



UNIVERSITAT
POLITÈCNICA
DE VALÈNCIA



Caracterización de la demanda de energía mediante patrones estocásticos en las Redes Eléctricas Inteligentes

Valencia, septiembre de 2020

Autor: Xavier Serrano Guerrero

Director: Guillermo Escrivá Escrivá

Agradecimientos

Es muy grato dirigirme en primer lugar a mi director de tesis, el Dr. Guillermo Escrivá, por su dirección eficaz, sus consejos, discusiones proactivas y todo el apoyo brindado en estos años de trabajo.

A la Universidad Politécnica Salesiana del Ecuador por el soporte dado durante estos años de estudio, a sus autoridades y de manera especial a los compañeros que me brindaron su apoyo incondicional cuando lo necesité, como son Antonio Barragán, Julio Viola, Eduardo Pinos, Flavio Quizhpi, Juan Inga, Marco Amaya, Arturo Peralta, Natalia González y Jorge Rojas. También es menester agradecer a Xavier Mora, Santiago Luna, Jaime Estrella y Danny Guachichullca, quienes como estudiantes de la UPS colaboraron de manera activa en el proyecto de investigación relacionado con esta tesis. A los compañeros de viaje con los que se ha compartido gratos momentos, Julio Zambrano, Jairo Sacoto, Victor Huilcapi, Marcelo Flores, Christian Sagbay y Christian Cobos.

También quiero recordar y expresar mi gratitud a los compañeros del Instituto de Ingeniería Energética de la UPV, de manera especial a Jean Michel Clairand por las conversaciones y discusiones mantenidas durante nuestros estudios. A Yago Rivera, Iván Ligardo, David Blanco, Paula Bastida, Alejandro Herrera, Yaisel Córdova, Iván Cuesta, Christian Stascheit y David Ribó, con los que compartí muy buenos momentos.

A mis padres y hermanos.

Finalmente, a mi esposa Rosana, por todo su apoyo, paciencia y comprensión en estos años de estudio.

Dedicatoria

A mi hija Luciana.

Abstract

The energy demand is increasing considerably every year worldwide. Moreover, the electrification of different sectors has caused the growth of electricity demand at an even higher rate. Thus, the electricity grids have technologically evolved in several aspects, one of them is the availability of demand data at different points and levels of the grid, such as in transmission and distribution systems, and large and small consumers. The data provided by the new smart grids are essential baseline information for the management and planning of electricity systems.

The data provided by smart meters on electricity grids is worthless if it is not properly analysed. In addition to the appropriate processing of such data, tools are needed to obtain useful information. The energy demand management systems associated with *pattern recognition* have been scarcely studied at present. Some limitations have been identified in this area of study. For example, the characterization of electricity demand through pattern recognition has not been used to identify and evaluate changes in energy consumption, whereas monitoring systems do not identify the possible causes of the anomalies detected in electricity demand.

Demand forecasting is an effective tool in the management of electricity supply systems. Currently, tools such as neural networks and deep learning are preferred for this purpose. However, they present remarkable drawbacks, such as the difficulty in quantifying uncertainties, the requirement of huge computational resources, and considerable effort to establish the structure of the neural network to provide adequate results.

Based on the limitations detected, this thesis proposes a new statistical methodology to characterize the behaviour of the energy demand of consumers and other points of the electricity grid by identifying and obtaining patterns. The use of these patterns allows the assessment and identification of changes in electricity load profiles. Besides, the evaluation of changes in electricity demand allows these values to be associated with

possible events in an installation. This methodology can be used to detect anomalies and to catalogue load profiles according to the changes they have had from their usual behaviour, which allows the identification of working modes of electrical systems. Concerning the prediction of demand, a simple application methodology is proposed to overcome the limitations detected in the instruments derived from artificial intelligence, which allows quantifying the uncertainty of the performed predictions. This information is useful for management since it is possible to generate alarms, reduce maintenance costs, and apply appropriate energy efficiency measures. On the other hand, a method is proposed for processing the data provided by the intelligent meters in order to adjust them to a known probability distribution, facilitating their interpretation and validating the statistical analysis. For this purpose, it has been related to the active electrical power absorbed by a consumer or a system in general with the existing theory of time-series data.

Finally, the application of the proposed methodologies is evaluated in several real case studies at different points and levels of the electricity grid, identifying the benefits that can be obtained in the management of each of these systems.

Resumen

La demanda de energía en todo el planeta continúa incrementándose de manera acelerada. Por otro lado, la electrificación de diferentes sectores ha hecho que la demanda de electricidad crezca a una tasa aún mayor. Las redes eléctricas han evolucionado tecnológicamente en muchos aspectos, uno de ellos se refiere a la disponibilidad de datos de la demanda en diferentes puntos y niveles de la red, como en redes de transmisión, redes de distribución y en los grandes y pequeños consumidores. Estos datos suministrados por las nuevas redes eléctricas inteligentes constituyen información de partida esencial para la gestión y planificación de los sistemas eléctricos.

Los datos proporcionados por los medidores inteligentes de las redes eléctricas no tienen ninguna utilidad si no se analizan adecuadamente. A más del procesamiento adecuado de esos datos, se requieren herramientas que permitan obtener información útil. Los sistemas de gestión de la demanda de energía asociados al *reconocimiento de patrones* actualmente se han estudiado de manera escasa. En esta área de estudio se han identificado algunas limitaciones. Por ejemplo, la caracterización de la demanda de electricidad mediante el reconocimiento de patrones no se ha utilizado para la identificación y valoración de cambios en el consumo de energía y los sistemas de monitorización no identifican posibles causas de las anomalías detectadas en la demanda de energía eléctrica.

La predicción de la demanda es también una herramienta eficaz en la gestión de los sistemas de suministro eléctrico. Actualmente, herramientas tales como las redes neuronales y el aprendizaje profundo son las preferidas para realizar esta labor. Sin embargo presentan algunos inconvenientes, tales como, la dificultad para cuantificar la incertidumbre, requieren un gasto computacional elevado y esfuerzo considerable para establecer la estructura de la red neuronal que proporcione resultados adecuados.

Con base en las limitaciones detectadas, en esta tesis se propone una nueva metodología estadística para caracterizar el comportamiento de la demanda de energía de los consumidores y otros puntos de la red eléctrica mediante la identificación y obtención de patrones. La utilización de estos patrones permite valorar e identificar cambios en perfiles de carga de electricidad. Además, la valoración de los cambios en la demanda eléctrica permite asociar estos valores a posibles eventos en una instalación. Esta metodología puede ser empleada para detectar anomalías y catalogar perfiles de carga de acuerdo al cambio que han tenido con respecto a su comportamiento habitual, lo que permite identificar modos de trabajo de los sistemas eléctricos. En cuanto a la predicción de la demanda, se propone una metodología de simple aplicación para afrontar las limitaciones detectadas en las herramientas derivadas de la inteligencia artificial, de tal manera que sea posible acotar la incertidumbre de las predicciones realizadas. Esta información resulta útil en la gestión, ya que es posible generar alarmas, reducir costos en el mantenimiento y aplicar medidas adecuadas de eficiencia energética. Por otro lado, se propone un método para tratar los datos proporcionados por los medidores inteligentes, de tal manera que se ajusten a una distribución de probabilidad conocida, facilitando su interpretación y validando el análisis estadístico. Para ello, se ha relacionado a la potencia eléctrica activa absorbida por un consumidor o un sistema en general con la teoría existente de las series temporales de datos.

Finalmente, se evalúa la aplicación del método y la metodología propuesta en diversos casos de estudio reales en diferentes puntos y niveles de la red eléctrica, identificando los beneficios que pueden obtenerse en la gestión de cada uno de ellos.

Resum

La demanda d'energia a tot el planeta continua incrementant-se de manera accelerada. D'altra banda, l'electrificació de diferents sectors ha fet que la demanda d'electricitat creixi a una taxa encara més gran. Les xarxes elèctriques han evolucionat tecnològicament en molts aspectes, un d'ells es refereix a la disponibilitat de dades de la demanda en diferents punts i nivells de la xarxa, com en les de transmissió, de distribució i en els grans i petits consumidors. Aquestes dades subministrades per les noves xarxes elèctriques intel·ligents constitueixen informació de partida essencial per a la gestió i planificació dels sistemes elèctrics.

Les dades proporcionades pels mesuradors intel·ligents de les xarxes elèctriques no tenen cap utilitat si no s'analitzen adequadament. A més del processament adequat d'aquestes dades, es requereixen eines que permetin obtenir informació útil. Els sistemes de gestió de la demanda d'energia associats al reconeixement de patrons actualment s'han estudiat de manera escassa. En aquesta àrea d'estudi s'han identificat algunes limitacions. Per exemple, la caracterització de la demanda d'electricitat mitjançant el reconeixement de patrons no s'ha utilitzat per a la identificació i valoració de canvis en el consum d'energia i els sistemes de monitorització no identifiquen possibles causes de les anomalies detectades en la demanda d'energia elèctrica.

La predicció de la demanda és també una eina eficaç en la gestió dels sistemes de subministrament elèctric. Actualment, eines com ara les xarxes neuronals i l'aprenentatge profund són les preferides per a realitzar aquesta tasca. Però presenten alguns inconvenients, com ara, la dificultat per quantificar la incertesa, requereixen una despesa computacional elevada i un esforç considerable per a establir l'estructura de la xarxa neuronal que proporcioni resultats adequats.

Amb base en les limitacions detectades, en aquesta tesi es proposa una nova metodologia estadística per caracteritzar el comportament de la demanda d'energia dels consu-

midors i altres punts de la xarxa elèctrica mitjançant la identificació i obtenció de patrons. La utilització d'aquests patrons permet valorar i identificar canvis en perfils de càrrega d'electricitat. A més, la valoració dels canvis en la demanda elèctrica permet associar aquests valors a possibles esdeveniments en una instal·lació. Aquesta metodologia pot ser emprada per detectar anomalies i catalogar perfils de càrrega d'acord al canvi que han tingut pel que fa al seu comportament habitual, el que permet identificar maneres de utilització dels sistemes elèctrics. En quant a la predicció de la demanda, es proposa una metodologia de simple aplicació per afrontar les limitacions detectades en les eines derivades de la intel·ligència artificial, de tal manera que sigui possible delimitar la incertesa de les prediccions realitzades. Aquesta informació és útil en la gestió, ja que és possible generar alarmes, reduir costos en el manteniment i aplicar mesures adequades d'eficiència energètica. D'altra banda, es proposa un mètode per a tractar les dades proporcionades pels mesuradors intel·ligents, de tal manera que s'ajustin a una distribució de probabilitat coneguda, facilitant la seva interpretació i validant l'anàlisi estadístic. Per a la qual cosa, s'ha relacionat a la potència elèctrica activa absorbida per un consumidor o un sistema en general amb la teoria existent de les sèries temporals de dades.

Finalment, s'avalua l'aplicació del mètode i la metodologia proposada en diversos casos d'estudi reals en diferents punts i nivells de la xarxa elèctrica, identificant els beneficis que es poden obtenir a la gestió de cada un d'ells.

Contenidos

1	Capítulo 1. Introducción y objetivos.....	1
1.1	Introducción.....	1
1.2	Justificación.....	1
1.3	Objetivos.....	3
1.4	Estructura de la tesis.....	4
2	Capítulo 2. Estado del arte.....	7
2.1	Introducción.....	7
2.2	Fundamentos de estadística y probabilidad.....	8
2.2.1	Fundamentos de estadística e introducción al análisis de datos.....	8
2.2.2	Distribuciones de probabilidad continuas.....	13
2.2.3	Distribuciones de muestreo y teorema del límite central.....	17
2.2.4	Inferencia estadística.....	18
2.2.5	Prueba de Hipótesis.....	21
2.3	Predicción de electricidad.....	24
2.3.1	Estudio de la literatura.....	24
2.3.2	Redes neuronales artificiales.....	28
2.3.3	Métricas para evaluar predicciones.....	31
2.4	Patrones de consumo.....	32
2.4.1	Estudio de la literatura.....	33
2.4.2	El reconocimiento de patrones.....	36

2.4.3	Técnicas de clasificación de patrones	37
2.5	Detección de anomalías en el consumo de electricidad.....	40
2.5.1	Estudio de la literatura.....	40
2.5.2	Etiquetas de datos.....	41
2.5.3	Técnicas estadísticas de detección de anomalías	42
2.5.4	Ventajas y desventajas de las técnicas estadísticas.....	45
2.6	Conclusiones del capítulo	46
3	Capítulo 3. Sistemas de gestión de energía eléctrica y análisis de los datos de partida.....	49
3.1	Introducción.....	49
3.2	Sistemas de medición de la energía eléctrica.....	50
3.2.1	Evolución de la red eléctrica tradicional.....	50
3.3	Sistema de medición inteligente en la UPS	58
3.3.1	La demanda de electricidad en la Universidad Politécnica Salesiana Sede Cuenca.....	58
3.3.2	Puesta en marcha del sistema de medición inteligente	60
3.4	La demanda de potencia activa como serie temporal	67
3.4.1	Variación estacional de una serie temporal	68
3.4.2	Promedio móvil.....	68
3.4.3	Estacionalidad de los datos y desestacionalización	69
3.5	Análisis de la muestra de datos.....	70
3.5.1	Muestra de los datos.....	70
3.5.2	Tratamiento de los datos de la muestra.....	73
3.5.3	Tratamiento de los datos de la muestra en el edificio 5E de la UPV	82
3.6	Conclusiones del capítulo	89
4	Capítulo 4. Metodología estadística para valorar e identificar cambios en perfiles de consumo de electricidad (SAICC)	93
4.1	Introducción.....	93
4.2	Metodología SAICC.....	94

4.2.1	Etapa 1: Adquisición y procesamiento de datos históricos	94
4.2.2	Etapa 2: Obtención de patrones estocásticos de consumo.....	101
4.2.3	Etapa 3: Cuantificación estadística	104
4.2.4	Etapa 4: Presentación de resultados y análisis físico.....	106
4.3	Influencia del tratamiento de la demanda de electricidad como serie temporal en la obtención de patrones y detección de anomalías.....	111
4.3.1	Metodología	114
4.3.2	Comparación de resultados.....	117
4.3.3	Discusión y análisis de los resultados.....	124
4.4	Conclusiones del capítulo	127
5	Capítulo 5. Aplicaciones de la Metodología SAICC	129
5.1	Introducción	129
5.2	Aplicación de la metodología SAICC en el Edificio 5E de la UPV	130
5.2.1	Etapa 1: Adquisición y procesamiento de datos históricos	132
5.2.2	Etapa 2: Obtención de patrones de consumo	132
5.2.3	Etapa 3: Cuantificación estadística	135
5.2.4	Etapa 4: Presentación de resultados y análisis físico.....	135
5.2.5	Análisis de los resultados.....	138
5.3	Aplicación de la metodología SAICC en la Universidad Politécnica Salesiana Sede Cuenca.....	139
5.3.1	Etapa 1: Adquisición y procesamiento de datos históricos	139
5.3.2	Etapa 2: Obtención de patrones de consumo	140
5.3.3	Etapa 3: Cuantificación estadística	143
5.3.4	Etapa 4: Presentación de resultados y análisis físico.....	144
5.3.5	Análisis de los resultados.....	145
5.4	Aplicación de la metodología SAICC en el Alimentador 0101 de la Empresa Eléctrica Regional Centrosur (EERCS).....	147
5.4.1	Etapa 1: Adquisición y procesamiento de datos históricos	147
5.4.2	Etapa 2: Obtención de patrones de consumo	148

5.4.3	Etapa 3: Cuantificación estadística	152
5.4.4	Etapa 4: Presentación de resultados y análisis físico	153
5.4.5	Análisis de los resultados	155
5.5	Aplicación de la metodología SAICC en el Alimentador 0425 de la EERCS .	156
5.5.1	Etapa 1: Adquisición y procesamiento de datos históricos	156
5.5.2	Etapa 2: Obtención de patrones de consumo	157
5.5.3	Etapa 3: Cuantificación estadística	160
5.5.4	Etapa 4: Presentación de resultados y análisis físico	162
5.5.5	Análisis de los resultados	164
5.6	Aplicación de la metodología SAICC en redes de transmisión eléctricas	164
5.6.1	Etapa 1: Adquisición y procesamiento de datos históricos	165
5.6.2	Etapa 2: Obtención de patrones de consumo	165
5.6.3	Etapa 3: Cuantificación estadística	167
5.6.4	Etapa 4: Presentación de resultados y análisis físico	169
5.6.5	Análisis de los resultados	172
5.7	Aplicación de la metodología SAICC en la predicción de la demanda de electricidad	176
5.7.1	Aplicación de la metodología SAICC en la predicción de la demanda de electricidad	177
5.7.2	Predicción de la demanda de electricidad mediante redes neuronales LSTM	195
5.8	Conclusiones del capítulo	206
6	Capítulo 6. Conclusiones	211
6.1	Conclusiones	211
6.2	Principales contribuciones	216
6.3	Trabajos futuros.....	216
6.4	Publicaciones.....	217
	Bibliografía.....	219

Índice de Imágenes

Imagen 2.1 Diagramas de dispersión.....	10
Imagen 2.2 Histograma.....	11
Imagen 2.3 a) Sesgo a la derecha b) Sin sesgo c) Sesgo a la izquierda	11
Imagen 2.4 Diagrama de cajas.....	12
Imagen 2.5 Función de densidad de probabilidad.....	13
Imagen 2.6 Distribución normal y normal estandarizada	15
Imagen 2.7 Coeficiente de confianza en una distribución normal estandarizada	19
Imagen 2.8 Función de distribución de probabilidad chi cuadrada	23
Imagen 2.9 Número de documentos en la base de datos de Scopus acerca de la predicción y pronóstico de electricidad [6]	25
Imagen 2.10 Modelo básico de una neurona artificial	29
Imagen 2.11 Estructura básica de una red LSTM [29].....	31
Imagen 2.12 Documentos en la base de datos de Scopus de patrones de consumo de electricidad.....	33
Imagen 2.13 Etapas del reconocimiento de patrones	36
Imagen 2.14 Un diagrama de caja para datos univariados	43
Imagen 3.1 Red eléctrica tradicional [60]	51
Imagen 3.2 Áreas funcionales en una red eléctrica inteligente [63]	53
Imagen 3.3 Estructura de una REI [63]	54
Imagen 3.4 Sistema de medición inteligente.....	55
Imagen 3.5 Evolución de los sistemas de medición inteligente [63].....	56

Imagen 3.6 Diagrama de bloques de un medidor inteligente [63]	57
Imagen 3.7 Estructura física de un medidor inteligente [63]	57
Imagen 3.8 Ubicación geográfica de la UPS Sede Cuenca (© 2020 Google Maps)	59
Imagen 3.9 Diagrama unifilar de la red de media tensión de la UPS Sede Cuenca.....	59
Imagen 3.10 Demanda de electricidad anual en la UPS Sede Cuenca	60
Imagen 3.11 Edificios escogidos para la instalación de los medidores inteligentes	61
Imagen 3.12 Medidores instalados en la de la UPS Sede Cuenca.....	63
Imagen 3.13 Ubicación física de los medidores inteligentes en la UPS Sede Cuenca (© 2020 Google Maps).....	64
Imagen 3.14 Ubicación de los medidores instalados en el diagrama unifilar	64
Imagen 3.15 Estructura de la red de medición inteligente de la UPS	65
Imagen 3.16 Entorno de la plataforma web de los medidores inteligentes	65
Imagen 3.17 Visualización de la potencia demandada en la plataforma web de los medidores inteligentes.....	66
Imagen 3.18 Variación estacional en la demanda de potencia eléctrica absorbida trimestral en la UPS Sede Cuenca y su promedio móvil	68
Imagen 3.19 PCE diarios en la UPS Sede Cuenca para un periodo de un año	71
Imagen 3.20 Matriz de perfiles diarios de consumo de electricidad y toma de la muestra	71
Imagen 3.21 Histogramas para diferentes intervalos cuarto horarios considerando todos los PCE de un año – UPS.....	73
Imagen 3.22 Segmentación de los datos del consumo de electricidad.....	74
Imagen 3.23 Distribución de los datos para los días miércoles laborables en la UPS..	75
Imagen 3.24 Ajuste de los datos a una distribución normal (10:00-10:15 am) para un miércoles laborable en la UPS, valor $p = 0.00013$	76
Imagen 3.25 Método SAEC	77
Imagen 3.26 Consumo de electricidad - Universidad Politécnica Salesiana (09-03-2017 hasta 08-03-2018).....	78
Imagen 3.27 Consumo de electricidad aplicando el método SAEC - Universidad Politécnica Salesiana (09-03-2017 hasta 08-03-2018)	79
Imagen 3.28 Distribución de los datos para los días miércoles laborables en la UPS aplicando el método SAEC	79

Imagen 3.29 Ajuste de los datos a una distribución normal para los miércoles laborables (10:00-10:15 am), al aplicar el método SAEC, valor $p = 0.07639$	80
Imagen 3.30 Distribución de los datos para el patrón de los días miércoles laborables en la UPS.....	81
Imagen 3.31 Ajuste a la distribución normal del patrón de los días miércoles (10:00-10:15 am) en la UPS, valor $p = 0.10208$	81
Imagen 3.32 Distribución de los datos para los días miércoles laborables en la UPV..	82
Imagen 3.33 Ajuste de los datos a una distribución normal para los miércoles laborables (10:00-10:15 am) en la UPV, valor $p=0.20972$	83
Imagen 3.34 Distribución de los datos para los días miércoles laborables en la UPV aplicando el método SAEC.....	83
Imagen 3.35 Ajuste de los datos a una distribución normal para los días miércoles laborables (10:00-10:15 am) aplicando el método SAEC, valor $p = 0.12007$	84
Imagen 3.36 Ajuste a la distribución normal del patrón de los días miércoles (10:00-10:15 am) aplicando el método SAEC, valor $p = 0.73952$	84
Imagen 3.37 Distribución de los datos para los días miércoles laborables en la UPV en el periodo de verano.....	85
Imagen 3.38 Ajuste de los datos a la distribución normal para los miércoles laborables (10:00-10:15 am) en el periodo de verano - UPV, valor $p = 0.217003$	86
Imagen 3.39 Ajuste de los datos a una distribución normal para los miércoles laborables (10:00-10:15 am) en el periodo de verano, al aplicar el método SAEC, valor $p = 0.37131$	86
Imagen 3.40 Ajuste a la distribución normal del patrón de los miércoles laborables (10:00-10:15 am) en la UPV en el periodo de verano, valor $p = 0.53981$	87
Imagen 3.41 Distribución de los datos para los días miércoles laborables en la UPV en el periodo de invierno	88
Imagen 3.42 Ajuste de los datos a la distribución normal para los miércoles laborables (10:00-10:15 am) en el periodo de invierno - UPV, valor $p = 0.31706$	88
Imagen 3.43 Ajuste de los datos a la distribución normal para los miércoles laborables (10:00-10:15 am) en el periodo de invierno aplicando el método SAEC, valor $p = 0.55183$	89
Imagen 4.1 Diagrama de flujo de la metodología SAICC	95
Imagen 4.2 Potencia media trimestral en el Edificio 5E de la UPV	97

Imagen 4.3 Consumo de electricidad - Edificio 5E de la UPV (29-11-2015 hasta 28-11-2016).....	97
Imagen 4.4 Consumo de electricidad aplicando el método SAEC - Edificio 5E de la UPV (29-11-2015 hasta 28-11-2016)	98
Imagen 4.5 Segmentación de los datos a través de matrices desagregadas	99
Imagen 4.6 Obtención de características del patrón, media y desviación estándar de la potencia (96 valores por día) del Edificio 5E de la UPV	102
Imagen 4.7 Obtención de características del patrón, valores Z_{max} y Z_{min} del Edificio 5E de la UPV.....	102
Imagen 4.8 Patrones estocásticos de consumo en el Edificio 5E de la UPV representados mediante diagramas de cajas y bigotes	104
Imagen 4.9 Cuatro días evaluados con su correspondiente patrón de consumo	107
Imagen 4.10 Patrón de potencia máxima, media y mínima en comparación con cuatro días diferentes (cuatro casos)	109
Imagen 4.11 Análisis de la estacionalidad de datos en la metodología SAICC.....	115
Imagen 4.12 Potencia media trimestral en la UPS (P1: febrero–abril, P2: mayo–julio, P3: agosto–octubre, P4: noviembre–enero).....	117
Imagen 4.13 Media de los patrones de consumo	118
Imagen 4.14 Desviación estándar de los patrones de consumo.....	119
Imagen 4.15 Valores de Z_{max} de los patrones de consumo	119
Imagen 4.16 Valores de Z_{min} de los patrones de consumo	120
Imagen 4.17 Patrón de consumo de electricidad representado por diagramas de cajas y bigotes para un día miércoles laborable, Edificio 5E de la UPV	120
Imagen 4.18 Patrón de consumo de electricidad representado por diagramas de cajas y bigotes para un miércoles laborable, UPS	121
Imagen 4.19 Día analizado comparado con su patrón aplicando el método SF.....	123
Imagen 4.20 Día analizado comparado con su patrón aplicando el método SAEC....	123
Imagen 4.21 Perfiles de consumo de electricidad anómalos de acuerdo al método SF	126
Imagen 4.22 Perfiles de consumo de electricidad anómalos de acuerdo al método SAEC.....	126
Imagen 5.1 Edificio 5E de la UPV.....	131

Imagen 5.2 Diagrama unifilar de baja tensión del centro de transformación que alimenta al edificio 5E de la UPV	132
Imagen 5.3 Patrón y perfiles de consumo de electricidad de 6 días de análisis.....	133
Imagen 5.4 Patrón de consumo (Z_{min} y Z_{max}) y los valores de Z para cada uno de los 6 días de análisis	134
Imagen 5.5 Valores de potencia máxima, media y mínima para los 6 días de análisis y su correspondiente patrón	134
Imagen 5.6 Media del patrón de cada día de la semana en la UPS Cuenca (28-05-2017 hasta 27-05-2018)	140
Imagen 5.7 Desviación estándar del patrón de cada día de la semana en la UPS Cuenca (28-05-2017 hasta 27-05-2018)	141
Imagen 5.8 Patrones estocásticos de consumo de cada día de la semana en la UPS – Cuenca.....	142
Imagen 5.9 Cuantificación estadística de 5 días miércoles versus su patrón de consumo	143
Imagen 5.10 Datos estandarizados para 5 días miércoles anómalos.....	144
Imagen 5.11 Potencia máxima, media y mínima de los 5 días analizados versus el patrón del día miércoles.....	144
Imagen 5.12 Transformador No. 1 de la Subestación 1 de la EERCS [66].....	147
Imagen 5.13 Diagrama Unifilar de la Subestación 1 de la EERCS [66].....	148
Imagen 5.14 Media del patrón de cada día de la semana en el alimentador 0101 (01-01-2017 hasta 31-12-2017)	149
Imagen 5.15 Desviación estándar del patrón de cada día de la semana del alimentador 0101 (01-01-2017 hasta 31-12-2017)	149
Imagen 5.16 Variación del índice de estacionalidad del alimentador 0101	151
Imagen 5.17 Patrones estocásticos de consumo de cada día de la semana en el Alimentador 0101	151
Imagen 5.18 Cuantificación estadística de 5 miércoles anómalos versus su patrón de consumo en el alimentador 0101	152
Imagen 5.19 Datos estandarizados de 5 miércoles anómalos y su patrón en el alimentador 0101	152
Imagen 5.20 Potencia máxima, media y mínima de 5 miércoles anómalos versus su patrón	153

Imagen 5.21 Subestación 4 EERCS [66]	156
Imagen 5.22 Diagrama unifilar de la Subestación 4 de la EERCS [66].....	157
Imagen 5.23 Variación del índice de estacionalidad del alimentador 0425 de la EERCS	158
Imagen 5.24 Media del patrón de cada día de la semana del alimentador 0425 (01-01- 2017 hasta el 31-12-2017).....	158
Imagen 5.25 Desviación estándar del patrón de cada día de la semana del alimentador 0425 (01-01-2017 hasta el 31-12-2017)	159
Imagen 5.26 Patrones estocásticos de cada uno de los días de la semana del Alimentador 0425.....	160
Imagen 5.27 Cuantificación estadística de 4 martes anómalos versus su patrón en el alimentador 0425.....	161
Imagen 5.28 Datos estandarizados de 4 martes anómalos y su patrón en el alimentador 0425	161
Imagen 5.29 Potencia máxima, media y mínima de 4 días martes anómalos en el alimentador 0425.....	162
Imagen 5.30 Línea de transmisión Molino – Pascuales en Ecuador	164
Imagen 5.31 Línea de transmisión Molino – Pascuales en Ecuador [107]	165
Imagen 5.32 Media del patrón de cada día de la semana. Flujo de potencia activa línea Molino-Pascuales (01-10-2017 hasta el 30-09-2018).....	166
Imagen 5.33 Desviación estándar del patrón de cada día de la semana. Flujo de potencia activa línea Molino-Pascuales (01-10-2017 hasta el 30-09-2018)	166
Imagen 5.34 Patrones estocásticos de cada uno de los días de la semana en la línea de transmisión Molino-Pascuales	168
Imagen 5.35 Clasificación de perfiles de carga de acuerdo al IoC de la línea de transmisión Molino-Pascuales (días martes y miércoles).....	170
Imagen 5.36 Flujos de potencia del clúster c versus su patrón.....	171
Imagen 5.37 Flujos de potencia máximo, medio y mínimo en contraste con el patrón	172
Imagen 5.38 Clústeres formados para los perfiles de carga de la línea de transmisión Molino-Pascuales (01-10- 2017 hasta 30-09- 2018) con la técnica k means.....	173
Imagen 5.39 Clústeres formados para los perfiles de carga de la línea de transmisión Molino-Pascuales para los días martes y miércoles (01-10- 2017 hasta 30-09- 2018) con la técnica k means.....	174

Imagen 5.40 Clasificación de los perfiles de carga de la línea de transmisión Molino-Pascuales Clústeres para los días martes y miércoles utilizando k means y SVM para el clúster 1	175
Imagen 5.41 Clasificación de los perfiles de carga de la línea de transmisión Molino-Pascuales Clústeres para los días martes y miércoles utilizando k means y SVM para el clúster 2	175
Imagen 5.42 Obtención de medias y desviaciones estándar en la matriz de patrón final	178
Imagen 5.43 Índice de estacionalidad para un periodo de 364 días	180
Imagen 5.44 Media del patrón de cada día de la semana desde el 16-10-2017 hasta 14-10-2018 (periodo de obtención de patrones 364 días)	180
Imagen 5.45 Desviación estándar del patrón de cada día de la semana desde el 16-10-2017 hasta 14-10-2018 (periodo de obtención de patrones 364 días).....	181
Imagen 5.46 Intervalos de predicción con el 60%, 80% y 95% de probabilidad desde el 15-10-2018 hasta el 21-10-2018 (periodo de obtención de patrones 364 días).....	182
Imagen 5.47 Intervalos de predicción con el 60%, 80% y 95% de probabilidad desde el 15-10-2018 hasta el 21-10-2018 (periodo de obtención de patrones 182 días).....	183
Imagen 5.48 Intervalos de predicción con el 60%, 80% y 95% de probabilidad desde el 15-10-2018 hasta el 21-10-2018 (periodo de obtención de patrones 119 días).....	184
Imagen 5.49 Porcentaje diario de valores reales que caen dentro de cada intervalo de predicción para periodos de obtención de patrones de 364 y 182 días	186
Imagen 5.50 Selección de datos para la obtención de intervalos de predicción para el caso de España peninsular (182 días de análisis)	187
Imagen 5.51 Media del patrón de cada día de la semana (periodo de obtención de patrones 182 días)	188
Imagen 5.52 Desviación estándar del patrón de cada día de la semana (periodo de obtención de patrones 182 días)	189
Imagen 5.53 Intervalos de predicción con el 60%, 80% y 95% de probabilidad desde el 21-01-2019 hasta el 27-01-2019 (periodo de obtención de patrones 182 días).....	190
Imagen 5.54 Intervalos de predicción con el 60%, 80% y 95% de probabilidad desde el 21-01-2019 hasta el 27-01-2019 (periodo de obtención de patrones 119 días).....	191
Imagen 5.55 Porcentaje diario de valores reales que están dentro de cada intervalo de predicción para periodos de obtención de patrones de 182 y 119 días	193
Imagen 5.56 Intervalo de predicción de la demanda de potencia activa en Ecuador continental desde el 15 al 28 de octubre de 2018	194

Imagen 5.57 Intervalo de predicción de la demanda de potencia activa en España desde el 04 al 17 de febrero de 2019.....	195
Imagen 5.58 Red neuronal artificial LSTM utilizada para la predicción.....	196
Imagen 5.59 Predicción desde el 15-10-2018 hasta el 21-10-2018 para Ecuador utilizando una RNA LSTM.....	197
Imagen 5.60 Predicción desde el 29-10-2018 hasta el 04-11-2018 para Ecuador utilizando una red neuronal LSTM.....	199
Imagen 5.61 Predicción desde el 29-10-2018 hasta el 04-11-2018 para Ecuador utilizando la media del intervalo de predicción del método SAICC.....	200
Imagen 5.62 Predicción desde el 29-10-2018 hasta el 04-11-2018 para Ecuador utilizando la RNA LSTM y la metodología SAICC.....	201
Imagen 5.63 Predicción desde el 24-09-2018 hasta el 30-09-2018 para Ecuador utilizando la RNA LSTM y la metodología SAICC.....	202
Imagen 5.64 Predicción del día viernes 28 de septiembre de 2018 mediante la RNA LSTM y la metodología SAICC.....	203
Imagen 5.65 Predicción desde el 21-01-2019 hasta el 27-01-2019 para España peninsular utilizando una RNA LSTM.....	204
Imagen 5.66 Predicción desde el 28-01-2019 hasta el 03-02-2019 para la España peninsular utilizando la RNA LSTM y la metodología SAICC.....	205

Índice de Tablas

Tabla 2.1 Prueba de hipótesis estadística	22
Tabla 2.2 Herramientas de predicción disponibles.....	26
Tabla 2.3 Aplicaciones de los patrones de consumo de electricidad a lo largo del tiempo.....	36
Tabla 3.1 Red eléctrica tradicional vs una red eléctrica inteligente [64].....	52
Tabla 3.2 Corrientes máximas medidas en los puntos de medida [65].....	62
Tabla 3.3 Medidores utilizados en la implementación del sistema de medición [65] ...	62
Tabla 3.4 Contrastación de mediciones en la acometida principal en el año 2017	67
Tabla 3.5 Variables categóricas y criterios considerados en la segmentación de datos	74
Tabla 4.1 Resultados del proceso de agrupamiento	103
Tabla 4.2 Variables calculadas	106
Tabla 4.3 Pruebas aplicadas en la metodología SAICC	107
Tabla 4.4 Tabla multicriterio de interpretación de las pruebas aplicadas (MCIAT) ...	108
Tabla 4.5 Catalogación del PCE diario usando los valores de IoC.....	110
Tabla 4.6 Revisión de la literatura de patrones de consumo	112
Tabla 4.7 Evaluación de las anomalías detectadas en el edificio 5E de la UPV	122
Tabla 4.8 Evaluación de las anomalías detectadas en la UPS.....	124
Tabla 4.9 Detección de anomalías para diferentes métodos de análisis de estacionalidad	125
Tabla 5.1 Resultados de la cuantificación estadística para el día de análisis 4.....	135
Tabla 5.2 Resultados de las 12 pruebas aplicadas al día de análisis 4.....	136

Tabla 5.3 Resumen de las pruebas aplicadas a los días de análisis	136
Tabla 5.4 Conclusiones del consumo de electricidad de los 6 días de análisis	137
Tabla 5.5 Conformación de los patrones de electricidad por día de la semana en la UPS Cuenca	141
Tabla 5.6 Resumen de las pruebas aplicadas a los días de análisis en la UPS Sede Cuenca	145
Tabla 5.7 Conclusiones de la metodología SAICC e IoC para los días de análisis	146
Tabla 5.8 Conformación de los patrones de electricidad para cada día de la semana en el alimentador 0101	150
Tabla 5.9 Conclusiones de la metodología SAICC e IoC (EERCS)	154
Tabla 5.10 Catalogación del perfil de consumo de electricidad usando los valores de IoC para subestaciones	155
Tabla 5.11 Conformación de los patrones de electricidad para cada día de la semana en alimentador 0425	159
Tabla 5.12 Resumen de las pruebas aplicadas a los días de análisis	162
Tabla 5.13 Conclusiones de la metodología SAICC e IoC (Alimentador 0425)	163
Tabla 5.14 Conformación de los patrones de electricidad por día de la semana (línea de Transmisión Molino-Pascuales).....	167
Tabla 5.15 Interpretación del IoC aplicada a líneas de transmisión	171
Tabla 5.16 Conformación de los patrones de electricidad para el periodo de 364 días	181
Tabla 5.17 Porcentaje de variación de los intervalos de predicción (periodo de obtención de patrones 364 días).....	182
Tabla 5.18 Resumen de los intervalos de predicción desde el 15 al 21 de octubre de 2018 (periodo de obtención de patrones 364 días)	183
Tabla 5.19 Resumen de la aplicación de los intervalos de predicción desde el 15 al 21 de octubre del 2018 (periodo de obtención de patrones 182 días)	184
Tabla 5.20 Resumen de la aplicación de los intervalos de predicción desde el 15 al 21 de octubre del 2018 (periodo de obtención de patrones 119 días)	185
Tabla 5.21 Resumen de la aplicación de los intervalos de predicción desde el 15 de octubre al 11 de noviembre del 2018	185
Tabla 5.22 Conformación de los patrones de electricidad para el periodo de 182 días	189

Tabla 5.23 Porcentaje de variación de los intervalos de predicción (periodo de obtención de patrones 182 días)	190
Tabla 5.24 Resumen de la aplicación de los intervalos de predicción desde el 21 al 27 de enero del 2019 (periodo de obtención de patrones 182 días)	191
Tabla 5.25 Resumen de la aplicación de los intervalos de predicción desde el 21 al 27 de enero del 2019 (periodo de obtención de patrones 119 días)	192
Tabla 5.26 Resumen de los intervalos de predicción obtenidos desde el 21 de enero al 17 de febrero del 2019 para España peninsular	192
Tabla 5.27 Parámetros de la RNA LSTM	196
Tabla 5.28 PEMA en la predicción de la demanda de electricidad en Ecuador para trece semanas mediante la RNA LSTM (10-09-2018 hasta 09-12-2018)	198
Tabla 5.29 PEMA en la predicción de la demanda de electricidad en Ecuador para trece semanas mediante el método SAICC (10-09-2018 hasta 09-12-2018).....	200
Tabla 5.30 Porcentaje de variación de los intervalos de predicción con respecto a su media (viernes 28 de septiembre de 2018)	203
Tabla 5.31 PEMA en la predicción de la demanda de electricidad en España peninsular para cuatro semanas utilizando la RNA LSTM (21-01-2019 hasta 17-02-2019).....	204
Tabla 5.32 PEMA en la predicción de la demanda de electricidad en España peninsular para cuatro semanas mediante la metodología SAICC (21-01-2019 hasta 17-02-2019)	205

Lista de Acrónimos

AMI	Advanced Metering Infrastructure
AMR	Advanced Meter Reading
ARIMA	Anglosajón Autoregressive Integrated Moving Average
ARMA	Anglosajón Autoregressive Moving Average
CUMSUM	Cumulative Sum
DL	Día Laborable
DM	Detrending Method
DNL	Día No Laborable
ECM	Error Cuadrático Medio
EERCS	Empresa Eléctrica Regional Centrosur
FN	Falso Negativo
FP	Falso Positivo
GMM	Gaussian Mixture Models
GD	Generación Distribuida
HL	Horas Laborables
HNL	Horas No Laborables
IA	Inteligencia Artificial
IoC	Index of Change
LSTM	Long Short Term Memory
LVQ	Learning Vector Quantization
MCIAT	Multi-Criteria Interpretation of Applied Test Table
MI	Medición Inteligente
NZEB	Nearly Zero Energy Building
PCE	Perfil de Consumo de Electricidad
PDF	Probability Distribution Function
PEMA	Porcentaje del Error Medio Absoluto
PMU	Phasor Measurement Unit

RECM	Raíz Cuadrática del Error Cuadrático Medio
REI	Red Eléctrica Inteligente
RIC	Rango Intercuartílico
RNA	Red Neuronal Artificial
RNAR	Red Neuronal Artificial Recurrente
SAEC	Seasonality Analysis of Electricity Consumption
SAICC	Statistical Assessment for Identifying Changes in Consumption
Sc	State Cell
SCADA	Supervisory Control and Data Acquisition
SF	Seasonal Filter
SIM	Subscriber Identity Module
SSE	Sistema de Suministro Eléctrico
SVM	Support Vector Machines
TFP	Tasa de Falsos Positivos
TFN	Tasa de Falsos Negativos
TCs	Transformadores de Corriente
UPS	Universidad Politécnica Salesiana
UPV	Universidad Politécnica de Valencia
VN	Verdadero Negativo
VP	Verdadero Positivo
WSA	Without Seasonality Analysis
ZEB	Zero Energy Buildings

Capítulo 1

Introducción y objetivos

1.1 Introducción

La demanda de energía en el planeta continúa incrementándose y parece que seguirá haciéndolo por muchos años más. Según la *Agencia Internacional de la Energía* [1] en el periodo de 1990 a 2016 la población mundial aumentó un 41%, la demanda de energía y las emisiones de CO₂ lo han hecho en un 55% aproximadamente, mientras que el consumo de electricidad creció un 111.8%. La electricidad en el mundo actual es un elemento indispensable para el desarrollo de las sociedades, su ausencia prolongada provocaría caos generalizado y consecuencias inimaginables.

Hoy en día existe una tendencia mundial a electrificar los diferentes sectores, por lo tanto cualquier esfuerzo es válido cuando se intenta mejorar el aprovechamiento y la eficiencia de las redes y de los sistemas eléctricos en general. Las nuevas redes eléctricas inteligentes (REI) han evolucionado tecnológicamente en muchos aspectos, uno de ellos se refiere a la disponibilidad de datos de la demanda en diferentes puntos y niveles de la red, como en transformadores en redes de transmisión, alimentadores, transformadores de distribución y en los grandes y pequeños consumidores. Estos datos suministrados por los medidores inteligentes constituyen información de partida fundamental para la gestión y planificación de los sistemas eléctricos.

1.2 Justificación

Los sistemas de gestión de la demanda de energía eléctrica asociados al *reconocimiento de patrones* se han estudiado de manera escasa hasta el momento. La revisión de la literatura ha permitido identificar algunas limitaciones en esta área, de entre las cuales se destacan:

- La caracterización de la demanda mediante el reconocimiento de patrones no se ha utilizado para la identificación y valoración de cambios en el consumo de energía eléctrica.
- Los sistemas de monitorización y vigilancia del consumo de electricidad no identifican las posibles causas de las anomalías detectadas.
- La detección de anomalías contextuales (datos atípicos en un contexto pero no en otro) en el consumo de electricidad se ha explorado de manera exigua, por lo tanto, es posible mejorar la precisión de los sistemas de vigilancia en este aspecto.

Las limitaciones detectadas han motivado el desarrollo de una nueva metodología estadística para caracterizar el comportamiento de la demanda de energía de los consumidores y otros puntos de la red eléctrica mediante la identificación y obtención de patrones. El comportamiento de la demanda de electricidad en un sistema común con n cargas puede considerarse no determinista, debido a la aleatoriedad con que ese conjunto de cargas demandan energía. El análisis estadístico de datos históricos del consumo energético permite definir comportamientos recurrentes y predecibles con presencia de incertidumbre, a los cuales en esta tesis se les ha denominado *patrones estocásticos*. La utilización de estos patrones permite valorar e identificar cambios en perfiles de carga de electricidad (PCE). La valoración de los cambios en la demanda de electricidad permite asociar estos valores a posibles eventos en una instalación. Esto puede ser empleado para detectar anomalías y catalogar perfiles de carga de acuerdo al cambio que han tenido con respecto a su comportamiento habitual, lo que permitiría identificar modos de trabajo de los sistemas eléctricos. Esta información resulta útil en la gestión, ya que es posible generar alarmas, reducir costos en mantenimiento y aplicar medidas adecuadas rápidamente cuando se presenten incidentes.

La predicción de la demanda representa una herramienta útil en la gestión de los sistemas de suministro eléctrico. En la actualidad herramientas tales como las redes neuronales y el aprendizaje profundo son las preferidas para realizar esta tarea, sin embargo, la revisión de la literatura y su utilización ha revelado las siguientes limitaciones:

- Tienen dificultad para cuantificar la incertidumbre de los resultados e interpretarlos físicamente, ya que la demanda de electricidad es una variable estocástica continua.
- Requieren un gasto computacional elevado.
- El establecimiento de la estructura y configuración de la red neuronal, esto es, número de neuronas, capas, algoritmo de optimización, etc., requiere experiencia y tiempo considerable.

Por lo expuesto, se considera necesario proponer estrategias y metodologías para solventar las limitaciones detectadas en las herramientas derivadas de la inteligencia artificial (IA), de tal manera que sea posible acotar la incertidumbre de las predicciones mediante una metodología de simple aplicación y bajos requerimientos computacionales.

Por otro lado, los datos proporcionados por los medidores inteligentes requieren un tratamiento previo para ajustarlos a una distribución de probabilidad conocida, facilitando su interpretación y validando el análisis estadístico. En este contexto, una adquisición adecuada de los datos, el posterior procesamiento y análisis exploratorio son tan importantes como aplicar cualquier técnica o herramienta posterior en el reconocimiento de patrones, detección de anomalías o predicción de la demanda de electricidad. Por ello, es de trascendental importancia relacionar a la potencia eléctrica activa absorbida por un consumidor o un sistema en general con la teoría existente de las series temporales de datos.

La investigación presentada en este documento ha sido realizada en el Instituto de Ingeniería Energética (IIE) de la Universitat Politècnica de València (UPV) y con el soporte del Grupo de Investigación en Energías (GIE) de la Universidad Politécnica Salesiana (UPS) de Cuenca, Ecuador mediante el proyecto de investigación titulado *Mejora de la Eficiencia Energética en Edificaciones de la Zona Ecuatorial de los Andes* con vigencia desde enero de 2016 hasta enero de 2020. El trabajo realizado en estos 4 años inició enfocado en la optimización de energía en edificaciones, sin embargo, las aportaciones realizadas se han extendido a los sistemas de suministro eléctrico en general.

1.3 Objetivos

El objetivo principal de esta tesis es el desarrollo de una metodología que permita una adecuada caracterización de la demanda de electricidad mediante la identificación de patrones estocásticos para mejorar la gestión de las nuevas REI.

El cumplimiento del objetivo principal demanda la consecución sistemática de los siguientes objetivos específicos:

1. Recopilar y evaluar los conceptos estadísticos, herramientas de predicción, reconocimiento de patrones y detección de anomalías que pueden ser aplicados a la caracterización y gestión de la demanda de energía eléctrica, para comprender adecuadamente el análisis, evaluación de datos y las metodologías que se han desarrollado en el área de estudio.
2. Desarrollar un método estadístico que permita un tratamiento adecuado de la demanda de energía eléctrica entendida como una serie temporal de datos de una va-

- riable estocástica continua. El método debe obtener una distribución de probabilidad definida que facilite el análisis e interpretación de esos datos.
3. Implementar una metodología estadística que pueda ser aplicada a los consumidores y otros puntos o niveles de una red eléctrica, con el objetivo de obtener patrones, caracterizar la demanda, valorar e identificar cambios en PCE, detectar anomalías e indicar sus posibles causas. Para esto se debe considerar la influencia del tratamiento de las componentes de la serie temporal de datos en la obtención de los patrones de consumo y la detección de anomalías. La detección de anomalías debe experimentar un incremento en la precisión al identificar las anomalías puntuales, colectivas y contextuales.
 4. Desarrollar una herramienta estadística para la predicción de la demanda de electricidad que permita cuantificar la incertidumbre de las predicciones obtenidas, utilizando el reconocimiento de patrones como punto de partida.
 5. Evaluar la aplicación de la metodología en diversos casos de estudio reales en diferentes puntos y niveles de la red eléctrica, identificando los beneficios que pueden obtenerse en la gestión de cada uno de esos sistemas. En este sentido se debe trabajar con una muestra de datos real y representativa, lo cual evita posibles sesgos en el análisis estadístico y problemas de generalización de la metodología.

1.4 Estructura de la tesis

Para alcanzar los objetivos especificados en el apartado anterior, la presente tesis se ha organizado de la siguiente manera:

En el *Capítulo 2. Estado del arte*, se realiza una recopilación de los conceptos de la teoría de probabilidad y estadística necesarios para comprender adecuadamente las metodologías de análisis y evaluación de datos que se estudian en los capítulos posteriores. Seguidamente, se analizan estudios realizados en la predicción de la demanda de electricidad y se estudian las técnicas de reconocimiento y clasificación de patrones. Para finalizar el capítulo, se presenta una recopilación de las técnicas de detección de anomalías utilizadas en el ámbito científico, así como su aplicación en los sistemas eléctricos.

De manera breve, en el *Capítulo 3. Sistemas de gestión de energía eléctrica y análisis de los datos de partida*, se describen los sistemas de medición inteligente (MI) disponibles en la actualidad y la evolución que han tenido. Además, se presenta la implementación y puesta en marcha de un sistema de MI, el cual permite la adquisición de los datos esenciales para el desarrollo de esta tesis. También, se describen las características de la potencia eléctrica activa absorbida en general, entendida como una serie temporal de datos. Se detalla la muestra estadística, las características y propiedades

que poseen. Por último, se propone un método para el tratamiento de esta serie temporal de datos que puede ser aplicado para cualquier tipo de consumidor. Este método permite ajustar los datos a una distribución normal, lo cual facilita su análisis e interpretación.

La caracterización de la demanda de electricidad de los consumidores hoy en día es posible gracias a los datos proporcionados por los medidores inteligentes. Tras la adquisición y tratamiento de la serie temporal de datos, en el *Capítulo 4. Metodología estadística para valorar e identificar cambios en perfiles de consumo de electricidad*, se presenta una nueva metodología estadística que puede ser aplicada a los consumidores y otros puntos o niveles de la red eléctrica con el objeto de obtener patrones, valorar e identificar cambios en los PCE, detectar anomalías e indicar sus posibles causas. En la segunda parte del capítulo se presenta un análisis de la influencia del tratamiento de la serie temporal de datos en la obtención de los patrones de consumo y la detección de anomalías.

Dada la disponibilidad de datos de la demanda en diferentes puntos y niveles de la red eléctrica, en el *Capítulo 5. Aplicaciones de la Metodología SAICC*, se aplica la metodología desarrollada en el Capítulo 4 a casos de estudio reales. En primer lugar se la aplica al edificio 5E de la UPV en Valencia, España. Posteriormente, se evalúa el consumo de electricidad de un conjunto de edificios al analizar los datos correspondientes a la UPS Sede Cuenca en Ecuador. También se presenta el análisis de la demanda de electricidad en dos alimentadores de diferentes subestaciones de la ciudad de Cuenca en Ecuador. El primer alimentador suministra energía a una zona residencial y comercial, en tanto que el segundo a una zona industrial. Posteriormente, el análisis se extiende a una de las líneas de transmisión eléctrica más importantes de Ecuador, en este caso la Molino-Pascuales. Por último, la metodología propuesta se utiliza para predecir la demanda de electricidad desde una perspectiva distinta a los estudios tradicionales al obtener un intervalo de predicción con una probabilidad asociada a que un valor real caiga dentro del intervalo, acotando de esta manera la incertidumbre de los valores predichos.

En el *Capítulo 6. Conclusiones*, se resumen las conclusiones obtenidas en la presente tesis, las aportaciones realizadas y se sugieren posibles desarrollos futuros relacionados. Para finalizar, se presenta el listado de las publicaciones realizadas durante el desarrollo de este trabajo.

Capítulo 2

Estado del arte

2.1 Introducción

La electricidad en el mundo actual es un recurso indispensable para el desarrollo de las sociedades, es tan importante que es muy difícil imaginar la vida cotidiana sin ella. Este tipo de energía final es usada en el sector doméstico, comercial, industrial, transporte, etc., y en aplicaciones diversas, por ejemplo productivas, de comunicación o de ocio. Encender una luz, tener aire acondicionado en casa o en la oficina, usar el teléfono móvil o un computador es tan natural que nos parecen servicios elementales. Su ausencia prolongada provocaría caos total y consecuencias inimaginables.

Una pequeña muestra de lo que podría pasar se dio el 14 de agosto de 2003 cuando un corte de energía eléctrica afectó al noreste de Estados Unidos y el sureste de Canadá. Alrededor de 50 millones de personas se quedaron sin suministro de electricidad durante al menos 24 horas, la situación se normalizó completamente 4 días después. El caos fue más evidente en Nueva York, gente atrapada en ascensores de los rascacielos, vagones del metro y trenes, aeropuertos fuera de servicio, atascos interminables causados por semáforos apagados, altas temperaturas con ausencia de aires acondicionados y agua fresca para beber. La telefonía móvil dejó de funcionar dejando incomunicadas a las personas, a esto hay que sumarle una serie de saqueos, robos y una psicosis colectiva por el temor a un atentado terrorista. La causa del apagón fue una sobrecarga de una línea de transmisión en Ontario (Canadá) hacia Estados Unidos. Los apagones de electricidad no son tan frecuentes en los países desarrollados, pero de vez en cuando ocurren recordando que la gestión de la electricidad no se puede descuidar ya que es necesaria para garantizar su suministro.

La gestión de los sistemas eléctricos tiene varias áreas de trabajo. En esta tesis se aborda el reconocimiento de patrones, la caracterización, identificación de cambios, detección de anomalías y la predicción en la demanda de electricidad. De manera transver-

sal, se ha utilizado la estadística y la teoría de probabilidad como herramientas fundamentales para el análisis de datos, por lo que son de vital importancia en este trabajo.

En este capítulo, inicialmente se presenta una recopilación de los conceptos estadísticos necesarios para comprender adecuadamente las metodologías de análisis y evaluación de datos que se exponen en los capítulos posteriores. Seguidamente, se analizan estudios realizados en la predicción de la demanda de electricidad, ya que esta se considera una herramienta poderosa en la gestión de la energía. A continuación se realiza un estudio de la literatura acerca de los patrones de consumo, así como las técnicas de reconocimiento y clasificación de patrones, las cuales son utilizadas para el desarrollo de la metodología SAICC presentada en el Capítulo 4. Al finalizar el capítulo se presenta una recopilación de las técnicas de detección de anomalías utilizadas en el ámbito científico, así como su aplicación en los sistemas eléctricos con el objeto de identificar consumos atípicos en un escenario en el que el volumen de datos crece día a día.

2.2 Fundamentos de estadística y probabilidad

La estadística es fundamental para respaldar hallazgos científicos y el tratamiento de datos de consumo de electricidad no es la excepción. El concepto de estadística puede definirse como la ciencia que recopila datos, para luego, organizarlos, presentarlos, analizarlos y finalmente interpretarlos, con el objeto de propiciar una toma de decisiones más eficaz [2].

El comportamiento del consumo de energía eléctrica se puede conocer en detalle mediante el análisis de sus datos. Sin embargo, es necesaria una interpretación adecuada para evitar errores y posibles sesgos en la presentación de los resultados. La ventaja principal de utilizar métodos estadísticos y la teoría de probabilidad en estos estudios, es la solución justificada en los resultados obtenidos, a diferencia de otros métodos utilizados actualmente como son los derivados de la inteligencia artificial.

En el Capítulo 4 se presenta una metodología estadística para valorar e identificar cambios en el perfil de consumo de electricidad, por lo que, en este capítulo se presenta una recopilación de conceptos de estadística y probabilidad necesarios para comprender adecuadamente la metodología propuesta.

2.2.1 Fundamentos de estadística e introducción al análisis de datos

La estadística descriptiva ayuda al estudio de la variabilidad de las observaciones, las cuales pueden clasificarse en medidas de localización y medidas de variabilidad.

2.2.1.1 Medidas de localización

Media: La media muestral está definida por la Ecuación 2.1:

$$\bar{x} = \sum_{i=1}^n \frac{x_i}{n} \quad \text{Ecuación 2.1}$$

En donde x_i es cada una de las i observaciones de la muestra y n es el número de observaciones.

Mediana: La mediana muestra la tendencia central de los datos sin ser influenciada por los valores extremos. Para encontrar la mediana muestral es necesario ordenar cada observación de la muestra de menor a mayor y aplicar la Ecuación 2.2.

$$\tilde{x} = \begin{cases} x_{(n+1)/2}, & \text{si } n \text{ es impar} \\ \frac{1}{2}(x_{n/2} + x_{n/2+1}), & \text{si } n \text{ es par} \end{cases} \quad \text{Ecuación 2.2}$$

Moda: La moda es el valor que más se repite en la muestra.

Existen otras medidas de localización, generalmente derivadas de la media y de la mediana, sin embargo no se explican en este documento ya que se utilizan en casos puntuales.

2.2.1.2 Medidas de variabilidad

La variabilidad de los datos es un ámbito de estudio muy importante en la estadística, ya que de ella dependen procesos y productos. Por ejemplo, al realizar mediciones de la potencia consumida en una instalación, se puede notar claramente que los datos presentan mayor variabilidad en las horas laborables (HL) en comparación con las horas no laborables (HNL) del día. Por lo general las medidas de localización se analizan en conjunto con las medidas de variabilidad para tener una lectura adecuada de la naturaleza del conjunto de datos [3].

Rango: Esta medida calcula la diferencia entre el valor máximo y el valor mínimo de la muestra [3].

Varianza: La varianza muestral se representa generalmente por s^2 y se calcula mediante la siguiente expresión:

$$s^2 = \sum_{i=1}^n \frac{(x_i - \bar{x})^2}{n - 1} \quad \text{Ecuación 2.3}$$

Desviación estándar: La desviación estándar es la medida de variabilidad más utilizada, esta se obtiene simplemente al calcular la raíz cuadrada de la varianza, tal como se muestra:

$$s = \sqrt{s^2} \quad \text{Ecuación 2.4}$$

2.2.1.3 Representación gráfica de datos

La representación gráfica de los datos a menudo es una forma efectiva de complementar el análisis de poblaciones estadísticas. La observación de gráficas en el análisis exploratorio de los datos permite tomar decisiones de cómo se realizará el estudio formal. Para este trabajo se han considerado diagramas de dispersión, histogramas y diagramas de cajas.

Diagramas de dispersión: El diagrama de dispersión se realiza generalmente con dos variables representadas en un plano cartesiano. En el eje de las abscisas se representa la variable independiente y en el eje de las ordenadas se representa la variable dependiente (ver Imagen 2.1). Este diagrama muestra cómo se relaciona una variable con otra [2].

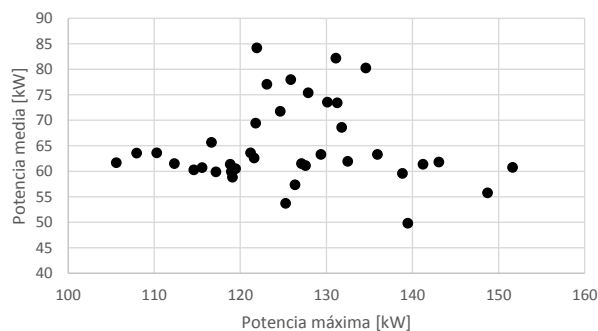


Imagen 2.1 Diagramas de dispersión

Histograma: El histograma es una gráfica que se utiliza regularmente para representar la función de distribución probabilidad PDF (del término anglosajón, probability distribution function) de una variable aleatoria continua. En el eje de las abscisas se señalan los intervalos de clase, mientras que, en el eje de las ordenadas la frecuencia relativa correspondiente (ver Imagen 2.2). La frecuencia relativa se obtiene al dividir el número de observaciones de cada clase (frecuencia de clase) entre el número total de observaciones.

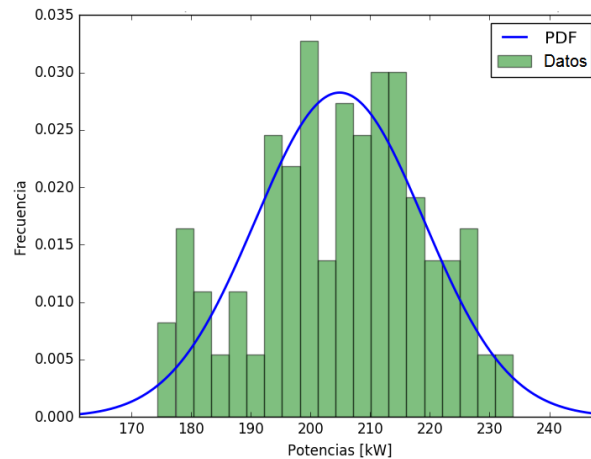


Imagen 2.2 Histograma

Este tipo de gráficos ayudan a comprender como están distribuidos los datos. Por ejemplo se dice que los datos están sesgados cuando su curva de distribución no es simétrica sobre un eje vertical, es decir, cuando una de las colas es más larga. Si no existe simetría, los datos pueden estar sesgados a la izquierda o a la derecha, tal como se indica en la Imagen 2.3.

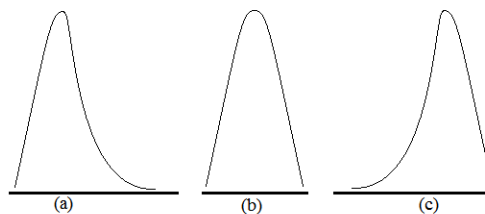


Imagen 2.3 a) Sesgo a la derecha b) Sin sesgo c) Sesgo a la izquierda

Diagramas de caja: Este diagrama se utiliza para variables continuas y se basa en cuartiles. Inicialmente se representa una caja cuya longitud es el rango intercuartil, es decir cuyo extremo inferior y superior son el percentil 25 (Q1) y 75 (Q3). Dentro de la caja, se dibuja con una línea recta la mediana o percentil 50 (Q2). Por fuera de la caja se representan dos bigotes cuyos extremos representan los valores mínimos y máximos (ver Imagen 2.4). Este tipo de gráfica es ideal para observar valores alejados de la masa de datos, por ejemplo es muy común considerar que si la distancia desde la caja excede 1.5 veces el rango intercuartil, la observación es anómala [3].

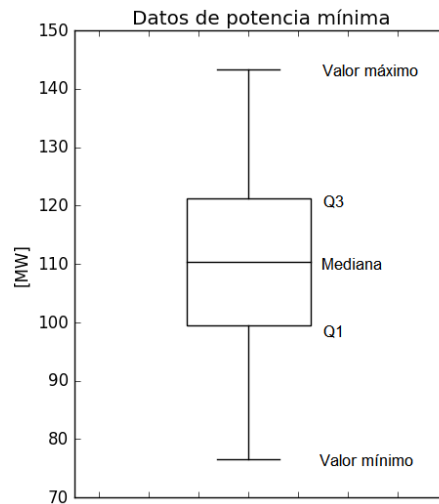


Imagen 2.4 Diagrama de cajas

2.2.1.4 Variables aleatorias

En estadística se utiliza la palabra “experimento” para definir un proceso que genera datos. El conjunto de todos los resultados posibles de ese experimento se denomina espacio muestral. A cada posible resultado se le llama punto muestral. Los puntos muestrales pueden ser continuos (si corresponden a un intervalo de los números reales) o discretos (si se pueden asociar a números naturales) [3].

El concepto de variable aleatoria aparece cuando a cada punto muestral se le asocia a un valor numérico real con determinadas probabilidades. Cuando los valores que puede tomar esta variable son infinitos dentro de un intervalo, se dice que su espacio muestral es continuo. En otras palabras, una variable aleatoria X es continua cuando puede tomar los valores de una escala continua, es decir, sus posibilidades son infinitas. Por otro lado, cuando un espacio muestral tiene un número finito de posibilidades, es decir resultados que sean contables, el espacio muestral es discreto. Una variable aleatoria es discreta cuando puede tomar una cantidad finita de valores en un intervalo [3].

Las variables principales que se analizan en este trabajo son, la potencia y la energía eléctrica absorbida, al ser éstas continuas, es necesario conocer las características de este tipo de variables y cómo están distribuidas. La probabilidad de que un valor se repita en una variable aleatoria continua es muy remota, incluso se le asigna un valor de 0, por este motivo en las distribuciones de probabilidad continuas se trabajan con intervalos y no con valores puntuales.

2.2.2 Distribuciones de probabilidad continuas

Las distribuciones de probabilidad continuas se pueden representar mediante funciones $f(x)$ que dependen de los valores de la variable aleatoria continua X . A estas se les llama funciones de densidad de probabilidad de X , y tienen una importancia muy grande a la hora de analizar datos, especialmente en aplicaciones prácticas.

El área bajo la curva de una función de densidad de probabilidad y limitada por el eje de las abscisas (*eje X*) es igual a 1, siempre y cuando se evalúe en el rango de X para el que se define $f(x)$ [3]. En la Imagen 2.5 se puede ver que la probabilidad de que X tenga valores entre a y b es igual al área bajo la curva de la función $f(x)$ y entre $x=a$ y $x=b$. La probabilidad descrita también se define mediante la Ecuación 2.5:

$$P(a < X < b) = \int_a^b f(x)dx \quad \text{Ecuación 2.5}$$

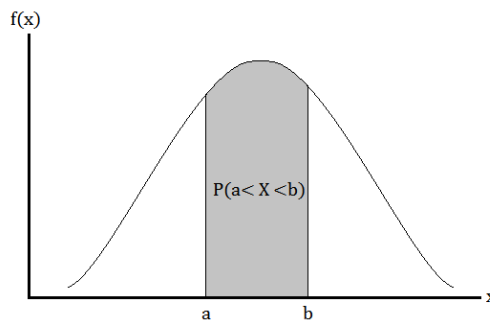


Imagen 2.5 Función de densidad de probabilidad

Una función continua de densidad de probabilidad tiene las siguientes características:

1. $f(x) \geq 0$, para toda $x \in R$.
2. $\int_{-\infty}^{\infty} f(x)dx = 1$.
3. $P(a < X < b) = \int_a^b f(x)dx$.

La función de distribución acumulativa $F(x)$ cuya función de densidad es $f(x)$, se define mediante la Ecuación 2.6 [3]:

$$F(x) = P(X \leq x) = \int_{-\infty}^x f(t)dt, \quad \text{para } -\infty < x < \infty \quad \text{Ecuación 2.6}$$

Como consecuencia de la definición anterior se puede deducir que:

$$P(a < X < b) = F(b) - F(a) \quad \text{y} \quad f(x) = \frac{dF(x)}{dx} \quad \text{Ecuación 2.7}$$

Existen varios tipos de funciones de densidad de probabilidad, cada una con sus propios parámetros y características. Por ello, es de trascendental importancia definir la función de densidad de probabilidad que mejor se ajuste a la distribución de los datos. A continuación se describen las distribuciones normal, gamma y chi cuadrada, ya que son utilizadas posteriormente.

2.2.2.1 Distribución normal

La distribución normal se considera la más importante de la estadística, ya que describe muy bien muchos fenómenos que ocurren en la realidad. La distribución de probabilidad normal se denota mediante $n(x; \mu, \sigma)$, ya que depende de μ y σ , la media y desviación estándar, respectivamente [3].

La función de distribución normal puede expresarse matemáticamente mediante la siguiente expresión:

$$n(x; \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2\sigma^2}(x-\mu)^2}, \quad \text{para } -\infty < x < \infty, \quad \text{Ecuación 2.8}$$

La curva normal o gaussiana como también se la conoce, tiene las siguientes propiedades:

- El punto más alto de la curva representa la moda y la media.
- La curva es simétrica en el eje vertical $x = \mu$.
- Los puntos de inflexión de la curva se encuentran en $x = \mu \pm \sigma$.
- La curva es asintótica al eje horizontal ya sea por la izquierda o por la derecha.
- El área bajo la curva es igual a uno.

La teoría desarrollada acerca de esta distribución resulta de mucha utilidad en esta tesis. Tal como se ve en el Capítulo 4, se utiliza la distribución normal como una distribución limitante, se aplica la inferencia estadística, test de hipótesis y detección de anomalías, en cuyos análisis se considera de manera sustentada que los datos se ajustan a una distribución normal.

Área bajo la curva normal

El área bajo la curva de cualquier distribución continua de probabilidad, acotada por $x=a$ y $x=b$ es igual a la probabilidad de que la variable aleatoria X tome un valor entre a y b . Esto se puede representar matemáticamente mediante la Ecuación 2.9:

$$P(a < X < b) = \int_a^b n(x; \mu, \sigma) dx = \frac{1}{\sigma\sqrt{2\pi}} \int_a^b e^{-\frac{1}{2\sigma^2}(x-\mu)^2} dx \quad \text{Ecuación 2.9}$$

Cuando se trabaja con experimentos reales, el valor de μ y σ es distinto para cada conjunto de datos, por lo cual, se requiere de cierto tiempo y esfuerzo para la resolución de la Ecuación 2.9. Este problema se supera al transformar los valores de la variable inicial X a una nueva variable Z con $\mu = 0$ y $\sigma = 1$. A esta transformación se le llama estandarización, la cual arroja una distribución continua llamada “normal estándar” [3]. Para ello es necesario aplicar la Ecuación 2.10:

$$Z = \frac{X - \mu}{\sigma}, \quad \text{Ecuación 2.10}$$

en donde μ es la media y σ es la desviación estándar correspondientes a la variable aleatoria X .

Para realizar el cálculo del área bajo la curva normal estándar equivalente a la Ecuación 2.9, es necesario hacer el cambio de variable correspondiente. De lo cual se obtiene:

$$P(a < X < b) = \frac{1}{\sigma\sqrt{2\pi}} \int_a^b e^{-\frac{1}{2\sigma^2}(x-\mu)^2} dx = \frac{1}{\sqrt{2\pi}} \int_{a_z}^{b_z} e^{-\frac{1}{2}z^2} dz \quad \text{Ecuación 2.11}$$

$$P(a < X < b) = \int_{a_z}^{b_z} n(z; 0, 1) dz = P(a_z < Z < b_z)$$

Como se puede notar en la Imagen 2.6, el área bajo la curva de la distribución original entre los valores $x = a$ y $x = b$ es igual al área bajo la curva normal estándar entre los valores de $z = a_z$ y $z = b_z$. La ventaja de usar la variable Z en lugar de la X , es que existen tablas y aplicaciones informáticas para calcular el área bajo la curva directamente.

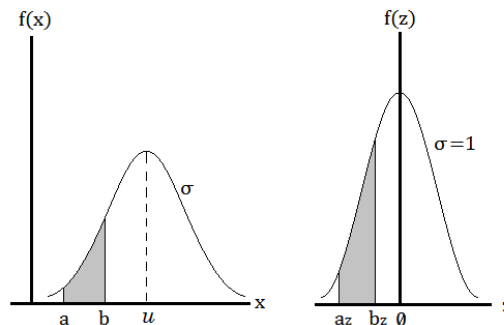


Imagen 2.6 Distribución normal y normal estandarizada

2.2.2.2 Distribución Gamma

La distribución gama tiene una importancia significativa en el mundo de la estadística, debido a su flexibilidad. Algunas distribuciones de probabilidad se derivan de esta, tal es el caso de la chi-cuadrada, de Earlang, exponencial, etc. La distribución chi-cuadrada es útil en la prueba de bondad de ajuste, la cual se detalla posteriormente.

La distribución gamma se basa en la función gama, la cual está definida por la Ecuación 2.12:

$$\Gamma(\alpha) = \int_0^{\infty} x^{\alpha-1} e^{-x} dx, \text{ para } \alpha > 0. \quad \text{Ecuación 2.12}$$

Esta función tiene algunas propiedades [3], las cuales pueden resumirse en:

- $\Gamma(n) = (n - 1)!$ para cualquier n , entero y positivo.
- $\Gamma(1) = 1$
- $\Gamma(1/2) = \sqrt{\pi}$

Una vez recordada la función gamma, es posible definir la distribución gamma:

$$f(x; \alpha, \beta) = \begin{cases} \frac{1}{\beta^\alpha \Gamma(\alpha)} x^{\alpha-1} e^{-x/\beta}, & x > 0, \\ 0, & \text{en otro caso,} \end{cases} \quad \text{Ecuación 2.13}$$

en donde x es la variable continua aleatoria que tiene una distribución gama, los parámetros α y β son mayores a cero.

La media y la varianza de la distribución gamma, están definidas mediante las siguientes ecuaciones:

$$\mu = \alpha\beta \text{ y } \sigma^2 = \alpha\beta^2. \quad \text{Ecuación 2.14}$$

Cuando $\alpha = 1$, la distribución toma el nombre de exponencial. En cambio, cuando $\alpha = v/2$ y $\beta = 2$, siendo v , un entero mayor que cero, a la distribución se le llama chi cuadrada.

2.2.2.3 Distribución Chi Cuadrada

La distribución chi cuadrada es un caso especial de la distribución gama, tiene un solo parámetro (v), al cual se le llama grados de libertad. Esta distribución tiene algunas aplicaciones interesantes, entre las cuales está la prueba de bondad de ajuste, la cual se utiliza en el análisis de la distribución de los datos de consumo de electricidad en el

Capítulo 4, cuando se aborda una nueva metodología estadística para el análisis de perfiles de electricidad.

La ecuación que define la distribución chi cuadrada es la siguiente:

$$F(x; v) = \begin{cases} \frac{1}{2^{v/2}\Gamma(v/2)} x^{v/(2-1)} e^{-x/2}, & x > 0, \\ 0, & \text{en otro caso,} \end{cases} \quad \text{Ecuación 2.15}$$

Se puede notar claramente que la distribución chi cuadrada es la distribución gamma pero con $\alpha = v/2$ y $\beta=2$. El parámetro v es un entero positivo.

La media y la varianza para esta distribución se define mediante las siguientes ecuaciones:

$$\mu = v \quad \text{Ecuación 2.16}$$

$$\sigma^2 = 2v \quad \text{Ecuación 2.17}$$

2.2.3 Distribuciones de muestreo y teorema del límite central

Los registros de consumo de electricidad que están disponibles, generalmente son el valor promedio de todas las mediciones tomadas en un intervalo de tiempo, es decir la media muestral. Usualmente ese intervalo es de 5, 10, 15, 30 minutos o 1 hora, aunque puede haber periodos mayores. Una distribución muestral de medias se considera que se ajusta bastante bien a una distribución normal. El teorema del límite central basado en esta consideración, se utiliza para establecer intervalos de confianza de la media de la población, así como también para realizar pruebas de hipótesis.

El teorema del límite central establece que una distribución muestral de la media se aproxima a una normal con media μ y varianza σ^2 . Mientras más grande sea el número de muestras mejor es el ajuste a una distribución normal. Aplicando la estandarización (Ecuación 2.10) se tiene:

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}, \quad \text{Ecuación 2.18}$$

en dónde, si n se acerca al infinito la distribución tiende a ser la normal estándar. Como se ve, X es reemplazado por \bar{X} , ya que la ecuación en este caso no se refiere a una observación sino a la media muestral. El siguiente cambio es reemplazar σ por σ/\sqrt{n} , es decir la desviación estándar es reemplazada por el error estándar de las n observaciones [3].

Cuando los datos tienen una distribución normal, el tamaño de la muestra no influye en la calidad del ajuste. Cuando la distribución de los datos no se ajusta a la normal o simplemente se desconoce, un número de muestras $n \geq 30$ garantiza que la distribución de la media muestral se ajusta a una normal [2]. El ajuste de los datos de la potencia eléctrica a la distribución normal se aborda en el siguiente capítulo y tiene sustancial importancia en el desarrollo de la metodología estadística desarrollada en el Capítulo 4.

2.2.4 Inferencia estadística

Mediante la inferencia estadística se pueden realizar generalizaciones de una población al analizar solamente una muestra. Para esto, se utilizan dos métodos, el clásico y el bayesiano. El método bayesiano se basa en el conocimiento subjetivo de parámetros desconocidos, por este motivo no se utiliza en esta tesis. La inferencia estadística mediante el método clásico se divide en dos áreas principales: la estimación y pruebas de hipótesis [3].

2.2.4.1 Estimación clásica

Un estimador puntual no es más que una variable, cuyo valor único sirve para representar un parámetro estadístico de la población en estudio. Así por ejemplo, el valor calculado \bar{x} del estadístico \bar{X} es un estimador puntual del parámetro μ de la población. Lógicamente la estimación presenta cierto error, el cual se espera sea el menor posible, para ello es necesario escoger el estimador más conveniente [3].

Un estimador insesgado es aquel cuyo estadístico $\bar{\theta}$ del parámetro θ cumple con lo siguiente:

$$\mu_{\bar{\theta}} = E(\bar{\theta}) = \theta, \quad \text{Ecuación 2.19}$$

siendo $E(\bar{\theta})$ el valor esperado de $\bar{\theta}$.

Cuando existen más de dos estimadores insesgados, a aquel que tiene menor varianza se le llama estimador más eficaz.

2.2.4.2 Estimación por intervalo

Cuando se trabaja con variables estocásticas continuas, tal es el caso de la demanda de electricidad, es difícil hacer una estimación puntual muy cercana al valor verdadero de un parámetro poblacional. A veces es preferible definir un intervalo dentro del cual se espera que se encuentre el valor a estimar, a esto se le llama estimación por intervalo. Esta problemática se trata con detalle en la predicción de la demanda de electricidad en el Capítulo 5.

Si θ es el parámetro a estimar, el intervalo es $\hat{\theta}_l < \theta < \hat{\theta}_s$, siendo $\hat{\theta}_l$ y $\hat{\theta}_s$ dependientes del estadístico y la distribución de muestreo de $\bar{\theta}$ [4]. También se puede asociar una probabilidad a que una muestra esté dentro de este intervalo, tal como se ve en la siguiente ecuación:

$$P(\hat{\theta}_l < \theta < \hat{\theta}_s) = 1 - \alpha \quad \text{Ecuación 2.20}$$

A este intervalo se le llama intervalo de confianza, α representa el valor de la integral bajo la curva de distribución de probabilidad que esta fuera del intervalo $\hat{\theta}_l < \theta < \hat{\theta}_s$. El valor $(1 - \alpha)$ es el coeficiente de confianza y los valores extremos $\hat{\theta}_l$ y $\hat{\theta}_s$ son los límites inferior y superior del intervalo, respectivamente. En estimaciones se prefiere un intervalo pequeño y un coeficiente de confianza alto [3].

2.2.4.3 Estimación de la media de una sola muestra

Si \bar{x} es la media muestral del estimador puntual de la media μ de la población y $\sigma_{\bar{x}}^2 = \sigma^2/n$, cuando el tamaño de la muestra n es grande, la estimación \bar{x} de μ probablemente es muy precisa [3]. Considerando el teorema del límite central y que \bar{X} (media muestral) tiene una distribución aproximadamente normal de media muestral $\mu_{\bar{x}} = \mu$, y desviación estándar muestral $\sigma_{\bar{x}} = \sigma/\sqrt{n}$, se puede decir que \bar{x} es un estimador de μ . A medida que el ancho del intervalo de confianza se acorta, la estimación puntual mejora.

Al estandarizar la variable X , mediante la Ecuación 2.18 y al definir un valor $z_{\alpha/2}$, de tal manera que por encima de este quede un área bajo la curva de $\alpha/2$ (ver Imagen 2.7), se puede definir la probabilidad de que un valor este dentro de un intervalo de confianza mediante:

$$P(-z_{\alpha/2} < Z < z_{\alpha/2}) = 1 - \alpha \quad \text{Ecuación 2.21}$$

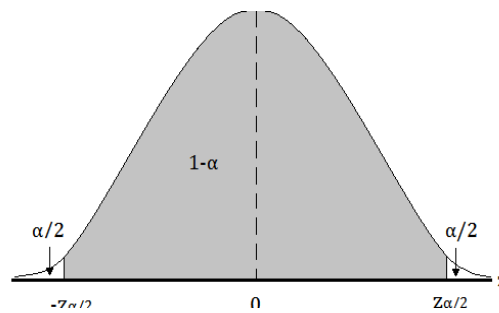


Imagen 2.7 Coeficiente de confianza en una distribución normal estandarizada

La Ecuación 2.21 se puede escribir de la siguiente manera:

$$P(-z_{\alpha/2} < \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} < z_{\alpha/2}) = 1 - \alpha \quad \text{Ecuación 2.22}$$

Al acomodar términos se tiene:

$$P(\bar{X} - z_{\alpha/2} \frac{\sigma}{\sqrt{n}} < \mu < \bar{X} + z_{\alpha/2} \frac{\sigma}{\sqrt{n}}) = 1 - \alpha \quad \text{Ecuación 2.23}$$

Se puede definir el siguiente intervalo:

$$\bar{x} - z_{\alpha/2} \frac{\sigma}{\sqrt{n}} < \mu < \bar{x} + z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \quad \text{Ecuación 2.24}$$

La Ecuación 2.24 establece el intervalo de confianza cuando se conoce σ , cuyo valor es $100(1 - \alpha)\%$, $z_{\alpha/2}$ es el valor de Z que deja un área de $\alpha/2$ a la derecha y por debajo de la curva normal. Los límites inferior y superior del intervalo de confianza quedan definidos de la siguiente manera:

$$\hat{\Theta}_l = \bar{x} - z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \quad \text{Ecuación 2.25}$$

$$\hat{\Theta}_s = \bar{x} + z_{\alpha/2} \frac{\sigma}{\sqrt{n}}. \quad \text{Ecuación 2.26}$$

Cuando se utiliza solo uno de ellos, se habla de límites unilaterales. Si se desconoce la desviación estándar real (σ), pero la muestra es lo suficientemente grande ($n > 30$) y la distribución de datos no es muy sesgada, el estadístico de la desviación estándar (s) se puede reemplazar por σ , ya que en ese caso s es similar a σ [3].

2.2.4.4 Intervalos y límites de predicción

Los intervalos de predicción sirven para estimar el valor de una observación futura. El intervalo de predicción de media desconocida μ y varianza conocida σ^2 se puede calcular mediante la siguiente ecuación:

$$\bar{x} - z_{\alpha/2} \sigma \sqrt{1 + 1/n} < x_0 < \bar{x} + z_{\alpha/2} \sigma \sqrt{1 + 1/n} \quad \text{Ecuación 2.27}$$

En dónde x_0 es el valor a predecir. Se puede apreciar que, la variación de la predicción es igual a la suma de la variación debida la estimación de la media, más la variación debido a una sola observación [3].

En el caso de que la varianza sea desconocida se puede utilizar una distribución t de Student en lugar de la normal. La distribución t tiene sentido cuando el tamaño de la muestra es pequeña y cuando se presume que los datos se ajustan a la normal. En este caso el intervalo de predicción está definido por:

$$\bar{x} - t_{\alpha/2} s \sqrt{1 + 1/n} < x_0 < \bar{x} + t_{\alpha/2} s \sqrt{1 + 1/n}, \quad \text{Ecuación 2.28}$$

En dónde $t_{\alpha/2}$ es el valor t con $n - 1$ grados de libertad que produce un área de $\alpha/2$ a la derecha y por debajo de la curva de distribución de probabilidad [3].

2.2.4.5 Detección de valores extremos usando límites de predicción

Un valor extremo es aquel que cae fuera del intervalo de predicción con probabilidad $1 - \alpha$, la observación extrema no se debe considerar para el cálculo del intervalo. Los valores extremos se suelen llamar también datos *atípicos*, *anómalos* o *defectuosos*. En el ámbito científico es muy común tratar con este tipo de observaciones, en el caso de este trabajo la identificación de los valores extremos es fundamental para identificar patrones y detectar cambios en la demanda de electricidad.

2.2.5 Prueba de Hipótesis

Una prueba de hipótesis es un procedimiento que se utiliza para obtener una conclusión relativa a poblaciones de datos. La aceptación o rechazo de una hipótesis estadística tiene trascendental importancia en la inferencia estadística.

La hipótesis como tal, es una aseveración que luego del análisis estadístico de la muestra debe ser aceptada o rechazada. Esta hipótesis debe estar planteada de tal forma que el rechazo de ella, sea la evidencia que argumente el hecho buscado. Por ejemplo, si se desea sustentar que existen mediciones anómalas de energía en el *edificio A*, la hipótesis debería expresar que: *No existen mediciones anómalas de energía en el edificio A*. A esta hipótesis se le llama *hipótesis nula*, normalmente representa el *status quo* y se le denota con H_0 ; cuando se rechaza H_0 , esto conlleva a que se acepte la *hipótesis alternativa*, denotada por H_1 . La hipótesis alternativa debería indicar lo siguiente: *Existen mediciones anómalas de energía en el edificio A*. Como se puede notar, estas dos hipótesis se contraponen y a la vez se complementan lógicamente.

Error tipo I y II

Las conclusiones que se pueden obtener a través de la prueba de hipótesis están sujetas a 2 tipos de errores. El error tipo I aparece cuando se rechaza la hipótesis nula cuando

es verdadera. El error tipo II en cambio se produce cuando no se rechaza la hipótesis nula siendo esta falsa. Estos conceptos se pueden aclarar mediante la Tabla 2.1 [3].

Tabla 2.1 Prueba de hipótesis estadística

	H ₀ es cierta	H ₀ es falsa
No rechazar H ₀	Conclusión acertada	Error tipo II
Rechazar H ₀	Error tipo I	Conclusión acertada

Al error tipo I se le llama también nivel de significancia (α). La probabilidad de cometer un error tipo I es directamente el nivel de significancia y se puede ajustar cambiando el valor crítico. Por otro lado, la probabilidad de incurrir en un error tipo II, está dada por β y se puede calcular solamente cuando la hipótesis alternativa esté bien definida. El valor de β es máximo cuando un valor de un parámetro coincide con el valor hipotético.

Siempre es conveniente que las probabilidades de cometer un error tipo I y tipo II sean muy bajas, para ello el procedimiento de prueba de hipótesis debe ser planteado con criterio y cierta habilidad; además, siempre ayuda trabajar con un tamaño de la muestra lo más grande posible. Estos errores están relacionados, ya que generalmente, cuando disminuye la probabilidad de cometer un error, aumenta la probabilidad de cometer el otro.

Ajuste a la distribución normal

La suposición de normalidad en la distribución de datos es muy común en la estadística, ya que muchos fundamentos teóricos han sido desarrollados para esta distribución. Resulta siempre atractivo suponer que los datos siguen una distribución normal. Para aceptar esta hipótesis es necesario realizar ciertas pruebas.

Prueba de Bondad de ajuste

Es muy común utilizar la *prueba de bondad de ajuste* para establecer si una población estadística se ajusta a determinada distribución de probabilidad. Esta prueba se basa en la comparación de la frecuencia de ocurrencia de los datos observados versus la frecuencia esperada de la distribución teórica a la cual se quieren ajustar.

En un experimento estadístico, a menudo a cada punto muestral, es decir a cada posible resultado se le llama *celda*. Por ejemplo, cuando se lanza una moneda existen dos celdas. La prueba de bondad de ajuste para w celdas se define de la siguiente manera:

$$\chi^2 = \sum_{i=1}^w \frac{o_i - e_i}{e_i}, \quad \text{Ecuación 2.29}$$

en dónde, χ^2 es una variable aleatoria, cuyos valores se aproximan a la distribución chi cuadrada con $w - 1$ grados de libertad. En tanto que, o_i y e_i son las frecuencias observada y esperada, respectivamente, para cada celda i [3]. La potencia eléctrica absorbida en cualquier punto de una red es una variable estocástica continua, en este caso se puede usar un histograma con $w - 1$ intervalos de clase. El procedimiento utilizado se detalla en el Capítulo 3 en donde se describe la distribución de probabilidad que tienen los datos de partida.

Si en la prueba de bondad de ajuste las frecuencias observadas y esperadas son cercanas, el ajuste es bueno y el valor de χ^2 es pequeño. Si χ^2 es lo suficientemente pequeño se acepta la hipótesis nula H_0 , lo cual indica que los datos se ajustan a la distribución de probabilidad propuesta. Para este caso la región crítica está en la cola derecha de la distribución chi cuadrada. Una vez establecido el nivel de significancia se encuentra el valor correspondiente de χ^2_α , el cual es un umbral. Si el valor resultante de χ^2 es mayor, este estará en la región crítica. En este proceso se recomienda siempre, que el valor mínimo de frecuencia esperada sea de 5 en cada intervalo de clase, si no es el caso, se deben juntar clases adyacentes [3].

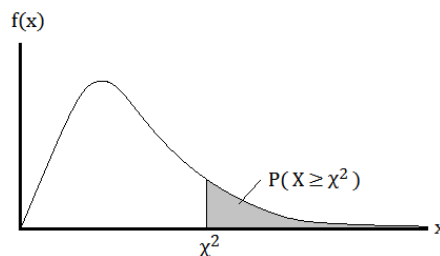


Imagen 2.8 Función de distribución de probabilidad chi cuadrada

Es importante recalcar que la prueba de chi cuadrada no es robusta si los resultados son sensibles a la normalidad de los datos, es decir que una probabilidad podría ser muy diferente a la real cuando los datos no tienen una distribución normal. Por lo tanto este procedimiento debe realizarse con cautela.

Valor p

Existe otro enfoque en las pruebas de hipótesis, el cual se refiere al uso del *valor p* o prueba de significancia. En esta prueba el rechazo o aceptación de H_0 se basa en el tamaño del valor p obtenido, el cual se define como “el valor más bajo en el que el

valor observado del estadístico de prueba es significativo” [3]. En otras palabras, el valor p se interpreta como la probabilidad de que exista un evento extremo dado que H_0 es verdadera. Por ende, si el valor p es menor o igual que un umbral establecido (α) es muy probable que H_0 sea falsa. En la mayor parte de problemas se utiliza un $\alpha = 0.05$, es decir, se asume una probabilidad del 5% de cometer un error tipo I.

En la prueba de normalidad de los datos, se define a H_0 como: *los datos siguen una distribución normal con $\alpha = 0.05$* . Es decir que, si en la prueba se obtiene un valor p mayor a α , se dice que no hay suficiente evidencia para rechazar H_0 , por lo tanto en este caso la conclusión será que los datos se ajustan a una distribución normal. Por el contrario, un valor p menor o igual que α rechaza H_0 , por lo tanto se dice que los datos no se ajustan a la distribución normal.

2.3 Predicción de electricidad

La predicción de la demanda es una herramienta muy importante para la administración de la electricidad, siendo clave en la toma de decisiones. La predicción en base a su horizonte temporal puede dividirse en tres tipos: corto, mediano y largo plazo. La predicción de electricidad a corto plazo puede ir desde una hora hasta una semana, tiene una utilidad importante en el despacho de la generación, gestión del mercado spot y en la detección de anomalías. La predicción a mediano plazo comprende un periodo de entre una semana y un año y se utiliza en la planificación, negociación de contratos y operación del sistema eléctrico. Mientras que, la predicción a largo plazo se ocupa de periodos superiores a un año y encuentra su mayor aplicación en la planificación de las redes de distribución, transmisión y nuevas centrales de generación [5]. En esta tesis se aborda la predicción a corto plazo y su enfoque se centra en el pronóstico de la demanda de electricidad de un día típico de la semana de cualquier época del año, el cual servirá de punto de referencia para el análisis de la demanda real.

El interés en la predicción y pronóstico de la demanda de electricidad ha experimentado un aumento considerable, especialmente en la última década. Pronosticar la demanda de electricidad en las redes inteligentes no es una tarea trivial, debido a que esta depende de diversos factores, tales como, el clima, factores sociales, económicos, laborales, etc.

2.3.1 Estudio de la literatura

Varias herramientas se han desarrollado y utilizado para predecir y pronosticar la demanda eléctrica. El estudio de la literatura realizado hasta marzo de 2020 indica que existen 810 documentos en la base de datos de Scopus relacionados con esta temática.

En la Imagen 2.9 se observa que el interés en este tema ha ido en aumento, en especial desde el año 2003.

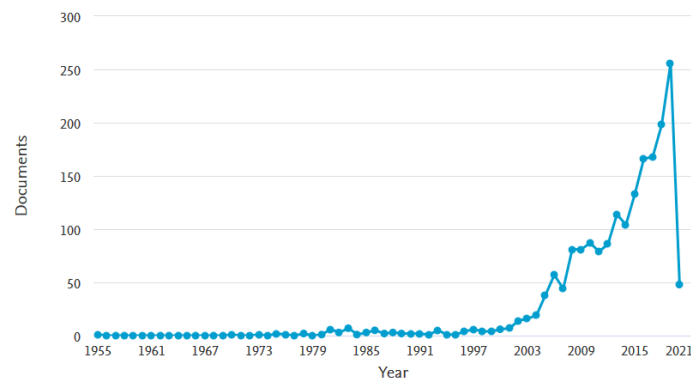


Imagen 2.9 Número de documentos en la base de datos de Scopus acerca de la predicción y pronóstico de electricidad [6]

El primer trabajo documentado acerca de predicción de electricidad data de 1955 [7] en donde el sistema predictor utiliza técnicas de interpolación y extrapolación de los datos históricos. En 1978 se mejora la predicción mediante métodos de regresión múltiple derivadas de teorías económicas [8]. En 1980 [9] se introdujo la utilización de modelos bayesianos discriminativos para mejorar la predicción de electricidad basándose en el error cuadrático medio como métrica para evaluar la calidad del método. En 1988 [10] se inicia el estudio de métodos adaptativos no estocásticos en la predicción de la demanda de electricidad con el algoritmo *Least mean square*.

En la década de los 90 se comienza a utilizar el aprendizaje automático para la predicción de electricidad. Es así que en 1996 se utilizan las redes neuronales artificiales (RNA) para predecir la demanda de electricidad [11], dando muy buenos resultados. A lo largo del tiempo la aplicación de las RNA ha evolucionado y se han ido realizando nuevas propuestas [12]–[14]. A más de las RNA otras herramientas basadas en aprendizaje automático se han utilizado para la predicción de consumos tales como las máquinas de soporte vectorial SVM (del término anglosajón support vector machines) [15] y árboles de decisión [16].

El avance de los algoritmos de las RNA, los datos masivos (big data) y la mayor capacidad de procesamiento de los ordenadores han dado lugar a lo que hoy se conoce como aprendizaje profundo (deep learning). Su utilización en la predicción y pronóstico ha sido muy reciente y ha presentado buenos resultados [17]–[19]. En este sentido, las redes neuronales recurrentes con memoria a corto y largo plazo LSTM (del término

anglosajón long short term memory) serán aplicadas en el Capítulo 5 debido al buen desempeño que evidencia la literatura.

En la Tabla 2.2 se sintetizan las herramientas más utilizadas para la predicción de la demanda de electricidad. Se distinguen las herramientas determinísticas, no determinísticas e híbridas. Las herramientas determinísticas se basan en la utilización de ecuaciones matemáticas para modelar fenómenos físicos, como por ejemplo la transferencia de calor, para luego estimar la energía que requieren los sistemas climatización y ventilación. A estas técnicas se les llama también de caja blanca o transparente ya que las relaciones entre la demanda de electricidad y las otras variables son conocidas. Las herramientas no determinísticas se pueden dividir en dos tipos, las que usan metodologías estadísticas y las que emplean el aprendizaje automático. El aprendizaje automático es una derivación de la IA que permite que las máquinas aprendan por si solas, a estas metodologías se las llama de caja negra debido a que no se conoce la relación entre el consumo de electricidad predicho y otras variables. Finalmente, están las herramientas de predicción híbridas o de caja gris, las cuales combinan las herramientas de caja blanca con las de caja negra.

Tabla 2.2 Herramientas de predicción disponibles

Herramientas de predicción	Metodologías	Ventajas	Desventajas
Determinísticas	Modelos matemáticos-físicos (ingeniería)	Modelan cualquier sistema físico o energético.	Son complejos, requieren conocimientos detallados de los sistemas físicos o energéticos. Errores elevados cuando los modelos no se ajustan a la realidad. Alto costo computacional.
Estadísticas	Modelos de regresión lineal	De simple aplicación. El modelo de predicción es descrito con una ecuación sencilla.	Se requiere datos de una o varias variables que tengan correlación importante con la variable dependiente. Resultados limitados cuando las variables independientes tienen relaciones no lineales con la variable de salida. Dificultad en el manejo de la multicolinealidad.
	Holt y Winters	Modelos de simple aplicación. Pueden modificarse para adaptarse a nuevas condiciones.	Predicción deficiente debido a la múltiple estacionalidad que tienen los datos de la demanda eléctrica.
	Box Jenkins (ARMA, ARIMA)	Admiten series de datos no estacionarias.	La eficacia disminuye cuando la serie de tiempo está dominada por la componente estacional [20]. Aplicación compleja.
	Modelos Bayesianos	De simple aplicación. Buenas predicciones cuando los datos se ajustan a una	No siempre los datos se ajustan a una distribución normal.

		distribución de probabilidad normal.	
Aprendizaje automático	Redes neuronales	Funcionan para problemas lineales o no lineales. La colinealidad entre variables de los datos de entrenamiento no representa un problema. Las anomalías en los datos no perjudican considerablemente los resultados de la predicción.	Utilizan muchos parámetros indeterminados ajustados sin reglas determinadas [21]. Para el entrenamiento requieren una completa y extensa base de datos del consumo de electricidad y sus variables relacionadas, concordantes en el tiempo. Es difícil acotar la incertidumbre de los resultados obtenidos e interpretarlos físicamente, especialmente en variables estocásticas.
	Máquinas de soporte vectorial	Es capaz de trabajar con bases de datos heterogéneas e incompletas. Se obtienen buenos resultados de predicción ya que utilizan un algoritmo de optimización.	Hay pocas pistas para seleccionar la mejor función kernel, sus parámetros correspondientes y dos constantes adicionales. Es difícil cuantificar la incertidumbre de los resultados obtenidos e interpretarlos físicamente.
	Árboles de decisión	Modelos simples de fácil interpretación. No es afectado por valores anómalos.	Predicción limitada para variables continuas. La confiabilidad de los resultados depende de la precisión en los valores de entrenamiento. Un pequeño cambio en la entrada puede causar grandes cambios en el árbol.
	Aprendizaje profundo	Alta adaptabilidad a los datos.	Gasto computacional elevado. Establecer la estructura de la red, esto es, número de neuronas, capas, algoritmo de optimización, etc, no es una tarea sencilla y puede requerir un tiempo considerable. Es difícil acotar la incertidumbre de los resultados obtenidos e interpretarlos físicamente.
	Algoritmos genéticos	Pueden resolver problemas no lineales. Alta adaptabilidad a los datos.	Gasto computacional elevado. Los resultados no siempre son óptimos ya que un adecuado ajuste del algoritmo depende del número de datos en la población, iteraciones, propiedades de los cromosomas y una correcta definición de la función <i>fitness</i> , además el tiempo de procesamiento puede ser elevado.
Híbridas	Combinan los modelos determinísticos con los de Ingeniería.	Un buen criterio en la selección de los modelos mejora las predicciones. Permiten mantener interpretaciones físicas sin la necesidad de un modelo matemático muy detallado y complejo.	Se requiere un experto para seleccionar los parámetros de los modelos no determinísticos. La implementación puede ser compleja.

Una consideración importante acerca de la predicción de la demanda de electricidad es que si bien, muchas variables pueden asociarse al consumo de electricidad, resulta poco práctico utilizarlas en un sistema de vigilancia y monitorización en tiempo real. Un sistema de predicción multi-variable eleva su requerimiento computacional y compleji-

dad y a la vez introduce mayor incertidumbre al requerir otras variables que también deben ser predichas. Se ha evidenciado que, para modelos de predicción a corto plazo, basta con un modelo univariado, ya que en estos casos variables externas tales como la temperatura tienen poca influencia [20]. Por este motivo, los aportes realizados en este documento se basan mayoritariamente en modelos univariados combinados con criterios de segmentación de datos.

La revisión de la literatura evidencia que la utilización y desarrollo de técnicas para la predicción de la demanda de electricidad, siendo el aprendizaje automático una de las metodologías preferidas por los investigadores en la actualidad. Sin embargo, en este tipo de metodologías resulta complicado acotar la incertidumbre y por otro lado, los resultados se ajustan para una instalación en particular, por lo tanto el mismo modelo podría no ser útil para otra instalación. En respuesta a estas debilidades identificadas, en el Capítulo 5 se presenta una metodología estadística que se puede utilizar para cualquier tipo de consumidor y que logra acotar la incertidumbre de la predicción al obtener un intervalo de predicción asociado a una probabilidad de ocurrencia. Para validar la funcionalidad del método se lo compara con una de las técnicas que mejores resultados han tenido recientemente, como son las RNA LSTM. En los siguientes apartados se detalla la teoría básica de las redes neuronales utilizadas en esta tesis para la predicción de la demanda de energía eléctrica.

2.3.2 Redes neuronales artificiales

2.3.2.1 Introducción

La IA es un campo de la ciencia y la ingeniería que busca entender y construir entidades o máquinas inteligentes. Se distinguen varios tipos de inteligencia artificial, los cuales se basan principalmente en los procesos de pensamiento, razonamiento y comportamiento. En base a estas consideraciones se pueden distinguir cuatro tipos de IA [22]:

- Sistemas que piensan como humanos
- Sistemas que actúan como humanos
- Sistemas que piensan racionalmente
- Sistemas que actúan racionalmente

Las RNA forman parte de los sistemas que piensan como humanos. Son modelos computacionales inspirados en el sistema nervioso de los seres vivos, los cuales se basan en la simulación del proceso de aprendizaje que tiene el cerebro [23]. Las RNA tienen la capacidad de adquirir información a través de unidades de procesamiento (neuronas) interconectadas entre sí [24]. El modelo más simple de neurona fue pro-

puesto por McCulloch and Pitts en 1943. Incluso hoy en día es el modelo más usado para describir la arquitectura de las RNA [25].

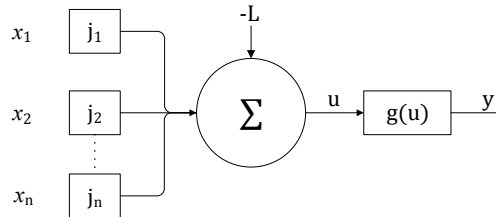


Imagen 2.10 Modelo básico de una neurona artificial

La Imagen 2.10 muestra la representación del mencionado modelo, en donde n señales de entrada se representan mediante un vector $[x_1, x_2, \dots, x_n]$. Cada entrada tiene un peso específico el cual es representado por el vector $[j_1, j_2, \dots, j_n]$. Mediante un agregador lineal se suma cada entrada (x_i) multiplicada por su peso correspondiente (j_i). A continuación se obtiene el potencial de activación (u) al hacer la diferencia entre la suma obtenida y un valor límite de activación (L). Finalmente, una función de activación (g) ajusta el valor de u , para obtener una salida (y) en un rango deseado de valores. La Ecuación 2.30 y Ecuación 2.31 resumen matemáticamente el modelo de propuesto por McCulloch and Pitts.

$$u = \sum_{i=1}^n j_i \cdot x_i, \quad \text{Ecuación 2.30}$$

$$y = g(u). \quad \text{Ecuación 2.31}$$

Las funciones de activación son de diferentes tipos, la más simple es la identidad en donde $g(u) = u$, sin embargo la utilidad de este tipo de función es limitada debido a problemas de convergencia ya que teóricamente el rango de u puede ser ilimitado [26].

Las funciones de activación más comunes limitan los valores de salida de la red artificial. Por ejemplo, la función escalón computa la salida mediante la Ecuación 2.32:

$$y = \begin{cases} 0 & \text{si } u < \theta \\ 1 & \text{si } u \geq \theta, \end{cases} \quad \text{Ecuación 2.32}$$

en donde θ es un valor umbral arbitrario.

Por otro lado, la función sigmoide suaviza a la tipo escalón con el objetivo de evitar las discontinuidades, tal como se muestra en la Ecuación 2.33:

$$y = \frac{1}{1 + e^{\frac{-u-\theta}{\tau}}}, \quad \text{Ecuación 2.33}$$

en donde τ es un parámetro que modifica la pendiente del suavizamiento.

Existen otras funciones de activación, sin embargo, en esta tesis se describen solamente aquellas que se han utilizado en las aplicaciones desarrolladas y presentadas en el Capítulo 5.

2.3.2.2 *Redes neuronales recurrentes con memoria a corto y largo plazo (LSTM)*

Las RNA LSTM son un tipo de redes neuronales con una arquitectura recurrente [27]. Esto significa que las salidas de las neuronas se realimentan hacia los ingresos de otras. Esta característica permite que estas redes sean utilizadas para procesar datos en sistemas variantes en el tiempo, tales como las predicciones de series temporales [24]. Además, están siendo utilizadas extensivamente en el campo del aprendizaje profundo (deep learning).

Las RNA recurrentes (RNAR) tradicionales operan con valores fijos de retrasos para procesar secuencias temporales de datos. Este aspecto hace que las RNAR no sean flexibles para predecir la demanda de energía eléctrica, debido a que el periodo en el que las observaciones están disponibles puede ser variable. La utilización de una red LSTM tiene ventajas significativas sobre las RNAR. Una de ellas se refiere a la factibilidad de realizar el entrenamiento con datos correspondientes a periodos largos y variables [28]. En las RNAR basadas en gradientes (retropropagación a través del tiempo o aprendizaje recurrente en tiempo real) la magnitud del error de retropropagación depende exponencialmente de la magnitud de los pesos, por lo que podría desvanecerse o incrementarse con facilidad. En este sentido, las redes LSTM solventan este problema al introducir un carrusel de error constante, a cuya activación se le llama estado de celda S_C (del término anglosajón state cell) [29].

La unidad básica en la capa oculta de una red LSTM es el bloque de memoria (con una o más celdas de memoria) y dos compuertas multiplicativas, una de entrada y una de salida. Las compuertas ajustan la magnitud de los datos desde y hacia la unidad LSTM. Cada celda de memoria realiza el seguimiento de las relaciones entre los datos de entrada y de salida, además consta del carrusel de error constante que mantiene el error de retropropagación en valores bajos [27]. La Imagen 2.11 muestra la estructura básica de una red LSTM con una sola celda de memoria. El S_C se actualiza en función de tres fuentes de entrada: red_c que son las entradas a la celda, red_{en} y red_{sal} que son las entra-

das de las compuertas de entrada y salida. Estas entradas son afectadas por los pesos j^c , j^{en} y j^{sal} . Las salidas de activación de la compuerta de entrada y de salida se representan mediante y^{en} y y^{sal} , respectivamente. Finalmente, la función g ajusta la entrada a la celda y la función h ajusta la salida de la celda [29].

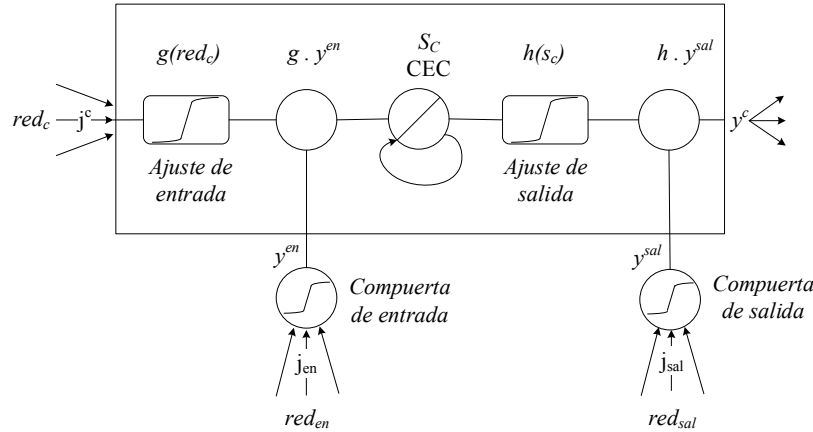


Imagen 2.11 Estructura básica de una red LSTM [29]

2.3.3 Métricas para evaluar predicciones

Una vez realizada una predicción es menester evaluarla, ya que no existe la predicción perfecta. La evaluación consiste en comparar la predicción realizada con los valores reales de las observaciones. La diferencia entre los valores estimados y los reales se puede cuantificar con un parámetro, al cual se le denomina error [30]. En la predicción de la demanda de electricidad se utilizan algunos tipos de errores de entre los que destacan los siguientes:

2.3.3.1 Error cuadrático medio

El error cuadrático medio (ECM) se define como:

$$ECM = \frac{1}{n} \sum_{i=1}^n (x_{o_i} - x_i)^2, \quad \text{Ecuación 2.34}$$

En donde n es el número de valores predichos, x_o es el valor de la predicción y x es el valor real de la observación.

2.3.3.2 Raíz cuadrada del error cuadrático medio

La raíz cuadrada del error cuadrático medio (RECM) se define mediante:

$$RECM = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_{o_i} - x_i)^2} \quad \text{Ecuación 2.35}$$

2.3.3.3 Porcentaje del error medio absoluto

El porcentaje del error medio absoluto (PEMA) es el error más utilizado para evaluar predicciones. Al ser un error relativo permite comparar resultados sin importar la magnitud de los valores:

$$PEMA = \frac{1}{n} \sum_{i=1}^n \frac{|x_{o_i} - x_i|}{x_i} \times 100 \% \quad \text{Ecuación 2.36}$$

2.4 Patrones de consumo

La demanda de electricidad está influenciada por diversas variables, tales como la hora, comportamientos sociales, el clima, la laboralidad, etc. Pese a la aleatoriedad de ciertas variables de las que depende esa demanda, se pueden identificar comportamientos recurrentes, a los cuales en esta tesis se los ha denominado patrones estocásticos debido a la incertidumbre que presentan. Conocer en detalle estos patrones de la demanda de electricidad en un determinado punto de una instalación o de la red eléctrica permite caracterizarla. La caracterización de la demanda tiene muchas ventajas, por ejemplo, posibilita realizar buenas predicciones, se pueden identificar errores en la medición, detectar anomalías, hurtos o encontrar oportunidades de mejora en la eficiencia energética.

Un enfoque interesante desde el punto de vista de la gestión y administración de electricidad es la detección de cambios del comportamiento de la demanda, los cuales pueden ser evidenciados en el PCE diario. En el Capítulo 4 se detalla una metodología que permite caracterizar la demanda, valorar e identificar cambios en los PCE en un determinado punto de una red eléctrica o consumidor. Esta metodología resulta en una herramienta interesante en la gestión y planificación de los sistemas de suministro eléctrico (SSE), ya que, a partir de sus resultados se pueden establecer estrategias de ahorro económico y energético. Estas estrategias pueden ser aplicadas en toda la cadena de suministro, es decir, en la generación, transmisión, distribución y en los propios consumidores, tal como se muestra en el Capítulo 5.

2.4.1 Estudio de la literatura

El estudio de la literatura realizado hasta julio de 2018 indica que existen 260 documentos en la base de datos de SCOPUS cuyo tema central son los patrones de consumo de electricidad, potencia o energía. Como se ve en la Imagen 2.12, el interés en su estudio ha ido en aumento, en especial desde el año 2005.

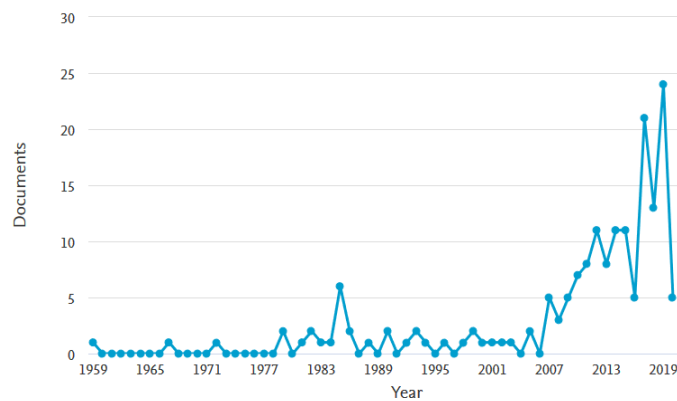


Imagen 2.12 Documentos en la base de datos Scopus de patrones de consumo de electricidad [6]

La primera publicación acerca de patrones de consumo data de 1959, realizada por E.W. Miller y publicada en Journal of Geography [31]; el autor aborda el rápido incremento de la demanda energética mundial debido al progreso de los estándares de vida. Es uno de los primeros artículos que expone la problemática de los combustibles fósiles en el abastecimiento de la demanda y sus reservas limitadas, se menciona la necesidad del desarrollo de nuevas tecnologías para el aprovechamiento de otras fuentes de energía, tales como la energía nuclear, geotérmica, solar, eólica, marina, etc.

Hasta 1982 las publicaciones sobre los patrones de consumo de energía se restringían a análisis y proyecciones de la demanda energética de diferentes países a nivel macro basados en factores económicos, sociales, ambientales. También se han analizado ciertos comportamientos a nivel rural o urbano pero de manera general. Sin embargo, en este año se publica un trabajo acerca de los patrones de consumo de energía en la industria del aluminio en India [32], es decir se inicia el estudio del comportamiento de la demanda de energía por usos finales. Es así que, desde 1985 se estudian en detalle los patrones de consumo de energía en diferentes sectores; tales como la industria textil [33], del papel, alimentaria [34], el sector doméstico [35], [36], sector agropecuario [37] y sector del transporte [38]. A partir del año 1993 la eficiencia energética se enlaza a los patrones de consumo. En este año en Turquía se proponen medidas de ahorro de energía en la industria mediante el estudio cuantitativo de patrones de consumo energé-

tico [39]. Al siguiente año, usando técnicas de regresión múltiple modificadas se estudian los datos de consumo de gas y electricidad en California (Estados Unidos), con el propósito de cuantificar potenciales ahorros incrementando la disponibilidad de gas en los vecindarios [40]. En 1999 se publica un trabajo en el cual se reconocen los patrones de consumo de los artefactos domésticos, con el objeto de desagregar la demanda de electricidad de toda la instalación, este trabajo es importante ya que es uno de los primeros que utilizan PCE diarios en la identificación de patrones.

El pronóstico del consumo eléctrico basado en el reconocimiento de patrones ha surgido de manera importante en la última década. En 2009 se realizan pronósticos de la demanda de electricidad agrupando datos primarios para disminuir el efecto de la variable periódica, se simplifica un patrón complejo en varios patrones simples, este estudio es aplicable a la toma de decisiones macroeconómicas y en procesos de ingeniería [41]. En 2014 se utilizan varios algoritmos de detección de patrones para predecir picos de consumo. El reconocimiento de patrones se basa en métodos clásicos de clasificación tales como, picos consistentes, SVM, modelos de mezcla de gaussianas GMM (del término anglosajón gaussian mixture models), clasificadores jerárquicos (combinación entre SVM y GMM) [42]. Se evidencia entonces que la predicción del consumo eléctrico mediante el reconocimiento de patrones es totalmente viable, sin embargo es una temática reciente que merece seguir siendo explorada, siendo un tópico de especial interés en el desarrollo de esta tesis.

La representación de patrones de consumo de electricidad como tal, es una problemática abordada recientemente. En 2013 se modelaron patrones de consumo de electricidad con GMM. Luego, a través de un agrupamiento jerárquico se extraen los patrones típicos de consumo por cada clúster. El *medoid* del clúster es el patrón representado [43]. Al siguiente año [44], definió a los patrones de consumo de energía térmica como una serie temporal horaria de energía demandada por cada metro cuadrado, este aporte es interesante ya que el patrón definido es función del tiempo, tal como los definidos en el Capítulo 4, con la salvedad de que en este caso no es energía térmica sino energía eléctrica.

La proliferación de big data en diferentes sectores y la aparición de nuevas aplicaciones informáticas no tradicionales con capacidad de procesar tal magnitud de datos ha dado paso al surgimiento de nuevas aplicaciones que podrían derivar en conflictos de privacidad de las personas. En el año 2012 se utilizó el reconocimiento de patrones de consumo de electricidad para identificar el comportamiento de los residentes en viviendas, en este caso los patrones de consumo son representados por curvas horarias [45]. Por otro lado, en [46] se utiliza el método de agrupamiento *fuzzy c-means* para descubrir patrones de consumo en usuarios residenciales (2017). Actualmente, estas aplicaciones

pueden ser útiles para fines de márketing, ya que es posible detectar requerimientos o hábitos de los potenciales clientes. Lamentablemente, esta información también podría ser utilizada de manera maliciosa al saber si las personas están o no en casa, por lo que las empresas eléctricas hoy en día hacen grandes esfuerzos para garantizar la seguridad de sus datos.

Por otro lado, el uso adecuado de la información permitiría a las empresas eléctricas segmentar a clientes y establecer tarifas más adecuadas para optimizar el sistema. Por ejemplo en 2017, se desarrolla un modelo de agrupamiento (clustering) borroso para la minería de patrones de consumo de electricidad en viviendas [47]. El modelo se utiliza para agrupar consumidores basados en las características de sus perfiles de carga de un mes. El estudio de la literatura indica que las técnicas de agrupamiento de datos pueden servir para obtener patrones de consumo.

Una aplicación reciente de los patrones de consumo es la detección de cambios y anomalías en el consumo. Es así que en 2007 se propone un análisis de series temporales de datos para identificar rápidamente cambios en patrones de consumo de electricidad en escuelas del Reino Unido [48], para ello se realiza una predicción con un simple modelo recurrente (el valor predicho es la media de los datos históricos) y luego una comparación del valor real mediante la técnica CUSUM (del término anglosajón cumulative sum). El método obtiene un gráfico de la CUSUM debiéndose interpretar visualmente las variaciones de los gradientes para identificar cambios en los patrones de consumo, lo cual resulta confuso y poco práctico. Un modelo más sofisticado fue propuesto en 2016, basándose en la premisa de que los PCE tienen información valiosa, Chijie et al., [49] propone un modelo para extraer características que pueden definir los patrones de consumo, luego aplica un análisis de componentes principales para asignar un plano de dos dimensiones a cada consumidor de energía, finalmente se calcula un factor de anomalías que puede indicar un consumo anómalo. En 2018 un estudio busca encontrar cambios en los patrones de consumo en el tiempo, para ello se utiliza un método de clustering jerárquico binario que agrupa perfiles de carga. Cada clúster de datos modela un patrón de consumo. Los clústeres luego se ensamblan para formar ventanas de tiempo, se analiza el cambio de un clúster al siguiente mediante un método de detección de cambios, de esta manera se identifican los cambios en los patrones de consumo [50]. Estos dos últimos métodos, son interesantes desde la perspectiva de la detección de cambios y anomalías, sin embargo no están pensados para analizar PCE diarios y tampoco tienen la capacidad de indicar las posibles causas de esas potenciales anomalías detectadas.

En la Tabla 2.3 se resumen las aplicaciones que han tenido los patrones de consumo de electricidad y el año en el cual tuvieron inicio.

Tabla 2.3 Aplicaciones de los patrones de consumo de electricidad a lo largo del tiempo

Aplicaciones	Año de inicio
Predicciones y pronósticos de la demanda de electricidad	1956
Análisis de usos finales de la energía	1982
Eficiencia energética	1993
Detección de anomalías y cambios en el consumo	2007
Predicción basada en reconocimiento de patrones propiamente dicho	2009
Identificar el comportamiento de los consumidores	2012
Representación de patrones de consumo	2013

El estudio de la literatura acerca de la utilización de patrones en el consumo de electricidad revela que tienen varias aplicaciones, las cuales se han ido desarrollando con el tiempo gracias al mejoramiento de ciertas técnicas y modelos matemáticos, estadísticos o derivados de la inteligencia artificial. Sin embargo, también se ha encontrado una brecha en la identificación de cambios en los PCE diarios, la detección de anomalías asociado a sus posibles causas. Consecuentemente es posible aportar a la gestión y vigilancia de los SSE y de los consumidores. En el Capítulo 4, se propone una novedosa metodología con la cual se pretende aportar en esta área del conocimiento.

En los siguientes apartados se presentan algunos conceptos con los que el lector debe estar familiarizado para comprender adecuadamente el desarrollo las metodologías propuestas.

2.4.2 El reconocimiento de patrones

El reconocimiento de patrones consiste en analizar objetos, extraer información, obtener propiedades y tomar decisiones sobre el conjunto de categorías de esos objetos. En el análisis de la demanda de electricidad, el objeto puede entenderse como la potencia activa media absorbida en un intervalo de tiempo, el manejo adecuado de este tipo de variables deriva también en la toma de decisiones de manera apropiada. El proceso del reconocimiento de patrones se muestra en la Imagen 2.13.

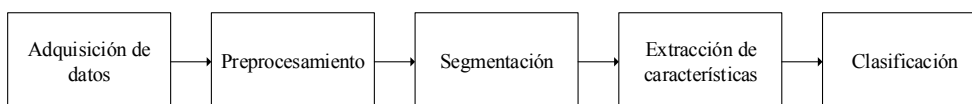


Imagen 2.13 Etapas del reconocimiento de patrones

El reconocimiento de patrones sigue las siguientes etapas [51]:

- a) *Adquisición de datos*: mediante un dispositivo sensor.
- b) *Preprocesamiento*: los datos disponibles son preparados o simplificados para su posterior uso, sin pérdida relevante de información.
- c) *Segmentación*: los datos con diferentes características son separados.
- d) *Extracción de características*: se extraen los valores de las características o propiedades (variables) más importantes, con el objetivo de reducir la información necesaria para llevar a cabo la clasificación o la representación de los objetos.
- e) *Clasificación*: Con las características extraídas, un clasificador toma una decisión final en base a la teoría de decisión (se utiliza un límite de decisión), la cual minimiza una función de costo. La clasificación utiliza procedimientos geométricos (clustering) o estadísticos, en este contexto la clasificación puede utilizar técnicas supervisadas, no supervisadas o parcialmente supervisadas.

2.4.3 Técnicas de clasificación de patrones

2.4.3.1 Técnicas supervisadas de clasificación

A las técnicas supervisadas de clasificación comúnmente se les llama clasificadores. Para el uso de estas técnicas es necesario contar con datos de entrenamiento o instancias de datos etiquetados, es decir, se requiere conocer a priori el tipo de datos existente. La clasificación puede ser multiclase cuando existen varias categorías (clases) o monoclasa cuando existe una sola categoría. Con base en los datos de entrenamiento, el modelo utilizado (clasificador) aprende a clasificar los datos, en las clases predefinidas [52]. Por ejemplo, en la demanda de electricidad, cada día de la semana pueden representar una clase distinta y cada día a su vez puede tener dos subclases, una para los días laborables (DL) y otra para los días no laborables (DNL).

Los clasificadores pueden implementarse con las siguientes técnicas:

Análisis discriminante

Los datos se dividen en grupos conocidos de acuerdo a una variable dependiente. Para esto se utilizan las funciones discriminantes, una por cada clase, las cuales dependen de varias variables independientes [53].

Vecino más cercano

Este método coloca un objeto en la clase en donde se encuentra el objeto más parecido de una muestra original [51]. Se basa en la premisa de que los datos se agrupan en

regiones de densidades altas. La técnica de vecinos más cercanos por lo general requiere de una medida de distancia o similitud entre instancias de datos, para variables continuas la métrica más usada es la distancia euclidiana, mientras que para variables categóricas es común usar el coeficiente simple de coincidencia. La técnica del vecino más cercano a su vez se divide en dos categorías [52]:

- a) Las que usan como umbral de clasificación la distancia de una instancia de datos al k vecino más cercano y,
- b) las que computan la densidad relativa de cada instancia de datos para realizar la clasificación o agrupamiento.

Redes neuronales artificiales

Una red neuronal maneja entradas y salidas deseadas. Mediante el ajuste de pesos en las neuronas del sistema se intenta disminuir la diferencia entre las salidas deseadas y las salidas del sistema (entrenamiento), de esta manera el sistema aprende. A lo largo del tiempo se han desarrollado varios tipos de redes neuronales, tales como el perceptrón simple, adaline, perceptrón multicapa, redes LVQ (del término anglosajón learning vector quantization) [54], etc.

Redes Bayesianas:

Una red bayesiana es un modelo probabilístico, la técnica básica para una variable categórica estima la probabilidad de observar una clase determinada, dado un conjunto de datos de entrenamiento. También se puede determinar la probabilidad de observar los datos, dada cierta clase. La técnica básica univariable se puede generalizar a un conjunto de datos categóricos multivariados agregando las probabilidades por atributo para cada instancia de prueba y utilizando el valor agregado para asignar una etiqueta de clase a la instancia de prueba [52].

Máquinas de soporte vectorial SVM

Las SVM de una clase se derivan de los modelos de regresión lineal y logística en las que se utiliza un margen para eliminar el sobreajuste [55]. Estas técnicas mediante kernels aprenden de una región (clase) que contienen las instancias de datos de entrenamiento, pudiendo incluso estas regiones ser complejas. Para cada prueba que realiza la técnica se determina si la instancia cae dentro o no de la clase. Una variación de esta técnica se basa en encontrar una pequeña hipersfera en el espacio del kernel que contengan las instancias de datos de entrenamiento, si los datos caen fuera de la hipersfera no pertenecen a esa clase [52].

2.4.3.2 Técnicas no supervisadas de clasificación

Estos métodos no requieren datos de entrenamiento etiquetados, hacen uso de técnicas de clustering, las cuales forman grupos con instancias de datos similares. Existen diferentes algoritmos de clustering los cuales pueden formar grupos distintos, incluso para los mismos datos [51]. En la demanda de electricidad se puede etiquetar con facilidad el día de la semana, la laboralidad, pero etiquetar una medición o conjunto de mediciones como anómalas no es una tarea sencilla, ya que estas dependen del contexto en el que se producen. En este sentido, la clasificación de PCE encaja dentro de las técnicas no supervisadas. Las técnicas de clustering fueron pensadas inicialmente como aprendizaje no supervisado, pero recientemente se han utilizado también como técnicas parcialmente supervisadas. Las técnicas no supervisadas de clasificación se pueden dividir en 3 categorías, en base a las siguientes suposiciones:

- a) Las instancias de datos típicos (no anómalos) pertenecen a un clúster, mientras que los datos anómalos no pertenecen a ninguno; de esta manera se obtienen dos grupos, una para datos típicos y otra para los anómalos [52].
- b) Las instancias de datos típicos se posicionan cerca del centroide del clúster más cercano, en tanto que los datos anómalos están más alejados de sus centroides más cercanos. Las técnicas más usadas en esta categoría son la k means y el clustering jerárquico.

b.1) K means

Esta es una técnica iterativa que divide un conjunto de datos en un número k de grupos o clústeres. El número k debe definirse inicialmente, la regla de Sturges es una herramienta que facilita su elección con la siguiente ecuación [56]:

$$k = 1 + \log_2(n), \quad \text{Ecuación 2.37}$$

en donde n representa la cantidad de datos. La técnica también requiere unos valores de centroides ($\mu_1, \mu_2, \dots, \mu_k$), cada elemento se posiciona en el clúster con el centroide más cercano, luego se vuelve a calcular el centroide correspondiente [51].

b.2) Clustering jerárquico

El clustering jerárquico se puede dividir en aglomerativo o divisivo. El clustering jerárquico aglomerativo inicia con n grupos, existiendo en cada uno de ellos una muestra. La siguiente partición tendrá $n - 1$ grupos y así en adelante. El clustering jerárquico divisivo en cambio es lo inverso, ya que se inicia con un solo grupo y con cada jerarquía se aumenta uno [57].

c) Las instancias de datos típicos pertenecen a clústeres grandes y densos, en tanto que los datos anómalos pertenecen a clústeres más dispersos y/o pequeños. De esta manera, se declaran como atípicas a instancias de datos que pertenecen a clústeres con una densidad o tamaño menor a un umbral establecido.

2.5 Detección de anomalías en el consumo de electricidad

La detección de anomalías es una herramienta poderosa para la vigilancia y administración de sistemas en la ingeniería. Por lo general en el sector eléctrico las anomalías producen interrupciones en el suministro con afectación a los usuarios, dando lugar a mantenimientos correctivos. El conocimiento detallado de los patrones estocásticos de la demanda de electricidad puede contribuir al mantenimiento preventivo en instalaciones y redes eléctricas. En este contexto, la vigilancia, administración y planificación del sector eléctrico puede beneficiarse sustancialmente de herramientas de detección de anomalías. De esta manera se propician medidas de mejora en el aprovechamiento de la infraestructura y de la energía, al identificar cambios repentinos en los PCE tales como robo de energía y pérdidas por fallos.

El desarrollo de un método automático para la detección de consumos anómalos de electricidad también puede reducir el tiempo y esfuerzo que los expertos técnicos dedican al análisis de grandes cantidades de datos, los cuales se generan en las nuevas REI. Por otro lado se evitan errores humanos. Por ejemplo, un sistema de monitorización podría eliminar automáticamente el encendido de una alarma debido a un alto consumo energético en un día caluroso de verano, ya que en esas condiciones ese consumo es normal. El método propuesto en el Capítulo 4 obtiene patrones estocásticos de consumo de electricidad y detecta anomalías, diferenciando los consumos energéticos normalmente altos o bajos, facilitando así la vigilancia de las redes eléctricas y de los consumidores.

2.5.1 Estudio de la literatura

Según la literatura consultada los patrones de consumo se han utilizado para detectar anomalías desde el año 2007. Para ello se han utilizado diferentes técnicas y herramientas tales como la predicción, técnicas de clasificación de patrones y otras más. No obstante, antes de continuar es importante recordar la base teórica que sustenta la detección de anomalías, ya que su definición, clasificaciones y técnicas empleadas no son tan intuitivas como se podría pensar.

Se considera que existe una anomalía cuando un dato es significativamente diferente del conjunto de datos restante [55]. Las anomalías se clasifican según su naturaleza en 3 categorías [58]:

a) *Anomalías puntuales*: Cuando un dato individual se considera anómalo con respecto a los demás datos [52].

b) *Anomalías contextuales o anomalías condicionales*: Se dan cuando los datos son anómalos en un contexto específico. Es decir, cuando los atributos de los datos y sus interrelaciones proporcionan un criterio adicional de detección de anomalías, ya que ciertas combinaciones usualmente no ocurren juntas [55]. Un día con un consumo elevado de energía dentro de una semana de consumo normalmente bajo, sería un ejemplo de este tipo de anomalía.

c) *Anomalías colectivas*: Las anomalías colectivas surgen cuando un grupo de datos es anómalo con respecto a todo el conjunto de datos. Los valores individuales del grupo de datos anómalos por si mismos pueden ser típicos, pero su ocurrencia colectiva representa una anomalía [52]. Por ejemplo un conjunto de mediciones con el mismo valor por lo general representa un error de medición.

La aplicación de diferentes técnicas de detección de anomalías depende del tipo de información que poseen los datos, sus características, cómo están distribuidos y si están o no etiquetados. Si los datos pueden ser asociados a distribuciones de probabilidad conocidas los resultados tendrán menor incertidumbre y por lo tanto serán más confiables.

El tratamiento que han tenido los datos de la demanda de energía eléctrica para identificar anomalías ha sido muy variado a lo largo del tiempo, sin embargo se han centrado en la detección de anomalías puntuales y colectivas, dejando de lado las contextuales. En este sentido, la metodología desarrollada en la presente tesis aborda los tres tipos de anomalías descritas anteriormente.

2.5.2 *Etiquetas de datos*

Las etiquetas sirven para diferenciar en el conjunto de datos a las anomalías y los datos considerados típicos. Esta tarea demanda gran esfuerzo, ya que por lo general se requiere de un experto humano que realice el etiquetado manualmente.

Las técnicas de detección de anomalías también se pueden clasificar en función de la disponibilidad de las etiquetas en los datos. De esta manera tenemos [52]:

a) *Técnicas de detección de anomalías supervisadas*: Para la utilización de estas técnicas es necesario contar con datos de entrenamiento etiquetados, de tal manera que sea posible diferenciar a los datos típicos de los anómalos. Estas técnicas normalmente utilizan un modelo predictivo para determinar si los datos son anómalos o no.

b) *Técnicas de detección de anomalías semisupervisadas*: En estas técnicas se asume que los datos tienen etiquetas solamente para los datos típicos, son ampliamente utilizadas ya que, generalmente etiquetar datos anómalos requiere de mayor esfuerzo.

c) *Técnicas de detección de anomalías no supervisadas*: Estas técnicas no requieren de datos etiquetados y trabajan bajo la premisa de que, los datos anómalos son mucho menos frecuentes que los típicos.

Las anomalías pueden ser detectadas con diversas técnicas, la mayoría de ellas fueron descritas en el apartado 2.4.3 cuando se abordó la clasificación de patrones. A continuación se presenta un enfoque estadístico para la identificación de anomalías, cuyos fundamentos no fueron descritos anteriormente.

2.5.3 Técnicas estadísticas de detección de anomalías

Una anomalía desde el punto de vista estadístico puede definirse como una observación sospechosa de ser parcial o totalmente anómala debido a que no fue generada por el modelo estocástico asumido. Las técnicas de detección de anomalías estadísticas se basan en la premisa de que las instancias de datos típicos se encuentran en regiones de alta probabilidad de un modelo estocástico, en tanto que, los datos anómalos se encuentran en las regiones de baja probabilidad [52]. En este contexto existen dos tipos de técnicas estadísticas para la detección de anomalías, las paramétricas y las no paramétricas.

2.5.3.1 Técnicas paramétricas

Las técnicas paramétricas asumen que las instancias de datos se ajustan adecuadamente a una distribución paramétrica, con parámetros Θ , por lo que tiene una función de densidad de probabilidad definida por $f(x, \Theta)$, en donde x es una variable aleatoria que representa a las observaciones y los parámetros Θ son los que definen la función de densidad. La puntuación (valoración) de la anomalía de una observación x , es la inversa de la función de densidad de probabilidad, es decir que una observación tendrá una puntuación de anomalía alta cuando $f(x, \Theta)$ tenga un valor pequeño. En las técnicas paramétricas, se pueden usar pruebas de hipótesis para identificar anomalías, en donde la hipótesis nula H_0 se podría definir por ejemplo como *los datos de la potencia eléctrica activa absorbida x se han generado con la distribución de parámetros Θ* . Entonces, si H_0 no se acepta luego de la prueba, x es declarada como una potencia anómala. De acuerdo al tipo de distribución asumida, las técnicas paramétricas se dividen en [52]:

a) Técnicas basadas en modelos gaussianos

En estas técnicas se considera que las instancias de datos se ajustan a una distribución normal o gaussiana. La distancia entre el dato y la media estimada de la instancia de datos es la puntuación de la anomalía, si la puntuación de la anomalía es mayor a un umbral, el dato es considerado anómalo [58].

En procesos de control de calidad es muy común utilizar el criterio 3σ , en donde σ es la desviación estándar de la distribución normal. Un dato es considerado anómalo cuando está a una distancia mayor de 3σ de la media μ . El área bajo la distribución gaussiana entre $\mu \pm 3\sigma$ contiene al 99.7% de los datos.

A finales del siglo XX e inicios del siglo XXI se extendió por algunas áreas del conocimiento una técnica simple de detección de anomalías univariable y multivariable. La técnica consiste en la utilización de la información que provee un diagrama de cajas y bigotes, en la cual se indica el valor mínimo no anómalo, el cuartil inferior Q1, la mediana, el cuartil superior Q3 y el valor máximo no anómalo (ver Imagen 2.14). Se define el rango intercuartílico (RIC) como la diferencia entre Q3 y Q1. Entonces, con este criterio una observación es considerada anómala cuando su posición es menor a $Q1 - 1.5 * RIC$ o mayor a $Q3 + 1.5 * RIC$, la región comprendida entre estos valores contienen el 99.3% de las observaciones en una distribución normal por lo que es equivalente al criterio 3σ descrito anteriormente [55].

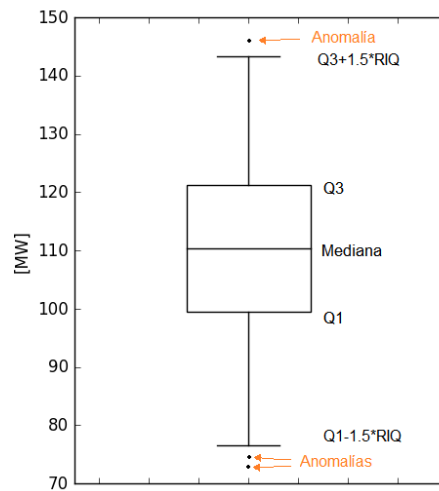


Imagen 2.14 Un diagrama de caja para datos univariados

El test de Grubb considera que los datos siguen una distribución normal. Esta técnica se utiliza para detectar anomalías en grupos de datos de una variable. Para cada valor x se calcula el valor de z , mediante [59]:

$$z = \frac{|x - \bar{x}|}{s}, \quad \text{Ecuación 2.38}$$

en donde \bar{x} y s son respectivamente, la media y la desviación estándar de la muestra de datos. Entonces, un dato es declarado anómalo cuando [59]:

$$z > \frac{N - 1}{\sqrt{N}} \sqrt{\frac{t_{\alpha/(2N), N-2}^2}{N - 2 + t_{\alpha/(2N), N-2}^2}}, \quad \text{Ecuación 2.39}$$

en donde $t_{\alpha/(2N), N-2}^2$ es el umbral que declara un dato como anómalo o típico. Este umbral es el valor tomado por una distribución t con un nivel de significancia de $\alpha/2n$ [59]. Existen otras técnicas derivadas del test del Grubb pero no serán consideradas ya que están fuera del alcance de este trabajo.

b) Técnicas basados en modelos de regresión

Los modelos de regresión han sido utilizados de forma extensa en la detección de anomalías en series temporales de datos, tal como es la demanda de electricidad. La técnica consiste en ajustar inicialmente un modelo de regresión a los datos, luego se cuantifica el residuo para el dato de prueba y se usa como un valor de puntuación de la anomalía, ya que ese residuo es parte del dato que no es explicado por el modelo de regresión [55]. En este contexto, algunas técnicas y modelos han sido utilizados, tales como, regresión robusta, modelo autoregresivo integrado de media móvil ARIMA (del término anglosajón autoregressive integrated moving average) y autoregresivo de media móvil ARMA (del término anglosajón autoregressive moving average).

c) Técnicas basadas en la mezcla de distribuciones paramétricas

En este tipo de técnicas se emplea una mezcla de distribuciones paramétricas, las cuales se pueden dividir en dos tipos. El primer grupo de técnicas modelan por separado a las instancias de datos típicos y a los datos anómalos, mientras que el segundo grupo de técnicas modelan solamente a los datos típicos [52].

El grupo de técnicas que modelan por separado los datos típicos y anómalos, asumen que los datos típicos o normales se ajustan a una distribución gaussiana con una media y varianza determinadas, mientras que los datos anómalos se ajustan a una distribución normal con la misma media pero con una varianza considerablemente mayor. Posteriormente, los datos se evalúan con la prueba de Grubb para etiquetar al dato como

anómalo o típico [59]. En estas técnicas también se han aplicado algoritmos de maximización esperada (expectation-maximization) para los modelos de dos clases, aquí se asume que cada dato tienen una probabilidad predefinida p de ser una anomalía y una probabilidad $1 - p$ de ser un dato normal [52].

El segundo grupo de técnicas modelan las instancias de datos normales considerándolos una mezcla de distribuciones paramétricas, de tal forma que un dato que no se ajuste a ninguno de los modelos aprendidos es considerado anómalo. La teoría estadística de valores extremos también puede ser catalogada dentro de este grupo de técnicas de detección de anomalías, sin embargo, los modelos gaussianos mixtos han sido los más utilizados [55]. Este tipo de técnicas es el aplicado en la detección de anomalías en la demanda de electricidad, tal como se expone posteriormente en el Capítulo 3 y Capítulo 4.

2.5.3.2 Técnicas No-paramétricas

En este tipo de técnicas, la distribución de las instancias de datos no está determinada inicialmente. Las técnicas no paramétricas más utilizadas se detallan a continuación:

a) Basadas en Histogramas

El histograma captura fácilmente la forma en la que se distribuyen los datos en un experimento estadístico, por esta razón se puede utilizar para identificar anomalías. El primer paso para detectar anomalías mediante histogramas es construir el histograma con intervalos de clase del tamaño adecuado para los datos típicos. Las anomalías se pueden detectar cuando los datos caen fuera de los intervalos de clase existentes, o cuando se ubican en intervalos de clase de baja frecuencia de ocurrencia [55].

b) Basadas en funciones Kernel

Las técnicas de detección de anomalías basadas en funciones Kernel son muy parecidas a las técnicas paramétricas descritas anteriormente, la diferencia es que se utiliza una función kernel para ajustar una instancia de datos a una función de densidad de probabilidad. Los datos que se ubican en zonas de baja densidad de probabilidad son considerados anómalos. La estimación por ventanas de parzen son un ejemplo de técnicas de ajuste de datos no paramétricos a funciones de densidad de probabilidad [55] [58].

2.5.4 Ventajas y desventajas de las técnicas estadísticas

Una vez sintetizadas las técnicas de detección de anomalías existentes vale la pena, recalcar las ventajas y desventajas de las técnicas estadísticas sobre las otras. Comprender de manera adecuada estas diferencias da valor al trabajo desarrollado en esta tesis.

2.5.4.1 *Ventajas* [52]:

- a) Las técnicas estadísticas proveen una solución justificada estadísticamente, siempre y cuando los datos se ajusten adecuadamente a una función de densidad de probabilidad. En el Capítulo 3 se recalca la importancia del análisis estadístico de los datos y de su ajuste a una distribución de probabilidad conocida.
- b) Los valores o puntuaciones dadas a las anomalías por técnicas estadísticas pueden ser asociadas al nivel de significancia, lo cual es valioso en la toma de decisiones.
- c) Cuando la instancia de datos se ajusta adecuadamente a una función de densidad de probabilidad, no es necesario etiquetar los datos. Las técnicas estadísticas pueden operar en modo no supervisado y obtener buenos resultados, tal como se evidencia en el Capítulo 5 cuando se aplica el método desarrollado a casos reales.

2.5.4.2 *Desventajas* [52]:

- a) Las técnicas estadísticas se basan en la suposición de que la instancia de datos se ajusta a una distribución específica. Esta suposición no siempre se mantiene cierta, especialmente para datos reales de alta dimensionalidad.
- b) Realizar el diseño de prueba de hipótesis, incluso con el mejor criterio no es una tarea sencilla.
- c) Si bien el uso de histogramas es una técnica simple, esta no permite determinar dependencias entre diferentes atributos. Por ejemplo, una anomalía podría tener valores de atributos individualmente muy frecuentes pero sus valores en combinación podrían ser muy raros, por lo que no se podría detectar. En la demanda de electricidad dos atributos pueden ser considerados, el primero es la potencia eléctrica activa y el segundo es la hora. En este contexto, el histograma de los valores de potencia en las HNL es muy diferente al histograma de las HL, ya que en el primero las observaciones tienen valores significativamente menores. Es importante entonces analizar estas combinaciones de atributos de manera adecuada para evitar errores en la interpretación de los datos. En el Capítulo 4 se presenta una metodología interesante para el análisis de anomalías con este tipo de peculiaridades.

2.6 Conclusiones del capítulo

En este capítulo se ha resaltado la importancia que tiene la electricidad para el desarrollo del mundo actual, por lo tanto los sistemas de gestión requeridos para el funcionamiento normal de los SSE tienen gran importancia hoy en día.

La estadística es una herramienta fundamental para respaldar hallazgos científicos y el sector eléctrico no es la excepción. La ventaja de la estadística en el análisis de la demanda de electricidad radica en la solución justificada de los resultados obtenidos, lo cual contrasta con otros métodos como son los derivados de la inteligencia artificial. Esto ha motivado al autor a realizar una metodología estadística capaz de caracterizar el consumo, detectar anomalías, valorando cambios en la demanda de electricidad e identificando sus posibles causas.

La predicción de la demanda tiene también gran importancia en la gestión de la electricidad, ya que permite la toma de decisiones a corto, mediano o largo plazo. En esta tesis, es de interés la predicción a corto plazo, por ejemplo la de un día típico de la semana, el cual sirve como punto de referencia para el análisis del consumo real. En este capítulo se sintetizaron diferentes herramientas de predicción, entre las que se destacan las estadísticas y las de aprendizaje profundo. Una importante consideración en la predicción de la demanda eléctrica es que si bien, muchas variables pueden relacionarse, resulta poco práctico utilizarlas en un sistema de monitorización en tiempo real. Un sistema de predicción multi-variable es más complejo, requiere un alto gasto computacional y además introduce mayor incertidumbre al requerir variables que también deben ser predichas. Estudios anteriores indican que para la predicción a corto plazo, basta con un modelo univariado, ya que variables externas tales como la temperatura tienen poca influencia en periodos cortos. Por este motivo, los aportes realizados por el autor se basan mayoritariamente en modelos univariados interrelacionados con criterios adicionales de segmentación como se sustenta en el siguiente capítulo.

El aprendizaje profundo logra adaptarse bastante bien a los datos para realizar predicciones, con la ventaja de que no se requiere conocer la distribución estadística de los datos y las relaciones que tienen variables externas con la variable a predecir. No obstante, se han identificado dos desventajas del aprendizaje profundo, la primera es la experticia necesaria para la adecuada puesta en marcha de este tipo de sistemas y la segunda, es la dificultad de acotar la incertidumbre de los resultados obtenidos e interpretarlos físicamente. Para intentar resolver esta última limitación, en el Capítulo 5 de esta tesis se presenta como una contribución un modelo de predicción por intervalos cuyos resultados se contrastan con los de un método de predicción de aprendizaje profundo, específicamente las RNA LSTM.

La identificación de comportamientos repetitivos en la demanda de electricidad, su caracterización y categorización constituye el reconocimiento de patrones. En este capítulo se han revisado los conceptos fundamentales del reconocimiento de patrones, así como las técnicas de clasificación más utilizadas. Se distinguen dos tipos de clasificadores, los que utilizan técnicas supervisadas y los que utilizan técnicas no supervisadas.

das. Debido al volumen de datos existentes en una red eléctrica inteligente, la utilización de técnicas supervisadas es limitada debido a que los datos deben estar etiquetados. Por lo tanto las técnicas no supervisadas resultan más adecuadas la identificación y obtención de patrones de consumo de electricidad.

Las anomalías en un conjunto de datos se clasifican en puntuales, colectivas y contextuales. La detección de las anomalías contextuales representa un desafío que en esta tesis se aborda en los próximos capítulos. Tal como se ha evidenciado, existen diferentes técnicas de detección de anomalías, de las cuales las estadísticas presentan resultados totalmente justificados siempre que las instancias de datos puedan ser ajustadas a una distribución de probabilidad conocida. Ajustar los datos a una distribución de probabilidad no siempre resulta sencillo. Por otro lado, la dependencia entre variables o atributos es difícil de cuantificar por lo que complica el análisis estadístico en la detección de anomalías. Es importante entonces analizar estas dependencias de atributos de manera adecuada para evitar errores de interpretación.

El estudio de la literatura acerca de la utilización de patrones en la demanda de electricidad revela que es posible aportar a la gestión y vigilancia de los SSE y de los consumidores. El esfuerzo de esta tesis se centra en desarrollar una metodología basada en el reconocimiento de patrones para caracterizar la demanda, valorar e identificar cambios y detectar anomalías en el consumo de electricidad de manera automática. La finalidad de la propuesta es evitar errores humanos y reducir el tiempo y esfuerzo que los expertos técnicos dedican al análisis de datos, en un escenario en el que el volumen de datos crece día a día.

En el siguiente capítulo se estudia con detalle la adquisición de los datos de la demanda de energía en las redes eléctricas inteligentes, su procesamiento, análisis exploratorio y segmentación, lo cual constituye las fases tempranas en el reconocimiento de patrones de consumo.

Capítulo 3

Sistemas de gestión de energía eléctrica y análisis de los datos de partida

3.1 Introducción

En el capítulo anterior se realizó una revisión del estado del arte acerca de diferentes técnicas que son utilizadas para mejorar la gestión de la electricidad, como son la predicción de la demanda, reconocimiento de patrones y detección de anomalías. Como herramientas transversales, la estadística y la teoría de la probabilidad también fueron estudiadas. Ninguna de estas técnicas o herramientas serían útiles sin datos como información de partida. El desarrollo de esta tesis aborda la caracterización de la demanda de electricidad mediante patrones estocásticos para la detección de anomalías, valoración de cambios y la predicción del consumo energético. Por lo tanto, los datos de potencia eléctrica absorbida, así como sus variables temporales relacionadas representan una información de partida esencial. En este sentido, una adquisición adecuada de los datos, su posterior procesamiento y análisis exploratorio son tan importantes como aplicar cualquier técnica o herramienta posterior en el reconocimiento de patrones, detección de anomalías o predicción de la demanda de electricidad.

En este capítulo, en primer lugar y de manera breve se describen los sistemas de medición inteligente disponibles en la actualidad y la evolución que han tenido, esto para entender cómo operan los sistemas de adquisición de datos, cómo procesan y organizan esos datos, para luego presentarlos como información útil.

Además, se presenta un sistema de MI implementado en la UPS Sede Cuenca en Ecuador, el cual permite la adquisición de los datos esenciales para el desarrollo de esta tesis. Con base en la experiencia adquirida, se recomienda un procedimiento para la implementación de este tipo de sistemas en medianos y grandes consumidores de electricidad.

Seguidamente, se describen las características de la potencia eléctrica activa absorbida por un consumidor en general, entendida como una serie temporal de datos. Para luego detallar la muestra estadística, las características y propiedades que poseen, su procesamiento y análisis exploratorio, identificando de esta manera la necesidad de obtener una distribución de probabilidad definida.

La ausencia de una distribución de probabilidad definida en las instancias de datos ha dado origen al planteamiento de tres procedimientos para ajustar esos datos a una distribución normal, lo cual facilita su análisis estadístico e interpretación. El primer procedimiento propuesto es la segmentación, el segundo es el tratamiento de las componentes de la serie temporal de datos y el tercero es la eliminación de anomalías. Para el tratamiento de la serie temporal se propone un nuevo método que elimina la tendencia y desestacionaliza los datos de la demanda de electricidad de cualquier tipo de consumidor, constituyendo un aporte importante en este trabajo. El análisis de la muestra de datos y su tratamiento sirven de base para el desarrollo de la metodología expuesta en el Capítulo 4, la cual se utiliza para obtener patrones estocásticos de consumo, identificar y valorar cambios y detectar anomalías en la demanda de electricidad.

3.2 Sistemas de medición de la energía eléctrica

Las nuevas tecnologías están provocando una transformación de las redes eléctricas en general y su administración es cada vez más compleja. En este contexto, una gestión adecuada de la energía eléctrica es fundamental, siendo la medición uno de sus aspectos fundamentales.

3.2.1 Evolución de la red eléctrica tradicional

Una red eléctrica tradicional tiene tres niveles principales, generación, transmisión y la distribución en donde se incluyen a los consumidores (ver Imagen 3.1). El flujo de potencia es unidireccional y va desde el generador hasta el consumidor. La electricidad recorre largas distancias, pasando por diferentes etapas de transformación de tensión, en este camino se producen pérdidas, principalmente debido al calentamiento de los conductores (efecto Joule). La gestión de este tipo de redes ha ido mejorando con el tiempo, sin embargo, todavía tienen funcionalidades limitadas, lo cual no favorece la incorporación de la generación por parte de los consumidores, además no permite monitorizar y peor aún controlar el sistema de manera integral.

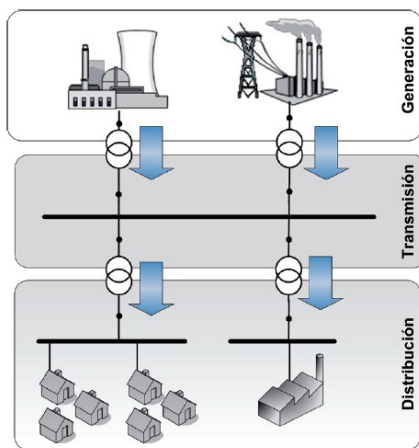


Imagen 3.1 Red eléctrica tradicional [60]

Las redes eléctricas se han expandido rápidamente en todo el planeta debido al crecimiento de la demanda. En muchos países también se ha observado un crecimiento acelerado de la generación distribuida (GD), entendida como generadores de menor potencia conectados en la red de distribución. La GD es conveniente, ya que reduce las pérdidas por efecto Joule, además, generalmente utiliza tecnologías limpias, en especial la energía solar fotovoltaica y la energía eólica, lo que ayuda a reducir las emisiones de CO₂. Sin embargo, la GD agrega mayor complejidad a las redes eléctricas, haciendo necesario un control más sofisticado para garantizar la estabilidad y confiabilidad del sistema [60]–[62].

La respuesta de la demanda es otro factor que está revolucionando las redes eléctricas. En muchos países se han implantado tarifas con discriminación horaria incluso en el sector residencial, con el objetivo de mejorar los índices de carga, manteniendo los perfiles de tensión. En este escenario, los consumidores requieren información precisa, oportuna y de ser posible en tiempo real, acerca de su consumo a lo largo del tiempo, lo que les permite mejorar su desempeño energético y económico. La carencia de este tipo de información limita el dinamismo y la evolución del mercado eléctrico y por lo tanto un mejor aprovechamiento de los recursos del sistema.

Los nuevos requerimientos técnicos y de los usuarios, hacen necesaria una transformación de las redes eléctricas tradicionales. La tecnología actual posibilita esta transformación a través de nuevos productos y servicios que integran supervisión, control, comunicación inteligente y tecnologías auto-reparadoras. Las redes eléctricas están evolucionando para ser más seguras, eficientes y confiables, ahorrar dinero, tener capacidad de almacenamiento energético, detectar fallos a mayor velocidad, restablecer el

servicio rápidamente ante perturbaciones del sistema, contaminar menos, permitir GD y un flujo bidireccional de potencia y finalmente proveer mayor protagonismo al consumidor [63]. Estos nuevos servicios en los sistemas eléctricos constituyen un nuevo concepto, el cual es llamado *redes eléctricas inteligentes* o *smart grids*.

3.2.1.1 *Redes eléctricas inteligentes*

Una REI puede definirse como un sistema inteligente de suministro de electricidad, combinada con nuevas tecnologías digitales y de información, el cual provee mayor confiabilidad, seguridad, eficiencia y otros beneficios a todos los actores del sistema [63]. La Plataforma Tecnológica Europea para las redes eléctricas en cambio, define a una REI como aquella que puede integrar inteligentemente las acciones de todos los usuarios conectados (generadores, consumidores y aquellos que hacen las dos cosas) con el objeto de entregar eficientemente suministros de electricidad sustentables, económicos y seguros [64]. En la Tabla 3.1 se detallan las diferencias más importantes entre una red eléctrica tradicional y una REI.

Tabla 3.1 Red eléctrica tradicional vs una red eléctrica inteligente [64]

Ámbito	Red eléctrica tradicional	Red Inteligente
Sensado y medición	Pocos sensores electromecánicos	Muchos sensores digitales
Comunicación	Comunicación unidireccional Baja seguridad	Comunicación bidireccional Alta seguridad
Control y optimización	Grandes centrales de generación Control centralizado limitado	Recursos de generación distribuidos Control descentralizado e integral
Rol del cliente	Pocas elecciones del cliente Ningún cuidado por la privacidad	Muchas elecciones del cliente Creciente preocupación por la privacidad

Una REI tiene diferentes áreas funcionales como se muestra en la Imagen 3.2, las cuales permiten gestionar por ejemplo la demanda, interrupciones en el servicio, la integración de energías renovables y/o vehículos eléctricos. Por otro lado facilitan la operación de la red gracias a la medición de variables eléctricas y el elevado grado de automatización que poseen, manteniendo o mejorando la calidad de la energía.

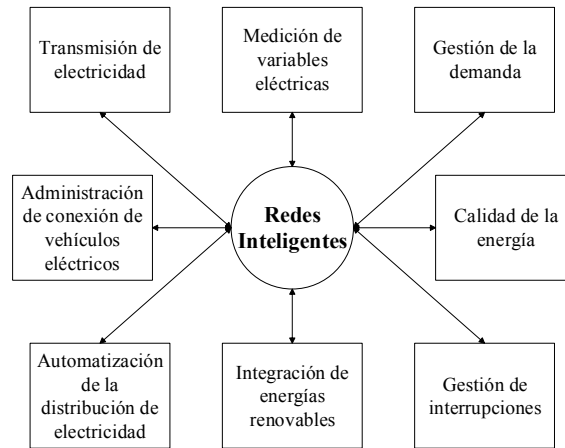


Imagen 3.2 Áreas funcionales en una red eléctrica inteligente [63]

Las nuevas funcionalidades de una REI se logran gracias al flujo de información desde y hacia las diferentes áreas de la propia red, esta información es muy sensible y por lo tanto debe manejarse en sistemas de comunicaciones robustos y seguros, ya que administran y controlan elementos tales como sensores, protecciones, controladores de carga, medidores inteligentes, etc. Los sistemas de comunicación se integran de manera transversal a los diferentes niveles de la red eléctrica, a los mercados, a la gestión de operaciones y a los proveedores de servicios del sistema, tal como se muestra en la Imagen 3.3, en donde se representa la estructura de una REI. Los sistemas de comunicación en las REI deben soportar el transporte de tres tipos de datos [64]:

- Datos provenientes del sistema SCADA (del término anglosajón supervisory control and data acquisition), por ejemplo parámetros ambientales o del sistema eléctrico de potencia.
- Datos desde las PMU (del término anglosajón phasor measurement unit), los cuales son medidas de parámetros eléctricos de la forma de onda.
- Datos de los consumidores, los cuales son básicamente la demanda de energía entregados por los medidores.

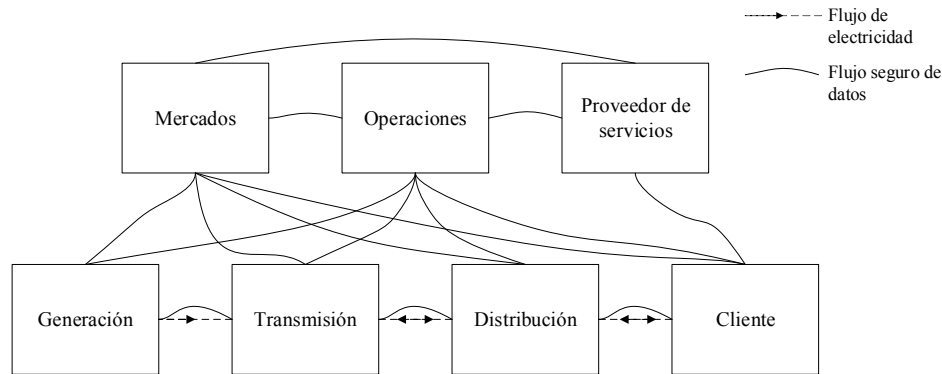


Imagen 3.3 Estructura de una REI [63]

La medición en una REI es fundamental, recientemente se han desarrollado nuevos sistemas de medición, por ejemplo AMI (del término anglosajón advanced metering infrastructure) provee una interacción bidireccional entre el consumidor y la red eléctrica [63]. En una REI los consumidores tienen la capacidad responder a estímulos del mercado para gestionar su consumo, generar o almacenar energía, permitiendo mejorar la estabilidad y aprovechar de manera eficaz los recursos existentes, a esto se le llama *respuesta de la demanda* [63].

3.2.1.2 Medición Inteligente

El sistema de medición es quizá la sección más importante de una REI, es la mayor fuente de datos del sistema y está preparada para medir, recopilar, analizar y procesar automáticamente datos de consumo energético [64]. Un sistema de MI consta de medidores inteligentes, una red de área local que comunica a una unidad de visualización del usuario, diferentes aplicaciones, nodos de comunicación y elementos de control de cargas. Esta red de área local a su vez se enlaza con una red de área extendida que permite la comunicación con un centro de control y la red de distribución (ver Imagen 3.4). De esta manera el sistema soporta funcionalidades que no estaban disponibles anteriormente.

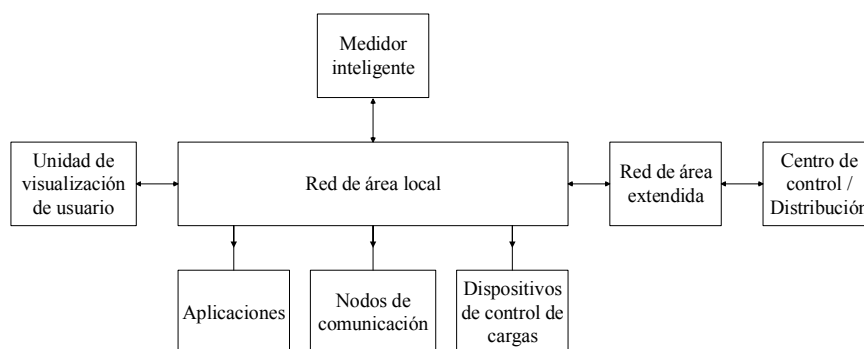


Imagen 3.4 Sistema de medición inteligente

Los medidores inteligentes son más precisos que los medidores convencionales, computan el flujo de potencia de manera bidireccional ya que el consumidor puede tener GD. Además proveen información detallada, continua y en tiempo real para determinar los usos que se le da a la electricidad, detectar fallos y caracterizar la demanda, de este modo se incentiva la gestión de la demanda, ahorro de energía y la eficiencia energética. Un consumidor por ejemplo, puede usar los datos del medidor para escoger una tarifa conveniente o decidir que artefacto o aplicación utilizar en determinadas horas para ahorrar dinero, incluso podría vender la energía almacenada en la batería de su vehículo eléctrico.

La información que proveen los medidores inteligentes no solo es útil para el consumidor, sino que entrega valiosa información a las empresas eléctricas. Se puede, conocer el patrón de consumo de cada cliente, establecer tarifas, predecir picos de demanda, mejorar la operación y administración del sistema, calcular costos y realizar anuncios o advertencias de manera remota. Conocer en detalle los patrones de demanda energética de los consumidores también genera ciertas preocupaciones a nivel de seguridad, como se mencionó anteriormente, ya que esta información podría ser empleada con fines maliciosos. Por ejemplo, el análisis de los datos medidos puede revelar si las personas están o no en casa. El uso de *minería de datos* puede sugerir comportamientos de los consumidores al descubrir patrones en el uso de electrodomésticos, pudiendo ser un insumo para los departamentos de márketing de las empresas. Entonces, un acceso, manipulación y control seguro a esta información es imperativo para proteger la privacidad de los consumidores, así como garantizar la eficaz operación del sistema [64].

3.2.1.3 Medidores de electricidad

Durante muchos años los medidores electromecánicos se utilizaron para contabilizar la demanda de energía de los usuarios. Gracias al avance de la tecnología se desarrollaron

medidores electrónicos capaces no sólo de medir energía sino también otros parámetros eléctricos. Los medidores de lectura automática AMR (del término anglosajón automatic meter reading) fueron introducidos en diferentes países entre los años 1970 y 2000, su gran ventaja es que envían información de las mediciones a las compañías eléctricas en periodos predefinidos, no obstante el flujo de datos es unidireccional. Los medidores inteligentes en cambio permiten una comunicación bidireccional, por un lado envían la información de los parámetros eléctricos medidos. Además reciben información importante, como por ejemplo, detalles del uso de la electricidad, costo, tarifas y otras notificaciones o servicios enviados desde otras áreas funcionales de la red. Si bien los medidores inteligentes ya se han utilizado por más de una década, todavía siguen evolucionando y agregando nuevas funcionalidades, mientras que diversas organizaciones, países y universidades unen esfuerzos para estandarizarlos [63].

La MI ha evolucionado en los últimos años, siempre agregando nuevos servicios. AMI es uno de los sistemas más extendidos, consta de una infraestructura de comunicación bidireccional y un sistema de administración de datos [63]. En la Imagen 3.5 se puede apreciar la evolución de la MI.

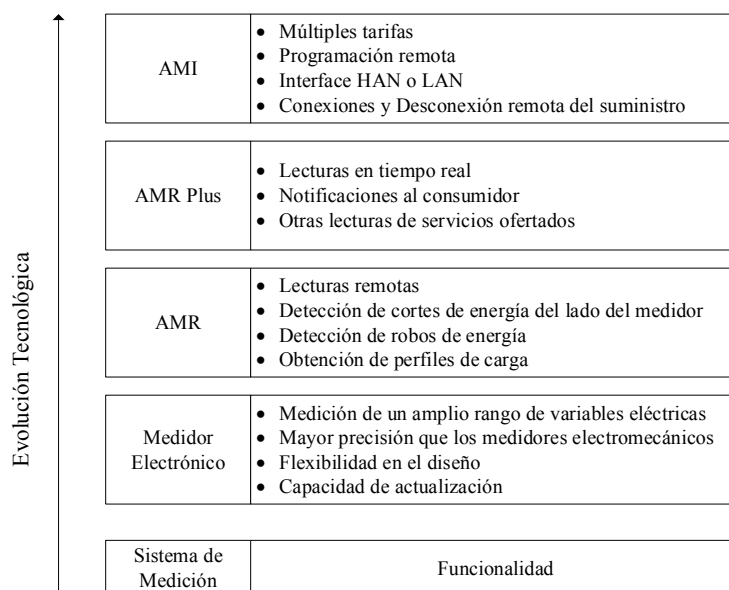


Imagen 3.5 Evolución de los sistemas de medición inteligente [63]

Un medidor inteligente se compone de los siguientes bloques funcionales: adquisición de señal, acondicionamiento, conversión analógica a digital, cálculos y comunicación, tal como se puede ver en la Imagen 3.6.

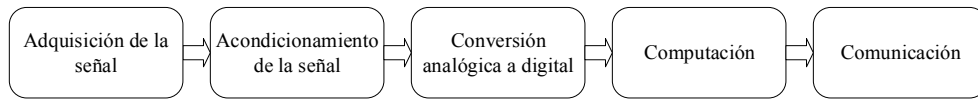


Imagen 3.6 Diagrama de bloques de un medidor inteligente [63]

La adquisición de señal se realiza mediante sensores de corriente y tensión, mientras que el acondicionamiento de la señal, conversión analógica a digital y los cálculos son realizados en un microcontrolador. Un medidor de este tipo requiere adicionalmente otros componentes, tales como una fuente de alimentación, baterías, reloj, memoria EEPROM, unidad de comunicación (Wi-Fi, GSM, GPRS) y un elemento de seccionamiento (Imagen 3.7) [63].

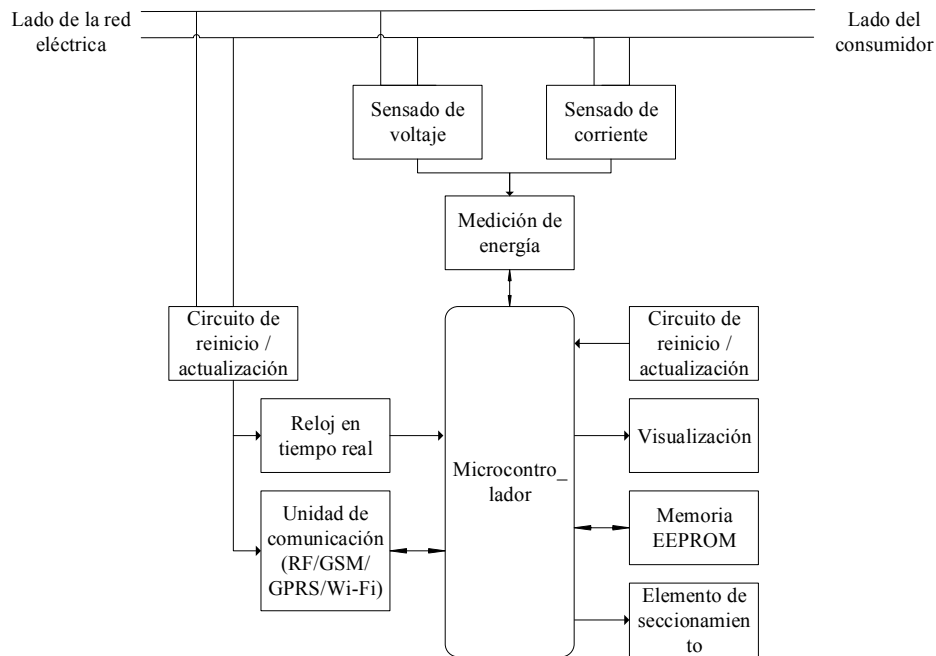


Imagen 3.7 Estructura física de un medidor inteligente [63]

Los datos de las variables provenientes de los sensores no son almacenados o enviados directamente debido a la cantidad de información que esto representaría. Estos datos son procesados en periodos predefinidos, los cuales son generalmente de: 1, 5, 10, 15, 30 minutos o 1 hora. Los periodos más utilizados son los de 15, 30 minutos y 1 hora. En cada periodo se computa la media, el valor máximo y mínimo de cada variable medida, de esta manera el requerimiento de capacidad de almacenamiento y/o transmisión de datos disminuye.

Una vez que se han revisado las características y la estructura de los sistemas de medición para las REI, en el siguiente apartado se presenta un sistema de MI implementado en la UPS Sede Cuenca [65].

3.3 Sistema de medición inteligente en la UPS

En la UPS se ha implementado un sistema de medición inteligente, cuyos datos obtenidos del consumo de energía eléctrica son esenciales en el desarrollo de este trabajo ya que proveen la información de partida para el desarrollo de la metodología propuesta en el Capítulo 4, además se utilizan en uno de los casos de estudio planteados en el Capítulo 5.

En Ecuador la implementación de medidores inteligentes ha sido escasa hasta el momento. Las empresas eléctricas todavía utilizan mayoritariamente medidores electro-mecánicos, mientras que los medidores electrónicos algunos de ellos AMR generalmente se utilizan para grandes consumidores. En este contexto, la puesta en marcha de un sistema de MI en la UPS es un interesante aporte al sector en el Ecuador, convirtiéndose en la actualidad en un laboratorio de medición y monitorización de electricidad. Este laboratorio permite el estudio de patrones de consumo, establecimiento de tarifas, predicción de la demanda, mejora en la operación y administración del sistema, evaluación de medidas de eficiencia energética, etc.

3.3.1 La demanda de electricidad en la Universidad Politécnica Salesiana Sede Cuenca

La UPS Sede Cuenca se emplaza en una extensión de 8 hectáreas, tiene 14 edificaciones y actualmente cuenta con aproximadamente 6000 estudiantes, 280 profesores y 200 empleados y personal administrativo. La ciudad de Cuenca está situada en la provincia del Azuay al sur de Ecuador. Se ubica en un valle en medio de la cordillera de los Andes, a una altura aproximada de 2500 msnm y cuenta con alrededor de 600.000 habitantes. La temperatura ambiente es poco variable a lo largo del año, con una media anual de 14.7°C; temperaturas máximas promedio de 21°C y mínimas promedio de 9°C.

La electricidad en la UPS Sede Cuenca es suministrada por la Empresa Eléctrica Regional Centro Sur (EERCS) [66] a 22 kV e internamente se distribuye mediante 5 transformadores de potencia trifásicos. En la Imagen 3.8 se puede ver la ubicación geográfica de la Universidad, mientras que la Imagen 3.9 muestra el diagrama unifilar de la red de media tensión.



Imagen 3.8 Ubicación geográfica de la UPS Sede Cuenca (© 2020 Google Maps)

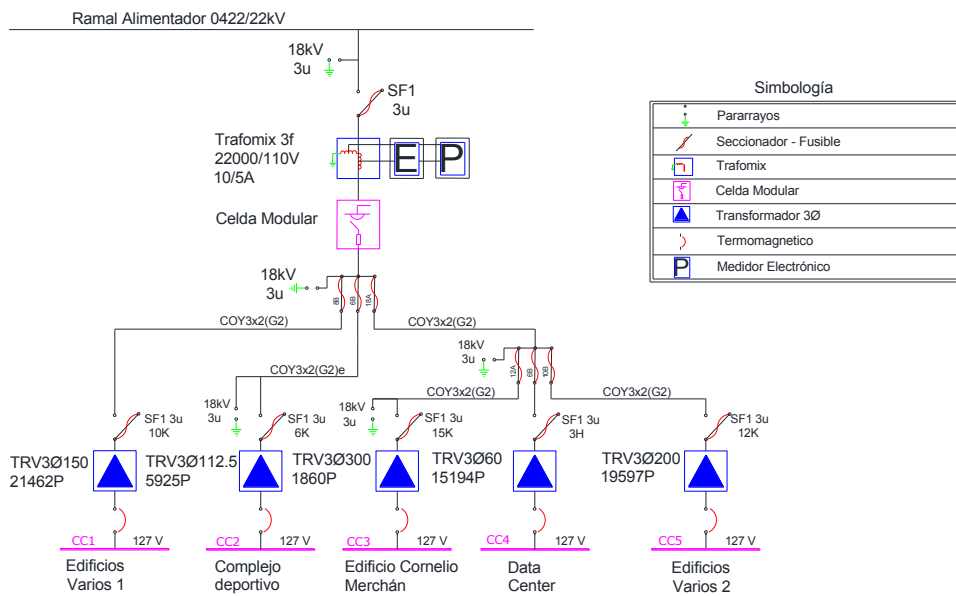


Imagen 3.9 Diagrama unifilar de la red de media tensión de la UPS Sede Cuenca

Los datos de consumo de electricidad recopilados desde el año 2013 en la UPS revelan que la demanda de energía crece por encima del 4% cada año (Imagen 3.10).

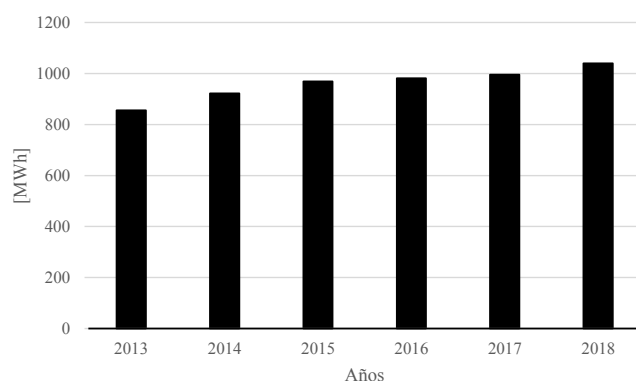


Imagen 3.10 Demanda de electricidad anual en la UPS Sede Cuenca

3.3.2 Puesta en marcha del sistema de medición inteligente

La instalación de un sistema de MI no es una tarea trivial. En base a la experiencia adquirida, se recomienda el siguiente procedimiento para la implementación de este tipo de sistemas:

- a) Levantamiento del sistema eléctrico y selección de los puntos de medida
- b) Selección del tipo de sistema de medición inteligente
- c) Medición inicial de variables eléctricas en los puntos seleccionados
- d) Selección de los medidores y transformadores de corriente
- e) Instalación y pruebas
- f) Puesta en marcha del sistema

A continuación se detalla el procedimiento llevado a cabo en la UPS Sede Cuenca.

3.3.2.1 Levantamiento del sistema eléctrico y selección de puntos de medida

En el apartado 3.3.1 se detalló el sistema eléctrico existente en la UPS Sede Cuenca. La selección de los puntos de medida se realizó en base a tres criterios: la cantidad de energía consumida, el tipo de cargas y la factibilidad de implementación de medidas de ahorro. Los puntos de medida seleccionados fueron los siguientes:

- a) *Acometida principal*: Conocer el consumo de electricidad de todo el campus universitario es fundamental. Además, es importante contrastar las mediciones realizadas por la empresa distribuidora y comercializadora de electricidad.

b) *Edificio Cornelio Merchán*: Este edificio es alimentado por un transformador de 300kVA y consume aproximadamente el 35% de toda la electricidad en la UPS.

c) *Biblioteca*: Demanda gran cantidad de energía en iluminación, sin embargo tiene muchas superficies acristaladas, por lo que las posibilidades de ahorro son considerables.

d) *Data Center*: El centro de datos concentra los servidores que dan soporte informático a la UPS a nivel nacional (Cuenca, Quito y Guayaquil). El data center esta alimentado por un transformador de 60 kVA.

e) *Edificio Guillermo Mensi*: En este edificio funcionan 21 aulas de clase y 4 auditorios, el comportamiento de este tipo de espacios es de gran interés en este estudio, ya que existen 5 edificios similares en el campus. En la Imagen 3.11 se muestran las fotografías de los lugares escogidos.



a) Acometida principal



b) Edificio Cornelio Merchán



c) Biblioteca



d) Data center



e) Edificio Guillermo Mensi

Imagen 3.11 Edificios escogidos para la instalación de los medidores inteligentes

3.3.2.2 Selección del tipo de sistema de medición inteligente

Se ha seleccionado un sistema de medición AMI debido a que es el más extendido en la actualidad, con una comunicación bidireccional segura y un sistema de administración de datos. Se ha optado por una comunicación a través de las redes de telefonía móvil, ya que es la que probablemente prevalecerá a mediano y largo plazo por sus bajos costes operativos y de acceso.

3.3.2.3 Medición inicial de variables eléctricas en los puntos seleccionados

La medición de la intensidad de corriente eléctrica en los puntos seleccionados permite seleccionar transformadores de corriente (TCs) adecuados para cada caso. Por otro lado el voltaje se mide de manera directa. La acometida principal a 22kV cuenta con un medidor de la EERCS y un transformador mixto de medida (trafomix), por lo que no es necesario instalar TCs. En los restantes puntos de medida seleccionados se tomaron mediciones de las diferentes variables eléctricas por un periodo de una semana, mediante un medidor de calidad de energía de la marca Fluke modelo 1735 [65]. Las corrientes máximas obtenidas en cada caso se detallan en la Tabla 3.2.

Tabla 3.2 Corrientes máximas medidas en los puntos de medida [65]

Punto de medida	Edificio	I ₁ [A]	I ₂ [A]	I ₃ [A]
b	Cornelio Merchán	556	447	521
c	Biblioteca	79	60	98
d	Data Center	146	123	158
e	Guillermo Mensi	129	99	147

3.3.2.4 Selección de los medidores y transformadores de corriente

Se adquirieron medidores inteligentes de la marca Discar. En la Tabla 3.3 se detallan los TCs seleccionados y el modelo del equipo de medición para cada punto de medida.

Tabla 3.3 Medidores utilizados en la implementación del sistema de medición [65]

Punto de medida	Lugar / Edificio	Marca y Modelo del medidor	TC seleccionado	Tipo de medición
a	Acometida principal	Discar DIMET3 GSM CT	Trafomix existente	Indirecta
b	Edificio Cornelio Merchán	Discar DIMET3 GSM CT	800/5	Indirecta
c	Biblioteca	Discar DIMET3 GSM CT	100/5	Indirecta
d	Data Center	Discar DIMET3 GSM CT	200/5	Indirecta
e	Edificio Guillermo Mensi	Discar DIMET3 GSM CT	150/5	Indirecta

3.3.2.5 Instalación y pruebas

El primer medidor se instaló en la acometida principal utilizando el mismo transformador mixto de medida de la EERCS. Los medidores restantes, sus transformadores de corriente y protecciones correspondientes se instalaron en los puntos indicados anteriormente. En la Imagen 3.12 se pueden observar los medidores instalados.



Imagen 3.12 Medidores instalados en la de la UPS Sede Cuenca

En la Imagen 3.13 se detalla la locación geográfica en dónde están instalados cada uno de los medidores en el campus universitario, mientras que en la Imagen 3.14 se muestra la ubicación de los medidores en el diagrama unifilar.



Imagen 3.13 Ubicación física de los medidores inteligentes en la UPS Sede Cuenca (© 2020 Google Maps)

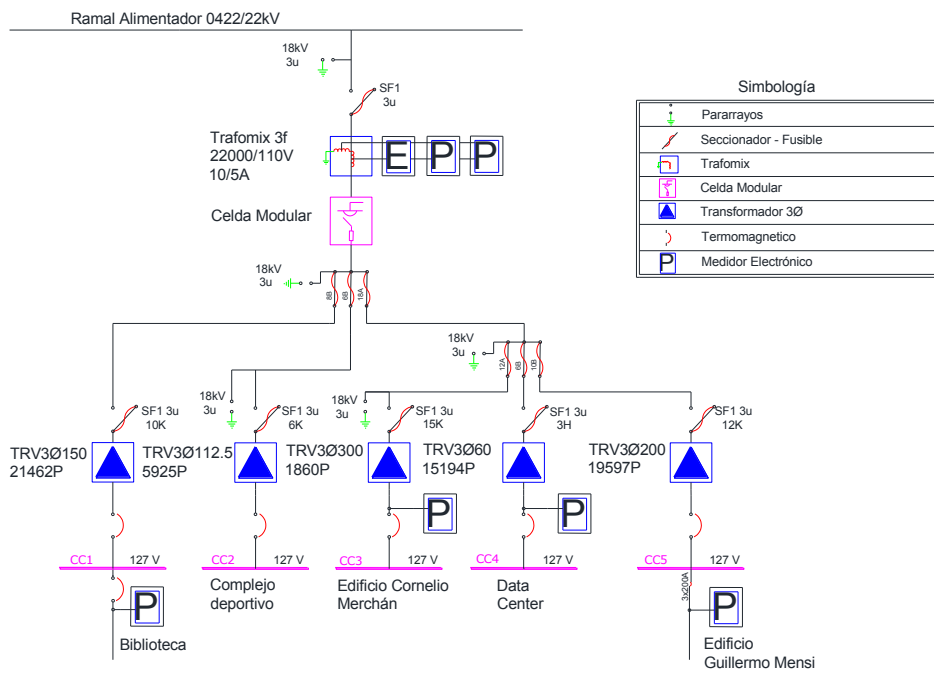


Imagen 3.14 Ubicación de los medidores instalados en el diagrama unifilar

3.3.2.6 Puesta en marcha del sistema

Los medidores instalados tienen módulos de comunicación GSM/GPRS, para ello disponen de una tarjeta SIM (del término anglosajón subscriber identity module). Una vez configurados los dispositivos y establecida la comunicación GSM/GPRS a través de la red de telefonía móvil proveedora de las tarjetas SIM. La red de comunicación establecida tiene la estructura que se muestra en la Imagen 3.15.

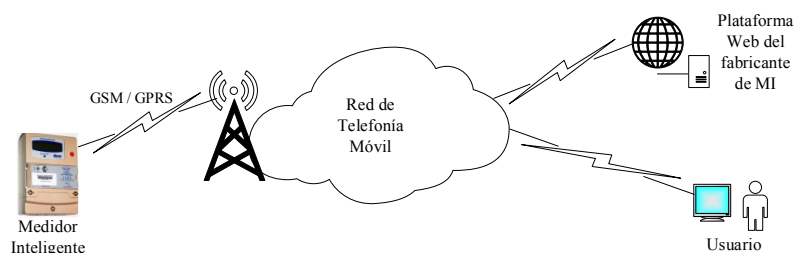


Imagen 3.15 Estructura de la red de medición inteligente de la UPS

El acceso a los datos y a la configuración de los medidores se realiza mediante una plataforma web. En la Imagen 3.16 se puede ver el entorno web que maneja dicha plataforma.



Imagen 3.16 Entorno de la plataforma web de los medidores inteligentes

Los datos están disponibles en tiempo real o en periodos definidos por el usuario, tal como se muestra en la Imagen 3.17.

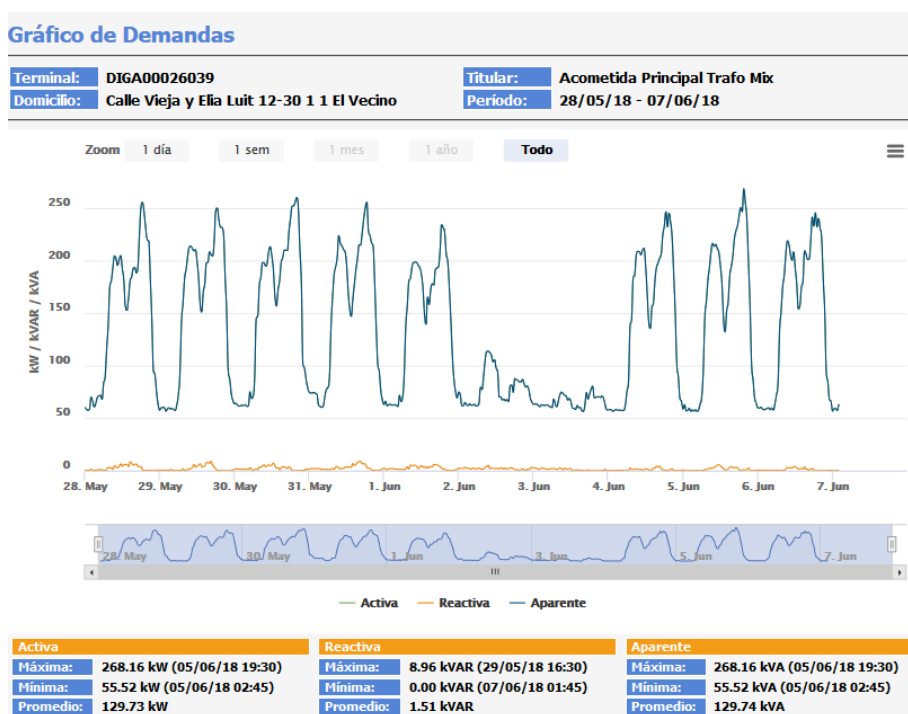


Imagen 3.17 Visualización de la potencia demandada en la plataforma web de los medidores inteligentes

La plataforma informática de los medidores provee un registro de datos con intervalos de 15 minutos. Cada dato a su vez representa la media de todas las mediciones realizadas en ese intervalo cuarto horario. Los medidores también permiten agregar tarifas diferenciadas por horario. Las variables eléctricas de las que se obtienen los registros son las siguientes:

- Voltajes (uno por cada fase) [V]
- Corrientes (una por cada fase) [A]
- Energía activa [Wh]
- Energía reactiva [VArh]
- Energía aparente [VAh]
- Potencia activa [W]
- Potencia reactiva [VAr]
- Potencia aparente [VA]
- Factor de potencia

- Demanda máxima por cada intervalo de 15min [W]

El sistema de MI instalado en la UPS provee los registros de las variables eléctricas mencionadas en cada punto de medida indicado desde el 26-03-2017 de manera continua y en tiempo real. Estos datos representan un recurso fundamental para la investigación de propuestas de mejora del desempeño energético en el campus universitario.

Las mediciones de energía realizadas por el medidor de la acometida principal se contrastan mes a mes con las realizadas por el medidor debidamente calibrado y certificado de la EERCS. De esta manera se verifican los valores facturados. Las contrastaciones realizadas indican que el error porcentual es inferior al 1%. En la Tabla 3.4 se presenta a modo de ejemplo, el contraste de mediciones para 3 meses del año 2017.

Tabla 3.4 Contratación de mediciones en la acometida principal en el año 2017

Mes - 2017	UPS [MWh]	EERCS [MWh]	Error porcentual [%]
Septiembre	72.3633	71.9320	-0.60
Octubre	88.1533	88.0250	-0.15
Noviembre	86.6540	86.5360	-0.14

3.4 La demanda de potencia activa como serie temporal

En esta tesis se trata la problemática acerca de la obtención de patrones estocásticos de consumo de electricidad y la identificación de cambios y/o anomalías en la demanda eléctrica. Para ello se utilizan los datos de la potencia activa absorbida y sus variables temporales relacionadas. Los sistemas de MI descritos en el apartado anterior representan una herramienta fundamental en la adquisición de este tipo de datos.

En las nuevas redes eléctricas inteligentes se adquieren, preprocesan y almacenan datos continuamente, sin embargo, estos por si solos son de poca utilidad. En el reconocimiento de patrones a más de las etapas de adquisición y preprocesamiento son necesarias las etapas de segmentación, extracción de características y clasificación lo cual proporciona información útil en la toma de decisiones. En este contexto, es fundamental comprender la naturaleza de estos datos y sus características como paso previo a la aplicación de cualquier técnica o herramienta de reconocimiento de patrones posterior.

La potencia eléctrica activa obtenida por un dispositivo de medición constituye un registro de datos secuenciales durante periodos definidos. Estos registros tienen características que hacen posible definirlos como una serie temporal. Una serie temporal constituye un grupo de datos registrados durante un periodo establecido y tiene 4 componentes: irregular, cíclica, estacional y la tendencia [2].

3.4.1 Variación estacional de una serie temporal

La variación estacional se refiere a la fluctuación de los valores de la serie de datos en un periodo. Por ejemplo, la potencia activa absorbida por un consumidor evidencia variaciones estacionales en periodos semanales y/o anuales. En la Imagen 3.18 se muestra la demanda media de potencia activa trimestral de algunos años en la UPS Sede Cuenca en la que se evidencia la variación estacional.

3.4.2 Promedio móvil

El promedio móvil permite suavizar una serie temporal, mostrando su tendencia claramente. El promedio móvil se aplica de manera válida cuando los valores de la serie temporal siguen una tendencia lineal y tienen fluctuaciones repetitivas en periodos definidos. Mediante este procedimiento se logra eliminar la componente cíclica y la irregular, siempre que la duración de los ciclos sea constante y los valores de amplitud en cada ciclo sean similares.

El número de elementos que se toman en cuenta para realizar un promedio móvil depende de las características de los datos. Por ejemplo, si los datos corresponden a periodos diarios podrían considerarse 7 elementos, ya que en una semana hay 7 días. Si los datos son semanales se podrían escoger 4 elementos, porque aproximadamente 4 semanas hacen un mes. Sin embargo, no hay reglas específicas, por lo que se recomienda realizar pruebas para establecer un número adecuado de elementos que suavicen las fluctuaciones. Por ejemplo, para la potencia media trimestral de la UPS se calcula el promedio móvil con 4 valores ya que 4 trimestres conforman un año, tal como se representa en la Imagen 3.18.

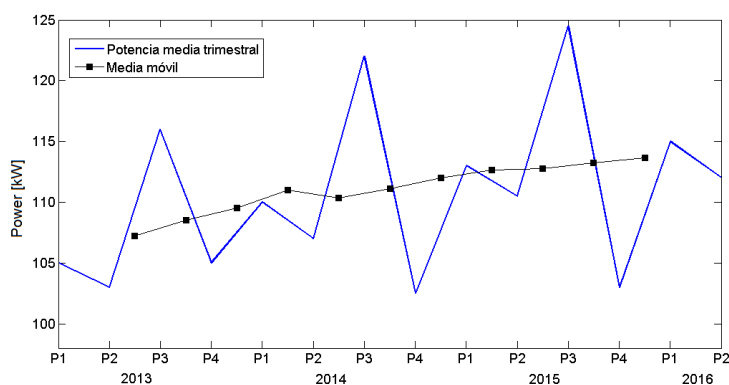


Imagen 3.18 Variación estacional en la demanda de potencia eléctrica absorbida trimestral en la UPS Sede Cuenca y su promedio móvil

3.4.3 Estacionalidad de los datos y desestacionalización

La demanda de electricidad es una serie de datos estacional que responde a un proceso estocástico. La fluctuación estacional de una serie temporal se puede medir con distintos métodos. El más usual es el conocido como *método de la razón con el promedio móvil*, en este se eliminan tendencias, ciclos e irregularidades que tienen los datos a analizar. Desestacionalizar los datos implica eliminar la variación estacional de la serie temporal, este proceso sigue los siguientes pasos [2]:

1. Definir un periodo adecuado para cada valor que se obtendrá de la serie temporal (diario, semanal, mensual, trimestral, etc). Su selección depende de las características de los datos y del tipo de análisis que se desea efectuar.
2. Definir el número de periodos para calcular el total móvil. El total móvil es la suma de los valores correspondientes cada periodo definido.
3. Determinar el total móvil de los periodos correspondientes.
4. Obtener el promedio móvil dividiendo el total móvil entre el número de periodos considerados en el punto 2.
5. Calcular el promedio móvil centrado. Esto se hace tomando los valores de los promedios móviles consecutivos y dividiéndolos entre 2. En un gráfico estos valores se representan entre las cantidades promediadas.
6. Calcular el índice de estacionalidad específico por cada periodo definido en el punto 1. Para ello es necesario dividir el valor original correspondiente a cada periodo (punto 1) entre el promedio móvil centrado obtenido en el punto 5.
7. Si se trabaja con más de un año de datos se debe calcular un índice de estacionalidad habitual. Estos índices se obtienen al realizar la media de los índices correspondientes a los mismos periodos de cada año (de cada día, semana, mes, trimestre, etc). Entonces habrá un índice específico por cada periodo en los que se ha dividido el año, este procedimiento elimina las variaciones irregulares de la serie.
8. Calcular el índice de estacionalidad ajustado. Se deben corregir los índices para que el promedio sea 1, para ello es necesario calcular un factor de corrección, mediante:

$$F_c = \frac{\text{Número de periodos que componen el índice}}{\sum \text{Índices de estacionalidad}} \quad \text{Ecuación 3.1}$$

El índice ajustado se obtiene al multiplicar el índice calculado en el punto 6 o 7 por el factor de corrección.

9. Obtener los valores desestacionalizados, al dividir los valores originales del punto 1 entre el índice ajustado calculado.

La obtención de una serie de datos desestacionalizados de electricidad en un periodo definido, permite realizar un análisis estadístico más robusto, evitando sesgar el análisis.

3.5 Análisis de la muestra de datos

En los apartados anteriores se ha descrito cómo y mediante qué equipos se realiza la adquisición de datos de la energía eléctrica. Además, se ha definido a la potencia activa absorbida como una serie temporal de datos estacionales que responde a un proceso estocástico. El detalle de la muestra de los datos, su procesamiento y análisis exploratorio se aborda en este apartado.

3.5.1 Muestra de los datos

Los PCE diarios están conformados por datos de potencia activa media en intervalos por lo general de 1 hora o cada 15 minutos. En la actualidad es muy común usar intervalos cuarto horarios, es decir que por cada día se tienen 96 observaciones.

La cantidad de días a considerar en el reconocimiento de patrones de consumo es un aspecto a tener en cuenta. El estilo de vida actual hace que los patrones de consumo de electricidad cambien constantemente debido a la evolución vertiginosa de la tecnología, por esta razón se considera conveniente trabajar con datos con una antigüedad hasta de un año. Bajo esta consideración el análisis inicia con el día de consumo más reciente, luego el anterior y así sucesivamente hasta completar el periodo seleccionado.

El primer problema en el reconocimiento de patrones en este tipo de datos se presenta al extraer sus características, ya que analizar todos los datos como una sola población estadística resulta inútil. Para ejemplificar lo expuesto se toman los datos de la acometida principal de la UPS Sede Cuenca desde el 28-05-2017 hasta el 27-05-2018. Resulta interesante ver los PCE diarios de todo un año en una sola figura. Como se ve en la Imagen 3.19 los perfiles son diversos y no se identifican características definidas.

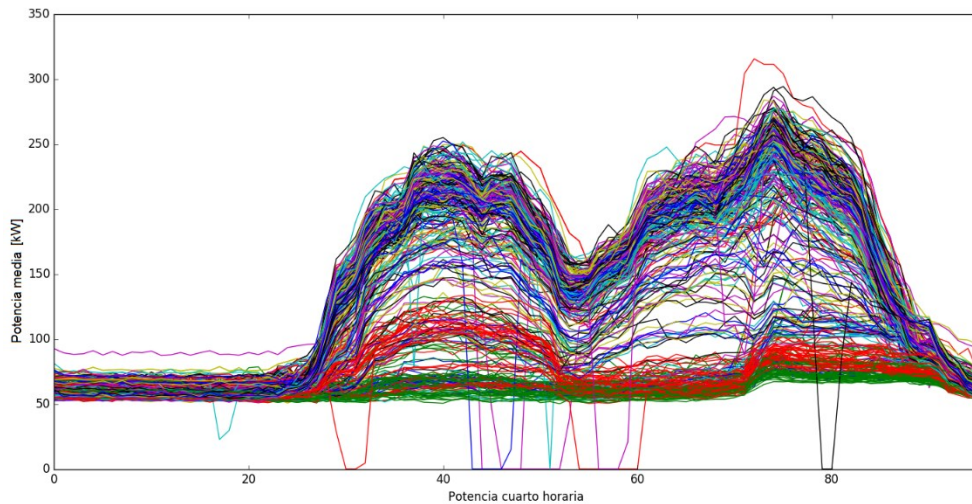


Imagen 3.19 PCE diarios en la UPS Sede Cuenca para un periodo de un año

Para verificar que tipo de distribución de probabilidad tienen las instancias de datos, los valores de potencia se han agrupado según un intervalo cuarto horario específico. Para esto se construye una matriz, en dónde cada fila representa un perfil de carga diario. Entonces, cada columna de esta matriz representa la potencia activa media absorbida cada 15 minutos para varios días. Es lógico pensar que el consumo energético en una instalación cambia según la hora del día, así por ejemplo, los valores de potencia son muy diferentes de 3:00 a 3:15 am que de 11:00 a 11:15 am. Es por esto que, el análisis de la muestra se realiza en intervalos cuarto horarios específicos. En la Imagen 3.20 se representa una matriz de datos de consumo para M días, también se ilustra cómo se agrupa la muestra datos para analizar su distribución de probabilidad.

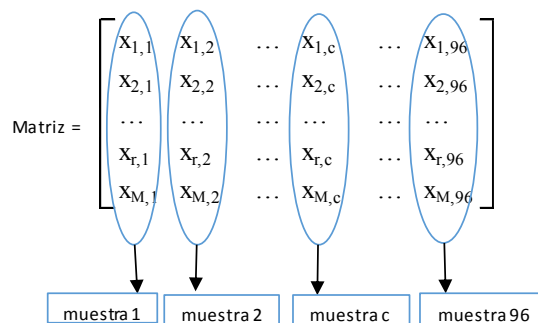


Imagen 3.20 Matriz de perfiles diarios de consumo de electricidad y toma de la muestra

Inicialmente, se analizan dos muestras en HNL y dos en HL para identificar la distribución de datos. Los intervalos seleccionados en HNL son de 2:00 a 2:15 y de 4:30 a 4:45 horas, mientras que los intervalos en HL son de 10:00 a 10:15 y de 17:30 a 17:45 horas. En la Imagen 3.21 se puede observar que las instancias de datos no tienen una distribución de probabilidad definida, esto se debe a que la demanda de electricidad depende del día de la semana, laboralidad y otras variables como por ejemplo la temperatura ambiente que influye directamente en el consumo de energía en la climatización de espacios interiores. La ausencia de una distribución de probabilidad definida en las instancias de datos dificulta su análisis e interpretación.

En este capítulo se plantean tres procedimientos para solventar la ausencia de una distribución definida en los datos, los cuales se abordan en los siguientes apartados. El primer procedimiento propuesto es la segmentación, el segundo es el tratamiento de datos de la serie temporal y el tercero es la eliminación de anomalías. El tratamiento de la serie temporal se propone en un nuevo método para eliminar la tendencia y desestacionalizar los datos de la demanda de electricidad de cualquier tipo de consumidor. A este método se le ha denominado SAEC por sus siglas en inglés (Seasonality Analysis of Electricity Consumption). Estos procedimientos han sido diseñados para facilitar el análisis estadístico de los datos, al conseguir que estos se ajusten a una distribución normal.

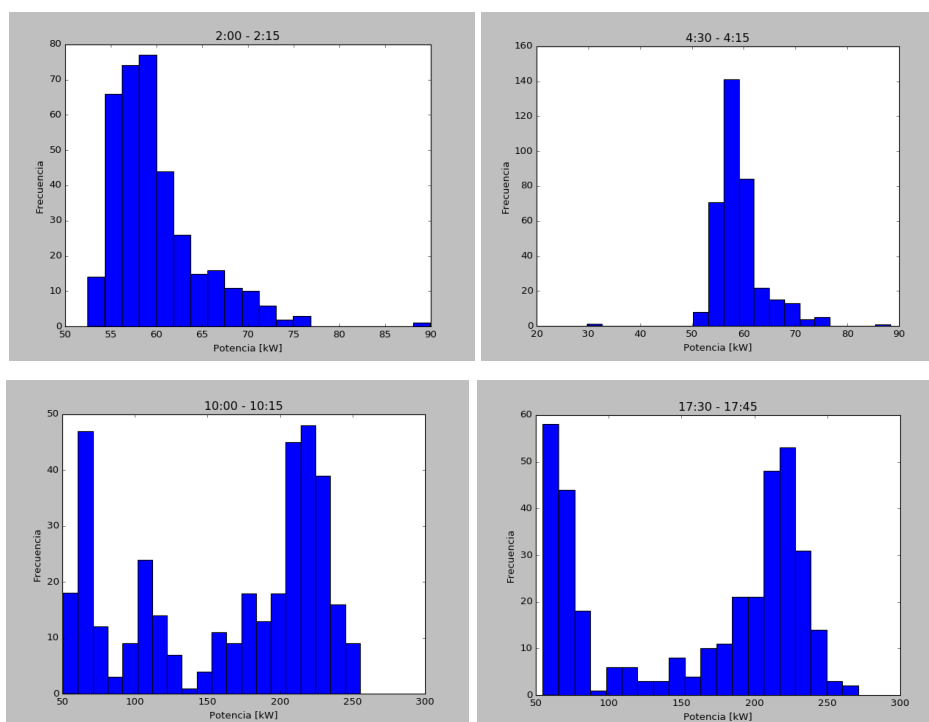


Imagen 3.21 Histogramas para diferentes intervalos cuarto horarios considerando todos los PCE de un año – UPS

3.5.2 Tratamiento de los datos de la muestra

El tratamiento de los datos de la muestra constituye una de las fases tempranas del reconocimiento de patrones, por lo que su adecuado análisis es tan importante como aplicar correctamente cualquier técnica o herramienta de clasificación posterior. A continuación se detalla cada uno de los procedimientos aplicados en esta labor.

3.5.2.1 Segmentación de datos

La segmentación de datos se realiza en función de dos criterios. El primer criterio considera variables categóricas temporales, mientras que el segundo considera una variable categórica laboral, todas ellas relacionadas a la demanda de electricidad. En la Tabla 3.5 se detallan las variables consideradas en la segmentación de las series temporales de electricidad.

Tabla 3.5 Variables categóricas y criterios considerados en la segmentación de datos

Variable categórica	Criterio temporal	Criterio laboral	Variable de segmentación	Variable informativa
Día de la semana	x		x	
Intervalo cuarto horario	x		x	
Laboralidad		x	x	
Fecha*	x			x

* La fecha es solo una variable informativa y no es utilizada en el procesamiento y tabulación de los datos

Los PCE así como las variables consideradas para la segmentación se organizan en una matriz, de tal manera que la demanda de electricidad de cada día está representado en una fila. La matriz tiene un tamaño de $M \times A$, siendo M el número de días que conforman el periodo de análisis, en tanto que A representa las 99 características del consumo de cada día (96 valores de potencia cuarto horaria y las 3 variables categóricas de segmentación indicadas en la Tabla 3.5. Posteriormente, la matriz $M \times A$ es desagregada en 14 matrices, dos por cada día de la semana. Una matriz contiene los DL y la otra matriz los DNL, tal como se muestra en la Imagen 3.22.

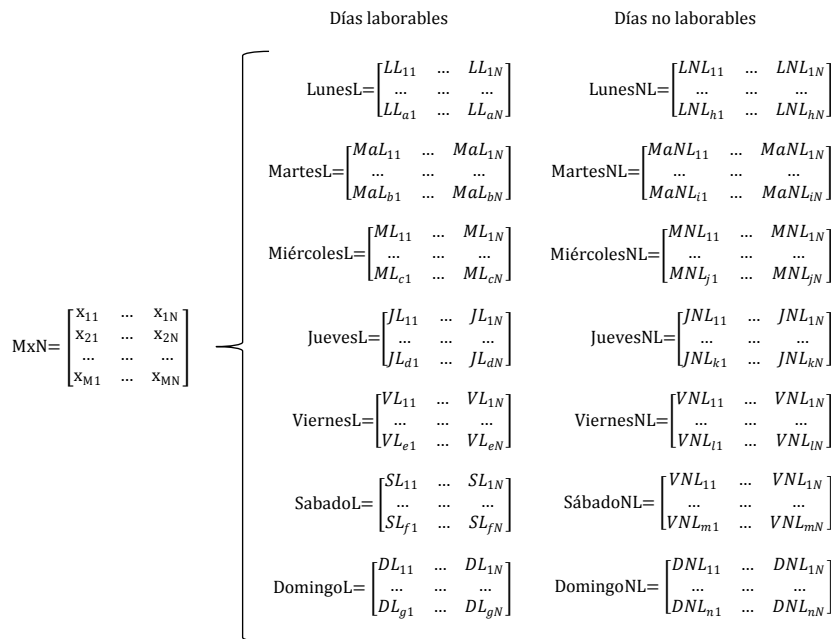


Imagen 3.22 Segmentación de los datos del consumo de electricidad

La segmentación aporta mayor uniformidad en la masa de datos de manera sencilla. En la Imagen 3.23 se muestra en histogramas la distribución de las observaciones de potencia activa eléctrica para los días miércoles laborables en diferentes intervalos cuarto horarios. Se ve claramente que los datos se agrupan más cerca de la media a diferencia de la Imagen 3.21. Luego de la segmentación los datos son más uniformes (ver Imagen 3.24), sin embargo su distribución no corresponde a una distribución de probabilidad normal, tal como lo indica su valor p (definido en el apartado 2.2.5).

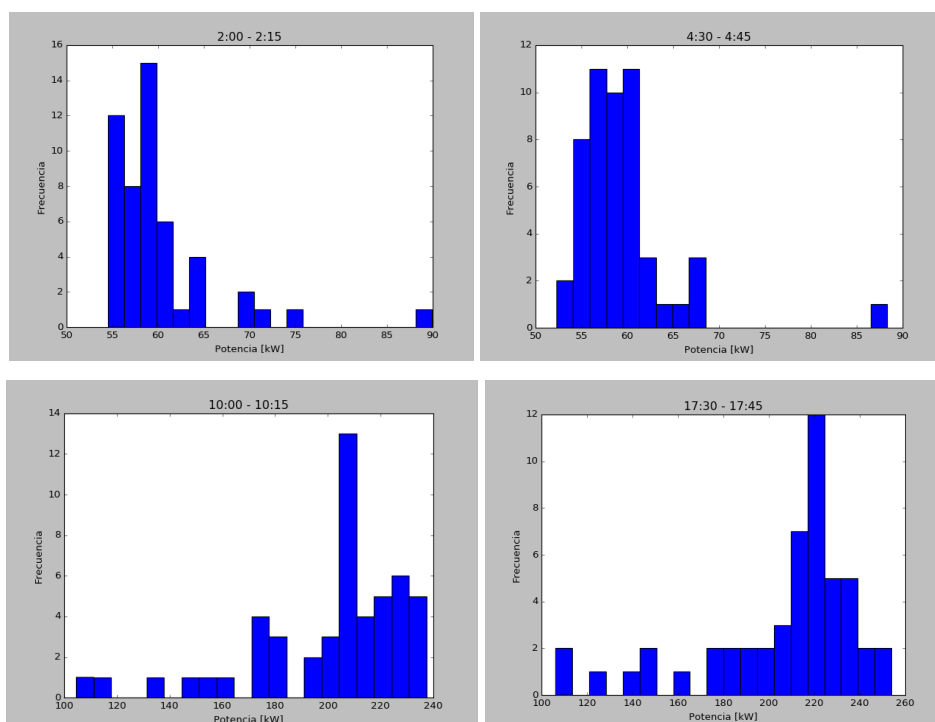


Imagen 3.23 Distribución de los datos para los días miércoles laborables en la UPS

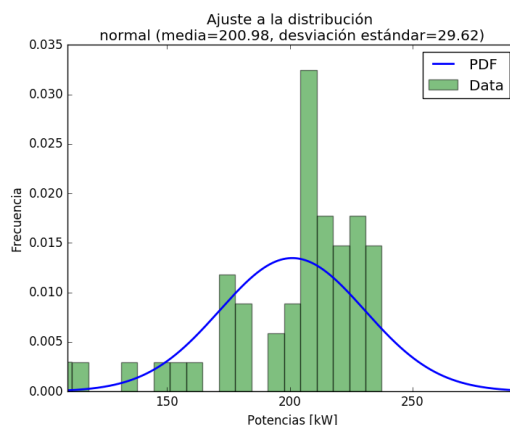


Imagen 3.24 Ajuste de los datos a una distribución normal (10:00-10:15 am) para un miércoles laborable en la UPS, valor $p = 0.00013$

3.5.2.2 Tratamiento de la serie temporal de datos con el método SAEC

La demanda de electricidad es una serie temporal con variaciones estacionales debido a cambios en los procesos de trabajo, temperatura ambiente, periodos de vacaciones, etc., tal como se ha expuesto anteriormente. En el apartado 3.4 se estudiaron las características de las series temporales de datos. Este estudio permite establecer las condiciones requeridas para llevar a cabo un tratamiento de una serie temporal, las cuales se resumen en las siguientes:

- Las fluctuaciones presentes en la serie de datos son repetitivas.
- Las fluctuaciones tienen periodos definidos.
- Los valores de la serie temporal siguen una tendencia lineal.
- Los valores de amplitud en cada ciclo tienen valores similares.

En el caso de la demanda de energía eléctrica de la UPS se evidencian fluctuaciones repetitivas en periodos de un año con una tendencia lineal positiva, tal como se observó en Imagen 3.18. La amplitud de los valores en cada ciclo es similar, obviamente afectada por la tendencia lineal. De esta manera, se observa de manera clara que las 4 condiciones para llevar a cabo el tratamiento de datos de la serie temporal se cumplen.

En este trabajo se propone un nuevo método para tratar la demanda de electricidad el cual ha sido publicado en la revista *Energies* [67]. Al método se le denomina SAEC y tiene la capacidad de eliminar satisfactoriamente la tendencia y la componente estacional de los datos. El método SAEC propuesto para el tratamiento de la serie temporal se ha resumido en 6 pasos, tal como se muestra en la Imagen 3.25.

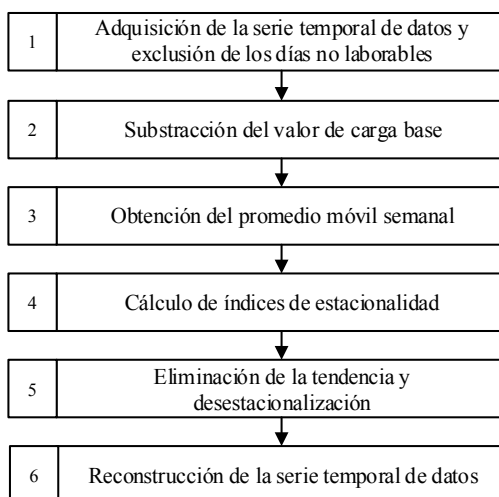


Imagen 3.25 Método SAEC

1. Adquisición de la serie temporal de datos y exclusión de los días no laborables

Se obtiene un vector de datos de potencia activa cuartal horaria solamente excluyendo los DNL correspondientes al periodo a analizar y se registra la cantidad de DL que conforman cada semana. Los datos correspondientes a los DNL no se consideran en este procedimiento debido a que la demanda de electricidad permanece más o menos constante cuando un edificio no tiene ocupación.

2. Substracción del valor de carga base

Se calcula el valor de carga base, para ello se computa el primer percentil de los datos del vector anterior. La estacionalidad tiene poca influencia en las HNL en donde se demandan valores de potencia cercanos a la carga base. Al analizar el vector se aprecia que la cantidad de anomalías de valor bajo, es menor que el 1% de todos los datos, por lo tanto, se ignoran mediciones que corresponden a eventos tales como cortes de energía, desconexiones, etc. A los datos del vector de los DL del paso 1, se le resta el valor de la carga base, así el vector resultante tiene valores cercanos a cero durante las HNL, esto es beneficioso en el análisis, ya que la estacionalidad tiene mínima influencia en estas horas.

3. Obtención del promedio móvil semanal

Se obtiene la potencia media de cada semana, considerando solamente el número de días registrados en el paso 1. Se calcula el promedio móvil de los valores obtenidos, para ello se utilizan 4 valores, es decir 4 semanas. A continuación se calcula el prome-

dio móvil centrado, esto se realiza al promediar dos valores consecutivos del promedio móvil.

4. Cálculo de índices de estacionalidad

Se calcula la media de todos los valores obtenidos al final del paso 3. Se obtienen los índices de estacionalidad al dividir cada promedio móvil centrado entre la media calculada. Así, se consigue que la media de los índices de estacionalidad sea 1.

5. Eliminación de la tendencia y desestacionalización

Se elimina la tendencia y se desestacionaliza el vector de datos obtenido en el paso 1 al dividir cada uno de sus valores entre los índices de estacionalidad correspondientes a cada semana que fueron obtenidos en el paso 4.

6. Reconstrucción de la serie temporal de datos.

Al vector de datos del paso anterior se le suma el valor de carga base calculado en el paso 2. Finalmente, se reconstruye el vector de datos de demanda eléctrica, juntando en la secuencia correcta los datos de los días laborables procesados y los datos de los días no laborables que fueron excluidos inicialmente.

En la Imagen 3.26 se muestra la demanda de electricidad (potencia media diaria) de un año para UPS Sede Cuenca, mientras que en la Imagen 3.27 se observan los datos tratados mediante con método SAEC.

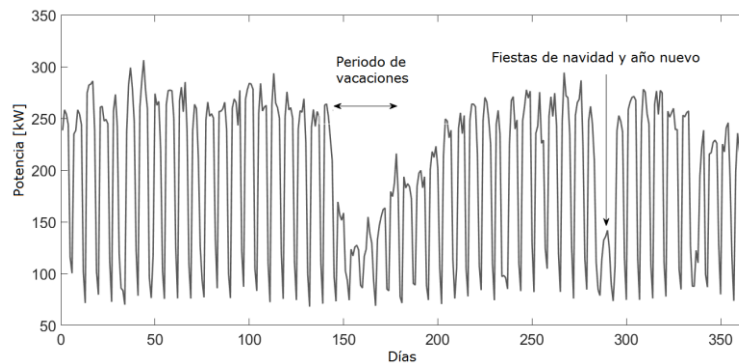


Imagen 3.26 Consumo de electricidad - Universidad Politécnica Salesiana (09-03-2017 hasta 08-03-2018)

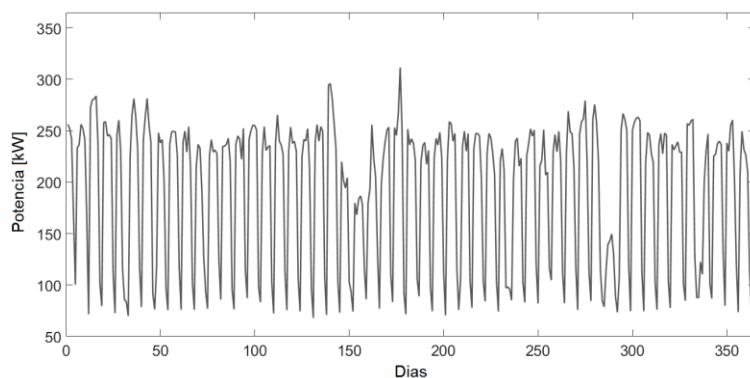


Imagen 3.27 Consumo de electricidad aplicando el método SAEC - Universidad Politécnica Salesiana (09-03-2017 hasta 08-03-2018)

A continuación se analiza la distribución de los datos después de realizar el tratamiento con el método SAEC (ver Imagen 3.28). Es evidente que el procedimiento realizado permite que los datos se ajusten de mejor manera a la distribución normal.

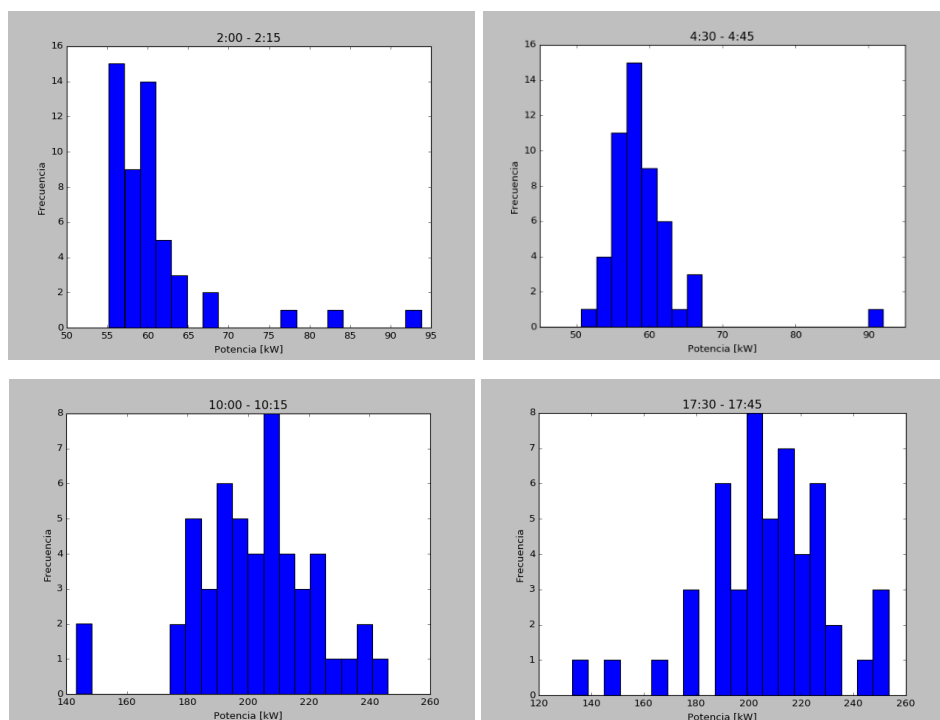


Imagen 3.28 Distribución de los datos para los días miércoles laborables en la UPS aplicando el método SAEC

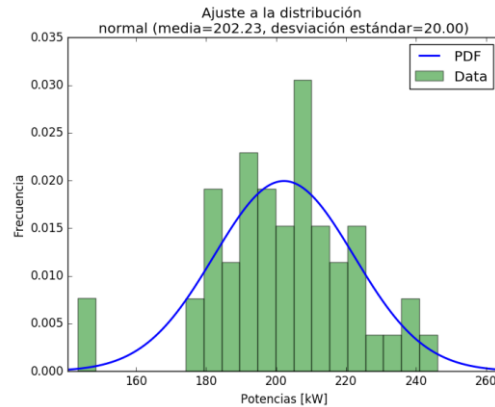


Imagen 3.29 Ajuste de los datos a una distribución normal para los miércoles laborables (10:00-10:15 am), al aplicar el método SAEC, valor $p = 0.07639$

En la Imagen 3.29 se muestra el ajuste de los datos de los días miércoles de 10:00 a 10:15 horas, en este caso el valor p es igual a 0.07639, por lo tanto se considera que los datos tienen una distribución normal.

3.5.2.3 Eliminación de anomalías y agrupamiento de datos

El ajuste de una instancia de datos a la distribución normal mejora cuando se eliminan anomalías y se agrupan varios tipos de días de consumos similares, aumentando de esta manera la cantidad de observaciones en la muestra. Este procedimiento se explica en detalle en el Capítulo 4, cuando se propone una nueva metodología para obtener patrones estocásticos de consumo. En la Imagen 3.30 se muestra la distribución de los datos resultante en varios intervalos cuarto horarios para un conjunto de PCE, los cuales representan el patrón de consumo para los días miércoles laborables.

La Imagen 3.31 muestra el ajuste a la distribución normal de los datos del patrón de los días miércoles laborables en la UPS entre las 10:00 y 10:15 am. El valor p obtenido es mayor que en los casos anteriores, confirmando que los datos se ajustan de mejor manera a la curva gaussiana.

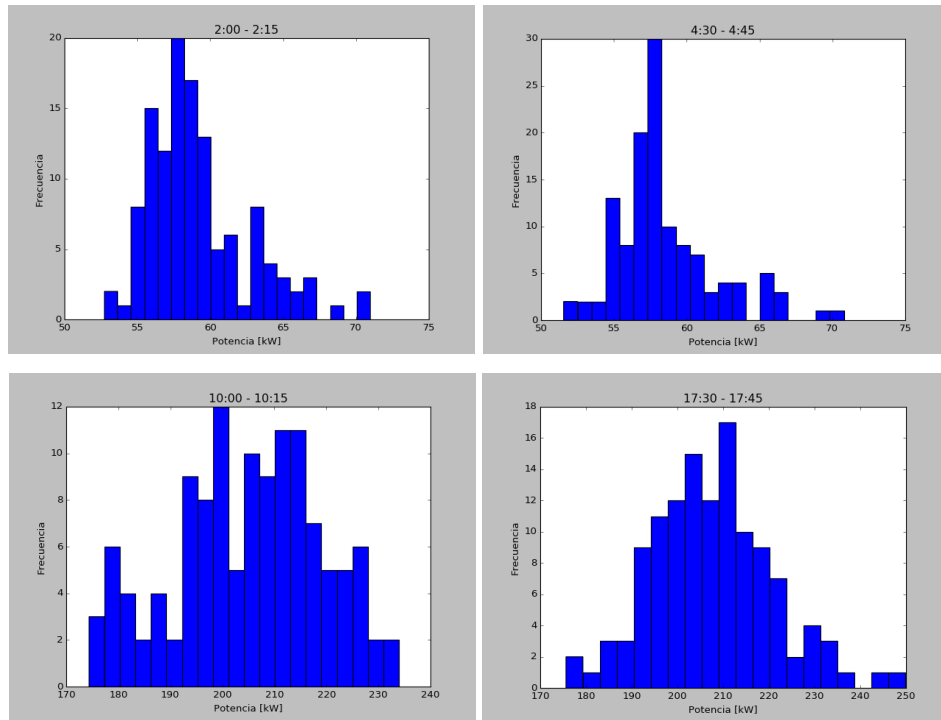


Imagen 3.30 Distribución de los datos para el patrón de los días miércoles laborables en la UPS

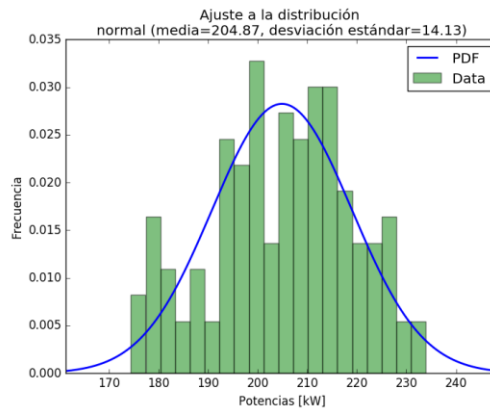


Imagen 3.31 Ajuste a la distribución normal del patrón de los días miércoles (10:00-10:15 am) en la UPS, valor $p = 0.10208$

3.5.3 Tratamiento de los datos de la muestra en el edificio 5E de la UPV

Adicionalmente, se utilizan los datos del consumo de electricidad de otro consumidor, como es el edificio 5E de la Universidad Politécnica de Valencia (UPV) en España desde el 01 de noviembre de 2014 hasta el 31 de octubre de 2015 con el objeto de confirmar la validez del análisis de la muestra de los datos de consumo eléctrico propuesto en este capítulo.

Inicialmente se verifica como están distribuidos los datos. En la Imagen 3.32 se observa la distribución de los datos para 4 diferentes intervalos cuarto horarios para los días miércoles laborables. Al igual que en el análisis anterior para la UPS, se toma el intervalo cuarto horario de entre las 10:00 y 10:15 am para verificar si los datos se ajustan a la distribución normal. En este caso en particular el ajuste es satisfactorio, incluso sin tratar los datos con el método SAEC aportado en esta tesis (ver Imagen 3.33).

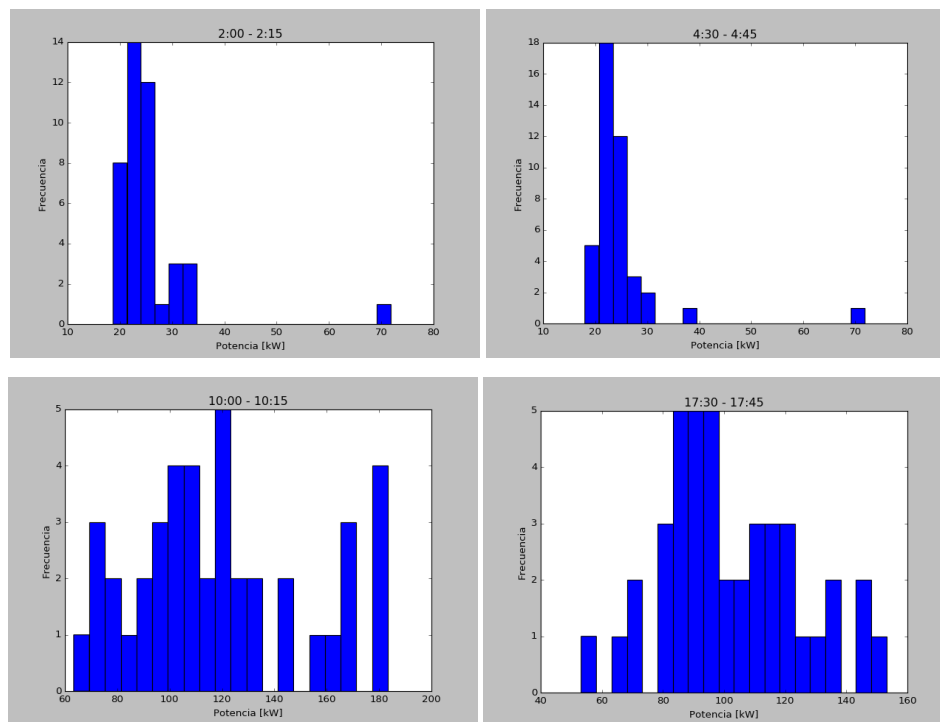


Imagen 3.32 Distribución de los datos para los días miércoles laborables en la UPV

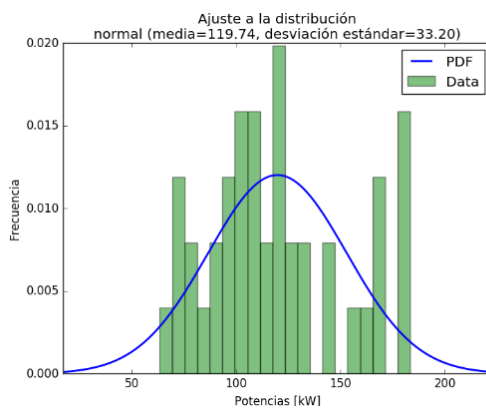


Imagen 3.33 Ajuste de los datos a una distribución normal para los miércoles laborables (10:00-10:15 am) en la UPV, valor $p=0.20972$

3.5.3.1 Tratamiento de datos con el método SAEC

En la Imagen 3.34 se presenta la distribución de los datos para los miércoles laborables a lo largo de un año en la UPV, después de aplicar el método SAEC.

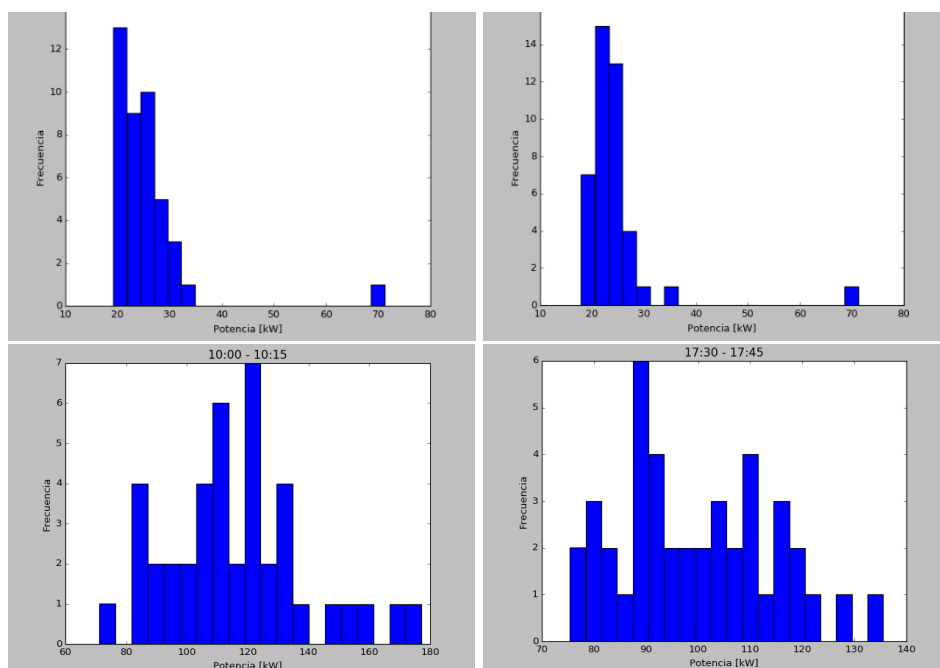


Imagen 3.34 Distribución de los datos para los días miércoles laborables en la UPV aplicando el método SAEC

Al observar la Imagen 3.35 se evidencia que el ajuste a la distribución normal desmejora al aplicar el método SAEC en la UPV. El valor p en este caso es menor que el indicado en la Imagen 3.33.

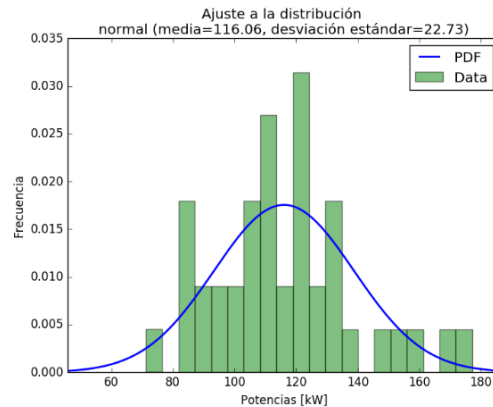


Imagen 3.35 Ajuste de los datos a una distribución normal para los días miércoles laborables (10:00-10:15 am) aplicando el método SAEC, valor $p = 0.12007$

3.5.3.2 Eliminación de anomalías y agrupamiento de datos

Al eliminar los PCE que contienen anomalías y realizar el agrupamiento de días con consumos similares se obtiene una muestra que representa el consumo típico del día de la semana seleccionado. Para el caso del miércoles seleccionado anteriormente, se evidencia que el ajuste a la distribución normal mejora notablemente (ver Imagen 3.36).

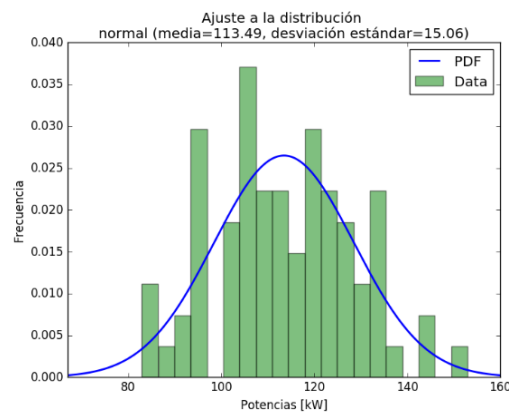


Imagen 3.36 Ajuste a la distribución normal del patrón de los días miércoles (10:00-10:15 am) aplicando el método SAEC, valor $p = 0.73952$

3.5.3.3 Tratamiento de los datos en periodos de verano e invierno

La climatización en los edificios de la UPV tienen dos modos de funcionamiento, en verano funcionan un tipo de dispositivos y en invierno otros. Los meses en los que los equipos de climatización de verano se activan son abril, mayo, junio, julio, agosto, septiembre y octubre, mientras que, los equipos de calefacción se activan en los meses de noviembre, diciembre, enero, febrero y marzo. A continuación se analizan por separado estos dos modos de funcionamiento con el objeto de verificar si los procedimientos de segmentación, tratamiento de la serie temporal de datos y la eliminación de anomalías mejoran el ajuste de los datos a una distribución normal.

Periodo de verano

Se analiza el periodo desde el 01 de abril hasta el 31 de octubre de 2015 (214 días). Los resultados para un día miércoles laborable sin aplicar el método SAEC se muestran en la Imagen 3.37.

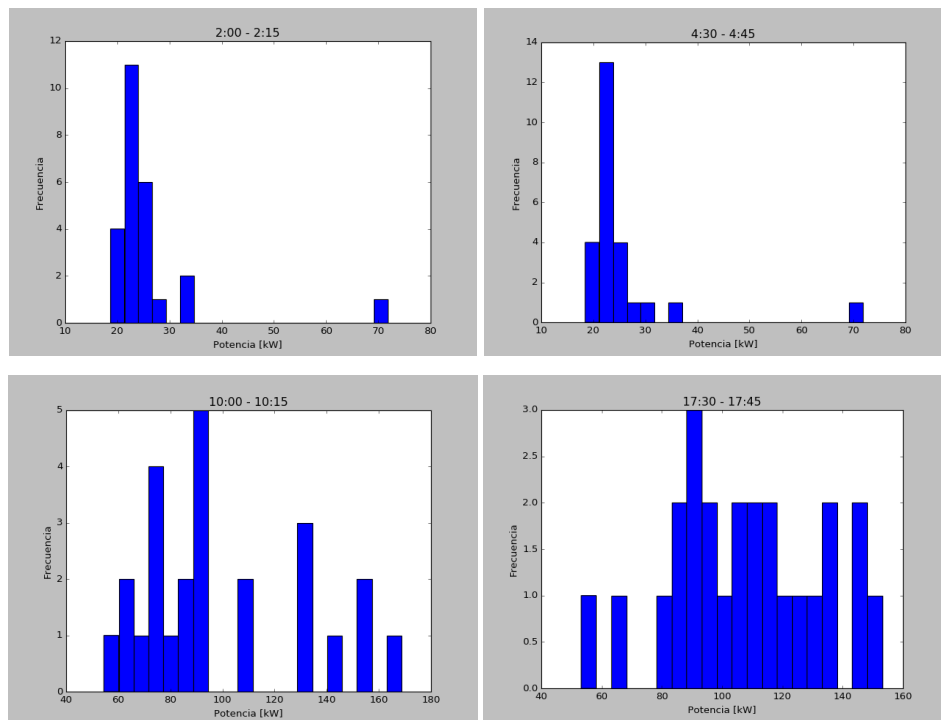


Imagen 3.37 Distribución de los datos para los días miércoles laborables en la UPV en el periodo de verano

Pese a que existen pocos datos en el intervalo de 10:00 a 10:15 am, estos se ajustan a una distribución normal con un valor $p=0.217003$, tal como se ve en la Imagen 3.38.

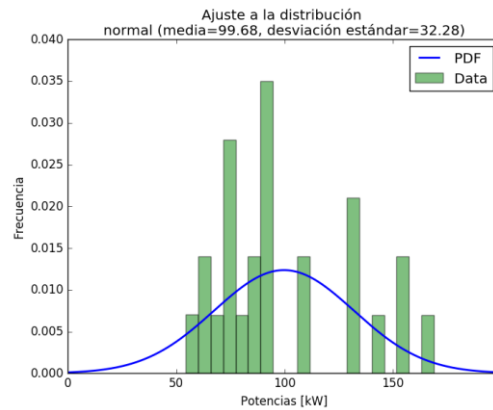


Imagen 3.38 Ajuste de los datos a la distribución normal para los miércoles laborables (10:00-10:15 am) en el periodo de verano - UPV, valor $p = 0.217003$

Al aplicar el método SAEC a la serie de datos el valor p aumenta (Imagen 3.39) por lo que se puede corroborar que este procedimiento también se puede utilizar cuando existen diferentes modos de funcionamiento en la climatización.

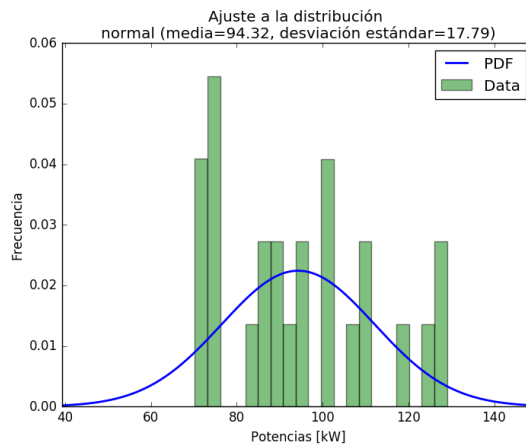


Imagen 3.39 Ajuste de los datos a una distribución normal para los miércoles laborables (10:00-10:15 am) en el periodo de verano, al aplicar el método SAEC, valor $p = 0.37131$

Al eliminar valores atípicos y agrupar los datos también se obtiene una distribución normal, tal como se observa en la Imagen 3.40.

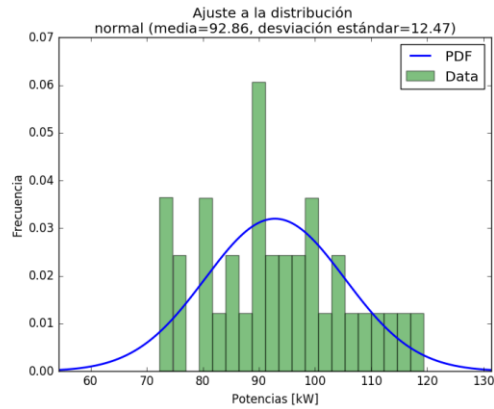


Imagen 3.40 Ajuste a la distribución normal del patrón de los miércoles laborables (10:00-10:15 am) en la UPV en el periodo de verano, valor $p = 0.53981$

Periodo de invierno

A continuación se analiza el modo de funcionamiento de los equipos de climatización para invierno. El intervalo de análisis es desde el 01 de noviembre de 2014 hasta el 31 de marzo de 2015 (151 días). La distribución de los datos sin aplicar el método SAEC se puede ver en la Imagen 3.41. En el periodo de invierno considerado existen 17 días miércoles, aún con esta pequeña cantidad de datos, estos se aproximan a la distribución normal. En la Imagen 3.42 se muestra el ajuste para el intervalo cuarto horario de 10:00 a 10:15 horas.

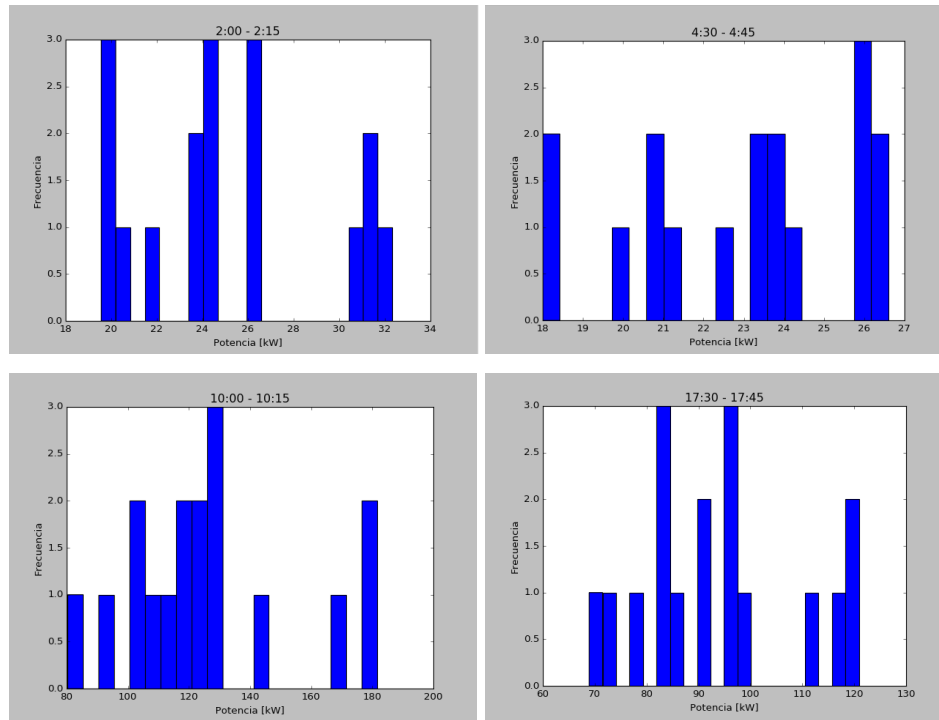


Imagen 3.41 Distribución de los datos para los días miércoles laborables en la UPV en el periodo de invierno

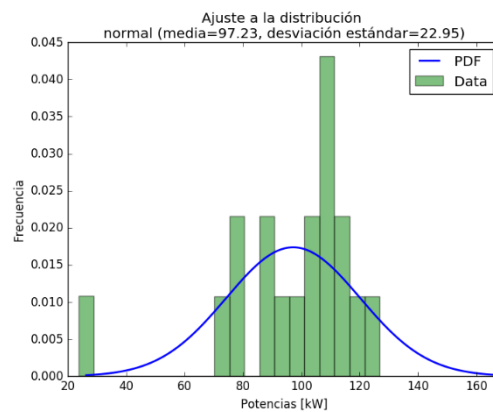


Imagen 3.42 Ajuste de los datos a la distribución normal para los miércoles laborables (10:00-10:15 am) en el periodo de invierno - UPV, valor $p = 0.31706$

Cuando se aplica el método SAEC a los datos, el ajuste a la distribución de probabilidad normal mejora. El ajuste a la normal del intervalo cuarto horario de 10:00 a 10:15 am se muestra en la Imagen 3.43.

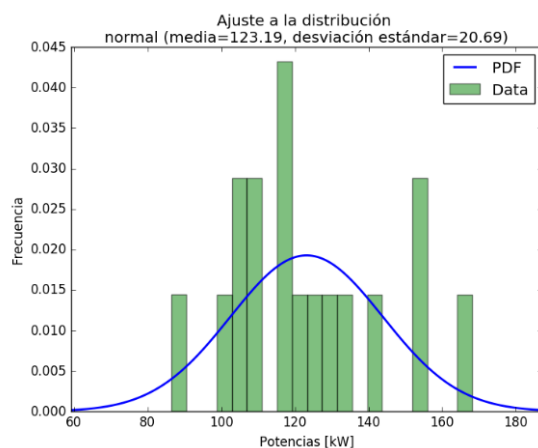


Imagen 3.43 Ajuste de los datos a la distribución normal para los miércoles laborables (10:00-10:15 am) en el periodo de invierno aplicando el método SAEC, valor $p = 0.55183$

Después de realizar la eliminación de anomalías y agrupar días de consumo similar, los datos se siguen ajustando una distribución de probabilidad normal. Los resultados obtenidos sugieren que los procedimientos de segmentación, tratamiento de la serie temporal y eliminación de anomalías también se pueden aplicar en una instalación en periodos interanuales siempre que existan dispositivos que funcionan en determinadas temporadas, lo cual influye en las componentes de la serie temporal de datos.

3.6 Conclusiones del capítulo

Las nuevas REI y sus sistemas de medición generan gran cantidad de datos, por ende se requiere una recopilación, procesamiento y análisis adecuado. La revisión de los sistemas de medición de electricidad ha permitido identificar algunas características y necesidades relevantes:

- El sistema de medición es una de las secciones más importante de una red eléctrica inteligente. La disponibilidad de información precisa y oportuna del consumo de electricidad en toda la cadena de suministro propicia el dinamismo y evolución del mercado eléctrico y por lo tanto un mejor aprovechamiento de los recursos de los consumidores y del sistema en general.

- La administración de las REI es cada vez más compleja. En este contexto, la medición de los parámetros eléctricos y el reconocimiento de patrones es imprescindible para la toma de decisiones en la gestión de este tipo de energía final.
- Las redes eléctricas enfrentan nuevos desafíos, ya que por un lado deben entregar información para detectar anomalías, aumentar la eficiencia, contaminar menos o aumentar la rentabilidad. Pero por otro lado deben evolucionar para, ser más seguras y confiables en el manejo de los datos, tener capacidad de almacenamiento energético, restablecer el servicio rápidamente ante perturbaciones en el sistema, permitir GD y un flujo bidireccional de potencia y finalmente proveer mayor protagonismo al consumidor.

En este capítulo, también se presentó la implementación de un sistema de MI en la UPS Sede Cuenca. Este sistema provee la información de partida para el desarrollo de la metodología estadística propuesta en el siguiente capítulo. Además, la puesta en marcha de este sistema es una interesante contribución al sector en el Ecuador, convirtiéndose en la actualidad en un laboratorio de medición y monitorización de electricidad. Este laboratorio permite la caracterización de la demanda, identificación de patrones de consumo basados en datos históricos, establecimiento de tarifas, predicción de la demanda, mejora en la operación y administración del sistema, evaluación de medidas de eficiencia energética, etc.

Los conceptos abordados acerca de las series temporales permitieron definir a la demanda de electricidad como una serie de datos estacional que responde a un proceso estocástico. Además, se describió el procedimiento seguido para el análisis de la muestra y se resaltó la importancia de definir la distribución de probabilidad que mejor se ajuste a las instancias de datos. Los datos adquiridos por los sistemas de medición por sí solos no se ajustan a una distribución de probabilidad específica, por lo que es necesario realizar un tratamiento previo.

En esta tesis se proponen tres procedimientos para afrontar la ausencia de una distribución de probabilidad definida en los datos de potencia cuarto horaria. El primer procedimiento es la segmentación, el segundo es el tratamiento de la serie temporal de datos y el tercero es la eliminación de anomalías. La segmentación se realiza en función de dos criterios. El primer criterio considera dos variables categóricas temporales (día de la semana e intervalo cuarto horario), mientras que el segundo considera a la laborabilidad como otra variable categórica de segmentación. Por otro lado, en el tratamiento de la serie temporal se aporta con un nuevo método para eliminar la tendencia y desestacionalizar los datos de la demanda de electricidad proporcionando mayor uniformidad en los datos de cualquier tipo de consumidor. A este método se le ha denominado

SAEC por sus siglas en inglés (Seasonality Analysis of Electricity Consumption). Las pruebas realizadas indican que la segmentación y el método SAEC también se pueden aplicar en periodos interanuales siempre que existan dispositivos que funcionen en determinadas temporadas, tal es el caso de la climatización. El ajuste a una distribución normal mejora cuando se eliminan datos anómalos y se agrupan PCE correspondientes a días de consumos similares.

Finalmente, se debe recalcar que un adecuado tratamiento de la muestra de los datos y su ajuste a una distribución de probabilidad conocida facilita el análisis e interpretación de datos. Estos datos procesados son la base de la metodología estadística propuesta en el siguiente capítulo.

Capítulo 4

Metodología estadística para valorar e identificar cambios en perfiles de consumo de electricidad (SAICC)

4.1 Introducción

La caracterización de la demanda de electricidad de los consumidores es posible gracias a los datos proporcionados por los contadores de energía, hoy en día medidores inteligentes. Según se ha visto en el Capítulo 3, estos datos requieren un tratamiento previo para que se ajusten a una distribución de probabilidad conocida, facilitando de esta manera el análisis e interpretación estadística de los datos.

En este capítulo se presenta una nueva metodología estadística que permite una adecuada caracterización de la demanda, por sus siglas en Inglés se le ha denominado metodología SAICC (statistical assessment for identifying changes in consumption methodology) y ha sido publicada en la revista *Energy and Buildings* [68]. Caracterizar la demanda de electricidad significa obtener su comportamiento típico, identificar y valorar sus cambios, catalogar perfiles de carga e incluso predecir comportamientos futuros. Esta caracterización se consigue gracias a un análisis estadístico de datos históricos de consumo de energía que permite identificar comportamientos recurrentes y predecibles con presencia de incertidumbre, a los cuales en esta tesis se les denomina *patrones estocásticos*. La metodología propuesta puede ser aplicada en cualquier tipo de consumidor y otros puntos o niveles de la red eléctrica. Valorar los cambios en la demanda de electricidad permite asociar estos valores a posibles eventos en una instalación. Esto puede ser empleado para detectar anomalías, generar alarmas, reducir costos en mantenimiento y aplicar medidas adecuadas rápidamente cuando se presenten incidentes. La metodología propuesta compara la demanda de energía de un *día de análisis* con el patrón estocástico obtenido. Posteriormente, los cambios detectados en la demanda se catalogan con el uso de una tabla de interpretación multicriterio, la cual

está basada en un algoritmo que presenta las posibles causas de esos cambios. Finalmente se calcula el índice de cambio IoC (del término anglosajón index of change).

En la segunda parte del capítulo se presenta un análisis de la influencia del tratamiento de la serie temporal de datos en la obtención de los patrones estocásticos de consumo y la detección de anomalías. En esta sección se utiliza el método SAEC propuesto en el Capítulo 3, para el tratamiento de los datos de la demanda de electricidad. Los resultados obtenidos evidencian que la utilización de este método aumenta la precisión en la detección de anomalías al identificar las anomalías contextuales.

4.2 Metodología SAICC

La metodología SAICC tiene cuatro etapas claramente definidas (ver diagrama de flujo en la Imagen 4.1), las dos primeras se basan en la teoría general del reconocimiento de patrones expuesta en el apartado 2.4.2. Cada una de estas etapas se detallan en los siguientes apartados.

4.2.1 Etapa 1: Adquisición y procesamiento de datos históricos

En la primera etapa de la metodología propuesta se adquieren y procesan los datos de los PCE, de tal manera que se ajusten a una distribución de probabilidad normal, permitiendo separar las anomalías. A continuación se describen cada una de sus subetapas.

E.1.1 Definición de parámetros iniciales

Para seleccionar los datos que posteriormente van a ser procesados, es necesario definir tres parámetros iniciales, el primero es el periodo de análisis (número de días que se usarán para reconocer los patrones), en donde cada día está caracterizado con 96 valores de medidas cuarta horarias formando un PCE. Además se define el tipo de día de análisis al seleccionar el día de la semana (lunes, martes, miércoles,...) y la laboralidad, es decir definir es laborable o no laborable.

E.1.2 Recolección de datos

Los datos de los PCE se organizan en una matriz de tamaño $M \times A$. En donde M indica el número de días del periodo seleccionado y A indica el número de características considerado en cada día.

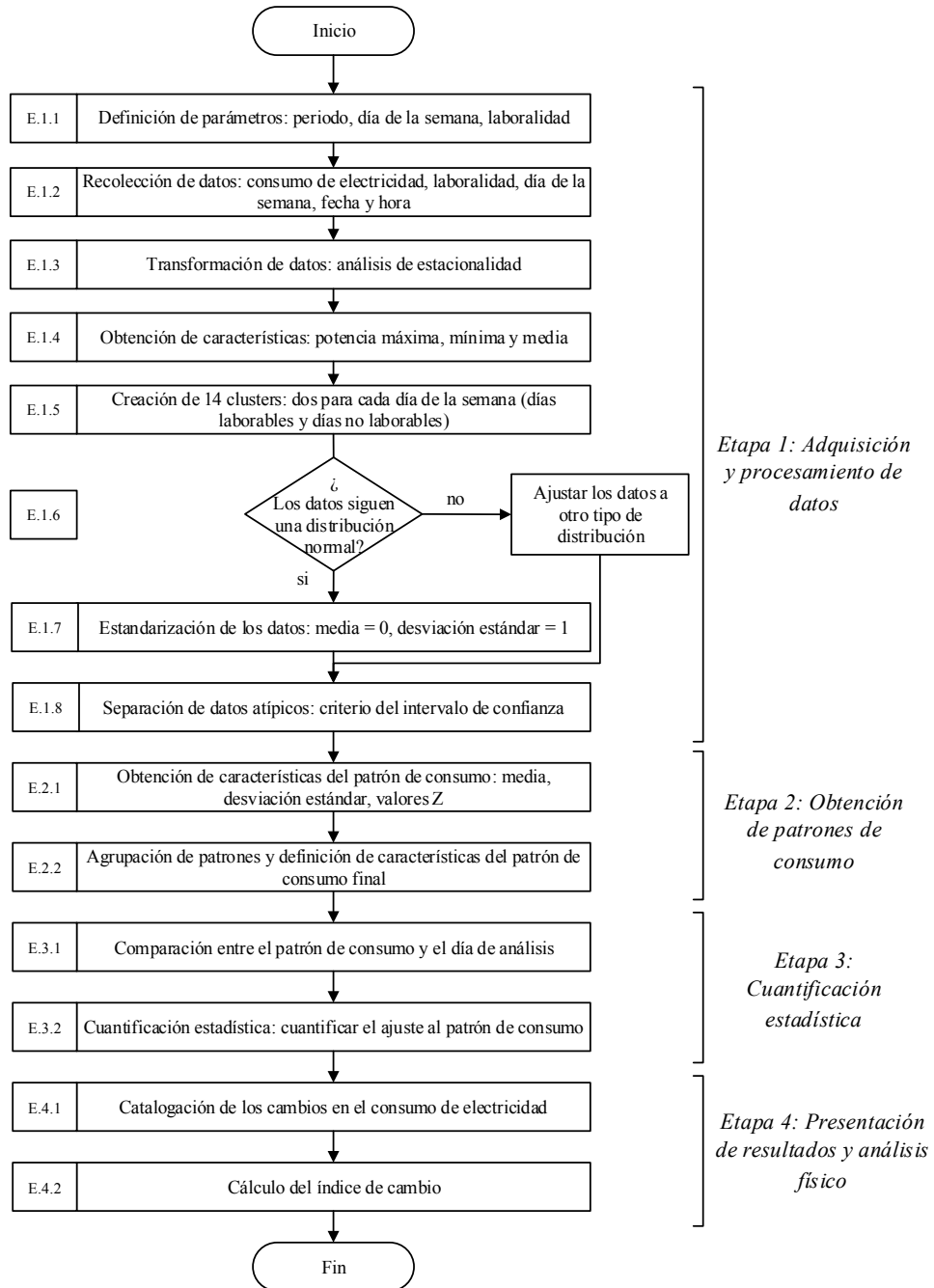


Imagen 4.1 Diagrama de flujo de la metodología SAICC

Debido a que generalmente la demanda de energía eléctrica en una instalación cambia a lo largo del tiempo, en esta tesis se utiliza un periodo máximo de un año, es decir, $M=365$. Las características consideradas para representar el consumo de electricidad de un día son los 96 datos de la potencia media cuarto horaria y tres variables que serán utilizadas para la segmentación de datos (fecha y hora, día de la semana y laboralidad), las cuales son descritas con detalle en el apartado 3.5.2.1 del capítulo anterior. Por tanto, el total de características es 99 ($A=99$).

E.1.3 Transformación de datos

La demanda de electricidad es una serie temporal con variaciones estacionales debido a cambios en los procesos de trabajo, temperatura ambiente, periodos de vacaciones, etc. En este tipo de datos es necesario evaluar si se requiere una transformación o tratamiento de los datos [69]. La transformación de datos en esta subetapa se realiza mediante la aplicación del método SAEC presentado en el apartado 3.5.2.2 del Capítulo 3, en donde también se establecieron las siguientes condiciones requeridas para su aplicación en la demanda de electricidad:

- Las fluctuaciones presentes en la serie de datos son repetitivas.
- Las fluctuaciones tienen periodos definidos.
- Los valores de la serie temporal siguen una tendencia lineal.
- Los valores de amplitud en cada ciclo tienen valores similares.

La aplicación del método SAEC elimina la tendencia y la componente estacional de los datos de tal manera que las instancias de datos se ajustan de mejor manera a la distribución de probabilidad normal, facilitando la interpretación de los resultados, tal como se evidenció en el apartado 3.5.2 y 3.5.3 del capítulo anterior.

En el caso de estudio del edificio 5E de la UPV se evidencian fluctuaciones repetitivas en periodos de un año con una tendencia lineal negativa, tal como se muestra en la Imagen 4.2, en la cual se representa la potencia activa media trimestral demandada. La amplitud de los valores en cada ciclo es similar, obviamente afectada por la tendencia lineal negativa. Se observa de manera clara que las 4 condiciones para llevar a cabo un análisis de estacionalidad se cumplen.

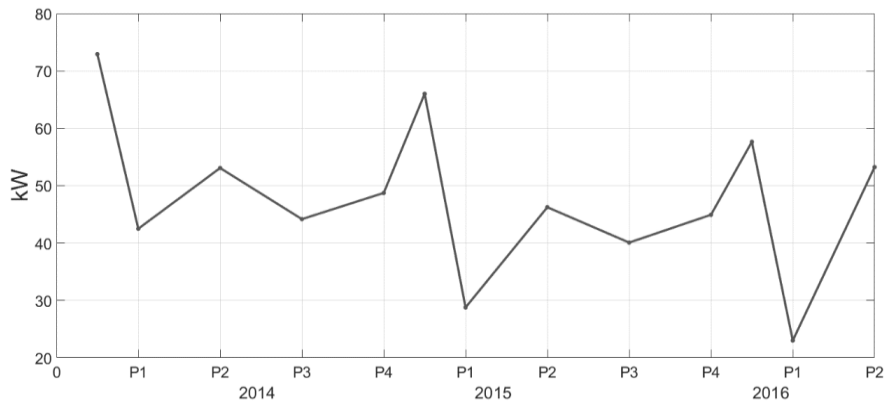


Imagen 4.2 Potencia media trimestral en el Edificio 5E de la UPV

Con la utilización del método SAEC propuesto se obtienen patrones estocásticos de consumo menos variables, mejorando la precisión en la detección de anomalías en los PCE, su aplicación en este ámbito se evalúa en el apartado 4.3 con mayor detalle.

La Imagen 4.3 muestra la demanda de electricidad (potencia media diaria) de un año para el Edificio 5E de la UPV. Mientras que, en la Imagen 4.4 se observan los datos tratados mediante el método SAEC.

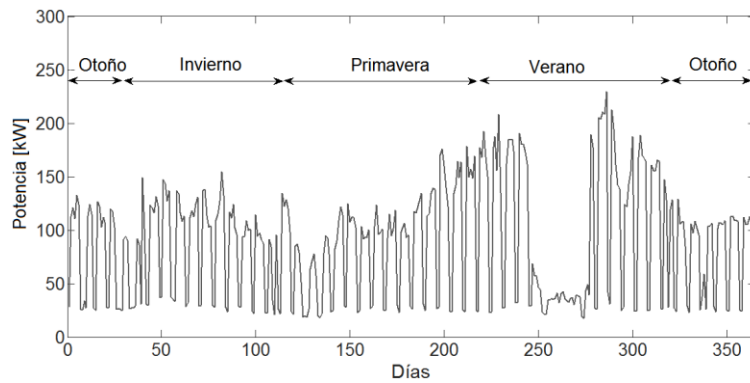


Imagen 4.3 Consumo de electricidad - Edificio 5E de la UPV (29-11-2015 hasta 28-11-2016)

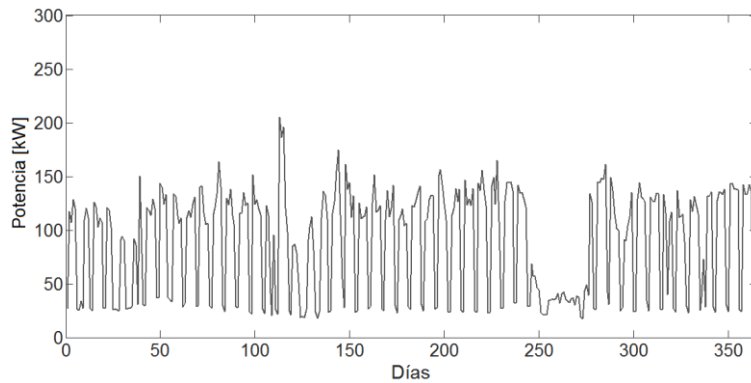


Imagen 4.4 Consumo de electricidad aplicando el método SAEC - Edificio 5E de la UPV (29-11-2015 hasta 28-11-2016)

E.1.4 Obtención de características

En el caso de que los datos hayan sido desestacionalizados se vuelven a organizar en la matriz de tamaño $M \times A$ descrita en el apartado 4.2.1. En el caso contrario se toma la matriz original. Seguidamente, se agregan tres columnas a la matriz de datos, en cada una de ellas se agrega la potencia máxima, media y mínima correspondiente a cada día. El cómputo del consumo de energía diario no se considera ya que se correlaciona directamente con la potencia media. La matriz de datos ahora tiene una dimensión de $M \times N$, en donde $N=102$.

E.1.5 Segmentación de datos

En el Capítulo 3 apartado 3.5 se abordó el análisis de la muestra de datos, llegándose a la conclusión de que los datos de la demanda de electricidad no tienen una distribución de probabilidad definida cuando carecen de un tratamiento previo. Entonces, se propusieron tres procedimientos para resolver este inconveniente, el tratamiento de los datos de la serie temporal (apartado 3.5.2.2), la eliminación de anomalías y la segmentación que se aborda en este apartado. La segmentación se realiza en función de tres variables de segmentación basadas en dos criterios. El primer criterio considera dos variables categóricas temporales (el intervalo cuarto horario y el día de la semana), mientras que el segundo considera a la laboralidad, como variable de segmentación.

Los PCE así como las variables consideradas para la segmentación están organizadas en la matriz $M \times N$ explicada en el apartado anterior. La matriz $M \times N$ es desagregada en 14 matrices, dos por cada día de la semana. Una matriz contiene los DL y la otra matriz los DNL, tal como se ve en la Imagen 4.5. El procedimiento descrito mejora la eficiencia del procesamiento de los datos ya que utiliza solamente la información de la o las matrices desagregadas de interés.

Una vez realizada la segmentación de datos las variables de segmentación son eliminadas. Las matrices desagregadas tienen entonces 99 columnas o características ($N'=99$), esto es, los 96 datos de la potencia media cuarto horaria y adicionalmente la potencia máxima, media y mínima correspondiente a cada fila.

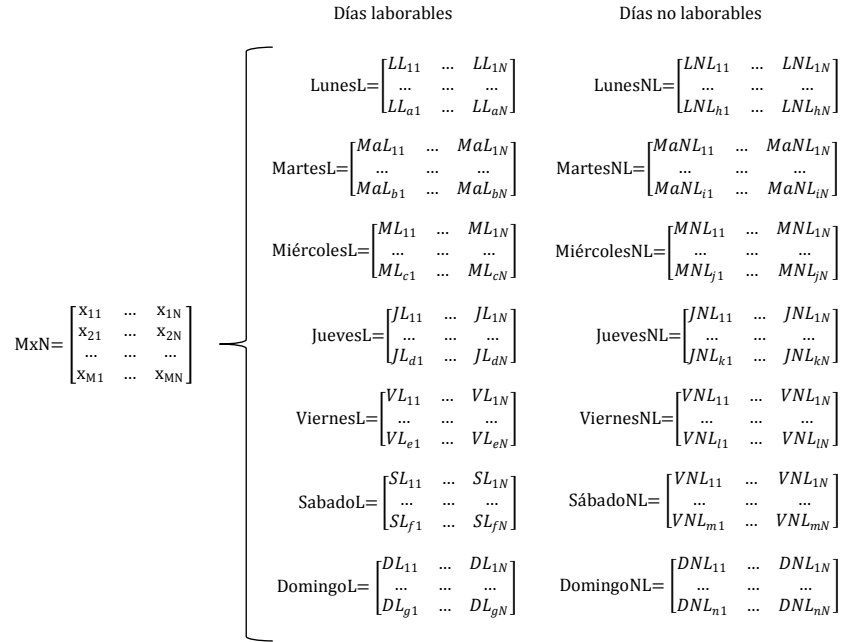


Imagen 4.5 Segmentación de los datos a través de matrices desagregadas

E.1.6 Análisis de la distribución de datos

Para el análisis estadístico propuesto posteriormente, se requiere comprobar si los datos en las matrices desagregadas siguen una distribución normal, esto se lo realiza por cada intervalo cuarto horario, es decir por cada columna de las submatrices. La Ecuación 4.1 describe la prueba de bondad de ajuste chi cuadrado utilizada para esta comprobación.

$$\chi^2 = \sum_{i=1}^k \frac{(o_i - e_i)^2}{e_i} \quad \text{Ecuación 4.1}$$

En donde χ^2 es el valor de la variable aleatoria, cuya distribución de muestreo tiene una distribución aproximada a la chi-cuadrado, con $k-1$ grados de libertad, k es el número de resultados posibles, por lo que $k-1$ representa el número de intervalos de clase. Finalmente o_i y e_i representan el valor de frecuencias observadas y esperadas respectivamente, para cada intervalo de clase i .

La teoría expuesta en el Capítulo 2 apartado 2.2.5 acerca de las pruebas de hipótesis permite establecer que los datos se ajustan a una distribución normal siempre y cuando la hipótesis nula H_0 sea rechazada. Definiéndose H_0 como: *los datos no se ajustan a una distribución normal*. Al rechazar H_0 se acepta la hipótesis alternativa H_1 que indica que: *los datos se ajustan a una distribución normal*. La prueba de hipótesis también se formaliza mediante el uso del *valor p*. El *valor p* es el valor más bajo en el que el estadístico de prueba es significativo y su utilidad ha sido ensayada en esta tesis en el apartado 3.5 del Capítulo 3 donde se analiza la muestra de datos.

Los resultados del análisis de la muestra de datos de la demanda de electricidad (apartado 3.5) indican que es factible suponer la normalidad de las instancias de datos gracias a la transformación y a la segmentación de datos. La ventaja de trabajar con datos ajustados a la normal es que muchos fundamentos teóricos ya han sido desarrollados, facilitando de esta manera el análisis estadístico.

E.1.7 Estandarización de los datos

Una vez probado que los datos se ajustan a una distribución normal, las matrices desagregadas son estandarizadas, haciendo la media de cero ($\mu=0$) y la desviación estándar uno ($\sigma=1$). La estandarización se realiza por cada columna de la matriz. Mientras más filas tenga la matriz, el estadístico de la desviación estándar tenderá a la desviación estándar real [3]. El valor de Z correspondiente a cada fila r y cada columna c , se calcula con la siguiente expresión:

$$Z_{rc} = \frac{x_{rc} - u_c}{\sigma_c}, \quad \text{Ecuación 4.2}$$

en donde x_{rc} es el valor de la variable X en la matriz desagregada de la fila r y la columna c , u_c es la media y σ_c es la desviación estándar de la variable X en la columna c . Los valores máximos de Z de cada columna se almacenan en un vector Z_{max} . Los valores mínimos de Z de cada columna se almacenan en el vector Z_{min} .

E.1.8 Separación de datos atípicos

En esta metodología la distribución normal actúa como una distribución limitante, de tal manera que un valor se considera anómalo cuando está fuera del intervalo de confianza del 95% de la distribución normal. Este procedimiento se realiza por cada columna de cada matriz desagregada. Se considera que, el consumo de un día es atípico cuando al menos uno de los N' valores de un día de análisis está fuera del intervalo de confianza.

La probabilidad de que un valor Z correspondiente a la columna c , esté dentro del intervalo de confianza puede ser expresado mediante:

$$P(Z_{\alpha_1,c} < Z_c < Z_{\alpha_2,c}) = 1 - \alpha_{1,c} - \alpha_{2,c} = 1 - \alpha \quad \text{Ecuación 4.3}$$

En donde $Z_{\alpha_1,c}$ y $Z_{\alpha_2,c}$, son los límites inferior y superior del intervalo de confianza de cada columna c , respectivamente. En tanto que $\alpha_{1,c}$ y $\alpha_{2,c}$ representan las áreas de la cola izquierda y derecha de la distribución normal estándar, respectivamente. Del Capítulo 2 es importante recordar que $\alpha_{1,c} + \alpha_{2,c} = \alpha$, siendo α el nivel de significancia y que $1 - \alpha$ representa el coeficiente de confianza.

Al final de esta etapa, en las matrices desagregadas de interés todos los datos correspondientes a los días atípicos son eliminados, quedando entonces solamente los datos de los días de demanda de electricidad típica.

4.2.2 Etapa 2: Obtención de patrones estocásticos de consumo

Una vez que las instancias de datos tienen una distribución normal y no contienen valores atípicos es posible obtener las características que representan estos datos, a las cuales se les ha llamado *patrones estocásticos de consumo*. A continuación se describe la etapa 2 de la metodología SAICC que describe este proceso.

E.2.1 Obtención de características del patrón de consumo

El patrón estocástico de consumo de cada día de la semana se obtiene a través de unas *matrices de patrón* que contienen los valores típicos del consumo de electricidad correspondiente a cada día. El tamaño de cada matriz es de $M_d \times 96$, siendo M_d el número de días que conforman cada *matriz de patrón* y 96 son los valores de potencia cuarta horaria que tiene cada día. El patrón se representa mediante 4 vectores que representan sus características: la media, desviación estándar y los valores de Z_{min} y Z_{max} de cada columna correspondiente a cada *matriz de patrón*.

En la Imagen 4.6 se representan los vectores de la media y desviación estándar correspondientes a los consumos típicos de cada día de la semana del Edificio 5E de la UPV tomando datos de un año. Por otro lado, en la Imagen 4.7 se muestran los valores Z_{min} y Z_{max} para el mismo periodo y edificio.

Caracterización de la demanda de energía mediante patrones estocásticos en las Redes Eléctricas Inteligentes

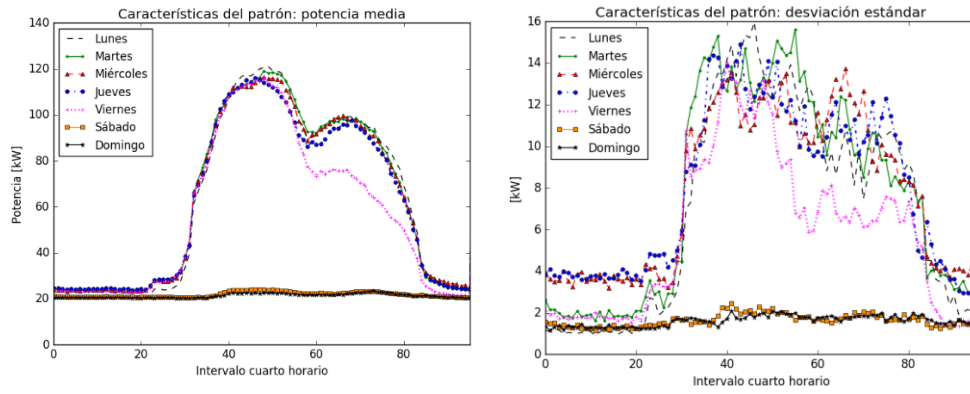


Imagen 4.6 Obtención de características del patrón, media y desviación estándar de la potencia (96 valores por día) del Edificio 5E de la UPV

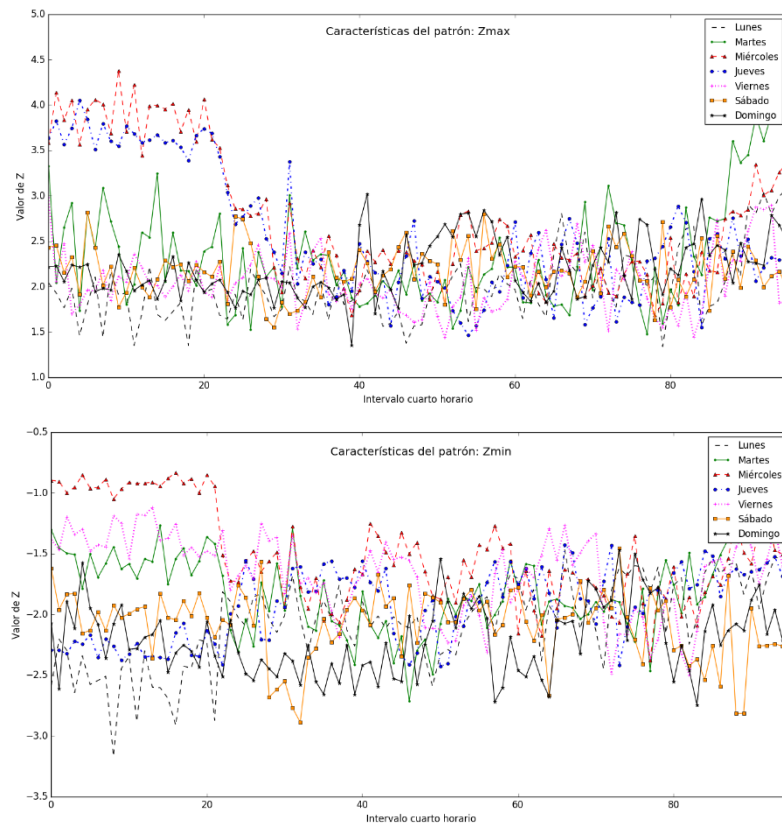


Imagen 4.7 Obtención de características del patrón, valores Z_{max} y Z_{min} del Edificio 5E de la UPV

E.2.2 Agrupación de patrones similares

En esta subetapa se agrupan los patrones que son similares al *día de análisis*. La agrupación resulta beneficiosa debido a que los métodos estadísticos trabajan mejor cuando se dispone de mayor cantidad de datos. Como métrica de agrupamiento se usa la distancia euclidiana, cuya ecuación es expresada mediante:

$$Distancia(A, B) = \sqrt{\sum_{i=1}^n (a_i - b_i)^2} \quad \text{Ecuación 4.4}$$

En donde A y B son vectores con el mismo número de elementos. En este caso $n = 96$ ya que los valores a_i y b_i , representan las características del patrón normalizadas, las cuales fueron definidas en el apartado anterior. La normalización en este caso se refiere simplemente a un cambio de escala al dividir cada valor a_i y b_i para el valor máximo de los dos vectores.

El proceso de agrupamiento inicia con el cálculo de la distancia euclidiana entre los vectores normalizados de la media y desviación estándar de cada *matriz de patrón* (7 matrices, una por cada día de la semana) obtenidos en el apartado anterior. Si la distancia calculada es menor a un valor umbral arbitrario, los patrones son considerados similares y seguidamente las matrices de patrón correspondientes se agrupan. Los resultados del proceso de agrupamiento en el Edificio 5E de la UPV expuesto en la Imagen 4.6, se detalla en la Tabla 4.1.

Tabla 4.1 Resultados del proceso de agrupamiento

N°	Matrices agrupadas
1	lunes
2	martes - miércoles - jueves
3	viernes
4	sábado - domingo

El *patrón final* de consumo es representado mediante una *matriz de patrón final*, la cual está formada por los datos de los días de demanda similar al día de análisis. Estos datos definen las características del patrón mediante la media, desviación estándar y valores Z . La información contenida en cada matriz de patrón final se aprecia de mejor manera cuando se la representa con diagramas de cajas y bigotes (ver Imagen 4.8), a los cuales también se los puede llamar patrones estocásticos.

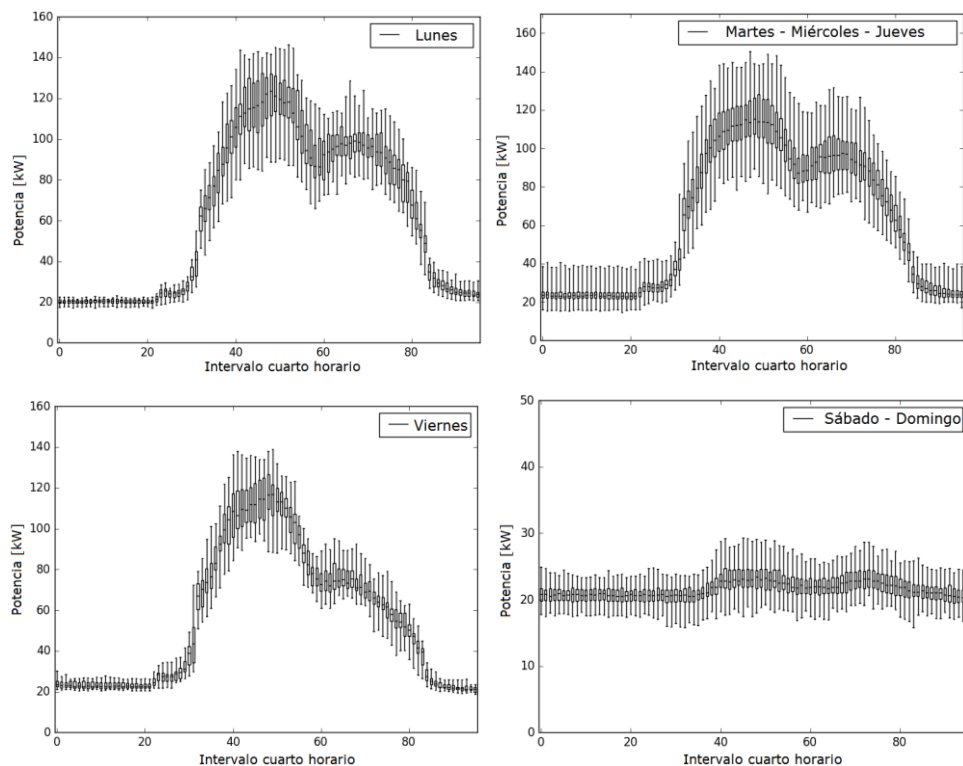


Imagen 4.8 Patrones estocásticos de consumo en el Edificio 5E de la UPV representados mediante diagramas de cajas y bigotes

4.2.3 Etapa 3: Cuantificación estadística

La tercera etapa de la metodología SAICC compara y cuantifica el ajuste que tiene la demanda de electricidad de un día de análisis al patrón estocástico obtenido en la etapa anterior. En los próximos apartados se detallan las dos subetapas que realizan esta tarea.

E.3.1 Comparación entre el patrón de consumo y el día de análisis

Un día de análisis puede compararse con el patrón obtenido. Como ya se ha indicado anteriormente mediante la *matriz de patrón final* se definen las características del patrón de consumo, las cuales sirven de referencia para comparar la demanda de electricidad de un día específico. El día de análisis se representa mediante un vector fila que tiene un número N' de columnas, cuyos valores fueron definidos en la subetapa E.1.5.

Los valores de consumo del día de análisis son estandarizados usando la siguiente expresión:

$$Z_i = \frac{x_i - u_i}{\sigma_i} \quad \text{Ecuación 4.5}$$

En donde Z_i es el valor Z , i es el índice de la columna, x_i es el valor original de la potencia activa en la posición i . En tanto que u_i es la media, y σ_i es la desviación estándar de los valores de la columna i de la matriz de patrón final.

Como se había indicado anteriormente, los datos transformados se ajustan a una distribución normal estándar. Además, es importante recalcar que los coeficientes de confianza no son fijos, esto debido a que los valores límite $Z_{min,i}$ y $Z_{max,i}$ son distintos para cada una de las N' columnas.

Se intuye claramente que, si Z_i está dentro del intervalo de confianza $Z_{min,i} < Z_i < Z_{max,i}$ este valor es considerado típico. Los coeficientes de confianza pueden ser asociados directamente a la probabilidad de que un valor sea típico, es decir que no existan anomalías en la demanda de energía eléctrica.

El test de hipótesis es empleado para aceptar o rechazar la existencia de valores atípicos. La hipótesis nula H_0 se define como: *la potencia consumida en el intervalo analizado no es anómala*, mientras que la hipótesis alternativa H_1 es definida como: *la potencia consumida en el intervalo analizado es anómala*.

Es importante recalcar que el nivel de significancia (α) no es fijo en cada intervalo analizado, ya que como se explicó con anterioridad los valores de $Z_{min,i}$ y $Z_{max,i}$ son diferentes. Lo anterior indica, que el test de hipótesis no se aproxima con una probabilidad fija del error tipo I (rechazar la hipótesis nula cuando esta es verdadera) [3].

E.3.2 Cuantificación del ajuste al patrón de consumo

En esta subetapa, se cuantifica estadísticamente el ajuste del día de análisis al patrón estocástico de consumo. Para ello se calculan diferentes variables: el valor de Z , coeficiente de confianza, error tipo I, error tipo II (no rechazar la hipótesis nula cuando esta es falsa) [3] y un índice de atípicos (0 si el valor es típico, 1 si el valor es atípico). Cada una de estas variables se computa para cada una de las 99 características de la demanda de electricidad. El resumen de las variables calculadas se muestra en la Tabla 4.2.

Tabla 4.2 Variables calculadas

Número de variable	Tipo de característica	Resultados
1-96	Potencia media consumida en un intervalo cuarto horario	
97	Potencia media del día	Z_i , coeficiente de confianza, error tipo I, error tipo II e índice de atípicos para cada una de las 99 variables
98	Potencia máxima del día	
99	Potencia mínima del día	

4.2.4 Etapa 4: Presentación de resultados y análisis físico

La última etapa de la metodología SAICC presenta como conclusiones las posibles causas de los cambios que son detectados en la demanda de electricidad del día de análisis. Además, se calcula un IoC que tiene la capacidad de catalogar los cambios que se han presentado. Las subetapas involucradas se detallan a continuación.

E.4.1 Catalogación de cambios en el consumo de electricidad

La metodología propuesta identifica consumos anómalos e indica cuál es su posible causa. Para ello, se aplican 12 pruebas diferentes en los datos del día de análisis, las cuales entregan información relevante. Estas pruebas se pueden resumir en la proporción de anomalías en todo el día, proporción anomalías en horarios específicos, duración de la demanda anómala, tiempo total en el que se repite el mismo valor de medición, finalmente se evalúan la potencia máxima, media y mínima mediante 3 pruebas. Los detalles de las pruebas aplicadas en la metodología SAICC se pueden ver en la Tabla 4.3.

Las 12 pruebas se analizan para presentar conclusiones acerca de las anomalías presentadas. Con este propósito se desarrolló un algoritmo que obtiene una tabla de interpretación multicriterio de las pruebas aplicadas MCIAT (del término anglosajón multicriteria interpretation of applied test table). De esta forma, se determinaron 569 resultados posibles, los cuales indican las posibles causas del consumo anómalo detectado. La Imagen 4.9 muestra el perfil de consumo de electricidad de cuatro días diferentes, que representan cuatro casos distintos, en comparación con su correspondiente patrón estocástico de consumo.

Tabla 4.3 Pruebas aplicadas en la metodología SAICC

Prueba número	Descripción	Valor de la Prueba
1	Número total de anomalías en el día, desde 1 a 96 (NTA)	$T1 = NTA/96 * 100\%$
2	Número de anomalías en HNL (NAHNL)	$T2 = NAHNL / 40 * 100\%$
3	Número de anomalías [0:00-7:00] (NA_0-7)	$T3 = NA_{0-7} / 28 * 100\%$
4	Número de anomalías [21:00-24:00] (NA_21-24)	$T4 = NA_{21-24} / 12 * 100\%$
5	Número de anomalías en HL [7:00-21:00] (NAHL)	$T5 = NAHL / 56 * 100\%$
6	Número de anomalías [7:00-14:00] (NA_7-14)	$NA_{7-14} / 28 * 100\%$
7	Número de anomalías [16:00-21:00] (NA_16-21)	$NA_{16-21} / 20 * 100\%$
8	Valores iguales consecutivos VIC	(VIC/4) [horas]
9	Números de anomalías consecutivas NAC	(NAC/4) [horas]
10	Potencia media	Mayor, menor, en rango y coeficiente de confianza.
11	Potencia máxima	Mayor, menor, en rango y coeficiente de confianza.
12	Potencia mínima	Mayor, menor, en rango y coeficiente de confianza.

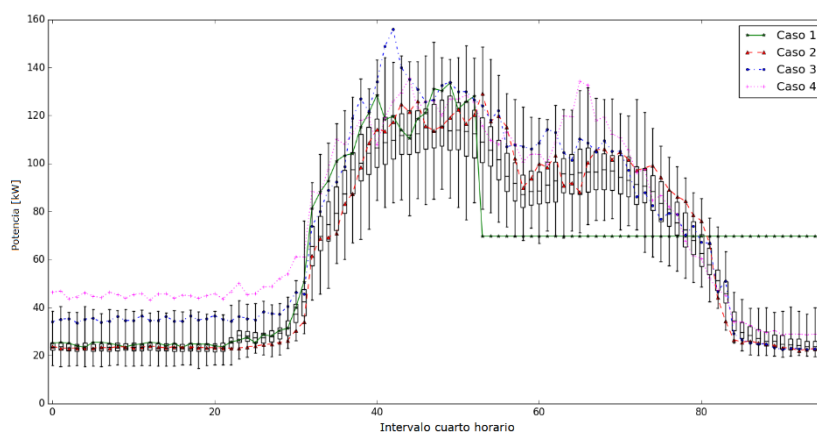


Imagen 4.9 Cuatro días evaluados con su correspondiente patrón de consumo

A modo de ejemplo, la Tabla 4.4 muestra la interpretación de las pruebas realizadas para esos cuatro días distintos representados como 4 casos.

Tabla 4.4 Tabla multicriterio de interpretación de las pruebas aplicadas (MCIAT)

Prueba	Caso 1	Caso 2	Caso 3	Caso 4
T1	x	0%	>0%	>0%
T2	x	x	<10%	>10% y <100%
T3	x	x	x	100%
T4	x	x	x	0
T5	x	x	<50%	>0 y <100%
T2/T5	x	x	x	>1.35
T3/T4	x	x	x	x
T8	>2	<2	<2	<2
T9	x	0	<1	x
T10	x	En el rango	En el rango	Mayor
T11	x	En el rango	Mayor	En el rango
T12	x	En el rango	En el rango	En el rango
Conclusión	Error de medición. Más de dos horas con el mismo valor de medida.	El consumo de energía en el día analizado se ajusta al patrón.	Menos de 10% de consumos anómalos, tanto en HL y HNL. Sin embargo, la potencia máxima consumida es mayor a la potencia máxima del patrón. Existe un consumo elevado en horas pico.	El consumo es anómalo en todas las mediciones de la madrugada. No hay consumos atípicos en las HNL de la noche. Posible carga se quedó conectada. El consumo de energía es mayor a la del patrón. Es probable que una carga adicional haya permanecido conectada, o existe un error de medición. La potencia mínima consumida es mayor a la del patrón. Es posible que una carga adicional haya permanecido conectada.

Nota: 'x' significa que la prueba puede tomar cualquier valor.

Las pruebas 10, 11 y 12 evalúan la potencia media, máxima y mínima respectivamente, siendo de trascendental importancia en el algoritmo que presenta las conclusiones del análisis de perfil de carga. Para estas variables se utilizan diagramas de cajas y bigotes. La Imagen 4.10 ilustra los patrones de potencia máxima, media y mínima en comparación con el consumo de los cuatro días diferentes analizados.

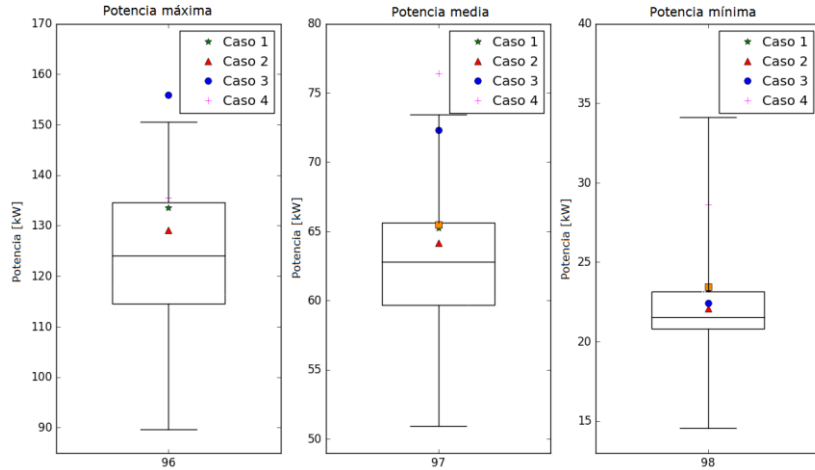


Imagen 4.10 Patrón de potencia máxima, media y mínima en comparación con cuatro días diferentes (cuatro casos)

E.4.2 Cálculo del índice de cambio

En esta metodología también se define un índice para valorar los cambios en el PCE diario con respecto al patrón estocástico obtenido con anterioridad. Esto permite valorar con una puntuación de forma clara e inequívoca que tan anómala es la demanda de energía en un día de análisis. Inicialmente se define I_{change} como:

$$I_{change} = P_1 T_1 I_{C1} + P_2 |I_2| I_{C2} + P_3 |I_3| I_{C3} + P_4 |I_4| I_{C4} , \quad \text{Ecuación 4.6}$$

en donde P_1 , P_2 , P_3 y P_4 son los pesos de cada término de I_{change} . Debido a que el primer y segundo término tienen gran influencia en la valoración de cambios del consumo, P_1 y P_2 toman el valor de $1/3$ cada uno, mientras que a P_3 y P_4 toman un valor de $1/6$. T_1 es el número diario de datos atípicos expresado como $T_1/100$. I_{C1} es la mediana de los valores del coeficiente de confianza de los 96 datos de potencia media cuarto horaria de la matriz de patrón final. I_{C2} , I_{C3} e I_{C4} son los valores del coeficiente de confianza de la potencia media, máxima y mínima de la matriz de patrón final, respectivamente. En tanto que I_2 , I_3 e I_4 son los índices de atípicos de la potencia media, máxima y mínima respectivamente, estos índices toman un valor de 1 cuando la observación está por encima del diagrama de cajas y bigotes, -1 cuando se encuentra por debajo y cero cuando se está dentro.

A continuación I_{change} es normalizado debido a que, por definición, siempre es menor que 1. Para ello se calcula I_{cmax} :

$$I_{cmax} = P_1 I_{C1} + I_{C2} P_2 + I_{C3} P_3 + I_{C4} P_4 \quad \text{Ecuación 4.7}$$

Como se puede notar I_{cmax} se obtiene de I_{change} , con las siguientes condiciones:

- $T_1 = 1$, es decir, se considera que todos los valores son atípicos.
- $|I_2| = |I_3| = |I_4| = 1$, se considera que las potencias media, máxima y mínima están fuera de rango.

Aun así, I_{cmax} no alcanza el valor de 1, ya que los coeficientes de confianza (I_{C2} , I_{C3} e I_{C4}) siempre son menores que la unidad. Finalmente, el IoC se define mediante:

$$IoC = \frac{I_{change}}{I_{cmax}} \quad \text{Ecuación 4.8}$$

El IoC puede variar entre 0 y 1 luego del ajuste realizado. Alcanzar un valor de 1 no es usual, debido a que deben cumplirse muchas condiciones simultáneamente. Por otro lado, un valor de 0 indica que el día de análisis se ajusta perfectamente a las características del patrón de consumo. La experiencia adquirida en la aplicación del método sugiere que el valor obtenido por el IoC puede ser utilizado para catalogar el nivel de cambio que ha tenido un PCE diario con respecto a su patrón. Esta catalogación se realiza acorde a la Tabla 4.5.

Tabla 4.5 Catalogación del PCE diario usando los valores de IoC

IoC	Cambio en el consumo	Interpretación
0	Ninguno	El consumo de energía del día analizado se ajusta al patrón.
(0-0.03]	Muy pequeño	Pocos datos anómalos, los cuales no indican cambios significativos en el consumo de electricidad.
(0.03-0.15]	Notorio	Existen varios datos anómalos, pero siempre menores que el 40% de todos los datos. Estos datos anómalos no provocan cambios en la potencia máxima, media y mínima. Es posible que una carga haya permanecido conectada durante las HNL.
(0.15-0.3]	Notable	Muchos datos anómalos, posible carga conectada durante todo el día. No se evidencia gran diferencia en la energía diaria demandada con respecto al patrón. Sin embargo, los valores de potencia máxima y mínima pueden estar fuera el rango.
(0.3-0.55]	Grande	Los datos anómalos ocasionan que la energía diaria demandada esté fuera de rango con respecto al patrón.
(0.55-0.78]	Muy grandes	La presencia de valores anómalos durante gran parte del día ocasiona que la energía diaria demandada esté fuera de rango. Asimismo la potencia máxima o mínima es anómala.
(0.78-1]	Extremo	Todos los parámetros analizados están fuera de rango. El perfil de carga de electricidad es totalmente anómalo.

4.3 Influencia del tratamiento de la demanda de electricidad como serie temporal en la obtención de patrones y detección de anomalías

El efecto del crecimiento y la estacionalidad en la demanda de electricidad es crucial para la planificación y operación de los SSE. Su estudio permite la detección de fraudes, anomalías y fallas. No considerar la estructura de la serie de datos puede provocar sesgo en el análisis [70].

En este apartado se evalúa las bondades e influencia del tratamiento de la demanda de electricidad como una serie temporal en la obtención de patrones de consumo y detección de anomalías en PCE diarios. Para ello se utiliza el método SAEC propuesto en el Capítulo 3. Los resultados obtenidos por el nuevo método son contrastados respecto a dos de los métodos más extendidos en el tratamiento de las series temporales, como son la *eliminación de tendencia* [2] y el *filtro estacional* [71]. Posteriormente, se analizan los resultados, discutiendo las diferencias entre los patrones obtenidos y la detección de anomalías en los PCE.

El estudio de la literatura de los patrones de electricidad reveló cinco grandes áreas de investigación: i) clasificación, ii) detección de anomalías, iii) predicción, iv) administración de cargas y energía y v) predicción de precios. Estos estudios han utilizado numerosas herramientas y técnicas, sin embargo se detectaron cinco tipos de debilidades en su aplicación:

- a) No se considera la estacionalidad de los datos.
- b) No se toma en cuenta información detallada acerca del consumo de electricidad.
- c) No se analiza adecuadamente el efecto de la estacionalidad, por lo que se podría mejorar la metodología.
- d) La selección de los datos de entrenamiento requieren un esfuerzo considerable.
- e) Existen problemas de generalización del experimento ya que solo un caso de estudio es abordado.

Los estudios más significativos sobre la clasificación de perfiles de electricidad evidencian debilidades del tipo *a*, *b* o *c*. Esto se debe a que varias técnicas de agrupamiento no tienen en cuenta los datos en su contexto de series temporales por lo que el efecto de la estacionalidad se omite, tal como se evidencia en [7–10]. Algunos autores como Seem [69], [76] y Li et al., [77] en la clasificación y detección de anomalías utilizan solamente la energía media diaria y el valor pico de consumo siendo imposible realizar un análisis detallado del PCE. Por otro lado Capazzoli et al., [78] caracteriza la serie de tiempo de energía en ventanas de tiempo lo cual implica no detectar las anomalías de corta duración. En cambio Park y Son [72] presentan una metodología para analizar

perfiles de carga basados en procesamiento de imágenes, sin embargo la eliminación de valores atípicos carece de un procedimiento estadístico robusto. El comportamiento de la demanda de electricidad también se ha utilizado para predecir precios de la electricidad, por ejemplo Borovkova y Geman [79] y Janczura et al., [80] quienes de cierta manera consideran el efecto de la estacionalidad, sin embargo un análisis más profundo y con un adecuado enfoque podría obtener mejores resultados.

A lo largo del tiempo los investigadores han desarrollado nuevas propuestas para la predicción de la demanda de electricidad [13], [81]–[83], siendo las técnicas de inteligencia artificial las preferidas actualmente. Sin embargo, su aplicación acarrea ciertas dificultades, como por ejemplo las debilidades del tipo *a*, *b*, *d* y *e* identificadas previamente. Por ejemplo, el efecto de la estacionalidad en la fase de entrenamiento de estos métodos es difícil de establecer y generalmente la selección y procesamiento de los datos de entrenamiento requieren un esfuerzo considerable, a menudo requiriendo variables que también deben ser predichas, lo cual incrementa la incertidumbre de las predicciones. Además, con mucha frecuencia la aplicación de métodos de inteligencia artificial no garantizan que estos puedan ser usados en otros casos, derivando en problemas de generalización del método, tal es el caso presentado en [83].

Los trabajos revisados en el área de la administración de cargas y energía evidencian debilidades del tipo *a*, *b* y *e*. En algunos casos como en [84] y [85] la estacionalidad e información detallada acerca del PCE no se tiene en cuenta, reduciéndose de esta manera la utilidad del método para la toma de decisiones a corto plazo. En la Tabla 4.6 se muestra el resumen de la revisión del estado del arte de los patrones de consumo, su área de aplicación, herramientas y técnicas utilizadas y debilidades detectadas.

Tabla 4.6 Revisión de la literatura de patrones de consumo

Área de aplicación	Tipo de debilidad	Herramientas y técnicas utilizadas	Autor
Clasificación	b	Estadística y clustering jerárquico	Seem (2005) [69]
		Minería de datos, PSO-kmeans y máquinas de soporte vectorial	Cai et al., (2019) [86]
	a	Clustering jerárquico, k-means, k-means borroso, método de cuantificación vectorial adaptativa, algoritmo follow the leader, mapa auto-organizado, redes neuronales probabilísticas	Chicco (2012) [87]
		K-means [88] [89]	Khan et al., (2016) [88] Al-Jarrah et al., (2017) [89]
Clasificación y detección de anomalías	a, c	Procesamiento de imágenes	Park and Son (2019) [72]
	b, c	Análisis de variables canónicas	Li et al., (2010) [77]
Detección de anomalías	b	K-means y máquinas de soporte vectorial	Mora et al., (2018) [75]
		Estadística y clustering jerárquico	Seem (2007) [76]
	a, b	Proceso de aproximación simbólica agregada	Capazzoli et al., (2018) [78]
		C-means basada en clustering borroso	dos Angelos et al., (2011) [73]

Capítulo 4. Metodología estadística para valorar e identificar cambios en perfiles de consumo de electricidad (SAICC)

		Máquinas de soporte vectorial and k-means [74]	Jokar et al., (2016) [74]
	a, c	Redes neuronales LSTM and estadística [90]	Fenza and Gallo (2019) [90]
	a, d	Redes neuronales artificiales	Escrivá et al., (2011) [13] Serrano et al., (2017) [81]
Predicción	a, e	Regresión de soporte vectorial	Jain et al., (2014) [83]
	b	Minería de datos, clustering no supervisado y redes de predicción bayesiana	Singh and Yasine (2018) [82]
	a, b	Clustering jerárquico	Jota et al., (2011) [84]
Administración de cargas y energía	a, e	Formulación de un problema de mochila y resolución a través de programación dinámica.	Kalid et al., (2017) [85]
	b	Redes neuronales artificiales, lógica borrosa, sistema de interferencia borrosa neuronal adaptativa y optimización heurística.	Shareef et al., (2018) [91]
Predicción de precios de la electricidad	c	Modelo de curva hacia adelante estacional, cuantificación de estacionalidades por una prima determinista [79]	Borovkova and Geman (2006) [79]
		Filtro recursivo en precios o diferencia de precios o un modelo de estimación recursivo estacional	Janczura et al., (2013) [80]

Como se ha visto, mediante la revisión de la literatura de los patrones de consumo de electricidad se han identificado debilidades. En este sentido, en esta tesis, se ha propuesto el método SAEC para eliminar la tendencia y desestacionalizar los datos, cuyas características se pueden resumir en las siguientes:

- Aborda la estacionalidad de los datos a través de una metodología robusta.
- Trabaja con una resolución de datos cuarto horaria con el objetivo de permitir la detección de anomalías en periodos cortos.
- Es simple de aplicar.
- Es capaz de trabajar con cualquier tipo de perfiles de carga de electricidad o consumidor.
- Es capaz de identificar anomalías puntuales, colectivas y contextuales.
- Reduce en su conjunto la TFP y la TFN (tasa de falsos negativos), aumentando la precisión en la detección de anomalías.

La TFP se entiende como el porcentaje de PCE erróneamente reportados (por un método o herramienta) como anómalos dividido por el número total de PCE realmente no anómalos. Por otro lado la TFN es el porcentaje de PCE erróneamente reportados como no anómalos partido para el número de PCE realmente anómalos [55]. La Ecuación 4.9 y la Ecuación 4.10 las define:

$$TFP = \frac{FP}{FP+VN} 100\%, \quad \text{Ecuación 4.9}$$

$$TFN = \frac{FN}{FN+VP} 100\%, \quad \text{Ecuación 4.10}$$

en donde FP es el número de falsos positivos, VN es el número de verdaderos negativos, FN es la cantidad de falsos negativos y VP el número de verdaderos positivos. Los falsos positivos son PCE erróneamente reportados como anómalos y los falsos negativos son los PCE anómalos que erróneamente han sido reportados como no anómalos.

En la detección de anomalías existen tres criterios de evaluación adicionales que se deben tener en cuenta. El primero es la precisión, la cual se define como el porcentaje de valores anómalos reportados que verdaderamente resultan ser anómalos. El otro criterio es la sensibilidad, muy conocido en la literatura de habla inglesa como *recall*, este en cambio representa el porcentaje de verdaderos anómalos que han sido reportados como anómalos. Finalmente, la especificidad es el porcentaje de PCE que se reportan como no anómalos entre el número total de PCE no anómalos [55]. La TFP y la especificidad así como la TFN y la sensibilidad son complementarios, es decir que su suma es el 100%. Las siguientes ecuaciones definen cada uno de ellos.

$$\text{Precisión} = \frac{VP}{VP+FP} 100\%, \quad \text{Ecuación 4.11}$$

$$\text{Sensibilidad} = \frac{VP}{VP+FN} 100\%, \quad \text{Ecuación 4.12}$$

$$\text{Especificidad} = \frac{VN}{VN+FP} 100\%, \quad \text{Ecuación 4.13}$$

4.3.1 Metodología

En este apartado se presenta una comparación entre los patrones de consumo obtenidos y la identificación de anomalías bajo dos escenarios. El primero sin realizar ningún tratamiento previo en los datos y el segundo mediante el tratamiento de la demanda de electricidad con tres métodos distintos, *eliminación de tendencia*, *filtro estacional* y el método SAEC propuesto. Posteriormente, se analizan y discuten las diferencias entre los resultados obtenidos. El tratamiento de los datos se lleva a cabo en la subetapa de transformación de datos de la primera etapa de la metodología SAICC explicada en el apartado 4.2.1 (ver Imagen 4.11).

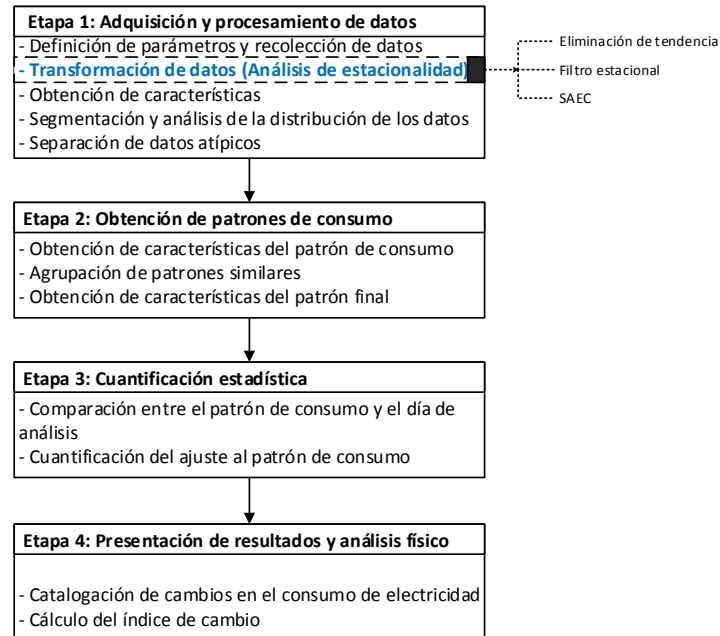


Imagen 4.11 Análisis de la estacionalidad de datos en la metodología SAICC

4.3.1.1 Método de eliminación de tendencia

Este método sustrae los valores de la línea de tendencia de la serie temporal a los datos originales, de tal manera que se obtiene un conjunto de datos con una media igual a cero. La mejor línea de ajuste se obtiene mediante la Ecuación 4.14 basada en el método de los mínimos cuadrados [92]:

$$\hat{p} = b + mt, \quad \text{Ecuación 4.14}$$

en donde t es el índice de tiempo en las unidades predefinidas, \hat{p} es el valor proyectado de la demanda de potencia p [kW] para un valor de t , en tanto que b es la intersección con el eje de las ordenadas, esto es el valor de p cuando $t = 0$. Finalmente, m es el valor de la pendiente de la línea o el cambio promedio de p por cada incremento de una unidad en t [2].

El método de los mínimos cuadrados consiste de los siguientes 4 pasos:

1. Se define la variable a analizar, en este caso la demanda de potencia p , la cual es variable en el tiempo t .

2. Se define el tamaño de los datos (n) y la frecuencia de muestreo con la que los datos son adquiridos (f). La variable t debe tener el mismo número n de datos que la potencia consumida (p).
3. Se calcula la suma de los n valores de p de acuerdo al tamaño de los datos y la frecuencia de muestreo definida.
4. Se calcula la pendiente (m) de la línea de tendencia, así como la intersección del eje de la variable p , esto es b de la Ecuación 4.14, por medio de la Ecuación 4.15 y Ecuación 4.16.

$$m = \frac{n \sum t \times p - \sum t \sum p}{n \sum t^2 - (\sum t)^2}, \quad \text{Ecuación 4.15}$$

$$b = \frac{\sum p - m \sum t}{n}. \quad \text{Ecuación 4.16}$$

4.3.1.2 *Filtro estacional*

El filtro estacional es un método estadístico basado en la observación de un periodo de datos finito, en donde la componente periódica de los datos se elimina, resultando una serie de datos desestacionalizada. El proceso es iterativo y puede ser ejecutado con datos de diferentes periodos. En este caso se analizan los datos de un año entero, por lo tanto el número de periodos es 1. El proceso se detalla a continuación [71]:

1. Se obtiene la potencia media de cada semana con los datos cuarto horarios disponibles.
2. Se calcula la media móvil con 13 términos, para eliminar la pérdida de información el primer y último valor de la media móvil obtenida se duplica.
3. Se obtiene un índice (descomposición multiplicativa) al dividir la potencia móvil de cada semana obtenida en el punto 1 entre la media móvil del punto 2.
4. Los datos originales cuarto horarios se dividen para el índice obtenido en el punto 3, consiguiendo así, una serie de datos desestacionalizada.

4.3.1.3 *Casos de estudio analizados*

Con el objetivo de analizar la influencia del tratamiento de la demanda de electricidad, en la obtención de patrones de consumo y detección de anomalías, se examina el consumo de electricidad de dos diferentes instalaciones. La primera instalación es el edificio 5E de la UPV en Valencia, España, cuyos datos fueron obtenidos a través del sistema DERD [30] (sistema de gestión implementado en toda la UPV que monitoriza los

consumos de todos los edificios del campus, el cual es del tipo presentado en el apartado 3.3 del Capítulo 3) desde el 29-11-2015 hasta el 28-11-2016 (ver Imagen 4.3)La segunda instalación corresponde a la acometida principal de la UPS en Cuenca, Ecuador con datos desde el 09-03-2017 hasta el 08-03-2018, los cuales fueron analizados previamente en apartado 3.5.2.2 del Capítulo 3. Estos dos casos de estudio resultan interesantes debido a que están en países distantes y sus patrones de consumo difieren significativamente.

En el apartado 4.2.1 se evidenció que la demanda de electricidad del Edificio 5E de la UPV cumple con las cuatro condiciones necesarias para llevar a cabo transformación de datos con el método SAEC (ver Imagen 4.3). En la Imagen 4.12 se muestra la potencia eléctrica media trimestral consumida en la UPS Sede Cuenca, se observa que los datos siguen una tendencia lineal positiva. La amplitud de los valores en cada periodo de cada ciclo es similar, incrementándose o disminuyendo de acuerdo a la línea de tendencia. De esta manera las cuatro condiciones mencionadas también se cumplen satisfactoriamente, por lo cual la transformación de datos se lleva a cabo para ambos casos. Si se observa en detalle la escala del eje vertical de las figuras, se comprueba que la componente estacional de la demanda de electricidad en la UPS es menor que la del caso en la UPV, esto se debe a la inexistencia de las 4 estaciones climáticas en zonas cercanas al ecuador.

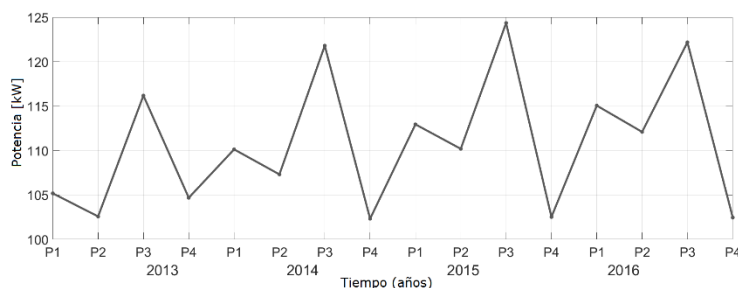


Imagen 4.12 Potencia media trimestral en la UPS (P1: febrero–abril, P2: mayo–julio, P3: agosto–octubre, P4: noviembre–enero)

4.3.2 Comparación de resultados

El objetivo de esta sección es evaluar la transformación de datos mediante el tratamiento de la serie temporal de datos en la obtención de patrones y la detección de anomalías en los PCE. Para ello se ha analizado el periodo de 365 días descrito en el apartado anterior, tanto para la UPV como para la UPS. Posterior a la obtención de los patrones y a la detección de anomalías se emplea la TFP, TFN, precisión, sensibilidad y especificidad como criterios de evaluación para cada uno de los siguientes métodos:

- Sin análisis de estacionalidad WSA (del término anglosajón without seasonality analysis).
- Eliminación de la tendencia DM (del término anglosajón detrending method).
- Filtro estacional SF (del término anglosajón seasonal filter).
- Método SAEC.

4.3.2.1 Patrones estocásticos obtenidos

En este apartado se presentan las diferencias entre los patrones obtenidos en cada método de análisis de estacionalidad y su interpretación acorde a la metodología SAICC. Si bien todos los días de la semana han sido analizados, a modo de ejemplo se presenta el caso de un día miércoles laborable para el edificio 5E de la UPV. En la Imagen 4.13 se muestran las medias del patrón obtenido por cada uno de los métodos. El patrón obtenido con el método WSA es similar al obtenido por el SF y el método SAEC. En HL el patrón obtenido por el método SAEC tiene valores levemente superiores a los obtenidos por los otros métodos. Por otro lado, el método DM obtiene un patrón desplazado con una media de cero.

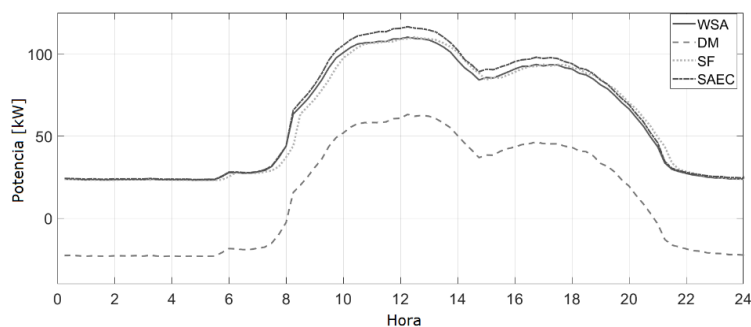


Imagen 4.13 Media de los patrones de consumo

La Imagen 4.14 muestra los valores de la desviación estándar del patrón de los días miércoles, se puede observar que no existen variaciones significativas cuando se comparan los métodos WSA, SF y DM. Por otro lado, cuando se aplica el método SAEC la desviación estándar de los datos disminuye considerablemente en las HL. Esto es relevante debido a que se consigue un patrón con menor variación, lo cual posibilita un aumento en la precisión al disminuir la TFP y TFN en la detección de anomalías.

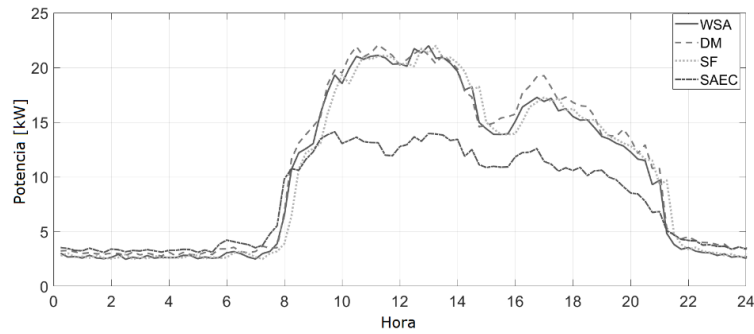


Imagen 4.14 Desviación estándar de los patrones de consumo

Los valores de Z_{max} y Z_{min} representan los límites superior e inferior de los intervalos de confianza de los patrones, respectivamente (un valor se considera anómalo cuando está fuera de este intervalo). La Imagen 4.15 e Imagen 4.16 muestran que los valores de Z_{max} y Z_{min} del filtro estacional son bastante similares a los obtenidos por el método WSA. El método DM obtiene intervalos de confianza más pequeños en las HL debido a su limitada capacidad para agrupar datos de días de consumos similares. Por otro lado, el método SAEC propuesto no siempre obtiene intervalos de confianza más pequeños. En conclusión, los valores de Z_{max} y Z_{min} por sí solos no revelan si un método es mejor que otro. Por este motivo la TFP, TFN y la precisión se consideran criterios de evaluación claves en este estudio.

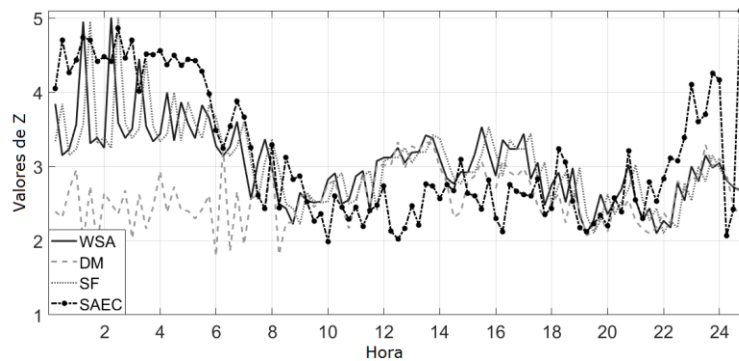


Imagen 4.15 Valores de Z_{max} de los patrones de consumo

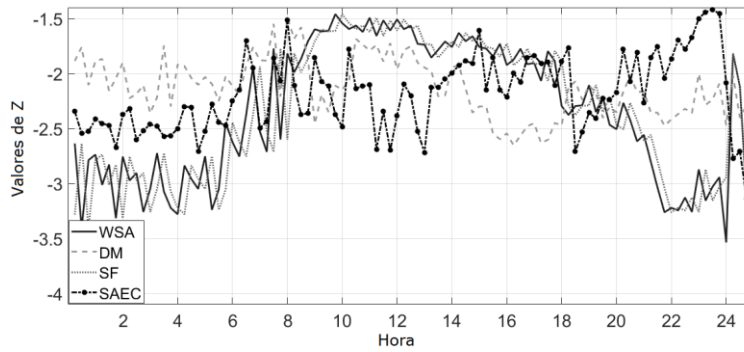


Imagen 4.16 Valores de Z_{min} de los patrones de consumo

Los patrones estocásticos obtenidos también pueden representarse con diagramas de cajas y bigotes (ver Imagen 4.17). Evidentemente, los métodos DM y SF producen patrones muy similares a WSA. Por otro lado, el método SAEC obtiene un patrón menos variable manteniendo un nivel de significancia similar, de tal manera que la detección de anomalías es más confiable. La bondad del método radica en aumentar la precisión incluso cuando el patrón tiene menor variación.

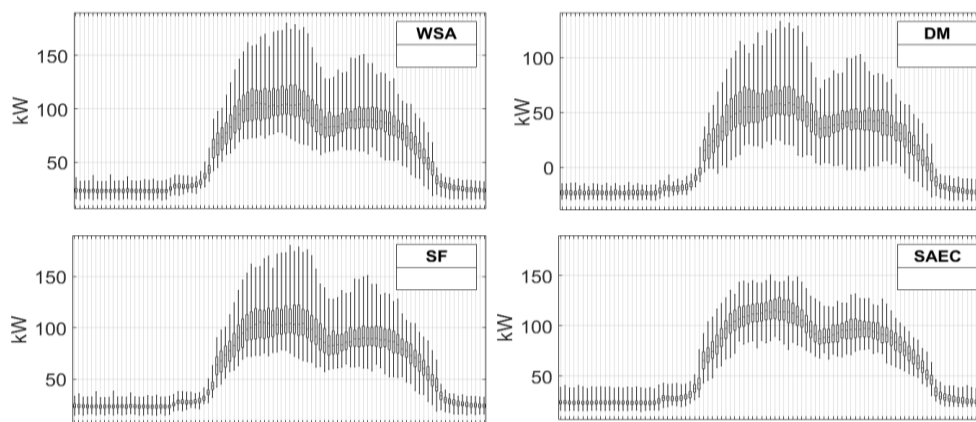


Imagen 4.17 Patrón de consumo de electricidad representado por diagramas de cajas y bigotes para un día miércoles laborable, Edificio 5E de la UPV

La aplicación del método SAEC aporta un patrón con menor variación incluso cuando la componente estacional de la demanda de electricidad es pequeña, tal como ocurre en la UPS Sede Cuenca (ver Imagen 4.18).

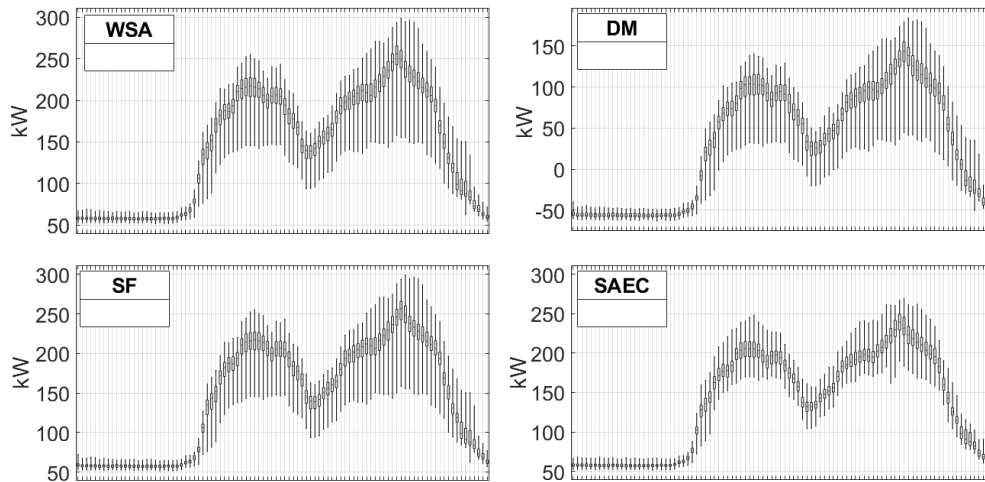


Imagen 4.18 Patrón de consumo de electricidad representado por diagramas de cajas y bigotes para un miércoles laborable, UPS

4.3.2.2 Detección de anomalías

Caso de estudio en el Edificio 5E de la UPV

Una vez definido el patrón de consumo, se lleva a cabo un análisis estadístico para detectar anomalías en un día de análisis. En la Tabla 4.7 se muestra el detalle de los patrones de consumo del edificio 5E para los días miércoles y viernes con los diferentes métodos propuestos. Los detalles incluyen el número de PCE reportados como típicos, el número de PCE reportados como anómalos, número de PCE anómalos reales (identificados por un humano experto), los PCE anómalos reportados como anómalos (VP), los PCE anómalos reportados como no anómalos, la TFP, TFN y la precisión.

Mediante el método SAEC se pueden identificar como típicos (VN) a PCE afectados por la estacionalidad que tienen un consumo de potencia lejano de la media en cada intervalo cuarto horario, los cuales hubiesen sido etiquetados como anómalos (FP). Esto se debe a que en el año hay semanas con una demanda de energía típicamente alta y otras semanas de bajo consumo. El método SAEC permite que este tipo de PCE no sea etiquetado como anómalo, reduciendo así de manera importante la TFP. Esto se evidencia en la Tabla 4.7 para el caso de los miércoles y en la Tabla 4.8 para los viernes.

Tabla 4.7 Evaluación de las anomalías detectadas en el edificio 5E de la UPV

Método	Clúster	PCE reportados como típicos (VN+FN)	PCE anómalos reportados (VP+FP)	PCE anómalos reales (VP+FN)	PCE anómalos (VP)	PCE anómalos (FN)	TFP [%]	TFN [%]	Precisión [%]
Patrón de consumo de los miércoles laborables									
WSA		28	14	9	8	1	18.2	11.1	57.1
DM	miércoles	29	13	9	7	2	18.2	22.2	53.8
SF		28	14	9	8	1	18.2	11.1	57.1
SAEC		33	9	9	7	2	6.1	22.2	77.8
Patrón de consumo de los viernes laborables									
WSA		28	13	17	9	8	16.7	47,1	69.2
DM	viernes	28	13	17	9	8	16.7	47,1	69.2
SF		28	13	17	9	8	16.7	47,1	69.2
SAEC		20	21	17	15	2	25.0	11,8	71.4

En comparación con los otros métodos, el método SAEC reporta con facilidad PCE anómalos cuando el consumo es alto en periodos donde la demanda energética es usualmente baja o viceversa (VP). En el caso del viernes laborable de la Tabla 4.7, con el método SAEC se reporta menor cantidad de PCE típicos, con lo cual se consigue un patrón con intervalos de confianza más pequeños. De esta manera, el número de PCE anómalos reportados por el método se incrementa, lo cual afecta de manera negativa la TFP. Sin embargo, la TFN baja del 47.1% al 11.8% y la precisión aumenta del 69.2% al 71.4%.

La aplicación de la transformación de datos con el método SAEC también permitió que el técnico experto catalogue como anómalo el consumo de 4 días viernes adicionales que no había percibido antes. En la Imagen 4.19 e Imagen 4.20 se muestra uno de los cuatro días viernes anómalos detectados y su patrón correspondiente obtenido con los métodos SF y SAEC respectivamente. En primera instancia el técnico experto indicó que la demanda de electricidad en ese día era típica debido a que el requerimiento energético fue bajo durante ese mes del año. Sin embargo, al comparar el PCE con el patrón obtenido con el método SAEC se perciben claramente anomalías entre la 1:30 a.m. y 3:30 a.m. y entre las 6:00 a.m. y 7:15 a.m., identificando de esta manera un incidente que no fue detectado anteriormente. La comparación se realizó únicamente entre

los métodos SF y SAEC debido a que los patrones obtenidos por los métodos WSA y DM son similares a los obtenidos por el SF.

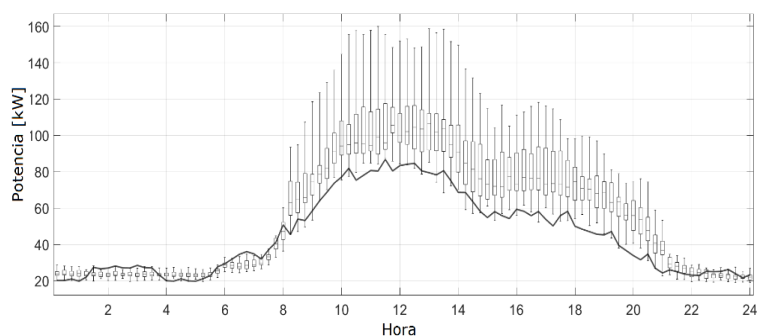


Imagen 4.19 Día analizado comparado con su patrón aplicando el método SF

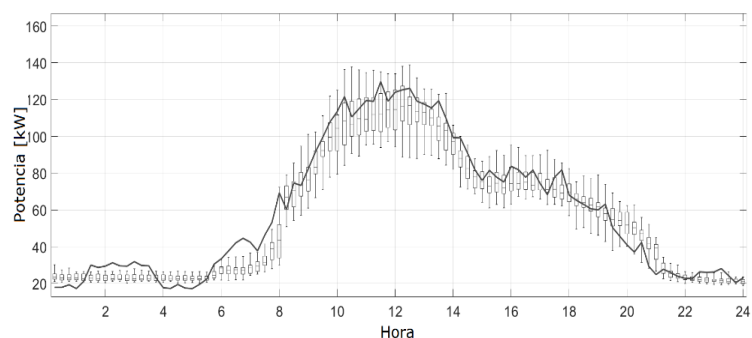


Imagen 4.20 Día analizado comparado con su patrón aplicando el método SAEC

Caso de estudio en la UPS Sede Cuenca

En la Tabla 4.8 se muestra el detalle de las anomalías detectadas en los días miércoles y viernes laborables en la acometida principal de la UPS.

La transformación de datos de la demanda de electricidad en la acometida principal de la UPS con el método SAEC reduce la cantidad de FN (PCE reportados como típicos cuando realmente son anómalos) para los días miércoles, tal como se muestra en la Tabla 4.8. En este sentido el método SAEC es capaz de identificar semanas de consumo normalmente bajas, mejorando de manera notable la precisión. Por otro lado, para el caso de los viernes no siempre es posible mejorar la TFN, sin embargo en este caso la TFP y la precisión mejoran considerablemente.

Tabla 4.8 Evaluación de las anomalías detectadas en la UPS

Método	Clúster	PCE reportados como típicos (VN+FN)	PCE anómalos detectados (VP+FP)	PCE anómalos reales (VP+FN)	PCE anómalos (VP)	PCE anómalos (FN)	TFP [%]	TFN [%]	Precisión [%]
Patrón de consumo de los miércoles laborables									
WSA	miércoles	35	16	17	12	5	11.8	29.4	75.0
DM		34	17	17	13	4	11.8	23.5	76.5
SF		35	16	17	13	4	8.8	23.5	81.3
SAEC		32	19	17	16	1	8.8	5.9	84.2
Patrón de consumo de los viernes laborables									
WSA	viernes	28	20	15	15	0	15.2	0.0	75.0
DM		28	20	15	15	0	15.2	0.0	75.0
SF		29	19	16	15	1	12.5	6.3	78.9
SAEC		33	15	15	14	1	3.0	6.7	93.3

4.3.3 Discusión y análisis de los resultados

Los resultados obtenidos sugieren que el método SAEC mejora la detección de PCE anómalos al diferenciar entre periodos de alta o baja demanda energética, identificando de esta manera a las anomalías contextuales, las cuales son difíciles de detectar. El método garantiza una mejora significativa de la precisión en la detección de anomalías cuando la demanda de electricidad cumple con los cuatro condiciones condiciones descritas en el apartado 4.2.1, subetapa E.1.3.

La cantidad de PCE anómalos reportados es susceptible de ser ajustada al variar el umbral de la puntuación de anomalías, en este método dicho ajuste se realiza al variar el tamaño de los intervalos de confianza detallados en el apartado 4.2.1. Así, intervalos grandes reportarán menor cantidad de anomalías, en tanto que, intervalos pequeños reportarán a más datos como anómalos.

La Tabla 4.9 muestra los resultados de la detección de anomalías en el Edificio 5E de la UPV para seis PCE, utilizando los 4 métodos de análisis referidos anteriormente. En los 4 primeros PCE seleccionados el método SAEC supera al resto de métodos al evitar los FP y FN en periodos en donde la demanda de energía es normalmente alta o baja (anomalías contextuales). Sin embargo, este método obtiene los mismos resultados que

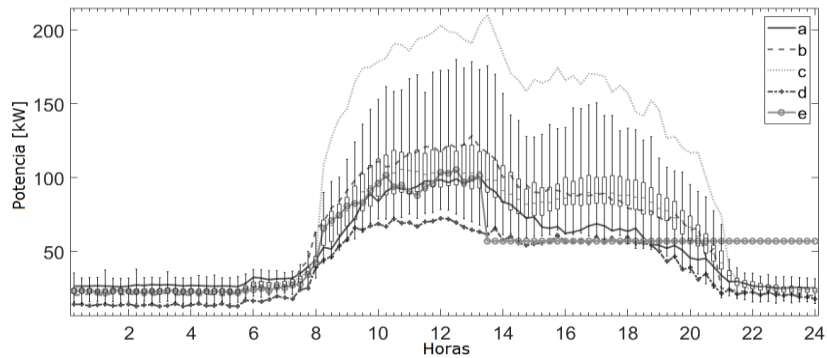
los otros cuando existen errores de medición o cuando el PCE tiene una forma anómala.

Tabla 4.9 Detección de anomalías para diferentes métodos de análisis de estacionalidad

Tipo de PCE diario	¿Es realmente anómalo?	¿Es anómalo con WSA?	¿Es anómalo con DM?	¿Es anómalo con SF?	¿Es anómalo con SAEC?
a – Anómalo de bajo consumo (bajo consumo en un periodo de consumo elevado)	Si	No (FN)	No (FN)	No (FN)	Si (VP)
b – Anómalo de alto consumo (alto consumo en un periodo de bajo consumo)	Si	No (FN)	No (FN)	No (FN)	Si (VP)
c – Consumo típico elevado (alto consumo en un periodo de elevado consumo)	No	Si (FP)	Si (FP)	Si (FP)	No (VN)
d - Consumo típico bajo (bajo consumo en un periodo de bajo consumo)	No	Si (FP)	Si (FP)	Si (FP)	No (VN)
e - El PCE tiene errores de medición (más de una hora con el mismo valor)	Si	Si (VP)	Si (VP)	Si (VP)	Si (VP)
El PCE tiene una forma anómala *	Si	Si (VP)	Si (VP)	Si (VP)	Si (VP)

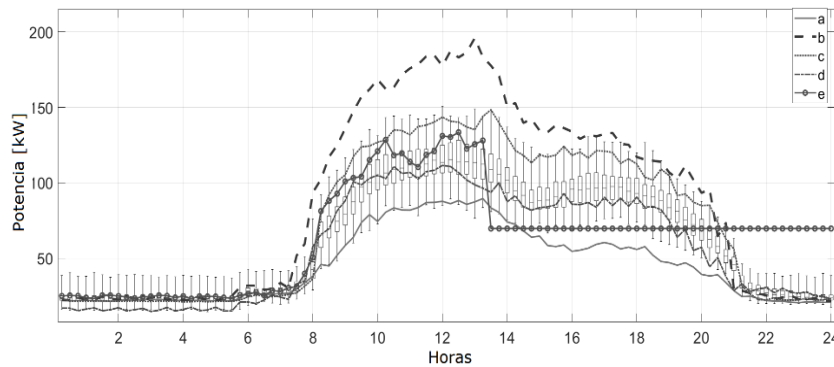
* Este PCE se mostró en la Imagen 4.20

La Imagen 4.21 e Imagen 4.22 muestran la diferencia de los métodos SF y SAEC en la detección de anomalías para cinco miércoles anómalos (cinco primeros PCE de la Tabla 4.9). Se presenta solamente la comparación entre el método SF y el SAEC debido a que los patrones de consumo obtenidos por el método WSA y DM son similares a aquellos obtenidos por el SF.



Nota: a) día de consumo anómalo bajo, b) día de consumo anómalo alto, c) día con alto consumo típico, d) día de bajo consumo típico y e) día con errores de medición

Imagen 4.21 Perfiles de consumo de electricidad anómalos de acuerdo al método SF



Nota: a) día de consumo anómalo bajo, b) día de consumo anómalo alto, c) día con alto consumo típico, d) día de bajo consumo típico y e) día con errores de medición

Imagen 4.22 Perfiles de consumo de electricidad anómalos de acuerdo al método SAEC

Un tratamiento adecuado de los datos mejora en la identificación de patrones y aumenta la precisión en la detección de anomalías en la demanda de energía eléctrica. Su aplicación en casos reales implica un avance en la vigilancia de sistemas preventivos y operacionales de electricidad, reduciendo el tiempo y esfuerzo tanto humano como computacional que involucra el análisis estadístico de los datos. Por ejemplo, un sistema que usa patrones para detectar anomalías en el consumo de electricidad podría prescindir del encendido de una alarma debido a un elevado consumo de energía en días calurosos, ya que en esa época del año las temperaturas y el consumo energético son generalmente altos.

Finalmente, es importante recalcar que la influencia del tratamiento de la serie temporal de datos de la demanda de electricidad en la obtención de patrones y en la detección

de anomalías es significativa, por este motivo merece especial atención. Los resultados obtenidos sugieren que al aplicar el método SAEC para transformar los datos, a más de identificar anomalías puntuales y colectivas se pueden identificar anomalías contextuales. Además, la precisión del método siempre mejora. En los casos en los que la TFP empeora, lo hace ligeramente, al mismo tiempo que la TFN mejora sustancialmente. Por otro lado, cuando la TFN empeora, lo hace ligeramente en tanto que la TFP mejora sustancialmente.

4.4 Conclusiones del capítulo

En el Capítulo 2 se identificó a la detección de anomalías como una herramienta fundamental para la vigilancia y administración de los sistemas eléctricos en general. Además, el conocimiento detallado de los patrones de consumo en una instalación o punto de suministro permite caracterizar la demanda para entre otros, detectar cambios y anomalías, identificar robos de energía y pérdidas debido a fallos.

En respuesta a las necesidades detectadas y como una contribución importante, en este capítulo se presenta una nueva metodología estadística denominada SAICC que tiene la capacidad de caracterizar la demanda de energía eléctrica mediante la identificación de patrones estocásticos de consumo para valorar e identificar cambios en PCE, así como detectar anomalías e indicar sus posibles causas. La aplicación de esta metodología puede derivar en beneficios en la gestión y vigilancia de los sistemas, tales como: generar alarmas, reducir costos en mantenimiento y aplicar medidas adecuadas de manera oportuna cuando se presenten incidentes. Al mismo tiempo puede reducir el tiempo y esfuerzo que los técnicos expertos invierten en el análisis de datos y evitar errores humanos.

Asimismo, se ha evaluado la aplicación del método SAEC propuesto en el Capítulo 3, especialmente diseñado para tratar datos de potencia eléctrica activa absorbida. El método elimina la tendencia y desestacionaliza la demanda de electricidad permitiendo identificar, a más de anomalías puntuales y colectivas a las anomalías contextuales, al reconocer periodos de consumo normalmente alto o bajo. En este capítulo también se ha verificado que el método SAEC propuesto proporciona patrones de consumo menos variables. Todos estos aspectos mejoran la precisión en la detección de anomalías reduciendo de forma global la cantidad de FP y FN.

Finalmente, en la metodología SAICC también se definió un índice que valora de 0 a 1 los cambios que ha tenido un PCE respecto a su patrón. La experiencia adquirida en la aplicación del método sugiere que el valor del índice obtenido puede emplearse para clasificar o catalogar los perfiles de consumo de electricidad.

Con el objetivo de validar la metodología SAICC propuesta, en el siguiente capítulo se realizan varios ejemplos de aplicación en diferentes tipos de consumidores y niveles del SSE. De esta manera se pone a prueba la funcionalidad de la metodología al aplicarse en casos reales.

Capítulo 5

Aplicaciones de la Metodología SAICC

5.1 Introducción

Hoy en día existe una tendencia mundial a electrificar los diferentes sectores, por lo tanto cualquier esfuerzo es válido cuando se intenta mejorar la eficiencia de las redes y de los sistemas eléctricos. Las redes eléctricas han evolucionado tecnológicamente en muchos aspectos, uno de ellos se refiere a la disponibilidad de datos de la demanda en diferentes puntos de la red, tales como transformadores en redes de transmisión, alimentadores, transformadores de distribución y en los grandes y pequeños consumidores. Como se ha comentado en capítulos anteriores esta información constituye un recurso de partida fundamental para la gestión y planificación de los sistemas eléctricos.

En los siguientes apartados del capítulo se muestran diversas aplicaciones de la metodología SAICC presentada en el Capítulo 4. Luego de caracterizar la demanda en diferentes puntos y niveles de la red eléctrica se exponen los beneficios que pueden obtenerse en la gestión de cada uno de esos sistemas. En este sentido se trabaja con una muestra de datos real y representativa, lo cual evita posibles sesgos en el análisis y problemas de generalización de la metodología.

En primer lugar, la metodología se aplica al edificio 5E de la UPV en Valencia, España. Seguidamente, se estudia el consumo de electricidad de un conjunto de edificios, al analizar los datos del consumo de la UPS Sede Cuenca en Ecuador. La metodología SAICC se presenta como una herramienta eficaz capaz de facilitar la gestión de la demanda de electricidad en pequeños o grandes consumidores. La metodología propuesta no sólo resulta útil para analizar la demanda de los consumidores, también puede aplicarse en otros niveles de la red eléctrica. En la presente tesis, se presenta el análisis de la demanda de electricidad en dos alimentadores de diferentes subestaciones de la ciudad de Cuenca en Ecuador. El primer alimentador se localiza en una zona residencial y

comercial, y el segundo en una zona industrial. Posteriormente, el análisis se extiende a otro nivel, en una de las líneas de transmisión eléctrica más importantes de Ecuador, en este caso la Molino-Pascuales. Por último, la metodología SAICC propuesta, a más de analizar perfiles de carga, se utiliza para predecir la demanda de electricidad desde una perspectiva distinta a los estudios tradicionales. Los métodos tradicionales predicen valores de potencia puntuales específicos, intentando minimizar el error, mientras que en el caso presentado se obtiene un intervalo de predicción con una probabilidad asociada a que un valor real caiga dentro del intervalo, cuantificando de esta manera la incertidumbre de la predicción.

5.2 Aplicación de la metodología SAICC en el Edificio 5E de la UPV

Los edificios consumen aproximadamente el 40% de las energías primarias de todo el planeta [93]. En los países desarrollados se estima que los edificios demandan entre el 20% y el 40% de la energía final [94], [95]. En Estados Unidos por ejemplo, los edificios consumen aproximadamente el 40% de toda la energía, de la cual el 68% es electricidad y son responsables de entre el 38% y el 41% de todas las emisiones de CO₂. Por otro lado en Europa, las edificaciones consumen el 40% del total de energía y emiten el 35% del CO₂ [96]–[99].

El aumento en la demanda de electricidad en edificaciones es evidente. En la actualidad una mayor demanda de energía se asocia a problemas tales como emisiones de CO₂ y otros gases contaminantes, necesidad de explotación de recursos energéticos en detrimento del medio ambiente, incremento de los costos de la energía, dificultades para cubrir la demanda en horas pico, entre otros. El considerable consumo de energía en edificios ha motivado la creación de leyes, regulaciones y normativas relacionadas a la eficiencia energética, actividades educativas, certificación de edificios, etc. [100]. Estudios recientes indican que utilizando estrategias adecuadas de operación y gestión se puede ahorrar entre el 20% y 30% de la energía, sin realizar cambios en la estructura y configuración del hardware del edificio [101]. Según Escrivá [102], existen acciones básicas para mejorar la eficiencia energética en edificaciones, entre las cuales están medición y almacenamiento de datos del consumo de energía, planificar los diferentes procesos, monitorización automática de la demanda eléctrica, elegir un profesional responsable del uso de la energía, realizar acciones proactivas para incrementar la eficiencia, modificar las instalaciones para un manejo sencillo, mejorar la comunicación entre el responsable del uso de energía y los usuarios.

Una de las propuestas más desafiantes para reducir la demanda de energía en edificios es el concepto de edificios de consumo cero ZEB (del término anglosajón zero energy buildings). En Estados Unidos se pretende que para el 2050 todos los edificios comer-

ciales sean ZEB. Por otro lado, en la Unión Europea se ha planteado que desde finales del 2020 todos los edificios nuevos tengan un consumo de energía cercano a cero NZEB (del término anglosajón nearly zero energy building) [103]. Es difícil disminuir a casi cero los requerimientos energéticos de una edificación, por lo que un NZEB se basa en mejorar el desempeño energético de la instalación y además generar su propia energía mediante microrredes de generación. Este tipo de tecnología busca principalmente ahorro económico y reducción de emisiones de CO₂ [101], [104].

La mejora considerable del desempeño energético es sólo posible con información detallada acerca del consumo del edificio. Esta información es cada vez más accesible gracias al uso de sistemas de MI como los presentados en el Capítulo 3, cuyos costos han ido disminuyendo paulatinamente en los últimos años. Un adecuado análisis de estos datos provee información relevante para implementar acciones de mejora. En este contexto, la metodología SAICC presentada en esta tesis pretende contribuir en la mejora de la gestión de la energía en edificios mediante la caracterización de la demanda.

En este apartado la metodología SAICC se aplica a los datos de consumo de electricidad del edificio 5E de la UPV (Imagen 5.1). El edificio 5E tiene un total de 9651 m² de construcción en el cual existen oficinas, laboratorios y aulas de clase. La energía se utiliza principalmente en iluminación, circuitos de fuerza que alimentan ordenadores, impresoras, equipos de laboratorio, finalmente un gran porcentaje de energía corresponde a los equipos de climatización [105]. El análisis de este tipo de edificaciones es importante ya que un edificio de oficinas requiere por metro cuadrado entre 10 y 20 veces más energía que uno destinado a vivienda [94]. A continuación se detalla cada una de las etapas de la metodología SAICC aplicadas en este caso de estudio.



Imagen 5.1 Edificio 5E de la UPV

5.2.1 Etapa 1: Adquisición y procesamiento de datos históricos

Los datos son adquiridos del contador m0560 del Sistema DERD, cuya ubicación en el diagrama unifilar puede verse en la Imagen 5.2.

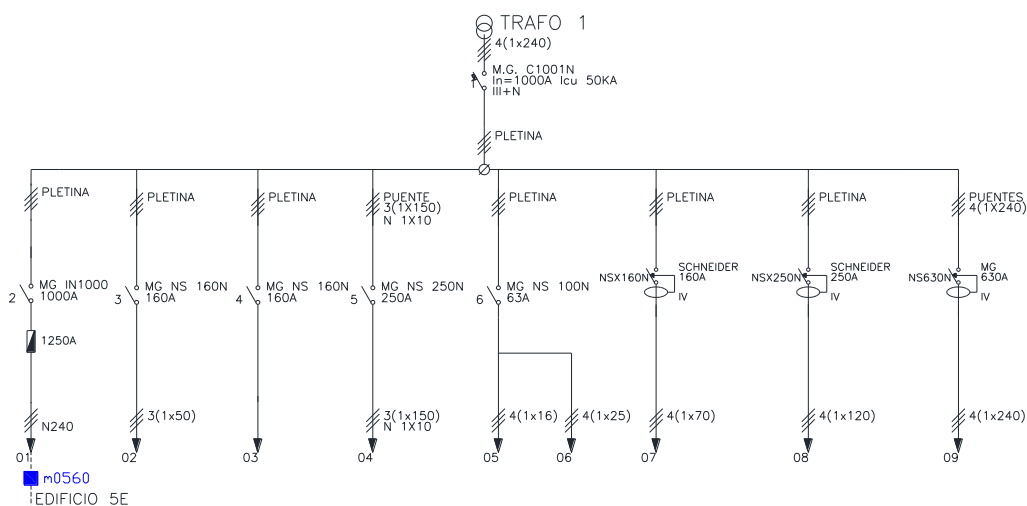


Imagen 5.2 Diagrama unifilar de baja tensión del centro de transformación que alimenta al edificio 5E de la UPV

En la primera etapa de la metodología SAICC se definen los siguientes parámetros:

- Periodo de análisis: 365 días (1 de julio de 2015 hasta 30 de junio de 2016).
- Día de la semana: se analizan todos los días de la semana.
- Laboralidad: días laborables.

El periodo de análisis es de un año, por lo cual se realiza la transformación de los datos mediante el método SAEC propuesto en el Capítulo 3. Si bien, todos los días fueron considerados, a continuación se presenta el análisis realizado para el tipo de día jueves laborable. Posteriormente, tal como se explicó, la metodología obtiene las características relevantes de los datos, los segmenta, analiza su distribución de probabilidad, separa los días con consumos anómalos y finalmente mediante la extracción de características define su patrón estocástico de consumo, caracterizando de esta manera la demanda de electricidad.

5.2.2 Etapa 2: Obtención de patrones de consumo

En la segunda etapa se toman como valores de partida solamente los datos típicos del consumo de electricidad correspondiente a cada día de la semana con el fin de extraer sus características. Luego, se agrupan los patrones similares al día de análisis determi-

nado. En la Imagen 5.3 se puede ver el patrón de consumo de los días jueves representado mediante un diagrama de caja y bigotes, además se pueden ver 6 días de análisis (todos ellos jueves laborables), en los que se analiza el consumo y se valora si han sido anómalos o no. Los días de análisis 1-4 y 6 corresponden a días con consumos atípicos, mientras que el día de análisis 5 corresponde a un día con consumos normales.

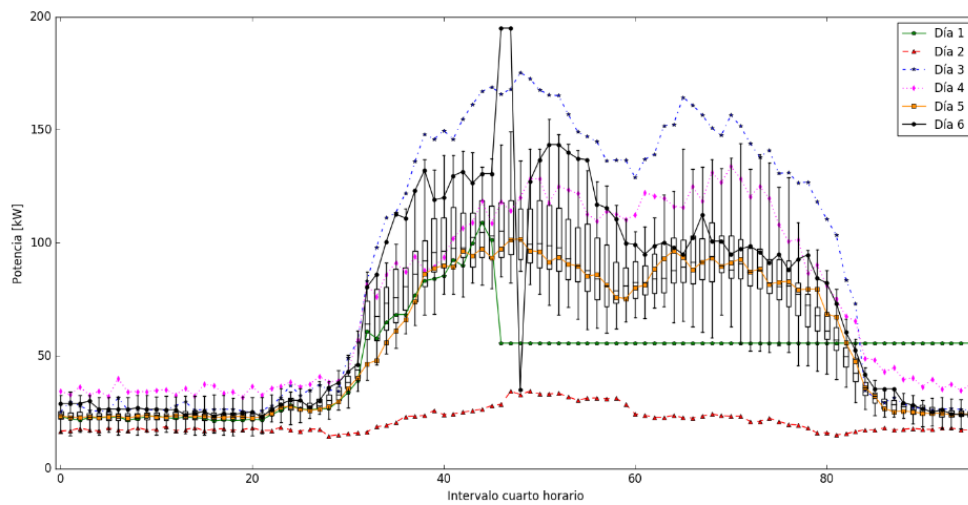


Imagen 5.3 Patrón y perfiles de consumo de electricidad de 6 días de análisis

En la metodología SAICC para apreciar de mejor manera la información presentada se estandarizan los datos, es decir, se sustituye la variable original (potencia activa media absorbida) por la variable Z . De esta manera, el patrón de consumo ahora se representa por el vector Z_{min} y Z_{max} , tal como se ve en la Imagen 5.4.

Caracterización de la demanda de energía mediante patrones estocásticos en las Redes Eléctricas Inteligentes

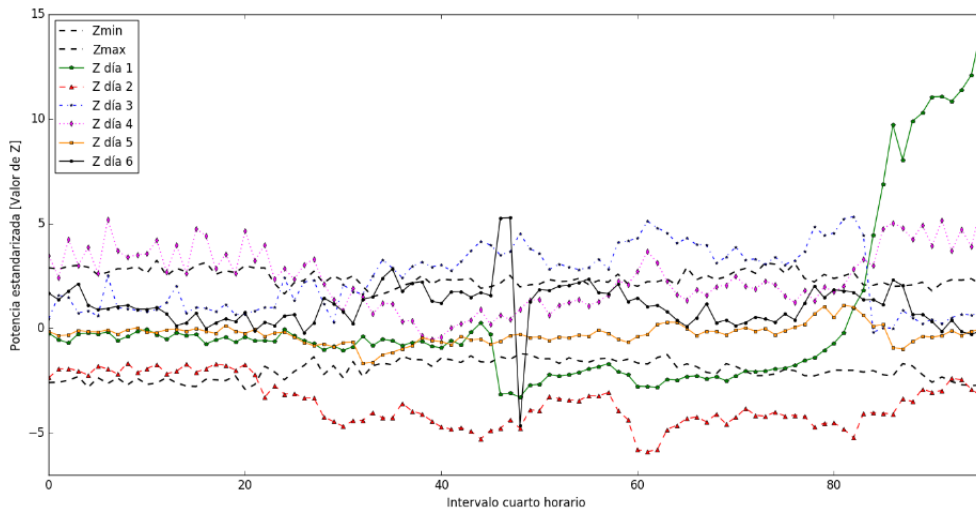


Imagen 5.4 Patrón de consumo (Z_{min} y Z_{max}) y los valores de Z para cada uno de los 6 días de análisis

En la Imagen 5.5 se pueden ver 3 características adicionales (potencia máxima, media y mínima) de los días de análisis y su correspondiente patrón representado como diagrama de caja y bigotes.

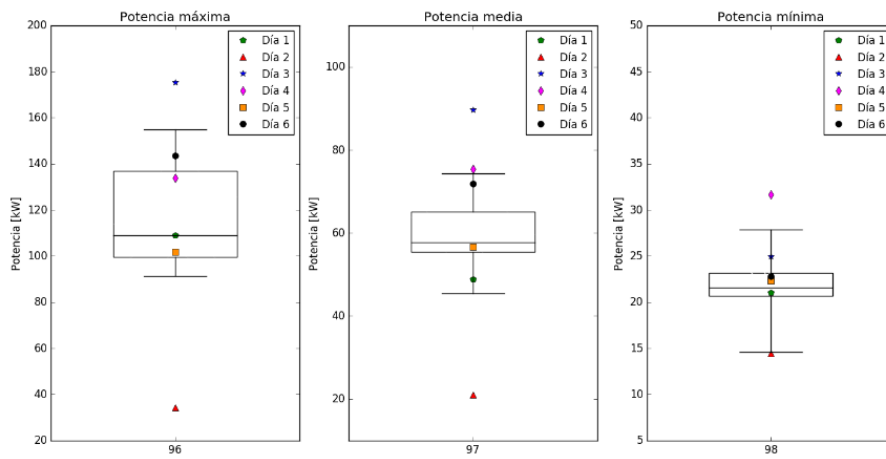


Imagen 5.5 Valores de potencia máxima, media y mínima para los 6 días de análisis y su correspondiente patrón

5.2.3 Etapa 3: Cuantificación estadística

En la tercera etapa de la metodología se lleva a cabo una cuantificación estadística del ajuste del día de análisis al patrón obtenido. Con el propósito de mostrar el procedimiento seguido, en la Tabla 5.1 se presentan los resultados de la cuantificación estadística del día de análisis 4. En la primera columna se representan cada una de las 99 características (96 intervalos cuarto horarios y la potencia media, máxima y mínima), en la siguiente columna se indica el intervalo cuarto horario correspondiente. Las variables de las columnas restantes se definieron en el Capítulo 4 cuando se presentó la metodología SAICC.

Tabla 5.1 Resultados de la cuantificación estadística para el día de análisis 4

Valor N°	Hora	Z _{min}	Z _{max}	Z _i	Coefficiente de confianza	Índice de atípicos	Error Tipo I	Error Tipo II
1	0:00-0:15	-2.599	2.881	3.447	0.9933	1	0.007	0.000
2	0:15-0:30	-2.580	2.821	2.433	0.9927	0	0.007	0.025
3	0:30-0:45	-2.493	2.953	4.243	0.9921	1	0.007	0.000
...
96	23:45-24:00	-2.7253	2.074	5.841	0.9777	1	0.022	0.000
97	-	-1.246	1.918	0.867	0.8661	0	0.134	0.000
98	-	-1.965	2.020	2.163	0.9536	1	0.046	0.000
99	-	-2.644	2.337	3.768	0.9862	1	0.014	0.000

5.2.4 Etapa 4: Presentación de resultados y análisis físico

En esta etapa de la metodología SAICC se llevan a cabo 12 pruebas (tests) con el objetivo de realizar el análisis físico. Con el resultado de estos tests se calcula el IoC para valorar el cambio del PCE con respecto al patrón del día en análisis. En la Tabla 5.2 se muestran los resultados de las pruebas aplicadas al día de análisis 4.

Tabla 5.2 Resultados de las 12 pruebas aplicadas al día de análisis 4

Prueba	Valor	Resultados de la prueba
Prueba 1: Total atípicos	42	43.8% de cambio en el consumo.
Prueba 2: Atípicos en HNL	37	92.5% de cambio en el consumo en HNL.
Prueba 3: Atípicos [0-7]	25	89% de cambio en el consumo en la madrugada.
Prueba 4: Atípicos [21-24]	12	100% de cambio en el consumo en HNL durante la noche.
Prueba 5: Atípicos [7-21]	5	8.93% de cambio en el consumo en HL.
Prueba 6: Atípicos (7-14)	0	0% de cambio en el consumo en HL durante la mañana.
Prueba 7: Atípicos (16-21)	2	10% de cambio en el consumo en HL durante la tarde y noche.
Prueba 8: Valores iguales consecutivos	0	0 horas de valores consecutivos iguales.
Prueba 9: Atípicos consecutivos	14	3.5 horas de consumo anómalo continuo.
Prueba 10: Potencia media	Mayor	El consumo de energía es mayor al del patrón. Coeficiente de confianza: 0.9536
Prueba 11: Potencia máxima	No	Potencia máxima demandada está dentro del rango del patrón de consumo. Coeficiente de confianza: 0.8661
Prueba 12: Potencia mínima	Mayor	Potencia mínima demandada es mayor a la del patrón de consumo. Coeficiente de confianza: 0.9862

Adicionalmente, en la Tabla 5.3 se muestra un resumen de la aplicación de las 12 pruebas a los 6 días de análisis considerados.

Tabla 5.3 Resumen de las pruebas aplicadas a los días de análisis

Día	T1 (%)	T2 (%)	T3 (%)	T4 (%)	T5 (%)	T2/T5	T3/T4	T8 (h)	T9 (h)	I2	I3	I4
1	38.5	30	0	100	44.6	0.67	0	12.5	6.25	0	0	0
2	75	40	21.4	83.4	100	0.40	0.26	0	17.5	-1	-1	-1
3	53.1	2.5	3.6	0	89.3	0.03	inf	0	10	1	1	0
4	43.8	92.5	89.3	100	8.9	10.4	0.89	0	3.5	1	0	1
5	0	0	0	0	0	0	0	0	0	0	0	0
6	8.3	5	0	16.7	10.7	0.46	0	0.50	0.75	0	1	0

* T1, T2, T3, T4 T5, T8 y T9 corresponden a las pruebas descritas en la Tabla 4.3; I2, I3 y I4 son descritas en la ecuación 4.6 del Capítulo 4.

Las conclusiones son generadas automáticamente por la metodología SAICC mediante el uso de la Tabla MCIAT. En la Tabla 5.4 se muestran estas conclusiones para los 6 días de análisis, además se incluye el diagnóstico de un técnico experto con el objeto de contrastar los resultados.

Tabla 5.4 Conclusiones del consumo de electricidad de los 6 días de análisis

Día de análisis	Conclusiones de la metodología SAICC	Diagnóstico del experto	Mejor interpretación	IoC
1	Error de medición. Más de 1 hora con el mismo valor de medición.	Error de medición debido a que se tienen valores consecutivos iguales.	Los dos entregan la misma información.	0.13
2	Existe mayor incidencia de datos atípicos en las HNL de la noche. Posible carga se quedó conectada en la noche. Existe mayor incidencia de consumos anómalos en las HL del día. La energía consumida es menor que la del patrón de consumo. Es probable que haya disminuido la actividad laboral. Puede haber ocurrido una falla, un mantenimiento o una desconexión de un circuito.	El día no fue laborable, este fue festivo.	Los dos dan información similar.	0.91
3	Existe mayor incidencia de consumos anómalos en las HL del día. La energía consumida es mayor que la del patrón de consumo. Es probable que haya incrementado la actividad laboral. La potencia máxima consumida es mayor que la del patrón. Es probable que la actividad laboral se haya incrementado.	Consumo elevado durante todo el día. Día con temperaturas extremas.	Los dos dan información similar.	0.66
4	El consumo es atípico en todas las mediciones de la noche, posible carga se quedó conectada. Existe mayor incidencia de datos atípicos en las HNL del día. La energía consumida es mayor a la del patrón. Probablemente algo ha sido conectado. La potencia mínima es mayor a la del patrón. Probablemente algo ha sido conectado o existe una falla.	Algo conectado en la noche. Alta temperatura ambiente.	El método propuesto brinda información adicional.	0.64
5	Día totalmente normal. Se ajusta al patrón de consumo.	Día con consumo normal de energía.	Los dos entregan la misma información.	0
6	Existen menos del 10% de datos anómalos; sin embargo, la potencia máxima es más alta que la del patrón. El consumo adicional se extiende en horas pico. El consumo analizado tiene un 5% de anomalías en HNL y un 10.7% en HL, lo cual no indica cambios significativos en el consumo de electricidad.	Problemas en la medición.	El experto define de mejor manera el problema.	0.20

Al analizar la Tabla 5.4 se evidencia que en 4 de los 6 casos, la metodología propuesta y el técnico experto entregan conclusiones con información similar, detectando por ejemplo errores de medición (día de análisis 1), bajo (día de análisis 2) o elevado consumo de energía (día de análisis 3) y por supuesto, identificando un consumo habitual de energía, en donde el IoC es igual a cero. En el día de análisis 4, la metodología propuesta entrega información de las anomalías detectadas con mayor detalle que el experto. Por otro lado, la metodología detecta un consumo atípico en el día de análisis 6, sin embargo tiene dificultades para interpretar la naturaleza de las anomalías, por lo que, en este caso el humano experto interpreta de mejor manera el consumo atípico presentado.

5.2.5 Análisis de los resultados

Para validar la metodología propuesta, los patrones obtenidos con un año de datos se utilizaron para analizar el PCE de 6 meses (01-06-2016 hasta 31-12-2016). En este periodo (180 días), 30 días fueron catalogados como anómalos por el método SAICC, en donde sus conclusiones fueron comparadas con el diagnóstico del técnico experto que administra el edificio 5E. En 19 de los 30 casos el método propuesto y el experto entregan resultados similares, en 9 casos el método propuesto entrega mayor información acerca de las posibles causas del consumo anómalo, mientras que en 2 casos el experto ofrece una mejor interpretación de los cambios del PCE con respecto a su patrón. Es importante recalcar que el experto requiere tiempo y esfuerzo para analizar la información, además es susceptible de cometer errores, por ejemplo elegir un patrón incorrecto para el día de análisis. Por el contrario, la metodología propuesta realiza el análisis de forma automática.

La metodología SAICC caracteriza la demanda de electricidad mediante la identificación de patrones estocásticos de consumo de acuerdo al periodo de datos históricos seleccionado, detectando PCE anómalos. Además, valora cambios en la demanda de energía diaria mediante un IoC y presenta las posibles causas de esos cambios con el uso de la Tabla MCIAT.

La metodología propuesta tiene la capacidad de aportar con los patrones de consumo en un sistema de monitorización en tiempo real para un edificio individual como el 5E. Su utilización puede ayudar a reducir costes y consumo de energía, detectar anomalías y fallas rápidamente, ahorrando tiempo que los administradores de energía y técnicos dedican al análisis de datos. Esta metodología también puede ayudar a identificar acciones de mejora para optimizar el consumo energético y valorar si las políticas de ahorro implementadas son efectivas.

5.3 Aplicación de la metodología SAICC en la Universidad Politécnica Salesiana Sede Cuenca

El consumo energético en países en vías de desarrollo se incrementa rápidamente, tanto es así que se espera que se duplique en los próximos 25 años [93]. En la región andina ecuatorial la situación no es muy diferente. Colombia, Perú y Ecuador actualmente tienen en conjunto una población de aproximada de 100 millones de personas. Entre 1990 y 2016 en estos países la población se ha incrementado un 46%, el consumo energético un 80%, las emisiones de CO₂ un 120% y finalmente la demanda de electricidad ha aumentado en un 207%. Para este mismo periodo en Ecuador los incrementos han sido mayores, la población lo ha hecho en un 61.7%, la demanda energética un 108%, las emisiones de CO₂ un 163% y el consumo de electricidad un 378% [1].

En Ecuador según Evans [106] los edificios demandan el 19% de la energía total del país. La *Agencia Internacional de la Energía* [1] indica que la demanda energética en el sector residencial, comercial y de servicios para Ecuador, Perú y Colombia es del 16%, 20% y 21% respectivamente. A diferencia de otras latitudes, la zona cercana al ecuador carece de periodos estacionales con grandes variaciones en la temperatura, esto ocasiona que el consumo de energía eléctrica tenga una componente estacional débil. Sin embargo, en estos países la caracterización del consumo de energía en edificaciones se ha tratado de manera escasa, por ello su estudio representa un valioso aporte.

En los siguientes apartados se aplica la metodología SAICC a los trece edificios de la UPS Sede Cuenca en Ecuador en su conjunto. La UPS ocupa un área aproximada de 8 hectáreas, cuenta con cuatro edificios con aulas de clase y auditorios, dos edificios de oficinas, un data center, una iglesia, un patio de comidas, un polideportivo, un edificio para la biblioteca general, un taller automotriz y un edificio de 25.000m² de uso mixto conformado por oficinas administrativas, salones de clase, departamento médico, laboratorios de usos múltiples, talleres de electricidad y de ingeniería mecánica y un teatro con capacidad para 510 personas. El caso de estudio presentado determina si la metodología es útil para analizar un grupo de edificaciones de diferentes usos. Los datos fueron adquiridos gracias al sistema de MI detallado en el Capítulo 3.

5.3.1 Etapa 1: Adquisición y procesamiento de datos históricos

Los datos de consumo de electricidad adquiridos corresponden al medidor de la acometida principal de la UPS, detallado en el apartado 3.3.2. El periodo de análisis es de 365 días, desde el 28-05-2017 hasta el 27-05-2018, analizándose todos los días de la semana, siendo de lunes a sábado laborables y los domingos no laborables.

El consumo de electricidad en la UPS muestra una componente estacional débil. Sin embargo, incluso en estos casos el método SAEC proporciona patrones menos variables, aumentando la precisión en la detección de anomalías, tal como se indicó en el capítulo anterior (apartado 4.3.2.1). Consecuentemente, en este caso también se lleva a cabo la transformación de datos.

5.3.2 Etapa 2: Obtención de patrones de consumo

Las características (media y la desviación estándar) de los patrones estocásticos definidos para cada día de la semana se muestran en la Imagen 5.6 e Imagen 5.7. Estas revelan una carga base un poco inferior a los 60kW en las horas de la madrugada y una potencia activa media absorbida pico de casi 250kW entre las 18:00 y 19:00 horas en los días laborables. La desviación estándar por su parte indica que el día con mayor variabilidad en la demanda de energía es el viernes con hasta 20kW, en tanto que el de menor variabilidad es el domingo con menos de 5kW.

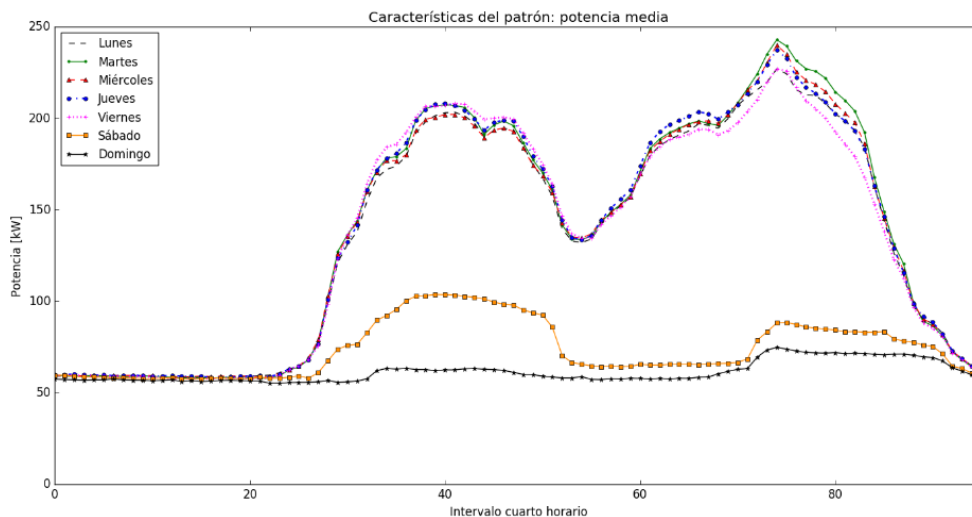


Imagen 5.6 Media del patrón de cada día de la semana en la UPS Cuenca (28-05-2017 hasta 27-05-2018)

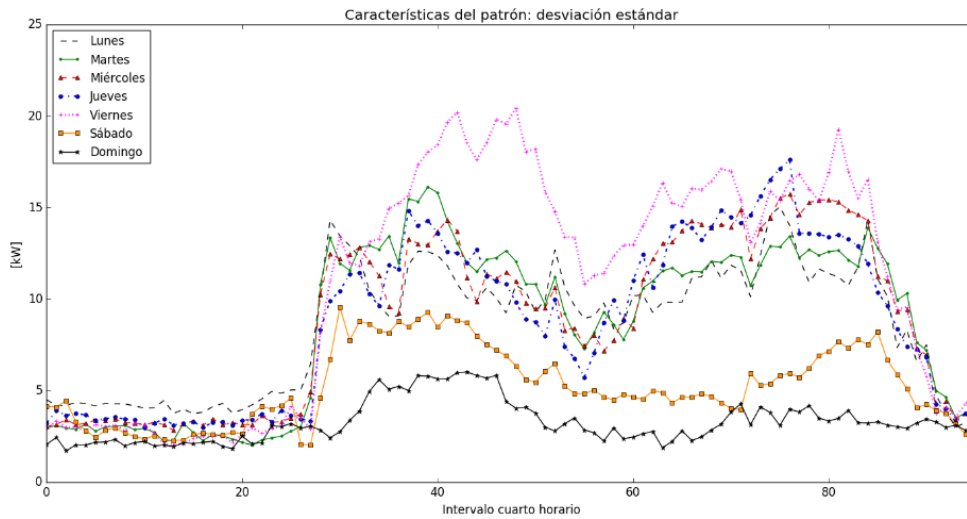


Imagen 5.7 Desviación estándar del patrón de cada día de la semana en la UPS Cuenca (28-05-2017 hasta 27-05-2018)

La conformación de patrones de electricidad para cada día de la semana se detalla en la Tabla 5.5. En dicha tabla se detalla el día de la semana, su laboralidad, los días con los que se agrupa debido a su similaridad, el número de PCE considerados típicos en ese día en concreto, el número de PCE atípicos y el número total de PCE considerados para conformar el patrón definitivo debido a la agrupación realizada automáticamente.

Tabla 5.5 Conformación de los patrones de electricidad por día de la semana en la UPS Cuenca

Día de la semana	Laboralidad	Días con los que se agrupa	Número de PCE típicos	Número de PCE atípicos	Número total de PCE en el patrón definitivo
lunes	Laborable	miércoles-jueves	33	15	94
martes	Laborable	miércoles-jueves	29	21	90
miércoles	Laborable	lunes-martes-jueves	33	18	123
jueves	Laborable	lunes-martes-miércoles	28	23	123
viernes	Laborable	-	30	18	30
sábado	Laborable	-	21	26	21
domingo	No Laborable	-	22	30	22

Al aplicar el método SAEC los patrones son menos variables incluso cuando el número de PCE típicos que conforman el patrón de cada día de la semana es mayor. La Imagen 5.8 muestra los patrones de consumo obtenidos para cada día de la semana.

Caracterización de la demanda de energía mediante patrones estocásticos en las Redes Eléctricas Inteligentes

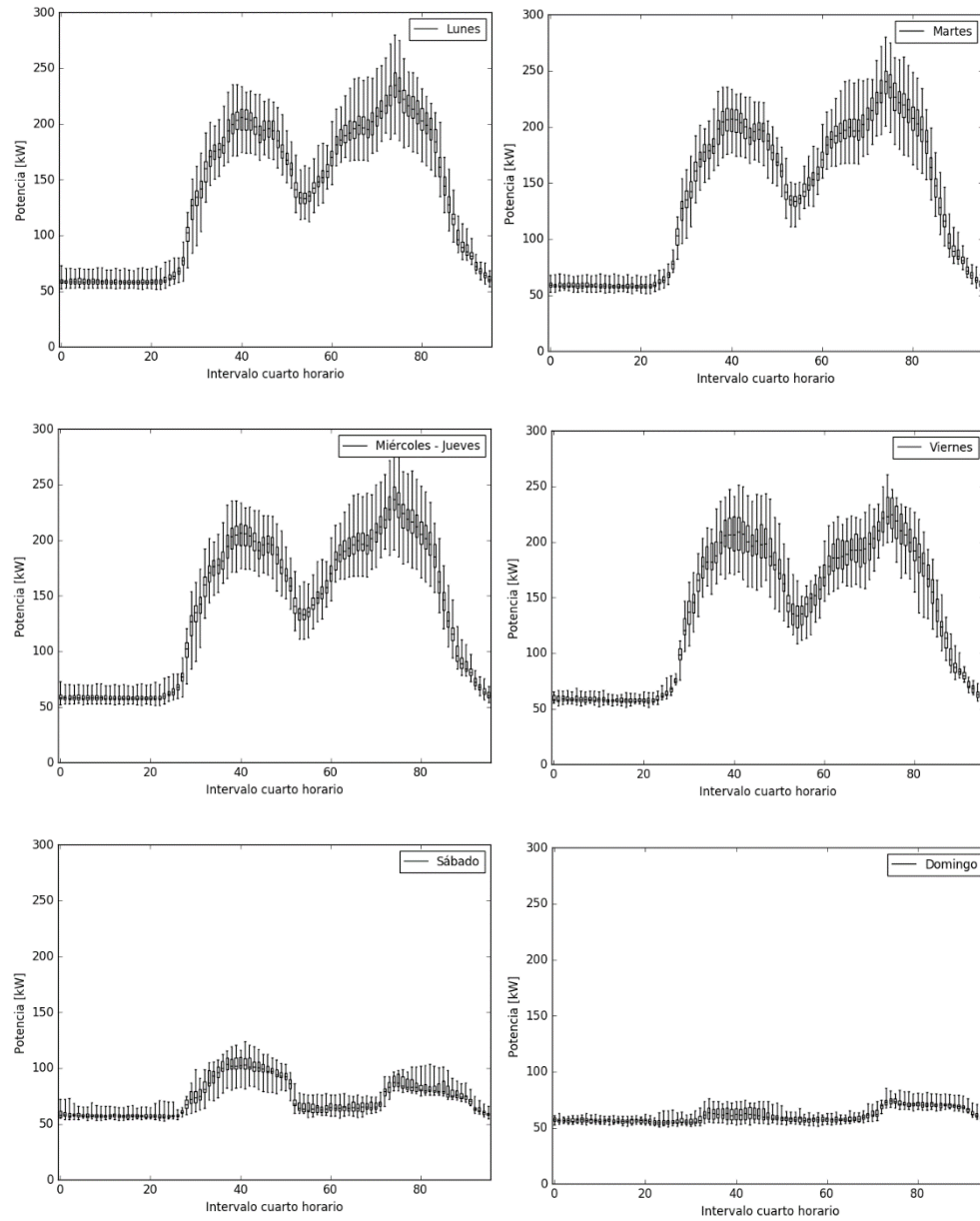


Imagen 5.8 Patrones estocásticos de consumo de cada día de la semana en la UPS – Cuenca

5.3.3 Etapa 3: Cuantificación estadística

La cuantificación estadística valora el ajuste de un día de análisis a su patrón correspondiente. En este apartado se presentan 5 días miércoles atípicos detectados por la metodología propuesta. La Imagen 5.9 muestra el patrón de consumo representado mediante diagramas de cajas y bigotes y los 5 días antes mencionados.

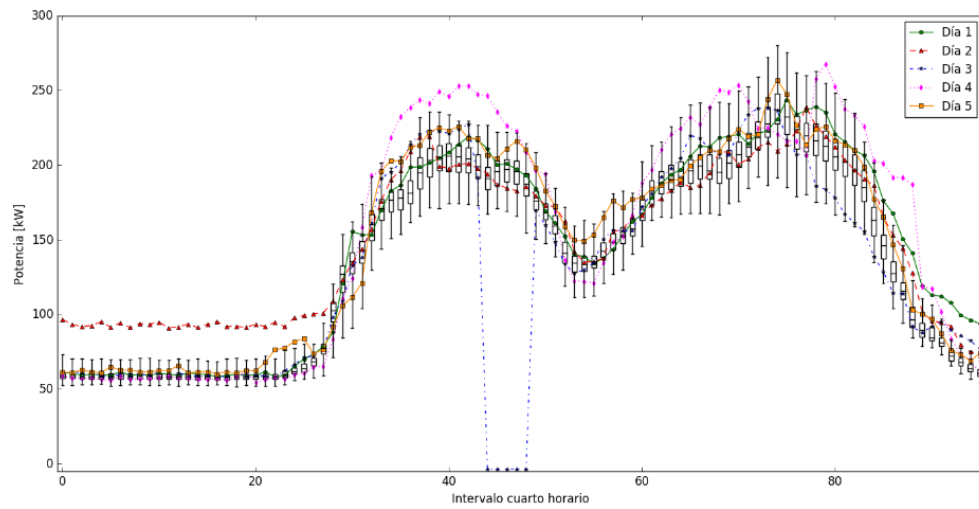


Imagen 5.9 Cuantificación estadística de 5 días miércoles versus su patrón de consumo

Gracias a la estandarización de datos las anomalías se aprecian de mejor manera ya que los valores del vector Z_{max} y Z_{min} muestran claramente los límites superior e inferior del patrón de consumo (ver Imagen 5.10).

Caracterización de la demanda de energía mediante patrones estocásticos en las Redes Eléctricas Inteligentes

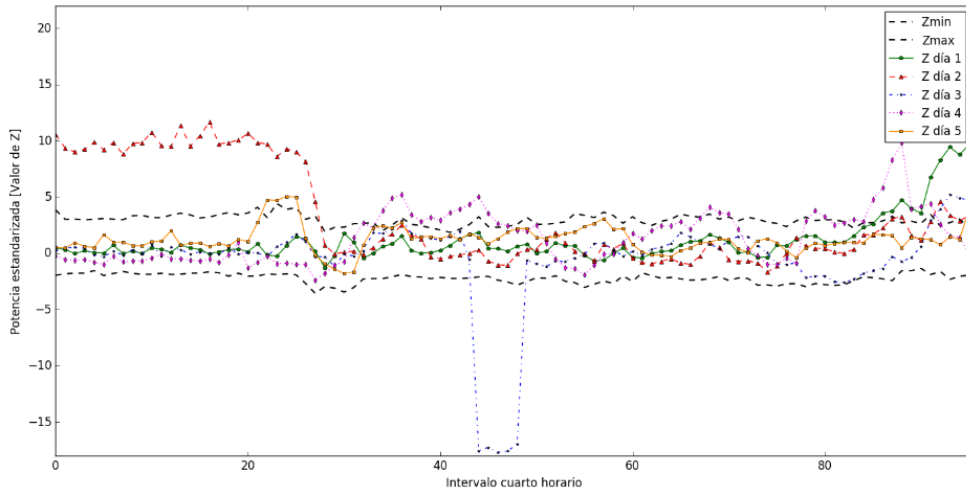


Imagen 5.10 Datos estandarizados para 5 días miércoles anómalos

También se presentan las potencias máximas, medias y mínimas de cada uno de los días miércoles atípicos comparadas con sus respectivos patrones (Imagen 5.11).

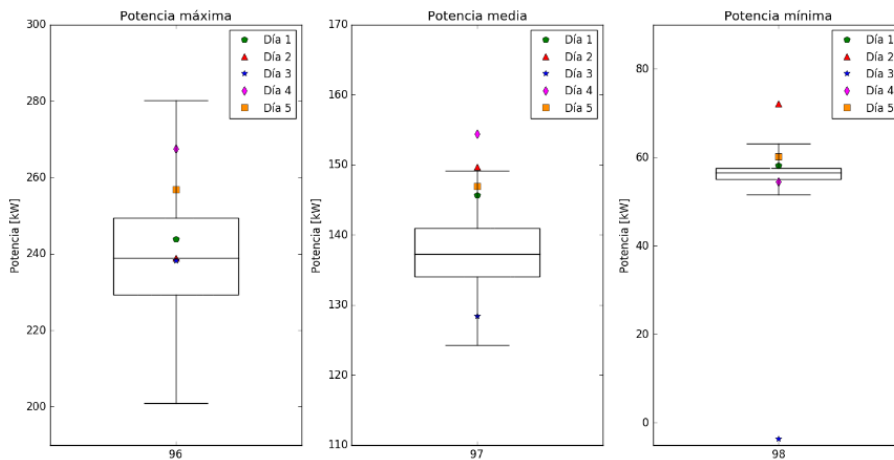


Imagen 5.11 Potencia máxima, media y mínima de los 5 días analizados versus el patrón del día miércoles

5.3.4 Etapa 4: Presentación de resultados y análisis físico

Los PCE correspondientes a los 5 días de análisis se evalúan mediante la aplicación de los tests descritos en el apartado 4.2.4 del Capítulo 4. El resumen de las pruebas aplicadas se muestra en la Tabla 5.6.

Tabla 5.6 Resumen de las pruebas aplicadas a los días de análisis en la UPS Sede Cuenca

Día	T1 (%)	T2 (%)	T3 (%)	T4 (%)	T5 (%)	T2/ T5	T3/ T4	T8 (h)	T9 (h)	I2	I3	I4
1	10.4	22.2	0	100	3.3	6	0	0	2.5	0	0	0
2	33.3	88.9	100	50	0	inf	2	0	7	1	0	1
3	9.4	11.1	0	50	8.33	1,33	0	1.25	1.25	0	0	1
4	31.3	13.9	0	62.5	41.68	0.33	0	0	3.5	1	0	0
5	4.2	8.3	11	0	1.67	4.97	inf	0	0.75	0	0	0

* T1, T2, T3, T4 T5, T8 y T9 corresponden a las pruebas descritas en la Tabla 4.3; I2, I3 y I4 son descritas en la Ecuación 4.6 del Capítulo 4.

Las conclusiones obtenidas por la metodología se muestran en la Tabla 5.7, en donde se las contrasta con el diagnóstico del administrador de los edificios de la UPS. Además, se muestra el IoC para cada PCE analizado.

5.3.5 Análisis de los resultados

Los resultados sugieren que la metodología propuesta es válida para analizar el consumo de un edificio o grupo de edificios o consumidores. De igual manera, se caracteriza la demanda, se obtienen patrones de consumo, se detectan anomalías y/o cambios en la demanda de energía y se presentan conclusiones acerca de las posibles causas de esos consumos atípicos. El IoC cataloga los PCE anómalos de acuerdo a la Tabla 4.5 presentada en el capítulo anterior.

Al igual que en el caso expuesto en el apartado 5.2, los resultados se contrastaron con el diagnóstico de un técnico experto. En este caso de estudio, los resultados obtenidos son similares, es decir que, cuando existen anomalías, más del 60% de las veces la metodología SAICC entrega información similar a la del experto de forma automática. En la mayoría de casos restantes la metodología proporciona mayor información que el experto, mientras que en casos puntuales el experto es más preciso. Los resultados evidencian que la metodología propuesta se puede aplicar a medianos o grandes consumidores que presentan una demanda de electricidad con componentes estacionales fuertes o débiles.

Tabla 5.7 Conclusiones de la metodología SAICC e IoC para los días de análisis

Día	Conclusiones de la metodología SAICC	Diagnóstico del experto	Mejor interpretación	IoC
1	<p>No hay consumos atípicos en las HNL de la madrugada, posible carga conectada en la noche.</p> <p>El consumo es atípico en todas las mediciones de la noche, posible carga se quedó conectada. Evaluar si existe error de medición.</p> <p>Existe mayor incidencia de datos atípicos en las horas no laborables del día. Sin embargo no hay cambios significativos en el consumo de energía diaria, potencia máxima y mínima consumida comparada con el patrón.</p>	<p>Carga conectada en la noche.</p> <p>Consumo de electricidad normal en las restantes horas del día.</p>	<p>Los dos dan información similar.</p>	0.03
2	<p>El consumo es atípico en todas las mediciones de la madrugada, posible carga se quedó conectada. Evaluar si existe error de medición.</p> <p>No hay consumo atípico en horas laborables del día. El consumo atípico se da en HNL.</p> <p>Existe mayor incidencia de datos atípicos en las HNL del día.</p> <p>La energía consumida es mayor a la del patrón. Es probable que algo se haya quedado conectado o exista alguna falla.</p> <p>La potencia mínima consumida es mayor a la del patrón. Es probable que algo se haya quedado conectado o exista alguna falla.</p>	<p>Carga conectada en toda la madrugada.</p>	<p>El método propuesto brinda información adicional.</p>	0.61
3	<p>Error de medición. Más de 1 hora con el mismo valor de medición.</p>	<p>Error de medición.</p>	<p>El experto indica información más precisa.</p>	0.20
4	<p>No hay consumos atípicos en las HNL de la madrugada, posible carga conectada en la noche.</p> <p>Existe mayor incidencia de consumos anómalos en las HL del día.</p> <p>La energía consumida es mayor a la del patrón. Es probable que la actividad laboral se incrementó (naturaleza de la ocupación)</p>	<p>Mayor consumo de energía en la Universidad debido a eventos realizados.</p>	<p>Los dos dan información similar.</p>	0.44
5	<p>El consumo analizado tiene 8.3 % de datos atípicos en HNL y 1.7 % en HL, los cuales no indican cambios representativos en el consumo de energía ni potencia.</p>	<p>Existe un mayor consumo de electricidad entre las 6 y 7 de la mañana. Causas no definidas</p>	<p>Los dos entregan información similar. Sin embargo el experto indica el intervalo exacto de la anomalía.</p>	0.01

5.4 Aplicación de la metodología SAICC en el Alimentador 0101 de la Empresa Eléctrica Regional Centrosur (EERCS)

Las empresas eléctricas, tanto de generación, transmisión y distribución son las encargadas de suministrar electricidad a los usuarios, cuya demanda crece en todo el mundo. La planificación de las redes eléctricas es muy importante ya que permite garantizar a corto, mediano y largo plazo el suministro de electricidad. En este apartado se analizan los PCE correspondientes al alimentador 0101 de la Subestación 1 de la EERCS en la ciudad de Cuenca, Ecuador (ver Imagen 5.12). La Subestación 1 es una de las más antiguas de la ciudad y con una potencia instalada de 19 MVA y una tensión de 6,3kV suministra energía a un importante sector comercial y residencial del centro histórico de la ciudad.



Imagen 5.12 Transformador No. 1 de la Subestación 1 de la EERCS [66]

El estudio de dicho alimentador que se presenta a continuación, permite conocer si la metodología SAICC resulta útil para caracterizar la demanda de energía eléctrica en todo un sector de una ciudad.

5.4.1 Etapa 1: Adquisición y procesamiento de datos históricos

El diagrama unifilar de la Subestación 1 se muestra en la Imagen 5.13, los datos de la demanda eléctrica del alimentador 0101 fueron suministrados por la empresa distribuidora mencionada. El periodo de análisis es de 365 días, desde el 01-01-2017 hasta el 31-12-2017. Se presenta el análisis de todos los días de la semana, considerando los días laborables de lunes a sábado, mientras que, todos los domingos se consideran no laborables.

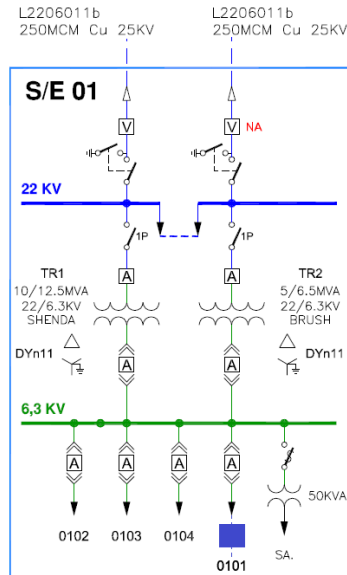


Imagen 5.13 Diagrama Unifilar de la Subestación 1 de la EERCS [66]

En este caso de estudio la demanda de electricidad tiene una componente estacional pequeña. No obstante, al realizar la transformación de datos a través del método SAEC el número de PCE típicos que son considerados en la conformación de los patrones se eleva en un 5.4%.

5.4.2 Etapa 2: Obtención de patrones de consumo

Tal como se realizó en los ejemplos anteriores, las características de los patrones obtenidos para cada día de la semana se muestran en la Imagen 5.14 e Imagen 5.15 en donde se pueden ver la media y la desviación estándar respectivamente.

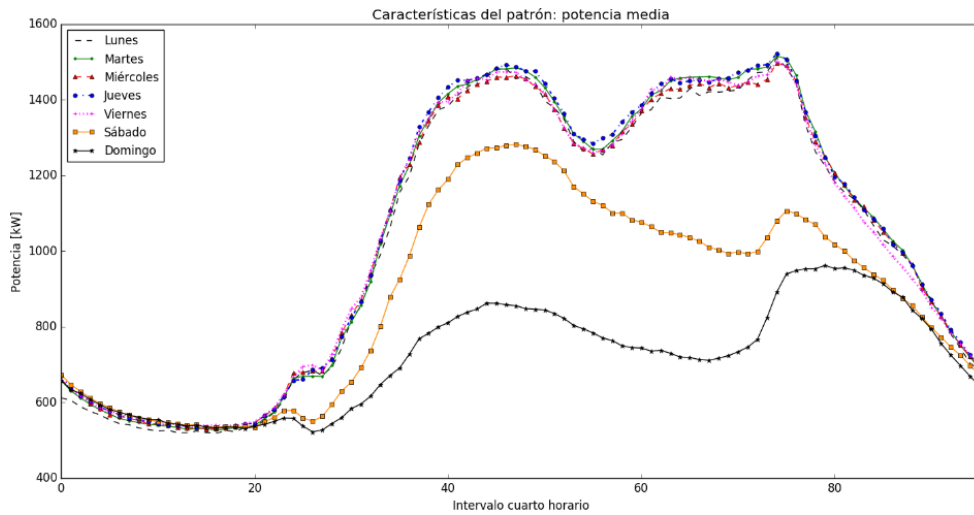


Imagen 5.14 Media del patrón de cada día de la semana en el alimentador 0101 (01-01-2017 hasta 31-12-2017)

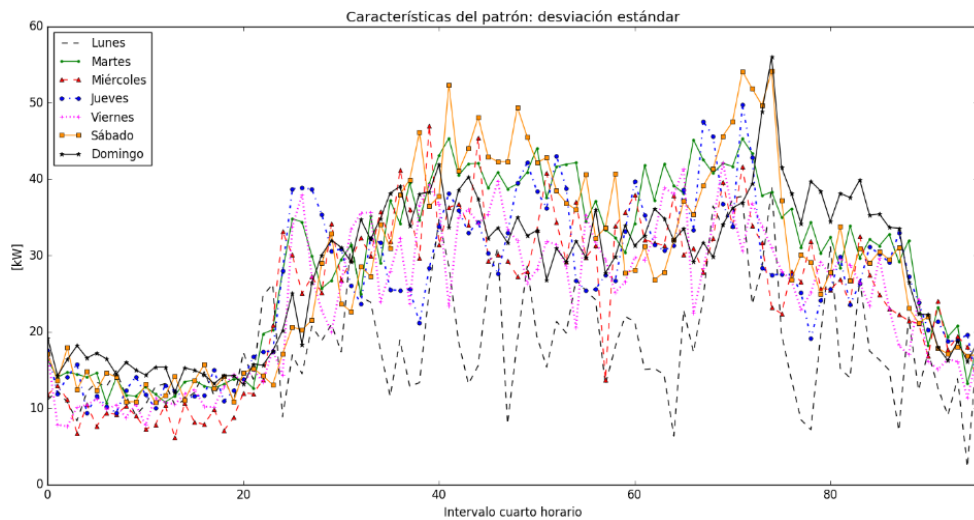


Imagen 5.15 Desviación estándar del patrón de cada día de la semana del alimentador 0101 (01-01-2017 hasta 31-12-2017)

El alimentador mencionado tiene una carga base de alrededor de 500kW que ocurre entre las 2:30 y 5:00 am. Los patrones de consumo son similares de lunes a viernes con una potencia activa media absorbida pico de alrededor 1500kW entre las 19:00 y 19:30 horas. El consumo de energía en los días sábados y domingos es menor, no obstante la

carga base se mantiene. Los detalles de la conformación de los patrones se muestran en la Tabla 5.8.

Tabla 5.8 Conformación de los patrones de electricidad para cada día de la semana en el alimentador 0101

Día de la semana	Laboralidad	Días con los que se agrupa	Número de PCE típicos	Número de PCE atípicos	Número total de PCE en el patrón definitivo
lunes	Laborable	martes-miércoles-jueves-viernes	10	36	91
martes	Laborable	lunes-miércoles-jueves-viernes	38	13	91
miércoles	Laborable	lunes-martes-jueves-viernes	17	34	91
jueves	Laborable	lunes-martes-miércoles-viernes	20	27	91
viernes	Laborable	lunes-martes-miércoles-jueves	14	27	91
sábado	Laborable	-	27	25	27
domingo	No Laborable	-	38	15	38
Total			164	177	341

En este caso, debido a la débil componente estacional que tiene la demanda de electricidad (ver Imagen 5.16), los patrones son similares a los obtenidos cuando no se aplica la transformación de datos con el método SAEC. Tampoco se evidencia intervalos de confianza más pequeños. Los patrones de todos los días de la semana representados mediante diagramas de cajas y bigotes se pueden ver en la Imagen 5.17.

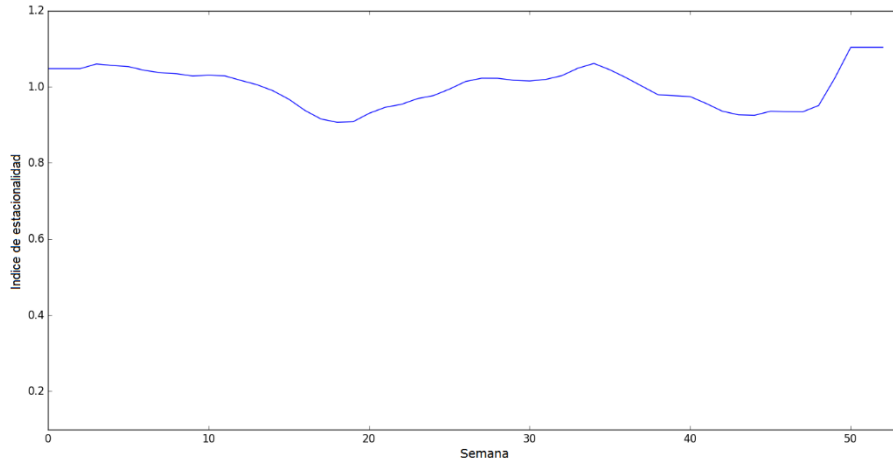


Imagen 5.16 Variación del índice de estacionalidad del alimentador 0101

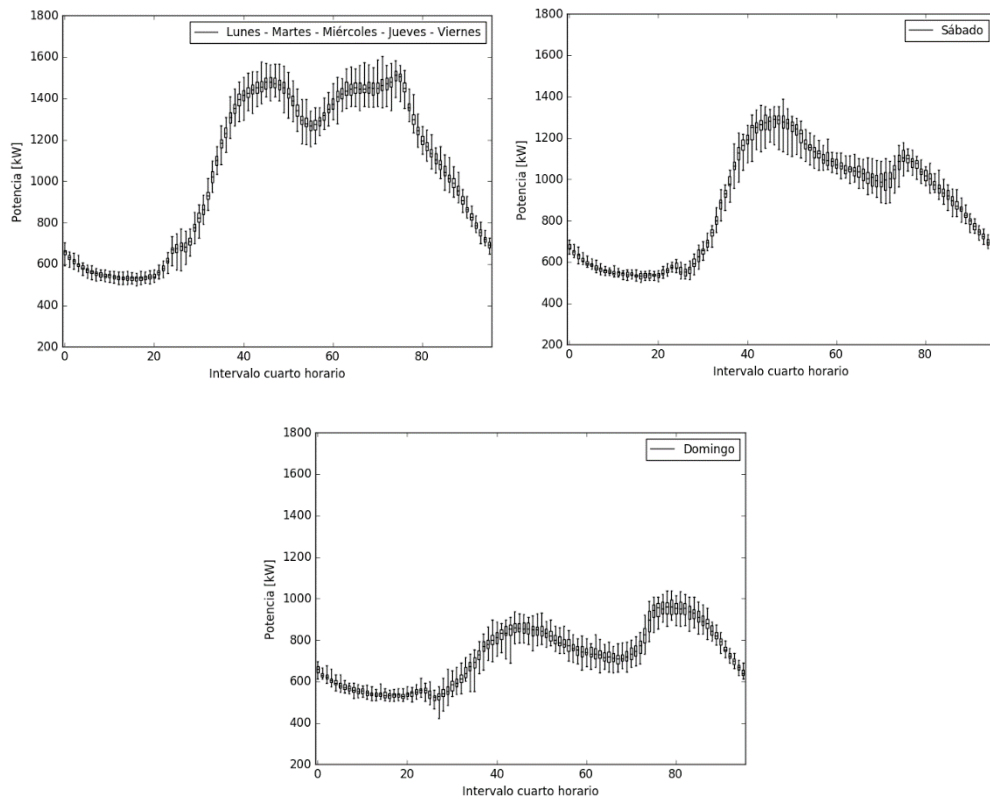


Imagen 5.17 Patrones estocásticos de consumo de cada día de la semana en el Alimentador 0101

5.4.3 Etapa 3: Cuantificación estadística

En este apartado se analizan 5 días miércoles atípicos detectados en el periodo definido en el apartado 5.4.1. En la Imagen 5.18 se muestra el patrón representado por los diagramas de cajas y bigotes y los días analizados.

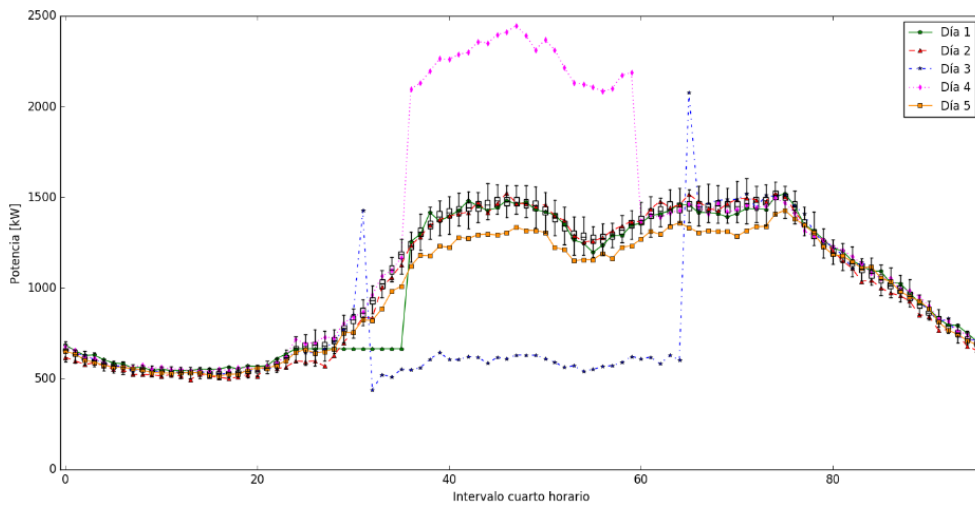


Imagen 5.18 Cuantificación estadística de 5 miércoles anómalos versus su patrón de consumo en el alimentador 0101

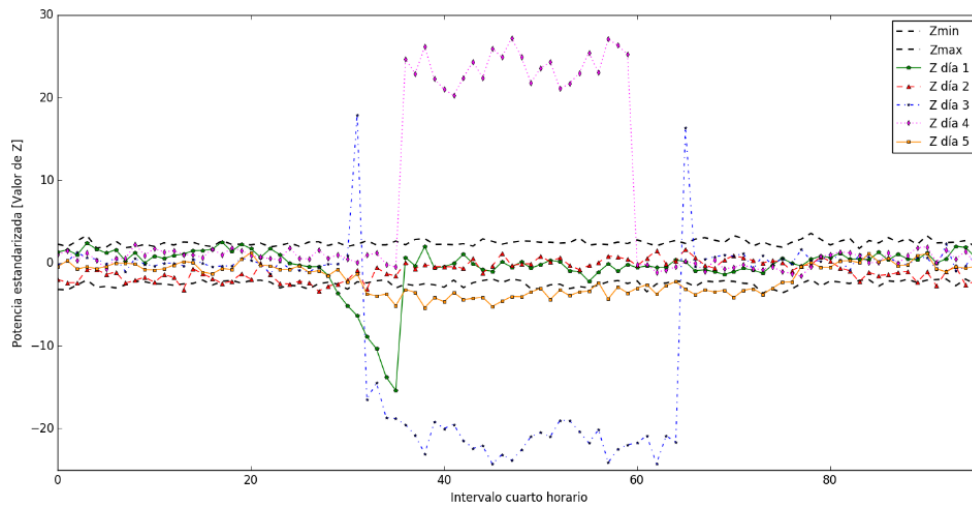


Imagen 5.19 Datos estandarizados de 5 miércoles anómalos y su patrón en el alimentador 0101

Las anomalías en la demanda se ven de manera clara cuando los datos son estandarizados. En la Imagen 5.19 se muestran los valores Z de cada día en análisis y los límites superior e inferior del patrón de consumo representado por los vectores Z_{max} y Z_{min} respectivamente.

En la Imagen 5.20 se puede observar el comportamiento de la potencia máxima, media y mínima de los días atípicos detectados con respecto a su patrón.

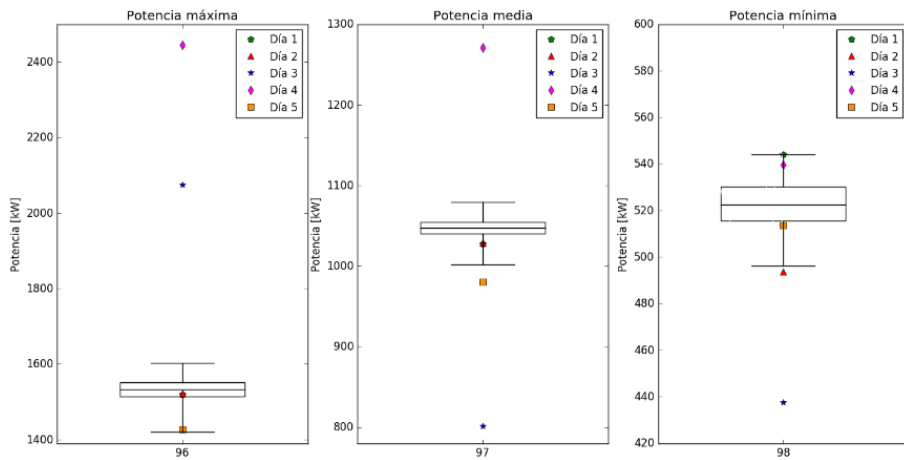


Imagen 5.20 Potencia máxima, media y mínima de 5 miércoles anómalos versus su patrón

5.4.4 Etapa 4: Presentación de resultados y análisis físico

Los PCE de los días anómalos seleccionados y descritos anteriormente, se evalúan mediante la aplicación de las pruebas descritas en el apartado 4.2.4 de este documento. El resumen de las pruebas aplicadas se muestra en la Tabla 5.8.

Tabla 5.8 Resumen de las pruebas aplicadas a los días de análisis

Día	T1 (%)	T2 (%)	T3 (%)	T4 (%)	T5 (%)	T2/T5	T3/T4	T8 (h)	T9 (h)	I2	I3	I4
1	7.29	0	0	0	11.7	0	0	3	1.75	0	0	1
2	9.38	16.7	14.3	25	5	3.33	0.57	0	0.25	0	0	-1
3	36.5	0	0	0	58.3	0	0	0	8.75	-1	1	-1
4	26.04	2.77	3.57	0	40	0.07	inf	0	6	1	1	0
5	40.6	0	0	0	65	0	0	0	4.5	-1	0	0

* T1, T2, T3, T4 T5, T8 y T9 corresponden a las pruebas descritas en la Tabla 4.3; I2, I3 y I4 son descritas en la Ecuación 4.6 del Capítulo 4.

Las pruebas aplicadas valoran los cambios en la demanda de electricidad y obtienen conclusiones acerca de las posibles causas de las anomalías detectadas. En la Tabla 5.9 se muestran estas conclusiones, el diagnóstico de un experto de la empresa distribuidora y el IoC de cada PCE analizado. El algoritmo que se utiliza para elaborar la Tabla MCIAT se mantiene intacto. Sin embargo, la terminología usada en las conclusiones se modifica levemente, por ejemplo en lugar de *desconexión de circuitos* se indica *transferencia de carga*, en lugar de *posible carga conectada* se indica *posible recepción de carga* o en lugar de *energía consumida* se indica *energía demanda*. Por ende, es importante resaltar que dicha tabla se debe adaptar levemente de acuerdo al nivel de la red eléctrica en donde se realiza el análisis.

Tabla 5.9 Conclusiones de la metodología SAICC e IoC (EERCS)

Día	Conclusiones de la metodología SAICC	Diagnóstico del experto	Mejor interpretación	IoC
1	Error de medición. Más de 1 hora con el mismo valor.	Error de medición, problema con el sistema SCADA.	Los dos dan información similar.	0.19
2	Existe mayor incidencia de datos atípicos en las HNL del día. La potencia mínima consumida es menor a la del patrón. Probable falla, error de medición, mantenimiento o transferencia de carga.	Se aprecia una disminución en la demanda de energía entre las 5:30 y 7:30 am debido al período de vacaciones de los estudiantes.	Experto	0.20
3	Existe mayor incidencia de consumos anómalos en las HL del día. La energía consumida es menor a la del patrón. La potencia máxima consumida es mayor a la del patrón. Posible disminución en actividad laboral o recepción de carga. La potencia mínima consumida es menor a la del patrón. Probable falla, error de medición, mantenimiento o transferencia de carga.	El alimentador en estudio transfirió carga a otros alimentadores debido a mantenimientos programados. Los picos que se presentan antes y después del evento son atípicos, deberían corregirse.	Experto	0.79
4	Existe mayor incidencia de consumos anómalos en las HL del día. La energía consumida es mayor a la del patrón. La potencia máxima consumida es mayor a la del patrón. Posible aumento en actividad laboral o recepción de carga.	El alimentador recibió carga de otro alimentador, debido a actividades de mantenimiento.	Los dos dan información similar.	0.59
5	Existe mayor incidencia de consumos anómalos en las HL del día. La energía consumida es menor a la del patrón. Probable disminución de la demanda.	Los primeros días del año se registra baja demanda.	Los dos dan información similar.	0.47

5.4.5 Análisis de los resultados

La metodología SAICC desarrollada inicialmente para analizar el consumo de electricidad en edificios puede utilizarse también para analizar la demanda en alimentadores de subestaciones, siendo capaz de detectar anomalías y/o cambios en los PCE. Las posibles causas de estas anomalías son válidas al hacer cambios únicamente en la terminología usada en la Tabla MCIAT, sin cambiar el algoritmo base.

Los resultados se contrastaron con el diagnóstico de un técnico experto de la empresa distribuidora, las conclusiones obtenidas tienen una confiabilidad similar a las aplicaciones expuestas anteriormente. En este caso de estudio se evidencia que la proporción de conclusiones en donde el experto entrega información más precisa que la del método SAICC aumenta en un pequeño porcentaje debido a eventos muy específicos, tales como mantenimientos programados.

Al igual que en los casos anteriores el IoC se puede utilizar para catalogar el cambio que ha tenido la demanda en el alimentador. La Tabla 5.10 se puede aplicar en alimentadores de subestaciones y resulta de modificar algunos términos usados en la Tabla 4.5 expuesta en el capítulo anterior.

Tabla 5.10 Catalogación del perfil de consumo de electricidad usando los valores de IoC para subestaciones

IoC	Cambio en el consumo	Interpretación
0	Ninguno	La demanda eléctrica del día analizado se ajusta al patrón.
(0-0.03]	Muy pequeño	No existe un cambio significativo en la demanda de electricidad.
(0.03-0.15]	Relativo	Existen varios datos atípicos, pero siempre menores que el 40% de todos los datos. Estos datos anómalos no provocan cambios en la potencia máxima, media y mínima.
(0.15-0.3]	Notable	Posible transferencia o recepción de carga en parte del día. No se evidencia gran diferencia en la demanda diaria con respecto al patrón. Sin embargo, los valores de potencia máxima y mínima pueden estar fuera el rango.
(0.3-0.55]	Grande	La demanda total de electricidad esta fuera de rango con respecto al patrón.
(0.55-0.78]	Muy grandes	La presencia de valores atípicos durante gran parte del día ocasiona que la demanda total de electricidad esté fuera de rango. Asimismo la potencia máxima o mínima es atípica.
(0.78-1]	Extremo	Todos los parámetros analizados están fuera de rango. La demanda de electricidad es totalmente atípica.

5.5 Aplicación de la metodología SAICC en el Alimentador 0425 de la EERCS

En este apartado se analizan los PCE correspondientes al alimentador 0425 de la Subestación 4 de la EERCS (Imagen 5.21). La Subestación 4 tiene una potencia instalada total de 44,5 MVA y suministra el fluido eléctrico a 22kV a un sector industrial en el norte de la ciudad. El estudio de este alimentador indica si la metodología es útil para evaluar cambios en PCE de consumidores industriales, cuya demanda difiere totalmente a la del anterior caso analizado.



Imagen 5.21 Subestación 4 EERCS [66]

5.5.1 Etapa 1: Adquisición y procesamiento de datos históricos

Como se ha indicado, los datos de la demanda de energía eléctrica corresponden al alimentador 0425, cuya ubicación en el diagrama unifilar de la subestación se muestra en la Imagen 5.22. El periodo de datos obtenido es de 365 días, desde el 01-01-2017 hasta el 31-12-2017. El análisis se realiza sobre todos los días de la semana.

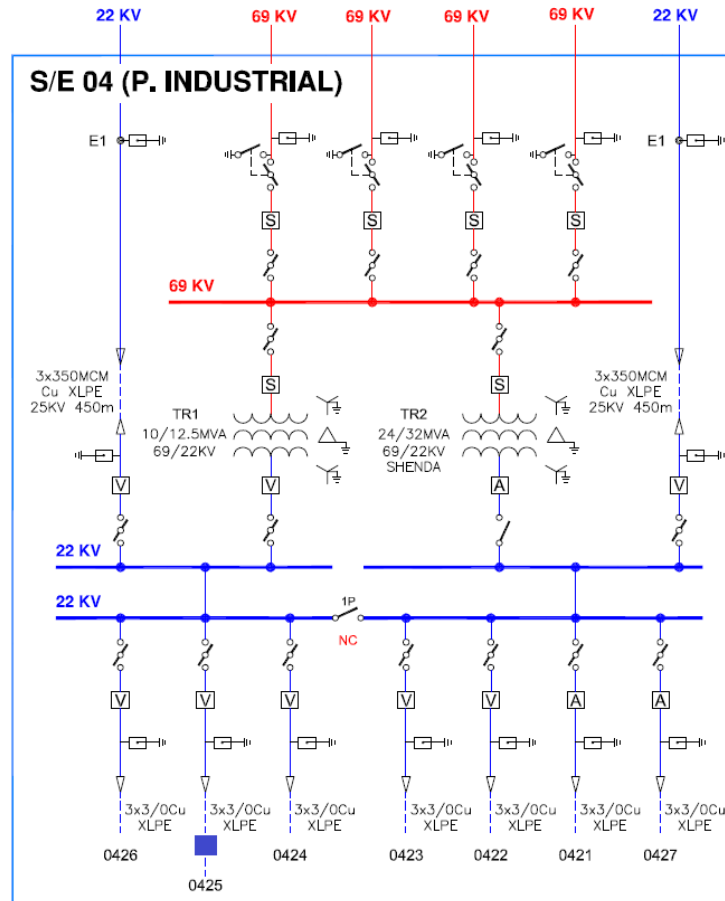


Imagen 5.22 Diagrama unifilar de la Subestación 4 de la EERCS [66]

5.5.2 Etapa 2: Obtención de patrones de consumo

En este caso de estudio, la serie de datos también tiene una componente estacional débil (ver Imagen 5.23). Sin embargo, la aplicación del método SAEC incrementa en un 6% el número de PCE considerados en la conformación de patrones. La media de los patrones obtenidos se muestra en la Imagen 5.24, en donde se aprecia que la potencia activa media absorbida llega hasta valores cercanos a los 6MW.

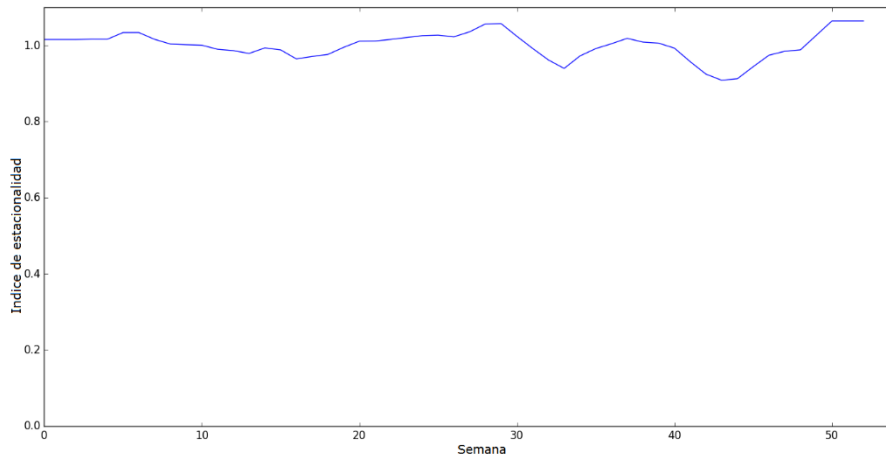


Imagen 5.23 Variación del índice de estacionalidad del alimentador 0425 de la EERCS

La demanda de energía del alimentador baja hasta unos 5.4MW el domingo luego de las 6:00 horas y vuelve a subir los días lunes luego de las 7:00 horas, esto debido al cese de actividades en una de las etapas de producción de la industria más grande conectada a este alimentador. Por otro lado, la desviación estándar (Imagen 5.25) indica que la variabilidad considerada típica oscila entre los 200kW y 300kW, aunque en ciertos días puede llegar hasta los 550kW.

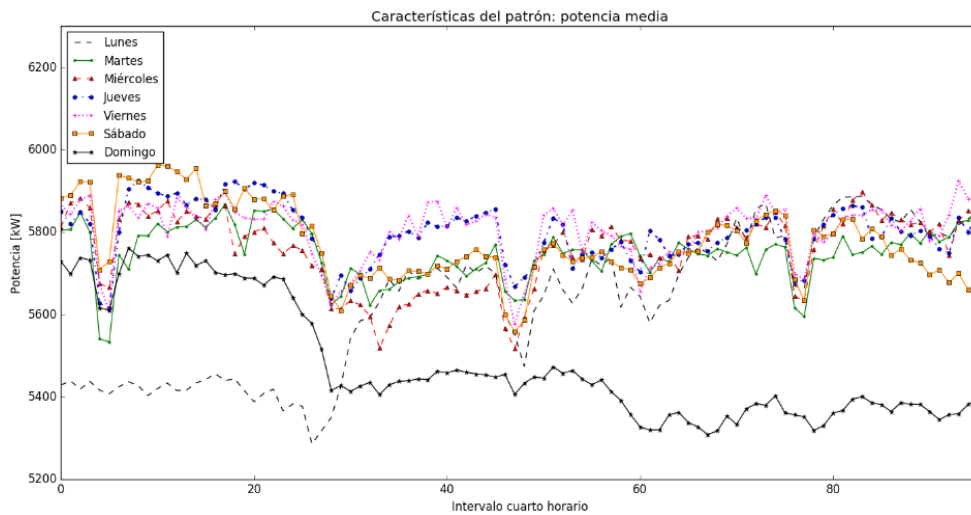


Imagen 5.24 Media del patrón de cada día de la semana del alimentador 0425 (01-01-2017 hasta el 31-12-2017)

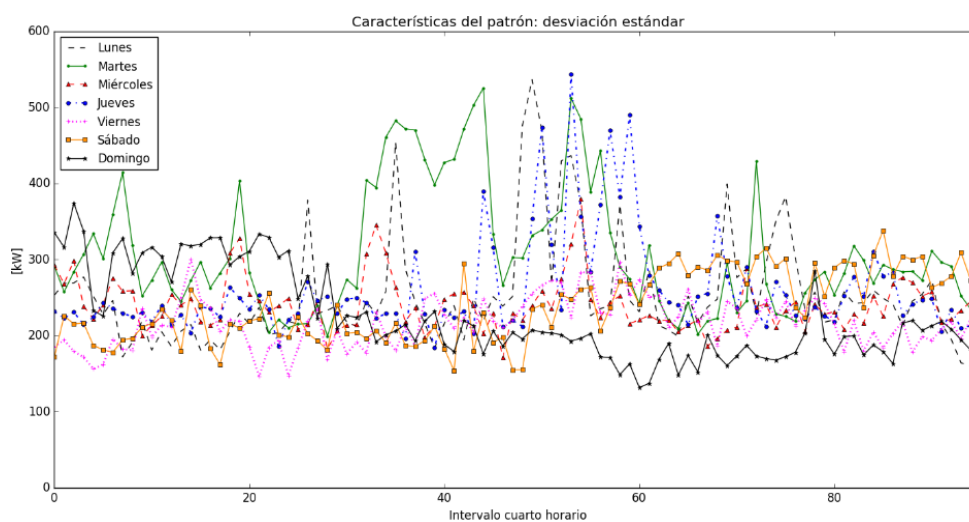


Imagen 5.25 Desviación estándar del patrón de cada día de la semana del alimentador 0425 (01-01-2017 hasta el 31-12-2017)

Los detalles de la conformación de los patrones se pueden muestran en la Tabla 5.11.

Tabla 5.11 Conformación de los patrones de electricidad para cada día de la semana en alimentador 0425

Día de la semana	Laboralidad	Días con los que se agrupa	Número de PCE típicos	Número de PCE atípicos	Número total de PCE en el patrón definitivo
lunes	Laborable	-	24	22	24
martes	Laborable	miércoles-jueves-viernes-sábado	36	15	137
miércoles	Laborable	martes-jueves-viernes-sábado	32	18	137
jueves	Laborable	martes-miércoles-viernes-sábado	27	21	137
viernes	Laborable	martes-miércoles-jueves-sábado	22	19	137
sábado	Laborable	martes-miércoles-jueves-viernes	23	29	137
domingo	No Laborable	-	33	20	33
Total			197	144	341

Al igual que en el caso del alimentador 0101 y a diferencia de los casos presentados en los apartados 5.2 y 5.3, al aplicar el método SAEC en los patrones identificados no se obtienen intervalos de confianza más pequeños, esto debido a la presencia de una componente estacional débil en los datos. Los patrones representados mediante diagramas de cajas y bigotes se muestran en la Imagen 5.26.

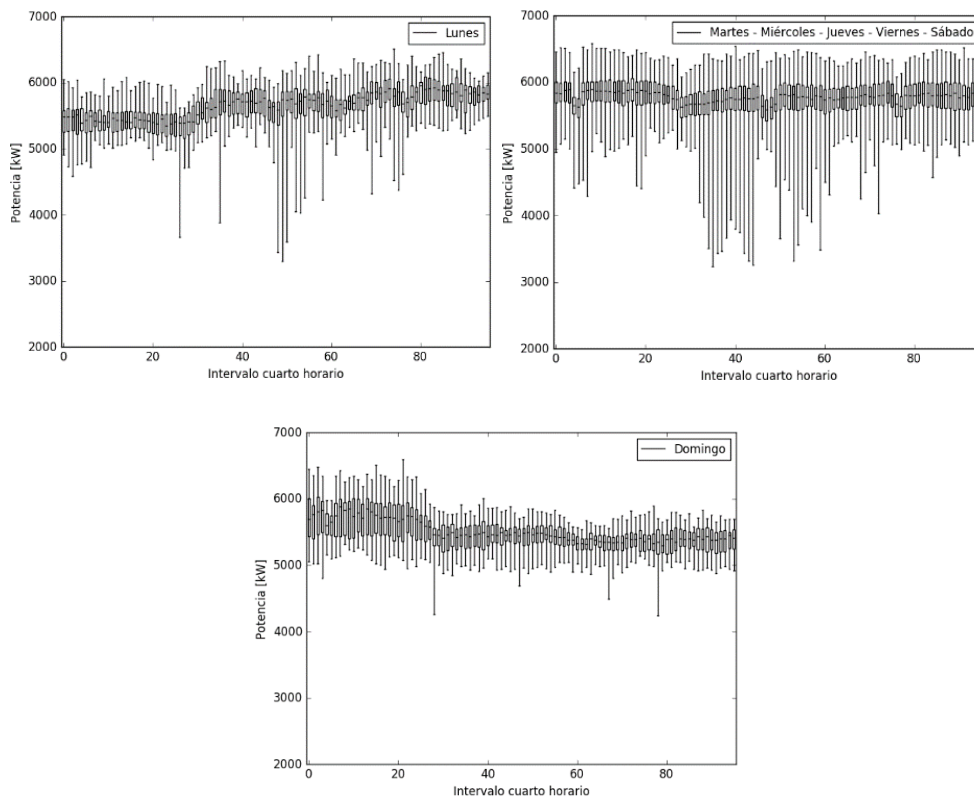


Imagen 5.26 Patrones estocásticos de cada uno de los días de la semana del Alimentador 0425

5.5.3 Etapa 3: Cuantificación estadística

A continuación, a modo de ejemplo se presenta el análisis de la demanda de electricidad de cuatro días martes atípicos detectados en el periodo indicado en el apartado 5.5.1. La Imagen 5.27 muestra los cuatro PCE y su patrón correspondiente.

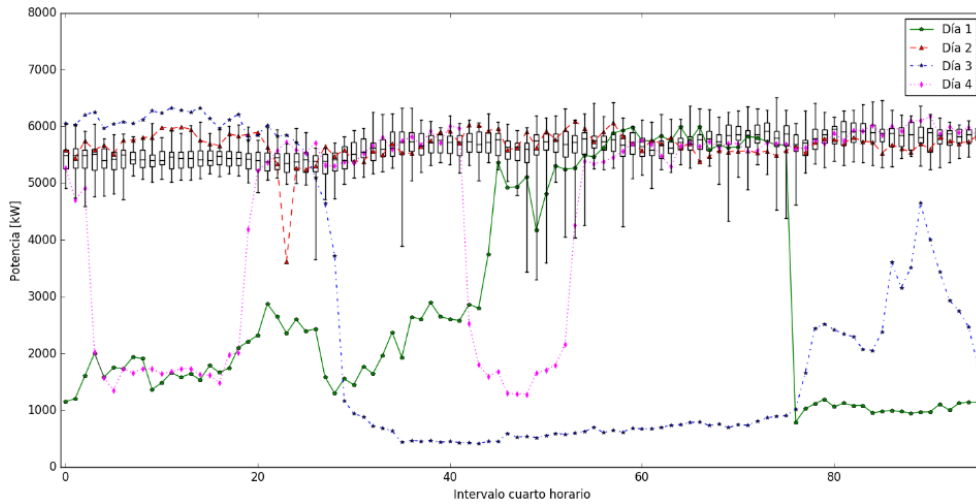


Imagen 5.27 Cuantificación estadística de 4 martes anómalos versus su patrón en el alimentador 0425

En la Imagen 5.28 se muestran los límites superior e inferior del patrón de consumo representado por los vectores Z_{max} y Z_{min} respectivamente, así como también los valores Z de cada día en análisis. En esta representación, las anomalías se observan claramente.

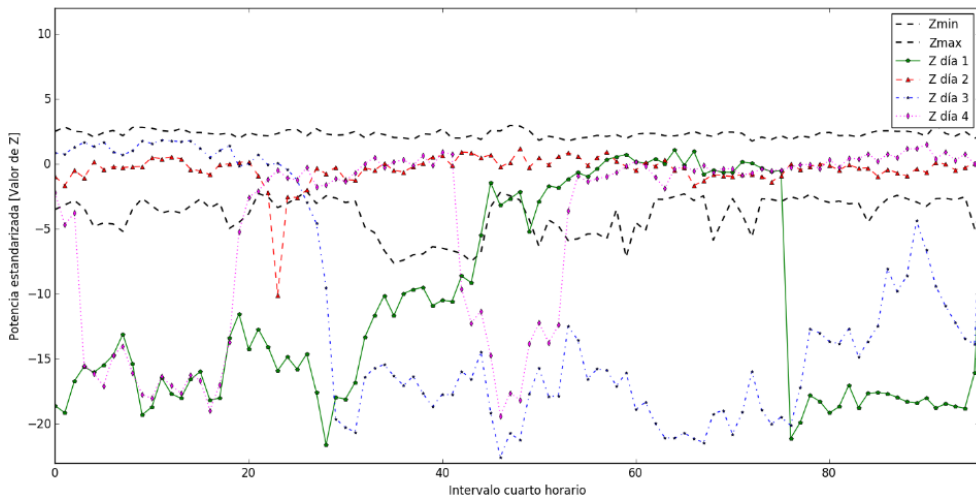


Imagen 5.28 Datos estandarizados de 4 martes anómalos y su patrón en el alimentador 0425

La potencia máxima, media y mínima de cada uno de los días atípicos detectados y su patrón de consumo se muestran en la Imagen 5.29.

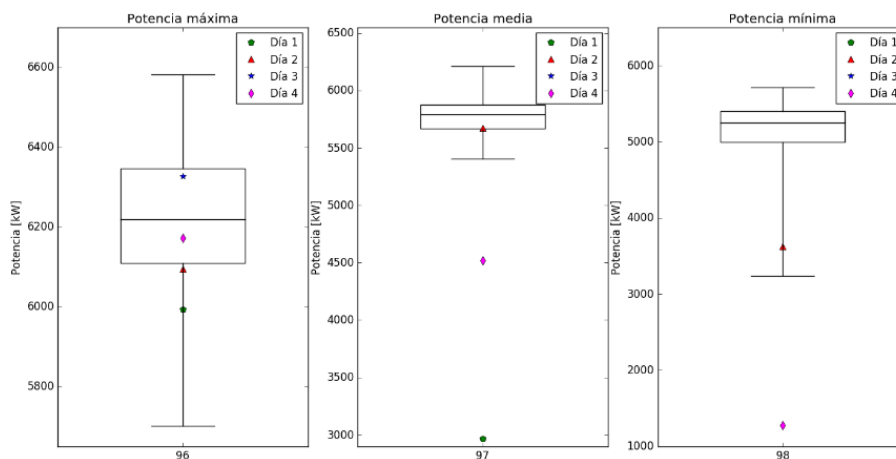


Imagen 5.29 Potencia máxima, media y mínima de 4 días martes anómalos en el alimentador 0425

5.5.4 Etapa 4: Presentación de resultados y análisis físico

El resumen de las pruebas realizadas en la cuarta etapa de la metodología SAICC para cada uno de los 4 días de análisis se muestra en la Tabla 5.12.

Tabla 5.12 Resumen de las pruebas aplicadas a los días de análisis

Día	T1 (%)	T2 (%)	T3 (%)	T4 (%)	T5 (%)	T2/T5	T3/T4	T8 (h)	T9 (h)	I2	I3	I4
1	69.8	100	100	100	51.7	1.93	1	0	11	-1	0	-1
2	1.04	2.77	3.57	0	0	inf	inf	0	0	0	0	0
3	72.9	27.8	7.14	100	100	0.28	0.07	0	17.5	-1	0	-1
4	31.25	52.8	67.9	0	18.3	2.88	inf	0	4.75	-1	0	-1

* T1, T2, T3, T4 T5, T8 y T9 corresponden a las pruebas descritas en la Tabla 4.3; I2, I3 y I4 son descritas en la Ecuación 4.6 del Capítulo 4.

Las conclusiones obtenidas se presentan en la Tabla 5.13 en conjunto con el IoC de cada PCE y el diagnóstico de un experto de la empresa distribuidora. Las modificaciones realizadas en la Tabla MCIAT para el alimentador 0101 (apartado 5.4.4), se mantienen para este caso de estudio.

Tabla 5.13 Conclusiones de la metodología SAICC e IoC (Alimentador 0425)

Día	Conclusiones de la metodología SAICC	Diagnóstico del experto	Mejor interpretación	IoC
1	La demanda es atípica todas las HNL del día. Posible recepción / transferencia permanente de carga.	El alimentador transfirió carga a otros alimentadores debido a mantenimientos programados.	Los dos entregan información similar.	0.73
	Existe mayor incidencia de datos atípicos en las HNL del día.			
	La energía demanda es menor a la del patrón. Posible transferencia de carga, mantenimiento o fallo.			
	La potencia mínima demandada es menor a la del patrón. Probable transferencia de carga, mantenimiento, falla o error de medición.			
2	El consumo analizado tiene 2.78% de datos atípicos en HNL y 0% en HL, los cuales no indican cambios representativos en la demanda de energía, ni potencia.	La demanda en el alimentador es normal.	Los dos entregan información similar.	0.003
3	La demanda es atípica en todas las mediciones de la noche, posible transferencia/recepción en esas horas. Evaluar si existe error de medición.	La demanda de energía en el alimentador es menor a lo habitual.	Los dos entregan información similar.	0.74
	Existe mayor incidencia de datos atípicos en las HNL de la noche. Posible transferencia de carga en la noche o falla.			
	Existe mayor incidencia de consumos anómalos en las HL del día.			
	La demanda de energía es menor a la del patrón. Posible transferencia de carga o disminución en la demanda.			
	La potencia mínima consumida es menor a la del patrón. Probable transferencia de carga, falla, error de medición o mantenimiento.			
4	No hay consumos atípicos en las horas no laborables de la noche. Posible transferencia / recepción de carga en horas de la madrugada.	El alimentador transfirió carga a otro alimentador, debido a actividades de mantenimiento.	Los dos dan información similar.	0.60
	Existe mayor incidencia de datos atípicos en las HNL de la madrugada. Posible transferencia / recepción de carga en las horas de la madrugada o falla o desconexión.			
	La energía demandada es menor a la del patrón. Posible transferencia de carga, mantenimiento o fallo.			
	La potencia mínima consumida es menor a la del patrón. Probable transferencia de carga, falla, error de medición o mantenimiento.			

5.5.5 Análisis de los resultados

Los resultados obtenidos indican que la metodología SAICC se puede utilizar para caracterizar la demanda de electricidad en diferentes tipos de alimentadores que suministran de energía a diversos sectores, ya sean estos residenciales, comerciales o industriales. Al igual que en los casos anteriores la metodología identifica patrones, detecta y valora anomalías en los PCE e indica las posibles causas de esas anomalías.

El IoC sigue siendo válido para catalogar o clasificar los PCE basado en la valoración del cambio que han tenido con respecto a su patrón correspondiente (Tabla 4.5).

La metodología SAICC propuesta se muestra como una herramienta útil en la gestión de la demanda de electricidad en empresas distribuidoras, propiciando la reducción de costes, detección temprana de anomalías en los alimentadores, tales como errores de medición o robo de energía. Además puede ahorrar el tiempo que los técnicos dedican al análisis de datos. Esta metodología también puede ayudar a planificar la repotenciación de las redes eléctricas de distribución al conocer en detalle los patrones de la demanda en cada alimentador.

5.6 Aplicación de la metodología SAICC en redes de transmisión eléctricas

La metodología SAICC se ha validado a nivel de consumidores y alimentadores de subestaciones que suministran electricidad a diversos sectores. En este apartado se analiza el flujo de potencia de una de las líneas de transmisión más importantes del Ecuador, en este caso la Molino-Pascuales del Sistema Nacional de Transmisión de la Empresa Celec EP Transelectric. La línea de transmisión tiene una capacidad de 2 x 375 MVA, una longitud de 188.4 km y una tensión de 230kV [107].



Imagen 5.30 Línea de transmisión Molino – Pascuales en Ecuador

En los siguientes apartados se describe la aplicación de la metodología SAICC a los datos del flujo de potencia activa de la línea de transmisión mencionada.

5.6.1 Etapa 1: Adquisición y procesamiento de datos históricos

Los datos de los flujos de potencia de la línea corresponden al periodo desde el 01-10-2017 hasta el 30-09-2018, es decir un año entero. Al igual que en los casos anteriores se analizan todos los días de la semana. En la Imagen 5.31 se puede ver la ubicación geográfica de la línea de transmisión Molino - Pascuales.

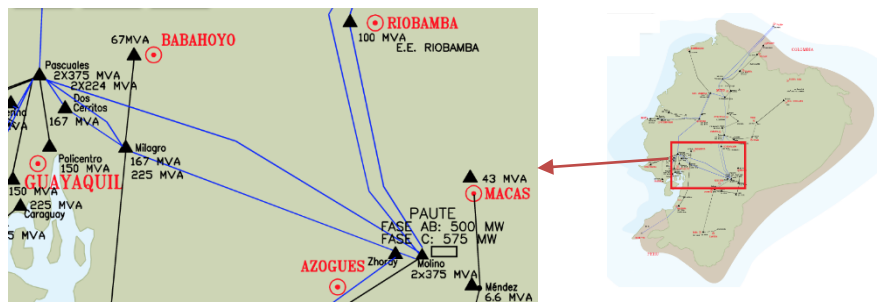


Imagen 5.31 Línea de transmisión Molino – Pascuales en Ecuador [107]

En este caso no se cumplen las 4 condiciones para llevar a cabo el tratamiento de datos con el método SAEC propuesto. El flujo de potencia a través de la línea no solo depende de la demanda, sino de la disponibilidad del recurso hídrico para generación, costos de producción y flujos de potencia en otras líneas de transmisión de la red eléctrica. Se pudo comprobar que, cuando se aplica el método SAEC los patrones obtenidos tienen mayor variabilidad, en tanto que la precisión en la detección de anomalías desmejora notablemente. Por tanto, en este caso de estudio, no se realizó la transformación de datos en la primera etapa del método SAICC.

5.6.2 Etapa 2: Obtención de patrones de consumo

Los patrones se obtienen sin eliminar la tendencia y desestacionalizar los datos, tal como se indicó en el apartado anterior. Las características de los patrones obtenidos se muestran en la Imagen 5.32 e Imagen 5.33 en donde se pueden ver los vectores de la media y la desviación estándar de cada uno de los patrones de los días de la semana, respectivamente.

La línea de transmisión tiene un flujo habitual de potencia activa que varía de entre 90MW y 170MW los días sábados y domingos. Los días lunes tienen un valor típico de carga mínima de 100MW en horas de la madrugada, en tanto que de lunes a viernes la potencia activa media pico ronda los 200MW. La desviación típica de la potencia acti-

va en la línea de transmisión oscila entre los 12MW y 17MW en horas de la madrugada y entre 15MW y 22MW en las horas restantes.

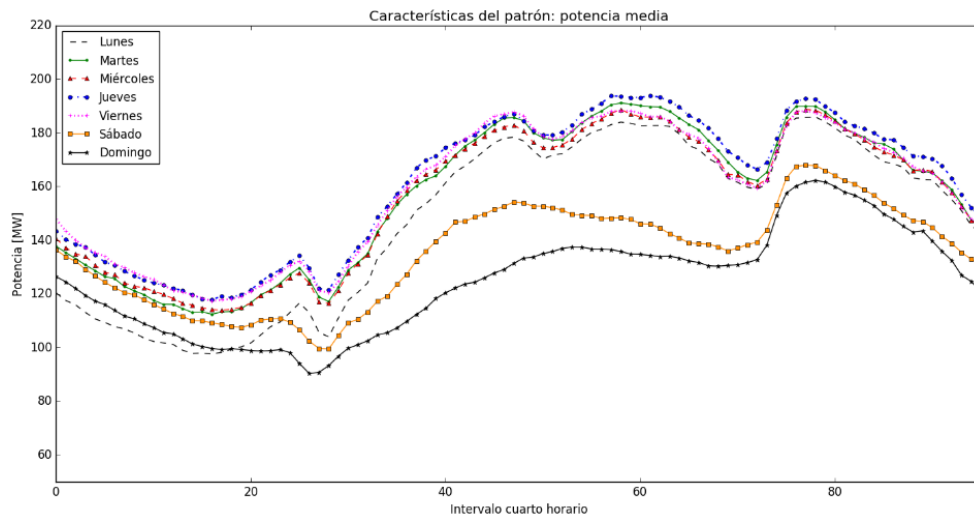


Imagen 5.32 Media del patrón de cada día de la semana. Flujo de potencia activa línea Molino-Pascuales (01-10-2017 hasta el 30-09-2018)

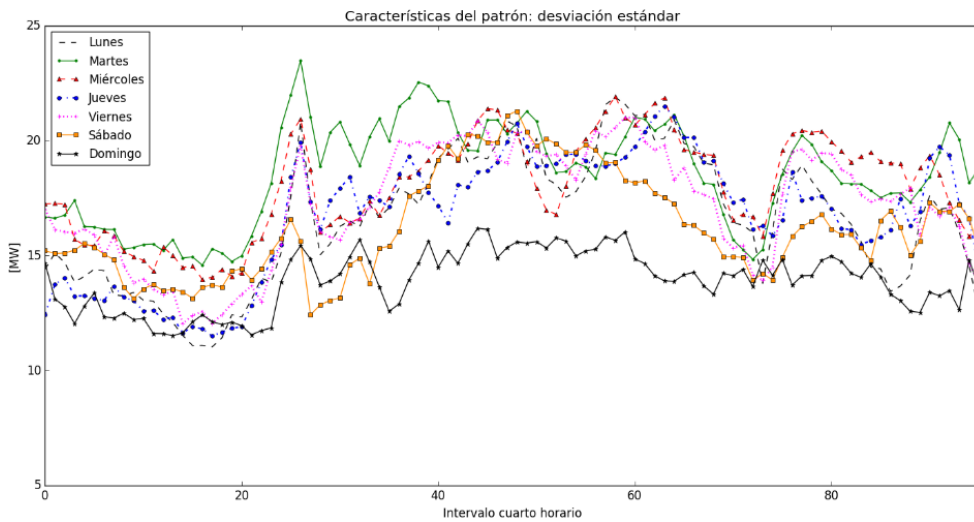


Imagen 5.33 Desviación estándar del patrón de cada día de la semana. Flujo de potencia activa línea Molino-Pascuales (01-10-2017 hasta el 30-09-2018)

La conformación de los patrones de electricidad de cada día de la semana se detalla en la Tabla 5.14 y su representación mediante diagramas de cajas y bigotes puede verse en la Imagen 5.34.

Tabla 5.14 Conformación de los patrones de electricidad por día de la semana (línea de Transmisión Molino-Pascuales)

Día de la semana	Laboralidad	Días con los que se agrupa	Número de PCE típicos	Número de PCE atípicos	Número total de PCE en el patrón definitivo
lunes	Laborable	-	33	15	33
martes	Laborable	miércoles-viernes	42	8	120
miércoles	Laborable	martes-viernes	41	11	120
jueves	Laborable	viernes	36	15	74
viernes	Laborable	martes-miércoles-jueves	38	10	156
sábado	Laborable	-	38	14	38
domingo	No Laborable	-	37	16	37
Total			265	89	354

5.6.3 Etapa 3: Cuantificación estadística

La cuantificación estadística se realiza según el procedimiento seguido en los casos de estudio anteriores al comparar un día de análisis con su patrón correspondiente (Imagen 5.34). En esta etapa es posible detectar y valorar los PCE anómalos de acuerdo a la metodología SAICC.

Caracterización de la demanda de energía mediante patrones estocásticos en las Redes Eléctricas Inteligentes

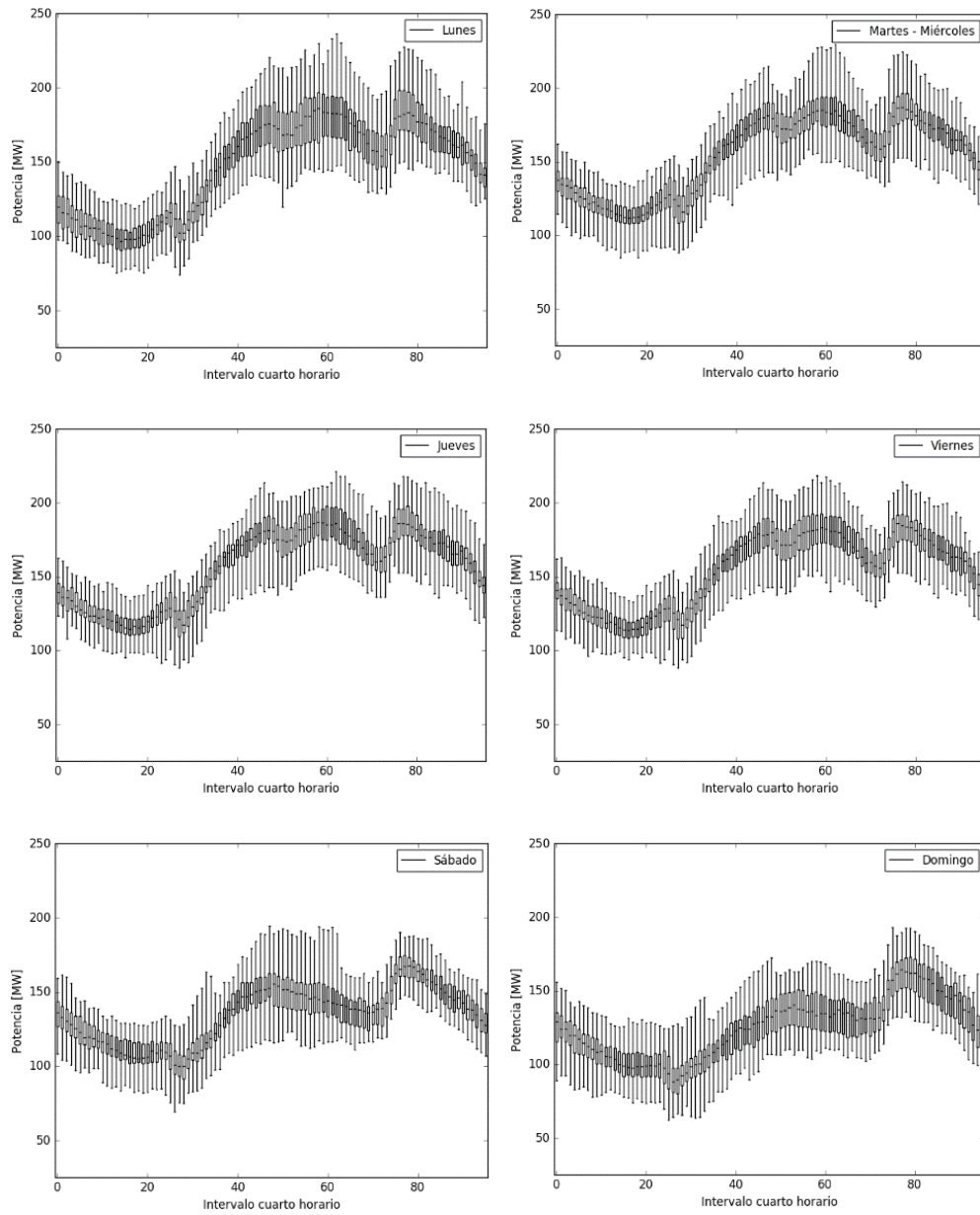


Imagen 5.34 Patrones estocásticos de cada uno de los días de la semana en la línea de transmisión Molino-Pascuales

5.6.4 Etapa 4: Presentación de resultados y análisis físico

Todos los flujos de potencia diarios en el periodo de estudio fueron evaluados mediante las pruebas de la cuarta etapa de la metodología SAICC. Sin embargo, como ejemplo de la aplicación se presentan únicamente los resultados de los perfiles de los días martes y miércoles.

Las conclusiones obtenidas acerca de las posibles causas de las anomalías en este caso no se ajustan siempre a la realidad debido a que el flujo de potencia activa de la línea de transmisión no depende únicamente de la demanda de energía, sino también de otros aspectos tales como la capacidad de la línea de transmisión, de la generación y de las condiciones del mercado.

La metodología en este caso de estudio es utilizada como un clasificador no supervisado de perfiles de carga diarios, agrupándolos en 7 clústeres de acuerdo a la categorización de su IoC de acuerdo a la Tabla 4.5 del Capítulo 4. Así por ejemplo, el primer clúster está conformado por los perfiles de carga típicos ($\text{IoC} = 0$), el segundo clúster contiene los perfiles de carga anómalos del tipo 1 ($0 < \text{IoC} \leq 0.03$) y así hasta llegar a los perfiles de carga anómalos del tipo 6, los cuales presentan cambios extremos con respecto al patrón ($0.78 < \text{IoC} \leq 1$). En la Imagen 5.35 se presentan las agrupaciones de los perfiles de carga de acuerdo a la metodología propuesta. Los resultados obtenidos sugieren que el método SAICC puede utilizarse como un potente clasificador no supervisado de perfiles de carga.

El valor del IoC puede interpretarse de acuerdo a la Tabla 5.15, la cual resulta de un leve cambio en los términos usados en la Tabla 4.5. Por ejemplo, en lugar de *consumo de electricidad* se indica *flujo de potencia*, en lugar de *potencia máxima* se indica *flujo máximo de potencia* y en lugar de *energía diaria consumida* se indica *energía diaria transmitida*.

Caracterización de la demanda de energía mediante patrones estocásticos en las Redes Eléctricas Inteligentes

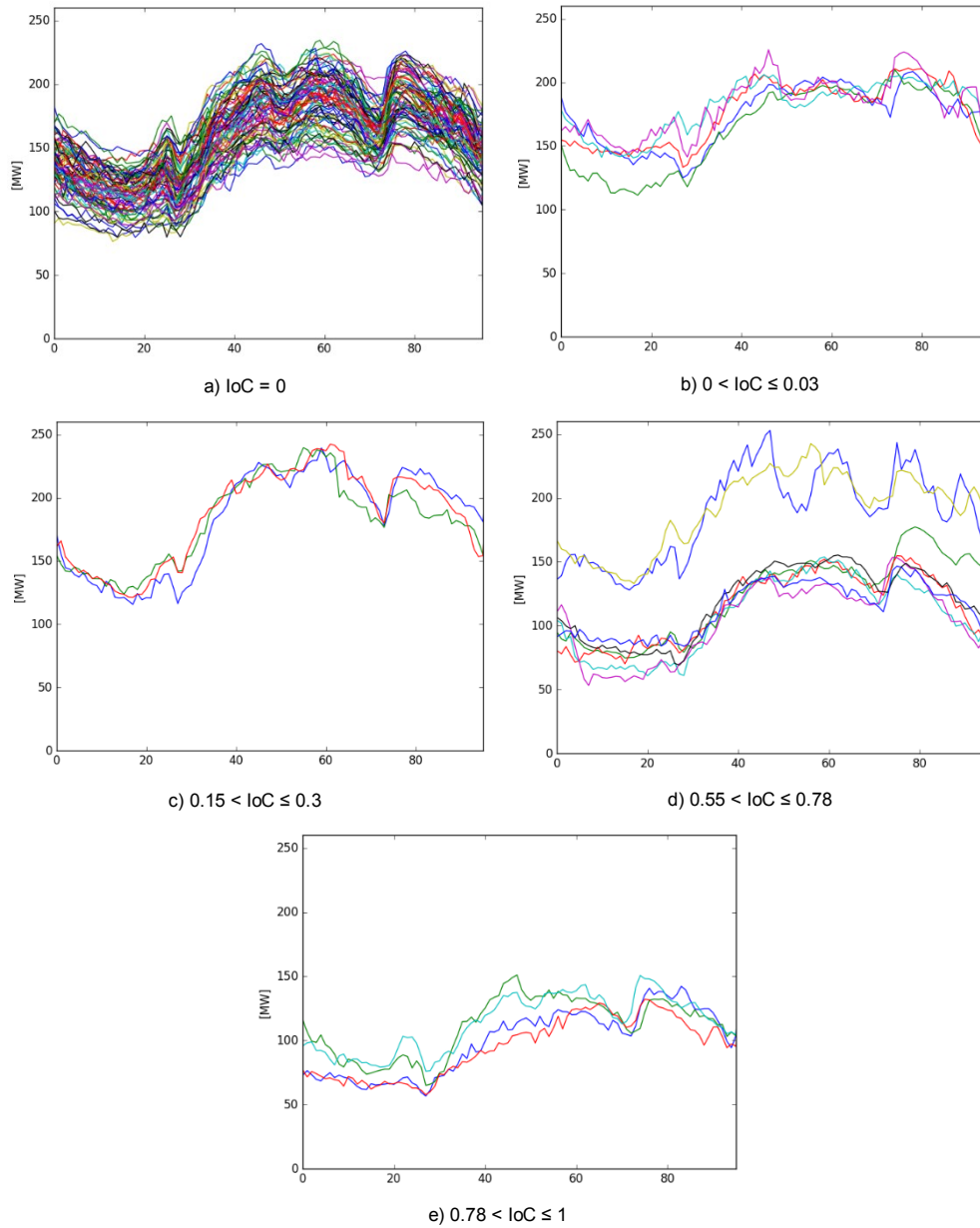


Imagen 5.35 Clasificación de perfiles de carga de acuerdo al IoC de la línea de transmisión Molino-Pascuales (días martes y miércoles).

Tabla 5.15 Interpretación del IoC aplicada a líneas de transmisión

IoC	Cambio en el consumo	Interpretación
0	Ninguno	El flujo de potencia del día analizado se ajusta al patrón.
(0-0.03]	Muy pequeño	Pocos datos atípicos, los cuales no indican cambios significativos en el flujo de potencia.
(0.03-0.15]	Notorio	Existen varios datos atípicos, pero siempre menores que el 40% de todos los datos. Estos datos anómalos no provocan cambios en el flujo de potencia máximo, medio y mínimo.
(0.15-0.3]	Notable	Muchos datos atípicos. No se evidencia gran diferencia en la energía diaria transmitida con respecto al patrón. Sin embargo, el valor del flujo de potencia máximo o mínimo está fuera el rango.
(0.3-0.55]	Grande	Los datos atípicos ocasionan que la energía diaria transmitida esté fuera de rango con respecto al patrón.
(0.55-0.78]	Muy grandes	La presencia de valores atípicos durante gran parte del día ocasiona que la energía diaria transmitida esté fuera de rang. Asimismo el flujo de potencia máximo o mínimo es atípico.
(0.78-1]	Extremo	Todos los parámetros analizados están fuera de rango. El flujo de potencia es totalmente atípico.

A modo de ejemplo, a continuación se presentan los tres perfiles de carga agrupados en el *clúster c* de la Imagen 5.35, cuyos índices de cambio están dentro del intervalo (0.15-0.3]. La Imagen 5.36 muestra estos tres perfiles en contraste con el patrón de consumo para los días martes y miércoles.

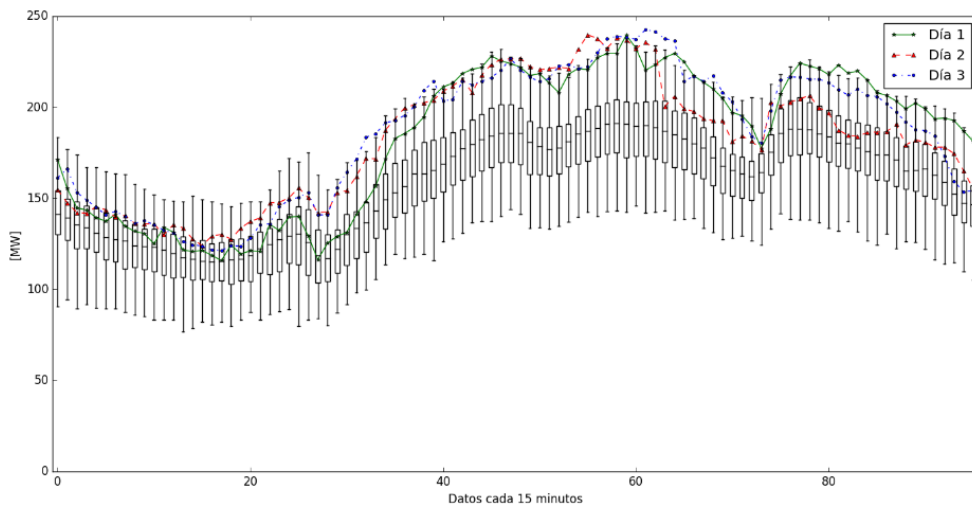


Imagen 5.36 Flujos de potencia del clúster c versus su patrón

Para este mismo caso, los flujos de potencia máximo, medio y mínimo se representan en la Imagen 5.37. Se observa que la potencia máxima de cada perfil excede el patrón, por tanto se confirma que el IoC cataloga y clasifica los perfiles de carga de manera adecuada de acuerdo a la Tabla 5.15.

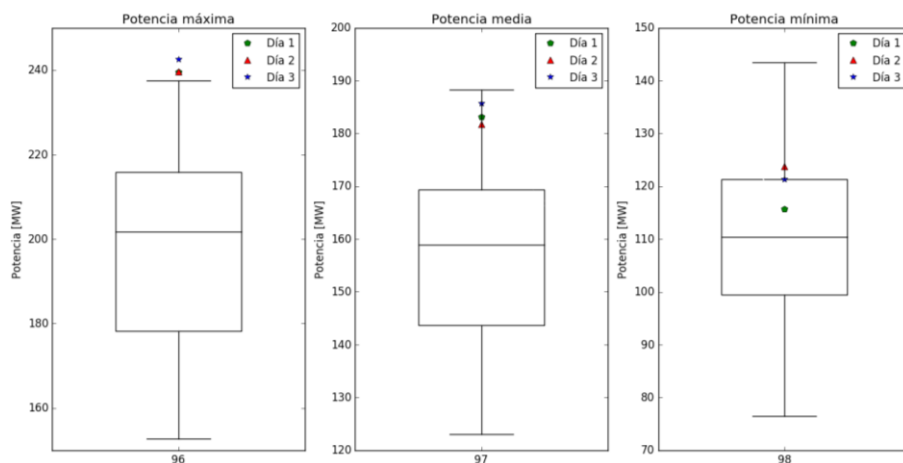


Imagen 5.37 Flujos de potencia máximo, medio y mínimo en contraste con el patrón

5.6.5 Análisis de los resultados

En este apartado se comparan los resultados de la clasificación de perfiles de carga de la línea de transmisión mediante la metodología SAICC y una de las técnicas no supervisadas más extendidas.

Si se considera que los perfiles de carga típicos se agrupan en un mismo clúster, en tanto que los perfiles de carga anómalos se agrupan en otros clústeres, es posible plantear una clasificación de perfiles de carga basados en la técnica *k means* descrita en el apartado 2.4.3.2 del Capítulo 2. Conviene recordar que el reconocimiento de patrones tiene cinco etapas, a) la adquisición de datos, b) preprocesamiento, c) segmentación, d) extracción de características y e) clasificación.

En este caso los datos de flujos de potencia activa cuarto horaria fueron entregados por Celec-Transelectric [107] y corresponden al mismo periodo descrito en el apartado 5.6.1. La etapa de extracción de características consiste en reducir los 96 datos de potencia media cuarto horaria a 23 factores de forma, cuyos detalles se presentaron y publicaron en la *3rd International Conference on Power and Renewable Energy (ICPRE 2018)* en Berlín, Alemania [75]. Por último, la etapa de clasificación se lleva a cabo mediante el algoritmo *k means* y SVM.

Las variables consideradas para la clasificación son los 23 factores de forma establecidos. Se define un número k de grupos y de manera arbitraria k centroides. Los centroides son recalculados hasta alcanzar un valor mínimo basado en una función de costo establecida con la distancia Euclidiana. El número k se afina con el uso de unos coeficientes de silueta. El valor del coeficiente de silueta es una medida de la cercanía de un perfil de carga a otro perfil en el mismo clúster. Cuando un valor es cercano a uno se dice que la instancia de datos está correctamente agrupada. Para determinar el número k de clústeres adecuado se determina el promedio de los valores del coeficiente de silueta de acuerdo al número de clústeres en el que fueron divididos los datos. El valor mínimo de k se establece en 2 y el máximo lo determina la regla de Sturges. El mayor promedio obtenido por el coeficiente de silueta indica el número adecuado de k .

La aplicación de esta técnica en el caso de estudio abordado obtiene 2 clústeres los cuales representan a días de consumo bajo (sábados, domingos y festivos) y días de consumo más elevado (Imagen 5.38). Sin embargo, en primera instancia no es posible conocer en detalle cuál es el comportamiento típico del flujo de potencia activa en la línea de transmisión en un día martes por ejemplo. Para abordar esta limitación, se segmentaron los datos de manera manual, considerando únicamente el flujo de potencia activa de los días martes y miércoles, de tal manera que sea posible comparar los resultados con los obtenidos por la metodología SAICC. En este caso también se obtiene como mejor solución 2 clústeres. El primer clúster está formado por los perfiles de mayor demanda (63 perfiles) y el segundo por los perfiles con potencias menores (41 perfiles), tal como se muestra en la Imagen 5.39.

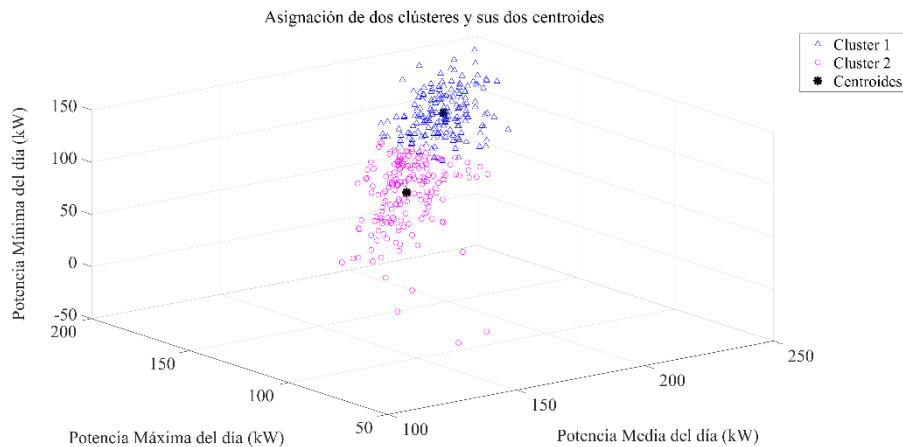


Imagen 5.38 Clústeres formados para los perfiles de carga de la línea de transmisión Molino-Pascuales (01-10- 2017 hasta 30-09- 2018) con la técnica k means

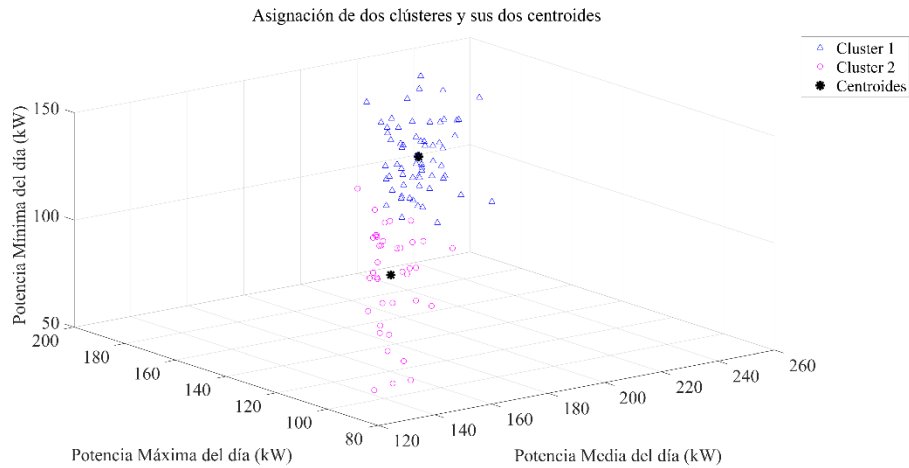


Imagen 5.39 Clústeres formados para los perfiles de carga de la línea de transmisión Molino-Pascuales para los días martes y miércoles (01-10- 2017 hasta 30-09- 2018) con la técnica k means

La clasificación de perfiles de carga mediante la técnica no supervisada propuesta no identifica perfiles anómalos. Por ello se aplican SVM, de tal manera que los perfiles cuyos factores de forma *caen* fuera de la hiperesfera definida se consideran atípicos dentro de cada clúster obtenido por la técnica k means. Con este procedimiento se obtienen 3 perfiles anómalos en el primer clúster y 8 en el segundo, tal como se muestra en la Imagen 5.40 e Imagen 5.41, respectivamente. Sin embargo, no se conoce información alguna acerca del tipo de anomalía detectada.

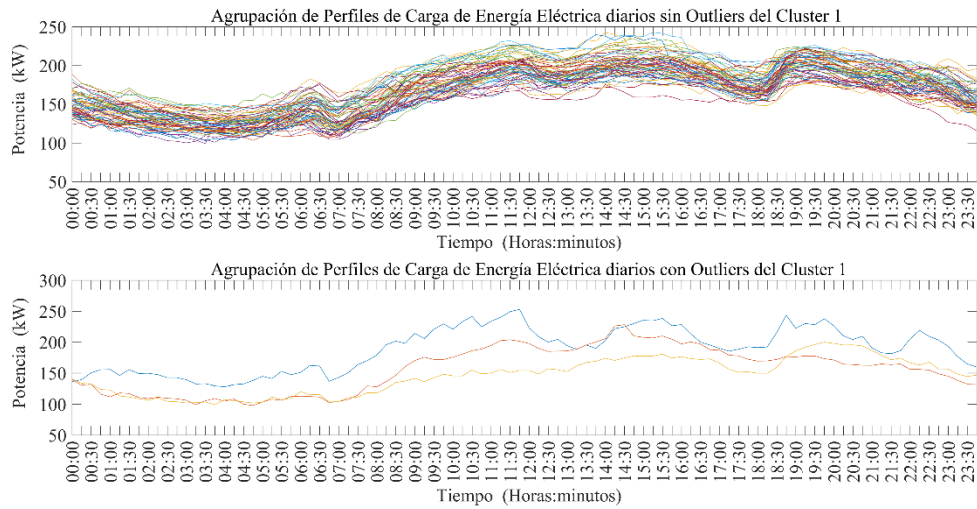


Imagen 5.40 Clasificación de los perfiles de carga de la línea de transmisión Molino-Pascuales Clústeres para los días martes y miércoles utilizando k means y SVM para el clúster 1

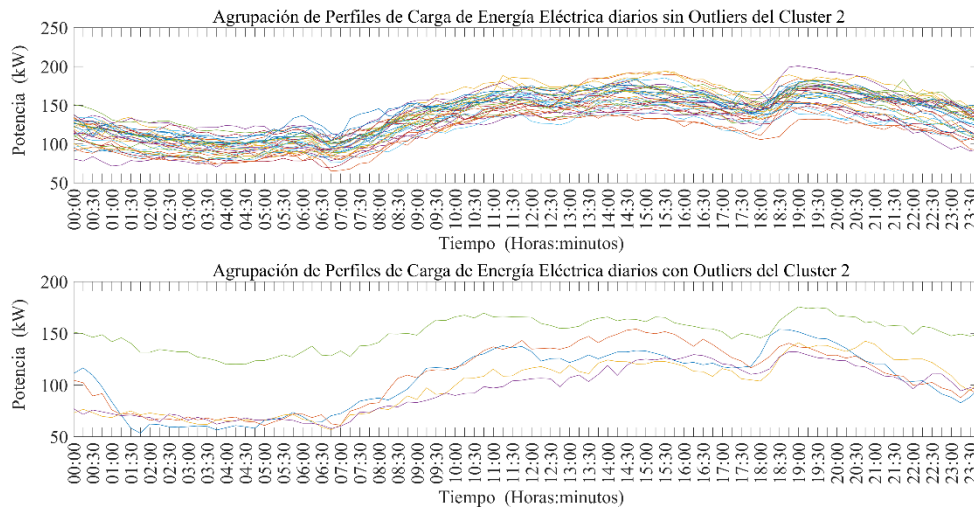


Imagen 5.41 Clasificación de los perfiles de carga de la línea de transmisión Molino-Pascuales Clústeres para los días martes y miércoles utilizando k means y SVM para el clúster 2

De lo anteriormente expuesto se pueden definir algunas ventajas del método SAICC con respecto a otras técnicas no supervisadas de clasificación:

- Las etapas de segmentación y extracción de características del reconocimiento de patrones se realizan automáticamente.

- Los perfiles de carga se clasifican en función de la valoración del cambio que han tenido con respecto a su patrón habitual.
- Las características de cada clúster están claramente definidas por la interpretación del IoC.

En resumen, el método SAICC es una herramienta interesante a la hora de analizar los datos que proporcionan las nuevas REI en sus diferentes puntos. En el caso de las líneas de transmisión es capaz de caracterizar el flujo de potencia al obtener sus patrones estocásticos. En consecuencia, tiene la capacidad de detectar y valorar cambios en el comportamiento de la línea de transmisión. Esta información puede ser utilizada para identificar modos de funcionamiento basados en aspectos como el despacho de energía, maniobras, fallos, variación de la demanda, cambios en la configuración del sistema o las restricciones técnicas o económicas existentes. El método también determina la cantidad de días al año que una línea de transmisión trabaja en determinadas condiciones, lo cual resulta útil para facilitar la operación y planificación de la red eléctrica.

5.7 Aplicación de la metodología SAICC en la predicción de la demanda de electricidad

La predicción de la demanda de electricidad es muy importante en la gestión del suministro de electricidad, siendo clave para la toma de decisiones. El alcance de esta tesis se centra en la predicción a corto plazo, la cual tiene aplicaciones en el despacho de la generación, gestión del mercado spot, aplicación de respuesta de la demanda y en la detección de anomalías, entre otros.

La revisión de la literatura realizada en el Capítulo 2 (apartado 2.3) evidencia que se han desarrollado varias herramientas para la predicción de la demanda de electricidad. Estas herramientas pueden ser determinísticas, no determinísticas (estadísticas o de aprendizaje automático) y mixtas. Dicho estudio también ha permitido identificar que las herramientas basadas en aprendizaje automático, ya sean RNA o de aprendizaje profundo son las preferidas en la actualidad, sin embargo, se han identificado las siguientes debilidades:

- Establecer la estructura de la red no es una tarea sencilla, además utilizan muchos parámetros ajustados sin reglas determinadas.
- La selección de los datos de entrenamiento requiere un esfuerzo considerable.
- Ajustan sus modelos a una determinada instalación, por tanto, el mismo modelo no puede utilizarse para otras.
- Tienen dificultades para acotar la incertidumbre de los resultados obtenidos.

En los siguientes apartados de esta tesis se da respuesta a las debilidades detectadas. Por un lado, se propone utilizar la metodología SAICC para obtener la predicción de la demanda de energía en determinado punto de la red. La predicción se realiza con intervalos de predicción asociados a probabilidades de ocurrencia, cuantificando de esta manera la incertidumbre de las predicciones realizadas. Por otro lado, se utiliza el aprendizaje profundo, específicamente una red neuronal LSTM (una de las técnicas que en la literatura ha obtenido excelentes resultados en la predicción de demanda de electricidad), para comparar sus resultados con los obtenidos por la metodología SAICC.

5.7.1 Aplicación de la metodología SAICC en la predicción de la demanda de electricidad

La revisión del estado del arte acerca de la predicción del consumo de electricidad realizada en el Capítulo 2, así como la experiencia adquirida por el propio autor en [108] indica que la demanda de electricidad puede asociarse a diferentes variables, tales como: la temperatura, precipitación, hora del día, mes, año, laboralidad, etc. Sin embargo, su utilización en sistemas de monitorización y vigilancia puede resultar poco práctica debido a su complejidad, elevado gasto computacional y el requerimiento de predicciones de las otras variables consideradas, lo cual también eleva la incertidumbre. Asimismo, Taylor en [20] indica que para sistemas de predicción a corto plazo basta un modelo univarido. En este contexto, en el presente apartado se utiliza la potencia eléctrica activa demandada como una serie temporal de datos para realizar predicciones. Las predicciones realizadas difieren del enfoque clásico, las cuales predicen valores puntuales intentando minimizar el error. En este caso, por cada día se predicen 96 intervalos de potencia activa absorbida, uno por cada periodo de 15 minutos y se define una probabilidad asociada a que los valores reales se encuentren dentro de cada intervalo obtenido.

5.7.1.1 Metodología

Para la predicción de la demanda de electricidad, la primera etapa de la metodología SAICC explicada en el apartado 4.2.1 del Capítulo 4 sufre dos modificaciones. En primer lugar, en la subetapa de *E.1.2 Recolección de datos*, se agrega una característica adicional, la cual se refiere a la temperatura ambiente promedio diaria asociada al lugar del consumo de energía. La segunda modificación se da en la subetapa de *E.1.8 Separación de datos atípicos*. A más de separar los PCE que tienen valores de potencia anómalos, se segmentan los PCE que tienen valores de temperatura promedio fuera de dos desviaciones estándar con respecto a la media. De esta manera se eliminan perfiles atípicos debido a temperaturas extremas.

La segunda etapa de la metodología SAICC no se modifica, por lo que provee los patrones estocásticos, tal como se detallada en el Capítulo 4. El patrón de consumo es representado mediante una matriz de patrón final para cada día de la semana, siendo estos datos esenciales para obtener los intervalos de predicción.

Después de llevar a cabo las dos primeras etapas de la metodología SAICC se define cada matriz de patrón final junto con sus características expresadas en vectores: la media, desviación estándar, Z_{min} y Z_{max} . Las medias y las desviaciones estándar muestrales se pueden aproximar a las medias y desviaciones estándar reales cuando el número de datos en la muestra es grande, en este caso se computan columna por columna en cada matriz de patrón final (ver Imagen 5.42).

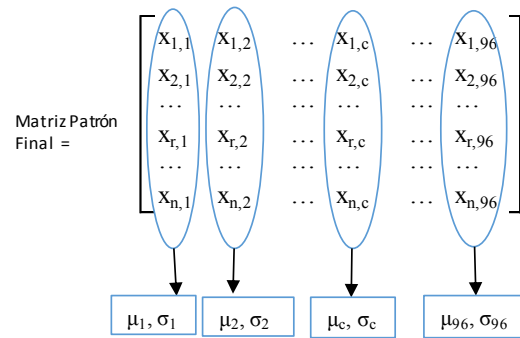


Imagen 5.42 Obtención de medias y desviaciones estándar en la matriz de patrón final

Previo al cálculo de los intervalos de predicción se realiza un ajuste a la media siempre que la transformación de datos mediante el método SAEC se haya llevado a cabo. El ajuste se realiza mediante el siguiente procedimiento:

1. A los valores de la media obtenidos se le sustrae el valor de la carga base, el cual es el percentil 1% de todas las observaciones.
2. A los valores obtenidos en el punto anterior se los multiplica por el valor del índice de estacionalidad (explicado en el apartado 3.5.2.2 del Capítulo 3) de la semana más reciente.
3. A los valores obtenidos en el punto anterior se les suma el valor de la carga base.
4. Se define la probabilidad $(1 - \alpha)$ asociada al intervalo de predicción. Para este estudio se selecciona 60%, 80% y 95%.
5. Se definen los valores $z_{\alpha/2}$ y $z_{-\alpha/2}$ de la distribución normal estándar que cumplen con la siguiente condición:

$$\int_{z_{-\alpha/2}}^{z_{\alpha/2}} n(z; 0,1) dz = 1 - \alpha, \quad \text{Ecuación 5.1}$$

6. Se calcula un intervalo de predicción a partir de la media y desviación estándar. La Ecuación 5.2 muestra la expresión utilizada.

$$\mu - z_{\alpha/2}\sigma\sqrt{1 + \frac{1}{n}} < X_0 < \mu + z_{\alpha/2}\sigma\sqrt{1 + \frac{1}{n}}, \quad \text{Ecuación 5.2}$$

en donde μ es la media, $Z_{\alpha/2}$ es el valor de Z que deja un valor de $\alpha/2$ bajo la curva normal, X_0 representa a la variable aleatoria a predecir y n es el número de datos, en este caso representa el número de PCE diarios dentro de la matriz de patrón final.

Si la transformación de datos no fue realizada, los valores de la media del patrón final no sufren ningún cambio y el procedimiento descrito continúa como se ha indicado.

5.7.1.2 Aplicación a la demanda nacional de Ecuador

En este apartado la metodología SAICC es utilizada para obtener intervalos de predicción de la demanda horaria de electricidad de toda la república del Ecuador. Los datos de la demanda de electricidad utilizados corresponden a las fechas comprendidas entre el 01-01-2017 y el 09-12-2018 y fueron obtenidos a través del Operador Nacional de Electricidad [109].

Los datos de la temperatura fueron obtenidos a través del Instituto Nacional de Meteorología e Hidrología [110]. Debido a que la predicción de la demanda de electricidad que se obtiene es de todo el país, la temperatura adquirida es la media ponderada de las tres ciudades más grandes de Ecuador: Guayaquil, Quito y Cuenca. Los pesos utilizados para este cálculo toman los valores correspondientes a la población total en cada ciudad.

Se ha seleccionado el día domingo 14-10-2018 como el día final para la obtención de los patrones de consumo. El periodo de análisis de datos se fija en 364, 182 y 119 días, debido a que estos valores contienen un número entero de semanas. En cada uno de los tres casos se obtienen los intervalos de predicción y el porcentaje de variación de los intervalos con respecto a su media. La evaluación de los resultados permite seleccionar el periodo de análisis más adecuado.

Intervalos de predicción con un periodo de 364 días para la obtención de patrones

En este caso, siendo el periodo de análisis de 364 días se analizan los datos desde el 16-10-2017 hasta el 14-10-2018. La estacionalidad del consumo es baja como se puede ver la Imagen 5.43.

Caracterización de la demanda de energía mediante patrones estocásticos en las Redes Eléctricas Inteligentes

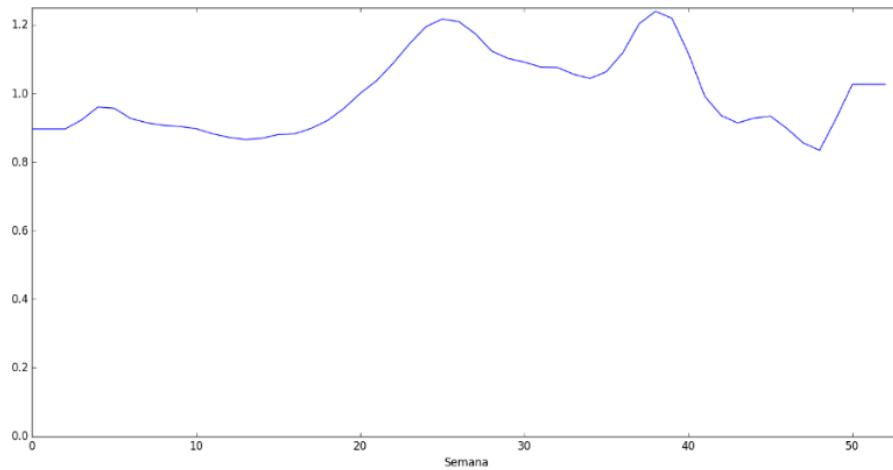


Imagen 5.43 Índice de estacionalidad para un periodo de 364 días

La media del patrón de la demanda de electricidad y la desviación estándar se muestran en la Imagen 5.44 e Imagen 5.45.

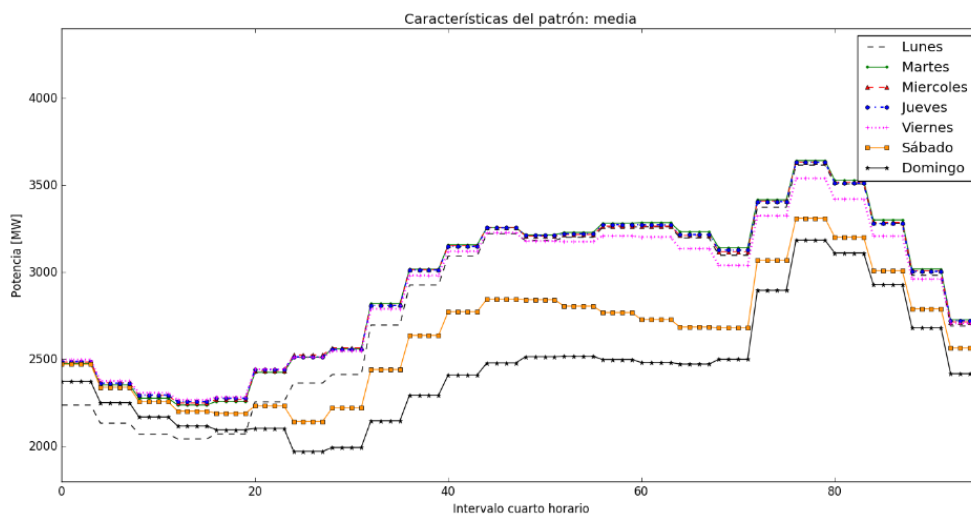


Imagen 5.44 Media del patrón de cada día de la semana desde el 16-10-2017 hasta 14-10-2018 (periodo de obtención de patrones 364 días)

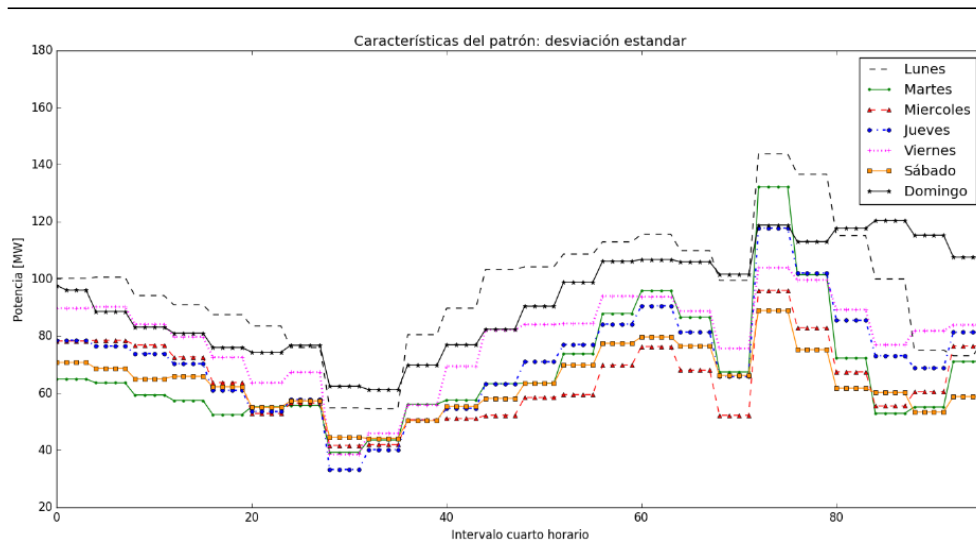


Imagen 5.45 Desviación estándar del patrón de cada día de la semana desde el 16-10-2017 hasta 14-10-2018 (periodo de obtención de patrones 364 días)

Los detalles de la conformación de patrones se pueden ver en la Tabla 5.16.

Tabla 5.16 Conformación de los patrones de electricidad para el periodo de 364 días

Día de la semana	Laboralidad	Días con los que se agrupa	Número de PCE típicos	Número de PCE atípicos	Número total de PCE en patrón definitivo
lunes	Laborable	-	42	6	42
martes	Laborable	miércoles-jueves	33	17	105
miércoles	Laborable	martes-jueves	33	19	105
jueves	Laborable	martes-miércoles	39	12	105
viernes	Laborable	-	35	13	35
sábado	Laborable	-	33	12	33
domingo	No Laborable	-	42	10	42
Total			257	89	346

La Imagen 5.46 muestra los intervalos de predicción obtenidos desde el 15 al 21 de octubre del 2018 en contraste con la demanda de electricidad real.

Caracterización de la demanda de energía mediante patrones estocásticos en las Redes Eléctricas Inteligentes

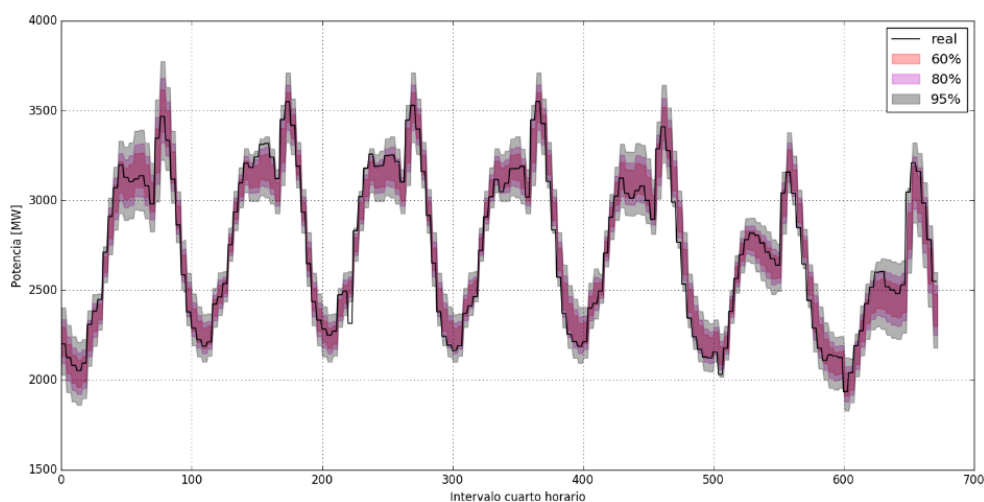


Imagen 5.46 Intervalos de predicción con el 60%, 80% y 95% de probabilidad desde el 15-10-2018 hasta el 21-10-2018 (periodo de obtención de patrones 364 días)

El porcentaje de variación de los intervalos de predicción se muestra en la Tabla 5.17.

Tabla 5.17 Porcentaje de variación de los intervalos de predicción (periodo de obtención de patrones 364 días)

	Rango medio con respecto a la media del patrón [%]			Rango máximo con respecto a la media del patrón [%]			Rango mínimo con respecto a la media del patrón [%]		
	Probabilidad			Probabilidad			Probabilidad		
	60%	80%	95%	60%	80%	95%	60%	80%	95%
Lunes	± 2.9	± 4.4	± 6.7	± 3.8	± 5.8	± 8.9	± 1.6	± 2.5	± 3.8
Martes- Miércoles- Jueves	± 2.1	± 3.1	± 4.8	± 2.9	± 4.4	± 6.8	± 1.4	± 2.2	± 3.3
Viernes	± 2.5	± 3.8	± 5.8	± 3.2	± 4.9	± 7.4	± 1.4	± 2.1	± 3.2
Sábado	± 2.2	± 3.4	± 5.1	± 2.8	± 4.2	± 6.5	± 1.6	± 2.5	± 3.8
Domingo	± 3.2	± 4.9	± 7.5	± 3.8	± 5.8	± 8.8	± 2.4	± 3.7	± 5.6
Total	± 2.4	± 3.7	± 5.7	± 3.8	± 5.8	± 8.9	± 1.4	± 2.1	± 3.2

La Tabla 5.18 muestra la cantidad de datos reales de la demanda de electricidad que *caen* dentro de cada intervalo de predicción obtenido. Así por ejemplo, en esta semana el 83.9% y el 99.4% de todos los valores reales están dentro del intervalo asociado a la probabilidad del 80% y 95% respectivamente.

Tabla 5.18 Resumen de los intervalos de predicción desde el 15 al 21 de octubre de 2018 (periodo de obtención de patrones 364 días)

Probabilidad	Valores dentro del intervalo	Porcentaje [%]	Variación media de los intervalos con respecto a la media del patrón [%]
60%	416	61.90	± 2.4
80%	564	83.93	± 3.7
95%	668	99.40	± 5.7

Intervalos de predicción con un periodo de 182 días para la obtención de patrones

En este caso se utilizan datos de 182 días para conformar los patrones (desde el 16 de abril hasta el 14 de octubre de 2018). El procedimiento seguido es el mismo mostrado para el periodo de análisis de 364 días. Es decir, se obtienen los índices de estacionalidad, media y desviación estándar del patrón para cada día de la semana. Los intervalos de predicción para la semana del 15 al 21 de octubre de 2018 y la demanda de electricidad real para este caso se muestran en la Imagen 5.47.

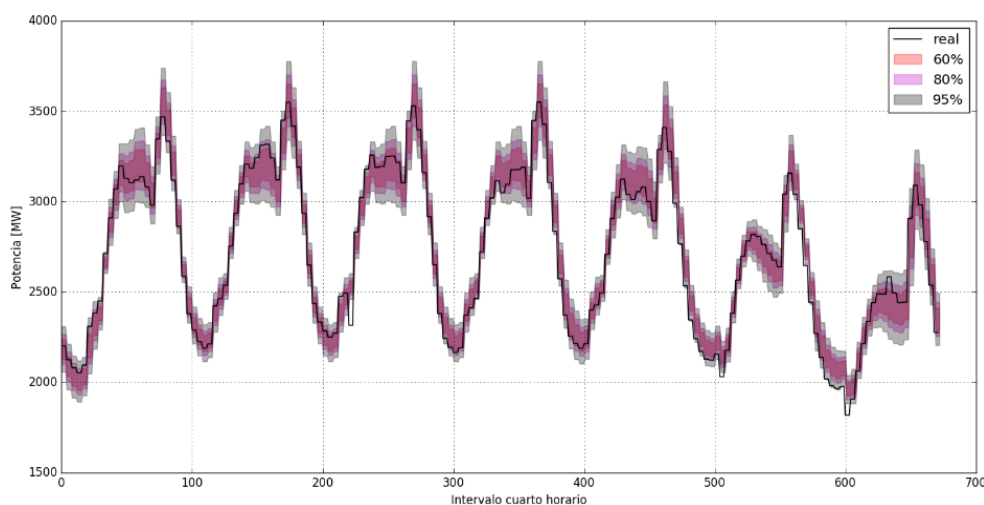


Imagen 5.47 Intervalos de predicción con el 60%, 80% y 95% de probabilidad desde el 15-10-2018 hasta el 21-10-2018 (periodo de obtención de patrones 182 días)

La Tabla 5.19 muestra la cantidad de datos reales de la demanda de electricidad que caen dentro de cada intervalo y el porcentaje de variación media de los intervalos de predicción con respecto a su media.

Tabla 5.19 Resumen de la aplicación de los intervalos de predicción desde el 15 al 21 de octubre del 2018 (periodo de obtención de patrones 182 días)

Probabilidad	Valores dentro del intervalo	Porcentaje [%]	Variación media de los intervalos con respecto a la media del patrón [%]
60%	376	55.95	± 2.1
80%	532	79.17	± 3.2
95%	648	96.43	± 4.9

Intervalos de predicción con un periodo de 119 días para la obtención de patrones

El periodo de obtención de datos con 119 días comprende desde el 18 de junio hasta el 14 de octubre de 2018. En este caso los intervalos de predicción obtenidos, así como la demanda de electricidad se muestran en la Imagen 5.48.

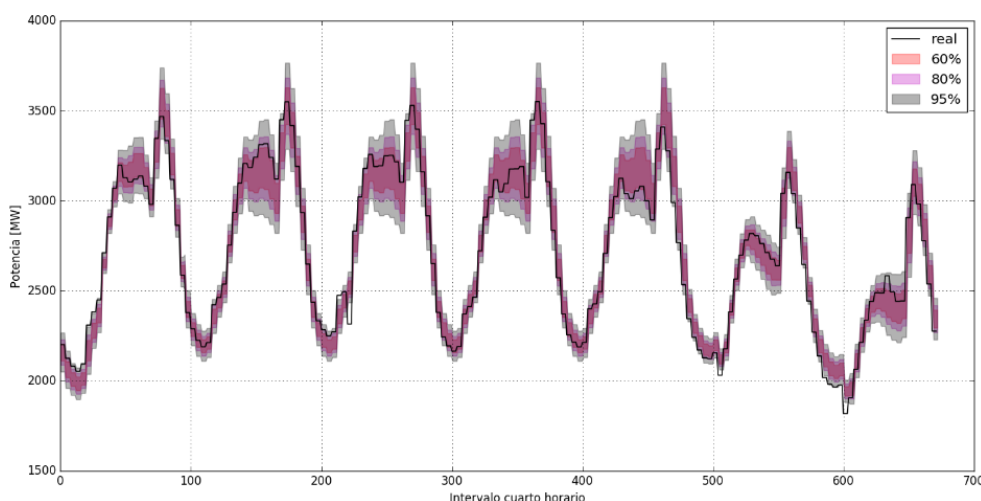


Imagen 5.48 Intervalos de predicción con el 60%, 80% y 95% de probabilidad desde el 15-10-2018 hasta el 21-10-2018 (periodo de obtención de patrones 119 días)

La Tabla 5.20 muestra el número de datos reales de la demanda de electricidad que están dentro de cada intervalo de predicción semanal y el porcentaje de la variación medio de los intervalos con respecto a su media. Se aprecia que el 94.05% de los valores caen dentro del intervalo de asociado a la probabilidad del 95%.

Tabla 5.20 Resumen de la aplicación de los intervalos de predicción desde el 15 al 21 de octubre del 2018 (periodo de obtención de patrones 119 días)

Probabilidad	Valores dentro del intervalo	Porcentaje [%]	Variación media de los intervalos con respecto a la media del patrón [%]
60%	364	54.1	± 2.2
80%	544	80.9	± 3.3
95%	632	94.0	± 5.0

El mismo procedimiento para la obtención de intervalos de predicción para los periodos definidos (364, 182 y 119 días) se aplicó para las 3 semanas siguientes. En la Tabla 5.21 se muestra el resumen de los resultados desde el 15 de octubre hasta el 11 de noviembre de 2018.

Tabla 5.21 Resumen de la aplicación de los intervalos de predicción desde el 15 de octubre al 11 de noviembre del 2018

No. días para el análisis	Probabilidad	Porcentaje de valores dentro del intervalo [%]	Variación media de los intervalos con respecto a la media del patrón [%]
364	60%	60.9	± 2.66
	80%	81.1	± 3.99
	95%	96.7	± 6.04
182	60%	60.0	± 2.32
	80%	77.1	± 3.46
	95%	91.7	± 5.21
119	60%	56.0	± 1.91
	80%	76.3	± 2.99
	95%	91.8	± 4.65

Como conclusión en este apartado, se resalta que el periodo de obtención de patrones con 364 días es el único que proporciona intervalos de predicción en los que el porcentaje de valores reales dentro del intervalo es mayor a la probabilidad asociada.

La Imagen 5.49 muestra mediante un diagrama de dispersión, el porcentaje diario de valores reales que caen dentro de cada intervalo de predicción para cada día predicho. Los cuadrados azules corresponden a una probabilidad de ocurrencia del 60%, los círculos rojos del 80% y los rombos verdes del 95%. El porcentaje de valores reales que caen dentro del intervalo de predicción obtenido con patrones de 364 y 182 días se

representan en el eje vertical y horizontal, respectivamente. Se puede observar que con los patrones formados con 364 días de análisis, el número de días con porcentaje de valores reales que están dentro del intervalo de predicción es sustancialmente mayor al obtenido por los patrones con los 182 días. Específicamente, para el caso de la probabilidad del 95%, se puede ver que todos los puntos están por encima de la línea diagonal. Esto indica que, en ningún caso, los intervalos de predicción obtenidos con patrones de 182 días han conseguido mejores resultados que los patrones obtenidos con 364 días. Los resultados obtenidos con periodos de obtención de patrones de 119 días son descartados debido al bajo porcentaje de valores reales dentro del intervalo de predicción.

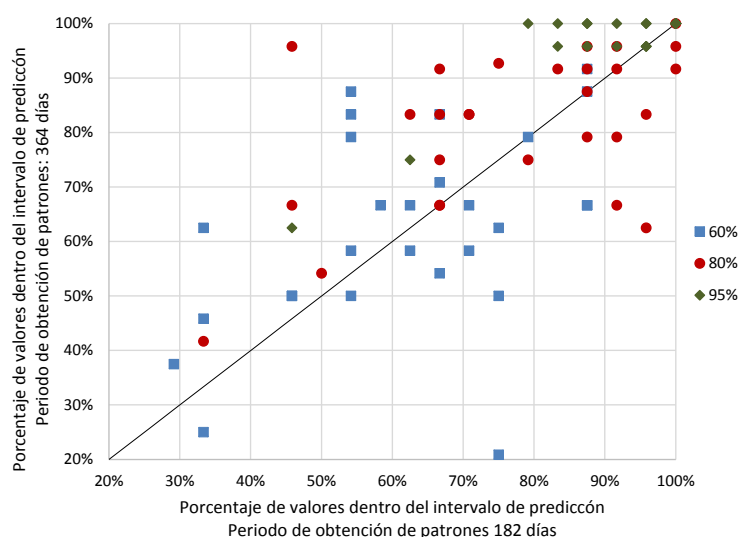


Imagen 5.49 Porcentaje diario de valores reales que caen dentro de cada intervalo de predicción para periodos de obtención de patrones de 364 y 182 días

5.7.1.3 Aplicación a la demanda de España peninsular

En este caso de estudio, la aplicación de la metodología SAICC para obtener intervalos de predicción se realiza sobre la demanda de electricidad de la España peninsular. Los datos de la demanda fueron obtenidos desde el Sistema de Información del Operador del Sistema de Red Eléctrica de España [111] en el periodo comprendido entre el 16-10-2015 hasta el 25-03-2019. Se ha seleccionado el día domingo 20-01-2019 como el día final para la obtención de los patrones de consumo.

En esta ocasión los datos de la temperatura fueron obtenidos desde la Agencia Estatal de Meteorología [112]. Como la predicción de la demanda de electricidad que se realiza es de toda España peninsular, la temperatura adquirida es la media ponderada de

cuatro de las ciudades más grandes y cuya ubicación geográfica dista considerablemente una de otra, como son Madrid, Barcelona, Sevilla y Bilbao. Al igual que en el caso de estudio anterior, los pesos utilizados para el cálculo de la media ponderada se establecen de acuerdo a la población total de cada ciudad.

Un primer análisis exploratorio indica que los intervalos de predicción obtenidos con periodos continuos de datos de 364, 182 y 119 días, como en el caso anterior, arrojan resultados poco satisfactorios. En España el consumo de electricidad tiene una fuerte componente estacional, además los perfiles de carga diarios cambian de forma de acuerdo a la época del año. Por este motivo el porcentaje de valores que están dentro del intervalo de predicción es mucho menor que la probabilidad asociada, siendo necesario definir una nueva forma de escoger los datos para la obtención de los patrones.

Intervalos de predicción con periodos de 9 semanas para la obtención de patrones

Los problemas expuestos anteriormente no ponen en duda la eficacia de la metodología planteada, más bien evidencian que cuando existe una serie temporal de datos altamente estacional se deben escoger los datos adecuadamente. Por tanto, los datos se recolectan de la siguiente manera:

- Se define la semana x a predecir, siendo x el número de la semana del año.
- Se toman los datos de las 4 semanas anteriores a la semana x a predecir.
- Para el año anterior se toman los datos de 4 semanas anteriores a la semana x de ese año, la semana x y 4 semanas posteriores a la semana x .
- Para el año más antiguo del cual se recolectan los datos, se consideran solamente las cuatro semanas posteriores a la semana x , de esta manera se aporta un balance adecuado en la cantidad de datos anteriores y posteriores a la semana x (ver Imagen 5.50).
- Los datos de las cuatro semanas anteriores a la semana x de cada año se evalúan mediante el método SAEC para establecer los índices de estacionalidad, de esta manera se ajusta el intervalo de predicción según la tendencia del consumo siguiendo los pasos detallados en el apartado 5.7.1.1.

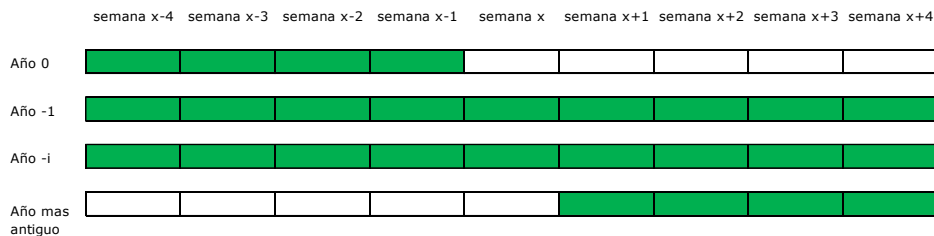


Imagen 5.50 Selección de datos para la obtención de intervalos de predicción para el caso de España peninsular (182 días de análisis)

Para predecir la demanda de electricidad desde del 21 al 27 de enero del 2019, en primera instancia se adquieren los datos correspondientes de 4 años, es decir de 26 semanas o 182 días, dispuestos de la siguiente manera:

- Año 0: 24-12-2018 hasta el 20-01-2019
- Año -1: 25-12-2017 hasta el 25-02-2018
- Año -2: 26-12-2016 hasta el 26-02-2017
- Año -3: 25-01-2016 hasta el 21-02-2016

En la Imagen 5.51 e Imagen 5.52 se puede ver la media y la desviación estándar del patrón de cada día de la semana, con los datos de las fechas seleccionadas.

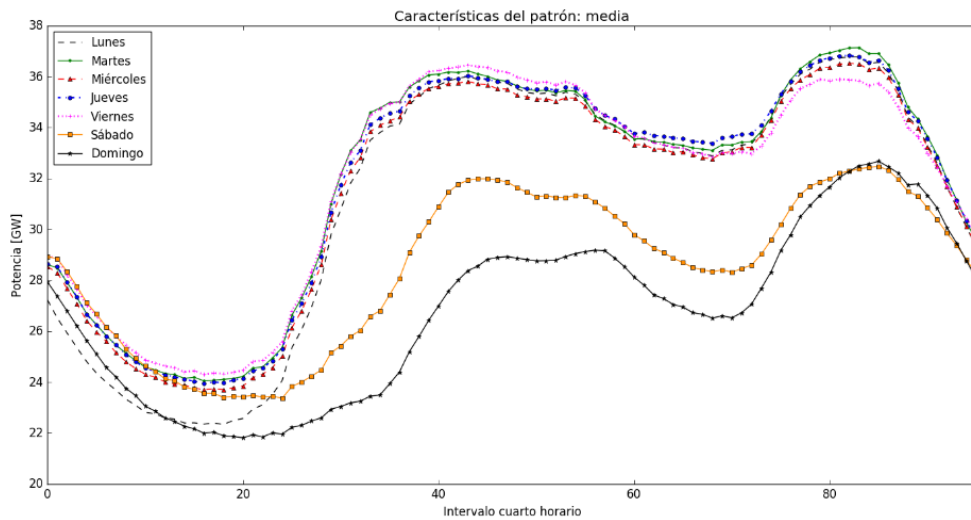


Imagen 5.51 Media del patrón de cada día de la semana (periodo de obtención de patrones 182 días)

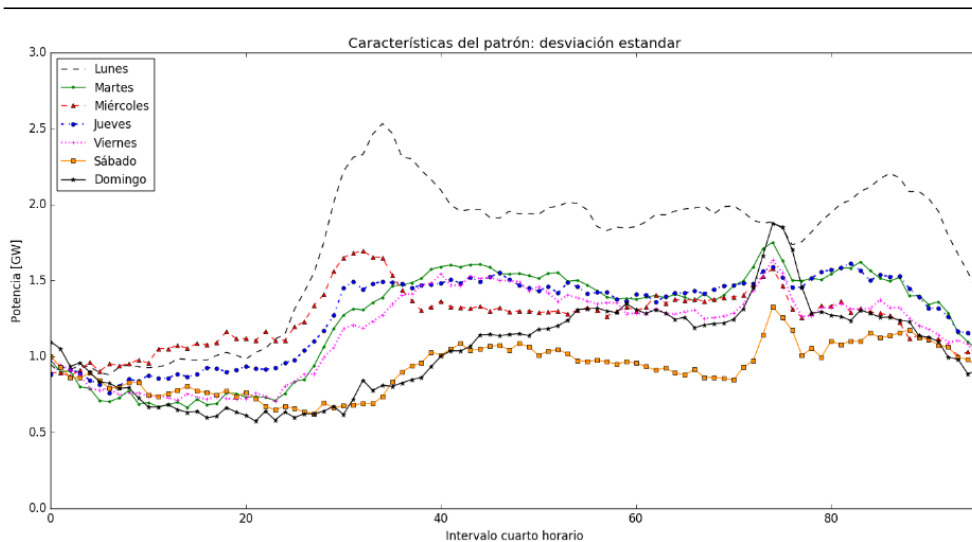


Imagen 5.52 Desviación estándar del patrón de cada día de la semana (periodo de obtención de patrones 182 días)

La conformación de los patrones se detallada en la Tabla 5.22.

Tabla 5.22 Conformación de los patrones de electricidad para el periodo de 182 días

Día de la semana	Laboralidad	Días con los que se agrupa	Número de PCE típicos	Número de PCE atípicos	Número total de PCE en el patrón definitivo
lunes	Laborable	-	21	3	21
martes	Laborable	miércoles-jueves-viernes	19	5	80
miércoles	Laborable	martes-jueves-viernes	21	5	80
jueves	Laborable	martes-miércoles-viernes	21	5	80
viernes	Laborable	martes-miércoles-jueves	19	6	80
sábado	Laborable	-	19	6	19
domingo	No Laborable	-	22	4	22
Total			122	34	156

Los intervalos de predicción obtenidos en este caso se muestran en la Imagen 5.53.

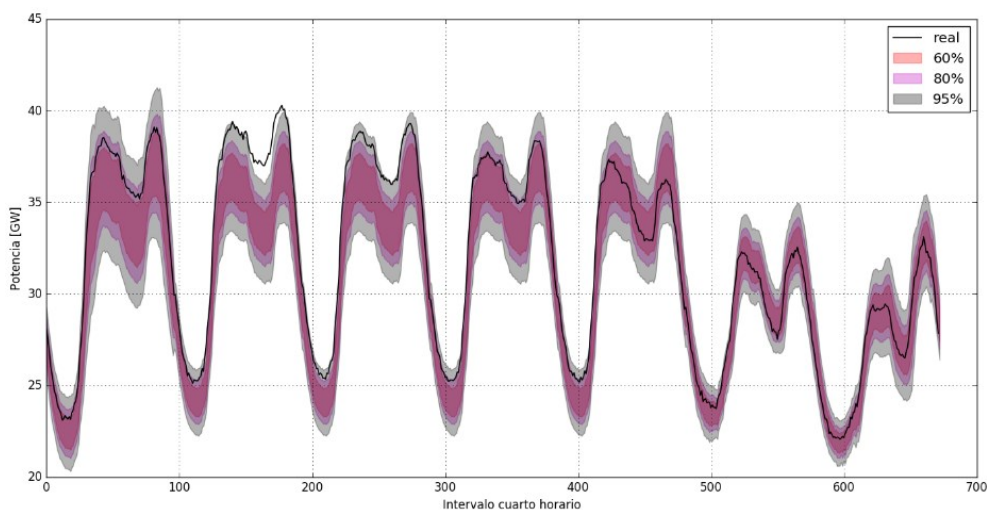


Imagen 5.53 Intervalos de predicción con el 60%, 80% y 95% de probabilidad desde el 21-01-2019 hasta el 27-01-2019 (periodo de obtención de patrones 182 días)

En la Tabla 5.23 se muestra el porcentaje de variación de los intervalos de predicción obtenidos para cada día de la semana con un periodo de obtención de patrones de 182 días. En tanto que en la se presenta el resumen de los resultados.

Tabla 5.23 Porcentaje de variación de los intervalos de predicción (periodo de obtención de patrones 182 días)

	Rango medio con respecto a la media del patrón [%]			Rango máximo con respecto a la media del patrón [%]			Rango mínimo con respecto a la media del patrón [%]		
	Probabilidad			Probabilidad			Probabilidad		
	60%	80%	95%	60%	80%	95%	60%	80%	95%
Lunes	± 4.6	± 7.0	±10.7	± 6.4	± 9.8	±14.9	± 2.9	± 4.4	± 6.7
Martes- Miércoles- Jueves-Viernes	± 3.3	± 5.1	± 7.8	± 4.1	± 6.2	± 9.5	± 2.7	± 4.1	± 6.3
Sábado	± 2.8	± 4.2	± 6.4	± 3.8	± 5.8	± 8.9	± 2.2	± 3.4	± 5.1
Domingo	± 3.3	± 5.0	± 7.6	± 5.6	± 8.6	±13.2	± 2.3	± 3.4	± 5.2
Total	± 3.5	± 5.3	± 8.1	± 6.4	± 9.8	±14.9	± 2.2	± 3.4	± 5.1

Tabla 5.24 Resumen de la aplicación de los intervalos de predicción desde el 21 al 27 de enero del 2019 (periodo de obtención de patrones 182 días)

Probabilidad	Valores dentro del intervalo	Porcentaje [%]	Variación media de los intervalos con respecto a la media del patrón [%]
60%	294	43.75	± 3.5
80%	472	70.24	± 5.3
95%	632	94.05	± 8.1

A continuación se realiza el mismo análisis con datos de 119 días correspondientes a 3 años, cuya distribución es la siguiente:

- Año 0: 24-12- 2018 hasta el 20-01-2019
- Año -1: 25-12-2017 hasta el 25-02-2018
- Año -2: 30-01-2017 hasta el 26-02-2017

Los intervalos de predicción obtenidos se muestran en la Imagen 5.54. Como se puede ver en la Tabla 5.25 el porcentaje de valores reales que caen dentro de los intervalos de predicción es similar al mostrado en la Tabla 5.24. No obstante, el porcentaje de variación de los patrones es visiblemente mayor debido a que una menor cantidad de datos produce un intervalo más grande (ver Ecuación 5.2).

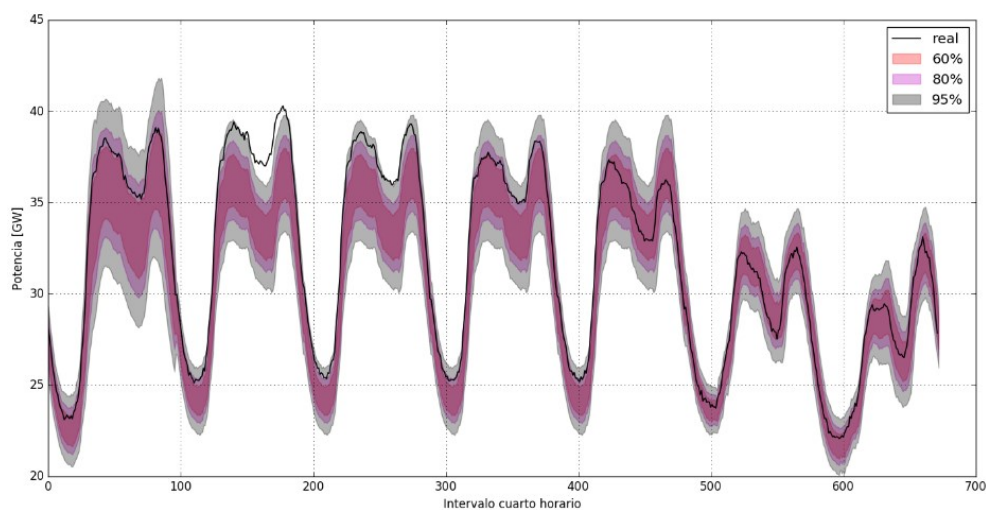


Imagen 5.54 Intervalos de predicción con el 60%, 80% y 95% de probabilidad desde el 21-01-2019 hasta el 27-01-2019 (periodo de obtención de patrones 119 días)

Tabla 5.25 Resumen de la aplicación de los intervalos de predicción desde el 21 al 27 de enero del 2019 (periodo de obtención de patrones 119 días)

Probabilidad	Valores dentro del intervalo	Porcentaje [%]	Variación media de los intervalos con respecto a la media del patrón [%]
60%	290	43.15	± 3.8
80%	481	71.58	± 5.7
95%	625	93.01	± 8.8

Con el fin de generalizar los resultados, el mismo procedimiento se aplicó para las 3 semanas siguientes. Se aplicaron los dos periodos para la obtención de patrones, el primero con 182 días y el segundo con 119 días. En la Tabla 5.26 se muestra el resumen de resultados de los intervalos de predicción obtenidos desde el 21 de enero hasta el 17 de febrero de 2019. En los dos periodos de análisis el porcentaje de valores reales que caen dentro del intervalo de predicción es superior a la probabilidad asociada.

Tabla 5.26 Resumen de los intervalos de predicción obtenidos desde el 21 de enero al 17 de febrero del 2019 para España peninsular

No. Días para el análisis	Probabilidad	Porcentaje de valores dentro del intervalo [%]	Variación media de los intervalos con respecto a la media del patrón [%]
182	60%	64.14	± 4.82
	80%	86.98	± 6.34
	95%	98.10	± 8.67
119	60%	69.01	± 5.23
	80%	89.43	± 6.88
	95%	98.16	± 9.41

Al observar el diagrama de dispersión de la Imagen 5.55 no es evidente qué periodo de obtención de patrones consigue mayor porcentaje de valores reales dentro. Sin embargo, los intervalos de predicción obtenidos con 182 días de análisis tienen menor variación, por lo que este periodo es utilizado para posteriores análisis.

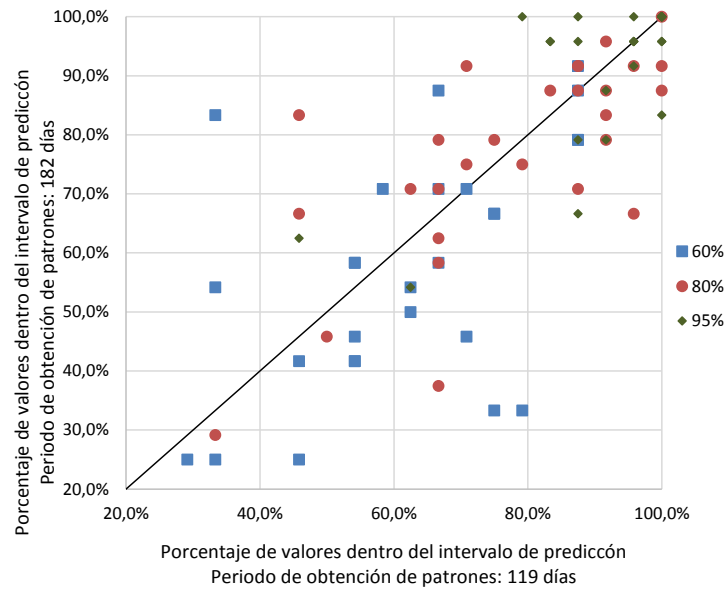


Imagen 5.55 Porcentaje diario de valores reales que están dentro de cada intervalo de predicción para periodos de obtención de patrones de 182 y 119 días

5.7.1.4 Análisis de los resultados

Los resultados obtenidos tanto para el caso de la demanda del Ecuador continental y de España peninsular indican que la metodología SAICC desarrollada puede ser utilizada para generar intervalos de predicción. Para el caso analizado en Ecuador, en donde los datos de la demanda de electricidad tienen una componente estacional débil, simplemente se selecciona la cantidad de días previos a la semana a predecir que ingresan al modelo. Al predecir 4 semanas con los datos de 364 días previos se obtiene un 81.14% de valores reales que caen dentro del intervalo de predicción del 80%, con una variación promedio del $\pm 3.99\%$ con respecto al valor medio del intervalo. El intervalo del 80% parece ser el más adecuado en la predicción de la demanda de electricidad.

En la Imagen 5.56 se muestra a modo de ejemplo el intervalo de predicción del 80% de dos semanas y la demanda de electricidad real en el Ecuador continental.

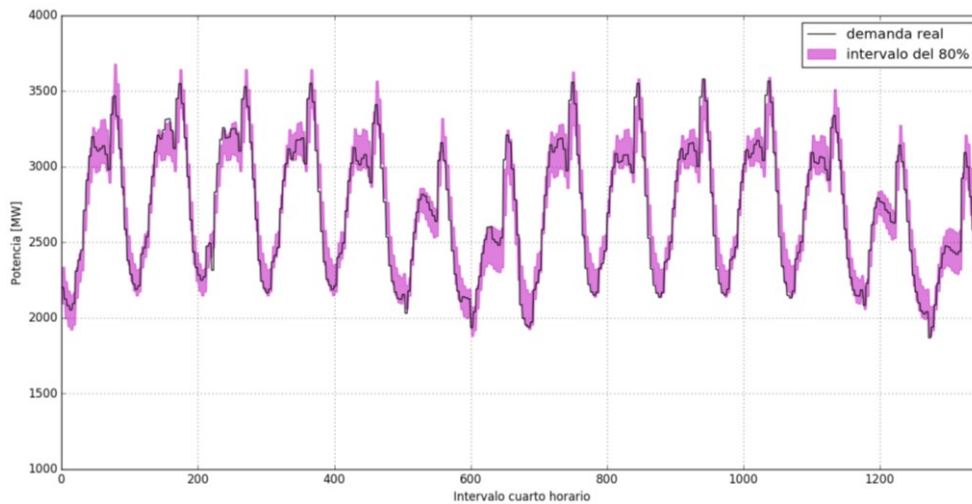


Imagen 5.56 Intervalo de predicción de la demanda de potencia activa en Ecuador continental desde el 15 al 28 de octubre de 2018

Del análisis realizado en la España peninsular se evidenció que, cuando la serie temporal tiene una fuerte componente estacional no es conveniente ingresar datos de periodos continuos prolongados. Para este caso se obtuvieron buenos resultados al ingresar al modelo los datos de las 4 semanas previas a la semana a predecir. Luego, se identifica y se obtienen datos de la semana equivalente a predecir pero del año anterior, de las cuatro semanas previas y las cuatro semanas posteriores (9 semanas). Si se escogen más años se seleccionan las mismas 9 semanas equivalentes. Para el año más antiguo considerado, se utilizan solamente los datos de las cuatro semanas posteriores a la semana equivalente a predecir. Los mejores resultados se obtienen cuando se utilizan datos de 4 años, es decir de 182 días. Para las 4 semanas predichas en este caso, el 86.98% de valores caen dentro del intervalo de predicción del 80%, con una variación promedio del intervalo de predicción del $\pm 6.3\%$ con respecto al valor medio del intervalo. En la Imagen 5.57 se muestra a modo de ejemplo, el intervalo de predicción del 80% y el valor de la demanda de potencia activa real de la España peninsular desde el 04 al 17 de febrero de 2019.

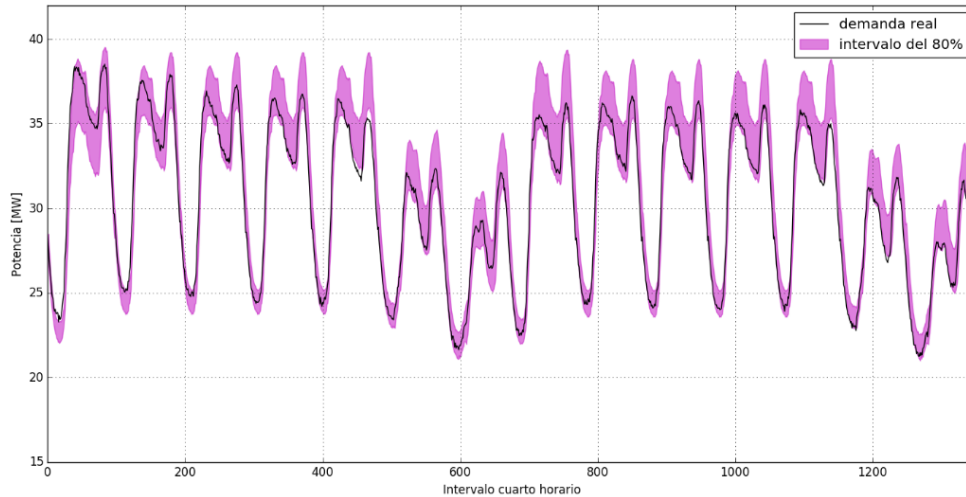


Imagen 5.57 Intervalo de predicción de la demanda de potencia activa en España desde el 04 al 17 de febrero de 2019

5.7.2 Predicción de la demanda de electricidad mediante redes neuronales LSTM

En este apartado se presenta un modelo de RNA LSTM y sus predicciones de la demanda de electricidad con el objetivo de compararlas con las de la metodología SAICC. Las redes LSTM se caracterizan por conseguir muy buenas predicciones en series temporales de datos de energía eléctrica.

5.7.2.1 Estructura de la red LSTM

La RNA LSTM se ha estructurado según el detalle de la Tabla 5.27. La red propuesta obtuvo los mejores resultados después de múltiples pruebas. Se evaluó el ECM para diferente número de capas intermedias y neuronas en cada capa, con el fin de encontrar la configuración que resulte en el menor error de predicción. Este procedimiento se justifica debido a que no existe un método analítico para determinar el número óptimo de capas intermedias y neuronas en una RNA.

El algoritmo de optimización de entrenamiento utilizado es el Adam. Adam fue propuesto en 2014 por Kingma y Ba [113] y se ha convertido en uno de los algoritmos más utilizados en el aprendizaje profundo. Entre sus mayores ventajas están su fácil implementación, bajo costo computacional, baja memoria requerida, finalmente su uso es apropiado cuando los datos presentan gradientes débiles o muy fuertes.

Tabla 5.27 Parámetros de la RNA LSTM

Parámetros	Configuración
Algoritmo de optimización para el entrenamiento	Adam
Número de capas	3
Número de neuronas en la capa de entrada	96
Número de neuronas en la intermedia	100
Números de neuronas en última capa	96
Look back	7 días

Las capas de entrada y de salida tienen 96 neuronas cada una. Las pruebas realizadas demostraron que, en la predicción de la demanda de electricidad el error disminuye cuando se aplica esta configuración. Si bien las redes LSTM son utilizadas ampliamente en sistemas variantes en el tiempo, se ha notado que cuando todos los datos de la serie temporal ingresan por la misma neurona, los errores son considerablemente mayores. Esta problemática se resuelve al introducir una matriz de entrada de datos a la red neuronal de dimensiones $96 \times M$, en donde cada fila representa un dato cuarto horario del PCE diario y M es el número de días que ingresan a la red neuronal. Tal como se ve en la Imagen 5.58, cada neurona de la capa de entrada recibe datos cuarto horarios con una periodicidad de 24 horas. Gracias a esta configuración la red obtiene errores más bajos ya que el consumo de energía en un momento específico del día no difiere en gran medida del consumo en otro día a la misma hora.

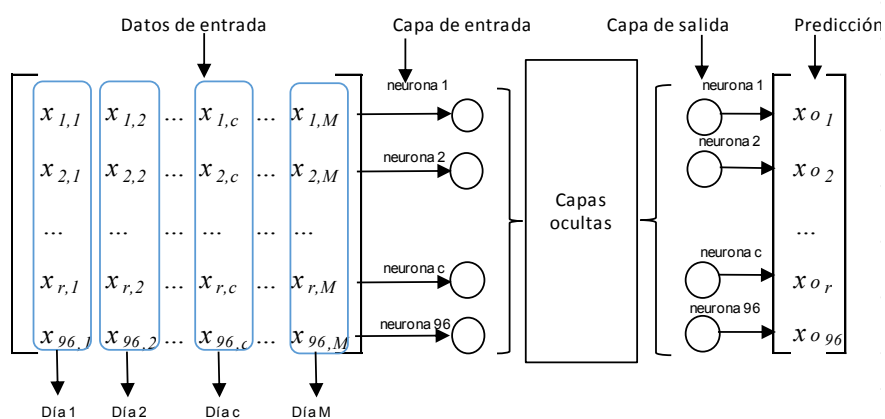


Imagen 5.58 Red neuronal artificial LSTM utilizada para la predicción

Por último, en el parámetro look back (bucle de retorno) se seleccionó 7 días, ya que los datos de la demanda de electricidad tienen un comportamiento cíclico semanal.

5.7.2.2 Resultados para el caso de estudio en Ecuador

El caso de estudio analizado es el mismo expuesto en el apartado 5.7.1.2, en el que se estudia la demanda de electricidad de Ecuador. Los datos disponibles desde el 01-01-2017 y hasta el 09-12-2018 (67970 registros) fueron segmentados de tal manera que, el 70% de datos iniciales se utilizan para el entrenamiento y el 30% restante para la evaluación de las predicciones.

Los resultados de las predicciones realizadas se presentan desde el 10 de septiembre hasta el 09 de diciembre de 2018 (trece semanas), en donde su desempeño se compara con las predicciones realizadas mediante la metodología SAICC. A modo de ejemplo, la Imagen 5.59 muestra las predicciones realizadas entre el 15 y el 21 de octubre de 2018. Las predicciones obtenidas son satisfactorias, obteniendo en este intervalo un valor promedio del PEMA de 2.19%.

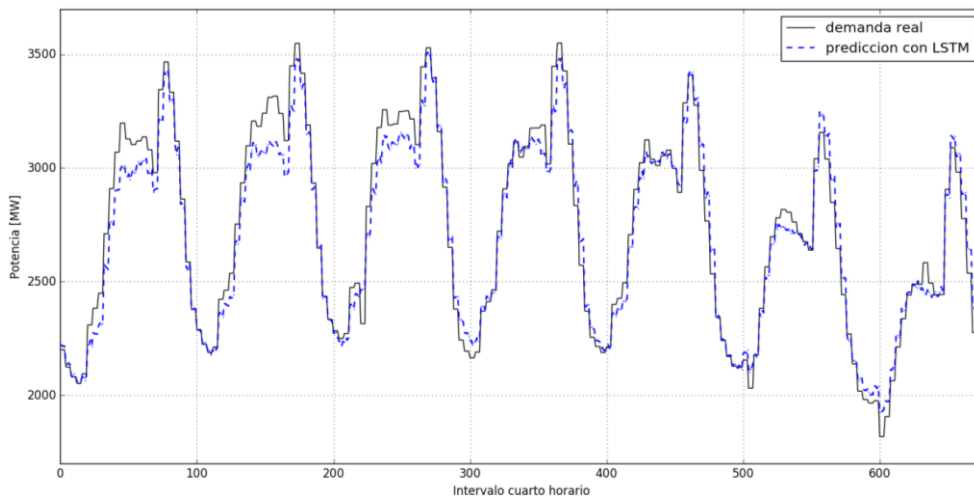


Imagen 5.59 Predicción desde el 15-10-2018 hasta el 21-10-2018 para Ecuador utilizando una RNA LSTM

En las trece semanas evaluadas se evidencia un valor medio del PEMA del 2.73%. El error obtenido es aceptable si lo comparamos con otros estudios. En contraste, existen días en los cuales la predicción de la RNA falla considerablemente. Por ejemplo, en las semanas de predicción 4, 5 y 8 de la Tabla 5.28 se obtienen valores del PEMA considerablemente superiores al valor medio total. Al examinar con mayor detenimiento la semana 8 en la que se producen los errores más elevados en la predicción, se denota

una debilidad de las redes LSTM. Como se ha indicado anteriormente, las redes LSTM tienen la capacidad de memorizar patrones. En este caso, esa característica se convierte en una desventaja, ya que al existir días festivos (del 01 al 04 de noviembre) la red comete errores en sus predicciones. Como se puede ver en la Imagen 5.60, el jueves al ser el primer día festivo se comporta como un sábado y el viernes como un domingo. La red entonces predice el consumo del jueves como un día laborable común, pero como el consumo de ese día es muy parecido a un sábado, el siguiente día es predicho como un domingo cuando en realidad es un viernes. La red neuronal predice entonces el siguiente día (sábado) como si fuera un lunes, de esta manera se acumulan errores, obteniendo una predicción deficiente.

Tabla 5.28 PEMA en la predicción de la demanda de electricidad en Ecuador para trece semanas mediante la RNA LSTM (10-09-2018 hasta 09-12-2018)

Semana	Fecha inicial	Fecha final	Lunes	Martes	Miércoles	Jueves	Viernes	Sábado	Domingo	Media
1	10/09/2018	16/09/2018	2.10%	1.34%	2.12%	1.91%	1.85%	2.60%	1.49%	1.92%
2	17/09/2018	23/09/2018	2.03%	1.79%	2.32%	1.66%	2.49%	2.23%	1.77%	2.04%
3	24/09/2018	30/09/2018	2.34%	1.07%	1.52%	2.52%	1.09%	2.22%	2.33%	1.87%
4	01/10/2018	07/10/2018	3.37%	5.97%	3.17%	3.65%	2.02%	2.12%	2.62%	3.28%
5	08/10/2018	14/10/2018	8.74%	8.10%	2.95%	1.90%	2.35%	4.48%	2.81%	4.48%
6	15/10/2018	21/10/2018	2.15%	2.81%	2.25%	1.43%	1.68%	2.84%	2.16%	2.19%
7	22/10/2018	28/10/2018	3.24%	2.45%	1.19%	1.33%	1.32%	2.68%	1.15%	1.91%
8	29/10/2018	04/11/2018	1.79%	1.32%	2.17%	8.06%	2.96%	11.57%	12.91%	5.82%
9	05/11/2018	11/11/2018	2.53%	2.24%	2.23%	2.92%	3.11%	3.36%	3.12%	2.79%
10	12/11/2018	18/11/2018	3.63%	1.18%	1.40%	1.56%	2.04%	3.33%	2.53%	2.24%
11	19/11/2018	25/11/2018	2.22%	2.55%	2.50%	1.75%	3.44%	4.41%	2.25%	2.73%
12	26/11/2018	02/12/2018	2.57%	1.87%	1.62%	2.19%	1.42%	2.25%	2.34%	2.04%
13	03/12/2018	09/12/2018	2.99%	1.78%	1.15%	2.21%	1.80%	2.65%	2.69%	2.18%
Media total			3.05%	2.65%	2.04%	2.54%	2.12%	3.60%	3.09%	2.73%

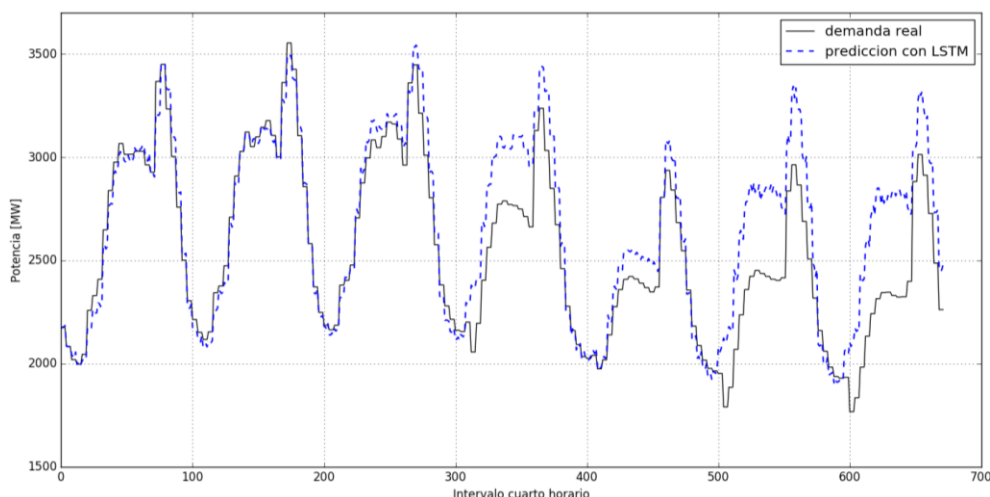


Imagen 5.60 Predicción desde el 29-10-2018 hasta el 04-11-2018 para Ecuador utilizando una red neuronal LSTM

Con la finalidad de contrastar el desempeño de las predicciones realizadas por la RNA LSTM y el método propuesto en el apartado 5.7.1, a continuación se presenta un resumen de los resultados obtenidos por el método derivado de la metodología SAICC. Se obtuvieron los intervalos de predicción de la demanda de electricidad desde el 10 de septiembre hasta el 09 de diciembre de 2018. Tal como se indicó anteriormente, la predicción por intervalos es un enfoque distinto al tradicional, por un lado las RNA obtienen valores puntuales de predicción y su desempeño se evalúa mediante el cálculo de un error. Por el otro, el método SAICC obtiene un valor mínimo y un máximo y una probabilidad asociada a que el valor real caiga dentro de ese intervalo, de tal manera que comparar el desempeño de cada método no es una tarea trivial. Entonces, se propone obtener la media de los intervalos y considerar este valor como la predicción para luego calcular el PEMA que puede compararse con el obtenido por la red LSTM (ver Imagen 5.61). El valor medio del PEMA para las trece semanas evaluadas es de 2.23%, es decir medio punto porcentual más bajo que el error conseguido por la RNA. Estos valores de error se muestran en la Tabla 5.29.

Caracterización de la demanda de energía mediante patrones estocásticos en las Redes Eléctricas Inteligentes

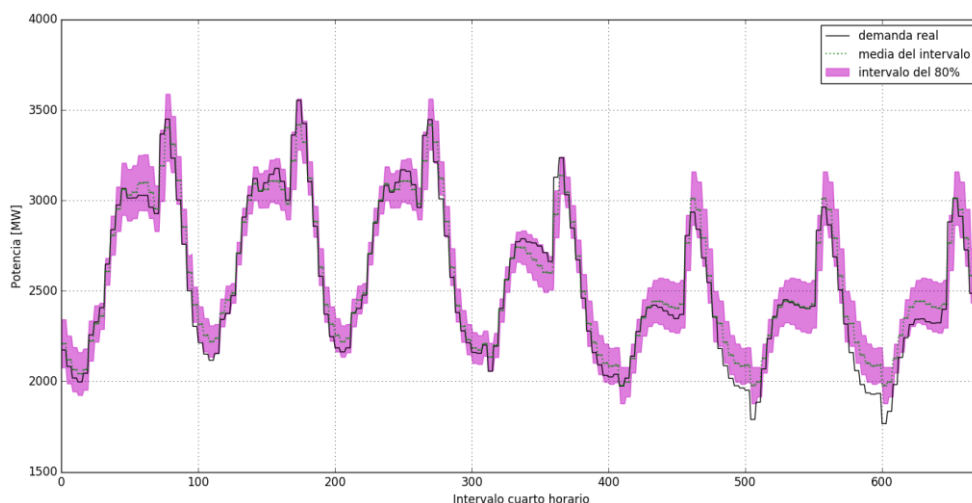


Imagen 5.61 Predicción desde el 29-10-2018 hasta el 04-11-2018 para Ecuador utilizando la media del intervalo de predicción del método SAICC

Tabla 5.29 PEMA en la predicción de la demanda de electricidad en Ecuador para trece semanas mediante el método SAICC (10-09-2018 hasta 09-12-2018)

Semana	Fecha inicial	Fecha final	Lunes	Martes	Miércoles	Jueves	Viernes	Sábado	Domingo	Media
1	10/09/2018	16/09/2018	1.17%	1.28%	1.39%	1.62%	2.30%	1.24%	1.66%	1.52%
2	17/09/2018	23/09/2018	2.54%	1.06%	1.05%	0.97%	1.37%	2.16%	3.17%	1.76%
3	24/09/2018	30/09/2018	1.39%	0.93%	1.71%	2.73%	1.52%	3.13%	3.07%	2.07%
4	01/10/2018	07/10/2018	2.60%	2.26%	1.42%	2.35%	2.71%	1.71%	2.51%	2.22%
5	08/10/2018	14/10/2018	6.83%	2.35%	2.00%	2.08%	1.96%	2.55%	1.47%	2.75%
6	15/10/2018	21/10/2018	1.57%	1.69%	1.94%	2.17%	2.16%	2.24%	3.00%	2.11%
7	22/10/2018	28/10/2018	2.82%	2.77%	2.18%	2.09%	1.84%	1.54%	1.74%	2.14%
8	29/10/2018	04/11/2018	1.95%	2.20%	1.72%	1.79%	1.82%	3.23%	5.37%	2.58%
9	05/11/2018	11/11/2018	2.57%	2.99%	1.70%	2.37%	2.40%	3.91%	1.88%	2.55%
10	12/11/2018	18/11/2018	2.01%	1.02%	1.38%	2.01%	2.05%	2.78%	2.15%	1.91%
11	19/11/2018	25/11/2018	1.40%	2.10%	1.84%	1.43%	2.20%	4.94%	5.05%	2.71%
12	26/11/2018	02/12/2018	3.05%	1.71%	2.51%	1.37%	1.15%	2.26%	3.09%	2.16%
13	03/12/2018	09/12/2018	1.33%	0.73%	1.53%	2.11%	2.64%	3.52%	5.92%	2.54%
Media total			2.40%	1.77%	1.72%	1.93%	2.01%	2.71%	3.08%	2.23%

Al evaluar las predicciones de los 91 días se evidencia que en 55 ocasiones la metodología SAICC obtiene un menor error que la RNA LSTM. En la semana de predicción 8 se consigue la mayor disminución del PEMA, pasando del 5.82% al 2.58%. La reducción se consigue gracias a que en la metodología SAICC se segmentan los datos, de tal forma que se conoce a priori cómo se comporta un día festivo y se lo agrupa generalmente con el consumo de los días sábados y/o domingos. En la Imagen 5.62 se muestra la demanda real de electricidad, la predicción realizada por la red neuronal y el intervalo de predicción del 80% desde el 29 de octubre al 04 de noviembre de 2018.

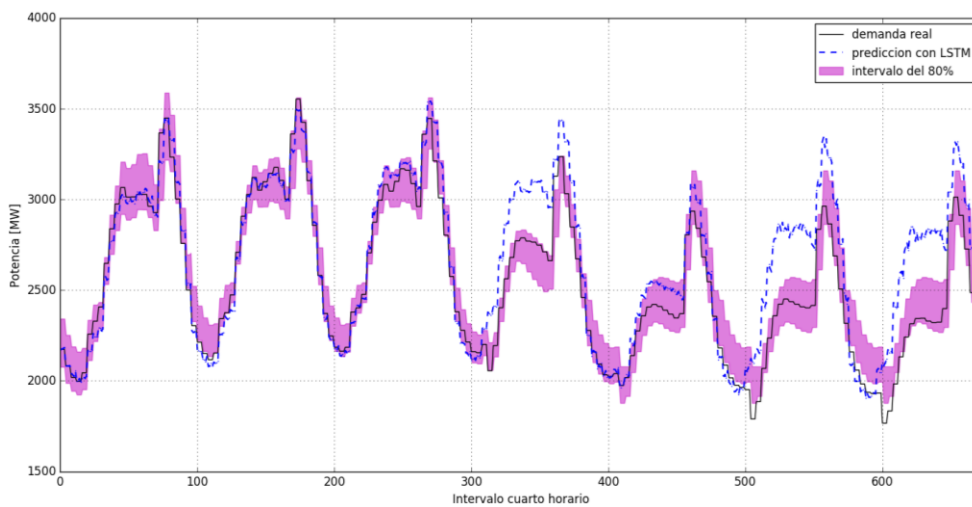


Imagen 5.62 Predicción desde el 29-10-2018 hasta el 04-11-2018 para Ecuador utilizando la RNA LSTM y la metodología SAICC

La predicción con la RNA LSTM desde el 24 al 30 de septiembre de 2018 es la que menor PEMA obtiene con un 1.87%. El PEMA obtenido con respecto a la media del intervalo de predicción en la misma semana es de 2.07%. En la Imagen 5.63 se muestra la predicción de la mencionada semana utilizando la RNA implementada y mediante la metodología SAICC. La predicción por intervalos presenta información adicional, tal como la probabilidad asociada y el porcentaje de variación del intervalo, acotando de esta manera la incertidumbre de la predicción. En esta semana, el 84.52% de los valores de la demanda real caen dentro del intervalo de predicción, en tanto que la variabilidad media semanal del intervalo de predicción del 80% es de $\pm 3.67\%$.

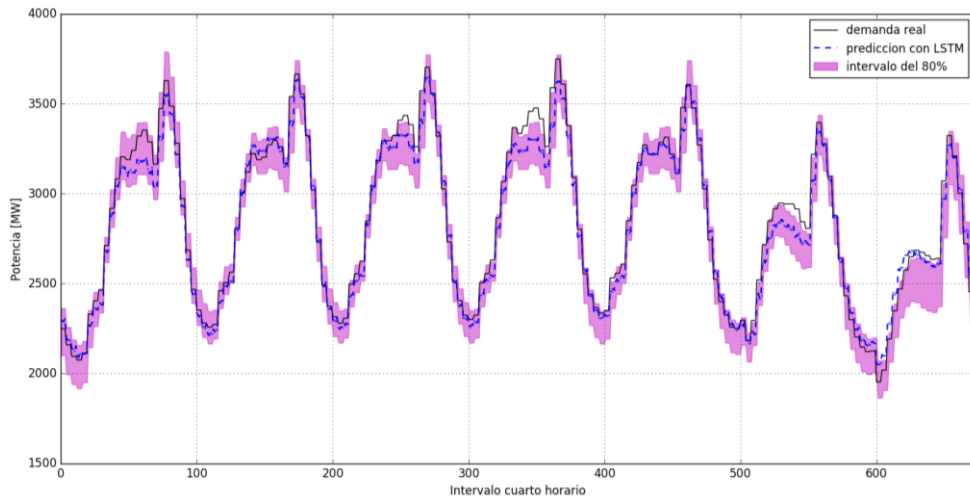


Imagen 5.63 Predicción desde el 24-09-2018 hasta el 30-09-2018 para Ecuador utilizando la RNA LSTM y la metodología SAICC

La metodología SAICC presenta información detallada acerca de las predicciones realizadas. Por ejemplo, para el día viernes 28 de septiembre de 2018, el método proporciona diferentes intervalos de predicción. A modo de ejemplo se han escogido intervalos asociados a probabilidades de ocurrencia de 60%, 80% y 95%, en donde la proporción de valores reales de demanda que han caído dentro de cada intervalo es del 66.7%, 95.8% y 100%, respectivamente (ver Imagen 5.64). También, se conoce el porcentaje de variación diario de cada intervalo de predicción con respecto a su media (Tabla 5.30). Finalmente, es posible conocer la variación cuarto horaria estimada, por ejemplo, para ese viernes entre las 9:30 y 9:45 horas la variación de los intervalos de predicción del 60%, 80% y 95% es de $\pm 1.82\%$, $\pm 2.77\%$ y $\pm 4.23\%$, respectivamente.

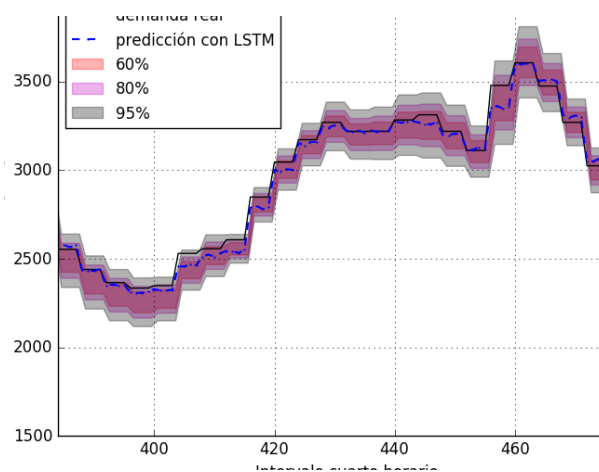


Imagen 5.64 Predicción del día viernes 28 de septiembre de 2018 mediante la RNA LSTM y la metodología SAICC

Tabla 5.30 Porcentaje de variación de los intervalos de predicción con respecto a su media (viernes 28 de septiembre de 2018)

Rango medio con respecto a la media del patrón [±]			Rango máximo con respecto a la media del patrón [±]			Rango mínimo con respecto a la media del patrón [±]		
Probabilidad			Probabilidad			Probabilidad		
60%	80%	95%	60%	80%	95%	60%	80%	95%
2.13%	3.24%	4.96%	2.99%	4.56%	6.97%	1.36%	2.08%	3.18%

5.7.2.3 Resultados para el caso de estudio en España

En este apartado se aplica la RNA LSTM propuesta y la metodología SAICC para predecir una demanda de energía eléctrica altamente estacional como la de la España peninsular. En este caso se cuentan con datos desde el 08 de marzo de 2017 hasta el 24 de marzo de 2019 (71712 registros). El 70% de los datos iniciales se utilizan para el entrenamiento de la red LSTM y el 30% restante para evaluar las predicciones.

A continuación se presentan los resultados de las predicciones realizadas desde el 21 de enero hasta el 17 de febrero de 2019 (el mismo caso de estudio del apartado 5.7.1.3). Al predecir las cuatro semanas con la RNA LSTM se obtiene un valor medio del PEMA del 1.44%. La Imagen 5.65 muestra la predicción realizada entre el 21 y el 27 de enero de 2019, en donde el valor medio del PEMA es de 0.99%. El error obtenido

en este caso de estudio es menor al de Ecuador, lo cual sugiere que este tipo de RNA obtiene buenos resultados con datos que tienen variaciones estacionales importantes.

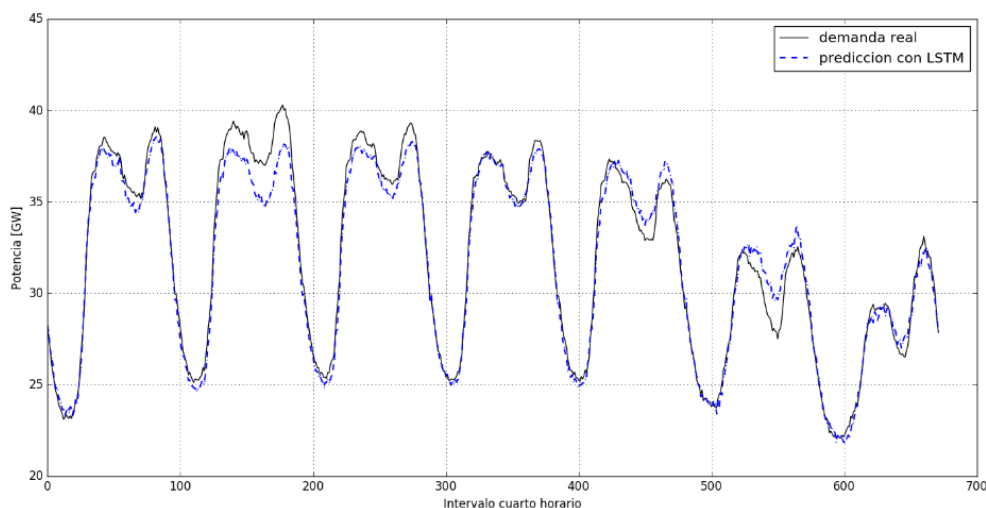


Imagen 5.65 Predicción desde el 21-01-2019 hasta el 27-01-2019 para España peninsular utilizando una RNA LSTM

En la Tabla 5.31 se muestra el PEMA diario obtenido en las cuatro semanas indicadas. Las predicciones tienen errores pequeños al no existir días festivos en este periodo.

Tabla 5.31 PEMA en la predicción de la demanda de electricidad en España peninsular para cuatro semanas utilizando la RNA LSTM (21-01-2019 hasta 17-02-2019)

Semana	Fecha inicial	Fecha final	Lunes	Martes	Miércoles	Jueves	Viernes	Sábado	Domingo	Media
1	21/01/2019	27/01/2019	1.22%	3.51%	1.78%	0.76%	1.89%	2.38%	1.31%	1.84%
2	28/01/2019	03/02/2019	0.90%	1.50%	1.13%	0.91%	1.02%	1.65%	1.73%	1.26%
3	04/02/2019	10/02/2019	1.28%	1.75%	1.52%	0.81%	1.43%	1.50%	1.09%	1.34%
4	11/02/2019	17/02/2019	1.79%	1.18%	1.29%	0.96%	1.39%	1.14%	1.52%	1.32%
Media total			1.30%	1.98%	1.43%	0.86%	1.43%	1.67%	1.41%	1.44%

Al aplicar la metodología SAICC se obtienen los intervalos de predicción. Al igual que en el apartado anterior, con el objetivo de comparar los resultados obtenidos se calcula el PEMA considerando la media del intervalo de predicción como el valor de las predicciones. En este caso el valor promedio del PEMA para las 4 semanas evaluadas es del 2.43%, tal como se indica en la Tabla 5.32. El porcentaje de valores reales que caen dentro del intervalo de predicción del 60%, 80% y 95% es de 64.1%, 87% y 98.1%,

respectivamente. Por otro lado, la variabilidad media de los intervalos de predicción del 60%, 80% y 95% es de $\pm 4.8\%$, $\pm 6.3\%$ y $\pm 8.7\%$, respectivamente.

Tabla 5.32 PEMA en la predicción de la demanda de electricidad en España peninsular para cuatro semanas mediante la metodología SAICC (21-01-2019 hasta 17-02-2019)

Semana	Fecha inicial	Fecha final	Lunes	Martes	Miércoles	Jueves	Viernes	Sábado	Domingo	Media
1	21/01/2019	27/01/2019	4.81%	6.83%	5.98%	4.03%	2.53%	1.29%	0.68%	3.73%
2	28/01/2019	03/02/2019	1.02%	1.24%	0.99%	1.01%	1.73%	1.58%	2.92%	1.50%
3	04/02/2019	10/02/2019	3.04%	1.32%	1.70%	1.75%	2.91%	1.78%	1.20%	1.96%
4	11/02/2019	17/02/2019	2.57%	1.47%	2.09%	2.48%	3.02%	2.94%	3.13%	2.53%
Media total			2,86%	2,71%	2,69%	2,32%	2,55%	1,90%	1,98%	2,43%

Si bien la predicción con la RNA en este caso resultó levemente más precisa, el enfoque mediante intervalos de predicción (apartado 5.7.1) ofrece información adicional que permite acotar la incertidumbre de la predicción, definiendo valores máximos y mínimos asociados a una probabilidad de ocurrencia. A modo de ejemplo en la Imagen 5.66 se muestra la predicción de la demanda de energía eléctrica de la España peninsular mediante la red LSTM y el intervalo de predicción del 80% obtenido mediante la metodología SAICC desde el 28 de enero hasta el 03 de febrero de 2019.

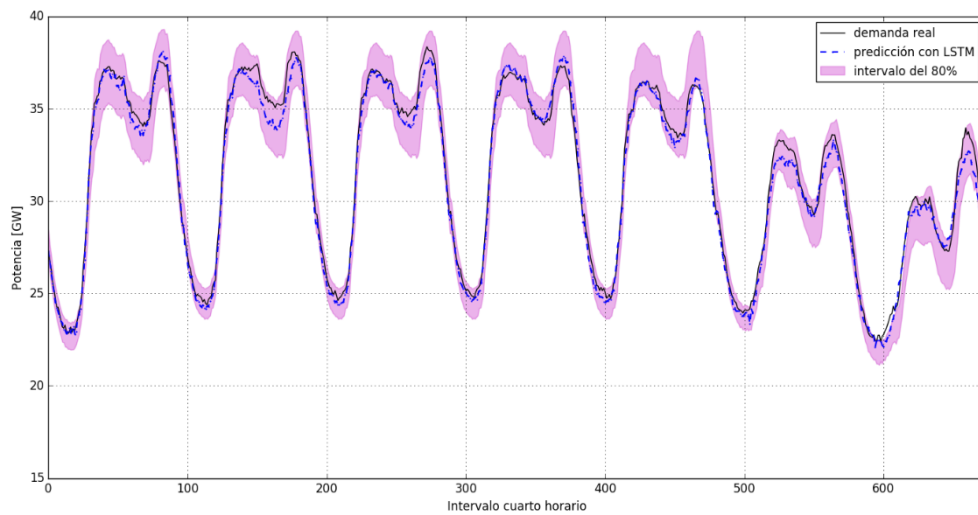


Imagen 5.66 Predicción desde el 28-01-2019 hasta el 03-02-2019 para la España peninsular utilizando la RNA LSTM y la metodología SAICC

5.7.2.4 *Análisis de los resultados*

Los resultados de los casos de estudio analizados han permitido identificar ventajas en la obtención de los intervalos de predicción utilizando la metodología SAICC con respecto a las predicciones realizadas con la RNA LSTM. Entre lo más destacable se encuentra:

- Los intervalos de predicción presentan información que permite acotar la incertidumbre de la predicción.
- Se define un intervalo de predicción y una probabilidad asociada a que un valor real caiga dentro del intervalo.
- Se define un porcentaje de variación del intervalo con respecto a su media para cada intervalo de predicción definido.

La media del intervalo de predicción puede definirse como un valor puntual de predicción, con lo cual es posible calcular el PEMA. Los valores de error son similares a los obtenidos cuando se realizan predicciones con la RNA LSTM.

Las predicciones realizadas por la metodología SAICC son sustancialmente superiores a las realizadas por la RNA LSTM cuando existen consumos atípicos tales como los que ocurren en días festivos. La ventaja radica en que en la metodología SAICC se segmentan los datos de tal forma que se conoce a priori cómo se comporta un día festivo; luego se lo agrupa generalmente con el consumo de los días sábados y/o domingos para obtener los patrones de consumo, reduciendo de esta manera los errores de la predicción. En tanto que, debido su capacidad de memorización la RNA LSTM comente errores cuando existen consumos atípicos.

5.8 Conclusiones del capítulo

Las REI han evolucionado tecnológicamente en muchos aspectos, uno de ellos se refiere a la disponibilidad de datos de la demanda de energía en diferentes puntos de la red. Estos datos son totalmente inútiles si no se obtiene información relevante de ellos. La metodología SAICC representa una potencial herramienta para facilitar la gestión de la demanda de electricidad. En este capítulo esta metodología se aplica a varios casos de estudio reales en diferentes puntos y niveles de las redes eléctricas y en dos países distintos. El análisis inicia con dos consumidores y luego continúa aguas arriba de la red, involucrando a dos alimentadores de diferentes subestaciones eléctricas y una línea de transmisión de 230 kV.

El análisis de la demanda de potencia activa de un edificio en España y la de un grupo de edificios en Ecuador sugiere que la metodología se puede aplicar a medianos o

grandes consumidores y con una demanda de electricidad con componentes estacionales débiles o fuertes. En ambos casos de estudio se obtuvieron resultados similares:

- La metodología obtiene patrones estocásticos e identifica consumos típicos y anómalos.
- Es posible cuantificar el cambio que ha tenido un PCE mediante el IoC.
- Se presentan conclusiones de manera automática, las cuales indican las posibles causas de los consumos anómalos. Alrededor del 60% de las veces el método entrega la misma información que un experto humano, el 30% de las veces el método entrega más información que el experto, en tanto que, en los casos restantes el experto humano entrega información más precisa.

La metodología SAICC puede utilizarse también para el análisis de la demanda de potencia activa en alimentadores de subestaciones, independientemente del tipo de consumidores aguas abajo. En estos casos la metodología es útil para:

- Identificar patrones estocásticos de consumo, detectar anomalías y/o cambios en los perfiles de carga exitosamente.
- Obtener conclusiones automáticas acerca de las posibles causas de las anomalías detectadas en los perfiles de carga.
- Catalogar los perfiles de carga de acuerdo al cambio que ha tenido la demanda en el alimentador con base en el IoC. Es decir, se puede conocer qué tipos de perfiles de carga existen y con qué frecuencia se presentan, lo cual resulta útil para la planificación y gestión de las redes de distribución eléctrica.

En una línea de transmisión el flujo de potencia no solo depende de la demanda, sino de la disponibilidad de los generadores, costos de producción, precios de la energía y flujos de potencia en otras líneas de transmisión de la red eléctrica. Consecuentemente, la aplicación de la metodología SAICC no proporciona conclusiones acertadas acerca de las posibles causas de las anomalías. Sin embargo, al obtener patrones de comportamiento es posible definir tipos de perfiles de carga del flujo de potencia activa de acuerdo a la categorización del IoC. De hecho, la clasificación de perfiles basada en el IoC mostró tres claras ventajas con respecto a otras técnicas no supervisadas de clasificación:

- Las etapas de segmentación y extracción de características del reconocimiento de patrones se realizan automáticamente.
- Los perfiles de carga se clasifican en función de la valoración del cambio que han tenido con respecto a su patrón habitual.
- Las características de cada clúster están claramente definidas por la interpretación del IoC.

La información resultante de la clasificación de perfiles de carga en líneas de transmisión se puede utilizar para:

- Identificar modos de funcionamiento en el sistema eléctrico basado en aspectos tales como, el despacho de energía, maniobras, fallos, variación de la demanda, cambios en la configuración del sistema o las restricciones técnicas o económicas existentes.
- Establecer la cantidad de días al año que una línea de transmisión trabaja en determinadas condiciones, lo cual es útil para facilitar la operación y planificación de la red.

Finalmente, se realizó una modificación a la metodología SAICC propuesta para obtener intervalos de predicción de la demanda total de electricidad en dos países, España y Ecuador. Los resultados fueron contrastados con una red neuronal LSTM muy utilizada en la actualidad en el aprendizaje profundo.

La aplicación de las redes neuronales en la predicción de la demanda de electricidad reveló las siguientes limitaciones:

- Dificultad para cuantificar la incertidumbre de los resultados e interpretarlos físicamente, ya que la demanda de electricidad es una variable estocástica continua.
- Gasto computacional elevado.
- Establecer la estructura y configuración de la red, esto es, número de neuronas, capas, algoritmo de optimización, etc., requiere experticia y tiempo considerable.

La metodología propuesta solventa las limitaciones de las redes neuronales al obtener un intervalo de predicción con una probabilidad asociada a que un valor real caiga dentro del intervalo. Además, presenta el porcentaje de variación del intervalo con respecto a su media, acotando de esta manera la incertidumbre de la predicción.

Las predicciones realizadas por la metodología SAICC son sustancialmente superiores a las realizadas por la RNA LSTM cuando existen consumos atípicos tales como los que ocurren en días festivos. La ventaja radica en que en la metodología propuesta se segmentan los datos de tal forma que se conoce a priori cómo se comporta un día festivo; luego se lo agrupa generalmente con el consumo de los días sábados y/o domingos para obtener los patrones de consumo, reduciendo de esta manera errores en la predicción.

La predicción de la demanda con la utilización de redes neuronales artificiales podría mejorarse al integrar criterios de segmentación y estacionalidad desarrollados en esta tesis, lo cual podría sugerirse como futuras investigaciones.

Los casos de estudio expuestos en este capítulo representan una muestra real de datos y además representativa, en este sentido se evitan posibles sesgos en el análisis, por lo tanto los resultados obtenidos pueden ser generalizados.

Capítulo 6

Conclusiones

En este capítulo se resumen las conclusiones más importantes. Seguidamente, se resaltan las contribuciones principales de este trabajo y sus implicaciones. También se plantean los posibles desarrollos futuros relacionados. Finalmente, se presenta un listado de publicaciones realizadas por el autor durante el desarrollo de la presente tesis doctoral.

6.1 Conclusiones

En esta tesis doctoral se presenta una nueva metodología estadística para caracterizar el comportamiento de la demanda de electricidad basada en el *reconocimiento de patrones*, con capacidad para identificar y obtener patrones estocásticos de consumo, valorar e identificar cambios mediante el análisis estadístico en los perfiles de carga de electricidad (PCE) de los consumidores y de otros puntos o niveles de las nuevas redes eléctricas inteligentes (REI). Esta metodología pretende de manera simple y con bajos requerimientos computacionales mejorar los sistemas de gestión de electricidad al evitar errores humanos y reducir el tiempo y esfuerzo que los técnicos expertos dedican al análisis de información en un escenario en el que el volumen de datos crece día a día.

La recopilación de la base teórica y revisión de la literatura realizada en el *Capítulo 2. Estado del arte*, ha permitido identificar puntos claves en el desarrollo de esta investigación. Los puntos a destacar son los siguientes:

- La ventaja de la estadística en el análisis de la demanda de electricidad radica en la solución justificada de los resultados obtenidos, lo cual contrasta con otros métodos como son los derivados de la inteligencia artificial.
- En la predicción de la demanda de electricidad resulta complejo utilizar muchas variables relacionadas al consumo en un sistema de monitorización en tiempo real. La predicción a corto plazo puede realizarse adecuadamente con

un modelo univariado interrelacionado con criterios adicionales de segmentación.

- La detección de las anomalías contextuales, las cuales son valores atípicos en un determinado contexto pero no en otro, representa un desafío al analizar la demanda de energía eléctrica. La dependencia entre variables o atributos es difícil de cuantificar, por tanto, pueden cometerse errores en la interpretación de los datos.
- Para el análisis de los datos de la demanda de electricidad mediante el reconocimiento de patrones resulta apropiado utilizar las técnicas de clasificación no supervisadas debido al volumen de datos existentes en una REI. La utilización de técnicas supervisadas es limitada debido a que los datos deben estar etiquetados.

Las nuevas REI y sus sistemas de medición generan gran cantidad de datos, por ende se requiere una recopilación, procesamiento y análisis adecuado. El estudio de estos sistemas en el *Capítulo 3. Sistemas de gestión de energía eléctrica y análisis de los datos de partida*, ha permitido identificar algunas de sus características y necesidades más relevantes:

- El sistema de medición es uno de los elementos más importante de una REI, cuya administración es cada vez más compleja. La disponibilidad de información oportuna del consumo de electricidad en toda la cadena de suministro propicia el dinamismo y evolución del mercado eléctrico y por lo tanto un mejor aprovechamiento de los recursos en general. En este contexto, la medición de los parámetros eléctricos y el reconocimiento de patrones es imprescindible para la toma de decisiones.
- Las REI enfrentan nuevos desafíos, por un lado deben entregar información para detectar anomalías, aumentar la eficiencia y rentabilidad. Pero por otro lado deben evolucionar para, ser más seguras en el manejo de los datos, tener capacidad de almacenamiento energético, restablecer el servicio rápidamente ante perturbaciones, permitir generación distribuida y un flujo bidireccional de potencia y finalmente proveer mayor protagonismo al consumidor.

Los datos de potencia eléctrica absorbida, así como sus variables temporales relacionadas representan una información de partida esencial para la gestión. En este sentido, una adquisición adecuada de los datos, su posterior procesamiento y análisis exploratorio son tan importantes como aplicar correctamente cualquier técnica o herramienta posterior en el reconocimiento de patrones, detección de anomalías o predicción de la demanda de electricidad.

Los conceptos abordados acerca de las series temporales permitieron definir a la demanda de electricidad como una serie de datos estacional que responde a un proceso estocástico. El análisis de la muestra requiere definir la distribución de probabilidad que mejor se ajuste las instancias de datos. Los datos adquiridos por los sistemas de medición por si solos no se ajustan a una distribución de probabilidad específica, por tanto es necesario realizar un tratamiento previo.

En esta tesis se proponen tres procedimientos para afrontar la ausencia de una distribución de probabilidad definida en los datos de potencia cuarto horaria:

- *Segmentación*: se realiza en función de dos criterios. El primer criterio considera dos variables categóricas temporales (día de la semana e intervalo cuarto horario), mientras que el segundo criterio utiliza a la laboralidad como una variable categórica de segmentación.
- *Tratamiento de la serie temporal de datos*: se aporta un nuevo método para eliminar la tendencia y desestacionalizar los datos de la demanda de electricidad proporcionando mayor uniformidad en los datos de cualquier tipo de consumidor. A este método se le ha denominado SAEC por sus siglas en inglés (seasonality analysis of electricity consumption).
- *Eliminación de datos atípicos*: la eliminación de valores extremos también contribuye a mejorar la distribución de los datos.

El conocimiento detallado de los patrones de consumo en una instalación o punto de suministro permite caracterizar la demanda, siendo fundamental en la vigilancia y administración de los sistemas eléctricos en general. En el *Capítulo 4* se presenta una nueva metodología estadística denominada SAICC por sus siglas en inglés (statistical assessment for identifying changes in consumption methodology) como contribución principal de esta tesis, la cual es capaz de reconocer patrones de consumo, valorar e identificar cambios en PCE, así como detectar anomalías en la demanda de energía e indicar sus posibles causas. La aplicación de esta metodología puede derivar en beneficios en la gestión y vigilancia de los sistemas, tales como: generar alarmas, reducir costes de mantenimiento y aplicar medidas adecuadas oportunamente cuando se presenten incidentes. Al mismo tiempo, puede reducir el tiempo y esfuerzo que los técnicos expertos invierten en el análisis de datos y evitar errores humanos.

La introducción del método SAEC en la primera etapa de la metodología SAICC mejora la detección de anomalías al identificar, a más de anomalías puntuales y colectivas a las anomalías contextuales, reconociendo periodos de consumo normalmente alto o bajo. El análisis realizado confirma que el método SAEC proporciona patrones de consumo menos variables. Todos estos aspectos mejoran la precisión en la detección de anomalías reduciendo de forma global la tasa de falsos positivos y falsos negativos.

La metodología SAICC también define un IoC (del término anglosajón index of change) que valora de 0 a 1 los cambios que ha tenido el consumo de electricidad respecto a su patrón. La experiencia adquirida en la aplicación del método sugiere que el valor del índice obtenido puede emplearse para clasificar o catalogar los PCE. De esta manera, el método propuesto se puede considerar como una nueva técnica de clasificación no supervisada basada en el cambio que los datos tienen con respecto su comportamiento habitual.

En el *Capítulo 5. Aplicaciones de la Metodología SAICC*, la metodología propuesta se aplicó en varios casos de estudios con datos reales en diferentes puntos y niveles de la red eléctrica y en dos países distintos, evidenciando beneficios en la gestión de cada uno de esos sistemas. El análisis inicia con dos consumidores y posteriormente continúa aguas arriba de la red, involucrando a dos alimentadores de diferentes subestaciones eléctricas y una línea de transmisión de 230 kV.

El análisis de la demanda de potencia activa de un edificio en Valencia, España, un grupo de edificios en Cuenca, Ecuador y dos alimentadores de subestaciones sugiere que la metodología se puede aplicar a medianos y grandes consumidores, a subestaciones y con una demanda de electricidad con estacionalidad débil o fuerte. De la evaluación de estos casos de estudio se concluye que:

- La metodología identifica patrones estocásticos de consumo, permite detectar anomalías y/o cambios en los PCE.
- Es posible obtener conclusiones confiables de manera automática, las cuales indican las posibles causas de los consumos anómalos.
- La metodología es útil para catalogar los PCE de acuerdo al IoC, es decir en función del cambio que han tenido con respecto a su patrón. De tal manera que se puede conocer los tipos de PCE que existen y con qué frecuencia se presentan, lo cual resulta útil en la planificación y gestión de los consumidores y de las redes de distribución eléctrica.

En una línea de transmisión el flujo de potencia no solo depende de la demanda, sino de la disponibilidad de los generadores, costos de producción, precios de la energía y flujos de potencia en otras líneas de transmisión de la red eléctrica. Al obtener patrones de comportamiento es posible definir tipos de perfiles de carga de acuerdo a la categorización dada por el IoC. De hecho, la clasificación de perfiles basada en el IoC mostró tres claras ventajas con respecto a otras técnicas no supervisadas de clasificación:

- Las etapas de segmentación y extracción de características del reconocimiento de patrones se realizan automáticamente.
- Los perfiles de carga se clasifican en función de la valoración del cambio que han tenido con respecto a su patrón.

- Las características de cada clúster están definidas por la interpretación del IoC.

La clasificación de perfiles de carga en líneas de transmisión se puede utilizar para:

- Identificar modos de funcionamiento en el sistema eléctrico basado en aspectos tales como, el despacho de energía, maniobras, fallos, variación de la demanda, cambios en la configuración del sistema o las restricciones técnicas o económicas existentes, etc.
- Establecer la cantidad de días al año que una línea de transmisión trabaja en determinadas condiciones, facilitando la operación y la planificación de la red.

La predicción es una herramienta importante en la gestión de la demanda de electricidad, ya que posibilita la planificación del sistema eléctrico para diferentes horizontes temporales. Al realizar una modificación a la metodología SAICC propuesta, es posible obtener intervalos de predicción asociados a una probabilidad de ocurrencia. Se realizaron pronósticos del consumo de electricidad en España y Ecuador, cuyos resultados fueron contrastados con una RNA (red neuronal artificial) LSTM (del término anglosajón long short term memory) muy utilizada en la actualidad en el aprendizaje profundo. La aplicación de las RNA en la predicción de la demanda de electricidad evidencia las siguientes limitaciones:

- Dificultad para acotar la incertidumbre de los resultados e interpretarlos físicamente.
- Requiere un gasto computacional elevado.
- Se requiere experticia y tiempo considerable para establecer la estructura y configuración de la red, esto es, número de neuronas, capas, algoritmo de optimización, etc.

La metodología propuesta resuelve las limitaciones de las RNA al obtener un intervalo de predicción con una probabilidad asociada a que un valor real caiga dentro del intervalo definido. Además, presenta el porcentaje de variación del intervalo con respecto a su media, acotando de esta manera la incertidumbre de la predicción. Las predicciones realizadas por la metodología SAICC son sustancialmente superiores a las realizadas por la RNA LSTM cuando existen consumos atípicos tales como los que ocurren en días festivos. La ventaja radica en que la metodología propuesta segmenta los datos de tal forma que se conoce a priori el comportamiento de la demanda en días laborables y festivos, reduciendo de esta manera errores en la predicción.

Las metodologías propuestas en esta tesis constituyen poderosas herramientas para la caracterización de la demanda de electricidad en las nuevas REI. Partiendo de la caracterización es posible mejorar los sistemas de gestión de electricidad mediante la vigilancia y monitorización. Los casos de estudio expuestos representan una muestra real

de datos y además representativa, en este sentido se evitan posibles sesgos en el análisis estadístico, por lo tanto los resultados obtenidos pueden ser generalizados.

6.2 Principales contribuciones

Como conclusión, las principales contribuciones de esta tesis se resumen en:

- El desarrollo de una metodología estadística para caracterizar la demanda de electricidad en las nuevas REI basada en el reconocimiento de patrones, con capacidad para identificar y obtener patrones estocásticos de consumo, valorar e identificar cambios en PCE, detectar anomalías en la demanda de energía e indicar sus posibles causas.
- Un método para el tratamiento de datos de consumo de energía eléctrica que aporta mayor uniformidad a los patrones de consumo, mejorando la detección de anomalías al reducir de forma global la tasa de falsos positivos y falsos negativos, gracias a la identificación de anomalías contextuales.
- Una nueva técnica de clasificación no supervisada basada en el cambio que tiene un conjunto de datos con respecto a su comportamiento habitual.
- Una metodología de predicción de la demanda de electricidad basada en intervalos y una probabilidad asociada a que un valor real caiga dentro del intervalo definido, acotando de esta manera la incertidumbre de los pronósticos.

6.3 Trabajos futuros

El trabajo realizado en esta tesis posibilita sugerir algunos futuros desarrollos en el área de estudio:

- En esta tesis se desarrolló una metodología estadística basada en el reconocimiento de patrones para obtener patrones de consumo. Dentro de la metodología se desarrolló un método para ajustar los datos a una distribución normal. En este sentido, resulta interesante explorar la estimación no paramétrica de funciones de densidad de probabilidad (métodos Kernel) para el reconocimiento de patrones.
- El desarrollo de una herramienta que en tiempo real monitorice la demanda de electricidad de una o varias instalaciones, de tal manera que sea posible valorar las mejoras en la gestión obtenidas.
- La planeación del dimensionamiento de subestaciones y redes de distribución mediante el reconocimiento de patrones y métodos de optimización.
- La detección de pérdidas no técnicas en las redes de distribución mediante sistemas de medición inteligente y el reconocimiento de patrones.

- La metodología utilizada en esta tesis para obtener intervalos de predicción presenta un enfoque distinto a los métodos tradicionales. Resulta interesante indagar en técnicas estadísticas o de inteligencia artificial para tratar de disminuir el tamaño de los intervalos de predicción manteniendo la misma probabilidad de ocurrencia.
- La predicción de la demanda con la utilización de redes neuronales artificiales que integren criterios de segmentación y estacionalidad desarrollados en esta tesis.

6.4 Publicaciones

El trabajo de investigación realizado durante el desarrollo de esta tesis se ha visto reflejado en varias publicaciones tanto de congresos como de revistas, las cuales han sido realizadas con la colaboración del Instituto de Ingeniería Energética de la UPV y del Grupo de Investigación en Energías de la UPS:

- Serrano-Guerrero, X., Escrivá-Escrivá, G., & Luna-Romero, S. (2020). A Time-Series Treatment Method to Obtain Electrical Consumption Patterns for Anomalies Detection Improvement in Electrical Consumption Profiles. *Energies*, 13(5), 1046 [67].
- Escrivá-Escrivá, G., Roldán-Blay, C., Roldán-Porta, C., & Serrano-Guerrero, X. (2019). Occasional Energy Reviews from an External Expert Help to Reduce Building Energy Consumption at a Reduced Cost. *Energies*, 12(15), 2929 [114].
- Serrano-Guerrero, X., Siavichay, L. F., Clairand, J. M., & Escrivá-Escrivá, G. (2019, March). Forecasting Building Electric Consumption Patterns through Statistical Methods. In *The International Conference on Advances in Emerging Trends and Technologies* (pp. 164-175). Springer, Cham [115].
- Serrano-Guerrero, X., Escrivá-Escrivá, G., & Roldán-Blay, C. (2018). Statistical methodology to assess changes in the electrical consumption profile of buildings. *Energy and Buildings*, 164, 99-108 [68].
- Mora-Álvarez, Milton., Contreras-Ortiz, P., Serrano-Guerrero, X., & Escrivá-Escrivá, G., (2018). Characterization and Classification of Daily Electricity Consumption Profiles: Shape Factors and k-Means Clustering Technique. In *E3S Web of Conferences* (Vol. 64, p. 08004). EDP Sciences [75].
- Serrano-Guerrero, X., Prieto-Galarza, R., Huilcatanda, E., Cabrera-Zeas, J., & Escrivá-Escrivá, G. (2017, November). Election of variables and short-term forecasting of electricity demand based on backpropagation artificial neural

- networks. In *2017 IEEE International Autumn Meeting on Power, Electronics and Computing (ROPEC)* (pp. 1-5). IEEE [81].
- Pesántez-Pesántez, F., Valdez-Salamea, A., Serrano-Guerreo, X., Orozco-Tupacyupanqui, W., & Escrivá-Escrivá, G. (2017, November). A digital control system for Lighting Energy Consumption Efficiency (LECE). In *2017 IEEE International Autumn Meeting on Power, Electronics and Computing (ROPEC)* (pp. 1-6). IEEE [116].
 - Serrano-Guerrero, X., González-Romero, J., Cárdenas-Carangui, X., & Escrivá-Escrivá, G. (2016, September). Improved variable step size P&O MPPT algorithm for PV systems. In *2016 51st International Universities Power Engineering Conference (UPEC)* (pp. 1-6). IEEE [117].
 - Serrano-Guerrero, X., Narváez, M., Urigüen, C., & Escrivá-Escrivá, G. (2016, September). Quantitative assessment of hybrid systems of heating domestic water based on solar energy in Andean zones of Ecuador. In *2016 51st International Universities Power Engineering Conference (UPEC)* (pp. 1-6). IEEE [118].
 - Serrano-Guerrero, X., & Escrivá-Escrivá, G. (2015). Simulation model for energy integration of distributed resources in buildings. *IEEE Latin America Transactions*, 13(1), 166-171 [119].

Bibliografía

- [1] International Energy Agency, “Statistics data browser.” [Online]. Available: <https://www.iea.org/statistics>. [Accessed: 04-Mar-2019].
- [2] D. A. Lind, W. G. Marchal, and S. A. Wathen, *Statistical Techniques in Business & Economics*. McGraw-Hill/Irwin, 2012.
- [3] R. E. Walpole, R. H. Myers, S. Myers, and K. Ye, *Probabilidad y estadística para ingeniería y ciencias*. 2007.
- [4] R. E. Walpole, R. H. Myers, S. L. Myers, and K. Ye, *Probabilidad y estadística para ingeniería y ciencias*, 9th ed., vol. 08, no. 2. Pearson Educación, 2012.
- [5] V. M. Rueda, V. Henao, J. David, F. Cardona, and C. Jaime, “Recent Advances in Load Forecasting Using Nonlinear Models,” *Dyna*, vol. 78, no. 167, pp. 36–43, 2011.
- [6] Scopus, “Document search,” 2018. [Online]. Available: www.scopus.com.
- [7] R. G. HOOKE, “Forecasting the Demand for Electricity,” *Electr. Eng.* 75(2), 132-132., 1956.
- [8] H. G. Lakhani and B. Bumb, “Forecasting Demand for Electricity in Maryland: An Econometric Approach.”
- [9] O. Derek W Bunn Department of Engineering Science, University of Oxford, Parks Road and U. OXI 3PJ, “Experimental study of a Bayesian method for daily electricity load forecasting,” 1979.
- [10] R. F. Engle, S. J. Brown, G. Stern, and J. Wiley, “A Comparison of Adaptive Structural Forecasting Methods for Electricity Sales,” *J. Forecast.*, vol. 7, pp. 149–172, 1988.
- [11] J. T. Connor, “A Robust Neural Network Filter for Electricity Demand Prediction,” *J. Forecast.*, vol. 15, pp. 437–458, 1996.

-
- [12] J. Ahmad, H. Larijani, R. Emmanuel, M. Mannion, A. Javed, and M. Phillipson, "Energy demand prediction through novel random neural network predictor for large non-domestic buildings," in *2017 Annual IEEE International Systems Conference (SysCon)*, 2017, pp. 1–6.
- [13] G. Escrivá-Escrivá, C. Álvarez-Bel, C. Roldán-Blay, and M. Alcázar-Ortega, "New artificial neural network prediction method for electrical consumption forecasting based on building end-uses," *Energy Build.*, vol. 43, no. 11, pp. 3112–3119, 2011.
- [14] C. Roldán-Blay, G. Escrivá-Escrivá, C. Álvarez-Bel, C. Roldán-Porta, and J. Rodríguez-García, "Upgrade of an artificial neural network prediction method for electrical consumption forecasting using an hourly temperature curve model," *Energy Build.*, vol. 60, pp. 38–46, 2013.
- [15] X. Wang, D. Luo, X. Zhao, and Z. Sun, "Estimates of energy consumption in China using a self-adaptive multi-verse optimizer-based support vector machine with rolling cross-validation," *Energy*, vol. 152, pp. 539–548, Jun. 2018.
- [16] M. W. Ahmad, M. Mourshed, and Y. Rezgui, "Trees vs Neurons: Comparison between random forest and ANN for high-resolution prediction of building energy consumption," *Energy Build.*, vol. 147, pp. 77–89, Jul. 2017.
- [17] S. Ryu, J. Noh, and H. Kim, "Deep Neural Network Based Demand Side Short Term Load Forecasting," *Energies*, vol. 10, no. 1, p. 3, Dec. 2016.
- [18] E. Mocanu, P. Nguyen, M. Gibescu, and W. L. Kling, "Deep learning for estimating building energy consumption," *Sustain. Energy, Grids Networks*, 2016.
- [19] A. Rahman, V. Srikumar, and A. D. Smith, "Predicting electricity consumption for commercial and residential buildings using deep recurrent neural networks," 2017.
- [20] J. W. Taylor, "Short-term electricity demand forecasting using double seasonal exponential smoothing," *J. Oper. Res. Soc.*, vol. 54, no. 8, pp. 799–805, 2003.
- [21] A. Fouquier, S. Robert, F. Suard, L. Stéphan, and A. Jay, "State of the art in building modelling and energy performances prediction: A review," *Renew. Sustain. Energy Rev.*, vol. 23, pp. 272–288, 2013.
- [22] S. Russell and P. Norving, *Artificial Intelligence A Modern Approach*, Third. 2016.
- [23] P. Joshi, *Artificial intelligence with python*. Packt Publishing, 2017.
- [24] I. Nunes da Silva, D. Hernane, L. Bartocci, and S. Franco dos Reis, *Artificial neural networks: a practical course*. Springer, 2017.
- [25] W. S. McCulloch and W. Pitts, "A logical calculus of the ideas immanent in nervous activity," *Bull. Math. Biophys.*, vol. 5, no. 4, pp. 115–133, 1943.
- [26] W. Ertel, *Introduction to artificial intelligence*, Second. Springer, 2017.
- [27] S. Hochreiter, "Long Short-Term Memory," vol. 1780, pp. 1735–1780, 1997.

- [28] X. Ma, Z. Tao, Y. Wang, H. Yu, and Y. Wang, "Long short-term memory neural network for traffic speed prediction using remote microwave sensor data," *Transp. Res. Part C Emerg. Technol.*, vol. 54, pp. 187–197, 2015.
- [29] F. A. Gers, J. Schmidhuber, and F. Cummins, "Learning to forget: Continual prediction with LSTM," *9th Int. Conf. Artif. Neural Networks ICANN '99*, 1999.
- [30] G. Escrivá-Escrivá, "Nuevas herramientas para facilitar la respuesta activa de consumidores en mercados eléctricos liberalizados: implementación y retribución," Universidad Politécnica de Valencia, 2009.
- [31] E. W. Miller, "World Patterns and Trends in Energy Consumption," *J. Geog.*, vol. 58, no. 6, pp. 269–279, 1959.
- [32] P. Sastri, P. K. Rohatgi, and G. N. K. Iyengar, "Energy Analysis of Aluminium Extraction in India: Part I - Energy Consumption Patterns.," *Trans. Indian Inst. Met.*, vol. 35, no. 6, pp. 552–557, 1982.
- [33] D. Balmforth, "ENERGY CONSUMPTION PATTERNS IN THE TEXTILE INDUSTRY : PART I . ENERGY USAGE PRIOR TO WET PROCESSING .," *Am. Dyest. Report.*, vol. 74, no. 7, 1985.
- [34] K. Blok, E. Nieuwlaar, A. J. M. van Wijk, and W. C. Turkenburg, "Industrial Energy Consumption Patterns: Stock-Taking and Application.[INDUSTRIELE ENERGIEVERBRUIKSPATRONEN: INVENTARISATIE EN TOEPASSING .]," *Elektrotechniek*, vol. 63, no. 12, pp. 1203–1208, 1985.
- [35] T. Sampath Kumar, P. Vasudevan, and S. V. Patwardhan, "Pattern of non-commercial energy consumption and availability in the Indian domestic sector- A case study," *Agric. Wastes*, vol. 12, no. 1, pp. 55–60, 1985.
- [36] J. R. McColl-Kennedy, L. D. Dahringer, and O. H. . Yau, "Energy Consumption Patterns of Mature Australian Consumers," *J. Int. Consum. Mark.*, vol. 4, no. 3, pp. 35–61, 1992.
- [37] A. K. Dahiya and P. Vasudevan, "A Field study of energy consumption pattern on small farms," *Energy*, vol. 11, no. 7, pp. 685–689, 1986.
- [38] M. Millar, R. Morrison, and A. Vyas, "TRAVEL CHARACTERISTICS AND TRANSPORTATION ENERGY CONSUMPTION PATTERNS OF MINORITY AND POOR HOUSEHOLDS .," *Transp. Res. Rec.*, pp. 26–38, 1986.
- [39] E. Taşdemiroğlu, "Industrial energy consumption patterns and possible savings in Türkiye," *Energy*, vol. 18, no. 3, pp. 251–258, 1993.
- [40] R. S. Lee and N. Singh, "Patterns in residential gas and electricity consumption: An econometric analysis," *J. Bus. Econ. Stat.*, vol. 12, no. 2, pp. 233–241, 1994.
- [41] H. R. Sadeghi Keyno, F. Ghaderi, A. Azade, and J. Razmi, "Forecasting electricity consumption by clustering data in order to decline the periodic variable's affects and simplification the pattern," *Energy Convers. Manag.*, vol. 50, no. 3, pp. 829–

- 836, 2009.
- [42] M. Goodwin and A. Yazidi, “A pattern recognition approach for peak prediction of electrical consumption,” *Integr. Comput. Aided. Eng.*, vol. 23, no. 2, pp. 101–113, 2016.
- [43] H. Hino, H. Shen, N. Murata, S. Wakao, and Y. Hayashi, “A Versatile Clustering Method for Electricity Consumption Pattern Analysis in Households,” *IEEE Trans. Smart Grid*, vol. 4, no. 2, pp. 1048–1057, 2013.
- [44] Z. Ma, H. Li, Q. Sun, C. Wang, A. Yan, and F. Starfelt, “Statistical analysis of energy consumption patterns on the heat demand of buildings in district heating systems,” *Energy Build.*, vol. 85, pp. 664–672, 2014.
- [45] J. M. Abreu, F. Câmara Pereira, and P. Ferrão, “Using pattern recognition to identify habitual behavior in residential electricity consumption,” *Energy Build.*, vol. 49, pp. 479–487, 2012.
- [46] K. Zhou, C. Yang, and J. Shen, “Discovering residential electricity consumption patterns through smart-meter data mining: A case study from China,” *Util. Policy*, vol. 44, pp. 73–84, 2017.
- [47] K. Zhou, S. Yang, and Z. Shao, “Household monthly electricity consumption pattern mining: A fuzzy clustering-based model and a case study,” *J. Clean. Prod.*, vol. 141, pp. 900–908, 2017.
- [48] G. Stuart, P. Fleming, V. Ferreira, and P. Harris, “Rapid analysis of time series data to identify changes in electricity consumption patterns in UK secondary schools,” *Build. Environ.*, vol. 42, no. 4, pp. 1568–1580, 2007.
- [49] Z. Chijie, Z. Bin, H. U. Jun, L. I. Qiushuo, and Z. Rong, “Anomaly Detection for Power Consumption Patterns Based on Unsupervised Learning,” vol. 36, no. 2, p. 13334, 2016.
- [50] I. Khan, J. Z. Huang, Z. Luo, and M. A. Masud, “CPLP : An algorithm for tracking the changes of power consumption patterns in load profile data over time,” *Inf. Sci. (Ny)*, vol. 429, pp. 332–348, 2018.
- [51] R. O. Duda, P. E. Hart, and D. G. Stork, “Pattern Classification,” *New York: John Wiley, Section*. p. 654, 2000.
- [52] V. Chandola, A. Banerjee, and V. Kumar, “Anomaly Detection: A Survey,” *ACM Comput. Surv.*, vol. 41, no. 3, 2007.
- [53] J. F. Hair, R. E. Anderson, R. L. Tatham, and W. C. Black, *Análisis multivariante*. Madrid: Prentice Hall, 1999.
- [54] L. Alonso Romero and T. Calonge Cano, “Redes neuronales y reconocimiento de patrones.” 2001.
- [55] C. C. Aggarwal, *Outlier analysis*, vol. 9781461463. Springer, 2017.

- [56] D. W. Scott, “Sturges’ rule,” *Wiley Interdiscip. Rev. Comput. Stat.*, vol. 1, no. 3, pp. 303–306, 2009.
- [57] S. Theodoridis and K. Koutroumbas, *Pattern Recognition*. 2006.
- [58] M. Gupta, J. Gao, C. Aggarwal, and J. Han, *Outlier Detection for Temporal Data*, vol. 5, no. 1. 2014.
- [59] F. E. Grubbs, “Procedures for detecting outlying observations in samples,” *Technometrics*, vol. 11, no. 1, pp. 1–21, 1969.
- [60] C. Diaz and J. Hernández, “Smart Grid: Las TICs y la modernización de las redes de energía eléctrica - Estado del Arte,” *Rev. S&T*, vol. 9, no. (18), pp. 53–81, 2011.
- [61] X. Serrano Guerrero, “Desarrollo de un modelo de simulación para la integración energética de recursos distribuidos mediante TRNSYS,” Universidad Politécnica de Valencia, 2012.
- [62] J. Xavier Serrano Guerrero and G. Escriva, “Simulation Model for Energy Integration of Distributed Resources in Buildings,” *IEEE Lat. Am. Trans.*, vol. 13, no. 1, pp. 166–171, Jan. 2015.
- [63] K. S. K. Weranga, S. Kumarawadu, and D. P. Chandima, *Smart Metering Design and Applications*. Springer, 2014.
- [64] H. Sun, N. D. Hatziargyriou, H. V. Poor, L. Carpanini, and M. A. Sánchez Fornié, *Smarter Energy: From Smart Metering to the Smart Grid*. The Institution of Engineering and Technology, 2016.
- [65] J. Estrella, “Implementacion de un Sistema de Medición Inteligente en la Universidad Politécnica Salesiana Sede Cuenca.” Cuenca, p. 101, 2018.
- [66] Empresa Eléctrica Regional Centro Sur C.A., “Centrosur – PORTAL DE LA CENTROSUR,” 2020. [Online]. Available: <http://www.centrosur.gob.ec/>.
- [67] X. Serrano-Guerrero, G. Escrivá-Escrivá, S. Luna-Romero, and J.-M. Clairand, “A Time-Series Treatment Method to Obtain Electrical Consumption Patterns for Anomalies Detection,” *Energies*, 2020.
- [68] X. Serrano-Guerrero, G. Escrivá-Escrivá, and C. Roldán-Blay, “Statistical Methodology to Assess Changes in the Electrical Consumption Profile of Buildings,” *Energy Build.*, vol. 164, pp. 99–108, 2018.
- [69] J. E. Seem, “Pattern recognition algorithm for determining days of the week with similar energy consumption profiles,” *Energy Build.*, vol. 37, no. 2, pp. 127–139, 2005.
- [70] L. C. Hunt, G. Judge, and Y. Ninomiya, “Underlying trends and seasonality in UK energy demand: A sectoral analysis,” *Energy Econ.*, vol. 25, no. 1, p. 93, 2003.
- [71] E. B. Dagum, *The X-II-ARIMA seasonal adjustment method*. Canada Statistics, Seasonal Adjustment and Time Series Staff, 1980.

-
- [72] K. Park and S. Son, "A Novel Load Image Profile-Based Electricity Load Clustering Methodology," *IEEE Access*, vol. 7, pp. 59048–59058, 2019.
- [73] E. W. S. Dos Angelos, O. R. Saavedra, O. A. C. Cortés, and A. N. De Souza, "Detection and identification of abnormalities in customer consumptions in power distribution systems," *IEEE Trans. Power Deliv.*, vol. 26, no. 4, pp. 2436–2442, 2011.
- [74] P. Jokar, N. Arianpoo, and V. C. M. Leung, "Electricity theft detection in AMI using customers' consumption patterns," *IEEE Trans. Smart Grid*, vol. 7, no. 1, pp. 216–226, 2016.
- [75] M. Mora-Alvarez, P. Contreras-Ortiz, X. Serrano-Guerrero, and G. Escrivá-Escriva, "Characterization and Classification of Daily Electricity Consumption Profiles: Shape Factors and k-Means Clustering Technique," *E3S Web Conf.*, vol. 64, p. 08004, 2018.
- [76] J. E. Seem, "Using intelligent data analysis to detect abnormal energy consumption in buildings," *Energy Build.*, vol. 39, no. 1, pp. 52–58, 2007.
- [77] X. Li, C. P. Bowers, and T. Schnier, "Classification of energy consumption in buildings with outlier detection," *IEEE Trans. Ind. Electron.*, vol. 57, no. 11, pp. 3639–3644, 2010.
- [78] A. Capozzoli, M. S. Piscitelli, S. Brandi, and D. Grassi, "Automated load pattern learning and anomaly detection for enhancing energy management in smart buildings," *Energy*, vol. 157, pp. 336–352, 2018.
- [79] S. Borovkova and H. Geman, "Studies in Nonlinear Dynamics & Econometrics Analysis and Modelling of Electricity Futures Prices," *Analysis*, vol. 10, no. 3, 2006.
- [80] J. Janczura, S. Trück, R. Weron, and R. C. Wolff, "Identifying spikes and seasonal components in electricity spot price data: A guide to robust modeling," *Energy Econ.*, vol. 38, pp. 96–110, 2013.
- [81] X. Serrano-Guerrero, R. Prieto-Galarza, E. Huilcatanda, J. Cabrera-Zeas, and G. Escrivá-Escriva, "Election of variables and short-term forecasting of electricity demand based on backpropagation artificial neural networks," *2017 IEEE Int. Autumn Meet. Power, Electron. Comput.*, pp. 1–5, Nov. 2017.
- [82] S. Singh and A. Yassine, "Big Data Mining of Energy Time Series for Behavioral Analytics and Energy Consumption Forecasting," *Energies*, vol. 11, no. 2, p. 452, 2018.
- [83] R. K. Jain, K. M. Smith, P. J. Culligan, and J. E. Taylor, "Forecasting energy consumption of multi-family residential buildings using support vector regression: Investigating the impact of temporal and spatial monitoring granularity on performance accuracy," *Appl. Energy*, vol. 123, pp. 168–178, Jun. 2014.
- [84] P. R. S. Jota, V. R. B. Silva, and F. G. Jota, "Building load management using

- cluster and statistical analyses,” *Int. J. Electr. Power Energy Syst.*, vol. 33, no. 8, pp. 1498–1505, 2011.
- [85] A. Khalid, N. Javaid, and S. Member, “Towards Dynamic Coordination Among Home Appliances Using Multi-Objective Energy Optimization for Demand Side Management in Smart Buildings,” *IEEE Access*, vol. 6, pp. 19509–19529, 2018.
- [86] H. Cai, S. Shen, Q. Lin, X. Li, and H. U. I. Xiao, “Predicting the Energy Consumption of Residential Buildings for Regional Electricity Supply-Side and Demand-Side Management,” *IEEE Access*, vol. 7, pp. 30386–30397, 2019.
- [87] G. Chicco, “Overview and performance assessment of the clustering methods for electrical load pattern grouping,” *Energy*, vol. 42, no. 1, pp. 68–80, 2012.
- [88] I. Khan, J. Z. Huang, A. Masud, and Q. Jiang, “Segmentation of Factories on Electricity Consumption Behaviors Using Load Profile Data,” *IEEE Access*, vol. 4, pp. 8394–8406, 2016.
- [89] O. Y. Al-Jarrah, Y. Al-Hammadi, P. D. Yoo, and S. Muhaidat, “Multi-Layered Clustering for Power Consumption Profiling in Smart Grids,” *IEEE Access*, vol. 5, pp. 18459–18468, 2017.
- [90] G. Fenza and M. Gallo, “Drift-Aware Methodology for Anomaly Detection in Smart Grid,” *IEEE Access*, vol. 7, pp. 9645–9657, 2019.
- [91] H. Shareef, M. S. Ahmed, A. Mohamed, and E. Al Hassan, “Review on Home Energy Management System Considering Demand Responses , Smart Technologies , and Intelligent Controllers,” *IEEE Access*, vol. 6, pp. 24498–24509, 2018.
- [92] P. J. Brockwell, R. A. Davis, and S. E. Fienberg, *Time series : theory and methods*. Springer Science & Business Media, 1991.
- [93] L. Pérez-Lombard, J. Ortiz, and C. Pout, “A review on buildings energy consumption information,” *Energy Build.*, vol. 40, no. 3, pp. 394–398, 2008.
- [94] J. Iwaro and A. Mwashu, “A review of building energy regulation and policy for energy conservation in developing countries,” *Energy Policy*, vol. 38, no. 12, pp. 7744–7755, 2010.
- [95] T. Babaei, H. Abdi, C. P. Lim, and S. Nahavandi, “A study and a directory of energy consumption data sets of buildings,” *Energy Build.*, vol. 94, pp. 91–99, 2015.
- [96] R. Gulbinas, A. Khosrowpour, and J. Taylor, “Segmentation and Classification of Commercial Building Occupants by Energy-Use Efficiency and Predictability,” *IEEE Trans. Smart Grid*, vol. 6, no. 3, pp. 1414–1424, 2015.
- [97] A. Thewes, S. Maas, F. Scholzen, D. Waldmann, and A. Zürbes, “Field study on the energy consumption of school buildings in Luxembourg,” *Energy Build.*, vol. 68, no. PARTA, pp. 460–470, 2014.

-
- [98] E. Annunziata, F. Rizzi, and M. Frey, “European legal status of the improvement of energy efficiency in buildings,” *9th Int. Conf. Eur. Energy Mark. EEM 12*, pp. 1–8, 2012.
- [99] K. Amasyali and N. M. El-gohary, “A review of data-driven building energy consumption prediction studies,” *Renew. Sustain. Energy Rev.*, vol. 81, no. September 2017, pp. 1192–1205, 2018.
- [100] A. B. R. González, J. J. V. Díaz, A. J. Caamaño, and M. R. Wilby, “Towards a universal energy efficiency index for buildings,” *Energy Build.*, vol. 43, no. 4, pp. 980–987, 2011.
- [101] X. Guan, Z. Xu, and Q. S. Jia, “Energy-efficient buildings facilitated by microgrid,” *IEEE Trans. Smart Grid*, vol. 1, no. 3, pp. 243–252, 2010.
- [102] G. Escrivá-Escrivá, “Basic actions to improve energy efficiency in commercial buildings in operation,” *Energy Build.*, vol. 43, no. 11, pp. 3106–3111, 2011.
- [103] A. J. Marszal *et al.*, “Zero Energy Building - A review of definitions and calculation methodologies,” *Energy Build.*, vol. 43, no. 4, pp. 971–979, 2011.
- [104] X. Serrano-Guerrero and G. Escrivá-Escrivá, “Simulation Model for Energy Integration of Distributed Resources in Buildings,” *IEEE Lat. Am. Trans.*, vol. 13, no. 1, pp. 166–171, 2015.
- [105] G. Escrivá-Escrivá, O. Santamaria-Orts, and F. Mugarra-Llopis, “Continuous assessment of energy efficiency in commercial buildings using energy rating factors,” *Energy Build.*, vol. 49, pp. 78–84, 2012.
- [106] J. M. Evans and S. De Schiller, “Promoción de Eficiencia Energética y Uso de Energía Solar en Vivienda del Ecuador,” in *Primer Congreso Internacional y Expo Científica*, 2016.
- [107] “Ecuador Corporación Eléctrica del Ecuador, Unidad de Negocio TRANSELECTRIC,” *Sistema de Transmisión*, 2019. [Online]. Available: <https://www.celec.gob.ec/transelectric/>. [Accessed: 29-Dec-2019].
- [108] X. Serrano-Guerrero, R. Prieto-Galarza, E. Huilcatanda, J. Cabrera-Zeas, and G. Escrivá-Escrivá, “Election of variables and short-term forecasting of electricity demand based on backpropagation artificial neural networks,” *2017 IEEE Int. Autumn Meet. Power, Electron. Comput.*, no. Ropec, pp. 1–5, 2017.
- [109] Operador Nacional de Electricidad, “CENACE,” 2018. [Online]. Available: <http://www.cenace.org.ec/>. [Accessed: 05-Nov-2018].
- [110] “Instituto Nacional de Meteorología e Hidrología,” “No Title,” *RED DE ESTACIONES METEOROLÓGICAS E HIDROLÓGICAS*, 2019. [Online]. Available: <http://www.serviciometeorologico.gob.ec/>. [Accessed: 02-Nov-2019].
- [111] “Red Electrica De España S.A.,” “Análisis | ESIOS electricidad · datos · transparencia.” [Online]. Available: <https://www.esios.ree.es/es/generacion-y->

- consumo/sistemas-no-peninsulares. [Accessed: 26-Mar-2019].
- [112] "Agencia Estatal de Meteorología - AEMET", "AEMET OpenData," *AEMET OpenData*, 2019. [Online]. Available: <http://www.aemet.es/es/portada>. [Accessed: 04-Nov-2019].
- [113] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv Prepr. arXiv1412.6980*, 2014.
- [114] Escrivá-Escrivá, Roldán-Blay, Roldán-Porta, and Serrano-Guerrero, "Occasional Energy Reviews from an External Expert Help to Reduce Building Energy Consumption at a Reduced Cost," *Energies*, vol. 12, no. 15, p. 2929, 2019.
- [115] X. Serrano-Guerrero, L. Siavichay, J. Clairand, and G. Escrivá-Escrivá, "Forecasting Building Electric Consumption Patterns Through Statistical Methods," in *Advances in Intelligent Systems and Computing*, 2020, vol. 1067.
- [116] F. Pesantez-Pesantez, A. Valdez-Salamea, X. Serrano-Guerrero, W. Orozco-Tupacyupanqui, and G. Escrivá-Escrivá, "A digital control system for Lighting Energy Consumption Efficiency (LECE)," *2017 IEEE Int. Autumn Meet. Power, Electron. Comput.*, pp. 1–6, Nov. 2017.
- [117] X. Serrano-Guerrero, J. Gonzalez-Romero, X. Cardenas-Carangui, and G. Escrivá-Escrivá, "Improved variable step size P&O MPPT algorithm for PV systems," in *Power Engineering Conference (UPEC), 2016 51st International Universities*, 2016, pp. 1–6.
- [118] X. Serrano, M. Narváez, C. Urigüen, and G. Escrivá, "Quantitative Assessment of Hybrid Systems of Heating Domestic Water Based on Solar Energy in Andean Zones of Ecuador," in *Power Engineering Conference (UPEC), 2016 51st International Universities*, 2016, pp. 1–6.
- [119] J. Serrano and G. Escrivá, "Simulation Model for Energy Integration of Distributed Resources in Buildings," *IEEE Lat. Am. Trans.*, vol. 13, no. 2015, pp. 166–171, 2016.