

ESTUDIO DEL PRONÓSTICO DE LA DEMANDA DE ENERGIA ELECTRICA, UTILIZANDO MODELOS DE SERIES DE TIEMPO

RESUMEN

El presente artículo describe un pronóstico de demanda de energía eléctrica, utilizando la metodología ARIMA (Autorregresive-Integrated-Moving Average), y el paquete estadístico SPSS (Statistical package for the social sciences) los datos históricos han sido suministrados por la Empresa de Energía de Pereira y van desde el primero de enero de 2001, hasta el 31 de diciembre de 2001.

PALABRAS CLAVES: Pronóstico, Modelos ARIMA, autocorrelación.

ABSTRACT

The present article describes a forecast of demand of electrical energy, using methodology ARIMA (Autorregresive-Integrated-Moving Average) and the statistical package SPSS (Statistical package the social sciences) the historical data they have been provided by the Company of Energy of Pereira and go from first of January of 2001, to the 31 of December of 2001.

KEYWORDS: Forecast, ARIMA models, autocorrelation.

JOAQUÍN MURILLO S

Profesor Facultad de Ingeniería Eléctrica
Universidad Tecnológica de Pereira
jmurillo@utp.edu.co

ALVARO TREJOS

Profesor
Facultad de Ingeniería Industrial
Universidad Tecnológica de Pereira
alvarot@utp.edu.co

PATRICIA CARVAJAL OLAYA

Tesista
Maestría en Investigación Operativa y Estadística
Universidad Tecnológica de Pereira
pacarva@utp.edu.co

1. INTRODUCCIÓN

En el actual marco de liberalización de los mercados de la energía eléctrica que se presenta en Colombia, donde las grandes industrias y superficies comerciales pueden suscribir contratos de compra de energía eléctrica, las empresas proveedoras de este servicio (comercializadoras o empresas de distribución locales), deben cumplir con unos criterios mínimos que la Comisión Reguladora de Energía y Gas, en asocio con el Centro Nacional de Despacho, dependiente de ISA, prevén para que estas empresas dentro de su estructura de operación, elaboren de manera óptima una predicción confiable de la carga, porque de no hacerlo y fallar el sistema, deben asumir las multas correspondientes por incumplimiento, y sobre todo las consecuencias que traería para el funcionamiento en condiciones óptimas del sistema de distribución local.

El presente trabajo intenta modelar los consumos de energía eléctrica aplicando los modelos ARIMA.

2. SERIES DE TIEMPO

Básicamente una serie de tiempo se le denomina a cualquier variable que conste de datos reunidos, registrados u observados sobre incrementos sucesivos de tiempo. Por lo tanto, se concluye que es una secuencia ordenada de observaciones sobre una variable en particular.

Una serie estacionaria es aquella cuyos momentos al origen y a la media no varían a través del tiempo. Estas situaciones se presentan cuando los patrones de demanda que influyen sobre la serie son relativamente estables.

Fecha de Recibo: 21 Agosto de 2003

Fecha de Aceptación: 25 Noviembre de 2003

3. COMPONENTES DE UNA SERIE DE TIEMPO

La descomposición clásica es un método que se basa en el supuesto que la serie de datos se pueden desagregar en componentes como: tendencia, ciclo, estacionalidad e irregularidad; que se describen a continuación:

3.1. Tendencia

Una serie de tiempo con tendencia es aquella que contiene un componente de largo plazo que representa el crecimiento o declinación de la serie a través de un período amplio.

3.2. Estacional

Se define como estacional una serie de tiempo con un patrón de cambio a si mismo año tras año. Por lo regular, el desarrollo de una técnica de pronóstico estacional comprende la selección de un método multiplicativo o uno de adición y estimar después índices estacionales a partir de la historia de la serie.

3.3. Ciclo

El efecto cíclico se define como la fluctuación en forma de onda alrededor de la tendencia. Los patrones cíclicos tienden a repetirse en los datos cada dos, tres o más años. Es difícil establecer un modelo para estos patrones cíclicos, ya que no son estables.

3.4. Irregular

El componente irregular de la serie de tiempo es el factor residual, es decir, "todo lo que sobra" y toma en consideración las desviaciones de los valores reales de la serie de tiempo en comparación con los esperados; es el elemento aleatorio.

4. PROCESO ESTOCÁSTICO DISCRETO

4.1. Proceso Estocástico (Univariante)

Secuencia de variables aleatorias ordenadas y equidistantes cronológicamente referidas a una característica de una única unidad observable en diferentes momentos o fechas. Representaciones: $(Z_t)_{t=0, \pm 1, \pm 2, \dots}$; (Z_t) donde Z_t es una variable aleatoria referida a la característica de la unidad observable considerada en el momento t .

4.2. Muestra

Una muestra de tamaño o longitud N (finito) procedente de un proceso estocástico (Z_t) es un subconjunto de N componentes consecutivos de (Z_t) . Representaciones de una muestra:

$$Z_1, Z_2, \dots, Z_N; Z \equiv (Z_1, Z_2, \dots, Z_N); (Z_t)_{t=1}^N$$

Proceso estocástico : $\dots, Z_{-1}, Z_0, Z_1, Z_2, \dots, Z_N; Z_{N+1}, \dots$

Muestra : Z_1, Z_2, \dots, Z_N

Serie temporal : z_1, z_2, \dots, z_N

Se elaborará un modelo estadístico para una muestra Z procedente de un proceso estocástico a partir de una única realización particular de Z (una serie temporal z). Bajo ciertas condiciones, un modelo para Z puede describir la evolución temporal de (Z_t) a lo largo de toda su historia reciente y no sólo a lo largo del intervalo muestral.

Se utilizará el modelo anterior para describir en términos probabilísticos algunos componentes futuros Z_{N+1} ($l \geq 1$) del proceso (Z_t) , es decir, prever Z_{N+1} Z_{N+2} . Se seleccionará un modelo dentro de una clase general de modelos de reconocida utilidad práctica (ARIMA), que implique ciertas propiedades teóricas compatibles con las propiedades observadas (muestrales) en z .

Para poder inferir la estructura probabilística de Z hay que suponer ciertas hipótesis sobre la misma, de manera que lo que no se suponga se pueda inferir a partir de una única realización particular (una única serie temporal z) de Z .

Se supone hipótesis de estacionariedad en la serie. Algunas propiedades teóricas de un proceso estocástico implicadas por esta hipótesis son las siguientes:

- **Media:** valor constante en el tiempo alrededor del cual evoluciona de manera aleatoria (errática) un proceso estocástico estacionario.

- **Varianza:** valor constante en el tiempo que mide la dispersión (variabilidad) de la evolución de un proceso estocástico estacionario alrededor de su media.

- **Autocorrelación:** valor constante en el tiempo que mide la relación existente entre cada par de componentes de un proceso separados por un intervalo temporal (retardo) dado.

Pero muchas series temporales no pueden considerarse generadas por un proceso estacionario (son series no estacionarias), ya que suelen presentar ciertas tendencias marcadas (no presentan afinidad hacia algún valor constante en el tiempo) y una dispersión creciente.

Para lograr la estacionariedad se realizan una serie de transformaciones en búsqueda de estabilizar el nivel y la dispersión de la serie.

El primer paso en la identificación del modelo consiste en determinar si la serie es o no estacionaria. En este último caso, se intenta convertirla en estacionaria a través del método de diferenciación.

Para verificar que componentes tiene la serie, necesariamente tiene que observarse el gráfico de dispersión de la serie original, y los correlogramas simples.

Además es conveniente observar el gráfico de la serie que resulta luego de diferenciar la serie tanto regular como estacionalmente.

A continuación se presenta la gráfica de dispersión de observaciones de la serie original en forma seccionada por cuatrimestres, ya que el programa SPSS presenta una limitante al mostrar los gráficos cuyos datos superen los 3000.

Tomemos por ejemplo el gráfico de dispersión del primer cuatrimestre del año 2001

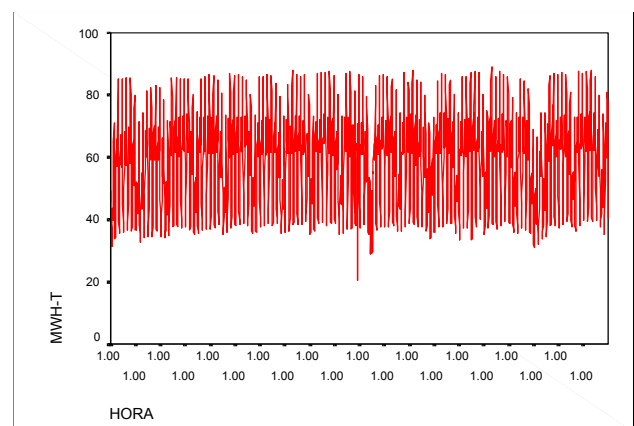


Figura 1. gráfica de dispersión del consumo de energía

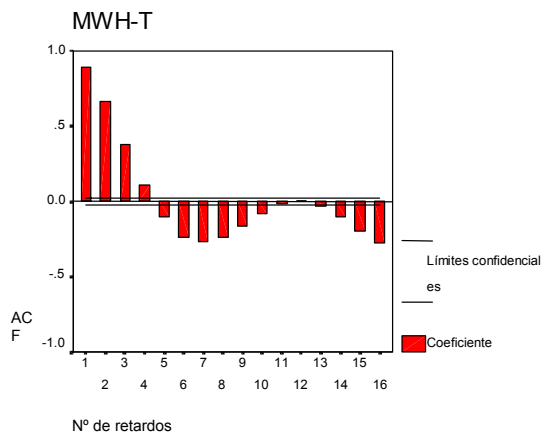


Figura 2. Gráfico de las autocorrelaciones simples

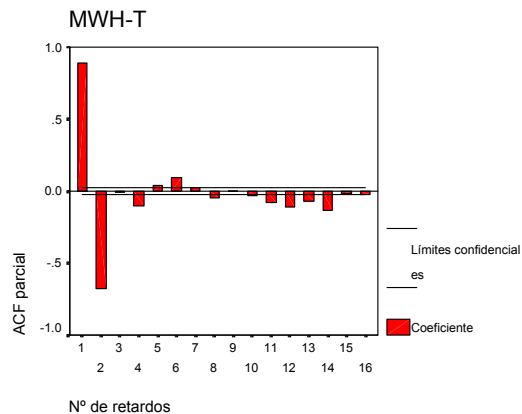


Figura 3. Gráfico de las autocorrelaciones parciales

La grafica de dispersión de la serie observada, aunque sea un 1/3 de periodo anual, va a ser exactamente igual en su configuración a los otros tercios de gráficas puesto que pertenecen a una misma serie de observaciones que es el conjunto del año 2001.

Por lo tanto se analizará el resultado de sus autocorrelaciones.

5. IDENTIFICACIÓN DE UN MODELO ARIMA

En la identificación de un modelo tentativo, observamos las autocorrelaciones simples y notamos que los retrasos caen hacia cero después del cuarto coeficiente de autocorrelación e inmediatamente, y después del quinto coeficiente es significativamente diferente de cero hasta el onceavo retraso, donde es cero hasta el treceavo, luego es significativamente diferente de cero y así sucesivamente en forma sinusoidal pero negativamente.

Según el anterior comportamiento basado en el análisis gráfico de las autocorrelaciones simples y parciales ACF y PACF, el estadístico Ljung-Box, entre otras la

teoría BOX-JENKINS nos dice que la serie no presenta un patrón de estacionariedad.

Por lo tanto, es necesario transformar la serie buscando estacionariedad, diferenciándola para lograr un modelo que se puede adaptar a nuestro objetivo. No sabemos qué tan exitoso pueda ser este modelo, esto se logra comparándolo con otros.

Se propone entonces construir un modelo para cada hora del día, todos los días del año. Es decir, se construirían 24 modelos, cada uno de estos lógicamente con 365 observaciones suficientes para realizar un pronóstico confiable.

En la siguiente sección se evaluará la *estacionalidad* de las series mediante las pruebas analíticas y gráficas.

Es decir, se mostrarán los modelos horarios de demanda de energía teniendo en cuenta que se construyó uno para cada hora del día, por se muestran 24 modelos diferentes con las respectivas pruebas que corroboran su idoneidad. Estas pruebas son seis, de las cuales tres son de tipo analítico y las otras tres de tipo gráfico. Las analíticas son: la prueba de iniciación del modelo, es decir, cuando desde el programa asignamos los parámetros numéricos al modelo ARIMA, determinamos la estacionalidad cíclica o regular y se involucra o no una constante. La segunda prueba es utilizar un estadístico descriptivo entre la observación de la serie original y el pronóstico que se genera con el modelo; este descriptivo hace una comparación en función de los valores mínimo y máximo de las series, la media de éstas y su desviación típica. El tercer modelo analítico es una prueba de normalidad del error de la observación de la serie original, denominada prueba de Kolmogorov-Smirnov de una muestra. Las tres pruebas de tipo gráfico es un multigráfico donde estará representada la observación de la serie original, el pronóstico arrojado por el modelo y los errores tanto por arriba como por abajo. Después se encontrará un histograma donde podemos ver la normalidad gráfica del error de la observación de la serie original y por último la autocorrelación simple del error de la observación de la serie original.

Se mostrarán entonces primero las pruebas de carácter analítico y posteriormente las pruebas de tipo gráfico.

Antes de esto se hará un análisis a las gráficas de autocorrelaciones a cualquier modelo horario, ya que éstas presentan una gran similitud en su comportamiento.

Análisis de la hora uno: la gráfica de autocorrelación muestra un patrón de no estacionariedad dado que sus rezagos están dentro de los límites y prácticamente son casi cero, por tanto, se probarán varios modelos tomando en cuenta las anteriores recomendaciones, para cada hora se probaron 6 modelos diferentes y luego se confrontaron entre sí para escoger el que mejor se adapte a nuestras

perspectivas de pronóstico a corto plazo. En un cuadro se mostrarán 5 modelos con sus respectivas pruebas teóricas, seguidamente aparece el análisis analítico y gráfico al modelo que cumplió todas las pruebas y objetivos de este proyecto. ARIMA (1,0,0)*(2,1,0)

Se tomará entonces un hora aleatoria del día y se realizará un análisis completo.

NO ESTACIONAL			ESTACIONAL			CONSTANTE	SCR	PARÁMETROS SIGNIFICATIVOS	CORRELACIÓN DE PARÁMETROS	AKAIKE
p	d	q	P	D	Q					
1	0	0	1	1	0	SI	11535	OK	OK	2265
2	0	0	2	1	0	NO	11061	SAR2 no es cero	OK	2252
1	1	0	2	1	0	SI	15039	SAR2, CTE; no son cero	OK	2357
2	0	0	1	1	0	SI	11464	AR2, CTE; no son cero	OK	2264
2	1	0	1	1	0	SI	14073	CTE no es cero	OK	2332

Tabla 1. Análisis para la hora 8 del día

NO ESTACIONAL			ESTACIONAL			SCHWARZ	RESIDUOS ALEATORIOS	NORMALIDAD	PRONÓSTICOS	# ITERACIONES
p	d	q	P	D	Q					
1	0	0	1	1	0	2277	NO	NO	OK	3
2	0	0	2	1	0	2267	NO	NO	OK	3
1	1	0	2	1	0	2372	NO	NO	OK	3
2	0	0	1	1	0	2280	NO	NO	OK	3
2	1	0	1	1	0	2348	NO	NO	OK	3

Tabla 1. Continuación

El cuadro anterior están representadas las pruebas teóricas, que permitirán saber si el modelo descrito allí es el óptimo para pronosticar.

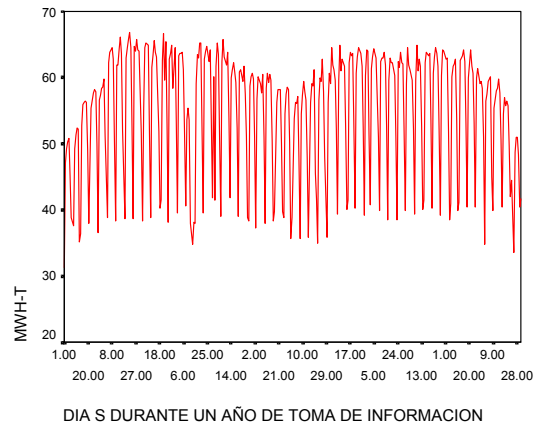


Figura 4. Gráfica de consumo durante un año

Estadísticos descriptivos

	N	Mínimo	Máximo	Media	Desv. típ
MWH-T	365	32.62	68.12	58.9294	7.6878
Fit for MWH-T from ARIMA, MOD_1	358	2.61963	8.83677	9.03617	459614
N válido (según lista)	358				

Tabla 2. Estadísticos Descriptivos

Prueba de Kolmogorov-Smirnov para una muestra

		Error for MWH-T from ARIMA, MOD_1 NOCON
N		358
Parámetros normales ^{a,b}	Media	6.033E-02
	Desviación típica	4.9733887
Diferencias más extremas	Absoluta	.172
	Positiva	.132
	Negativa	-.172
Z de Kolmogorov-Smirnov		3.245
Sig. asintót. (bilateral)		.000

- a. La distribución de contraste es la Normal.
- b. Se han calculado a partir de los datos.

Tabla 3. Prueba de Kolmogorov Smirnov

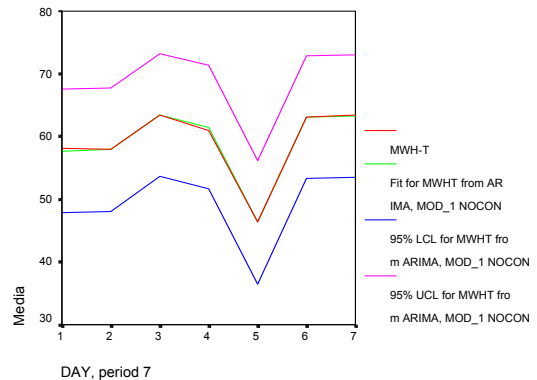


Figura 5. Intervalos de confianza para los pronósticos

Para la gráfica anterior, se puede observar como es la curva de observación real (datos reales e históricos), contra la curva de pronóstico (resultado del modelo buscado y del programa SPSS). Las curvas restantes, dan indicio del comportamiento del pronóstico, tanto por encima como por debajo.

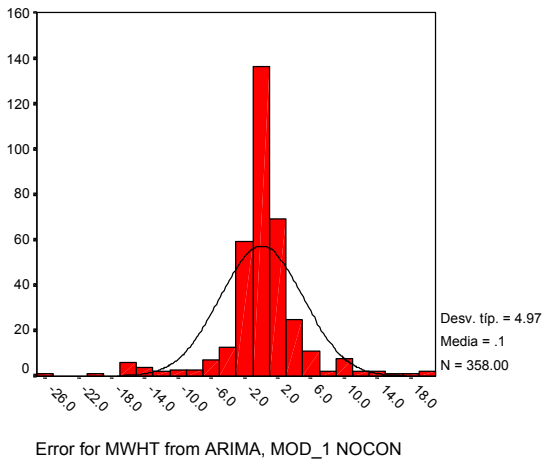


Figura 6. Distribución de los errores

La gráfica interior indica la normalidad de los errores, es visible como es la concentración de datos, circunscrita a la curva gaussiana, lo que indica una gran “afinidad” de los datos, sin importar la magnitud de estos.

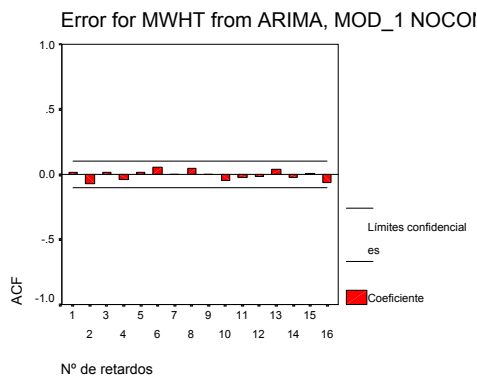


Figura 7. Gráfico de autocorrelaciones de los errores

El gráfico anterior nos muestra la Función de Autocorrelación (ACF en inglés), indica el error del pronóstico, se encuentra dentro del límite de ajuste, es decir, la desviación de datos no existe, aunque para otras horas algún valor pueda salir, pero ese depende de la magnitud de los datos.

Después de disponer de un modelo ARIMA: (1,0,0)*(2,1,0) individual para cada una de las series

temporales, y en función del modelo seleccionado se diseña una ecuación siguiendo la metodología Box-Jenkins.

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-7} + \phi_3 y_{t-14}$$

- y_t = Es la serie temporal de indicador que se modeliza.
- ϕ_1 = Coeficiente autorregresivo de la parte no estacional
- ϕ_2, ϕ_3 = Corresponden a los coeficientes autorregresivos de la parte.
- y_{t-7} = Corresponde a la observación (estacional) de la hora t siete días atrás.
- y_{t-14} = Corresponde a la observación (estacional) de la hora t catorce días atrás.

Pronóstico hora 8:

Día Domingo:

$$\begin{aligned} \phi_1 &= 0.31818 \\ \phi_2 &= -0.31687 \\ \phi_3 &= -0.10414 \\ \tilde{N} &= 58.9294 \end{aligned}$$

$$Y_t = 0.31818(54.03) - 0.31687(47.08) - 0.10414(46.26) + 58.9294 = 56.30 \rightarrow \text{Resultado matemático.}$$

$$Y_t = 55.43 \rightarrow \text{Resultado SPSS}$$

Pronóstico hora 9:

Día sábado:

$$\begin{aligned} \phi_1 &= 0.24032 \\ \phi_2 &= -0.30469 \\ \phi_3 &= -0.10427 \\ \tilde{N} &= 60.3587 \end{aligned}$$

$$Y_t = 0.24032(61.50711) - 0.30469(63.62) - 0.10427(63.42) + 60.3587 = 62.4370 \rightarrow \text{Resultado matemático.}$$

$$Y_t = 51.84549 \rightarrow \text{Resultado SPSS}$$

HORA 8		
DAY	DATE	FIT
1	SUN	55.45000
2	MON	43.60810
3	TUE	58.06043
4	WED	59.85782
5	THU	60.01899
6	FRI	55.48592
7	SAT	46.05784

6. CONCLUSIONES

Es posible modelar los consumos de energía eléctrica en los municipios de Colombia utilizando la metodología propuesta por Box y Jenkins, dado que el comportamiento de estas series no es completamente aleatorio y se pueden describir como series de tiempo con una alta probabilidad de éxito en la modelación a través de dicha metodología.

Para facilitar la búsqueda de un buen modelo que represente la serie consumo de energía en el Municipio de Pereira durante el año 2001, se toma la decisión de dividir los datos en 24 series correspondientes a los promedios de consumo por cada hora del día; con ello se aíslan las perturbaciones causadas por el mes, la semana y la hora.

Como ilustración del uso de la metodología propuesta por Box y Jenkins se modela la serie correspondiente al consumo de energía eléctrica a las 8 de la mañana. Los datos originales no muestran estacionariedad, por tanto se aplica una diferenciación estacional y con ello se suaviza la serie logrando estacionariedad de tal forma que así se pueda ajustar un modelo ARIMA a los datos en estudio.

El modelo ARIMA que mejor ajusta describe el consumo de energía eléctrica a las 8 a.m., luego de analizar las autocorrelaciones simples y parciales es ARIMA(1,0,0)(2,1,0)

El modelo ARIMA(1,0,0)(2,1,0) matemáticamente se expresa de la siguiente manera:

$y_t = \phi_1 y_t - 1\phi_2 y_{t-7} + \phi_3 y_{t-14}$ Los errores producidos por este modelo cumplen con los supuestos de normalidad y aleatoriedad, lo que permite confiar en la bondad de ajuste del mismo.

7. BIBLIOGRAFÍA

- [1] ABRIL, J.C. Análisis estadístico de series de tiempo basado en modelos de espacio de Estado. E.U.D.E.B.A., 1999.
- [2] AJUB, Alberto. Ecuaciones en diferencias finitas. Editorial del Colognia, 1985.
- [2] CARVAJAL, Patricia; TREJOS, Alvaro. Trabajo de tesis en series de tiempo. UNIVALLE, 1991.
- [3] CHAO, Lincoln L. Estadística para Ciencias Sociales y Administrativas. Bogotá: McGraw-Hill.
- [4] CHATFIELD, Cliristopher. The analysis of time series. An introduction. Fourth edition. Cliprnan and Hall, London New York.
- [5] DASH, P.K.; SATPANTHY, H.P.; LIEW, A.C.; RAHMAN, S.A. Real-time short-term load forecasting system using functional link network,k IEEE. Transactions on Power Systems. Vol. 12, Nro. 2. May 1997.