostico,sim(net,estimacion(:,6)));
    for i=7:12
        X=sim(net,estimacion(:,i));
        pronostico=cat(2,pronostico,X);
    end
    pronostico=pronostico*maximo;
    matriz(:,j)=pronostico;
    j=j+1;
end
mf=zeros(6,1);

for k=1:6
    suma=0;
    for i=1:500
        suma=suma+matriz(k,i);
        i=i+1;
    end
    promedio=suma/500;

```

```
mf(k,1)=promedio;  
k=k+1;  
end  
disp(mf)
```