

Modelo de Predicción de la Demanda Energía Eléctrica en Colombia por Regresiones Múltiples

John William Grimaldo Guerrero

Trabajo Final presentado como requisito parcial para optar al Título de
“Magíster en Ingeniería Eléctrica”

Director: José Daniel Soto Ortiz
Ingeniero Electricista. MSc en Ingeniería Eléctrica



Fundación Universidad del Norte
Departamento de Ingeniería Eléctrica y Electrónica
Maestría en Ingeniería Eléctrica
Barranquilla - Atlántico
2012

Nota de aceptación

Firma del presidente del jurado

Firma del jurado

Firma del jurado

Barranquilla, Mayo del 2013

"No habrá desarrollo sin educación, ni progreso sin cultura"

Prof. Alberto Assa

Tabla de contenido

Resumen.....	10
Abstract	10
1. INTRODUCCIÓN.....	11
1.1. Justificación.....	12
1.2. Antecedentes	13
1.3. Hipótesis	17
1.4. Objetivos	17
1.5. Alcances.....	17
1.6. Metodología	17
2. MARCO TEÓRICO.....	19
2.1. Pronósticos	19
2.2. Metodologías Cuantitativas	21
2.3. Precisión del pronóstico	22
2.4. Coeficiente Estadístico Cp de Mallows.....	23
2.5. Revisión Bibliográfica.....	24
2.6. Modelo de Regresión Lineal.....	27
2.7. Modelo de Regresión Lineal Múltiple	27
2.8. Regresión y Análisis de Varianza en Excel	29
2.9. Adquisición de Datos	31
3. MODELO DE ENERGÍA ELÉCTRICA.....	32
3.1. Modelo Lineal	32
3.2. Modelo de dos Variables	36
3.3. Modelo Multi-Variable	39
3.4. Comparación en los Pronósticos.	47
4. PRONÓSTICOS DEL MODELO MULTI-VARIABLE.....	51
4.1. Intervalos por Sector	52
4.2. Intervalo Global.....	53
5. CONCLUSIONES	56

6.	REFERENCIAS	57
40.	ANEXOS	60

Índice de Tablas

Tabla 1. Plantas Térmicas asociadas a la ANDEG.....	15
Tabla 2. Proyectos en Desarrollo. Fuente: Revista Semana [27].	16
Tabla 3. Diferencia entre la Metodología Cualitativa y la Cuantitativa [18].	20
Tabla 4. Modelos Internacionales.	24
Tabla 5. Modelos Colombianos. [29].....	25
Tabla 6. Resultados de la Regresión Entregada por Excel (1).....	29
Tabla 7. Resultados de la Regresión Entregada por Excel (2).....	29
Tabla 8. Resultados de la Regresión Entregada por Excel (3).....	30
Tabla 9. Estadísticas de la Regresión del Modelo Lineal.	32
Tabla 10. Análisis de Varianza del Modelo Lineal.....	32
Tabla 11. Coeficientes del Modelo Lineal.	33
Tabla 12. Resultados de los Pronósticos del Modelo Lineal.....	34
Tabla 13. Comparación del modelo 2000-2009 y el modelo 2000-2010.....	35
Tabla 14. Comparación de los Errores Porcentuales.....	35
Tabla 15. Comparación de los Modelos de 2 y 1 Variable.	36
Tabla 16. Análisis de Varianza del Modelo de Dos Variables.	36
Tabla 17. Coeficientes del Modelo de Dos Variables.	37
Tabla 18. Comparación de Resultados entre el Modelo de 2 Variables y de 1 Variable	37
Tabla 19. Comparación de los Errores Máximos del Modelo de 1 y 2 Variables.	38
Tabla 20. Correlación R^2 de los Sectores Económicos.....	39
Tabla 21. Nomenclatura para los Sectores.	41
Tabla 22. Selección de los 10 Modelos con las Mejores Estadísticas.....	42
Tabla 23. Conteo de Repetición de las Variables.....	42
Tabla 24. Resultados del Modelo ADH.	43
Tabla 25. Comparación de los Modelos Múltiples.	43
Tabla 26. Análisis de Varianza para el Modelo ADH	44
Tabla 27. Coeficientes para el Modelo ADH.....	44
Tabla 28. Análisis de Varianza del Modelo ADEH.	44
Tabla 29. Coeficientes del Modelo ADEH.	45
Tabla 30. Análisis de Varianza del Modelo ABEFH.	45
Tabla 31. Coeficientes del Modelo ABEFH.	45
Tabla 32. Análisis de Varianza del Modelo ACDH.	46
Tabla 33. Coeficientes del Modelo ACDH.....	46
Tabla 34. Pronostico de los Modelos Multi-Variable.....	47
Tabla 35. Residuos de los Pronósticos.	47
Tabla 36. Error Porcentual de los Pronósticos.....	48
Tabla 37. Errores de los Modelos Diseñados.	48
Tabla 38. Error Máximo Negativo y Positivo.	49
Tabla 39. Máximo y Mínimo Error Absoluto.	49

<i>Tabla 40. Puntuación para Máximo Error Positivo y Negativo.</i>	49
<i>Tabla 41. Puntuación para la Mayor y Menor Desviación Porcentual.</i>	50
<i>Tabla 42. Intervalos Generados para cada Sector.</i>	52
<i>Tabla 43. Pronósticos del Modelo ADEH con Intervalos para cada Sector.</i>	52
<i>Tabla 44. Tasas de Crecimiento para el PIB Global.</i>	54
<i>Tabla 45. Pronósticos del Modelo ADEH con Intervalos para el PIB Global.</i>	54

Índice de Ilustraciones

<i>Ilustración 1. Evolución Trimestral de la Demanda de Energía Eléctrica y el PIB</i>	<i>13</i>
<i>Ilustración 2. Evolución del Parque Generador Colombiano 2008 – 2018.</i>	<i>14</i>
<i>Ilustración 3. Metodología para la Formulación del Modelo.</i>	<i>18</i>
<i>Ilustración 4. Nube de Puntos de la Serie de Datos.</i>	<i>33</i>
<i>Ilustración 5. Modelamiento de los Pronósticos Vs los Datos Reales.</i>	<i>34</i>
<i>Ilustración 6. Modelamiento de los Pronósticos Vs los Datos Reales.</i>	<i>38</i>
<i>Ilustración 7. Correlación de Minería Vs DEE.</i>	<i>39</i>
<i>Ilustración 8. Correlación de Servicios Públicos Vs DEE.</i>	<i>39</i>
<i>Ilustración 9. Correlación de Construcción Vs DEE.</i>	<i>40</i>
<i>Ilustración 10. Correlación de Agricultura Vs DEE.</i>	<i>40</i>
<i>Ilustración 11. Correlación de Manufactura Vs DEE</i>	<i>40</i>
<i>Ilustración 12. Correlación de Comercio Vs DEE.</i>	<i>40</i>
<i>Ilustración 13. Correlación de Financiero Vs DEE.</i>	<i>40</i>
<i>Ilustración 14. Correlación de Sociales Vs DEE.</i>	<i>40</i>
<i>Ilustración 15. Histograma y Gráfico de Densidad para el Sector de Agricultura.</i>	<i>51</i>
<i>Ilustración 16. Histograma y Gráfico de Densidad para el Sector de Minería.</i>	<i>51</i>
<i>Ilustración 17. Evolución de la Demanda de Energía Eléctrica Usando el Modelo ADEH.</i>	<i>53</i>
<i>Ilustración 18. Histograma y Gráfico de Densidad para el PIB Global.</i>	<i>53</i>
<i>Ilustración 19. Evolución de la Demanda de Energía Eléctrica.</i>	<i>54</i>
<i>Ilustración 20. Comparacion de la Evolucion de la DEE.</i>	<i>55</i>

Índice de Anexos

<i>Anexo 1. Demanda Anual de Energía Eléctrica.....</i>	<i>60</i>
<i>Anexo 2. Datos de la Población Colombiana.....</i>	<i>61</i>
<i>Anexo 3. PIB de las Ramas Económicas, dadas en miles de millones de pesos.</i>	<i>62</i>
<i>Anexo 4. Modelos Obtenidos al Combinar las Diferentes Variables.</i>	<i>63</i>

Resumen

En la presente memoria se muestra la investigación de un modelo de regresión lineal múltiple aplicado al pronóstico anual de la demanda de energía eléctrica, sus resultados fueron validados mediante estadísticas de regresión y el coeficiente Cp de Mallows, usando datos reales del Sistema Eléctrico Colombiano.

Palabras Claves: *Modelo de Regresión, Demanda Energética, Pronóstico.*

Abstract

In the present report is submitted the investigation of a linear multiple regression model. It was applied to the annual forecast of the electrical load. The results took account of the regression's statistics and the Mallows Cp coefficient, using real data of the Colombian System.

Keywords: *Regression Model, Energetic Demand, Forecast.*

1. INTRODUCCIÓN

La presente tesis de investigativa plantea la implementación de un modelo de regresión lineal múltiple para la predicción de la demanda de energía eléctrica, DEE, anual de Colombia; el cual tendrá en cuenta los factores del producto interno bruto (PIB) de los diferentes sectores económicos definidos por el Banco de la República de Colombia y la población. El objetivo del proyecto consiste en presentar una alternativa de los pronósticos anuales de la DEE con ayuda del PIB sectorial, con el fin de proporcionar evidencia experimental sobre la posibilidad de plantear y realizar políticas que promuevan la conservación de la energía eléctrica. El presente trabajo de investigación dejará una base para el diseño de un modelo para el pronóstico de la demanda anual de energía eléctrica en Colombia, con el fin de mejorar el modelo que se tiene actualmente y así mismo mejorar el desempeño del sistema eléctrico en el sector de la planeación. La validación del modelo está sujeta a la información suministrada por los entes gubernamentales y empresas encargadas del sector eléctrico.

Durante el presente capítulo se explicará cuál es el problema a solucionar así como la metodología utilizada para cumplir satisfactoriamente los objetivos trazados. El documento se encuentra organizado de la siguiente manera:

Capítulo 1: Se realiza una presentación del panorama actual, sobre el proyecto y las pautas realizadas, para el diseño de esta investigación. A su vez se presenta un breve resumen de lo que se encontrara posteriormente en el documento.

Capítulo 2: Se explican los conceptos como el coeficiente estadístico C_p Mallows, cálculos de los errores de pronóstico, se muestran las investigaciones sobre la misma línea que se han realizado. Además como realizar un modelo de regresión lineal simple y múltiple, entender los resultados que se obtienen y por último se presentan los datos con los que se han trabajado durante la investigación.

Capítulo 3: Se presentan los resultados obtenidos en la generación de los diferentes modelos, empezando con el modelo de regresión lineal simple y los modelos de regresión lineal de dos y los multi-variable. El diseño del modelo de regresión fue realizado con las cifras históricas del consumo anual de energía eléctrica, crecimiento de la población y PIB de Colombia correspondientes al periodo comprendido entre el año 2000 al 2011

Capítulo 4: En este capítulo se muestran los pronósticos obtenidos a partir del modelo escogido, que cumplió con todos los requisitos de validación. El óptimo desempeño del modelo de predicción y su posible eficiencia en la aplicación como modelo de

predicción en la industria del sector energético, fue comprobado al examinar los errores porcentuales absolutos inferiores al 1% obtenido en la etapa de prueba del modelo de predicción

Capítulo 5: Contiene las conclusiones de la tesis y los trabajos futuros recomendados para continuar con la investigación.

1.1. Justificación

La estimación energética es uno de los puntos más importante en el ámbito de la planificación, dado que estos dan un punto inicial para poder decidir sobre la oferta que poseemos actualmente, es decir si ésta llega a ser insuficiente para los años próximos, podremos evitar inconvenientes de abastecimiento energético diseñando un plan que involucra la creación de proyectos de ampliación de la oferta, control de crecimiento de la demanda o crear acuerdos para la importación y exportación energética.

Debido a los problemas que se presentaron en el racionamiento de energía eléctrica pasado, causado principalmente por los malos manejos y la falta regulación del sector. En este momento nuestro país no es ajeno a un posible racionamiento, debido a que presenta los siguientes inconvenientes:

- ✓ Atrasos en los proyectos de las nuevas plantas de generación [30].
- ✓ Falta de claridad en la regulación del sector del gas [15] [23] [27].
- ✓ Contratos de venta de gas en firme a las termoeléctricas hasta el 2014 [23].
- ✓ El aumento y la inestabilidad de los precios de los combustibles fósiles [17].
- ✓ La fuerte incidencia del fenómeno del niño y la niña que tienen sobre el parque generador Colombiano [15].

Lo anterior hace conveniente tener un manejo más preventivo en las propuestas de mediano y largo plazo para evitar futuros problemas. Hay soluciones llamadas de segunda mano como es la importación de gas, pero ésta no es una solución totalmente factible, debido a que en esos momentos se podría tener una volatilidad en el precio o peor aún no permitan la venta debido a la falta del energético en el país oferente, provocando una gran incertidumbre en el gremio de las termoeléctricas a la posibilidad de no poder responder a la demanda.

Si bien el modelo que comúnmente se ha estado usando en la planificación del crecimiento de la DEE es un modelo que se ajusta al crecimiento del PIB [29], gracias

a la alta correlación que existe entre ellos dos, el cual ha funcionado correctamente, pero a partir del año 2003 la correlación que hay entre ellos ha venido en declive. Esta correlación puede ser observada en la ilustración 1, según una publicación de XM [39] dicen que esta relación “se ha venido desacoplando debido probablemente a:

- ❖ *La alta contribución en el PIB de ciertas actividades que tienen un menor consumo de electricidad en sus procesos productivos (ej: actividad de la construcción).*
- ❖ *La mayor eficiencia en los procesos.*
- ❖ *Sustitución de la energía eléctrica por otros energéticos.*
- ❖ *El aumento de la autoproducción de energía.”*

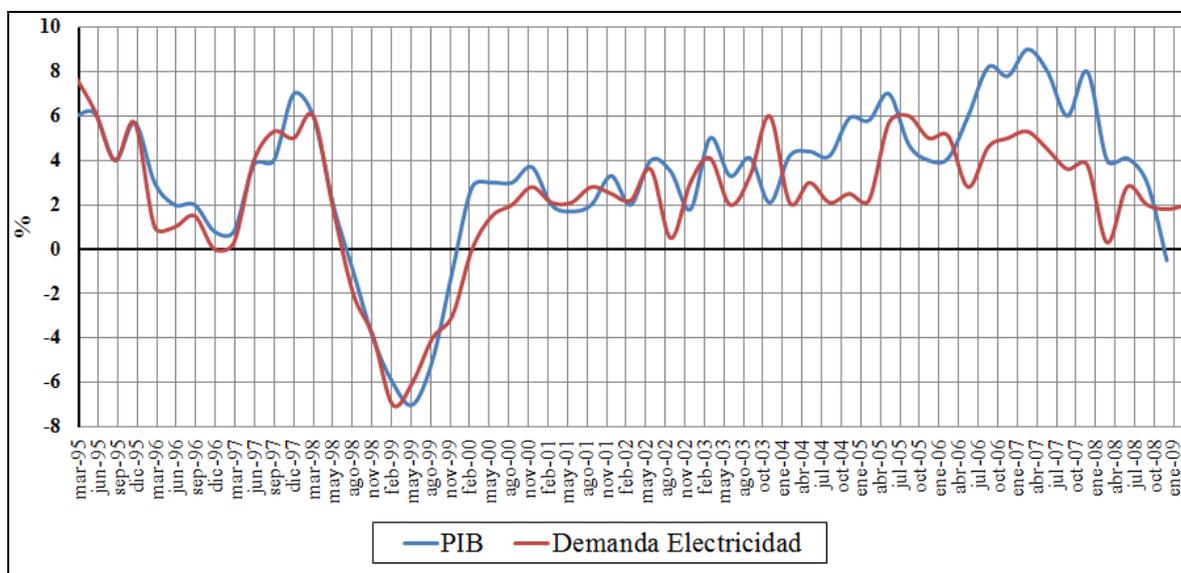


Ilustración 1. Evolución Trimestral de la Demanda de Energía Eléctrica y el PIB

Fuente: Tomada de XM [39], Adaptado.

Con la investigación realizada se quiso dar un modelo alternativo para el diseño de pronósticos anuales de largo plazo. Este modelo presenta mejor comportamiento en sus resultados que los modelos anteriores que se han usado hasta la fecha. La tesis estableció utilizar regresiones multi-variables, para dar diferentes enfoques de crecimiento, con el fin de involucrar los sectores de la economía según la clasificación que ha dado el Banco de la Republica de Colombia.

1.2. Antecedentes

La planeación es un tema de amplio interés para las empresas, porque con esto se alcanza a tener el mejor desempeño, bajo los estándares de calidad, continuidad y

confiabilidad exigidos por los usuarios y los entes reguladores. Los procesos de toma de decisiones y las estrategias en la planeación, deben ser apoyados con herramientas que permitan analizar diversos escenarios de comportamiento y su evolución, con el propósito de establecer planes o realizar los ajustes correspondientes que brinden la información, para la toma de decisiones en los nuevos proyectos que fortalezcan al sistema en estudio. En el caso de la demanda de energía eléctrica es necesario tener una metodología de proyección para el crecimiento de ésta; con los resultados que se obtienen permite diseñar planes de expansión del parque generador, las redes de transmisión y de distribución.

Revisando la historia de Colombia se sabe que durante el año de 1890 se creó el primer servicio de energía eléctrica, a partir ese momento empezó la creación de nuevas empresas y la carrera por la expansión del sistema eléctrico, que dieron una nueva cara a la nación. Pero debido a su falta de control y regulación provocó tres racionamientos, en las fechas de 1976, 1981 y 1992, los dos primeros producidos por los retrasos en las obras de ampliación del parque generador, y el último por una cadena de sucesos, como la alta dependencia de las hidroeléctricas que relevaron a las termoeléctricas, y gracias a una fuerte sequía y al deficiente mantenimiento de estas últimas, se produjo el peor racionamiento que ha vivido Colombia. La experiencia obtenida en 1992, enseñó que debe existir un equilibrio entre la generación térmica y la hidráulica, además de llevar un buen plan de regulación y control. Al no tener en cuenta esto podría implicar un retroceso en el sistema eléctrico Colombiano.

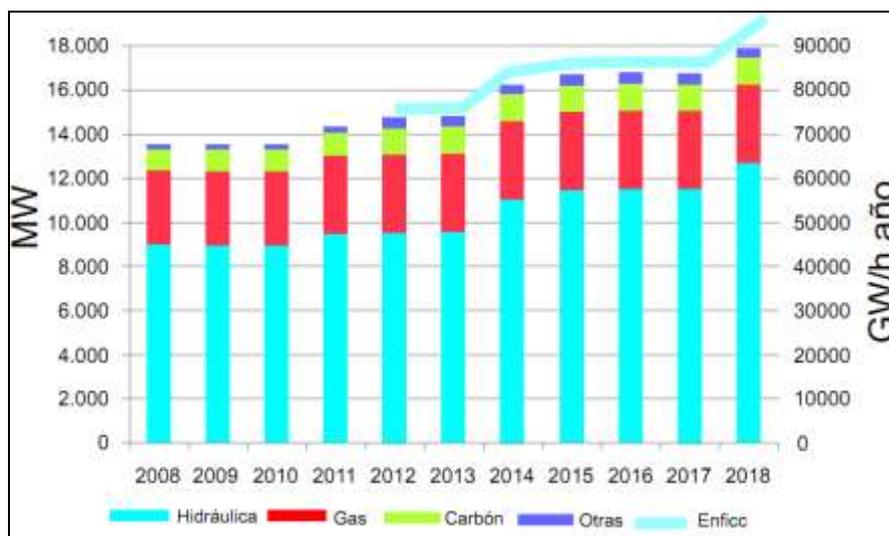


Ilustración 2. Evolución del Parque Generador Colombiano 2008 – 2018.

Fuente: ACOLGEN [03].

En ilustración 2 se observa la evolución del parque generador, podemos notar que el aporte por parte de las hídricas es el mayor, ocupando históricamente un porcentaje superior al 60% y las térmicas a un valor superior al 30%; siendo un buen indicador en tema de costos ya sea de generación y para el mismo usuario, pero en tiempos de sequía la operación del sistema cambia, las hidráulicas no pueden trabajar a todo su potencial y las térmicas pasan de entregar el 22% al 67% de la demanda de la electricidad [27], dejando al descubierto un problema de respaldo por parte del gas, combustible utilizado por las térmicas debido a su bajo costo; se tiene asegurado el suministro de gas para el 2012-2013, pero no se han realizado más contratos de suministro de gas dejando en incertidumbre a las generadoras térmicas. En tiempos de sequía estas son las plantas térmicas que estarían a disposición para cumplir con la demanda:

EMPRESA	CAPACIDAD (MW)
Termobarranquilla¹	918
Flores¹	610
Termopaipa² (I, II, III, IV)	321
Termocandelaria¹ (I, II)	314
Guajira¹² (I, II)	320
Termoemcali¹	229
Termovalle¹	205
Merielectrica¹	157
Termotasajero²	155
Proelectrica¹	90
Termoyopal¹	29

Tabla 1. Plantas Térmicas asociadas a la ANDEG.

Fuente: Asociación Nacional de Empresas Generadoras, ANDEG [27].

Con las termoeléctricas mencionadas en la tabla 1 se lograría tener una capacidad efectiva de 3.348 Megavatios, capaces de responder con total claridad a una alta demanda, siempre y cuando se tenga el combustible a disposición y los mantenimientos preventivos al día; esto regala un parte de tranquilidad, sumado que los embalses nacionales se encuentran en sus máximos debido a la pasada ola invernal; pero como se puede notar 8 de las empresas expuestas anteriormente funcionan con

¹ Tienen al Gas como principal combustible.

² Tienen al Carbón como principal combustible.

gas como principal combustible, dejando a 2552 Megavatios en una posible deriva si no hay contratos de gas en firme para el 2014.

Gracias a la crisis del 92 hubo una reestructuración del sistema eléctrico, ley 141 y 142 de 1992, creando nuevas entes que velan por la seguridad energética de la nación, durante cada año analizan la situación y generan pronósticos de la oferta y de la demanda de los energéticos, a su vez abren licitaciones de inversión para la construcción de nuevas plantas generadoras o expansión del sistema interconectado; a continuación se presenta los proyectos que se tienen en curso y su fecha para entrada en operación.

#	PLANTA EMPRESA ENCARGADA	ENTRADA EN OPERACIÓN	ENERGETICO PRINCIPAL
1	Gecelca III - Gecelca	2012	Gas
2	Barroso - Century Energy	2012	Agua
3	Termocol - Poliobras	2012	Gas
4	Alto Tulua - EPSA	2012	Agua
5	Amoya - Isagen	2013	Agua
6	Popal - Century Energy	2013	Agua
7	Bajo Tuluá - EPSA	2013	Agua
8	Hidrosogamoso - Isagen	2014	Agua
9	El Quimbo - Emgesa	2014	Agua
10	Cucuana - EPSA	2014	Agua
11	Gecelca - Gecelca	2015	Gas
12	Termotasajero II - Termotasajero	2015	Carbón
13	Rio Ambeima - Empresa de Energía de los Andes	2015	Agua
14	Carlos Lleras Restrepo - Hidroeléctrica Alto Porce	2015	Agua
15	San Miguel - Century Energy	2015	Agua
16	Termonorte - Termonorte	2017	Gas
17	Hidroituango - EPM	2018	Agua
18	Porvenir II - Producción Energía	2018	Agua

Tabla 2. Proyectos en Desarrollo. Fuente: Revista Semana [27].

Actualmente no se han presentado ningún plan para un posible racionamiento, pero con los últimos retrasos en las obras de ampliación del parque generador, y a la sequía que se prevé en el año 2014[11], presenta un gran desafío que a la vez abre la puerta a una crisis parecida a la del año 1992, de aquí la necesidad de realizar modelos que generen menores errores de pronóstico y no tener demandas subestimadas que coloquen en riesgo la estabilidad en la producción y suministro de energía eléctrica.

1.3.Hipótesis

Se puede lograr un modelo tal que mejore los errores porcentuales en los pronósticos, en comparación al modelo de regresión lineal de variable PIB y el modelo de dos variables población y PIB.

1.4.Objetivos

Generales:

- ✓ Diseñar un modelo para pronosticar la demanda de energía eléctrica a largo plazo.

Específicos:

- ✓ Determinar las variables para realizar la predicción de la demanda.
- ✓ Plantear un modelo que pueda validar datos anteriores al 2009.
- ✓ Validar el modelo para los valores posteriores a los datos usados.
- ✓ Realizar pronósticos la demanda de energía de los años siguientes.

1.5.Alcances

- ❖ Se realizara un modelo para el pronóstico de la demanda anual de energía eléctrica.
- ❖ La base de datos serán tomados de las pagina oficiales del Banco de Republica de Colombia, UPME, el Banco Mundial y entes reguladores.
- ❖ Se tendrá un modelo de regresión lineal múltiple que podrá pronosticar la demanda de energía eléctrica de Colombia con un error inferior al 2%.
- ❖ Se tendrá una metodología a la cual se podrá extrapolar para el uso en otros países.

1.6.Metodología

En aras de completar los objetivos planteados anteriormente se realizaron las siguientes fases de trabajo:

Fase 1: Revisión del estado del arte y la técnica.

Se recolecta la información acerca de avances teóricos y técnicos, relacionados con los modelos de predicción, principalmente modelos de series anuales. Se revisarán también tendencias relacionadas con la selección de modelos de regresiones multivariadas. Para el desarrollo satisfactorio de esta fase se consultaran referencias bibliográficas, revistas, artículos, bases de datos, etc.

Fase 2: Selección de los datos y del modelo.

En esta fase se crea el banco de datos con las variables escogidas, según la recomendación de la literatura, se escoge el tipo de modelo que se diseñará. Es necesario hacer un estudio para descartar tipos de modelos que no pueden ser utilizados.

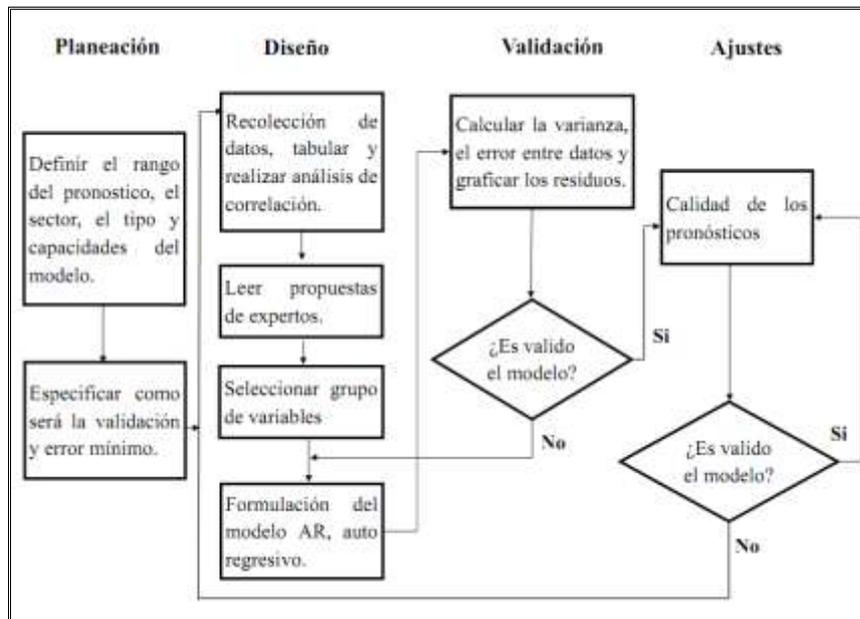


Ilustración 3. Metodología para la Formulación del Modelo.

Fase 3: Modelamiento y simulación del modelo.

Al llegar a esta fase deben encontrarse definidos todos los datos, el tipo de modelo y la metodología de escogencia; se realizará los diferentes modelos para obtener sus resultados en los pronósticos para así continuar con la fase de validación.

Fase 4: Validación del modelo.

En esta fase se usará una validación correspondiente al tipo de modelo. En este caso se realizará primero una selección por medio del uso del coeficiente estadístico Cp Mallows y las estadísticas de regresión; y luego se validara con los errores de pronóstico en datos anteriores que no hayan sido usados en la fase de diseño. En la ilustración 3 muestra el procedimiento realizado.

Fase 5: Generación de pronósticos.

Al finalizar la validación del modelo se procede a la generación de los pronósticos, estos se realizan con pronósticos de intervalo, gracias a la facilidad de tomar la diferente gama de escenarios.

2. MARCO TEÓRICO

A continuación se hallara información que ayudara al lector a entender mejor la tesis.

2.1. Pronósticos

Pronosticar es “*adelantarse a lo que va a suceder en el futuro a partir de algunos indicios, datos o informaciones*” [32]. Con ayuda de esta información da la posibilidad de diseñar planes ya sean de inversión, prevención o dependiendo del dato pronosticado para estar mejor preparados para ese futuro incierto, se debe notar que las previsiones muy rara vez llegan a ser perfectas. Para generar pronósticos se debe tener primero en cuenta el plazo al que se realizara, estos se dividen en tres tipos de horizontes temporales, ellos son:

- ✓ **Previsión a corto plazo:** Son generalmente inferiores a los tres meses, son usados especialmente para la programación de actividades, trabajos y tareas de corta duración.
- ✓ **Previsión a medio plazo:** Su plazo está entre tres meses a tres años. Son los más comunes y son usados en la planificación de las ventas, de la producción y del presupuesto.
- ✓ **Previsiones a largo plazo:** Sus periodos son superiores a los tres años. Estas tiene un impacto sobre la dirección de los sistemas de producción y que a su vez deben ser consistentes con las metas a largo plazo de las organizaciones.

En los generadores permite localizar los periodos para los cuales habrá faltante o excesos de capacidad de generación, y así planificar las inversiones de expansión del sistema, estas inversiones son proyectos de gran tiempo de ejecución y la entrada a tiempo puede generar mayor confiabilidad al sistema. En los distribuidores y transmisores permite definir los mantenimientos y a su vez programar la expansión del sistema de transmisión.

En los comercializadores, es un punto fundamental para el estudio del comportamiento del mercado y generar los planes operativos, que a su vez ayudaran al operador para programar la operación de las generadores minimizando los ajustes de entrada de operación evitando sobrecostos en la producción, en Colombia las generadoras térmicas poseen un costo adicional de arranque y parada debido al funcionamiento de sus motores.

A la secuencia de datos uniformemente espaciados, es decir es la toma de datos de una variable en periodos de tiempo regulares se les llama Series Temporales. Para la

realización de pronósticos a partir de estos se supone que los datos pasados han sido influenciados por factores, y que a su vez el cambio de estos factores influenciaran en el pasado y lo seguirán haciendo en el futuro. Por lo general las series temporales tienen cuatro características, estas son:

- ❖ **Tendencia:** Es el cambio gradual de los datos a lo largo del tiempo, este es influenciado por la variación de los factores asociados a la variable en estudio.
- ❖ **Estacionalidad:** Son los ascensos o los descensos de la variable en estudio durante un periodo de tiempo definido; estas son producidas por los factores que tienen variaciones durante dicho periodo, un ejemplo muy común son las estaciones climáticas, o las horas de un día.
- ❖ **Ciclos:** Son las repeticiones de una estacionalidad durante toda la serie de datos. para los ejemplos anteriores los ciclos serian un año y un día respectivamente.
- ❖ **Variaciones Aleatorias:** Son variaciones en los datos causados por el azar y situaciones inusuales. Un ejemplo de estas son las huelga, desastres naturales por lo general son de corta duración y no son repetitivos.

Dependiendo del enfoque del pronóstico se clasifican en investigación cualitativa y cuantitativa, en la siguiente tabla se muestran las diferencias entre estos tipos de investigación:

INVESTIGACIÓN CUALITATIVA	INVESTIGACIÓN CUANTITATIVA
Centrada en la fenomenología y comprensión	Basada en la inducción probabilística del positivismo lógico
Observación naturista sin control	Medición penetrante y controlada
Subjetiva	Objetiva
Inferencias de sus datos	Inferencias más allá de los datos
Exploratoria, inductiva y descriptiva	Confirmatoria, inferencial, deductiva
Orientada al proceso	Orientada al resultado
Datos “ricos y profundos”	Datos “sólidos y repetibles”
No generalizable	Generalizable
Holista	Particularista
Realidad dinámica	Realidad estática

Tabla 3. Diferencia entre la Metodología Cualitativa y la Cuantitativa [18].

2.2. Metodologías Cuantitativas

Para el manejo de las series de datos es mucho mejor el uso de las metodologías numéricas, a continuación presentamos algunos de los modelos cuantitativos que se usan [12]:

- ❖ **Media Simple y Móviles:** Se fundamenta en realizar una media entre los valores pasados, la diferencia entre las dos radica que en la media móvil solamente se toman los n periodos anteriores para predecir el siguiente periodo.
- ❖ **Alisado Exponencial:** Es un caso similar de media móvil, la diferencia radica en darle una ponderación mayor a los periodos recientes, y esta disminuirá exponencialmente para los periodos muy anteriores.
- ❖ **Proyección de Tendencia:** Es conocido comúnmente como regresión; su fundamento radica en buscar una tasa de crecimiento regular para cada periodo de tiempo. De este tipo de modelo se desprende los modelos de regresión lineal, exponencial, entre otros.
- ❖ **Modelos Asociativos:** Son modelos de regresiones asociadas a una variable de entrada y una variable de salida, en este tipo de modelos se toma la suposición de una relación entre las dos o más variables involucradas.
- ❖ **Método Winters:** Consiste en realizar una tendencia lineal, pero a esta se le agregara un patrón estacional; es decir su función es analizar la serie temporal y luego aplicarle un suavizado exponencial.
- ❖ **Técnica de Box-Jenkins:** Se investigan generalmente tres patrones: auto-regresión, promedios móviles y tendencias. El procedimiento requiere una secuencia de tres pasos que son la identificación, donde se busca el modelo apropiado, la estimación, en el que se registran en una secuencia temporal los valores estimados de los coeficientes y el diagnóstico, en el que se verifica la conveniencia del ajuste para ver si es el adecuado [24].
- ❖ **Serie de tiempo de Shiskin:** Es útil cuando se trata de descomponer una serie de tiempo en sus características básicas que son estacionalidad, tendencia e irregulares, su punto débil es la necesidad de una gran cantidad de datos históricos.
- ❖ **ARIMA:** Su sigla significa modelo Auto Regresivos Integrados de Medias Móviles, su estudio esta sobre una serie temporal, sin uso de variables externas, es el método más común usado por los expertos en el área.

2.3. Precisión del pronóstico

La calidad de un modelo se basa en principalmente en el alto grado de correlación que hay entre el modelo y la variable respuesta, además de la calidad que tiene para pronosticar los datos futuros; a continuación se presentan las medidas de error que se utilizan para evaluar la calidad de los modelos producidos.

El error de un pronóstico (E) es la diferencia que hay entre el valor real (VR) y el valor calculado (VC) para cualquier punto t de la serie temporal,

$$E_t = VR_t - VC_t \quad (1)$$

El cual indica la diferencia que hay entre los puntos, también llamado residuo, o calcular el error porcentual, este da una mejor imagen del pronóstico ya que indica el porcentaje de desfase que se tiene,

$$\%E_t = \frac{VR_t - VC_t}{VR_t} * 100 \quad (2)$$

A partir de estos existen diferentes medidas para interpretar el error, entre ellas tenemos [37]:

Error absoluto de la media (MAD)

$$MAD = \frac{\sum_{t=1}^N |E_t|}{N} \quad (3)$$

Error absoluto porcentual de la media (MAPE)

$$MAPE = \frac{\sum_{t=1}^N \left| \frac{E_t}{VR_t} \right|}{N} \quad (4)$$

Desviación porcentual absoluta de la media (PMAD)

$$PMAD = \frac{\sum_{t=1}^N |E_t|}{\sum_{t=1}^N |VR_t|} \quad (5)$$

Error cuadrático de la media (MSE)

$$MSE = \frac{\sum_{t=1}^N E_t^2}{N} \quad (6)$$

Raíz del error cuadrático de la media (RMSE)

$$RMSE = \sqrt{\frac{\sum_{t=1}^N E_t^2}{N}} \quad (7)$$

2.4. Coeficiente Estadístico Cp de Mallows

Al momento de escoger un modelo de regresión se tiene una idea de cómo seleccionar en primera instancia al modelo, buscando que la relación entre las variables regresoras y la variable respuesta sea alta. Pero siempre se hace necesario tener un criterio estadístico que permita comparar los criterios de bondad de ajuste y de esta manera comparar los distintos modelos generados por un grupo de variables, uno de ellos es el coeficiente Cp de Mallows [34]. Este coeficiente tiene la forma:

$$Cp = \frac{SC_{Error}(p)}{s^2} - (n - 2p) \quad (8)$$

Donde;

p Es el número de variables, incluyendo la constante.

$SC_{Error}(p)$ Es la suma de cuadrados del error de un modelo

s^2 Es el cuadrado medio del error del ajuste obtenido cuando se incluyen todas las variables predictoras posibles.

Al analizar la relación del coeficiente Cp tenemos que:

$$\begin{aligned} E(Cp) &\approx E \left[\frac{SC_{Error}(p)}{s^2} - (n - 2p) \right] \\ &\approx E \left[\frac{SC_{Error}(p)}{s^2} \right] - (n - 2p) \\ &\approx E[(n - p)] - (n - 2p) \\ &\approx (n - p) - (n - 2p) = p \\ E(Cp) &\approx p \end{aligned}$$

Se concluye que si el modelo de p variables tiene un Cp muy parecido a p , este tendrá una mejor capacidad para predecir nuevos datos. Este procedimiento llega a ser bueno si no hay muchas variables predictoras. Cuando existen muchas variables predictoras dentro del modelo es preferible usar otros procedimientos entre ellos se recomienda los procedimientos backward o forward.

2.5.Revisión Bibliográfica

Durante la fase de revisión bibliográfica, se estudiaron artículos publicados en revistas indexadas, se complementó con la revisión de tesis doctorales buscando por “*pronósticos de demanda de energía eléctrica*”. Dentro de la revisión se encontró que la técnica más usada durante estos últimos tiempos son las redes neuronales, la segunda es los modelos ARIMA y luego las técnicas tradicionales como las regresiones lineales y no lineales. A continuación se presenta un resumen de los modelos encontrados, realizados a nivel internacional, con su respectiva descripción.

NOMBRE	DATOS	PAÍS	AUTOR	SERIE
Redes Neuronales (ANN)	Producción de Petróleos, Clima y Población	Arabia Saudita	Al-Saba y El-Amin (1999)	Anual
ANN	Precio de la Electricidad, IVA, Población	Irán	Azadeh, Ghaderi, Tarverdian y Saberi (2007)	Anual
Regresión Múltiple	Precio de la Electricidad, Población y Turistas	Ciprés	Egelioglu, Mahamad y Guven	Anual
Regresión Lineal	PIB y Población	Italia	Bianco, Manca y Nardini	Anual
Regresión Lineal Múltiple	PIB, Precio de la Electricidad, Población y Exportaciones	Nueva Zelanda	Mohamed y Bodger	Anual
Redes Neuronales	Temperatura	Turquía	Yalcinoz y Eminoglu (2005)	Diaria y Mensual
Regresión Múltiple	Variables Climáticas	Grecia	Mirasgedis, Sarafidis et al	Diario y Mensual
Modelo Híbrido Neuro-Difuso	Temperatura, Estación del año e índice de comodidad y consumo	Brasil	Velasco, Pacheco, Ribeiro y de Souza (2004)	Horario
Redes Neuronales	Temperatura	Japón	Senjyu, Mandal, Uezato y Funabashi (2004)	Horario
Redes Abductivas	Parámetros Climáticos e Indicadores Económicos y Demográficos	Arabia Saudita	Abdel, Al-Garni y Al-Nassar (1997)	Mensual
MLP	Datos Históricos	Irán	Azadeh, Ghaderi y Sohrabkhani (2007)	Mensual
DAN2	Clima	Taiwán	Ghiassi, Zimbra y Saidane	Mensual
AR	PIB e Importaciones	Líbano	Nasr, Bard y Saab	Mensual

Tabla 4. Modelos Internacionales.

Según lo observado se concluye que las investigaciones para la predicción de la demanda de energía eléctrica han usado en mayor proporción las redes neuronales y los modelos ARIMA. Además la mayor parte de los estudios se han realizado para

generar modelos de series diarias u horarias, existiendo muy pocos documentos que referencien series mensuales o con las series anuales. A continuación se presenta los modelos que han sido realizados en Colombia:

MODELO	SERIE	VARIABLES	AUTORES
ARIMA	Horas	Datos de Consumo para Horas, Día y Meses	Barrientos, Olaya y Gonzales (2007)
ARIMA	Horas		Murillo, Trejos y Carvajal (2003)
Perceptrón Multicapa	Horas		Sarmiento y Villa (2008)
Modelo de Componentes no Observables	Mensual		Franco, Velásquez y Olaya (2008)
ANFIS (Sistemas de Inferencia Borroso Adaptativo)	Mensual	Tarifa promedio de la Energía, Consumo de Combustibles, PIB, entre otras.	Medina y García (2005)
ARIMA, Perceptrón Multicapa, Red Neuronal Auto-regresiva	Mensual		Velásquez, Franco y García (2009)

Tabla 5. Modelos Colombianos. [29]

Dentro de la revisión solamente se encontraron tres documentos, en el cual trabajaron con series anuales y su metodología de trabajo fue precisamente regresiones múltiples, en algunos usando variables atípicas o poco convencionales.

A continuación se presenta un comentario de los trabajos antes mencionados:

A. *Economic Variables and Electricity Consumption in Northern Cyprus* [14].

- ✓ Usa los datos históricos de los años 1988 a 1997, para predecir el consumo anual de electricidad.
- ✓ Usa solamente 10 datos, no realiza pruebas de predicción.
- ✓ El estudio nace por el repentino crecimiento del precio de la energía eléctrica, el cual triplico el precio en un solo año, además de la falta de generación en la isla.
- ✓ Utiliza las variables demanda anual de electricidad, consumidores o número de suscriptores, población, salario, precio de la energía eléctrica y número de turistas
- ✓ La prueba de Valor F y F critico dieron aceptables, el R^2 y el ajustado dieron por encima del 90%.

B. Modelling and Forecasting Electricity Consumption of Malaysian Large Steel Mills [31].

- ✓ El modelo calcula la demanda máxima diaria y anual de energía eléctrica para Malasia mediante el uso de dos modelos el Top-Down y el Bottom-Up, el cual usa modelo de regresiones para calcular la demanda anual.
- ✓ Los datos usados fueron la producción y el consumo de acero, el producto interno bruto y el precio de la energía eléctrica, estas series van desde 1992 al 2008.
- ✓ Usa 14 datos para la generación del modelo y 3 datos para hacer la validación de este.
- ✓ Los resultados en la prueba de valor F y valor P dieron respuestas que pueden ser aceptadas, su MAPE fue igual a 4,27%, y el R^2 y el ajustado dieron alrededor del 99%.
- ✓ El modelo se validó para los últimos tres años, 2006-2008, y los errores obtenidos fueron -1,6%, 2,6% y 2,1% respectivamente.

C. Forecasting Electricity Consumption in New Zealand Using Economic and Demographic Variables [20].

- ✓ Las variables utilizadas fueron población, precio de la energía eléctrica y PIB, para proponer un modelo de pronóstico de la DEE.
- ✓ Las series de datos van desde el año 1965 hasta 1999, 35 datos en total.
- ✓ No realiza validación del modelo por medio de pronóstico de años anteriores, todos los datos se usan para generar la regresión.
- ✓ Las pruebas estadísticas, exactamente la prueba de F, realizadas dieron aceptables.

De lo anterior se concluye que se usaran los datos de los años 2000 hasta el 2009, y se validará su capacidad de pronóstico con los valores de la DEE de los años 2010 y 2011. Se comprobara que la prueba F y los errores MAPE y R^2 esten dentro de un rango óptimo, y que a su vez los errores de pronósticos sean inferiores al 1%.

2.6. Modelo de Regresión Lineal

Es el método que se aplica con mayor frecuencia en economía, donde se analiza el comportamiento de una variable, Y, a través de una variable X. Al aplicar el método de regresión lineal se obtiene una ecuación lineal del tipo:

$$Y = (m * X) + b \quad (9)$$

Pero al ser una ecuación de correlación, rara vez llegara a ser exacta y generara un error o una perturbación, entonces la ecuación queda como:

$$Y = (m * X) + b + u_x \quad (10)$$

$$\text{Donde } u_x = Y_{Real} - Y_{Calculada}$$

El objetivo de este modelo es minimizar el valor de u para todos los valores de X que son datos de entrada, para conseguirlo se encuentran los valores de m y b tales que u sea un mínimo para todo X, estos pueden ser hallados por medio de las siguientes ecuaciones:

$$m = \frac{\sum x \sum y - n \sum xy}{(\sum x)^2 - n \sum (x)^2} \quad (11)$$

$$b = \frac{\sum y - m \sum x}{n} \quad (12)$$

2.7. Modelo de Regresión Lineal Múltiple

El método es usado para analizar el comportamiento de una variable, Y, a través de un conjunto de variables X, este tipo de modelo genera una correlación con más de una variable y son asociadas en forma lineal. La forma de este modelo es:

$$y(x_1, x_2 \dots x_n) = b_0 + b_1 x_{11} + b_2 x_{21} + \dots + b_n x_{n1}$$

Para diseñar un modelo de regresión múltiple, se tiene un conjunto de ecuación similar al siguiente [21]:

$$\begin{aligned} y_1^* &= b_0 + b_1 x_{11} + b_2 x_{21} + \dots + b_n x_{n1} \\ y_2^* &= b_0 + b_1 x_{12} + b_2 x_{22} + \dots + b_n x_{n2} \\ &\dots \\ y_k^* &= b_0 + b_1 x_{1k} + b_2 x_{2k} + \dots + b_n x_{nk} \end{aligned} \quad (13)$$

Expresándolo en forma matricial tenemos el siguiente sistema:

$$Y^* = X * B$$

Donde los vectores y la matriz tienen la forma:

$$Y^* = \begin{bmatrix} y_1^* \\ y_2^* \\ \vdots \\ y_k^* \end{bmatrix} \quad X = \begin{bmatrix} 1 & x_{11} & x_{21} & \cdots & x_{n1} \\ 1 & x_{12} & x_{22} & \cdots & x_{n2} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_{1k} & x_{2k} & \cdots & x_{nk} \end{bmatrix} \quad B = \begin{bmatrix} b_0 \\ b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix} \quad (14)$$

Los valores Y^* son los valores calculados, la matriz X son los valores de las variables que toman para cada dato de Y , en este sistema debemos hallar los valores del vector B , que son los coeficientes que representarían la ecuación de la regresión múltiple. De lo anterior tenemos que

$$Error = E = Y - Y^* = \begin{bmatrix} e_1 = y_1 - y_1^* \\ e_2 = y_2 - y_2^* \\ \vdots \\ e_k = y_k - y_k^* \end{bmatrix} \quad (15)$$

Donde Y es el valor real y Y^* es el valor calculado; para determinar el conjunto de valores que minimicen el Error, hacemos:

$$\sum_{i=1}^k E_i^2 = \sum_{i=1}^k (y_i - y_i^*)^2 = E^t E \quad (16)$$

De lo cual sabemos que:

$$E = Y - XB, \text{ reemplazando tenemos que:}$$

$$E^t E = (Y - XB)^t * (Y - XB)$$

$$E^t E = Y^t Y - 2Y^t X B + B^t X^t X B$$

Para minimizar el error usamos el primer criterio de la derivada, entonces a la relación anterior, la derivamos e igualamos a cero:

$$\frac{\partial}{\partial B} E^t E = -2X^t Y + 2X^t X B = 0$$

Despejando al vector B tenemos que los mejores coeficientes están dados siempre y cuando la matriz $X^t X$ tenga inversa por la relación:

$$B = (X^t X)^{-1} X^t Y \quad (17)$$

2.8. Regresión y Análisis de Varianza en Excel

En esta sección del trabajo se explicará como leer los resultados que se entregaran próximamente en la sección de modelos. Para la obtención de los resultados, *diseño de los modelos de pronóstico*, se usó el software Microsoft Excel, este programa cuenta con una opción que se encuentra en análisis de datos llamada regresión, el cual calcula los coeficientes óptimos para realizar un modelo de regresión.

Hay que notar que si se utiliza esta opción, Excel entrega resultados numéricos y no formulas, es decir que si se llega a cambiar un dato de entrada tocaría hacer nuevamente el uso del comando. Al aplicar esta opción los resultados que arroja comúnmente son:

ESTADÍSTICAS DE LA REGRESIÓN	
Coefficiente de correlación múltiple	0,999655439
Coefficiente de determinación R²	0,999310997
R² ajustado	0,998759794
Error típico	159,2220757
Observaciones	10

Tabla 6. Resultados de la Regresión Entregada por Excel (1).

Con ayuda de la tabla 5 se calculan los estadísticos mostrados en la tabla 4, recuerde que estos valores están relacionados con la varianza y covarianza de los datos utilizados. No se muestran las ecuaciones porque no es parte de este trabajo, se recomienda la bibliografía [36] para el entendimiento de estos estadísticos.

ANÁLISIS DE VARIANZA					
	<i>Grados de libertad</i>	<i>Suma de cuadrados</i>	<i>Promedio de los cuadrados</i>	<i>F</i>	<i>Valor crítico de F</i>
Regresión	4	183846725,7	45961681,41	1812,964689	4,35921E-08
Residuos	5	126758,347	25351,66939		
Total	9	183973484			

Tabla 7. Resultados de la Regresión Entregada por Excel (2)

La tabla de resultados que entrega Excel, tabla 4, como su título lo indica son las estadísticas de la regresión, estos indicadores expresan que tan correlacionados se encuentran las variables. El primer número es el coeficiente de correlación o de Pearson, este puede tomar valores en el intervalo [-1,1], si este valor es negativo significa que con inversamente proporcionales y si por el contrario es positivo

entonces son directamente proporcionales. El coeficiente de determinación R^2 nos indica la cantidad que se logra explicar a través del modelo realizado, si este valor se multiplica por 100 nos indicara el porcentaje de explicación por parte de las variables de entrada.

El coeficiente de determinación R^2 ajustado es la correlación que existe incluyendo los grados de libertad presentes en el modelo, de este modo se tiene la precaución que si aumenten las variables explicativas, el valor del coeficiente y del error típico disminuirá si estas no hacen mucho peso dentro del modelo. De esta manera si al aumentar el número de variables explicativas no resulta una mejora en el modelo tendríamos una alerta. El Valor de F y el Valor crítico de F nos dan la seguridad de aceptar el modelo, siempre y cuando el Valor crítico de F sea menor de 0,05. La tercera tabla de resultados que entrega Excel son los coeficientes del modelo buscado.

	<i>Coefficientes</i>	<i>Error Típico</i>	<i>Estadístico T</i>	<i>Probabilidad</i>	<i>Inferior 95%</i>	<i>Superior 95%</i>
Intercepción	19988,0009	3326,7138	6,0083	0,0018	11436,4108	28539,5910
Var 1	0,3686	0,1444	2,5523	0,0511	-0,0026	0,7398
Var 2	0,1753	0,1096	1,5994	0,1706	-0,1065	0,4571
Var 3	-0,1579	0,0778	-2,0302	0,0981	-0,3578	0,0420
Var 4	0,9976	0,4546	2,1945	0,0796	-0,1710	2,1661

Tabla 8. Resultados de la Regresión Entregada por Excel (3)

Con el resultado obtenido en la tabla 6 entenderemos que nuestro modelo es:

$$Y = 19988,0009 + 0,3686 * X_1 + 0,1753 * X_2 - 0,1579 * X_3 + 0,9976 * X_4$$

Al entregarnos este modelo el software realiza una prueba de significancia o prueba t para tener un criterio de decisión. En el que debemos entender que si tomamos el modelo según los resultados dados por la tabla 4 estaríamos dispuestos a correr el riesgo de un 5%, valor usado tradicionalmente; pero a la vez nos calcula la probabilidad de falla de los coeficientes, estos están dados por el valor que indica la columna probabilidad. En el ejemplo que se muestra la variable 2, Var 2, posee un valor de probabilidad del 17,06% es decir una tasa superior de error al 5% que comúnmente se espera; de aquí se concluye que el modelo es bueno pero hay un riesgo a que el modelo falle por culpa a ese coeficiente.

2.9. Adquisición de Datos

En esta sección del trabajo se presentan los datos de entrada y salida usados para generar los modelos de regresiones. Como ha indicado con antelación fueron la población, el producto interno bruto neto y sectorial y la demanda de energía eléctrica.

Demanda de Energía Eléctrica, DEE.

Los datos de la demanda de energía eléctrica fueron tomados de la página oficial de XM [38], los datos están dados en Giga-Vatios. *Ver Anexo 1.*

Población

Los datos de la población y sus respectivos pronósticos fueron adquiridos de la página oficial del Departamento Administrativo Nacional de Estadística (DANE), que es la entidad responsable de la planeación, levantamiento, procesamiento, análisis y difusión de las estadísticas oficiales de Colombia [13]. *Ver Anexo 2.*

Producto Interno Bruto, PIB.

Los datos del producto interno bruto anual a precios constantes de 2005 fueron datos extraídos del banco de la república [05], estos datos están dados en miles de millones de pesos. *Ver Anexo 3.*

3. MODELO DE ENERGÍA ELÉCTRICA

Para la realización de los modelos se usaron los softwares Excel y Statgraphics, usando series de datos que van desde el 2000 hasta el 2009, y se validará su capacidad de pronóstico con los valores de la DEE de los años 2010 y 2011.

Se compararon tres clases de modelos, el primero un modelo lineal a partir del PIB, el segundo un modelo de dos variables asociado a la población y el PIB; y el tercero es un modelo multi-variable a partir de los PIB de los diferentes sectores clasificados por el Banco de la República de Colombia. A continuación se presentaran los resultados obtenidos y la comparación entre estos modelos.

3.1. Modelo Lineal

Este modelo es el usado comúnmente para generar pronósticos de mayor rapidez, debido a la hipótesis sobre la relación de *dualidad que hay entre el consumo energético y el PIB de una nación* [10]. Este modelo es el más aceptado debido a su simplicidad y facilidad en su manejo [01]. A continuación presentamos los resultados obtenidos con este modelo sobre los datos recopilados, ver cap 2.9.

ESTADÍSTICAS DE LA REGRESIÓN	
Coefficiente de correlación múltiple	0,995814947
Coefficiente de determinación R²	0,99164741
R² ajustado	0,990603336
Error típico	438,2714858
Observaciones	10

Tabla 9. Estadísticas de la Regresión del Modelo Lineal.

ANÁLISIS DE VARIANZA					
	<i>Grados de libertad</i>	<i>Suma de cuadrados</i>	<i>Promedio de los cuadrados</i>	<i>F</i>	<i>Valor crítico de F</i>
Regresión	1	182436828,8	182436828,8	949,7866971	1,33536E-09
Residuos	8	1536655,162	192081,8952		
Total	9	183973484			

Tabla 10. Análisis de Varianza del Modelo Lineal.

Según los resultados obtenidos en la tabla 7 y 8, el modelo es de alta calidad, debido a que los coeficientes de correlación son altos, tomando valores aproximados al 99%.

Esto además nos indica que el PIB es una variable de mucho peso para la predicción de la demanda de energía eléctrica. Por otro lado el análisis de varianza nos confirma que podemos aceptar el modelo con un 95% de confianza que no va a fallar.

Los coeficientes entregados por esta regresión son:

	<i>COEFICIENTES</i>	<i>ERROR TÍPICO</i>	<i>ESTADÍSTICO T</i>
Intercepción	15856,80921	1064,339781	14,89825853
PIB	0,095496972	0,003098679	30,81860959

Tabla 11. Coeficientes del Modelo Lineal.

Por la prueba t vemos que los coeficientes entregados son de una gran fiabilidad, así que el modelo propuesto sería:

$$DEE(t) \approx 15856,8092 + 0,0955 * PIB(t)$$

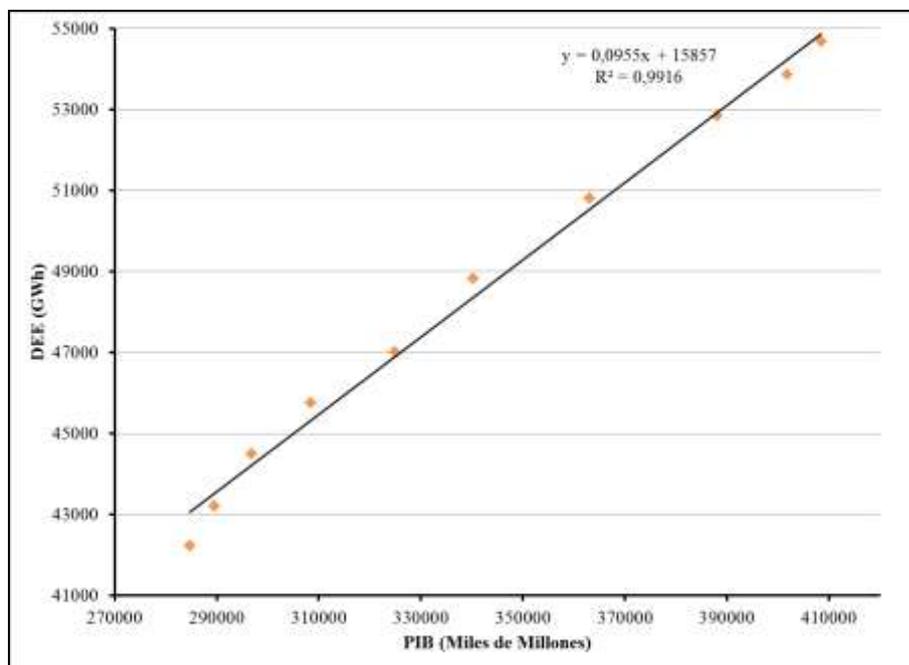


Ilustración 4. Nube de Puntos de la Serie de Datos.

De la ilustración 4 podemos observar que la nube puntos creada a partir de la correlación de las series de datos del PIB y DEE posee una fuerte tendencia lineal, de ahí que sus estadísticas de correlaciones tienen valores altos.

A continuación se presenta la tabla de resultados y el gráfico comparativo entre los datos reales y los pronosticados, los datos del 2010 y 2011 no se tuvieron en cuenta para la generación del modelo y estos se usaran para verificar la calidad del pronóstico:

<i>Año</i>	<i>Real</i>	<i>Pronóstico</i>	<i>Residuo</i>	<i>Residuos Estándares</i>	<i>% de Error</i>
2000	42246	43050,6223	-804,6223	-1,947265	-1,9046
2001	43216	43506,9069	-290,9069	-0,704023	-0,6731
2002	44499	44199,2599	299,7401	0,725401	0,6736
2003	45768	45309,7942	458,2058	1,108903	1,0011
2004	47017	46880,5284	136,4716	0,330275	0,2903
2005	48829	48340,6771	488,3229	1,181790	1,0001
2006	50815	50516,2891	298,7109	0,722910	0,5878
2007	52851	52908,0107	-57,0107	-0,137972	-0,1079
2008	53870	54222,1446	-352,1445	-0,852224	-0,6537
2009	54679	54855,7670	-176,7670	-0,427793	-0,3233
2010	56148	56416,1875	-268,1875		-0,4776
2011	57156	58814,8804	-1658,8804		-2,9024

Tabla 12. Resultados de los Pronósticos del Modelo Lineal.

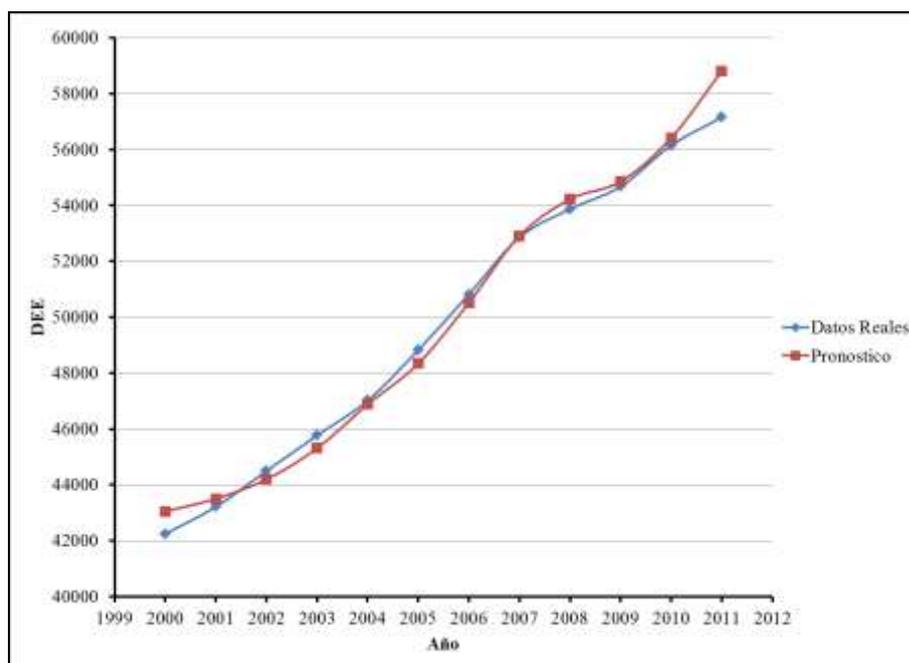


Ilustración 5. Modelamiento de los Pronósticos Vs los Datos Reales.

El modelo pronostico los datos con una tasa inferior al 2%, excepto en el último año donde tuvo su mayor desfase con un valor aproximado al 3%, si bien el modelo no supero el error del 5% de confiabilidad, cuando se realicen proyecciones de mayor periodo de tiempo el modelo tendería a fallar.

De la ilustración 5 podemos notar que el modelo no logra predecir el comportamiento, mostrando un abrupto crecimiento frente al incremento moderado presentado por la serie de datos reales. A partir de esta idea se hará un modelo con los datos para los años del 2000 al 2010, y hallar la tasa de falla para el dato 2011.

<i>ESTADÍSTICAS DE LA REGRESIÓN</i>	<i>2000-2010</i>	<i>2000-2009</i>
Coefficiente de correlación múltiple	0,996674058	0,995814947
Coefficiente de determinación R²	0,993359178	0,99164741
R² ajustado	0,992621309	0,990603336
Error típico	419,8040886	438,2714858
Observaciones	11	10

Tabla 13. Comparación del modelo 2000-2009 y el modelo 2000-2010.

El nuevo modelo generado con los datos del 2000 al 2010 muestra una gran mejoría en las estadísticas de regresión, especialmente al disminuir el error típico; a continuación mostramos una tabla comparativa de los errores porcentuales.

<i>Año</i>	<i>Real</i>	<i>2000-2009</i>	<i>% de error</i>	<i>2000-2010</i>	<i>% de error</i>
2000	42246	43050,6223	-1,9046	43075,4735	-1,9634
2001	43216	43506,9069	-0,6731	43528,0506	-0,7221
2002	44499	44199,2599	0,6736	44214,7780	0,6387
2003	45768	45309,7942	1,0011	45316,2888	0,9870
2004	47017	46880,5284	0,2903	46874,2602	0,3036
2005	48829	48340,6771	1,0001	48322,5446	1,0372
2006	50815	50516,2891	0,5878	50480,4790	0,6583
2007	52851	52908,0107	-0,1079	52852,7670	-0,0033
2008	53870	54222,1446	-0,6537	54156,2230	-0,5313
2009	54679	54855,7670	-0,3233	54784,6970	-0,1933
2010	56148	56416,1875	-0,4776	56332,4385	-0,3285
2011	57156	58814,8804	-2,9024	58711,6412	-2,7217

Tabla 14. Comparación de los Errores Porcentuales.

El error en el año 2011 disminuyó en aproximadamente dos décimas, podemos decir a favor del modelo que este error es provocado por un factor externo, como se había comentado anteriormente debido a “*La alta contribución en el PIB de ciertas actividades que tienen un menor consumo de electricidad en sus procesos productivos*”.

Se concluye que si la demanda sigue este comportamiento de desacople, el modelo que acabamos de presentar sería inútil dentro de un plazo de tiempo, si bien sus errores

presentados son menores del 5%, que pasaría si el próximo año hay un desacople del 7%, de aquí la necesidad de investigar un nuevo modelo de pronóstico anual.

3.2. Modelo de dos Variables

En el siguiente modelo se usaron dos variables, la primera el PIB, manteniendo la hipótesis usada en el primer modelo, y la segunda variable es la población neta de Colombia, tomándola bajo la hipótesis que el aumento de la población daría una mayor estabilidad en el modelo, debido a que si aumenta la población habrá un aumento de la demanda per cápita de Energía Eléctrica.

A continuación se presentan los datos obtenidos, los datos fueron ingresados como PIB en miles de millones y la población neta en millones.

<i>ESTADÍSTICAS DE LA REGRESIÓN</i>	<i>2 Variables</i>	<i>1 Variable</i>
Coefficiente de correlación múltiple	0,998977	0,995815
Coefficiente de determinación R²	0,997956	0,991647
R² ajustado	0,997371	0,990603
Error típico	231,792836	438,271486
Observaciones	10	10

Tabla 15. Comparación de los Modelos de 2 y 1 Variable.

ANÁLISIS DE VARIANZA					
	<i>Grados de libertad</i>	<i>Suma de cuadrados</i>	<i>Promedio de los cuadrados</i>	<i>F</i>	<i>Valor crítico de F</i>
Regresión	2	183597388,6	91798694,28	1708,584592	3,86278E-10
Residuos	7	376095,4319	53727,91884		
Total	9	183973484			

Tabla 16. Análisis de Varianza del Modelo de Dos Variables.

Según los resultados obtenidos y comparándolo con el modelo lineal inicial, es de mejor calidad, no solo debido a que los coeficientes de correlación, estos tienen valores un poco más superiores a los anteriores, sino que el mismo error típico mejoro en un 47%. Esto además indica que el PIB es una variable de mucho peso para la predicción de la demanda de energía eléctrica. Y al igual que en modelo pasado el análisis de varianza confirma que es aceptado. Los coeficientes entregados por esta regresión están dados por:

	<i>COEFICIENTES</i>	<i>ERROR TÍPICO</i>	<i>ESTADÍSTICO T</i>
Intercepción	-27001,30686	9238,617465	-2,92265666
Población	1364,079363	293,498509	4,647653468
PIB	0,050589641	0,009800361	5,162018062

Tabla 17. Coeficientes del Modelo de Dos Variables.

A partir de la prueba t se aceptan los coeficientes, se debe notar que los valores de probabilidad aumentaron considerablemente en comparación del modelo lineal. Considerando los coeficientes obtenidos, el modelo sería:

$$DEE(t) \approx 15856,8092 + 0,0506 * PIB(t) + 1364,0794 * Población$$

Recuérdese que los datos deben ser ingresados en la siguiente forma, el PIB en miles de millones y la población neta en millones. Notar que el coeficiente de la variable población es mayor que el del PIB, esto se debe simplemente a que los valores del PIB son del orden de los cientos de miles, mientras que la población no llega a los cientos. Si ambas variables estuvieran en la misma potencia los coeficientes entregados serían del mismo orden.

A continuación se presenta la tabla de resultados:

<i>Año</i>	<i>Real</i>	<i>2 Var</i>	<i>Residuos</i>	<i>Residuos Estándares</i>	<i>% Error 2 Var</i>	<i>1 Var</i>	<i>% Error 1 Var</i>
2000	42246	42370,9959	-124,9959	-0,6115	-0,2959	43050,6223	-1,9046
2001	43216	43319,2763	-103,2763	-0,5052	-0,2390	43506,9069	-0,6731
2002	44499	44388,9381	110,0619	0,5384	0,2473	44199,2599	0,6736
2003	45768	45686,7505	81,2495	0,3975	0,1775	45309,7942	1,0011
2004	47017	47227,5291	-210,5291	-1,0299	-0,4478	46880,5284	0,2903
2005	48829	48710,5064	118,4936	0,5797	0,2427	48340,6771	1,0001
2006	50815	50568,7652	246,2348	1,2045	0,4846	50516,2891	0,5878
2007	52851	52546,4313	304,5687	1,4899	0,5763	52908,0107	-0,1079
2008	53870	53957,6703	-87,6703	-0,4289	-0,1627	54222,1446	-0,6537
2009	54679	55013,1368	-334,1368	-1,6345	-0,6111	54855,7670	-0,3233
2010	56148	56563,7594			-0,7405	56416,1875	-0,4776
2011	57156	58564,2757			-2,4639	58814,8804	-2,9024

Tabla 18. Comparación de Resultados entre el Modelo de 2 Variables y de 1 Variable

De los resultados se resalta que:

Error	1 Var	2 Var
Negativo	-2,9024	-2,4639
Positivo	1,0011	0,5763

Tabla 19. Comparación de los Errores Máximos del Modelo de 1 y 2 Variables.

Se concluye que el modelo con la segunda variable mejora los pronósticos en los diferentes aspectos, en no subestimar y no sobreestimar la demanda ya sea en correlación y en pronóstico. Además debe notar que este nuevo modelo sigue teniendo la desviación en el año 2011, aunque disminuyó en un 0.3% al error pasado, dando más peso a la afirmación que el PIB ya no estaría siendo una buena variable para generar el modelo de predicción. A continuación se presenta la ilustración 6, donde se muestra un gráfico comparativo entre los datos reales y los pronosticados.

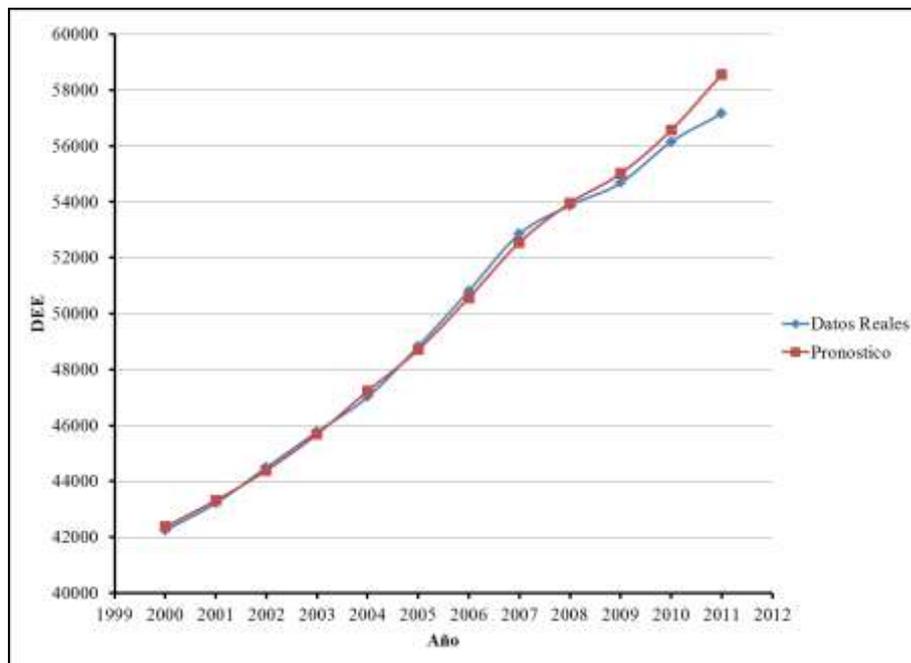


Ilustración 6. Modelamiento de los Pronósticos Vs los Datos Reales.

Podemos observar que al igual que en la ilustración 5, se presenta el mismo error, el modelo toma un crecimiento superior al presentado por los datos reales. De aquí podemos concluir que no se puede realizar pronósticos con ayuda de un PIB global; si bien este modelo no está presentando errores superiores al 5%, si presenta un desfase para los años próximos, provocando una gran incertidumbre en los pronósticos que se lleguen a realizar con esta variable.

3.3. Modelo Multi-Variable

Este modelo toma como variables a los diferentes sectores de la economía, que se describieron anteriormente, se basa en la hipótesis que si sectorizamos el PIB total se lograría mejorar la correlación y los pronósticos debido a que hay sectores que dan mayor y mejor uso a la Energía Eléctrica. A continuación se presenta el grafico de nube de puntos correspondiente a cada sector:

Sector	R ²
Minería	0,6141
Servicios Públicos	0,9906
Construcción	0,9905
Transporte	0,9872
Agricultura	0,9444
Manufactura	0,9525
Comercio	0,9885
Financieros	0,9861
Sociales	0,9705

Tabla 20. Correlación R² de los Sectores Económicos.

De los coeficientes podemos observar, que solo el sector de la minería se desacopla con la DEE; el sector transporte tiene una alta correlación, pero a su vez no se logra entender una hipótesis para relacionarla con la demanda de energía eléctrica a diferencia de los demás sectores donde se puede ver una clara relación debido a las actividades que representan cada uno.

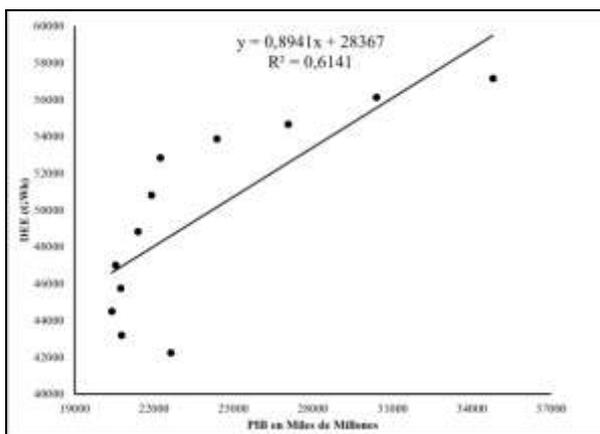


Ilustración 7. Correlación de Minería Vs DEE.

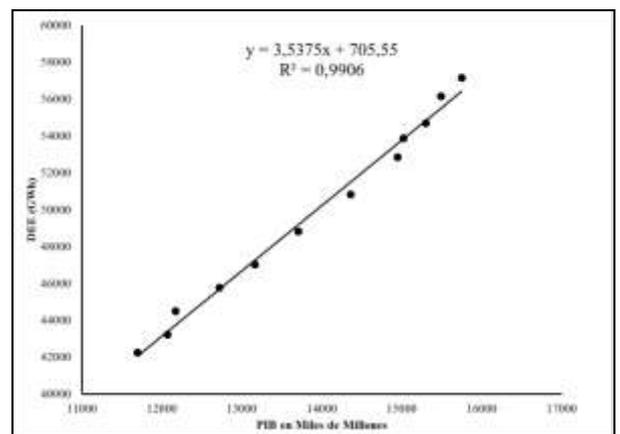


Ilustración 8. Correlación de Servicios Públicos Vs DEE.

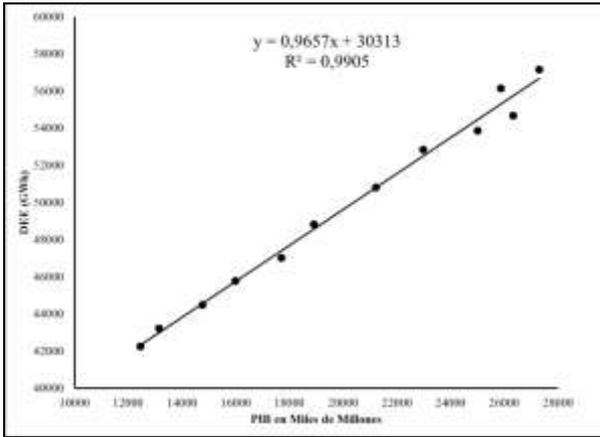


Ilustración 9. Correlación de Construcción Vs DEE.

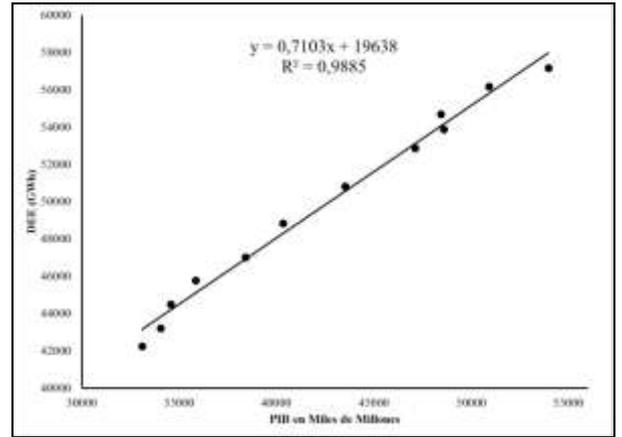


Ilustración 12. Correlación de Comercio Vs DEE.

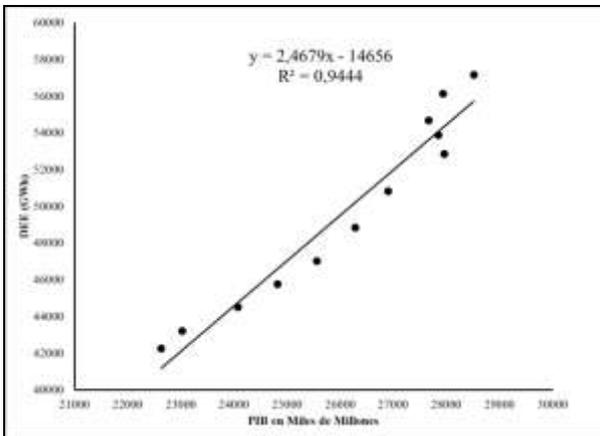


Ilustración 10. Correlación de Agricultura Vs DEE.

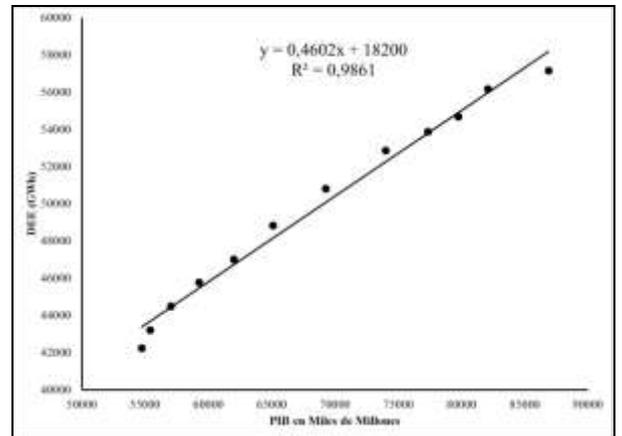


Ilustración 13. Correlación de Financiero Vs DEE.

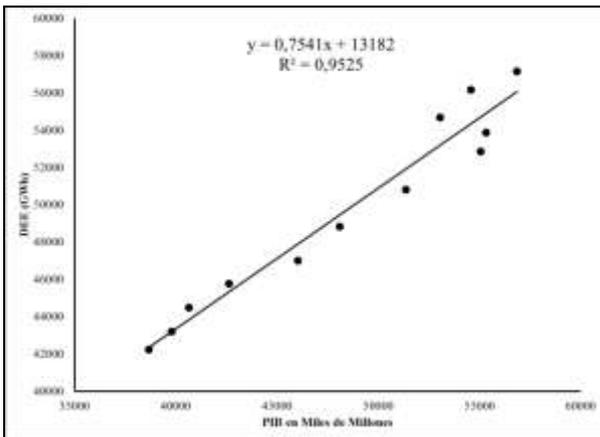


Ilustración 11. Correlación de Manufactura Vs DEE.

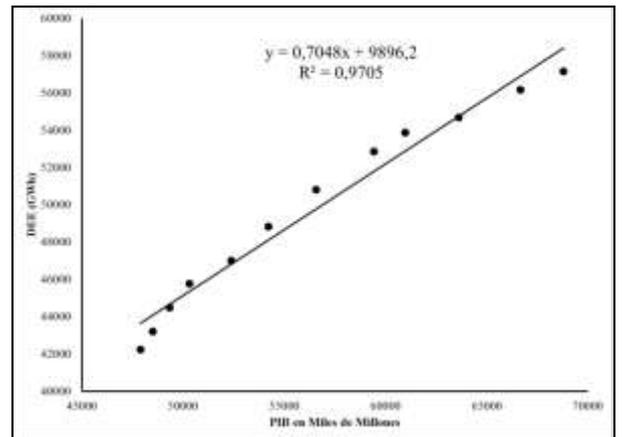


Ilustración 14. Correlación de Sociales Vs DEE.

Para generar los modelos se pueden usar cierto grupo de variables empezando desde 1 hasta llegar a las 9 que se han expuesto, pero tomar 1 no cumpliría el objetivo de generar un modelo multi-variable, y escoger grupo de 6 a 9 variables no generaría un modelo confiable debido a la baja cantidad de datos que se tienen, es decir las variables serían iguales a las ecuaciones iniciales que usamos para generar el modelo, así que no existiría ningún error o más bien la correlación sería perfecta, de aquí se decidió trabajar con un máximo de 5 variables. Las series de datos van desde el año 2000 al 2009, y usaremos los datos del 2010 y 2011 para realizar los análisis comparativos.

Letra	Sector
A	Agricultura
B	Comercio
C	Construcción
D	Financieros
E	Manufactura
F	Minería
G	Sociales
H	Servicios públicos

Tabla 21. Nomenclatura para los Sectores.

Usando el software de análisis estadístico Statgraphic se generaron todas las combinaciones posibles para las 8 variables, se excluyó transporte debido a la poca claridad en la relación que habría con la DEE. En la tabla 19 se presenta la nomenclatura que se utilizó para en la columna variable de los resultados del *Anexo 4*, en la cual se indica que variables se utilizan para generar el modelo número n, además de los coeficientes de correlación, de correlación ajustada y de Cp Mallows.

En total se generaron 219 modelos, empezando con uno que no poseyera ninguna variable hasta las combinaciones posibles con grupo de 5. Luego estos resultados pasaron por un filtro donde primero se comparó el Cp contra el número de variables usadas, recordemos que $Cp \approx \#variables + 1$, y luego el siguiente filtro fue su R² ajustada; escogimos los 10 modelos que tuvieron las mejores estadísticas.

Para comparar el Cp se realizó la resta en valor absoluto entre Cp y el número de variables más uno; este resultado se muestra en la columna 7 de la tabla 24. Resaltemos que no importa el signo, sino que el valor del Cp sea próximo al j+1.

Puesto	Modelo	R-cuadrada	Ajustada	Cp	Incluidas	Valor Absoluto de Cp-(j+1)
1	121	99,9365	99,8856	5,0408	ADEH	0,0408
2	181	99,9507	99,8890	5,9138	ABEFH	0,0862
3	112	99,9358	99,8845	5,0897	ACDH	0,0897
4	176	99,9524	99,8930	5,7730	ABDEH	0,2269
5	153	99,9311	99,8760	5,4656	CDFH	0,4656
6	188	99,9433	99,8725	6,4967	ACDFH	0,4967
7	182	99,9561	99,9011	5,4863	ABEGH	0,5137
8	203	99,9571	99,9035	5,4014	BCDFH	0,5986
9	196	99,9420	99,8694	6,6044	ADEGH	0,6044
10	186	99,9417	99,8689	6,6221	ACDEH	0,6220

Tabla 22. Selección de los 10 Modelos con las Mejores Estadísticas.

De los resultados presentados en la tabla 20 podemos ver que los diez mejores modelos, presentan una diferencia entre el valor del coeficiente de Mallows y el número de variables incluidas más 1 desde 0,04 hasta 0,62; en algunos caso se sobrepasó el valor de 1, para este caso solamente se examinaran los tres primeros.

MODELO	VARIABLE							
	A	B	C	D	E	F	G	H
1	1			1	1			1
2	1	1			1	1		1
3	1		1	1				1
4	1	1		1	1			1
5			1	1		1		1
6	1		1	1		1		1
7	1	1			1		1	1
8		1	1	1		1		1
9	1			1	1		1	1
10	1		1	1			1	1
Repeticiones	8	4	5	8	5	4	3	10

Tabla 23. Conteo de Repetición de las Variables.

Analizando las variables incluidas, columna 6 tabla 20, se puede notar un resultado singular; que hay una repetición muy seguida de un grupo de variables. Al realizar un conteo, tabla 21, tomando solamente los 10 modelos considerados anteriormente tenemos los resultados indican que el sector de servicios públicos es una variable

fundamental, luego le siguen el sector agricultura y financieros. Si revisamos las estadísticas del modelo ADH sus resultados fueron:

Modelo	R²	R² Ajustada	Cp	Variables
52	99,9067	99,8601	5,4007	ADH

Tabla 24. Resultados del Modelo ADH.

Si bien el modelo tuvo excelentes estadísticas de regresión, cuando se calcula el desvío del Cp nos da que es 1,4; un valor muy alto para los modelos considerados como mejores. Solo por la casualidad dada por el conteo de la repetición de variables este modelo será tenido en cuenta, con los tres primeros modelos; los cuales obtuvieron un desvío de Cp inferiores al 0,09. A continuación se comparara las estadísticas de regresión de los 4 modelos elegidos:

ESTADÍSTICAS DE LA REGRESIÓN	ADH	ADEH	ABEFH	ACDH
Coefficiente de correlación múltiple	0,999533	0,999682	0,999753	0,999679
Coefficiente de determinación R²	0,999067	0,999365	0,999507	0,999358
R² ajustado	0,998601	0,998856	0,998890	0,998845
Error típico	169,134081	152,909923	150,638491	153,649746
Observaciones	10	10	10	10

Tabla 25. Comparación de los Modelos Múltiples.

Se puede decir que los cuatro modelo son aptos en cuestión de regresión, ya que los valores de correlación son superiores al 99,8%, se mira el R² ajustado ya que esta variable no tiene en cuenta cuantas variable utiliza el modelo. A continuación se presenta los análisis de varianza uno de los modelos escogidos.

3.3.1. Modelo ADH

Para el modelo que fue incluido debido a la repetitividad de las variables explicativas, obtuvimos los siguientes resultados.

ANÁLISIS DE VARIANZA					
	<i>Grados de libertad</i>	<i>Suma de cuadrados</i>	<i>Promedio de los cuadrados</i>	<i>F</i>	<i>Valor crítico de F</i>
Regresión	3	183801846	61267281,9900	2141,73808	1,77571E-09
Residuos	6	171638,024	28606,3373		
Total	9	183973484			

Tabla 26. Análisis de Varianza para el Modelo ADH

	<i>Coefficientes</i>	<i>Error Típico</i>	<i>Estadístico T</i>
Intercepción	6290,927851	1214,347865	5,180499
Agricultura (A)	0,623878	0,179021	3,484944
Financieros (D)	0,266900	0,046234	5,772848
S. Publico (H)	0,637406	0,512416	1,243923

Tabla 27. Coeficientes para el Modelo ADH

Según los resultados, se puede aceptar el modelo de regresión múltiple por la prueba valor F, el cual daría de la siguiente manera:

$$Y = 6290,927851 + 0,623878 * A + 0,266900 * D + 0,637406 * H$$

3.3.2. Modelo ADEH

Este modelo ocupó el primer lugar en la desviación del Cp, está conformado por la serie de datos de agricultura, financieros, manufactura y servicios públicos. Pertenece a la familia del modelo pasado, simplemente agregando la variable manufactura. Los resultados fueron los siguientes:

ANÁLISIS DE VARIANZA					
	<i>Grados de libertad</i>	<i>Suma de cuadrados</i>	<i>Promedio de los cuadrados</i>	<i>F</i>	<i>Valor crítico de F</i>
Regresión	4	183856576,8000	45964144,1900	1965,838506	3,56112E-08
Residuos	5	116907,2232	23381,44463		
Total	9	183973484			

Tabla 28. Análisis de Varianza del Modelo ADEH.

	<i>Coefficientes</i>	<i>Error Típico</i>	<i>Estadístico T</i>
Intercepción	3975,402240	1869,717850	2,126204
Agricultura (A)	0,734549	0,177278	4,143494
Financieros (D)	0,260440	0,042011	6,199259
Manufactura (E)	-0,087071	0,056911	-1,529961
S. Publico (H)	0,932912	0,501914	1,858710

Tabla 29. Coeficientes del Modelo ADEH.

Según los resultados se puede aceptar el modelo de regresión múltiple por la prueba valor F, el cual daría de la siguiente manera:

$$Y = 3975,402240 + 0,734549 * A + 0,260440 * D - 0,087071 * E + 0,932912 * H$$

3.3.3. Modelo ABEFH

El modelo ABEFH posee 5 variables, estos datos son los de agricultura, comercio, manufactura, minería y servicios públicos.

ANÁLISIS DE VARIANZA					
	<i>Grados de libertad</i>	<i>Suma de cuadrados</i>	<i>Promedio de los cuadrados</i>	<i>F</i>	<i>Valor crítico de F</i>
Regresión	5	183882716,2000	36776543,2400	1620,686413	1,06443E-06
Residuos	4	90767,8202	22691,95505		
Total	9	183973484			

Tabla 30. Análisis de Varianza del Modelo ABEFH.

	<i>Coefficientes</i>	<i>Error Típico</i>	<i>Estadístico T</i>
Intercepción	1410,584901	2555,820106	0,551911
Agricultura (A)	1,116895	0,218546	5,110571
Comercio (B)	0,594376	0,144391	4,116425
Manufactura (E)	-0,425298	0,115533	-3,681193
Minería (F)	0,000748	0,087984	0,008499
S. Publico (H)	1,056255	0,483903	2,182781

Tabla 31. Coeficientes del Modelo ABEFH.

Según los resultados se puede aceptar el modelo de regresión múltiple por la prueba valor F, el cual daría de la siguiente manera:

$$Y = 1410,584901 + 1,116895 * A + 0,594376 * B - 0,425298 * E + 0,000748 * F + 1,056255 * H$$

Según el coeficiente entregado por minería, este valor tiende a no ser representativo, de ahí que su probabilidad de error sea muy alta, sería interesante mirar que resultados nos da si no incluimos a esta variable en la ecuación, es decir tener el modelo de la siguiente manera:

$$Y = 1410,584901 + 1,116895 * A + 0,594376 * B - 0,425298 * E + 1,056255 * H$$

3.3.4. Modelo ACDH

El modelo ACDH, quien ocupo la tercera plaza usa las variables agricultura, construcción, financieros y servicios públicos.

ANÁLISIS DE VARIANZA					
	<i>Grados de libertad</i>	<i>Suma de cuadrados</i>	<i>Promedio de los cuadrados</i>	<i>F</i>	<i>Valor crítico de F</i>
Regresión	4	183855442,8000	45963860,6900	1946,941073	3,64809E-08
Residuos	5	118041,2220	23608,2444		
Total	9	183973484			

Tabla 32. Análisis de Varianza del Modelo ACDH.

	<i>Coefficientes</i>	<i>Error Típico</i>	<i>Estadístico T</i>
Intercepción	11552,677054	3662,251285	3,154529
Agricultura (A)	0,442808	0,202214	2,189794
Construcción (C)	0,247624	0,164344	1,506737
Financieros (D)	0,143734	0,091903	1,563982
S. Publico (H)	0,842726	0,485039	1,737441

Tabla 33. Coeficientes del Modelo ACDH.

Según los resultados se puede aceptar el modelo de regresión múltiple por la prueba valor F, el cual daría de la siguiente manera:

$$Y = 11552,677054 + 0,442808 * A + 0,247624 * C + 0,143734 * D + 0,842726 * H$$

3.4. Comparación en los Pronósticos.

Para comparar los pronósticos que generan los modelos, se calculara y analizara el error porcentual de cada uno de ellos. A continuación en la tabla 32 encontraremos los valores de pronósticos generado con los modelos elegidos anteriormente, 5 en total, y los datos reales del sistema; y en la tabla 33, los residuos de los pronósticos..

Año	Real	ADH	ADEH	ABEFH	ABEFH - F	ACDH
2000	42246	42469,1709	42394,4522	42276,4675	42267,7208	42378,6435
2001	43216	43133,3256	43110,4656	43210,5231	43201,4982	43134,4666
2002	44499	44292,9805	44331,1672	44428,9895	44419,8890	44322,5317
2003	45768	45706,8499	45802,7530	45762,6376	45753,1244	45742,4695
2004	47017	47177,2584	47172,4177	47124,1169	47114,2747	47258,2123
2005	48829	48797,5038	48834,0183	48775,8262	48765,5757	48781,6415
2006	50815	50720,8802	50707,1160	50660,9708	50650,2327	50778,6852
2007	52851	53011,3923	52935,0289	53024,5937	53013,4167	52855,8148
2008	53870	53886,9013	53774,6830	53744,5163	53733,2839	53854,9808
2009	54679	54593,7372	54727,8983	54716,4821	54705,0396	54682,5541
2010	56148	55496,5343	55567,7223	56030,2762	56018,6917	55178,6417
2011	57156	57303,6571	57286,3099	57800,0685	57788,2925	56694,9106

Tabla 34. Pronostico de los Modelos Multi-Variable.

Año	ADH	ADEH	ABEFH	ABEFH - F	ACDH
2000	-223,17	-148,45	-30,47	-21,72	-132,64
2001	82,67	105,53	5,48	14,50	81,53
2002	206,02	167,83	70,01	79,11	176,47
2003	61,15	-34,75	5,36	14,88	25,53
2004	-160,26	-155,42	-107,12	-97,27	-241,21
2005	31,50	-5,02	53,17	63,42	47,36
2006	94,12	107,88	154,03	164,77	36,31
2007	-160,39	-84,03	-173,59	-162,42	-4,81
2008	-16,90	95,32	125,48	136,72	15,02
2009	85,26	-48,90	-37,48	-26,04	-3,55
2010	651,47	580,28	117,72	129,31	969,36
2011	-147,66	-130,31	-644,07	-632,29	461,09

Tabla 35. Residuos de los Pronósticos.

En la tabla 34 calculamos los errores porcentuales:

Año	Real	ADH	ADEH	ABEFH	ABEFH - F	ACDH
2000	42246	-0,53%	-0,35%	-0,07%	-0,05%	-0,31%
2001	43216	0,19%	0,24%	0,01%	0,03%	0,19%
2002	44499	0,46%	0,38%	0,16%	0,18%	0,40%
2003	45768	0,13%	-0,08%	0,01%	0,03%	0,06%
2004	47017	-0,34%	-0,33%	-0,23%	-0,21%	-0,51%
2005	48829	0,06%	-0,01%	0,11%	0,13%	0,10%
2006	50815	0,19%	0,21%	0,30%	0,32%	0,07%
2007	52851	-0,30%	-0,16%	-0,33%	-0,31%	-0,01%
2008	53870	-0,03%	0,18%	0,23%	0,25%	0,03%
2009	54679	0,16%	-0,09%	-0,07%	-0,05%	-0,01%
2010	56148	1,16%	1,03%	0,21%	0,23%	1,73%
2011	57156	-0,26%	-0,23%	-1,13%	-1,11%	0,81%

Tabla 36. Error Porcentual de los Pronósticos

En la tabla 35 los errores usados para evaluar los modelos de pronóstico:

Error	ADH	ADEH	ABEFH	ABEFH - F	ACDH
MAD	160,0474	138,6437	126,9991	128,5374	182,9081
MAPE	0,32%	0,27%	0,24%	0,24%	0,35%
PMAD	0,32%	0,28%	0,26%	0,26%	0,37%
MSE	51487,3564	39217,5040	43323,6619	42504,8494	105858,3462
RMSE	226,9083	198,0341	208,1434	206,1670	325,3588

Tabla 37. Errores de los Modelos Diseñados.

Según los resultados obtenidos tenemos:

- ❖ Que el modelo ACDH ocupó las peores estadísticas de error y el modelo ADH siendo un modelo de tres variables ocupó puestos intermedios y con estadísticas dentro de lo esperado.
- ❖ El modelo ADEH y ABEFH tuvieron buenos lugares, era de esperarse debido a que el modelo ABEFH tiene 5 variables, teniendo así una variable que explica un porcentaje más del modelo. Pero comparando ADEH y ABEFH-F, este segundo tiene mejores estadísticas estando en la misma condición de 4 variables.
- ❖ Ninguno de los modelos tuvo errores superiores al 1%, dando un punto muy positivo a todos los modelos.

En las desviaciones presentadas en los modelos, se deben revisar dos aspectos, el primero cuanto fue el error máximo positivo y el máximo negativo, con el fin de establecer qué modelo tiene mayor sobre estimación y sub estimación de la DEE. Y segundo hallar quien presenta la peor y la mejor desviación. Para esto se realizaron dos tablas, en la número 36 y la 37 encontraremos las dos especificaciones dichas anteriormente:

ERROR	ADH	ADEH	ABEFH	ABEFH-F	ACDH
Positivo	1,16%	1,03%	0,30%	0,32%	1,73%
Negativo	-0,53%	-0,35%	-1,13%	-1,11%	-0,51%

Tabla 38. Error Máximo Negativo y Positivo.

ERROR	ADH	ADEH	ABEFH	ABEFH-F	ACDH
Máximo	1,16%	1,03%	1,13%	1,11%	1,73%
Mínimo	0,03%	0,01%	0,01%	0,03%	0,01%

Tabla 39. Máximo y Mínimo Error Absoluto.

Recordemos que una desviación positiva indica que está subestimando la demanda, si por el contrario es negativa esta la sobreestimo. De la tabla anterior podemos decir:

- ✓ La peor desviación positiva fue por el modelo ACDH, en el año 2010, es decir en uno de los datos de validación, es decir que este modelo hubiera puesto en riesgo el sistema debido a que esperaba menos de lo que se consumió realmente.
- ✓ La peor desviación negativa se presentó en el modelo ABEFH y estuvo seguida del ABEFH-F, aunque este error de sobre estimación no es tan grave, se espera que no se tenga grandes errores.

Si realizamos una calificación entregando 5 puntos al mejor pronóstico y 1 al peor, y dando peso de un 60% a la sub-estimación y 40% a la sobre-estimación

ERROR	ADH	ADEH	ABEFH	ABEFH-F	ACDH
Positivo	2	3	5	4	1
Negativo	3	5	1	2	4
Puntuación Neta	2,4	3,8	3,4	3,2	2,2

Tabla 40. Puntuación para Máximo Error Positivo y Negativo.

Los modelos quedarían organizados como ADEH, ABEFH, ABEFH-F, ADH y por ultimo ACDH.

- ✓ Con los resultados obtenidos anteriormente el ADEH y ABEFH fueron los que tuvieron menores errores, ahora el ADEH está en mejor ubicación y presentando solamente 4 variables, y podría ser considerado como el mejor modelo.

Si realizamos el mismo tipo de evaluación pero evaluando la mayor y la menor desviación; esta vez dando un porcentaje de 60% a la mayor desviación y 40% a la menor desviación, tendríamos lo siguiente.

ERROR	ADH	ADEH	ABEFH	ABEFH-F	ACDH
Máximo	2	5	3	4	1
Mínimo	4	5	5	4	5
Puntuación Neta	2,8	5	3,8	4	2,6

Tabla 41. Puntuación para la Mayor y Menor Desviación Porcentual.

Los modelos quedarían organizados como ADEH, ABEFH-F, ABEFH, ADH y por ultimo ACDH. Con los resultados obtenidos anteriormente se debe:

- ✓ Descartar al modelo ACDH por estar entre el último lugar, en las diferentes pruebas.
- ✓ Al modelo ADH y ABEFH se descartan por no tener 4 variables, vemos que 4 variables sería un numero óptimo para la cantidad de datos.
- ✓ Con las pruebas realizadas y con los resultados obtenidos se tiene que el modelo ADEH es el mejor modelo multi-variable que se puede obtener, cumpliendo con valor Cp, estadísticas de regresión óptimas y ocupa los primeros lugares en los estudios realizados en las desviaciones individuales.

4. PRONÓSTICOS DEL MODELO MULTI-VARIABLE

Los pronósticos se realizan en intervalo, de esta forma se puede asegurar que se cumple cualquier posible combinación de eventos; para tener un escenario bajo, medio y alto debemos realizar un intervalo de confianza, para cada sector, donde el límite inferior, la media y el límite superior serían las tasas de crecimiento para los escenarios respectivamente mencionados.

Debido a la poca cantidad de datos que se tienen, no se puede realizar un intervalo de confianza muy certero, debido especialmente a la aleatoriedad del PIB y más a los cambios negativos que se presentaron en algunos de los sectores. Este último se presenta de manera inesperada y por lo general el peor escenario sería uno donde no hubiese cambio, es decir una tasa cero. A continuación presentamos dos ejemplos de los gráficos de densidad para el sector de la Agricultura y la Minería:

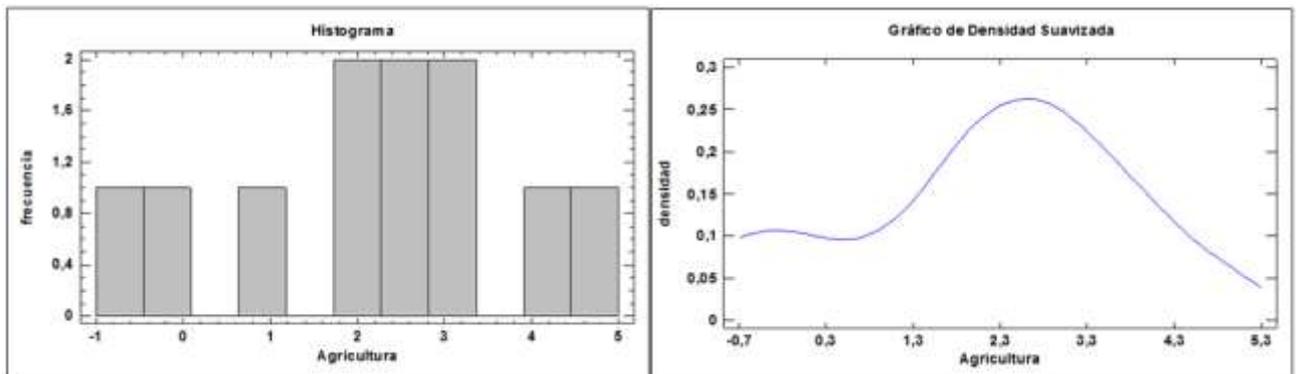


Ilustración 15. Histograma y Gráfico de Densidad para el Sector de Agricultura.

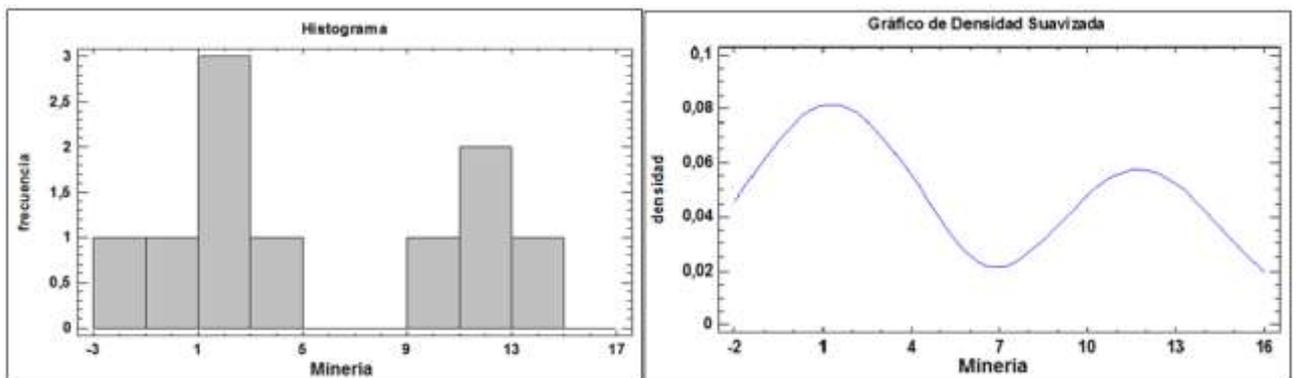


Ilustración 16. Histograma y Gráfico de Densidad para el Sector de Minería.

De lo anterior podemos notar que si bien en casos como el de la agricultura se presenta una densidad de forma tipo campana, el cual facilita el diseño de intervalo de confianza, está el caso donde no hay una forma definida y no genera una densidad

uniforme. Entonces se optó por dos vías la primera generar intervalos excluyendo las tasas negativas, corriendo el riesgo de tener menos datos, y el segundo suponer que el crecimiento del PIB no importa el sector será igual al crecimiento del PIB global.

4.1. Intervalos por Sector

Se calculará un intervalo de crecimiento para cada sector de la economía descrito, con un 95% de confianza, se retiraran los crecimientos negativos, para así tener tasas de crecimiento y no de decrecimiento. Se presenta a continuación la nomenclatura de los sectores y las tasas de crecimiento individuales:

- | | | |
|-----------------|------------------|-----------------|
| a) Agricultura | d) Servicios | g) Transporte. |
| b) Minería. | Públicos. | h) Financieros. |
| c) Manufactura. | e) Construcción. | i) Sociales. |
| | f) Comercio. | |

Escenario	A	B	C	D	E	F	G	H	I
Bajo	1,0361	1,2423	1,3171	1,6966	4,8003	2,7670	3,2607	3,1662	2,4791
Medio	2,1347	5,4487	3,6182	2,7513	7,4643	4,5766	5,6816	4,3045	3,3540
Alto	3,2333	9,6551	5,9192	3,8059	10,1283	6,3862	8,1026	5,4427	4,2288

Tabla 42. Intervalos Generados para cada Sector.

Tomando las tasas de crecimiento mostradas anteriormente y el modelo ADEH conformado por las series de datos agricultura, financieros, manufactura y servicios públicos generamos los pronósticos para los años venideros; de esta manera tenemos que el consumo sería:

Año	Bajo	Medio	Alto
2010	56148	56148	56148
2011	57156	57156	57156
2012	58403,7734	58932,4349	59461,0974
2013	59549,5386	60634,6678	61728,7436
2014	60724,4075	62395,0883	64093,3876
2015	61929,206	64215,8583	66559,3603
2016	63164,7848	66099,2247	69131,1934
2017	64432,0205	68047,5237	71813,6283
2018	65731,8155	70063,1835	74611,6271

Tabla 43. Pronósticos del Modelo ADEH con Intervalos para cada Sector.

Gráficamente tendríamos que la evolución de la demanda de energía eléctrica sería:

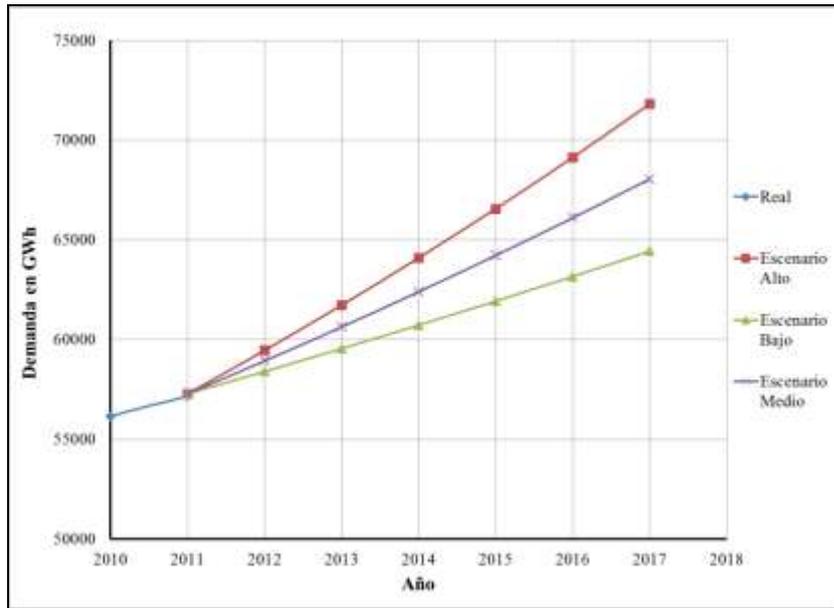


Ilustración 17. Evolución de la Demanda de Energía Eléctrica Usando el Modelo ADEH.

4.2. Intervalo Global

Se calcula un intervalo de crecimiento para el PIB global, y se trabaja con la suposición que los PIB sectoriales tendrán la misma tasa de crecimiento; esta idea se usa debido a la poca cantidad de datos para generar un intervalo de confianza más preciso. A continuación se presenta el histograma y gráfico de densidad del PIB:

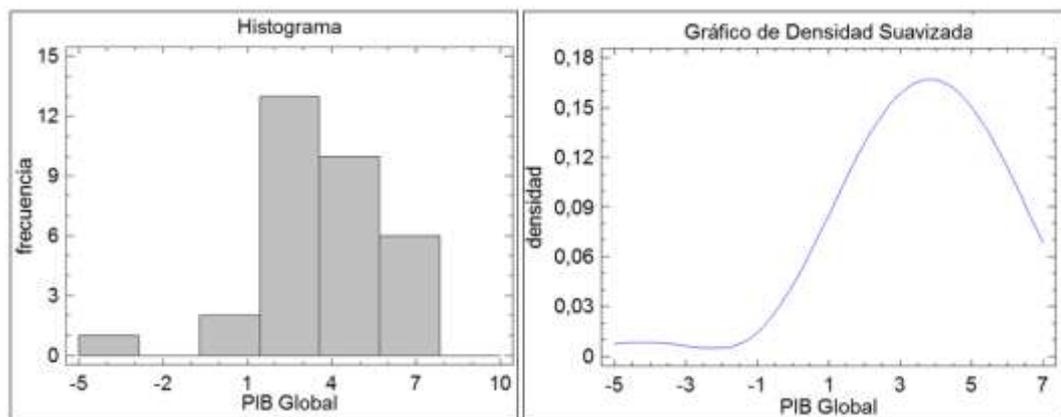


Ilustración 18. Histograma y Gráfico de Densidad para el PIB Global.

Para los datos de PIB tomados para los años de 1980 hasta el 2011, datos tomados de la página del Banco Mundial, tenemos que el intervalo de confianza del 95,0% es de:

Para la media: $3,55937 \pm 0,796871$ es decir $[2,7625, 4,3562]$

En otras palabras tendríamos que:

Escenario	Tasa de Crecimiento
Bajo	2,7625
Medio	3,5594
Alto	4,3562

Tabla 44. Tasas de Crecimiento para el PIB Global.

Tomando las tasas de crecimiento mostradas anteriormente y el modelo ADEH generamos los pronósticos para los años venideros; de esta manera tenemos que los pronósticos y su evolución sería:

Año	Bajo	Medio	Alto
2010	56148	56148	56148
2011	57156	57156	57156
2012	58759,0237	59183,8583	59608,6396
2013	60272,4213	61148,9481	62032,1347
2014	61827,6264	63183,9833	64561,2021
2015	63425,7941	65291,4535	67200,4407
2016	65068,1112	67473,9371	69954,6499
2017	66755,7973	69734,1039	72828,8378
2018	68490,1057	72074,7192	75828,2312

Tabla 45. Pronósticos del Modelo ADEH con Intervalos para el PIB Global.

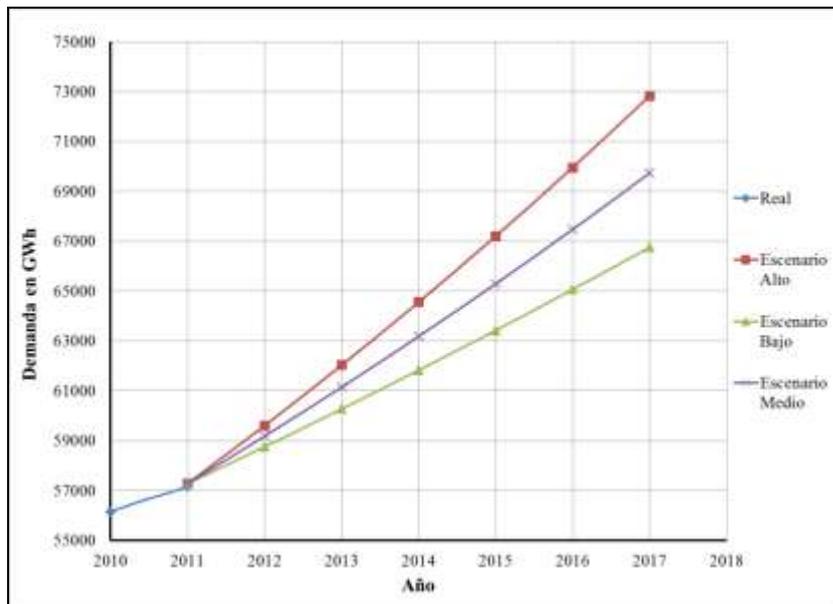


Ilustración 19. Evolución de la Demanda de Energía Eléctrica.

Si comparamos los resultados obtenidos, siendo la secuencia A para los intervalos sectoriales o líneas azules y la secuencia B o líneas naranjas al intervalo general; podemos observar del que la secuencia B presenta mayor crecimiento.

Esto claramente lo podemos observar en la ilustración 20.

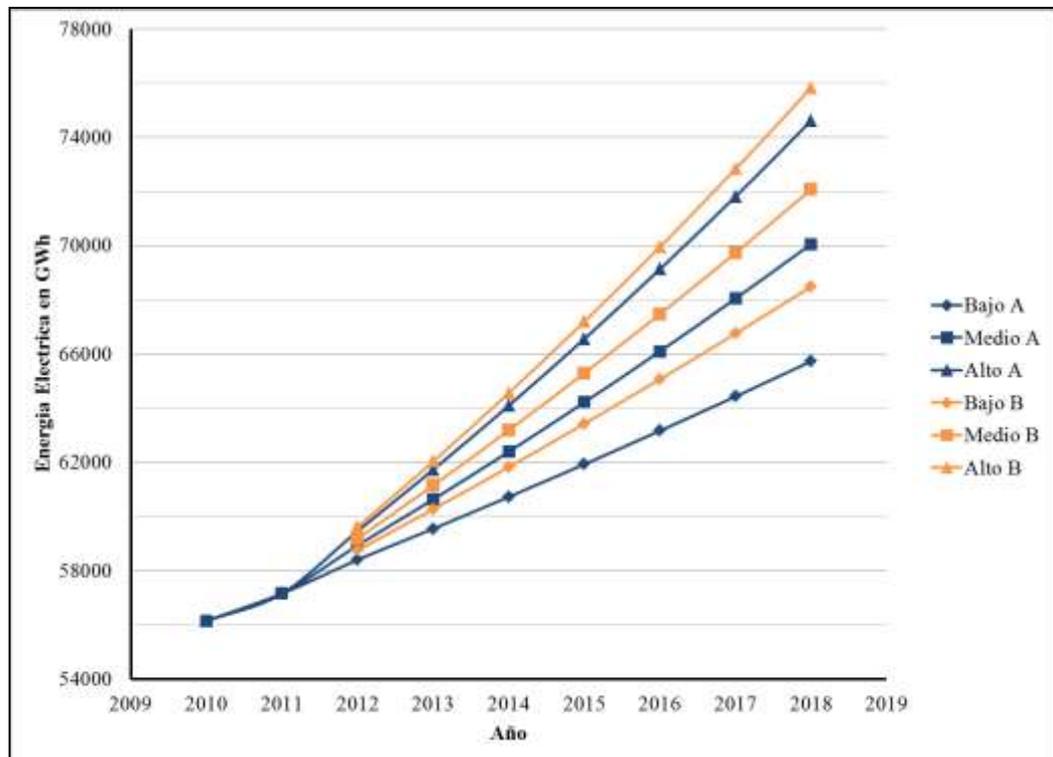


Ilustración 20. Comparacion de la Evolucion de la DEE.

5. CONCLUSIONES

Durante la investigación que dio como resultado este documento se presentó una introducción a los conceptos y temática utilizada para el diseño de modelos de pronóstico, a su vez se estudiaron tres modelos para pronosticar la DEE anual en Colombia; en el tercer modelo se utilizó una metodología que emplea el coeficiente estadístico C_p de Mallows, la cual permite escoger el modelo más apropiado para la generación de pronósticos.

Los modelos se generaron con valores que van del año 2000 al 2009 y a su vez validaron, primero con el coeficiente de Mallows, segundo con los estadísticos de regresión y por ultimo con el pronóstico de los valores de los años 2010 y 2011.

Se mostró la falencia que se está produciendo al usar un modelo con la variable PIB global, puede ser no muy notorio pero que poco a poco podría convertirse en un gran problema para los diseños en los planes de expansión del parque generador.

Entre las conclusiones más relevantes del proyecto tenemos que:

- Se presentó un modelo diseñado a partir de los PIB de los diferentes sectores de la economía, con el fin de pronosticar la demanda de energía eléctrica a largo plazo, usando la metodología de regresiones múltiples.
- El modelo logra relacionar los datos que van del año 2000 al 2009 con un error no superior al 0,3%. Y validar los datos 2010 y 2011 con un error máximo de 1%.
- Se realizaron los pronósticos de intervalo para los años 2012 hasta el 2017, por medio de dos forma, la primera realizando intervalos individuales para cada sector y la segunda realizando un intervalo de crecimiento global.

Como trabajo futuro tenemos que:

- Se debe diseñar estudios de la demanda de energía eléctrica de los sectores de la economía, y tratarlos de manera individual, debido a la modernización de sus procesos y a la inclusión de generación distribuida de las empresas.
- Disminuir el tamaño de la serie de datos, para hacer revisiones mensuales o trimensuales.

6. REFERENCIAS

1. Acquirente Unico, *Rapporto di previsione della domanda elettrica*, Italia, 2003.
2. Anderson D.R y Sweeney D.J, *Statistics For Business And Economics*, CENGAGE Learning , 10 Ed, 2008.
3. Asociación Colombiana de Generadores de Energía Eléctrica, ACOGEN, *La Expansión de la Generación Eléctrica en Colombia: 2012-2018*, 2008.
4. Aucoin Frédéric, *Analyse De La Demande D'Électricité Du Secteur Résidentiel Du Quebec*, Université du Québec à Montréal, Febrero 2007.
5. Banco de la Republico de Colombia, 17 de octubre de 2012, <http://www.banrep.gov.co/series-estadisticas/>
6. Barón FJ y Tellez F, *Apuntes de Bioestadística*, Universidad de Málaga, Junio 2011, Capitulo 6.
7. Barrientos A.F, Olaya J. y González V.M, *Un Modelo Spline para el Pronóstico de la Demanda de Energía Eléctrica*, Revista Colombiana de Estadística, 2007.
8. Bernard Jean Thomas, *Un modèle intégré de la demande totale d'énergie Application à la province de Québec*, Université Laval, 1998.
9. Campo Jacobo y Sarmiento Viviana, *Relación Consumo de Energía Eléctrica y PIB: Evidencia desde un Panel Cointegrado de 10 Países de América Latina entre 1971-2007*, Colombia, Junio 2011.
10. Campo Robledo J y Sarmiento Guzmán Viviana, *Relación Consumo de energía Eléctrica y PIB: Evidencia desde un Panel Cointegrado de 10 Países de América Latina entre 1971 - 2007*, Junio 2011.
11. CaracolRadio, <http://www.caracol.com.co/noticias/ecologia/peligrosas-sequias-podrian-vivirse-en-menos-de-dos-anos-en-colombia /20120306/nota/1650195.aspx>, 13 Marzo 2012
12. Cohen Roger, *Pronósticos*, Gestión de Recursos - FI UBA, Octubre 2003.
13. Departamento Administrativo Nacional de Estadística, DANE, 17 de octubre de 2012, <http://www.dane.gov.co>
14. Egelioglu F, Mohamad A.A. y Guven H, *Economic Variables and Electricity Consumption in Northern Cyprus*, Septiembre 1999.

15. ELESPECTADOR.COM, *Colombia podría sufrir problemas con suministro de energía en 2014*, <http://www.americaeconomia.com/negocios-industrias/colombia-podria-sufrir-problemas-con-suministro-de-energia-en-2014>, 17 de octubre de 2012.
16. Franco Carlos, Velásquez Juan y Olaya Yris, *Caracterización de la Demanda Mensual de Electricidad en Colombia usando un Modelo de Componentes no Observables*, Colombia, 2008.
17. González Rodríguez Hernán, *Precios de la Energía Eléctrica en Colombia*, <http://www.elespectador.com/impreso/opinion/columna-322779-precios-de-energia-electrica-colombia>, 17 de octubre de 2012.
18. Magliano Fernando Ignacio, 17 de octubre de 2012, <http://conocimientopractico.wordpress.com/article/caracteristicas-de-la-metodologia-2sr10788nwjjj-26/>
19. Medina Santiago y Garcia Josefina, *Predicción de Demanda de Energía en Colombia Mediante un Sistema de Inferencia Difuso Neuronal*, Revista Energética, Colombia, 2005.
20. Mohamed Zaid y Pat Dodger, *Forecasting Electricity Consumption in New Zealand Using Economic and Demographic Variables*, Octubre 2003.
21. Mohamed Zaid, *Forecasting Electricity Consumption: A Comparison of Growth Curves, Econometric and ARIMA Models for Selected Countries and World Regions*, University of Canterbury, Noviembre 2004.
22. Murillos Joaquin, Trejos Alvaro y Carvajal Patricia, *Estudio del Pronóstico de la Demanda de Energía Eléctrica Utilizando Modelos de Series de Tiempo*, Scientia et Technica, Diciembre 2003.
23. Name Jose David, *Es urgente que Gobierno defina racionamiento programado de gas natural*, <http://www.senado.gov.co/sala-de-prensa/noticias/item/15104-es-urgente-que-gobierno-defina-racionamiento-programado-de-gas-natural-dice-senador-jose-david-name-cardozo?tmpl=component&print=1>, 17 de octubre de 2012.
24. Petroff John, Traducido por Arce Miguel, *Análisis de series de tiempo según Box-Jenkins*, 17 de octubre de 2012, <http://www.peoi.org/Courses/Coursessp/finanal/ch/ch5f1.html>

25. Prasad Anurag, *Forecasting and Model Selection*, Marzo 2008.
26. Renatas K y Ángel A. J., *Modelo de Regresión Múltiple*, Diciembre del 2006.
27. Revista Semana, *Potencia Eléctrica “Radiografía de un Sector Modelo y Estratégico en el Mundo”*, Julio 2012.
28. Rueda Viviana, *Predicción del Consumo de Energía en Colombia con Modelos no Lineales*, 2011.
29. Rueda Viviana, Velazquez Juan y Franco Carlos, *Recent Advances in Load Forecasting Using Nonlinear Models*, Dyna, Colombia, Junio 2001.
30. Soto Ortiz José Daniel y Grimaldo Guerrero John William, *¿Nuevo Racionamiento de Energía? Atrasos en los proyectos de generación: Riesgos en el Sector Eléctrico*, Informativo Un Norte, Barranquilla: Universidad del Norte, 2012, pág. 9.
31. Syazwani Aman, Hew Wooi Ping y Marizan Mubin, *Modelling and Forecasting Electricity Consumption of Malaysian Large Steel Mills*, Abril 2011.
32. The free dictionary, <http://es.thefreedictionary.com/anunciar>, 13 Marzo 2012
33. Unidad de Planeación Minero Energética UPME, 17 de octubre de 2012, <http://www.upme.gov.co>
34. Universidad Nacional de Colombia, 17 de octubre de 2012, http://virtual.uptc.edu.co/ova/estadistica/docs/libros/2007315/lecciones_html/capitulo_8/leccion1/leccion1-4/cpmallows.html
35. UPME, *Proyecciones Regionales de la Demanda de Energía Eléctrica y Potencia Máxima en Colombia*, Diciembre 2009.
36. Vélez Pareja Ignacio, *Apuntes de Probabilidad y Estadística para Ingeniería y Administración*, Octubre 2003.
37. Wikipedia, *Precisión del Pronóstico*, 17 de octubre de 2012, [http://es.wikipedia.org/wiki/Pron%C3%B3stico_\(Estad%C3%ADstica\)](http://es.wikipedia.org/wiki/Pron%C3%B3stico_(Estad%C3%ADstica))
38. XM, 17 de octubre de 2012, <http://www.xm.com.co>
39. XM, *Análisis del PIB y la Demanda de Energía en Colombia en Momentos de Crisis Económica*, 17 de octubre de 2012, <http://www.xm.com.co/BoletinXM/Pages/MECjun24.aspx>

40.ANEXOS

Anexo 1. Demanda Anual de Energía Eléctrica

AÑO	GWH
1995	41974
1996	42300
1997	43033
1998	43734
1999	41503
2000	42246
2001	43216
2002	44499
2003	45768
2004	47017
2005	48829
2006	50815
2007	52851
2008	53870
2009	54679
2010	56148
2011	57156
2012	59369
2013³	25531

³ Leído el 06 de mayo de 2013.

Anexo 2. Datos de la Población Colombiana.

AÑO	TOTAL	HOMBRES	MUJERES
1991	34.830.570	17.183.454	17.647.116
1992	35.520.940	17.524.117	17.996.823
1993	36.207.108	17.862.788	18.344.320
1994	36.853.905	18.182.235	18.671.670
1995	37.472.184	18.488.018	18.984.166
1996	38.068.050	18.783.203	19.284.847
1997	38.635.691	19.063.910	19.571.781
1998	39.184.456	19.335.555	19.848.901
1999	39.730.798	19.605.580	20.125.218
2000	40.295.563	19.883.774	20.411.789
2001	40.813.541	20.139.879	20.673.662
2002	41.328.824	20.395.339	20.933.485
2003	41.848.959	20.653.560	21.195.399
2004	42.368.489	20.913.566	21.454.923
2005	42.888.592	21.169.835	21.718.757
2006	43.405.956	21.426.954	21.979.002
2007	43.926.929	21.683.071	22.243.858
2008	44.451.147	21.942.355	22.508.792
2009	44.978.832	22.203.708	22.775.124
2010	45.509.584	22.466.660	23.042.924
2011	46.044.601	22.731.299	23.313.302
2012	46.581.823	22.997.087	23.584.736
2013	47.121.089	23.264.039	23.857.050
2014	47.661.787	23.531.670	24.130.117
2015	48.203.405	23.799.679	24.403.726
2016	48.747.708	24.069.035	24.678.673
2017	49.291.609	24.337.747	24.953.862
2018	49.834.240	24.605.796	25.228.444
2019	50.374.478	24.873.329	25.501.149
2020	50.911.747	25.138.964	25.772.783

Anexo 3. PIB de las Ramas Económicas, dadas en miles de millones de pesos.

Se presentan los sectores o ramas económicas expuestas por el banco de la república:

- a. Agricultura: Comprende las actividades de agricultura, ganadería, caza, silvicultura y pesca.
- b. Minería: Explotación de minas, yacimientos y canteras.
- c. Manufactura: Industrias manufactureras.
- d. Servicios Públicos: suministro de electricidad, gas y agua.
- e. Construcción: construcción.
- f. Comercio: Comercio, reparación, restaurantes y hoteles.
- g. Transporte: Transporte, almacenamiento y comunicaciones.
- h. Financieros: Establecimientos financieros, seguros, actividades inmobiliarias y servicios a las empresas.
- i. Sociales: Actividades de servicios sociales, comunales y personales.

	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010
A	22.629	23.029	24.077	24.821	25.560	26.279	26.902	27.954	27.849	27.666	27.932
B	22.620	20.750	20.381	20.726	20.538	21.371	21.890	22.217	24.362	27.062	30.384
C	38.665	39.791	40.642	42.632	46.019	48.082	51.353	55.051	55.332	53.045	54.580
D	11.697	12.069	12.170	12.722	13.162	13.708	14.360	14.947	15.021	15.302	15.492
E	12.451	13.136	14.756	15.981	17.698	18.915	21.213	22.982	25.009	26.326	25.868
F	33.097	34.061	34.571	35.859	38.402	40.335	43.517	47.119	48.597	48.440	50.911
G	18.100	18.694	19.210	20.078	21.598	23.285	25.792	28.614	29.925	29.511	30.981
H	54.720	55.385	57.039	59.279	62.010	65.096	69.289	74.010	77.359	79.764	82.071
I	47.895	48.495	49.335	50.317	52.381	54.210	56.572	59.416	60.959	63.613	66.662
Total	284.761	289.539	296.789	308.418	324.866	340.156	362.938	387.983	401.744	408.379	424.719

Anexo 4. Modelos Obtenidos al Combinar las Diferentes Variables.

Modelo	R ²	R ² Ajustada	Cp	Variables
1	0,0000	0,0000	7924,6200	
2	95,6838	95,1443	336,3850	A
3	98,6612	98,4938	100,2040	B
4	99,3427	99,2605	46,1415	C
5	98,6265	98,4548	102,9580	D
6	96,0679	95,5764	305,9210	E
7	48,6207	42,1983	4069,7200	F
8	97,9326	97,6741	158,0020	G
9	99,3881	99,3116	42,5413	H
10	99,3038	99,1049	51,2268	AB
11	99,7371	99,6620	16,8561	AC
12	99,8826	99,8491	5,3093	AD
13	96,6472	95,6892	261,9670	AE
14	99,4523	99,2958	39,4484	AF
15	99,8472	99,8036	8,1173	AG
16	99,3885	99,2138	44,5066	AH
17	99,5909	99,4740	28,4515	BC
18	98,9446	98,6430	79,7247	BD
19	98,6761	98,2979	101,0170	BE
20	98,6916	98,3177	99,7918	BF
21	98,8988	98,5841	83,3571	BG
22	99,5041	99,3624	35,3398	BH
23	99,3452	99,1581	47,9448	CD
24	99,7931	99,7340	12,4101	CE
25	99,6277	99,5214	25,5299	CF
26	99,3531	99,1683	47,3170	CG
27	99,8742	99,8383	5,9792	CH
28	99,3838	99,2077	44,8826	DE
29	99,7705	99,7049	14,2063	DF
30	98,7381	98,3776	96,1016	DG

Modelo	R²	R² Ajustada	Cp	Variables
31	99,7179	99,6373	18,3809	DH
32	98,1918	97,6751	139,4400	EF
33	99,3960	99,2234	43,9118	EG
34	99,4267	99,2629	41,4754	EH
35	99,6172	99,5078	26,3651	FG
36	99,5075	99,3668	35,0673	FH
37	99,6361	99,5321	24,8655	GH
38	99,8720	99,8080	8,1544	ABC
39	99,8881	99,8321	6,8788	ABD
40	99,8889	99,8333	6,8158	ABE
41	99,7187	99,5781	20,3117	ABF
42	99,8612	99,7919	9,0071	ABG
43	99,5339	99,3009	34,9739	ABH
44	99,8971	99,8457	6,1626	ACD
45	99,8238	99,7357	11,9797	ACE
46	99,7486	99,6229	17,9433	ACF
47	99,8717	99,8076	8,1754	ACG
48	99,9044	99,8567	5,5797	ACH
49	99,8925	99,8388	6,5239	ADE
50	99,8892	99,8338	6,7912	ADF
51	99,8935	99,8403	6,4476	ADG
52	99,9067	99,8601	5,4007	ADH
53	99,5423	99,3134	34,3115	AEF
54	99,8505	99,7757	9,8625	AEG
55	99,4480	99,1720	41,7856	AEH
56	99,8501	99,7752	9,8904	AFG
57	99,7405	99,6108	18,5844	AFH
58	99,8639	99,7958	8,8001	AGH
59	99,7289	99,5933	19,5078	BCD
60	99,7938	99,6908	14,3538	BCE
61	99,8114	99,7172	12,9578	BCF
62	99,6815	99,5223	23,2649	BCG
63	99,8779	99,8169	7,6835	BCH

Modelo	R²	R² Ajustada	Cp	Variables
64	99,7211	99,5816	20,1264	BDE
65	99,8827	99,8240	7,3079	BDF
66	98,9467	98,4201	81,5509	BDG
67	99,7320	99,5979	19,2621	BDH
68	98,6999	98,0499	101,1290	BEF
69	99,5588	99,3382	33,0000	BEG
70	99,6003	99,4005	29,7031	BEH
71	99,6322	99,4483	27,1759	BFG
72	99,5458	99,3188	34,0266	BFH
73	99,6399	99,4598	26,5670	BGH
74	99,7932	99,6898	14,4047	CDE
75	99,8647	99,7971	8,7298	CDF
76	99,4231	99,1346	43,7641	CDG
77	99,8743	99,8115	7,9710	CDH
78	99,8041	99,7062	13,5361	CEF
79	99,7960	99,6941	14,1794	CEG
80	99,8901	99,8352	6,7175	CEH
81	99,8085	99,7128	13,1876	CFG
82	99,8958	99,8438	6,2618	CFH
83	99,8826	99,8239	7,3115	CGH
84	99,7978	99,6967	14,0408	DEF
85	99,4114	99,1171	44,6900	DEG
86	99,7183	99,5774	20,3496	DEH
87	99,7971	99,6957	14,0950	DFG
88	99,8413	99,7620	10,5865	DFH
89	99,7416	99,6124	18,4971	DGH
90	99,6598	99,4897	24,9860	EFG
91	99,5077	99,2616	37,0510	EFH
92	99,6462	99,4693	26,0640	EGH
93	99,7259	99,5889	19,7430	FGH
94	99,8975	99,8155	8,1330	ABCD
95	99,9103	99,8385	7,1174	ABCE
96	99,8740	99,7731	9,9974	ABCF

Modelo	R²	R² Ajustada	Cp	Variables
97	99,9003	99,8206	7,9057	ABCG
98	99,9216	99,8589	6,2170	ABCH
99	99,9029	99,8252	7,7038	ABDE
100	99,9241	99,8634	6,0195	ABDF
101	99,8947	99,8105	8,3517	ABDG
102	99,9254	99,8657	5,9177	ABDH
103	99,8919	99,8054	8,5756	ABEF
104	99,9014	99,8225	7,8218	ABEG
105	99,9507	99,9112	3,9138	ABEH
106	99,8615	99,7507	10,9850	ABFG
107	99,7835	99,6103	17,1727	ABFH
108	99,8688	99,7639	10,4040	ABGH
109	99,8975	99,8155	8,1313	ACDE
110	99,9110	99,8398	7,0607	ACDF
111	99,9024	99,8243	7,7418	ACDG
112	99,9358	99,8845	5,0897	ACDH
113	99,8300	99,6939	13,4880	ACEF
114	99,8888	99,7998	8,8240	ACEG
115	99,9066	99,8319	7,4073	ACEH
116	99,8818	99,7872	9,3799	ACFG
117	99,9046	99,8284	7,5643	ACFH
118	99,9148	99,8466	6,7617	ACGH
119	99,9201	99,8562	6,3361	ADEF
120	99,8971	99,8148	8,1617	ADEG
121	99,9365	99,8856	5,0408	ADEH
122	99,9063	99,8313	7,4368	ADFG
123	99,9089	99,8360	7,2292	ADFH
124	99,9098	99,8376	7,1555	ADGH
125	99,8519	99,7335	11,7467	AEFG
126	99,7417	99,5350	20,4933	AEFH
127	99,8641	99,7554	10,7802	AEGH
128	99,8640	99,7551	10,7912	AFGH
129	99,8077	99,6539	15,2538	BCDE

Modelo	R²	R² Ajustada	Cp	Variables
130	99,8848	99,7927	9,1366	BCDF
131	99,7321	99,5177	21,2547	BCDG
132	99,8863	99,7954	9,0177	BCDH
133	99,8149	99,6669	14,6811	BCEF
134	99,8164	99,6696	14,5618	BCEG
135	99,8903	99,8026	8,6997	BCEH
136	99,8443	99,7198	12,3502	BCFG
137	99,9091	99,8365	7,2069	BCFH
138	99,8965	99,8136	8,2140	BCGH
139	99,9265	99,8678	5,8274	BDEF
140	99,7371	99,5268	20,8552	BDEG
141	99,8183	99,6729	14,4168	BDEH
142	99,8863	99,7954	9,0179	BDFG
143	99,9565	99,9217	3,4526	BDFH
144	99,7988	99,6378	15,9601	BDGH
145	99,6835	99,4302	25,1103	BEFG
146	99,6285	99,3313	29,4685	BEFH
147	99,6532	99,3758	27,5093	BEGH
148	99,7262	99,5071	21,7224	BFGH
149	99,8650	99,7570	10,7091	CDEF
150	99,8041	99,6474	15,5393	CDEG
151	99,8902	99,8023	8,7112	CDEH
152	99,8729	99,7713	10,0799	CDFG
153	99,9311	99,8760	5,4656	CDFH
154	99,9007	99,8212	7,8787	CDGH
155	99,8598	99,7476	11,1250	CEFG
156	99,9002	99,8203	7,9177	CEFH
157	99,8923	99,8062	8,5429	CEGH
158	99,9021	99,8238	7,7662	CFGH
159	99,8103	99,6586	15,0470	DEFG
160	99,9080	99,8344	7,2963	DEFH
161	99,7583	99,5649	19,1737	DEGH
162	99,8444	99,7200	12,3401	DFGH

Modelo	R²	R² Ajustada	Cp	Variables
163	99,7264	99,5076	21,7021	EFGH
164	99,9113	99,8004	9,0374	ABCDE
165	99,9241	99,8293	8,0193	ABCDF
166	99,9042	99,7845	9,5982	ABCDG
167	99,9365	99,8570	7,0399	ABCDH
168	99,9103	99,7981	9,1173	ABCEF
169	99,9146	99,8079	8,7735	ABCEG
170	99,9614	99,9132	5,0610	ABCEH
171	99,9037	99,7833	9,6393	ABCFG
172	99,9217	99,8239	8,2093	ABCFH
173	99,9253	99,8320	7,9220	ABCGH
174	99,9267	99,8351	7,8141	ABDEF
175	99,9060	99,7886	9,4532	ABDEG
176	99,9524	99,8930	5,7730	ABDEH
177	99,9299	99,8424	7,5574	ABDFG
178	99,9638	99,9186	4,8711	ABDFH
179	99,9263	99,8342	7,8440	ABDGH
180	99,9016	99,7785	9,8076	ABEFG
181	99,9507	99,8890	5,9138	ABEFH
182	99,9561	99,9011	5,4863	ABEGH
183	99,8689	99,7049	12,4026	ABFGH
184	99,9213	99,8229	8,2454	ACDEF
185	99,9024	99,7805	9,7393	ACDEG
186	99,9417	99,8689	6,6221	ACDEH
187	99,9209	99,8219	8,2786	ACDFG
188	99,9433	99,8725	6,4967	ACDFH
189	99,9362	99,8563	7,0646	ACDGH
190	99,8942	99,7619	10,3940	ACEFG
191	99,9069	99,7905	9,3849	ACEFH
192	99,9164	99,8118	8,6338	ACEGH
193	99,9171	99,8134	8,5796	ACFGH
194	99,9254	99,8321	7,9202	ADEFG
195	99,9630	99,9167	4,9371	ADEFH

Modelo	R²	R² Ajustada	Cp	Variables
196	99,9420	99,8694	6,6044	ADEGH
197	99,9151	99,8089	8,7360	ADFGH
198	99,8643	99,6946	12,7666	AEFGH
199	99,9285	99,8392	7,6704	BCDEF
200	99,8190	99,5927	16,3604	BCDEG
201	99,8907	99,7542	10,6674	BCDEH
202	99,8887	99,7496	10,8283	BCDFG
203	99,9571	99,9035	5,4014	BCDFH
204	99,9008	99,7768	9,8695	BCDGH
205	99,8615	99,6883	12,9883	BCEFG
206	99,9140	99,8066	8,8182	BCEFH
207	99,8974	99,7691	10,1390	BCEGH
208	99,9104	99,7984	9,1059	BCFGH
209	99,9361	99,8563	7,0650	BDEFG
210	99,9625	99,9156	4,9761	BDEFH
211	99,8271	99,6111	15,7115	BDEGH
212	99,9613	99,9130	5,0670	BDFGH
213	99,7266	99,3848	23,6878	BEFGH
214	99,8747	99,7180	11,9407	CDEFG
215	99,9397	99,8643	6,7845	CDEFH
216	99,9008	99,7768	9,8709	CDEGH
217	99,9316	99,8461	7,4261	CDFGH
218	99,9060	99,7884	9,4596	CEFGH
219	99,9194	99,8187	8,3926	DEFGH