

Análisis Exploratorio de la Demanda Eléctrica Utilizando Herramientas Informacionales

1st Barberia, Juan Luis

CPSI

UTN-FRBA

Buenos Aires, Argentina

jbarberia@frba,utn.edu.ar

1nd Esangui, Sebastian

CPSI

UTN-FRBA

Buenos Aires, Argentina

—@frba,utn.edu.ar

1nd Fernandez Biancardi, Juan Facundo

CPSI

UTN-FRBA

Buenos Aires, Argentina

jfernandezbiancardi@frba,utn.edu.ar

1nd Legnani, Walter Edgardo

CPSI

UTN-FRBA

Buenos Aires, Argentina

—@frba,utn.edu.ar

1nd Licata Caruso, Lorenzo

CPSI

UTN-FRBA

Buenos Aires, Argentina

—@frba,utn.edu.ar

1nd Lottero, Giancarlo

CPSI

UTN-FRBA

Buenos Aires, Argentina

—@frba,utn.edu.ar

1nd Paglia, Adrián Ezequiel

CPSI

UTN-FRBA

Buenos Aires, Argentina

—@frba,utn.edu.ar

Resumen—De acuerdo a que la entropía de permutación permite analizar el grado de incertidumbre que tiene una señal, se presenta una propuesta original de análisis para descubrir características relevantes y eventos particulares de la señal de demanda de potencia eléctrica en el Sistema Interconectado Argentino (SADI). Como resultado se presenta la posibilidad de detectar ciertos patrones en los registros horarios de demanda eléctrica del SADI, lo cual hace que la tarea de exploración de datos sea más veloz, eficiente y accesible de implementar. La simple observación e interpretación de esta métrica permite una mejora de la predicción de futuros valores en un 7 %. Junto con la inclusión de datos de temperatura, mediante un modelo polinomial de orden dos, representa una mejora total del 21 % respecto a un modelo lineal simple de regresión sobre los datos de la señal original.

Index Terms—demanda de energía eléctrica, modelos de predicción, entropía de permutación de Shannon.

I. INTRODUCCIÓN

La estimación de demanda eléctrica es una variable de vital importancia en el estudio de un sistema de potencia. Afecta directamente a la toma de decisiones estratégicas y de despacho de energía, para mejorar su confiabilidad y asegurar la disponibilidad de la potencia necesaria en determinado momento. Por tal motivo es esencial conocer tanto su magnitud como su variabilidad a lo largo del día, sus respectivos máximos y mínimos, y su interacción con las distintas variables exógenas como lo pueden ser la desviación de la temperatura respecto al valor estándar, de que día de la semana se trata, si son días feriados o no, estación estación del año, entre otros.

En la comunidad científica se ha abordado desde distintos enfoques, incluyendo utilizando [modelos de regresión no paramétrica](#) [1] y [modelos no lineales](#) [2] o bien [técnicas estadísticas](#) [3]

El presente trabajo propone un enfoque de análisis mediante la utilización de herramientas informacionales basados en conceptos de entropía para el estudio de la curva de demanda eléctrica y su predicción. Se utiliza un modelo simple de regresión y no se exploran regresores avanzados, ya que la motivación de esta investigación es la de incorporar nuevas herramientas al análisis y no obtener una mejor predicción frente a una determinada métrica. Los resultados de la presente propuesta permiten identificar variables e identificar segmentos de las señales que tienen una directa influencia en el rendimiento de los modelos de predicción.

II. MATERIALES Y MÉTODOS

II-A. Origen y selección de los datos

Para este trabajo se utilizaron datos horarios, de libre acceso, de la demanda eléctrica y temperatura ambiente. La utilización complementaria de esta última, corresponde a que la demanda eléctrica esta directamente relacionada con esta variable meteorológica de naturaleza exógena. [4].

Los datos de la demanda eléctrica se tomaron de informes y estadísticas de la Compañía Administradora del Mercado Mayorista Eléctrico (CAMMESA). Los mismos corresponden a la totalidad de la demanda sobre el Sistema Argentino de Interconexión (SADI). Por otro lado, los datos meteorológicos corresponden a las bases de datos del Servicio Meteorológico Nacional (SMN).

Para evitar una gran dimensionalidad de los datos utilizados, en este trabajo se seleccionaron estaciones meteorológicas que están dentro del área de mayor demanda eléctrica del SADI, tomando un promedio de los registros de las estaciones meteorológicas Villa Ortuzar y Aeropuerto Jorge Newbery. Ambas estaciones están localizadas sobre la Ciudad Autónoma de Buenos Aires, ubicación que representa el 36 % de la demanda eléctrica del SADI [5].

El total de registros relevados corresponde al periodo del 1° de enero del 2021 hasta el 31 de enero del 2024, tanto

para los valores de demanda de energía eléctrica como para los de temperatura ambiente. En total se contabilizan 26278 registros, donde se utilizaron los primeros 25533 registros a modo de entrenamiento de modelos de pronóstico y los 745 registros restantes, correspondientes al mes de enero de 2024 para la verificación de los mismos.

II-B. Procesamiento

Con los datos obtenidos se construye una matriz de **variables observaciones** de 3 columnas, correspondientes a la fecha, la demanda eléctrica y la temperatura para una determinada hora. Se detectó la presencia de datos faltantes y ello se resolvió tomándose un promedio entre muestras adyacentes. Se crearon columnas adicionales para indicar el día de la semana al que corresponde un dado datos, mediante una variable binaria que indica si la fecha corresponde a un día laborable o no. También, se relevaron las temperaturas medias mensuales de la Ciudad Autónoma de Buenos Aires, incorporándose estas a la matriz de datos indicando sí la temperatura horaria superaba o no ese valor.

Fecha	Demanda [MW]	T° [°C]	Día	Supera T_{med}
31/1/2024 21:00	23.977	31.80	H	No
31/1/2024 22:00	22.934	30.85	H	No
31/1/2024 23:00	23.471	30.25	H	No

Cuadro I
EXTRACTO DE LA BASE DE DATOS ELABORADA

II-C. Herramientas utilizadas

II-C1. Pronóstico: La realización de un pronóstico consiste en estimar un valor de la señal x_t $t = 1, 2, \dots, N$, utilizando una ponderación de valores anteriores $\{x_{k-1}, x_{k-2}, \dots, x_{k-i}\}$ para k e $i \in \mathbb{Z}^+$. Si se tiene una serie de tiempo completa, es posible organizar los datos del pronóstico en una forma matricial como se muestra en 1, donde la predicción de un valor \hat{x}_i se realiza observando las 3 muestras anteriores de la señal.

$$\begin{bmatrix} x_1 & x_2 & x_3 \\ x_2 & x_3 & x_4 \\ x_3 & x_4 & x_5 \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{bmatrix} = \begin{bmatrix} \hat{x}_4 \\ \hat{x}_5 \\ \hat{x}_6 \end{bmatrix}. \quad (1)$$

En este problema, el objetivo es minimizar el error entre \hat{x}_t y x_t lo cual se puede interpretar como una tarea de regresión, con la minimización de una función de costo empírico J_{emp} como se expresa en 2:

$$J_{emp} = \sum_{i \in D} L(x_i, \hat{x}_i), \quad (2)$$

siendo D la totalidad de filas que tenga el conjunto de datos analizado y L una función de distancia que bien puede ser lineal como no lineal. Para mejorar el pronóstico se pueden utilizar fuentes de información externas a la señal (exógenas). Estas se introducen como columnas nuevas en la matriz de observaciones, con el agregado de un peso extra (β_i) como se muestra en 3. Siendo $y_t = t = 1, 2, \dots, N$ una señal externa a la señal x_t , cuyo tipo de dato puede ser numérico, binario

o categórico, permitiendo de esta forma incorporar una gran cantidad de información al modelo de predicción.

$$\begin{bmatrix} x_1 & x_2 & x_3 & y_3 \\ x_2 & x_3 & x_4 & y_4 \\ x_3 & x_4 & x_5 & y_5 \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \\ \beta_1 \end{bmatrix} = \begin{bmatrix} \hat{x}_4 \\ \hat{x}_5 \\ \hat{x}_6 \end{bmatrix}. \quad (3)$$

II-C2. Correlación: Una de las herramientas más utilizadas para poder analizar la estructura y relación entre puntos vecinos en el entorno de una señal es el análisis de correlación [6]. El objetivo de esta herramienta matemática es encontrar un desplazamiento que tenga gran similitud con los datos que se busque estudiar.

II-C3. Entropía de una Señal: La entropía de información de una señal y sus variantes es un método para analizar la irregularidad y el desconocimiento que se tiene de la evolución de una serie de datos que describe. Como es un tipo de medida de la irregularidad, la misma se comporta de manera similar a los exponentes de Lyapunov, teniendo la ventaja de que su forma de cálculo es más simple y su comportamiento es robusto frente a diferentes tipos de transformaciones.

Bandt y Pompe [7] introdujeron el concepto de entropía de permutación de acuerdo a la comparación de la señal con sus valores locales. El método se basa en elegir una cantidad de puntos de la serie de datos, denominada dimensión de embedding, asignarle un alfabeto que puede ser de carácter alfanumérico, y contabilizar su orden de aparición secuencial en la serie temporal. Con la estadística de estos patrones se puede estimar una probabilidad y de allí el cálculo de la entropía correspondiente. En el caso de la entropía en este trabajo se trabajó con la formulación de Shannón. Un ejemplo de la metodología propuesta por Bandt y Pompe como se muestra en la figura 1, donde la frecuencia relativa de la aparición de un patrón dado se emplea para estimar la probabilidad de ocurrencia de un patrón ordenado.

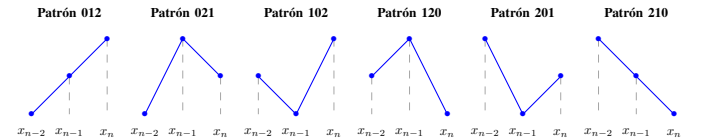


Figura 1. Esquema demostrativo de patrones ordinales para una dimensión de embedding de tres puntos ($n = 3$).

En una serie de datos compuesta por T valores, al considerar patrones cuya longitud (o sea su dimensión de embedding) es n , su distribución se notará como π entonces se estima su probabilidad $p(\pi_i)$ de aparición, como la cantidad de secuencia de símbolos presentes en la señal respecto de la cantidad de símbolos posibles. Matemáticamente queda expresado de la siguiente manera:

$$p(\pi) = \frac{\#\{t | 0 \leq t \leq T - n, (x_{t+1}, \dots, x_{t+n}) \text{ es tipo } \pi\}}{T - n + 1}. \quad (4)$$

Una vez que se tiene la distribución de probabilidades, se procede al cálculo de la entropía través de la formulación de Shannon:

$$H(n) = - \sum_{i=1,2,\dots,n!} p(\pi_i) \log p(\pi_i). \quad (5)$$

Tomando segmentos de longitud n sobre la señal se tendrán $n!$ distintos símbolos (permutaciones) posibles, con lo cual la expresión 5 sumará estas $n!$ posibles probabilidades. Se observa que la entropía de permutación esta numéricamente restringida a los valores $0 \leq H(n) \leq \log n!$. Por este motivo se puede dividir por el último valor para obtener una entropía normalizada, independientemente de la dimensión de embedding utilizada.

Al avanzar las investigaciones con la aplicación de los patrones ordinales se incorporaron variantes y modificaciones que permiten discriminar ciertos comportamientos de los sistemas dinámicos. Surgen variantes como la entropía de permutación ponderada introducida por [8] que permite no solo contabilizar la ocurrencia de los patrones ordinales sino que también considerar las variaciones de amplitud de la señal.

En resumen, esta herramienta es utilizada como indicador de que tan "predecible", o no, una señal puede ser, brindando una herramienta para el análisis de series de datos que no requiere de hacer ninguna suposición del proceso subyacente. La entropía es equivalente a tener una indicación de "divertencia" o "sorpresa" sobre la información que contiene los datos de una serie de tiempo. Es precisamente esta característica lo que indujo su utilización para detectar segmentos claves de la señal de demanda de potencia eléctrica donde se requiere un análisis especial o se presentan situaciones anómalas en la misma.

II-C4. Implementación computacional del análisis exploratorio: Se utilizó el lenguaje de programación *Python* mediante el uso de un notebook colaborativo para ir explorando diferentes combinaciones de los datos de demanda de energía, la temperatura, y otras variables. De esta manera, se llegaron a los resultados que permitieron formular las conclusiones del estudio y establecer las características propias de la evolución de la demanda eléctrica. En particular, para el cálculo de las entropías se utilizaron los paquetes de especializados *ordpy* [9] y *EntropyHub* [10]. Para la confección de los gráficos se utilizó la librería *Plotly* [11].

III. PROPUESTA

Para la tarea de análisis y desarrollos de modelos de predicción de un determinado sistema, es necesario que el analista tenga herramientas que lo guíen a utilizar el mejor conjunto de variables y parámetros para el desarrollo del mismo. En la sección siguiente se explorarán diferentes aplicaciones de las métricas informacionales para poder descubrir diferentes aspectos relevantes de la evolución de la demanda eléctrica del SADI. Se utilizaran datos de entrenamiento correspondientes a los anteriores tres años (2021-2023), para poder predecir la demanda del mes de enero de 2024, tal como se muestra en la figura 2.

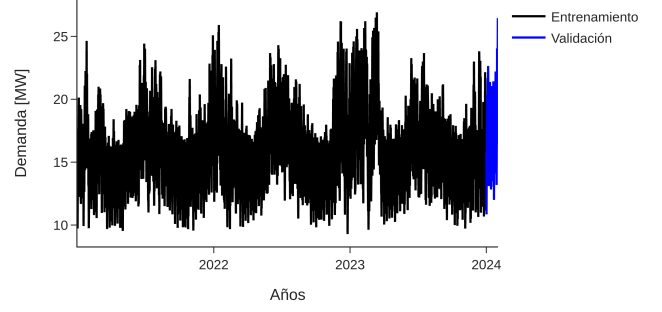


Figura 2. Ejemplo de datos para entrenar el modelo en color azul mientras que en color rojo se indican los valores de demanda de energía eléctrica predichos.

IV. PRUEBAS Y RESULTADOS

IV-A. Exploración de datos

Cuando se tiene una señal proveniente de una medición o proceso, es conveniente en primer medida visualizarla gráficamente, como también emplear métricas exploratorias que indiquen la presencia de errores en la toma de medición o valores espurios. Para esta tarea se seleccionó la entropía ponderada por su característica de resaltar anomalías, no solo en la estructura morfológica de la señal, sino que también en su amplitud.

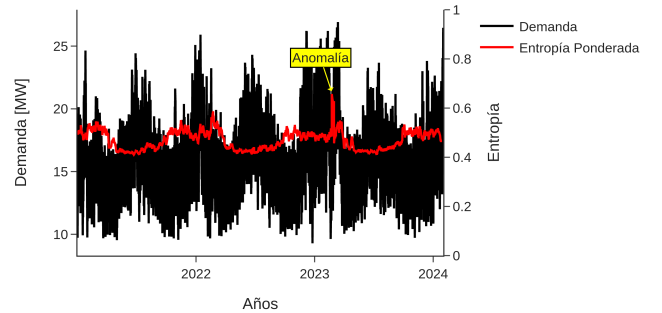


Figura 3. Demanda eléctrica y su correspondiente entropía de permutación ponderada. Esta última puede ser interpretada como un indicador de anomalías en la señal.

En la figura 3 se observa claramente un gran cambio en la entropía para un determinado momento, mientras que en la curva de demanda no se observa a priori un gran cambio estructural debido a la extensión de los datos adquiridos. Si se agranda ese periodo de tiempo en particular, figura 4, se

observa que el cambio en la entropía corresponde al momento del colapso parcial del SADI [12]

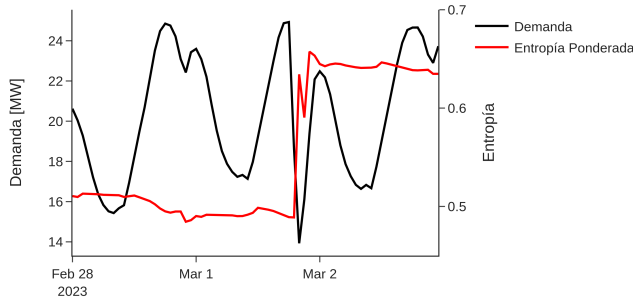


Figura 4. Un cambio significativo en la entropía ponderada es indicador de anomalías en el sistema, tal como lo fue el colapso del SADI durante el 1º de marzo de 2023.

Este análisis preliminar muestra que la detección de anomalías a partir de utilizar esta herramienta informacional, aporta un indicador simple para iniciar análisis ad-hoc detallados de momentos específicos de la señal.

Al graficar la señal de demanda y la entropía de permutación de forma bimestral, figura 5, puede observarse es en el comienzo de una semana, el valor de entropía muestra elevarse de manera considerable al compararlo con el día anterior y que al transcurrir la semana este valor tiende a "estabilizarse" hasta el comienzo de la próxima semana, donde vuelve a repetirse dicho patrón. Se nota que cada cinco días de baja entropía hay dos días posteriores de alta entropía. Este cambio indica que los días hábiles hacen el sistema más predecible durante los días laborales. Llegado al fin de semana, el comportamiento de la población, y por ende, del consumo eléctrico, varía respecto a los días laborales, haciendo menos predecible al sistema.

De lo expuesto anteriormente, se podría suponer que la entropía de la señal, revela patrones regulares y puntos atípicos que podrían llegar a ser de utilidad a la hora de realizar una predicción. Como principales observaciones se aprecia que para la confección de un modelo de pronóstico, será útil disponer una variable exógena para poder capturar la condición del tipo de día (laborable o no).

IV-B. Análisis de auto correlación

Siguiendo, con un análisis clásico de la señal. Se realizó un análisis de correlación de la misma. Esto nos permitió definir qué longitud de la serie temporal bajo estudio es apropiada utilizar para maximizar la similitud entre las variaciones de la demanda. Puede observarse que el modelo a utilizar deberá ser más sensible frente a los registros que sucedieron la hora anterior y el día anterior a la misma hora del pronóstico, como se muestra en la figura 6. Lo cual, en este estudio se emplearon datos horarios, con una ventana de siete días de análisis, para

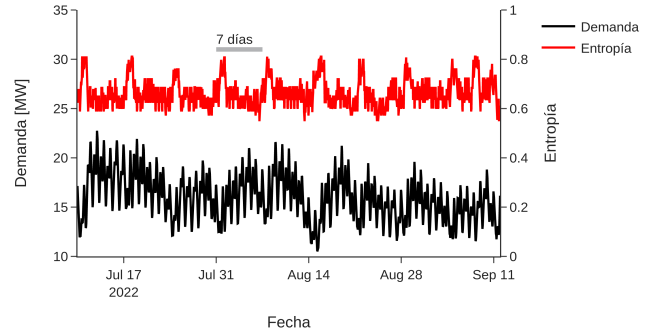


Figura 5. Cambios en el comportamiento de la demanda son capturados por la entropía de permutación. En este caso corresponde al cambio en la demanda eléctrica entre días hábiles y no hábiles.

eliminar las influencias de las variaciones intersemanales en la predicción.

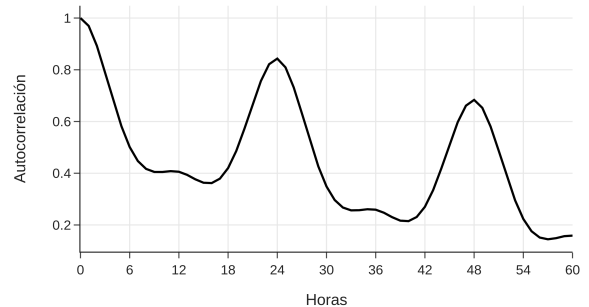


Figura 6. Ejemplo de auto correlación de la señal de demanda de energía eléctrica, en el eje horizontal se indican las horas de día a partir de una fecha predeterminada.

IV-C. Variables Exógenas

También se debe en cuenta la inclusión de otras variables que sean sensibles al modelo que de predicción. Para esto se incluyeron las mediciones de temperatura representativas del área de mayor demanda eléctrica [13].

Sin embargo se debe tener en cuenta que la demanda aumentará tanto con un aumento como con un decremento de la temperatura. Pensando en el sector domiciliario, cuanto más se aleje la temperatura de un punto de confort, más energía se consumirá en artefactos de climatización. Asimismo, como sugiere un artículo del ministerio de energía de la república Argentina [?], no dependerá linealmente con la variación de temperatura, sino que seguirá una forma polinómica. Para

evaluar la variación percibida de la demanda en función de la temperatura, se realiza un ajuste mediante un polinomio de segundo grado, obteniendo la siguiente expresión para el ajuste de la temperatura de toda la serie de datos disponibles:

$$D(T) = 31,9 \cdot T^2 - 1081 \cdot T + 23620. \quad (6)$$

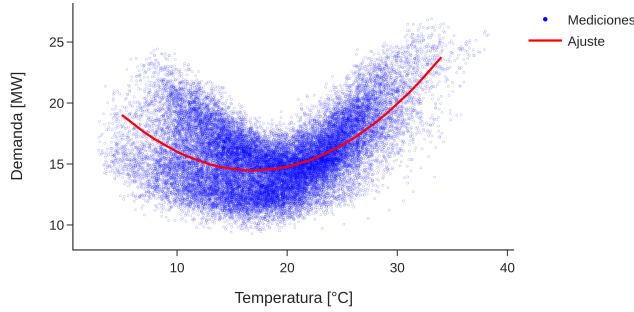


Figura 7. Ajuste de la demanda en función de la temperatura utilizando un polinomio de segundo grado.

V. RESULTADOS

Se desarrolló con un modelo simple de demanda de potencia eléctrica, donde la única variable es el dato original provisto por CAMMESA, para determinar el valor próximo de la señal se tomó la medición del día anterior de acuerdo a la gran auto correlación observada en la figura 6. Este modelo se tomó como referencia frente a las subsiguientes mejoras que resultaron del estudio.

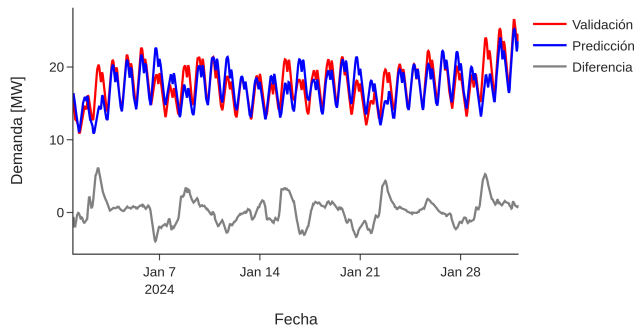


Figura 8. Modelo inicial ingenuo

Luego se procede a realizar un modelo de predicción lineal, donde se le añadieron las variables exógenas analizadas en las secciones anteriores. Se tomaron como variables externas la

temperatura (T), a la demanda estimada (DE) en función de la temperatura modelada en IV-C, la indicación binaria (DH) de día hábil y no hábil observada en IV-A, como también a la entropía (H) muestreada semanalmente, con un valor de embedding $m = 3$ y delay time embedding $\tau = 1$ para su respectivo cálculo. En este caso se tomó el valor de la dimensión de embedding de acuerdo a la longitud de la serie de datos procesada.

Se realizaron diferentes modelos para las diferentes combinaciones de variables exógenas, listándose los resultados en la tabla II.

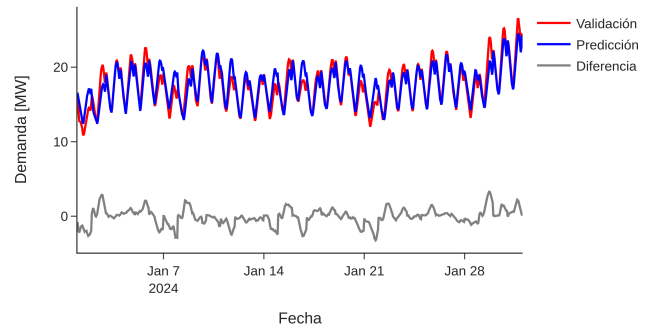


Figura 9. Modelo predictivo desarrollado al considerar todas las variables exógenas

Modelo	Variable Exógena	ECM	Mejora
Base	-	3119371,47	-
Líneal	-	1530491,74	49 %(*)
1	T	1476649,45	4 %
2	H	1545521,44	-1 %
3	DE	1265580,96	17 %
4	DH	1417696,85	7 %
5	T, H	1493297,11	2 %
6	T, DE	1337558,37	13 %
7	T, DH	1394660,23	9 %
8	H, DE	1270497,29	17 %
9	H, DH	1421323,7	7 %
10	DE, DH	1243984,92	19 %
11	T, H, DE	1353784,27	12 %
12	T, H, DH	1397266,1	9 %
13	T, DE, DH	1246617,62	19 %
14	H, DE, DH	1243526,69	19 %
15	T, H, DE, DH	1247569,37	18 %

Cuadro II

RESULTADOS OBTENIDOS AL UTILIZAR DIFERENTES VARIABLES EXÓGENAS, RESPECTO AL MODELO LÍNEAL. *(MEJORA RESPECTO AL MODELO BASE)

En la tabla II T es la temperatura, TC es la temperatura corregida mediante un modelo polinomial cuadrático, DH es una variable binomial que indica si es día hábil o no, EP es la entropía de permutación ponderada.

La mejora acumulada total fue de 20.28 % desde el modelo de predicción de referencia.

VI. CONCLUSIONES

La exploración de datos por medio de la entropía permitió encontrar rápidamente anomalías que pueden disparar análisis ad-hoc de situaciones particulares del sistema bajo estudio. Se detectó de forma clara y eficiente un disturbio mayor en el SADI, que frente a inspección de los datos originales resultaba imperceptible. Respecto a los modelos implementados, se observó una mejora del 19 % respecto de un modelo lineal. Siendo las mejoras la incorporación de magnitudes derivadas como la demanda estimada (DE) y el indicador del tipo de día (DH). Se observó que utilizar directamente la temperatura (T) y la entropía (H) sin ningún tipo de análisis, provoca una mejora de solo el 2 %. Esto muestra la necesidad clara de disponer herramientas de preprocesamiento y visualización para poder seleccionar correctamente las variables y obtener un mejor pronóstico, antes que la inclusión de un gran volumen de información al modelo sin un análisis previo.

~~Como conclusión, podemos se puede mencionar que el uso de variables exógenas mejora el modelo en gran medida, sobre todo teniendo en cuenta una caracterización fenomenológica de la variable, como se hizo por ejemplo con el modelo de temperaturas, que también podría depender de cada ciudad y es necesario tener información que caracterice cada zona energética.~~

~~¿qué decimos de la entropía...? es la segunda que menos mejora introduce... pero de alguna manera es la mas original~~
‡

REFERENCIAS

- [1] F. Barrientos, J. Olaya, and V. González, "Un modelo spline para el pronóstico de la demanda de energía eléctrica. universidad del valle," *Facultad de Ingenierías. Escuela de Ingeniería Industrial y Estadística. Cali, Colombia*, 2007.
- [2] V. RUEDA, J. D. Velásquez Henao, and C. J. Franco Cardona, "Avances recientes en la predicción de la demanda de electricidad usando modelos no lineales," *Dyna*, vol. 78, no. 167, pp. 36–43, 2011.
- [3] J. F. T. Muñoz, C. A. V. Galvis, and M. V. Cárdenas, "Comparación de técnicas estadísticas de pronóstico para la demanda de energía eléctrica," *Revista Ingeniería Industrial*, vol. 13, no. 1, p. 2, 2014.
- [4] L. Mastronardi, M. A. Sfeir, and S. Sánchez, "La temperatura y su influencia en la demanda de energía eléctrica: Un análisis regional para argentina usando modelos econométricos," *LI Reunión Anual. Asociación Argentina de Economía Política*, 2016.
- [5] D. de Ingeniería de Operación, "Resumen de datos operativos del día viernes 02/02/2024 a la hora pico del sadi (14:31 hs.)," 2024.
- [6] D. C. Montgomery, C. L. Jennings, and M. Kulahci, *Introduction to time series analysis and forecasting*. John Wiley & Sons, 2015.
- [7] C. Bandt and B. Pompe, "Permutation entropy: a natural complexity measure for time series," *Physical review letters*, vol. 88, no. 17, p. 174102, 2002.
- [8] B. Fadlallah, B. Chen, A. Keil, and J. Principe, "Weighted-permutation entropy: A complexity measure for time series incorporating amplitude information," *Physical Review E*, vol. 87, no. 2, p. 022911, 2013.
- [9] A. A. Pessa and H. V. Ribeiro, "ordpy: A python package for data analysis with permutation entropy and ordinal network methods," *Chaos: An Interdisciplinary Journal of Nonlinear Science*, vol. 31, no. 6, 2021.
- [10] M. W. Flood and B. Grimm, "Entropyhub: An open-source toolkit for entropic time series analysis," *PloS one*, vol. 16, no. 11, p. e0259448, 2021.
- [11] A. Schroeder, C. Mayer, and A. M. Ward, *The Book of Dash: Build Dashboards with Python and Plotly*. No Starch Press, 2022.
- [12] CAMMESA, "Falla triple por incendio de campos: Desenganche de línea 500 kv atucha ii - rodríguez estando f/s líneas campana - rodríguez y belgrano - rodríguez," <https://cammesaweb.cammesa.com>, consultado: 16 de Mayo de 2024.
- [13] G. Heinemann, D. Nordmian, and E. Plant, "The relationship between summer weather and summer loads-a regression analysis," *IEEE Transactions on Power Apparatus and Systems*, no. 11, pp. 1144–1154, 1966.