

**UNIVERSIDAD POLITÉCNICA SALESIANA**  
**SEDE CUENCA**

**CARRERA DE INGENIERÍA ELÉCTRICA**

*Trabajo de titulación  
previo a la obtención del  
título de Ingeniero Eléctrico*

**PROYECTO TÉCNICO CON ENFOQUE INVESTIGATIVO:**

**“PRONOSTICO DE PERFILES DE CONSUMO DE ELECTRICIDAD A  
TRAVÉS DE MODELOS ESTADÍSTICOS Y DESARROLLO DE UNA  
METODOLOGÍA COMPARATIVA”**

**AUTORES:**

PAUL ESTEBAN PAUCAR UYAGUARI

LUIS AGUSTIN YUBI QUITO

**TUTOR:**

ING. JOHNNY XAVIER SERRANO GUERRERO, M.Sc.

CUENCA – ECUADOR

2020

## CESIÓN DE DERECHOS DE AUTOR

Nosotros, Paul Esteban Paucar Uyaguari con documento de identificación N° 0105264493 Luis Agustín Yubi Quito con documento de identificación N° 0302635008, manifestamos nuestra voluntad y cedemos a la Universidad Politécnica Salesiana la titularidad sobre los derechos patrimoniales en virtud de que somos autores del trabajo de titulación: **“PRONOSTICO DE PERFILES DE CONSUMO DE ELECTRICIDAD A TRAVÉS DE MODELOS ESTADÍSTICOS Y DESARROLLO DE UNA METODOLOGÍA COMPARATIVA”**, mismo que ha sido desarrollado para optar por el título de: *Ingeniero Eléctrico*, en la Universidad Politécnica Salesiana, quedando la Universidad facultada para ejercer plenamente los derechos cedidos anteriormente.

En aplicación a lo determinado en la Ley de Propiedad Intelectual, en nuestra condición de autores nos reservamos los derechos morales de la obra antes citada. En concordancia, suscribimos este documento en el momento que hacemos la entrega del trabajo final en formato digital a la Biblioteca de la Universidad Politécnica Salesiana.

Cuenca, enero de 2020



---

Paul Esteban Paucar Uyaguari

C.I. 0105264493



---

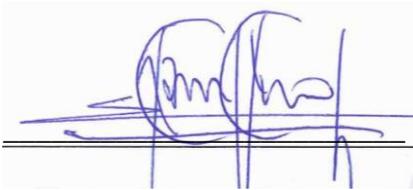
Luis Agustín Yubi Quito

C.I. 0302635008

## CERTIFICACIÓN

Yo, declaro que bajo mi tutoría fue desarrollado el trabajo de titulación: **“PRONOSTICO DE PERFILES DE CONSUMO DE ELECTRICIDAD A TRAVÉS DE MODELOS ESTADÍSTICOS Y DESARROLLO DE UNA METODOLOGÍA COMPARATIVA”**, realizado por Paul Esteban Paucar Uyaguari y Luis Agustín Yubi Quito, obteniendo el *Proyecto Técnico con enfoque investigativo*, que cumple con todos los requisitos estipulados por la Universidad Politécnica Salesiana.

Cuenca, enero de 2020

A handwritten signature in blue ink, appearing to be 'Johnny Xavier Serrano Guerrero', written over a horizontal line.

Ing. Johnny Xavier Serrano Guerrero, M.Sc.

C.I. 0104983382

## DECLARATORIA DE RESPONSABILIDAD

Nosotros, Paul Esteban Paucar Uyaguari con documento de identificación N° 0105264493 y Luis Agustín Yubi Quito con documento de identificación N° 0302635008, autores del trabajo de titulación: **“PRONOSTICO DE PERFILES DE CONSUMO DE ELECTRICIDAD A TRAVÉS DE MODELOS ESTADÍSTICOS Y DESARROLLO DE UNA METODOLOGÍA COMPARATIVA”**, certificamos que el total contenido del *Proyecto Técnico con enfoque investigativo*, es de nuestra exclusiva responsabilidad y autoría.

Cuenca, enero de 2020



Paul Esteban Paucar Uyaguari

C.I. 0105264493



Luis Agustín Yubi Quito

C.I. 0302635008

## **DEDICATORIA**

Quiero dedicar todo el esfuerzo contenido en este trabajo a mi esposa Joselyn en especial que siempre creyó en mí y a mi hija Paulita que han sido mi mayor motivación en cada etapa de este trabajo.

**Paul Esteban Paucar**

## **DEDICATORIA**

Este trabajo va dedicado a Dios por darme la vida y las fuerzas necesarias para llegar hasta este momento tan importante de mi formación profesional. A mi madre y a mi padre por ser los pilares fundamentales de mi vida y brindarme todo el apoyo para culminar esta etapa de formación universitaria, por su cariño y comprensión porque sin ellos no hubiese llegado a cumplir esta meta.

**Luis Agustín Yubi**

## **AGRADECIMIENTO**

Quiero agradecer en primer lugar a Dios que ha puesto todo en su debido momento y me ha dado la sabiduría necesaria para poder terminar mis estudios.

A mi esposa Joselyn, por creer en mí y por su apoyo incondicional en los momentos más complicados. A mi hija Paula, que es el motor de mi vida.

Familiares que de una u otra manera han colaborado para este fin y muy especialmente a mi mami Leo por sus valores y principios que siempre los llevaré conmigo.

Un agradecimiento especial al Ing. Xavier Serrano que siempre supo guiarnos y apoyarnos hasta el fin de este trabajo.

Finalmente, a mi amigo Agustín Yubi por su sacrificio en la culminación de este trabajo.

**Paul Esteban Paucar**

## **AGRADECIMIENTO**

Quiero agradecer a Dios por brindarme todo el conocimiento, inteligencia y sabiduría necesaria para culminar esta etapa de formación, por protegerme y darme fuerzas para superar los obstáculos y dificultades a lo largo de toda mi vida.

A mis padres por ser los que me guían en mi vida, por sus consejos y por enseñarme a ser persistente y a nunca darme por vencido. A su apoyo ya que sin esto no sería posible la culminación de esta etapa de formación profesional. A mis hermanos, por brindarme su ayuda y palabras de ánimo en cada momento.

Familiares y amigos que han sabido brindarme la ayuda de una u otra forma durante toda esta etapa.

Un agradecimiento especial al Ing. Xavier Serrano por brindar la ayuda y guía necesaria para culminar este trabajo de titulación.

A mi compañero de tesis Paul Paucar por brindarme la confianza y formar parte de este equipo para la elaboración de este trabajo.

**Luis Agustín Yubi**

## **RESUMEN**

Este proyecto investigativo está orientado al pronóstico de consumo de energía eléctrica mediante la aplicación de métodos estadísticos como son: Regresión Lineal Simple (RLS), Regresión Lineal Múltiple (RLM) y Modelo Autorregresivo Integrado de Media Móvil (ARIMA). Para este propósito se cuenta con información del consumo de energía eléctrica y variables climáticas de la Universidad Politécnica Salesiana sede Cuenca y la Universidad Politécnica de Valencia. Se efectúa una segmentación horaria de los datos y se realiza un análisis descriptivo, además se forman agrupaciones de días de consumo similar mediante la distancia euclidiana.

Para la aplicación de los métodos estadísticos se crea un programa en el software Matlab el cual facilita el procesamiento de los datos como también del pronóstico. Se establece una metodología de evaluación mediante el cálculo y análisis de los errores como son: error absoluto medio (MAE), raíz del error cuadrático medio (RMSE) y porcentaje de error absoluto medio (MAPE).

Como resultado del análisis de los errores se determina que los métodos de regresión tienen mejores resultados que los obtenidos con el método ARIMA.

## **ABSTRACT**

This research project is oriented to the forecast of electric power consumption through the application of statistical methods such as: Simple Linear Regression (RLS), Multiple Linear Regression (RLM) and Integrated Mobile Media Autoregressive Model (ARIMA). For this purpose, there is information on the consumption of electrical energy and climatic variables of the Salesian Polytechnic University, Cuenca headquarters and the Polytechnic University of Valencia. Hourly segmentation of the data is carried out and a descriptive analysis is carried out, in addition groups of days of similar consumption are formed by the Euclidean distance. For the application of statistical methods a program is created in the Matlab software which facilitates the processing of the data as well as the forecast. An evaluation methodology is established through the calculation and analysis of errors such as: mean absolute error (MAE), root of the mean square error (RMSE) and percentage of absolute absolute error (MAPE). As a result of the analysis of the errors it is determined that the regression methods have better results than those obtained with the ARIMA method.

## **PREFACIO**

Esta investigación da a conocer los resultados obtenidos de la aplicación de métodos estadísticos en el pronóstico de perfiles de consumo de energía eléctrica de la Universidad Politécnica Salesiana sede Cuenca y de la Universidad Politécnica de Valencia.

Para realizar esta investigación, se inició con una revisión del estado del arte de todos los métodos estadísticos que se utilizan para el pronóstico y análisis del consumo de energía eléctrica. Luego de una segmentación horaria de los datos, se realiza un análisis estadístico a través de la mediana, desviación estándar, valores máximos y mínimos, etc. Adicional a esto, se forman agrupaciones de datos que corresponden a los perfiles de consumo similar entre los días en base a la distancia euclidiana.

Luego se aplican los métodos de pronóstico con ayuda de un programa elaborado en el software Matlab. Finalmente, con los resultados obtenidos se establece una metodología de evaluación de los errores para determinar el o los métodos adecuados para cada base de datos.

# ÍNDICE GENERAL

INTRODUCCIÓN.....	1
JUSTIFICACIÓN .....	2
GRUPO OBJETIVO (BENEFICIARIOS) .....	3
Objetivo general .....	3
Objetivos específicos .....	3
1.    CAPÍTULO 1. ESTUDIO DEL ESTADO DEL ARTE.....	4
1.1.    CONCEPTOS TEÓRICOS.....	6
1.1.1.    Carga eléctrica de un sistema.....	6
1.1.2.    Diagrama de la carga eléctrica de un sistema.....	6
1.1.3.    Consumo de Energía Eléctrica.....	6
1.1.4.    Pronóstico de la carga o demanda eléctrica .....	6
1.2.    MEDIDAS ESTADÍSTICAS.....	7
1.2.1.    Medidas de Centralización .....	7
1.2.2.    Media.....	7
1.2.3.    Mediana.....	7
1.2.4.    Moda.....	7
1.2.5.    Varianza .....	7
1.2.6.    Desviación Estándar.....	7
1.3.    ANÁLISIS DE LOS DATOS.....	8
1.3.1.    Análisis de correlación .....	8
1.3.2.    Valores atípicos.....	8
1.4.    AGRUPACIÓN DE DATOS .....	9
1.4.1.    Análisis de la varianza ANOVA.....	9
1.4.2.    Análisis discriminante.....	9
1.4.3.    Algoritmo Particional .....	9
1.4.4.    Distancia Euclidiana .....	9
1.5.    NORMALIZACIÓN DE DATOS.....	10
1.5.1.    Min Max (MM).....	10
1.5.2.    Decimal-scaling (DS).....	10
1.6.    ANÁLISIS DE LA DEMANDA FRENTE A LA VARIABLE CLIMÁTICA.....	10
1.7.    MÉTODOS DE PRONÓSTICO CONVENCIONALES .....	11
1.7.1.    Regresión Lineal Simple .....	11
1.8.    Regresión Lineal Múltiple .....	11
1.9.    SERIES DE TIEMPO .....	13
1.9.1.    Componentes de una serie de tiempo .....	13

1.9.2.	Procesos o modelos de series de tiempo.....	13
1.9.3.	Modelos Autorregresivos para series de tiempo.....	14
1.10.	<b>MÉTODOS BASADOS EN INTELIGENCIA ARTIFICIAL .....</b>	19
1.10.1.	Lógica difusa.....	19
1.10.2.	Redes neuronales.....	19
1.10.3.	Máquina de vectores de soporte.....	20
1.10.4.	Sistemas expertos basados en el conocimiento.....	20
1.11.	<b>MEDIDAS DE EVALUACIÓN .....</b>	20
1.11.1.	Error absoluto Medio (MAE).....	20
1.11.2.	Raíz del error cuadrado medio (RMSE).....	21
1.11.3.	Porcentaje de error absoluto medio (MAPE).....	21
1.12.	<b>Criterios para evaluar un modelo ARIMA.....</b>	21
1.12.1.	Criterio AKAIKE INFORMATION CRITERION (AIC).....	21
1.12.2.	Criterio de información Bayesiano de Schwarz (BIC).....	21
2.	<b>CAPITULO 2. RECOPIACIÓN DE DATOS DE CONSUMO DE ELECTRICIDAD Y ANALISIS ESTADÍSTICO.....</b>	23
2.1.	<b>RECOPIACIÓN DE DATOS.....</b>	23
2.2.	<b>PROCESAMIENTO DE LOS DATOS.....</b>	23
2.3.	<b>AGRUPACIÓN DE LOS DATOS.....</b>	24
2.4.	<b>NORMALIZACIÓN DE LOS DATOS.....</b>	24
2.5.	<b>BASE DE DATOS UPS.....</b>	26
2.6.	<b>BASE DE DATOS UPV.....</b>	31
3.	<b>CAPÍTULO 3. APLICACIÓN DE MÉTODOS ESTADÍSTICOS.....</b>	40
3.1.	<b>REGRESIÓN LINEAL SIMPLE.....</b>	40
3.2.	<b>REGRESIÓN LINEAL MÚLTIPLE.....</b>	44
3.3.	<b>MODELO ARIMA.....</b>	47
4.	<b>CAPITULO 4. METODOLOGÍA COMPARATIVA DE LOS MÉTODOS ESTADÍSTICOS .....</b>	53
4.1.	<b>BASE DE DATOS DE LA UNIVERSIDAD POLITECNICA DE VALENCIA.....</b>	53
4.2.	<b>BASE DE DATOS DE LA UNIVERSIDAD POLITÉCNICA SALESIANA .....</b>	62
4.3.	<b>SENSIBILIDAD DE LOS ERRORES EN FUNCIÓN DEL NÚMERO DE DÍAS UTILIZADOS.....</b>	68
4.3.1.	Base de datos UPS.....	68
4.3.2.	Base de datos UPV.....	69
	<b>CONCLUSIONES.....</b>	73
	<b>RECOMENDACIONES.....</b>	74

<b>ANEXO 1. MANUAL DE USUARIO DE ENTORNO DE PROGRAMACIÓN EN MATLAB PARA PRONÓSTICO DE CONSUMO DE ELECTRICIDAD .....</b>	<b>75</b>
<b>ANEXO 2. DATOS DE PRONÓSTICO .....</b>	<b>85</b>
<b>BIBLIOGRAFÍA.....</b>	<b>90</b>

## ÍNDICE DE FIGURAS

Figura 2.1 Perfil de consumo de electricidad para días laborables de la UPV en la primavera ..	25
Figura 2.2 Entorno de programación visual desarrollado en MATLAB .....	25
Figura 2.3 Representación gráfica de los datos correspondientes a días laborables en la UPS ..	26
Figura 2.4 Representación gráfica de los datos correspondientes a días no laborables en la UPS .....	27
Figura 2.5 Representación gráfica de la relación potencia-temperatura para las a) 09:00 y b) 12:00 horas del día. ....	28
Figura 2.6 Representación gráfica de la relación potencia-temperatura para las a) 18:00 y b) 01:00 horas del día. ....	28
Figura 2.7 Representación gráfica de la relación potencia-temperatura para las a) 09:00 y b) 12:00 horas del día sin valores atípicos.....	28
Figura 2.8. Representación gráfica de la relación potencia-temperatura para las a) 18:00 y b) 01:00 horas del día sin valores atípicos.....	29
Figura 2.9 Perfil de consumo de electricidad de días laborables de la UPS. ....	30
Figura 2.10 Perfil de consumo de electricidad de días no laborables en la UPS. ....	30
Figura 2.11 Gráfica descriptiva de los datos para los días laborables de la UPV. ....	31
Figura 2.12 Gráfica descriptiva para las estaciones de a) primavera y b) verano de la UPV .....	32
Figura 2.13 Gráfica descriptiva para las estaciones de a) otoño e b) invierno de la UPV .....	32
Figura 2.14 Gráfica descriptiva de los datos para los días no laborables de la UPV. ....	34
Figura 2.15 Relación no lineal potencia-temperatura de la UPV, a) 09:00 y b) a la 12:00.....	35
Figura 2.16 Relación no lineal potencia-temperatura de la UPV, a) 18:00 y b) a la 01:00.....	35
Figura 2.17 Relación lineal potencia-temperatura de la UPV a las a) 09:00 y b) 12:00. ....	37
Figura 2.18 Relación lineal potencia-temperatura de la UPV a las a) 18:00 y b) 01:00. ....	37
Figura 2.19 Perfiles de consumo de electricidad por estaciones en la UPV, a) primavera, b) verano.....	38
Figura 2.20 Perfiles de consumo de electricidad por estaciones en la UPV, a) otoño, b) invierno. ....	38
Figura 2.21 Perfiles de consumo de electricidad de días no laborables de la UPV. Autores.....	38
Figura 3.1 Ajuste de la recta de regresión para el pronóstico del día lunes 30 de mayo de 2016 en la UPV a las a) 08:00 y b) 16:00 horas.....	40
Figura 3.2 Potencia pronosticada del día lunes 30 de mayo de 2016 en la UPV .....	41
Figura 3.3 Ajuste de la recta de regresión para el pronóstico del día lunes 30 de mayo de 2016 en la UPV a las a) 08:00 y b) 16:00 horas.....	42
Figura 3.4 Pronóstico usando RLS del día lunes 30 de mayo de 2016 para un edificio en la UPV .....	43
Figura 3.5 Pronóstico del lunes 4 de diciembre de 2017 en un edificio de la Universidad Politécnica Salesiana.....	44
Figura 3.6 Valor pronosticado del consumo de potencia para la UPV a las 12:00 del día lunes 30 de mayo de 2016. ....	46
Figura 3.7 Pronóstico usando RLM del día lunes 30 de mayo de 2016 para la UPV. ....	46
Figura 3.8 Pronóstico usando RLM del día lunes 4 de diciembre de 2017 para la UPS.....	47
Figura 3.9 Análisis de estacionariedad del conjunto de datos del grupo 1 de la primavera.....	48
Figura 3.10 Autocorrelaciones parcial y simple correspondiente al grupo 1 del día lunes en la UPV.....	48
Figura 3.11 Autocorrelaciones simples y parciales.....	49
Figura 3.12 Pronóstico de consumo de energía eléctrica de la UPV correspondiente al día lunes 23 de mayo de 2016 .....	50

Figura 3.13 Autocorrelaciones Simples y Parciales .....	51
Figura 3.14 Pronóstico de consumo de energía eléctrica de la UPS del día lunes 4 de diciembre de 2017 .....	52
Figura 4.1 Pronóstico de consumo de electricidad para días laborables desde el miércoles 4 de mayo al martes 10 de mayo de 2016. ....	53
Figura 4.2 Pronóstico de consumo de electricidad para días laborables desde el miércoles 11 de mayo al martes 17 de mayo de 2016. ....	54
Figura 4.3 Pronóstico de consumo de electricidad para días laborables desde el miércoles 18 de mayo al martes 24 de mayo de 2016. ....	54
Figura 4.4 Pronóstico de consumo de electricidad para días laborables desde el miércoles 25 de mayo al martes 31 de mayo de 2016. ....	55
Figura 4.5 Pronóstico de consumo de electricidad para días laborables desde el lunes 25 de julio de 2016 al viernes 29 de julio de 2016. ....	56
Figura 4.6 Pronóstico de consumo de electricidad para días laborables desde el martes 22 de noviembre de 2016 al lunes 28 de noviembre de 2016. ....	58
Figura 4.7 Pronóstico de consumo de electricidad para días laborables desde el martes 23 de febrero de 2016 al lunes 29 de febrero de 2016. ....	59
Figura 4.8 Pronóstico de consumo de electricidad para el día sábado 26 de noviembre y domingo 27 de noviembre de 2016. ....	61
Figura 4.9. Pronóstico de consumo de electricidad para días laborables desde el viernes 10 de diciembre de 2017 al jueves 16 de diciembre de 2017 en la UPS.....	62
Figura 4.10 Pronóstico de consumo de electricidad para días laborables desde el viernes 17 de diciembre de 2017 al jueves 23 de diciembre de 2017 en la UPS.....	63
Figura 4.11 Pronóstico de consumo de electricidad para días laborables desde el viernes 24 de diciembre de 2017 al jueves 30 de diciembre de 2017 en la UPS.....	63
Figura 4.12 Pronóstico de consumo de electricidad para días laborables desde el viernes 1 de diciembre de 2017 al jueves 7 de diciembre de 2017 en la UPS.....	64
Figura 4.13 Pronóstico de consumo de electricidad para el día domingo 3 de diciembre de 2016. ....	66
Figura 4.14 Variación de los errores de pronóstico en función de la cantidad de días usados para la UPS con RLS (temperatura).....	68
Figura 4.15 Variación de los errores de pronóstico en función de la cantidad de días usados para la UPS con RLM. ....	69
Figura 4.16 Variación de los errores de pronóstico en función de la cantidad de días usados para la UPS con ARIMA. ....	69
Figura 4.17 Variación de los errores de pronóstico en función de la cantidad de días usados para la UPV con RLS (temperatura). ....	70
Figura 4.18 Variación de los errores de pronóstico en función de la cantidad de días usados para la UPV con RLM. ....	70
Figura 4.19 Variación de los errores de pronóstico en función de la cantidad de días usados para la UPV con ARIMA.....	71

## ÍNDICE DE TABLAS

Tabla 1.1 Características de los modelos autorregresivos mediante los patrones de autocorrelación [55].	17
Tabla 1.2 Modelos ARIMA[58]	17
Tabla 2.1 Valores de distancia euclidiana entre los perfiles de consumo de electricidad de la UPV.	24
Tabla 2.2 Valores de distancia euclidiana normalizados de la tabla 2.1.	24
Tabla 2.3 Valores de medidas descriptivas para días laborables en la UPS.	26
Tabla 2.4 Valores de medidas descriptivas para días no laborables en la UPS.	27
Tabla 2.5 Coeficiente de determinación de varias variables con respecto a la potencia demandada para días laborables y no laborables.	29
Tabla 2.6 Clasificación de grupos para días laborables y no laborables de la UPS.	31
Tabla 2.7 Medidas descriptivas para el análisis de los días laborables de la UPV.	31
Tabla 2.8 Medidas descriptivas para el análisis de la primavera en días laborables de la UPV.	33
Tabla 2.9 Medidas descriptivas para el análisis del verano en días laborables de la UPV.	33
Tabla 2.10 Medidas descriptivas para el análisis del otoño en días laborables de la UPV.	33
Tabla 2.11 Medidas descriptivas para el análisis del invierno en días laborables de la UPV.	34
Tabla 2.12 Medidas descriptivas para el análisis del otoño en días no laborables de la UPV.	34
Tabla 2.13 Coeficientes de determinación para días laborables y no laborables de la UPV.	36
Tabla 2.14 Coeficiente de determinación por estaciones para días laborables y no laborables de la UPV.	36
Tabla 2.15 Clasificación de grupos para días laborables y no laborables de la UPV.	39
Tabla 3.1 Valores de los coeficientes en la recta de regresión.	40
Tabla 3.2 Valores de los coeficientes $b_0$ y $b_1$ de las rectas de regresión.	42
Tabla 3.3 Errores de pronóstico de la Figura 3.4.	43
Tabla 3.4 Coeficientes del modelo de regresión múltiple.	45
Tabla 3.5 Errores del pronóstico de la Figura 3.7.	46
Tabla 3.6 Errores en el pronóstico de la Figura 3.8.	47
Tabla 3.7 Valores de P, D y Q para los grupos de la base de datos UPV, días laborables.	49
Tabla 3.8 Valores de P, D y Q para los grupos de la base de datos UPV, Grupo No Laborables.	50
Tabla 3.9 Errores de pronóstico de la Figura 3.12.	51
Tabla 3.10 Valores de P, D y Q para los datos de los perfiles de carga de la UPS, días laborables Grupo Días Laborables.	51
Tabla 3.11 Valores de P, D y Q para los datos de los perfiles de carga de la UPS, días no laborables Grupo Días No Laborables.	51
Tabla 3.12 Errores de pronóstico con el método ARIMA en la UPS.	52
Tabla 4.1 Errores de pronóstico promedio para días laborables de la UPV en la primavera.	55
Tabla 4.2 Errores de pronóstico para días laborables del 25 al 29 de julio de 2016 de la UPV en el verano.	57
Tabla 4.3 Errores de pronóstico para días laborables del 22 al 28 de noviembre de 2016 de la UPV en el otoño.	58
Tabla 4.4 Errores de pronóstico para días laborables del 23 al 29 de febrero de 2016 de la UPV en el invierno.	59
Tabla 4.5. Errores de pronóstico para días no laborables sábado 26 y domingo 27 de noviembre de 2016.	61
Tabla 4.6 Errores de pronóstico promedio para días laborables en la UPS.	64

Tabla 4.7 Errores de pronóstico para el domingo no laborable 3 de diciembre de 2016. ....	66
Tabla 4.8 Errores promedio del pronóstico de días laborables de la UPS y UPV.....	67
Tabla 4.9 Errores promedio del pronóstico de días no laborables .....	68
Tabla 4.10. Rango de datos de consumo para realizar el pronóstico de consumo de electricidad .....	71
Tabla 4.11 Comparación de resultados desde el 1 al 7 de diciembre de 2017 .....	72
Tabla 4.12 Comparación de resultados de la UPV.....	72

## **GLOSARIO**

RLS	Regresión Lineal Simple
RLM	Regresión Lineal Múltiple
ARIMA	Modelo Autorregresivo Integrado de Media Móvil
MAE	Error Absoluto Medio
RMSE	Raíz del error cuadrático medio
MAPE	Porcentaje de error absoluto medio
UPS	Universidad Politécnica Salesiana
UPV	Universidad Politécnica de Valencia
CEE	Consumo de energía eléctrica
FAC	Función de autocorrelación
FACP	Función de autocorrelación parcial

## INTRODUCCIÓN

Hoy en día el avance tecnológico ha permitido el desarrollo de los pueblos y regiones a gran escala, pero cuando hablamos de progreso, sabemos que va de la mano con la energía eléctrica. Esto supone un incremento en la demanda, producido por mayores niveles de consumo que representan picos más elevados en la curva de demanda. Para esto es necesario disponer de una mayor capacidad de generación, misma que debe estar en equilibrio con la capacidad demandada de energía eléctrica. De esta manera se logra evitar sobre generación o déficit de energía ocasionando impacto y afecciones a los consumidores, por esto es de vital importancia el tema de la planificación y gestión de la generación, distribución y consumo de energía eléctrica [1][2].

Las empresas eléctricas escatiman esfuerzos en el pronóstico o estimación del consumo de energía eléctrica a largo, mediano y corto plazo, dependiendo de la aplicación que se prevé dar a los datos [3][1]. El pronóstico a corto plazo se utiliza para estimar consumos de energía eléctrica desde días hasta semanas considerando bases de datos con diversas variables de estudio, las cuales deberán ser analizadas y segmentadas por la presencia de consumos anómalos conocidos como valores atípicos que influyen en la precisión del pronóstico [1][4][5].

El pronóstico permite conocer la demanda que será requerida en un tiempo futuro y así definir una estrategia de planificación incentivando el uso racional de la energía eléctrica y motivando a los usuarios a tener consumos en periodos distintos a las horas pico.

Entre los métodos estadísticos de pronóstico tenemos los métodos regresivos, los cuales se exponen en el artículo [6] y aplicado en el consumo de los hogares de Italia brindaron resultados bastante satisfactorios con una alta precisión.

Otro método de predicción son las series de tiempo, el método más usado es el ARIMA como se constata en los artículos [7][8][9]. Sin embargo, existen variables y factores que influyen al consumo de energía eléctrica como son la temperatura, humedad, radiación solar, etc. Los artículos [10][11] emplean la regresión lineal múltiple (RLM) para predecir consumos tomando en cuenta esas variables climáticas.

## **JUSTIFICACIÓN**

Cuando no se conoce el comportamiento del Consumo de Energía Eléctrica (CEE), crece la incertidumbre para tomar decisiones operativas y estratégicas, esto limita a la planificación de una manera precisa del comportamiento del sistema eléctrico [12].

La falta de conocimiento de los perfiles de CEE en centrales, subestaciones, distribuidoras o edificaciones, limita la gestión energética y económica disminuyendo la capacidad de optimizar los recursos [12].

Al no establecer comportamientos futuros del CEE a corto plazo en edificios o a nivel industrial o de generación no es posible definir con anticipación el comportamiento del perfil de CEE para poder establecer medidas de control o de prevención en caso de ocurrir consumos anómalos o problemas en las instalaciones eléctricas [13]. Por esto es importante realizar el pronóstico de la demanda de energía eléctrica con el mínimo margen de error a corto plazo, usando para ello un modelo adecuado a la naturaleza de las variables de una base de datos y así anticipar un posible déficit de energía y considerar ante ello el aumento de la potencia instalada en la planificación a corto plazo.

## **GRUPO OBJETIVO (BENEFICIARIOS)**

Este trabajo investigativo permitirá a la Universidad Politécnica Salesiana sede Cuenca conocer el comportamiento futuro de los consumos de energía eléctrica con ayuda de un programa elaborado en el software Matlab. Esto ayudará a una mejor planificación del uso de la energía, como también a optimizar los recursos energéticos.

## **OBJETIVOS**

### **Objetivo general**

- Pronosticar perfiles de consumo de electricidad a través de los modelos de Regresión Lineal, Regresión Lineal Múltiple y Modelo Autorregresivo Integrado de Media Móvil y desarrollar una metodología para comparar los resultados obtenidos.

### **Objetivos específicos**

- Realizar un estudio del estado del arte acerca de los modelos de pronóstico de consumo de electricidad.
- Recopilar bases de datos de consumo de electricidad y realizar un análisis estadístico.
- Aplicar la Regresión lineal, Regresión lineal múltiple y modelo Autorregresivo integrado de media móvil (ARIMA) para pronosticar el consumo de energía eléctrica en intervalos horarios durante el día.
- Desarrollar una metodología para comparar los resultados obtenidos por cada modelo.

# 1. CAPÍTULO 1. ESTUDIO DEL ESTADO DEL ARTE

El aumento de la población y del número de dispositivos dependientes de la energía eléctrica ha ocasionado que en la actualidad se tenga un especial interés en la predicción de la carga eléctrica con métodos más sofisticados [14]. En la India el consumo anual de energía presentó un incremento de 418 GWh en 2005 a 948 GWh en 2015. Por otro lado, en Hong Kong, se demandaron 5718 GWh en 1991 y 9111 GWh en 2001 solo en el sector doméstico. Para garantizar la confiabilidad del suministro, las empresas públicas obtienen el registro del consumo doméstico el cual es usado para la predicción mediante diversos métodos [14], [15].

En un edificio la energía eléctrica es la principal fuente de energía que se utiliza. Es por ello que un adecuado pronóstico de consumo brinda beneficios, tanto para los propietarios como para los ocupantes del edificio [16]. Una limitada capacidad para abastecer la demanda o por el contrario el sobredimensionamiento del abastecimiento de energía, genera problemas a mediano y largo plazo, por lo que, minimizar el error en la predicción es de vital importancia [17]. Además, un pronóstico acertado de la misma permite a las empresas distribuidoras llevar una planificación adecuada del suministro de energía eléctrica, por otro lado, tener un mayor control sobre las pérdidas de energía [16].

La predicción de consumo de energía eléctrica puede estar asociada con otras variables tales como: el clima, temperatura, área de influencia (sector residencial, comercial o industrial), jornadas laborables y no laborables, etc. Para realizar dicha predicción es necesario disponer de una base de datos que represente el historial de consumo durante un determinado periodo [6].

A continuación, se realiza la revisión de algunos artículos que han usado métodos estadísticos para la predicción del consumo de energía eléctrica como son: la regresión lineal múltiple (RLM), regresión lineal simple (RLS) y modelo autorregresivo integrado de media móvil (ARIMA).

Los métodos estadísticos de predicción que se aplicó en el artículo [6] son RLM y RLS con el fin de predecir el consumo de energía eléctrica en los hogares de Italia, tomando como variables de estudio el consumo doméstico, no doméstico y su costo. Los autores de este artículo llegaron a deducir que el costo no es una variable representativa para la predicción por lo que, obtuvieron predicciones precisas y porcentajes de error estándar de 1% y en el peor de los casos el 11%.

En el artículo [7], se realizó un estudio en Ecuador pronosticando el consumo de energía eléctrica en el sector residencial mediante los métodos ARIMA y ARIMAX. Se realizó una comparación entre los dos métodos a través de los valores de correlación y del error estándar, teniendo como resultado que el método ARIMA es el más adecuado para realizar pronósticos en zonas residenciales.

En [8] se empleó el método ARIMA para el pronóstico del consumo de la energía eléctrica en Colombia, partiendo de una base de datos de la Empresa Eléctrica de Pereira en el año 2001, comprendida desde enero hasta diciembre del mismo año. El modelo que más se ajustó y tuvo menos errores fue el SARIMA (1,0,0) (2,1,0) otorgando mayor confiabilidad en su pronóstico.

Una forma de mejorar la exactitud del pronóstico empleando el método ARIMA es aumentar el grado de precisión en la determinación de la demanda, lo cual se consiguió al utilizar modelos de modificación de previsión comprendidos por el método óptimo de Fourier y el modelo ARIMA estacional como se pudo constatar en el artículo [18]. De igual manera en China, partiendo del histórico de consumo desde 1999 hasta 2009, empleando el mismo método, se obtuvo un incremento en la precisión del pronóstico comparado con el método ARIMA estacional.

Otros autores como es el caso de los artículos [19][20] realizan comparaciones con distintos métodos de predicción tales como: red neuronal artificial (ANN), RLM, máquina de vectores de soporte (MVS), regresión del proceso gaussiano (GRP). Llegando a concluir que con la ANN se tiene una mayor precisión, pero demanda un mayor tiempo de cálculo, mientras que el método RLM tiene ofrece una mayor precisión en menos tiempo y esfuerzo de cálculo.

Los métodos de predicción de consumo de energía eléctrica pueden verse afectados por picos de potencia, los cuales aumentan el error en la predicción, por lo que, en el artículo [21] proponen un método híbrido para la predicción del consumo de energía eléctrica a corto plazo mediante la fusión de un modelo regresivo y un compensador de picos de potencia que hace que la predicción sea más exacta.

El artículo [16] agrupa los métodos de predicción en dos grupos: el primero basados en inteligencia artificial (IA) y el segundo en métodos convencionales, dentro de los cuales se encuentran las series temporales y los modelos de regresión que tradicionalmente han sido usados para realizar la predicción sobre el consumo de energía [17]. Las series temporales en la actualidad siguen siendo usadas ampliamente ya que producen resultados satisfactorios con los problemas lineales [14].

Se sabe que los modelos de predicción para patrones de consumo no lineales han sido poco desarrollados para ofrecer resultados robustos. Sin embargo, los modelos basados en inteligencia artificial en la actualidad son los más populares ya que han logrado tener un buen manejo de la no linealidad que se presenta en la carga eléctrica [14], [16]. Algunos de ellos se basan en el conocimiento, lógica difusa, redes neuronales y máquinas de vectores de soporte [17].

Por otro lado, los métodos convencionales son fáciles de aplicar, pero no funcionan bien con la no linealidad que pueden causar los factores climáticos, condiciones estructurales del edificio, costumbres de los ocupantes, etc. La no linealidad es el principal problema que puede afectar la

predicción del consumo de energía, por lo que se deben integrar con otros métodos para obtener mejores resultados [16].

Para poder hacer frente a los problemas que presentan los métodos convencionales con la no linealidad es frecuente que se los utilice en combinación con métodos de inteligencia artificial, lo cual se conoce como métodos híbridos. También es posible que se combinen métodos de IA para obtener mejores resultados [16].

## **1.1. CONCEPTOS TEÓRICOS**

A continuación, se presentan todos los fundamentos teóricos necesarios para realizar el pronóstico de perfiles de consumo de energía eléctrica.

### **1.1.1. Carga eléctrica de un sistema**

La carga de un sistema se define como no estacionaria por su tendencia creciente o decreciente en ciertos intervalos del día. Además, que sus valores no están alrededor de una media. Estas variaciones se deben a parámetros como: la hora, día, clima, etc., y pueden estar influenciados también por factores económicos y sociales [22]. Por otra parte, existen ciertos patrones comparables que se repiten en ciclos diarios, es decir, tienen estacionalidades definidas [23].

### **1.1.2. Diagrama de la carga eléctrica de un sistema**

Es una representación gráfica del comportamiento de la demanda de potencia eléctrica en un determinado periodo.

### **1.1.3. Consumo de Energía Eléctrica**

Es la sumatoria de la energía consumida de uno o más usuarios en un intervalo de tiempo determinado. También se la puede definir como la energía que el usuario emplea para realizar una actividad determinada en un periodo definido [24], [25].

### **1.1.4. Pronóstico de la carga o demanda eléctrica**

Realizar un pronóstico de la demanda de energía eléctrica permite establecer la cantidad de potencia que requerirá el usuario en un tiempo futuro, por lo que se requiere conocer la demanda inicial y su comportamiento en el tiempo [26].

Dicho pronóstico ayuda a obtener una adecuada planificación del sistema eléctrico de distribución, permitiendo establecer de una manera más exacta periodos de mantenimiento que contribuyen a reducir la aplicación de redespacho [12].

En las predicciones de la carga existen pronósticos a corto plazo (días hasta semanas), mediano plazo (desde un mes hasta un año) y a largo plazo (desde uno hasta veinte años) [27].

## 1.2. MEDIDAS ESTADÍSTICAS

### 1.2.1. Medidas de Centralización

Las medidas de centralización son aquellas que permiten conocer las características de la distribución de los datos, [28] entre ellas tenemos:

### 1.2.2. Media

Consiste en la suma de todas las observaciones de un conjunto de datos dividido para el número total de estas observaciones [28].

$$\bar{X} = \frac{\sum_{i=1}^k X_i}{N} \quad (1)$$

Donde:  $N$  es el número total de datos,  $X$  es el dato de la muestra.

### 1.2.3. Mediana

Es aquella que divide en dos partes iguales al conjunto de datos por lo tanto es el centro de una secuencia ordenada de datos, no se ve afectada por los valores externos [28], [29].

$$\bar{x} = \frac{x_{n+1}}{2} \quad n \rightarrow \text{Impar} \quad (2)$$

$$\bar{x} = \frac{x_{n/2} + x_{n/2+1}}{2} \quad n \rightarrow \text{Par} \quad (3)$$

### 1.2.4. Moda

Es el valor que se presenta con mayor frecuencia en todo el conjunto de datos. En una muestra o conjunto de observación se puede encontrar dos o más datos repetidos [28].

### 1.2.5. Varianza

La varianza es el promedio aritmético de la diferencia de cada uno de los valores y la media del conjunto elevado al cuadrado [29].

$$s^2 = \sigma^2 = \frac{\sum_{i=1}^k (x_i - \bar{x}_i)^2}{n - 1} \quad (4)$$

### 1.2.6. Desviación Estándar

También llamada como desviación típica, se define como la raíz cuadrada de la varianza [29].

$$s = \sqrt{s^2} = \sqrt{\frac{\sum_{i=1}^k (x_i - \bar{x}_i)^2}{n - 1}} \quad (5)$$

### 1.3. ANÁLISIS DE LOS DATOS

El análisis de datos univariantes está enfocado en el comportamiento individual de cada variable dentro de un conjunto de datos, centrándose en dos aspectos como son: la distribución y dispersión. Mientras que, si nos encontramos con datos bivariantes los cuales perciben el efecto de una variable sobre otra se aplican pruebas paramétricas y no paramétricas. En caso de tener datos multivariantes de más de dos variables se aplican varias técnicas estadísticas y algoritmos de cálculo.

Para analizar el comportamiento de los datos a lo largo de su trayectoria, como se relacionan entre sí y definir sus semejanzas, se aplica diferentes técnicas, las cuales se explican en los siguientes apartados [30][31].

#### 1.3.1. Análisis de correlación

Es aquella que nos permite establecer la relación entre dos variables mediante una representación visual de los datos. Para una representación numérica, esta relación se la realiza con el cálculo del coeficiente de correlación, el cual toma valores de entre 1 y -1. [32].

$$r = \frac{\sum(x - \bar{x})(y - \bar{y})}{(n - 1)s_x s_y} \quad (6)$$

#### 1.3.2. Valores atípicos

Los valores atípicos son un conjunto de datos que son muy diferentes a las observaciones o datos dentro de un mismo conjunto. Estos valores pueden generar problemas a la hora de realizar el pronóstico debido a que nacen de un error de digitación o de un error a la hora de recolectar los datos [33]. Los valores atípicos se los puede determinar mediante los valores Z.

##### 1.3.2.1. Valores Z

Los valores z son aquellos que nos permite discriminar los datos que se encuentran fuera de un rango establecido por quien los está analizando. Estos van a depender de cambios de la media de una muestra a otra.

Por medio de la ecuación (7) podemos calcular estos valores.

$$\text{Valores Z} = \frac{(X_{valor} - Promedio)}{Desviacion Est.} \quad (7)$$

Para determinar si un valor es atípico o no, se establece un margen máximo y mínimo dentro del cual se encuentra un porcentaje del conjunto de datos, es decir, los datos que se encuentren fuera de este rango son determinados como atípicos y son eliminados [34]–[36].

## 1.4. AGRUPACIÓN DE DATOS

También conocido como clustering, es utilizado para establecer patrones similares de consumo dentro de un conjunto de datos. Esta agrupación se la puede realizar tanto con variables numéricas como con variables categóricas. Para variables numéricas, las agrupaciones se pueden definir mediante la distancia entre los puntos, lo cual no se puede realizar con las variables categóricas [37][38].

### 1.4.1. Análisis de la varianza ANOVA

Es aquella que permite analizar las diferencias entre las medias de diferentes grupos de datos. Si se tiene solamente dos grupos de datos se procede a analizar las medias, en caso de tener más de dos grupos de datos se procede a analizar la varianza [30].

### 1.4.2. Análisis discriminante

Esta técnica está basada en el uso de la regla de Bayes y permite diferenciar las variables de un conjunto de datos para luego clasificarlas y agruparlas según su semejanza o relación que ocupen dentro del conjunto de datos. [30].

### 1.4.3. Algoritmo Particional

También llamado K-means, es el encargado de agrupar datos mediante las distancias entre los centros de cada grupo y los puntos. Esta distancia debe ser la mínima entre el centro y un punto, los valores que estén más cercanos al centro forman un grupo. Previamente, se debe establecer el número de grupos a formar. Para conocer la distancia entre el centro del grupo y el punto se utiliza la distancia euclidiana.

### 1.4.4. Distancia Euclidiana

Es aquella que nos permite establecer la distancia entre dos puntos, se basa en el teorema de Pitágoras y se representa por la ecuación (8).

$$D_{a \leftrightarrow b} = \sqrt{\sum_i (a_i - b_i)^2} \quad (8)$$

Para detectar grupos con características similares mediante la distancia euclidiana es necesario tener una muestra homogénea. La distancia euclidiana en forma de matriz se puede presentar de la siguiente manera, ecuación (9) [39], [40].

$$d_{ij} = \|x_i - x_j\|_2^2 \quad (9)$$

## 1.5. NORMALIZACIÓN DE DATOS

La normalización de datos es un proceso encargado en convertir estructuras de datos complejos en estructuras de datos más simples, para facilitar el manejo y operación de los mismos debido a que algunas estructuras de datos, presentan diferentes escalas numéricas como también diferentes distribuciones [41], [42]. Los Métodos de Normalización de datos más empleados son:

### 1.5.1. Min Max (MM)

Este método utiliza el valor máximo y mínimo de la base de datos para obtener valores de entre 0 y 1 [42], [43].

$$\hat{s}_{ij} = \frac{s_{ij} - \min_j}{\max_j - \min_j} \quad (10)$$

Donde:  $\hat{s}_{ij}$  es el dato transformado del conjunto de datos,  $s_{ij}$  es el dato del conjunto de datos,  $\min_j$  es el valor mínimo del conjunto de datos,  $\max_j$  es el valor máximo del conjunto de datos.

### 1.5.2. Decimal-scaling (DS)

Normaliza los datos con valores entre 1 y -1. El valor k de la ecuación (11) es el menor número con el que se cumpla la condición  $(|\hat{s}_{ij}|) < 1$  [42].

$$\hat{s}_{ij} = \frac{s_{ij}}{10^k} \quad (11)$$

## 1.6. ANÁLISIS DE LA DEMANDA FRENTE A LA VARIABLE CLIMÁTICA

La demanda eléctrica se ve afectada por diversas variables externas y condiciones climáticas definidas por la estacionalidad del tiempo, lo cual causa que la demanda no presente un patrón definido. En España en la época de invierno aumenta la demanda de electricidad por el uso de calefactores y en verano por el uso de aire acondicionado haciendo que su gráfica anual adopte una forma de U o valle [44], [45].

Para el caso de Ecuador, la variación de la demanda va a depender de la región en la cual queremos realizar el estudio, es así que, en la región costa se tiene consumos elevados de energía en la época de verano debido al incremento de la temperatura y la necesidad del uso de aires acondicionados. La región oriental presenta un consumo en forma exponencial ascendente, debido al menor consumo en horas de la madrugada y un consumo máximo en las horas de la tarde y noche. La región sierra presenta un consumo casi lineal por tener un clima templado y pocas variaciones de temperatura [46].

## 1.7. MÉTODOS DE PRONÓSTICO CONVENCIONALES

### 1.7.1. Regresión Lineal Simple

Es un método de predicción compuesto por dos variables a pronosticar, una independiente ( $x$ ) y otra dependiente ( $Y$ ) [47], [48].

Ecuación del modelo de regresión lineal simple (12):

$$Y = \beta_0 + \beta_1 x + \varepsilon \quad (12)$$

Donde:

$\beta_0$  es la intersección con el eje Y,  $\beta_1$  es la pendiente de la recta,  $\varepsilon$  es el error aleatorio.

Para determinar los coeficientes de regresión  $\beta_0$  y  $\beta_1$  es necesario ajustar la recta a los puntos, lo cual permite estimar los valores pronosticados. El ajuste de la recta se realiza mediante la ecuación (13):

$$\hat{Y} = b_0 + b_1 x \quad (13)$$

Donde:

$b_0 = \beta_0$  es la intersección de la muestra con el eje Y,  $b_1 = \beta_1$  es la pendiente de la muestra

Esta ecuación es llamada ecuación de regresión lineal simple o recta de predicción. Para determinar los valores de  $b_0$  y  $b_1$  es necesario aplicar el método de mínimos cuadrados.

#### 1.7.1.1. Método de mínimos cuadrados

El método de mínimos cuadrados ayuda a determinar los valores de  $b_0$  y  $b_1$  lo que reduce al mínimo la suma de las diferencias al cuadrado alrededor de la línea de predicción, es decir entre los valores estimados y los valores reales [47].

Las ecuaciones (14) y (15) para el cálculo de  $b_0$  y  $b_1$  son:

$$b_1 = \frac{n \sum_{i=1}^n x_i y_i - (\sum_{i=1}^n x_i)(\sum_{i=1}^n y_i)}{n \sum_{i=1}^n x_i^2 - (\sum_{i=1}^n x_i)^2} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2} \quad (14)$$

$$b_0 = \frac{\sum_{i=1}^n y_i - b_1 \sum_{i=1}^n x_i}{n} = \bar{y} - b_1 \bar{x} \quad (15)$$

## 1.8. Regresión Lineal Múltiple

Es un método de predicción que analiza de forma estadística la relación que tiene una variable dependiente con un conjunto de variables independientes. Tiene áreas de aplicación muy amplias,

pero centrándonos en este estudio, es el método más utilizado para la predicción de la carga eléctrica.

La ecuación de la regresión múltiple se expresa mediante (16).

$$L = X\pi + \varepsilon \quad (16)$$

Donde:

$L$  es el vector  $n \times 1$  de las variables dependientes.,  $X$  es la matriz  $n \times m$  que describe las variables.,  $\pi$  es el vector  $m \times 1$  de los coeficientes de regresión.,  $\varepsilon$  es el error del modelo.

Las variables independientes pueden llegar a ser el tipo de día laborable o no laborable, la carga de la red, las condiciones climáticas, etc. La variable dependiente siempre será la carga ya que es la variable que se va a pronosticar.

Los coeficientes de regresión son aquellos que se ajustan a través de mínimos cuadrados. El error del modelo es el que se analiza luego de la estimación del modelo [17], [49].

Para facilitar el ajuste del modelo de regresión múltiple se lo puede realizar mediante la ayuda de matrices, teniendo en cuenta el número de  $k$  variables independientes y  $N$  observaciones, la ecuación de la regresión múltiple se observa como en (17).

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \dots + \beta_k x_{ki} + \varepsilon_i \quad (17)$$

En forma matricial de esta ecuación se plantea como:

$$y = X\beta + \varepsilon \quad (18)$$

$$y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, X = \begin{bmatrix} 1 & x_{11} & x_{21} & \dots & x_{k1} \\ 1 & x_{12} & x_{22} & \dots & x_{k2} \\ \vdots & \vdots & \vdots & \dots & \vdots \\ 1 & x_{1n} & x_{2n} & \dots & x_{kn} \end{bmatrix}, \beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix}, \varepsilon = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix} \quad (19)$$

Para calcular los coeficientes de regresión se utiliza el mismo método que en la regresión lineal simple, es decir, por mínimos cuadrados que en este caso quedaría representada como matriz de la siguiente manera [47]:

$$b = A^{-1}g = (X'X)^{-1}X'y \quad (20)$$

Donde:

$$X'X = \begin{bmatrix} n & \sum_{i=1}^n x_{1i} & \sum_{i=1}^n x_{2i} & \dots & \sum_{i=1}^n x_{ki} \\ \sum_{i=1}^n x_{1i} & \sum_{i=1}^n x_{1i}^2 & \sum_{i=1}^n x_{1i} x_{2i} & \dots & \sum_{i=1}^n x_{1i} x_{ki} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \sum_{i=1}^n x_{ki} & \sum_{i=1}^n x_{ki} x_{1i} & \sum_{i=1}^n x_{ki} x_{2i} & \dots & \sum_{i=1}^n x_{ki}^2 \end{bmatrix} \quad (21)$$

$$X'y = \begin{bmatrix} g_0 = \sum_{i=1}^n x_{1i} \\ g_1 = \sum_{i=1}^n x_{1i} y_i \\ \vdots \\ g_k = \sum_{i=1}^n x_{ki} y_i \end{bmatrix} \quad (22)$$

## 1.9. SERIES DE TIEMPO

Como series de tiempo entendemos a un conjunto de datos registrados en diferentes intervalos de tiempo. Estos datos pueden comportarse de diferente forma y son utilizados para desarrollar modelos estadísticos que permiten comprender como se comportan los datos en función del tiempo [32], [50], [51].

### 1.9.1. Componentes de una serie de tiempo

**Tendencia (T).** - Es la trayectoria uniforme que tiene una serie de tiempo a largo plazo. Esta puede ser constante, lineal, cuadrática, exponencial, etc. [32], [50].

**Estacionalidad (E).** - Describe a una serie de tiempo que tiene patrones que se repiten durante un periodo corto de tiempo. Estos pueden ser factores climáticos o institucionales [50], [51].

**Ciclo (C).** - Se caracteriza por presentar a lo largo de su trayectoria oscilaciones de larga duración, se puede presentar en fenómenos climáticos que duran ciclos largos durante el año [32], [51].

**Aleatorio (A).** - Es una serie de tiempo que no presenta ningún patrón ni forma establecida, es decir que es impredecible debido a factores externos no controlados [50], [51].

### 1.9.2. Procesos o modelos de series de tiempo

Para que las series de tiempo tengan un enfoque predictivo deben tener procesos o modelos en los cuales el pronóstico se enfoque en el comportamiento de la variable de interés. Estos procesos son:

**Procesos deterministas.** - Son métodos sencillos, con menor precisión. Se caracterizan por el promedio móvil. Es decir, que el pronóstico de una variable se la realiza mediante el promedio de todos los valores anteriores [51].

**Procesos estocásticos.** – Es un conjunto de variables aleatorias difíciles de determinar, se extrae de la serie observada, por ello, para realizar el pronóstico se necesita construir modelos aproximados [51]. Los procesos estocásticos se dividen en:

**Procesos estocásticos aleatorios.** - Es el conjunto de variables aleatorias que están agrupadas y ordenadas en el tiempo. Su media, varianza y covarianza cambian con el tiempo.

**Procesos estocásticos estacionarios.** - Se identifican porque su media y varianza son constantes en el tiempo, mientras que su covarianza va a depender de la distancia entre dos periodos [52].

Por lo tanto podemos decir que una serie de tiempo, pertenece a un proceso estocástico debido a que se escoge un conjunto de datos por cada variable aleatoria que se establece en el estudio [50].

### **1.9.3. Modelos Autorregresivos para series de tiempo**

Para analizar las series de tiempo desde un punto de vista estocástico, utilizaremos métodos más complejos, los cuales toman en cuenta la distancia que existe entre los datos, es decir, que van a ser tomados de observaciones anteriores.

El modelo genérico para tratar las series de tiempo es el modelo ARIMA (Modelo Autorregresivo Integrado de Media Móvil) del cual se deriva los tres componentes que da lugar a sus siglas ARIMA, estos componentes son AR (Autorregresivo), I (Integrado) y MA (Medias móviles). Este método es el más usado para los pronósticos en el área de ingeniería eléctrica para el procesamiento de señales y energía en general [16], [53].

#### **1.9.3.1. Modelo Autorregresivo AR**

Este modelo se ha venido utilizando desde 1996, trata de una suma ponderada de observaciones basada en la variable escogida. El indicador  $p$  es el que representa el orden del modelo autorregresivo [16], [52].

El modelo autorregresivo se denota de la siguiente manera:

$$(Y_t - \delta) = a_1(Y_{t-1} - \delta) + a_2(Y_{t-2} - \delta) + \dots + a_p(Y_{t-p} - \delta) + u_t \quad (23)$$

Donde  $Y_t$  es la demanda de la energía eléctrica,  $\delta$  es la media de la demanda,  $u_t$  representa el error aleatorio y  $a_p$  es el parámetro que define el modelo (valor a estimar).

### 1.9.3.2. Modelo de Medias Móviles (MA (q))

El modelo de media móvil es una combinación lineal de términos de errores presentes y pasados [12], [52]. El modelo de media móvil está representado por la ecuación (24).

$$Y_t = \mu + \beta_0 u_t + \beta_1 u_1 + \beta_2 u_2 + \dots + \beta_q u_{t-q} \quad (24)$$

$\mu$  es el valor constante,  $u$  es el termino de error,  $q$  es el orden de media móvil.

### 1.9.3.3. Modelo autorregresivo y de media móvil (ARMA)

Fue introducido en 1951 como la combinación de términos autorregresivos y de media móvil, por lo que está compuesto de términos autorregresivos como también de medias móviles [16], [52]. El modelo autorregresivo y de media móvil viene dada por la siguiente ecuación (25):

$$Y_t = \theta + \alpha_t Y_{t-1} + \beta_0 u_t + \beta_1 u_{t-1} \quad (25)$$

Donde  $\theta$  es el valor constante,  $Y_{t-1}$  es la demanda de la energía eléctrica,  $\alpha_t$  es el parámetro a estimar,  $u_t$  término de error y  $\beta_t$  es el término a estimar.

### 1.9.3.4. Modelo autorregresivo integrado de media móvil (ARIMA)

El modelo ARIMA es una combinación de un término autorregresivo (AR) y un término de media móvil (MA) con un elemento diferenciador identificado por la letra I. Debe usarse siempre y cuando el proceso sea no estacionario, el cual será transformado a estacionario para así poder diferenciar si es un modelo ARMA, AR o MA. El modelo ARIMA es un modelo univariado, es decir, que compara los valores presentes con los pasados para así realizar el pronóstico [12], [16], [54].

El modelo ARIMA viene dada por la ecuación (26):

$$Y_t = \phi_1 Y_{t-1} + \dots + \phi_p Y_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \dots - \theta_q \varepsilon_{t-q} \quad (26)$$

Donde:

$\phi$  es el coeficiente autorregresivo,  $\theta$  es el coeficiente de media móvil,  $\varepsilon$  es el error,  $Y_{t-p}$  es el registro normalizado de la serie a modelar.

#### 1.9.3.4.1. Método Box Jenkins

Para determinar si un modelo es AR, MA, ARMA o ARIMA es necesario determinar los valores de  $p$ ,  $d$  y  $q$ . Para ello se emplea el método de Box Jenkins, el cual consta de 4 pasos detallados a continuación: [52].

#### Identificación del modelo

Para determinar el orden de  $p$ ,  $d$  y  $q$  es necesario usar la función de autocorrelación (FAC) y autocorrelación parcial (FACP) [52]. Antes de definir los conceptos de FAC y FACP es necesario conocer el concepto de estacionariedad.

### **Estacionariedad**

Un modelo estocástico es estacionario si y solo si sus vectores tienen la misma distribución de probabilidad. Por lo tanto, si calculamos sus medias y varianzas y no hay cambios significativos, es decir, que no hay tendencias crecientes ni decrecientes a lo largo de los años [53], [55]. En caso de no existir estacionariedad en los datos se procede a realizar una diferenciación, transformando el conjunto de datos con media y varianza cero. La diferenciación se realiza con la siguiente ecuación:

$$y_t = x_t - x_{t-n} \quad (27)$$

### **Estacionalidad**

Un modelo estocástico tiene estacionalidad cuando se observan patrones que se repiten a lo largo del tiempo en diferentes rangos. Este tipo de patrones se observa en grupos de datos que corresponden a datos de meses y años [56], [57].

### **Autocorrelación**

Esta función nos permite calcular la estacionariedad de una serie temporal. Está dada por la ecuación (28).

$$\rho_k = \frac{y_k}{y_0} = \frac{\text{Covarianza al rezago } k}{\text{varianza}} \quad (28)$$

Sus valores varían entre  $+1$  y  $-1$ . Para definir si es un modelo AR, MA o ARMA debemos emplear las funciones de autocorrelación FAC y FACP. La FACP determina la correlación entre los datos actuales y los datos pasados. En forma muestral la FACP está expresada por las siguientes ecuaciones:

$$\hat{y}_k = \frac{\sum(Y_t - \bar{Y})(Y_{t+k} - \bar{Y})}{n} \quad (29)$$

$$\hat{y}_0 = \frac{\sum(Y_t - \bar{Y})^2}{n} \quad (30)$$

Donde:

$\bar{Y}$  es la media muestral,  $n$  es el tamaño de la muestra,  $\hat{y}_k$  es la covarianza muestral,  $\hat{y}_0$  es la varianza muestral

Al graficar la función  $r_k$  (31), denominada como correlograma muestral permite observar de manera gráfica los coeficientes de autocorrelación de la serie de tiempo [53], [55].

$$r_k = \frac{\sum_{t=1}^{n-k} (Y_t - \bar{Y})(Y_{t+k} - \bar{Y})}{\sum_{t=1}^n (Y_t - \bar{Y})^2} \quad (31)$$

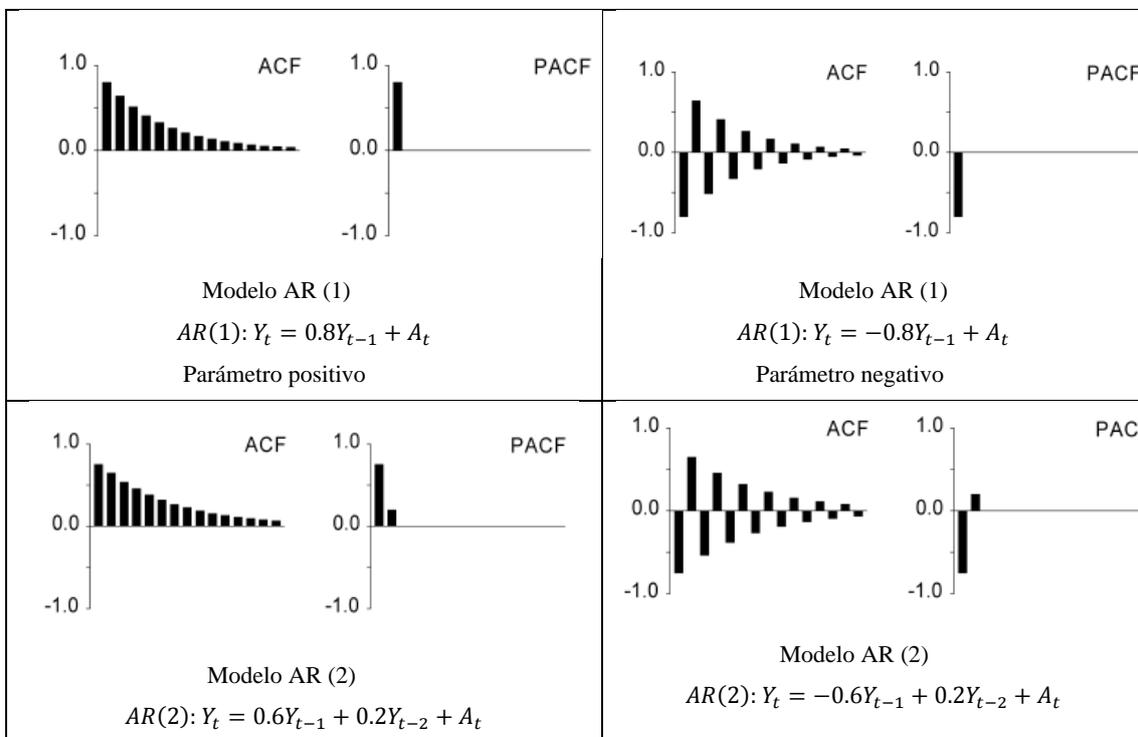
Para identificar los modelos autorregresivos de acuerdo a las correlaciones se usa la Tabla 1.1

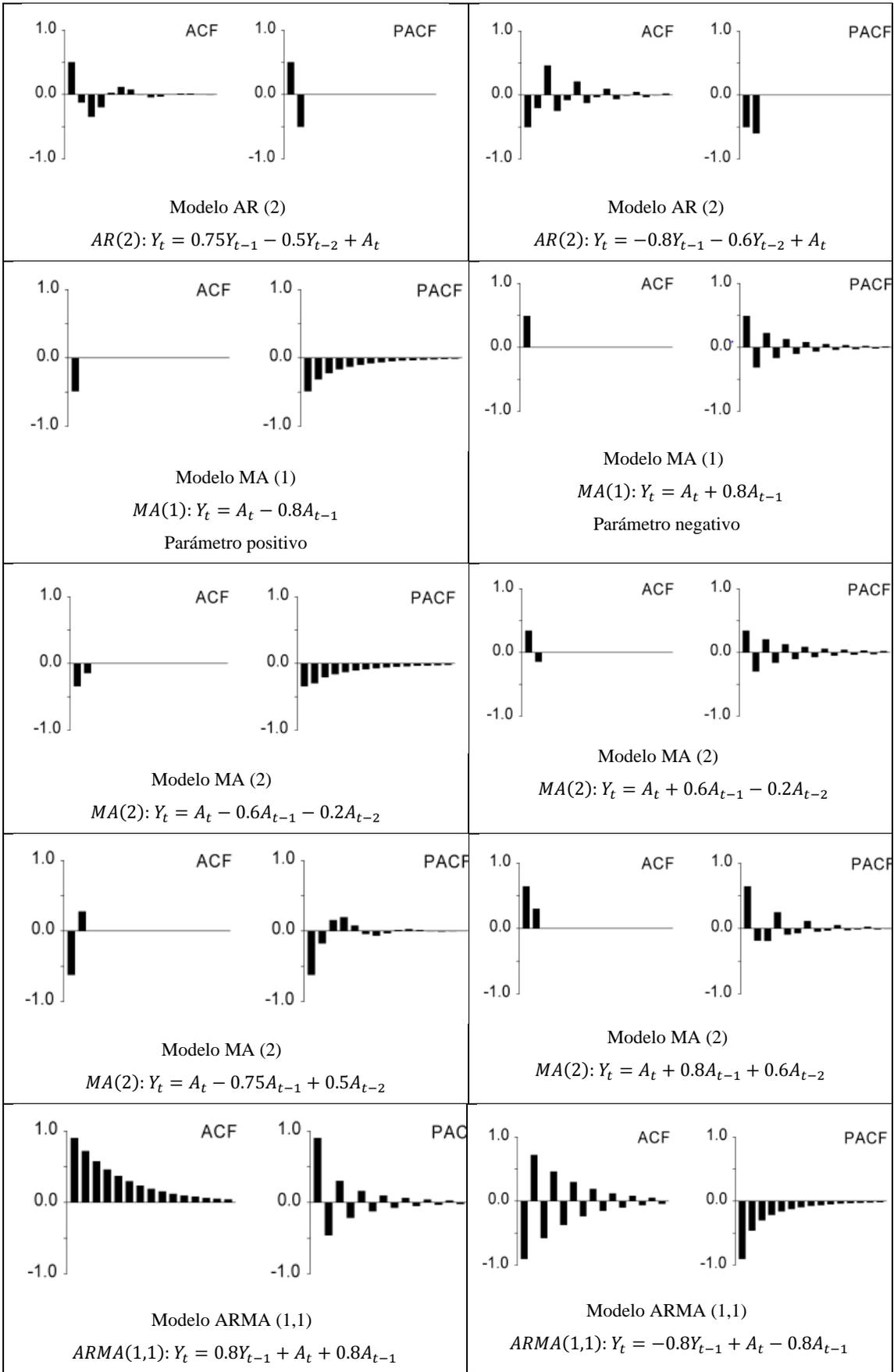
Tabla 1.1 Características de los modelos autorregresivos mediante los patrones de autocorrelación [55].

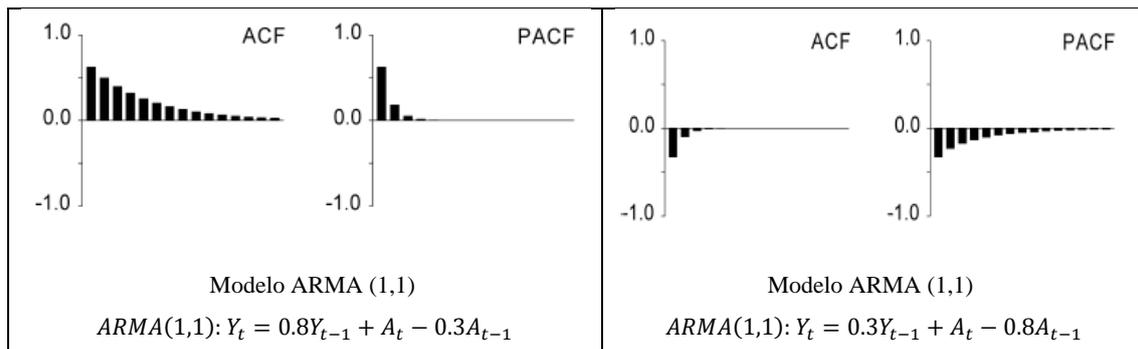
Modelo	Patrón Típico para FAC	Patrón Típico para FACP
AR (p)	Decrece de forma exponencial, o sinusoidal o ambos a la vez.	Picos significativos a lo largo de su trayectoria.
MA (q)	Picos grandes a lo largo de su trayectoria.	Decrece de forma exponencial sin llegar a anularse.
ARMA (p, q)	Disminuye de forma exponencial sin anularse.	

La elección de los modelos p, d y q serán elegidos en base a la Tabla 1.2.

Tabla 1.2 Modelos ARIMA[58]







### Estimación del modelo

Conocidos los valores de  $p$  y  $q$ , se procede a determinar los parámetros de los términos autorregresivos y de promedios móviles mediante el desarrollo de mínimos cuadrados o estimación no lineal [52], [55].

### Comprobación de diagnóstico

Una vez determinado el modelo ARIMA y establecidos los valores de  $p$  y  $q$ , el siguiente paso es verificar si este modelo seleccionado se ajusta a los datos de una forma razonable. En caso de no llegar a ajustarse correctamente el modelo, el proceso debe ser repetido, ya que puede existir otro modelo ARIMA que se ajuste de mejor manera [52].

### Pronóstico

El método ARIMA dado por Box Jenkins es el más eficaz y el más confiable a diferencia de los demás métodos de predicción debido a su alta precisión en especial para predicciones a corto plazo [52].

## 1.10. MÉTODOS BASADOS EN INTELIGENCIA ARTIFICIAL

### 1.10.1. Lógica difusa

Es una lógica basada en valores de incertidumbre generados por la falta de información, puede tomar valores desde 0 (totalmente falso) hasta 1 (totalmente verdadero).

Para poder estimar variables de interés, el proceso de lógica difusa consta de cuatro partes. La primera es la fusificación, que consiste en asignar valores difusos a la variable de interés, la segunda es construir las reglas de lógica difusa proporcionadas por un experto en el tema, el tercer paso es ejecutar todas las reglas y combinarlas para obtener una salida difusa y por último convertir la salida difusa en un valor no difuso. [59]

### 1.10.2. Redes neuronales

Este enfoque está basado en el cerebro que puede adaptarse a un proceso de aprendizaje para luego mediante el recuerdo poder realizar una tarea minimizando el riesgo empírico [16]. En el

proceso de aprendizaje, la red neuronal se entrena hasta establecer las relaciones entre los datos de entrada. Al ser un sistema de autoaprendizaje no es necesario tener conocimiento previo en el modelado para obtener resultados precisos [17].

Una desventaja de este método es que se requiere de un proceso de entrenamiento que puede durar mucho tiempo y por la gran cantidad de datos de entrada que necesita se produce un efecto de adaptación excesiva en la mayoría de ANN [16]. Por otra parte son muy inflexibles para establecer una relación específica no automática [17].

Su principal aplicación se basa en la efectividad de sus resultados ante la no linealidad de los datos, como la temperatura, radiación solar, velocidad del viento, humedad, etc. es por ello que se vuelve muy popular en la predicción de la carga de edificios [16].

### **1.10.3. Máquina de vectores de soporte**

Al igual que las redes neuronales este método ha sido popular en las últimas décadas por la efectividad de los resultados ante la no linealidad. A diferencia del primero, este método usa una pequeña cantidad de datos de los cuales puede identificar los patrones de carga, minimizando el riesgo estructural. Se presenta como una alternativa ante la adaptación excesiva que se produce con la mayoría de redes neuronales. Los antecedentes de este método han motivado su uso al compararlos con redes neuronales y otros modelos de predicción [16].

### **1.10.4. Sistemas expertos basados en el conocimiento**

Los sistemas expertos son sistemas informáticos que basan su funcionamiento en el conocimiento progresivo de acuerdo a nueva información proporcionada. La denominada base de conocimiento es extraída de un experto en el campo que, mediante relaciones entre los cambios en la carga del sistema y los cambios entre los factores ambientales, identifica hechos y reglas que modifican el uso de la electricidad. Estas relaciones lógicas se obtienen diariamente para establecer el pronóstico y han sido analizadas ampliamente para obtener diferentes enfoques entre la carga y el clima. [22].

## **1.11. MEDIDAS DE EVALUACIÓN**

### **1.11.1. Error absoluto Medio (MAE)**

Es una medida de error, la cual suma todos los valores absolutos pronosticados y los divide para el número de datos. Es de gran utilidad para evaluar procesos de pronóstico debido a que se da en las unidades originales [60], [61].

$$EAM = \frac{\sum_{t=1}^n |Y_t - \hat{Y}_t|}{n} \quad (32)$$

Donde:

$Y_t$  es la demanda real u original,  $\hat{Y}_t$  es la demanda Pronosticada,  $n$  es el número de datos

### 1.11.2. Raíz del error cuadrado medio (RMSE)

Consiste en calcular el error a través de la sumatoria de la diferencia de los valores reales y pronosticados elevados al cuadrado. Permite conocer el grado de dispersión del error de predicción [60], [62].

$$RECM = \sqrt{\frac{\sum_{t=1}^n (Y_t - \hat{Y}_t)^2}{n}} \quad (33)$$

### 1.11.3. Porcentaje de error absoluto medio (MAPE)

Nos permite medir el error en términos de porcentaje, su fácil interpretación hace que sea usado por quién realiza pronósticos. En este método de evaluación no es necesario que los datos tengan las mismas magnitudes [60].

$$PEAM = \frac{\sum_{t=1}^n \frac{100|Y_t - \hat{Y}_t|}{Y_t}}{n} \quad (34)$$

## 1.12. Criterios para evaluar un modelo ARIMA

### 1.12.1. Criterio AKAIKE INFORMATION CRITERION (AIC)

Este criterio permite determinar con qué eficiencia dichos modelos se ajustan a la base de datos y se puede aplicar a modelos ARMA o ARIMA, por lo que, el criterio que mejor se ajusta a los datos es el que tiene menor valor de AIC [12], [63].

$$AIC = -\frac{2L}{T} + \frac{2(p + d + q)}{T} \quad (35)$$

Donde:

$p$  son los parámetros Autorregresivos,  $d$  son las medias Móviles,  $q$  son las diferenciaciones para la serie,  $T$  son los instantes considerados en la serie,  $L$  es el logaritmo de la función de verosimilitud

### 1.12.2. Criterio de información Bayesiano de Schwarz (BIC)

Está basado en el teorema de Bayes, se diferencia del criterio anterior ya que acepta información externa al estudio y en conjunto con los datos observados, estima una distribución de probabilidad.

Al igual que el criterio AIC, el valor más bajo indica un mejor ajuste.

$$BIC = G * gl * \ln(n) \quad (36)$$

Donde:

$G$  es el cociente de verosimilitud,  $gl$  son los grados de libertad,  $n$  es el tamaño de la muestra.

## **2. CAPITULO 2. RECOPIACIÓN DE DATOS DE CONSUMO DE ELECTRICIDAD Y ANALISIS ESTADÍSTICO**

### **2.1. RECOPIACIÓN DE DATOS**

La base de datos para el análisis y pronóstico del consumo de energía eléctrica está constituida por datos meteorológicos y datos de consumo de energía eléctrica. Se recopilaron dos bases de datos de las cuales, la primera corresponde al edificio Cornelio Merchán de la Universidad Politécnica Salesiana sede Cuenca y la segunda base de datos corresponde a un edificio de la Universidad Politécnica de Valencia en España.

La Universidad Politécnica Salesiana ha implementado un sistema de medición inteligente cuyo análisis y estudio se encuentra en el trabajo [64]. En [65] se implementan mediciones inteligentes y aplicaciones Android para facilitar el manejo de información. En [66] se desarrolla mediante clúster una distribución de bases para redes inteligentes [67][68].

Con la ayuda del medidor trifásico DiMET3 instalado en la Universidad Politécnica Salesiana sede Cuenca se adquieren los datos de consumo de energía eléctrica del edificio Cornelio Merchán. Un total de 26404 datos registrados en el periodo comprendido entre el 8 de marzo y el 8 de diciembre de 2017.

La estación meteorológica de la UPS sede Cuenca proporciona datos meteorológicos con intervalos horarios. Los datos que serán empleados para el estudio son: potencia eléctrica, humedad relativa, radiación solar global, temperatura del aire, dirección y velocidad del viento.

También se utiliza la base de datos de un edificio de la UPV que comprende de 84,649 datos de consumo de energía eléctrica con sus respectivos valores de variables meteorológicas con intervalos de 15 minutos comprendidos entre el 1 de julio de 2014 y el 28 de noviembre de 2016. Las variables que se estudian en este caso son: potencia eléctrica, temperatura y humedad.

### **2.2. PROCESAMIENTO DE LOS DATOS**

Los datos han sido clasificados de acuerdo a la laborabilidad (día laborable y no laborable), día de la semana y su hora correspondiente. Para cada hora la base de datos cuenta con un registro cuarto horario, por lo que, se realizó un promedio de dichos valores para obtener uno por hora con cada variable implicada en el estudio.

El estudio se realiza para días laborables y no laborables. Estos se clasifican de acuerdo al consumo similar que puedan tener, para cada día se obtienen 24 grupos de datos, los cuales se han analizado para eliminar atípicos y su posterior obtención del perfil de consumo de electricidad.

Los valores atípicos para todos los casos considerados a continuación son los que tienen un valor absoluto mayor a 1.98, que corresponde aproximadamente al 95% del total de los datos en una distribución normal.

### 2.3 AGRUPACIÓN DE LOS DATOS

En este apartado se agrupan los datos de días que tengan un consumo de electricidad similar, a esto también se le llama clusterización. En nuestro estudio la agrupación de los datos está basado al consumo por hora, es decir, que existen gráficas de perfiles de consumo eléctrico de todos los días laborables y no laborables por hora, en donde se observa los consumos similares de dichos días.

Para demostrar y analizar de forma numérica los agrupamientos se toman datos de dos días y se realiza el cálculo de la distancia euclidiana entre los valores de potencia de cada hora. En la Tabla 2.1 se presenta el cálculo de la distancia euclidiana entre los perfiles de consumo de electricidad en una semana.

Tabla 2.1 Valores de distancia euclidiana entre los perfiles de consumo de electricidad de la UPV.

Distancia euclidiana	Lunes [kW]	Martes [kW]	Miércoles [kW]	Jueves [kW]	Viernes [kW]
Lunes	0	398.22	391.61	440.67	<b>867.00</b>
Martes	398.22	0	226.08	332.05	549.15
Miércoles	391.61	<b>226.08</b>	0	304.92	580.25
Jueves	440.67	332.05	304.92	0	680.96
Viernes	867.00	549.15	580.25	680.96	0

### 2.4. NORMALIZACIÓN DE LOS DATOS

Los valores que se obtienen de la distancia euclidiana se normalizan con ayuda del método de normalización Min Max o normalización lineal mediante la ecuación (10). Se estableció un valor umbral de 0.2, basado en las observaciones de los perfiles de carga, luego de múltiples pruebas. Si el valor es mayor a 0.2 los datos no se agrupan. En la tabla 2.2 se observan los valores normalizados de la Tabla 2.1.

Tabla 2.2 Valores de distancia euclidiana normalizados de la tabla 2.1.

Distancia euclidiana	Lunes	Martes	Miércoles	Jueves	Viernes
Lunes	0	0.27	0.26	0.33	1
Martes	0.26	0	0	0.17	0.50
Miércoles	0.25	0	0	0.12	0.55
Jueves	0.33	0.17	0.12	0	0.71
Viernes	1	0.50	0.55	0.71	0

Como se observa en la Tabla 2.2 el valor normalizado de la distancia euclidiana entre el día lunes y el resto de los días es mayor al umbral 0.2. Los datos que se usan para el pronóstico de los días lunes son los datos de consumo de ese día. Se puede observar que la distancia euclidiana del día miércoles con respecto al martes y jueves es de 0 y 0.1230 respectivamente, estos días pertenecen al mismo grupo. La distancia normalizada entre martes y miércoles es cero, pues como se indica en la tabla 2.1 este valor mínimo, es tomado como referencia para la normalización.

La distancia normalizada del día viernes con respecto al resto de días es mayor a 0.2. Para el pronóstico de los días viernes se usan solamente los datos de consumo de ese día. Con este análisis se forman los grupos para su posterior pronóstico. En este caso, el primer grupo lo conforma el lunes, el segundo el viernes y un tercer grupo está conformado por el martes, miércoles y jueves. En la Figura 2.1 se observa la clasificación de los días de consumo.

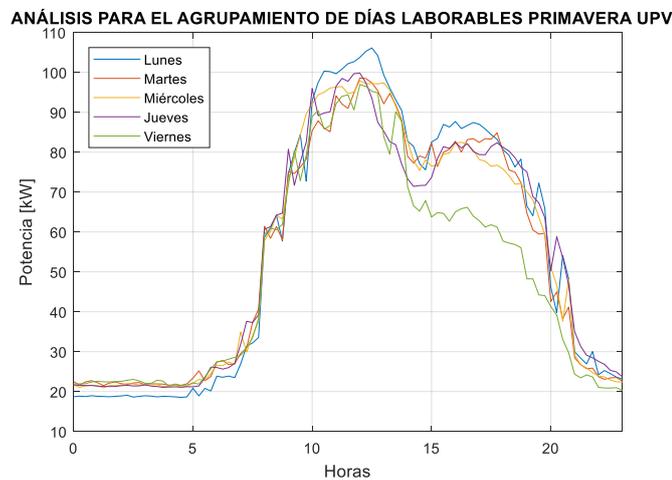


Figura 2.1 Perfil de consumo de electricidad para días laborables de la UPV en la primavera

Un entorno de programación en el software Matlab fue desarrollado para el cálculo del pronóstico como se observa en la Figura 2.2.

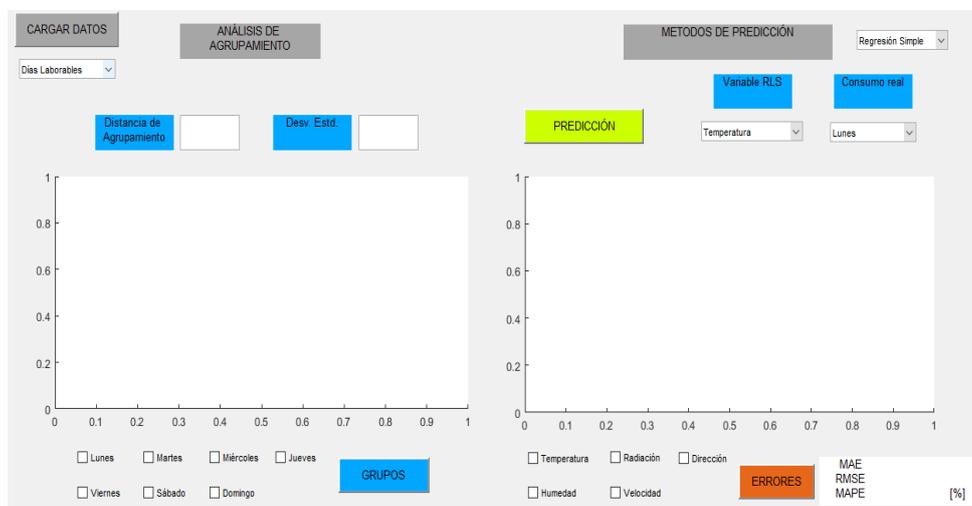


Figura 2.2 Entorno de programación visual desarrollado en MATLAB

A continuación, se realiza el análisis de la base de datos de la UPS y de la UPV. En ningún caso el coeficiente de determinación supera el 80%. Sin embargo la variable climática temperatura muestra la mejor relación con la potencia. Es por ello que se utiliza como principal variable independiente para los análisis.

## 2.5 BASE DE DATOS UPS

Se realiza inicialmente una exploración de los datos para días laborables y no laborables mediante el boxplot de Matlab. La mediana, los cuartiles y el valor máximo y mínimo de cada grupo de datos se muestran en la Figura 2.3. En la Tabla 2.3 se muestran estos valores.

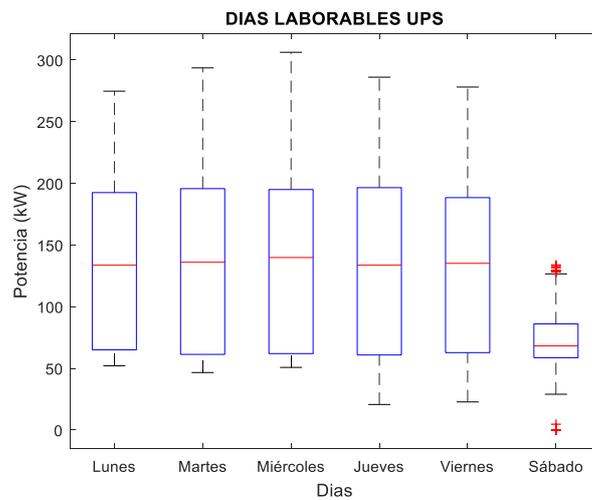


Figura 2.3 Representación gráfica de los datos correspondientes a días laborables en la UPS

La mediana es uniforme desde el día lunes al viernes. De lunes a viernes los valores del primer cuartil están agrupados y los valores entre el cuartil 3 y los máximos se encuentran dispersos. El día sábado es considerado como laborable, sin embargo, sus parámetros de análisis son diferentes al resto de días.

Tabla 2.3 Valores de medidas descriptivas para días laborables en la UPS.

Parámetro/Día	Lunes	Martes	Miércoles	Jueves	Viernes	Sábado
Media [kW]	132.47	135.18	135.25	133.61	131.48	73.83
Mediana [kW]	133.60	136.00	139.12	133.60	135.12	68.16
Desviación Estándar	63.74	66.97	65.91	65.87	62.26	17.62
Mínimo [kW]	52.00	46.56	50.72	20.64	22.72	24.64
Máximo [kW]	274.72	293.60	306.24	286.08	278.08	133.76

La mediana, el valor máximo y mínimo varían considerablemente entre los días no laborables. En la Figura 2.4 se muestra la representación de estos datos y en la tabla 2.4 se muestran los valores.

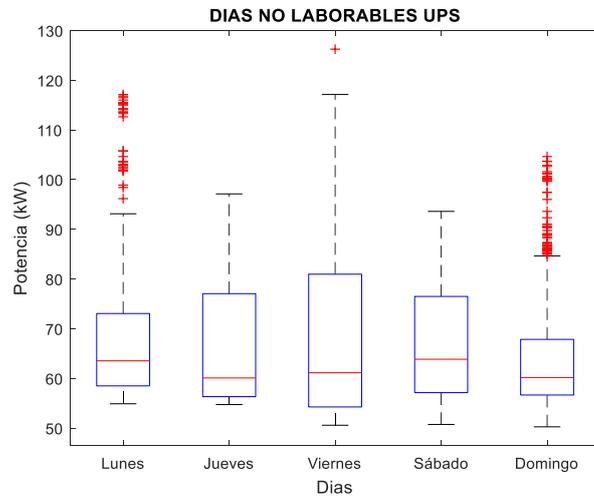
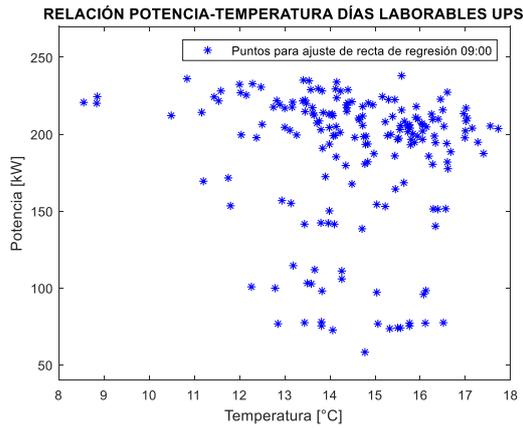


Figura 2.4 Representación gráfica de los datos correspondientes a días no laborables en la UPS

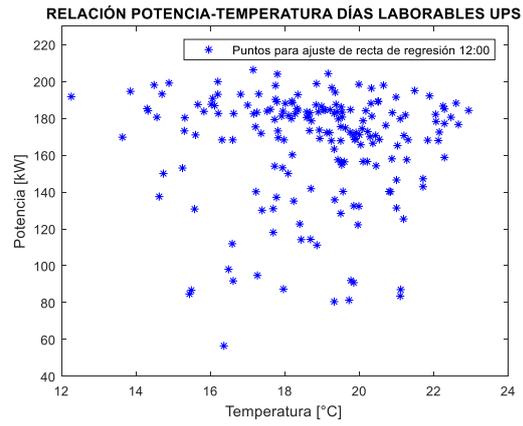
Tabla 2.4 Valores de medidas descriptivas para días no laborables en la UPS.

Medida/Día	Lunes	Jueves	Viernes	Sábado	Domingo
Media [kW]	70.86	67.41	69.00	67.08	62.62
Mediana [kW]	63.52	60.08	61.12	63.84	60.16
Desviación Estándar	18.426	14.103	17.793	11.668	8.017
Mínimo [kW]	54.880	54.720	50.560	50.720	50.240
Máximo [kW]	117.120	97.120	126.240	93.600	104.640

La demanda de energía eléctrica no se ve influenciada por la temperatura en la UPS sede Cuenca, tal como se revisó en el Capítulo 1. En la Figura 2.5 y 2.6 se puede observar de mejor manera la relación existente entre potencia y temperatura.

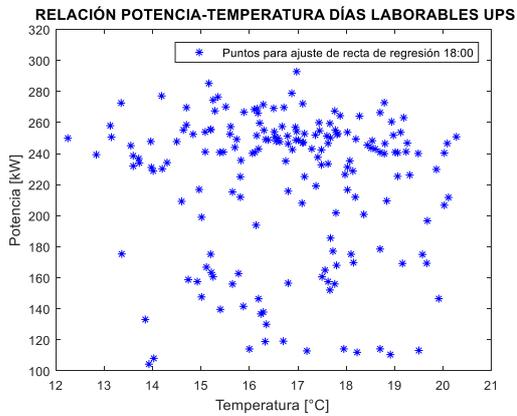


a)

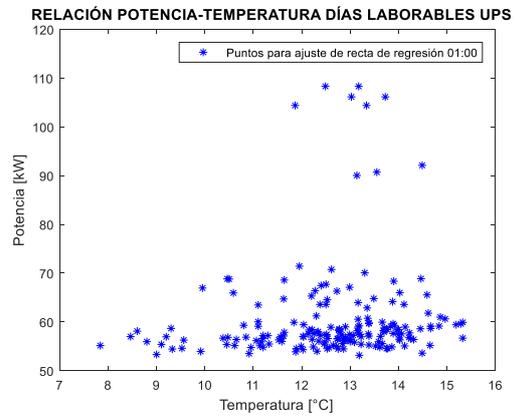


b)

Figura 2.5 Representación gráfica de la relación potencia-temperatura para las a) 09:00 y b) 12:00 horas del día.



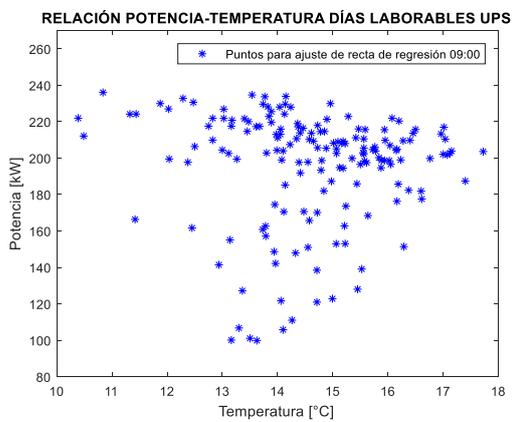
a)



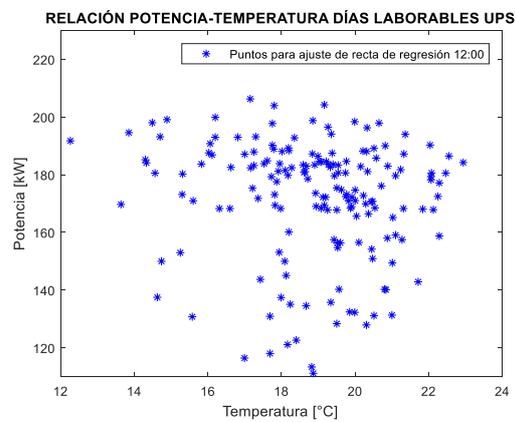
b)

Figura 2.6 Representación gráfica de la relación potencia-temperatura para las a) 18:00 y b) 01:00 horas del día.

Previo al cálculo de los coeficientes de las rectas de regresión se realiza la eliminación de valores atípicos usando puntuaciones z. En la gráfica 2.7 y 2.8 se pueden observar los valores calculados sin valores atípicos para una base de datos de la UPS en días laborables.

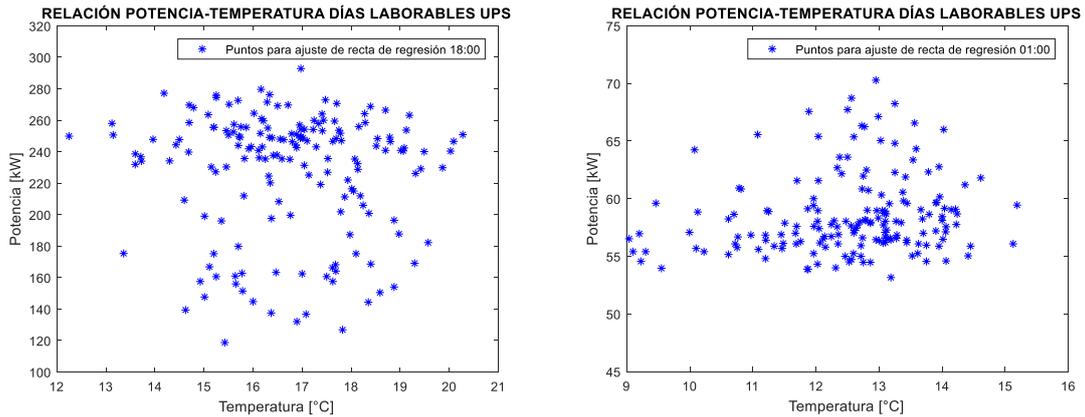


a)



b)

Figura 2.7 Representación gráfica de la relación potencia-temperatura para las a) 09:00 y b) 12:00 horas del día sin valores atípicos



a) b)  
 Figura 2.8. Representación gráfica de la relación potencia-temperatura para las a) 18:00 y b) 01:00 horas del día sin valores atípicos

El coeficiente de determinación  $R^2$  es calculado para determinar la relación entre la potencia y las variables climáticas. En la Tabla 2.5 se muestran los resultados. Por el bajo porcentaje de relación entre la potencia y la dirección del viento no se involucra en el análisis esta variable.

Tabla 2.5 Coeficiente de determinación de varias variables con respecto a la potencia demandada para días laborables y no laborables

Laborabilidad /Var. Climática	Temperatura [%]	Humedad [%]	Radiación Solar [%]	Velocidad del Viento [%]	Dirección del Viento [%]
Laborable	21.030	15.90	11.35	12.73	0.12
No Laborable	3.10	1.90	1.35	2.39	0.17

En las Figuras 2.9 y 2.10 se conocen los perfiles de consumo de los días laborables y no laborables respectivamente, el día sábado se considera como laborable.

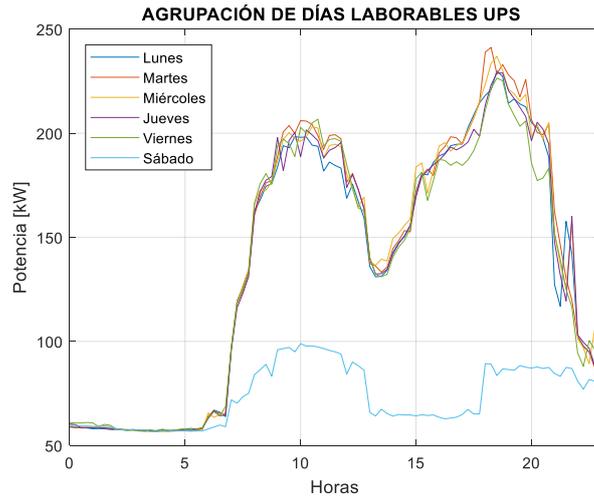


Figura 2.9 Perfil de consumo de electricidad de días laborables de la UPS.

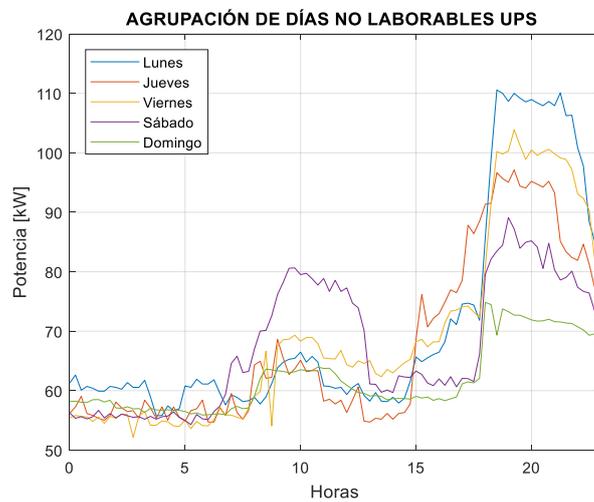


Figura 2.10 Perfil de consumo de electricidad de días no laborables en la UPS.

En la Tabla 2.6 se muestra la forma en la que se agrupan los días laborables y no laborables. Como se ve, los días laborables se dividen en dos grupos. Dentro del primero grupo se encuentran los días de lunes a viernes y en el segundo grupo está el día sábado. Por otro lado, los días no laborables se dividen en 4 grupos. El primero corresponde al lunes, el segundo al jueves y viernes, el tercero al sábado y el cuarto al domingo.

Tabla 2.6 Clasificación de grupos para días laborables y no laborables de la UPS.

U P S	Grupo 1							Grupo 2							Grupo 3							Grupo 4						
	L	M	M	J	V	S	D	L	M	M	J	V	S	D	L	M	M	J	V	S	D	L	M	M	J	V	S	D
Lab	x	x	x	x	x								x															
No Lab	x										x	x								x					x			

## 2.6 BASE DE DATOS UPV

La gráfica exploratoria de datos mediante el comando boxplot para la UPV que corresponden a los días laborables sin separar en estaciones se muestra en la Figura 2.11.

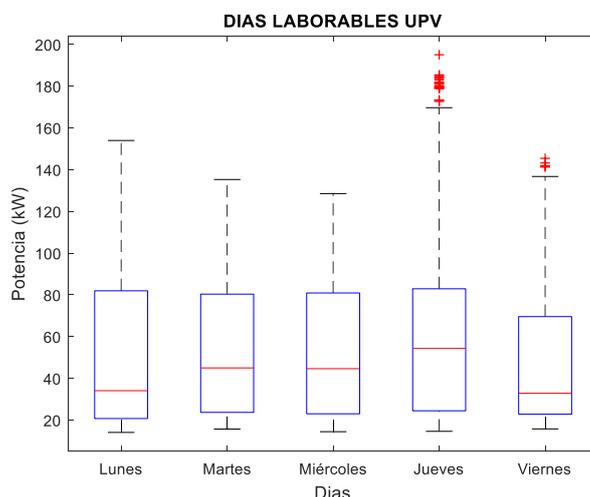


Figura 2.11 Gráfica descriptiva de los datos para los días laborables de la UPV.

Las características observadas en la Figura 2.11 son: Valores de mediana muy variables entre cada día. Dispersión de datos entre el tercer cuartil y el valor máximo. Agrupación de datos entre el mínimo y el primer cuartil. Presencia de valores atípicos en los días jueves y viernes.

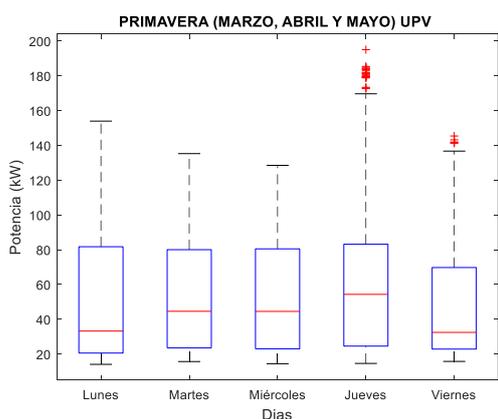
En la Tabla 2.7 se indican algunos valores de medida para la base de datos de la UPV en los días laborables.

Tabla 2.7 Medidas descriptivas para el análisis de los días laborables de la UPV.

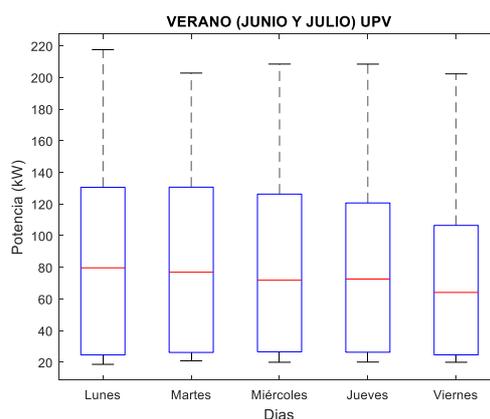
Medida/Día	Lunes	Martes	Miércoles	Jueves	Viernes
Media [kW]	51.98	53.00	52.54	56.70	47.21
Mediana [kW]	34.06	44.98	44.62	54.32	32.83

Desviación Estándar	34.80	30.96	31.38	34.67	29.11
Mínimo [kW]	14.07	15.68	14.42	14.60	15.76
Máximo [kW]	153.92	135.24	128.48	170.00	140.36

Con el agrupamiento de los datos por estaciones se han obtenido los resultados que se muestran en las Figuras 2.12 y 2.13.

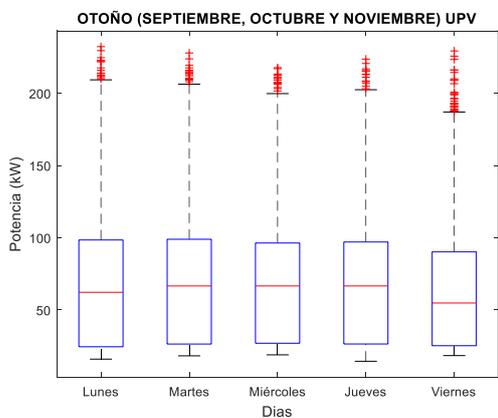


a)

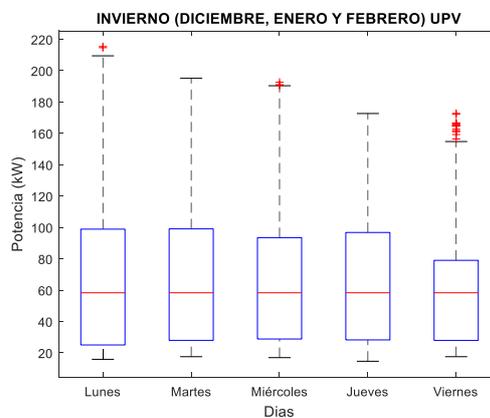


b)

Figura 2.12 Gráfica descriptiva para las estaciones de a) primavera y b) verano de la UPV



a)



b)

Figura 2.13 Gráfica descriptiva para las estaciones de a) otoño e b) invierno de la UPV

En las Figuras 2.12 y 2.13 la mediana es más uniforme principalmente en el verano y en el invierno. Por otra parte, los valores máximos y mínimos también se mantienen similares entre un día y otro. En el elemento a) de la Figura 2.12 resalta la presencia de valores atípicos correspondientes al otoño. Otros valores de medida descriptiva de los datos de cada estación se indican en las Tablas 2.8 a 2.11.

Tabla 2.8 Medidas descriptivas para el análisis de la primavera en días laborables de la UPV

Medida/Día	Lunes	Martes	Miércoles	Jueves	Viernes
Media [kW]	51.83	52.82	52.39	57.01	47.11
Mediana [kW]	33.22	44.58	44.52	54.32	32.44
Desviación Estándar	34.92	30.91	31.35	35.00	29.23
Mínimo [kW]	14.07	15.68	14.42	14.60	15.76
Máximo [kW]	153.92	135.24	128.48	167.00	145.36

Tabla 2.9 Medidas descriptivas para el análisis del verano en días laborables de la UPV

Medida/Día	Lunes	Martes	Miércoles	Jueves	Viernes
Media [kW]	82.33	82.57	81.25	77.83	70.59
Mediana [kW]	79.64	76.88	71.80	72.56	64.10
Desviación Estándar	57.96	55.03	54.11	51.95	48.10
Mínimo [kW]	18.64	20.88	19.96	20.08	19.96
Máximo [kW]	217.52	202.76	208.44	208.40	202.24

Tabla 2.10 Medidas descriptivas para el análisis del otoño en días laborables de la UPV

Medida/Día	Lunes	Martes	Miércoles	Jueves	Viernes
Media [kW]	68.13	69.64	69.76	68.77	62.51
Mediana [kW]	62.24	66.62	66.64	66.64	54.84
Desviación Estándar	47.95	47.05	44.71	45.00	41.68
Mínimo [kW]	16.00	18.28	18.96	14.52	18.44
Máximo [kW]	210.28	205.88	198.68	203.64	195.28

Tabla 2.11 Medidas descriptivas para el análisis del invierno en días laborables de la UPV.

Medida/Día	Lunes	Martes	Miércoles	Jueves	Viernes
Media [kW]	65.48	67.00	65.47	65.92	59.01
Mediana [kW]	58.32	58.32	58.32	58.32	58.32
Desviación Estándar	42.82	39.74	37.62	37.70	33.54
Mínimo [kW]	15.88	17.56	16.96	14.64	17.56
Máximo [kW]	215.36	195.16	192.64	172.68	162.72

Para los días no laborables se tiene los resultados de la Figura 2.14. Valores como la mediana, máximos y mínimos son muy variables para cada día.

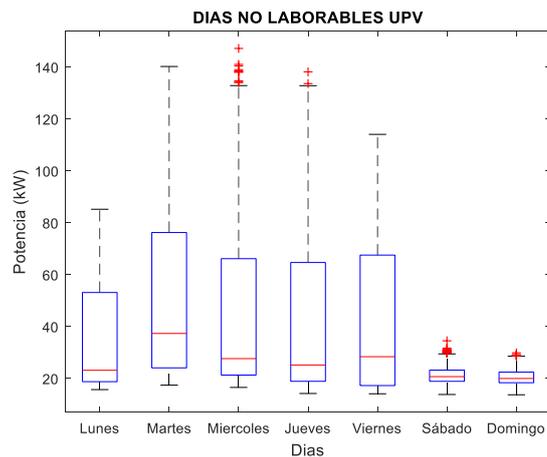


Figura 2.14 Gráfica descriptiva de los datos para los días no laborables de la UPV.

En la Tabla 2.12 se muestran otras variables adicionales de los días no laborables.

Tabla 2.12 Medidas descriptivas para el análisis del otoño en días no laborables de la UPV.

Medida/Día	Lunes	Martes	Miércoles	Jueves	Viernes	Sábado	Domingo
Media [kW]	33.94	53.91	46.22	42.51	43.67	21.15	20.42
Mediana [kW]	23.14	37.31	27.62	25.16	28.36	20.68	19.97
Desviación Estándar	22.09	35.38	32.07	32.30	28.86	3.23	2.94
Mínimo [kW]	15.68	17.40	16.52	14.18	14.07	13.79	13.69
Máximo [kW]	85.18	140.20	135.16	138.12	114.04	34.48	29.84

La representación gráfica de la potencia-temperatura mostrada en la Figura 2.15 y Figura 2.16 permite observar la relación entre estas variables. Los elementos a), b) de la Figura 2.15 y el elemento a) de la Figura 2.16 corresponden a horarios académicos que muestran una relación no lineal entre la potencia y temperatura. Esta curva en forma de U obedece a la influencia que tiene la temperatura en el uso de equipos como aire acondicionado y calefacción.

El elemento b) de la Figura 2.16 representa la relación entre la potencia y la temperatura a la 01:00 am, existiendo una relación mayormente lineal. Por la ausencia de estudiantes en el edificio la potencia no se ve afectada ante los cambios de temperatura.

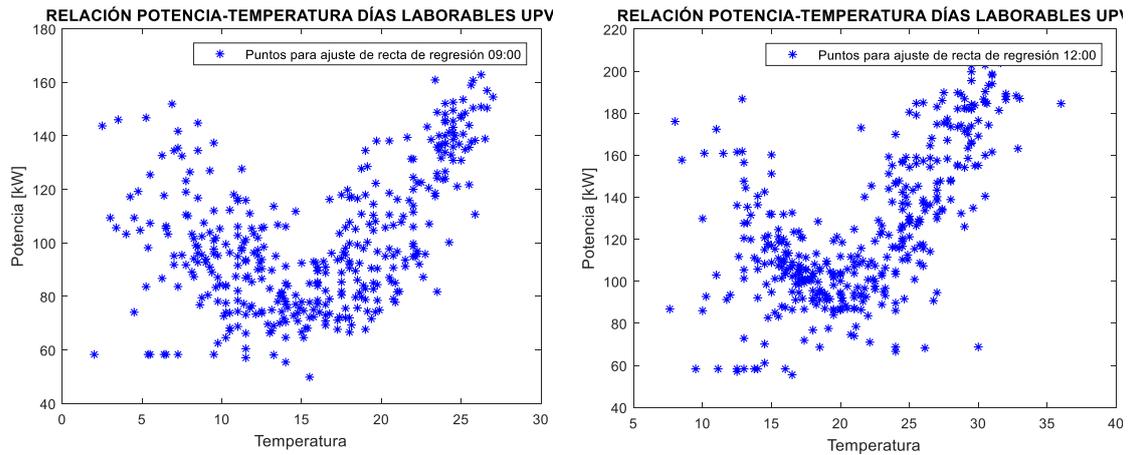
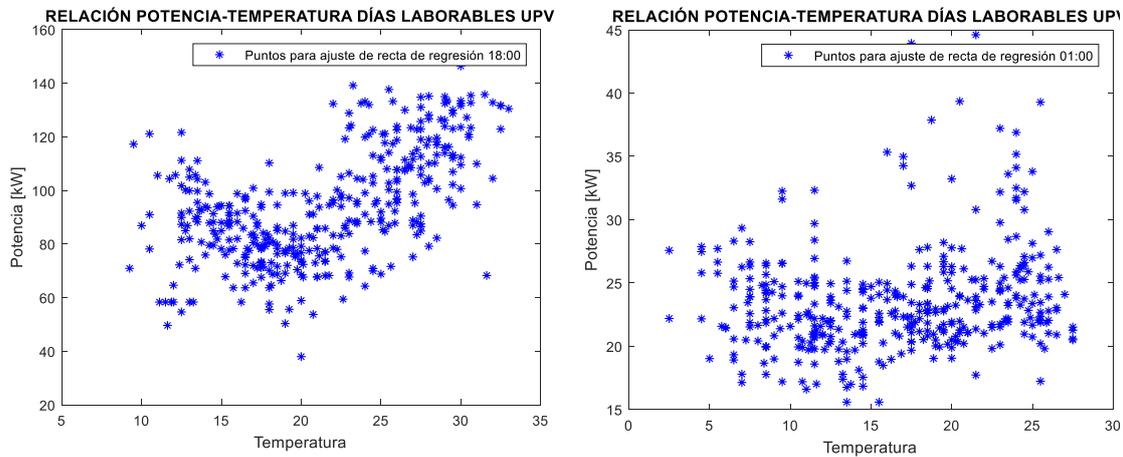


Figura 2.15 Relación no lineal potencia-temperatura de la UPV, a) 09:00 y b) a la 12:00



a)

b)

Figura 2.16 Relación no lineal potencia-temperatura de la UPV, a) 18:00 y b) a la 01:00.

Para conocer la relación de las variables climáticas temperatura y humedad con la potencia se usa el coeficiente de determinación  $R^2$ . Los resultados se muestran en la Tabla 2.13.

Tabla 2.13 Coeficientes de determinación para días laborables y no laborables de la UPV.

Laborabilidad /Var. Climática	Temperatura [%]	Humedad [%]
Laborable	19,42	15,33
No Laborable	1,20	3,03

Para mejorar el ajuste de la recta de regresión a los valores de potencia, los datos de la UPV fueron segmentados por los meses que corresponden a las estaciones de Valencia en España de la siguiente manera:

- Primavera: marzo, abril y mayo
- Verano: junio y julio
- Otoño: septiembre, octubre y noviembre
- Invierno: diciembre, enero y febrero

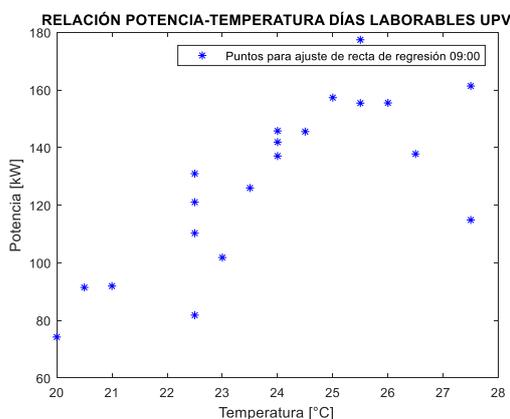
El fin de la segmentación es realizar un ajuste adecuado de los datos a una recta lineal de regresión, ya que inicialmente la gráfica potencia-temperatura mostraba una relación no lineal. En la Tabla 2.14 se indican los valores del coeficiente de determinación con la separación en estaciones de los datos.

Tabla 2.14 Coeficiente de determinación por estaciones para días laborables y no laborables de la UPV.

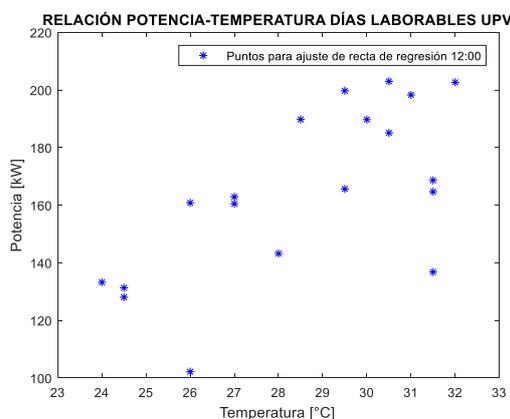
Laborabilidad	Estación	Temperatura [%]	Humedad [%]
Laborable	Primavera	23,92	19,69
	Verano	35,31	25,22
	Otoño	27,26	19,87
	Invierno	10,72	8,41
No Laborable	Todo el año	1,20	3,03

El coeficiente de determinación para cada estación de la Tabla 2.14 es mayor en relación al obtenido en la Tabla 2.13 que corresponden a datos de todo el año. Hay que tomar en cuenta que en las horas donde no existen estudiantes en el edificio la relación  $R^2$  es baja, lo que afecta el valor total obtenido.

En las Figuras 2.17 y 2.18 se puede observar la relación de las variables en el verano.

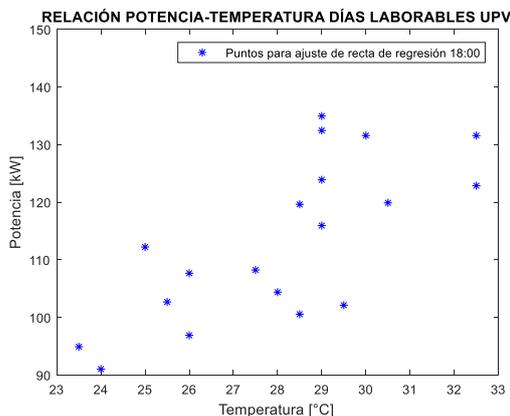


a)

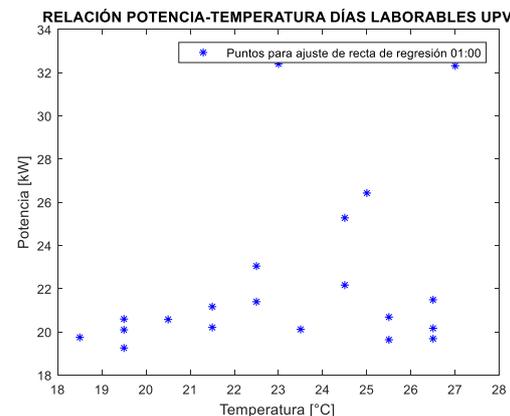


b)

Figura 2.17 Relación lineal potencia-temperatura de la UPV a las a) 09:00 y b) 12:00.



a)

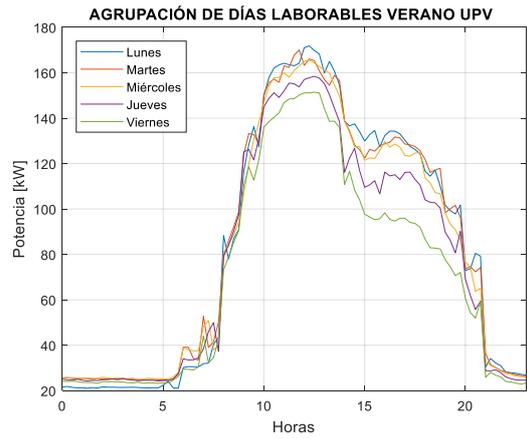
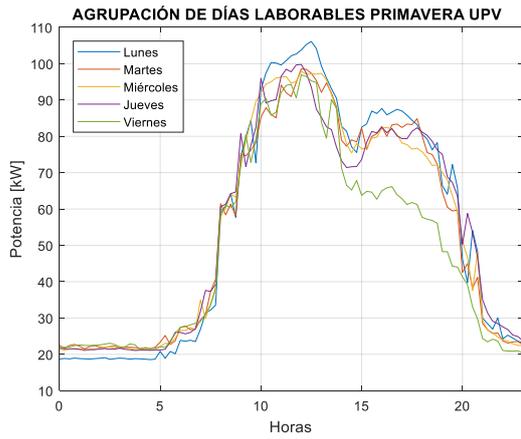


b)

Figura 2.18 Relación lineal potencia-temperatura de la UPV a las a) 18:00 y b) 01:00.

Los elementos a), b) de la Figura 2.17 y a) de la Figura 2.18 corresponden a datos en horarios académicos donde existe una relación proporcional directa de la potencia-temperatura. El elemento b) de la Figura 2.18 corresponde a un horario donde no existen estudiantes lo que evidencia una pobre relación entre las dos variables.

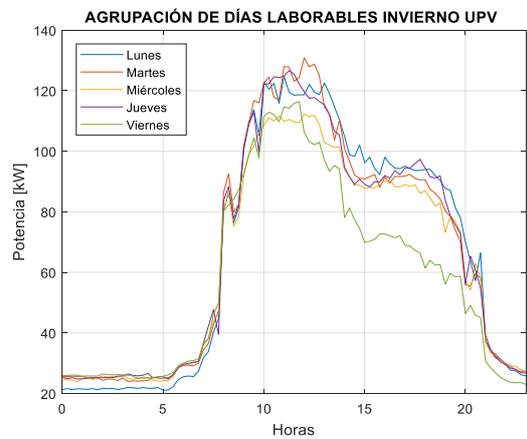
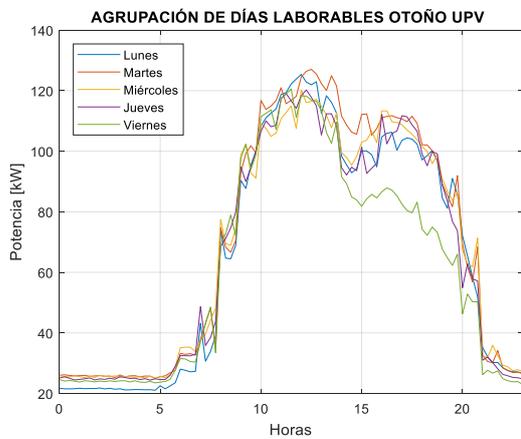
El perfil de consumo de días laborables y no laborables se muestran en las Figuras 2.19. y 2.20.



a)

b)

Figura 2.19 Perfiles de consumo de electricidad por estaciones en la UPV, a) primavera, b) verano



a)

b)

Figura 2.20 Perfiles de consumo de electricidad por estaciones en la UPV, a) otoño, b) invierno.

En la Figura 2.21 se muestra el perfil de los días no laborables correspondientes a la UPV.

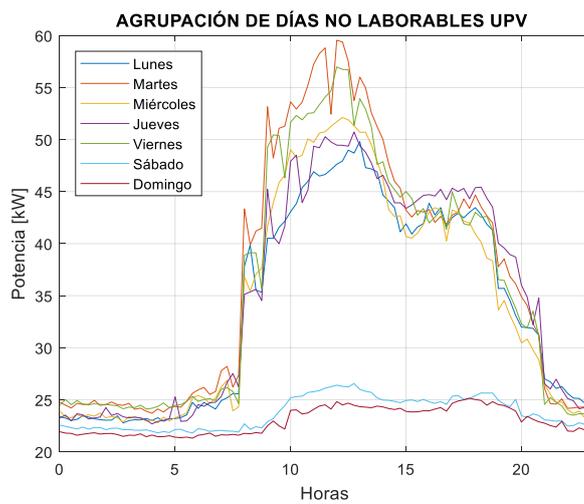


Figura 2.21 Perfiles de consumo de electricidad de días no laborables de la UPV. Autores

La forma en la que se agrupan los días laborables para cada estación y de los días no laborables se muestra en la Tabla 2.15.

Tabla 2.15 Clasificación de grupos para días laborables y no laborables de la UPV

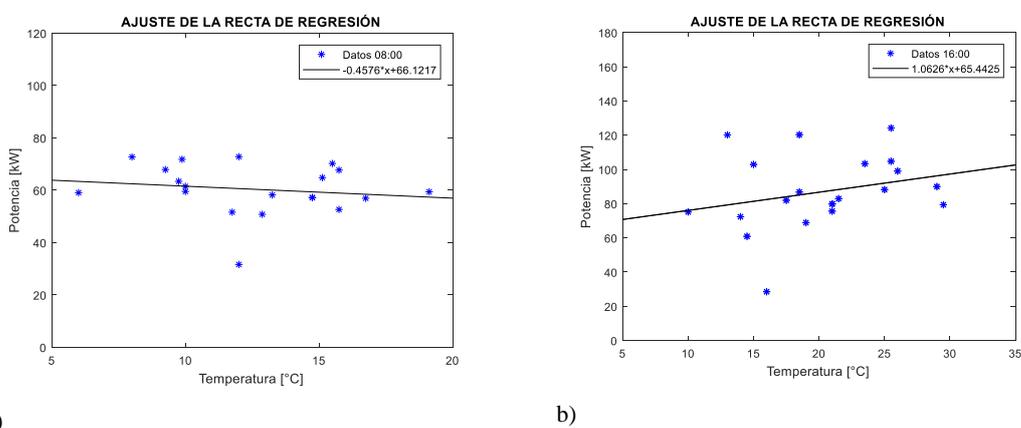
UPV		Grupo 1							Grupo 2							Grupo 3							
		L	M	M	J	V	S	D	L	M	M	J	V	S	D	L	M	M	J	V	S	D	
Lab	Primavera	x								x	x	x										x	
	Verano	x	x	x								x										x	
	Otoño	x			x					x	x											x	
	Invierno	x	x	x	x									x									
No Lab	-	x	x	x	x	x									x	x							

### 3. CAPÍTULO 3. APLICACIÓN DE MÉTODOS ESTADÍSTICOS

En este capítulo se aplican los métodos de Regresión Lineal Simple, Regresión Lineal Múltiple y ARIMA. Por medio de un entorno de programación en MATLAB se realiza la predicción con cada método, el detalle del software se presenta en el *manual de usuario* adjunto en el Anexo 1.

#### 3.1. REGRESIÓN LINEAL SIMPLE

Con los datos de potencia y de las variables climáticas se ajusta una recta de regresión lineal para cada hora del día. Para la predicción de un día se tiene 24 rectas de regresión. La Figura 3.1 muestra el ajuste de la recta de regresión para las 08:00am y 16:00pm para los días lunes considerando un periodo de datos de 24 días. El objeto de este ajuste es predecir la demanda de energía para el día lunes 30 de mayo de 2016 en la UPV con la temperatura como variable independiente.



a)

b)

Figura 3.1 Ajuste de la recta de regresión para el pronóstico del día lunes 30 de mayo de 2016 en la UPV a las a) 08:00 y b) 16:00 horas.

En la Tabla 3.1 se muestran los valores de los coeficientes  $b_0$  y  $b_1$  de la ecuación 13 para todas las horas del ejemplo de la Figura 3.1.

Tabla 3.1 Valores de los coeficientes en la recta de regresión

Hora	$b_0$	$b_1$	Potencia pronosticada [kW]
0:00	14.35	0.30	20.64
1:00	13.88	0.34	20.77
2:00	12.64	0.45	21.42
3:00	13.39	0.41	21.38
4:00	13.55	0.39	20.81
5:00	16.31	0.30	21.33
6:00	20.22	0.29	25.26
7:00	25.32	0.48	33.18
8:00	66.12	-0.46	57.89

9:00	85.28	-0.51	74.84
10:00	101.83	-0.26	96.05
11:00	101.93	-0.03	101.10
12:00	104.35	0.02	104.98
13:00	87.22	0.45	97.90
14:00	56.46	1.23	87.33
15:00	74.87	0.49	87.42
16:00	65.44	1.06	92.01
17:00	70.23	0.74	87.73
18:00	64.17	0.70	81.07
19:00	46.36	1.09	72.60
20:00	28.45	1.04	52.28
21:00	20.16	0.53	31.47
22:00	18.13	0.40	26.27
23:00	16.95	0.42	25.07

Para el cálculo del pronóstico también se necesitan los datos pronosticados horarios de la variable independiente. Estos valores son reemplazados en las rectas de regresión para obtener el valor de consumo de electricidad pronosticado. A modo de ejemplo se realiza el cálculo del valor pronosticados para las 11:00 am.

La recta de regresión para las 11:00am es  $\hat{Y} = -0.0337x + 101.9307$ . Para esa hora, se espera que la temperatura sea de 24.5°C. En la Figura 3.2 se muestra la potencia pronosticada con el valor calculado a continuación.

$$\hat{Y} = -0.0337(24.5) + 101.9307 = 101.105 \text{ kW}$$

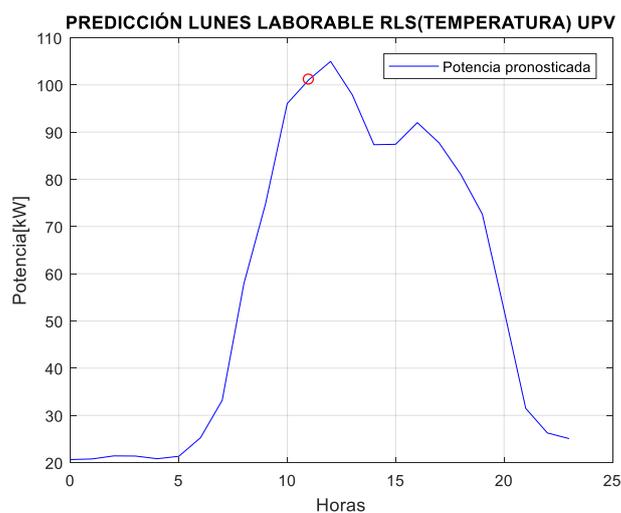


Figura 3.2 Potencia pronosticada del día lunes 30 de mayo de 2016 en la UPV

El mismo análisis se realiza con la humedad, en la Figura 3.3 se obtiene las rectas de regresión para las 08:00am y 16:00pm.

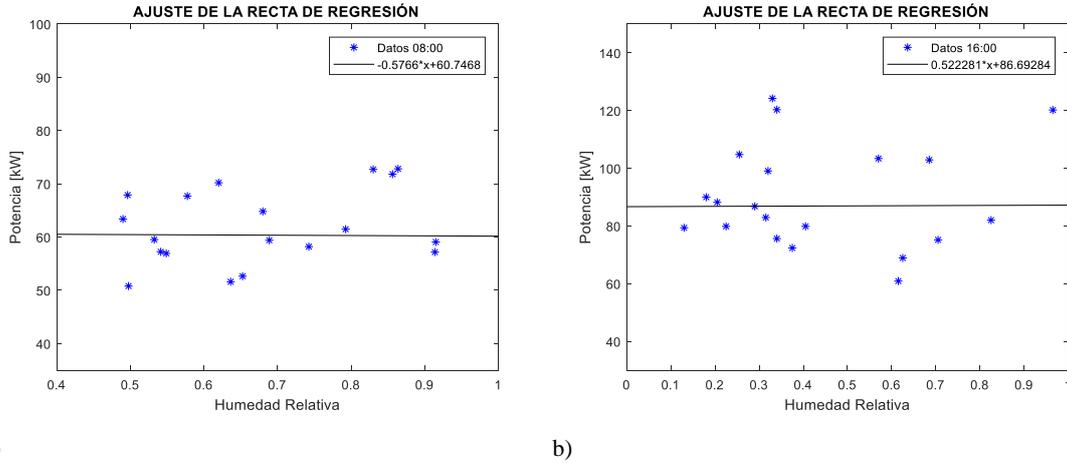


Figura 3.3 Ajuste de la recta de regresión para el pronóstico del día lunes 30 de mayo de 2016 en la UPV a las a) 08:00 y b) 16:00 horas.

Los valores de los coeficientes de regresión se muestran en la Tabla 3.2. Previo al cálculo del pronóstico de la potencia se necesita conocer los valores pronosticados de la humedad por hora.

Tabla 3.2 Valores de los coeficientes b0 y b1 de las rectas de regresión

Hora	b0	b1	Potencia pronosticada [kW]
0:00	19.92	-1.86	19.43
1:00	20.22	-2.26	19.38
2:00	20.85	-2.99	19.62
3:00	20.60	-2.59	19.53
4:00	20.96	-3.35	19.49
5:00	23.30	-4.48	21.06
6:00	27.88	-5.84	25.11
7:00	31.98	-1.35	31.32
8:00	60.75	-0.58	60.48
9:00	77.05	1.39	77.57
10:00	83.15	26.92	91.89
11:00	92.67	17.31	96.82
12:00	99.01	12.31	101.84
13:00	89.37	14.44	93.63
14:00	92.66	-26.85	85.27
15:00	81.04	8.55	83.61
16:00	86.69	0.52	86.85
17:00	85.23	-0.28	85.12
18:00	81.13	-7.03	78.85

19:00	72.62	-11.63	68.90
20:00	65.04	-31.84	52.62
21:00	33.50	-7.21	29.68
22:00	31.28	-10.41	25.04
23:00	32.30	-13.01	24.56

Los resultados obtenidos del pronóstico de potencia con la temperatura y la humedad como variables independientes se muestran en la Figura 3.4. También se muestra el consumo real, el cual se usa para comparar y obtener los errores de pronóstico.

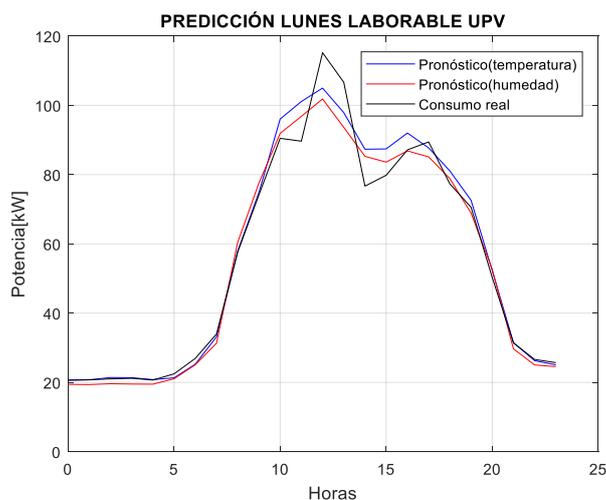


Figura 3.4 Pronóstico usando RLS del día lunes 30 de mayo de 2016 para un edificio en la UPV

Los errores que se utilizan para evaluar los métodos son: error medio absoluto medio (MAE), raíz del error cuadrático medio (RMSE) y el porcentaje del error absoluto medio (MAPE). Para el cálculo de los errores se utiliza el promedio de los errores calculados por cada hora en todo el día.

Los errores que se obtienen del ejemplo de la Figura 3.4 se indican en la Tabla 3.3.

Tabla 3.3 Errores de pronóstico de la Figura 3.4

Errores	RLS	
	Temperatura como variable independiente	RLS Humedad como variable independiente
MAE [kW]	3.12	3.37
RMSE [kW]	4.87	4.86
MAPE [%]	4.21	5.94

Para el caso de la Universidad Politécnica Salesiana se sigue el mismo procedimiento y se utilizan las variables climáticas como son: temperatura, humedad, radiación solar y velocidad del viento como se muestra en la Figura 3.5.

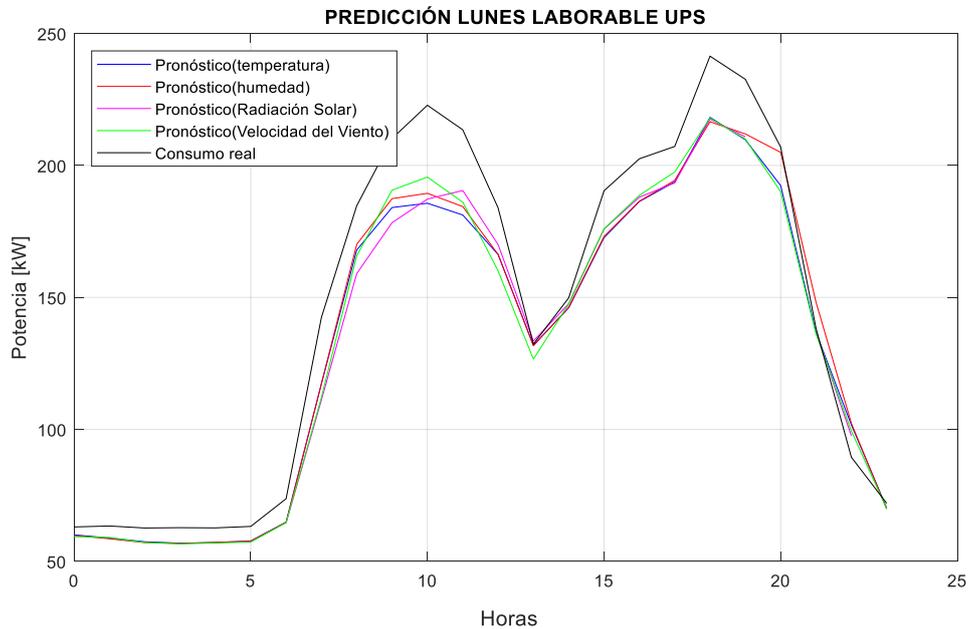


Figura 3.5 Pronóstico del lunes 4 de diciembre de 2017 en un edificio de la Universidad Politécnica Salesiana

Para el caso de la radiación solar, el pronóstico de la potencia de una hora se obtiene siempre que esta variable climática tenga un valor de radiación en dicha hora. Cuando la radiación solar sea cero, el valor de la potencia no se puede determinar pues se obtiene un valor infinito en el coeficiente de la recta de regresión que representa la pendiente de la recta. El pronóstico de las horas faltantes se puede obtener con el uso de otra variable climática.

### 3.2.REGRESIÓN LINEAL MÚLTIPLE

Este método usa todas las variables climáticas independientes en conjunto para calcular el pronóstico. Se realiza el ajuste a un modelo de regresión múltiple por medio de coeficientes que se ajustan a través de mínimos cuadrados y que son iguales al número de las variables independientes más uno.

Para el caso de la UPV las variables independientes son la temperatura y la humedad. A modo de ejemplo se realiza el pronóstico del día lunes 30 de mayo de 2016 para un edificio de la UPV.

Los coeficientes obtenidos para la predicción están resumidos en la Tabla 3.4.

Tabla 3.4 Coeficientes del modelo de regresión múltiple

Hora	b0	b1	b2	Potencia Pronosticada [kW]
0:00	14.70	0.29	-0.42	20.73
1:00	14.12	0.33	-0.27	20.81
2:00	12.98	0.44	-0.37	21.47
3:00	13.81	0.40	-0.47	21.46
4:00	15.16	0.37	-1.92	21.17
5:00	19.22	0.27	-3.50	21.88
6:00	24.29	0.22	-4.62	25.97
7:00	24.68	0.48	0.77	33.03
8:00	69.11	-0.50	-3.56	58.44
9:00	87.76	-0.55	-2.96	75.27
10:00	76.13	0.31	30.40	92.82
11:00	75.26	0.71	26.53	98.94
12:00	77.16	0.82	25.21	103.97
13:00	43.62	1.66	43.55	96.41
14:00	56.20	1.24	0.28	87.32
15:00	25.66	1.96	44.62	88.97
16:00	20.40	2.38	42.14	92.48
17:00	26.03	2.04	39.56	90.09
18:00	53.50	0.98	11.71	80.73
19:00	30.39	1.52	15.78	71.80
20:00	55.77	0.31	-25.09	53.02
21:00	11.22	0.75	8.02	31.66
22:00	23.25	0.28	-5.01	26.04
23:00	25.10	0.25	-8.14	25.20

Con los valores de los coeficientes y de las variables climáticas por cada hora se obtiene el pronóstico. Como ejemplo, a continuación, se calcula el valor de la potencia pronosticada para las 12:00pm:

$$\hat{Y} = b0_{12:00} + b1_{12:00} * temperatura + b2_{12:00} * humedad$$

$$\hat{Y} = 77.161 + 0.824 * 25.5 + 25.208 * 0.23$$

$$\hat{Y} = 103.97 \text{ kW}$$

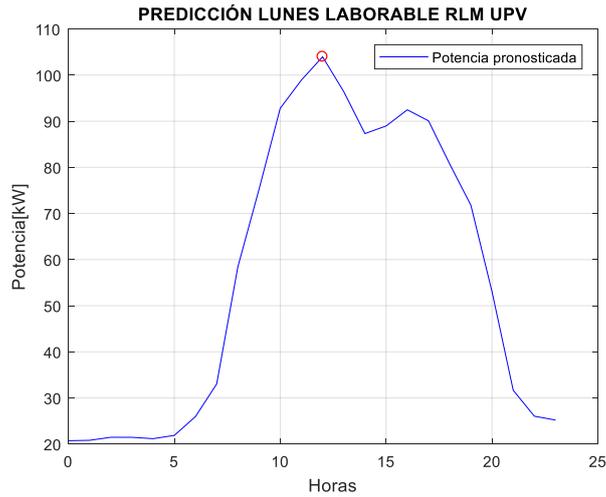


Figura 3.6 Valor pronosticado del consumo de potencia para la UPV a las 12:00 del día lunes 30 de mayo de 2016. Los datos de consumo real se usan para obtener los errores de pronóstico como son: MAE, RMSE y MAPE. Los errores de pronóstico para el día lunes 30 de mayo de 2016 se indican en la Tabla 3.5.

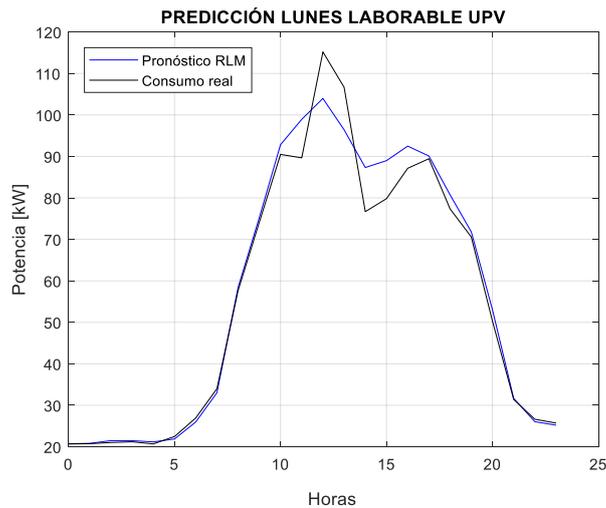


Figura 3.7 Pronóstico usando RLM del día lunes 30 de mayo de 2016 para la UPV.

Tabla 3.5 Errores del pronóstico de la Figura 3.7

Errores	Regresión
	Lineal Múltiple
MAE [kW]	3.05
RMSE [kW]	4.89
MAPE [%]	4.13

El mismo procedimiento se usa para el pronóstico del día lunes 4 de diciembre de 2017 en la UPS como se muestra en la Figura 3.8. Los errores de pronóstico obtenidos se indican en la Tabla 3.6.

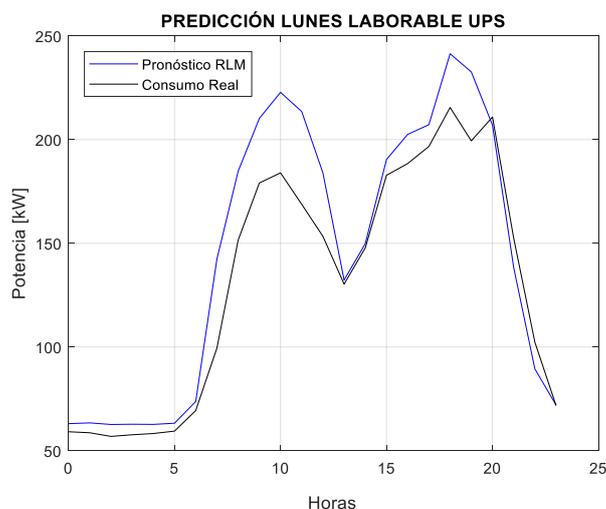


Figura 3.8 Pronóstico usando RLM del día lunes 4 de diciembre de 2017 para la UPS.

Tabla 3.6 Errores en el pronóstico de la Figura 3.8

Errores	Regresión
	Lineal Múltiple
MAE [kW]	11.52
RMSE[kW]	15.25
MAPE [%]	7.8

Los errores de pronóstico obtenidos con la aplicación del método de RLS y RLM son similares para la base de datos de la UPV. En el caso de la UPS estos errores son mayores.

### 3.3.MODELO ARIMA

El modelo ARIMA usado para el pronóstico del consumo de electricidad se desarrolla siguiendo la Metodología de Box Jenkins revisada en el Capítulo 1, con algunas características en particular que se indican a continuación. El primer paso que se realiza en la metodología de Box Jenkins es la segmentación y el tratamiento de la base de datos. En la segmentación de datos diarios existe mucha variabilidad, esto dificulta la elección de los parámetros del modelo, por ello se utilizan todos los datos disponibles.

Para aplicar el método ARIMA, se observa si el conjunto de datos presenta estacionariedad y estacionalidad, para seguir paso a paso el desarrollo de la predicción de consumo de energía eléctrica de la UPV se escoge como ejemplo el grupo 1 de la primavera. La estacionariedad puede

ser analizada de dos maneras. La primera forma es visual y se observa que la gráfica del conjunto de datos tenga una tendencia definida, como se observa en la Figura 3.9.

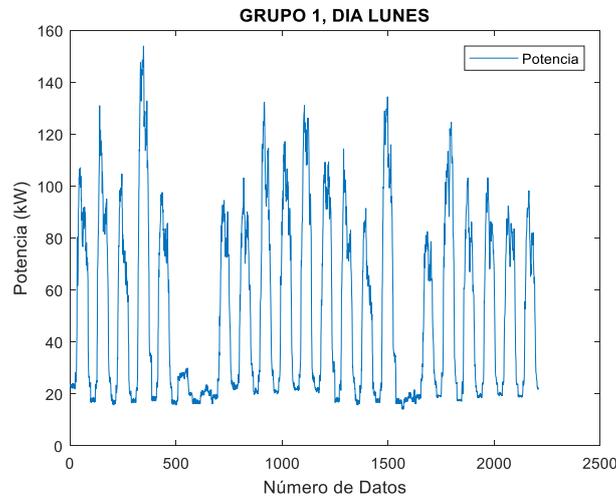


Figura 3.9 Análisis de estacionariedad del conjunto de datos del grupo 1 de la primavera

La segunda forma de determinar si un conjunto tiene estacionariedad es usar la prueba de Dickey-Fuller con el código en Matlab adfest revisado en el Capítulo 1.

El análisis de la estacionalidad se realiza de manera visual, en este caso observamos la existencia de patrones que se repiten cada 24 horas. En el caso de no existir estacionariedad se realiza una diferenciación para que su media y su varianza sean cero. Las autocorrelaciones se obtienen cuando el conjunto de datos tiene estacionalidad y estacionariedad. Para determinar las autocorrelaciones se utiliza el código de Matlab autocorr y parcorr que permite observar de manera gráfica el comportamiento de los datos, como se observa en la Figura 3.10.

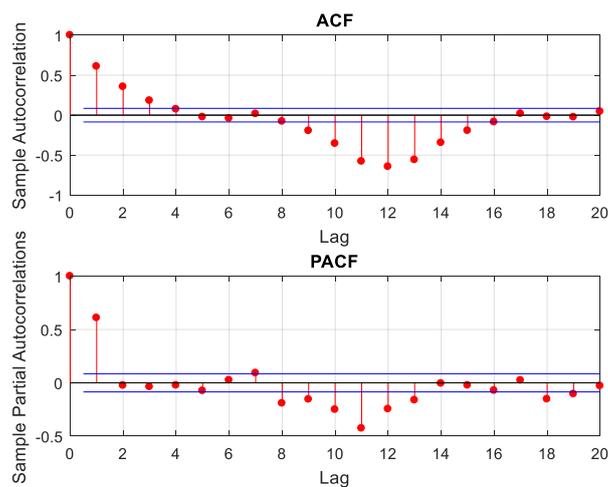


Figura 3.10 Autocorrelaciones parcial y simple correspondiente al grupo 1 del día lunes en la UPV.

Para el caso de la base de datos de la UPV se tiene un modelo ARIMA estacional y un modelo ARIMA estacionario. El modelo ARIMA estacionario es (1,1,1) y el modelo ARIMA estacional es (1,0,1) con estacionalidad 24 quedando expresado de la siguiente manera:

$$ARIMA_{UPV} = (1,1,1) \times (1,0,1)_{24}$$

Los valores de P, D y Q se obtienen al observar las gráficas de las autocorrelaciones y compararlas con los modelos de la Tabla 1.2 del Capítulo 1, como se indica en la Figura 3.11.

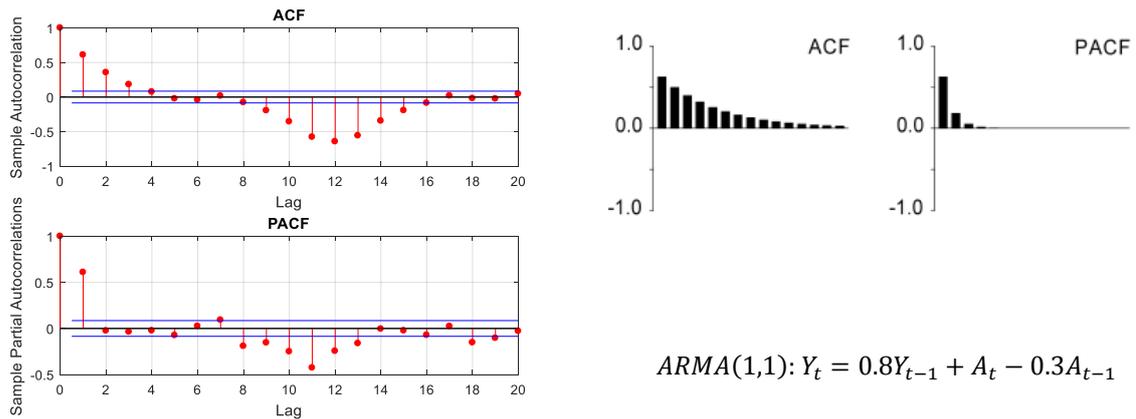


Figura 3.11 Autocorrelaciones simples y parciales.

En la Figura 3.11 los valores de P, D y Q son 0.8, 0 y -0.3 respectivamente. Cabe recalcar que los valores de P, D y Q no deben ser mayores a la unidad debido al modelo (1,0,1)<sub>24</sub>. Los valores de P, D y Q son diferentes para cada agrupación de acuerdo a las características de cada grupo de datos. Los valores se indican en las Tablas 3.7 y 3.8.

Tabla 3.7 Valores de P, D y Q para los grupos de la base de datos UPV, días laborables

Grupo Marzo, Abril y Mayo					
	Lunes	Martes	Miércoles	Jueves	Viernes
Valor P	0.8	0.8	0.8	0.8	0.9
Valor D	0	0	0	0	0
Valor Q	-0.3	-0.9	-0.9	-0.9	0.7
Grupo Junio y Julio					
	Lunes	Martes	Miércoles	Jueves	Viernes
Valor P	0.5	0.5	0.5	0.5	0.5
Valor D	0	0	0	0	0
Valor Q	-0.6	-0.6	-0.6	-0.7	-0.7

Grupo Septiembre, Octubre y Noviembre					
	Lunes	Martes	Miércoles	Jueves	Viernes
Valor P	0.9	0.9	0.9	0.9	0.1
Valor D	0	0	0	0	0
Valor Q	-0.8	-0.8	-0.8	-0.8	-0.6
Grupo Diciembre, Enero y Febrero					
	Lunes	Martes	Miércoles	Jueves	Viernes
Valor P	0,9	0,9	0,9	0,9	0,9
Valor D	0	0	0	0	0
Valor Q	-0,5	-0,5	-0,5	-0,5	-0,5

Tabla 3.8 Valores de P, D y Q para los grupos de la base de datos UPV, Grupo No Laborables

Valores de P, D y Q para los grupos de la base de datos UPV, Grupo No Laborables							
	Lunes	Martes	Miércoles	Jueves	Viernes	Sábado	Domingo
Valor P	0.8	0.8	0.8	0.8	0.8	0.8	0.8
Valor D	0	0	0	0	0	0	0
Valor Q	-0.3	-0.3	-0.3	-0.3	-0.3	-0.3	-0.3

El resultado de la predicción se muestra en la Figura 3.12. La curva azul corresponde al pronóstico mientras que la negra es la curva de consumo real. Los errores de pronóstico se indican en la Tabla 3.9.

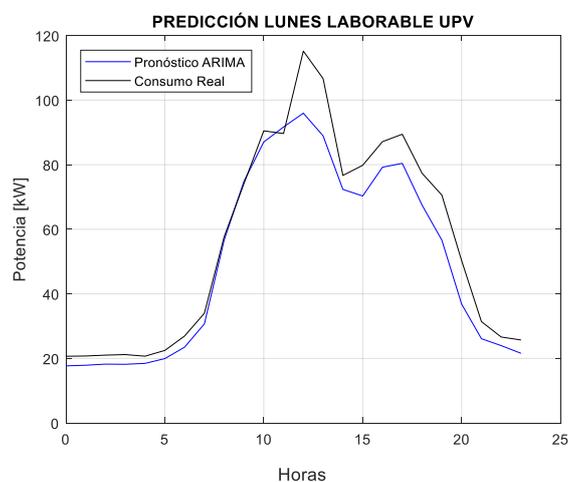


Figura 3.12 Pronóstico de consumo de energía eléctrica de la UPV correspondiente al día lunes 23 de mayo de 2016

Tabla 3.9 Errores de pronóstico de la Figura 3.12

Errores	Arima
MAE [kW]	6.14
RMSE [kW]	8.03
MAPE [%]	11.71

Para la base de datos de la UPS el procedimiento se realiza de la misma manera, la diferencia es el análisis de las autocorrelaciones simples y parciales para establecer los parámetros P, D y Q.

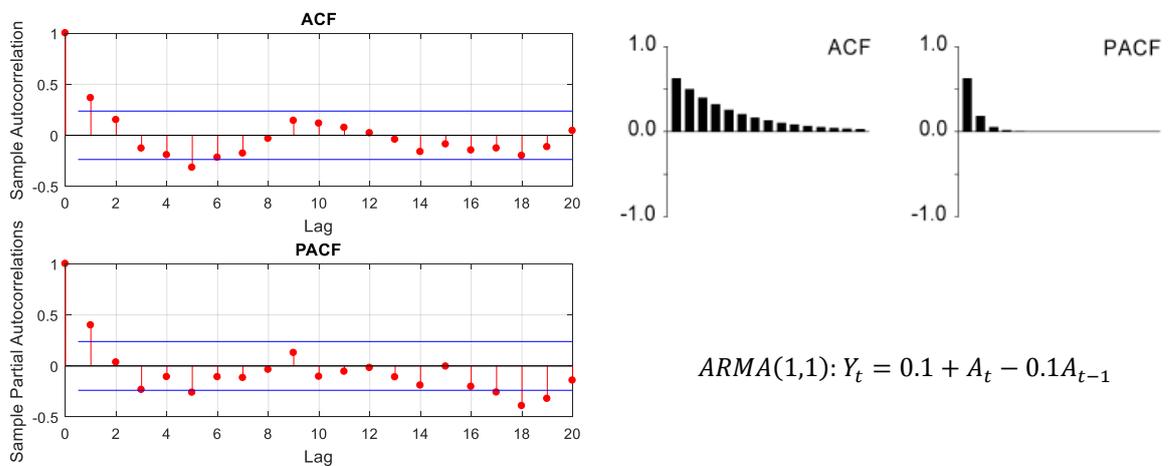


Figura 3.13 Autocorrelaciones Simples y Parciales

En base a las diferentes agrupaciones de perfiles de carga de la UPS mostradas en la Tabla 2.6, en la Tabla 3.10 se muestran los parámetros P, D y Q.

Tabla 3.10 Valores de P, D y Q para los datos de los perfiles de carga de la UPS, días laborables Grupo Días Laborables

Valores de P, D y Q para los datos de los perfiles de carga de la UPS, días laborables Grupo Días Laborables						
	Lunes	Martes	Miércoles	Jueves	Viernes	Sábado
Valor P	0.7	0.7	0.7	0.7	0.7	0.1
Valor D	0	0	0	0	0	0
Valor Q	-0.2	-0.2	-0.2	-0.2	-0.2	-0.4

Tabla 3.11 Valores de P, D y Q para los datos de los perfiles de carga de la UPS, días no laborables Grupo Días No Laborables

Valores de P, D y Q para los datos de los perfiles de carga de la UPS, días no laborables Grupo Días No Laborables					
	Lunes	Jueves	Viernes	Sábado	Domingo
Valor P	0.1	0.1	0.1	0.1	0.1
Valor D	0	0	0	0	0
Valor Q	-0.5	-0.5	-0.5	-0.5	-0.5

El resultado del pronóstico se muestra en la Figura 3.14. Los errores de pronóstico están en la Tabla 3.12.

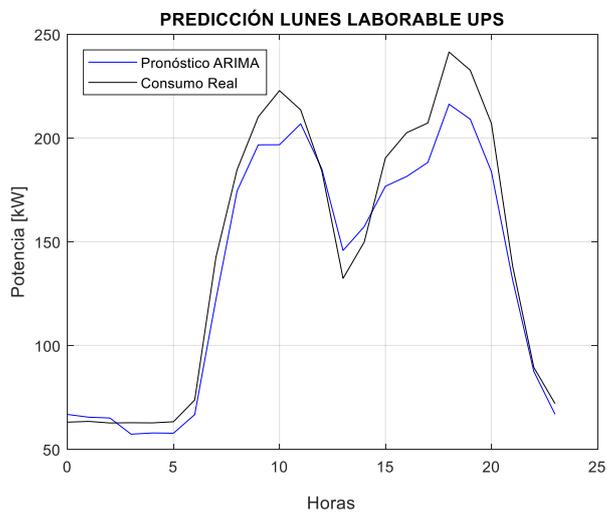


Figura 3.14 Pronóstico de consumo de energía eléctrica de la UPS del día lunes 4 de diciembre de 2017

Tabla 3.12 Errores de pronóstico con el método ARIMA en la UPS

Errores	Arima
MAE [kW]	11.23
RMSE [kW]	13.87
MAPE [%]	7.38

El análisis de los errores para cada grupo de datos se muestra en el Capítulo 4.

## 4. CAPITULO 4. METODOLOGÍA COMPARATIVA DE LOS MÉTODOS ESTADÍSTICOS

En este capítulo se realiza una comparación entre los métodos estadísticos, para determinar el más adecuado para el pronóstico del consumo de electricidad a corto plazo mediante la evaluación de sus errores como son MAE, RMSE y MAPE. Para la UPV se usan los datos de la primavera para pronosticar cuatro semanas, para el caso de la UPS se pronostica 4 semanas correspondientes a días laborables.

### 4.1. BASE DE DATOS DE LA UNIVERSIDAD POLITECNICA DE VALENCIA

Para realizar el pronóstico de una semana se utilizan todos los datos registrados del consumo de electricidad exceptuando los datos de la semana que se quiere pronosticar. Los datos de consumo real que no son incluidos para el pronóstico se usan para comparar con los resultados obtenidos.

Desde la Figura 4.1 hasta la Figura 4.4 se muestra el pronóstico de 4 semanas de la primavera obtenidos con la RLS, temperatura y la humedad como variables independientes, la RLM y el método ARIMA. También se muestran los perfiles de consumo real. En la Figura 4.1 se puede observar el pronóstico de días laborables desde el miércoles 4 al martes 10 de mayo de 2016. La secuencia de los días comienza el lunes y termina el viernes.

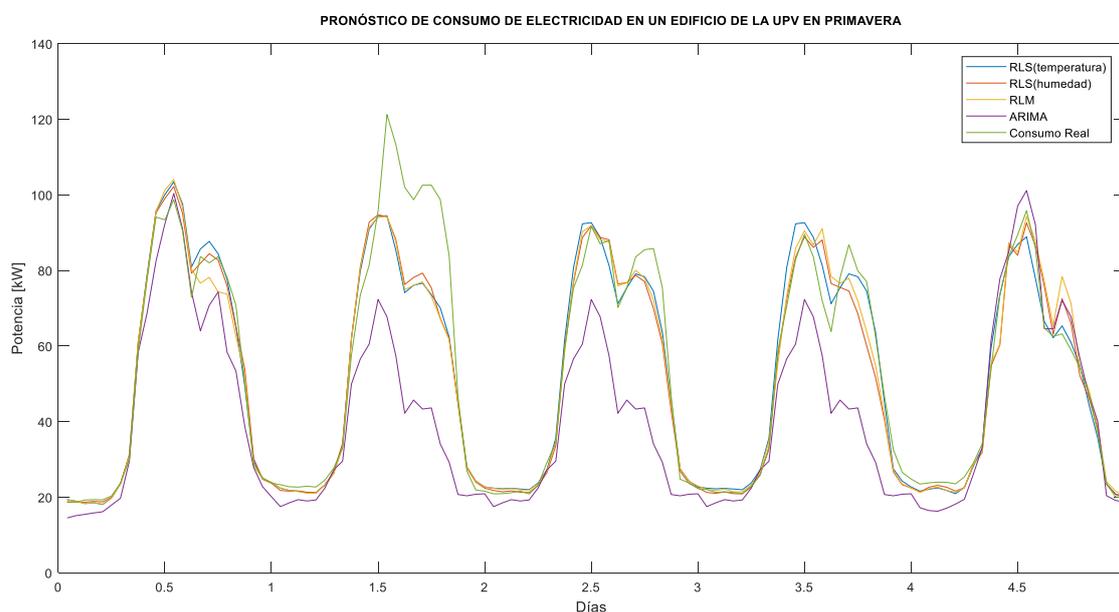


Figura 4.1 Pronóstico de consumo de electricidad para días laborables desde el miércoles 4 de mayo al martes 10 de mayo de 2016.

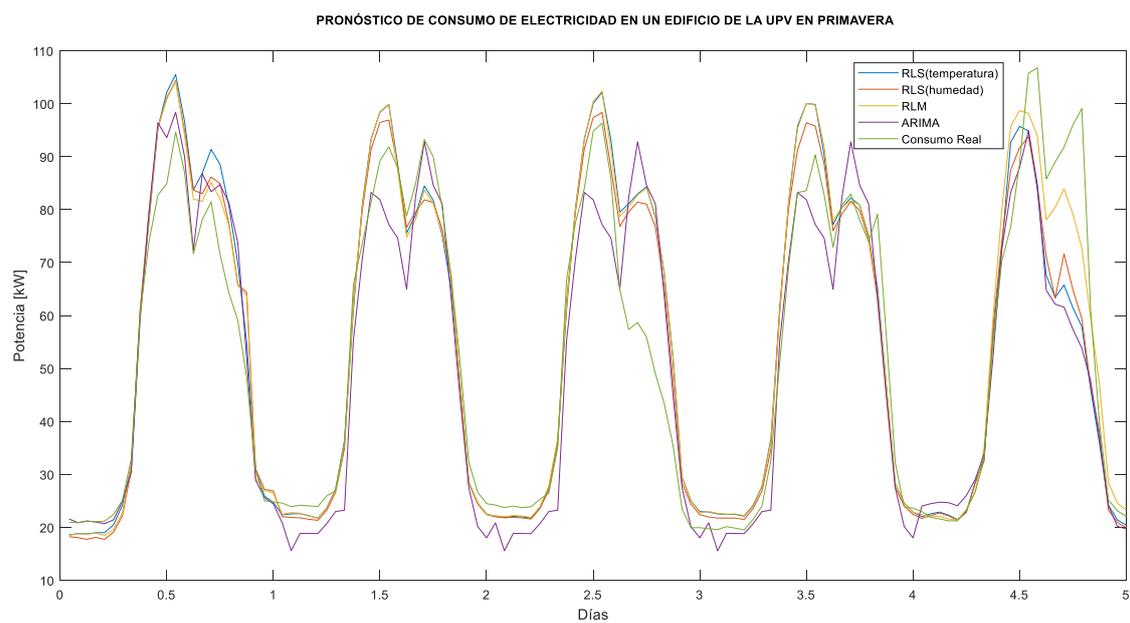


Figura 4.2 Pronóstico de consumo de electricidad para días laborables desde el miércoles 11 de mayo al martes 17 de mayo de 2016.

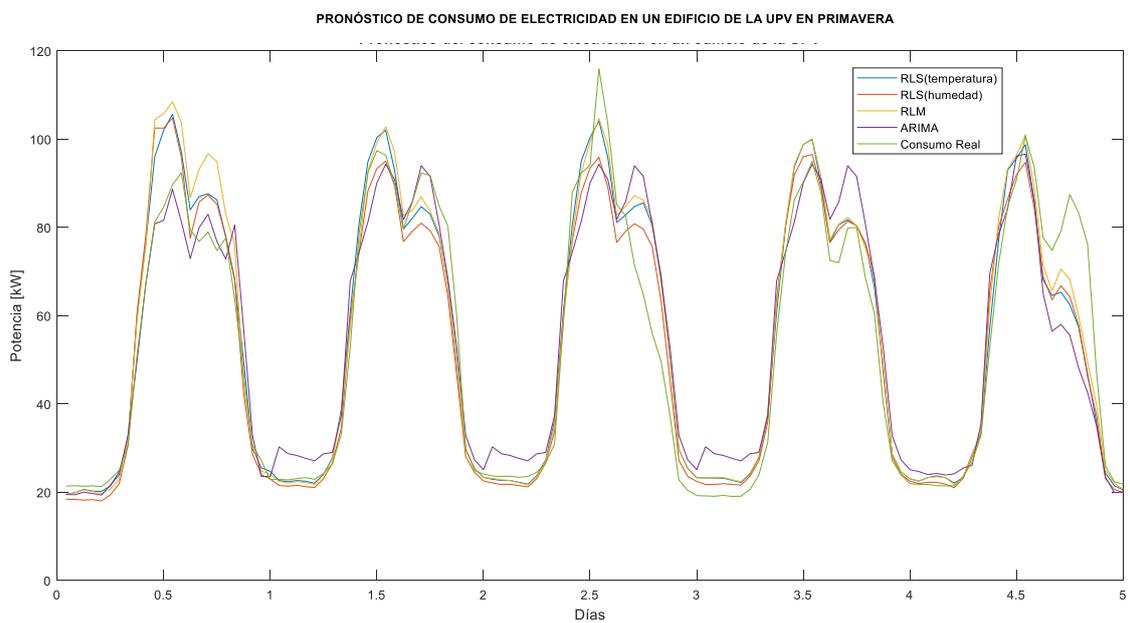


Figura 4.3 Pronóstico de consumo de electricidad para días laborables desde el miércoles 18 de mayo al martes 24 de mayo de 2016.

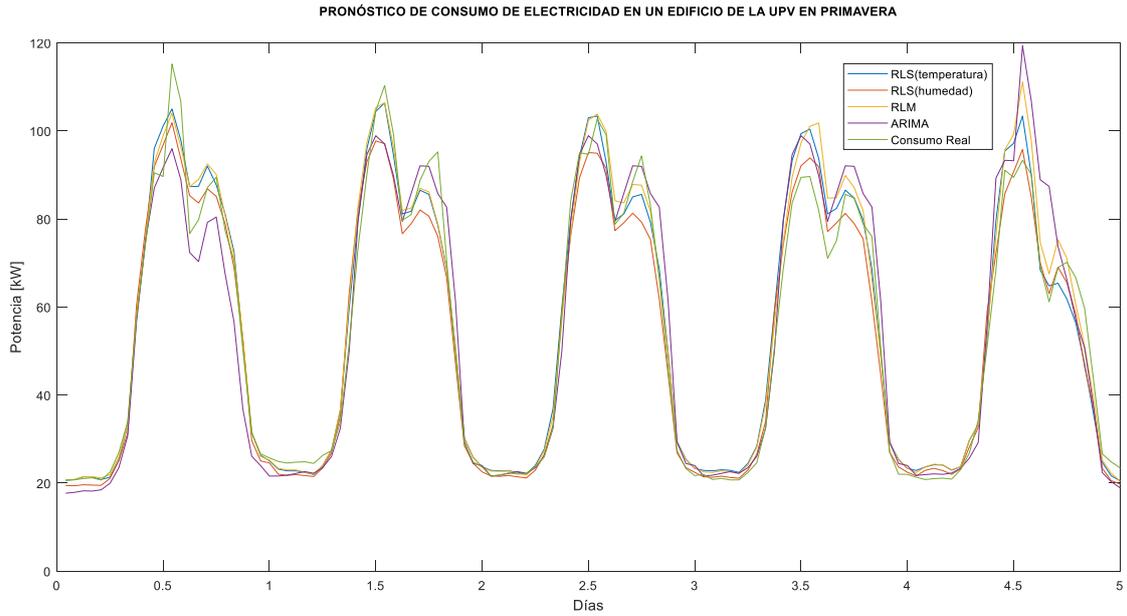


Figura 4.4 Pronóstico de consumo de electricidad para días laborables desde el miércoles 25 de mayo al martes 31 de mayo de 2016.

Los datos de pronóstico del consumo de electricidad se adjuntan en el Anexo 2. En la Tabla 4.1 se muestra el promedio de los errores MAE, RMSE y MAPE de las cuatro semanas correspondientes al pronóstico de la UPV.

Tabla 4.1 Errores de pronóstico promedio para días laborables de la UPV en la primavera

Método	Error	Lunes	Martes	Miércoles	Jueves	Viernes
RLS (var. temp.)	MAE [kW]	4.17	6.04	3.90	4.14	3.63
	RMSE [kW]	5.81	8.67	5.49	5.59	5.19
	MAPE [%]	6.63	8.88	6.84	7.46	7.13
RLS (var. hum.)	MAE [kW]	4.08	6.15	3.15	4.10	3.87
	RMSE [kW]	5.55	8.83	4.80	5.80	5.65
	MAPE [%]	7.04	8.61	5.36	7.17	7.44

RLM	MAE [kW]	4.80	6.12	3.76	4.51	5.19
	RMSE [kW]	6.79	8.72	5.36	6.28	7.46
	MAPE [%]	7.54	8.88	6.60	7.80	10.32
ARIMA	MAE [kW]	5.26	14.25	12.39	11.89	5.28
	RMSE [kW]	7.25	16.42	14.67	13.88	7.55
	MAPE [%]	10.13	36.82	36.58	399.82	9.98

Se realiza un promedio los errores de la Tabla 4.1 para cada método. La RLS con la humedad como variable independiente tiene los errores de pronóstico mas bajos, 7.12% de MAPE, 6.12 kW de RMSE y 4.27 kW para MAE. Se sigue el mismo procedimiento para realizar el pronóstico para el verano, otoño e invierno. En la Figura 4.5 se muestra el pronóstico de los días laborables del verano. En la Tabla 4.2 se muestra los errores de pronóstico.

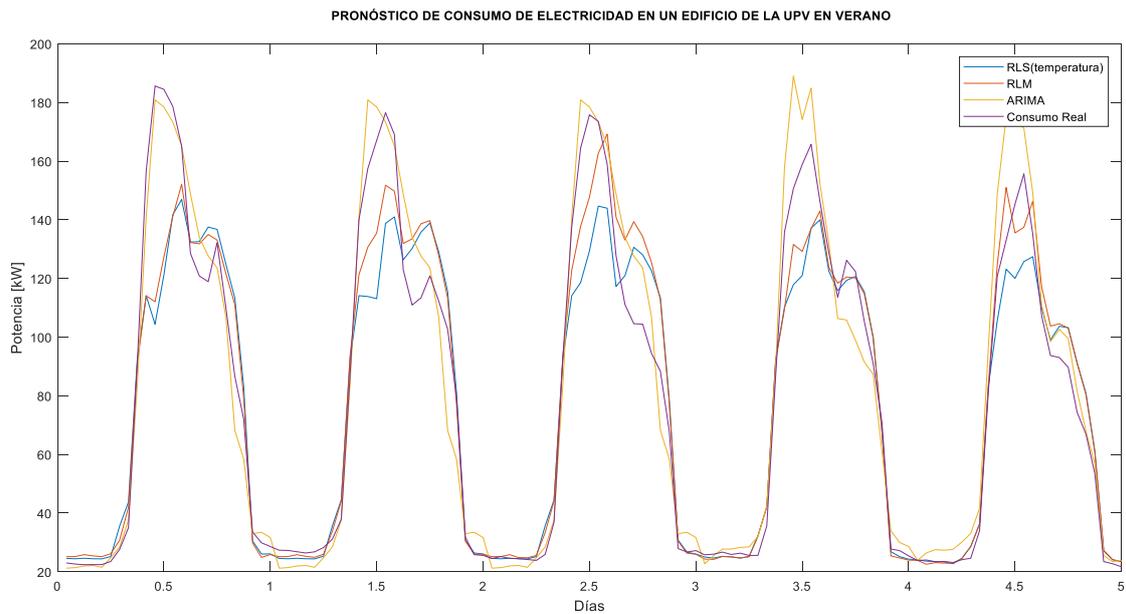


Figura 4.5 Pronóstico de consumo de electricidad para días laborables desde el lunes 25 de julio de 2016 al viernes 29 de julio de 2016.

Tabla 4.2 Errores de pronóstico para días laborables del 25 al 29 de julio de 2016 de la UPV en el verano

Método	Error	Lunes	Martes	Miércoles	Jueves	Viernes
RLS (var. temp.)	MAE [kW]	15.51	13.22	13.23	8.33	7.29
	RMSE [kW]	25.71	19.88	19.37	13.60	10.86
	MAPE [%]	15.16	12.59	13.30	8.94	9.04
RLS (var. hum.)	MAE [kW]	6.64	5.74	7.74	4.70	6.86
	RMSE [kW]	9.44	8.15	11.77	5.91	10.54
	MAPE [%]	8.77	7.18	9.03	7.43	8.60
RLM	MAE [kW]	14.04	10.54	11.65	7.40	6.76
	RMSE [kW]	23.62	14.57	16.41	11.46	9.07
	MAPE [%]	14.28	10.20	12.73	8.73	9.01
ARIMA	MAE [kW]	5.88	9.20	8.37	9.27	9.37
	RMSE [kW]	8.39	12.98	11.20	13.07	13.80
	MAPE [%]	7.10	13.69	12.07	11.62	13.66

Los valores de la Tabla 4.2 indican que el método de RLS con la variable humedad tiene los errores mas bajos para el pronóstico de consumo de electricidad en el verano. Su promedio de error MAPE es 8.20%, RMSE tiene 9.16 kW y MAE con 6.33 kW. En la Figura 4.6 se muestra

el pronóstico para los días laborables del otoño. En la Tabla 4.3 se muestra los errores de pronóstico.

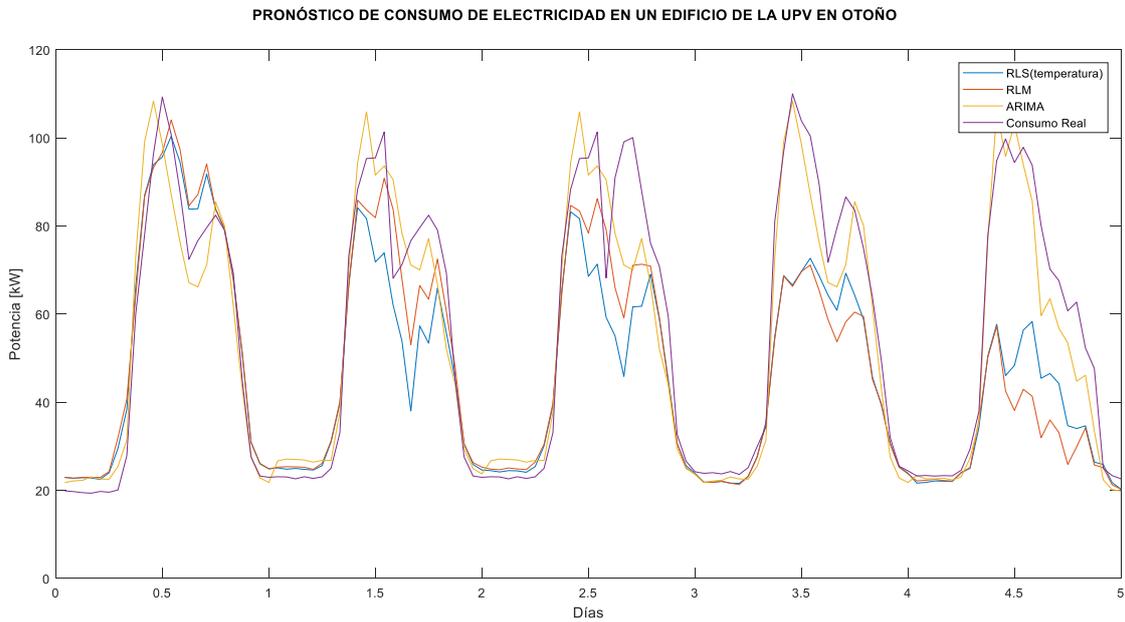


Figura 4.6 Pronóstico de consumo de electricidad para días laborables desde el martes 22 de noviembre de 2016 al lunes 28 de noviembre de 2016.

Tabla 4.3 Errores de pronóstico para días laborables del 22 al 28 de noviembre de 2016 de la UPV en el otoño

Método	Error	Lunes	Martes	Miércoles	Jueves	Viernes
RLS (var. temp.)	MAE [kW]	5.27	10.18	21.86	12.72	18.19
	RMSE [kW]	6.51	14.48	25.72	17.71	24.77
	MAPE [%]	13.46	17.85	33.58	16.72	25.52
RLS (var. hum.)	MAE [kW]	11.87	21.45	25.22	9.44	8.36
	RMSE [kW]	14.33	27.82	31.75	13.44	12.25
	MAPE [%]	23.34	38.02	42.08	13.11	12.66
RLM	MAE	5.92	9.85	21.05	14.11	21.97

	[kW]					
	RMSE [kW]	7.30	13.10	23.99	19.28	30.41
	MAPE [%]	14.94	18.36	33.49	18.45	29.96
ARIMA	MAE [kW]	6.10	11.35	23.09	5.01	5.76
	RMSE [kW]	8.04	14.81	26.15	6.61	8.00
	MAPE [%]	11.66	21.70	37.65	8.96	10.06

La recta de regresión no se ajusta adecuadamente a los datos de consumo para los días laborables del otoño por la dispersión de datos, tal como se observa en la Figura 2.13. Debido a ello, no se tiene un adecuado pronóstico con los métodos de regresión. Los errores mas bajos se obtienen con el método ARIMA con 18% de MAPE, 12.17 kW de RMSE y 10.02 kW de MAE. En la Figura 4.7 se muestra el pronóstico para los días laborables del invierno. En la Tabla 4.4 se muestran los errores de pronóstico.

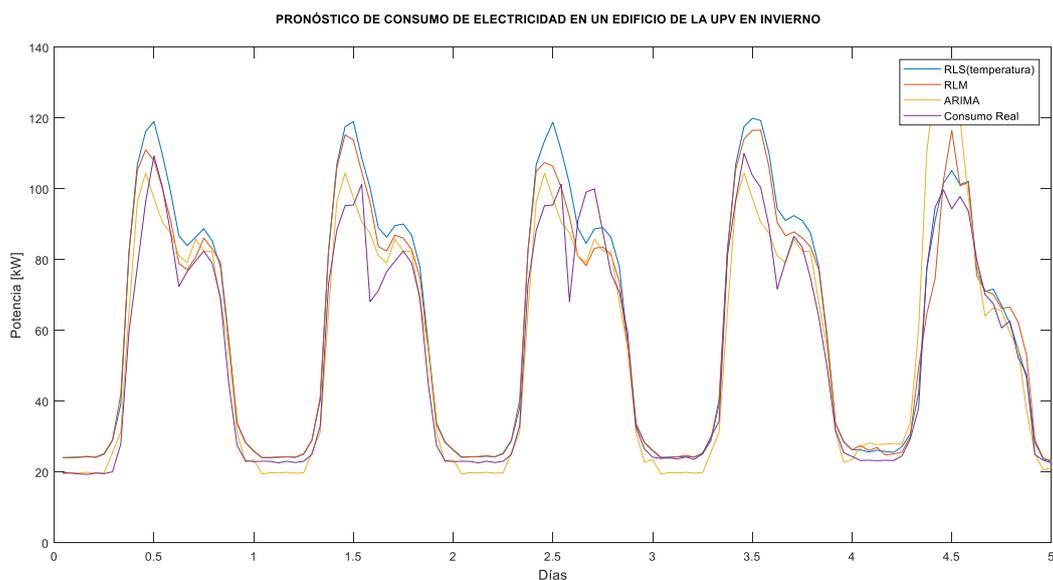


Figura 4.7 Pronóstico de consumo de electricidad para días laborables desde el martes 23 de febrero de 2016 al lunes 29 de febrero de 2016.

Tabla 4.4 Errores de pronóstico para días laborables del 23 al 29 de febrero de 2016 de la UPV en el invierno

Método	Error	Lunes	Martes	Miércoles	Jueves	Viernes
RLS (var. temp.)	MAE [kW]	7.22	5.40	9.61	16.34	7.06
	RMSE [kW]	9.58	7.94	12.89	21.43	9.72
	MAPE [%]	12.46	9.89	13.86	27.47	12.51
RLS (var. hum.)	MAE [kW]	6.47	5.02	6.72	14.62	6.85
	RMSE [kW]	8.14	7.29	8.98	19.15	10.38
	MAPE [%]	11.62	9.13	10.07	24.77	11.97
RLM	MAE [kW]	6.08	5.21	6.30	14.68	6.39
	RMSE [kW]	7.73	7.50	8.38	19.17	9.63
	MAPE [%]	11.12	9.16	9.47	24.90	11.34
ARIMA	MAE [kW]	7.09	9.78	6.91	10.50	15.86
	RMSE [kW]	10.41	12.64	8.02	13.47	22.32
	MAPE [%]	15.10	17.29	14.35	21.92	27.41

Para el invierno, los métodos de regresión obtienen los mejores resultados. Los errores de pronóstico para RLS y RLM son similares. El MAPE promedio es de 13.2%, RMSE con 10.5 kW y MAE con 7.8 kW.

El pronóstico para días no laborables se observa en la Figura 4.8. Los errores de pronóstico se muestra en la Tabla 4.5.

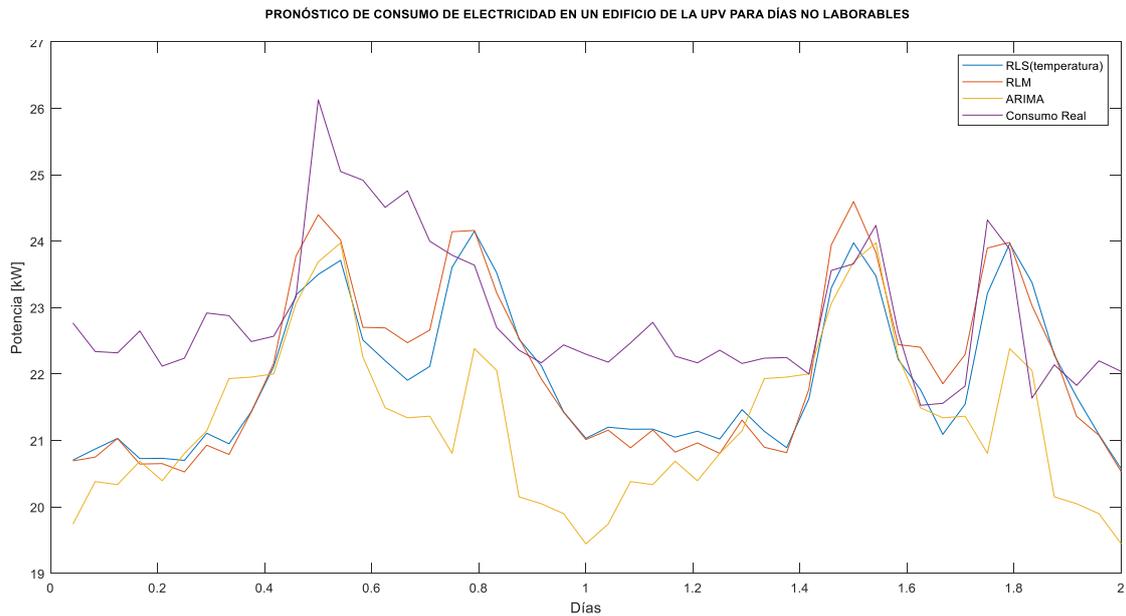


Figura 4.8 Pronóstico de consumo de electricidad para el día sábado 26 de noviembre y domingo 27 de noviembre de 2016.

Tabla 4.5. Errores de pronóstico para días no laborables sábado 26 y domingo 27 de noviembre de 2016.

Método	Error	Sábado	Domingo
RLS (var. temp.)	MAE [kW]	1.34	0.81
	RMSE [kW]	1.57	0.96
	MAPE [%]	5.74	3.63
RLS (var. hum.)	MAE [kW]	0.67	1.08
	RMSE [kW]	0.78	1.38
	MAPE [%]	2.90	4.85
RLM	MAE	1.28	0.87

	[kW]		
	RMSE		
	[kW]	1.44	1.01
	MAPE		
	[%]	5.50	3.92
ARIMA	MAE		
	[kW]	1.91	1.22
	RMSE		
	[kW]	2.11	1.58
	MAPE		
	[%]	8.19	5.43

De acuerdo a los datos obtenidos en la Tabla 4.5 el método de RLS con la variable humedad tiene el error mas bajo, 3.87% de MAPE.

#### 4.2.BASE DE DATOS DE LA UNIVERSIDAD POLITÉCNICA SALESIANA

Para la UPS se utilizan los datos de consumo de electricidad de los días laborables para realizar el pronóstico de 4 semanas. Desde la Figura 4.9 hasta la Figura 4.12 se muestran los resultados obtenidos para cada método. Los errores de pronóstico de la Tabla 4.6 son un promedio de las cuatro semanas pronosticadas.

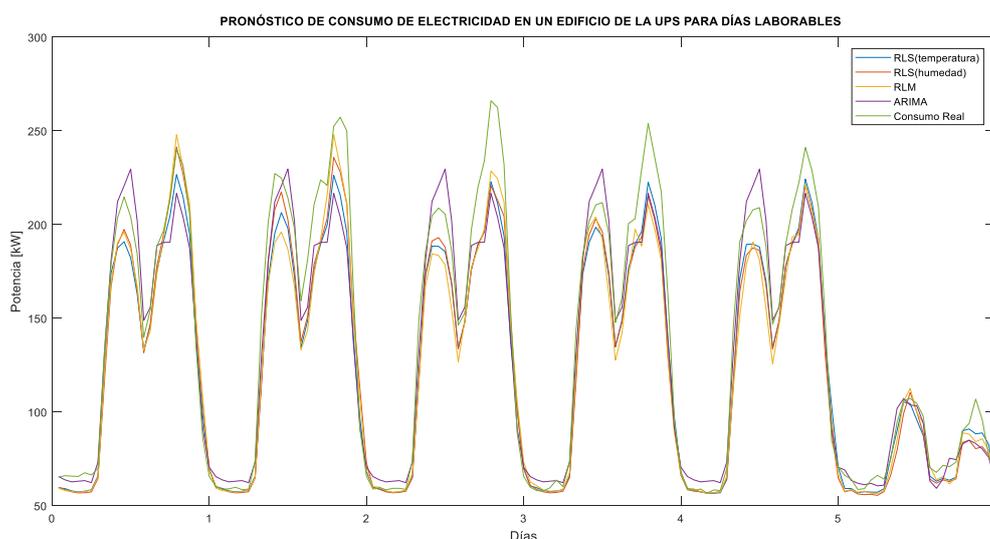


Figura 4.9. Pronóstico de consumo de electricidad para días laborables desde el viernes 10 de diciembre de 2017 al jueves 16 de diciembre de 2017 en la UPS.

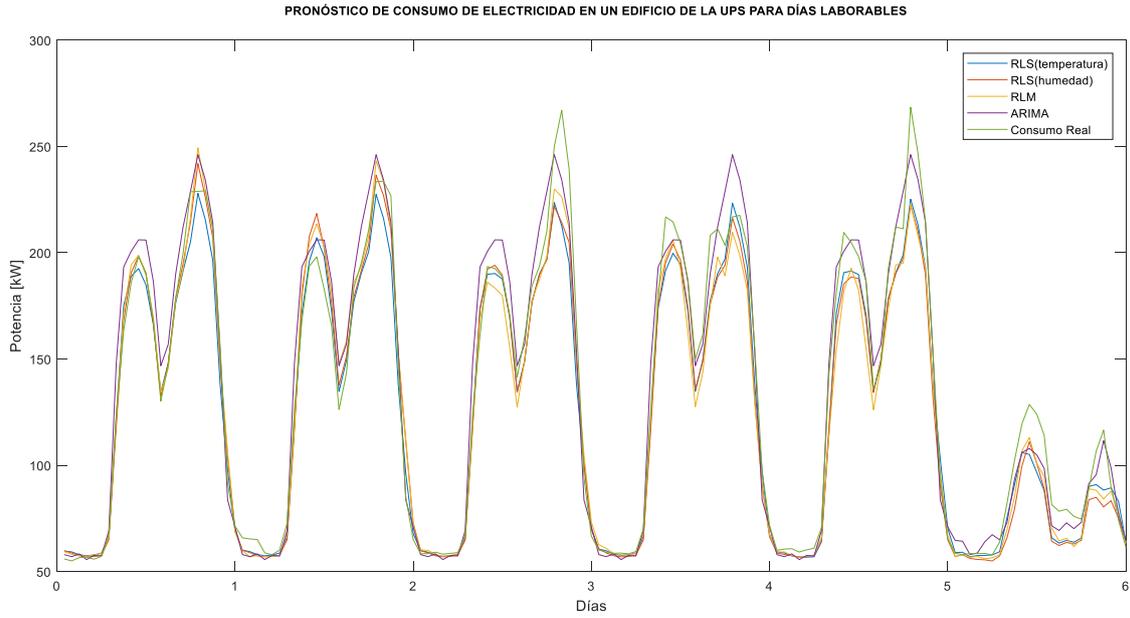


Figura 4.10 Pronóstico de consumo de electricidad para días laborables desde el viernes 17 de diciembre de 2017 al jueves 23 de diciembre de 2017 en la UPS.

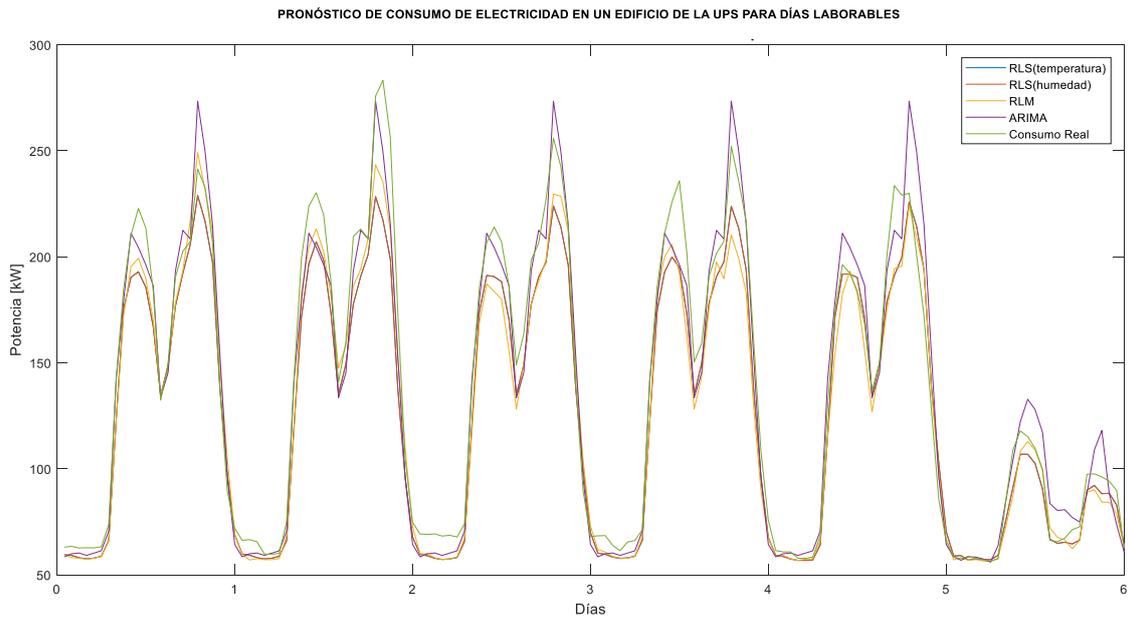


Figura 4.11 Pronóstico de consumo de electricidad para días laborables desde el viernes 24 de diciembre de 2017 al jueves 30 de diciembre de 2017 en la UPS.

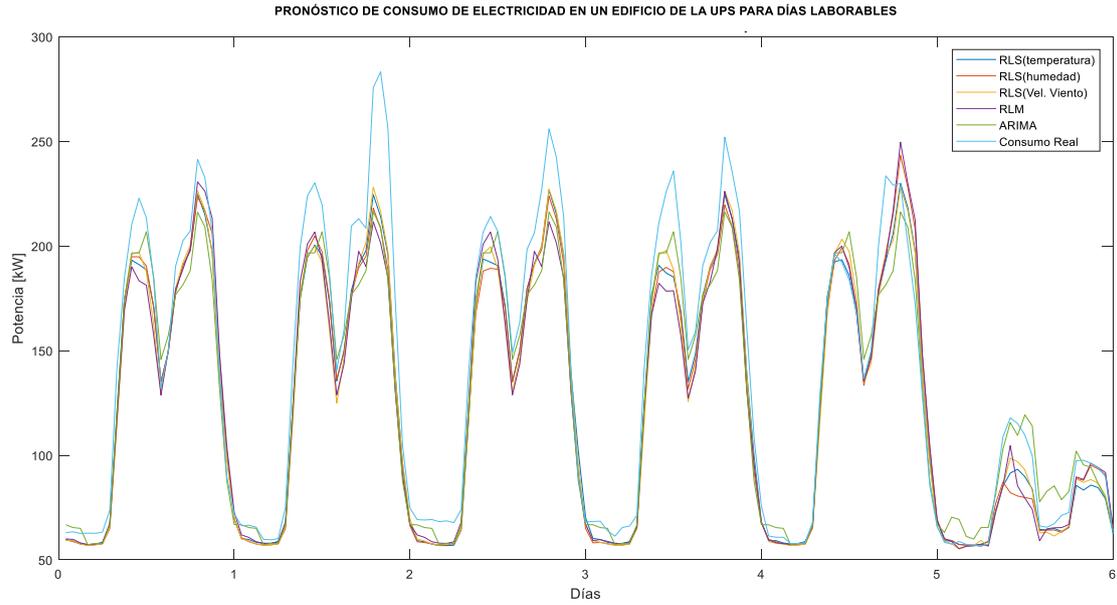


Figura 4.12 Pronóstico de consumo de electricidad para días laborables desde el viernes 1 de diciembre de 2017 al jueves 7 de diciembre de 2017 en la UPS.

Tabla 4.6 Errores de pronóstico promedio para días laborables en la UPS..

Método	Error	Lunes	Martes	Miércoles	Jueves	Viernes	Sábado
RLS (var. temp.)	MAE [kW]	9.46	14.62	13.55	13.16	9.42	8.32
	RMSE [kW]	11.56	19.23	17.99	17.26	12.80	10.77
	MAPE [%]	6.72	8.29	8.22	7.72	6.14	9.22
RLS (var. hum.)	MAE [kW]	8.17	13.90	13.18	14.56	10.45	9.93
	RMSE [kW]	9.94	18.14	17.45	19.01	13.97	12.92
	MAPE [%]	6.41	8.72	8.04	8.82	6.51	11.06
RLS (var. vel. viento)	MAE [kW]	10.51	21.18	21.18	17.02	9.38	7.73

	RMSE [kW]	13.15	27.41	27.41	20.01	13.64	10.21
	MAPE [%]	7.32	11.87	11.87	11.38	5.93	8.36
RLM	MAE [kW]	8.53	14.41	14.30	15.83	11.97	8.09
	RMSE [kW]	10.72	19.15	18.14	20.93	16.35	10.77
	MAPE [%]	6.53	8.82	8.82	9.30	7.36	9.15
ARIMA	MAE [kW]	9.60	13.99	11.38	10.42	22.54	18.41
	RMSE [kW]	12.75	19.54	15.85	13.99	30.21	23.37
	MAPE [%]	6.27	8.02	6.91	6.58	21.49	14.08

Los errores mas bajos se muestran con los métodos de regresión. Para la RLS y RLM los errores de pronóstico con similares. Un promedio de 8% en MAPE, 10.5 kW para RMSE y 9 kW para MAE. El resultado del pronóstico con cada método estadístico de los días no laborables se muestra en la Figura 4.13 y los errores en la Tabla 4.7.

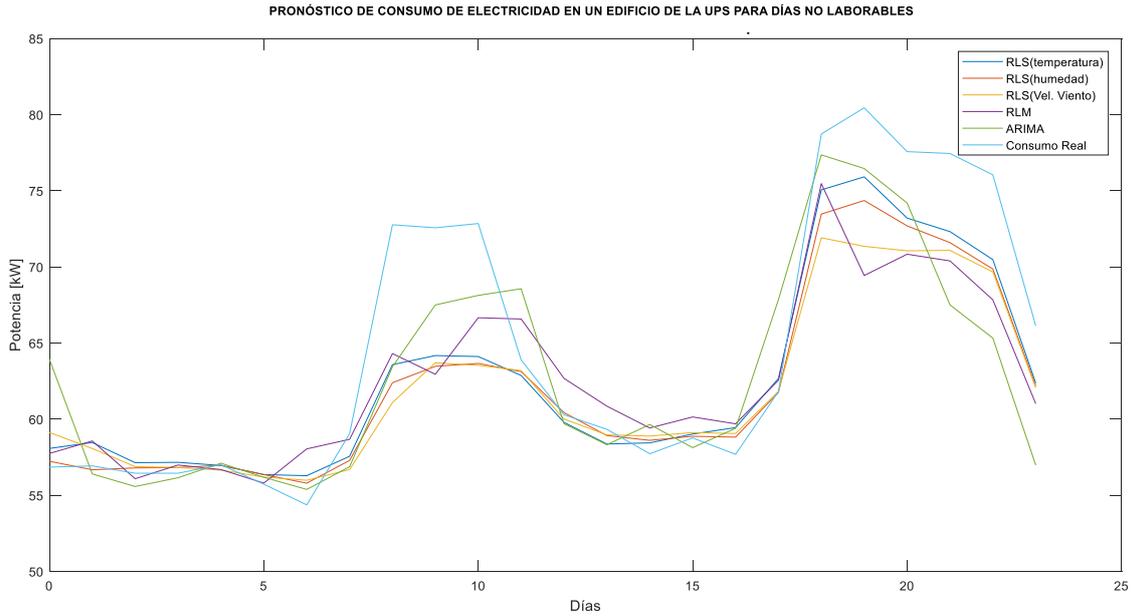


Figura 4.13 Pronóstico de consumo de electricidad para el día domingo 3 de diciembre de 2016.

Tabla 4.7 Errores de pronóstico para el domingo no laborable 3 de diciembre de 2016.

Método	Error	Domingo
RLS (var. temp.)	MAE [kW]	2.81
	RMSE [kW]	3.94
	MAPE [%]	4.02
RLS (var. hum.)	MAE [kW]	2.89
	RMSE [kW]	4.38
	MAPE [%]	4.02
RLS (var. vel. viento)	MAE [kW]	3.30

	RMSE [kW]	4.78
	MAPE [%]	4.62
RLM	MAE [kW]	2.87
	RMSE [kW]	3.99
	MAPE [%]	4.07
ARIMA	MAE [kW]	11.84
	RMSE [kW]	14.85
	MAPE [%]	17.79

A continuación se realiza una comparación con los resultados de error promedio de cada método para los días laborables de la UPS y UPV. Los resultados se muestran en la Tabla 4.8, en donde se determina que el método que tiene los mejores resultados en el pronóstico de los días laborables es la RLS con la variable temperatura y humedad. Estas variables explican de mejor manera la variación de la potencia, tal como se muestra la Tabla 2.5 y Tabla 2.14.

Tabla 4.8 Errores promedio del pronóstico de días laborables de la UPS y UPV

Días laborables			
Método	Error		
	MAE [kW]	RMSE [kW]	MAPE [%]
RLS (var. temperatura)	10.54	14.24	10.84
RLS (var. humedad)	10.08	13.37	10.97
RLM	10.75	14.41	11.10
ARIMA	12.01	15.88	23.72

Para los días no laborables que tienen poca dispersión de datos los métodos de regresión tienen buenos resultados como es el caso de la RLS y RLM, como se muestra en la Tabla 4.9.

Tabla 4.9 Errores promedio del pronóstico de días no laborables

Días no laborables			
Método	Error		
	MAE [kW]	RMSE [kW]	MAPE [%]
RLS (var. temperatura)	1.94	2.60	3.83
RLS (var. humedad)	1.88	2.88	3.95
RLM	1.97	2.61	4.00
ARIMA	6.70	8.35	11.61

### 4.3. SENSIBILIDAD DE LOS ERRORES EN FUNCIÓN DEL NÚMERO DE DÍAS UTILIZADOS

A continuación, se realiza el análisis de sensibilidad de los errores en función de la cantidad de datos. Esto se utiliza para determinar el rango más favorable. El análisis se realiza para cada caso de estudio en días laborables. En la UPV se utilizan los datos del grupo 1 correspondiente a la primavera de la Tabla 2.15, en la UPS también se usan los datos del grupo 1 de la Tabla 2.6.

#### 4.3.1. Base de datos UPS

En la Figura 4.14 se muestra la variación de los errores cuando se realiza el pronóstico del día lunes 4 de diciembre de 2017 en la UPS. El día lunes forma parte de la agrupación de los días lunes, martes, miércoles, jueves y viernes, que suman en total 185 días. Según la Figura 4.14, para realizar el pronóstico usando RLS con la temperatura como variable independiente, se puede usar datos de consumo de entre 40 y 100 días para obtener un buen pronóstico en función de su error.

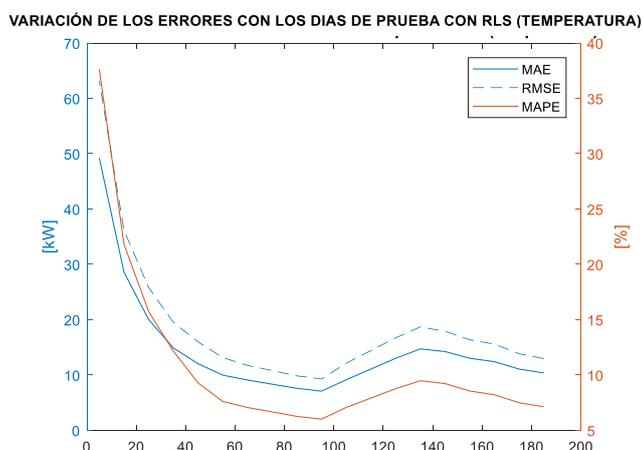


Figura 4.14 Variación de los errores de pronóstico en función de la cantidad de días usados para la UPS con RLS (temperatura).

En la Figura 4.15 se muestra la variación del error usando el método de RLM. Los datos que corresponden de entre 40 y 100 días de consumo son adecuados para tener un buen pronóstico.

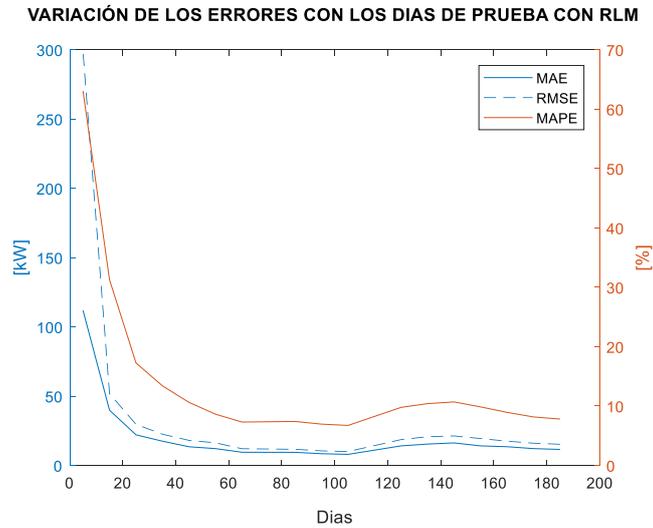


Figura 4.15 Variación de los errores de pronóstico en función de la cantidad de días usados para la UPS con RLM.

La Figura 4.16 muestra un comportamiento del error muy variable en función de los días utilizados para el pronóstico. Cuando se utilizan de 40 a 100 días en la obtención del modelo se obtienen los errores más bajos.

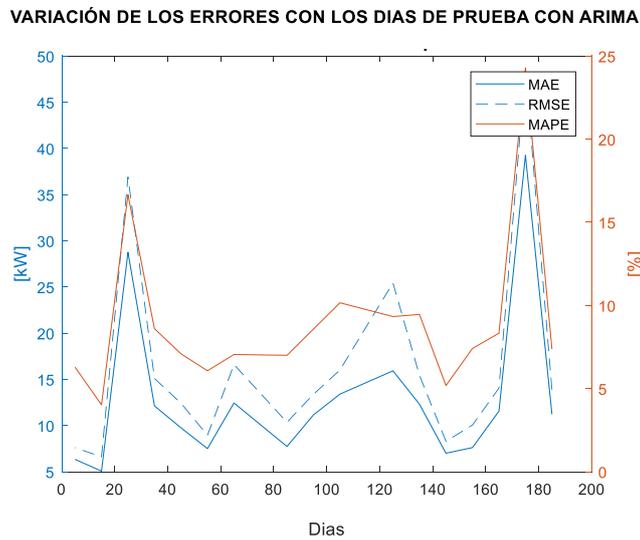


Figura 4.16 Variación de los errores de pronóstico en función de la cantidad de días usados para la UPS con ARIMA.

### 4.3.2. Base de datos UPV

Dentro del grupo 1 de la primavera están los días lunes que suman en total 60 días. Se analiza la variación de los errores para determinar un rango adecuado de días para el pronóstico.

La Figura 4.17 muestra la RLS con la temperatura como variable independiente, se nota que los errores de pronóstico se reducen cuando se utilizan 40 días para la obtención del modelo.

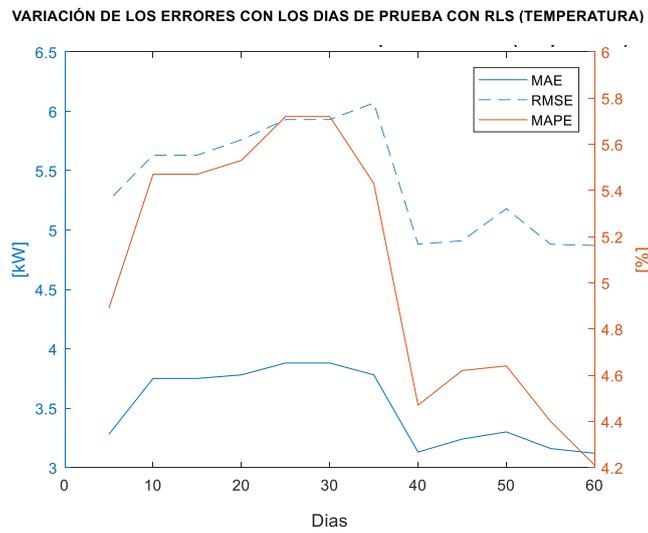


Figura 4.17 Variación de los errores de pronóstico en función de la cantidad de días usados para la UPV con RLS (temperatura).

Con la RLM, la variación de los errores indica que, el modelo con 30 días o más son adecuados para el pronóstico, tal como se observa en la Figura 4.18.

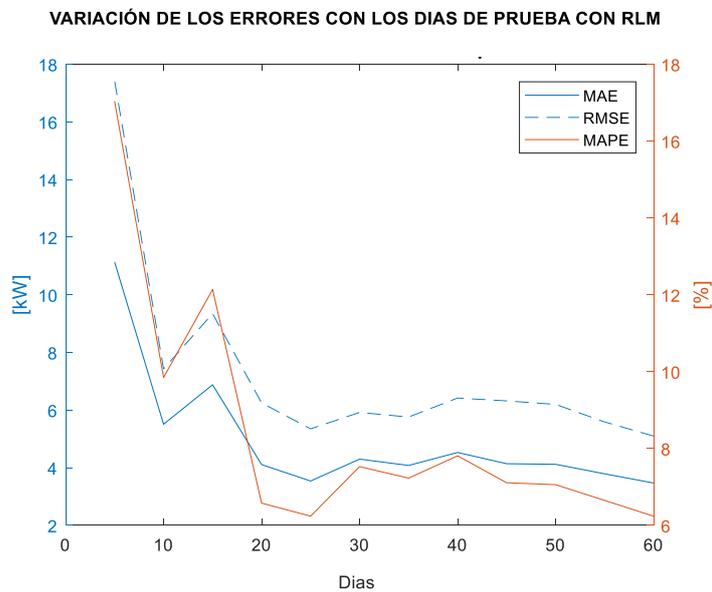


Figura 4.18 Variación de los errores de pronóstico en función de la cantidad de días usados para la UPV con RLM.

El modelo ARIMA, conformado con datos de 10 a 30 días presenta los menores errores de pronóstico.

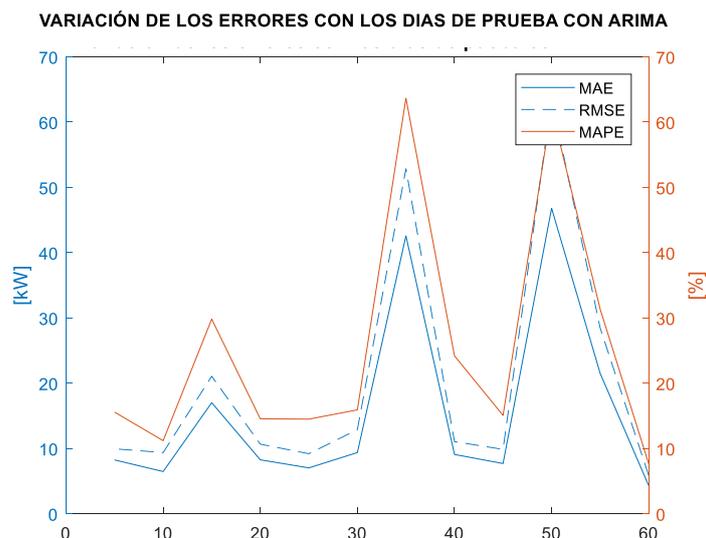


Figura 4.19 Variación de los errores de pronóstico en función de la cantidad de días usados para la UPV con ARIMA.

En la Tabla 4.10 se muestra un resumen del rango de días que deben ser usados para la obtención de cada modelo de pronóstico. En el caso de la UPS, los métodos de RLS, RLM y ARIMA conformados con datos entre 40 y 100 días tienen los mejores resultados.

Para la UPV, el método de RLS tiene los errores más bajos cuando se utiliza datos iguales o mayores a 40 días. Con RLM es ideal el pronóstico con datos mayores a 30 días. En el caso del método ARIMA la cantidad de datos más favorable es entre 10 y 30 días.

Tabla 4.10. Rango de datos de consumo para realizar el pronóstico de consumo de electricidad

Método	Número de días adecuado	
	UPS	UPV
RLS (temp.)	40-100	+40
RLM	40-100	+30
ARIMA	40-100	10-30

Para la UPV, el método de RLS tiene los errores más bajos cuando se utiliza datos iguales o mayores a 40 días. Con RLM es ideal el pronóstico con datos mayores a 30 días. En el caso del método ARIMA la cantidad de datos más favorable es entre 10 y 30 días.

En la Tabla 4.11 se realiza una comparación con los resultados obtenidos para la UPS en [12] con base al error porcentual absoluto medio (MAPE).

Tabla 4.11 Comparación de resultados desde el 1 al 7 de diciembre de 2017

	Trabajo de titulación [12]			Metodología propuesta		
	MAPE [%]					
	RLS	RLM	ARIMA	RLS	RLM	ARIMA
Lunes	49,46	48,26	14,09	6,56	6,53	6,27
Martes	39,19	39,78	9,9	8,5	8,82	8,02
Miércoles	24,16	65,18	7,29	8,13	8,82	6,91
Jueves	34,01	56,85	14,61	8,27	9,3	6,58
Viernes	84,86	85,45	10,6	6,32	7,36	21,49
Sábado	65,79	59,73	13,68	10,14	9,15	14,08
Domingo	64,98	94,66	16,8	4,02	4,07	17,79

En la Tabla 4.12 se compara los resultados obtenidos de la UPV en [12] para una semana del mes de julio correspondiente al verano.

Tabla 4.12 Comparación de resultados de la UPV

	Trabajo de titulación [12]			Metodología propuesta		
	MAPE [%]					
	RLS	RLM	ARIMA	RLS	RLM	ARIMA
Lunes	44,25	38,28	13,21	11,96	14,28	7,1
Martes	32,27	41,17	13,25	9,88	10,2	13,69
Miércoles	58,48	75,06	12,87	11,16	12,73	12,07
Jueves	81,04	92,73	13,01	8,18	8,73	11,62
Viernes	101,75	71,13	11,39	8,82	9,01	13,66
Sábado	28,31	9,25	5,12	4,32	5,5	3,92
Domingo	12,02	13,92	5,69	4,24	8,19	5,43

De acuerdo a la Tabla 4.11 y 4.12, los errores obtenidos en los métodos de regresión al realizar la segmentación horaria mejoran satisfactoriamente en relación a [12]. Para el método ARIMA, los errores obtenidos con el entorno de programación en MATLAB son similares en relación a los obtenidos en el software SPSS de [12].

## 5. CAPITULO 5

### CONCLUSIONES

Los resultados del pronóstico realizado en los dos casos de estudio (UPS y UPV) han permitido conocer la precisión de los métodos estadísticos y las condiciones bajo las cuales se pueden utilizar para obtener errores bajos.

El análisis estadístico permitió identificar las características de la demanda de energía de los días laborables y no laborables para posteriormente realizar la agrupación de días de consumo similar y el tratamiento por horas. La agrupación de datos favorece en la reducción de los errores, al aumentar la cantidad de datos que se utilizan para conformar el modelo. Gracias al cálculo del coeficiente de determinación, se determinó que la dirección del viento tiene una pobre relación con la potencia eléctrica, por lo cual se excluyó del análisis.

En la UPS sede Cuenca la potencia eléctrica no se ve influenciada por los cambios de temperatura. Los datos de consumo en relación a la variable independiente tienen baja dispersión en la segmentación horaria. Los resultados obtenidos para la RLS y RLM son similares, el MAPE promedio obtenido fue del 8%, RMSE de 10.35 kW y MAE de 9 kW.

Para los días no laborables de la UPS sede Cuenca la ausencia de estudiantes ocasiona que la potencia sea constante en todas las horas permitiendo un adecuado ajuste de la recta de regresión. Los errores obtenidos de los métodos de regresión son muy bajos y similares entre RLS y RLM. El error promedio es de 2.9kW de MAE, 4.2kW en RMSE y 4.1% para MAPE. Por otra parte, el método ARIMA obtiene errores más altos.

Al separar los datos de consumo de electricidad de la UPV de acuerdo a las estaciones climáticas, se define una relación mayormente lineal entre la potencia y las variables independientes. En la primavera y el verano se obtienen errores de pronóstico más bajos con los métodos de regresión, además la potencia eléctrica tiene mayor relación con las variables de temperatura y humedad.

La presencia de valores atípicos en todos los días del otoño como se observa en la Figura 2.13 afecta el ajuste con los métodos de regresión. Los errores más bajos en esta época se dan con el método ARIMA, 10.26 kW de MAE, 12.72 en el RMSE y 18% P para el MAPE. En el invierno, según la Tabla 2.14, la potencia y las variables temperatura y humedad tiene menor porcentaje de relación con respecto al resto de estaciones. Con los métodos de regresión se obtienen los errores más bajos, 13.2% en MAPE, 10.5 kW en RMSE y 7.8 kW para MAE.

Los métodos de regresión son adecuados para grupos de datos con baja dispersión y una relación lineal entre las variables. Además, cuando el coeficiente de determinación es más alto, el

pronóstico es más preciso y por lo tanto los errores son menores. De acuerdo a los resultados, la cantidad de datos de días de consumo que se deben usar para generar los modelos de regresión con bajo error es de entre 40 y 100 días.

El método ARIMA obtiene buenos resultados con grupos de datos más dispersos como en el caso del otoño en la UPV. Sin embargo, se debe prestar especial atención en la cantidad de datos que se usa para el pronóstico. De acuerdo a los resultados obtenidos, se determinó que datos de consumo de entre 40 y 60 días es adecuado para este caso.

La utilización de una cantidad de datos adecuada reduce el tiempo y gasto computacional requerido.

En este trabajo se presenta una aplicación informática desarrollada en el software Matlab para el pronóstico de consumo de energía eléctrica. Esta aplicación tiene la capacidad de realizar el tratamiento y análisis de bases de datos relacionadas a la demanda de energía eléctrica de una manera eficaz. La aplicación desarrollada facilitará el procesamiento, agrupación y pronóstico de la demanda de electricidad en futuras investigaciones.

## **RECOMENDACIONES**

Para posteriores investigaciones se sugiere realizar el estudio de la base de datos de la UPV usando métodos de pronóstico basados en inteligencia artificial debido a la relación no lineal entre la temperatura y la demanda de energía.

Para el correcto funcionamiento del programa desarrollado en el software Matlab se debe tener en cuenta el modelo de las bases de datos utilizadas en este trabajo investigativo y seguir el manual de usuario tal como se indica en el Anexo 1.

# ANEXO 1. MANUAL DE USUARIO DE ENTORNO DE PROGRAMACIÓN EN MATLAB PARA PRONÓSTICO DE CONSUMO DE ELECTRICIDAD

## Regresión lineal simple y regresión lineal múltiple.

1. Para usar correctamente el entorno de programación que ha sido diseñado en Matlab para RL y RLM, se debe contar con todos los archivos de Excel que el programa necesita, estos son:

- Base de datos históricos

Contiene los datos históricos del consumo eléctrico, estos deben ser ingresados tal y como se muestra en la siguiente tabla.

Para los días de la semana se ingresará su equivalente numérico, siendo 1 para lunes y 7 para domingo, en el caso de la laborabilidad será 1 para días laborable y 3 para no laborables.

Tabla 1. Modelo para ingreso de la base de datos históricos.

Hora	Día Semana	Laboralidad	Potencia [kW]	Temperatura [°C]	Humedad	Var. Climática	Var. Climática	Var. Climática
0	1	1	23.72	18	0.41	0	0	0

```
global A Hora Dia Laborabilidad  
A=xlsread('C:\Users\Usuario\Desktop\PROCESO TITULACION\DOCUMENTOS MATLAB\DatosUPV.xlsx',6);
```

Figura 1. Línea de código para cargar la base de datos históricos asignado al botón “CARGAR DATOS”.

- Datos de valores predichos

Estos datos corresponden a los valores predichos de las variables climáticas que se están analizando, se deberá ingresar un valor por cada hora y para cada día de la semana.

- Datos de consumo real

Un valor para cada hora y por cada día de la semana deberá ser ingresado en el consumo real.

Cada uno de estos archivos de Excel de preferencia deberán ser colocados en una misma carpeta, en el código del script se deberá colocar la dirección del documento y su nombre para que pueda ser leído.

```

Valores_Predicidos=xlsread('C:\Users\Usuario\Desktop\PROCESO TITULACION\DOCUMENTOS MATLAB\Valores predicidosRLI
Consumo_real=xlsread('C:\Users\Usuario\Desktop\PROCESO TITULACION\DOCUMENTOS MATLAB\Consumo_real.xlsx',1);

```

Figura 2. Línea de código para lectura de valores predichos y de consumo real asignado al botón “PREDICCIÓN”.

```

sum_porcentaje_error_medio=0;
Consumo_real=xlsread('C:\Users\Usuario\Desktop\PROCESO TITULACION\DOCUMENTOS MATLAB\Consumo_real4.xlsx',1);

```

Figura 3. Línea de código para la lectura del consumo real asignado al botón “ERRORES”.

2. Dentro del entorno de programación, se dará clic sobre el botón “CARGAR DATOS” esperando la confirmación de carga a través de un mensaje en un cuadro de diálogo.

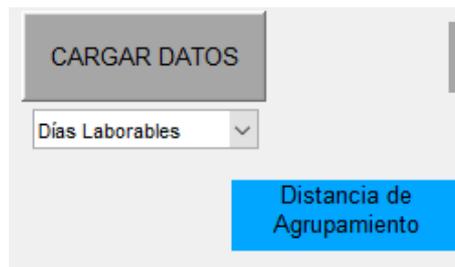


Figura 4. Botón “CARGAR DATOS”

3. Ingresar el valor para la distancia de agrupamiento entre 0 y 1, además el valor Z para la eliminación de valores atípicos (valor recomendado de 1.98 para el 95% de los datos). Seleccionar la laborabilidad que se quiera analizar dentro de la lista desplegable. Una vez realizado este paso, se deberá esperar que se muestre los perfiles de consumo de electricidad de cada uno de los días.

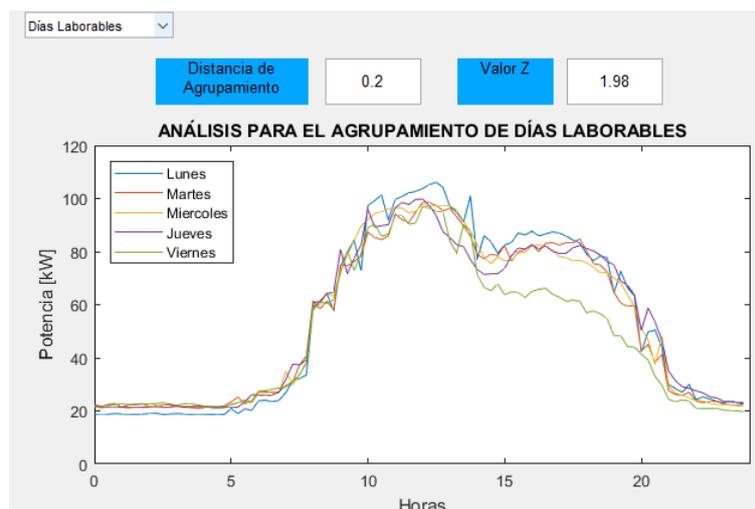


Figura 5. Ingreso de datos para la distancia de agrupamiento y el valor Z.

4. Para conocer el agrupamiento de los días de consumo similar de electricidad, haga clic en el botón “GRUPOS”, se mostrará un cuadro de diálogo que indica la formación de los grupos.

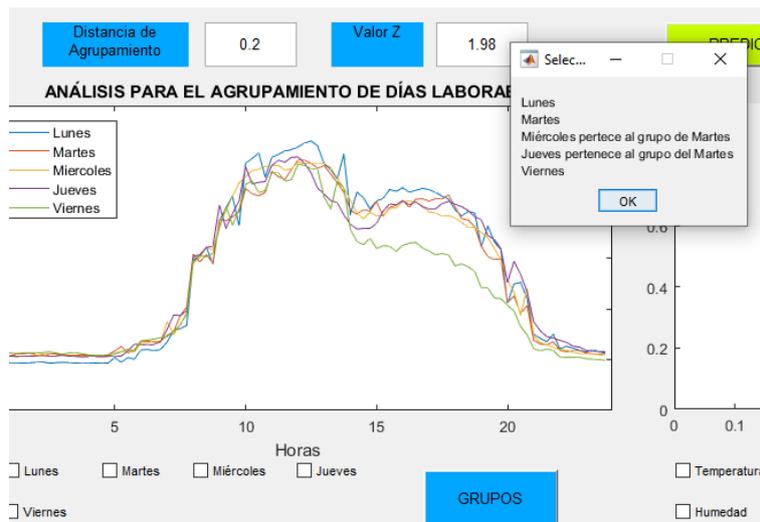


Figura 6. Botón “GRUPOS”

- De acuerdo al agrupamiento que se quiera analizar, se deberá seleccionar los checkbox que corresponden a los días requeridos del grupo.

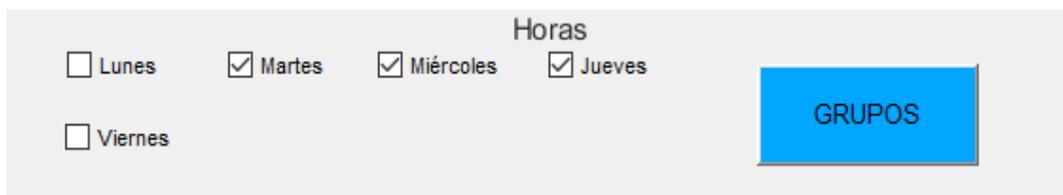


Figura 7. Selección de grupos de acuerdo al agrupamiento

- Dentro de la sección de métodos de predicción, se selecciona el método, ya sea Regresión Lineal o Regresión Lineal Múltiple. A continuación, se selecciona también el día que se quiere conocer el pronóstico.

En el caso de que se trate de una Regresión lineal simple, adicionalmente se selecciona la variable climática que se quiere involucrar en la lista desplegable.

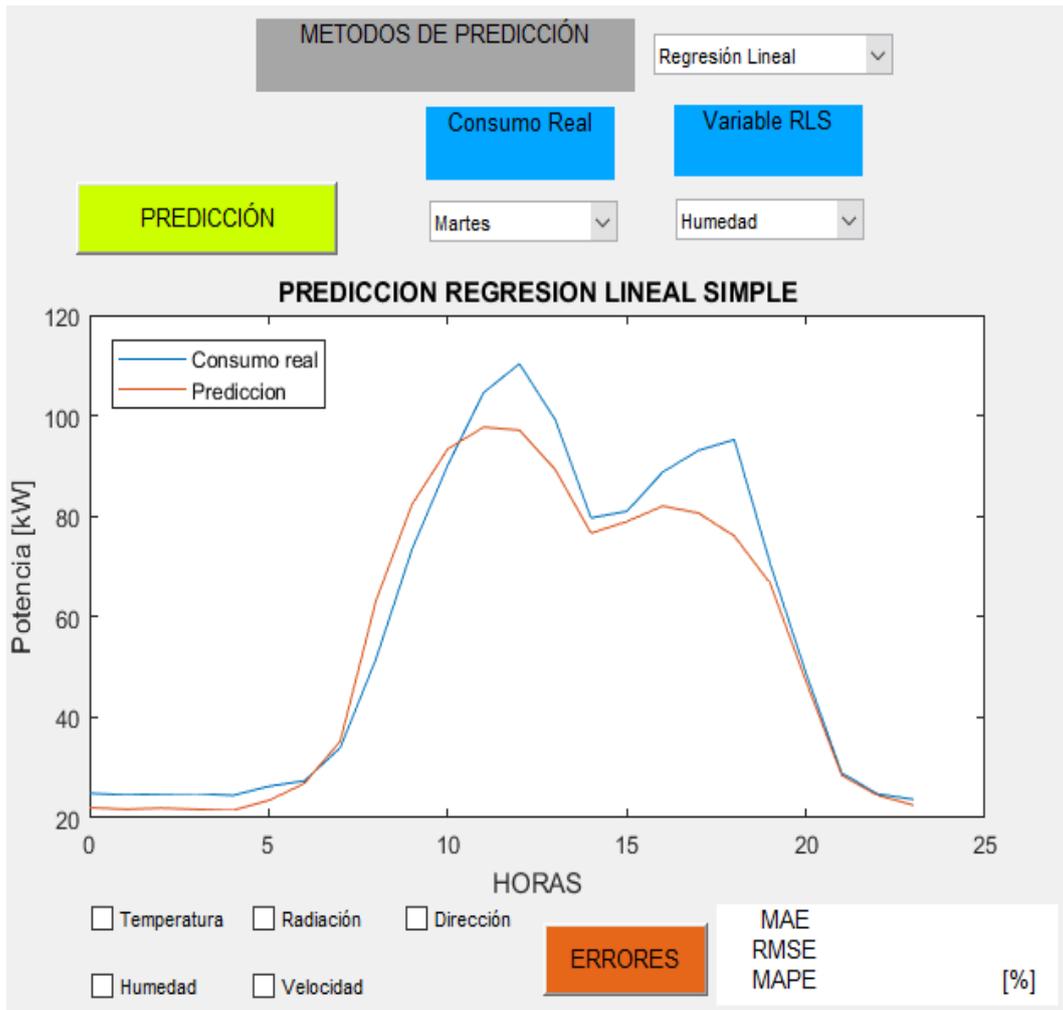


Figura 8. Selección de una variable climática para RLS.

Para cuando se trate de una Regresión Lineal Múltiple, se seleccionan a la vez todas las variables climáticas que se quieren considerar en los checkbox.

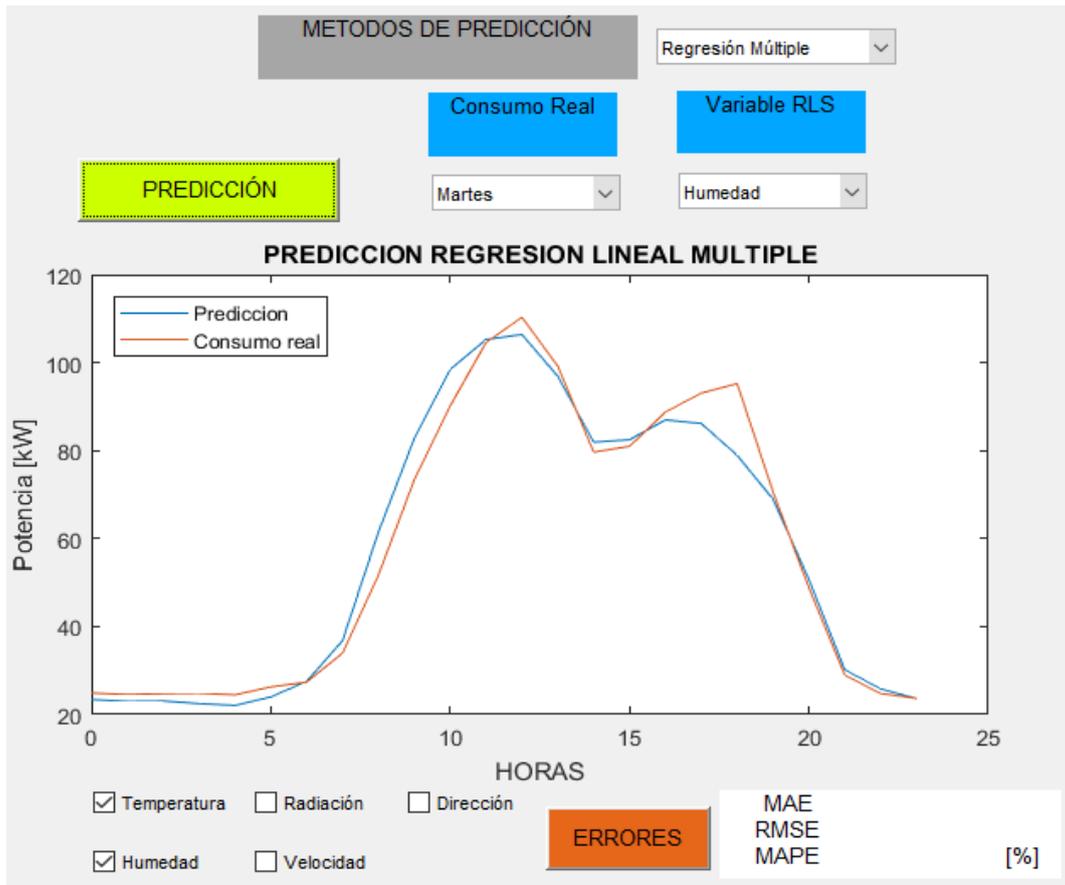


Figura 8. Selección de las variables climática para RLM.

- Finalmente, para conocer los errores entre el consumo real y la predicción, se hace clic sobre el botón “ERRORES”.

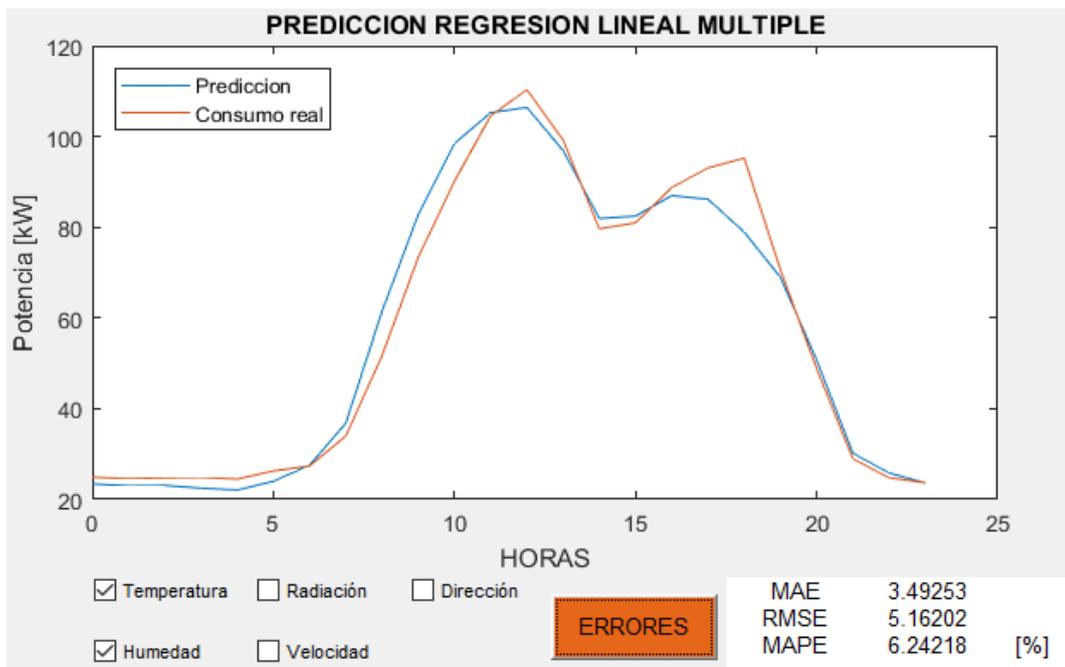


Figura 9. Botón “ERRORES”

## Modelo Autorregresivo Integrado de Media Móvil (ARIMA)

Para utilizar correctamente este programa se deben seguir los pasos que se detallan a continuación. En los ejemplos se usa la base de datos de la UPV.

1. Para usar correctamente el entorno de programación que ha sido diseñado en Matlab para el método ARIMA, se debe contar con todos los archivos de Excel que el programa necesita, estos son:

- Base de datos históricos

La base de datos a ingresar debe seguir el formato de la tabla mostrada a continuación. Tiene la opción de trabajar con 5 variables climáticas, en caso de no requerir hacerlo se debe poner el valor de 0 en todas las celdas.

Tabla 2. Modelo para ingreso de datos

Hora	Día Semanal	Laboralidad	Potencia [kW]	Temperatura [°C]	Humedad	Var. Climática	Var. Climática	Var. Climática
0	1	1	23,72	18	0,41	0	0	0

- Datos de consumo real

La base de datos de consumo real se usa para evaluar los errores de pronóstico, por ello se deberá tener el consumo real de todos los días de la semana que se esté analizando ya sea para días laborables o no laborables.

```
sum_porcentaje_error_medio=0;
Consumo_real=xlsread('C:\Users\Usuario\Desktop\PROCESO TITULACION\DOCUMENTOS MATLAB\Consumo_real4.xlsx',7);
```

Figura 10. Línea de código para lectura del consumo real asignado al botón “ERRORES”.

```
Consumo_real=xlsread('C:\Users\Usuario\Desktop\PROCESO TITULACION\DOCUMENTOS MATLAB\Consumo_real4.xlsx',7);
y =(Puntos_potenciagrupo0);
```

Figura 11. Línea de código para la lectura del consumo real asignado al botón “PREDICCIÓN”.

2. Cargar la base de datos presionando “CARGAR DATOS”. Se muestra un mensaje en un cuadro de texto para confirmar la carga de datos.



Figura 12. Botón “CARGAR DATOS”

- Ingresar los valores para la distancia de agrupamiento y el valor Z.

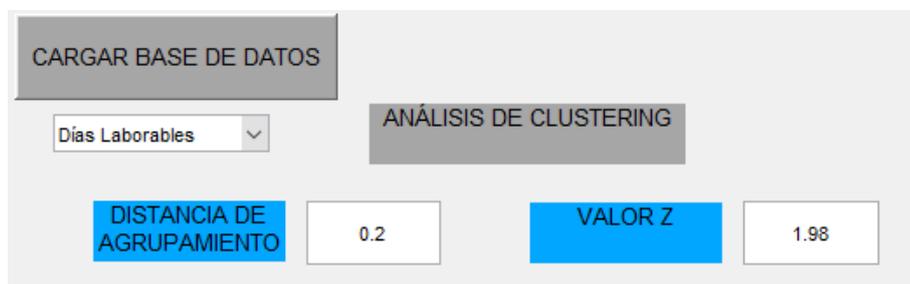


Figura 13. Ingreso de valores de distancia de agrupamiento y valor Z

- Seleccionar en la pestaña días laborables o no laborables, se mostrará los perfiles de consumo de electricidad de los días que corresponde.

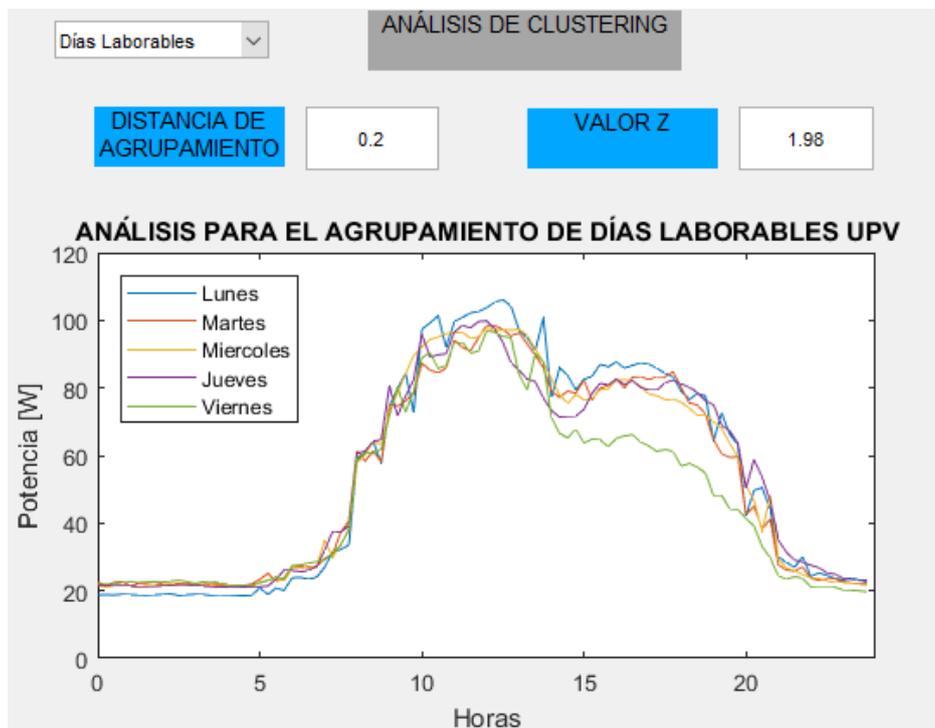


Figura 14. Perfiles de consumo de días laborables de la UPV en la primavera.

5. Al presionar el botón “GRUPOS” se mostrará una ventana, en la cual se indica el agrupamiento de los días de consumo similar.

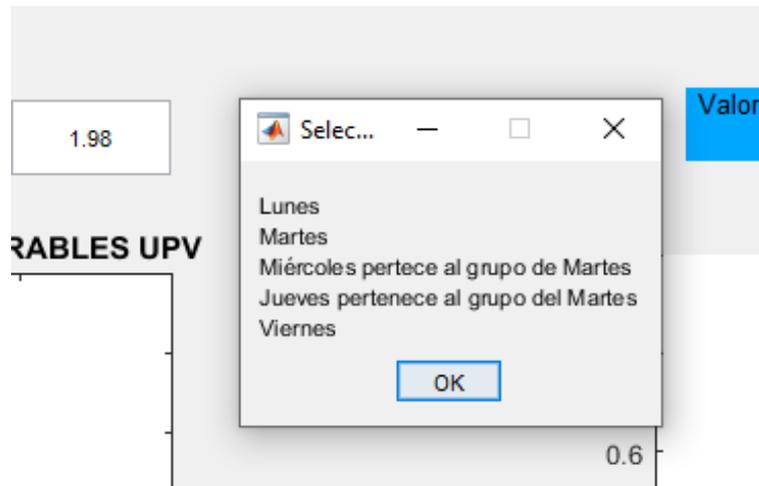


Figura 15. Grupos obtenidos del consumo similar de electricidad

6. De acuerdo a la información obtenida en el paso anterior se deberá marcar las casillas correspondientes de los días.

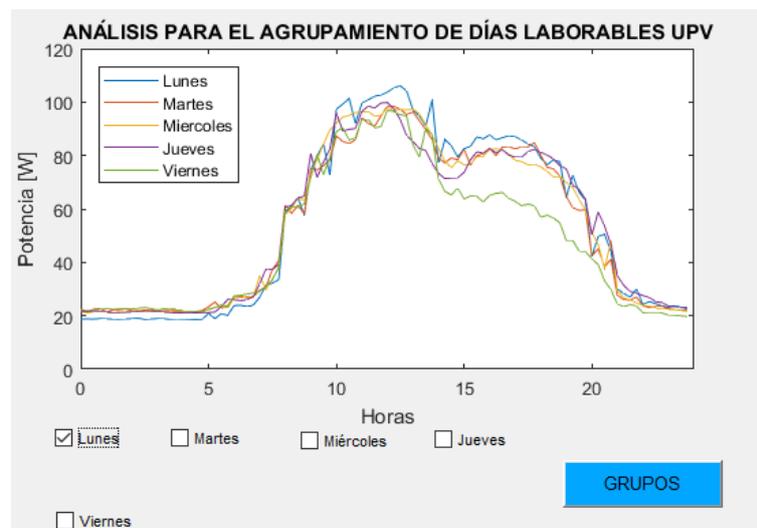


Figura 16. Días seleccionados de acuerdo al agrupamiento obtenido.

7. Se presiona el botón “AUTOCORRELACIÓN”, de lo cual se obtienen las gráficas de las autocorrelaciones, donde se deberá verificar a que modelo corresponde para posterior el ingreso de los valores P, D y Q del modelo ARIMA.

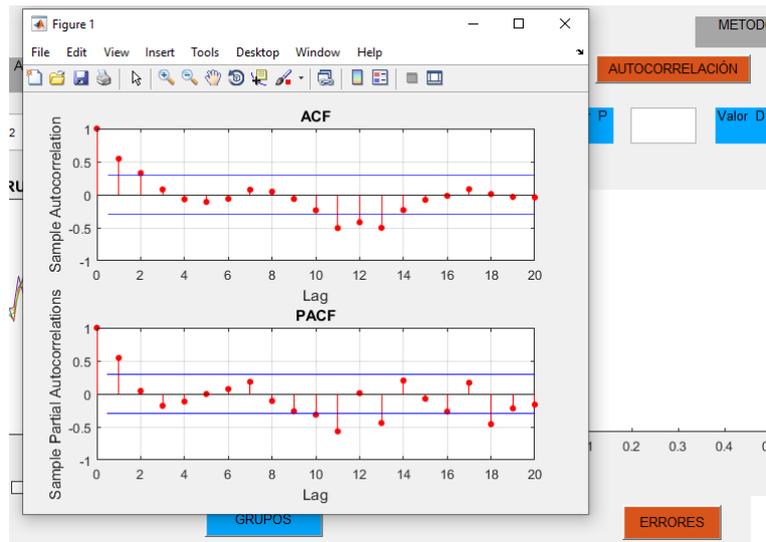


Figura 17. Autocorrelación simple y parcial

8. Una vez que se han ingresado los valores P, D y Q, se selecciona el día que se quiere pronosticar y se presiona el botón “PREDICCIÓN” en la ventana del guide.

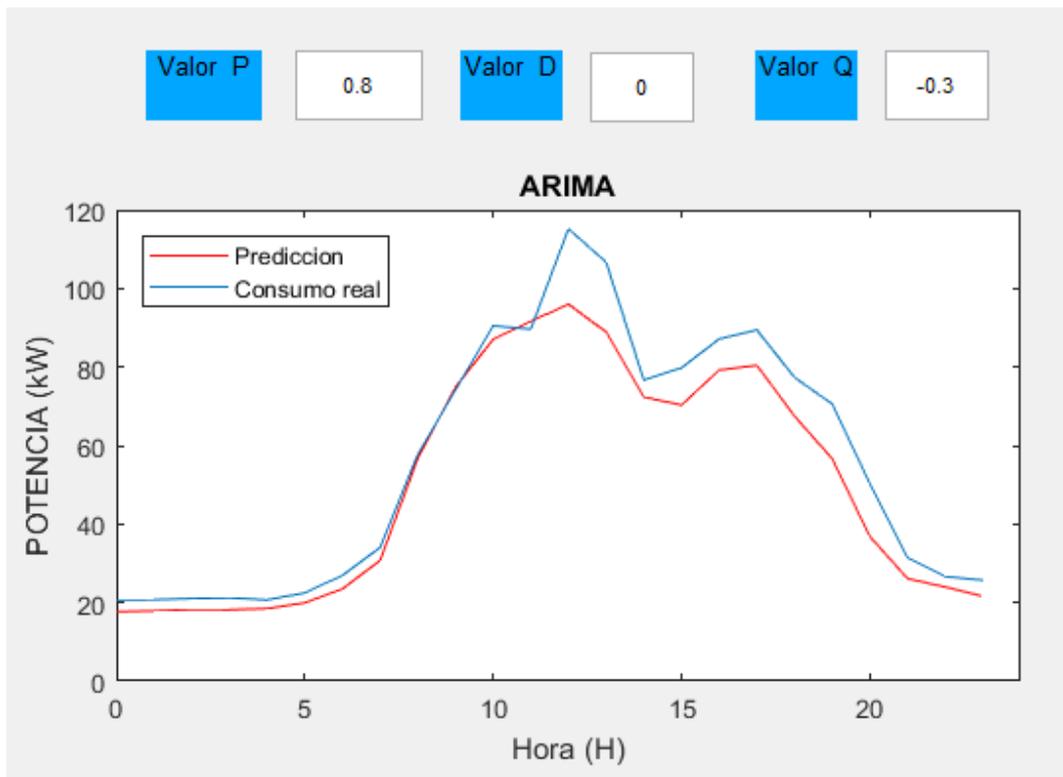


Figura 18. Pronóstico del consumo de electricidad obtenido

9. Finalmente, al obtener las gráficas del consumo real y pronóstico del día requerido se obtienen los errores de pronóstico presionando el botón “ERRORES”.

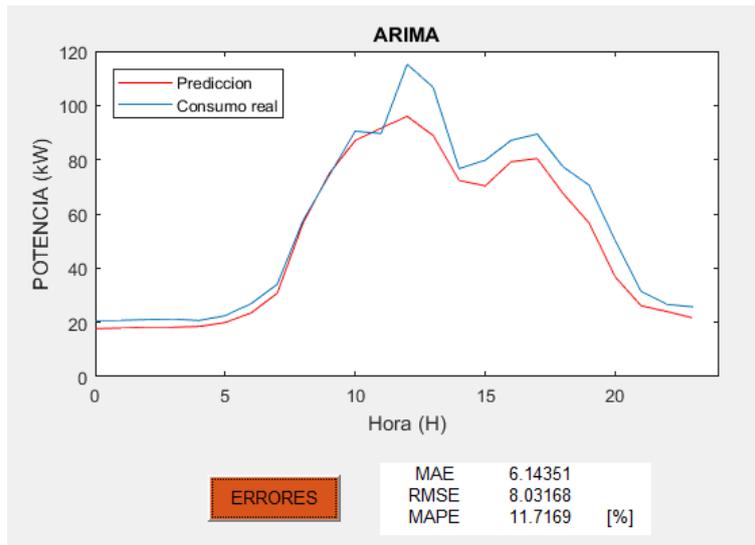


Figura 19. Errores de pronóstico obtenidos

## ANEXO 2. DATOS DE PRONÓSTICO

Tabla 1. Potencias pronosticadas de la UPV para la semana del 25 de mayo al 31 de mayo de 2016

Potencias pronosticadas					
Día/Método	RLS (temperatura)	RLS (humedad)	RLM	Arima	Consumo Real
LUNES	19,51	18,40	19,44	19,46	21,37
	19,90	18,37	19,86	19,42	21,46
	20,59	18,20	20,50	20,01	21,34
	20,26	18,30	20,15	19,69	21,44
	20,14	18,01	19,68	19,34	21,25
	21,41	19,37	20,66	21,58	22,89
	24,66	21,92	22,86	23,94	24,97
	32,21	30,69	32,38	33,10	30,73
	59,67	60,52	58,88	49,47	50,90
	76,70	78,33	76,02	67,28	67,30
	96,06	102,53	104,36	80,79	81,26
	101,98	102,49	105,79	81,61	84,58
	105,65	104,81	108,48	88,64	89,67
	97,09	95,85	103,75	81,31	92,37
	83,96	77,61	86,77	72,99	79,42
	86,99	85,79	93,16	79,97	76,80
	87,62	87,33	96,67	82,97	78,93
	86,25	85,24	94,90	76,81	74,74
	77,75	78,01	82,92	72,82	77,60
	68,03	68,27	76,06	80,52	63,78
49,83	42,42	44,54	57,12	46,78	
29,86	28,60	30,53	32,93	30,70	
25,48	23,93	24,80	23,57	27,14	

	24,73	22,97	24,01	23,54	22,83
MARTES	22,64	21,52	22,44	30,27	22,89
	22,38	21,35	22,08	28,71	22,79
	22,64	21,55	22,40	28,31	23,07
	22,47	21,24	22,15	27,65	23,30
	22,00	21,05	21,75	27,09	22,90
	24,08	23,07	23,84	28,70	24,23
	27,82	26,60	27,58	28,98	26,44
	38,79	34,54	38,16	37,31	33,18
	60,69	59,07	57,27	67,83	52,81
	80,97	74,45	75,63	74,40	78,65
	94,95	88,32	92,77	81,29	92,45
	100,35	93,31	99,18	90,08	97,39
	102,05	95,03	102,79	94,31	96,39
	92,94	89,46	97,17	90,87	89,77
	79,64	76,76	82,72	81,80	79,62
	82,07	79,14	83,90	85,82	85,73
	84,72	80,95	86,92	93,98	92,33
	82,98	79,28	83,60	91,55	91,75
	78,10	75,72	78,75	80,66	84,91
	67,11	64,37	66,57	68,47	80,23
	49,41	46,33	49,41	52,79	59,73
	29,62	28,05	29,62	32,83	29,91
	25,39	24,45	25,43	27,31	24,87
23,37	22,52	23,30	25,09	24,12	
MIERCOLES	22,92	22,09	23,13	30,27	23,67
	22,66	21,74	22,81	28,71	23,56
	22,64	21,78	22,65	28,31	23,62
	22,24	21,50	22,13	27,65	23,36

	21,80	21,20	21,66	27,09	23,54
	23,74	23,07	23,52	28,70	24,53
	27,31	26,61	27,08	28,98	26,56
	36,26	33,72	35,08	37,31	30,93
	60,88	60,47	59,15	67,83	58,41
	80,54	76,95	77,83	74,40	87,97
	94,95	87,80	92,41	81,29	92,29
	100,35	93,45	99,24	90,08	94,05
	104,16	95,87	104,66	94,31	115,91
	95,91	89,11	98,65	90,87	103,56
	81,12	76,58	82,77	81,80	85,23
	82,80	79,14	84,75	85,82	82,61
	84,72	80,83	87,19	93,98	71,48
	85,58	79,61	86,23	91,55	64,73
	80,53	75,59	81,87	80,66	55,95
	68,94	63,23	67,44	68,47	49,60
	49,95	45,15	49,92	52,79	37,35
	29,62	27,18	29,53	32,83	22,73
	25,39	23,54	25,28	27,31	20,40
	23,25	22,43	23,38	25,09	19,23
JUEVES	23,21	21,75	23,16	30,27	19,15
	23,22	21,74	23,37	28,71	19,07
	23,16	21,92	23,32	28,31	19,27
	22,70	21,76	22,78	27,65	19,03
	22,20	21,62	22,35	27,09	19,09
	24,08	23,60	24,31	28,70	20,58
	27,82	26,99	28,00	28,98	23,87
	37,53	35,85	38,04	37,31	31,30
	60,69	63,77	62,49	67,83	55,80

	80,11	80,13	80,68	74,40	73,91
	93,78	91,90	94,26	81,29	86,31
	98,76	96,01	98,78	90,08	90,26
	99,94	96,54	100,07	94,31	94,95
	89,97	89,11	91,60	90,87	87,85
	76,67	76,56	77,32	81,80	72,50
	80,62	79,46	80,70	85,82	72,03
	81,73	81,39	82,29	93,98	79,87
	80,39	80,41	80,35	91,55	79,95
	76,28	76,36	75,29	80,66	68,49
	66,50	69,18	68,57	68,47	60,51
	48,32	46,98	48,34	52,79	40,13
	28,59	28,00	28,59	32,83	27,20
	24,61	24,06	24,60	27,31	23,93
	23,01	22,48	23,02	25,09	21,93
VIERNES	22,51	21,95	22,53	24,68	21,76
	23,28	22,24	23,30	23,98	21,75
	23,68	22,22	23,48	24,24	21,46
	23,26	21,82	23,31	23,88	21,45
	22,07	21,01	21,68	24,15	21,44
	23,36	23,16	23,45	25,45	23,54
	28,58	27,90	28,46	26,10	26,82
	33,26	33,61	33,95	35,44	32,79
	57,23	65,71	62,36	69,62	52,95
	77,20	80,19	82,02	78,50	71,81
	92,97	86,46	93,28	84,21	84,72
	95,66	91,94	96,31	96,14	90,77
	98,71	94,72	100,93	96,64	100,76
	85,77	84,34	88,79	86,10	93,59

68,09	68,85	71,27	64,99	77,68
64,54	63,62	65,67	56,45	74,79
65,33	66,79	70,55	58,02	79,17
62,49	64,27	68,18	55,56	87,46
57,37	57,90	59,91	48,08	83,22
46,84	46,20	49,47	42,57	76,22
35,70	37,44	39,86	35,11	47,34
24,35	23,11	25,04	23,48	26,00
21,48	20,59	22,00	19,90	22,39
20,54	19,96	20,70	20,02	21,77

## BIBLIOGRAFÍA

- [1] 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. ROPEC 2017*, vol. 2018-Janua, no. Ropec, pp. 1–5, 2018.
- [2] J. Berrezueta and A. Encalada, "Análisis de los factores que afectan la demanda de energía eléctrica y su estimación. Sector residencial del Área de concesión de la Empresa Eléctrica Regional Centro Sur C.A., Periodo 2002-2012.," pp. 1–145, 2014.
- [3] T. Ahmad and H. Chen, "Nonlinear autoregressive and random forest approaches to forecasting electricity load for utility energy management systems," *Sustain. Cities Soc.*, vol. 45, pp. 460–473, 2019.
- [4] J. Yuan, C. Farnham, C. Azuma, and K. Emura, "Predictive artificial neuronal network models to forecast the seasonal hourly electricity consumption for a University Campus," *Sustain. Cities Soc.*, 2018.
- [5] H. Son and C. Kim, "Short-term forecasting of electricity demand for the residential sector using weather and social variables," *Resour. Conserv. Recycl.*, vol. 123, pp. 200–207, 2017.
- [6] V. Bianco, O. Manca, and S. Nardini, "Electricity consumption forecasting in Italy using linear regression models," *Energy*, vol. 34, no. 9, pp. 1413–1421, 2009.
- [7] M. Chong and R. Aguilar, "Proyección de Series de Tiempo para el Consumo de la Energía Eléctrica a Clientes Residenciales en Ecuador," vol. 29, no. Julio, pp. 56–76, 2016.
- [8] C. P. Murillos Joaquin, Trejos Alvaro, "Estudio del pronóstico de la demanda de energía eléctrica, utilizando modelos de series de tiempo," no. 23, pp. 37–42, 2003.
- [9] S. Noureen, S. Atique, V. Roy, and S. Bayne, "A Comparative Forecasting Analysis of ARIMA Model vs Random Forest Algorithm for A Case Study of Small-Scale Industrial Load A Comparative Forecasting Analysis of ARIMA Model vs Random Forest Algorithm for A Case Study of Small-Scale Industrial Load," no. October, 2019.
- [10] M. Safa, B. Kc, and M. Safa, "Linear Model To Predict Energy Consumption Using Historical Data From Cold Stores," *Int. J. Adv. Sci. Eng. Technol. Spl. Issue-3*, no. 3, pp. 2321–9009, 2015.
- [11] M. Safa, J. Allen, and M. Safa, "Predicting energy usage using historical data and linear models," *31st Int. Symp. Autom. Robot. Constr. Mining, ISARC 2014 - Proc.*, no. Isarc, pp. 531–535, 2014.
- [12] L. F. Siavichay Cumbe and J. X. Serrano Guerrero, "Aplicación de métodos estadísticos para el análisis y predicción de perfiles de consumo de energía eléctrica," Universidad Politécnica Salesiana, 2018.
- [13] J. P. Muñoz-Vizñay, M. V. Rojas-Moncayo, and C. R. Barreto-Calle, "Incentivo a la generación distribuida en el Ecuador," *Ingenius*, no. 19, p. 60, 2018.
- [14] J. Bedi and D. Toshniwal, "Deep learning framework to forecast electricity demand," *Appl. Energy*, pp. 1312–1326, Mar. 2019.
- [15] G. K. F. Tso and K. K. W. Yau, "Predicting electricity energy consumption: A comparison of regression analysis, decision tree and neural networks," *Energy*, vol. 32, no. 9, pp. 1761–1768, 2007.
- [16] M. Azhar *et al.*, "Building electrical energy consumption forecasting analysis using

- conventional and artificial intelligence methods: A review,” 2017.
- [17] K. Berk, *Modeling and Forecasting Electricity Demand*, vol. 6, no. 2. 1393.
- [18] Y. Wang, J. Wang, G. Zhao, and Y. Dong, “Application of residual modification approach in seasonal ARIMA for electricity demand forecasting: A case study of China,” *Energy Policy*, vol. 48, pp. 284–294, 2012.
- [19] K. Panklib, C. Prakasvudhisarn, and D. Khummongkol, “Electricity Consumption Forecasting in Thailand Using an Electricity Consumption Forecasting in Thailand Using an Artificial Neural Network and Multiple Linear Regression,” no. May, 2015.
- [20] A. Zeng, S. Liu, and Y. Yu, “Energy & Buildings Comparative study of data driven methods in building electricity use prediction,” *Energy Build.*, vol. 194, pp. 289–300, 2019.
- [21] X. Z. Gao, A. Kaarna, L. Lensu, and S. Honkapuro, “A hybrid method for short-term electricity consumption prediction,” *Proc. IECON 2017 - 43rd Annu. Conf. IEEE Ind. Electron. Soc.*, vol. 2017-Janua, pp. 7393–7398, 2017.
- [22] H. K. Alfares and M. Nazeeruddin, “Electric load forecasting: Literature survey and classification of methods,” *Int. J. Syst. Sci.*, vol. 33, no. 1, pp. 23–34, 2002.
- [23] 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.
- [24] L. Y. R. Primarias, I. Segundo, and A. Horna, *ELECTRIFICACIÓN RURAL Y URBANA*.  
.
- [25] S. Castaño Ramirez, *Redes de distribución de energía*. 2004.
- [26] V. Hinojosa, “Pronostico de Demanda a corto plazo en Sistemas de Suministro de Energia Electrica utilizando Inteligencia Artificial,” p. 167, 2007.
- [27] M. Ortiz and E. M. Garcia Torres, “Pronóstico de la demanda eléctrica residencial basado en el modelo de regresión adaptativa multivariante spline (MARS),” *Tesis*, p. 32, 2015.
- [28] J. Garcia, N. Lopez, and J. Calvo, *Estadística Basica Para Estudiantes de Ciencias*, 1st ed. Madrid, 2011.
- [29] E. Morales, *Estadística Y Probabilidades*. 2012.
- [30] M. Arriaza, *Guía práctica de análisis de datos*, no. January 2006. 2017.
- [31] W. Menhenhall, R. Beaver, and B. Beaver, *Introducción a la probabilidad y estadística*. 2016.
- [32] D. A. Lind, W. G. Marchal, and S. A. Wathen, *Negocios Y La Economía*, 15th ed. 2014.
- [33] L. M. Dicovskiy Riobóo, “Estadística Basica,” in *Estadística Basica*, 2008, p. 146.
- [34] C. H. Cramér, “Valores Atípicos,” 1985, pp. 1–17.
- [35] F. Ocaña, “Datos Atípicos o Outliers Datos ausentes o missing,” in *Técnicas estadísticas en Nutrición y Salud Tratamiento estadístico de outliers y datos faltantes*, pp. 1–6.
- [36] 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.
- [37] R.-M. Ștefan, “Cluster Type Methodologies for Grouping Data,” *Procedia Econ. Financ.*, vol. 15, no. 14, pp. 357–362, 2014.

- [38] P. S. Contreras Ortiz and X. M. Alvarez Mora, “Aplicación De Técnicas De Clustering Y Detección De Anomalías Para La Clasificación De Consumos De Energía Eléctrica,” p. 230, 2018.
- [39] P. Tarazaga and H. Kurata, “On cell matrices: A class of Euclidean distance matrices,” *Appl. Math. Comput.*, vol. 238, pp. 468–474, 2014.
- [40] J. E. Beaver and R. M. Dean, “Using Euclidean distance in the comparative analysis of taxonomic abundance,” *J. Archaeol. Sci. Reports*, vol. 25, no. September 2018, pp. 331–340, 2019.
- [41] L. Aldaz and B. Eguía, “Diseño conceptual y Normalización de datos,” in *Organización y Estructura de la Información*, 2009, pp. 1–21.
- [42] J. Gil, J. Mora, and S. Perez, “Análisis del procesamiento de los datos de entrada para un localizador de fallas en sistemas de distribución,” vol. 18, no. 41, pp. 64–75, 2013.
- [43] T. Zhang and X. You, “Improvement of the training and normalization method of artificial neural network in the prediction of indoor environment,” *Procedia Eng.*, vol. 121, pp. 1245–1251, 2015.
- [44] J. Moral-Carcedo and J. Pérez-García, “Time of day effects of temperature and daylight on short term electricity load,” *Energy*, vol. 174, pp. 169–183, 2019.
- [45] J. Moral-Carcedo and J. Pérez-García, “Temperature effects on firms’ electricity demand: An analysis of sectorial differences in Spain,” *Appl. Energy*, vol. 142, pp. 407–425, 2015.
- [46] N. G. Jara Cobos, F. Z. Reinoso AVECILLAS, C. A. Isaza Roldan, and J. L. Espinoza Abad, “Impacts on the consumption of electric power by the use of efficient refrigerators - Ecuador case,” *Ingenius*, no. 18, p. 53, Jul. 2017.
- [47] R. Walpole, R. Myers, and S. Myers, *Probabilidad y Estadística para ingeniería y ciencias*, 9th ed. Mexico, 2012.
- [48] L. ORELLANA, “Análisis De Regresión,” *Regresión Lineal Simple*, pp. 1–49, 2008.
- [49] M. J. Rodríguez Jaume and R. Catala Modar, “Análisis de Regresión Simple,” *Estadística Informática casos y ejemplos con el SPSS*, pp. 3–17, 2001.
- [50] C. Hernandez, L. F. Pedraza, and A. E. Diaz, “Aplicaciones De Las Series De Tiempo En Modelos De Tráfico Para Una Red De Datos,” *Sci. Tech.*, vol. 14, no. 38, pp. 31–36, 2008.
- [51] G. Rios, “Series de Tiempo,” *Univ. Chile. Fac. Ciencias Fis. y Mat.*, p. 52, 2008.
- [52] D. Gujarati and D. Porter, *ECONOMETRIA*, 5th ed. Mexico, 2010.
- [53] S. Fuente, “MODELO ARIMA,” *Univ. Auton. Madrid*, p. 100, 2008.
- [54] G. Amaris, H. Ávila, and T. Guerrero, “Applying ARIMA model for annual volume time series of the Magdalena River,” *Tecnura*, vol. 21, no. 52, pp. 88–101, 2017.
- [55] L. Siavichay and X. Serrano, “APLICACIÓN DE MÉTODOS ESTADÍSTICOS PARA EL ANÁLISIS Y PREDICCIÓN DE PERFILES DE CONSUMO DE ENERGÍA ELÉCTRICA,” 2018.
- [56] P. González, “Análisis de series temporales económicas: modelos ARIMA,” p. 169, 2007.
- [57] M. Gallardo and H. Rubio, “ESTUDIOS ECONÓMICOS ESTADÍSTICOS Diagnóstico de estacionalidad con X-12-ARIMA STUDIES IN ECONOMIC STATISTICS

- CENTRAL BANK OF CHILE,” 2009.
- [58] G. Corres, A. Esteban, J. García, and C. Zárate, “Análisis de series temporales,” *Rev. Ing. Ind.*, vol. 8, no. 1, pp. 21–33, 2009.
- [59] M. Naderi and E. Khamsehchi, “Fuzzy logic coupled with exhaustive search algorithm for forecasting of petroleum economic parameters,” *J. Pet. Sci. Eng.*, vol. 176, no. September 2018, pp. 291–298, 2019.
- [60] A. Ariza, “Metodos utilizados para el pronostico de demanda de energia electrica en sistemas de distribucion,” p. 143, 2013.
- [61] T. Chai and R. R. Draxler, “Root mean square error (RMSE) or mean absolute error (MAE)? -Arguments against avoiding RMSE in the literature,” *Geosci. Model Dev.*, vol. 7, no. 3, pp. 1247–1250, 2014.
- [62] P. A. Negrón Baez, “Redes Neuronales Sigmoidal Con Algoritmo Lm Para Pronostico De Tendencia Del Precio De Las Acciones Del Ipsa Pablo Andrés Negrón Baez,” p. 52, 2014.
- [63] S. Gomez, V. Torres, and Y. Garcia, “Procedimientos estadísticos mas utilizados en el analisis de medidas repetidas en el tiempo en el sector agropecuario,” *Rev. Cuba. Cienc. Agrícola*, vol. 46, no. 1, 2012.
- [64] J. Estrella, “Implementacion de un sistema de medicion inteligente en la Universidad Politecnica Salesiana sede Cuenca,” 2018.
- [65] L. Cruz and E. Correa, “APLICACIÓN ANDROID PARA MEDICIÓN INTELIGENTE DE ENERGÍA ELÉCTRICA EN USUARIOS RESIDENCIALES URBANOS,” UNIVERSIDAD POLITECNICA DE QUITO, 2013.
- [66] E. Inga, “Despliegue óptimo de redes inalámbricas para la infraestructura de medición inteligente de energía eléctrica,” 2015.
- [67] A. Peralta, E. Inga, and R. Hincapié, “Optimal Scalability of FiWi Networks Based on Multistage Stochastic Programming and Policies,” *J. Opt. Commun. Netw.*, vol. 9, no. 12, p. 1172, 2017.
- [68] A. G. Peralta, “Planning of FiWi Networks to Support Communications Infrastructure of SG and SC,” in *Broadband Communications Networks - Recent Advances and Lessons from Practice*, InTech, 2018.