

# Análisis de Variables Temporales para la Predicción del Consumo Eléctrico

D. Lizondo<sup>3</sup> V. Jiménez<sup>1</sup> F. Villacis<sup>1</sup> A. Will<sup>2</sup> S. Rodríguez<sup>1</sup>

<sup>1</sup>Grupo de Investigación en Tecnologías Informáticas Avanzadas - U.T.N. - F.R.T.  
<sup>2</sup>Facultad de Ciencias Exactas y Tecnología - Universidad Nacional de Tucumán  
<sup>3</sup>Consejo Nacional de Investigaciones Científicas y Técnicas  
E-mail: diego.lizondo@gitia.org; adrian.jimenez@gitia.org; adrian.will@gitia.org;  
sebastian.rodriguez@gitia.org, fervillacis@gmail.com

## Resumen

El problema de la predicción de consumo eléctrico a corto plazo o Short Term Load Forecasting (STLF), es un tema de capital importancia para las empresas de energía en la actualidad, ya que permite un manejo más eficiente, permitiendo un mejor aprovechamiento de los equipos y recursos. La predicción de la demanda es un problema complejo, ya que está relacionada a factores económicos, climáticos, temporales, y su comportamiento varía de una sociedad a otra. Cada uno de estos factores aporta determinadas variables que pueden ser representadas de diferentes maneras, en particular las temporales. Se plantea en este trabajo la hipótesis que el método utilizado para presentar las variables temporales a un sistema de predicción de consumo eléctrico afecta los resultados. Para verificar la hipótesis planteada, consideramos diferentes métodos de representación de estas variables, aplicados al problema de predicción de valores diarios de consumo eléctrico en la provincia de Tucumán, Argentina. La división de la variable temporal en variables día, día de la semana, mes y año en forma individual para cada periodo involucrado en el problema, resultó ser el método más conveniente, obteniendo una mejora de hasta el 10,56% respecto de otros métodos considerados.

**Palabras clave**—Consumo Eléctrico, Short Term Load Forecasting, Selección de variables, Regresión Lineal.

## Abstract

Short Term Load Forecasting (STLF) currently is a major importance issue for Energy Companies. STFL allows a more efficient manage and use of resources and equipment. The electric demand prediction is a complex issue, since it depends or is related to economic factors, climate and time to mention a few. Furthermore, its behaviour changes from one society to another. Each factor provides a particular variable that could be presented in different forms, particularly the time variables. In this paper we present the hypothesis that the way an input variable is introduced to an energy prediction system affects the result. To validate this hypothesis, different methods to represent time variables were considered and applied to the prediction problem of daily electric consumption in Tucumán, a province of Argentina. The separation of the time variables into single variables representing the day, day of the week, month and year for each period involved into the problem, was the most convenient method. The improvement of this method was about the 10 % in comparison to the others.

**Index terms**—Energy Consumption, Short Term Load Forecasting, Variable Selection, Linear Resegion.

Recibido: 08-09-2014, Aprobado tras revisión: 21-11-2014

Forma sugerida de citación: Lizondo, D.; Jiménez, V.; Villacis, F.; Will, A.; Rodríguez, S. (2015). "Análisis de Variables Temporales para la Predicción del Consumo Eléctrico". Revista Técnica "energía". N° 11, Pp. 5-12. ISSN 1390-5074.

## 1. INTRODUCCIÓN

En el nuevo escenario global y local, el advenimiento de las redes de distribución inteligente o *Smart Grids*, permite la recolección a tiempo real de datos sobre el estado de funcionamiento de la red eléctrica. En base a esta disponibilidad de datos, resulta factible y conveniente la predicción de consumo a corto plazo, desde unas horas y hasta una semana. La predicción a corto plazo está estrechamente relacionada con el problema de los picos de consumo, en los cuales se observa un fuerte aumento de la demanda en un breve período de tiempo. Esto resulta en un serio problema que, de no ser adecuadamente previsto y gestionado, provoca pérdidas económicas, daños en los equipos a diferentes niveles, y cortes de servicio o aumento de precio (ambos como una forma de desalentar el consumo en economías que permiten consumo por franjas horarias, y debido a la necesidad de reparar o reponer equipos dañados). Así, esta predicción juega un importante rol para las empresas del sector eléctrico especialmente en lo que se refiere a distribución, debido a que permite tener un mejor panorama para tomar decisiones estratégicas y operativas [4].

La demanda de energía es entonces un tema de gran importancia pero de difícil solución por distintos motivos: al ser la demanda eléctrica la suma de todos los consumos individuales de los nodos de la red, determinar los niveles de demanda eléctrica son un proceso no estacionario, aleatorio, compuesto de muchos y diversos componentes individuales. Los factores que pueden llegar a influir se pueden clasificar en: factores económicos, temporales, climáticos y efectos aleatorios tanto separados como compuestos [5]. El comportamiento del sistema se ve fuertemente afectado también por la escala a la que se realiza la medición, siendo más suave a medida que se aumenta el nivel de agregación de consumidores: El consumo eléctrico a nivel ciudades o superior resulta razonablemente estable, mientras que el consumo eléctrico de un grupo pequeño de consumidores o casas aisladas es un proceso fuertemente aleatorio y ruidoso [6], [7].

Existen diversos trabajos que abordan el problema de la determinación de las variables de entrada necesarias para STLF. Wang et al. [8] para seleccionar las variables más significativas. Emplea una clasificación en grupos con límites “blandos” en base a similitudes con otros, sin transiciones bruscas entre agrupaciones. Esta propiedad permite que un *elemento* pueda pertenecer a más de un grupo. Estos conjuntos son los que se utilizan como entradas para el sistema de predicción. Por su parte, en [3] para seleccionar las variables de entrada se utiliza la *maximum conditional entropy*, que determina la relevancia de cada variable en la predicción de la demanda eléctrica, seleccionando las mejores. No obstante, en [9] se emplea el método de *phase-*

*space embedding* para identificar un conjunto de variables “independientes” que están relacionadas con la serie de tiempo de la demanda eléctrica. En los casos mencionados, los métodos fueron utilizados para determinar las variables de entradas a sistemas basados en Redes Neuronales Artificiales. En [10], se propone un sistema híbrido PSO-SVR, utilizando un algoritmo de *Particle Swarm Optimization* (PSO), para seleccionar las variables óptimas de entrada, limitando el número de variables de entradas en Support Vector Regression (SVR) para STLF. Por otra parte, Ceperic et al. [2] para realizar la selección de variables. A su vez, Borges et al. [11], plantea que en la selección de variables las *m* mejores características no conducen necesariamente a una buena selección, y que parte de la información proporcionada por las variables eliminadas se pierde. Por lo tanto, los autores emplean *Random Forest Method* para reducir el conjunto de variables de entrada.

Restringiéndonos al aspecto temporal, existen numerosos métodos para representar variables relacionadas a la fecha utilizada en los modelos de predicción de consumo eléctrico. Entre los más usados, se encuentran diversas variaciones de *DOY* (Day-Of-Year), como el número de muestra, numeración secuencial y variables no lineales derivadas de la fecha. En base a esto introducimos la hipótesis de que el método utilizado para incluir variables temporales en un sistema de predicción de consumo eléctrico influye en los resultados. Se propone entonces una comparativa de diferentes métodos de representación de variables temporales para determinar cuál es la más conveniente para la predicción del consumo eléctrico a corto plazo. Como caso de estudio se eligió el consumo eléctrico de la provincia de Tucumán en el período 2006 a 2011. Se emplea un método de regresión lineal simple para validar la hipótesis planteada, ya que el objetivo del presente artículo es determinar el impacto en el modo en que se presentan las variables temporales y no construir un sistema de predicción.

Este trabajo está organizado de la siguiente manera: en la sección se describen los materiales y la metodología utilizada, los errores empleados, datos y la herramienta seleccionada para la predicción a corto plazo; en la sección 3 se detallan los resultados obtenidos utilizando diferentes *delays* en la predicción *one-day-ahead*; finalmente en la sección 4 se presentan las conclusiones y propuestas de trabajos futuros.

## 2. MATERIALES Y METODOLOGÍA

En este trabajo proponemos como hipótesis que la forma de incluir variables temporales afecta la calidad de la predicción del consumo eléctrico. Para validarla, proponemos una comparativa entre los resultados de los experimentos de STLF utilizando diferentes

métodos para representar variables temporales. El análisis se realiza sobre datos reales de la provincia de Tucumán.

Basándonos en el uso recurrente de la variable *DOY* (ver sección ), usaremos este enfoque como primer método y punto de partida, con el fin de establecer el primer punto de comparación. Como segundo método proponemos una separación de la variable fecha, logrando descomponerla en cuatro variables individuales: año, mes, día, y día de la semana. Por último, se emplea una adaptación del método descrito en [12], donde el autor utiliza un

pre-procesamiento (operaciones no lineales) sobre las variables temporales de entrada. Este último enfoque se asemeja al anterior, ya que denota una adición de variables temporales derivadas de las originales.

De esta manera fueron conformados tres grupos de variables para las diferentes pruebas detalladas en la Tabla 1. El cuarto grupo se corresponde a las variables climáticas y variables con información de los días atípicos que serán utilizadas en todas las pruebas (para más detalles, ver subsección ). Se procede entonces a probar la combinación de éstos tres métodos siguiendo lo realizado por [13].

Tabla 1: Descripción estadística de la base de datos utilizados

Grupo	Origen	Variable	Mínimo	Máximo	Promedio	STD
1	Fecha	Día del año (DOY)	1	366	188,5	107,38
2	Fecha	Año	2006	2011	2008,30	1,72
		Mes	1	12	6,70	3,52
		Día (DOM)	1	31	15,72	8,82
		Día de la semana (DOW)	1	7	3,99	2,00
3	Fecha	$sDOY = \sin((2\pi/365)*DOY)$	0	1	0,28	0,45
		$cDOY = \cos((2\pi/365)*DOY)$	0	1	0,04	0,19
		$sDOM = \sin((2\pi/31)*DOM)$	0	1	0,04	0,19
		$sDOM = \cos((2\pi/31)*DOM)$	0	1	0,04	0,19
		$sDOW = \sin((2\pi/7)*DOW)$	0	1	0,04	0,19
		$sDOW = \cos((2\pi/7)*DOW)$	0	1	0,04	0,19
4	Días atípicos	Fin de semana	0,00	1,00	0,28	0,45
		Feriado Nacional	0,00	1,00	0,04	0,19
	Clima – Santa Ana	Temp. Máxima	2,3	42,7	26,16	6,18
		Temp. Máxima al cuadrado	5,29	1823,29	722,69	312,77
		Temp. Mínima	-3,6	24,5	13,22	6,31
		Temp. Mínima al cuadrado	0,01	600,25	214,56	151,52
		Presión	949,1	989,6	966,07	5,75
		Humedad Relativa Media	31	96	71,97	12,93
		Radiación Solar	12,3	337,5	166,80	83,88
	Clima – Pueblo Viejo	Temp. Máxima	2,5	42,8	25,91	6,22
		Temp. Máxima al cuadrado	6,25	1831,84	709,93	310,86
		Temp. Mínima	-2,4	25,6	13,97	5,69
		Temp. Mínima al cuadrado	0,01	655,36	227,67	145,8
		Presión	944,6	984,4	961,20	5,68
		Humedad Relativa Media	29	95	69,19	13,39
		Radiación Solar	12,9	362,3	170,41	85,23
	Clima – Monte Redondo	Temp. Máxima	3,5	67,62	27,43	6,52
		Temp. Máxima al cuadrado	12,25	2007,04	792,76	347,88
		Temp. Mínima	-6,7	25,6	12,86	6,62
		Temp. Mínima al cuadrado	0	655,36	209,27	153,82
		Presión	0	989,4	964,32	29,57
		Humedad Relativa Media	0	97	67,31	15,63
		Radiación Solar	0	965,29	196,28	87,01
	Clima – Casas Viejas	Temp. Máxima	0,17	180,43	26,67	7,38
		Temp. Máxima al cuadrado	0	1944,81	748,59	329,68
		Temp. Mínima	-5,8	207,39	13,00	8,23
		Temp. Mínima al cuadrado	0	635,04	208,94	153,8
		Presión	-374,74	991,3	961,64	71,13
Humedad Relativa Media		-28,79	311,28	66,22	15,78	
Radiación Solar		0	828,92	193,42	87,92	
Clima – El Colmenar	Temp. Máxima	0,86	65	26,08	6,36	
	Temp. Máxima al cuadrado	0,4	1927,21	717,70	317,04	
	Temp. Mínima	-1,5	27,7	14,53	5,72	
	Temp. Mínima al cuadrado	0	767,29	243,97	155,3	
	Presión	0	978,9	953,16	49,41	
	Humedad Relativa Media	0	959,11	65,45	26,83	
	Radiación Solar	0	948,4	184,50	87,12	
5	Consumo Eléctrico	Potencia [MWh]	3615,78	10644,06	5907,13	875,46

Dado que el consumo eléctrico varía continuamente en el tiempo, resulta factible aplicar diferentes técnicas y metodologías de series de tiempo para predecir el consumo basado en datos históricos disponibles. Una de estas técnicas consiste en el uso de información actual y/o anterior para predecir variables futuras [14]. En nuestro caso de estudio proponemos el uso de datos de días anteriores, tanto de la demanda de energía eléctrica como de variables climatológicas de la región y datos relacionados a fechas.

A su vez, en pos de mejorar la precisión de la predicción, se aprovechó el hecho que existen variables cuyos valores futuros pueden ser conocidos con antelación. Este es el caso de las variables de tipo fecha y datos relacionados (días feriados, festivos o eventos especiales regionales). Cabe destacar que la presencia de días atípicos como los fines de semana y feriados influyen en el consumo de electricidad [15]. Por tal motivo, proponemos agregar esta información como variables de entrada comunes para todas las pruebas. La ventaja que nos brinda la posibilidad de agregar información del día de la predicción es que produce una disminución del error que pueda generarse por posibles cambios de tendencias que se vean reflejados o alterados por la presencia de éste tipo de información.

En cuanto a la separación de datos, al tratarse de una serie de tiempo, el conjunto de validación se conformó del 30% de las últimas muestras de la serie en forma consecutiva, es decir las muestras más recientes. A su vez, el restante 70% fue empleado para el entrenamiento. Como parte de las pruebas realizadas, se empleó el procedimiento de separación de variables descripto en [13] que se acerca más al uso real de las empresas de electricidad. Sin embargo, las mejoras encontradas no son significativas.

Con el objetivo de evaluar los resultados obtenidos con el modelo seleccionado, se analiza el error obtenido a través de distintas métricas: Root Mean Squared Error (RMSE) que al estar expresado en la misma unidad de medida que la variable a estimar facilita su interpretación; Relative Error (RE) expresado porcentualmente; Mean Bias Error (MBE) que permite analizar si existe una subestimación o una sobreestimación en la predicción del consumo eléctrico; y el Coeficiente de Correlación Lineal de Pearson (R) que ayuda a determinar el grado con que los datos siguen la tendencia general del modelo.

## 2.1. Datos Utilizados

Para el desarrollo de las pruebas, fue empleada información de bases de datos provenientes de diferentes fuentes. Los datos climáticos fueron recolectados de 5 diferentes estaciones meteorológicas, pertenecientes a la Estación Experimental Agroindustrial Obispo Colombres (E.E.A.O.C.). Los datos de consumo

eléctrico, por otro lado, fueron suministrados por la Empresa de Distribución Eléctrica de Tucumán S.A. (E.D.E.T.s.a.), y corresponden a mediciones de la demanda de consumo eléctrico de la provincia. Por último, los datos referentes a fechas y días feriados de la República Argentina, fueron adquiridos a través del sitio web de la Dirección Nacional de Asuntos Políticos, dependiente del Ministerio del Interior y Transporte de la Nación Argentina.

Los datos empleados corresponden a períodos comprendidos entre el 01-01-2006 al 29-12-2011 y contiene muestras diarias de las variables detalladas en la. Además se agregan las temperaturas mínimas y máximas al cuadrado perteneciente a cada una de las estaciones de manera similar a lo observado en [13]. Pruebas preliminares determinaron que las variables cúbicas no influyen mayormente en este caso.

En el procedimiento de análisis inicial de los datos, y como es usual en las redes de sensores distribuidas, se encontraron registros con valores ceros o faltantes, tanto en los datos climáticos como los de consumo eléctrico. Esto se debe posiblemente a problemas en los equipos de medición o en los procesos de adquisición de información. Debido a que la cantidad de datos afectados no es significativa, se optó por realizar un preprocesamiento a través del cual se eliminaron los registros que presenten alguna anomalía. El porcentaje de celdas con datos faltantes es del 8,6%, dejando como resultado 1885 muestras utilizables de un total de 2188.

## 2.2. Regresión Lineal para STLF

La Regresión Lineal es uno de los métodos estadísticos más utilizados, en parte debido a su fácil y sencilla interpretación sobre el modelo que se desea adquirir. Debido a tales características y teniendo en cuenta que el objetivo del presente trabajo es validar nuestra hipótesis, no obtener la mejor predicción posible, se utilizó ésta técnica para estimar los valores de consumo eléctrico, de igual modo que [13], [16], [17], entendiendo que para versiones definitivas del sistema es necesario utilizar *Weighted o Multivariate Linear Regression* [14].

Resulta importante aclarar el hecho de que en el problema abordado, muchas de las variables de entrada presentan una correlación lineal, de manera que la mayoría de los sistemas lineales involucrados en el trabajo son mal condicionados. Este tipo de sistemas lineales producen una fuerte variación en la salida ante pequeños cambios en la entrada, haciendo que la solución no sea adecuada en muchos casos. Para resolver este problema, utilizamos pseudoinversa de Moore-Penrose [18], que es capaz de obtener buenas soluciones en el caso de sistemas mal condicionados.

### 3. RESULTADOS OBTENIDOS

En las pruebas realizadas se contó con muestras diarias, teniendo como objetivo la predicción de un día hacia adelante (One-Day Ahead) sobre la variable consumo eléctrico. Para mejorar la estimación, además de tomar muestras actuales, se considera las muestras de varios días anteriores consecutivos o delay (series de tiempo auto-regresivas).

En la Fig. 1, se muestra la variación del error utilizando la variable *DOY* y de días atípicos, y las variables climáticas, para los 4 tipo de errores utilizados como métricas de comparación. Se puede observar que el mejor resultado se logra con un delay 2, es decir considerando muestras actuales y las de 1 día hacia atrás. Sin embargo, para valores mayores de delay, el error de entrenamiento disminuye mientras que el de validación tiende a aumentar. Esto se produce debido a que el aumento de variables que implica un delay grande provoca un aumento en la complejidad del modelo de regresión lineal, lo que a su vez genera un sobreentrenamiento (overfitting) del sistema, tal como lo expuesto en [19].

En la Tabla 2 se detallan los errores obtenidos a nivel provincia, para los conjuntos de entrenamiento y validación respectivamente. Se puede denotar, tanto en entrenamiento como en validación, que el mejor resultado se obtuvo mediante el uso de las variables de los grupos 2-4 como entrada del sistema. Fue lograda una mejora de un 10,56 % en el Error Relativo respecto a la no utilización de variables temporales. Las otras alternativas no aportan una mejora significativa.

Como se puede ver en la Fig. 2, la tendencia del consumo eléctrico puede ser estimada correctamente por el sistema lineal. En la Fig. 3 y Fig. 4 se muestra que el sistema capta correctamente la linealidad del problema y que muy pocos datos están por fuera de la franja de error del 10% (zona entre las líneas punteadas) que son de esperarse en épocas de calor en Tucumán. Estos valores externos corresponden a los picos de carga. El tratamiento de estos datos anómalos requiere el desarrollo de sistemas específicos que escapen al objetivo de este trabajo.

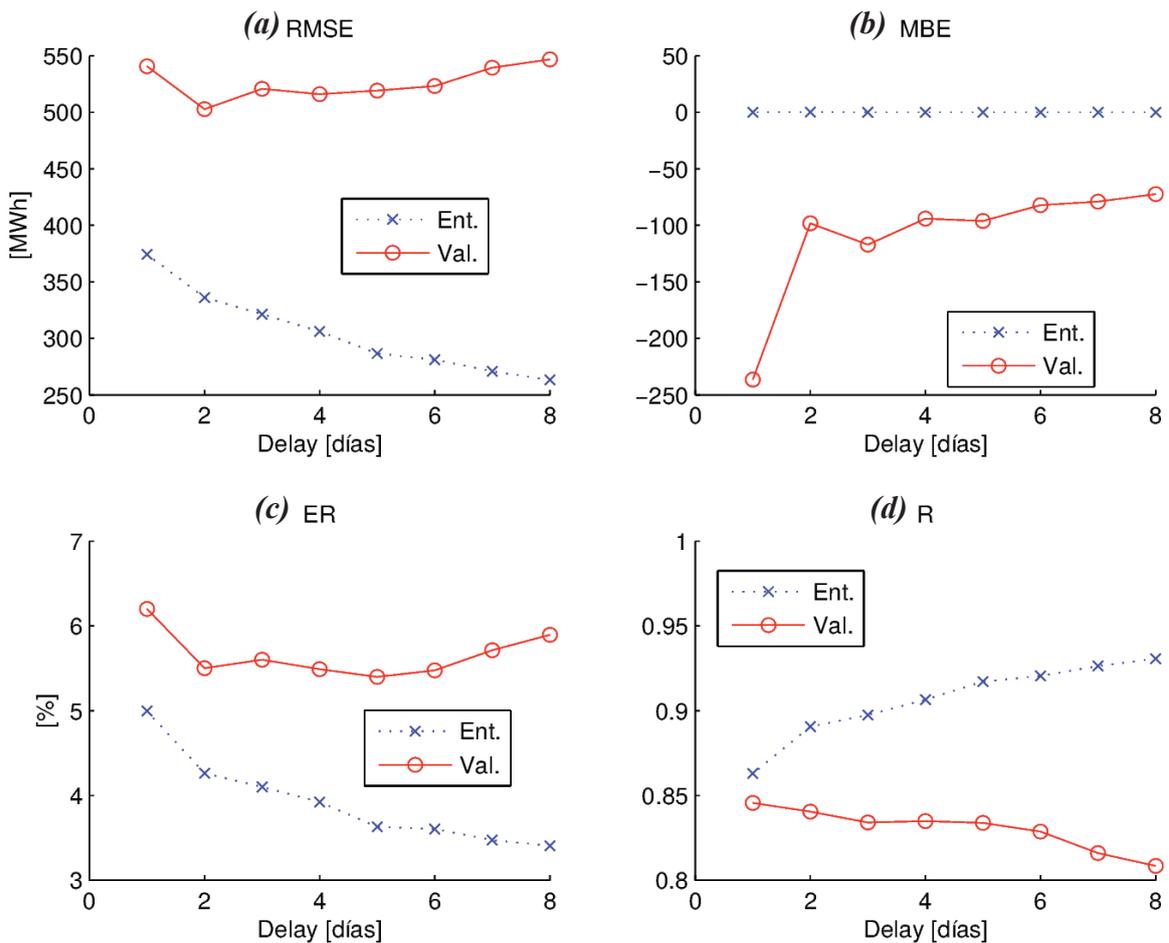


Figura 1: Errores obtenidos para diferentes valores de delay

Tabla 2: Errores de Entrenamiento y Validación

Grupos de variables	Descripción	SSE	MBE	R	RMSE	Err.Rel
4	Variables Básicas	148568740,4	0,0245	0,8896	337,2812	4,26
2-4	Descomposición en fechas	127837417,4	0,0000	0,9059	312,8653	3,81
3-4	Componentes no lineales	160526358,8	0,0171	0,8802	350,5917	4,26
1-4	DOY	147462055,6	0,0302	0,8905	336,0226	4,26
1-2-3-4	Todas	125168035,6	0,0002	0,9079	309,5816	3,78

(a) Entrenamiento.

Grupos de variables	Descripción	SSE	MBE	R	RMSE	Err.Rel
4	Variables Básicas	141107092,6	-87,0013	0,8398	501,5255	5,48
2-4	Descomposición en fechas	118930725,0	4,0445	0,8627	460,4321	4,91
3-4	Componentes no lineales	148209330,8	-121,3175	0,8358	513,992	5,63
1-4	DOY	141810197,4	-98,3265	0,8403	502,7734	5,50
1-2-3-4	Todas	118063505,7	12,0437	0,8639	458,7503	4,92

(b) Validación.

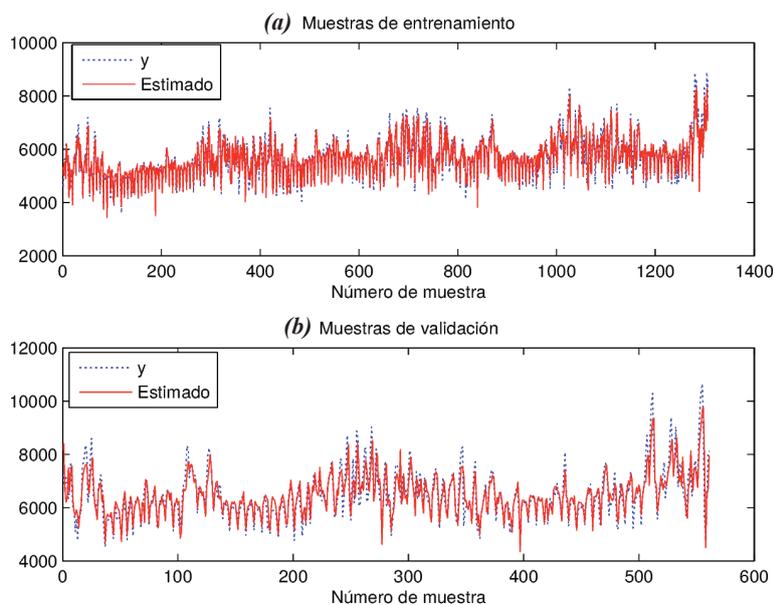


Figura 2: Perfil de curva - Predicción a nivel Provincia

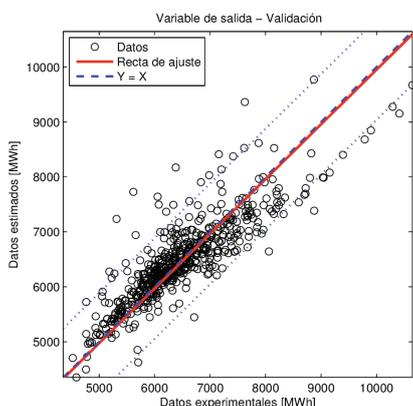


Figura 3: Dispersión del Entrenamiento

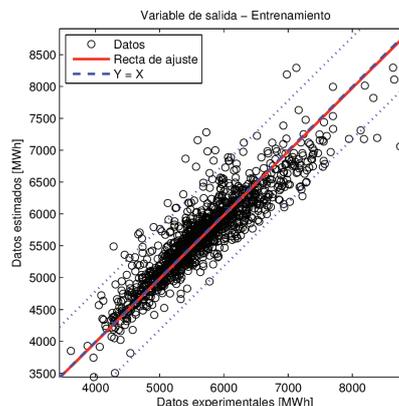


Figura 4: Dispersión de la Validación

#### 4. CONCLUSIONES Y TRABAJO FUTURO

En este trabajo propusimos como hipótesis que la forma como son presentadas las variables temporales a un sistema de predicción de consumo eléctrico influye en la calidad de los resultados. Esta hipótesis fue validada utilizando distintos métodos para representar éstas variables, aplicándolo en conjunto con datos climáticos y consumo eléctrico de la provincia de Tucumán. La división de la variable temporal en día, día de la semana, mes y año en forma individual para cada periodo involucrado en el problema, resultó ser el método más conveniente, obteniendo una mejora de hasta el 10,56 % respecto de otros métodos considerados.

La utilización de todas las variables temporales disponibles en conjunto, produjo que el error de entrenamiento y validación se redujeran. Esto implica que ninguna de las variables temporales introduce ruido en la predicción para el conjunto de datos utilizado.

Es deseable generar modelos más simples, que dependan de menos variables y posean el mismo nivel de error. Esto se debe a que los modelos excesivamente complejos tienen una buena capacidad para ajustar los datos de entrenamiento, pero pueden no adaptarse bien a nuevos datos (overfitting) [19]. Por lo tanto, aplicar técnicas de selección de variables ayudará a reducir la complejidad del modelo, eliminando aquellas variables que resulten innecesarias, redundantes y con demasiado ruido.

Cabe destacar, que nos centramos en la comparativa del uso de diferentes conjuntos de variables sin dar demasiada importancia al método de predicción. Por tal motivo y por simplicidad se utilizó regresión lineal, obteniendo un error del orden del 3%, lo que significa que el problema es lineal.

Se hace importante destacar que la escala en la cual se realiza el análisis también puede influir en los resultados, haciendo que la forma idónea en que se presenta los datos cambie.

Como trabajos futuros, para mejorar la predicción es necesario utilizar métodos que permitan captar los componentes no lineales del problema, si los hay. A su vez, se torna de vital importancia efectuar experimentos en otras escalas temporales y de consumo eléctrico, como así también aplicar técnicas de selección de variables.

#### AGRADECIMIENTOS

Este trabajo fue soportado parcialmente por subsidios PID-UTN 25/P052 - UTI 1781, PID-UTN 25/P051 UTI 1757 y el subsidio CIUNT 26/E457. También deseamos extender los agradecimientos a la Empresa de Distribución Eléctrica de Tucumán S. A.

(E.D.E.T.s.a.) y a la Estación Experimental Obispo Colombes (E.E.A.O.C.), Tucumán - Argentina por proporcionar los datos necesarios para la realización del trabajo. Particularmente hacemos referencia al Ing. Adrián Avella por su apoyo y predisposición durante todo el proceso desarrollado.

#### REFERENCIAS BIBLIOGRÁFICAS

- [1] T. Haida and S. Muto, "Regression based peak load forecasting using a transformation technique," *Power Syst. IEEE Trans.On*, vol. 9, no. 4, pp. 1788–1794, 1994.
- [2] E. Ceperic, V. Ceperic, and A. Baric, "A Strategy for Short-Term Load Forecasting by Support Vector Regression Machines," *IEEE Trans. Power Syst.*, vol. 28, no. 4, pp. 4356–4364, Nov. 2013.
- [3] Y.-J. He, Y.-C. Zhu, D.-X. Duan, and W. Sun, "Application of Neural Network Model Based on Combination of Fuzzy Classification and Input Selection in Short Term Load Forecasting," in *2006 International Conference on Machine Learning and Cybernetics*, 2006, pp. 3152–3156.
- [4] D. R. Palacio, "Simple model for load forecast weather sensitive," in *Electricity Distribution*, 2001. Part 1: Contributions. CIRED.16th International Conference and Exhibition on (IEE Conf. Publ No. 482), 2001, vol. 4, p. 3 pp. vol.4–.
- [5] H. K. Alfares and M. Nazeeruddin, "Electric load forecasting: literature survey and classification of methods," *Int. J. Syst. Sci.*, vol. 33, no. 1, pp. 23–34, 2002.
- [6] A. Tidemann, B. A. Hoverstad, H. Langseth, and P. Ozturk, "Effects of scale on load prediction algorithms," in *Electricity Distribution (CIRED 2013)*, 22nd International Conference and Exhibition on, 2013, pp. 1–4.
- [7] R. Sevljan and R. Rajagopal, "Short Term Electricity Load Forecasting on Varying Levels of Aggregation," *ArXivPrepr.ArXiv14040058*, 2014.
- [8] Z. Y. Wang, C. X. Guo, and Y. J. Cao, "A new method for short-term load forecasting integrating fuzzy-rough sets with artificial neural network," in *Power Engineering Conference*, 2005. IPEC 2005. The 7th International, 2005, pp. 1–173.
- [9] I. Drezga and S. Rahman, "Input variable selection for ANN-based short-term load forecasting," *IEEE Trans. Power Syst.*, vol. 13, no. 4, pp. 1238–1244, Nov. 1998.
- [10] Y.-C. Guo, "An integrated PSO for parameter determination and feature selection of SVR and its application in STLF," in *2009 International Conference on Machine Learning and Cybernetics*, 2009, vol. 1, pp. 359–364.

- [11] Y.-Y. Cheng, P. P. K. Chan, and Z.-W. Qiu, "Random forest based ensemble system for short term load forecasting," in 2012 International Conference on Machine Learning and Cybernetics (ICMLC), 2012, vol. 1, pp. 52–56.
- [12] P. D. Brierley, Some Practical Applications of Neural Networks in the Electricity Industry. Cranfield University, 1998.
- [13] C. E. Borges, A. Peña, and Y. K. Peña, "On the influence of surrounding load demand to improve primary substation STLF," in Industrial Electronics Society, IECON 2013 - 39th Annual Conference of the IEEE, 2013, pp. 8166–8171.
- [14] N. Abu-Shikhah, "Medium-Term Electric Load Forecasting Using Multivariable Linear and Non-Linear Regression," Smart Grid Renew. Energy, vol. 02, no. 02, pp. 126–135, 2011.
- [15] C. Sotomane, L. Asker, and V. Massingue, "ICT for automated forecasting of electrical power consumption: A case study in Maputo," in IST-Africa Conference Proceedings, 2011, pp. 1–8.
- [16] V. Bianco, O. Manca, and S. Nardini, "Electricity consumption forecasting in Italy using linear regression models," Energy, vol. 34, no. 9, pp. 1413–1421, Sep. 2009.
- [17] T. Hong, Short term electric load forecasting. North Carolina State University, 2011.
- [18] L. Eldén, Matrix Methods in Data Mining and Pattern Recognition. Philadelphia, PA, USA: Society for Industrial and Applied Mathematics, 2007.
- [19] K. Varmuza and P. Filzmoser, Introduction to multivariate statistical analysis in chemometrics. CRC press, 2008.



**Diego Fernando Lizondo.-** Received his degree in Systems Information Engineering at the Facultad Regional Tucumán, Universidad Tecnológica Nacional in 2011. He became a member of the GITIA in April 2010. Currently, he is a PhD candidate and became a CONICET grant holder. His main research interests are Multi-Agent Systems, Artificial Immune Systems, Smart Grids and Energy Optimization Techniques.



**Victor Adrián Jimenez.-** Received his degree in Computer Engineering at the Facultad de Ciencias Exactas y Tecnología, Universidad Nacional de Tucumán in 2012. He became a member of the GITIA in April 2013. Currently, he is a PhD candidate. His main research interests are Genetic and Evolutionary Algorithms and its applications to Climate and Energy.



**Fernando Villacis Postigo.-** Is an Adjoint Professor in Computing Science Department at Facultad de Ciencias Naturales at Universidad Nacional Tucumán. He received his degree in Systems Information Engineering from the Facultad Regional Tucumán at Universidad Tecnológica Nacional in 2001. Currently, he is a Mg. candidate and became GITIA's researcher in the mid-2013. His main research interest deal with energy.



**Adrian Will.-** Is an Adjoint Professor in the Mathematics Department at U.N.T. He is currently the CoDirector of the Advanced Technology Research Center of Tucumán, Argentina (C.I.T.A.T. – F.R.T. – U.T.N.). He received a degree in Mathematics from FaMAF, Córdoba, Argentina, and a PhD in Mathematics from the same university, working under the direction of Carlos Olmos. He moved to Tucumán and has been working in Genetic Algorithms and Neural Networks for Industrial Applications ever since.



**Sebastián Alberto Rodríguez.-** Is a Full Professor of the Department of Computer Science, National Technology University (NTU), Argentina. He is also the founder and Head of the Advanced Informatics Technology Research Group (GITIA), Argentina, and an associate researcher of the Systems and Transportation Laboratory at the University of Technology of Belfort-Montbéliard (UTBM), France. He received a Computer Engineer degree for the National University of Tucumán, Argentina, a M.S. degree in computer science from the University of Franche-Comté and a Ph.D. degree in computer science of the UTBM.