

COMPARACIÓN DE TÉCNICAS ESTADÍSTICAS DE PRONÓSTICO PARA LA DEMANDA DE ENERGÍA ELÉCTRICA

COMPARISON OF STATISTICAL FORECASTING TECHNIQUES FOR ELECTRICAL ENERGY DEMAND

José Fernando Tabares Muñoz¹, Carlos Andrés Velásquez Galvis¹, Marisol Valencia Cárdenas^{2,*}

RESUMEN

En este artículo se muestra una comparación de tres métodos para realizar pronóstico, aplicados sobre la demanda de energía eléctrica diaria de Colombia. La caracterización del comportamiento y pronóstico de la demanda energética diaria tiene gran importancia para las entidades reguladoras de la energía en Colombia. Cada vez son más las técnicas exploradas con el fin de mejorar los pronósticos de consumo de energía del país; por ello, en este trabajo se muestra una comparación de tres técnicas estadísticas aplicadas a su pronóstico, determinando las ventajas y desventajas de cada una, a partir de sus resultados y cuál es la más eficiente acorde con el indicador de error absoluto medio porcentual (MAPE). Entre los resultados se encuentran eficientes los modelos: econométrico y bayesiano utilizado, pero no muy acertado el modelo estimado SARIMA.

Palabras Clave: Demanda de Energía, series de tiempo, regresión Gaussiana bayesiana, Colombia, método MCMC.

ABSTRACT

This paper shows a comparison of three methods to do forecasts, applied to electric energy daily demand of Colombia. The characterization of the behavior and forecast of daily energy demand of the country is very important for energy regulators in Colombia. Increasingly explored techniques are created in order to improve forecasts of energy consumption in the country, so this work is a comparison of three statistical techniques applied to forecast it, determining the advantages and disadvantages of each of these from their results and the most efficient according to the indicator: mean absolute percentage error (MAPE). SARIMA techniques are found to be inadequate, whereas econometric and Bayesian method are efficient.

Keywords: Energy demand, time series, bayesian Gaussian regression, Colombia, MCMC method.

¹Ingeniero Industrial, Universidad Pontificia Bolivariana. Medellín. Colombia.

²MSc en Estadística. PhD(C) en Ingeniería-Industria y organizaciones. Universidad Nacional de Colombia. Medellín. Colombia.

*Autor para correspondencia: mvalencia@unal.edu.co

INTRODUCCIÓN

Existen diversas investigaciones sobre el comportamiento habitual de la energía en Colombia (Medina & Garcia, 2005; Gómez *et al.*, 2011; Rueda *et al.*, 2011); sin embargo, se han encontrado en algunas de ellas falencias como: la falta de acierto, falta de integración de variables exógenas, la falta de datos para estimar modelos, entre otras.

En Colombia, el Centro Nacional de Despacho (CND), departamento de XM Compañía de Expertos en Mercados S.A.E.S.P, filial de ISA, es el encargado de la operación y administración de todo el Sistema Interconectado Nacional de Colombia (SIN) (Gómez *et al.*, 2011). Es decir, su tarea es planear, supervisar y controlar los recursos de generadores, transmisores, distribuidores y comercializadores de energía. El CND debe realizar un plan de maniobras para las empresas generadora, indicando la cantidad de potencia que deben producir diariamente; por este motivo, el pronóstico de demanda energética es uno de los suministros más importantes de este proceso. Por ello es realmente necesaria una predicción efectiva, garantizando calidad, seguridad y confiabilidad en el servicio de los usuarios. Al respecto, Rueda *et al.* (2010) expresan: “*La predicción de la demanda es un problema de gran importancia para el sector eléctrico, ya que, a partir de sus resultados, los agentes del mercado de energía toman las decisiones más adecuadas para su labor*”. Por su parte, Codensa S.A ESP, empresa distribuidora y comercializadora de energía eléctrica en Colombia, ha elaborado pronósticos basados en regresión lineal, suavización exponencial y media móvil (Gómez *et al.*, 2011), comparándolos por medio del criterio el MAPE.

En la última década se han realizado investigaciones muy importantes con relación al pronóstico de energía en Colombia. Murillos *et al.* (2003) llevaron a cabo un proyecto cuya finalidad era predecir la demanda de energía eléctrica, con modelos ARIMA, basado en datos históricos que van desde el primero de enero hasta el 31 de diciembre de 2001, y el paquete estadístico SPSS.

Los modelos de pronóstico tradicionales han sido ampliamente usados para la toma de decisiones a nivel de producción, finanzas y la energía, como se ha mencionado; tal es el caso de los modelos ARIMA, desarrollados en los años 70 por George Box y Gwilym Jenkins (Box & Jenkins, 1976; Makridakis *et al.*, 2011); estos modelos incorporan características del pasado de la misma serie, acorde con su autocorrelación.

La filial de ISA, XM utiliza los modelos ARIMA estacionales multiplicativos para efectuar pronósticos a corto plazo; por ello se incluirá el estudio de un modelo SARIMA para compararlo con otros dos. En este trabajo se efectuará una caracterización, análisis y comparación de los modelos: modelo SARIMA, modelo econométrico (Caridad y Ocerin, 1998; Montgomery *et al.*, 2012), y una técnica bayesiana: regresión gaussiana con simulación Monte Carlo por Cadenas de Markov (MCMC) (Andrieu *et al.*, 2003; Martin *et al.*, 2013), que permitan luego de la estimación de pruebas, validación y la medición del indicador de error absoluto medio porcentual (MAPE) con datos de ajuste y de pronóstico para cada uno, y con estos dos determinar la mejor para efectuar la predicción de dicha demanda para el país.

Modelo ARIMA estacional multiplicativo

El modelo ARIMA estacional multiplicativo $ARIMA(p,d,q) \times (P,D,Q)_s$, (Box & Jenkins, 1976) se representa habitualmente en la forma:

$$(1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p)(1 - \Phi_1 B^s - \Phi_2 B^{2s} - \dots - \Phi_P B^{Ps}) \nabla^d \nabla_s^D y_t = (1 + \theta_1 B + \theta_2 B^2 + \dots + \theta_q B^q)(1 + \Theta_1 B^s + \Theta_2 B^{2s} + \dots + \Theta_Q B^{Qs}) a_t \quad (1)$$

Siendo B el operador de retardos.

Modelo econométrico

Según Caridad y Ocerin (1998), un modelo econométrico uniecuacional está formado por una ecuación, en la que la variable explicada o endógena depende de una o más variables explicativas (p.3). La estimación se realiza mediante el método de máxima verosimilitud, estudiado, entre otros, para los modelos de regresión comunes Montgomery *et al.* (2002).

La siguiente es una ecuación general del modelo:

$$Y_t = \alpha + \beta_0 X_t + \gamma Y_{t-1} + \varepsilon_t \quad (2)$$

Dónde:

β_0 : es el efecto de la variable explicativa X_t sobre la respuesta Y_t .

γ : es el efecto de la variable explicativa Y_{t-1} (retardo de la endógena) sobre la respuesta Y_t .

Los supuestos básicos de los modelos clásicos de series de tiempo se establecen sobre las perturbaciones ε_t , que son variables aleatorias y que se supone tendrán: esperanza cero, varianza constante, no autocorrelacionadas y con distribución normal (Jarque & Bera, 1987; Ljung, 1978).

La prueba de Ljung Box es la prueba usada para determinar incorrelación en los residuos, que son los valores numéricos obtenidos posterior a la estimación.

Proceso Bayesiano

Las técnicas bayesianas son muy aplicables a procesos de pronósticos, como el utilizado en este trabajo. La estadística bayesiana parte de supuestos que difieren sólo un poco con respecto a modelos clásicos (Barrera & Correa 2008; Gill, 2002; Martin *et al.*, 2011), por ejemplo, consideran parámetros de distribuciones de probabilidad o coeficientes específicos de modelos como variables aleatorias θ , sobre la que se dispone de una información a priori cuantificada en una distribución de probabilidad, cuya función de densidad, $\xi(\theta)$, se supone conocida, y, además, se dispone de información muestral, y_1, y_2, \dots, y_n , tomada en la población y objeto de estudio, información que se resume en su función de verosimilitud $L(y_1, y_2, \dots, y_n | \theta)$. Con el producto de la distribución de probabilidad a priori $\xi(\theta)$ y la función de verosimilitud y utilizando el teorema de Bayes, es posible la estimación de la función a posteriori $\xi(\theta | \text{datos})$, como se detalla a continuación.

La función densidad posterior del parámetro θ : $\xi(\theta | \text{datos})$ es proporcional al producto entre la a priori $\xi(\theta)$ y la verosimilitud $f(\text{datos} | \theta)$.

$$\xi\left(\frac{\theta}{\text{datos}}\right) = \frac{L(\text{datos} | \theta) \xi(\theta)}{f_x(x)} \quad (3)$$

Finalmente, el denominador será siempre un número constante que no dependerá del parámetro; por tanto, el teorema queda como sigue:

$$((\theta / \text{datos}) \propto L(\text{datos} / \theta) \xi(\theta) \quad (4)$$

Lo cual constituye una distribución a posteriori de la variable aleatoria θ , a partir de la cual se realizan inferencias.

En muchas ocasiones esta distribución final tiene formas que a menudo son complejas o imposibles de resolver analíticamente; por ello se recurre a procesos de muestreo de la variable aleatoria θ , así como para la distribución usada para hacer inferencia predictiva bayesiana.

En este trabajo se usará la distribución a priori normal para los parámetros de un modelo de regresión gaussiana bayesiana, y la gamma inversa para las perturbaciones aleatorias.

MonteCarlo por Cadenas de Markov (MCMC)

La simulación de Montecarlo realiza una elección aleatoria de muestras para reducir un problema combinatorio complejo a uno más simple (Andrieu *et al.*, 2003). Dicha simulación genera valores independientes, tomando como base una distribución de probabilidad deseada. Por otro lado, las cadenas de Markov postulan que los estados anteriores son irrelevantes para la predicción de los estados siguientes, con conocimiento del estado actual, pero existe dependencia entre los valores simulados consecutivamente. *Monte Carlo por Cadenas de Markov (MCMC)* consiste en generar un muestreo a partir de distribuciones de probabilidad basadas en la construcción de cadenas de Markov, donde cada valor simulado tiene dependencia con el dato anterior, llegando a una convergencia a la distribución deseada (Gill, 2002). Después de una larga cantidad de corridas, estos resultados constituyen una muestra incorrelacionada de la distribución deseada. En algunas ocasiones, la generación de variables aleatorias bajo una distribución de probabilidad compleja, se hace muy difícil; para ello, se recurre a algoritmos de apoyo basados en MCMC, como el de *Gibbs* o el de *Metropolis Hastings* (Geyer & Thompson, 1995; Gill, 2002), cuya eficacia en la simulación puede apoyarse en la elaboración de pruebas específicas.

Muestreador de Gibbs

El muestreador Gibbs es un algoritmo iterativo de Monte Carlo por Cadenas de Markov, diseñado para extraer muestras de los parámetros bajo la distribución a posteriori de cada uno, a partir de una cadena estocástica. Este algoritmo es flexible y confiable para generar cadenas de valores. Este es un *kernel* de transición (Gill, 2002), creado por una serie de distribuciones condicionales en un esquema Markoviano. Si la distribución límite de interés es $f(b)$, donde b es un vector k dimensional de coeficientes a estimar, entonces el objetivo es producir una cadena de Markov que genere ciclos alrededor de estas condicionales y que converja a dicha distribución.

Regresión gaussiana bayesiana por Cadenas de Markov

El modelo general de regresión gaussiana bayesiana toma la siguiente forma (Martin *et al.* 2011):

$$y_i = x_i' \beta + \varepsilon_i \quad (5)$$

Para realizar el modelo específico para este trabajo, el programa R cuenta con un paquete: MCMCPack, que con la función MCMC regress estima los parámetros, aplicando el muestreador de Gibbs, bajo las siguientes distribuciones:

- Distribución a priori Normal para los betas (coeficientes del modelo de regresión).
- Distribución normal para los datos.
- Distribución a priori Gamma inversa para la varianza del error condicional.
- Distribución Normal para las perturbaciones ε_t

MATERIALES Y MÉTODOS

El proceso partió del análisis exploratorio de la serie de tiempo descrita como un conjunto de datos históricos de la demanda energética diaria en Colombia, desde el 15 de diciembre de 2010 hasta el 31 de diciembre de 2012, facilitados por el departamento encargado de pronósticos de demanda de energía en XM filial de ISA, información de carácter público. Se compararon tres modelos para determinar el mejor: modelo SARIMA, modelo econométrico con variables exógenas, modelo de regresión gaussiana bayesiana.

La figura 1 muestra el comportamiento de la demanda de energía eléctrica diaria, para el transcurso de un mes.

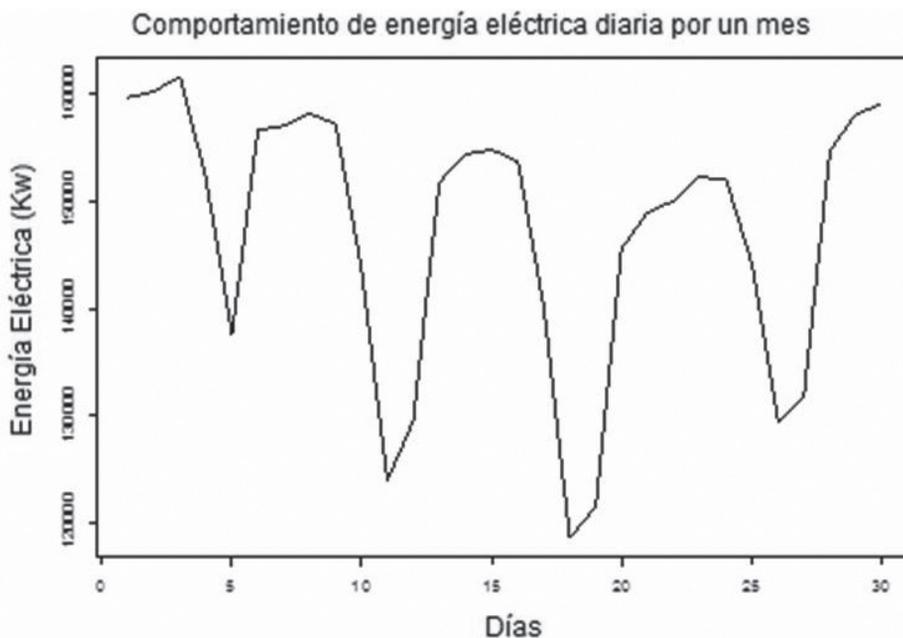


Figura 1. Demanda de energía eléctrica diaria, en un mes

Acorde con la figura 1 puede notarse un comportamiento muy similar cada 7 días, con demanda alta al inicio de la semana y baja al final en el fin de semana; así puede identificarse un comportamiento estacional 7. Además de lo anterior, a partir de entrevistas con el personal experto de la empresa, se comprobó esta hipótesis y se determinaron otras variables que pueden tener impactos importantes como la temperatura y nivel de energía, información que fue útil para estimar el modelo econométrico con variables exógenas y endógenas, así como el modelo SARIMA para efectuar la comparación de los tres modelos.

Lo anterior, además, sirvió de insumo para incorporar algunas variables explicativas importantes en el modelo de regresión gaussiana bayesiana por MCMC.

Modelo econométrico

En el modelo fue considerada la variable endógena, Z_t , el consumo de electricidad, y variables explicativas como: la endógena con retardos Z_{t-k} , la temperatura diaria, y otras exógenas como las variables ficticias para incluir intervenciones en la serie (como: días de fiesta, día de la semana), buscando estimar un modelo de la forma:

$$Z_t = \beta_0 + \beta_1 Z_{t-1} + \dots + at + \alpha_1 \text{sen}(2\pi t/7) + \alpha_2 \text{cos}(2\pi t/7) + \dots + \text{día semana} + \text{variables artificiales} + \varepsilon_t \quad (6)$$

En el proceso se exploraron otros, agregando o eliminando variables si éstas fuesen significativas o no al nivel del 5%.

Modelo de Regresión Gaussiana Bayesiana por Cadenas de Markov

El modelo general de regresión gaussiana bayesiana usado fue:

$$Z_t = \beta_0 + \beta_1 Z_{t-1} + \dots + at + \dots + \text{día semana} + \varepsilon_t \quad (7)$$

Donde:

Z_t es la demanda de energía eléctrica

Z_{t-1} es la variable con retardo de la demanda de energía eléctrica

t es la tendencia temporal

Día de la semana: son variables indicadoras (0/1) acorde con el día, para que se agreguen a la constante β_0 para el día respectivo.

Las estadísticas que fueron usadas para la estimación de cada coeficiente corresponden al valor medio del muestreo MCMC realizado.

Criterio de evaluación final de los modelos

En este análisis de los modelos se usaron pruebas de hipótesis para analizar la significancia de los parámetros de los modelos clásicos, basados en los supuestos de normalidad, pruebas para determinar el cumplimiento de los supuestos en los residuales. Estas no fueron las mismas para los procesos bayesianos, en los que fue necesario determinar si la simulación generó muestras independientes, como se espera. Pero sí se estimó otro indicador como el criterio de elección final del mejor modelo.

Se midió entonces el indicador MAPE (ecuación (8)), de dos tipos: de ajuste, con datos usados para estimación y de pronóstico; con datos no usados para estimar, ya que, debido a que el objetivo final era encontrar un modelo con adecuada capacidad para pronosticar, se consideró un adecuado indicador de capacidad de acierto en los modelos y, por tanto, el criterio final de elección. El MAPE fue calculado de la siguiente manera:

$$MAPE = \frac{1}{N} \left| \frac{\hat{Z}_t - Z_t}{Z_t} \right| \quad (8)$$

Donde:

\hat{Z}_t = Valor pronosticado de energía eléctrica en el periodo t

Z_t = Valor real de energía eléctrica en el periodo t

Siendo N el total de datos final para el ajuste, sean $N-k$ acorde a los retardos o para los pronósticos, los que se dejaron para pronosticar.

RESULTADOS

Se presentan aquí las estimaciones de los modelos estadísticos usados, con la respectiva caracterización del consumo energético diario de Colombia, usando las pruebas de validación y criterios de comparación necesarios para un adecuado análisis y elección del que permite optimizar de manera más integral, el pronóstico al mostrar menor nivel de error absoluto relativo.

Modelo SARIMA

Al analizar los valores de autocorrelación de la serie de consumo energético, se detectó una estacionalidad de orden $s=7$ muy evidente, ya que en los períodos 7 y 14 tomó valores muy altos (0,7639 y 0,7070 respectivamente). Asimismo, la autocorrelación de orden 1 es muy alta (0,4527). Esto indica que los retardos más evidentes que tienen efectos sobre la serie son Z_{t-1} , Z_{t-7} y Z_{t-14} . Además de estas evidencias, se usó el método automático de R para detectar de mejor forma los parámetros indicados para el modelo. Se probará qué tan adecuado puede ser el análisis con este modelo; posteriormente se introducen cambios usando el modelo econométrico.

Luego de estimar varios modelos, se obtuvo que el ARIMA (1,1,1)X(2,0,1) fue el de menor valor de MAPE.

Tabla 1. Coeficientes SARIMA

Coeficientes				
ar1	ma1	sar1	sar2	sma1
0,4950	-0,9321	0,9842	0,0075	-0,8028
0,0416	0,0196	0,0407	0,0402	0,0275

Se estimaron intervalos de confianza al 95%, usando el error estándar de la última línea de la tabla 1. El único coeficiente no significativo es el segundo término estacional. Sin embargo, al eliminarlo, el modelo pierde capacidad de acierto, resultando con mejor MAPE el presentado en la tabla 1.

El modelo SARIMA (tabla 1) no cumplió adecuadamente el supuesto de normalidad de residuales acorde con Jarque Bera, pero sí la de incorrelación de éstos, acorde con la prueba de Ljung Box. Además, acorde con la prueba de Levene, los residuales presentan heterocedasticidad. Además, el ajuste tiene un MAPE de 2,05% y MAPE de pronóstico de 6,20%. Por lo anterior, no fue muy acertado el uso de este modelo para pronosticar el consumo energético diario.

Este resultado indicó que existen fuentes de variación importantes, que deben ser consideradas antes de estimar el modelo SARIMA, aspectos considerados para el modelo econométrico.

Modelo econométrico

La ecuación siguiente refleja el modelo final estimado, que fue mejorado usando pruebas de significancia al 5%

$$Z_t = \beta_0 + \beta_1 Z_{t-1} + \beta_2 Z_{t-7} + \dots + \alpha_1 t + \alpha_2 t^3 + \dots + \beta_i \text{Indicadoras(día semana)} + \beta_j \text{Indicadoras niveles} + \beta_k \text{Temperatura} + \varepsilon_t \quad (9)$$

En la estimación del modelo econométrico se involucraron variables como:

- Tiempo, para determinar impactos de la tendencia.
- Día, dado que el período estacional en la serie es 7.
- Comportamiento sinusoidal, para estudiar el mejoramiento de la estacionalidad.
- Tres rezagos (variables del pasado), por la autocorrelación detectada en la serie.
- Temperatura diaria.

Las variables artificiales exploradas fueron:

1. En la variable "Nivel", se catalogó la demanda en 4 intervalos (1, 2, 3, 4) de menor a mayor respectivamente. Estos intervalos tienen un rango de más o menos 14000 MWh cada uno.
2. En la variable Nivel 2 se catalogaron las fechas especiales, que son: 24, 25 y 31 de diciembre. Esto se hizo con el fin de que el modelo identificara los picos de la demanda energética Colombiana.
3. En Nivel 3 se identificaron los picos más bajos de los datos; generalmente lo representó el día domingo.

Estos niveles fueron incorporados al modelo, con el fin de poder captar picos y datos atípicos a la hora de pronosticar la demanda. Los coeficientes del modelo final se muestran en la segunda columna de la tabla 2.

Las pruebas de hipótesis para contrastar la significancia de las variables cuantitativas se realizaron analizando el valor P de la última columna de la tabla 2, del estadístico t, indicando con valores menores del nivel de significancia del 5%, que el parámetro tenía significancia. Para las variables explicativas cualitativas (artificiales o indicadoras) fue evidente que más de uno de sus coeficientes tuvo un valor P menor de 5%. Esto también se comprobó con la tabla ANOVA tipo III, que mostró valores p de la prueba F, menores al 5%.

Tabla 2. Coeficientes Estimados finales del modelo econométrico final.

Coeficientes	Estimado	Error estándar	Valor t	Pr(> t)
Constante	2,99E+02	8,30E+00	36,066	< 2e-16
t	8,08E-02	1,12E-02	7,201	1,52E-12
t ³	-3,88E-06	1,01E-06	-3,86	0,000124
Día lunes	3,65E+00	8,01E-01	4,555	6,16E-06
Día martes	7,15E+00	1,11E+00	6,465	1,88E-10
Día miércoles	5,06E+00	9,25E-01	5,473	6,15E-08
Día jueves	4,64E+00	8,84E-01	5,249	2,03E-07
Día viernes	4,34E+00	8,69E-01	4,994	7,46E-07
Día sábado	1,46E-01	7,54E-01	0,193	0,847106
Temperatura	2,83E-01	8,51E-02	3,327	0,000924
Indicador(nivel)2	1,63E+01	1,09E+00	15,032	< 2e-16
Indicador(nivel)3	2,35E+01	1,29E+00	18,309	< 2e-16
Indicador(nivel)4	3,28E+01	1,39E+00	23,606	< 2e-16
Indicador-m(Nivel)3	1,27E+01	9,21E-01	13,785	< 2e-16
Indicador-s(Nivel)3	1,98E+01	9,88E-01	20,054	< 2e-16
$\sqrt{Z_{t-7}}$	3,10E-02	1,58E-02	1,964	0,04989
$\sqrt{Z_{t-2}}$	8,56E-02	1,63E-02	5,241	2,11E-07

Las variables que tienen la palabra indicador, toman valor 0 ó 1; esto lleva a que los coeficientes estimados sean desviaciones con respecto al consumo energético de los domingos, pues se suman al intercepto.

Sin embargo, los residuos de este modelo no resultaron ser incorrelacionados; por ello fue necesario realizar una corrección. Este proceso fue realizado al modelar los residuales con un modelo SARIMA y, posteriormente, su estimación se unificó al modelo econométrico estimado para la respuesta Y_t .

Acorde con el alto comportamiento de dependencia de orden 1 y de orden 7 (estacional) fue evidente que estos valores debían incorporarse como rezagos explicativos en el modelo, lo que fue usado para encontrar un modelo con coeficiente estacional (tabla 3): SARMA(1,0)(2,1)₇

Tabla 3. Coeficientes del modelo SARMA (1,0)x(2,1)₇ para los residuales del modelo econométrico.

	Coeficientes:			
	ar1	sar1	sar2	sma1
	0,2876	-0,1040	0,1309	0,1555
S.E	0,0360	0,3958	0,0411	0,3998

Acorde con la tabla 4, el Valor P de la prueba de Ljung Box resultó mayor de 0,01. Por ello se puede afirmar que los residuos cumplen con el supuesto de incorrelación en los residuales. El Valor P de la prueba de JarqueBera fue menor de este nivel; por ello no cumplió normalidad, pero el Valor P de la prueba de Levene es de 1,24%, superior al 1%, cumpliendo homocedasticidad.

Tabla4. Pruebas de incorrelación, normalidad y homocedasticidad del modelo SARMA (1,0)x(2,1)₇ para los residuales del modelo econométrico.

Prueba	Estadístico	Valor P
Box-Ljung test	X-squared = 3,4695, df = 7,	0,8384
JarqueBera Test	X-squared = 174,3578,df = 2	2,2e-16
Levene's Test for Homogeneity of Variance	F value=2,6899, Df=6	0,01373

El error del ajuste encontrado fue muy bajo, MAPE ajustado: 1,21%. Asimismo, el indicador de la capacidad elegido para este trabajo, MAPE de pronóstico fue de 0,99%; indicando hasta ahora el mejor de todos los modelos estimados es este: Econométrico con residuales modelados con SARIMA.

Modelo de regresión Gaussiana vía MonteCarlo por Cadenas de Markov.

Se realizó una simulación Montecarlo con 12000 iteraciones, y se hizo un quemado de las primeras 2000 muestras obtenidas por medio del paquete estadístico MCMC Regress, con el fin de garantizar independencia en la simulación, que será comprobada con la prueba de Ljung-Box.

La tabla 5 muestra la media y desviación estándar para cada variable, con el error estándar. La media fue el valor usado para efectuar pronósticos.

Tabla 5. Coeficientes Estimados

	Media	Desviación estándar	Error estándar previo (Naive)	Series de tiempo SE
Intercepto	-6,10E+06	7,75E+05	7,75E+03	7,75E+03
T	3,06E+03	3,87E+02	3,87E+00	3,87E+00
Día lunes	2,94E+04	8,74E+02	8,74E+00	8,74E+00
Día martes	2,41E+04	7,07E+02	7,07E+00	7,07E+00
Día miércoles	2,18E+04	7,57E+02	7,57E+00	7,57E+00
Día jueves	2,09E+04	7,68E+02	7,68E+00	7,46E+00
Día viernes	2,05E+04	7,68E+02	7,68E+00	7,42E+00
Día sábado	1,13E+04	7,66E+02	7,66E+00	7,66E+00
Z_{t-1}	5,56E-01	3,12E-02	3,12E-04	3,18E-04
sigma2	2,58E+07	1,37E+06	1,37E+04	1,40E+04

La estimación del modelo, usando MCMC, permitió simular los parámetros de regresión, y para cada uno fueron generados 10000 datos, mostrando las estadísticas básicas en la tabla 5. La media es el estimador elegido como el coeficiente a incorporar en el modelo de regresión vía MCMC, usando el muestreador de Gibbs. Ahora, es necesario realizar una prueba que indique que los resultados de la simulación no presentaron dependencia alguna. Para ello, se aplicó el test de Ljung Box.

Tabla 6. Pruebas de incorrelación de las muestras de parámetros

Prueba Ljung Box	
Variable	Valor P
Constante	0,5912
T	0,2834
Día lunes	0,5784
Día martes	0,1106
Día miércoles	0,4774
Día jueves	0,2834
Día viernes	0,5343
Día sábado	0,06226
Retardo Z_{t-1}	0,2498
Sigma2	0,1011

En la tabla 6 se observa que las muestras de los parámetros que fueron simulados son incorrelacionados, ya que los valores p están por encima del 5%; además, los indicadores de error obtenidos son: MAPE ajuste=1,90% y el MAPE pronóstico=4,61%.

Al sintetizar todos los modelos (tabla 7), se observó que el modelo econométrico con residual SARIMA fue el más asertivo para pronosticar este tipo de demanda; esto se debe a que esta técnica permitió incorporar las variables exógenas, admitiendo algunas indicadores como las variables de nivel que condujeron a una buena estimación de los datos atípicos. No puede dejarse a un lado la técnica bayesiana, que por medio de la simulación MCMC permitió estimar

un modelo con el cual se elaboraron pronósticos con un valor MAPE de 4,6%, menor al del modelo SARIMA. Modelo que puede actualizarse con nuevos datos, corriendo de nuevo la simulación en cada tiempo t .

La técnica bayesiana mostró ser una alternativa, cuando no se hace necesario el cumplimiento de premisas o, incluso, cuando no se tienen muchos datos históricos.

Tabla 7. Síntesis de resultados

Modelo	MAPE pronóstico (%)	Premisas	Ventaja/desventaja encontrada
SARIMA	6,2	Residuales incorrelacionados, normalidad y heterocedasticidad.	Cumplió incorrelación. No cumplió normalidad, error de pronóstico poco aceptable.
Modelo econométrico con residuos modelados con SARIMA	0,99	Residuales incorrelacionados, normalidad y homocedasticidad.	No cumplió normalidad. Error de pronóstico óptimo (el mejor de todos). Buen resultado para estimar atípicos.
MCMCregress	4,62	Distribución normal a priori para los parámetros, la perturbación aleatoria y gamma inversa para la varianza de las perturbaciones aleatorias. Incorrelación en los parámetros.	Error aceptable. Se encontró que la simulación presentó independencia, usando el muestreador de Gibbs del paquete MCMCregress del software R. No se requieren los mismos supuestos de los modelos clásicos. Los pesos de cada día fueron un importante insumo para la técnica usada en el pronóstico de demanda de energía diaria. Se encontró muy buen resultado para estimar atípicos.

CONCLUSIONES

De todos los modelos estimados y analizados, se encontró que el modelo econométrico con errores SARIMA fue el óptimo para lograr el menor error de pronóstico, acercándose más a la realidad del comportamiento de la energía diaria. Esto se debe a que al incorporar variables significativas como los retardos, las variables ficticias o de nivel, explicaron mejor los cambios estructurales en la serie, logrando con ello proporcionar una herramienta que puede facilitar la optimización final del abastecimiento energético diario de Colombia, por la confianza del pronóstico.

El modelo SARIMA para el consumo energético, que era uno de los utilizados para estimar el pronóstico de demanda de energía por el Centro Nacional de Despacho (CND), mostró un problema con la validación de supuestos requerida y, además, fue superado por la eficiencia

del modelo econométrico, lo cual sugiere replantearlo, con la herramienta proporcionada en este trabajo.

Una gran ventaja del método de regresión bayesiana gaussiana por MCMC, es que no requiere un cumplimiento de los mismos supuestos de un modelo clásico, y facilita la actualización de datos nuevos con la simulación para cada tiempo $t+1$ que desee pronosticarse. En este trabajo, el método simuló los parámetros de un modelo de regresión, llegando a generarlos sin dependencia temporal. Se mostró, además, una eficiencia en el pronóstico aceptable, brindando una alternativa cuando existen pocos datos y modelos difíciles de ajustar.

Recomendaciones

Las alternativas de pronósticos en la industria no siempre son fácilmente aplicables. Este trabajo puede extenderse al pronóstico de demanda energética de regiones específicas del país, incluso aquellas que apenas inician con la distribución energética controlada por ISA.

Sobre los modelos bayesianos, aún quedan otros por explorar, donde se consideren otras distribuciones a priori o procesos que incorporen las no informativas, extendiendo una posibilidad de aumentar la investigación en un campo de alta necesidad y cuidado para la regulación económica: la demanda de energía eléctrica.

REFERENCIAS

ANDRIEU, Christophe., *et al.* An Introduction to MCMC for machine learning. *Machine learning* [en línea], 2003, vol. 50, p. 5–43. [citado Octubre, 2013]. Disponible en internet: <http://link.springer.com/article/10.1023/A:1020281327116>

BARRERA, Carlos Javier., and CORREA, Juan Carlos. Distribución predictiva bayesiana para modelos de pruebas de vida vía MCMC. *Revista Colombiana de Estadística*. 2008. vol. 31, no 2, p. 145–155.

BOX, George., and JENKINS, Gwilym. *Time Series Analysis: Forecasting and Control*. San Francisco: Holden-Day, 1976

CARIDAD Y OCERIN, José M. *Econometria: Modelos Econométricos y series temporales*. Ed. Reverté S.A. 1998.

GEYER, Charles., and THOMPSON, Elizabeth. Annealing Markov Chain Monte Carlo with applications to ancestral inference. *Journal of the American Statistical Association* [En línea]. 1995, vol. 90, no. 431, p. 909–920. [citado octubre, 2013]. [en línea] <<http://amstat.tandfonline.com/doi/full/10.1080/01621459.1995.10476590>>

GILL, Jeff. *Bayesian Methods-A social and Behavioral Sciences Approach*. 2002. 459 p.

GÓMEZ, Gonzalo., CARREÑO, Juan Carlos., and ZAMBRANO, Juan. Herramienta de software para el pronóstico de demanda horaria de potencia eléctrica en el sistema eléctrico de Codensa S.A. ESP. *Tecnura*. 2011, vol. 15, no. 28, p. 7-22.

JARQUE, Carlos., and BERA, AK. Test for Normality of Observations and Regression Residuals. *International Statistical*, 1987, p.163-172.

LJUNG, Box. On a Measure of a Lack of Fit in Time Series Models. *Biometrika*, 1978. p. 297-303.

MAKRIDAKIS, Spyros., HIBON, Michel., and MOSER, Claus. Accuracy of Forecasting : An Empirical Investigation of Forecasting : An Empirical Accuracy Investigation. *Journal of the Royal Statistical Society*, 2011. vol.142, no. 2, p.97–145.

MARTIN, Andrew., QUINN, Kevin., and PARK, Jong Hee. MCMC pack: Markov Chain Monte Carlo in R. En: *Journal of Statistical Software*, [en línea] 2011. vol. 42, no 9, p. 1–21. [citado 2013]. [en línea] < <https://www.law.berkeley.edu/files/jstatsoftMCMCpack.pdf> >

MEDINA, Santiago.,andGARCIA, Josefina. Predicción de demanda de energía en Colombia mediante un sistema de inferencia difuso neurona. 2005. *ISSN 0120-9833*, p. 15-24.

MONTGOMERY, Douglas., PECK, Elizabeth., and VINNING, Geoffry. *Introduction to Linear Regression Analysis*. John Wiley and Sons. 2012. 5th Ed.

MURILLOS, Joaquín., TREJOS, Alvaro., and CARVAJAL, Patricia. Estudio del pronóstico de la demanda de energía eléctrica, utilizando modelos de series de tiempo. *Scientia et Technica*, 2003, vol. 23, p. 37-40.

RUEDA, Viviana, VELÁSQUEZ, Juan David. and FRANCO, Carlos Jaime. Avances recientes en la predicción de la demanda de electricidad usando modelos no lineales. *DYNA*, 2011. vol. 78, no. 167, p. 36–43.