

PRONÓSTICO DE LA DEMANDA
ELÉCTRICA RESIDENCIAL BASADO EN EL
MODELO DE REGRESIÓN ADAPTATIVA
MULTIVARIANTE SPLINE (MARS)

PRONÓSTICO DE LA DEMANDA ELÉCTRICA RESIDENCIAL BASADO EN EL MODELO DE REGRESIÓN ADAPTATIVA MULTIVARIANTE SPLINE (MARS)

Miguel Eduardo Ortiz Alvarado
Egresado de la Carrera de Ingeniería Eléctrica
Facultad de Ingenierías
Universidad Politécnica Salesiana

Dirigido por:
Edwin Marcelo García Torres
Facultad de Ingenierías
Universidad Politécnica Salesiana



Quito - Ecuador

Miguel Eduardo Ortiz Alvarado

**PRONÓSTICO DE LA DEMANDA ELÉCTRICA RESIDENCIAL BASADO
EN EL MODELO DE REGRESIÓN ADAPTATIVA MULTIVARIANTE
SPLINE (MARS)**

Universidad Politécnica Salesiana
Ingeniería Eléctrica

Breve reseña historia e información de contacto:



generación distribuida.

mortiz@est.ups.edu.ec

Miguel Eduardo Ortiz Alvarado (Y'1988-M'01). Bachiller Técnico Industrial, especialidad Electrónica del Instituto Tecnológico Superior Central Técnico. Egresado de la Carrera de Ingeniería Eléctrica de la Universidad Politécnica Salesiana. Miembro Estudiantil del Grupo de Investigación en Redes Eléctricas Inteligentes-GIREI. Su trabajo se basa en investigación de técnicas y modelos para predicción de la demanda eléctrica residencial, procesos estocásticos y redes eléctricas inteligentes. Entre sus intereses están las redes eléctricas de distribución y

Dirigido por:



Marcelo García (A'1978-M'10). Nació en Ambato, Ecuador. Se graduó en Ingeniería Eléctrica de la Universidad Politécnica Salesiana, Ecuador en 2010 y actualmente se encuentra cursando un Master en Gestión de la Energía en la Universidad Técnica de Cotopaxi. Es profesor e investigador de la Universidad Politécnica Salesiana - Quito Ecuador. En la actualidad es miembro del Grupo de Investigación en Redes Eléctricas Inteligentes-GIREI.

egarcia@ups.edu.ec

Todos los derechos reservados:

Queda Prohibida, salvo excepción prevista en la Ley, cualquier forma de reproducción, distribución, comunicación pública y transformación de esta obra para fines comerciales, sin contar con la autorización de los titulares de propiedad intelectual. La infracción de los derechos mencionados puede ser constitutiva de delito contra la propiedad intelectual. Se permite la libre difusión de este texto con fines académicos o investigativos por cualquier medio, con la debida notificación a los autores.

DERECHOS RESERVADOS

©2015 Universidad Politécnica Salesiana

QUITO – ECUADOR

DEDICATORIA.

El presente trabajo se lo dedico al todopoderoso que guía mis pasos, que me dio la luz primera y la capacidad para cumplir mis metas. A mi padre Miguel (†) por su amor incondicional, por su lucha incansable y sobre todo por su ejemplo de sacrificio, el cual nos dejó un legado vida a toda su familia. Esta meta es parte de una promesa de amor y un homenaje póstumo a su memoria. A mi madre Delia, que por difíciles que fuesen los tiempos, cuidó de su familia y sin ese apoyo fundamental no sería posible este sueño. A mi hermana Vero, siempre pendiente de todos en casa y que estuvo allí dándome su mano en momentos decisivos. A mis hermanos Estefy y Luis por ser inspiración y mostrarles que el camino del conocimiento es lo que los llevará lejos. A Javier, por sus consejos y compañía desde el inicio de este sueño, por su cariño. A mis amigos, compañeros de estudio y trabajo, que siempre tuvieron una mano amiga y la mejor disposición para ayudarme a cumplir una meta más en mi vida. A todos ustedes les dedico mi lucha y mi compromiso de ser más para servir mejor.

AGRADECIMIENTO.

A todos mis maestros, que supieron se guía y senda de conocimiento en este gran desafío de la ingeniería, a mi tutor de proyecto, el Ing. Marcelo García por su apoyo a la consecución del mismo y a la gran familia Salesiana, por su obra inspirada en San Juan Bosco para cambiar la vida de los jóvenes.

DECLARATORIA DE COAUTORÍA DEL DOCENTE TUTOR/A

Yo, Edwin Marcelo García Torres declaro que bajo mi dirección y asesoría fue desarrollado el trabajo de titulación *Pronóstico de la Demanda Eléctrica Residencial Basado en el Modelo de Regresión Adaptativa Multivariante Spline (Mars)* realizado por Miguel Eduardo Ortiz Alvarado, obteniendo un producto que cumple con todos los requisitos estipulados por la Universidad Politécnica Salesiana para ser considerados como trabajo final de titulación.

Quito, Diciembre 2015

.....

Edwin Marcelo García Torres

Cédula de identidad: 1803087400

CESIÓN DE DERECHOS DE AUTOR

Yo, Miguel Eduardo Ortiz Alvarado, con documento de identificación N° 0502232655, manifiesto mi voluntad y cedo a la Universidad Politécnica Salesiana la titularidad sobre los derechos patrimoniales en virtud de que soy autor/es del trabajo de grado/titulación intitulado: “Pronóstico de la Demanda Eléctrica Residencial Basado en el Modelo de Regresión Adaptativa Multivariante Spline (Mars)”, mismo que ha sido desarrollado para optar por el título de: Ingeniero Eléctrico, en la Universidad Politécnica Salesiana, quedando la Universidad facultada para ejercer plenamente los derechos cedidos anteriormente. En aplicación a lo determinado en la Ley de Propiedad Intelectual, en mi condición de autor me reservo los derechos morales de la obra antes citada. En concordancia, suscribo este documento en el momento que hago entrega del trabajo final en formato impreso y digital a la Biblioteca de la Universidad Politécnica Salesiana.

Firma

.....

Nombre: Miguel Eduardo Ortiz Alvarado

Cédula: 0502232655

Fecha: Diciembre 2015

GLOSARIO

A

ADMINISTRACIÓN DE LA DEMANDA: Es una serie de acciones encaminadas a optimizar el uso de los recursos instalados y tiene como fin reducir o controlar la demanda de energía.

C

CONFIABILIDAD: es la probabilidad de que el sistema cumpla con su objetivo, bajo condiciones determinadas de uso en un lapso de tiempo.

D

DEMANDA MÁXIMA: es la máxima coincidencia de cargas en un lapso definido.

E

EFICIENCIA ENERGÉTICA: consiste en una serie de acciones para reducir el consumo de energía eléctrica y está relacionado con la cantidad de energía útil que se puede obtener de un sistema.

F

FORECASTING: Es el término usado para predicción o pronóstico de la demanda de energía en un lapso determinado.

G

GENERACIÓN DISTRIBUIDA: Etapa de generación que se ubica cerca de la carga y es una fuente adicional para suministrar energía. Su característica principal es la interconexión con la etapa de distribución o sub transmisión del sistema eléctrico.

GENERADOR PÚBLICO: persona natural o jurídica cuyo título habilitante le faculta realizar actividades de generación.

I

INTELIGENCIA ARTIFICIAL: Rama de la computación que asocia un fenómeno natural, con una analogía artificial por medio de un software.

M

MARS: Regresión Adaptativa Multivariante Spline, por sus siglas en inglés, es una técnica estocástica de pronóstico propuesta por Friedman en los años 90.

P

PROCESO ESTOCÁSTICO: Se define como un proceso en el cual se involucren probabilidades. Es un concepto matemático que sirve para caracterizar una sucesión de variables aleatorias, que dependen de otra variable.

R

REGRESIÓN: Es un proceso estadístico que estima una relación entre variables, generalmente una variable independiente con una o más variables independientes.

S

SERIES DE TIEMPO: Se denominan a las variables que tengan datos registrados con intervalos sucesivos de tiempo.

SMART GRID: Red eléctrica inteligente, es un modelo de red con eficiencia y confiabilidad energética.

SISTEMA DE DISTRIBUCIÓN: Parte del sistema eléctrico de potencia que tiene como objetivo la entrega de energía al consumidor final.

INDICE GENERAL

<i>GLOSARIO</i>	1
<i>INDICE GENERAL</i>	3
<i>INDICE DE FIGURAS</i>	4
<i>INDICE DE TABLAS</i>	4
<i>TÍTULO DE LA INVESTIGACIÓN</i>	5
<i>RESUMEN DEL PROYECTO</i>	5
<i>ABSTRACT</i>	5
1. <i>INTRODUCCIÓN</i>	6
<i>Planteamiento del Proyecto</i>	6
<i>Estado del Arte</i>	9
2. <i>OBJETIVOS</i>	13
<i>OBJETIVO GENERAL</i>	13
<i>OBJETIVOS ESPECÍFICOS</i>	13
3. <i>METODOLOGÍA</i>	13
4. <i>DESARROLLO DE LA INVESTIGACIÓN</i>	14
4.1 <i>MODELOS Y MÉTODOS DE PARA PREDICCIÓN DE LA DEMANDA</i> <i>ELECTRICA</i>	14
4.1.1 <i>Modelos de regresión lineal</i>	16
4.1.2 <i>Series de tiempo</i>	16
4.1.3 <i>Modelos ARIMA</i>	17
4.2 <i>MULTIVARIATE ADAPTIVE REGRESSION SPLINES</i>	18
4.2.1 <i>Método spline - mínimos cuadrados penalizado</i>	24
4.2.2 <i>Splines Cúbicos Naturales</i>	24
5. <i>RESULTADOS ESPERADOS</i>	25
6. <i>ESTRATEGIA PARA LA DIVULAGACIÓN DE LOS RESULTADOS</i>	29

7. <i>SECTORES BENEFICIADOS</i>	29
8. <i>CONCLUSIONES</i>	29
9. <i>RECOMENDACIONES</i>	30
10. <i>REFERENCIAS</i>	31

INDICE DE FIGURAS

Ilustración 1 Factores que influyen en el pronóstico de la demanda de energía eléctrica [7].....	8
Ilustración 2 Temática de la investigación.....	11
Ilustración 3 Formulación del problema.	11
Ilustración 4 Comparativo de soluciones.	12
Ilustración 5 Clasificación de los métodos de pronóstico de demanda de energía eléctrica.	15
Ilustración 6 Las funciones base usadas por MARS.....	19
Ilustración 7 Metodología para pronóstico utilizando regresión.....	23
Ilustración 8 Función demanda año 2013.	26
Ilustración 9 Función demanda año 2014.	27
Ilustración 10 Método de interpolación spline para el año 2015.	28

INDICE DE TABLAS

Tabla 1 Matriz de estado del arte.....	9
Tabla 2 Histórico de demanda residencial para los años 2013 y 2014.....	25

TÍTULO DE LA INVESTIGACIÓN

PRONÓSTICO DE LA DEMANDA ELÉCTRICA RESIDENCIAL BASADO EN EL MODELO DE REGRESIÓN ADAPTATIVA MULTIVARIANTE SPLINE (MARS)

RESUMEN DEL PROYECTO

Un problema que suscita mucho interés a resolver en ingeniería es determinar el comportamiento de la demanda de energía eléctrica de los usuarios a través del tiempo, debido a que esta no tiene una característica lineal y además está sujeta a una amplia variedad de variables exógenas, incluyendo las condiciones prevalecientes del tiempo, temporada del año, variables demográficas y económicas, así como un carácter aleatorio inherente en su uso individual. La predicción de la demanda de energía de los consumidores tiene como propósito mejorar la confiabilidad, tomar decisiones y planificación del sistema, entre otras. El método más adecuado para resolver este problema no está claramente definido, sin embargo, en este trabajo se hace una revisión general de los modelos y métodos comúnmente empleados para su pronóstico, con énfasis en la técnica de Regresión Adaptativa Multivariante Splines (MARS) -por sus siglas en inglés-, que es un modelo de regresión no paramétrica y su principal objetivo es predecir los valores de una variable dependiente o resultado, de un conjunto de variables independientes o predictoras. Además se toma en cuenta que se debe contar con suficiente información histórica de la curva de carga diaria para realizar estudios estadísticos y de predicción, con el fin de planificar recursos y en el futuro sea una herramienta para la implementación de una red inteligente, ya que mejorará las características de administración de la demanda y las tareas de gestión en la distribución. La aplicación del modelo en la investigación propone delimitar un patrón de datos históricos de consumo residencial y extrapolar a predicciones futuras.

ABSTRACT

A problem to be solved much interest in engineering is to determine the behavior of the electricity demand of users over time, because this does not have a linear characteristic and is also subject to a wide variety of exogenous variables, including the prevailing weather conditions, time of year, demographic and economic variables, as well as an inherent randomness in individual use. The prediction of energy demand of consumers has the purpose to improve reliability, make decisions and system planning, among others. The most

appropriate method to resolve this issue is not clearly defined, however, this work is carried out a general review of the models and methods commonly employed for prognosis, with emphasis on technical Multivariate Adaptive Regression Splines (MARS) which it is a nonparametric regression model and its main objective is to forecast the values of a dependent variable or outcome of a set of independent or predictor variables. Also take note that you must have sufficient historical information on the daily load curve for statistical and forecasting studies, in order to plan future resources and will be a tool for the implementation of an intelligent network because improve the characteristics of demand management and management tasks in distribution. Applying the model research proposes to define a pattern of residential consumption historical data and extrapolate to future predictions.

1. INTRODUCCIÓN

Planteamiento del Proyecto

La predicción de la demanda de electricidad es un problema de suma importancia en ingeniería y uno de los principales retos es escoger la técnica o método más adecuado para dicha predicción, ya que de los resultados obtenidos se tomarán las decisiones más acertadas para garantizar el suministro de energía, obtener el menor número de interrupciones en el servicio y al menor costo posible.

Existen una gran cantidad de artículos que plantean dar solución a la predicción de la demanda de electricidad y algunos métodos han sido empleados para el pronóstico la demanda de energía eléctrica, pero en el presente documento se analiza un modelo estadístico como se evidencia en la sección 4.2.

En general se los puede clasificar en tres principales grupos [1]:

1. Modelos estocásticos.
2. Inteligencia artificial.
3. Aquellos en base al criterio humano.

Además la predicción de la demanda es de vital importancia para los servicios públicos y esta puede dividirse en tres categorías:

1. Predicción a corto plazo.
2. Predicción a mediano plazo.
3. Predicción a largo plazo.

La predicción a corto plazo predice la demanda de carga en un intervalo de tiempo que va desde un día hasta varias semanas. Puede ayudar a estimar los flujos de carga y tomar decisiones que puedan prevenir una sobrecarga. La predicción de carga a medio plazo predice la demanda de carga en el rango de un mes a varios años y proporciona información para la planificación y operación del sistema eléctrico. La predicción de carga a largo plazo predice la demanda de carga en un intervalo de tiempo que va desde un año hasta veinte años [2].

Al realizar pronósticos de la demanda de energía existen varias razones por las que dicho estudio puede complicarse, por ejemplo la característica de los datos, el horizonte de pronóstico y las propias limitantes de la metodología empleada, pero esa necesidad de pronóstico es fundamental para todos los agentes del mercado y se justifica por algunos motivos, dependiendo el tipo de agente, por ejemplo en el caso de los generadores nos ayuda a determinar la demanda a corto plazo, planificar la cantidad de energía a ofrecer el día siguiente, además de realizar un programa para la operación de unidades y fijar precios. En el mediano plazo permite definiciones y análisis de evolución del mercado eléctrico. Para el largo plazo nos da la posibilidad de detectar periodos con exceso o faltante en generación con el fin de tomar acciones y plantear medidas para un ahorro en el consumo de energía, etcétera [3].

Para la distribución y transmisión como se menciona en [3] y [4] nos da la posibilidad de monitorear la red e identificar puntos críticos en el sistema y tomar las acciones necesarias para evitar una falla. Con un estudio de pronóstico (entre otros adicionales) se puede definir un programa de expansión del sistema de transmisión, ya que en la comercialización de energía es de vital importancia analizar el comportamiento del mercado en un mediano y largo plazo. Sirve para definir planes estratégicos para la comercialización de energía. Para el operador del mercado, el estudio del pronóstico de la demanda es de especial importancia por ser una herramienta para la programación y operación de las unidades de generación y con ello se evita sobrecostos en la producción de electricidad [5].

El concepto de la predicción de la demanda no es nuevo y han sido pocos los estudios que han empleado un enfoque sistemático para identificar el modelo idóneo para el problema de la predicción. Una debilidad presente en la mayoría de los enfoques es que, a pesar de las

especificaciones del modelo, tienen cierta parte intuitiva y no son en realidad en base a los datos, pero resulta evidente que al permitir que los datos que determinan la forma del modelo, además del valor de los parámetros, se pueden obtener pronósticos mejorados [6].

Para realizar un estudio de predicción de la demanda de energía eléctrica se debe tener en cuenta también ciertos factores que influyen en el proceso, debido a que esta tiene una característica no lineal y está sujeta a una variedad de variables externa, como condiciones climáticas, temporada del año, variables económicas y demográficas así como las condiciones propias de uso. Esto se evidencia de mejor manera en la Ilustración 1, donde se hace una breve explicación de los factores que influyen en la predicción de la demanda de electricidad.



Ilustración 1 Factores que influyen en el pronóstico de la demanda de energía eléctrica [7]

Además según [8] es fundamental la predicción de la de la carga eléctrica para planificar recursos y para la realización de una futura red inteligente, ya que mejora las características de administración de la demanda y las tareas de gestión en la distribución, por ejemplo con el balance de carga y la reducción del pico de demanda, la capacidad de recuperación a fallas y apagones, entro otras.

En la parte del desarrollo se hace una revisión general de algunos de los métodos que existen para la predicción de la demanda, destacando algunas de sus características principales, ventajas y desventajas.

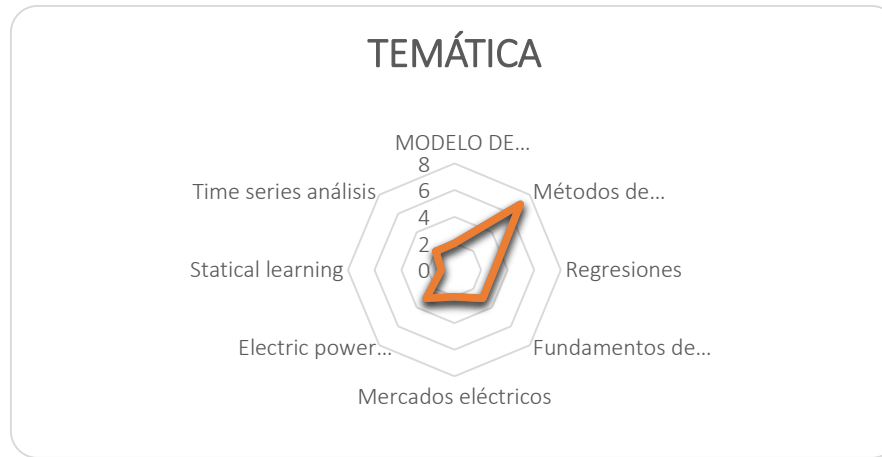


Ilustración 2 Temática de la investigación. Fuente: Autor

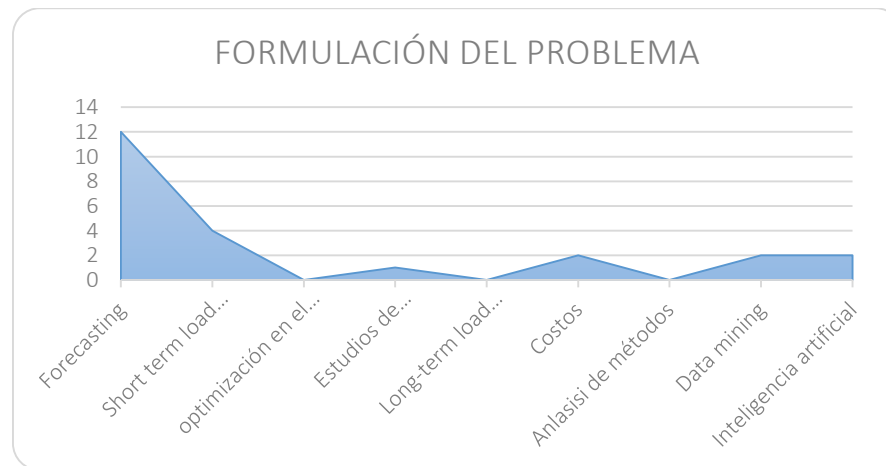


Ilustración 3 Formulación del problema. Fuente: Autor

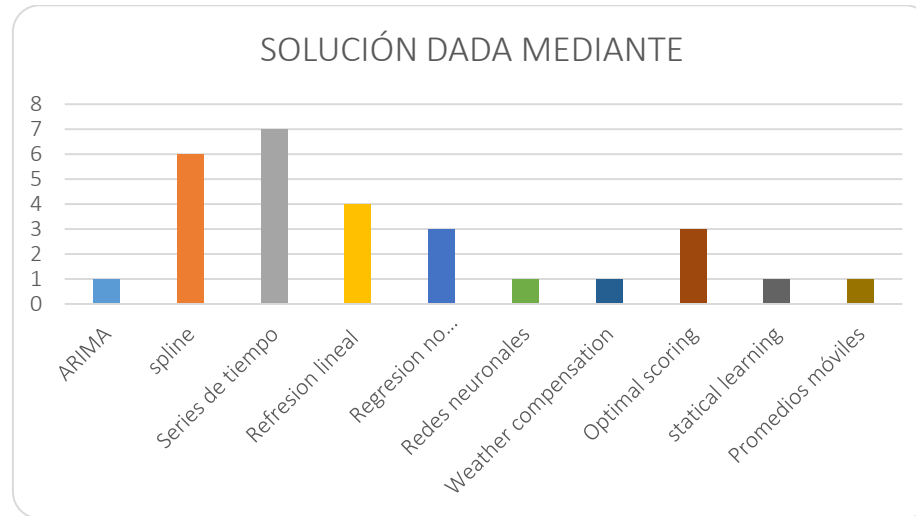


Ilustración 4 Comparativo de soluciones. Fuente: Autor

2. OBJETIVOS

OBJETIVO GENERAL

- Analizar los modelos y métodos actualmente más empleados en la predicción de la demanda residencial de electricidad y conocer la importancia de dicho pronóstico para garantizar la confiabilidad en un sistema eléctrico.

OBJETIVOS ESPECÍFICOS

- Investigar los principales modelos para la predicción de la demanda de energía, tomando en cuenta los factores que influyen en el proceso y las limitaciones propias de los métodos.
- Determinar la importancia de las predicciones de la demanda de electricidad a corto plazo, mediano y largo plazo para el control y programación de los sistemas de energía como dato entrada para algoritmos de planificación de generación y transmisión de energía.
- Analizar los resultados obtenidos en investigaciones para pronóstico de la demanda usando métodos de regresión adaptativa multivariante Spline.

3. METODOLOGÍA

Para analizar el problema de la predicción de la demanda residencial de electricidad se hará énfasis especialmente en la técnica propuesta por Friedman en los años 90, la regresión adaptativa multivariante Spline, empleando un análisis formal del modelo estadístico propuesto, detallando sus ecuaciones y limitaciones propias, así como el desarrollo de una técnica de pronóstico. Para ello se hace uso de herramientas de edición gráfica como visio y paquetes de cálculo computacional como Matlab.

4. DESARROLLO DE LA INVESTIGACIÓN

4.1 MODELOS Y MÉTODOS DE PARA PREDICCIÓN DE LA DEMANDA ELECTRICA

La predicción de la demanda es un aspecto importante en el desarrollo de cualquier modelo para la planificación de un sistema eléctrico y se pueden representar como una demanda máxima, curvas de carga diaria, semanal o anual. Las predicciones a corto plazo son fundamentales para el control y programación de los sistemas de energía y ellas son necesarias como dato entrada para los algoritmos de planificación, generación y transmisión de electricidad.

En el corto plazo, la carga está influenciada principalmente por las condiciones meteorológicas, los efectos estacionales (diarios y ciclos semanales, días festivos) y eventos especiales, además de los cambios climáticos que son importantes en la predicción de la demanda de electricidad [9]. En los últimos años varias técnicas han sido desarrolladas para el pronóstico de la demanda de electricidad. Los modelos estadísticos son ampliamente adoptados para el problema de la predicción de carga, que incluyen modelos de regresión lineal, modelos de procesos estocásticos, suavización exponencial y modelos ARIMA (por sus siglas en inglés) [10]. En esta sección -como se mencionó en la introducción- se hará una revisión general de los modelos enunciados con sus principales características.

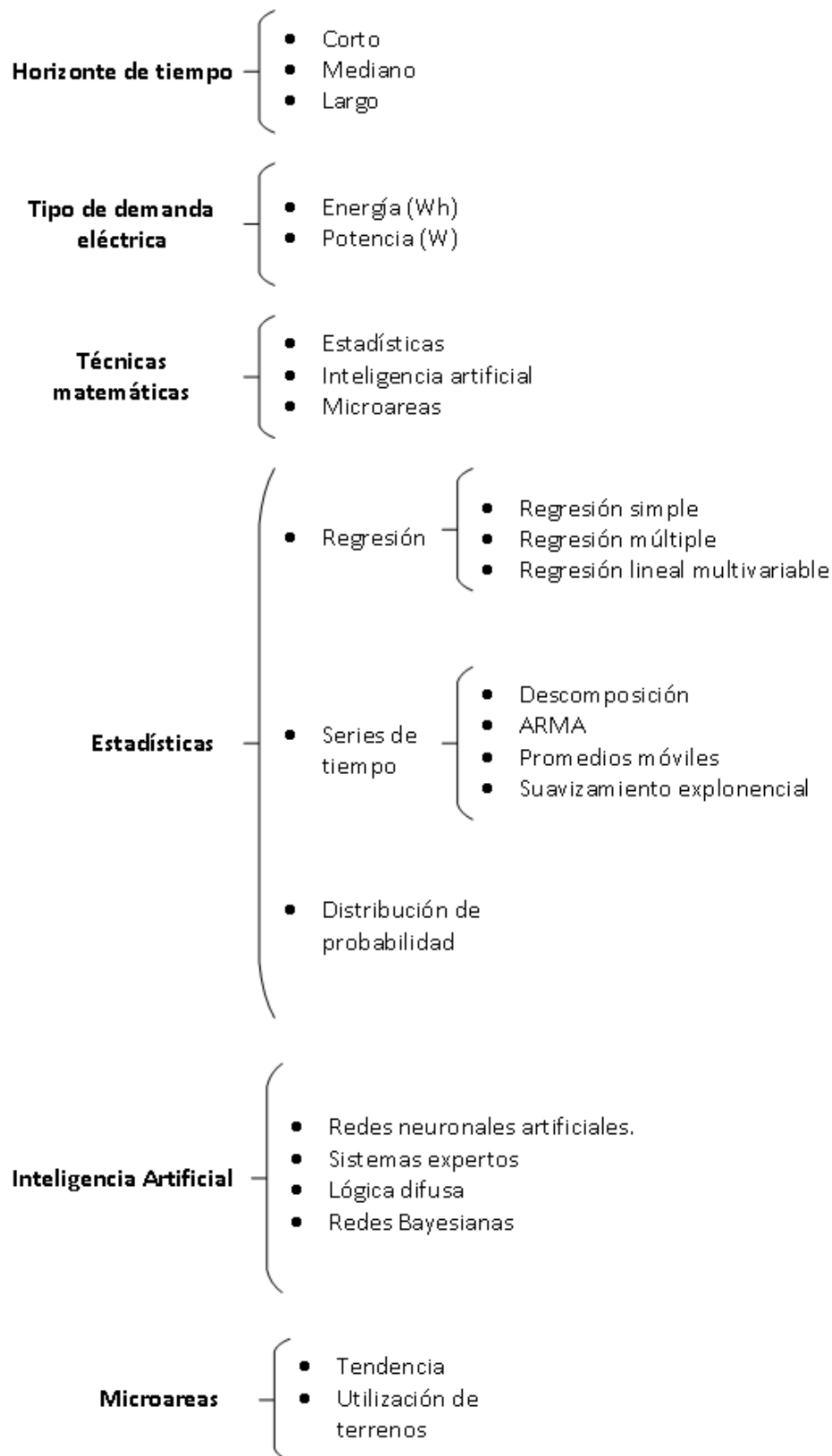


Ilustración 5 Clasificación de los métodos de pronóstico de demanda de energía eléctrica.

Fuente: Autor

4.1.1 Modelos de regresión lineal

Gráficamente el modelo más sencillo para relacionar una variable dependiente con una independiente es una línea recta, debido a ello su denominación de regresión lineal. Un modelo de regresión lineal tiene como objetivo encontrar la ecuación de la recta que ajuste correctamente al conjunto de puntos de datos X , es decir se debe calcular la suma de las distancias al cuadrado entre sus puntos reales y los definidos por la recta estimada en base a las variables introducidas en el modelo, de forma que la estimación ideal será la que minimice estas distancias. En consecuencia, a esta recta se conoce como línea de regresión [11]. La ecuación que describe la recta tiene la forma:

$$y = b_0 + bX + \varepsilon \quad (1)$$

Donde:

b_0 Intersección con el eje Y , cuando $X = 0$

b Pendiente de la recta y es la variación en Y al incremento de X en una unidad.

ε Es la diferencia entre el valor real y el valor pronosticado.

Para hallar los valores de b_0 y b se emplea el método conocido como Método de mínimos cuadrados para encontrar la mejor relación lineal entre la variable dependiente y la variable independiente.

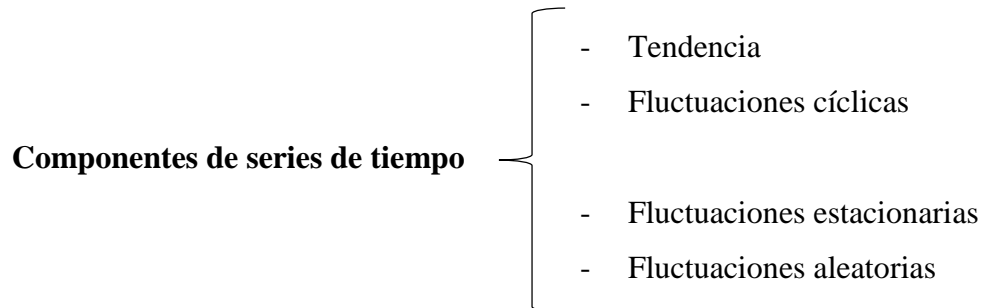
4.1.2 Series de tiempo

La demanda eléctrica *per se* varía continuamente con en el tiempo y constituye una secuencia ordenada de observaciones sobre una variable en particular. Series de tiempo se le denomina a cualquier variable con datos registrados sobre intervalos consecutivos de tiempo [12]. Es un método que requiere la menor información posible, debido a que la única variable independiente es el tiempo y su propósito es cuantificar el consumo de energía eléctrica en intervalos uniformemente distribuidos.

Para analizar una serie de tiempo se grafican sus valores, con lo que es posible realizar un estudio completo de la misma. Obtenidos dichos gráficos nos ayudan a detectar irregularidades o valores fuera de rango en caso de haberlos, por ejemplo la aparición de un nivel, efectos del clima o cambios periódicos. Las discontinuidades en una serie de tiempo resultan primordiales para el análisis de los datos por intervalos. Para un análisis con series de tiempo es importante estudiar sus componentes y realizar cambios para modificarlas o

eliminarlas según sea el caso, con ello se obtienen series de tiempo estacionarias que permiten un análisis de mayor precisión.

En el modelo de series de tiempo algunas componentes:



Una manera para analizar una serie de tiempo, es con un promedio simple, el cual es un método que usa datos recientes para hacer pronósticos de carga y evaluar de igual manera a todos los usuarios. El resultado se obtiene aplicando la media a los valores válidos y se usa para el pronóstico del siguiente periodo. Se recomienda su uso en el caso de tener demandas estables y no existe estacionalidad. Su fórmula es:

$$P_S = \frac{\sum_{t=1}^n Y_t}{n} \quad (2)$$

Donde:

n = Es la cantidad de periodos a determinar el promedio.

Y_t = Demanda en un periodo i .

El término n nos indica el grado de suavización de la predicción, es decir si n es mayor, existe mayor suavización, en cuyo caso es recomendable para datos poco aleatorios, ya que como resultado se obtiene una proyección constante y reacciona más lentamente a las variaciones en la demanda.

4.1.3 Modelos ARIMA

Definido como proceso autorregresivo integrado de media móvil –por sus siglas en inglés- ARIMA es un modelo estocástico utilizado para analizar series de tiempo y su aplicación se debe a Box y Jenkins [13]. Dicho modelo se ha aplicado con éxito en algunos países, por ejemplo en [14] un caso de estudio en Ghana en donde el método ARIMA fue el modelo

usado para proporcionar un pronóstico de siete años de la demanda de electricidad en la ciudad y entre otros, se llegó a la conclusión de que la demanda doméstica y comercial fue aumentando más rápidamente que la demanda en el sector industrial. Una gran desventaja de los modelos ARIMA es que en su implementación se necesita hacer 24 modelos sólo para realizar el pronóstico de un día y realiza una verificación de supuestos en base a cada uno de dichos modelos. Por consiguiente se evidencia que los ARIMA no son tan viables en la predicción de la demanda de energía, ya que desde un punto de vista práctico es una tarea complicada y que requiere mucho tiempo.

4.2 MULTIVARIATE ADAPTIVE REGRESSION SPLINES

La Regresión Adaptativa Multivariable Spline, más conocida por sus siglas en inglés como MARS, fue presentada por Friedman [15] en los años noventa y su principal objetivo es predecir los valores de una variable dependiente o resultado, de un conjunto de variables independientes o predictoras. La propuesta de Friedman es un procedimiento para una regresión no paramétrica adaptativa [16] y no hace ninguna suposición sobre la relación funcional entre las variables dependientes e independientes, en su lugar construye una relación de un conjunto de coeficientes y funciones base que provienen de los datos de la regresión. Fundamentalmente, el método se basa en el concepto de dividir el espacio de entrada en regiones, cada una con su propia ecuación de regresión. Esto la hace particularmente adecuada para problemas con dimensiones superiores de entrada, es decir, con más de 2 variables, que con otras técnicas de regresión seguramente tendríamos inconvenientes.

MARS utiliza expansiones en funciones base lineales por partes o trozos de la forma:

$(x - t)_+$ $(t - x)_+$. El "+" identifica a la parte positiva, por lo que:

$$(x - t)_+ \begin{cases} x - t, & \text{if } x > t, \\ 0, & \text{caso contrario} \end{cases}$$

$$(t - x)_+ \begin{cases} t - x, & \text{if } x < t, \\ 0, & \text{caso contrario} \end{cases}$$

Cada función es lineal a trozos, con un punto de unión en t (obsérvese la Ilustración 6). En [17] se las denomina Splines lineales y la idea es formar pares reflejados para cada X_j de entrada con los puntos de unión del valor observado en cada x_{ij} de esa entrada. Por lo tanto, el grupo de funciones base será:

$$C = \{(X_j - t)_+, (t - X_j)_+\} \quad (3)$$

$$t \in \{x_{1j}, x_{2j}, \dots, x_{Nj}\}$$

$$j = 1, 2, \dots, p.$$

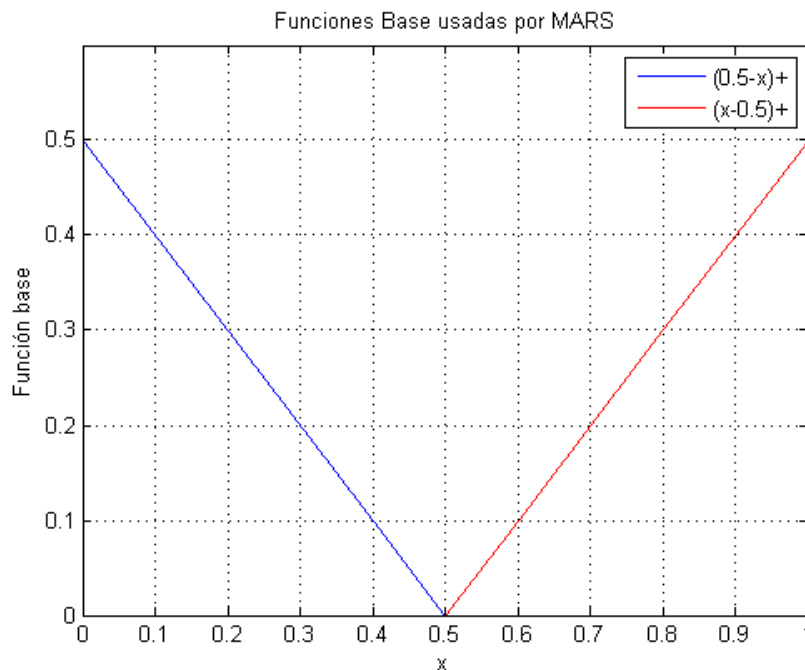


Ilustración 6 Las funciones base $(0.5-x)_+$ (azul) y $(x-0.5)_+$ (rojo) usadas por MARS.

Fuente: Autor

Si todos los valores de entrada son distintos, hay $2Np$ funciones base en total. Hay que tener en cuenta que aunque cada función base depende sólo de un único X_j , es considerada como una función sobre todo el dominio de entrada de \mathbb{R}^p .

Para la construcción de modelo es una regresión lineal por pasos hacia adelante, pero en lugar de utilizar las entradas originales, es posible utilizar las funciones de la serie C , descrita en la ecuación (3) y sus productos. Así, el modelo tiene la forma:

$$f(X) = \beta_0 + \sum_{m=1}^M \beta_m h_m(X) \quad (4)$$

Donde cada $h_m(X)$ es una función en C o un producto de dos o más de dichas funciones. Dada una alternativa para el h_m , los coeficientes β_m se estiman minimizando el residuo de la suma de cuadrados, es decir, mediante una regresión lineal estándar. El desafío en realidad, está en construir las funciones h_m .

Iniciamos únicamente con la función constante $h_0(X) = 1$ en nuestro modelo, todas las funciones en el conjunto C son funciones elegibles y en cada etapa se considera como un nuevo par de funciones base a todos los productos de una función h_m en el conjunto modelo M , con uno de los pares reflejados en C . Entonces añadimos al modelo M el término de forma que se obtenga la mayor disminución en el error de su formación. Aquí $\hat{\beta}_{M+1}$ y $\hat{\beta}_{M+2}$ son coeficientes estimados mediante mínimos cuadrados, junto con todos los otros coeficientes del modelo $M+1$.

$$\hat{\beta}_{M+1} h_l(X) \cdot (X_j - t)_+ + \hat{\beta}_{M+2} h_l(X) \cdot (t - X_j)_+ \quad (5)$$

$$h_l \in M$$

Entonces los resultados obtenidos son agregados al modelo y el proceso continúa hasta que el modelo establecido M contenga un número máximo predeterminado de términos. Al final de este proceso tenemos un modelo grande de la forma de la ecuación (4). Este modelo típicamente sobreajusta los datos, y así se aplica un procedimiento de eliminación en reversa.

El término cuya remoción causa el menor incremento en el residuo del error cuadrático, es borrado del modelo en cada etapa, obteniendo un mejor estimado del modelo \hat{f}_λ para cada número de términos λ .

Se podría usar una validación cruzada para estimar un valor óptimo de λ , pero para los optimizar recursos computacionales del procedimiento MARS, en su lugar utiliza una validación cruzada generalizada. Este concepto se define como:

$$GCV(\lambda) = \frac{\sum_{i=1}^N (y_i - \hat{f}_\lambda(x_i))^2}{\left(\frac{1 - M(\lambda)}{N}\right)^2} \quad (6)$$

Donde:

GCV = validación cruzada generalizada

El valor $M(\lambda)$ es el número efectivo de parámetros en el modelo: esto se explica tanto por el número de términos en los modelos, como el número de parámetros usados en la selección de las posiciones óptimas de los puntos de unión (nodos).

Algunos resultados matemáticos y de simulación sugieren que se debería compensar un valor de tres parámetros para la selección de un nodo en una regresión lineal por tramos, de este modo si hay r funciones base linealmente independientes en el modelo y se seleccionaron K nodos en el proceso, la fórmula es:

$$M(\lambda) = r + cK \quad (7)$$

Donde $c = 3$

El parámetro c del modelo se lo conoce también como penalización.

En consecuencia elegimos el modelo según la secuencia de regresión que minimice la validación cruzada generalizada (GCV).

Para entender porque usar las funciones base lineales a trozos miremos la Ilustración 6, ya que una propiedad fundamental de las funciones en dicha figura es su habilidad para operar localmente y son cero en una parte del rango. Cuando ambas se multiplican, el resultado es diferente de cero solamente sobre una pequeña parte del rango de la función, donde tienen componentes diferentes de cero. Como resultado, la superficie de regresión se construye, usando componentes diferentes de cero sólo donde son necesarias.

La importancia de esta parte radica en que hay que “usar” estos parámetros cuidadosamente cuando tenemos altas dimensiones en la entrada, ya que dichos recursos pueden agotarse rápidamente. El uso de otras funciones base tales como polinomios, produciría un producto distinto de cero en todas partes y no funcionaría igual.

Por otra parte nos ayuda a reducir la extensión del cálculo, consideremos por ejemplo la función M , con cada N de sus pares reflejados para una entrada X_j . Esto requerirá un ajuste

al modelo de N de una sola entrada del modelo de regresión lineal y cada una de ellas usa $Q(N)$ operaciones dando como resultado un total de $Q(N^2)$ operaciones. Sin embargo, podemos aprovechar la forma simple de la función lineal por partes. Primero encajamos el par reflejado con el nodo más a la derecha. Como el nodo se desplaza sucesivamente una posición a la izquierda a la vez, las funciones base difieren de cero sobre la parte izquierda del dominio, y por una constante sobre la parte derecha. Por tanto, luego de cada desplazamiento podemos actualizar el ajuste en $Q(1)$ operaciones y con ello se hará posible probar cada nodo en solamente $Q(N)$ operaciones.

En la Ilustración 7 se muestra un flujograma para realizar pronósticos utilizando regresiones.

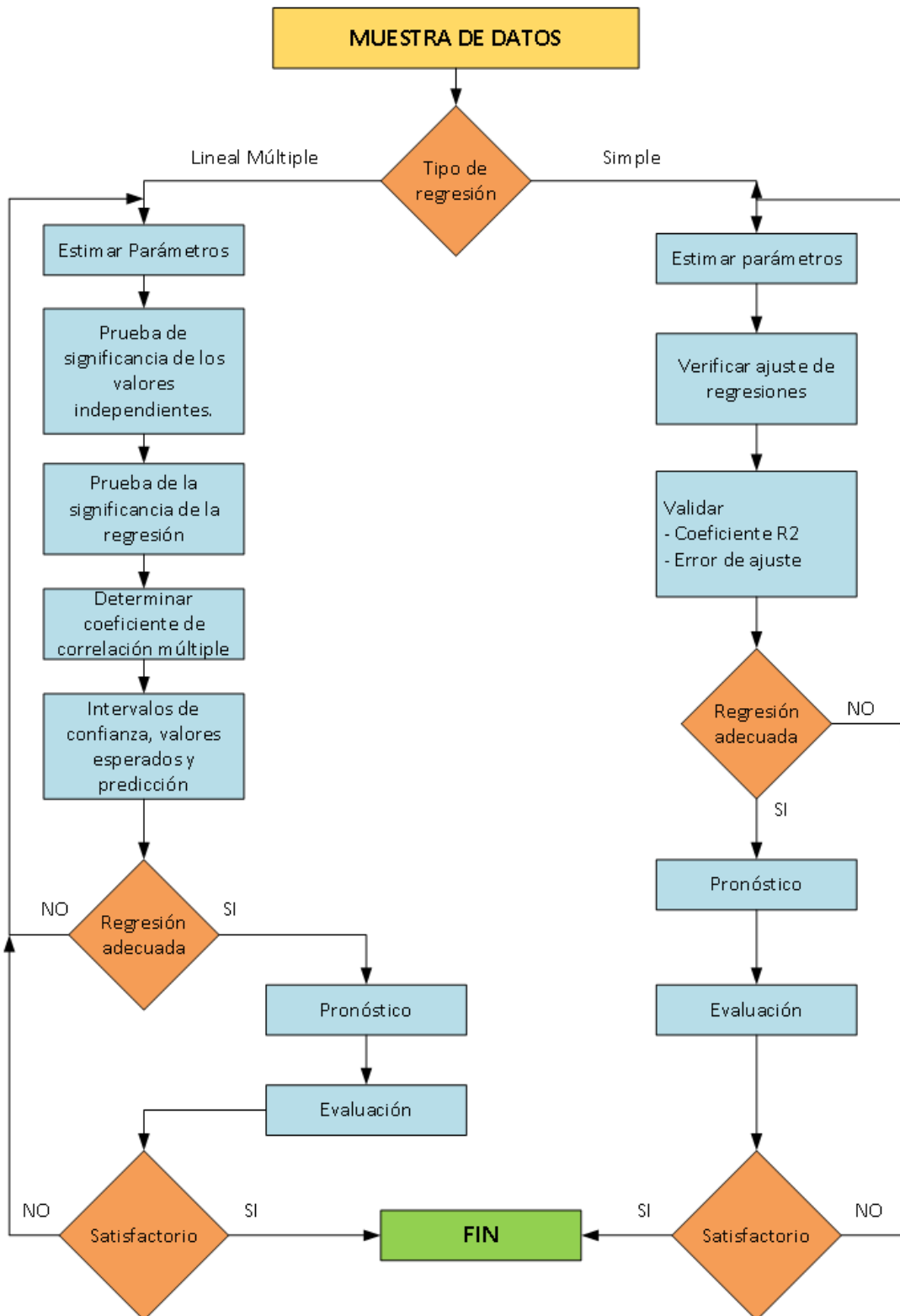


Ilustración 7 Metodología para pronóstico utilizando regresión. Fuente: Autor

4.2.1 Método spline - mínimos cuadrados penalizado

Una medida de ajuste para la función f está dada por $n^{-1} \sum_{i=1}^n (y_i - f(x_i))^2$, mientras que la medida de suavidad relacionada con f , con condiciones de continuidad de funciones elegibles como funciones de regresión W_2^m , está dada por $\int_0^1 (f^m(x))^2 dx$ y se conoce como “penalidad por aspereza”. En el modelo se usa dos funciones, para estimar una función Spline única f_λ con el propósito de minimizar una combinación convexa.

En la expresión (2), el parámetro de suavización es $\lambda = \frac{q}{q-1}$ y representa la tasa de intercambio entre el error residual y la variación local. Entonces si $f_\lambda(x)$ es un Spline natural, para minimizar (2) sobre todas las funciones en W_2^m queda reducido a un espacio finito n-dimensional de splines naturales.

$$(1 - q)n^{-1} \sum_{i=1}^n (y_i - f(x_i))^2 + \lambda \int_0^1 (f^m(x))^2 dx \quad (8)$$

4.2.2 Splines Cúbicos Naturales

El comportamiento de los polinomios al ajustarse a los datos tiende a ser irregular cerca de las fronteras, por ende aplicar una extrapolación puede no resultar conveniente y esto se puede agravar aplicando splines. Los polinomios se ajustan más allá de los nodos frontera se extienden más que los correspondientes polinomios globales en esa región, pero esto puede ser conveniente para realizar un ajuste de splines por el método de mínimos cuadrados. Los splines cúbicos naturales añaden restricciones adicionales, sabiendo que la función es lineal más allá de los nodos de frontera y ello permite liberar cuatro grados de libertad (dos restricciones cada una de las dos regiones de frontera), que se pueden “gastar” de manera más eficiente esparciendo más nodos en la región interior.

Un spline cúbico natural con nodos está representado por k funciones base. Se puede empezar a partir de una base de splines cúbicos, y obtener la base reducida al imponer las restricciones de frontera.

Donde:

$$N_1(x) = 1, N_2(x) = X, N_{k+2}(x) = d_k(X) - d_{k-1}(x) \quad (9)$$

$$d_k(x) = \frac{(X - \xi_k)_+^3 (X - \xi_K)_+^3}{\xi_K - \xi_k} \quad (10)$$

Cada una de estas funciones base puede tender a cero en la segunda y tercera derivada para $X > \xi_K$.

5. RESULTADOS ESPERADOS

Para el problema de predicción se tomará datos de registros de lecturas anteriores, usando para la interpolación el método spline descrito en la sección 4.2. Dicho modelo se lo realiza con el paquete Adaptive Regression Splines para Matlab, un toolbox que sirve para calcular los parámetros especificados del modelo. En el parámetro c de la ecuación (7), para un modelo aditivo, se tomará el valor para dicha variable que se recomienda en [17], pág. 325 correspondiente a 2. Los demás parámetros incluidos los coeficientes, la valoración cruzada generalizada, los nodos, funciones base, etcétera, nos da como resultado del paquete computacional de Matlab.

Para los datos de la regresión se analiza un consumo residencial promedio contando con lecturas de consumo consecutivas de dos años antes, provisto por la empresa de distribución local mostrado en la Tabla 2

Tabla 2 Histórico de demanda residencial para los años 2013 y 2014. Fuente: Autor

2013	DEMANDA (kWh)	2014	DEMANDA (kWh)
Enero	176	Enero	193
febrero	160	febrero	163
Marzo	194	Marzo	164
Abril	176	Abril	152
Mayo	177	Mayo	138
Junio	192	Junio	162
Julio	197	Julio	193

Agosto	197	Agosto	143
Septiembre	243	Septiembre	148
Octubre	190	Octubre	190
Noviembre	219	Noviembre	197
Diciembre	262	Diciembre	218

El objetivo es establecer una predicción de consumo de energía para el año 2015 con el método Spline propuesto. Se muestra primero el patrón de consumo de los años previos en sus respectivas curvas.

En la Ilustración 8 se observa el comportamiento de la carga para el año 2013, conforme a la Tabla 2. Se evidencia que su pico anual está en el mes 12 (diciembre) y el menor en el mes 2 (febrero).

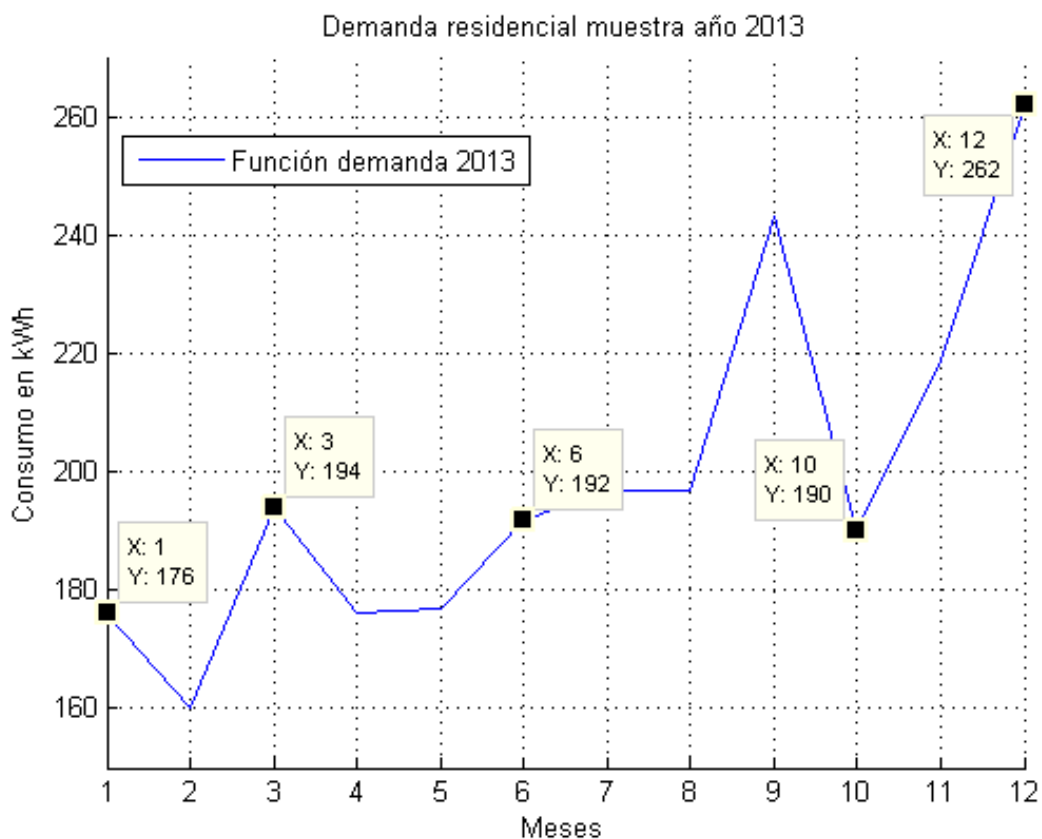


Ilustración 8 Función demanda año 2013. Fuente: Autor

En la Ilustración 9 se muestra el comportamiento de la carga durante el año 2014, en base a la Tabla 2, teniendo su pico en el mes 12 (diciembre), con 218kWh y su más bajo en el mes 5 (mayo) con 138kWh.

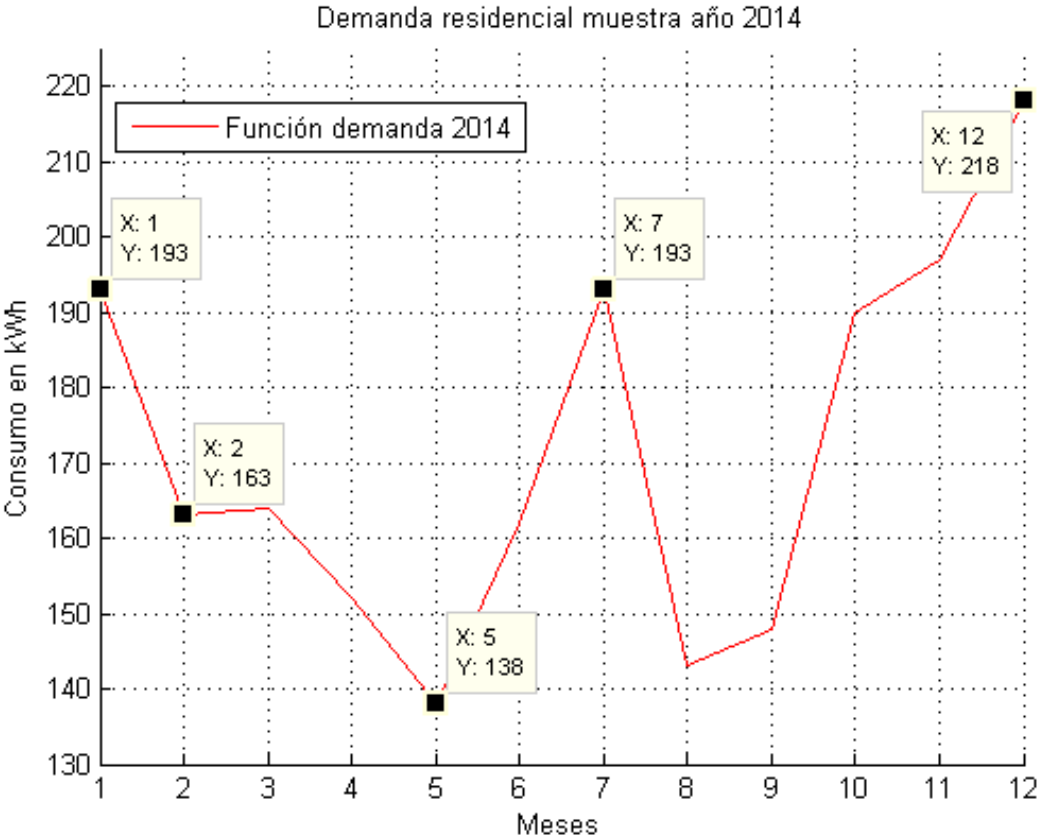


Ilustración 9 Función demanda año 2014. Fuente: Autor

En la Ilustración 10 se hace una comparación de resultados con la demanda prevista para el año 2015 y su interpolación basado en el método spline, usando matlab.

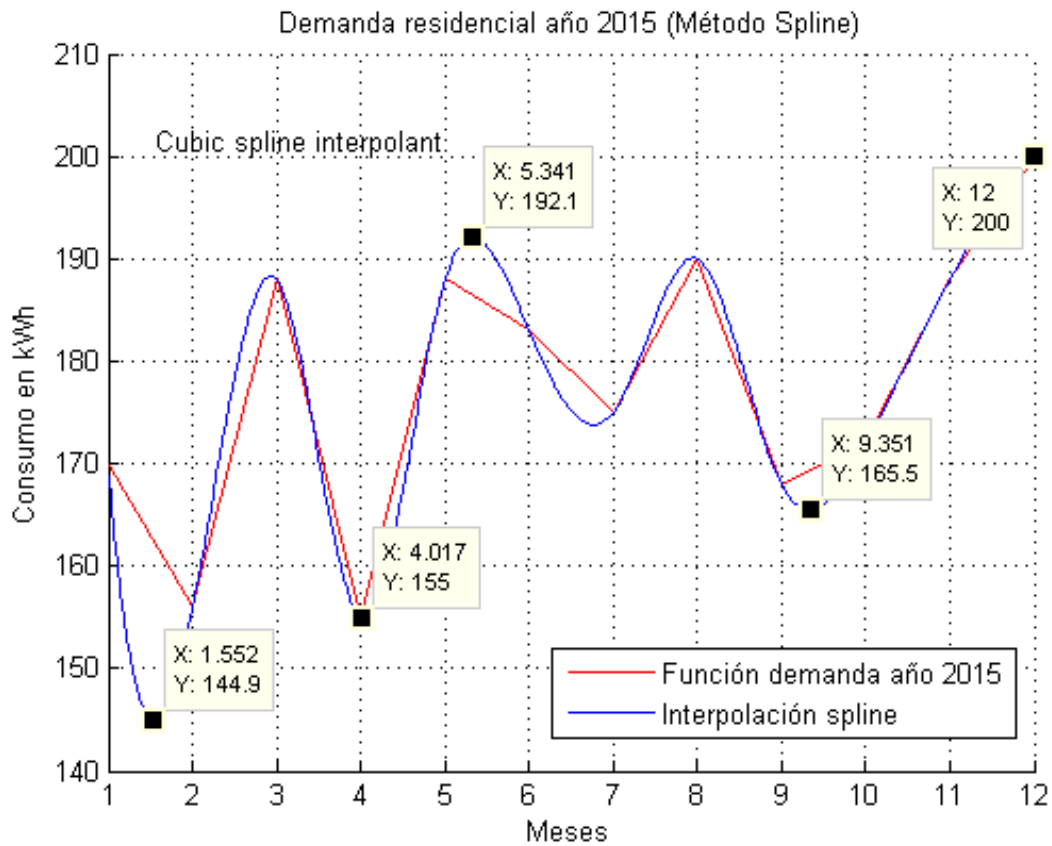


Ilustración 10 Método de interpolación spline para el año 2015. Fuente: Autor

Se observa que el pico mantiene una tendencia de comportamiento de los años 2013 y 2014, ubicándose en el mes 12 (diciembre) con 200kWh. Con la interpolación Spline, el punto de menor consumo se ubica en el mes 1.552 del año, con 144.9kWh. La suavización considera restricciones entre los nodos de la curva, el parámetro c definido y los parámetros adicionales propios del modelo explicado anteriormente.

6. ESTRATEGIA PARA LA DIVULGACIÓN DE LOS RESULTADOS

Dentro de las estrategias de divulgación se plantean las siguientes:

1. Informe escrito final como producto de la tesis. El informe será entregado a biblioteca UPS
2. Jornada Científica de Ingeniería Eléctrica.

7. SECTORES BENEFICIADOS

Los sectores beneficiados con el proyecto son los siguientes:

1. Universidad Politécnica Salesiana
2. Grupos de investigación vinculados con la optimización de la demanda.
3. Sectores públicos y privados relacionados al tema de investigación.

8. CONCLUSIONES

Para realizar un estudio de pronóstico de la demanda eléctrica se debe tener en cuenta factores que intervienen en el análisis, como el nivel socioeconómico, sector donde se hará el pronóstico, número de clientes, condiciones climáticas entre otras. El objetivo es llegar a una estimación que incluya la mayoría de sus variantes relacionadas.

De la comparación de los modelos descritos en este estudio se concluye que es determinante la elección correcta del modelo, además de otras consideraciones. Dimensionar sus alcances y contar con los datos históricos del consumo en kWh es fundamental al momento de llevar a cabo el análisis.

Como se pudo observar, cada modelo tiene sus alcances, limitaciones y formulación para obtener predicciones que se acerquen lo más posible a la realidad. Si bien pueden diferenciar, es debido a los ajustes en los parámetros, tipo de regresión, suavización y forma polinomial de los mismos.

Considerar la literatura técnica existente, analizando los estudios y casos previos en los que se han realizado similares pronósticos nos ayuda a determinar los beneficios, limitaciones y recomendaciones de los autores de dichos estudios. En las referencias mencionadas se obtiene información valiosa para seguir en el desarrollo del modelo, considerando sus experiencias.

9. *RECOMENDACIONES*

Efectuar un seguimiento adecuado de la evolución en la curva de carga a través del tiempo nos dará una mejor estimación para el proceso. Se puede contar con datos fiables y en el volumen requerido, contando en lo posible con un programa informático y una base de datos exportable para el cálculo computacional.

El intervalo de registros del consumo eléctrico residencial debería en un futuro ser de al menos cada 15 minutos. Con la implementación de una red inteligente se lograría este objetivo. La factibilidad de dicha implementación sería un objeto de estudio para futuros trabajos.

Realizar registros de carga por parte de la empresa de distribución local a los clientes para – entre otros propósitos- cumplir estándares de calidad que influirán en los datos proporcionados para los estudios y mejora de resultados.

10. REFERENCIAS

- [1] A. F. Barrientos, J. Olaya, and V. M. González, “Un modelo spline para el pronóstico de la demanda de energía eléctrica,” *Rev. Colomb. Estad.*, vol. 30, no. 2, pp. 187–202, 2007.
- [2] I. Azmira, A. Razak, S. Majid, and H. A. Rahman, “Short Term Load Forecasting Using Data Mining Technique,” no. PECon 08, pp. 139–142, 2008.
- [3] M. García, “Predicción de demanda de energía en colombia mediante un sistema de inferencia difuso neuronal,” pp. 15–24, 2005.
- [4] A. Valencia, C. Lozano, and C. Moreno, “Modelo de promedios móviles para el pronóstico horario de potencia y energía eléctrica,” no. 29, pp. 96–105, 2007.
- [5] V. M. Rueda, J. David, V. Henao, C. Jaime, and F. Cardona, “Electricidad Usando Modelos No Lineales Recent Advances in Load Forecasting Using Nonlinear Models,” *Dyna*, vol. 167, no. June 2015, pp. 36–43, 2011.
- [6] B. S. Maybee, “Time series forecasting of utility load duration curves,” vol. 4, no. 1, pp. 4–8, 2000.
- [7] T. Gonen, *Electric Power Distribution Engineering, Third Edition*, 3 edition. {CRC} Press.
- [8] M. Rossi and D. Brunelli, “Electricity demand forecasting of single residential units,” *2013 IEEE Work. Environ. Energy Struct. Monit. Syst. EESMS 2013 - Proc.*, no. Dii, 2013.
- [9] T. W. S. Chow and C. T. Leung, “Neural network based short-term load forecasting

using weather compensation,” vol. 11, no. 4, pp. 1736–1742.

- [10] S. Fan, S. Member, and R. J. Hyndman, “Forecasting Electricity Demand in Australian National Electricity Market,” no. February, pp. 1–4, 2012.
- [11] A. Ariza, “Métodos Utilizados Para el Pronóstico de Demanda de Energía Eléctrica en Sistemas de Distribución,” 2013.
- [12] J. Murillos, a Trejos, and P. Carvajal, “Estudio del pronóstico de la demanda de energia electrica, utilizando modelos de series de tiempo,” *Sci. Tech.*, vol. 23, no. 23, pp. 37–40, 2003.
- [13] G. Box, George; Jenkins, *Time Series Analysis: Forecasting and Control*, 4 edition. Wiley.
- [14] S. Katara, A. Faisal, and G. M. Engmann, “A Time Series Analysis of Electricity Demand in Tamale , Ghana,” vol. 4, no. 6, pp. 269–275, 2014.
- [15] H. Friedman, “MULTIVARIATE ADAPTIVE REGRESSION SPLINES *,” no. August, 1990.
- [16] R. T. and A. B. Trevor Hastie, “Flexible Discriminant Analysis by Optimal Scoring.” 1993.
- [17] J. F. Trevor Hastie, Robert Tibshirani, “The Elements of statistical learning.”