

UNIVERSIDAD DE EL SALVADOR
FACULTAD DE INGENIERÍA Y ARQUITECTURA
ESCUELA DE INGENIERÍA ELÉCTRICA



**MODELO PREDICTIVO DE LA DEMANDA DE ENERGÍA
ELÉCTRICA EN EL CAMPUS CENTRAL DE LA
UNIVERSIDAD DE EL SALVADOR**

PRESENTADO POR:

FLORES CLAROS IGNACIO FERNANDO

HERRERA CALDERON JIMMY ADONAY

PARA OPTAR AL TITULO DE:

INGENIERO ELECTRICISTA

CIUDAD UNIVERSITARIA, OCTUBRE 2018

UNIVERSIDAD DE EL SALVADOR

RECTOR:

MSC. ROGER ARMANDO ARIAS ALVARADO

SECRETARIO GENERAL:

MSc. CRISTOBAL HERNAN RIOS BENITEZ

FACULTAD DE INGENIERIA Y ARQUITECTURA

DECANO:

ING. FRANCISCO ANTONIO ALARCON SANDOVAL

SECRETARIO:

ING. JULIO ALBERTO PORTILLO

ESCUELA DE INGENIERÍA ELÉCTRICA

DIRECTOR:

ING. ARMANDO MARTÍNEZ CALDERÓN

UNIVERSIDAD DE EL SALVADOR
FACULTAD DE INGENIERÍA Y ARQUITECTURA
ESCUELA DE INGENIERÍA ELCTRICA

Trabajo de Graduación previo a la opción al Grado de:

INGENIERO ELECTRICISTA

Título:

**MODELO PREDICTIVO DE LA DEMANDA DE ENERGÍA
ELÉCTRICA EN EL CAMPUS CENTRAL DE LA
UNIVERSIDAD DE EL SALVADOR**

Presentado por:

FLORES CLAROS IGNACIO FERNANDO

HERRERA CALDERON JIMMY ADONAY

Trabajo de Graduación Aprobado por:

Docente Asesor:

MSC. LUIS ROBERTO CHEVEZ PAZ

SAN SALVADOR, OCTUBRE 2018

Trabajo de Graduación Aprobado por:

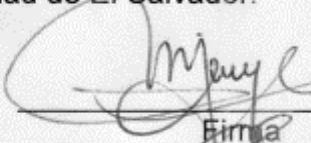
Docente Asesor:


MSC. LUIS ROBERTO CHEVEZ PAZ

ACTA DE CONSTANCIA DE NOTA Y DEFENSA FINAL

En esta fecha, jueves 27 de septiembre de 2018, en la Sala de Lectura de la Escuela de Ingeniería Eléctrica, a las 10:00 horas, en presencia de las siguientes autoridades de la Escuela de Ingeniería Eléctrica de la Universidad de El Salvador:

1. Ing. Armando Martínez Calderón
Director


Firma


MSC. José Wilber Calderón Urrutia
Secretario
Firma

Y, con el Honorable Jurado de Evaluación integrado por las personas siguientes:

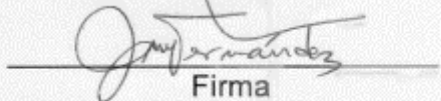
- MSC. LUIS ROBERTO CHÉVEZ PAZ
(Docente Asesor)


Firma

- MSC. JORGE ALBERTO ZETÍNO CHICAS


Firma

- ING. JOSE MIGUEL HERNÁNDEZ


Firma

Se efectuó la defensa final reglamentaria del Trabajo de Graduación:

MODELO PREDICTIVO DE LA DEMANDA DE ENERGÍA ELÉCTRICA EN EL CAMPUS CENTRAL DE LA UNIVERSIDAD DE EL SALVADOR

A cargo de los Bachilleres:

- FLORES CLAROS IGNACIO FERNANDO

- HERRERA CALDERÓN JIMMY ADONAY

Habiendo obtenido en el presente Trabajo una nota promedio de la defensa final: 8.3

(ocho punto tres)

Tabla de contenido

Tabla de contenido.....	6
INTRODUCCION.....	11
OBJETIVOS:.....	12
ALCANCES:.....	12
LIMITACIONES:	12
ANTECEDENTES:	13
PLANTEAMIENTO DEL PROBLEMA:	13
JUSTIFICACIÓN:	14
1 GENERALIDADES	15
1.1 CONCEPTOS.....	15
1.2 FACTORES QUE INFLUYEN EN EL PRONÓSTICO DE DEMANDA DE ENERGÍA ELÉCTRICA	18
1.3 CLASIFICACIÓN DE LOS MÉTODOS DE PRONÓSTICO DE DEMANDA DE ENERGÍA ELÉCTRICA	20
1.3.1 PREDICIONES A LARGO Y MEDIO PLAZO	20
1.3.2 PREDICIONES A CORTO PLAZO	20
1.3.3 OTRA CLASIFICACIÓN.....	20
1.4 CRITERIOS DE VALIDACIÓN DEL MÉTODO Y ANÁLISIS.....	21
2 CAPITULO II. METODOS DE ANALISIS PARA ESTIMAR LA DEMANDA DE ENERGIA ELECTRICA.....	23
2.1 REDES NEURONALES.....	23
2.1.1 Neurona biológica	23
2.2 FUNCIONES BASE Y DE ACTIVACIÓN	24
2.3 Topología de redes neuronales.....	26
2.3.1 Según la estructura en capas.....	26
2.3.2 Según el flujo de datos en la red	26
2.4 Mecanismos de aprendizaje	28
2.4.1 El entrenamiento supervisado	28
2.5 Etapas en el diseño de una red neuronal	29
2.5.1 Procesamiento de los datos	30
2.5.2 Diseño de la red.....	30
2.5.3 Implementación	30

2.5.4	Validación	30
2.6	PERCEPTRÓN MULTICAPA	31
2.7	PASOS PARA LA IMPLEMENTACIÓN DE REDES NEURONALES EN EL PRONÓSTICO DE DEMANDA DE ENERGÍA ELÉCTRICA.....	32
2.8	Algoritmo para la creación de redes neuronales	32
2.9	Métodos de gradiente descendente	33
2.10	Estructura del perceptron.....	33
2.11	El algoritmo Backpropagation	35
2.11.1	Deducción del algoritmo Backpropagation	35
3	PRONOSTICO A TRAVES DEL ANALISIS DE SERIES TEMPORALES.	39
3.1.1	Series temporales y procesos estocásticos univariantes.	39
3.1.1.1	Introducción.....	39
3.1.2	Clases de procesos estocásticos	39
3.1.3	Proceso de ruido blanco.....	40
3.1.4	Proceso lineal general.....	41
3.2	Identificación de una serie de tiempo	43
3.2.1	Pruebas de estacionariedad.	43
3.2.2	FUNCION DE AUTOCORRELACION PARCIAL	45
3.2.3	Identificación de modelos estacionarios a partir de las fas y fap.....	46
3.2.4	Procesos estocásticos no estacionarios.....	46
3.3	Construcción del modelo estocástico.	48
3.3.1	Introducción.....	48
3.3.2	Enfoque iterativo para la construcción de modelos.....	48
3.3.3	IDENTIFICACION DE MODELO.....	50
3.3.4	Uso de las funciones de autocorrelación y autocorrelación parcial para la identificación del orden del proceso	50
3.3.5	Estimación de los parámetros del modelo.	51
3.3.6	Relación entre Autocorrelaciones estimadas y teóricas.....	51
3.4	PRONÓSTICOS	52
3.4.1	Pronósticos con el modelo	52
3.4.2	Tres formas explicitas para el modelo	52
3.4.3	Tres formas básicas para el pronóstico.	53

3.4.4	Uso de coeficientes ψ	54
3.5	Modelos estacionales.....	55
3.5.1	¿Cómo identificar la estacionalidad?	55
3.5.2	Operador Simplificador $1 - Bs$	56
3.5.3	Determinación de la diferenciación estacional.....	58
3.5.4	Notacion de Modelos estacionales ARIMA y estacionalidad multiplicativa.	58
3.6	Metodología Box-Jenkins para ajustar los modelos a series reales.....	58
3.6.1	Algoritmo para la aplicación del método.	58
3.6.2	Determinación de la transformación para estabilizar la varianza.....	61
3.6.3	Determinación de la transformación para estabilizar la media del proceso estacional	62
3.6.4	Diagnos del modelo.....	62
3.6.5	Contraste de media cero	63
3.6.6	Contraste de varianza estable (homocedasticidad)	63
3.6.7	Contraste de normalidad	64
3.7	Valores atípicos	64
3.7.1	Atípicos aditivos (AO)	65
3.7.2	Atípicos innovativos (IO).....	65
4	Redes neuronales en Matlab	67
4.1	Creación de la red	67
4.2	Simulación de la red.....	67
4.3	Entrenamiento.....	68
4.4	Caso de estudio	70
5.	Modelo definitivo para el pronóstico de demandas.....	91
5.1	METODOLOGIA GENERAL:.....	91
5.2	OBTENCION DE MODELOS.....	94
5.2.1	Transformación para heterocedasticidad de la serie	94
5.2.2	Determinación de los órdenes de diferenciación d y D	95
5.2.3	Identificación de posibles modelos.....	96
5.2.4	Refinado de modelo	97
5.3	Prueba de hipótesis para los residuos de los modelos definidos	98
5.3.1	Hipótesis de media cero.....	98

5.3.2	Hipótesis de varianza constante	98
5.3.3	Contraste de correlaciones - Test de Ljung-Box.....	99
5.3.4	Prueba de normalidad	100
6	Conclusiones	101
7	Referencias.....	102
8	ANEXOS	103

INDICE DE FIGURAS

Figura 1.1.	Factores que influyen en el pronóstico de demanda de energía eléctrica.....	18
Figura 1.2.	metodos de validación	21
Figura 2.1	Relación entre una neurona biológica y artificial.....	24
Figura 2.2.	Función sigmoideal.....	25
Figura 2.3.	Red unidireccional propagación hacia adelante.....	27
Figura 2.4.	Propagación hacia atrás.....	27
Figura 2.5.	Esquema general de un Perceptrón Multicapa.....	31
Figura 2.6.	Metodología para pronóstico utilizando Redes Neuronales Artificiales.....	32
Figura 2.7.	Para creación de redes neuronales.....	32
Figura 2.8.	Estructura del perceptron simple	34
Figura 2.9.	Estructura del perceptron múltiple.....	34
Figura 3.1.	44
Figura 3.2.	Distintos casos de estacionalidad.....	48
Figura 3.3.	Comportamiento de la función de predicción estacional para varias alternativas del operador general autor regresivo estacional.....	57
Figura 3.4.	Metodología de Box-Jenkins para la selección de un modelo adecuado	60
Figura 4.1	resultado de la red neuronal.....	74
Figura 4.2	resultado del entrenamiento.....	74
Figura 4.3	resultados de la simulación del entrenamiento y prueba de simulación.....	75
Figura 4.4	se compara la salida con el pronóstico que se muestra en la figura 4.3	75
Figura 4.5	gráficas del pronóstico y la salida que se quiere obtener.....	76
Figura 4.6	resultados de la red neuronal para nuevos valores.....	78
Figura 4.7	resultado del entrenamiento de la red con nuevas entradas y salidas.....	78
Figura 4.8	resultados de la simulación para predecir nuevos valores diferentes lunes.....	79
Figura 4.9	salida que se quiere obtener.....	79
Figura 4.10	resultados de los errores que se obtienen con la salida respecto a la predicción.....	80
Figura 4.11	grafica de predicción vrs la salida nueva.....	80
Figura 4.12	grafica obtenidas de diferentes días lunes para medidor 3 ubicado en la facultad de humanidades.....	82
Figura 4.13	resultados de la red neuronal para la facultado de odontología.....	84
Figura 4.14	graficas obtenidas para diferentes martes de la facultad de odontología.....	84
Figura 4.15	red neuronal creada para la faculta de odontología medidor 3 para un día miércoles.....	86

Figura 4.16 graficas obtenidas para diferentes miércoles pronostico y predicción.	86
Figura 4.17 resultados de la red neuronal para un día lunes para la facultad de periodismo.	88
Figura 4.18 resultados de las gráficas para diferentes lunes de la facultad de periodismo.	88
Figura 4.19 resultados de la red neuronal para un día martes facultad de química.	90
Figura 4.20 graficas de un día martes de pronostico y predicción para un día martes facultad de química.....	90
Figura 5.1. Graficas punta, valle y resto.	94
Figura 5.2. Graficas obtenidas de la FAS y FAP de las series.....	95
Figura 5.3. FAS y FAP luego de aplicar una diferencia regular y una estacional.	96
Figura 5.4. Graficas de residuos del modelo (ruido blanco característico del proceso).	97
Figura 5.5. Resultados luego de retirar atípicos de la serie original.	98
Figura 5.6. Graficas de auto correlación simple y parcial de los residuos.	99

INDICE DE TABLAS

Tabla 1.1 clasificación de los distintos métodos de estimación de la demanda.....	20
Tabla 3.1 Estimacion de los momentos de procesos estacionarios.....	40
Tabla 3.2 Comportamiento de las FAS y FAP para la identificacion de procesos e identificacion de condiciones para la realizacion válida del modelo.....	46
Tabla 3.3 Identificación de modelos por medio de sus funciones de autocorrelación simple y autocorrelación parcial.	51
Tabla 4.1 parámetros de entrenamiento.....	69
Tabla 4.2 factores de variación horaria para el día lunes del medidor 2 ubicado en la facultad de humanidades.....	72
Tabla 4.3 días de entrenamiento y de prueba.	73
Tabla 4.4 datos de validación de la red son nuevos datos para ver cómo se comporta la red con diferentes días.....	77
Tabla 4.5 datos de variación horaria del día lunes medidor 3 de la facultad de humanidades.	81
Tabla 4.6 factores de variación horaria para el día martes para el medidor 2 de la facultad de odontología.	83
Tabla 4.7 factores de variación horaria para el día miércoles corresponden al medidor 3 de la facultad de odontología.	85
Tabla 4.8 factores de variación horaria para un día lunes corresponde al medidor de la facultad de periodismo.	87
Tabla 4.9 factores de variación horaria para un día martes facultad de química.....	89
Tabla 5.1. Resultados de la regresión para determinar si existe relación media-varianza.....	95
Tabla 5.2. Determinación de coeficientes para los posibles modelos identificados.	96
Tabla 5.3. Rendimiento de cada modelo identificado.	97
Tabla 5.4. Datos atípicos identificados en las gráficas de residuos.	97
Tabla 5.5. Resultados de prueba de hipótesis “media de los residuos igual a cero”.....	98
Tabla 5.6. Resultados de prueba de hipótesis “varianza de los residuos constante”.....	99
Tabla 5.7. Resultados de prueba de hipótesis “incorrelacion de residuos”.	100
Tabla 5.8. Resultados de prueba de hipótesis “distribución normal de los residuos”.	100

INTRODUCCION.

El pronóstico de la demanda de energía eléctrica es un procedimiento sistemático que permite definir cuantitativamente la demanda futura y por ser de vital importancia debe procurarse que dicha información sea lo más exacta posible, sin olvidar la presencia de incertidumbres. Dependiendo del período de tiempo que se adopte en el análisis se podría considerar dentro del pronóstico tres técnicas de predicción: de corto, mediano y largo plazo, todo depende del tipo de análisis a ejecutarse, lo cual será detallado más adelante. Así mismo el pronóstico de demanda puede referirse a la potencia pico, a la energía total demandada o al comportamiento de la curva de carga. El pronóstico de la demanda de energía eléctrica puede ser de dos formas, energía (utiliza demanda de electricidad en unidades de Wh) y potencia (Demanda eléctrica en unidades de W).

El objetivo principal del sistema eléctrico de potencia es suministrar el servicio de energía de manera continua, segura y confiable, por esta razón debe adaptarse continuamente a la demanda. Sin embargo, los sistemas eléctricos no son sistemas adaptativos, y para suplir dicha demanda requiere contar con un estimado del valor demandado por dicho sistema para poder realizar las modificaciones y cumplir con las exigencias del servicio. La predicción de la demanda es una actividad de mucha importancia para las empresas destinadas a la generación de energía eléctrica. Una correcta predicción de la demanda de energía juega un papel importante ya que puede servir para la toma de decisiones de las empresas eléctricas respecto a qué cantidad y a qué precios sería conveniente realizar transacciones con otras empresas del sector. Si se llegara a tener un pronóstico subestimado de la demanda de energía, podría provocar que el sistema no esté preparado ante un alza de la demanda, y como consecuencia, se estaría comprometiendo la confiabilidad y la seguridad del sistema, así como, en un caso extremo, tener que realizar un racionamiento de la demanda. Es por ello que en este documento se analizan los métodos para el pronóstico de demanda a corto y mediano plazo, evaluando sus ventajas y desventajas que permitan tomar decisiones que puedan prevenir sobrecargas, evitar fallas de los equipos o apagones que implican grandes pérdidas y de esta manera proporcionar un mejor servicio.

OBJETIVOS:

GENERAL:

- ✓ Aplicar diversos métodos que nos permitan determinar un pronóstico de la demanda de energía en el campus central de la Universidad de El Salvador.

ESPECIFICOS:

- ✓ Determinar los factores que inciden en el pronóstico de demanda de energía eléctrica
- ✓ Identificar ventajas y desventajas de cada método en mención según su horizonte de tiempo.
- ✓ Realizar la validación estadística de los métodos que cumplan con los criterios de aceptación.

ALCANCES:

- ✓ Comprobar la confiabilidad y precisión del método a aplicar en la pronosticación de corto y mediano plazo.
- ✓ Realizar mediciones para validar algunas instalaciones.
- ✓ Investigar una aplicación scilab u octave que maneje el algoritmo empleado.
- ✓ Comparar los métodos antes mencionados, y llegar a una conclusión sobre su factibilidad de aplicación para hacer un uso eficiente de las fuentes energéticas.

LIMITACIONES:

- ✓ Para realizar las predicciones se necesitan datos históricos de consumo de energía, estos datos históricos están limitados al consumo del campus central de la Universidad de El Salvador.
- ✓ se obtendrán mediciones del perfil de carga por medio de la distribuidora.
- ✓ En caso de no obtener acceso a las instalaciones no se podrá realizar la validación de datos

ANTECEDENTES:

El pronóstico de la demanda de energía eléctrica es un procedimiento sistemático que permite definir cuantitativamente la demanda futura y por ser de vital importancia debe procurarse que dicha información sea lo más exacta posible, sin olvidar la presencia de incertidumbres. Dependiendo del período de tiempo que se adopte en el análisis se podría considerar dentro del pronóstico tres técnicas de predicción: de corto, mediano y largo plazo, todo depende del tipo de análisis a ejecutarse y de la complejidad en la modelación. Además, para satisfacer la carga futura se debe considerar los proyectos de nuevas centrales de generación. Así mismo el pronóstico de demanda puede referirse a la potencia pico, a la energía total demandada o al comportamiento de la curva de carga. El pronóstico de la demanda de energía eléctrica puede ser de dos formas, energía (utiliza demanda de electricidad en unidades de Wh) y potencia (Demanda eléctrica en unidades de W), usualmente los operadores de red realizan proyecciones de potencia horaria (energía) en el orden de los MWh.

PLANTEAMIENTO DEL PROBLEMA:

El objetivo principal del sistema eléctrico de potencia es suministrar el servicio de energía de manera continua, segura y confiable, por esta razón debe adaptarse continuamente a la demanda. Sin embargo, los sistemas eléctricos no son sistemas adaptativos, y para suplir dicha demanda requiere contar con un estimado del valor demandado por dicho sistema para poder realizar las modificaciones y cumplir con las exigencias del servicio.

Un pronóstico más acertado reduce el riesgo de realizar re-despachos donde se deban usar unidades de generación más costosas o, por el contrario, se deba prescindir de generaciones programadas anteriormente. Es por ello que en este documento se analizan los métodos para el pronóstico de demanda a mediano plazo, evaluando sus ventajas y desventajas que permitan tomar decisiones que puedan prevenir sobrecargas, evitar fallas de los equipos o apagones que implican grandes pérdidas y de esta manera proporcionar un mejor servicio

JUSTIFICACIÓN:

- ✓ La importancia del pronóstico de demanda de energía eléctrica se incrementa en la medida que los cumplimientos de los objetivos trazados dependan lo menos posible del azar, disminuyendo así la incertidumbre sobre los resultados de las decisiones tomadas. Por esto, es de vital importancia conocer a priori el crecimiento de la demanda de energía eléctrica, de una manera segura y confiable lo cual se aproxime lo más cercano a la realidad.
- ✓ El pronóstico de la demanda de electricidad es una herramienta fundamental para la toma de decisiones operativas y estratégicas, cuya falta de precisión puede traer altos costos económicos.

1 GENERALIDADES

1.1 CONCEPTOS

Demanda. Es la cantidad de potencia que un consumidor utiliza en un periodo de tiempo. La demanda de una instalación eléctrica en los terminales receptores, es tomada como un valor medio en un intervalo determinado. El período durante el cual se toma el valor medio se denomina intervalo de demanda. La duración que se fije en este intervalo dependerá del valor de demanda que se desee conocer. Para establecer una demanda es indispensable indicar el intervalo de demanda ya que sin él no tendría sentido práctico. La demanda se puede expresar en kVA, kW, kVAR, A, etc. La variación de la demanda en el tiempo para una carga dada origina el ciclo de carga que es una curva de carga (Demanda vs tiempo). (1)

Demanda o carga máxima. Corresponde a la carga mayor que se presenta en un sistema o instalación en un período de trabajo previamente establecido. Para establecer la demanda máxima se debe especificar el intervalo de demanda para medirla. La carga puede expresarse en p.u de la carga pico del sistema. (1)

Carga instalada. Es la suma de potencias nominales de aparatos y equipos de consumo que se encuentran conectados a un sistema o a parte de él, se expresa generalmente en kVA, MVA, kW o MW. (1)

Capacidad instalada. Corresponde a la suma de las potencias nominales de los equipos (transformadores, generadores), instalados a líneas que suministran la potencia eléctrica a las cargas o servicios conectados. Es llamada también capacidad nominal del sistema. (1)

Demanda diversificada o coincidente. Es la suma de varias cargas en un momento determinado. Se considera el mismo intervalo de tiempo. (1)

$$D_{Div} = \sum_{i=1}^n D_i x t_i \quad (1.1)$$

Dónde:

D_i = Demanda de energía eléctrica de usuario

t_i = Periodo en que se registra la demanda

i = numero de usuarios

Demanda máxima no coincidente. Es la suma de la demanda máxima de cada usuario sin importar que ocurran o no al mismo tiempo. (1)

Curvas de carga diaria. Las curvas de carga diaria están formadas por los picos obtenidos en intervalos de una hora para cada hora del día. Las curvas de carga diaria dan una indicación de las características de la carga en el sistema, para el sector residencial, comercial o industrial y de la forma en que se combinan para producir el pico. (1)

Curvas de carga anual. Las curvas de carga anual están formadas por los valores de la demanda a la hora pico en cada mes, permiten una visualización de los crecimientos y variaciones de los picos mensuales y anuales. El análisis de las causas de estas variaciones debe conducir a conclusiones prácticas sobre el comportamiento del sistema y los factores que lo afectan. (1)

Factor de demanda. El factor de demanda en un intervalo de tiempo t , de una carga, es la razón entre la $Carga_{MAX}$ y la carga total instalada. El factor de demanda por lo general es menor que 1. En el caso de máquinas eléctricas recibe el nombre de factor de uso o de utilización. (1)

$$F_{Dem} = \frac{\Delta Carga_{maxima}}{\Delta Carga\ instalada\ total} \leq 1 \quad (1.2)$$

Factor de carga. Es la razón entre la D_{Prom} en un intervalo de tiempo dado y la D_{max} observada en el mismo intervalo de tiempo. (1)

$$F_c = \frac{D_{prom}}{D_{max}} \quad 0 < F_c \leq 1 \quad (1.3)$$

Factor diversidad. Es la razón entre la sumatoria de las demandas máximas individuales y la demanda máxima del conjunto o grupo de usuarios (llamada también demanda máxima coincidente). (1)

$$F_{Div} = \frac{\sum D_{Max\ individuales}}{D_{Max\ promedio\ por\ consumidor\ para\ n\ consumidores}} = \frac{D_{Max\ no\ coincidentes}}{D_{Max\ coincidentes}} \quad (1.4)$$

Serie de tiempo: Una serie temporal o cronológica es una secuencia de datos, observaciones o valores, medidos en determinados momentos y ordenados cronológicamente. (2)

Proceso estocástico: En proceso estocástico es un concepto matemático que sirve para caracterizar una sucesión de variables aleatorias (estocásticas) que evolucionan en función de otra variable, generalmente el tiempo.

Media: La media aritmética (también llamada promedio o media) de un conjunto finito de números es el valor característico de una serie de datos cuantitativos, objeto de estudio que parte del principio de la esperanza matemática o valor esperado.

Varianza σ : Es una medida de dispersión definida como la desviación estándar al cuadrado de la desviación de dicha variable respecto a su media.

Varianza muestral s: Caso especial de la varianza poblacional, en el cual solo se toma una muestra grande del total de datos para determinar la dispersión de la serie temporal.

Covarianza (Autocovarianza): es un valor que indica el grado de variación conjunta de dos variables aleatorias respecto a sus medias (relación de variación o tendencia mutua).

Correlación (Autocorrelación): indica la fuerza y la dirección de una relación lineal y proporcionalidad entre dos variables estadísticas.

Límites de confianza: se llama límites o intervalo de confianza a un par o varios pares de números entre los cuales se estima que estará cierto valor desconocido con una determinada probabilidad de acierto.

FAS: función que presenta de forma gráfica las autocorrelaciones entre los valores de una serie de tiempo respecto al valor en un origen.

FAP: función que representa en forma gráfica a cada uno de los k-ésimos coeficientes de cada uno de los procesos autorregresivos que dan origen a la k-ésima autocorrelación de la FAS.

Estacionariedad: Se dice que un proceso estocástico es estacionario si sus propiedades estadísticas son invariantes ante una traslación del tiempo.

Estacionalidad: Es la variación periódica y predecible de la misma con un periodo s .

Tendencia: Variación determinística de un proceso estocástico en el tiempo.

Estacionario en sentido estricto: Se dice que un proceso es estrictamente estacionario si sus propiedades (media, varianza, distribución de probabilidad), no se ven afectadas por cambios de origen temporal, esto es, cuando al realizar un mismo desplazamiento en el tiempo de todas las variables de cualquier distribución conjunta finita, resulta que esta distribución no varía.

Estacionario en sentido débil: La estacionariedad en sentido estricto conlleva, entre otras cuestiones, la necesidad de conocer la función de distribución del proceso estocástico, lo cual no es posible. Por ello el concepto de estacionariedad que se asume es en un sentido débil, el cual se obtiene cuando la media y varianza son constantes e independientes del

tiempo, y que la autocovarianza γ , por tanto, los coeficientes de autocorrelación sólo dependen del desfase, pero no del momento de tiempo.

Parsimonioso: Modelo que considera solamente la cantidad mínima de variables necesarias a tomar en cuenta.

Valor P (nivel observado de significancia): Es el mínimo nivel de significancia en el cual H_0 sería rechazada cuando se utiliza un procedimiento de prueba especificado con un conjunto dado de información. Una vez que el valor P se haya determinado, la conclusión en cualquier nivel α particular resulta de comparar P con α ; si $P \leq \alpha$ se rechaza H_0 , y si $P > \alpha$ no se rechaza H_0 .

1.2 FACTORES QUE INFLUYEN EN EL PRONÓSTICO DE DEMANDA DE ENERGÍA ELÉCTRICA

La demanda de energía eléctrica está sujeta a alteraciones debido a un gran número de factores los cuales inciden directamente en su proyección. En la figura 1.1 se hace una breve explicación de los factores que se deben tener en cuenta en la proyección de demanda de energía eléctrica.

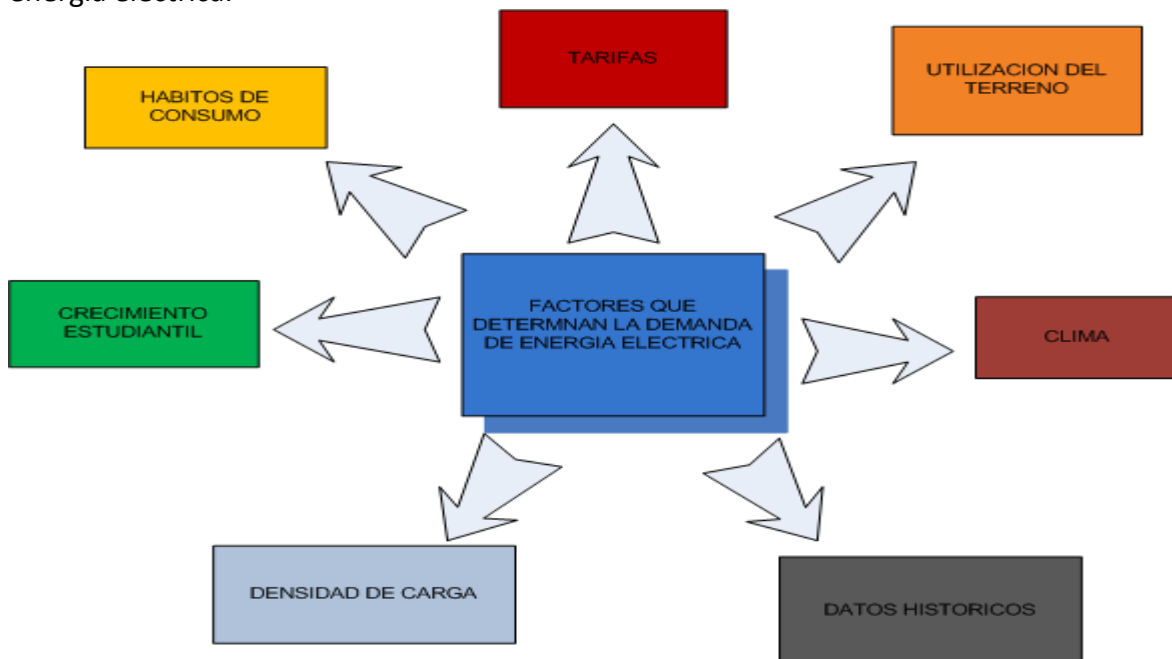


Figura 1.1. Factores que influyen en el pronóstico de demanda de energía eléctrica.

(2) Fuente: Basado en GÖNEN, Turan. Electric Power Distribution System Engineering. McGraw-Hill, 1986.

Clima los factores climáticos inciden el comportamiento de la demanda de electricidad. El Salvador es un país que se encuentra en el trópico por lo que se desarrolla solo 2 estaciones climáticas que son periodos secos o lluviosos, pero su relieve le permite diversas temperaturas.

Hábitos de consumo. Este factor abarca el comportamiento de la población en cuanto al consumo de electricidad debido a la carencia de cultura de ahorro o consumo eficiente del servicio, frecuencia de utilización de dispositivos de cómputo y comunicación principalmente, también incluye las conductas adoptadas por los consumidores en días festivos, laborales, fines de semana, fines de semana con festivos, semana santa o vacaciones.

Utilización de terrenos. Este aspecto es importante pues se parte del hecho que al construirse o utilizarse más terrenos entonces habrá un incremento de la demanda de electricidad, en este factor se relacionan los datos de oferta y demanda del suelo para crear más edificios, laboratorios entre otros y posteriormente se transforma en información de demanda eléctrica, a través de curvas de carga características.

Densidad de carga. Este parámetro nos indica cuánta es la carga por unidad de área. Es frecuentemente útil para medir las necesidades eléctricas de un área determinada. Se puede medir en kVA por metro cuadrado y se define como la carga instalada por unidad de área; el término voltio-amperio por pie cuadrado es usado también al referirse a la densidad de la carga, sin embargo, es limitado a edificios comerciales o plantas industriales. Conociendo la densidad de carga y el área de la sección en estudio, se puede conocer el valor de la carga instalada. (1)

Crecimiento Estudiantil. En la medida en que aumenta la población, se incrementa el número de consumidores de electricidad quienes requieren del uso de la misma para sus electrodomésticos y dispositivos electrónicos. Además, este crecimiento demográfico incide en el aumento de fabricación de productos en general, para satisfacer la demanda de los mismos y en esta misma relación se incrementa el consumo de energía eléctrica dado que estos precisan mayoritariamente de este recurso en su proceso de elaboración.

Datos históricos. La información histórica es de vital importancia en la proyección de demanda de electricidad pues con estos podemos generar patrones de consumo, analizar su comportamiento año tras año y proyectarlo al futuro.

Tarifas. Para los usuarios regulados o residenciales este factor es indiferente puesto que el consumo es del orden de los cientos de kWh y están obligados a pagar la energía según la tarifa dispuesta por la empresa de energía y regulada por las empresas que distribuyen la energía eléctrica. (3) (4) (5)

1.3 CLASIFICACIÓN DE LOS MÉTODOS DE PRONÓSTICO DE DEMANDA DE ENERGÍA ELÉCTRICA

Horizonte de tiempo	Tipo de demanda eléctrica	Técnicas matemáticas
<ul style="list-style-type: none">• Corto plazo• Mediano plazo• Largo plazo	<ul style="list-style-type: none">• Energía (Wh)• Potencia (W)	<ul style="list-style-type: none">• Estadísticas• Inteligencia artificial

Tabla 1.1 clasificación de los distintos métodos de estimación de la demanda.

Según su horizonte de tiempo. El pronóstico de demanda de energía eléctrica según su horizonte de tiempo se clasifica en corto, mediano y largo plazo, se dispone de esta manera de acuerdo a su aplicación en la operación de las unidades de generación y el despacho económico, planeación del uso de energéticos y programación del mantenimiento de redes entre otras. (3)

1.3.1 PREDICIONES A LARGO Y MEDIO PLAZO

Una predicción precisa de la demanda eléctrica a medio-largo plazo juega un papel fundamental en la planificación de sistemas de potencia eléctrica. Estas **predicciones a largo plazo** se encargan de la predicción de la demanda con un tiempo suficiente de antelación como para poder efectuar una planificación energética a medio-largo plazo, afectando de manera activa a la planificación para el desarrollo de nuevas instalaciones de generación, la compra de unidades de generación y el desarrollo de la red de transporte y distribución.

La exactitud de la predicción a medio-largo plazo tiene un impacto significativo en el desarrollo de la planificación de los sistemas de generación y distribución. Una sobreestimación de la demanda conlleva en un incremento sustancial de la inversión para la construcción de instalaciones con excesiva potencia, en cambio, una subestimación implica el descontento de los consumidores (por cortes de suministro). El rango de tiempo para las **predicciones a medio y largo plazo** va desde unas pocas semanas hasta unos cuantos años, por desgracia, es difícil tener una buena exactitud en las predicciones con periodos de tanto tiempo.

1.3.2 PREDICIONES A CORTO PLAZO

La mayoría de los métodos de predicción, están dedicados a la **predicción a corto plazo**, desde unos pocos minutos hasta 24 horas. Las predicciones a muy corto plazo, de minuto en minuto, se utilizan para una asignación satisfactoria de cargas que beneficie a las unidades de generación. La predicción a corto plazo que oscila entre una hora y un día es valiosa para la gestión de las unidades de consumo energético, esta previsión permitirá planificar y controlar la puesta en marcha y el apagado de las unidades de generación, a su vez permite conocer si estas unidades de generación son capaces de asumir las variaciones de demanda provenientes de la unidad consumidora.

1.3.3 OTRA CLASIFICACIÓN

Aparte del tipo de predicciones relacionadas con el periodo de tiempo predicho también podemos distinguir las predicciones según se quiera una predicción exacta de la potencia

consumida en cada instante de tiempo definido **predicción horaria**, o solo interesen los puntos críticos de consumo de potencia, es decir, unos puntos clave que den toda la información necesaria, este tipo de predicción es conocida como **predicción por picos**.

1.4 CRITERIOS DE VALIDACIÓN DEL MÉTODO Y ANÁLISIS

Existen diferentes tipos de validez estadística que son relevantes para la investigación y la experimentación. Cada una es importante para que el experimento brinde predicciones precisas y arroje conclusiones válidas. Algunas de estas son:

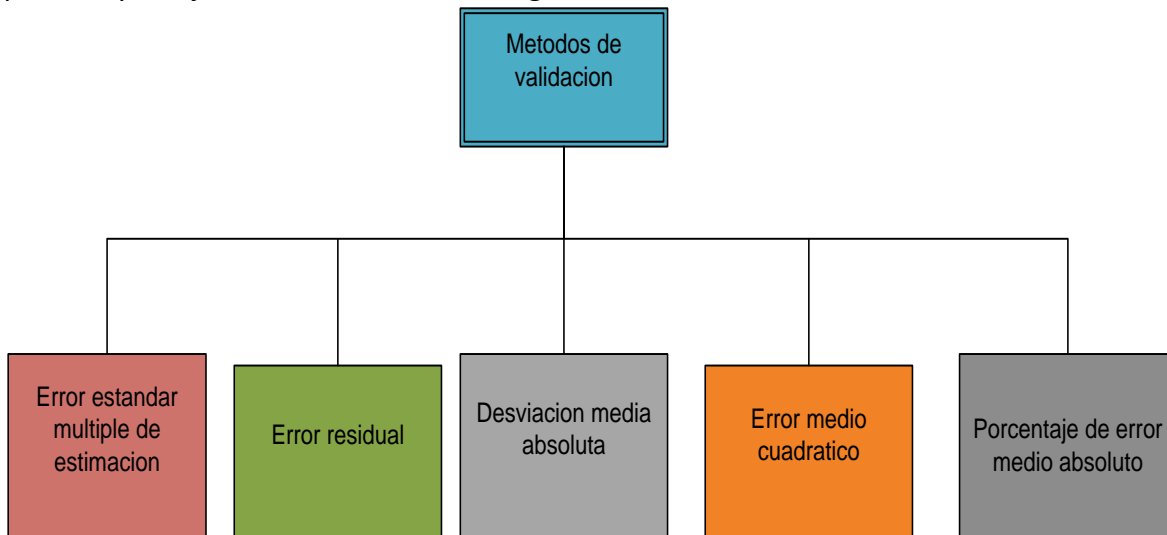


Figura 1.2. metodos de validación

El error estándar de la regresión múltiple

Es una medida de dispersión la estimación se hace más precisa conforme el grado de dispersión alrededor del plano de regresión se hace más pequeño. Para medirla se utiliza la fórmula:

$$S_{xy} = \sqrt{\frac{\sum(Y - \hat{Y})^2}{n - m - 1}} \quad (1.5)$$

Dónde:

Y = Es el valor observado de Y en los datos

\hat{Y} = Es el valor estimado de Y de la ecuación de regresión

n = Número de datos

m = Número de variables independientes X

Error residual

Se define error residual como la diferencia entre el valor del pronosticado y lo que realmente ocurrió en dicho período. A partir del concepto que la demanda tiene un componente aleatorio, todos los pronósticos contienen con certeza algún error. (4)

$$e_{re} = Y - \hat{Y} \quad (1.6)$$

Dónde:

Y: Es la demanda real u observada en el periodo

\hat{Y} : Demanda pronosticada para el periodo

Desviación media absoluta

La desviación media absoluta es una medida que se utiliza para calcular cuánto varían de su media los valores de un conjunto de datos. También se conoce como desviación media o desviación absoluta media. (4) (3)

$$MAD = \frac{\sum |Y - \hat{Y}|}{n} \quad (1.7)$$

Dónde:

Y: Demanda real u observada

\hat{Y} : Demanda pronosticada

n: numero de datos.

MSE (Mean Square Error - Error Medio Cuadrático).

Cada error o residual se eleva al cuadrado; luego, estos valores se suman y se divide entre el mismo número de observaciones. Este enfoque penaliza los errores mayores de pronóstico ya que eleva cada uno al cuadrado. Esto es importante pues en ocasiones pudiera ser preferible una técnica que produzca errores moderados a otra que por lo regular tenga errores pequeños, pero que ocasionalmente arroje algunos en extremo grandes. Idealmente, se desea que el resultado del EMC sea 0, lo que indicaría un pronóstico perfecto. En general, mientras más pequeño sea el valor de EMC, mejor es el modelo de pronóstico. La diferencia se produce debido a la aleatoriedad o porque el estimador no tiene en cuenta la información que podría producir una estimación más precisa. (4) (6)

$$EMC = \frac{\sum_{t=1}^n (\hat{Y}_t - Y_t)^2}{n} \quad (1.8)$$

MAPE (Mean Absolute Percentage Error- Porcentaje de Error Medio Absoluto).

En ocasiones, resulta más útil calcular los errores de pronóstico en términos de porcentaje y no en cantidades. El Porcentaje de Error Medio Absoluto (PEMA) se calcula encontrando el error absoluto en cada período, dividiendo éste entre el valor real observado para ese período y después promediando estos errores absolutos de porcentaje. Este enfoque es útil cuando el tamaño o magnitud de la variable de pronóstico es importante en la evaluación de la precisión del pronóstico. El PEMA proporciona una indicación de qué tan grandes son los errores de pronóstico comparados con los valores reales de la serie. También se puede utilizar el PEMA para comparar la precisión de la misma u otra técnica sobre dos series completamente diferentes. (6) (3)

$$MAPE = \frac{\sum_{t=1}^n \frac{|Y_t - \hat{Y}_t|}{Y_t}}{n} \quad (1.9)$$

2 CAPITULO II. METODOS DE ANALISIS PARA ESTIMAR LA DEMANDA DE ENERGIA ELECTRICA.

2.1 REDES NEURONALES

2.1.1 Neurona biológica

El cerebro humano puede ser visto como una entidad de procesamiento de la información. Este recibe información de un entorno externo por medio de los sentidos y los procesa para formar modelos internos a partir de fenómenos externos. El cerebro es capaz de ajustar estos modelos, así como de interpolarlos y extrapolarlos para adaptarse a nuevas situaciones con mayor agilidad que puede hacer que este tome decisiones más confiables e incluso a reconocimiento de patrones, entendimientos de conceptos y realización de predicciones a partir de una información parcial o incompleta. (7)

Las unidades de procesamiento local de información del cerebro son las neuronas, las cuales forman un gran número de redes paralelas en la corteza del cerebro y son fundamentales para esta actividad. La corteza es una delgada capa externa del cerebro que contiene un gran número de neuronas, entre 100 y 500 billones. El cerebro tiene una gran variedad de neuronas especializadas, y dependiendo del tipo, cada neurona puede enviar señales a cualquiera de cientos de miles de otras neuronas. Ahora se sabe que una repetida excitación de neuronas conduce a un crecimiento de nuevas conexiones entre ellas y entonces se crea y se expande una gran interconectada que es capaz de poseer memoria. La memoria o conocimiento es almacenado como fuertes conexiones entre neuronas.

Una neurona biológica está conformado por tres componentes principales: las dendritas, que se encargan de canalizar las señales de entrada o estímulos, las cuales son ponderadas por las fuerzas de las conexiones, al cuerpo celular; un cuerpo celular que acumula las señales de entrada ponderadas y además procesa estas señales; y un axón que transmite la señal de salida a otras neuronas que están conectadas a este. (8) (7)

La conexión entre neuronas recibe el nombre de sinapsis. Las conexiones sinápticas pueden excitar o inhibir la señal que se quiera transmitir. Una sinapsis fuerte y con gran número de repeticiones tiende a reforzarse y a ser más eficiente cada vez en la transmisión; mientras que una sinapsis débil o de poca utilidad tiende a disminuir su eficiencia de transmisión y por lo tanto podría desaparecer. (9) (8)

Neurona artificial. Son elementos simples de cálculo que a partir de un vector de entrada procedente del mundo exterior, y/o a partir del estímulo recibido de otras neuronas proporcionan una respuesta única (salida).

Una red neuronal es un conjunto de unidades procesadoras (neuronas artificiales) distribuidas paralelamente y conectadas entre sí que intentan imitar la actividad cerebral. La red es capaz de aprender de los estímulos de su entorno (datos) y percibe sus características y propiedades.

Las neuronas presentan una tendencia natural a almacenar conocimientos por experiencia y lo ponen a disposición para su uso. Estas redes guardan ciertas similitudes con el cerebro humano, siendo principalmente estas dos:

- El conocimiento es adquirido por la red a través de un proceso de aprendizaje.
- Las interconexiones entre neuronas, que se conocen como pesos sinápticos o simplemente pesos, son usados para almacenar el conocimiento.

De la observación detallada del proceso llevado a cabo por neuronas biológicas se ha propuesto un modelo de neurona artificial que se muestra a continuación

Capa de entrada: es la capa que recibe directamente la información proveniente de las fuentes externas de la red.

Capa Ocultas: Son internas a la red y no tienen contacto directo con el entorno exterior. El número de niveles ocultos puede estar entre cero y un número elevado. Las neuronas de las capas ocultas pueden estar interconectadas de distintas maneras, lo que determina, junto con su número, las distintas topologías de redes neuronales.

Capa de salida: En esta capa se transfiere la información de la red hacia el exterior.

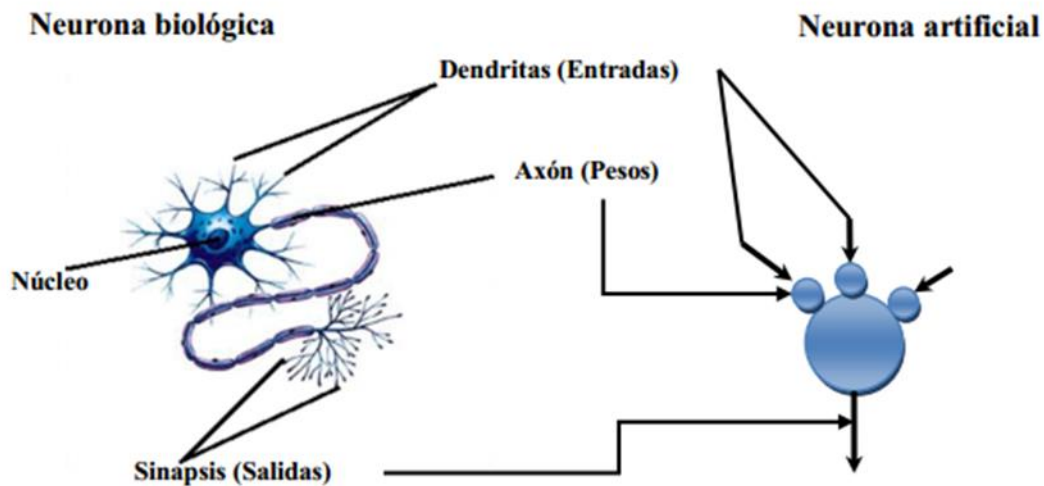


Figura 2.1 Relación entre una neurona biológica y artificial.

2.2 FUNCIONES BASE Y DE ACTIVACIÓN

Una red neuronal típica se puede caracterizar por la función de base y la función de activación. Cada nodo (unidad de proceso), suministra un valor a su salida. Este valor se propaga a través de la red mediante conexiones unidireccionales hacia otros nodos de la red.

Función Base

Todas las neuronas de la red presentan cierto estado inicial, de reposo o excitación que depende de su valor de activación. Este valor puede ser continuo o discreto, es decir, suelen tomar un conjunto discreto de valores binarios, así un estado activo se indicaría con el número uno (1) y un estado pasivo estaría representado por el número cero (0); pueden ser limitado o ilimitado, según la entrada total recibida y el umbral de la propia neurona. Si se designa como la activación de la i -ésima unidad respecto al momento de tiempo t , resulta posible definir el vector.

Si se designa como $a_i(t)$ la activación de la i -ésima unidad U_i respecto al momento de tiempo t , resulta posible definir el vector:

$$A(t) = [a_1(t), \dots, a_i(t), \dots, a_n(t)] \quad (2.1)$$

Que representa el estado de activación total de todas las neuronas de la red (de entrada, ocultas y de salida).

Función lineal de tipo hiperplano: El valor de red es una combinación lineal de las entradas.

$$u_i(w, x) = \sum_{j=1}^n w_{ij} x_j = w_j^t \cdot x_i \quad (2.2)$$

Donde w_j^t , vector transpuesto representativo de los N pesos de entrada que llegan a la j -ésima neurona, siendo $w_j^t = w_{1j}, \dots, w_{ij}, \dots, w_{iN}$ y x_i la i -ésima entrada a la j -ésima neurona. Las conexiones que unen a las neuronas que forman una RNA tiene asociado un peso, que es el que hace que la red adquiera conocimiento. Se considera que el efecto de cada señal es aditivo, de tal forma que la entrada neta que recibe una neurona (potencial postsináptico) es la suma del producto de cada señal individual por el valor de la sinapsis que conecta ambas neuronas y es lo que se conoce como regla de propagación.

Función de Activación (Función de neurona)

El valor de red, expresado por la función de base, se transforma mediante una función de activación no lineal. Las funciones de activación más comunes son la función sigmoideal y gaussiana (8):

Función sigmoideal

$$f(u_i) = \frac{1}{1 + e^{-\frac{u_i}{\sigma^2}}} \quad (2.3)$$

Se conoce con este nombre debido a que la gráfica correspondiente es similar a la letra "S". Esta función es comúnmente usada en redes multicapa, entrenadas con el algoritmo de retropropagación, en parte porque se trata de una función diferenciable.

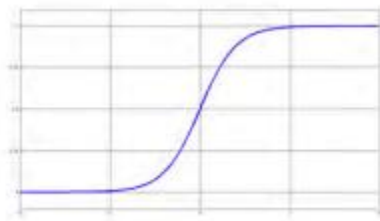


Figura 2.2. *Función sigmoideal.*

Función gaussiana

Su rango está entre 0 y 1. Se utiliza en redes neuronales de función de base radial, las cuales pueden aplicarse a problemas de predicción. Su expresión matemática es:

$$f(u_i) = ce^{-\frac{u_i^2}{\sigma^2}} \quad 2.4$$

2.3 Topología de redes neuronales.

Una red neuronal consiste en una organización de neuronas conectadas entre sí, formando capas o agrupaciones interconectadas entre la entrada y la salida. Los componentes fundamentales de la red son: el número de capas, el número de neuronas de cada capa, el grado de conectividad y el tipo de función de transferencia de cada capa. (10)

Una red neuronal se determina por las neuronas y la matriz de pesos. Hay tres tipos de capas de neuronas (10):

- La capa de entrada
- La capa oculta
- La capa de salida

De acuerdo al número de capas, las redes pueden ser clasificadas en redes de simple o de múltiples capas. Las capas existentes entre la capa de entrada y la de salida se llaman capas ocultas.

2.3.1 Según la estructura en capas.

2.3.1.1 Redes monocapa:

Compuestas por una única capa de neuronas, entre las que se establecen conexiones laterales y en ocasiones autorrecurrentes. Este tipo de redes suele utilizarse en problemas de auto asociación (10).

2.3.1.2 Redes multicapa:

Las neuronas están dispuestas en varias capas (entrada, oculta(s), salida). La capa a la que pertenece la neurona puede distinguirse mediante la observación del origen de las señales que recibe y el destino de las señales que genera. El número de elementos de procesado puede variar en unas capas respecto a otras. Las funciones (tanto de activación como de salida) de los elementos de procesado pueden ser diferentes. En general, cuanto mayor sea la diferencia entre los elementos de entrada y los de salida, será necesario dotar a la estructura de más capas ocultas. Estas capas ocultas crean una representación interna de los patrones de entrada. La habilidad de la red para procesar información crece en proporción directa al número de capas ocultas (10).

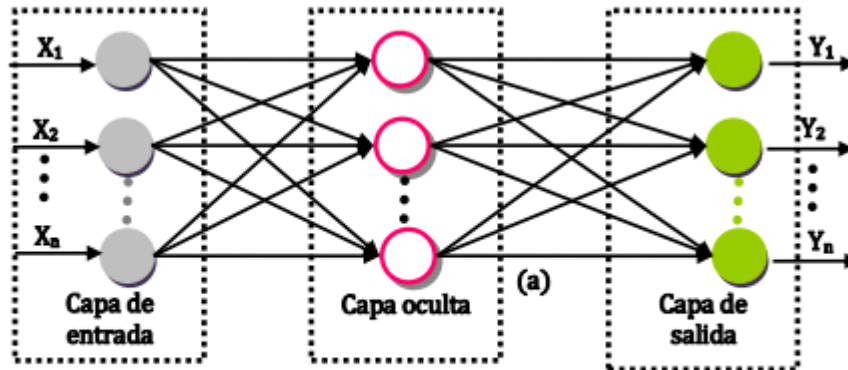
2.3.2 Según el flujo de datos en la red

Las capas de neuronas se interconectan mediante una red de pesos que puede ser de los siguientes tipos (10):

- Hacia adelante
- Hacia atrás
- Lateral
- De retardo

2.3.2.1 Redes unidireccionales o de propagación hacia adelante (feedforward)

Ninguna salida neuronal es entrada de unidades de la misma capa o de capas precedentes. La información circula en un único sentido desde las neuronas de entrada hacia las neuronas de salida de la red. (10)



Donde:

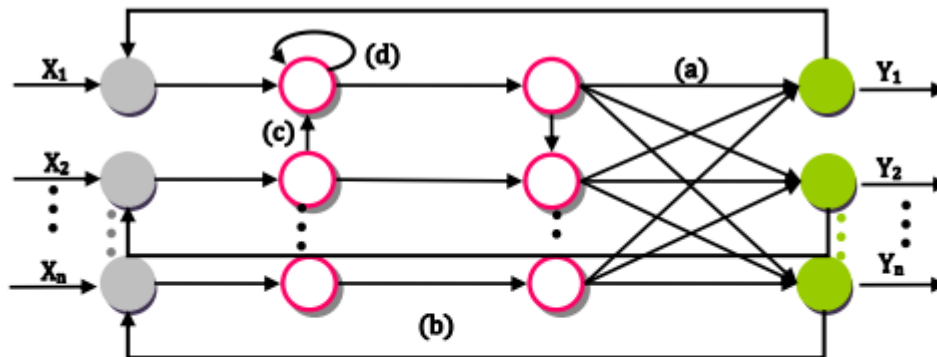
(a) = Conexiones hacia adelante

Figura 2.3. Red unidireccional propagación hacia adelante.

En una red multicapa de propagación hacia delante, puede haber una o más capas ocultas entre las capas de entrada y salida. El tamaño de las redes depende del número de capas y del número de neuronas ocultas por capa. El número de unidades ocultas está directamente relacionado con las capacidades de la red. Para que el comportamiento de la red sea correcto, se tiene que determinar apropiadamente el número de neuronas de la capa oculta.

2.3.2.2 Redes de propagación hacia atrás (feedback)

Las salidas de las neuronas pueden servir de entrada a unidades del mismo nivel (conexiones laterales) o de niveles previos. Las redes de propagación hacia atrás que presentan lazos cerrados se denominan sistemas recurrentes (10).



Donde:

(a) = Conexiones hacia adelante

(b) = Conexiones hacia atrás

(c) = Conexiones laterales

(d) = Conexiones autorrecurrentes

Figura 2.4. Propagación hacia atrás.

2.3.2.3 Redes de propagación lateral:

A la neurona de salida que da el valor más alto se le asigna el valor total (por ejemplo, 1), mientras que a todas las demás se le da un valor de 0.

2.3.2.4 Redes de propagación con retardo:

Los elementos de retardo se incorporan en las conexiones para implementar modelos dinámicos y temporales, es decir, modelos que precisan de memoria.

2.4 Mecanismos de aprendizaje

El propósito del proceso de aprendizaje es entrenar la red para que ésta pueda desempeñar algunas tareas. Hay varios tipos de reglas de aprendizaje de redes neuronales que pueden ser clasificadas en dos categorías principales: supervisado, no supervisado. (8) (7)

2.4.1 El entrenamiento supervisado

Se realiza mediante un entrenamiento controlado por un agente externo (supervisor o maestro) que determina la respuesta que debería generar la red a partir de una entrada determinada. El supervisor comprueba la salida de la red y en el caso de que ésta no coincida con la deseada, se procederá a modificar los pesos de las conexiones, con el fin de conseguir que la salida se aproxime cada vez más a la deseada. (8) (7)

Se consideran tres formas de llevar a cabo este tipo de aprendizaje:

2.4.1.1 Aprendizaje por corrección de error

Este tipo de entrenamiento consiste en presentar al sistema un conjunto de datos de entrada y la salida deseada para dicha entrada ajustando los pesos en función de la diferencia entre los valores deseados y los obtenidos como salida de la red. El objetivo, es minimizar el error entre la salida deseada y la salida que se obtiene. (8) (7)

2.4.1.2 Aprendizaje por refuerzo

Es más lento que el anterior. No se dispone de un ejemplo completo del comportamiento deseado pues no se conoce la salida deseada exacta para cada entrada sino que se conoce el comportamiento de manera general para diferentes entradas. La relación entrada-salida, se realiza a través de un proceso de éxito o fracaso, produciendo este una señal de refuerzo que mide el buen funcionamiento del sistema. La función del supervisor es más la de un crítico que la de un maestro. (8) (7)

2.4.1.3 Aprendizaje estocástico

Este tipo de aprendizaje consiste básicamente en realizar cambios aleatorios de los valores de los pesos y evaluar su efecto a partir del objetivo deseado y de distribuciones de probabilidad. (8) (7)

2.4.1.4 No supervisado.

Estas redes no requieren influencia externa para ajustar los pesos de las conexiones entre neuronas. La red no recibe ninguna información por parte del entorno que le indique si la salida generada es o no correcta, así que existen varias posibilidades en cuanto a la interpretación de la salida de estas redes (8) (7), tales como:

- **Familiaridad o Similitud:** Semejanzas entre la información actual y la información pasada.
- **Codificación:** En este se realiza una codificación de los datos de entrada, generando a la salida una versión codificada de la entrada, con menos bits, pero manteniendo la información relevante de los datos.
- **Mapeo:** Algunas redes con aprendizaje no supervisado realizan un mapeo de características, obteniéndose en las neuronas de salida una disposición geométrica que representa un mapa topográfico de las características de los datos de entrada, de tal forma que si se presentan a la red informaciones similares, siempre sean afectadas neuronas de salidas próximas entre sí, en la misma zona del mapa.
- **Clusterización o establecimiento de clases:** La red se encarga de encontrar las características propias de cada clase.

En general en el aprendizaje no supervisado se suelen considerar dos tipos:

2.4.1.5 Aprendizaje Hebbiano

Consiste básicamente en el ajuste de los pesos de las conexiones a partir de la correlación existente entre las salidas generadas por cada celda (8), (7):

$$\Delta W_{ij} = y_i * y_j \quad (2.5)$$

La regla de Hebb es de tipo no supervisado, pues la modificación de los pesos depende de los estados (salidas) de las neuronas obtenidas tras la presentación de un estímulo determinado, con independencia de que coincidan o no con las deseadas. De esta forma en el aprendizaje hebbiano múltiples neuronas de salida pueden activarse simultáneamente.

2.4.1.6 Aprendizaje Competitivo y Cooperativo

Las neuronas compiten (y cooperan) unas con otras con el fin de llevar a cabo una tarea dada. Con este tipo de aprendizaje se pretende que cuando se presente a la red cierta información de entrada, solo una de las neuronas de salida se active (alcance su valor de respuesta máximo). Por tanto las neuronas compiten por activarse, quedando finalmente una, o una por grupo, como neurona vencedora. El procedimiento de entrenamiento es muy simple: cuando se presenta un patrón de entrada se selecciona una única unidad de salida (la unidad ganadora) y se modifican los pesos de esa unidad. (8), (7)

2.5 Etapas en el diseño de una red neuronal

Existen algunos pasos asociados a la creación de una red neuronal que no necesariamente son rigurosos porque tienden a traslaparse en la práctica, pero resultan útiles para establecer cierto orden y tener una idea estructurada del proceso de construcción de la red. Las etapas de diseño se pueden dividir en cuatro (8), (7):

- ✓ Procesamiento de los datos
- ✓ Diseño de la red
- ✓ Implementación
- ✓ Validación

2.5.1 Procesamiento de los datos

Los datos iniciales pueden obtenerse de registros históricos, mediante la simulación por computadora, de tablas de resultados o de mediciones directas. Antes que estos datos sean utilizados como entradas efectivas en ocasiones es necesario un procesamiento inicial con el objetivo de limitar y/o reducir la dimensión del vector de entrada y con ello el crecimiento exponencial de la complejidad del problema vinculada a la magnitud de ese vector. También el procesamiento de los datos puede ser necesario para "limpiarlos", remover errores, valores faltantes o alguna irregularidad.

2.5.2 Diseño de la red

El primer paso es seleccionar una arquitectura apropiada para la red de acuerdo con la tarea que se desea realizar. Luego, se debe decidir sobre el número de capas, neuronas por capas y el tipo de función de transferencia. La entrada a la red responde al número de variables que representen al sistema y para ello se debe tener determinado conocimiento del comportamiento del mismo, así como de las variables independientes que condicionan su salida. El número de neuronas a la salida es igual al número de variables de respuesta requeridas. El número de neuronas en las capas ocultas es más difícil de determinar y existen pocas reglas heurísticas para ello. Lo más común es recurrir a ensayos de prueba y error, si se seleccionan pocas, el modelo puede ser poco flexible; y si se seleccionan muchas, tiende a conducir a problemas de memorización o sobre fijación de datos así como al incremento del tiempo de cómputo lo cual constituye una dificultad que debe conocerse.

2.5.3 Implementación

Después que la red ha sido diseñada, se debe entrenar utilizando alguno de los métodos existentes. Cuando el entrenamiento es iterativo, se debe definir un criterio para detener este proceso si así se requiere. Este puede ser fijar un número máximo de iteraciones o cierta tolerancia alcanzada en la función de error utilizada.

2.5.4 Validación

Es la etapa final y debe prestarse atención al hecho de que la buena convergencia entre los datos reales y los predichos por la red durante el entrenamiento no es un indicador suficiente de que la red responda de forma adecuada a todas las entradas que puedan suceder. Una de las medidas más comunes para confirmar la calidad de la red es seccionar los datos disponibles en tres grupos, uno de ellos es el de entrenamiento, otro se dedica a la simulación de determinada zona donde algunos de sus valores pueden traslaparse con los de entrenamiento, y finalmente, uno dedicado a la validación. Los datos de validación no son presentados a la red durante su entrenamiento.

2.6 PERCEPTRÓN MULTICAPA

Dentro de los modelos de RNA existen diversos tipos de estructuras dependiendo de la disposición de los enlaces que conectan las neuronas o si existen bucles dentro de la red. Una estructura ampliamente utilizada en el pronóstico de series de tiempo es el Perceptrón Multicapa (PM), el cual presenta las siguientes características:

- No existen bucles ni conexiones entre las neuronas de una misma capa.
- Las funciones de activación son iguales para cada neurona de una misma capa.
- Tiene una sola neurona en la capa de salida.

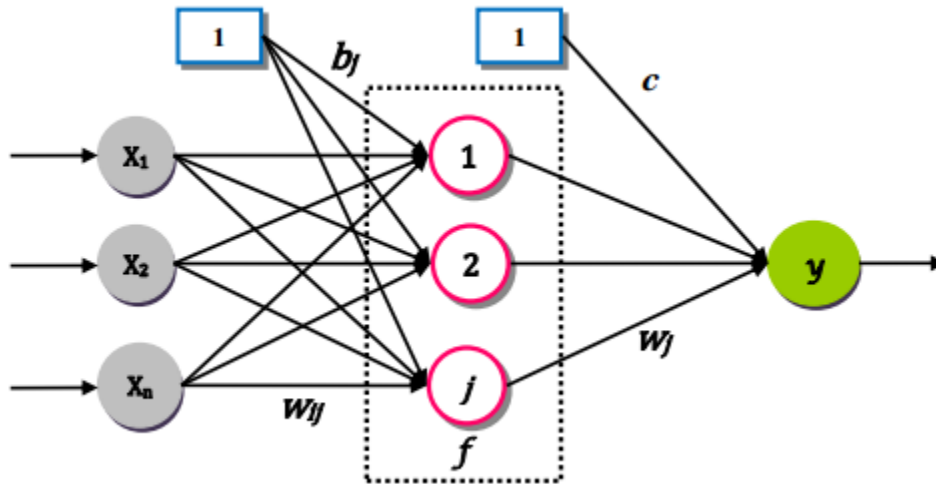


Figura 2.5. Esquema general de un Perceptrón Multicapa

La representación matemática de la **figura 2.5**, se indica en la ecuación

$$Y = \sum_{j=1}^m w_j \cdot f_j \left(\sum_{i=1}^n a_{ij} x_i + b_j \right) + c \quad (2.6)$$

Dónde:

Y: Variable pronosticada

x_i : Variables de entrada

w_j : Pesos que conectan la j-ésima salida de la capa oculta a la capa de salida

w_{ij} : Pesos que conectan la i-ésima entrada al modelo con la j-ésima neurona de la capa oculta.

b_j : Sesgos (bias) o intercepto de las j-ésima neurona oculta c es el sesgo (bias) o intercepto de la neurona y de salida.

w_j, a_{ij}, b_j y c son los parámetros del modelo que deben ser ajustadas mediante el algoritmo de entrenamiento. En total son $(n+1)(m+1)$ parámetros a ajustar, donde n es el número de entradas al modelo y m es el número de neuronas en la capa oculta de la red. Interpretando esta ecuación, el modelo resulta ser una regresión caracterizada por la función f, la cual puede ser lineal o no lineal y determina la forma en que se relacionan las entradas al modelo con la salida. (8), (7)

2.7 PASOS PARA LA IMPLEMENTACIÓN DE REDES NEURONALES EN EL PRONÓSTICO DE DEMANDA DE ENERGÍA ELÉCTRICA

1. Selección de la Variable
2. Recolección de Datos
3. Procesamiento de Datos.
4. Definir conjunto de Entrenamiento, Validación y Prueba
5. Selección topología de la red neuronal
 - a. Número de Neuronas de Entrada.
 - b. Número de Capas Ocultas.
 - c. Número de Neuronas Ocultas.
 - d. Número de Neuronas de Salida.
 - e. Función de Transferencia.

Figura 2.6. Metodología para pronóstico utilizando Redes Neuronales Artificiales.

2.8 Algoritmo para la creación de redes neuronales

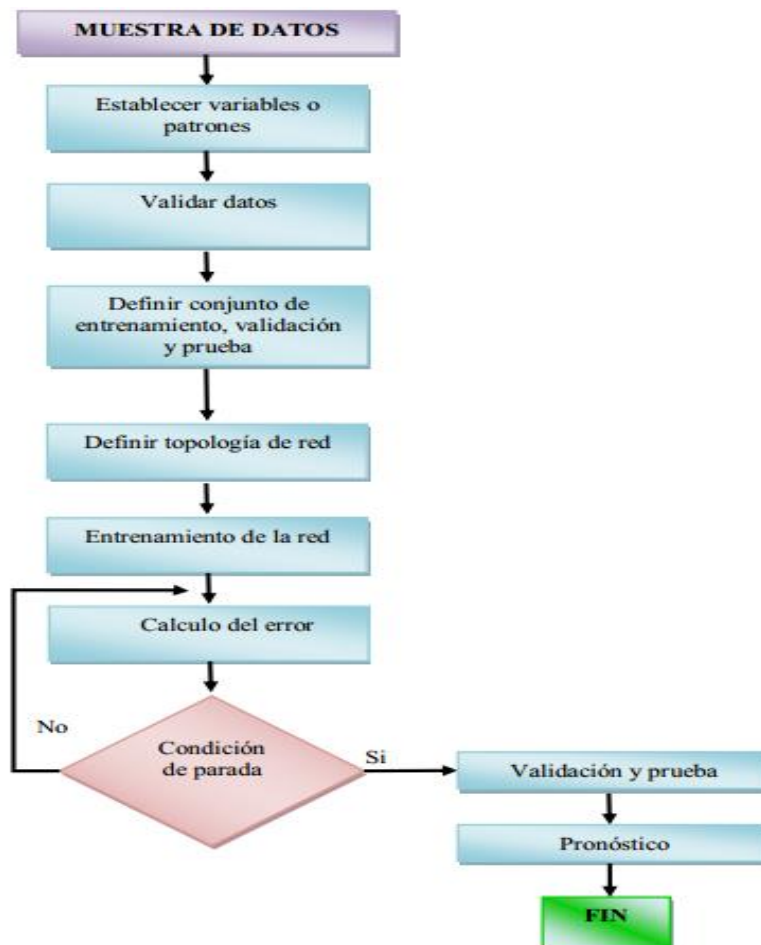


Figura 2.7. Para creación de redes neuronales.

2.9 Métodos de gradiente descendente

El método de entrenamiento más utilizado es el método del gradiente descendente este método define una función del error al cuadrado con respecto a los pesos de la red $E(W)$ que proporciona el error que comete la red en función del conjunto de pesos sinápticos W . el objetivo del aprendizaje será encontrar la configuración de pesos que corresponda al mínimo global de la función de error, aunque en muchos casos es suficiente encontrar un mínimo local lo suficientemente bueno. (11)

El principio general del método es el siguiente: dado un conjunto de pesos W para un instante de tiempo $t=0$, se calcula la dirección de máxima variación del error. La dirección de máximo crecimiento de la función $E(W)$ en $W(t=0)$ viene dado por el gradiente $\vec{\nabla}E(W)$. Luego se actualizan los pesos siguiendo el sentido contrario al indicado por el gradiente $\vec{\nabla}E(W)$, dirección que indica el sentido de máximo crecimiento. De este modo se va produciendo un descenso por la superficie de error hasta alcanzar un mínimo local. (11)

Para actualizar el nuevo peso en un tiempo $t=1$

$$W(T + 1) = W(t) - \alpha \vec{\nabla}E(w) \quad (2.7)$$

Donde α indica el tamaño del paso tomando en cada iteración, pudiendo ser diferente para cada peso e idealmente debería ser infinitesimal. El tamaño del paso es un factor importante a la hora de diseñar un método de estas características. Si se toma un paso muy pequeño el proceso de entrenamiento resulta muy lento, mientras que si el tamaño del paso es muy grande se producen oscilaciones den torno al punto mínimo. (11)

2.10 Estructura del perceptron

Cada elemento del vector input \vec{P} es conectado para cada neurona a travez de la matriz de pesos \vec{W} .

La neurona i -esima procesa la información usando los elementos inputs ponderados y agregando el valor de bias generando de esta forma el inputs que será tomado por la función de transferencia, dando como resultado el output $a(i)$. (11), (5)

$$z(i) = W_{i1} * p1 + W_{i2} * p2 + W_{i3} * p3 \dots \dots W_{iR} * pR + b \quad (2.8)$$

$$a(i) = F(z(i)) \quad (2.9)$$

La notación usada por los pesos es la siguiente W_{ij} , haciendo referencia al peso que hacia la neurona i -esima desde el elemento j -esimo del vector input:por ejemplo W_{12} es la fuerza con que viaja desde el elemento 2 del vector input hacia la primera neurona.

Es importante de resaltar que el número de elementos inputs no necesariamente tiene que ser igual al número de neuronas.

Estructura de una red neuronal con una sola capa

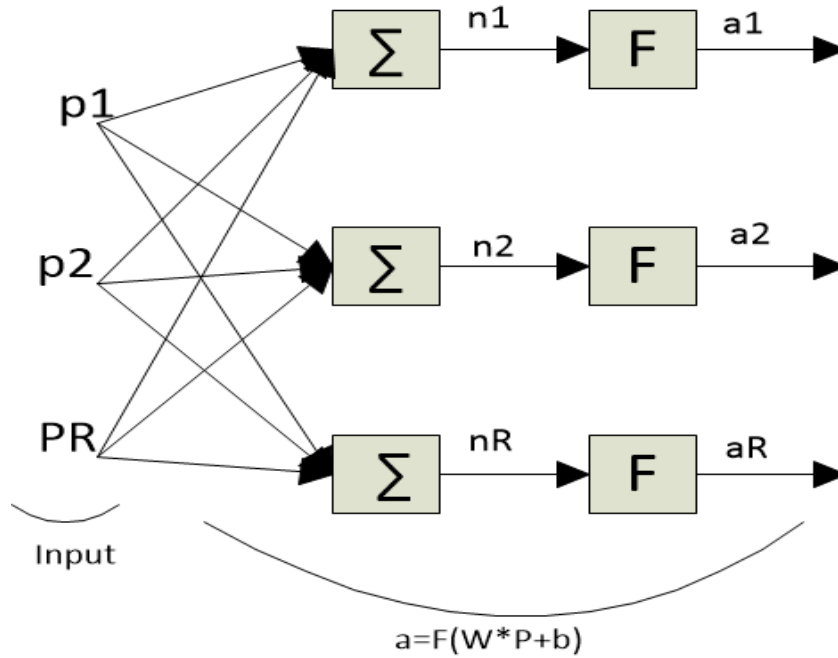


Figura 2.8. Estructura del perceptron simple

Las salidas o los outputs de cada capa intermedia son los inputs para las siguientes capas, de esta manera la capa 2 puede ser analizada como una red de una capa con inputs S1 y S2 neuronas, el vector de entrada es a1 y el de salida es a2. (11)

Estructura para varias capas.

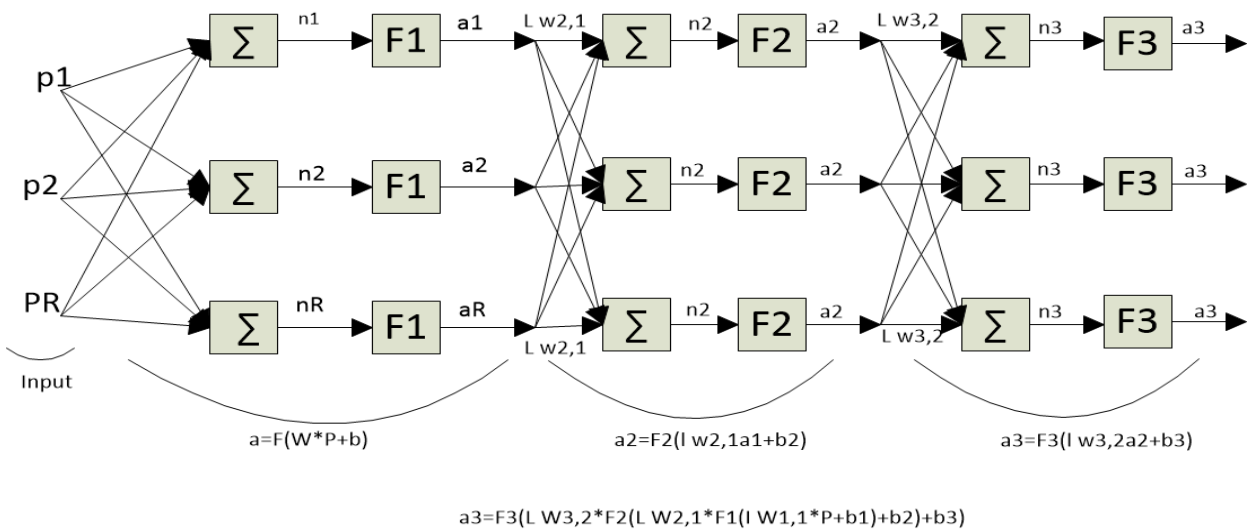


Figura 2.9. Estructura del perceptron múltiple.

2.11 El algoritmo Backpropagation

El algoritmo backpropagation es el método de entrenamiento más utilizado en redes con conexión hacia adelante. Es un método de aprendizaje supervisado de gradiente descendente, en el que se distingue claramente dos fases (5), (11):

1. Se aplica un patrón de entrada, el cual se propaga por las distintas capas que componen la red hasta producir la salida de la misma. Esta salida se compara con la salida deseada y se calcula el error cometido por cada neurona de salida. Estos errores se transmiten hacia atrás, partiendo de la capa de salida, hacia todas las neuronas de las capas intermedias.
2. Cada neurona recibe su error que es proporcional a su contribución sobre el error total de la red. Basándose en el error recibido, se ajustan los errores de los pesos sinápticos de cada neurona.

2.11.1 Deducción del algoritmo Backpropagation

El algoritmo propone una actualización iterativa de los pesos de la siguiente manera (11):

$$W(T + 1) = W(t) - \Delta W(t) \quad (2.10)$$

Si tomamos una variación proporcional al gradiente de una función de error $E(W)$ tenemos que (11):

$$W(T + 1) = W(t) - \alpha \vec{\nabla} E[W(t)] \quad (2.11)$$

Como se explicó anteriormente el primer paso de este algoritmo consiste en propagar hacia adelante un patrón de entrada o_k y obtener la salida de la red o_h . (5)

La regla de propagación más simple y utilizada consiste en realizar una suma de las entradas ponderadas con sus pesos sinápticos correspondientes. (5)

$$net_h = \sum_{j=0}^n w_{ij} * o_k \quad (2.12)$$

Al aplicarle la función de activación a la regla de propagación queda de la siguiente manera que es la salida de la neurona de la capa oculta:

$$o_h = f(net_h) = f(\sum_{j=0}^n w_{ij} * o_k) \quad (2.13)$$

La entrada ' o_h ' a una neurona es la suma ponderada de las salidas o_h de las neuronas anteriores. Si está en la primera capa después de la capa de entrada la o_h de la capa de entrada son simplemente las entradas o_k a la red. (5)

La función de activación:

$$\varphi(net_h) = \frac{1}{1 + e^{-net_h}} \quad (2.14)$$

Tiene una derivada (5):

$$\frac{\partial \varphi(net_h)}{\partial o_h} = -\frac{e^z}{(1+e^z)^2} = \varphi(net_h)(1 - \varphi(net_h)) \quad (2.15)$$

$$\varphi\left(\sum_{j=0}^n w_{ij} * o_k\right)' = -\frac{1}{(1+e^{\sum_{j=0}^n w_{ij}*o_k})^2} \quad (2.16)$$

De acuerdo a la ecuación 2.11 la variación de los pesos sinápticos será proporcional al gradiente de la función de error (5):

$$\Delta W_{ji} = -\alpha \frac{\partial E}{\partial w_{ji}} \quad (2.17)$$

El error viene dado por:

$$E = \frac{1}{2}(t - y)^2 \quad (2.18)$$

t: es el objetivo para una muestra de entrenamiento

y: es la salida real de la neurona de salida

El factor de $\frac{1}{2}$ esta incluido para cancelar el exponente al diferenciar.

Se procederá a encontrar la derivada del error con respecto a un peso w_{ij} aplicando la regla de la cadena 2 veces (5).

$$\frac{\partial E}{\partial w_{ij}} = \frac{\partial E}{\partial o_h} \frac{\partial o_h}{\partial net_h} \frac{\partial net_h}{\partial w_{ij}} \quad (2.19)$$

Encontrando el último factor de la ecuación anterior

$$\frac{\partial net_h}{\partial w_{ij}} = \frac{\partial (\sum_{j=0}^n w_{ij} * o_k)}{\partial w_{ij}} = \frac{\partial w_{ij} * o_k}{\partial w_{ij}} = o_k \quad (2.20)$$

La derivada de la salida de la neurona j con respecto a su entrada es simplemente la derivada parcial de la función de activación.

$$\varphi(net_h) = \frac{1}{1+e^{net_h}} \quad (2.21)$$

$$\frac{\partial(\varphi(net_h))}{\partial net_h} = \frac{\partial}{\partial net_h} \left(\frac{1}{1+e^{net_h}} \right) = \varphi(net_h)(1 - \varphi(net_h)) = -\frac{e^{net_h}}{(1+e^{net_h})^2} \quad (2.22)$$

Esta es la razón por la que backpropagation requiere de a función de activación sea diferenciable.

Si escribimos los dos primeros términos de la ecuación 2.19 como:

$$-\delta_h = \frac{\partial E}{\partial o_h} \frac{\partial net_h}{\partial w_{ij}} \quad (2.23)$$

Tenemos que:

$$\frac{\partial E}{\partial w_{ij}} = -\delta_h o_k \quad (2.24)$$

Esta es la regla delta para una red de una capa. (5)

Tendríamos un resultado de la ecuación 2.10 como se muestra continuación:

$$\Delta W_{ji} = -\alpha \delta_j o_k \quad (2.25)$$

El producto de la tasa de aprendizaje y el gradiente, multiplicado por -1 garantiza que W_{ji} tendrá cambios de una manera que siempre disminuya E. (11), (5)

Para actualizar los pesos W_{ji} usando un gradiente descendente, uno debe elegir una tasa de aprendizaje $\alpha > 0$. El cambio en el peso debe reflejar el impacto en el Error de un aumento o disminución de W_{ji} . (11)

Si

$$\frac{\partial E}{\partial w_{ij}} > 0 \text{ Un aumento en } w_{ij} \text{ aumenta } E \quad (2.26)$$

Por otra parte, si:

$$\frac{\partial E}{\partial w_{ij}} < 0 \text{ Aumenta } w_{ij} \text{ disminuye } E \quad (2.27)$$

El primer factor de la ecuación 2.19 es simplemente para evaluar si la neurona está en la capa de salida o en una capa intermedia (5), (11).

En el caso si está en la capa de salida:

Para este caso $y = o_h$

$$\frac{\partial E}{\partial o_h} = \frac{\partial E}{\partial y} = \frac{\partial (\frac{1}{2}(t-y)^2)}{\partial y} = y - t \quad (2.28)$$

Sin embargo, si j se encuentra en una capa interna arbitraria de la red, encontrándola derivada de E con respecto a " o_h ".

Donde "h" es el subíndice de las neuronas que pertenecen a la próxima capa.

$$-\delta_h = -\frac{\partial E}{\partial net_h} \quad (2.29)$$

$$-\frac{\partial E}{\partial net_h} = \frac{\partial E}{\partial o_h} \frac{\partial o_h}{\partial net_h} \quad (2.30)$$

$$\frac{\partial o_h}{\partial net_h} = -\frac{e^{net_h}}{(1+e^{net_h})^2} \quad (2.31)$$

Ahora vamos a derivar el primer factor de la ecuación anterior debemos señalar que la derivación de la función de error según la salida de una capa de una neurona interna depende del vector de todas las entradas de la red de la siguiente capa. (5), (11)

$$-\frac{\partial E}{\partial o_h} = -\frac{\partial E(net_{l1}, \dots, net_{l1L1})}{\partial o_h} \quad (2.32)$$

De acuerdo con la definición de la regla de cadena multidimensional, obtenemos inmediatamente la ecuación 2.33:

$$-\frac{\partial E}{\partial o_h} = \sum_{l \in L} \left(-\frac{\partial E}{\partial net_l} \cdot \frac{\partial net_l}{\partial o_h} \right) \quad (2.33)$$

La suma en la ecuación 2.33 contiene dos factores. Ahora queremos discutir que estos factores se agreguen a la capa L posterior. Simplemente calculamos el segundo factor en la siguiente ecuación 2.33:

$$\frac{\partial net_l}{\partial o_h} = \frac{\partial (\sum_{j=0}^n w_{lj} * o_h)}{\partial o_h} \quad (2.34)$$

$$\frac{\partial net_l}{\partial o_h} = w_{h.L} \quad (2.35)$$

Para el primer factor

$$-\frac{\partial E}{\partial net_l} = \delta_l \quad (2.36)$$

Por la ecuación 2.33 tenemos que:

$$\frac{\partial E}{\partial o_h} = \sum_{l \in L} \delta_l w_{h.L} \quad (2.37)$$

$$-\delta_h = -\frac{e^{net_h}}{(1+e^{net_h})^2} \sum_{l \in L} \delta_l w_{h.L} \quad (2.38)$$

Así, la variación de los pesos de una conexión que va desde una capa hacia otra capa de la red que no sea la externa se calcula como (5), (11):

$$\Delta W_{ji} = \alpha \delta_h o_k$$

$$\Delta W_{ji} = \alpha -\frac{e^{net_h}}{(1+e^{net_h})^2} \sum_{l \in L} \delta_l w_{h.L} o_k \quad (2.39)$$

En la implementación del algoritmo, se toma una amplitud de paso que viene dado por la tasa de aprendizaje α . A mayor tasa de aprendizaje el proceso será más rápido. Sin embargo, si la tasa de aprendizaje es muy alta puede dar lugar a oscilaciones en torno a un mínimo local. (5), (11)

3 PRONOSTICO A TRAVES DEL ANALISIS DE SERIES TEMPORALES.

3.1.1 Series temporales y procesos estocásticos univariantes.

3.1.1.1 Introducción.

Una serie temporal es el resultado de observar los valores de una variable a lo largo del tiempo en intervalos regulares (cada día, cada mes, cada año, etc.). Además, se puede considerar que una serie temporal es la observación de la realización de un proceso entre los instantes 1 y N: $X_1, X_2, X_3, \dots, X_N$.

Un proceso estocástico univariante es un conjunto de variables aleatorias $\{X_t\}$ donde el índice t toma valores en un cierto conjunto Ω .

Para cada valor de t del conjunto está definida una variable aleatoria, X_t , y los valores observados de las variables aleatorias en distintos instantes forman una serie temporal. Es decir, una serie de N datos, (X_1, \dots, X_N) , es una muestra del vector de N variables aleatorias ordenadas en el tiempo correspondientes a los momentos $t = 1, \dots, N$, y la serie observada se considera una realización del proceso estocástico.

De esta manera conociendo $X_1, X_2, X_3, \dots, X_N$ se tendrá que identificar el proceso $\{X_t\}$ por lo que habrá que inferir el proceso estocástico que ha generado los datos que hemos observado, el cual será caracterizado por su función de distribución.

3.1.2 Clases de procesos estocásticos

3.1.2.1 Clasificación.

Los fenómenos dinámicos que observamos mediante series temporales pueden clasificarse en dos clases. Los primeros son los que toman valores estables en el tiempo alrededor de un nivel constante, sin mostrar una tendencia a crecer o a decrecer a largo plazo, estos procesos se denominan estacionarios. Además, hay una segunda clase de procesos, que son los procesos no estacionarios, que son aquellos que pueden mostrar tendencia, estacionalidad y otros efectos evolutivos en el tiempo.

3.1.2.2 Procesos estocásticos estacionarios

Se dice que un proceso estocástico es *estacionario* en *sentido estricto* si:

1. las distribuciones marginales de todas las variables son idénticas;
2. las distribuciones finito-dimensionales de cualquier conjunto de variables sólo dependen de los retardos entre ellas.

La primera condición establece que, en particular, la media y la varianza de todas las variables son las mismas.

La segunda condición impone que la dependencia entre las variables sólo depende de sus retardos, es decir, la misma dependencia existe entre las variables X_t, X_{t+j}, X_{t+j+h} , que entre las variables $X_{t+k}, X_{t-j+k}, X_{t-j-h+k}$.

Estas dos condiciones pueden resumirse estableciendo que la distribución conjunta de cualquier conjunto de variables no se modifica si trasladamos las variables en el tiempo, es decir:

$$F(t_1, t_2, \dots, t_N) = F(t_{1+k}, t_{2+k}, \dots, t_{N+k})$$

Una propiedad más débil, pero más fácil de contrastar en la práctica, es la *estacionariedad en sentido débil*, que implica la estabilidad de la media, la varianza y la estructura de covarianzas a lo largo del tiempo.

Un proceso es estacionario en sentido débil si, para todo t:

$$\begin{aligned} E[X_t] &= \mu < +\infty \\ V[X_t] &= \gamma_0 < +\infty \\ cov(X_t, X_{t+k}) &= \gamma_k \end{aligned}$$

Las dos primeras condiciones indican que la media y la varianza son constantes. La tercera, que la covarianza entre dos variables depende sólo de su separación. En un proceso estacionario las autocovarianzas y autocorrelaciones sólo dependen del retardo entre las observaciones y, en particular, la relación entre X_t y X_{t-k} , es siempre igual a la relación entre X_t y X_{t+k} . En consecuencia, en los procesos estacionarios:

$$cov(X_t, X_{t+k}) = cov(X_{t+j}, X_{t+j+k}) = \gamma_k \quad \forall t, \forall k, \forall j$$

3.1.2.3 Estimación de los momentos de procesos estacionarios

A continuación, se detalla la manera de cómo estimar la media, la varianza, las covarianzas y las autocorrelaciones del proceso a partir de la única realización disponible.

Estimadores:

$\bar{X} = \sum_{t=1}^N \frac{X_t}{N}$	$\widehat{\gamma}_k = \sum_{t=1}^{N-k} \frac{(X_t - \bar{X})(X_{t+k} - \bar{X})}{N} \quad (\text{autocovarianza})$
$\widehat{\gamma}_0 = \sum_{t=1}^N \frac{(X_t - \bar{X})^2}{N}$	$\widehat{\rho}_k = r_k = \frac{\widehat{\gamma}_k}{\widehat{\gamma}_0} \quad (\text{autocorrelacion})$

Tabla 3.1 Estimación de los momentos de procesos estacionarios

3.1.3 Proceso de ruido blanco

3.1.3.1 Definición.

El proceso de ruido blanco es un proceso estacionario importante, el cual está definido por las siguientes condiciones:

$$\begin{aligned} E[a_t] &= 0, \forall t \\ V[a_t] &= \sigma_a^2, \forall t \\ cov(a_t, a_{t+k}) &= 0, \forall t, \forall k \neq 0 \end{aligned}$$

La primera condición establece que la esperanza es constante e igual a cero, la segunda condición establece que la varianza es constante y la tercera establece que las variables del proceso de ruido blanco no están correlacionadas para todos los retardos.

Un proceso de ruido blanco no es necesariamente estacionario en sentido estricto ni debe estar formado por variables independientes, ya que sólo se exige la incorrelación. Si se impone la condición adicional de que las variables del proceso sean independientes, y no sólo incorrelacionadas, se le conoce como *proceso de ruido blanco estricto*. Si se supone que las variables tienen distribución normal, la incorrelación garantiza la independencia, y la normalidad la igualdad de las distribuciones marginales, con lo que el proceso será de

ruido blanco estricto y con variables normales y se conoce al proceso como: *proceso de ruido blanco normal*. A partir del proceso de ruido blanco se puede pasar a generar procesos estocásticos más complejos.

3.1.4 Proceso lineal general.

3.1.4.1 Definición.

Constituye la clase más interesante de procesos estacionarios que depende de un número infinito de parámetros, y se puede considerar como un filtro lineal sobre el proceso de ruido blanco $\{a_t\}_{t \in \mathbb{Z}}$ con las siguientes propiedades:

$$X_t = \sum_{j=0}^{\infty} \psi_j a_{t-j} \qquad \sum_{j=0}^{\infty} \psi_j^2 < +\infty$$

$$\psi_0 = 1 \qquad = a_t + \psi_1 a_{t-1} + \psi_2 a_{t-2} + \dots$$

$$\psi_j \in \mathbb{R}, \forall j$$

Ésta se puede representar de la siguiente manera:

$$X_t = \sum_{j=0}^{\infty} \psi_j a_{t-j} = \psi(B) a_t \quad (3.1)$$

$$\psi(B) = \sum_{j=0}^{\infty} \psi_j B^j$$

Donde a B se conoce como operador de retardo:

$$BX_t = X_{t-1}$$

$$B^k X_t = X_{t-k}$$

3.1.4.2 Modelos Autorregresivos y de Media Móvil

3.1.4.2.1 Modelos ARMA(p,q)

Son una clase de procesos lineales generales que dependen de un número *finito* de parámetros, así teniendo:

$$\psi(B) = \sum_{j=0}^{\infty} \psi_j B^j$$

Dicha función se puede expresar como una función desarrollable en series de potencia, así:

$$\psi(B) = \frac{\theta(B)}{\phi(B)} \quad (3.2)$$

Siendo:

$$\theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \theta_3 B^3 \dots - \theta_q B^q$$

$$\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \phi_3 B^3 \dots - \phi_p B^p$$

$\psi(B)$ es el operador que verifica:

$$\phi(B)\psi(B) = \theta(B) \quad (3.3)$$

Dado que $\psi(B)$ depende de un número infinito de parámetros, el problema es el denominador para considerar que el proceso es estacionario y se considerarán dos

condiciones: la primera es la condición de estacionariedad (CE) donde las soluciones del polinomio $\phi(B) = 0$ deben estar fuera de $|B| \leq 1$. La segunda es la condición de invertibilidad (CI) donde las soluciones del polinomio $\theta(B) = 0$ deben estar fuera de $|B| \leq 1$.

El proceso autorregresivo de media móvil ARMA(p,q) será:

$$(1 - \phi_1 B - \phi_2 B^2 - \phi_3 B^3 \dots - \phi_p B^p) X_t = (1 - \theta_1 B - \theta_2 B^2 - \theta_3 B^3 \dots - \theta_q B^q) a_t$$

O, en notación compacta,

$$\phi_p(B) X_t = \theta_q(B) a_t \quad (3.4)$$

El proceso será estacionario si las soluciones de $\phi_q(B) = 0$ están fuera del círculo unidad, e invertible si lo están las de $\theta_q(B) = 0$.

El proceso ARMA(p,q) se puede expresar de la siguiente manera:

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} \dots + \phi_p X_{t-p} + a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} \dots - \theta_q a_{t-q}$$

A partir de este proceso se pasa a dos casos particulares:

Procesos Autorregresivos ARMA(p,0) ~ AR(p) expresados como una combinación lineal de valores anteriores de la serie:

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} \dots + \phi_p X_{t-p} \quad (3.5)$$

Procesos de Media Móvil ARMA(0,q) ~ MA(q) expresados como una combinación lineal de un proceso de ruido blanco:

$$X_t = a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} \dots - \theta_q a_{t-q} \quad (3.6)$$

3.1.4.2.2 Modelos AR(p)

Una serie de tiempo estacionaria sigue un proceso autorregresivo de orden p si:

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} \dots + \phi_p X_{t-p} + a_t$$

Donde a_t es un proceso de ruido blanco. Utilizando la notación de operadores de retardo, se puede expresar como:

$$(1 - \phi_1 B - \phi_2 B^2 - \phi_3 B^3 \dots - \phi_p B^p) X_t = a_t$$

O en forma compacta:

$$\phi_p(B) X_t = a_t \quad (3.7)$$

Los procesos autorregresivos AR(p) expresan que el valor actual del proceso está expresado como una combinación lineal de p valores previos del proceso y un ruido aleatorio.

Este tipo de proceso puede considerarse siempre como invertible, sin embargo, se debe demostrar que el proceso es estacionario.

Llamando ecuación característica del proceso a:

$$\phi_p(B) = 0$$

Considerada como función de B. Ésta tendrá p raíces G_1, G_2, \dots, G_p , en general distintas, de manera que los coeficientes son los factores de la ecuación característica, y se demuestra que el proceso es estacionario si $|G_i| < 1, \forall i$

3.1.4.2.3 Modelos MA(q)

Son procesos cuyo valor actual depende de q errores aleatorios. El proceso MA(q) tiene la siguiente representación:

$$X_t = a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} \dots - \theta_q a_{t-q}$$

Y donde es a_t es un proceso de ruido blanco. Utilizando la notación de operadores de retardo, se puede expresar como:

$$X_t = (1 - \theta_1 B - \theta_2 B^2 \dots - \theta_q B^q) a_t$$

O en forma compacta:

$$X_t = \theta_q(B) a_t \quad (3.8)$$

Un proceso MA(q) es siempre estacionario, por ser la suma de procesos estacionarios. El proceso será invertible si las soluciones del operador $\theta_q(B) = 0$, en módulo, son mayores que la unidad.

3.2 Identificación de una serie de tiempo

3.2.1 Pruebas de estacionariedad.

Los modelos clásicos de Box-Jenkins describen series temporales estacionarias. Por lo tanto, es necesario primero determinar si la serie temporal es estacionaria; Sí no es así, se debe transformar la serie.

1. Propiedades estadísticas De manera intuitiva una serie es estacionaria si la media y varianza son esencialmente constantes a través del tiempo, es decir, que sus n valores fluctúan con variación constante respecto de la media.

2. Prueba gráfica Es aconsejable graficar la serie de tiempo bajo estudio, para observar su comportamiento y tendencia.

3. Función de autocorrelación y correlograma La función de autocorrelación al retardo k se define como: $\widehat{\rho}_k = r_k = \frac{\widehat{\gamma}_k}{\widehat{\gamma}_0} = \frac{\text{covarianza estimada al retardo } k}{\text{varianza estimada}}$

La varianza y covarianza están medidas en las mismas unidades por lo que es un número sin unidad de medida y se encuentra entre -1 y +1, igual que el coeficiente de correlación. Si se grafica frente a k la gráfica obtenida se conoce como *correlograma poblacional*. En la práctica solo se tiene una realización de un proceso estocástico, es decir, la muestra, solamente se puede calcular la *función de autocorrelación muestral*. Por consiguiente, se debe calcular primero la covarianza muestral al retardo k, y la varianza muestral, que fueron definidas anteriormente en este capítulo. La grafica de frente a retardo k se le conoce como **correlograma muestral**. Una regla práctica consiste en calcular la FAC hasta un tercio o una cuarta parte de la longitud de la serie de tiempo. Una forma de saber si la serie tiene Tendencia, Estacionalidad, es una serie Aleatoria o una serie Estacionaria es mediante la observación del correlograma.

El correlograma es una herramienta gráfica que se emplea para exhibir las autocorrelaciones para varios retardos en una serie de tiempo.

Matriz de autocorrelación

El coeficiente de autocovarianza γ_k , al retardo k , mide la covarianza entre dos valores X_t y X_{t+k} a una distancia de tiempo k . La gráfica de γ_k versus k se llama función de autocovarianza $\{\gamma_k\}$ del proceso estocástico. Similarmente la gráfica del coeficiente de correlación ρ_k como función del retardo k es llamada función de autocorrelación $\{\rho_k\}$ del proceso. Note que la función de autocorrelación es adimensional, lo que significa, independiente de la escala de medición de la serie de tiempo. Dado que $\gamma_k = \rho_k \sigma_X^2$.

Se llamará matriz de autocorrelación a:

$$\mathbf{P}_n = \begin{bmatrix} 1 & \rho_1 & \rho_2 & \cdots & \rho_{n-1} \\ \rho_1 & 1 & \rho_1 & \cdots & \rho_{n-2} \\ \rho_2 & \rho_1 & 1 & \cdots & \rho_{n-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho_{n-1} & \rho_{n-2} & \rho_{n-3} & \cdots & 1 \end{bmatrix}$$

La función de autocorrelación, mostrada en la figura 1 como una gráfica de las diagonales de la matriz de autocorrelación, revela como la correlación entre dos valores cualquiera de la serie mientras cambia su separación. Dado que $\rho_k = \rho_{-k}$, la función de autocorrelación es necesariamente simétricamente alrededor de cero, y en la práctica es necesario solamente graficar la mitad positiva de la función. En la figura también se muestra la mitad positiva de la función de autocorrelación dada.

De ahora en adelante cuando hablamos de función de autocorrelación, nos referimos a la parte positiva.

En la práctica para obtener un estimado de la función de autocorrelación, necesitaremos por lo menos 50 observaciones, y las autocorrelaciones estimadas r_k serán calculadas para $k = 0, 1, 2, \dots, k$, donde k no será necesariamente mayor que $N/3$.

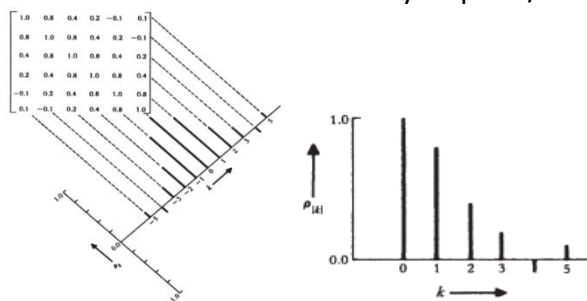


Figura 3.1

Error estándar en la función de autocorrelación estimada.

Para identificar un modelo, es útil tener una comprobación práctica que ρ_k sea efectivamente cero después de cierto retardo. Para este propósito, se puede hacer uso de una aproximación de la varianza de los coeficientes de autocorrelación de un proceso normal estacionario, gracias a Bartlett (11)

$$var[r_k] \cong \frac{1}{N} \sum_{v=-\infty}^{\infty} (\rho_v^2 + \rho_{v+k}\rho_{v-k} - 4\rho_k\rho_v\rho_{v-k} + 2\rho_v^2\rho_k^2) \quad (3.9)$$

Para cualquier proceso para el cual todas las autocorrelaciones ρ_v son cero para $v > q$, todos los términos excepto el primero del lado derecho de (3.9) son cero cuando $k > q$. Entonces, la varianza de las autocorrelaciones estimadas r_k , para k mayores que un valor q , más allá del cual la función de autocorrelación teórica pueda ser considerada que ha “muerto”, la aproximación de Bartlett nos da:

$$var[r_k] \cong \frac{1}{N} \left(1 + 2 \sum_{v=1}^q (\rho_v^2) \right) \quad k > q \quad (3.10)$$

En la práctica las autocorrelaciones aproximadas r_k ($k = 1, 2, \dots, q$) sustituyen a las autocorrelaciones teóricas ρ_k . Un caso de interés particular ocurre cuando $q = 0$, por lo tanto, cuando las autocorrelaciones teóricas se toman como cero para todos los retardos (excepto el retardo cero), y por lo tanto la serie es completamente aleatoria o ruido blanco. Entonces la varianza para autocorrelaciones estimadas r_k toman la forma simple:

$$var[r_k] \cong \frac{1}{N} \quad k > 0 \quad (3.11)$$

Dado que se conoce que r_k esta aproximadamente distribuida normalmente para una muestra N grande, tienden a ser independientes y normalmente distribuidas con media 0 y varianza $1/N$.

Los intervalos de confianza del 95% son: $\pm 1.96(\sqrt{var[r_k]})$

3.2.2 FUNCION DE AUTOCORRELACION PARCIAL

Inicialmente, podemos no conocer el orden del proceso autorregresivo que se ajusta a una serie de tiempo observada. El problema es análogo a decidir el número de variables independientes a ser incluidas en una regresión múltiple.

La función de autocorrelación parcial es un método que explota el hecho que donde un proceso AR(p) tiene una función de autocorrelación que es de longitud infinita, puede por su naturaleza ser descrita en p funciones de autocorrelaciones distintas de cero.

Denotamos ϕ_{kj} al j -ésimo coeficiente en una representación autorregresiva de orden k , de manera que ϕ_{kk} es el último coeficiente. De la ecuación (3.5), ϕ_{kj} satisface que:

$$\rho_j = \phi_{k1}\rho_{j-1} + \dots + \phi_{k(k-1)}\rho_{j-(k-1)} + \phi_{kk}\rho_{j-k} \quad j = 1, 2, \dots, k$$

Llevando a las ecuaciones de Yule-Walker, que pueden ser escritas:

$$\begin{bmatrix} 1 & \rho_1 & \rho_2 & \dots & \rho_{n-1} \\ \rho_1 & 1 & \rho_1 & \dots & \rho_{n-2} \\ \rho_2 & \rho_1 & 1 & \dots & \rho_{n-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho_{n-1} & \rho_{n-2} & \rho_{n-3} & \dots & 1 \end{bmatrix} \begin{bmatrix} \phi_{k1} \\ \phi_{k2} \\ \phi_{k3} \\ \vdots \\ \phi_{kk} \end{bmatrix} = \begin{bmatrix} \rho_1 \\ \rho_2 \\ \rho_3 \\ \vdots \\ \rho_k \end{bmatrix}$$

Y resolviendo para $k=1, 2, \dots$ sucesivamente se obtienen $\phi_{11}, \phi_{22}, \dots, \phi_{kk}$

Existe un método recursivo simple para hacer esta estimación, gracias a Durbin (11). Sin embargo, estas estimaciones obtenidas se vuelven muy sensibles a los errores de redondeo y no deberían utilizarse si los valores de los parámetros están cerca de los límites no estacionarios. Fue demostrado por Quenouille (11) que en la hipótesis que el proceso es

autorregresivo de orden p , las autocorrelaciones parciales estimadas de orden $p + 1$, y superior, son aproximadamente independientes y normalmente distribuidos con media cero. Además, si n es el número de observaciones utilizadas en la aproximación de la autocorrelación estimada $\hat{\phi}_{kk}$.

$$var[\hat{\phi}_{kk}] \cong \frac{1}{n} \quad k \geq p + 1 \quad (3.12)$$

Al igual que con la función de autocorrelación, la comprobación de la hipótesis que $\hat{\phi}_{kk}$ sea efectivamente cero para un límite de confianza del 95% es $\pm 1.96 \sqrt{var(\hat{\phi}_{kk})}$

3.2.3 Identificación de modelos estacionarios a partir de las fas y fap

En la tabla 3.2 se muestran los distintos casos posibles para la identificación de modelos estacionarios, la interpretación grafica de “muchos coeficientes no nulos” a menudo se presenta como exponenciales amortiguadas, o como una combinación de senoidales y/o exponenciales amortiguadas.

	PROCESO AUTORREGRESIVO	PROCESO DE MEDIA MÓVIL	PROCESO MIXTO
CONDICION DE ESTACIONARIEDAD	Raíces de $\phi(B) = 0$ caen fuera del circulo unitario	Siempre estacionario	Raíces de $\phi(B) = 0$ caen fuera del circulo unitario
CONDICION DE INVERTIBILIDAD	Siempre invertible	Raíces de $\theta(B) = 0$ caen fuera del circulo unitario	Raíces de $\theta(B) = 0$ caen fuera del circulo unitario
FUNCION DE AUTOCORRELACION	Infinita (exponenciales amortiguadas y/o senoidales amortiguadas), se desvanece	Se corta después de q retardos	Infinita (exponenciales amortiguadas y/o senoidales amortiguadas después de los primeros $q-p$ retardo), se desvanece
FUNCION DE AUTOCORRELACION PARCIAL	Se corta después de p retardos	Infinita, exponenciales y/o senoidales amortiguadas, se desvanece	Infinita (exponenciales amortiguadas y/o senoidales amortiguadas después de los primeros $p-q$ retardo), se desvanece

Tabla 3.2 Comportamiento de las FAS y FAP para la identificación de procesos e identificación de condiciones para la realización válida del modelo.

3.2.4 Procesos estocásticos no estacionarios.

Un proceso puede ser no estacionario en la media, en la varianza, en las autocorrelaciones, o en otras características de la distribución de las variables. Cuando el nivel de la serie no es estable en el tiempo, pudiendo tener tendencia creciente o decreciente, la serie es no

estacionaria en la media. Cuando la variabilidad o las autocorrelaciones se modifican con el tiempo, entonces la serie es no *estacionaria en la varianza* o en las autocorrelaciones. Finalmente, si la distribución de la variable en cada instante varía con el tiempo, la serie no es *estacionaria en distribución*.

3.2.4.1 Procesos integrados

Los procesos más importantes son los integrados, que tienen la propiedad fundamental de que al diferenciarlos se obtienen procesos estacionarios. Una propiedad importante que diferencia a los procesos integrados de los estacionarios es la forma en que desaparece la dependencia con el tiempo. En los procesos estacionarios ARMA(p,q) las autocorrelaciones disminuyen geométricamente, y se hacen prácticamente cero a los pocos retardos. En los procesos integrados las autocorrelaciones disminuyen linealmente con el tiempo y es posible encontrar coeficientes de autocorrelación distintos de cero hasta para retardos muy altos.

3.2.4.1.1 Procesos integrados ARIMA

Un proceso es integrado de orden $d \geq 0$, cuando al diferenciarlo h veces se obtiene un proceso estacionario. Un proceso estacionario es, por tanto, siempre de orden $d = 0$. Se obtienen considerando que el operador autorregresivo contiene la raíz $B=1$ con multiplicidad d , es decir, contiene el factor: $(1 - B)^d$. El resto de las soluciones están fuera de $|B| \leq 1$.

El proceso integrado ARIMA(p,d,q) se representa como:

$$\phi(B)(1 - B)^d X_t = \theta(B)a_t \quad (3.13)$$

Donde $\phi(B)$ es un operador autorregresivo estacionario, y $\nabla = (1 - B)$ es el operador diferenciador, podemos escribir la ecuación como:

$$\phi(B)\nabla^d X_t = \theta(B)a_t \quad (3.14)$$

Es decir, después de diferenciar d veces, se obtiene un proceso ARMA(p,q) estacionario. El nombre ARIMA proviene de las iniciales en inglés de los procesos autorregresivos integrados de media móvil (autoregressive integrated moving average).

3.2.4.1.2 Dos interpretaciones del modelo ARIMA

Para aproximar el modelo desde un punto de vista diferente, considere la situación donde $d=0$. El requerimiento de que los ceros de $\phi(B)$ caigan fuera del círculo unitario aseguraría no solo que el proceso X_t sea estacionario con media cero, también que $\nabla X_t, \nabla^2 X_t, \nabla^3 X_t, \dots$ sean cada uno de media cero. La figura 2(a) muestra un tipo de serie no estacionaria. Esta serie es homogénea excepto en nivel, exceptuando eso, para una traslación vertical, una parte de ella luce igual a la otra. Podemos representar dicho comportamiento conservando el requerimiento que cada una de las diferencias sea estacionaria con media cero, pero dejando al nivel "ser libre". Esto lo hacemos usando el modelo

$$\phi(B)\nabla X_t = \theta(B)a_t$$

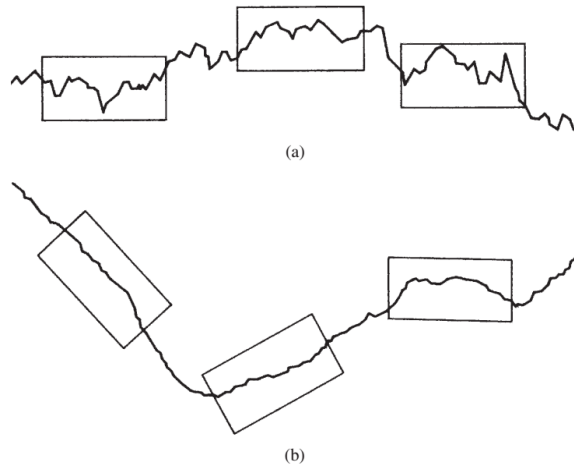


Figura 3.2. Distintos casos de estacionalidad

La figura 2(b) muestra otro tipo de serie no estacionaria de muy común ocurrencia. La serie no tiene ni un valor fijo, ni una pendiente fija, pero su comportamiento es homogéneo si permitimos diferencias en estas características. Podemos representar dicho comportamiento por el modelo

$$\phi(B)\nabla^2 X_t = \theta(B)a_t$$

El cual asegura estacionariedad y media cero para todas las diferencias después de la primera o segunda, pero permite al nivel y a la pendiente “ser libre”

Para la representación de series de tiempo, rara vez parece encontrar situaciones para las cuales p , d o q debe ser mayor que 2. Frecuentemente, los valores de cero o unidad serán apropiados para uno o más de estos órdenes.

3.3 Construcción del modelo estocástico.

3.3.1 Introducción.

Hemos visto que un modelo ARIMA de orden (p, d, q) proporciona una clase de modelos capaces de representar series temporales que, aunque no necesariamente estacionarias, son homogéneas y en equilibrio estadístico en muchos aspectos.

El modelo ARIMA está definido por la ecuación

$$\phi(B)(1 - B)^d X_t = \theta_0 + \theta(B)a_t \quad (3.15)$$

donde $\phi(B)$ y $\theta(B)$ son operadores en B de grado p y q , respectivamente, cuyos ceros se encuentran fuera del círculo unitario. Hemos notado que el modelo es muy general, incluye modelos autorregresivos, modelos de media móvil, modelos mixtos autorregresivo y de media móvil, y las formas integradas de los tres.

3.3.2 Enfoque iterativo para la construcción de modelos

Relacionar un modelo de este tipo con datos generalmente se logra mejor mediante un procedimiento iterativo de tres etapas basado en la identificación, estimación y comprobación de diagnóstico.

1. Por **identificación** nos referimos al uso de los datos, y de cualquier información sobre cómo se generó la serie, para sugerir una subclase de modelos parsimoniosos digno de ser considerado.
2. Por **estimación**, nos referimos al uso eficiente de los datos para hacer inferencias sobre los parámetros condicionados a la adecuación del modelo considerado.
3. Por **comprobación de diagnóstico** nos referimos a verificar el modelo ajustado en su relación con los datos con la intención de revelar las insuficiencias del modelo y así lograr la mejora del modelo.

3.3.2.1 Rol del término constante en el pronóstico

En general, sin embargo, podemos desear incluir una función determinista del tiempo $f(t)$ en el modelo. En particular permite una tendencia polinómica determinista de grado d , que se puede agregar permitiendo que θ_0 sea distinto de cero, lo que es equivalente a permitir:

$$E[w_t] = E[\nabla^d X_t] = \mu_w = \frac{\theta_0}{(1 - \phi_1 - \phi_2 - \dots - \phi_p)} \quad (3.16)$$

Una alternativa más general de representar el modelo es en la forma de un proceso ARMA invertible sustituyendo $\tilde{w}_t = w_t - \mu_w$

Como se discutió en el anteriormente, un valor distinto de cero para θ_0 en la ecuación (3.15) $\phi(B)w_t = \theta_0 + \theta(B)a_t$, donde $w_t = (1 - B)^d X_t$, implica la existencia de una tendencia polinómica sistemática de grado d . También notamos el impacto en el pronóstico de permitir un término constante distinto de cero θ_0 en el modelo ARIMA(p, d, q), un término determinístico polinomial de grado d estará presente. Esto es porque en la relación $\phi(B)\nabla^d \hat{X}_t(l) = 0$, el pronóstico ahora satisface $\phi(B)\nabla^d \hat{X}_t(l) = \theta_0$, para todo $l > q$, y el polinomio determinista de grado d representa una solución particular para esta ecuación en diferencia no homogénea. Por lo tanto, en caso de una constante distinta de cero, el modelo ARIMA también se puede expresar como $\phi(B)(\nabla^d \hat{X}_t(l) - \mu_w) = \theta(B)a_t$, $\mu_w \neq 0$ y el pronóstico puede ser visto de la forma $\tilde{X}_{t+l} = X_{t+l} - f(t+l)$, donde $f(t+l) = \left(\frac{\mu_w}{d!}\right)(t+l)^d + g(t+l)$, y $g(t)$ es cualquier polinomio determinístico de grado menor o igual a $d - 1$ [incluyendo la posibilidad $g(t) = 0$]. Note que, en el caso especial de un proceso estacionario, con $d = 0$, el término determinístico adicional, se reduce a la media del proceso $\mu = E[X_t]$.

En muchas aplicaciones, donde no existe una razón física para un componente determinista, se puede suponer que la media de w_t es cero a menos que tal suposición demuestre lo contrario a los hechos presentados en los datos. Está claro que para muchas aplicaciones, la suposición de una tendencia estocástica es a menudo más realista que la suposición de una tendencia determinista. Esto es de especial importancia para pronosticar una serie temporal, ya que una tendencia estocástica no requiere que la serie siga el patrón de tendencia idéntico que ha desarrollado en el pasado. En lo consiguiente, cuando $d > 0$, a menudo asumiremos que $\mu_w = 0$, o equivalentemente, que $\theta_0 = 0$, a menos que sea claro a partir de los datos o de la naturaleza del problema que una media distinta de cero, o más generalmente una componente determinista de forma conocida, es necesario.

3.3.3 IDENTIFICACION DE MODELO.

Los métodos de identificación son procedimientos aproximados aplicados a un conjunto de datos para indicar el tipo de modelo representacional que merece una investigación más a fondo. El objetivo específico aquí es obtener una idea de los valores de p , d y q necesarios en el modelo ARIMA lineal general y para obtener estimaciones iniciales para los parámetros. El modelo provisional así obtenido proporciona un punto de partida para la aplicación del método de estimación más formal.

Nuestro enfoque será como sigue:

1. Para evaluar la estacionariedad del proceso X_t y, si es necesario, para diferenciarlo tantas veces como sea necesario para producir estacionariedad, con la esperanza de reducir el proceso en estudio a un proceso mixto autoregresivo y de media móvil:

$$\phi(B)w_t = \theta_0 + \theta(B)a_t \quad (3.17)$$

Dónde:

$$w_t = (1 - B)^d X_t = \nabla^d X_t \quad (3.18)$$

2. Identificar el modelo ARMA de w_t

Nuestras herramientas principales para efectuar los pasos 1 y 2 serán la función de autocorrelación simple y la función de autocorrelación parcial. Se usan no solo para ayudar a adivinar la forma del modelo, sino también para obtener aproximaciones de los parámetros. Tales aproximaciones son a menudo útiles en la etapa de estimación para proporcionar valores iniciales para los procedimientos iterativos empleados en esa etapa. La identificación del modelo requiere identificar la estructura no estacionaria, si existe, y después la estructura ARMA estacionaria. La identificación de la estructura no estacionaria consiste en detectar qué transformaciones hay que aplicar para conseguir un proceso ARMA estacionario con varianza y media constante.

En particular, consiste en determinar si es necesario:

- a) transformar la serie para que tenga varianza constante
- b) diferenciar la serie para que tenga media constante. Posteriormente se identifica la estructura ARMA para la serie estacionaria.

3.3.4 Uso de las funciones de autocorrelación y autocorrelación parcial para la identificación del orden del proceso

3.3.4.1 El grado de diferenciación.

En la práctica, d es normalmente 0, 1 o 2, y es generalmente suficiente inspeccionar las primeras 20 o más autocorrelaciones estimadas de la serie original, y de su primera y segunda diferencia si es necesario.

3.3.4.2 Determinación de la diferenciación regular

Si la serie tiene tendencia, o muestra cambios de nivel en la media, mediante la aplicación de diferenciación regular $(1 - B)$ se puede transformar en una serie estacionaria. Cuando la decisión de diferenciar no sea clara, una característica que identifica una serie no estacionaria en la función de autocorrelación simple (*fas*) es que presenta un decrecimiento

lento en forma lineal. Peña (14) indica que si la *fas* no se amortigua para retardos altos (15 ó 20), en general es necesario diferenciar para obtener un proceso estacionario.

3.3.4.3 El proceso ARMA estacionario resultante.

En pocas palabras, mientras que la función de autocorrelación de un proceso autorregresivo de orden p disminuye hasta desvanecerse, su función de autocorrelación parcial tiene un punto de corte después del retardo p . Por el contrario, la función de autocorrelación de un proceso de media móvil de orden q tiene un corte después del retardo q , mientras que su función de autocorrelación parcial disminuye hasta desvanecerse. Si ambas funciones de autocorrelación y autocorrelación parcial disminuyen, se sugiere un proceso mixto.

Además, la función de autocorrelación para un proceso mixto, que contiene una componente autorregresivo de orden p y un componente de media móvil de orden q , es una mezcla de exponenciales y ondas sinusoidales amortiguadas después de los primeros retardos de $q - p$. Por el contrario, la función de autocorrelación parcial para un proceso mixto está dominada por una mezcla de exponenciales y ondas sinusoidales amortiguadas después de los primeros retardos $p - q$ (ver Tabla 3.3)

$f(k)$	(1,d,0)	(0,d,1)	(2,d,0)	(0,d,2)	(1,d,1)
ρ_k	Decaimiento exponencial.	Solo ρ_1 distinto de cero.	Mezcla de exponenciales u ondas senoidales amortiguadas.	Solo ρ_1 y ρ_2 distintas de cero.	Decaimiento exponencial desde el primer retardo.
ϕ_{kk}	Solo ϕ_{11} distinto de cero.	Amortiguamiento exponencial dominante.	Solo ϕ_{11} y ϕ_{22} distintas de cero.	Mezcla de exponenciales u ondas senoidales amortiguadas dominante.	Decaimiento exponencial desde el primer retardo dominante.

Tabla 3.3 Identificación de modelos por medio de sus funciones de autocorrelación simple y autocorrelación parcial.

3.3.5 Estimación de los parámetros del modelo.

Una vez que se haya determinado la posible estructura del modelo, se procede a estimar los parámetros del modelo.

Supuesto que la serie ha sido generada por un proceso estacionario e invertible, Habrá que estimar: ϕ , θ , σ_a .

Los métodos usualmente utilizados son los siguientes:

- a) Método de los momentos.
- b) Mínimos cuadrados.
- c) Máxima verosimilitud.

3.3.6 Relación entre Autocorrelaciones estimadas y teóricas.

Al emplear la función de autocorrelación estimada como una herramienta para la identificación, generalmente es posible estar bastante seguro acerca de las características

generales, pero más indicaciones sutiles pueden o no representar efectos reales, y dos o más modelos relacionados pueden ser necesitados para ser tomados en cuenta e investigarlos más a fondo en la etapa de estimación y etapa de comprobación de diagnóstico de la construcción del modelo.

Deben mencionarse otros tres puntos relacionados con este procedimiento de identificación:

1. La diferenciación simple del tipo que hemos utilizado no producirá estacionariedad en series que contengan componentes estacionales.

2. Para algunas series, un valor de $\theta_0 = 0$ puede perfectamente explicar el comportamiento estocástico de la serie. Ocasionalmente, sin embargo, habrá algunos fenómenos físicos que requiere la provisión de dicho componente determinístico. En otros casos podría no estar claro si se debería o no hacer tal disposición.

Alguna indicación de la evidencia provista por los datos, para la inclusión de θ_0 en el modelo, se puede obtener en la etapa de identificación comparando la media \bar{w} de $w_t = \nabla^d X_t$ con su error estándar aproximado, usando $\sigma^2(\bar{w}) = n^{-1} \sigma_w^2 [1 + 2\rho_1(w) + 2\rho_2(w) + \dots]$.

3. Se observó en la sección anterior que, para cualquier proceso ARMA(p, q) con $p - q > 0$, la mitad positiva completa de la función de autocorrelación será una mezcla de ondas sinusoidales amortiguadas y exponenciales. Esto, por supuesto, no nos impide identificar tentativamente q, porque (a) la función de autocorrelación parcial mostrará p - q valores "anómalos" antes de comportarse como un proceso MA (q), y (b) q debe ser tal de modo que la función de autocorrelación podría tomar, como valores iniciales siguiendo el patrón general, ρ_q de vuelta a $\rho_{-(p-q-1)}$.

3.4 PRONÓSTICOS

3.4.1 Pronósticos con el modelo

Peña (14), define los predictores óptimos como aquellos que minimizan en promedio los errores de predicción al cuadrado. Estos predictores se obtienen calculando los valores esperados de las observaciones futuras, condicionadas a los datos observados. La función de predicción de un modelo ARIMA tiene una estructura simple: los operadores no estacionarios, es decir, las diferencias y la constante si existe, determinan la predicción a largo plazo, mientras que los operadores estacionarios, AR y MA, determinan la predicción a corto plazo. Además, en la práctica existen cuatro tipos de incertidumbre en la predicción, asociadas al desconocimiento de: a) los errores aleatorios futuros, b) la distribución de los errores aleatorios, c) los verdaderos valores de los parámetros, y d) el modelo que ha generado los datos.

3.4.2 Tres formas explícitas para el modelo

Una observación X_{t+l} generada por el proceso ARIMA puede ser expresada de las siguientes maneras:

1. Directamente de la ecuación de diferencias donde $\varphi(B) = \nabla^d \phi(B)$.

$$X_{t+l} = \varphi_1 X_{t+l-1} + \dots + \varphi_{p+d} X_{t+l-(p+d)} - \theta_1 a_{t+l-1} - \dots - \theta_q a_{t+l-q} + a_{t+l} \quad (3.19)$$

2. Como una suma infinita ponderada de ruidos a_j actuales y previos.

$$X_{t+l} = \sum_{j=0}^{\infty} \psi_j a_{t+l-j} \quad (3.20)$$

3. Como una suma infinita ponderada de observaciones previas del proceso y un término de ruido blanco aleatorio.

$$X_{t+l} = \sum_{j=1}^{\infty} \pi_j X_{t+l-j} + a_{t+l} \quad (3.21)$$

Ahora suponemos, en el origen t , hacemos un pronóstico $\hat{X}_t(l)$ de X_{t+l} , que será una función lineal de observaciones actuales y previas $X_t, X_{t-1}, X_{t-2}, \dots$ entonces también será una función lineal de ruidos actuales y previos $a_t, a_{t-1}, a_{t-2}, \dots$, entonces tenemos

$$X_{t+l} = e_t(l) + \hat{X}_t(l) \quad (3.22)$$

Donde $e_t(l)$ es el error del pronóstico $\hat{X}_t(l) = \mathbf{E}[X_{t+l}|X_t, X_{t-1}, \dots]$ a un tiempo adelantado l , dado que se conocen los valores de X_t hasta un tiempo t . Y asumimos también que a_t es una secuencia de variables aleatorias independientes y que $\mathbf{E}[a_{t+j}] = 0 \forall j > 0$. El error del pronóstico a un paso de adelanto es entonces

$$e_t(l) = X_{t+1} - \hat{X}_t(1) = a_{t+1} \quad (3.23)$$

3.4.3 Tres formas básicas para el pronóstico.

1. *Predicción desde la ecuación de diferencias.*
2. *Predicción de forma integrada.*
3. *Predicción como un promedio ponderado de observaciones previas y predicción hecha en tiempos de adelanto previos desde el mismo origen*

Teniendo en cuenta lo siguiente para sustituir en cualquiera de las tres formas básicas para el pronóstico.

Si definimos a j como un entero no negativo.

1. Los X_{t-j} ($j = 0, 1, 2, \dots$), que ya han ocurrido en el origen t , se dejan sin cambios.
2. Los X_{t+j} ($j = 1, 2, 3, \dots$), que aún no han ocurrido, se reemplazan por sus predicciones $\hat{X}_t(j)$ en el origen t .
3. Los a_{t-j} ($j = 0, 1, 2, \dots$), que han ocurrido, están disponibles de $a_{t-j} = X_{t-j} - \hat{X}_{t-j-1}(1)$.
4. Los a_{t+j} ($j = 1, 2, 3, \dots$), que aún no han ocurrido, se reemplazan con ceros.

3.4.4 Uso de coeficientes ψ

3.4.4.1 Cálculo de los coeficientes ψ

Supongamos que se requieren los pronósticos a los tiempos $1, 2, \dots, L$. Para obtener límites de probabilidad para estos pronósticos y para permitir que nuevos pronósticos se calculen mediante un proceso de actualización de la anterior, es necesario calcular los pesos $\psi_1, \psi_2, \dots, \psi_{L-1}$. Esto se logra usando (3.3), es decir

$$\varphi(B)\psi(B) = \theta(B)$$

Donde

$$\varphi(B) = \nabla^d \phi(B)$$

Esto es, igualando coeficientes en potencias de B

$$(1 - \varphi_1 B - \dots - \varphi_{p+d} B^{p+d})(1 + \psi_1 B + \psi_2 B^2 + \dots) = (1 - \theta_1 B - \dots - \theta_q B^q)$$

Conociendo los valores de φ y de θ , los valores de ψ pueden obtenerse igualando coeficientes de B así:

$$\begin{aligned} \psi_1 &= \varphi_1 - \theta_1 \\ \psi_2 &= \varphi_1 \psi_1 + \varphi_2 - \theta_2 \\ \psi_j &= \varphi_1 \psi_{j-1} + \dots + \varphi_{p+d} \psi_{j-p-d} - \theta_j \end{aligned} \quad (3.24)$$

Donde $\psi_0 = 1$, $\psi_j = 0$ para $j < 0$ y $\theta_j = 0$ para $j > q$. Si K es el más grande de los enteros $(p + d - 1)$ y q , entonces para $j > K$

$$\psi_j = \varphi_1 \psi_{j-1} + \varphi_2 \psi_{j-2} \dots + \varphi_{p+d} \psi_{j-p-d} \quad (3.25)$$

Así los coeficientes ψ son fácilmente calculados de forma recursiva usando las ecuaciones (3.24) y (3.25).

3.4.4.2 Uso de los coeficientes ψ para actualizar un pronóstico.

Es interesante considerar otra manera de generar el pronóstico. De la ecuación (3.20) obtenemos.

$$\hat{X}_t(l) = \psi_l a_t + \psi_{l+1} a_{t-1} + \dots = E[X_{t+l}]$$

Podemos expresar los pronósticos $X_{t+1}(l)$ y $X_t(l+1)$, de la observación futura X_{t+l+1} , realizadas en los orígenes $t+1$ y t , como:

$$\begin{aligned} \hat{X}_{t+1}(l) &= \psi_l a_{t+1} + \psi_{l+1} a_t + \psi_{l+2} a_{t-1} + \dots \\ \hat{X}_t(l+1) &= \psi_{l+1} a_t + \psi_{l+2} a_{t-1} + \dots \end{aligned}$$

Sustrayendo las dos anteriores ecuaciones obtenemos

$$\hat{X}_{t+1}(l) = \hat{X}_t(l+1) + \psi_l a_{t+1} \quad (3.26)$$

Explícitamente, el pronóstico de origen t de X_{t+l+1} puede ser actualizado para convertirse en el pronóstico de origen $t+1$ del mismo X_{t+l+1} , sumando una constante múltiplo del error del pronóstico a un paso adelantado $a_{t+1} \equiv X_{t+1} - X_t(1) = e_t(l)$, con multiplicador ψ_l .

Esto lleva a una conclusión bastante notable. Supongamos que actualmente tenemos pronósticos en origen t para los adelantos $1, 2, \dots, L$. Entonces, tan pronto como se obtenga X_{t+1} , podemos calcular $a_{t+1} \equiv X_{t+1} - \hat{X}_t(1)$ y actualizar proporcionalmente para obtener pronósticos $X_{t+1}(l) = X_t(l+1) + \psi_l a_{t+1}$ en el origen $t+1$, para los adelantos $1, 2, \dots, L-1$. El nuevo pronóstico $X_{t+1}(L)$, para el tiempo de adelanto L , no se puede

calcular por este método, pero se obtiene fácilmente de los pronósticos de tiempos de adelanto más cortos, utilizando la ecuación diferencial.

3.4.4.3 *Calculo de los límites de probabilidad para un pronóstico en cualquier tiempo hacia adelante.*

De la ecuación (3.24) se tiene para un pronóstico con errores mínimos cuadrados:

$$\hat{X}_t(l) = \psi_l a_t + \psi_{l+1} a_{t-1} + \dots = E[X_{t+l}]$$

Una conclusión, que es un caso especial de resultados más generales en la teoría de predicción gracias a Wold, Kolmogoroff, Wiener, y Wittle, entonces tenemos:

$$\begin{aligned} X_{t+l} &= (a_{t+l} + \psi_1 a_{t+l-1} + \dots + \psi_{l-1} a_{t+1}) + \psi_l a_t + \psi_{l+1} a_{t-1} + \dots \\ &= e_t(l) + \hat{X}_t(l) \quad (3.27) \end{aligned}$$

El error del pronóstico para cualquier tiempo adelantado es entonces:

$$e_t(l) = a_{t+l} + \psi_1 a_{t+l-1} + \dots + \psi_{l-1} a_{t+1} \quad (3.28)$$

También la varianza del pronóstico es:

$$V(l) = var[e_t(l)] = (1 + \psi_1^2 + \psi_2^2 + \dots + \psi_{l-1}^2) \sigma_a^2 = \left(1 + \sum_{j=1}^{l-1} \psi_j^2 \right) \sigma_a^2 \quad (3.29)$$

Asumiendo una distribución normal en las componentes de ruido blanco, la distribución de probabilidad condicional $p(X_{t+l}|X_t, X_{t-1}, \dots)$ de un valor futuro X_{t+l} del proceso será normal con media $\hat{X}_t(l)$ y una desviación estándar dada por:

$$\sigma(l) = \left(1 + \sum_{j=1}^{l-1} \psi_j^2 \right)^{1/2} \sigma_a$$

Se demostró anteriormente como estimar s_a^2 , de la varianza σ_a^2 , puede ser obtenida de datos de la serie de tiempo, cuando el número de observaciones de las cuales se basa la estimación es mayor a 50, los límites de probabilidad estarán determinados por

$$X_{t+l}(\pm) = \hat{X}_t(l) \pm u_{\varepsilon/2} \left(1 + \sum_{j=1}^{l-1} \psi_j^2 \right)^{1/2} s_a \quad (3.30)$$

Donde $u_{\varepsilon/2}$ es la desviación que excede una proporción $\varepsilon/2$ de la distribución normal unitaria.

3.5 Modelos estacionales.

3.5.1 ¿Cómo identificar la estacionalidad?

Se considera la estacionalidad como un tipo de falta de estacionariedad, que hace que la media de las observaciones no sea constante, pero evoluciona de forma previsible de acuerdo a un patrón cíclico. Se dice que una serie es estacional cuando su valor esperado no es constante, pero varía con una pauta cíclica, de manera que:

$$E[X_t] = E[X_{t+s}]$$

Se dice que la serie tiene una estacionalidad de período s .

Además, se puede convertir una serie estacional en estacionaria aplicando una diferencia estacional, el operador diferencia estacional de período s será $(1 - B^s)$. En condiciones bastante generales para la estructura estacional, y tanto para procesos con estacionalidad determinista como estocástica, el operador convierte un proceso estacional en estacionario. El caso más habitual es que podamos incorporar la estacionalidad dentro del modelo ARIMA de forma multiplicativa, con lo que resulta el modelo ARIMA estacional multiplicativo.

3.5.2 Operador Simplificador $1 - B^s$

El hecho fundamental acerca de las series temporales estacionales con el período s es que las observaciones separadas por s intervalos son similares. Por lo tanto, se podría esperar que la operación $B^s X_t = X_{t-s}$ jugaría un papel particularmente importante en el análisis de series estacionales y, además, dado que no se espera estacionariedad en las series $X_t, X_{t-s}, X_{t-2s}, \dots$, la operación de simplificación:

$$\nabla_s X_t = (1 - B^s)X_t = X_t - X_{t-s} \quad (3.31)$$

Podría ser útil. Este operador no estacionario $(1 - B^s)$ tiene s ceros en $e^{-i(2\pi k/s)}$ ($k = 0, 1, \dots, s - 1$) espaciado uniformemente en el círculo unitario. Además, la función de predicción eventual satisface $(1 - B^s)\hat{X}_t(l) = 0$ y, por lo tanto, puede (pero no necesita) representarse mediante un complemento completo de senos y cosenos, así:

$$\hat{X}_t(l) = b_0^{(t)} + \sum_{j=1}^{s/2} \left[b_{1j}^{(t)} \cos\left(\frac{2\pi jl}{s}\right) + b_{2j}^{(t)} \sin\left(\frac{2\pi jl}{s}\right) \right] \quad (3.32)$$

Donde las b 's son coeficientes adaptativos, y donde $[s/2] = \frac{1}{2}s$ si s es par y $[s/2] = \frac{1}{2}(s - 1)$ si s es impar.

Con respecto a los datos anuales, el efecto estacional implica que una observación para un mes en particular, por ejemplo, abril, está relacionada con las observaciones de los abrils anteriores. Supongamos que la t -ésima observación X_t es para el mes de abril. Podríamos vincular esta observación X_t a las observaciones en abrils anteriores mediante un modelo de la forma

$$\Phi_P(B^s)\nabla_s^D X_t = \Theta_Q(B^s)\alpha_t \quad (3.33)$$

donde $s = 12$, $\nabla_s = (1 - B^s)$, y $\Phi(B^s), \Theta(B^s)$ son polinomios en B^s grados P y Q , respectivamente, y condiciones satisfactorias de estacionariedad e invertibilidad. Del mismo modo, un modelo

$$\Phi_P(B^s)\nabla_s^D X_{t-1} = \Theta_Q(B^s)\alpha_{t-1} \quad (3.34)$$

podría usarse para vincular el comportamiento actual de marzo con las observaciones de marzo anteriores, y así sucesivamente, para cada uno de los 12 meses. Además, sería usualmente razonable asumir que los parámetros Θ y Φ contenidos en estos modelos mensuales serían aproximadamente los mismos para cada mes.

Ahora las componentes de error $\alpha_t, \alpha_{t-1}, \dots$, en estos modelos podría en general no estar correlacionado. Por ejemplo, el comportamiento anual de una serie de tiempo, mientras está relacionado con los totales de abril previos, también estaría relacionado con los totales en los meses anteriores, marzo, febrero, enero, y así sucesivamente. Por lo tanto, esperaríamos que α_t en la ecuación (3.33) estaría relacionado con α_{t-1} en (3.34) y α_{t-2} , y así sucesivamente. Por lo tanto, para cuidar de tales relaciones, presentamos un segundo modelo

$$\phi(B)\nabla^d\alpha_t = \theta(B)a_t \quad (3.35)$$

donde ahora a_t es un proceso de ruido blanco y $\phi(B)$ y $\theta(B)$ son polinomios en B de grados p y q, respectivamente, y condiciones de estacionariedad e invertibilidad satisfactorias, y $\nabla = \nabla_1 = (1 - B)$. Sustituyendo (3.35) en (3.33), finalmente obtenemos un modelo multiplicativo general

$$\Phi_P(B^S)\phi_p(B)\nabla^d\nabla_S^D X_t = \Theta_Q(B^S)\theta_q(B)a_t \quad (3.36)$$

Dónde:

$\Phi_P(B^S) = (1 - \Phi_1 B^S - \Phi_2 B^{2S} \dots - \Phi_P B^{PS})$ es el operador AR estacional de orden P.

$\phi_p(B) = (1 - \phi_1 B - \phi_2 B^2 - \phi_3 B^3 \dots - \phi_p B^p)$ es el operador AR regular de orden p.

$(1 - B^S)^D$ representa las diferencias estacionales; y $(1 - B)^d$ representa las diferencias regulares.

$\Theta_Q(B^S) = (1 - \Theta_1 B^S - \Theta_2 B^{2S} \dots - \Theta_Q B^{QS})$ es el operador MA estacional de orden Q.

$\theta_q(B) = (1 - \theta_1 B - \theta_2 B^2 - \theta_3 B^3 \dots - \theta_q B^q)$ es el operador MA regular de orden q.

a_t es un proceso de ruido blanco.

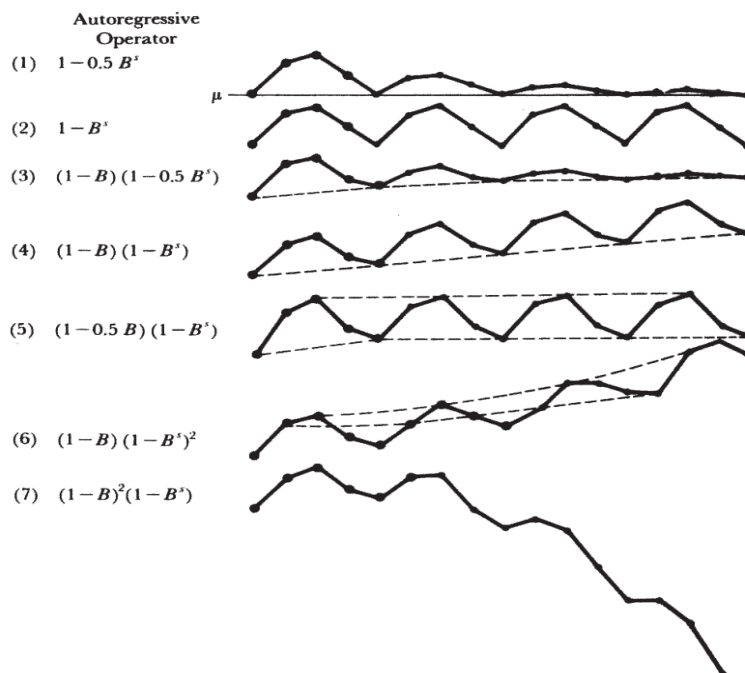


Figura 3.3. Comportamiento de la función de predicción estacional para varias alternativas del operador general autor regresivo estacional.

3.5.3 Determinación de la diferenciación estacional

Si la serie presenta estacionalidad, se procede a aplicar una diferencia estacional $(1 - B^s)$, para conseguir una serie estacionaria. La estacionalidad se manifiesta:

- a) En el gráfico de la serie, se presenta una pauta repetida de período s .
- b) En la función de autocorrelación simple, se presentan coeficientes positivos que decrecen lentamente en los retardos $s, 2s, 3s, \dots$

3.5.4 Notación de Modelos estacionales ARIMA y estacionalidad multiplicativa.

Dentro de estos modelos se encuentran los modelos estacionales ARIMA que se denotan anteponiéndoles la letra S y el orden de sus parámetros se escribe con mayúscula, así:

- Modelos Autorregresivos Estacionales SAR(P)
- Promedios Móviles Estacionales SMA(Q)
- Modelo Autorregresivo y de Promedios Móviles Estacionales SARMA (P, Q).

Los modelos SARMA son análogos al proceso ARMA pero considerando los retardos del ruido blanco y del proceso de s en s . Sin embargo, estos modelos SARMA no son capaces de explicar todos los movimientos estacionales, pues si estos crecieran de año en año, los SARMA serían incapaces de recoger esta evolución pues al igual que los ARMA son estacionarios.

Esta dificultad se resuelve a través de los modelos autorregresivos de promedios móviles integrados estacionales **SARIMA (P, D, Q)**.

La unión de modelos estacionales con modelos no estacionales conduce a un modelo de gran capacidad de adaptación que puede reflejar tanto la tendencia como la estacionalidad de una serie (enfoque de Box y Jenkins). La combinación de estos modelos se logra a través de la multiplicación de los operadores polinomiales que caracterizan a cada modelo obteniendo los modelos conocidos como **ARIMA(p,d,q)*SARIMA(P,D,Q)**. Por lo general los parámetros que no se consideran son de orden cero, como se describe a continuación:

- p : orden de un modelo autorregresivo AR
- d : orden de diferenciación en la parte regular o no estacionaria de la serie
- q : orden de un modelo promedio móvil MA
- P : orden de un modelo autorregresivo estacional SAR
- D : orden de diferenciación en la parte estacional de la serie
- Q : orden de un modelo promedio móvil estacional SMA

3.6 Metodología Box-Jenkins para ajustar los modelos a series reales.

3.6.1 Algoritmo para la aplicación del método.

Uno de los procedimientos más comúnmente discutidos para realizar pronósticos de demanda de energía es el análisis de series de tiempo, normalmente utilizando la metodología descrita por Box y Jenkins (15), donde se propone realizar un ajuste de los modelos en tres etapas.

La primera etapa consiste en *identificar* el posible modelo ARIMA que sigue la serie de tiempo, lo que requerirá:

- a) decidir qué transformaciones aplicar para convertir la serie observada en una serie estacionaria.
- b) determinar un modelo ARMA para la serie estacionaria, es decir, los órdenes p y q de su estructura autorregresiva y de media móvil y , si el proceso es estacional, los órdenes P , Q de la estructura ARMA estacional.

Una vez se haya seleccionado un posible modelo para la serie estacionaria, en la segunda etapa se procede a la *estimación*, donde los parámetros AR y MA del modelo se estiman por máxima verosimilitud y se obtienen sus errores estándar y los residuos del modelo. La tercera etapa consiste en realizar la *diagnos*, donde se comprueba que los residuos no tienen estructura de dependencia y siguen un proceso de ruido blanco. Si los residuos no contienen información, se toma como válido el modelo para posteriormente utilizarlo para la previsión o la toma de decisiones. Si los residuos muestran estructura, se procederá a modificar el modelo para incorporarla y repetir las tres etapas anteriores hasta obtener un modelo adecuado. En la figura 4 se presenta el algoritmo a seguir.

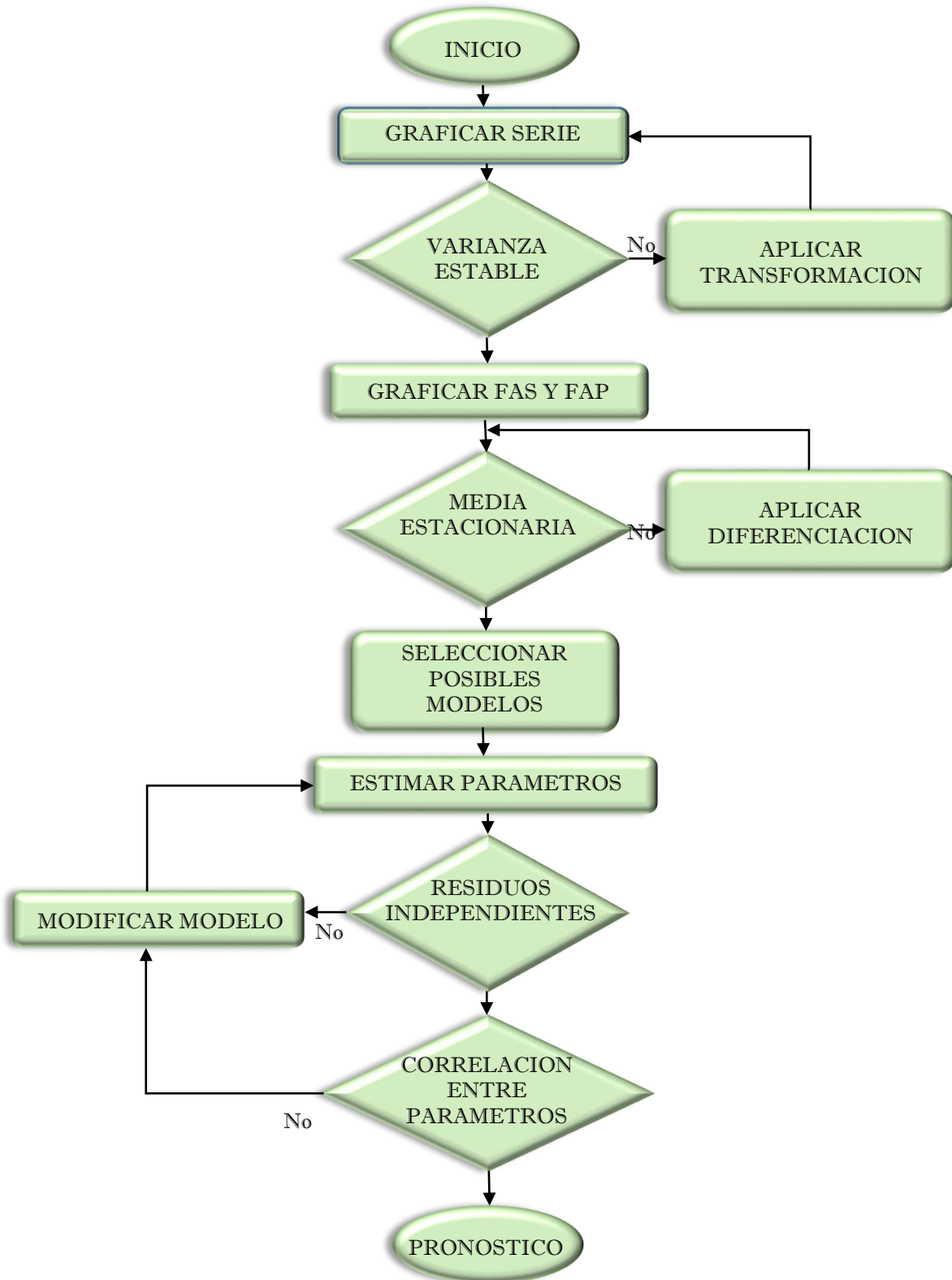


Figura 3.4. Metodología de Box-Jenkins para la selección de un modelo adecuado

3.6.2 Determinación de la transformación para estabilizar la varianza

El procedimiento para la determinación de la transformación a realizar en la serie, consiste en realizar una comparación homogénea, entre los diferentes años haciendo un gráfico entre una medida de variabilidad, como la desviación típica, y una medida del nivel, como la media local. Dado que una serie puede presentar una variabilidad mayor cuando la serie toma valores altos. Para hacer comparaciones homogéneas, si la serie es estacional de período s , se deberá hacer grupos de ese tamaño, para incluir todos los períodos estacionales en todos los grupos. Para series no estacionales, se toma un tamaño de grupo de cinco o seis observaciones contiguas. Una vez realizados los grupos homogéneos se procede a calcular en cada período la desviación típica y la media.

A continuación, se representa la desviación típica de cada período como función de la media y se procede a analizar la relación de dependencia entre ambas variables, para ello se realiza una transformación de logaritmo natural a la desviación típica y a la media de los grupos homogéneos, para definir la siguiente relación de dependencia:

$$\hat{a}_t = \tilde{w}_t - \sum_{j=1}^p \hat{\phi}_j \tilde{w}_{t-j} + \sum_{j=1}^q \hat{\theta}_j \hat{a}_{t-j} \quad t = 1, 2, 3, \dots, n \quad (3.37)$$

$$\log S_i = \alpha_0 + \alpha \log \bar{X}_i \quad (3.38)$$

Dónde:

S_i es la desviación típica del i -ésimo grupo homogéneo.

\bar{X}_i es la media del i -ésimo grupo homogéneo.

α_0 es la constante del modelo de dependencia.

α es la relación de dependencia entre la desviación típica y la media.

La dependencia de la variabilidad del nivel de la serie puede ser el resultado de que la serie se genera como producto de un componente sistemático o predecible, que define el nivel de la serie. Cuando en la serie se observa una relación entre el nivel y la variabilidad, se estima el valor de α mediante la aplicación de regresión lineal, y una vez obtenido el valor de α , la valoración a realizar es:

a) Si, $\alpha \approx 1$, entonces puede considerarse que no hay relación de dependencia del valor de la desviación con respecto al nivel de la serie, por lo que no habría necesidad de realizar transformación sobre la misma.

b) Si $\alpha \approx 0$, entonces puede considerarse que sí hay una relación de dependencia del valor de la variabilidad de la serie con el nivel de la misma, por lo que habría que proceder a realizar una transformación logaritmo natural para obtener una serie con varianza constante.

c) Para un valor intermedio obtenido para α puede considerarse proceder a realizar un tipo de transformación Box-Cox con el objetivo de lograr que la varianza sea independiente de la media.

3.6.3 Determinación de la transformación para estabilizar la media del proceso estacional

Se realiza en 2 pasos:

3.6.3.1 Determinar la diferenciación Regular d

Si la serie tiene tendencia, o muestra cambios de nivel en la media, mediante la aplicación de diferenciación regular $(1 - B)$ se puede transformar en una serie estacionaria. Cuando la decisión de diferenciar no sea clara, una característica que identifica una serie no estacionaria en la función de autocorrelación simple (fas) es que presenta un decrecimiento lento en forma lineal. Peña (2005) indica que si la fase no se amortigua para retardos altos (15 ó 20), en general es necesario diferenciar para obtener un proceso estacionario.

3.6.3.2 Determinar la diferenciación estacional D

Si la serie presenta estacionalidad, se procede a aplicar una diferencia estacional $1-B_s$, para conseguir una serie estacionaria. La estacionalidad se manifiesta:

- En el gráfico de la serie, se presenta una pauta repetida de período s .
- En la función de autocorrelación simple, se presentan coeficientes positivos que decrecen lentamente en los retardos $s, 2s, 3s, \dots$

3.6.4 Diagnósis del modelo.

3.6.4.1 Finalidad de las pruebas.

La diagnósis del modelo requiere comprobar que las hipótesis básicas realizadas respecto a los residuos son ciertas. Estos deben tener: a) media igual a cero; b) varianza constante; c) falta de correlación para cualquier retardo; d) distribución normal. Adicionalmente, habrá que realizar un análisis de los resultados de la estimación, que son: a) estimadores significativos; b) condiciones de estacionariedad e invertibilidad; c) correlación entre parámetros. No se puede enfatizar demasiado que la inspección visual de una parte de los residuos en sí, es un primer paso indispensable en el proceso de verificación.

3.6.4.2 Prueba de autocorrelaciones

Suponga que un modelo del cual han sido estimados los parámetros $(\hat{\phi}, \hat{\theta})$, entonces nos referimos a \hat{a}_t como residuos. Estos son calculados recursivamente de $\hat{\theta}(B)\hat{a}_t = \hat{\phi}(B)\tilde{w}_t$:

$$\hat{a}_t = \tilde{w}_t - \sum_{j=1}^p \hat{\phi}_j \tilde{w}_{t-j} + \sum_{j=1}^q \hat{\theta}_j \hat{a}_{t-j} \quad t = 1, 2, 3, \dots, n \quad (3.39)$$

Mientras la longitud de la serie se incrementa, \hat{a}_t se aproxima al ruido blanco a_t . Ahora, en la práctica no conocemos los verdaderos valores de los parámetros. Solo conocemos sus estimaciones $(\hat{\phi}, \hat{\theta})$, de las cuales usando (3.37) calculamos \hat{a}_t . Las autocorrelaciones $r[\hat{a}_t]$, pueden conducir a valiosa evidencia de falta de ajuste en el modelo y la posible naturaleza de la inadecuación del modelo. Sin embargo, Durbin (16) señaló que podría ser peligroso evaluar la significancia estadística de aparentes discrepancias de estas autocorrelaciones $r_k(\hat{a}_t)$ a partir de sus valores cero teóricos sobre la base de un error estándar $n^{-1/2}$, apropiado para el $r_k(\hat{a}_t)$.

Si para los retardos bajos usamos el error estándar $n^{-1/2}$ para los $r_k(\hat{a}_t)$, podemos subestimar seriamente la importancia de las aparentes discrepancias.

Prueba de falta de ajuste: Contraste sobre las autocorrelaciones.

En lugar de considerar las $r_k(\hat{a}_t)$ individualmente, un indicador a menudo es requerido de ya sea, las primeras 20 autocorrelaciones de \hat{a}_t tomadas como un total indica inadecuación del modelo. Suponga que tenemos las primeras K autocorrelaciones $r_k(\hat{a}_t)$ ($k = 1, 2, 3, \dots, K$) para cualquier modelo ARIMA(p, d, q); entonces es posible mostrar si el modelo ajustado es apropiado. Observando el gráfico de autocorrelación de los residuos se puede determinar independencia o no del mismo, adicionalmente se debe realizar una prueba estadística que lo fundamente matemáticamente. Una prueba estadística adecuada es la hipótesis de independencia de Ljung-Box, el cual permite contrastar la hipótesis nula de que todos los coeficientes de correlación son nulos. Si el p-valor asociado al estadístico Ljung-Box es menor que α , se rechazará la hipótesis nula al nivel de significancia α . El estadístico Q de Ljung-Box está definido como:

$$\tilde{Q} = N(N + 2) \sum_{k=1}^K \frac{\hat{r}_k^2(\hat{a})}{(N - k)} \quad (3.40)$$

Que tiene aproximadamente la media $E(Q) \approx K - n$ distribuido como una $\chi^2(K - n)$, donde N es el número de valores de la serie diferenciada o estacionaria usados para ajustar el modelo. El estadístico se distribuye, asintóticamente, como una χ^2 con grados de libertad igual al número de coeficientes en la suma K menos el número de parámetros estimados, n. Para modelos no estacionales ($n=p+q+1$), o ($n=p+q$), según que el modelo tenga o no constante θ_0 , y para los estacionales, que habitualmente no tienen constante, ($n = p+q+P+Q$). Se considerará un modelo como inadecuado si el valor de obtenido es mayor que el percentil 0.95 de la distribución con (K-n) grados de libertad, que se denota por $\chi_{0.95}^2(K - n)$. En general, rechazaremos la hipótesis de incorrelación de los residuos cuando la probabilidad:

$$P(\chi^2(K - n)) > Q(h)$$

Sea pequeña (menor que 0.05 o 0.01).

3.6.5 Contraste de media cero

Se debe determinar que la media de los residuos de las componentes de ruido blanco es cero, donde aplicando el contraste estadístico “t-student” para estimar la media de una población normalmente distribuida cuando el tamaño de la muestra es pequeño, tal que:

$$t_{\mu a} = \frac{\mu_a}{\sigma_a} \sim N(0,1) \quad (3.41)$$

Y donde la hipótesis nula $H_0: \mu_a = 0$ es aceptada para un nivel de significancia α , cuando $|t_{\mu a}| < N_{\frac{\alpha}{2}, K-1}(0,1)$, caso contrario es rechazada.

Esta prueba se realiza para un número K relativamente pequeño de muestras de valores residuales.

3.6.6 Contraste de varianza estable (homocedasticidad)

La estabilidad de la varianza de los residuos se comprueba estudiando el gráfico de los residuos a lo largo del tiempo. Peña (2005), expone que, si a la vista de los residuos

estimados parece existir un cambio de varianza a partir de un punto, $t=n1$, se puede dividir el intervalo muestral en dos partes y aplicar un contraste de varianzas. En la hipótesis de que ambos tramos tienen la misma varianza, el estadístico:

$$F = \frac{\sum_{t=1}^{n1} a_t^2 / n1}{\sum_{t=n1+1}^N a_t^2 / (N - n1)} = \frac{S_1^2}{S_2^2} \quad (3.42)$$

Se distribuirá, aproximadamente, como una F-Snedcor con $n1$ y $N - n1$ grados de libertad. Esto también se conoce como “test F” para evaluación de homocedasticidad.

Se puede ocupar como alternativa al test F, el test de Levene. Este se caracteriza por comparar 2 o más poblaciones, y por permitir elegir entre diferentes estadísticos de centralidad (mediana, media, media truncada).

3.6.7 Contraste de normalidad

La hipótesis de que los residuos tienen una distribución normal se comprueba con cualquiera de los contrastes habituales. La prueba de Lilliefors (Kolmogorov-Smirnov), es una prueba no paramétrica de “buen de ajuste”. Se utiliza para contrastar si los residuos se distribuyen según una distribución normal, el contraste trata de verificar la hipótesis nula H_0 : La distribución observada se ajusta a una distribución normal, contra la hipótesis alternativa H_1 : La distribución observada no se ajusta a la distribución normal. Cuando la prueba de (Kolmogorov-Smirnov) se aplica para contrastar la hipótesis de normalidad de la población, el estadístico de prueba es la máxima diferencia:

$$D = \max |F_n(X) - F_0(X)| \quad (3.43)$$

Siendo $F_n(X)$ la función de distribución de los residuos y $F_0(X)$ la función teórica o correspondiente a la población normal especificada en la hipótesis nula. La distribución del estadístico de Kolmogorov-Smirnov es independiente de la distribución poblacional especificada en la hipótesis nula y los valores críticos de este estadístico están tabulados, de esto se obtiene el valor p con el cual se toma la decisión sobre la hipótesis planteada al inicio, si el valor p es mayor o igual que α , no se rechaza H_0 .

3.7 Valores atípicos

Con frecuencia ocurren en las series hechos puntuales que no han sido generados igual que los demás. Las observaciones afectadas por esas intervenciones pueden representar una estructura diferente de las demás y aparecer como datos atípicos.

Pena(15), Señala que es importante ser capaces de identificar estas situaciones desconocidas y separarlas de la dinámica habitual de la serie porque:

- a) Si sus efectos son grandes, pueden sesgar la estimación de los parámetros, lo que producirá malas predicciones futuras.
- b) Si el suceso ha ocurrido en la última parte de la serie y alguna observación afectada se utiliza para generar predicciones estas no serán buenas, incluso aunque los parámetros estén bien estimados.
- c) Si estos sucesos atípicos pueden volver a aparecer en el futuro y los identificamos y estimamos, podemos incorporar esta información en las predicciones y obtener intervalos de predicción más realistas.

La identificación de atípicos consiste en detectar observaciones que parecen haber sido generadas de forma distinta al resto. Posteriormente, se investigan las causas de estos valores atípicos y, se procede a incorporar estas observaciones en un modelo explícito. Esta identificación se realiza mediante la huella que los atípicos dejan en la serie, se pueden considerar dos tipos de atípicos: los aditivos y los innovativos.

3.7.1 Atípicos aditivos (AO)

Se considera que ha ocurrido un atípico aditivo (AO) sobre una serie temporal en el instante h si el valor de la serie se genera en ese instante de manera distinta al resto. El modelo que seguirá la serie observada si ha sido afectada por un AO en t es, sustituyendo en la ecuación (3.20):

$$X_t = \begin{cases} y_t, & t \neq h \\ y_t + W_A, & t = h \end{cases} \quad (3.44)$$

$$X_t = W_A I_t^h + \psi(B) a_t$$

Donde y_t es la serie no contaminada por atípicos y:

$$I_t = \begin{cases} 0, & t \neq h \\ 1, & t = h \end{cases}$$

El nivel de la serie observada resulta afectado en $t=h$ en que se produce el atípico. Una forma equivalente de escribir el modelo ARIMA para la serie es con la ecuación (3.21):

$$\pi(B)y_t = a_t$$

Donde se verifica que $\psi(B)\pi(B) = 1$, y el modelo resultante para un AO es:

$$\pi(B)(X_t - W_A I_t^h) = a_t \quad (3.45)$$

La huella que un atípico aditivo deja en la serie es debida a la alteración de su valor en un punto, cuando se desconozca su presencia y se construya un modelo ARIMA su efecto puede detectarse con los residuos del modelo.

3.7.2 Atípicos innovativos (IO)

Un atípico innovativo (IO) ha ocurrido en la serie temporal en $t=h$ cuando la innovación en ese punto está directamente afectada por una cantidad desconocida debida a un suceso imprevisto. La innovación de una serie temporal representa el efecto agregado no sistemático de todas las variables que influyen sobre la serie. Un atípico sobre la innovación puede interpretarse como un cambio imprevisto en una o varias de las variables que afectan a la serie. Como la innovación representa el error de previsión a un paso, se puede decir que un IO es un incremento del error de previsión de la serie en un punto debido a un cambio en las condiciones externas. El modelo obtenido para la serie sería:

$$X_t = \psi(B)(W_I I_t^h + a_t) \quad (3.46)$$

Si se toma que $e_t = \pi(B)X_t$, de manera similar a los atípicos aditivos, a los residuos calculados suponiendo que la serie está libre de atípicos y que se conocen los parámetros del modelo, la relación entre estos residuos y los verdaderos es:

$$e_t = W_I I_t^h + a_t \quad (3.47)$$

Que se puede representar como:

$$e_t = W_I + a_h$$

Y para $j > h$

$$e_{h+j} = a_{h+j}$$

Cuando ocurre un IO en $t=h$ los residuos estimados (conociendo los parámetros del proceso) serán iguales a las innovaciones en todos los puntos excepto en $t=h$. El efecto de un IO sobre la serie es muy distinto que el de un AO porque este último produce un efecto fijo, una alteración en la observación, mientras que el efecto de un IO sobre la serie depende del modelo. El efecto en la serie de un IO se propaga siguiendo el mismo patrón dinámico que esta sigue.

4 Redes neuronales en Matlab

4.1 Creación de la red

En Matlab se crea una red perceptron multicapa con conexiones hacia delante utilizando la función *newff*. Esta función devuelve una variable que representa a la red. Los argumentos de entrada de esta función (descritos en el orden en el que se colocan) son (10):

Sintaxis

net = newff (PR, [S1 S2 ... SNI], {TF1 TF2 ... TFN1}, BTF, BLF, PF)

- PR que es una matriz cuyo número de filas es el número de entradas, y su número de columnas es igual a 2. Esta matriz debe señalar los valores mínimos y máximos que pueden tomar cada una de las entradas a la red.
- [*Slayer1, Slayer2,...*] es un vector que describe el número de neuronas que tienen todas las capas ocultas de la red y el número de neuronas de salida. Asimismo el tamaño de dicho vector permite conocer al programa el número de capas que debe tener la red que se va a crear.
- {*Funciones*} es un vector de varias cadenas de caracteres en el que se señala las funciones de activación que va a poseer todas las neuronas de una capa. Así, la primera cadena señalará la función de activación de las neuronas de la primera capa oculta, la segunda señalará la función de las neuronas de la segunda capa, y así sucesivamente. Por ello, este vector debe tener tantas cadenas como capas vaya a tener nuestra red. Para que la función de activación sea lineal, la cadena debe ser '*purelin*', para que ésta sea una función tangente hiperbólica se debe escribir '*tansig*', y para indicar que queremos una función sigmoideal escribiremos '*logsig*'. Por defecto Matlab utiliza la función tangente hiperbólica.
- BTF - Función de entrenamiento de red Backprop, default = 'trainlm'.
- BLF - Función de aprendizaje del peso / sesgo de Backprop, default = 'learnngdm'.
- PF - Función de rendimiento, default = 'mse'.

4.2 Simulación de la red

Para obtener las salidas de una determinada red ante unas ciertas entradas se utiliza la función *sim*. Esta función devuelve un vector cuyas componentes son las salidas que se obtienen de cada neurona de salida de la red. A dicha función se le introducen como argumentos el nombre del objeto que representa a la red en Matlab, y el vector de entradas a la red. Se pueden obtener varias simulaciones de un tiempo introduciendo una matriz, cuyos vectores sea cada uno de los vectores de entrada de los que se quiere obtener su salida. De esta manera, si se tiene una red "*net*" de la que se quiere obtener su respuesta ante unas entradas "*a*", se escribe en Matlab la siguiente línea de código. (10)

Sintaxis

[Y,Pf,Af,E,perf] = sim(net,P,Pi,Ai,T)

Dónde:

Net: red.

P: Entradas de red.

Pi: condiciones iniciales de retraso de entrada, por defecto = ceros.

Ai: Condiciones de retraso de la capa inicial, por defecto = ceros.

T: Objetivos de red, predeterminado = ceros.

Y regresa,

Y: salidas de red.

Pf: Condiciones finales de retraso de entrada.

Af: Condiciones de retraso de capa final.

E: Errores de red.

Perf: rendimiento de la red.

4.3 Entrenamiento

Como se ha comentado anteriormente, el entrenamiento de la red consiste en presentarle unas entradas y sus correspondientes salidas (que son conocidas por nosotros), para que la red vaya reajustando su salida mediante la modificación de sus pesos y valores umbrales, de manera que el error de actuación de la red se minimice. Dicho valor Matlab lo almacena en la variable *net.performFcn*. La medida del error por defecto en Matlab es el error medio cuadrático. (10)

Para entrenar una red en Matlab se utiliza la función *train*. Dicha función utiliza como argumentos el nombre de la red que se quiere entrenar (y que anteriormente ha debido ser creada) y los patrones, compuestos por unos vectores de entradas (unidos en una matriz "a") y sus correspondientes salidas (todas ellas en una matriz "p"). De esta manera la línea de código que se debería escribir es (10):

[net,pr] = train (net,a,p);

La variable *pr* contiene información sobre el proceso de entrenamiento y la variable *net* contiene a la red ya entrenada, es decir con sus pesos y valores umbrales ajustados. Por otro lado, destacar que esta función utiliza una serie de variables para definir el entrenamiento y que pueden definirse con anterioridad. Estas variables varían según el tipo de entrenamiento utilizado por lo que se hablará de ellas cuando se describan como se definen los diferentes algoritmos de aprendizaje en Matlab. (10)

Para utilizar este entrenamiento básico (regla delta generalizada), se introduce en el argumento correspondiente al entrenamiento la cadena 'traingd'. Existen siete parámetros asociados con este tipo de entrenamiento. Estos son (10):

- *epochs*: Define el máximo número de épocas de entrenamiento que puede tener nuestro proceso de aprendizaje.
- *show*: Indica a Matlab la forma de visualización que deseamos tener durante el entrenamiento de la red. Si su valor es *Nan* quiere decir que no se quiere ningún tipo de visualización.
- *goal*: Este variable indica un valor mínimo límite que puede alcanzar la función error de la red. Si ésta alcanza dicho valor el entrenamiento se parará automáticamente.
- *time*: Este parámetro indica el tiempo máximo en segundos que puede durar el entrenamiento de la red. Una vez que el tiempo del proceso alcance dicho valor el entrenamiento se detendrá.
- *min_grad*: Determina el valor mínimo necesario que debe tener el gradiente para detener el algoritmo.
- *max_fail*: Es el máximo número de iteraciones que puede incrementarse el error de validación antes de detenerse el entrenamiento.
- *lr*: Es el ratio de aprendizaje α .

trainidx es una función de entrenamiento en red que actualiza los valores de peso y sesgo de acuerdo con el ímpetu de gradiente de descenso y una tasa de aprendizaje adaptativo. con el ímpetu de gradiente de descenso y una tasa de aprendizaje adaptativo. (10)

net.trainParam.epochs	1000	Número máximo de épocas para entrenar
net.trainParam.goal	0	Objetivo de rendimiento
net.trainParam.lr	0.01	Tasa de aprendizaje
net.trainParam.lr_inc	1.05	Relación para aumentar la tasa de aprendizaje
net.trainParam.lr_dec	0.7	Relación para disminuir la tasa de aprendizaje
net.trainParam.max_fail	6	Máximo de fallas de validación
net.trainParam.max_perf_inc	1.04	Aumento de rendimiento máximo
net.trainParam.mc	0.9	Constante del impulso
net.trainParam.min_grad	1e-5	Gradiente de rendimiento mínimo
net.trainParam.show	25	Épocas entre las pantallas (NaNpara no mostrar)
net.trainParam.showCommandLine	false	Generar salida de línea de comando
net.trainParam.showWindow	true	Mostrar GUI de entrenamiento
net.trainParam.time	inf	Máximo tiempo para entrenar en segundos

Tabla 4.1 parámetros de entrenamiento

4.4 Caso de estudio

Se aplicarán los pasos para aplicar el método de redes neuronales que se muestran en la sección 2.7. Los datos que se muestran a continuación son datos de los medidores instalados en las diferentes facultades que hay en la universidad de el salvador que corresponden a las facultades de:

- Humanidades 2: Medidor ubicado en la Facultad de Humanidades en la esquina del parqueo atrás del edificio de Ciencias y Matemática.
- Humanidades 3: Medidor ubicado en la Facultad de Humanidades cerca de las canchas de tenis y voleibol.
- Odontología 2: Medidor ubicado en la Facultad de Odontología.
- Odontología 3: Medidor ubicado en la Facultad de Odontología.
- Periodismo: Medidor ubicado en la parte posterior del edificio de periodismo.
- Psicología: Medidor ubicado atrás del edificio de Psicología en la zona de verde. La subestación está compuesta por 3 transformadores monofásicos de 50kVA cada uno.
- Química: Medidor ubicado al costado oriente de la facultad de Química y Farmacia.

Los datos obtenidos comprenden del año 2017 comprenden de los meses de julio a noviembre del año 2017 los datos que se obtuvieron son datos en una toma de 10 minutos se pasó a formato de variación horaria para poder estimar la de curva de demanda diaria que poseen los edificios que comprenden los medidores antes mencionados se estimara la curva de demanda diaria aplicando redes neuronales

- ✓ Selección de la variable: factor de variación horaria en p.u.
- ✓ Recolección de datos: datos obtenidos a través de una base de datos los datos tomados de los medidores son tomados cada 10 minutos se sacó un promedio de cada hora para pasar a formato de variación horaria comprenden de los meses de julio hasta noviembre.
- ✓ Procesamiento de datos: Se realiza un análisis y transformación de las variables de entrada y salida para minimizar el ruido, detectar tendencias y establecer perfiles para ayudar a la red neuronal en el aprendizaje de patrones relevantes, debido a que las redes neuronales son buscadores de patrones, la representación de los datos es crítica en el diseño de una red exitosa. Las variables de entrada son raramente ingresadas a la red en su forma original.

- ✓ Conjunto de Entrenamiento: Es el conjunto de datos que utiliza la red neuronal para aprender los patrones presentes en los datos.
- ✓ Conjunto de validación: Conjunto de datos utilizados para verificar el aprendizaje de la red donde los datos empleados son los más recientes consecuentemente al último valor de la muestra.
- ✓ Conjunto de prueba: Es el conjunto de datos encargados de evaluar la capacidad de predicción y precisión de la red.
- ✓ Número de Neuronas de Entrada: Corresponde a las mediciones de demanda de energía eléctrica de un día. Por tanto, la cantidad de neuronas de entrada son 24 horas.
- ✓ Numero de capas ocultas: son las encargadas de brindarle a la red la capacidad de generalizar por lo general se usan una o dos capas, el incremento en número de capas incrementa el tiempo de procesamiento y la dificultad sobre el ajuste lo que conduce a un pobre desempeño en la predicción fuera de la muestra.
- ✓ Numero de neuronas en la capa oculta: en este caso serán 25 neuronas en la capa ocultas.
- ✓ Numero de neuronas de salida: El número de neuronas de la capa de salida de nuestra red es una, porque tan solo intentamos predecir el valor de la variable seleccionada para un día.
- ✓ Función de transferencia: El propósito de esta función es prevenir a las salidas de alcanzar valores muy elevados que puedan paralizar la red y detener el entrenamiento de la misma. Como función de transferencia se utiliza la función Sigmoïdal (rango de salida entre -1 y +1).

A continuación, vamos a ir mostrando los resultados obtenidos de las diferentes simulaciones, vamos ir aplicando los mismos pasos para poder entrenar la red neuronal para las diferentes facultades el procedimiento es el mismo para la simulación, pero vamos a ir variando los días se simulará los días lunes, martes y miércoles para ver las distintas curvas de demanda de diferentes días a pronosticar.

En la siguiente tabla vamos a mostrar diferentes lunes con sus respectivos factores de variación horaria vamos a crear la red y vamos a entrenarla para que pueda a prender el comportamiento de los días.

La tabla 4.1 siguiente muestra los datos de demanda de un día del edificio de humanidades solo de los días lunes.

	fecha											
hora	17/07/2017	24/07/2017	31/07/2017	07/08/2017	14/08/2017	21/08/2017	28/08/2017	04/09/2017	11/09/2017	18/09/2017	25/09/2017	09/10/2017
0	0.6352566	0.07905988	0.39573768	0.05095224	0.07030158	0.07760317	0.09329679	0.0770088	0.0887144	0.07196044	0.08785123	0.09266504
1	0.92410309	0.08093175	0.37958532	0.05599833	0.08252261	0.1023263	0.08471925	0.08420586	0.08382902	0.08648838	0.09214576	0.07867678
2	0.88758112	0.07771665	0.36221752	0.05120501	0.08527364	0.08372159	0.08575882	0.08260105	0.08060742	0.07219641	0.09461596	0.08401171
3	0.90811986	0.07838691	0.34054389	0.06173555	0.07213885	0.08279823	0.09952701	0.07283103	0.09145521	0.07174065	0.09442264	0.08111923
4	0.54859765	0.0746292	0.40708858	0.05939237	0.08723119	0.07949484	0.09481688	0.07967614	0.09229678	0.07985245	0.09632911	0.10001487
5	0.35344422	0.30222848	0.38775574	0.05617919	0.0694275	0.08173158	0.09060056	0.08857486	0.09015476	0.0766058	0.08734811	0.09044079
6	0.19768259	1	0.40714156	0.12562253	0.15421867	0.16504116	0.20474043	0.11197692	0.14897245	0.13014202	0.13766109	0.11534148
7	0.14146357	0.86359063	0.68450119	0.48480164	0.40206898	0.52303408	0.57454444	0.35308452	0.45924676	0.38397119	0.44201464	0.40100407
8	0.10515959	0.45277512	0.71518394	0.85541184	0.80685915	0.8013237	0.99048088	0.86062694	0.8151689	0.79025128	0.84980545	0.75115107
9	0.12052308	0.2524119	1	1	0.98521609	0.99863786	0.89914858	0.95260008	1	1	1	0.9563562
10	0.08753565	0.16383462	0.83363927	0.91678121	0.90358253	1	0.82689812	0.8670878	0.82500608	0.87548181	0.90038137	0.91899648
11	0.08190365	0.1297691	0.74488654	0.9272707	0.91077597	0.91093724	0.81724596	0.96153443	0.88892256	0.71417176	0.90632892	0.99818841
12	0.0827224	0.13623351	0.44967361	0.72713619	0.70842086	0.73370103	0.71484603	0.71416972	0.65440732	0.63010372	0.70269099	0.69627888
13	0.07523299	0.11734466	0.39800474	0.73179372	0.8785301	0.85427768	0.91153687	0.79650994	0.74344764	0.77471382	0.68786195	0.74510608
14	0.07929461	0.12187979	0.43607659	0.87902256	1	0.92314297	0.89757313	1	0.84994319	0.86123084	0.93537541	0.87012099
15	0.07359176	0.07314662	0.57717705	0.9083816	0.95710735	0.86759565	1	0.92043262	0.80277836	0.81203445	0.89499595	1
16	0.07820537	0.05686311	0.96324101	0.66847124	0.70128977	0.46153046	0.60310753	0.60866645	0.49972164	0.5975478	0.56340605	0.73093374
17	0.52948438	0.0543651	0.73776331	0.32607528	0.40378673	0.30554345	0.30702117	0.53092374	0.29454805	0.32625511	0.44784646	0.32669541
18	1	0.05040055	0.47835845	0.17889863	0.2622174	0.16961793	0.27971269	0.31798885	0.18602065	0.26443212	0.26843182	0.22160502
19	0.75054074	0.06140128	0.53088334	0.14987472	0.16428552	0.12693228	0.18530558	0.19738142	0.12793944	0.15155561	0.15081722	0.19066885
20	0.34816159	0.05766417	0.59307683	0.12626301	0.14446744	0.10167237	0.15825389	0.1377152	0.1122849	0.10230277	0.14215563	0.13351502
21	0.22491883	0.06296333	0.59293335	0.13290858	0.13463876	0.09255269	0.12142039	0.10328018	0.11574934	0.09708978	0.11866206	0.11023099
22	0.15000979	0.07908398	0.47097184	0.09790684	0.10214789	0.07789981	0.11244233	0.09897274	0.08464751	0.09007112	0.09965886	0.09252601
23	0.12916576	0.10935341	0.40953836	0.08897426	0.10865248	0.08776662	0.09077563	0.09245733	0.08419131	0.07916784	0.09297386	0.08966072

Tabla 4.2 factores de variación horaria para el día lunes del medidor 2 ubicado en la facultad de humanidades.

El objetivo es crear una red neuronal que pueda predecir la curva de demanda diaria de los días antes visto en la tabla 4.1 tendremos que entrenarla para que pueda predecir la curva de demanda diaria de un día como podemos ver en la tabla 4.1 los días lunes tiene diferentes valores máximos a distintas horas del día estos valores son atípicos en dichos valores atípicos se tendrá que entrenar la red para que pueda aprender a predecir esos días.

Para crear la red neuronal se seleccionan los primeros días de la tabla 4.1 que corresponden 17/07/2017, 24/07/2017 y 31/07/2017 con estos datos vamos a entrenar a la red y vamos a probar si la red ha sido capaz de pronosticar la curva de los días antes mencionados. Como conjunto de validación y de prueba se toman los días lunes 17/07/2017, 24/07/2017 y 31/07/2017.

Procedemos a crear la red neuronal con los datos que se muestran en la tabla 4.3 la red consta de un retardo de 24 datos y 25 neuronas en la capa oculta.

Hora	17/07/2017	24/07/2017	31/07/2017
0	0.6352566	0.07905988	0.39573768
1	0.92410309	0.08093175	0.37958532
2	0.88758112	0.07771665	0.36221752
3	0.90811986	0.07838691	0.34054389
4	0.54859765	0.0746292	0.40708858
5	0.35344422	0.30222848	0.38775574
6	0.19768259	1	0.40714156
7	0.14146357	0.86359063	0.68450119
8	0.10515959	0.45277512	0.71518394
9	0.12052308	0.2524119	1
10	0.08753565	0.16383462	0.83363927
11	0.08190365	0.1297691	0.74488654
12	0.0827224	0.13623351	0.44967361
13	0.07523299	0.11734466	0.39800474
14	0.07929461	0.12187979	0.43607659
15	0.07359176	0.07314662	0.57717705
16	0.07820537	0.05686311	0.96324101
17	0.52948438	0.0543651	0.73776331
18	1	0.05040055	0.47835845
19	0.75054074	0.06140128	0.53088334
20	0.34816159	0.05766417	0.59307683
21	0.22491883	0.06296333	0.59293335
22	0.15000979	0.07908398	0.47097184
23	0.12916576	0.10935341	0.40953836

Tabla 4.3 días de entrenamiento y de prueba.

Con estos datos entrenamos la red neuronal y probamos si la red neuronal funciona adecuadamente los datos de entrada en este caso son los mismo que los de salida porque estamos entrenando la red.

```
>> net=newfftd(datos,datos,(1:24),25)
```

Una vez creada la red procedemos entrenarla con los parámetros antes mencionados en la tabla 4.1 una vez creada la red ingresamos a la línea de comando de matlab y nos muestra cómo se creó la red neuronal nos muestra la cantidad de iteraciones cantidad de tiempo que se llevó acabo el entrenamiento el rendimiento del entrenamiento como el gradiente en la figura 1 y figura 2 muestra los resultados del entrenamiento.

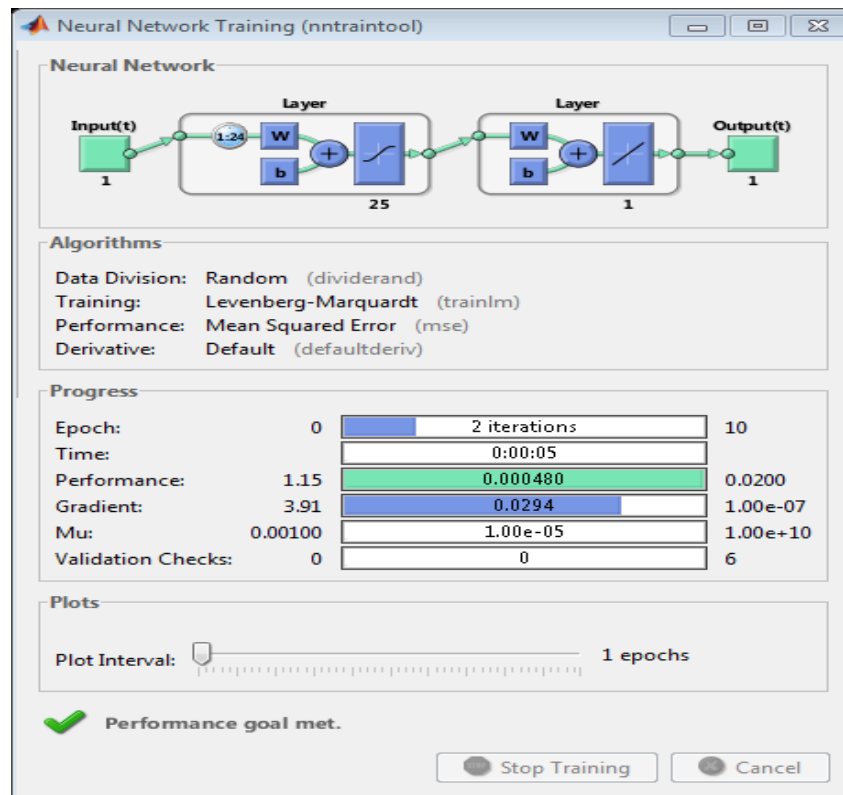


Figura 4.1 resultado de la red neuronal.

```
>> [net,tr]=train(net,entrada,salida,valorinicial)
```

Training with TRAINLM.

Epoch 0/10, Time 3.865, Performance 1.1524/0.02, Gradient 3.9128/1e-07, Mu 0.001/10000000000, Validation Checks 0/6

Epoch 1/10, Time 5.069, Performance 0.024353/0.02, Gradient 0.25812/1e-07, Mu 0.0001/10000000000, Validation Checks 0/6

Epoch 2/10, Time 5.148, Performance 0.00048025/0.02, Gradient 0.029412/1e-07, Mu 1e-05/10000000000, Validation Checks 0/6

Training with TRAINLM completed: Performance goal met.

Figura 4.2 resultado del entrenamiento.

Ya entrenada la red procedemos a simular la red los resultados que nos muestra son comparados con la salida deseada que queremos obtener en la figura 4.3 y figura 4.4 muestra valores cercanos a la salida que queremos obtener.

```
>> pronostico=sim(net,entrada,valorinicial)

pronostico =

Columns 1 through 7
    [0.0817]    [0.0805]    [0.0712]    [0.0692]    [0.0772]    [0.3209]    [1.0121]

Columns 8 through 14
    [0.8860]    [0.4117]    [0.2326]    [0.1700]    [0.1359]    [0.1372]    [0.1179]

Columns 15 through 20
    [0.1171]    [0.0663]    [0.0546]    [0.0657]    [0.0759]    [0.0736]

Columns 21 through 24
    [-0.0112]    [0.0136]    [0.0814]    [0.1159]
```

Figura 4.3 resultados de la simulación del entrenamiento y prueba de simulación.

```
>> salida=cell2mat(salida)

salida =

Columns 1 through 8
    0.0791    0.0809    0.0777    0.0784    0.0746    0.3022    1.0000    0.8636

Columns 9 through 16
    0.4528    0.2524    0.1638    0.1298    0.1362    0.1173    0.1219    0.0731

Columns 17 through 24
    0.0569    0.0544    0.0504    0.0614    0.0577    0.0630    0.0791    0.1094
```

Figura 4.4 se compara la salida con el pronóstico que se muestra en la figura 4.3

En la figura 4.5 siguiente muestra la gráfica del pronóstico junto con la salida que queremos tener.

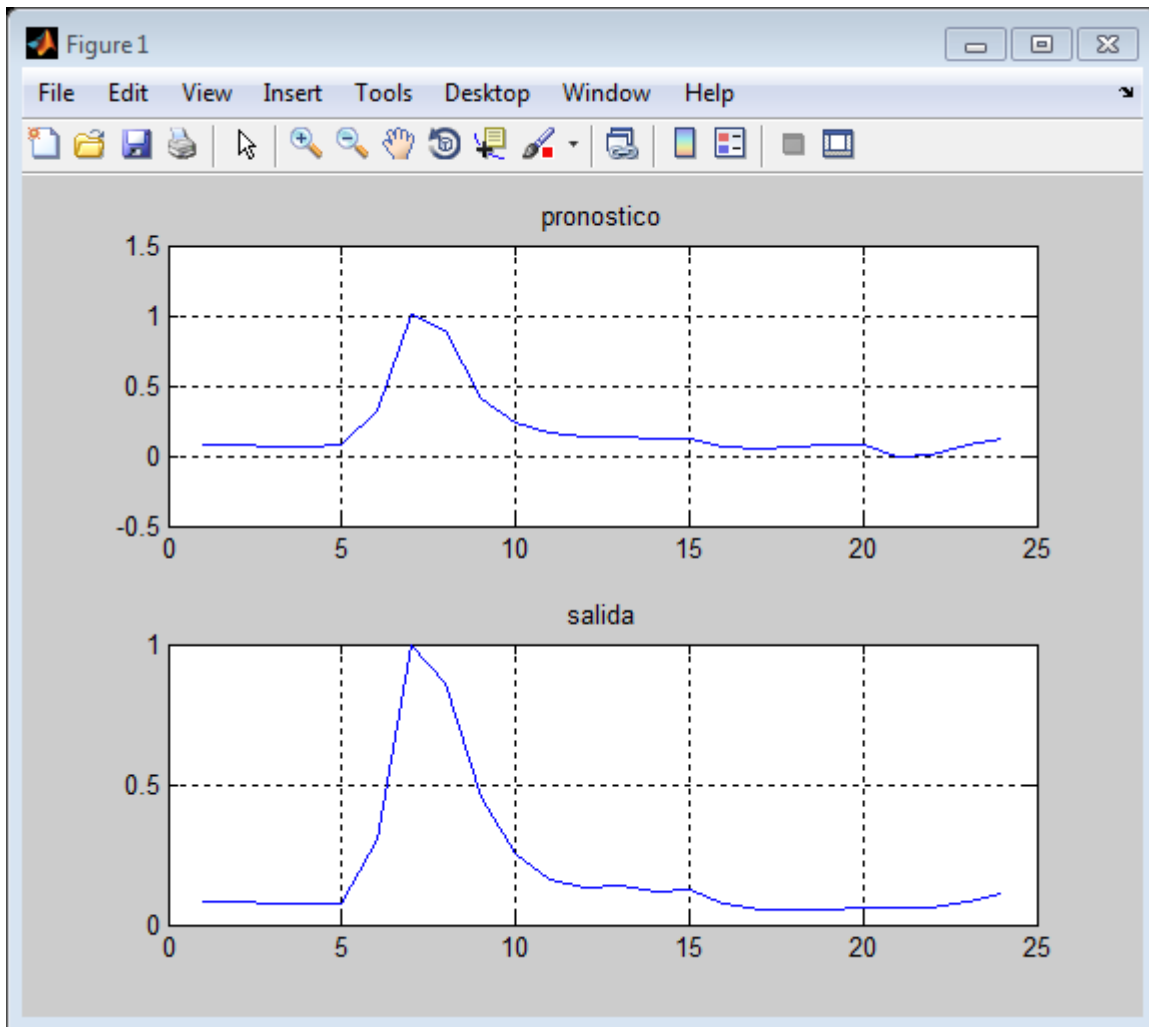


Figura 4.5 gráficas del pronóstico y la salida que se quiere obtener.

Ahora procedemos a entrenar a la red con diferentes días los días ingresados son: 04/09/2017 este dato se toma como parámetro de tiempo de retardo como dato de entrada le ingresamos el día 11/09/2017 como salida le ingresamos el día 18/09/2017.

Hora	04/09/2017	11/09/2017	18/09/2017
0	0.0770088	0.0887144	0.07196044
1	0.08420586	0.08382902	0.08648838
2	0.08260105	0.08060742	0.07219641
3	0.07283103	0.09145521	0.07174065
4	0.07967614	0.09229678	0.07985245
5	0.08857486	0.09015476	0.0766058
6	0.11197692	0.14897245	0.13014202
7	0.35308452	0.45924676	0.38397119
8	0.86062694	0.8151689	0.79025128
9	0.95260008	1	1
10	0.8670878	0.82500608	0.87548181
11	0.96153443	0.88892256	0.71417176
12	0.71416972	0.65440732	0.63010372
13	0.79650994	0.74344764	0.77471382
14	1	0.84994319	0.86123084
15	0.92043262	0.80277836	0.81203445
16	0.60866645	0.49972164	0.5975478
17	0.53092374	0.29454805	0.32625511
18	0.31798885	0.18602065	0.26443212
19	0.19738142	0.12793944	0.15155561
20	0.1377152	0.1122849	0.10230277
21	0.10328018	0.11574934	0.09708978
22	0.09897274	0.08464751	0.09007112
23	0.09245733	0.08419131	0.07916784

Tabla 4.4 datos de validación de la red son nuevos datos para ver cómo se comporta la red con diferentes días.

En este caso ya tenemos creada la red neuronal solo tendremos que ingresar nuevas entradas y retardos con una nueva salida como podemos comparar la tabla 4.3 con la tabla 4.4 los valores que ingresamos son distintos en este caso son datos atípicos entonces en este caso se vuelve a entrenar la red neuronal para que se adapte a nuevas entradas y salidas en las figuras siguientes nos muestra el entrenamiento de la red, la red su estructura es la misma lo que cambia son sus pesos adaptivos de la red neuronal .

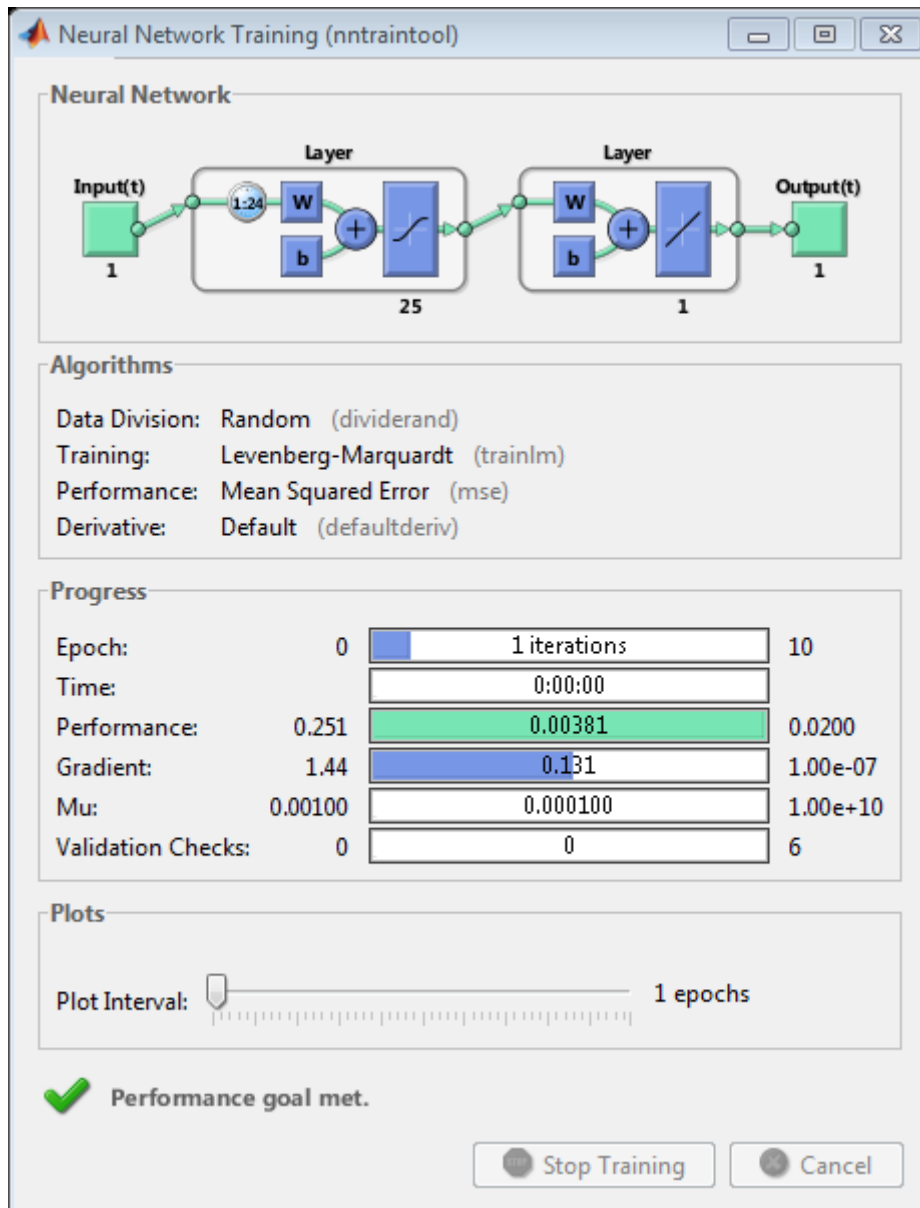


Figura 4.6 resultados de la red neuronal para nuevos valores.

```
>> [net,tr]=train(net,entrada,salida,valorinicial)
```

Training with TRAINLM.

Epoch 0/10, Time 0.079, Performance 0.25102/0.02, Gradient 1.4383/1e-07, Mu 0.001/10000000000, Validation Checks 0/6

Epoch 1/10, Time 0.202, Performance 0.0038108/0.02, Gradient 0.13124/1e-07, Mu 0.0001/10000000000, Validation Checks 0/6

Training with TRAINLM completed: Performance goal met.

Figura 4.7 resultado del entrenamiento de la red con nuevas entradas y salidas.

En la siguiente figura nos muestra los datos de simulación que se obtienen al simular la red con las nuevas entradas y salidas como podemos la predicción es parecida a la salida que buscamos.

```

>> prediccion=sim(net,entrada,valorinicial)

prediccion =

Columns 1 through 7

    [0.0826]    [0.0776]    [0.0971]    [0.0874]    [0.1173]    [0.0491]    [0.2106]

Columns 8 through 14

    [0.4656]    [0.8222]    [0.9669]    [0.8137]    [0.8295]    [0.6218]    [0.7270]

Columns 15 through 21

    [0.7636]    [0.9173]    [0.6211]    [0.3384]    [0.2830]    [0.1744]    [0.2074]

Columns 22 through 24

    [0.2209]    [0.1745]    [0.1273]

```

Figura 4.8 resultados de la simulación para predecir nuevos valores diferentes lunes.

```

>> salida

salida =

Columns 1 through 7

    [0.0887]    [0.0838]    [0.0806]    [0.0915]    [0.0923]    [0.0902]    [0.1490]

Columns 8 through 14

    [0.4592]    [0.8152]    [1]    [0.8250]    [0.8889]    [0.6544]    [0.7434]

Columns 15 through 21

    [0.8499]    [0.8028]    [0.4997]    [0.2945]    [0.1860]    [0.1279]    [0.1123]

Columns 22 through 24

    [0.1157]    [0.0846]    [0.0842]

```

Figura 4.9 salida que se quiere obtener.

Calculamos los errores entre la salida y la predicción como vemos el error no es mucho la red ha sido entrenada bien para que pueda predecir valores de días diferentes en las figuras siguiente nos muestra los resultados.

```
>> error=abs(salida-prediccion)

error =

Columns 1 through 8

    0.0061    0.0062    0.0165    0.0041    0.0250    0.0411    0.0616    0.0064

Columns 9 through 16

    0.0070    0.0331    0.0113    0.0595    0.0326    0.0165    0.0864    0.1146

Columns 17 through 24

    0.1213    0.0438    0.0970    0.0464    0.0951    0.1051    0.0899    0.0431
```

Figura 4.10 resultados de los errores que se obtienen con la salida respecto a la predicción.

Como podemos ver la gráfica obtenida son muy parecidas la red neuronal muestra valores parecidos a cualquier día lunes que ingresamos a la red neuronal en casos que haya días lunes que varían muchos a otros tendremos que entrenar la red para poder ajustar las entradas y las salidas deseadas.

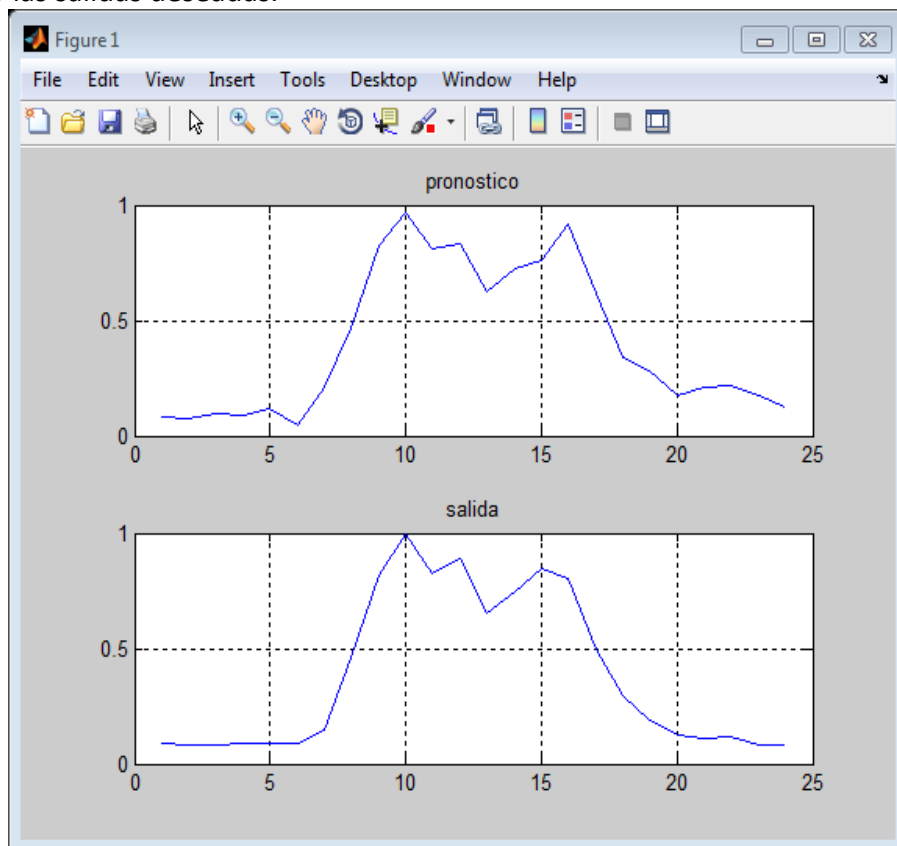


Figura 4.11 grafica de predicción vrs la salida nueva.

A continuación, mostramos los factores de variación horaria para el medidor de humanidades 3.

En la siguiente tabla 4.5 muestra los datos tabulados de los días lunes del medidor de humanidades 3

	fechas												
hora	29/05/2017	05/06/2017	12/06/2017	26/06/2017	03/07/2017	31/07/2017	07/08/2017	14/08/2017	21/08/2017	28/08/2017	04/09/2017	11/09/2017	25/09/2017
0	0.12856058	0.12869083	0.08769328	0.09686012	0.07805959	0.40044475	0.05519483	0.09383307	0.10275715	0.08318532	0.1213109	0.09691633	0.06812711
1	0.12729905	0.13226545	0.09941821	0.1083495	0.07160108	0.38094226	0.054449	0.09393721	0.10013095	0.08002358	0.12591711	0.088195	0.06880466
2	0.1324372	0.1287931	0.09340359	0.09881559	0.0739715	0.40376412	0.05883363	0.09606681	0.10692551	0.08016386	0.12927761	0.08782418	0.06835736
3	0.14138833	0.12091297	0.09999853	0.10202975	0.07267708	0.36534796	0.05636523	0.09053896	0.10686554	0.08019208	0.12977466	0.09115946	0.07240905
4	0.12333189	0.13433819	0.11355217	0.09994892	0.0760594	0.41851105	0.05996792	0.10213407	0.10594898	0.09060166	0.12028636	0.0995964	0.0890581
5	0.22410656	0.32504915	0.22812254	0.20861433	0.13593743	0.36028084	0.20743422	0.22847549	0.37080664	0.30585834	0.37350621	0.29880475	0.31346644
6	0.51211677	0.71854275	0.63549682	0.52490476	0.36018149	0.3455713	0.51237084	0.61448128	0.62926848	0.65609271	0.80591529	0.6529195	0.63387538
7	0.63681109	0.72932829	0.76577235	0.61439023	0.5207576	0.61182117	0.63195949	0.74615516	0.74740054	0.768087	0.92619811	0.9033079	0.80663442
8	0.98054006	0.97013804	0.93782104	0.81832462	0.62603562	0.70397946	0.80146495	0.94198783	0.91478162	0.85840886	0.98764839	0.99204362	0.90342134
9	0.99797675	1	0.97230631	0.9604748	0.92440654	1	0.8561628	0.92743713	0.94077802	0.93585399	0.98390062	0.97198432	0.82878976
10	0.97104919	0.95414101	0.97929689	1	1	0.93888811	1	1	1	0.93464377	1	1	0.96378347
11	1	0.83498766	0.88062431	0.8434501	0.90544382	0.84636878	0.87620796	0.87980934	0.90979846	1	0.92690657	0.92940006	1
12	0.69227905	0.7009116	0.6801821	0.78939165	0.71067986	0.6852182	0.62282975	0.65445321	0.77897519	0.68974918	0.84917432	0.81214426	0.77831775
13	0.85515175	0.91367786	0.9510423	0.87449324	0.71883997	0.67933033	0.81112002	0.81693761	0.88877959	0.78187472	0.91330184	0.91650061	0.81950107
14	0.78944393	0.94009948	1	0.86448331	0.72626513	0.6911774	0.7441726	0.92551965	0.9065133	0.87009594	0.93673932	0.97176455	0.84931859
15	0.75182769	0.87004258	0.89004876	0.88873109	0.57443202	0.58318063	0.78181705	0.75618996	0.8337625	0.91484198	0.8578877	0.85251363	0.77926314
16	0.71241186	0.84781195	0.72306054	0.73374123	0.48252161	0.32288988	0.5639356	0.71455854	0.77439419	0.76095119	0.80979673	0.74835551	0.72963311
17	0.63133411	0.81423657	0.72148101	0.65996639	0.47172391	0.24015642	0.45245508	0.61738994	0.69249444	0.72405837	0.70063611	0.62595267	0.72311796
18	0.49901892	0.80131565	0.61183565	0.69321306	0.43525068	0.29720364	0.30261723	0.44475114	0.51514355	0.56145694	0.59175902	0.47402809	0.44793203
19	0.11726993	0.38710008	0.32646138	0.40565157	0.23801595	0.25833312	0.18524101	0.26094891	0.36001632	0.33854288	0.38495919	0.26382879	0.27928192
20	0.09183924	0.11770365	0.12369495	0.13475747	0.09581823	0.34939622	0.13587165	0.08191109	0.19094098	0.11187243	0.09318689	0.09276808	0.08416528
21	0.09174639	0.11131223	0.11102523	0.1498944	0.09933486	0.29373609	0.14629742	0.0851285	0.11585702	0.1125593	0.08640334	0.09017938	0.09094482
22	0.08943385	0.11463532	0.11244289	0.14103689	0.09110058	0.32267318	0.13211126	0.08939361	0.12010052	0.10844174	0.09865131	0.09716821	0.08898374
23	0.10358927	0.10942812	0.10631212	0.12779179	0.10356343	0.29195844	0.12522351	0.09236725	0.11438306	0.11789404	0.09818729	0.07941869	0.08144452

Tabla 4.5 datos de variación horaria del día lunes medidor 3 de la facultad de humanidades.

La siguiente grafica muestra la gráfica de la facultad de humanidades 3 nos muestra la curva de demanda diaria para el día 06/06/2017 y para el día 28/08/2017.

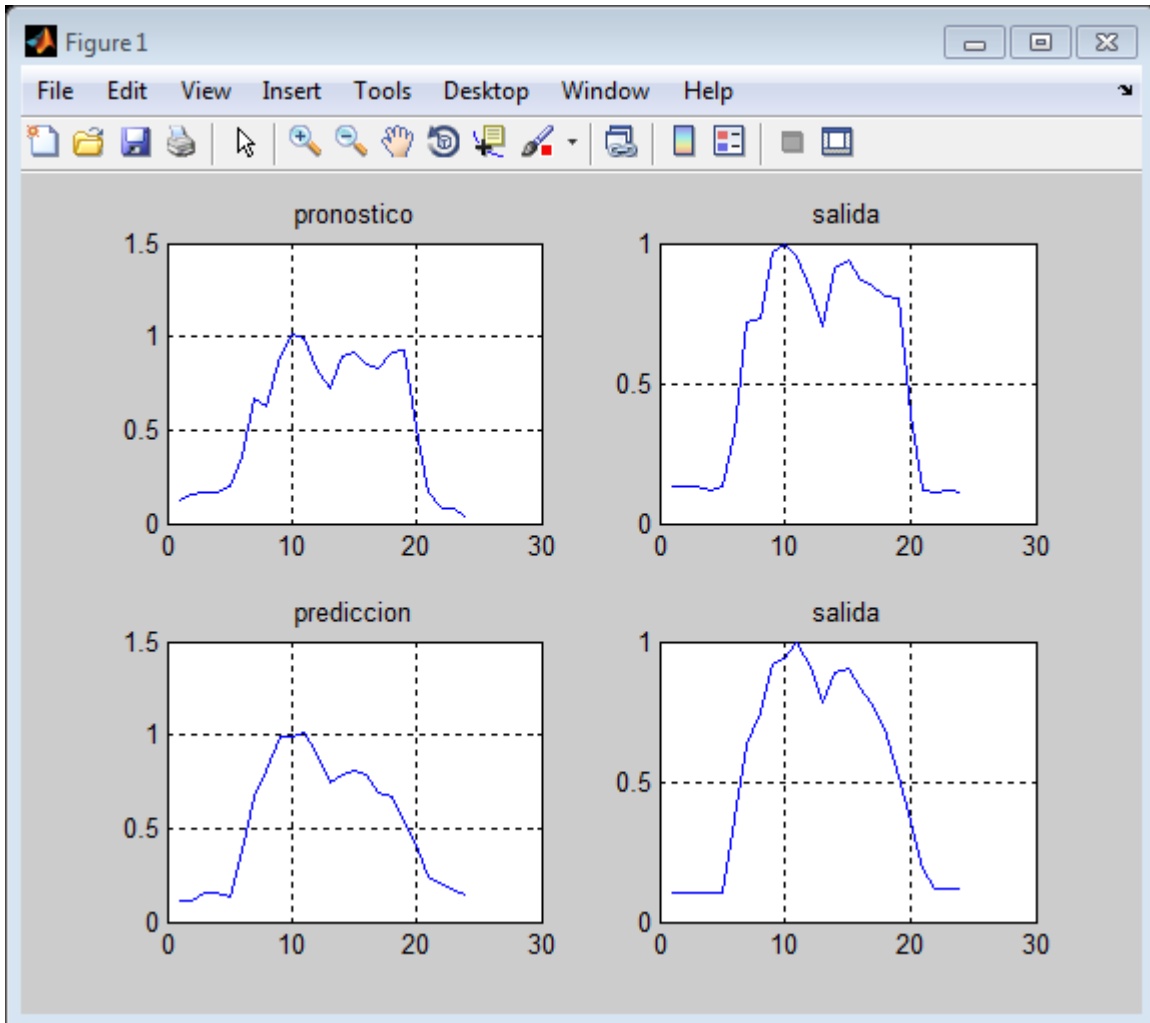


Figura 4.12 grafica obtenidas de diferentes días lunes para medidor 3 ubicado en la facultad de humanidades.

La grafica pronostico muestra el resultado del entrenamiento y prueba de la red neuronal la gráfica de predicción muestra el resultado de validación de la red podemos comprobar con cualquier día de la tabla para humanidades 3.

La siguiente tabla muestra los factores de variación horaria para para la facultad de odontología corresponde al medidor 2 ubicado en dicha facultad los datos mostrados corresponden al día martes.

hora	fecha												
	18/07/2017	25/07/2017	01/08/2017	08/08/2017	15/08/2017	22/08/2017	29/08/2017	12/09/2017	19/09/2017	26/09/2017	03/10/2017	10/10/2017	17/10/2017
0	0.13821381	0.12361587	0.92837409	0.12852317	0.13002455	0.14001655	0.17775309	0.13593785	0.16141623	0.14966459	0.13614984	0.15252972	0.13631388
1	0.13809246	0.12725428	0.9214631	0.13041314	0.13269245	0.13566906	0.17971218	0.14211774	0.15135825	0.15275812	0.15083641	0.14673182	0.13723426
2	0.13699508	0.12264759	0.95379043	0.13045194	0.14426538	0.13703664	0.17219919	0.14273002	0.16511399	0.15846636	0.14347456	0.15174097	0.14299174
3	0.13468631	0.12769103	0.93507247	0.13175573	0.1331421	0.12440996	0.17141993	0.13549785	0.15469582	0.15242896	0.14586283	0.14749972	0.13752137
4	0.13346049	0.12294171	0.9548562	0.12989969	0.13400313	0.13053188	0.17306326	0.14898399	0.1550948	0.15313113	0.14199342	0.16559834	0.13697081
5	0.12464411	0.11361433	0.86644948	0.12631169	0.13413923	0.11873154	0.1729573	0.13729479	0.15183049	0.14225968	0.14819976	0.14085966	0.1390721
6	0.35254221	0.35091094	0.67459679	0.30455223	0.33489856	0.32870649	0.44987615	0.36826528	0.45482802	0.40359846	0.37597989	0.52908338	0.35493933
7	0.79095781	0.96134095	0.58559612	0.81283002	0.77703244	0.83386883	0.94964179	0.77002644	0.98140219	0.6708013	0.85723105	0.97471709	0.83622976
8	0.94923817	1	0.53911457	0.89226567	0.9737906	0.93545839	1	0.97579594	1	0.81630301	1	0.95891273	1
9	0.89905001	0.86038264	0.57192827	1	0.82209291	0.88023977	0.85307252	0.85822858	0.95012986	0.81938951	0.85084341	0.84613213	0.97647392
10	1	0.93851189	0.60691666	0.85447873	1	1	0.97552191	1	0.98028976	1	0.84200662	1	0.93074682
11	0.86471608	0.95017713	0.54435175	0.72865568	0.97530388	0.86650838	0.88917443	0.94325029	0.83055328	0.86880888	0.79535166	0.91322873	0.8452104
12	0.67743558	0.75591466	0.54233922	0.54750051	0.74427729	0.70273919	0.62556342	0.69966798	0.64320367	0.6957015	0.6216878	0.61397544	0.54914137
13	0.65022088	0.7713129	0.55464284	0.2220739	0.63817553	0.72257554	0.60712828	0.67844673	0.86137538	0.76823027	0.64038612	0.81523854	0.63273215
14	0.84774151	0.77693285	0.60133877	0.16177351	0.70551672	0.82360178	0.78776274	0.78844526	0.88417603	0.79038658	0.67985063	0.75218911	0.86189759
15	0.40771331	0.3834845	0.59521963	0.11749082	0.35024175	0.34255933	0.51309274	0.39891271	0.4457751	0.41582749	0.32521618	0.33643001	0.47312887
16	0.16568736	0.15436867	0.60039018	0.13797103	0.188783	0.17112544	0.40891757	0.19376326	0.28414568	0.24507912	0.18559099	0.20518559	0.30732805
17	0.12372364	0.12106626	0.66663755	0.14293277	0.12093481	0.13214261	0.20703278	0.12857401	0.17893927	0.1532017	0.13151915	0.14976801	0.12683045
18	0.11456209	0.1174587	0.79145549	0.13819254	0.13175089	0.11801365	0.16019733	0.15455045	0.17244482	0.15985627	0.15800917	0.1764633	0.13662254
19	0.12746985	0.13588747	0.93206765	0.14717441	0.13923951	0.13383388	0.17241585	0.15175463	0.18399548	0.15326852	0.15774346	0.18011077	0.14864398
20	0.13346529	0.12449193	0.94139647	0.14139583	0.13783279	0.13171091	0.15855853	0.14434914	0.1795141	0.14612336	0.14048317	0.15935722	0.14754111
21	0.13355852	0.12603772	1	0.13891515	0.14866935	0.13557181	0.15331934	0.14955215	0.15751548	0.15099708	0.15043519	0.16167594	0.13977166
22	0.13525567	0.12280241	0.92953867	0.13542276	0.13405686	0.12476901	0.15723076	0.13668126	0.17099452	0.1493319	0.15572272	0.16030712	0.14343991
23	0.13139588	0.12416785	0.96267614	0.14102086	0.14748443	0.13405811	0.15275397	0.14873718	0.18331814	0.15886969	0.15209938	0.15476654	0.13890521

Tabla 4.6 factores de variación horaria para el día martes para el medidor 2 de la facultad de odontología.

A continuación, mostramos los resultados de las simulaciones de la red neuronal que se creó para la facultad de odontología que posee el medidor 2. Los resultados mostrados corresponden a diferente martes en la figura 7 muestra como quedo el diseño de la red neuronal en la figura 8 muestra los resultados de las gráficas de los días que se hizo el entrenamiento de la predicción

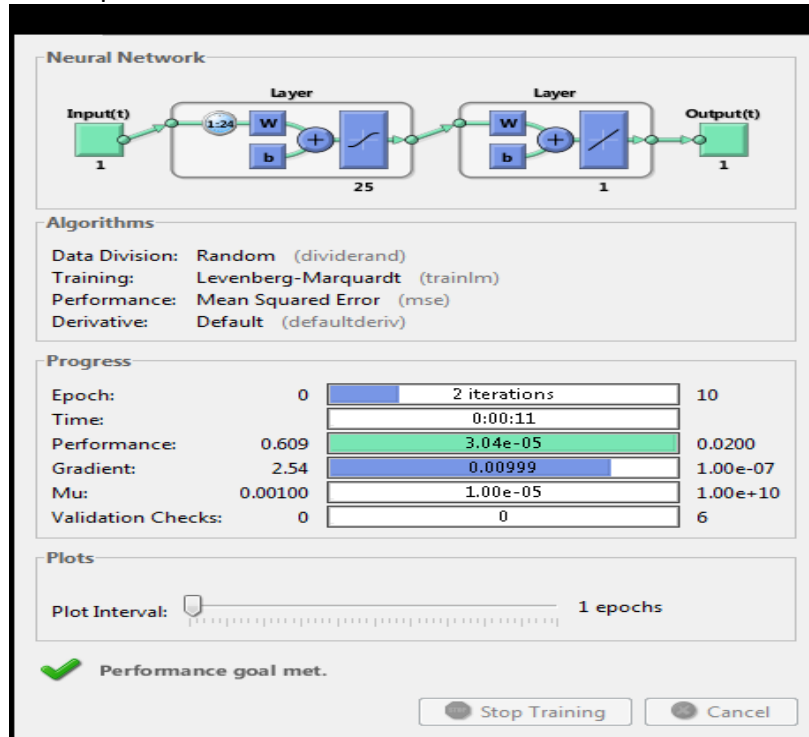


Figura 4.13 resultados de la red neuronal para la facultado de odontología.

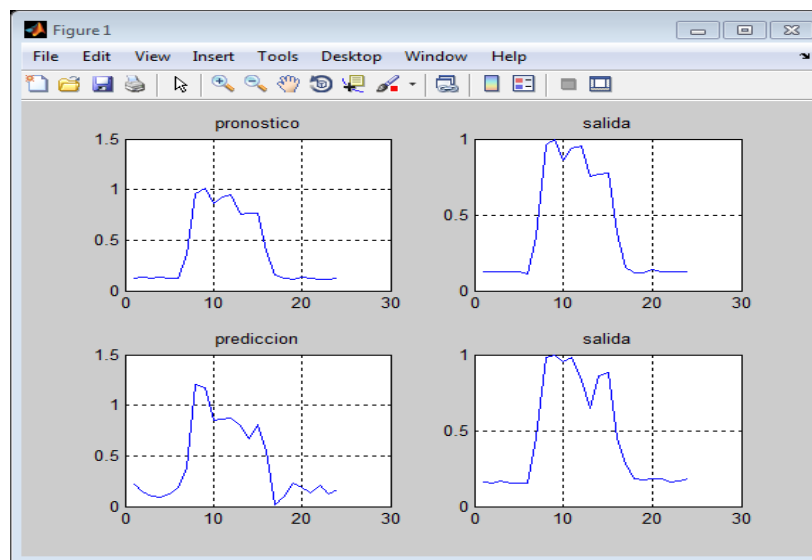


Figura 4.14 graficas obtenidas para diferentes martes de la facultad de odontología.

La siguiente tabla muestra los factores de variación horaria para el día miércoles los datos mostrados corresponden al medidor de la facultad de odontología 3

hora	fecha												
	12/07/2017	19/07/2017	26/07/2017	02/08/2017	09/08/2017	16/08/2017	23/08/2017	30/08/2017	06/09/2017	13/09/2017	20/09/2017	27/09/2017	04/10/2017
0	0.09062189	0.09258027	0.07792746	1	0.07238942	0.12252618	0.06913935	0.0610943	0.06372832	0.08363571	0.06802565	0.0724111	0.43488732
1	0.0922361	0.09388777	0.08019936	0.97081559	0.07351608	0.12418935	0.07330022	0.05940379	0.06328232	0.0846296	0.06996691	0.07122373	0.43110703
2	0.0929342	0.09200169	0.08205488	0.96396038	0.07300833	0.12495664	0.07206083	0.06178674	0.06354147	0.08273224	0.0672138	0.06791245	0.44028024
3	0.08906975	0.0921275	0.08207389	0.97912307	0.07283018	0.12109075	0.06997854	0.06083681	0.06407623	0.08864647	0.07129344	0.0730402	0.42938554
4	0.08939918	0.09068891	0.07888175	0.94439362	0.07165388	0.12413538	0.072915	0.06032924	0.06074553	0.08531333	0.06966816	0.070876	0.43036868
5	0.08785785	0.09210985	0.07877897	0.96830143	0.07144011	0.1256982	0.068371	0.06174168	0.06292667	0.08165446	0.07319981	0.06849796	0.40781678
6	0.09922061	0.18180978	0.10538002	0.43369975	0.10424138	0.17859198	0.15486744	0.12355731	0.11841081	0.12667017	0.11294093	0.1331498	0.41820638
7	0.7729151	0.87229632	0.76177908	0.33433655	0.56271339	0.40787935	0.74258144	0.81092165	0.7636257	0.98074746	0.88580311	1	0.74003924
8	1	1	0.90848284	0.37617295	0.75695571	0.03024169	0.9088514	0.78921361	0.9921908	0.98983291	0.89485595	0.91116209	1
9	0.8082313	0.78321033	0.70125677	0.33623347	0.85611583	0.03317017	1	0.80572393	0.71302471	1	0.84121926	0.99278411	0.86648964
10	0.73751558	0.96004641	0.76421371	0.34754972	0.75539545	0.04084571	0.88290634	1	0.80048694	0.96919515	0.8637631	0.99191295	0.71893959
11	0.89498217	0.84742073	1	0.35266621	1	0.68336072	0.81614064	0.77662329	1	0.9934726	1	0.89612461	0.22391458
12	0.46315195	0.21335943	0.40621612	0.35680385	0.29512753	1	0.22630567	0.36720542	0.2185256	0.71606406	0.38037257	0.22912501	0.22720885
13	0.25082086	0.80788847	0.53869928	0.34763701	0.51570985	0.72382949	0.65484513	0.80715563	0.7502201	0.76705983	0.87353431	0.77200543	0.2666612
14	0.21376211	0.88051754	0.77869153	0.33243976	0.61739629	0.10487591	0.76143012	0.7663169	0.86083451	0.9062731	0.88764596	0.85013306	0.46250067
15	0.10510995	0.11852658	0.09275826	0.34606993	0.42388305	0.13573822	0.66045461	0.56235239	0.11091585	0.15841316	0.5637132	0.44716558	0.45983064
16	0.06031871	0.05558042	0.05152065	0.33947124	0.50516142	0.14482165	0.55408575	0.37376478	0.0579957	0.06343549	0.64926394	0.57596235	0.46208623
17	0.06398219	0.04535469	0.43399702	0.43085543	0.04677752	0.13627061	0.08771111	0.06842045	0.06390238	0.06218658	0.07135644	0.05205504	0.41062729
18	0.09775626	0.10172323	0.41235251	0.8940827	0.06679504	0.13018173	0.07845028	0.07173737	0.07377521	0.09233038	0.09029565	0.07016357	0.42847979
19	0.10378942	0.11422725	0.28864914	0.86919895	0.07965553	0.13070206	0.07844069	0.07112012	0.07485736	0.0992875	0.08513939	0.07703278	0.43818864
20	0.10416308	0.09730071	0.08616717	0.93477009	0.07961523	0.10330882	0.07604879	0.06993359	0.07280813	0.09900629	0.07852796	0.07662003	0.50007913
21	0.09793581	0.09548156	0.07969564	0.97401555	0.07509214	0.09812301	0.07176488	0.06266404	0.06672334	0.0856439	0.07246447	0.07208207	0.47562497
22	0.09340166	0.08955986	0.07583309	0.92883021	0.07334453	0.09741118	0.0686778	0.061096	0.0643774	0.08445383	0.07012731	0.06884058	0.49137236
23	0.09229162	0.08933871	0.07542058	0.94262829	0.07547274	0.09483704	0.06879929	0.0616983	0.06254341	0.08356543	0.07524879	0.06859711	0.48021299

Tabla 4.7 factores de variación horaria para el día miércoles corresponden al medidor 3 de la facultad de odontología.

Resultados que se obtuvieron de crear una red neuronal del día miércoles para la facultad de odontología 3

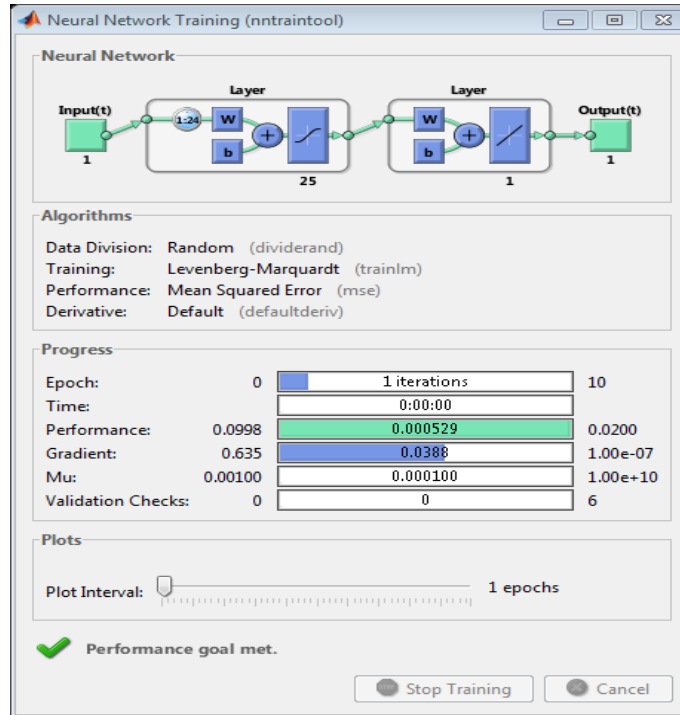


Figura 4.15 red neuronal creada para la faculta de odontología medidor 3 para un día miércoles.

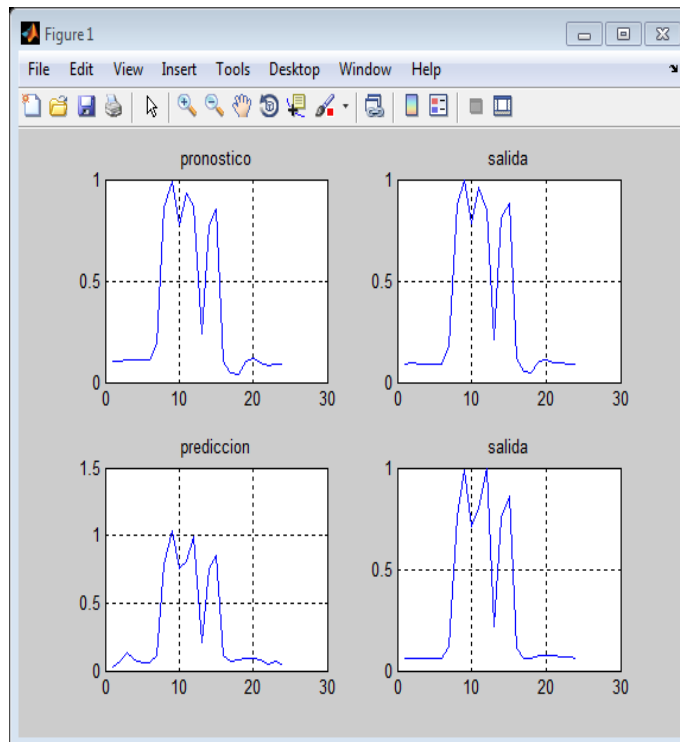


Figura 4.16 graficas obtenidas para diferentes miércoles pronostico y predicción.

La siguiente tabla los factores de variación horaria para el día lunes corresponde a la facultad de periodismo

hora	fecha												
	24/07/2017	31/07/2017	07/08/2017	14/08/2017	21/08/2017	28/08/2017	04/09/2017	11/09/2017	18/09/2017	25/09/2017	09/10/2017	16/10/2017	23/10/2017
0	0.07882157	0.60044889	0.04714413	0.07748284	0.05694202	0.08012102	0.06882298	0.06872997	0.05911836	0.06404051	0.09033267	0.04355607	0.05439703
1	0.08165973	0.59688798	0.04810398	0.07550847	0.05809862	0.07982158	0.06662416	0.07498894	0.05546827	0.06488146	0.09666068	0.04174837	0.05190155
2	0.07933549	0.62697577	0.04754116	0.07854586	0.05644842	0.08620375	0.06883417	0.07451857	0.05376512	0.07042798	0.08881326	0.0450154	0.05249843
3	0.08158311	0.59945884	0.04806472	0.0751912	0.06233287	0.08030444	0.06632004	0.07188694	0.05583457	0.06918014	0.09263405	0.04638751	0.05364243
4	0.11032305	0.95831677	0.07054586	0.10607526	0.0787737	0.11120855	0.09466345	0.09988121	0.0818265	0.08929786	0.12364113	0.06189684	0.07310534
5	0.3196429	0.97366327	0.08550325	0.1261963	0.10945119	0.1241136	0.10178787	0.11562274	0.09993272	0.1281968	0.13627987	0.07353724	0.0818029
6	1	0.64271368	0.24647323	0.23295388	0.19638627	0.27715034	0.30786585	0.3789594	0.28391237	0.29090686	0.25676979	0.1718066	0.31749944
7	0.72704419	0.59041154	0.59400449	0.69241908	0.64802873	0.65593507	0.65308661	0.64763909	0.44229365	0.62106748	0.55428565	0.55869684	0.54491281
8	0.26704639	0.61107439	0.84620442	0.85993217	0.92416066	0.79067742	0.79950371	0.73580531	0.86665983	0.76921611	0.76463507	0.73534462	0.76906652
9	0.16918	0.67001412	0.95248471	1	0.94292812	0.94087606	1	0.79861579	0.99689654	0.97723844	0.94540822	0.87024683	0.83790478
10	0.13894859	0.67552442	1	0.91819923	1	0.92447216	0.93360117	0.88066081	1	1	0.83700257	0.95030948	0.86108952
11	0.12871552	0.66081334	0.91268442	0.77053942	0.96001108	0.84896934	0.82987589	0.9567147	0.96129777	0.85865014	0.9275705	0.76309376	0.87539921
12	0.12903709	0.63840501	0.90567105	0.80088129	0.87621293	0.89920638	0.90973637	0.92444041	0.89657278	0.70178304	0.87074795	0.88885715	0.92963062
13	0.13117409	0.66878143	0.88229552	0.87466108	0.81910373	0.83254183	0.94145341	0.84141685	0.72382693	0.78352507	0.96285739	0.84581972	0.99527364
14	0.10384585	0.62407892	0.88299817	0.96376891	0.94389537	0.97555408	0.87845942	1	0.42350582	0.8411759	1	1	1
15	0.06789102	0.59207841	0.7675323	0.78382889	0.82952537	1	0.78438361	0.89931769	0.40189427	0.77882396	0.88136286	0.83252232	0.969557
16	0.06144615	0.60331366	0.48952396	0.63959483	0.60901764	0.87525146	0.63131916	0.65702524	0.29495813	0.67492022	0.6441045	0.62747622	0.65155091
17	0.06432359	0.5737976	0.29904841	0.38998372	0.35158271	0.46420426	0.38718412	0.57210996	0.23032906	0.38987905	0.4076063	0.40306042	0.37338942
18	0.06296253	0.91285113	0.18498474	0.20946698	0.30406233	0.26637896	0.23769321	0.21466128	0.16625777	0.27713508	0.25460182	0.22293284	0.28892743
19	0.08975004	1	0.16838583	0.16604573	0.15644994	0.17425811	0.13940301	0.15031895	0.15001427	0.16833082	0.13284037	0.14556596	0.15448821
20	0.10270387	0.95594488	0.11920163	0.1252749	0.12458738	0.11781521	0.09642647	0.10739322	0.09462441	0.10710142	0.10438392	0.09186948	0.09497965
21	0.07389991	0.96090414	0.10457271	0.11872576	0.0945686	0.1014319	0.09561718	0.10514492	0.07556827	0.11085426	0.10962151	0.09379669	0.09760382
22	0.06002198	0.93489779	0.10775052	0.11535685	0.0883509	0.1047373	0.09480695	0.10851119	0.07471796	0.11061734	0.10330267	0.0698474	0.08631472
23	0.06331954	0.63348235	0.08159273	0.08116118	0.06701209	0.07096344	0.07471816	0.08298594	0.04827518	0.1756192	0.07984473	0.05175557	0.07241732

Tabla 4.8 factores de variación horaria para un día lunes corresponde al medidor de la facultad de periodismo.

Resultados obtenidos de crear la red neuronal y predicciones de de un día lunes de la facultad de periodismo.

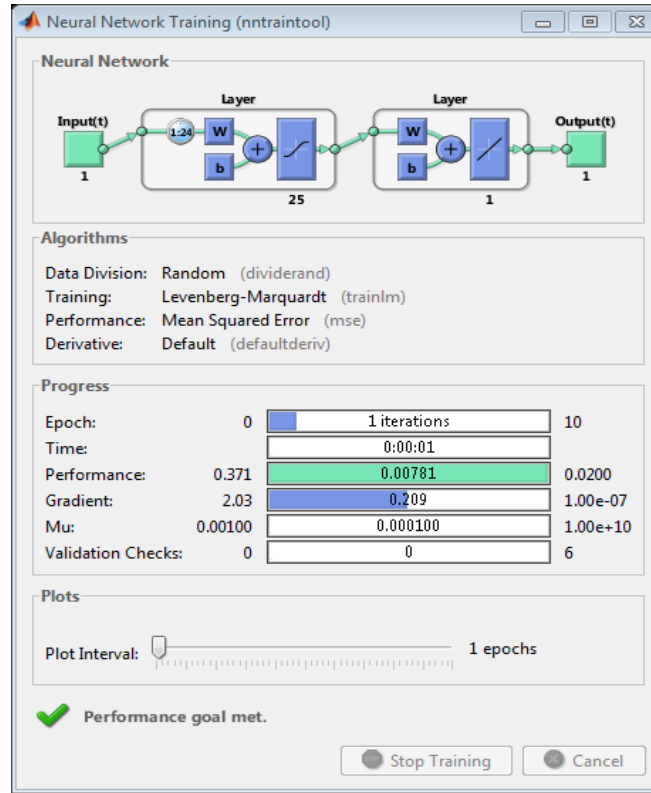


Figura 4.17 resultados de la red neuronal para un día lunes para la facultad de periodismo.

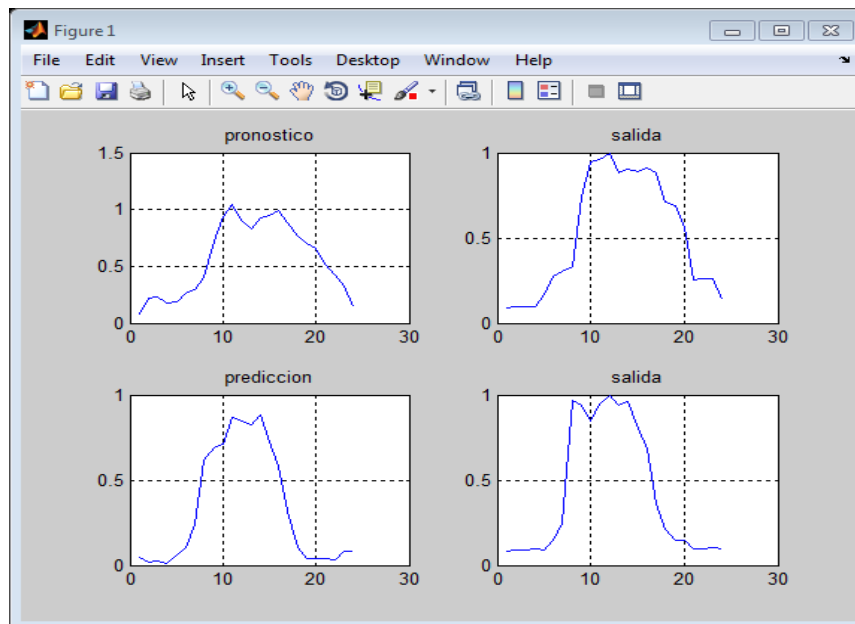


Figura 4.18 resultados de las gráficas para diferentes lunes de la facultad de periodismo.

La siguiente tabla corresponde a la facultad de química corresponde a los días martes

hora	fecha												
	18/07/2017	25/07/2017	08/08/2017	15/08/2017	22/08/2017	29/08/2017	05/09/2017	12/09/2017	19/09/2017	26/09/2017	03/10/2017	10/10/2017	17/10/2017
0	0.14030101	0.15468605	0.18558873	0.09458845	0.11649711	0.14222251	0.1598315	0.11400859	0.14494724	0.12496522	0.11851992	0.12982356	0.10988482
1	0.15642526	0.15721829	0.18629541	0.09998793	0.12632939	0.09342089	0.14306736	0.11742652	0.17605784	0.13744249	0.12182168	0.13392587	0.10088328
2	0.13093777	0.1560756	0.17825716	0.0940142	0.13416953	0.13245679	0.15322389	0.11408997	0.17109308	0.12020477	0.12127192	0.11046478	0.12907665
3	0.15544115	0.14983124	0.19461127	0.0919708	0.11215746	0.12593776	0.14524385	0.11423086	0.12999825	0.10811757	0.12186932	0.13178026	0.09600857
4	0.13487011	0.15109583	0.17348516	0.07824942	0.13305299	0.13044458	0.14911313	0.10846382	0.17823229	0.13910848	0.12337725	0.12985803	0.11989302
5	0.18931229	0.19462036	0.22627636	0.12634161	0.17029964	0.17410304	0.18237331	0.19064752	0.23466843	0.17629738	0.2042933	0.19434027	0.1367895
6	0.24424145	0.29018747	0.29402033	0.1642877	0.18959322	0.18330033	0.25490306	0.30874428	0.30760357	0.24265736	0.28089193	0.23591943	0.22180222
7	0.53142936	0.4703656	0.46368361	0.3698114	0.49640709	0.3259	0.44959229	0.4379171	0.44687499	0.41739411	0.5045192	0.48260911	0.34299561
8	0.7434863	0.81150246	0.71577136	0.58914121	0.64032289	0.54578821	0.62108169	0.73273375	0.45498254	0.52804572	0.53099971	0.50126737	0.51492425
9	0.78115021	0.89426245	0.82367609	0.7630073	0.82869422	0.6688055	0.67328778	0.76687737	0.58676465	0.68759909	0.76257374	0.62621738	0.52566185
10	0.73713766	0.90982628	0.8213763	0.76498776	1	0.8780081	0.51959892	0.79886586	0.59073153	0.65617728	1	0.70108615	0.5912873
11	0.78029486	0.9424836	1	0.63963917	0.95930895	0.87980758	0.6928289	0.89317593	0.60992626	0.63227283	0.88201316	0.67938027	0.66707997
12	0.78971974	0.83787955	0.90281838	0.62187398	0.76131935	0.8064393	0.81558189	0.84236127	0.60252337	0.61689682	0.80676141	0.56849487	0.59712737
13	0.86000209	0.88549727	0.79238444	0.80603253	0.90155374	0.94407155	0.93069193	1	0.96579159	0.70727673	0.68665802	0.73402566	0.8683016
14	0.87194106	1	0.23955191	1	0.81671147	1	1	0.93149526	1	1	0.79509874	1	1
15	1	0.93881059	0.17055941	0.8012378	0.65273928	0.83993727	0.97213091	0.75834914	0.83405094	0.83428459	0.65264288	0.72634828	0.8507806
16	0.64234915	0.59963949	0.218191	0.57616327	0.43193493	0.48041068	0.50343372	0.4764828	0.5028262	0.52631914	0.5340017	0.57004822	0.6454788
17	0.19152694	0.18018519	0.18919036	0.23304446	0.24958211	0.23911892	0.25023863	0.2293026	0.2790844	0.19356541	0.22050538	0.25647782	0.24051666
18	0.19363851	0.17365542	0.2166031	0.13197686	0.1359868	0.14706263	0.15828491	0.17425444	0.17889411	0.14183418	0.1884174	0.15531517	0.1175443
19	0.21577612	0.18553912	0.17428506	0.15413111	0.13745281	0.11482829	0.17730567	0.19302839	0.16572225	0.11625645	0.17816637	0.12594087	0.12524397
20	0.2088618	0.17809141	0.19687346	0.12915249	0.14887405	0.11912672	0.16676045	0.14367624	0.13803337	0.15106557	0.1592033	0.13877636	0.1025469
21	0.19047522	0.1707791	0.15724159	0.13074422	0.10922881	0.09802883	0.1465151	0.1474058	0.15470464	0.11739323	0.13175373	0.11832707	0.08395138
22	0.18496837	0.16576046	0.11836605	0.11010173	0.12736323	0.11793964	0.15317027	0.15953107	0.12893926	0.1269489	0.16750063	0.1084598	0.07740891
23	0.16763676	0.16907218	0.1252147	0.11886284	0.1213748	0.11286929	0.14231454	0.14262577	0.1258326	0.10299523	0.15093004	0.13152317	0.09020462

Tabla 4.9 factores de variación horaria para un día martes facultad de química.

Resultados obtenidos de la red neuronal creada para un día martes muestra la red y el entrenamiento además de eso la gráfica para diferentes días simulados muestra las gráficas.

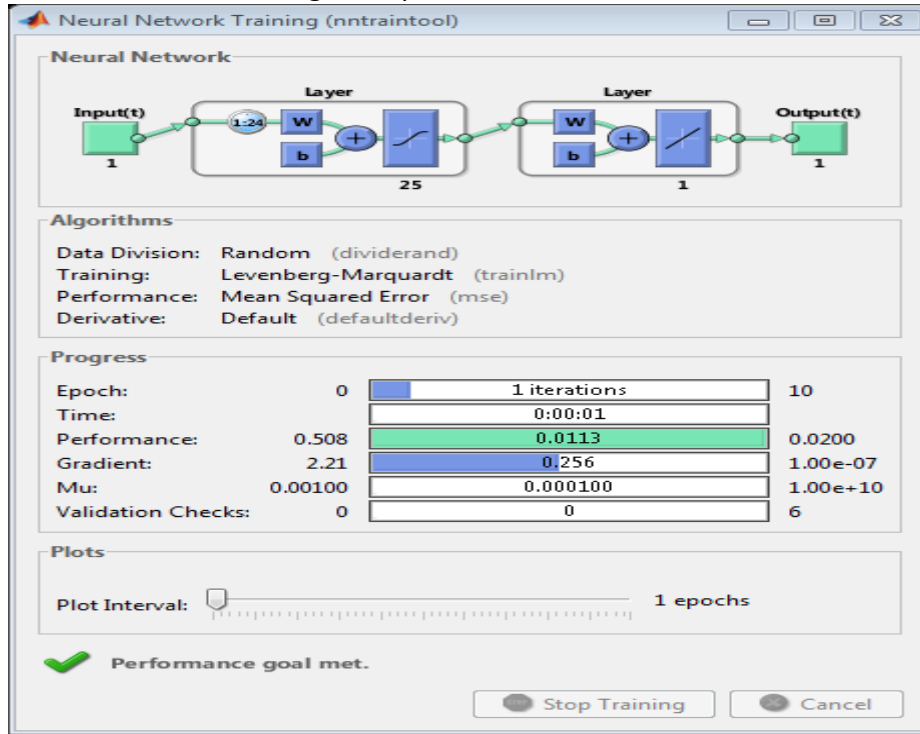


Figura 4.19 resultados de la red neuronal para un día martes facultad de química.

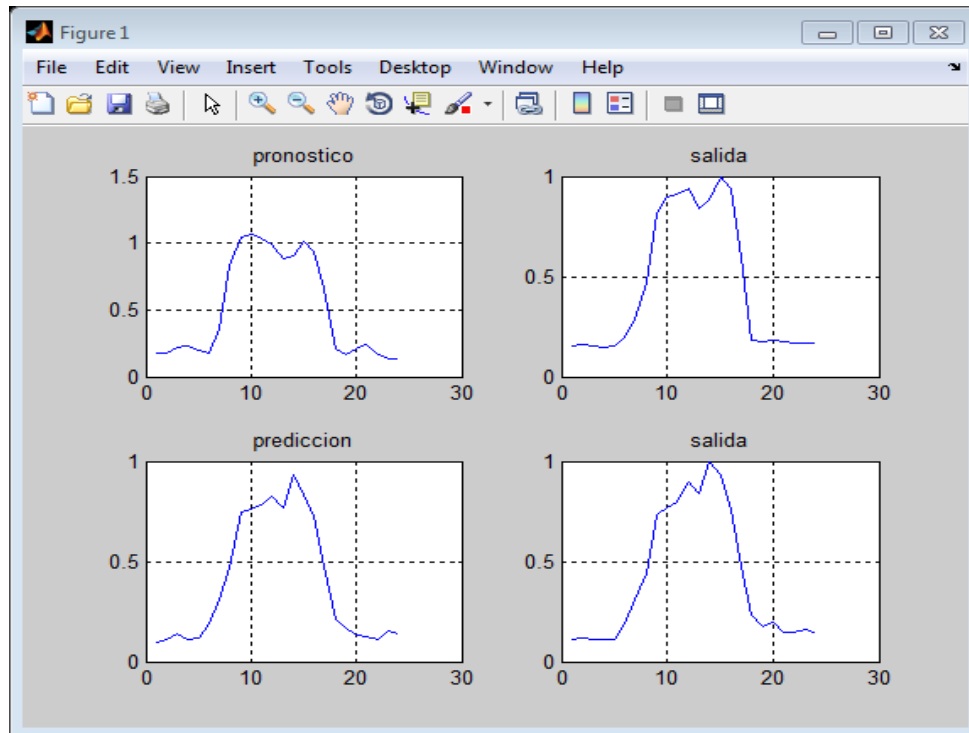


Figura 4.20 graficas de un día martes de pronóstico y predicción para un día martes facultad de química

5. Modelo definitivo para el pronóstico de demandas.

5.1 METODOLOGIA GENERAL:

1. Representar gráficamente la serie, su función de autocorrelación simple (FAS) y su función de autocorrelación parcial (FAP). La gráfica de la serie nos indica si la serie es estacionaria o no. Según los motivos por los que la serie no es estacionaria, tendremos que aplicar uno de los siguientes procedimientos hasta hacerla estacionaria.
 - a. Si tiene tendencia: Tomaremos diferencias regulares hasta que desaparezca. Normalmente el orden de la diferencia será 1, y raramente será mayor a 3.
 - b. Si es heterocedástica, es decir, no tiene varianza constante, habrá que transformar la serie. Con tomar el logaritmo es suficiente en nuestro caso, aunque existen algunas transformaciones más complejas, como las de Box-Cox.
 - c. Si es estacional: Tomaremos diferencias estacionales hasta que desaparezca el patrón que se repite. En la práctica es muy raro tener que aplicar más de una diferencia estacional.
2. Una vez que el gráfico de la nueva serie (una transformación de la original) indica que es estacionaria, podemos intentar deducir su estructura (¡no la de la serie original!) observando la FAS y la FAP, teniendo en cuenta que:
 - a. El decrecimiento de la altura de las barras (los coeficientes de la función de autocorrelación simple y parcial) debe ser exponencial. También suelen aparecer una alternancia de los signos o una forma sinusoidal; la alternancia depende de los signos de los parámetros, y la forma sinusoidal es debida a que hay raíces complejas en la ecuación característica del proceso. Si no hay decrecimiento exponencial, suele deberse a que la serie es integrada; para éstas el decrecimiento es lineal, no exponencial. (Hay un tipo de series estacionarias que tienen memoria mucho más larga que la de los AR, se las llama de «memoria larga», que muestran ese decrecimiento lineal; pero no se han visto y por tanto debéis pensar en las integradas cuando haya decrecimiento lineal.) Una serie es integrada cuando al tomar diferencias («diferenciarla») se convierte en una serie estacionaria; si esta serie estacionaria es un modelo ARMA(p,q) decimos que son modelos ARIMA(p,d,q).
 - b. Una de las dos funciones de autocorrelación va a mostrar una estructura como la descrita en a), mientras que la otra mostrará algunas barras que salen de la banda de significatividad. La posición de la última de estas barras, que suele coincidir con el número de barras que sobresalen, nos dicen el

orden del proceso. Según sea la FAS la que presenta la estructura de a) o sea la FAP, diremos que el proceso es AR o MA, respectivamente.

- c. Hay situaciones que no son muy claras, en estos casos puede tratarse de un modelo con las dos partes, la AR y la MA. Estos son los modelos ARMA(p,q). Sus funciones FAS y FAP son combinaciones de las de ambas partes, por lo que son difíciles de identificar a simple vista.
- d. La FAS y la FAP de un ruido blanco (serie de datos procedentes de variables sin ninguna estructura común, es decir, independientes entre sí) tiene todos los coeficientes nulos en teoría, o no significativos en nuestras gráficas. Respecto a la banda de significatividad, se advierte en el libro «Análisis de series temporales», de Daniel Peña, que: «como [...] estos límites sobrestiman la varianza en los retardos pequeños, un valor próximo a los límites de confianza $\pm 2/\sqrt{T}$ en los retardos iniciales debe considerarse un indicio claro de que el modelo es inadecuado» (T es la longitud de la serie).

Hay que tener en cuenta que las líneas de significatividad que aparecen en las gráficas de las funciones de autocorrelación suelen determinar un límite de un nivel de confianza del 95%, es decir, el coeficiente que sobresale se puede considerar no nulo con un 95% de seguridad. Esto implica que puede haber un 5% de barras que sobresalgan sin motivo real. Debemos pensar esto cuando hay alguna barra aislada que no cuadra con el resto de la estructura de barras. Debido a los estimadores utilizados para esas barras, y para garantizar que las estimaciones se basan en un número suficiente de datos, el programa elige el número de barras que representa. Cuando hay 20, por ejemplo, puede suceder que 1 se haya salido erróneamente de la banda de significatividad.

- 3. Observando las tres gráficas anteriores podemos hacernos una idea de qué modelo subyace a nuestra serie, o al menos de cuáles son los primeros candidatos que debemos probar. Para comprobar analíticamente (no visualmente) un modelo, vamos a intentar ajustar un modelo ARIMA(p,d,q) y observaremos dos cosas que miden lo bien que ajusta el modelo: lo necesarios que son todos sus términos, y lo que se parecen al ruido blanco los residuos que nos quedan. Estos son los dos análisis que hemos visto en la Práctica 4. Ajustamos el modelo ARIMA(p,d,q) según lo que hayamos deducido de todos los apartados anteriores: el valor de «d» es el número de diferencias regulares que hemos tenido que tomar en «1.» para hacer la serie estacionaria, «p» es el orden de la parte AR y «q» el de la parte MA. Frecuentemente vamos a tener que probar varios modelos ARIMA(p,d,q) hasta que encontremos el modelo que nos convence, incluso cuando se cumplan las condiciones que se exigen en los siguientes análisis:

- a. Las condiciones en el primero son que, para cada coeficiente del modelo propuesto, el valor absoluto del estadístico «t» sea mayor que 2, a la vez que

el p-valor de ese contraste (que mide si se puede considerar cero o no) es menor que 0,05. Los coeficientes que no cumplan estas dos condiciones no son necesarios en el modelo, porque no hay evidencia de que deban ser no nulos. Esto nos ayuda en la búsqueda de nuestro modelo lo más sencillo posible.

- b. El otro análisis nos proporciona varios contrastes de hipótesis. Vamos a fijarnos en el de Box-Pierce, aplicado a los residuos después de ajustar el modelo. En este caso el contraste confirma con más fuerza que los residuos son ruido blanco cuanto mayor es su p-valor. El nivel mínimo es 0,05, es decir, debe suceder que el p-valor sea mayor que 0,05; y cuando mayor sea, más evidencia hay de que los residuos son ruido blanco (y mejor hemos captado en el modelo la parte que no se debe al azar).
 - c. Otro recurso (gráfico) es el de observar, tras ajustar el modelo ARIMA, las funciones FAS y FAP de los residuos, lo que nos dará una información gráfica de la no correlación de estos.
4. Finalmente, si se ha transformado la serie, el modelo encontrado es el de la transformación, no el de la serie original. Esto se puede indicar diciendo que, después de hacer las transformaciones, la serie resultante tiene estructura transformada. Puesto que no hay un único modelo correcto (que explica suficientemente bien los datos), de la misma manera que no hay una única función que pase por varios puntos que tracemos en un sistema de ejes XY, algunos criterios que se pueden seguir para elegir un modelo satisfactorio son:
- a. Debe verificar las condiciones gráficas y numéricas necesarias.
 - b. Sus residuos después de ajustar el modelo deben ser ruido blanco.
5. Cuando el orden de una parte, AR ó MA, es muy alto, podemos probar a añadir algún término de la otra parte y restarle orden a ésta.
6. En último término, la elección del modelo final debe ser un equilibrio entre simplicidad y eficacia, porque siempre sería posible añadir parámetros indefinidamente para obtener mejoras...

5.2 OBTENCION DE MODELOS.

5.2.1 Transformación para heterocedasticidad de la serie

Primeramente, para tener una idea a priori se grafica la serie tal y como está, con esto se puede observar la estacionalidad que está presente y la tendencia de los datos, una manera de tener más claro la naturaleza de las componentes de la serie es descomponiéndola en sus partes esenciales, tendencia, estacional, y aleatoria, como se muestra en la figura 5.1.

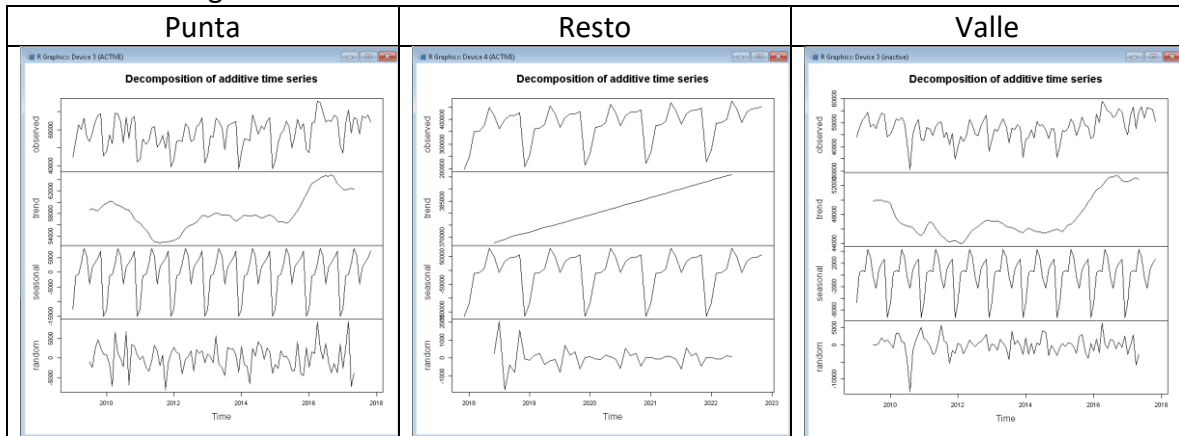


Figura 5.1. Graficas punta, valle y resto.

Para cada uno de los bloques, se cuenta con datos desde 2009 hasta 2017. Se ha separado en 2 partes. La primera desde 2009 hasta 2016, que se utiliza para determinar el modelo. La parte restante es desde 2016 hasta 2017, la cual se utiliza para la comparación de los pronósticos realizados con cada modelo que se determine.

Comenzamos analizando la heterocedasticidad de la serie. Para esto se realiza una prueba de relación varianza versus media, hacen grupos anuales de media y varianza y procedemos a analizar si existe una relación lineal entre ambas cantidades propias de la serie.

Para esto se realiza una prueba de t-student que determine matemáticamente la existencia de relación lineal entre la varianza y la media, obteniendo el siguiente resultado para cada serie.

En R primero se crea un dataframe que contenga los datos agrupados anualmente de media y varianza de la serie, luego se crea un modelo lineal con `lm` (fitting linear models), del cual presenta el resumen del resultado para cada uno de los modelos en la tabla. Esto calcula una pendiente y un intercepto con un error mínimo cuadrado, la hipótesis nula para cada intercepto y pendiente es “Los coeficientes que relacionan las variables son cero (variables son independientes)”, para un nivel de confianza del 95%, el valor de prueba p debe que ser mayor a 0.05.

Punta	Resto	Valle
Intercept: -1.110e+04	Intercept: -1.197e+05	Intercept: 14570.1011
Pendiente: 3.301e-01	Pendiente: 6.147e-01	Pendiente: -0.2228
Valor p: 0.03304	Valor p: 0.04393	Valor p: 0.3978
Conclusión: Existe dependencia entre varianza y media	Conclusión: Existe dependencia entre varianza y media	Conclusión: No existe dependencia entre varianza y media

Tabla 5.1. Resultados de la regresión para determinar si existe relación media-varianza.

De las series temporales en las cuales existe dependencia entre varianza y media se toma el logaritmo de éstas para proceder a analizar los posibles modelos, en casos contrarios simplemente se ocupa la serie sin ningún cambio.

5.2.2 Determinación de los órdenes de diferenciación d y D

A continuación se hará el análisis de estacionariedad de la serie, para esto se grafica las funciones de autocorrelación simple y autocorrelación parcial de la serie, como se muestra en la figura, para esto usamos las funciones en R `acf()` y `pacf()` respectivamente.

Se debe tener en cuenta que las líneas de significatividad que aparecen en las gráficas de las funciones de autocorrelación suelen determinar una zona de una significatividad del 95%, es decir, el coeficiente que sobresale se puede considerar no nulo con un 95% de seguridad. Esto implica que puede haber un 5% de barras que sobresalgan sin motivo real.

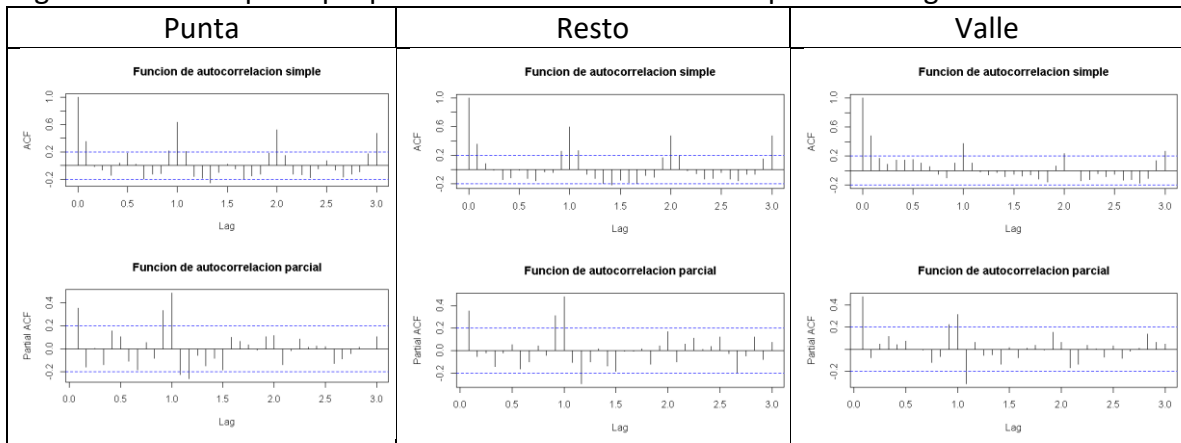


Figura 5.2. Graficas obtenidas de la FAS y FAP de las series.

Ahora nos enfocamos en adecuar la serie para aplicar el modelado de la serie, es decir, hacerla estacionaria.

Para esto nos valemos de lo observado en la serie original y sus funciones de autocorrelación simple y parcial, sea aplica una diferencia regular a las tres series para eliminar la tendencia, y una diferencia estacional de periodo 12 para eliminar la estacionalidad.

5.2.3 Identificación de posibles modelos

Ahora se procede a la identificación de posibles modelos que puedan representar la serie, se analizarán únicamente dos modelos para cada serie, a partir de las gráficas de autocorrelación simple y parcial observamos los siguientes posibles modelos para cada serie

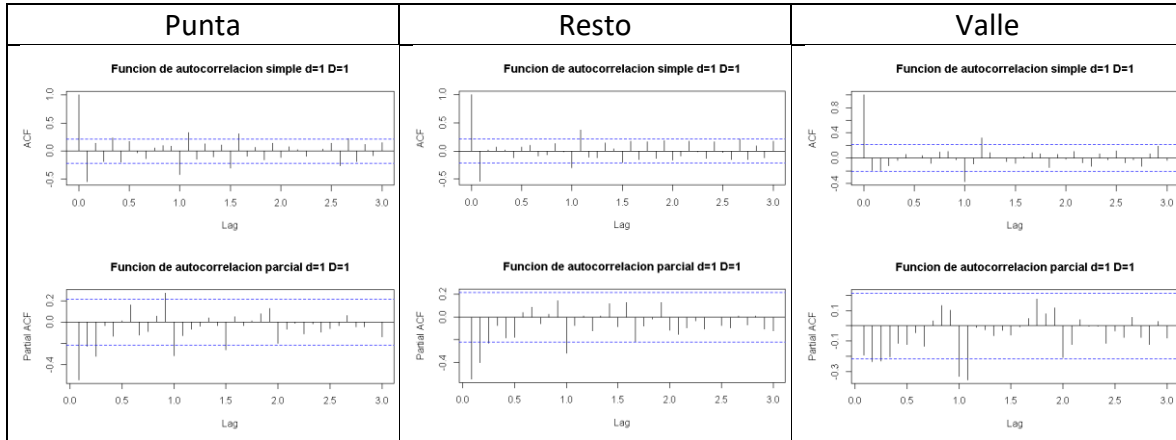


Figura 5.3. FAS y FAP luego de aplicar una diferencia regular y una estacional.

Punta	Resto	Valle
<p>Modelo 1: ARIMA(0,1,1)x(0,1,1) ma1 -0.6903 sma1 -1.0000</p>	<p>Modelo 1: ARIMA(0,1,1)x(0,1,1) ma1 -0.8323 sma1 -0.5564</p>	<p>Modelo 1: ARIMA(2,1,1)x(0,1,1) ar1 0.3241 ar2 0.0324 ma1 -0.8289 sma1 -0.7847</p>
<p>Modelo 2: ARIMA(0,1,1)x(1,1,1) ma1 -0.6858 sar1 -.2297 sma1 -0.9999</p>	<p>Modelo 2: ARIMA(0,1,2)x(1,1,1) ma1 -0.8314 ma2 0.0757 sar1-0.0207 sma1 -1.0000</p>	<p>Modelo 2: ARIMA(0,1,2)x(0,1,1) ma1 -0.5214 ma2 -0.1761 sma1 -0.7806</p>

Tabla 5.2. Determinación de coeficientes para los posibles modelos identificados.

Ahora se comprueban cada uno de los posibles modelos y se contrasta con los valores reales del año 2017. El indicador de desempeño que se utilizará será el MAPE el cual se define como.

$$MAPE = \frac{\sum_{l=1}^n \frac{|X(l) - \hat{X}(l)|}{|X(l)|}}{n} \times 100\% \quad (5.1)$$

Donde $X(l)$ es el valor real de energía en un tiempo $t + l$ hacia adelante, y $\hat{X}(l)$ es el pronóstico generado por el posible modelo. El resultado se resume en la tabla.

Punta	Resto	Valle
Modelo 1: ARIMA(0,1,1)x(0,1,1) MAPE 6.32%	Modelo 1: ARIMA(0,1,1)x(0,1,1) MAPE 10.26%	Modelo 1: ARIMA(2,1,1)x(0,1,1) MAPE 5.71%
Modelo 2: ARIMA(0,1,1)x(1,1,1) MAPE 6.62%	Modelo 2: ARIMA(0,1,2)x(1,1,1) MAPE 10.71%	Modelo 2: ARIMA(0,1,2)x(0,1,1) MAPE 5.89%

Tabla 5.3. Rendimiento de cada modelo identificado.

5.2.4 Refinado de modelo

Ahora que se ha seleccionado el modelo que produce un menor error porcentual absoluto medio, se procede a mejorar los resultados que pueda arrojar, para esto se modifica la serie de tiempo retirando datos atípicos que esta pueda contener, para ello observamos cada una de las gráficas que corresponden a los residuos que produce el modelo, es decir la componente aleatoria de la serie.

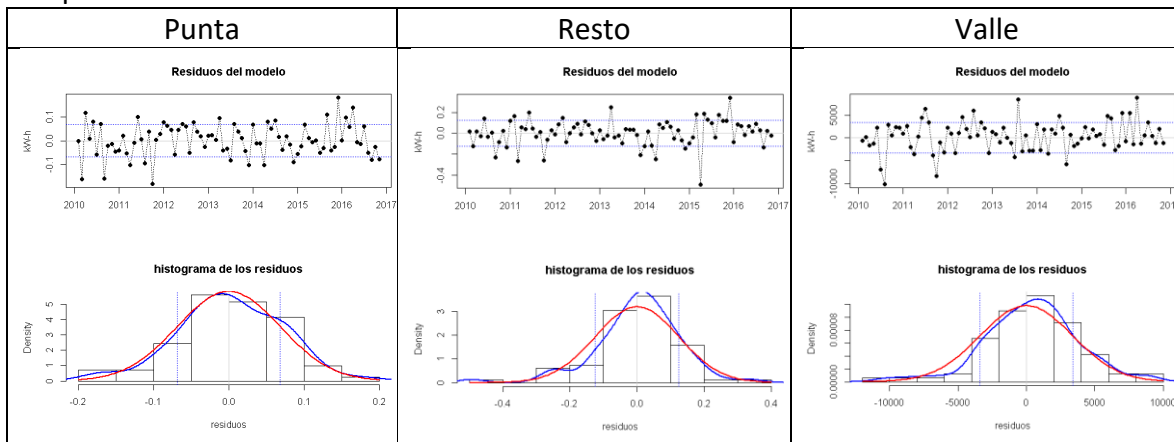


Figura 5.4. Gráficas de residuos del modelo (ruido blanco característico del proceso).

Por cuestiones de simplicidad se retirarán únicamente los datos que se identifiquen visualmente como atípicos (los que más se alejan de las líneas de varianza).

Punta	Resto	Valle
Marzo 2010	Abril 2015	Agosto 2010
Septiembre 2010	Diciembre 2015	Agosto 2013
Octubre 2011		Abril 2016
Diciembre 2015		

Tabla 5.4. Datos atípicos identificados en las gráficas de residuos.

Con estos atípicos retirados, se procede nuevamente a calcular los coeficientes el MAPE y los residuos para los modelos seleccionados. Para el caso del modelo para valle, al retirar los valores atípicos de la serie, ahora si se aplica transformación de logaritmo a la serie.

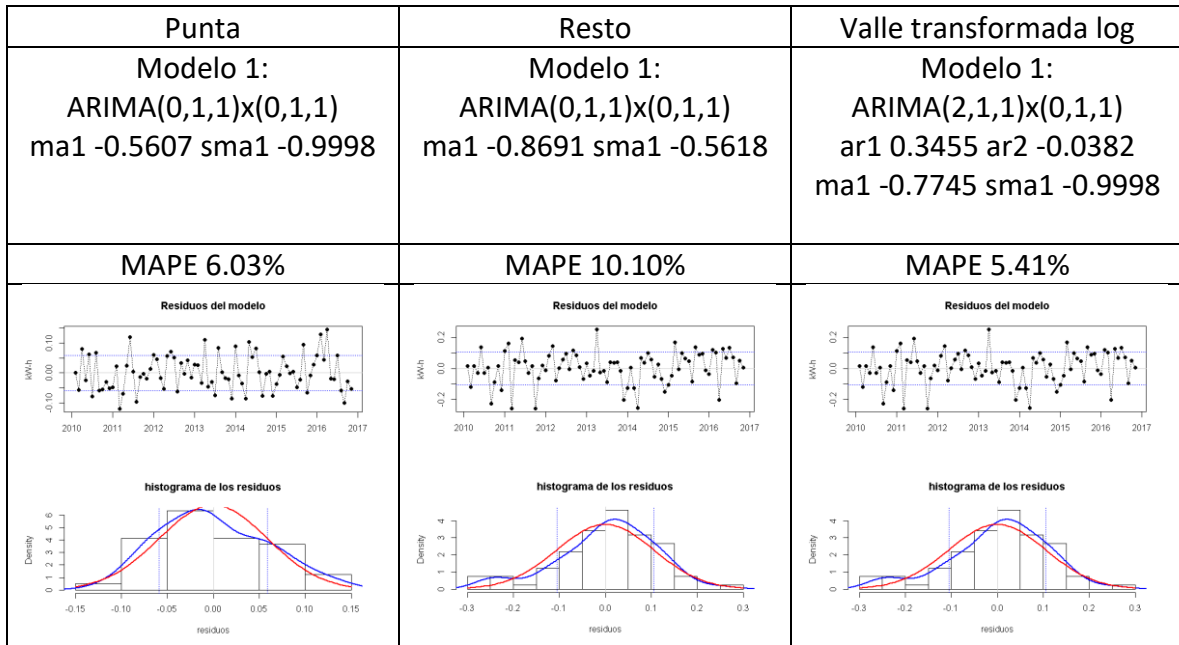


Figura 5.5. Resultados luego de retirar atípicos de la serie original.

5.3 Prueba de hipótesis para los residuos de los modelos definidos

Ahora con los modelos obtenidos, procedemos a comprobar las hipótesis supuestas al aplicar este método, estas fueron:

- Media igual a cero
- Varianza constante
- Falta de correlación para cualquier retardo
- Distribución normal

5.3.1 Hipótesis de media cero

En este caso el programa R cuenta con la prueba t-student $t.test(residuos, mu=0)$ para media de una serie de muestras. Se muestran los resultados para todas las series en la tabla

Punta	Resto	Valle
$t=0.018917, df=81$	$t=0.32567, df=81$	$t=1.012, df=81$
valor-p=0.985	valor-p=0.7455	valor-p=0.3146
Media=0.0001235345	Media=0.003813675	Media=0.006320587

Tabla 5.5. Resultados de prueba de hipótesis "media de los residuos igual a cero".

Ahora prestamos atención al valor-p de la prueba de t-student, el cual es mayor a 0.05, por lo que aceptamos la hipótesis nula: la media de los residuos es igual a cero.

5.3.2 Hipótesis de varianza constante

Para esta prueba de varianza constante de los residuos, tenemos que dividir en dos grupos los residuos generados por el modelo. Hacemos en R lo siguiente para el total de 82 residuos.

residuos1=residuos[1:41] (numerador)
 residuos2=residuos[42:82](denominador)

Luego hacemos la prueba de varianzas conocida como test F `var.test(residuos1,residuos2)` para los dos grupos de residuos.

Punta	Resto	Valle
F=0.8129, num df=40 den df=40	F=1.1222, num df=40 den df=40	F=1.0247, num df=40 den df=40
valor-p=0.5154	valor-p=0.7171	valor-p=0.9388
Razón de varianzas= 0.8129028	Razón de varianzas= 1.122246	Razón de varianzas= 1.024746

Tabla 5.6. Resultados de prueba de hipótesis “varianza de los residuos constante”.

El valor p de la prueba es mayor a 0.05, y por lo tanto se acepta la hipótesis nula: las varianzas de ambos grupos son iguales, y por lo tanto constantes para todos los residuos. Además de la prueba de contraste de varianzas se grafica las funciones de autocorrelación simple y parcial para la serie residuos generada por cada modelo. En esto vemos que para cualquier retardo mayor que cero se encuentren debajo de los límites de significancia, lo que nos dice con un 95% de seguridad que se trata de ruido blanco.

5.3.3 Contraste de correlaciones - Test de Ljung-Box

Ahora se realiza una prueba que fundamenta matemáticamente la hipótesis de incorrelación de los residuos.

Para comprobar la falta de correlación de los residuos, comprobamos primero graficando su función de autocorrelación simple y su función de autocorrelación parcial. Estas deben estar por debajo de la línea de significancia cero, esto quiere decir con un nivel de confianza del 95% que no existe correlación en las componentes de ruido blanco.

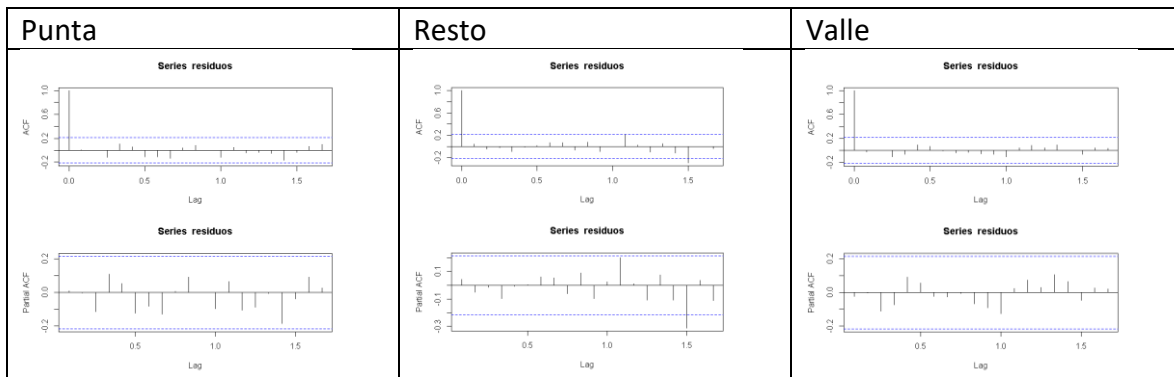


Figura 5.6. Graficas de auto correlación simple y parcial de los residuos.

Llevando esta prueba a un nivel más detallado, aplicamos la prueba de Ljung-Box como se explicó en la sección 3.6.4.2

En R utilizamos la función `Box.test(residuos, lag=30, type="Ljung", fitdf)`, lo anterior lo realizamos para los primeros 30 residuos obtenidos con el modelo y con dos parámetros tomados en cuenta para ajustar el modelo $fitdf=p+q+P+Q$.

Punta	Resto	Valle
χ -cuadrada=18.004, df=28	χ -cuadrada=26.8, df=28	χ -cuadrada=17.278, df=26
valor-p=0.9261	valor-p=0.5292	valor-p=0.9005

Tabla 5.7. Resultados de prueba de hipótesis "incorrelacion de residuos".

En este caso tomamos la hipótesis como válida, ya que el valor p es mayor a 0.05 y por lo tanto esta fuera del área de rechazo.

5.3.4 Prueba de normalidad

El test de Kolmogórov-Smirnov tiene la finalidad de probar la diferencia entre dos formas características de distribución provenientes de dos muestras de datos, en lugar de comparar tendencia, dispersión u otros parámetros, teniendo como hipótesis nula H_0 : las dos muestras tienen una forma igual de distribución.

En nuestro caso se realiza una prueba con una muestra `ks.test(residuos,"pnorm",mean,sd)` la cual se compara con una curva de distribución normal con media y desviación estándar iguales.

Punta	Resto	Valle
D=0.090816	D=0.080432	D=0.074382
valor-p=0.4805	valor-p=0.6341	valor-p=0.7623

Tabla 5.8. Resultados de prueba de hipótesis "distribución normal de los residuos".

El valor p para la prueba de normalidad es mayor a 0.05 por lo que se acepta la hipótesis nula formulada: la distribución de las muestras se ajusta a la forma de distribución normal.

6 Conclusiones

La presente tesis tuvo como objetivo la investigación de cómo crear una red neuronal para que pueda predecir la curva de variación horaria de un día cualquiera, el modelo de red neuronal pareciera ser el mismo para las diferentes facultades, pero, posee diferentes entradas y diferentes salidas, el tipo de entrenamiento es el mismo, sin embargo, lo que varía son el valor de los pesos para diferentes entradas con el fin de, poder lograr la obtención de la salida deseada.

Es importante realizar un muestreo de los datos a la hora de seleccionar los datos a analizar podrían existir datos atípicos que tengan efectos negativos a la hora de mostrar los resultados, para el método de redes neuronales se muestran varios datos que tienen datos atípicos si hay datos atípicos se tendrá que entrenar la red neuronal para poder adaptar a una nueva salida en este caso sería la curva de variación horaria.

Cabe resaltar que, para poder predecir una curva de variación horaria es importante tener datos de días, meses o años atrás y así poder estimar la salida deseada, es importante tener una base de datos de estos días con el fin de, poder entrenar la red neuronal con diferentes datos para que se vaya adaptando a la salida deseada, en este documento se ocupó un entrenamiento supervisado pues conocíamos datos de días atrás que nos ayudaron a obtener la salida deseada.

No existe un único modelo matemáticamente correcto para la representación de una serie de tiempo, por lo tanto, pueden existir otros modelos, aparte de los presentados, que modelen la serie de tiempo de forma aceptable. Los modelos aquí expuestos son los que toman en cuenta la mínima cantidad de variables necesarias.

Tener en cuenta a la hora de realizar pronósticos, que estos tienen una incertidumbre, como se mencionó en el capítulo 3, la cual crece exponencialmente. La metodología de series de tiempo, tiene la ventaja de ser precisa, y se puede extender a una variedad de casos de los cuales se pueda disponer de una serie de datos lo suficientemente larga para poder dar validez a pronósticos de hasta mediano plazo.

Los indicadores de desempeño obtenidos fueron los siguientes:

- Serie punta ARIMA(0,1,1)x(0,1,1) pronóstico año 2017, MAPE: 6.03%.
- Serie resto ARIMA(0,1,1)x(0,1,1) pronóstico año 2017, MAPE: 10.10%.
- Serie valle ARIMA(2,1,1)x(0,1,1) pronóstico año 2017, MAPE: 5.41%.

7 Referencias

1. **Castano, Samuel Ramirez.** *Redes De Distribucione Electrica.* Colombia : Centro de Publicaciones Universidad Nacional de Colombia Sede Manizales, 2004.
2. **GÖNEN.** *Electric Power Distribution System Engineering.* Turan : McGraw-Hill, 1986.
3. **DEL BARRIO, Tomás and CLAR, Miguel y Suriñach, Jordi.** *Modelo de regresión lineal múltiple: especificación, estimación y contraste.* Barcelona Espana : UOC Papers, 2002.
4. **ANDERSON, David and SWEENEY, Dennis y WILLIAMS, Thomas.** *Estadística para administración y economía.* . México : Cengage Learning Editores S.A., 2008.
5. **David Kriesel, Doctor Aleman.** *A Brief Introduction to Neural Networks.* Estados Unidos : s.n., 2005.
6. **BOWERMAN, Bruce and O'CONNELL, Richard y KOEHLER.** *Pronósticos, series de tiempo y regresión: Un enfoque aplicado.* México : Internacional Thomson, 2007.
7. **HILERA, José y MARTÍNEZ, Víctor.** *Redes Neuronales Artificiales: Fundamentos, Modelos y Aplicaciones.* . Madrid : Alfa Omega-Rama, 2000.
8. **FLÓREZ, Raquel y FERNÁNDEZ José Miguel.** *Las Redes Neuronales Artificiales: Fundamentos teóricos y aplicaciones prácticas.* España : Netbiblo, 2008.
9. **Pedro Isasi Vinuela, Ines M. Galvan Leon.** *Redes Neuronales Artificiales Un Enfoque Practico.* madrid Espana : Pearson Educacion., 2004.
10. **Mark Hudson Beale, Martin T. Hagan, Howard B. Demuth.** *Neural Network Toolbox™*
7. U.S : s.n., 1992–2010.
11. **Bartlett, M. S.** *Stochastic Processes.* Cambridge : Cambridge University Press, 1955.
12. **Durbin, J.** *The fitting of time-series models.* s.l. : Rev. Internat. Statist. Inst.
13. **Quenouille.** *Approximate tests of correlation in time-series.* s.l. : J. Royal Statist. Soc.
14. **Peña, Daniel.** *Analisis de series de tiempo.* s.l. : Alianza Editorial, 2005.
15. **George E. P. Box, Gwilym M. Jenkins, Gregory C. Reinsel.** *Time Series Analysis 4th Ed.* s.l. : WILEY, 2008.
16. **Durbin.** *An alternative to the bounds test for testing for serial correlation in least-squares regression.*
17. **Foundation, Neural Newtworks: A Comprehensive.** *Simon Haykin.* s.l. : 2 Edicion.

8 ANEXOS

```
%código de cómo crear una red neuronal que pronostique y predicción de la
%curva de demanda de un día cualquiera
datos=xlsread('Mediciones humanidades 3','Lunes entrenamiento','A2:BU2')
%se llama a los datos que se encuentran en la hoja de Excel llamada lunes
de entrenamiento.

datos=con2seq(datos) % se convierte los datos a formato de matriz

datos=datos(1:48) % se llaman los primeros 48 horas de variacion horaria

net=newfftd(datos,datos,(1:24),25) % se crea la red neuronal con un
retardo de (1:24) horas y 25 neuronas en la capa oculta.

% se ingresan parámetros de entrenamiento
net.trainParam.epochs=10 %número de iteraciones

net.trainParam.goal=0.02 %error de 0.02 para que se detenga las
iteraciones

net.trainParam.show=1 %que muestre iteración paso a paso

net.trainParam.showCommandLine=1 % que muestre resultados línea por línea

entrada=datos(25:48) % datos de entrada a la red son 24 datos

Salida=datos(25:48) % datos de salida son los mismos que las entrada para
el entrenamiento

valorinicial=datos(1:24)% valor de retardo de la red corresponde a las
datos anteriores las primeras 24 horas

[net,tr]=train(net,entrada,salida,valorinicial) %entrenamiento de la red

pronostico=sim(net,entrada,valorinicial) % simulación de la red

salida

pronostico=cell2mat(pronostico) % pasamos de formato de matriz a forma de
vector fila para poder graficar los resultados

salida=cell2mat(salida) % pasamos de formato de matriz a forma de vector
fila para poder graficar los resultados

subplot(2,2,1); plot(pronostico)
grid
title('pronostico')
subplot(2,2,2); plot(salida)
grid
```

```

title('salida')
save('net.mat', 'net') %se guarda la red neuronal y los parametros de
entrenamiento de ella

-----

load net.mat %se carga la red neuronal y los parametros anteriores

salida=xlsread('Mediciones humanidades 3','Lunes totales','A10:X10') %
estas entradas estan en un una hoja llamada lunes totales que se encuentran
todos los dias lunes

salida=con2seq(salida) % se convierten los datos a matriz

entrada=xlsread('Mediciones humanidades 3','Lunes totales','A9:X9') %
estas entradas estan en un una hoja llamada lunes totales que se encuentran
todos los dias lunes
entrada=con2seq(entrada)% se convierten los datos a matriz

valorinicial=xlsread('Mediciones humanidades 3','Lunes totales','A8:X8')
% estas entradas estan en un una hoja llamada lunes totales que se
encuentran todos los dias lunes

valorinicial=con2seq(valorinicial)% se convierten los datos a matriz.

%se ingresan parámetros de entrenamiento para las nuevas entradas y
%salidas
net.trainParam.epochs=10;

net.trainParam.goal=0.02;

net.trainParam.show=1;

net.trainParam.showCommandLine=1;

[net,tr]=train(net,entrada,salida,valorinicial) % entrenamiento de la
red neuronal
prediccion=sim(net,entrada,valorinicial) % simulación de la red neuronal
prediccion=cell2mat(prediccion) % pasamos de formato de matriz a forma
de vector fila para poder graficar los resultados
salida=cell2mat(salida) % pasamos de formato de matriz a forma de vector
fila para poder graficar los resultados

error=abs(salida-prediccion) % se calculan el error de las salidas vrs
las predicciones.

subplot(2,2,3); plot(prediccion)
grid
title('prediccion')
subplot(2,2,4); plot(salida)
grid
title('salida')

save('net.mat', 'net') % se guardan los procesos anteriores.

```

```

1. #hacer primero el analisis de estabilizacion de varianza, luego
   diferenciaciones
2. #luego acf y pacf para encontrar modelos y realizarlos
3. #luego, hacer un analisis de residuos para identificar atipicos en
   la serie
4. #hacer tratamiento de atipicos para hacer una serie con el modelo
   final identificado
5. #esto se hara para cada uno de los bloques horarios, y para las
   tarifas de distribucion.
6. #no se si hacerlo tambien con el cargo por potencia.
7.
8. #Analisis de varianza
9. #crear tabla de promedios y desviaciones estandar anuales
10.     #para analizar estabilidad en varianza
11.
12.     print("Introducir Datos de bloque horario a
analizar")
13.     Nombre_Bloque = scan(file="", what=character(), 1)
14.     Datos = read.table(paste(getwd(), "/", Nombre_Bloque, ".
csv", sep=""), header=TRUE, sep=";")
15.
16.     DatosComparativos = read.table(paste(getwd(), "/", Nomb
re_Bloque, " para confrontar modelo", ".csv", sep=""), header=TRUE,
sep=";")
17.     B = ts(DatosComparativos[2], frequency=12, start=c(201
6, 12))
18.     #crear variables donde se almacenaran los datos de
s(varianza) y u (media)
19.
20.     DesvStd = matrix(0, nrow(Datos)/12, 1)
21.     Promedio = matrix(0, nrow(Datos)/12, 1)
22.     for (k in 1:(nrow(Datos)/12)){
23.         Promedio[k, 1] = mean(Datos[(12*(k-
1)+1):(12*k), 2])
24.         DesvStd[k, 1] = sd(Datos[(12*(k-
1)+1):(12*k), 2])*sqrt(11/12)
25.     }
26.     #crear objeto data frame para aplicar el modelo de regresion
lineal
27.     Dispersion = data.frame(DesvStd, Promedio)
28.     Regresion = lm(Dispersion)
29.     print(summary(Regresion))
30.     print("¿Existe dependencia entre varianza y media?")
31.     decision=scan(file="", what=integer(), 1)
32.     if (decision == TRUE){
33.         serie=log(Datos[2])
34.         unidades="log kW-h"
35.     }else{
36.         serie=Datos[2]
37.         unidades="kW-h"
38.     }
39.     print("ingrese año inicial de la serie")
40.     aqo = scan(file="", what=integer(), 1)
41.     print("ingrese mes inicial de la serie")
42.     mes = scan(file="", what=integer(), 1)
43.     serie = ts(serie, frequency = 12, start=c(aqo, mes))
44.     dev.set(2)#ventana grafica 2 activa

```

```

45.         plot(serie, main=Nombre_Bloque, xlab="",
ylab=unidades, type="n")
46.         lines(serie, col="green")
47.         points(serie, bg="limegreen", pch=21)
48.         #grafica la serie observada y sus componentes
aleatorias, estacional, y tendencia
49.         win.graph()
50.         serie.descompuesta=decompose(serie)
51.         plot(serie.descompuesta)
52.
53.     #hacer una pausa
54.     print("presione enter para continuar")
55.     scan(file="",1)
56.     #determinar ACF y PACF
57.     win.graph()
58.     par(mfcol=c(2,1))
59.     plot(acf(serie, 36, plot=F), main="")
60.     title(main="Funcion de autocorrelacion simple")
61.     plot(pacf(serie, 36, plot=F), main="")
62.     title(main="Funcion de autocorrelacion parcial")
63.     print("introduzca el orden de diferenciacion regular
d")
64.     d = scan(file="", what=integer(),1)
65.     print("introduzca el orden de diferenciacion
estacional D")
66.     D = scan(file="", what=integer(),1)
67.     if (d == 0){
68.         serieD=serie
69.     }else{
70.         serieD = diff(serie, lag=1, differences=d)
71.     }
72.     if (D == 0){
73.         serieDS = serieD
74.     }else{
75.         serieDS = diff(serieD,lag=12, differences=D)
76.     }
77.     #graficando serie con d diferencias regulares y D diferencias
estacionales
78.     win.graph()
79.     ts.plot(serieDS)
80.     win.graph()
81.     par(mfcol=c(2,1))
82.     plot(acf(serieDS, 36, plot=F), main="")
83.     title(main=paste("Funcion de autocorrelacion simple
d=", d," D=", D, sep=""))
84.     plot(pacf(serieDS, 36, plot=F), main="")
85.     title(main=paste("Funcion de autocorrelacion parcial
d=", d," D=", D, sep=""))
86.     print("introduzca los ordenes p y q de la parte
regular de la serie")
87.     orden_regular = scan(file="", what=integer(), 2)
88.     print("introduzca los ordenes P y Q de la parte
estacional de la serie")
89.     orden_estacional = scan(file="", what=integer(), 2)
90.
91.     #creando modelo SARIMA multiplicativo

```

```

92.      modelo1 = arima(serie, order=c(orden_regular[1],d,orden_regular[2]),seasonal=list(order=c(orden_estacional[1],D,orden_estacional[2])))
93.      print(modelo1)
94.      s_a = sqrt(modelo1[[2]])
95.      residuos = residuals(modelo1)[(12*D+d+1):(length(serie))]
96.      residuos = ts(residuos, frequency=12, start=c(aqo+D,mes+d))
97.      win.graph()
98.      par(mfcol=c(2,1))
99.      plot(residuos, type="l", lty="dotted", main="", ylab="", xlab="")
100.     abline(h=0, col=gray(.8))
101.     abline(h=s_a, lty="dotted", col="blue")
102.     abline(h=-s_a, lty="dotted", col="blue")
103.     points(residuos, bg="black", pch=21)
104.     title(main="Residuos del modelo", ylab="kW-h", xlab="")
105.     hist(residuos, freq=F, main="histograma de los residuos")
106.     abline(v=0, col=gray(.8))
107.     abline(v=s_a, lty="dotted", col="blue")
108.     abline(v=-s_a, lty="dotted", col="blue")
109.     lines(density(residuos), col="blue", lwd=2)
110.     #crea lista con distribucion normal para hacer una curva
111.     x = rnorm(200,mean=0, sd=s_a)
112.     curve(dnorm(x, sd=s_a),col="red", lwd=2, add=TRUE)
113.     #main xlab ylab para etiquetar graficas dev.off(3)
114.     #predict() o forecast() para pronosticos tsdiag para diagnostico grafico
115.     #plot(density(residuals(modelo1)[(12*D+d+1):(length(serie))]))
116.     print("introduzca la cantidad de meses que desea pronosticar con el modelo")
117.     L = scan(file="", what=integer(), 1)
118.     A=predict(modelo1, n.ahead=L)
119.     limite.superior = A[[1]]+1.96*A[[2]]
120.     limite.inferior = A[[1]]-1.96*A[[2]]
121.     #si la serie fue transformada
122.     if (desicion==TRUE){
123.     #retorna los valores del pronostico y de la serie a kW-h
124.     A[[1]]=exp(A[[1]])
125.     serie=exp(serie)
126.     #tambien los limites del 5%
127.     limite.superior = exp(limite.superior)
128.     limite.inferior = exp(limite.inferior)
129.     }
130.     win.graph()
131.     plot(serie, xlim=c(aqo,(aqo+(length(serie)+L)/12)), col="black")
132.     #agregar area sombreada de limites del 95%
133.     polygon(x=c(seq(aqo+length(serie)/12,aqo+length(serie)/12+(L-1)/12, length.out=L),seq(aqo+length(serie)/12+(L-1)/12,

```

```

aqo+length(serie)/12, length.out=L)),
y=c(limite.superior[1:length(A[[1])],
limite.inferior[length(A[[1]):1], col=gray(.9), lty="blank")
134.     lines(B, col="black", lty="solid", lwd=2)
135.     #points(B, bg="limegreen", pch=21)
136.     lines(A[[1]], col="red", lty="solid", lwd=2)
137.     #creando archivo de salida tipo csv
138.     derp=data.frame(A[1])
139.     pred=ts(derp, frequency=12, start=c(2016,12))
140.     MAPE=sum(abs(B-pred)/B)/12
141.     print(paste("El error porcentual absoluto medio
(MAPE) es ",MAPE*100, "%"))
142.     write.table(cbind(residuos,serie,B,pred),paste(getwd(
),"/",Nombre_Bloque," resultado
",paste(orden_regular[1],orden_regular[2],sep=","), "x",paste(orden_
estacional[1],orden_estacional[2],sep=","), ".csv", sep=""), row.name
s=FALSE, sep=";")
143.

```

EJEMPLO PARA USO DEL PROGRAMA

Paso 1.

Colocamos el archivo script.r en el directorio de trabajo de R, por defecto es la carpeta documentos

Paso 2.

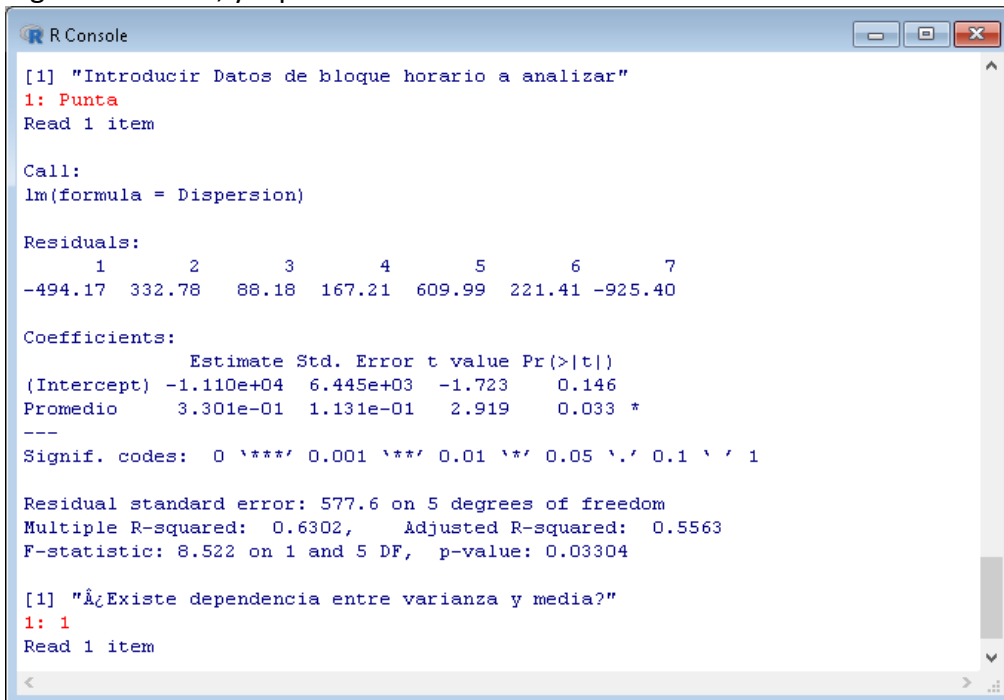
Se copian de la misma forma los archivos csv de consumo en cada bloque horario.

Paso 3.

Desde la consola de R digitamos:

```
source(script.r)
```

Ahora digitamos el nombre del archivo csv que contiene la serie de datos, para este caso digitamos Punta, ya que el archivo se llama Punta.csv.



```
R Console
[1] "Introducir Datos de bloque horario a analizar"
1: Punta
Read 1 item

Call:
lm(formula = Dispersion)

Residuals:
    1     2     3     4     5     6     7
-494.17  332.78  88.18  167.21  609.99  221.41 -925.40

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) -1.110e+04  6.445e+03  -1.723   0.146
Promedio     3.301e-01  1.131e-01   2.919   0.033 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

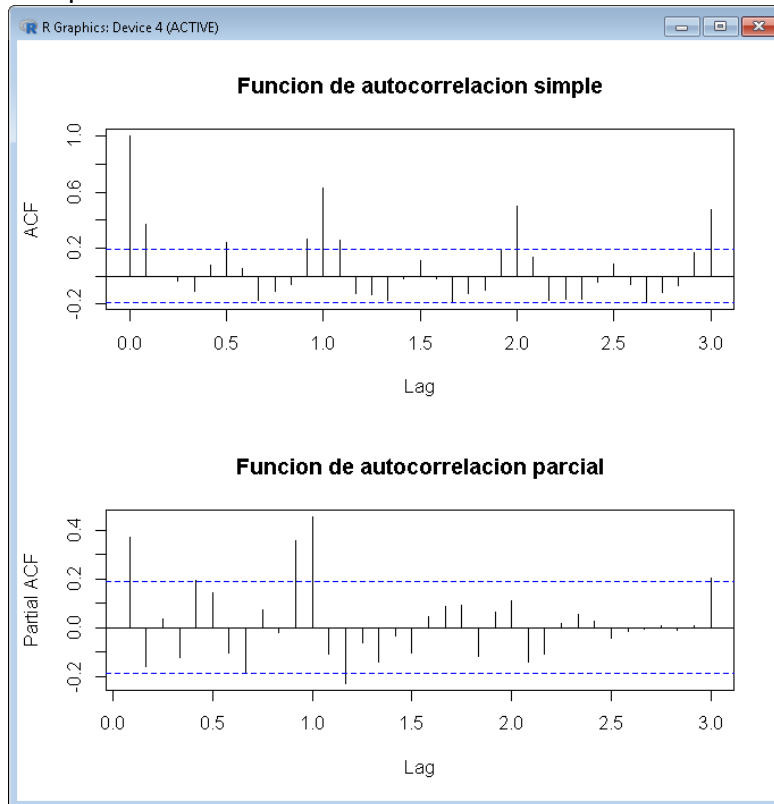
Residual standard error: 577.6 on 5 degrees of freedom
Multiple R-squared:  0.6302,    Adjusted R-squared:  0.5563
F-statistic: 8.522 on 1 and 5 DF,  p-value: 0.03304

[1] "¿Existe dependencia entre varianza y media?"
1: 1
Read 1 item
```

Ahora nos preguntará si existe dependencia entre media y varianza, esto se comprueba si el valor $Pr(>|t|)$ para la variable promedio es mayor a 0.05, si no lo es, quiere decir que la hipótesis Nula "Ho: No existe relación entre media y varianza de la serie" se rechaza. Y entonces se concluye que hay que transformar la serie para lograr estacionariedad.

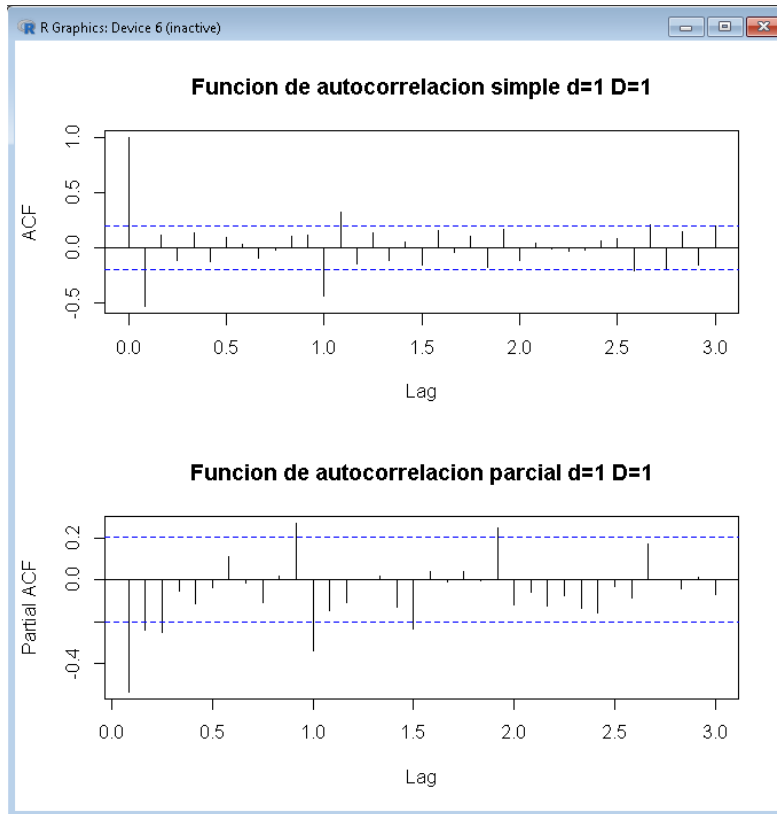
Paso 4.

Ahora nos muestra las gráficas de autocorrelación simple y de autocorrelación parcial de la serie resultante del paso 3.



Observamos que se comportan como senoidales y que existe un decrecimiento lineal de periodo 1 o sea 12 meses cada periodo.

Ahora tomamos la decisión de las diferenciaciones necesarias para lograr estacionariedad, para ello consideramos tomar una diferencia regular "d" de orden 1, y una diferencia estacional "D" de orden 1, y volvemos a observar cómo se comportan las gráficas de autocorrelación simple y parcial de la serie diferenciada.

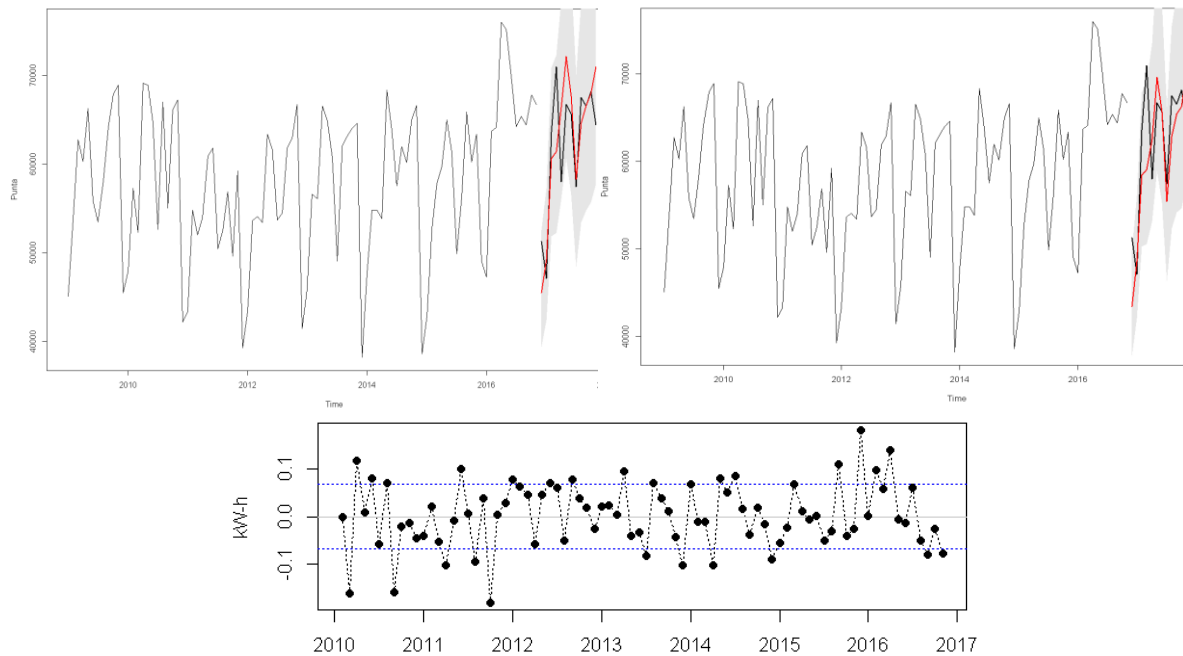


Ahora comenzamos a probar con distintos modelos que puedan representar la serie, observamos la gráfica de autocorrelación simple, se observa que sobresale un término regular y un término estacional seguidos de términos menos significativos. Por otra parte, en la gráfica de autocorrelación parcial se observa un notable decaimiento exponencial en ambas partes, regular y estacional. Por lo tanto la parte de media móvil podría tratarse de un proceso $ARMA(0,1) \times (0,1)$.

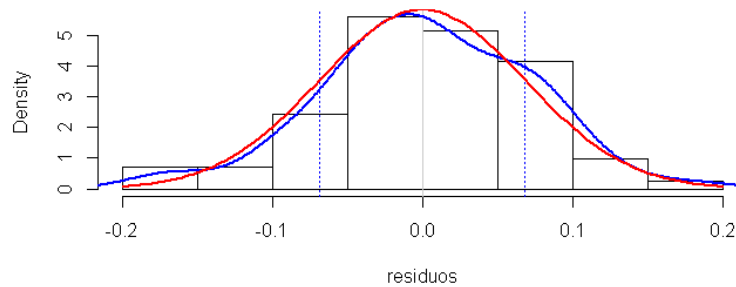
También se considerará el modelo $ARMA(0,1) \times (1,1)$ dada la posibilidad de que el segundo término estacional en realidad no sea significativo.

Paso 5.

Ahora se prueban ambos modelos haciendo doce meses de pronósticos del año 2017 y tomando como referencia el error porcentual promedio, y eligiendo el que menor error produce.



histograma de los residuos



A través de la gráfica se retiran estos atípicos del archivo csv por medio del archivo que se crea llamado "Punta resultado 0,1x0,1", el cual contiene la serie introducida, los residuos del modelo, y los pronósticos, desde una hoja de cálculo, ahí se toman los valores de los residuos y se les restan a los valores correspondientes de los meses atípicos seleccionados, tener en cuenta si la serie fue transformada al inicio, en ese caso se toman los exponentes de los residuos.

Paso 7.

Se reemplazan los atípicos en el archivo csv de entrada y se repite nuevamente el proceso para obtener el mismo modelo a partir de la serie de datos sin atípicos.

Ahora se procede a guardar el objeto arima creado para futuros pronósticos.

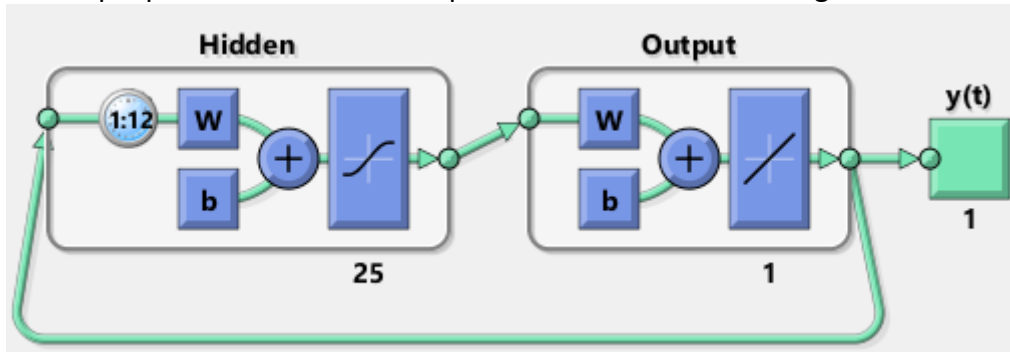
Este modelo puede ser válido para pronósticos de hasta 5 años, debido al error creciente inherente del proceso de la serie. Además, se deberá verificar su desempeño continuamente para asegurar su validez.

UNA RED NEURONAL AUTORREGRESIVA NO LINEAL PARA REALIZAR PRONOSTICOS DE UNA SERIE DE TIEMPO

Ahora se mostrará la forma de realizar pronósticos de una serie de tiempo solamente entrenando una red neuronal autorregresiva.

Una red neuronal autorregresiva no lineal de lazo cerrado, consiste en dicha red ya entrenada, se utiliza para predecir una serie temporal a partir de valores pasados y realimentados de dicha serie.

Su forma simple para 25 neuronas en capa oculta se muestra en la figura XXXX



Usando los datos disponibles de consumo energético del campus central, desde enero de 2009 hasta noviembre de 2017

Se utilizarán los datos de la serie para entrenamiento hasta noviembre de 2016, dejando 12 datos para confrontarlos con sus respectivos pronósticos.

El código para la creación de la red en cada uno de los casos es prácticamente el mismo, se le asignan un retardo de realimentación de 24 el cual corresponde a un periodo estacional de la serie. a la misma vez, evaluaremos el rendimiento del entrenamiento para 25, 50, y 100 neuronas en la capa oculta, compararemos los desempeños de cada red creada usando el indicador MAPE. Esto se repite para cada uno de los bloques horarios y se resume en la tabla XXXXX.

Punta		Resto		Valle	
Neuronas	MAPE %	Neuronas	MAPE %	Neuronas	MAPE %
10	11.9473	10	19.9929	10	15.0840
25	12.0583	25	21.7868	25	8.5916
50	12.8266	50	22.3469	50	13.2505
100	13.0467	100	23.0067	100	12.6823

Ahora para predecir la salida 12 pasos hacia adelante, simulamos la red en lazo cerrado se muestran las gráficas de los pronósticos para hacer un contraste con las mediciones reales. Se presentan únicamente las gráficas con las que el rendimiento fue mejor.