

Statistical Analysis and SARIMA Forecasting Model Applied to Electrical Energy Consumption in University Facilities

Análisis estadístico y modelo de pronóstico SARIMA aplicado al consumo de energía eléctrica en instalaciones universitarias

José Luis **Reyes Reyes**¹, Guillermo **Urriolagoitia Sosa**², Francisco Javier **Gallegos Funes**³,
Beatriz **Romero Ángeles**⁴, Israel **Flores Baez**⁵, Misael **Flores Baez**⁶

Instituto Politécnico Nacional

¹<https://orcid.org/0000-0002-7645-7280> | jreyesre@ipn.mx

²<https://orcid.org/0000-0001-7867-7386> | guiurri@hotmail.com

³<https://orcid.org/0000-0002-4854-6438> | fgallegosf@ipn.mx

⁴<https://orcid.org/0000-0001-6345-3726> | bromero@ipn.mx

Universidad Politécnica de Tecámac | Instituto Politécnico Nacional, MÉXICO

⁵<https://orcid.org/0000-0002-3339-1912> | israelb364@yahoo.com.mx

⁶<https://orcid.org/0000-0001-7657-5298> | misaelfloresbaez@yahoo.com.mx

Recibido 24-05-2022, aceptado 30-08-2022

Abstract

Analyzing the energy consumption behavior in buildings is essential for implementing energy-saving and efficient energy use measures without losing attention to the comfort inside the buildings. In this study, a statistical analysis and time series forecast of the energy situation of a group of buildings in a university academic unit in Mexico City was conducted. Seasonal Autoregressive Integrated Moving Average (*SARIMA*) models were used for the forecast with electrical energy consumption data from 55 months. Training and test partitions were created with these data to generate two *SARIMA* models. The results showed a strong dependence on the school cycle of electricity consumption, in addition to a shift in the cycle in the first year of the study. The mean absolute percentage error (*MAPE*) for the training partitions created shows that the best fit is provided by the *SARIMA* (3,1,1) (1,0,0)₁₂ model for the 48-month separation. In comparison, the *SARIMA* (2,1,2) (1,0,0)₁₂ model does so for the 43-month test partition. The confidence intervals for the 7- and 12-month forecast are less wide for the *SARIMA* (3,1,1) (1,0,0)₁₂ model than for the *SARIMA* (2,1,2) (1,0,0)₁₂ model. Statistical analysis and time series modeling allows a better understanding of the building stock's energy performance and strengthens the energy audit to design or implement energy saving or efficient energy use measures.

Index terms: energy consumption, scholar buildings, time series forecasting, *SARIMA* models.

Resumen

Analizar el comportamiento del consumo energético en edificios es fundamental para la implementación de medidas de ahorro y uso eficiente de la energía, sin perder atención al confort al interior de estos. En este estudio se realizó un análisis estadístico y de pronóstico con series de tiempo de la situación energética de un conjunto de edificios de una unidad académica universitaria de la Ciudad de México. Para el pronóstico se utilizaron modelos Estacionales Autorregresivos Integrados y de Medias Móviles (*SARIMA*) con datos del consumo de energía eléctrica de 55 meses y con estos se crearon particiones de entrenamiento y prueba que generaron dos modelos *SARIMA*. Los resultados mostraron una gran dependencia en el ciclo escolar del consumo de electricidad, además de un corrimiento en el ciclo en el primer año de estudio. El porcentaje de error absoluto medio (*MAPE*) para las particiones de entrenamiento creadas muestra que el mejor ajuste lo tiene el modelo *SARIMA* (3,1,1) (1,0,0)₁₂ para la partición de 48 meses, mientras que el modelo *SARIMA* (2,1,2) (1,0,0)₁₂ lo hace para la partición de prueba de 43 meses. Los intervalos de confianza para el pronóstico a 7 y 12 meses son menos amplios para el modelo *SARIMA* (3,1,1) (1,0,0)₁₂ que para el modelo *SARIMA* (2,1,2) (1,0,0)₁₂. Finalmente, el análisis estadístico y el modelado de series de tiempo permiten un mejor entendimiento del comportamiento energético del conjunto de edificios y fortalece la auditoría energética con miras a diseñar o aplicar medidas de ahorro o uso eficiente de la energía.

Palabras clave: consumo de energía, edificios educativos, pronóstico de series de tiempo, modelos *SARIMA*.

I. INTRODUCTION

Global warming and climate change result from fossil fuel consumption as a source of energy. The energy demand is increasing day by day; the development of countries requires the consumption of energy from different carriers. However, the predominant sources of primary energy are still oil (33.1%), coal (27%), and natural gas (24.2%) [1]. Among the different sectors, the buildings sector contributes almost one-third of final energy consumption and continues to grow, driven by the economic development of the countries [2]. However, the E.U. has reduced energy consumption in buildings through energy efficiency policies. This is not the case in the U.S. [3] and Canada [4], where there are increases in demand for commercial and residential buildings. In the case of Mexico, final consumption in commercial, public, and residential buildings has remained relatively stable [5]. For 2018, CO₂ emissions from energy consumption in buildings were 29% of a total of 33.9 Gt of CO₂ [6]. To reduce energy consumption in buildings, it is necessary to implement measures conducive to this end, thereby reducing the negative impacts on the planet. The IEA establishes five measures applicable to buildings, including energy efficiency [7]. Energy efficiency in buildings allows better management of economic and material resources; it leads to maintenance improvements, achieving both environmental and economic benefits. However, to establish which energy efficiency actions should be taken, an energy diagnosis of the current situation of the building or group of buildings is necessary. Knowing what type of energy and how it is consumed is essential to implement energy-saving and efficient energy use measures, always ensuring that the comfort conditions inside the buildings are adequate for the performance of human activities. In the case of public buildings, such as schools, it is necessary to ensure that actions to achieve energy efficiency and economic savings do not somehow decimate the conditions suitable for the realization of the activities of each type or educational level [8]. Before opting for any measure according to the circumstances or existing ones [9] in the refurbishment of buildings to improve their energy efficiency, it is important to analyze the consumption pattern and forecast it.

For this purpose, energy forecasting models are used to establish energy-saving and efficient energy-use measures without altering the proper operation or service of the facilities. It is essential that the ability of the forecasting model can learn from past energy consumption patterns and accurately predict the future. This would allow the management and maintenance of the building to find corrective measures to possible variations in demand. Various methods and models are used to model energy behavior [10], [11], [12]. There are statistical methods to analyze the energy performance of buildings [13], [14], [15], [16], regression models [17], [18] and others that use specialized software for energy analysis [19], [20], [21], [22]. But for some years now, methods based on time series have been used [23], [24], and among these, the autoregressive integrated moving average (ARIMA) and SARIMA models that by themselves require fewer parameters and resources in their application. [25], [26]. These models have been used in combination with physical models [27] and others, such as artificial neural networks (ANN) and supported vector machine (SVM) [28], [29] and machine learning (ML) [30].

Among the range of data prediction techniques, some are the most recurrent in the energy analysis of buildings, all of which have advantages and disadvantages in their application. For example, ANN models can be applied to nonlinear processes without knowing the relationship between input and output variables. However, evaluating the estimated parameters' relevance is impossible since there are no p-values. ARMA and ARIMA models are characterized by their ease of application and interpretation of parameters; they are more accurate than regression models, provide more reliable confidence intervals in predictions, require few computational resources, and use historical data. However, many models may need to be tested to fit, and although it is possible to determine the relationship between variables, their causal mechanism is not. Finally, these models are affected by outliers, and the forecast horizon may be short. In Decision Tree models, DT, rules are obtained that can be interpreted together with logical statements. However, they do not work well for nonlinear processes; they are susceptible to noise and unsuitable for time series. In the case of SVM models, they easily adapt to various problems, and optimal solutions are obtained; they can transform a nonlinear problem into a linear one. However, it is sometimes difficult to determine the kernel function and they can be computationally inefficient. Another type of model is the Fuzzy model, which, among its advantages, is its ability to be conducted without

a training phase. This means that it can be used on data not contained in the training set. *Fuzzy* logic is derived from Boolean logic, and its rules for a model are usually not difficult to structure.

On the other hand, its disadvantages are that the model cannot be better than the expert because it cannot be trained and is challenging to fit with noise, in addition to its high computational complexity and lack of stability. The k-nearest neighbor (*k-NN*) models are characterized by the fact that they do not require prior training, which results in faster processing and ease of implementation for various problems. However, how the nearest neighbors are calculated, i.e., the distance function, is difficult to determine. Therefore, they are unsuitable for large data sets and overly sensitive to outliers and noise. Table 1 shows some previously described methodologies used in building energy analysis, classifying them by energy scale, energy type, time scale and type of input data for building energy analysis.

TABLE 1. MODEL, TIME SCALE, TYPE OF ENERGY ANALYZED, LENGTH OF MEASUREMENT, AND TYPE OF MODEL INPUT DATA USED IN ENERGY ANALYSIS.

Model	Energy type	Time scale	Data source	Measure length
ANN	<ul style="list-style-type: none"> • Electric Consumption ^{[31], [32]} • Cooling energy consumption ^[33] • Heating and cooling consumption ^[34] • Cooling demand ^[35] 	<ul style="list-style-type: none"> • Hourly ^[32] • Daily ^{[31], [33]} 	<ul style="list-style-type: none"> • Day type/Temp/Hum/Wind/Rad • Historical data 	<ul style="list-style-type: none"> 1-2 years 2 months 1 year 45 weekdays
SVM	<ul style="list-style-type: none"> • Electric consumption ^{[36], [37]} • Electric consumption ^{[38], [39]} • Cooling load ^[40] 	<ul style="list-style-type: none"> • Hourly ^[37] • Monthly ^[36] • Daily and Half-hourly ^[38] • Daily ^[40] • Monthly ^[39] 	<ul style="list-style-type: none"> • Calendar/historical data/Temp • Sensor/Temp/holidays indicator • Meters • Measure data • Historical data 	<ul style="list-style-type: none"> 2 years 1 year 6 months 3 years
ARIMA model	<ul style="list-style-type: none"> • Thermal load ^[41] • HVAC ^{[42], [43]} • Electric consumption ^[44] • Electricity demand ^[45] 	<ul style="list-style-type: none"> • Hourly ^{[41], [43]} • Daily ^[43] • Yearly ^[42] • Monthly ^[44] • Monthly ^[45] 	<ul style="list-style-type: none"> • Tem/Hum/Wind/Rad/Occu • Building performance • Temperature 	<ul style="list-style-type: none"> 6 months 1 year 7 years 10 years, 10 months
Fuzzy model	<ul style="list-style-type: none"> • Electric consumption ^[46] • Heating demand, electricity demand ^[47] • Energy consumption ^[48] 	<ul style="list-style-type: none"> • Daily ^[46] • Hourly ^[47] 	<ul style="list-style-type: none"> • Day type/Temp/Rad • Energy bills • Survey 	<ul style="list-style-type: none"> 4-6 months 3 years 45 years
k-NN	<ul style="list-style-type: none"> • Electric consumption ^[49] • Power demand ^[50] • Electric load ^[51] 	<ul style="list-style-type: none"> • Minutes-daily ^[49] • Daily ^[50] • Minutes ^[51] 	<ul style="list-style-type: none"> • Meter and meteorological variables ^[1] • Smart Meter ^{[2], [5]} 	<ul style="list-style-type: none"> 1 year 19 months

Temp: Temperature; Hum: Humidity; Rad: Radiation; Occu: Occupancy

For example, Chae [31] uses *ANN* to forecast electricity consumption in commercial buildings. The data collected for the study came from a management system; power and electricity consumption were measured at one-minute and 15-minute intervals, respectively. In addition, weather variables and operating conditions were incorporated, which requires a large data set and significant computational power [32], [33], [34], [35]. Zhang [38], using *SVR* develops an electrical load forecasting model for a university building using time series of electrical energy consumption with two types of intervals: daily and half-hourly. The information from the management system corresponds to one year of consumption. Dong [36] forecasts electricity consumption using an *SVR* algorithm for a set of commercial buildings. The input variables were the monthly electric service billing and weather data for four years [37], [38], [39], [40], [41], [42], [43]. Kaur and Ahuja [44] predict the electricity consumption of a healthcare institution with *ARIMA* models using monthly, bimonthly, and quarterly periods of historical consumption data for more than 10 years [45]. Li [46] uses two databases with electricity consumption and meteorological variables to predict electricity consumption using *Fuzzy+ANN* models. In this case, the data of the environmental variables do not correspond to the same period of the consumption data [47], [48], [49]. Finally, Valgaev [50] uses hourly meter load data to forecast the next day's load using *k-NN* models applicable to all buildings. This work aims to study the energy performance of the buildings that constitute the academic unit based on statistical analysis and energy consumption forecasting using univariate *SARIMA* models. The advantage of this technique over others is that the modeling can be built with few parameters, it does not require special personnel and equipment, nor significant computational capacity. Historical consumption data are used as predictive variables, which could facilitate the preliminary energy use analysis (PEA), or a level 1 audit [51], [52], [53].

II. METHODOLOGY

The study was carried out using statistical research methods divided into two phases. The first is the seasonal and correlation analysis, to analyze the seasonal behavior of the data using descriptive statistics that allow characterizing the data and examining the existence of patterns in the structure of the data over time to determine the seasonal component and its frequency. In addition, the relationship between the data set was primarily related to the immediate past. The second phase consisted of modeling the data as a time series using *SARIMA* processes to obtain a univariate predictive model of the series. Training and test partitions were created with the data, and to establish the partitions, the criterion followed was that the length of the test partition should not exceed 30% of the data. Otherwise, there would be a risk of not having enough information for the model training process. The two phases of the study were conducted with the statistical programs R [54] and RStudio [55].

The *SARIMA* models are derived from autoregressive and moving average models. The autoregressive models are based on the idea that the current value of the time series, X_t , can be explained as a function of a linear combination of p past values $X_t, X_{t-1}, X_{t-2}, \dots, X_{t-p}$, where p determines the number of lags needed to forecast a current value [56]. The autoregressive models of order p , *AR*(p), are expressed as equation (1)

$$X_t = \varphi_1 X_{t-1} + \varphi_2 X_{t-2} + \dots + \varphi_p X_{t-p} + \varepsilon_t \quad (1)$$

where ε_t is an error term, which is assumed to be approximately a white-noise process, and $\varphi_1, \varphi_2, \dots, \varphi_p$ are the parameters of the model, being applicable in a time series if, and only if, the series in question is stationary, that is, a time series whose properties do not depend on the time in which it is observed. On many occasions, time series present patterns that the model used for their prediction cannot represent. In this case, a known moving average process with a number q of past error terms can capture the patterns in the series. Moving average models are defined by an external information source, where the actual value of the series X_t , is determined or influenced by values from a random white noise process [56]. These moving average models of order q , *MA*(q), are defined by the equation (2)

$$X_t = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \dots + \theta_q \varepsilon_{t-q} \quad (2)$$

And where ε_t is a white-noise process of the series and $\theta_1, \theta_2, \dots, \theta_q$ are the model parameters, and like the AR models, the MA are applicable to seasonal series. On the other hand, although both AR and MA processes can be used in time series separately, the combination of both allows working with more complex time series. The combination of AR(p) and MA(q) processes is known as ARMA (p, q) processes and can be written as

$$X_t = \varphi_1 X_{t-1} + \varphi_2 X_{t-2} + \dots + \varphi_p X_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \dots + \theta_q \varepsilon_{t-q} \quad (3)$$

5

where X_t in equation (3) represents the time series, p defines the number of lags for the regression, q the number of past error terms used in the equation and $\varphi_1, \varphi_2, \dots, \varphi_p, \theta_1, \theta_2, \dots, \theta_q$ the parameters to be determined from the model. However, ARMA (p, q) models, like AR(p) and MA(q) are limited in their application to seasonal time series. To deal with the problem of the non-stationarity of a series, techniques such as the logarithmic transformation and differencing, which consists of differentiating the time series using its lags, and where parameter estimation is not required. The first differencing, for example, is represented as $\nabla X_t = X_t - X_{t-1}$, for the second differencing $(X_t - X_{t-1}) - (X_{t-1} - X_{t-2})$, which removes linear and quadratic trends from the series. In general terms, this differentiation process can be written as equation (4)

$$X_d = (X_t - X_{t-1}) - \dots - (X_{t-d+1} - X_{t-d}) \quad (4)$$

where X_d is the d differentiation of the series. These differences can be worked out through an operator known as the backward shift operator and defined in the form

$$BX_t = X_{t-1} \quad (5)$$

And that by multiplying equation (5) with itself, the second differentiation is obtained, resulting in equation (6).

$$B^2 X_t = BX_{t-1} = X_{t-2} \quad (6)$$

And that for a number d of differentiations, one has

$$B^d X_t = X_{t-d} \quad (7)$$

From equation (7), the first differentiation can be rewritten as equation (8)

$$\nabla X_t = (1 - B)X_t \quad (8)$$

And that, in general, for differentiation of order d with the operator B, we have the expression of equation (9)

$$\nabla^d = (1 - B)^d \quad (9)$$

If the operator B is applied to the process AR(p) represented by equation (1), one has

$$(1 - \varphi_1 B - \varphi_2 B^2 - \dots - \varphi_p B^p)X_t = \varepsilon_t \quad (10)$$

Or equation (10) can be simplified in the form of equation (11)

$$\varphi(B)X_t = \varepsilon_t \quad (11)$$

where

$$\varphi(B) = 1 - \varphi_1 B - \varphi_2 B^2 - \dots - \varphi_p B^p \quad (12)$$

Equation (12) is known as the autoregressive operator. A similar result can be obtained for the processes $MA(q)$, which can be written from equation (2) in the form

$$\theta(B)\varepsilon_t = X_t \quad (13)$$

6

Defining the moving average operator by means of equation (13) as

$$\theta(B) = 1 + \theta_1 B + \theta_2 B^2 + \dots + \theta_q B^q \quad (14)$$

Therefore, if the time series with which we are working is not stationary and we want to model it with $ARMA(p, q)$ processes, we add a differentiation process called the integration process. The model that arises from this integration is known as $ARIMA(p, d, q)$ [57], [58] and using the autoregressive and moving average operators of equations (12) and (14) in equation (3), equation (15) is obtained.

$$\varphi(B)(1 - B)^d X_t = \theta(B)\varepsilon_t \quad (15)$$

Nevertheless, if the time series contains seasonal variations between periods, then the series ε_t will not be white noise since it contains correlations between periods. However, a time series with a seasonal component strongly related to its seasonal lags can be modeled with an $ARIMA$ model using these lags, being represented in the form of equation (16)

$$\Phi(B^S)(1 - B^S)^D \varepsilon_t = \Theta(B^S)\omega_t \quad (16)$$

The coefficient D , represents the past seasonal degree lag of the seasonal differencing of the series, while S denotes the seasonality of the model and ω_t is a white noise process with mean zero.

$$\Phi(B^S) = 1 - \Phi_1 B^S - \Phi_2 B^{2S} - \dots - \Phi_P B^{PS} \quad (17)$$

$$\Theta(B^S) = 1 + \Theta_1 B^S + \Theta_2 B^{2S} + \dots + \Theta_Q B^{QS} \quad (18)$$

Equations (17) and (18) derived from equation (16) represent the seasonal regressor and moving average operator, respectively, with Φ_P y Θ_Q being the coefficients of the seasonal autoregressive and moving average processes, $SAR(P)$ and $SMA(Q)$, where P represents the past seasonal lags and Q are the past error terms. We denote the parameterization of these models as $SARIMA(p, d, q)(P, D, Q)$ [57] and where p and q are the parameters of the nonstationary AR and MA processes, respectively. In contrast, d and D define the degree of differencing for nonstationary and seasonal lags, respectively. Similarly, P and Q are the order of the $SAR(P)$ and $SMA(Q)$ processes for seasonal lags. By combining both models to model the time series, the general expression for a $SARIMA$ model is obtained in the form

$$\Phi(B^S)\varphi(B)(1 - B)^d(1 - B^S)^D X_t = \theta(B)\Theta(B^S)\omega_t \quad (19)$$

Equation (19) represents the combination of seasonal and non-seasonal autoregressive and moving average models used to model the electricity consumption series of this study.

II. 1 DESCRIPTION OF THE ACADEMIC BUILDINGS

7 The facilities considered in the study are part of the professional unit of the National Polytechnic Institute located in Mexico City, in the center area of the Valley of Mexico, whose geographical coordinates are 19.5° N and 99.14° W. The unit was built more than 60 years ago, although not all the buildings were constructed simultaneously. The facilities studied have ten buildings of linear geometry with four stories and an annex building with a different purpose, where academic, research and administrative activities are carried out. The structure of the buildings is made of steel, and the walls are made of prefabricated material. Fig. 1 shows the distribution and location of all the buildings of the unit that make up the study. All of them were built at distinct stages according to academic and professional needs. In the red box, you can see nine buildings in parallel and one transverse building, the longest one. They all have the same building structure, as shown in Fig. 2. The smaller red boxes show classroom buildings, offices, and teachers' cubicles on three-stories and a four-stories foreign language teaching center. The laboratories and workshops for maintenance and miscellaneous services in the green box are single-story buildings with high walls and a laminated roof. The administrative buildings in the blue box are one-story, and only the building shown in the lower part of the same box has two stories. In the yellow box, the national library was built on three levels, a basement, and a two-levels auditorium, with glass envelopes. Finally, there are five coffee shops, one on each side of the parallel buildings. Although the unit has more buildings and various facilities, the set under study was considered because it is connected to the same electrical system. In contrast, the rest has three independent electrical supply networks.

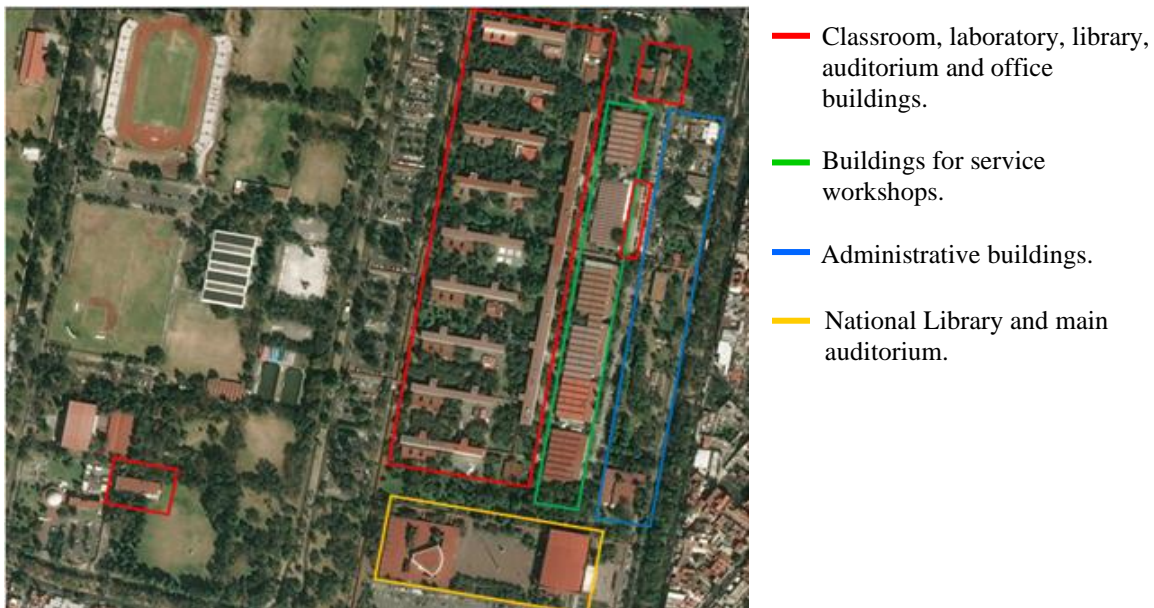


Fig. 1. Satellite image of the Unidad Profesional Adolfo López Mateos, Zacatenco.



Fig. 2. Classroom, laboratory, library, auditorium and office buildings, and offices with linear steel structure.

On the other hand, since it is an educational institution, its operation is determined by the seasonality of the academic periods, in this case, two-semester periods; one takes place from September to February, while the other from March to August. The buildings with classrooms are used Monday through Friday from 7:00 am to 10:00 pm, while the library operates Monday through Friday from 8:30 am to 8:00 pm; Saturday and Sunday from 9:00 am to 4:30 pm; the foreign language center Monday through Friday from 7:00 am to 9:00 pm; Saturday and Sunday from 7:00 am to 12:00 pm. In the case of laboratories and services, only Monday through Friday from 7:00 am to 8:00 pm. The spaces dedicated to research do not have a limited and established schedule of activities.

II.2 DESCRIPTION OF THE DATA

The professional unit is supplied with electrical energy. Other energy sources, such as gas or fuels, are of specific consumption and are not relevant to the study, so only information from the electrical energy source was collected. Electricity consumption data were obtained from the General Services Department of the Institute and came from the billing provided by the electricity supply company. These are monthly for a period of five years from 2015 to 2019, meeting the requirements for a building energy analysis; monthly billing data and covering a period of two or more years, sufficient for a level 0 or 1 audit, as established by the American Society of Heating, Refrigerating and Air-Conditioning Engineers (ASHRAE) [52]. On the other hand, information before 2015 was not considered due to events inside the Institute that forced the closure of the facilities for more than two months in 2014 and distorted the behavioral pattern of electricity consumption, producing harmful outliers for the prediction models. As shown in Fig. 3, the annual electricity consumption in the academic unit has a strong dependence on school periods and a tendency to decrease from year to year.

On the other hand, there was an irregularity in the annual consumption cycle in 2015, caused by the events mentioned in 2014 that forced changes in the school calendar. As a result, in March 2015, activities concluded, and a one-week vacation began in April, leading to a drop in consumption. When activities resumed, consumption increased until July, and in August, there was a drop in consumption due to the summer vacation period. With the start of activities in September, consumption increased until it dropped significantly in December because only the first two weeks of the year were worked.

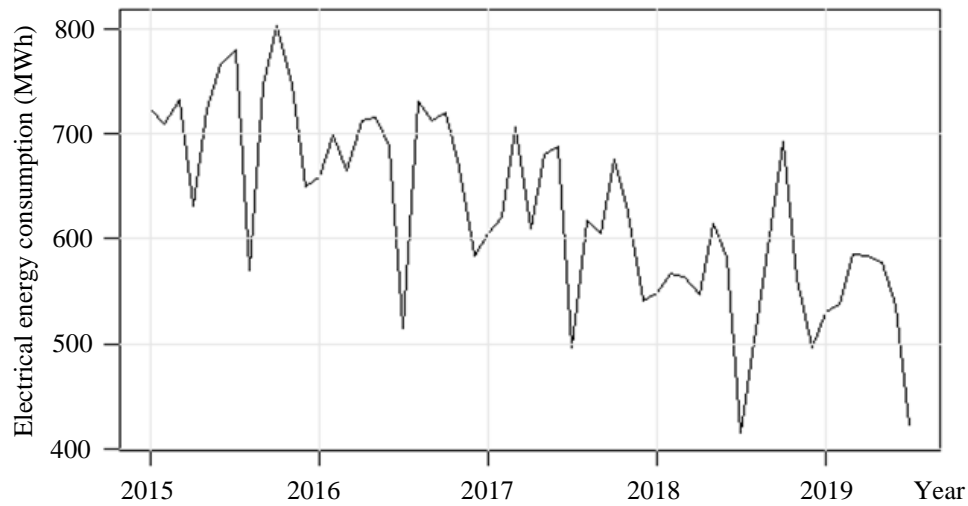


Fig. 3. Electricity consumption of the academic unit from 2015 to 2019.

It is not until the following year, 2016, that the annual cycle, due to calendar adjustments, begins to regularize, and the minimums in consumption occur in March 2016, April 2017, and April 2018 because of the one-week vacation period. Not so the case of April 2019, where the decrease in energy consumption due to the holiday period is marginal. This is because there were more academic and work activity days in April. From May to June, consumption increased due to the activities, concluding at the end of June. July saw the summer vacation period, reducing annual consumption from 2016 to 2019. After the break, by August, consumption increases until reaching its maximums, which corresponds to the beginning of activities of the second school period of the year. After that, consumption decreases until the December holiday as only the first two weeks are worked.

III. RESULTS

The results of this study were divided into three parts: seasonal analysis, autocorrelation analysis, and modeling.

III.1 Seasonal analysis

Fig. 4 shows the grouping of the frequency unit, i.e., the same month of each year. Thus, the average in each frequency group was examined, showing that, on average, monthly consumption varies except in May and June, where it is similar. On the other hand, the annual behavior during the school cycles; July and December were the months where activities are reduced; the vacation periods of April and December decrease consumption, and the peak occurs in October, the month of greatest academic, research, and administrative activity in the academic unit. The variations in the monthly frequency averages have their origin in the trend of the series. The decreasing trend presented by the series did not significantly modify the monthly difference since it decreased the series in the same proportion. In addition, the variance (standard deviation) of the monthly averages of each frequency decreased since it aligns each monthly observation closer to its frequency peers.

10

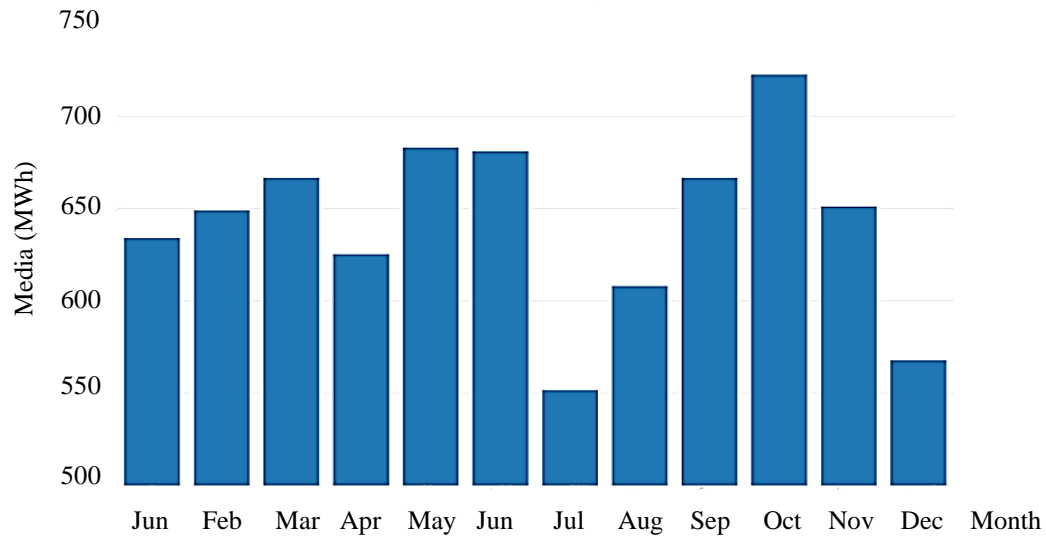


Fig. 4. Graph of the monthly average by frequency

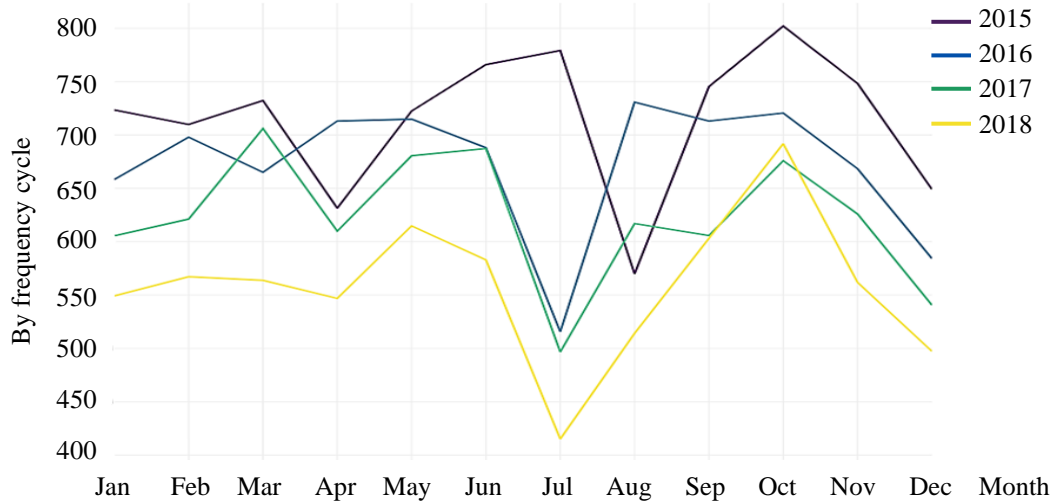


Fig. 5. Graph of consumption behavior per cycle

However, the series of monthly averages did not show how consumption behaved by month and year. By month, July and August presented significant volatility, produced by the events of 2014, Fig. 5, while the effect due to the cycle shift in the first quarter of 2015 was presented from June to August 2015, as shown in Fig. 6. The cyclical pattern is also distinguishable with the 2015 shift. This shift was the major contributor to the adjustment error due to the large dispersion of observations for July and August, as shown in Fig. 7.

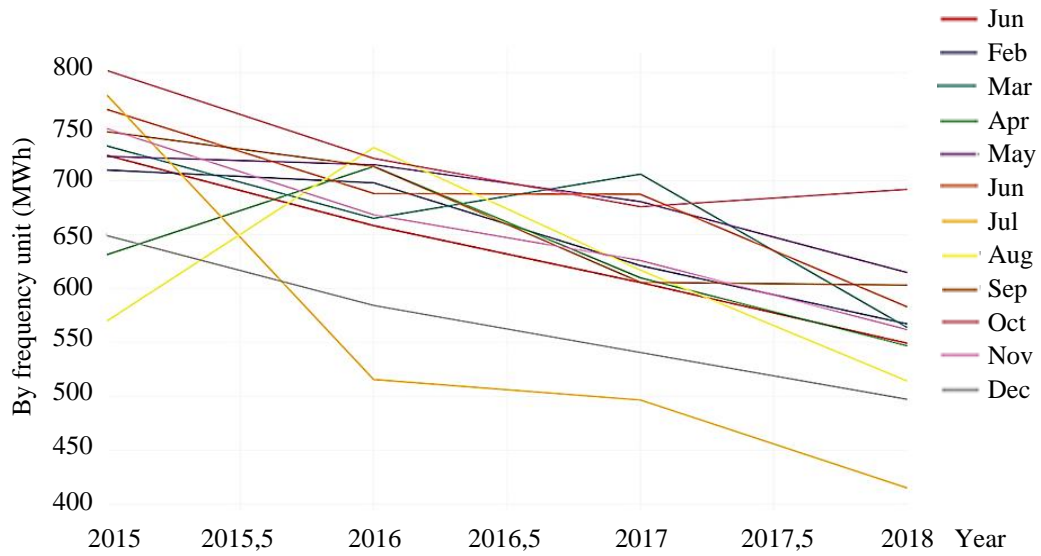


Fig. 6. Consumption behavior graph by month and cycle

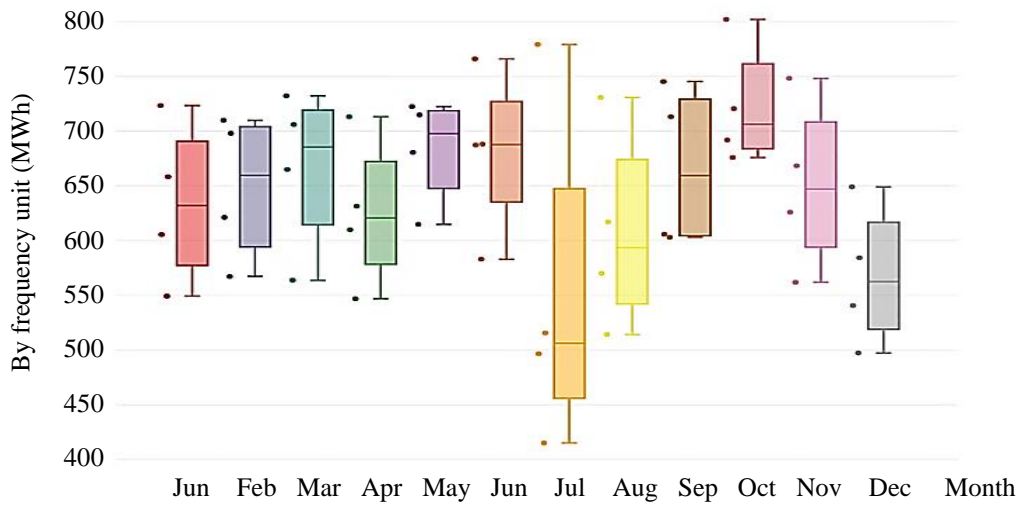


Fig.7. Box plot: Shows the dispersion of observations and median

III.2 Autocorrelation analysis

The autocorrelation function, *ACF* [56], [57], [58] is a tool for analyzing the linear dependence, stationarity, and trend of variables in a time series. The autocorrelation function measures the correlation between two variables separated by k periods, i.e., the lags of the series or distance between periods.

$$\rho_k = \text{corr}(X_t, X_{t+k}) = \frac{\text{Cov}(X_t, X_{t+k})}{\sqrt{V(X_t)}\sqrt{V(X_{t+k})}} \quad (20)$$

12

Equation (20) shows the form of the autocorrelation function in which $\text{Cov}(X_t, X_{t+k})$ is the covariance between X_t and X_{t+k} , $V(X_t)$ and $V(X_{t+k})$ the variances of X_t and X_{t+k} , respectively. Along with the *ACF* there is also the partial autocorrelation function, *PACF*, which, unlike the *ACF*, the partial autocorrelation function measures the correlation between two variables separated by k periods when the effect of other lags is removed, i.e., the dependence created by the lags existing between the two variables is not considered. Partial autocorrelation is defined in the form of equation (21)

$$\phi_k = \frac{\text{Cov}(X_t - \hat{X}_t, X_{t+k} - \hat{X}_{t+k})}{\sqrt{V(X_t - \hat{X}_t)}\sqrt{V(X_{t+k} - \hat{X}_{t+k})}} \quad (21)$$

Where the expressions \hat{X}_t and \hat{X}_{t+k} represent the regressions for X_t and X_{t+k} .

By applying this correlation analysis, it was found that the time series of the consumption data has a trend (Fig. 3), the series is not stationary. Fig. 8 shows the *ACF* correlogram for the series, where the lags 1, 2, 3, and 4 represent seasonal lags and correspond to the periods of 12, 24, 36, and 48 months. As can be seen, the first seven lags presented significant correlation, where the blue line in the graph shows the 5% critical values at $\pm 1.96n^{1/2}$ under the null hypothesis of white noise, being n the sample size. The way in which the peaks are decreasing is due to the trend of the series, while the change in the direction of the peaks demonstrates their seasonality. The *PACF* in Fig. 9 confirms the seasonal behavior of the series and its non-stationarity.

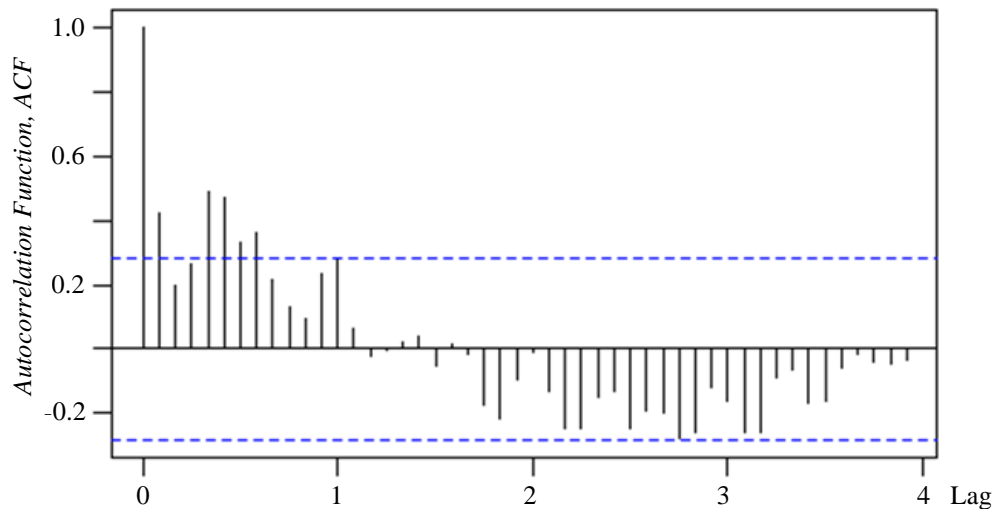


Fig. 8. ACF correlogram of electricity consumption time series.

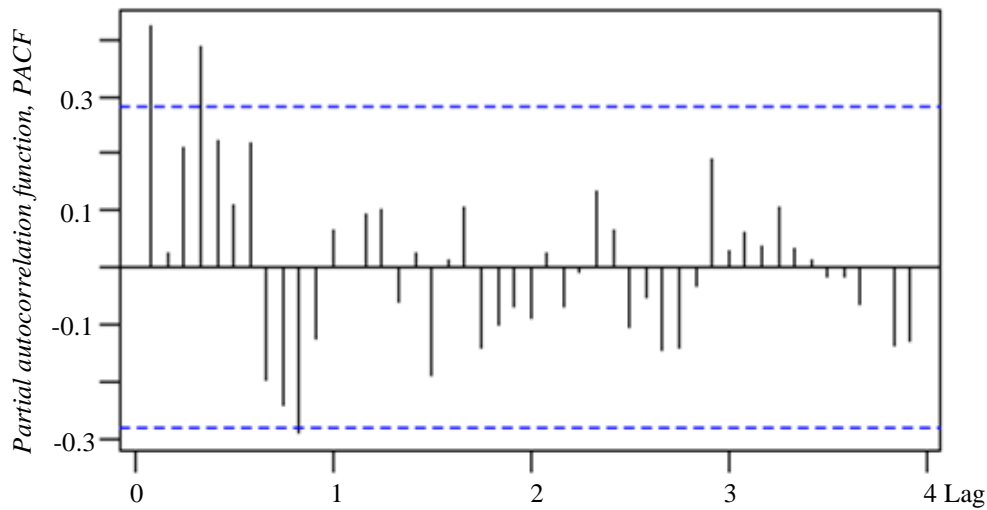


Fig. 9. PACF correlogram of the electricity consumption time series

III.3 Modeling

The modeling strategy consisted of building a forecast model from training and test partitions. First, the series of consumed electric power data underwent a differencing process, as described in the methodology, to stabilize the mean and variance. Then, the data were exponentially smoothed to minimize the impact of the irregular behavior of the first months of 2015. Subsequently, training and test partitions were created, following the criterion that the length of the test partition should not exceed 30% of the time series data. Due to the number of observations in the time series (55 months), two models were built: one with a training partition of 48 months (from January 2015 to December 2018) and a test partition of 7 months (from January to July 2019). Another was with a training partition of 43 months (from January 2015 to July 2018) and a test partition of 12 months (from August 2018 to July 2019). The procedure described above was performed using R statistical software and the RStudio platform. Once the order of the *SARIMA* model (i.e., the values of p , d , q , P , D and Q) was found, the estimation of the parameters followed, using the maximum likelihood estimation (*MLE*). This technique finds those parameter values that have the maximum likelihood of obtaining the observed data, employing the Bayesian Information Criterion, $BIC = -2\ln(\text{maximum likelihood estimate}) + z\ln(n)$, where z represents the number of model parameters and n the number of observations used in the model. The best model is obtained by minimizing the value of *BIC*.

So, among the fit models tested in the 48-month training partition, the best fitting model was *SARIMA* (3,1,1) (1,0,0)₁₂, which showed the *BIC* value = -63.77039. Given the model, the residuals were analyzed, and it was found that they have a normal distribution and the *ACF* (Fig. 10) that the lags showed no correlation. On the other hand, to test whether a series of observations in a specific period are independent, the Ljung-Box test was used. The Ljung-Box test showed a p-value of 0.1128 above 0.01 as the significance test value and thus confirmed that there is no correlation in the residuals. The results show that, even though the model worked with the training partition data, the forecast accuracy of the test data is better, as demonstrated by all the metrics. The same procedure was performed to model the 43-month training partition, and the *SARIMA* (2,1,2) (1,0,0)₁₂ model was obtained, with a *BIC* = -59.10216. Fig. 11 corresponds to the residual analysis of the model. Its distribution is normal, and the Ljung-Box test is p-value = 0.03941, which confirms that there is no correlation between lags.

14

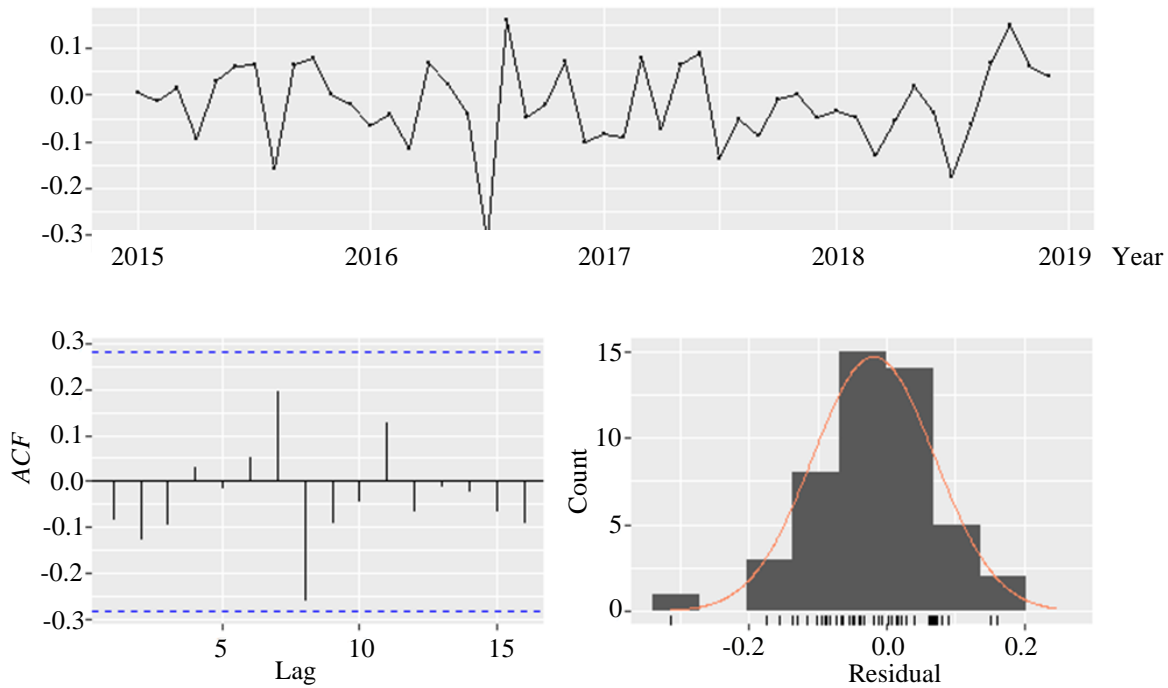


Fig. 10. Graph of the residual analysis of the 48-month partition from SARIMA (3, 1, 1) (1, 0, 0)₁₂

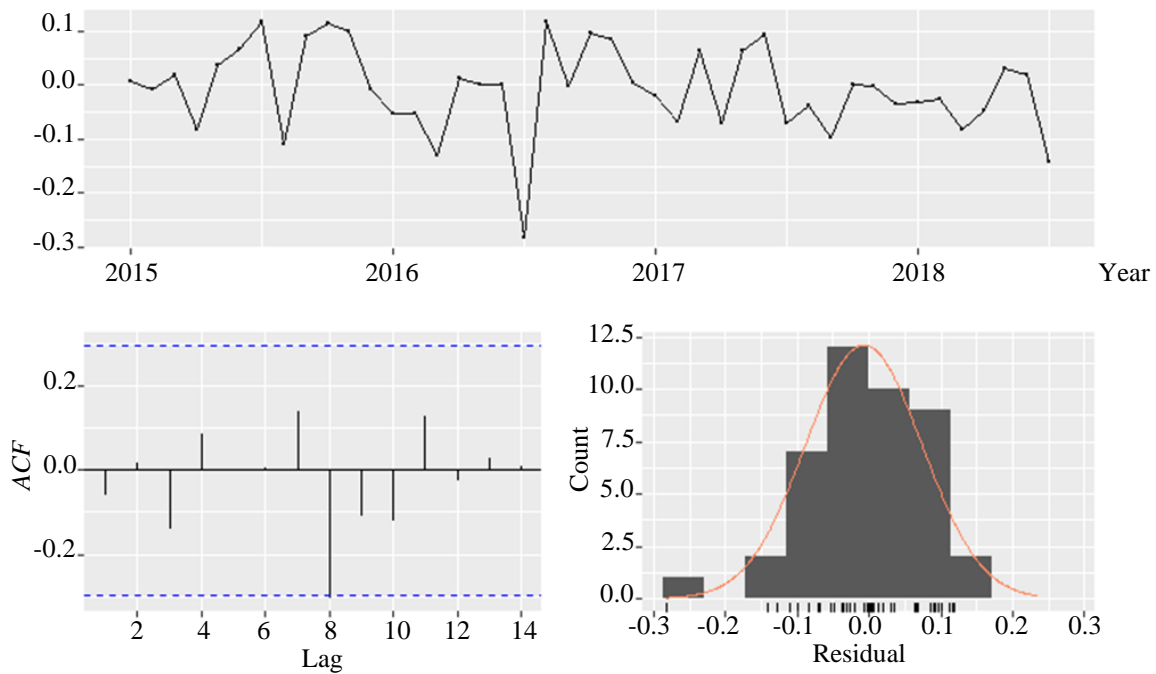


Fig. 11. Residual analysis of the 43-month training model from ARIMA (2, 1, 2) (1, 0, 0)₁₂ with shift

III.4 Prediction precision criteria

Once the models were built, their predictive capacity was analyzed, and for this purpose, the mean absolute percentage error, *MAPE*, was used, which is calculated by a term-to-term comparison of the relative error of the prediction value with respect to the real observed value, as shown in equation (22).

$$MAPE = \frac{1}{n} \sum_{t=1}^n \left| \frac{X_t - \hat{X}_t}{X_t} \right| \quad (22)$$

Where X_t are the actual values and \hat{X}_t are the predicted values, and n is the number data observations considered. Table 2 shows the error metric of both the training and test partitions for the two models. When analyzing the values in the error metric, it is observed that the fit to the training partition was better for the *SARIMA* (2,1,2) (1,0,0)₁₂ model. In comparison, the test partition was better fitted by the *SARIMA* (3,1,1) (1,0,0)₁₂ model. This result is confirmed by analyzing the plots in Fig. 12 and Fig. 13. The test values are closer to the actual values of the series for the *SARIMA* (3,1,1) (1,0,0)₁₂ model. While the fitted values for the training partition are better represented by the *SARIMA* (2,1,2) (1,0,0)₁₂.

TABLE 2. ERROR METRIC FOR THE TRAINING AND TEST PARTITIONS

Model	SARIMA (3,1,1) (1,0,0) ₁₂		SARIMA (2,1,2) (1,0,0) ₁₂	
	Training	Test	Training	Test
MAPE	1.09405146	0.9320904	0.93786799	1.62500404

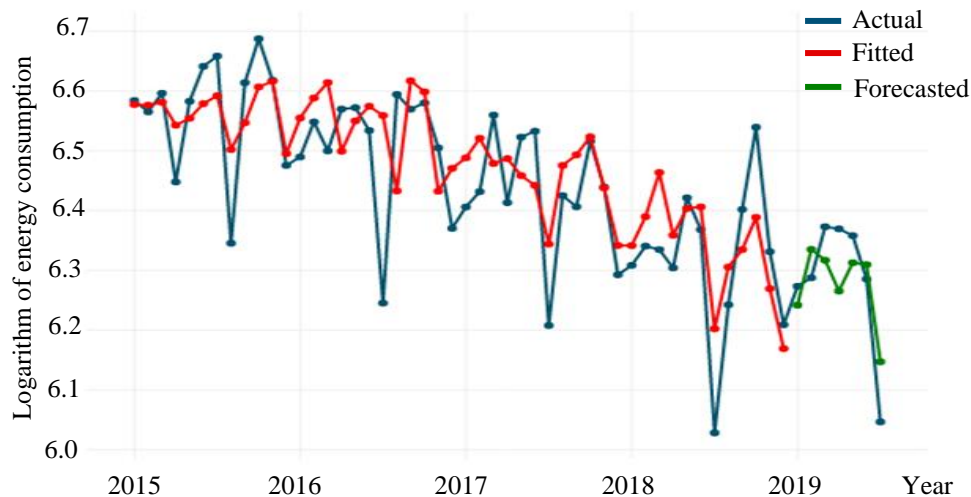


Fig. 12. Current vs. forecast values and adjustment; *SARIMA* (3,1,1) (1,0,0)₁₂

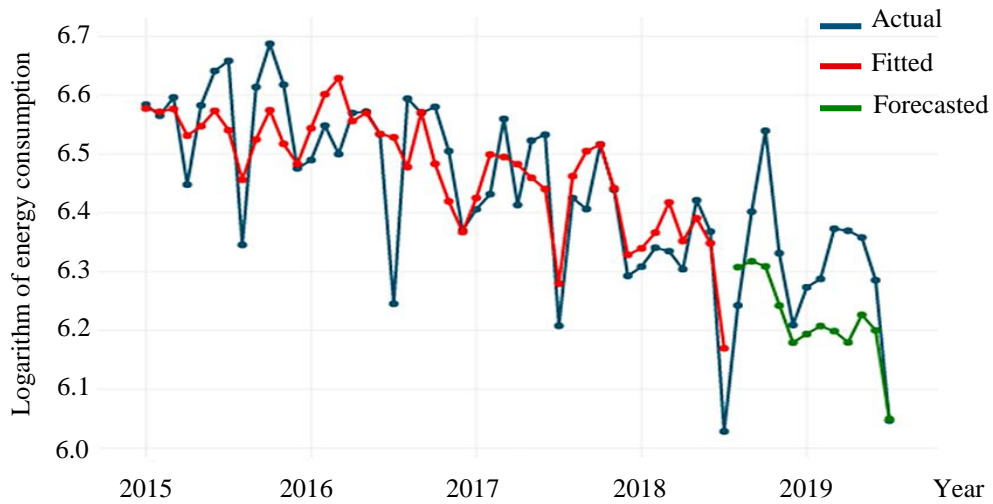


Fig. 13. Current vs. forecast values and adjustment; $SARIMA(2, 1, 2)(1, 0, 0)_{12}$

III.5 Confidence intervals

To show how accurate the forecast model is, confidence intervals were used. This is a statistical approximation method to express a range of possible values in which the observed value of the series lies with a certain degree of certainty, i.e., with a given probability. However, any percentage of probability can be used in the confidence interval. For this study, the usual gaps of 80% and 95% were considered. As can be seen in Fig. 14, the confidence intervals for the 7-month forecast of the $SARIMA(3,1,1)(1,0,0)_{12}$ model is extensive, both for the 80% and 95% levels and the same situation occurs for the $SARIMA(2,1,2)(1,0,0)_{12}$ model (Fig. 15).

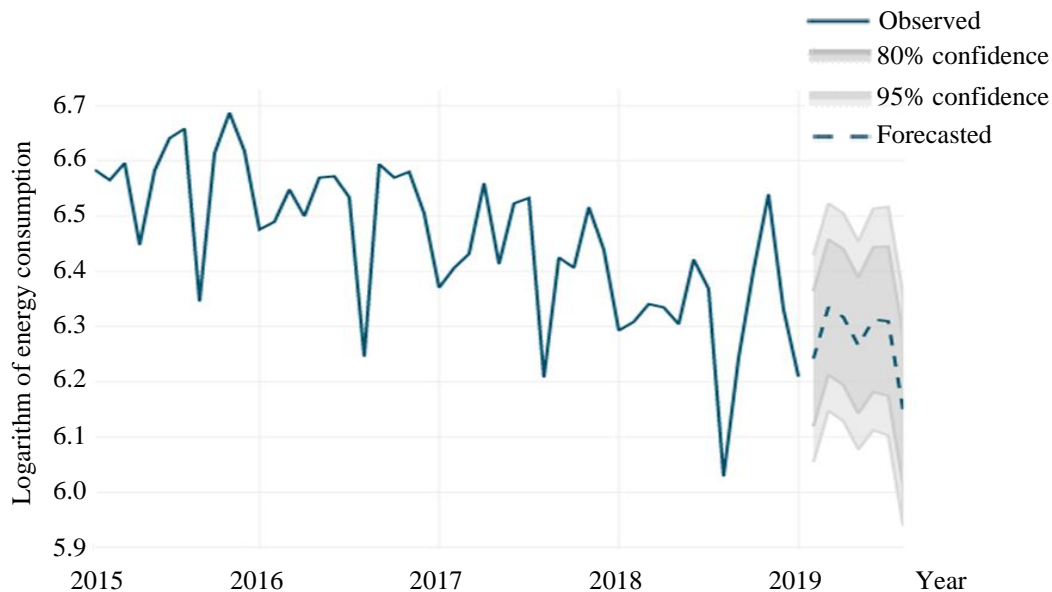


Fig. 14. $SARIMA(3,1,1)(1,0,0)_{12}$ 7-month forward consumption forecast.

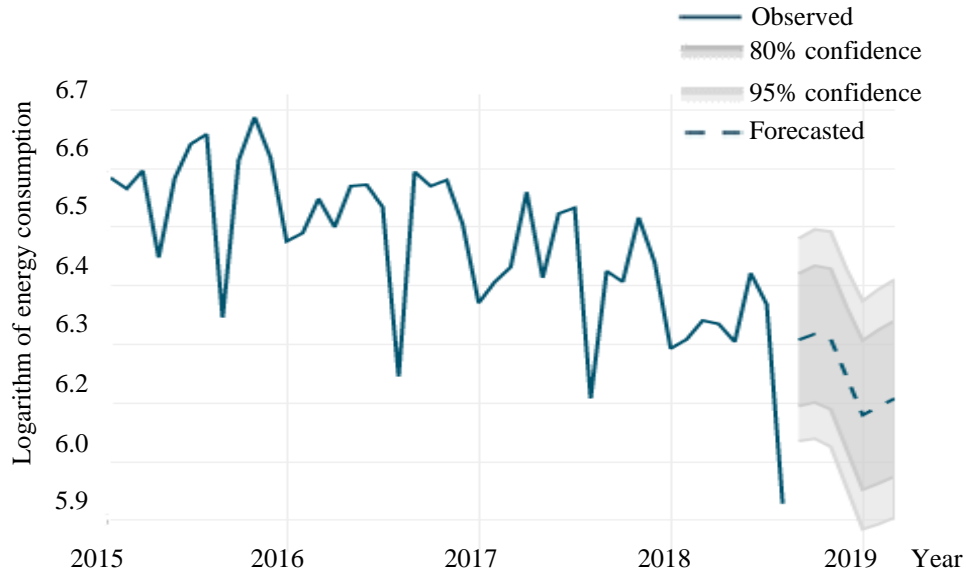


Fig. 15. SARIMA (2,1,2) (1,0,0)₁₂ 7-month forward consumption forecast

For the 12-month forecast horizon, as shown in Fig. 16, the SARIMA (3,1,1) (1,0,0)₁₂ model predicted the trend and seasonal behavior of the original series. This was not the case for the SARIMA (2,1,2) (1,0,0)₁₂ model, which showed a downward trend and did not reproduce the seasonal pattern of the original series. In Fig. 17, its confidence intervals are more extensive than the SARIMA (3,1,1) (1,0,0)₁₂ model, which means greater uncertainty in the expected value.

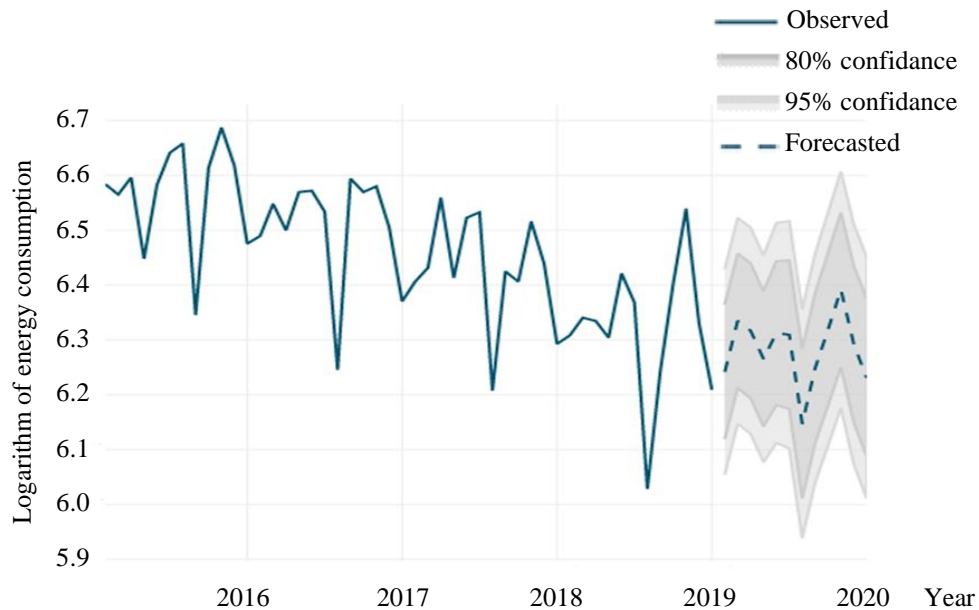


Fig. 16. SARIMA (3,1,1) (1,0,0)₁₂ 12-month forward consumption forecast.

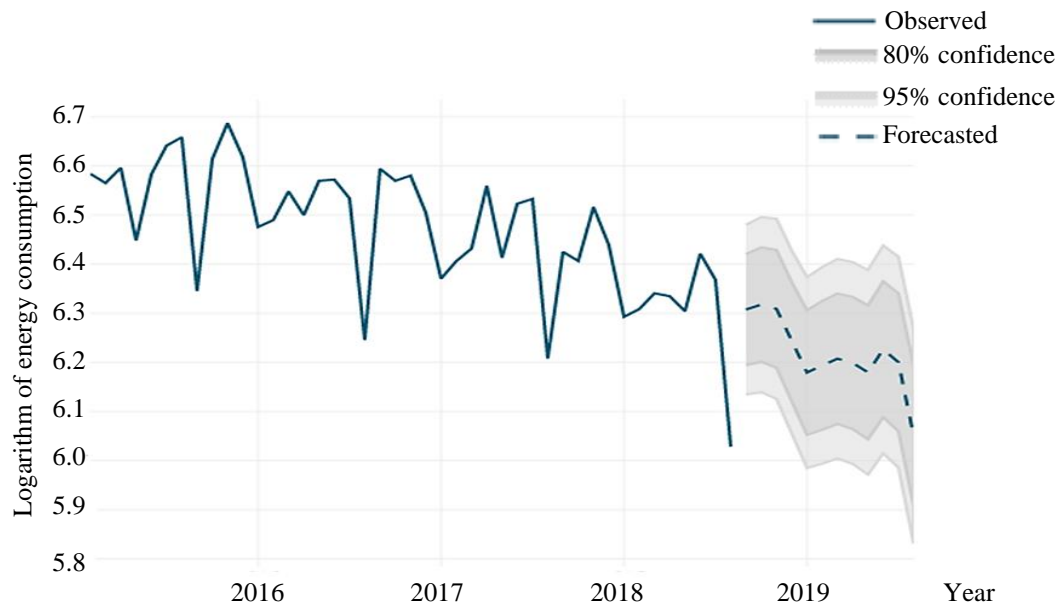


Fig. 17. SARIMA (2,1,2) (1,0,0)₁₂ 12-month forward consumption forecast

IV. DISCUSSION

Since the academic unit is in a temperate climate region, the usual use of *HVAC* systems is limited to certain spaces, and its power source is electricity, the main energy source of the buildings. There is a direct relationship between electrical energy consumption and the activities within the academic cycle, but not this trend. The monthly average values show that a seasonal cycle in consumption is maintained, while the monthly and annual trend is to decrease. In the case of 2015, there is a shift explained by the fact that the school period underwent adjustments in 2014. July and August are the months with the most significant disparity in the data because of the shift in the school period, which forced the continuity of activities where there are usually vacations. The trend presented by electricity consumption has a steeper decreasing slope from January 2015 to May 2018, indicating a steep decline in electricity consumption. Subsequently, the slope decreased; consumption declined much lower than before June 2018. Although the institution has increased its enrollment, electricity consumption has been reduced due to measures such as replacing lighting fixtures and office equipment with lower consumption. Classrooms have been fitted with sensors to control lighting, energy saving, and efficient energy use measures that the Institute implemented based on the Comprehensive Energy Diagnosis conducted by the Mexican Center for Cleaner Production. In conjunction with the National Commission for the Efficient Use of Energy, conducted during 2015 and 2016.

Despite the above, no measures respond to seasonal consumption behavior, especially when consumption is higher. One possible explanation for this behavior is that between May and June, consumption increases due to the need to conclude academic and administrative activities before the first vacation period. While for August, the increase in consumption is due to the extra administrative work that is added to the other activities since it is that month that students enter the institution for the first time. From the results obtained by the models, the SARIMA (3,1,1) (1,0,0)₁₂ model had the best fit for the test data, with the most significant minimum values being those that the model could not fit more accurately. In its 12-month forecast, it is observed that it maintains the trend of the series at a constant mean and that it represents the behavior that would be expected given that consumption could not continue to be reduced under the current operating conditions of the facilities. The SARIMA (2,1,2) (1,0,0)₁₂ model for the 43-month training partition fits the most significant minimum values better than the previous model. However, its accuracy in predicting the test values is not good. Furthermore,

when examining the 7- and 12-month forecasts, the model does not present the natural trend of the series, indicating that electricity consumption would continue its downward trend, which would be unrealistic. These results suggest that larger test partitions would improve the forecasting model, optimizing the buildings' energy audit.

V. CONCLUSIONS

19 In this study, we have worked with two approaches to analyze and forecast electrical energy consumption in educational buildings; the statistical approach and the univariate modeling with *SARIMA* processes. The univariate modeling of the time series of electricity consumption shows that, of the two best-evaluated models, *SARIMA* (3,1,1) (1,0,0,0)₁₂ best fits the real values, maintaining the seasonal behavior and the trend, which demonstrates its predictive capacity. Furthermore, in the medium-term projection of the model, it establishes that electric energy consumption will be a stationary process where its mean will be constant, which means that the trend will decay in such a way that it will cancel out. This is what would be expected in electricity consumption if the conditions of use do not change. For this model, a training partition of 48 months was used, indicating that a larger number of input data would result in a better-fitting model. However, it should be considered that as input data increases, the number of parameters to be calculated also increases.

On the other hand, from the statistical analysis, it is concluded that although there are actual values for 2015 that are presented as unusual due to the school calendar adjustment, the data do not contain outliers that could significantly affect the capacity of the predictive model. Furthermore, the monthly consumption averages project a seasonal behavior that can be used to establish electric energy efficiency strategies. For example, implementing *ASHRAE Standard 100-2018*, the months with the highest electricity consumption, May and June, have the highest daylighting, which means that artificial lighting time could be reduced through a building energy management system. Also, it would be possible to reduce lighting through devices that can vary light levels or dim when appropriate, along with implementing task lightings where needed, such as in offices and libraries. If possible, use occupancy, presence, or motion sensors in corridors and stairwells whose operation allows manual activation or turning on lighting at no more than 50% of capacity. Finally, upgrade indoor and outdoor lighting systems to provide demand response capacity to reduce lighting loads during peak electricity demand periods such as October.

The advantage of modeling with *SARIMA* processes is the ease of building and adjusting the model, making it efficient and a viable option to be implemented in building energy control and management systems. Also, to form a part of the processes for conducting energy audits that require an energy performance model. The complexity of the model and scope will depend on the needs of the audit.

ACKNOWLEDGEMENTS

This work has been possible thanks to the General Services Department of the Instituto Politécnico Nacional, which provided the information, and to the Sección de Estudios de Posgrado e Investigación of the Escuela Superior de Ingeniería Eléctrica y Mecánica of the Instituto Politécnico Nacional (SEPI-ESIME) for lending the facilities for the study. This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

REFERENCES

- [1] EIA, "International Energy Outlook 2019", *EIA*, 2019, https://www.eia.gov/outlooks/ieo/tables_side.php
- [2] IEA, "World Energy Balances 2019", *IEA*, 2019, <https://www.iea.org/data-and-statistics/data-product/world-energy-balances>
- [3] EIA, "Annual Energy Outlook 2018", *IEA*, 2018, <https://www.eia.gov/outlooks/archive/aeo18/>
- [4] CCEI, "Report on Energy Supply and Demand in Canada", *CCEI*, 2017, <https://www150.statcan.gc.ca/n1/pub/57-003-x/57-003-x2020001-eng.htm>

- [5] SENER, "Balance Nacional de Energía 2018", *SENER*, 2018, <https://www.gob.mx/sener/documentos/balance-nacional-de-energia-2018>
- [6] BP, "Statistical Review of the World Energy 2020", *BP*, 2020, <https://www.bp.com/en/global/corporate/news-and-insights/press-releases/bp-statistical-review-of-world-energy-2020-published.html>
- [7] IEA, "Energy and Climate Change 2015", *IEA*, 2015, <https://iea.blob.core.windows.net/assets/8d783513-fd22-463a-b57d-a0d8d608d86f/WEO2015SpecialReportonEnergyandClimateChange.pdf>
- [8] M. O. Fadeyi, K. Alkhaja, M. B. Sulayem, B. Abu-Hijleh, "Evaluation of indoor environmental quality conditions in elementary schools' classrooms in the United Arab Emirates", *Frontiers of Architectural Research*, vol. 3, no. 2, pp. 166-177, 2014, <https://doi.org/10.1016/j.foar.2014.03.001>
- [9] Z. Ma, P. Cooper, D. Daly, L. Ledo, "Existing building retrofits: Methodology and state-of-the-art", *Energy and buildings*, vol. 55, pp. 889-902, 2012, <https://doi.org/10.1016/j.enbuild.2012.08.018>
- [10] W. Chung, "Review of building energy-use performance benchmarking methodologies," *Applied Energy*, vol. 88, no. 5, pp. 1470-1479, 2011, <https://doi.org/10.1016/j.apenergy.2010.11.022>
- [11] T. Nikolaou, D. Kolokotsa, G. Stavrakakis, "Review on methodologies for energy benchmarking, rating and classification of buildings," *Advances in Building Energy Research*, vol. 5, no. 1, pp. 53-70, 2011, <https://doi.org/10.1080/17512549.2011.582340>
- [12] K. P. Amber, R. Ahmad, M. W. Aslam, A. Kousar, M. Usman, M. S. Khan, "Intelligent techniques for forecasting electricity consumption of buildings," *Energy*, vol. 157, pp. 886-893, 2018, <https://doi.org/10.1016/j.energy.2018.05.155>
- [13] J. Zhao, Y. Xin, D. Tong, "Energy consumption quota of public buildings based on statistical analysis," *Energy Policy*, vol. 43, pp. 362-370, 2012, <https://doi.org/10.1016/j.enpol.2012.01.015>
- [14] M. Raatikainen, J. P. Skön, K. Leiviskä, M. Kolehmainen, "Intelligent analysis of energy consumption in school buildings," *Applied energy*, vol. 165, pp. 416-429, 2016, <https://doi.org/10.1016/j.apenergy.2015.12.072>
- [15] H. Xiao, Q. Wei, Y. Jiang, "The reality and statistical distribution of energy consumption in office buildings in China," *Energy and Buildings*, vol. 50, pp. 259-265, 2012, <https://doi.org/10.1016/j.enbuild.2012.03.048>
- [16] T. Sekki, M. Airaksinen, A. Saari, "Measured energy consumption of educational buildings in a Finnish city," *Energy and Buildings*, vol. 87, pp. 105-115, 2015, <https://doi.org/10.1016/j.enbuild.2014.11.032>
- [17] A. Thewes, S. Maas, F. Scholzen, D. Waldmann, A. Zürbes, "Field study on the energy consumption of school buildings in Luxembourg," *Energy and Buildings*, vol. 68, pp. 460-470, 2014, <https://doi.org/10.1016/j.enbuild.2013.10.002>
- [18] B. Arregi, R. Garay, "Regression analysis of the energy consumption of tertiary buildings," *Energy Procedia*, vol. 122, pp. 9-14, 2017, <https://doi.org/10.1016/j.egypro.2017.07.290>
- [19] L. Brady, M. Abdellatif, "Assessment of energy consumption in existing buildings," *Energy and Buildings*, vol. 149, pp. 142-150, 2017, <https://doi.org/10.1016/j.enbuild.2017.05.051>
- [20] H. Ma, N. Du, S. Yu, W. Lu, Z. Zhang, N. Deng, C. Li, "Analysis of typical public building energy consumption in northern China," *Energy and Buildings*, vol. 136, pp. 139-150, 2017, <https://doi.org/10.1016/j.enbuild.2016.11.037>
- [21] S. S. Amiri, M. Mottahedi, S. Asadi, "Using multiple regression analysis to develop energy consumption indicators for commercial buildings in the US," *Energy and Buildings*, vol. 109, pp. 209-216, 2015, <https://doi.org/10.1016/j.enbuild.2015.09.073>
- [22] M. Mottahedi, A. Mohammadpour, S. S. Amiri, D. Riley, S. Asadi, "Multi-linear regression models to predict the annual energy consumption of an office building with different shapes," *Procedia Engineering*, vol. 118, pp. 622-629, 2015, <https://doi.org/10.1016/j.proeng.2015.08.495>
- [23] C. Deb, F. Zhang, J. Yang, S. E. Lee, K. W. Shah, "A review on time series forecasting techniques for building energy consumption," *Renewable and Sustainable Energy Reviews*, vol. 74, pp. 902-924, 2017, <https://doi.org/10.1016/j.rser.2017.02.085>
- [24] H. X. Zhao, F. Magoulès, "A review on the prediction of building energy consumption," *Renewable and Sustainable Energy Reviews*, vol. 16, no. 6, pp. 3586-3592, 2012, <https://doi.org/10.1016/j.rser.2012.02.049>
- [25] P. Chujai, N. Kerdprasop, K. Kerdprasop, "Time series analysis of household electric consumption with ARIMA and ARMA models," In *Proceedings of the International Multiconference of Engineers and*

Computer Scientists, vol. 1, pp. 295-300, 2013, http://www.iaeng.org/publication/IMECS2013/IMECS2013_pp295-300.pdf

- [26] M. Bourdeau, X. Qiang Zhai, E. Nefzaoui, X. Guo, P. Chatellier, "Modeling and forecasting building energy consumption: A review of data-driven techniques," *Sustainable Cities and Society*, vol. 48, pp. 101533, 2019, <https://doi.org/10.1016/j.scs.2019.101533>
- [27] X. Lü, T. Lu, C. J. Kibert, M. Viljanen, "Modeling and forecasting energy consumption for heterogeneous buildings using a physical–statistical approach," *Applied Energy*, vol. 144, pp. 261-275, 2015, <https://doi.org/10.1016/j.apenergy.2014.12.019>
- [28] A. S. Ahmad, M. Y. Hassan, M. P. Abdullah, H. A. Rahman, F. Hussin, H. Abdullah, R. Saidur, "A review on applications of ANN and SVM for building electrical energy consumption forecasting," *Renewable and Sustainable Energy Reviews*, vol. 33, pp. 102-109, 2014, <https://doi.org/10.1016/j.rser.2014.01.069>
- [29] D. Liu, Q. Chen, K. Mori, "Time series forecasting method of building energy consumption using support vector regression," In *2015 IEEE international conference on information and automation*, pp. 1628-1632, Ago. 2015, <https://doi.org/10.1109/ICInfA.2015.7279546>
- [30] J. Hwang, D. Suh, M. O. Otto, "Forecasting Electricity Consumption in Commercial Buildings Using a Machine Learning Approach," *Energies*, vol. 13, no. 22, pp. 5885, 2020, <https://doi.org/10.3390/en13225885>
- [31] Y. Chae, R. Horesh, Y. Hwang, Y. Lee, "Artificial neural network model for forecasting sub-hourly electricity usage in commercial buildings," *Energy and Buildings*, vol. 111, pp. 184-194, 2016, <https://doi.org/10.1016/j.enbuild.2015.11.045>
- [32] R. Mena, F. Rodríguez, M. Castilla, M. Arahál, "A prediction model based on neural networks for the energy consumption of a bioclimatic building," *Energy and Buildings*, vol. 82, pp. 142-155, 2014, <https://doi.org/10.1016/j.enbuild.2014.06.052>
- [33] C. Deb, L. Eang, J. Yang, M. Santamouris, "Forecasting diurnal cooling energy load for institutional buildings using Artificial Neural Networks," *Energy and Buildings*, vol. 121, pp. 284-297, 2016, <https://doi.org/10.1016/j.enbuild.2015.12.050>
- [34] Y. Cheng-wen, Y. Jian, "Application of ANN for the prediction of building energy consumption at different climate zones with HDD and CDD," In *2010 2nd International Conference on Future Computer and Communication*, vol. 3, pp. V3-286-289, May. 2010, <https://doi.org/10.1109/ICFCC.2010.5497626>
- [35] R. Yokoyama, T. Wakui, R. Satake, "Prediction of energy demands using neural network with model identification by global optimization," *Energy Conversion and Management*, vol. 50, no. 2, pp. 319-327, 2009, <https://doi.org/10.1016/j.enconman.2008.09.017>
- [36] B. Dong, C. Cao, L. Lee, "Applying support vector machines to predict building energy consumption in tropical region," *Energy and Buildings*, vol. 37, no. 5, pp. 545-553, 2005, <https://doi.org/10.1016/j.enbuild.2004.09.009>
- [37] R. Jain, K. Smith, P. Culligan, J. Taylor, "Forecasting energy consumption of multi-family residential buildings using support vector regression: Investigating the impact of temporal and spatial monitoring granularity on performance accuracy," *Applied Energy*, vol. 123, pp. 168-178, 2014, <https://doi.org/10.1016/j.apenergy.2014.02.057>
- [38] F. Zhang, C. Deb, S. Lee, J. Yang, K. Shah, "Time series forecasting for building energy consumption using weighted Support Vector Regression with differential evolution optimization technique," *Energy and Buildings*, vol. 126, pp. 94-103, 2016, <https://doi.org/10.1016/j.enbuild.2016.05.028>
- [39] F. Wahid, D. Kim, "A prediction approach for demand analysis of energy consumption using k-nearest neighbor in residential buildings," *International Journal of Smart Home*, vol. 10, no. 2, pp. 97-108, 2016, <https://doi.org/10.14257/ijsh.2016.10.2.10>
- [40] L. Xuemei, D. Yuyan, D. Lixing, J. Liangzhong, "Building cooling load forecasting using fuzzy support vector machine and fuzzy C-mean clustering," In *2010 international conference on computer and communication technologies in agriculture engineering*, vol. 1, pp. 438-441, Jun. 2010, <https://doi.org/10.1109/CCTAE.2010.5543577>
- [41] K. Yun, R. Luck, P. Mago, H., Cho, "Building hourly thermal load prediction using an indexed ARX model," *Energy and Buildings*, vol. 54, pp. 225-233, 2012, <https://doi.org/10.1016/j.enbuild.2012.08.007>

- [42] I. Korolija, Y. Zhang, L. Marjanovic-Halburd, V. Hanby, "Regression models for predicting UK office building energy consumption from heating and cooling demands," *Energy and Buildings*, vol. 59, pp. 214-227, 2013, <https://doi.org/10.1016/j.enbuild.2012.12.005>
- [43] Y. Zhang, Z. O'Neill, B. Dong, G. Augenbroe, "Comparisons of inverse modeling approaches for predicting building energy performance," *Building and Environment*, vol. 86, pp. 177-190, 2015, <https://doi.org/10.1016/j.buildenv.2014.12.023>
- [44] K. Jeong, C. Koo, T. Hong, "An estimation model for determining the annual energy cost budget in educational facilities using SARIMA (seasonal autoregressive integrated moving average) and ANN (artificial neural network)," *energy*, vol. 71, pp. 71-79, 2015, <https://doi.org/10.1016/j.energy.2014.04.027>
- [45] H. Kaur, S. Ahuja, "Time Series Analysis and Prediction of Electricity Consumption of Health Care Institution Using ARIMA Model," *Proceedings of Sixth International Conference on Soft Computing for Problem Solving. Advances in Intelligent Systems and Computing*, vol. 547, pp. 347-358, 2017, https://doi.org/10.1007/978-981-10-3325-4_35
- [46] K. Li, H. Su, J. Chu, "Forecasting building energy consumption using neural networks and hybrid neuro-fuzzy system: A comparative study," *Energy and Buildings*, vol. 43, no. 10, pp. 2893-2899, 2011, <https://doi.org/10.1016/j.enbuild.2011.07.010>
- [47] M. Santamouris, G. Mihalakakou, P. Patargias, et al., "Using intelligent clustering techniques to classify the energy performance of school buildings," *Energy and buildings*, vol. 39, no. 1, pp. 45-51, 2007, <https://doi.org/10.1016/j.enbuild.2006.04.018>
- [48] W. Chung, "Using the fuzzy linear regression method to benchmark the energy efficiency of commercial buildings," *Applied energy*, vol. 95, pp. 45-49, 2012, <https://doi.org/10.1016/j.apenergy.2012.01.061>
- [49] C. Fan, F. Xiao, S. Wang, "Development of prediction models for next-day building energy consumption and peak power demand using data mining techniques," *Applied Energy*, vol. 127, pp. 1-10, 2014, <https://doi.org/10.1016/J.APENERGY.2014.04.016>
- [50] O. Valgaev, F. Kupzog, "Building power demand forecasting using K-nearest neighbors' model – initial approach," *2016 IEEE PES Asia-Pacific power and energy engineering conference (APPEEC)*, pp. 1055-1060, 2016, <https://doi.org/10.1109/APPEEC.2016>
- [51] W. Ho, F. Yu, "Measurement and verification of energy performance for chiller system retrofit with k nearest neighbors regression," *Journal of Building Engineering*, vol. 46, pp. 103845, 2022, <https://doi.org/10.1016/j.jobe.2021.103845>
- [52] M. P. Deru, J. Kelsey, D. Pearson, *Procedures for commercial building energy audits*, 2nd ed. Atlanta, GA, E. U., ASHRAE, 2011, https://www.techstreet.com/ashrae/ashrae_books.html
- [53] T. Lawrence, A. K. Darwich, J. K. Means, D. Macauley, *ASHRAE green guide: Design, construction, and operation of sustainable buildings*, 5th ed. Atlanta, GA, E. U., ASHRAE, 2018, https://www.techstreet.com/ashrae/ashrae_books.html
- [54] R Foundation, *R project, 2020*, <https://www.r-project.org/>
- [55] Team RStudio, *RStudio Desktop*. Boston, MA, E. U., RStudio, 2022, <https://www.rstudio.com/>
- [56] R. H. Shumway, D. S. Stoffer, *Time Series Analysis and Its Applications with R Examples*, Fourth edition, New York, E. U., Springer Science+Business Media, 2017.
- [57] G. E. P. Box, G. M. Jenkins, G. C. Reinsel, G. M. Ljung, *Time series analysis: forecasting and control*, fifth edition, Hoboken, New Jersey, E. U., John Wiley & Sons, Inc., 2016.
- [58] C. Chatfield, H. Xing, *The Analysis of Time Series: An Introduction with R*, seventh edition, New York, E. U., Chapman & Hall, 2019.