Available online at www.sciencedirect.com
ScienceDirect
Energy Reports 8 (2022) 250–258
www.elsevier.com/locate/egyr

TMREES22-Fr, EURACA, 09 to 11 May 2022, Metz-Grand Est, France

Model of monthly electricity consumption of healthcare buildings based on climatological variables using PCA and linear regression

Ernesto Pérez-Montalvo, Manuel-Eduardo Zapata-Velásquez, Laura-María Benítez-Vázquez, Juan-Manuel Cermeño-González, Jose Alejandro-Miranda, Miguel-Ángel Martínez-Cabero, Álvaro de la Puente-Gil*

Departamento de Ingeniería Eléctrica, Sistemas y Automática, Universidad de León, Campus de Vegazana s/n, 24071, León, España

Received 14 June 2022; accepted 25 June 2022
Available online 6 July 2022

Abstract

At this time, due to the global pandemic that has occurred, public administrations want to optimize resources and reduce greenhouse gases with more interest than before. It is the case of the Energy Regional Entity of the Junta de Castilla y León (Spain) that pursues the optimization of the energy consumption in particular of healthcare sector buildings. For this purpose, this work focuses on estimating electricity consumption for each month, for which different scenarios will be generated and the corresponding model is obtained for each scenario. This model has been developed considering the historical monthly data of consumption and climatic variables for the last 3 years. Electricity consumption in public sanitary buildings is related to their climatology, due to the use of air conditioning to adjust the indoor temperature. Subsequently, from the models obtained, the results will be analyzed. Significant differences have been observed in the estimation of electricity consumption with respect to the real data provided by the Junta de Castilla y León. The results obtained show how the availability of climatic variables increases the accuracy of the model obtained by about 30%.

Peer-review under responsibility of the scientific committee of the TMREES22-Fr, EURACA, 2022.

Keywords: PCA; Data mining; Regression models; Building energy index; Smart metering

1. Introduction

The European Union is committed to establishing a sustainable, competitive, safe, and decarbonized energy system. Within the framework of action on climate and energy of the EU until the year 2030, ambitious commitments are established to continue reducing greenhouse gas emissions (at least 55% by 2030, compared to 1990). Furthermore, the Commission's proposal for the first European Climate Law aims to come true the European Green Deal: to make the European economy and society climate neutral by 2050.

* Corresponding author.
E-mail address: alvaro.puente@unileon.es (Á. de la Puente-Gil).

https://doi.org/10.1016/j.egyr.2022.06.117

Peer-review under responsibility of the scientific committee of the TMREES22-Fr, EURACA, 2022.

Achieving the target of 55% greenhouse gas emissions by 2030 will require action across all sectors. The building sector is responsible for 40% of final energy use. Because of this, this sector has a large and cost-effective potential to become more energy-efficient and reduce emissions. The goal is to reduce greenhouse gas emissions by 36% [1].

As part of the clean energy for all Europeans package, the Energy Performance of Buildings Directive (EPBD) 2010/31/EU and the Energy Efficiency Directive (EED) 2012/27/EU were revised in 2018 by the Directive (EU) 2018/844. Some of the aspects considered are the commitment of the member states to carry out long-term renovation strategies as well as the obligation to rehabilitate public buildings with 3% of their surface annually [2].

In the long-term renovation strategy, each Member State shall establish: (i) an overview of the national building stock, (ii) policies and actions to target all public buildings, and (iii) an evidence-based estimate of expected energy savings and wider benefits, establishing measurable progress indicators. In addition, it is stated that it will be mandatory to determine the energy performance of a building based on actual or calculated energy consumption and should reflect typical energy use, not only for heating, cooling, or domestic hot water but also for lighting and other technical building systems [3].

In this context, it will be necessary to establish advanced energy indexes within public buildings. Energy indexes seem: (i) to be a useful tool for monitoring energy consumption and greenhouse gas emissions; (ii) to make available information to register historical energy consumption and develop efficient energy policies in public administrations, from the local, regional and national level; (iii) quantify the real energy savings obtained derived from energy saving and efficiency measures or established in energy service contracts or energy performance contracting (EPC) [4].

EPC is a form of 'creative financing' for capital improvement which allows funding for energy upgrades from cost reductions [5]. Demonstrable energy savings are agreed upon in EPCs. This is where savings measurement and verification protocols such as the International Performance Measurement and Verification Protocol (IPMVP) developed by the efficiency valuation organizations (EVO) are of vital importance [6].

Models for calculating savings are typically developed using simple or multiple linear regressions on a building-by-building basis and independently [7]. However, when the EPC or indexes must be done for a large number of buildings, the clustering techniques would make sense. On the one hand, the buildings with similar energy behaviors or the same user would be delimited and on the other hand, it would allow the groups to be compared. The protocol proposes different methodologies to calculate energy savings. Some authors use the calibrated simulation option [8]. Other authors use the option of the whole facility verification by comparing linear regressions and neural networks [9]. The facility verification option is used when there is no building consumption data, for example, a new building. The option of a simulated model is used when multiple building improvements are made and the savings are significant [10].

Clustering techniques were used first, delimiting the buildings under study [4]. Now, in this article and for each cluster, different techniques were developed to predict the electrical consumption of public health buildings. The techniques used are principal component analysis and linear regression. Therefore, the techniques developed could help detect anomalies and serve as a support to calculate the energy savings of EPCs of public buildings.

Traditionally, energy efficiency labeling in the building sector has been carried out through energy simulations, establishing efficiency labels in comparison with the results obtained by a reference building. This methodology seems to work well with the electrical consumption of certain buildings and good results are obtained [11]. However, it does not seem to be the most appropriate in large buildings, such as hospitals, since many complex installations influence and the simulations are not sufficiently accurate [12].

Other comparative energy benchmarks that can be used in public buildings based on real electricity consumption are the UK's Display Energy Certificate (DEC) [13] and the German VDI3807 [14,15]. Other authors propose the use of the real consumption of buildings to calibrate the energy simulations carried out [16].

To calculate energy savings in building retrofitting projects, methodologies based on the IPMVP protocol are also used, based on real consumption data. Many software tools are suggested in IPMVP to analyze energy consumption by simulation and subsequent calibration with the real data; a long calculation time, the complexity and cost of implementing the model, and the uncertainty of the model parameters are the barriers to predicting the energy consumption of existing buildings. Some authors propose simplified simulation models to solve these problems [8].

Some of the applications that this research may have is to be able to use these consumption estimation profiles to propose energy efficiency measures, vehicle to grid, self-consumption installations or even to determine which buildings are energy inefficient [17,18].

Difficulties in the simulation of the power demand behavior of buildings can be overcome with access to real consumption measurements and the application of the so-called "Big Data" and "data mining" techniques.

Thus, some works were carried out to propose a first approach to identify the energy efficiency of buildings and predict their energy demand profiles [19], identifying reference electrical energy consumption profiles (in terms of final energy use) by comparing several clustering techniques [4]. Some authors propose clustering with regression techniques for the prediction of consumption in buildings [20,21]. Other studies use MARS for consumption predictions [22,23].

The choice of the algorithm must be done following some criteria such as the accuracy needed in the prediction or the volume of the available data. For example, the prediction can be done with a linear regression algorithm [24]. This is a simple manner to solve the problem. The problem can be solved with a neural network algorithm, but the volume of data must be enough to get a good solution as commented in [25].

Nowadays there is a wide range of possibilities when it comes to solving this type of problem. Many algorithms can solve the problem, but it is necessary to determine some requirements to validate the result obtained.

It can be observed that the majority percentage of annual electricity consumption belongs to the hospitals (approximately 85% of total consumption) the rest of the categories are close to 5% as shown in Table 1. As can be seen in the data shown in Table 1, the remaining categories that are not considered as hospitals represent about 16 GWh-year^{-1} of annual electricity consumption. On the other hand, the variation in the total electricity consumption, evaluated through the standard deviation, is relatively small on an annual basis, considering the evaluated period, which lasts from January 2016 to December 2019.

Table 1. Electrical energy consumption description of the Public Health System's building stock of the Castilla y León region in Spain.
Source: Junta de Castilla y León.

Building type	Inventory	Average consumption (MWh yr^{-1})	Sd. deviation (MWh yr^{-1})	Consumption share (%)
Hospitals	25	100 256 453	1 268 194	86%
Health centers without emergencies	146	8 817 623	189 674	8%
Health centers with emergencies	66	6 088 346	164 626	5%
Others	20	1 550 965	63 596	1%
Total	257	116 713 388	1 686 092	100%

Finally, let it is noted that the classification of buildings provided is valid for administrative purposes, but inefficient for energy analysis. Therefore, the research conducted on these buildings has been used to be able to properly use the data from an energy point of view, which may differ from a purely administrative classification.

As shown in the image Fig. 1, we carry out a step-by-step process to arrive at the model for estimating the energy consumption of the different health buildings, which is detailed as follows:

1. Data collection: we have conducted an analysis of the various data sources and collected the data from the different sources that are needed for the model.

2. Data Set: We perform a data cleaning and refinement process for our data set using statistical analysis. At this point, we also perform Principal Component Analysis (PCA) to check if it is possible to achieve a reduction in the number of variables. Then we generate the different scenarios that we are going to compare with our estimate.

3. Machine learning: with the Data Set of our different scenarios, we carry out our estimation model with the Linear Regression technique.

4. Prediction Result: After having the different results of the estimates for each result, we validate from which we obtain our best models with a shared analysis and pressure indicators, to obtain our conclusion.

2. Materials and methods

The above data flow explains how the work was carried out. Starting from data collection to the selection of an estimation model of the energy consumption of buildings within the healthcare sector. In the first step, the data collection, all the data sources to be used are requested. In the second step, we begin to compose the data set and refine it, this is done by gathering all the sources and doing the cleaning process, and then performing a Principal Components Analysis (PCA) to check if it is possible to reduce the number of variables. For each of the proposed scenarios, the data is normalized and pre-processed to obtain the cleanest and most orderly data set possible and subsequently obtain the linear regression models for the estimation of monthly consumption. Finally, each of the models that were obtained to estimate the energy consumption is checked against the real value to verify its behavior.

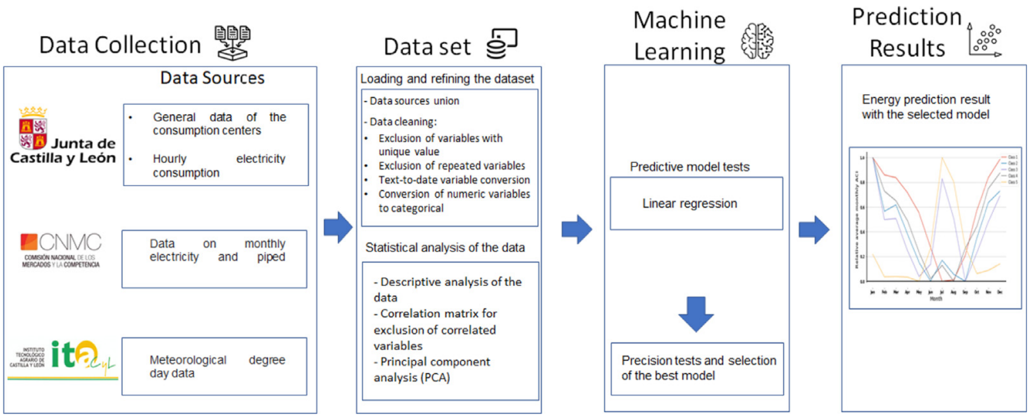


Fig. 1. Data flow forecast hospital energy.

2.1. Scenarios

By applying the linear regression technique, we intend to compare how having different variables influences a monthly estimation analysis. Having multiple different variables (monthly electricity consumption, climatic variables, clustering, etc.) with different sources, we decided to divide our data into three scenarios, to conclude which data behave better in our models. In each scenario, the set of variables used is indicated.

Each scenario includes the complete data for each cluster and the data for each cluster separately used in the previous research [4]:

Scenario 1: uses the dataset with all-climate variables (Temperature, humidity, velocity, direction, radiation and precipitation), as well as date variables, electricity consumption, and building floor area.

Scenario 2: uses the dataset with the results of the application of the PCA technique on the climatic variables mentioned above, as well as the date variables, electricity consumption, and building area.

Scenario 3: uses the data set with the variables gd_20 and gd_26, as well as the date variables, electricity consumption, and building area.

2.2. Linear regression

Multiple regression models are used to estimate the monthly consumption of electricity. This method allows us to estimate the value of the dependent variable from a set of independent variables. The model is represented by the following equation [26]:

$$Y_i = (\beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \cdots + \beta_n X_{ni}) + e_i, \tag{1}$$

where Y_i is the electricity consumption dependent variable, X_{ni} are the independent variables, β_0 is the intercept, β_n are the regression coefficients and e_i is the error.

The simple linear regression model shows as a result a line that passes as close as possible to all points in the point cloud. In the case of the Multiple regression model, the result is more complex, it is a space with as many dimensions as independent variables there are [27].

To apply the Multiple Linear Regression Model it is necessary to ensure that the data meet different requirements (Non-collinearity, linear relationship between the numerical predictors and the dependent variable, normal distribution of residuals, homoscedasticity).

The proposed linear regression method has been used previously in different investigations to estimate and predict the electricity consumption of a building [22].

2.3. Statistical test

Different statistical tests are used to validate the models and estimate their accuracy.

• R-squared: statistical measure that represents the proportion of the variance for a dependent variable that is explained by an independent variable or variables in a regression model. This method is used in [28,29].

$$R^2 = 1 - \frac{\sum_{t=1}^{T} \left(\widehat{Y}_t - \underline{Y}\right)^2}{\sum_{t=1}^{T} \left(Y_t - \underline{Y}\right)^2}, \tag{2}$$

where $\sum_{t=1}^{T} \left(\widehat{Y}_t - \underline{Y}\right)^2$ is the sum of residual squares and $\sum_{t=1}^{T} \left(Y_t - \underline{Y}\right)^2$ is the sum of total squares.

• Root Mean Squared Error (RMSE): standard deviation of the residuals (prediction errors). Residuals are a measure of how far from the regression line data points are. This accuracy method is used in [8].

$$RMSE = \sqrt{\frac{1}{n} \sum \left(y_j - \widehat{y_j}\right)^2}, \tag{3}$$

where n is the number of samples, y_j are the predictions and $\widehat{y_j}$ are the actual observations

• Mean absolute error (MAE): measures the average magnitude of the errors in a set of predictions, without considering their direction. It is the average over the test sample of the absolute differences between prediction and actual observation where all individual differences have equal weight [30].

$$MAE = \frac{1}{n} \cdot \sum \left| y_j - \widehat{y_j} \right|, \tag{4}$$

where n is the number of samples, y_j are the predictions and $\widehat{y_j}$ are the actual observations

3. Results and analysis

3.1. Correlation and principal components analysis

A correlation matrix is made to discard variables from our data set that are very similar in their behavior. The objective is to avoid duplicate contributions to the estimation model. Thus, variables with a correlation higher than 0.8 in the matrix will be excluded, since above this value they are highly correlated, as indicated by several studies and books [31], and may overestimate the model. Then, it is necessary to determine the optimal number of principal components for our dataset, which will tell us the percentage of explained variance by each number of principal components. This will give the number of principal components where the rise in the percentage of explained variance starts to reduce.

The principal component analysis is performed on the climatic variables to determine which ones provide information in the model and to compare the results with other scenarios that analyze other climatic variables.

Based on Fig. 2 we keep five principal components since the rise in the percentage of explained variance between taking 5 and 6 Principal Components is not substantial enough. With 5 principal components it already has a variance higher than 90% as shown in figure a.

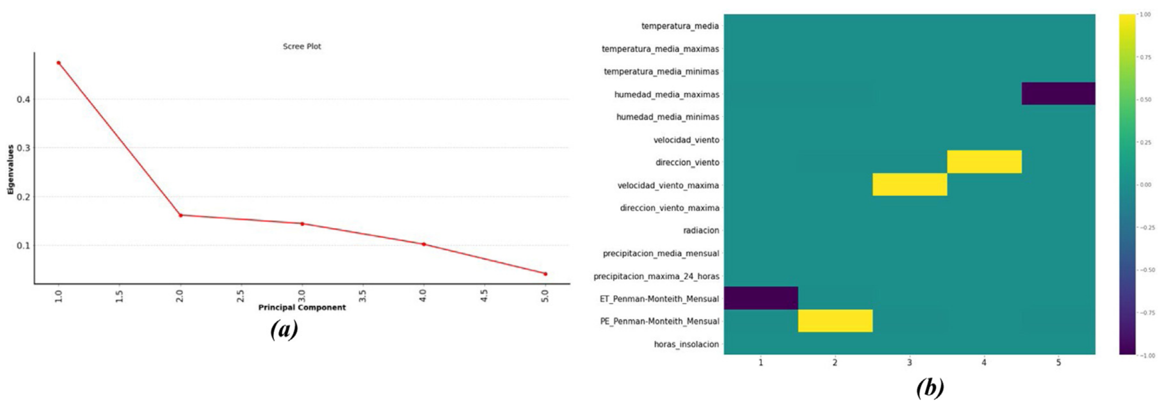


Fig. 2. (a) PCA scree plot. (b) PC coefficients graph with varimax rotation.

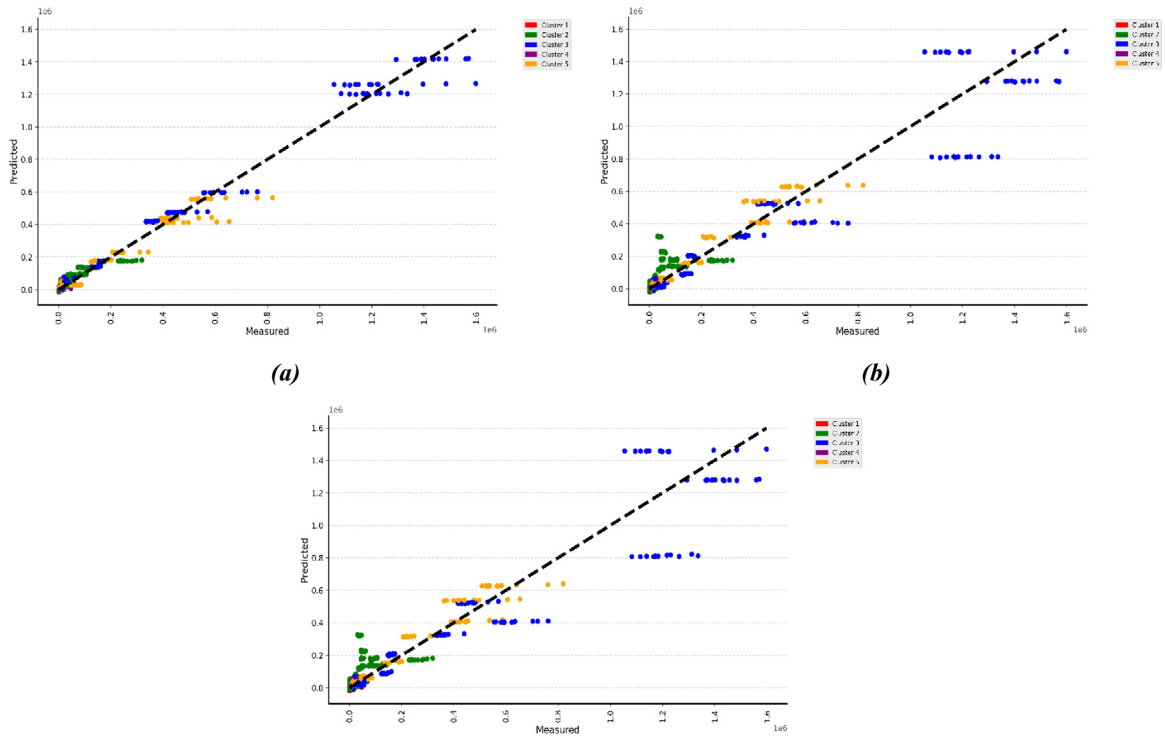
After applying Varimax rotation on our PCA is obtain the simplified graph of Fig. 2 b to determine the principal influence of each of the principal components on the variables analyzed [32].

Based on the plot after applying the rotation is reach the following conclusions:

1. Principal Component 1 is explained based on the reference evapotranspiration and temperature.
2. Principal Component 2 is explained based on potential evapotranspiration.
3. Principal Component 3 is explained based on the maximum wind speed.
4. Principal Component 4 is explained based on the direction of the wind.
5. Principal Component 5 is explained based on the maximum mean humidity variable.

3.2. Linear regression results

Fig. 3(a) shows how the estimated record fits the regression line. It can be seen visually how the points fit better as consumption is lower, but the relative errors are lower in the buildings with higher consumption. The buildings that best fit the regression line belong to groups 1 and 4. Fig. 3(b) shows how much accurate is the estimation of the consumption for scenario 2.



(a) *(b)* *(c)*

Fig. 3. Regression graph for every building [kWh month^{-1}], linear regression method. (a) scenario 1, (b) scenario 2, (c) scenario 3.

The distribution is considerably similar to scenario 1. This means that the use of Principal Component Analysis does not imply an improvement in the results of Linear Regression, since by using a smaller set of variables the results will be less accurate. This loss of precision may be admissible when the set of variables used in the model is reduced from 32 to 5 variables. Finally, the above Fig. 3(c) shows the estimation for scenario 3. Which is the most accurate for the different groupings of buildings.

Although visually Fig. 3(a, b, and c) are similar. The results obtained in Table 2, show differences in the adjustments obtained from the selected indicators RMSE, MAE, and R^2.

Table 2. Statistic estimators result for the reference electric energy consumption profiles linear regression method (scenarios 1, 2 and 3).

	Cluster	R^2	RMSE (kWh yr^{-1})	MAE (kWh yr^{-1})	Samples
Scenario 1	Class 1	0.61	1 362	1 046	4
	Class 2	0.52	22 517	6 856	97
	Class 3	0.94	74 121	30 587	57
	Class 4	0.74	3 551	1 992	46
	Class 5	0.95	25 091	12 927	53
	Global	0.91	46 242	17 701	257
Scenario 2	Class 1	0.49	2 030	1 616	4
	Class 2	0.51	36 420	14 928	97
	Class 3	0.94	430 332	189 171	57
	Class 4	0.74	7 280	4 287	46
	Class 5	0.95	170 442	76 511	53
	Global	0.91	46 356	17 677	257
Scenario 3	Class 1	0.52	1 322	1 124	4
	Class 2	0.48	18 932	5 950	97
	Class 3	0.93	77 482	29 914	57
	Class 4	0.69	2 707	1 652	46
	Class 5	0.95	25 775	12 762	53
	Global	0.91	45 919	17 669	257

Table 2 shows the numerical results for scenarios 1, 2 and 3. It can be highlighted that the cluster with the lowest R-squared is class 2. Comparing class 2 with the best class, which is class 5, there is a 45% difference between their r-squared values, in this case, the value of RMSE and MAE is considerably different among the classes. This can be explained since there is a lot of variation in the consumption of the different classes. For that reason, it is better to analyze the results with a percentual indicator as r-squared. Table 2 shows the results for scenario 2. For this scenario, the lowest value of R-squared is again for class 2 and the best for class 5, the difference among both r-squared values is 46%. Analyzing scenario 2 r-squared results against r-squared scenario 1 results, it can be highlighted that the r-squared class 1 value is 19% lower in scenario 2 as can be seen in Table 3. The other r-squared values are almost equal in both scenarios. Comparing the MAE and RSME values between scenarios 1 and 2, it can be concluded that MAE and RSME values are higher in scenario 2 than in scenario 1 for all the classes. This means that results for MAE and RSME are better in scenario 1.

Table 3. Comparison of the R^2 estimator for the reference electric energy consumption (scenarios 1, 2 and 3).

	Comparative scenarios 1 and 3		Comparative scenarios 1 and 2		Comparative scenarios 2 and 3	
	R^2	% R^2	R^2	% R^2	R^2	% R^2
Class 1	0.09	14.75%	0.12	19.67%	−0.03	−6.12%
Class 2	0.04	7.69%	0.01	1.92%	0.03	5.88%
Class 3	0.01	1.06%	0	0%	0.01	1.06%
Class 4	0.05	6.76%	0	0%	0.05	6.76%
Class 5	0	0%	0	0%	0	0%

Finally, Table 2 shows the results for scenario 3. As in the other, two scenarios the lowest value of R-squared is for class 2 and the best for class 5, the variation, in this case, is the same as in scenario 1 is 45%. Analyzing scenario 3 RSME and MAE results against scenario 1, it depends on the class which scenario is better than the other, but it can be concluded that there is no substantial difference between them. The models obtained have adequate accuracy values in general terms (R2 above 90%), improving the previous results in which the climatic variables were not available (above 30%).

4. Conclusions

After having analyzed the three-monthly estimation models obtained through linear regression with the different scenarios. The following conclusions can be observed:

The buildings belonging to class 1 would lead us to conclude that scenario 2 is the one in which the lowest accuracy is obtained, but concluding only by analyzing this class is incorrect, since, as can be seen, in general terms they are not very different in any of the different scenarios.

The model obtained in scenario 1, where all the existing climate variables are included, shows higher precision values than in the other two models (scenarios 2 and 3).

In addition, to determine whether it is better to perform a principal component analysis of the climatic variables or to use two indexes (degree days in base 20 and base 26), it can be verified that the RSME and MAE indexes are lower in scenario 3, so it can be concluded that to estimate the electricity consumption of the buildings it is better to use degree days in base 20 and base 26 than to perform a principal component analysis of the climatic variables obtained from meteorological stations.

The values obtained in the quality indexes by means of the linear regression model can be visualized as valid for these investigations, they are like those obtained by other researchers in their investigations when trying to estimate the value of the energy consumption of a building, improving by more than 30% the previous investigations.

From the results obtained, it is possible for early detection of anomalies in electricity consumption, evaluations, and energy efficiency, as well as to determine the cost and evaluate the possibility of centralized energy purchasing.

CRediT authorship contribution statement

Ernesto Pérez-Montalvo: Investigation, Formal analysis, Software, Writing – review & editing. **Manuel-Eduardo Zapata-Velásquez:** Investigation, Formal analysis, Software, Writing – review & editing. **Laura-María Benítez-Vázquez:** Investigation, Formal analysis, Writing – review & editing. **Juan-Manuel Cermeño-González:** Investigation, Formal analysis, Writing – review & editing. **Jose Alejandro-Miranda:** Investigation, Formal analysis, Writing – review & editing. **Miguel-Ángel Martínez-Cabero:** Validation, Writing – original draft, Investigation, Conceptualization, Supervision, Writing – review & editing. **Álvaro de la Puente-Gil:** Conceptualization, Methodology, Validation, Investigation, Software, Formal analysis, Validation, Writing – original draft, Writing – review & editing, Supervision.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The data used are available on another platform: https://analisis.datosabiertos.jcyl.es/explore/?sort=modified.

References

[1] European Commission, Ed. Energy roadmap 2050. Luxembourg: Publications Office of the European Union; 2012.

[2] Directiva (UE) 2018/844 del parlamento europeo y del consejo, de 30 de mayo de 2018, por la que se modifica la directiva 2010/31/UE relativa a la eficiencia energética de los edificios y la directiva 2012/27/UE relativa a la eficiencia energética, p. 17.

[3] Papadopoulos S, Bonczak B, Kontokosta CE. Pattern recognition in building energy performance over time using energy benchmarking data. Appl Energy 2018;221:576–86. http://dx.doi.org/10.1016/j.apenergy.2018.03.079.

[4] De La Puente Gil Á, González Martínez A, Borge Diez D, Martínez Cabero MÁ, de Simón Martín M. True power consumption labeling and mapping of the health system of the Castilla y León region in Spain by clustering techniques. Energy Procedia.

[5] T E3P. Energy performance contracting. 2013.

[6] Tanguay D. International performance measurement and verification protocol (IPMVP). Efficiency Valuation Organization (EVO); 2021, https://evo-world.org/en/products-services-mainmenu-en/protocols/ipmvp (accedido 14 de julio de 2021).

[7] Bianco V, Manca O, Nardini S. Linear regression models to forecast electricity consumption in Italy. Energy Sources B 2013;8(1):86–93. http://dx.doi.org/10.1080/15567240903289549.

[8] Piccinini A, Hajdukiewicz M, Keane MM. A novel reduced order model technology framework to support the estimation of the energy savings in building retrofits. Energy Build 2021;244:110896. http://dx.doi.org/10.1016/j.enbuild.2021.110896.

[9] Ye KK, Demirezen G, Fung AS, Janssen E. The use of artificial neural networks (ANN) in the prediction of energy consumption of air-source heat pump in retrofit residential housing. IOP Conf Ser Earth Environ Sci 2020;463:012165. http://dx.doi.org/10.1088/1755-1315/463/1/012165.

[10] Panda S, et al. Residential demand side management model, optimization and future perspective: A review. Energy Rep 2022;8:3727–66. http://dx.doi.org/10.1016/j.egyr.2022.02.300.

[11] Zorita AL, Fernández-Temprano MA, García-Escudero L-A, Duque-Perez O. A statistical modeling approach to detect anomalies in energetic efficiency of buildings. Energy Build 2016;110:377–86. http://dx.doi.org/10.1016/j.enbuild.2015.11.005.
[12] Kaur H, Ahuja S. Time series analysis and prediction of electricity consumption of health care institution using ARIMA model. In: Proceedings of sixth international conference on soft computing for problem solving. Singapore; 2017, p. 347–58. http://dx.doi.org/10.1007/978-981-10-3325-4_35.
[13] Hong S-M, Paterson G, Burman E, Steadman P, Mumovic D. A comparative study of benchmarking approaches for non-domestic buildings: Part 1 – Top-down approach. Int J Sustain Built Environ 2013;2(2):119–30. http://dx.doi.org/10.1016/j.ijsbe.2014.04.001.
[14] VDI - Verein Deutscher Ingenieure. Verbrauchskennwerte für gebäude, grunldlangen. characteristic consumption values for buildings. Fundamentals. VDI 3807. Blatt 1/Part 1. 2013.
[15] VDI - Verein Deutscher Ingenieure. Verbrauchskennwerte für gebäude. Verbrauchskennwerte für t heizenergie, strom und wasser. Characteristic consumption values for buildings. Characteristic heating-energy, electrical-energy and water consumption values. VDI 3807. Blatt 2/Part 2. 2014.
[16] Maile T, Bazjanac V, Fischer M. A method to compare simulated and measured data to assess building energy performance. Build Environ 2012;56:241–51. http://dx.doi.org/10.1016/j.buildenv.2012.03.012.
[17] de Simón-Martín M, de la Puente-Gil Á, Juan Blanes-Peiró J, Bracco S, Delfino F, Piazza G. Smart charging of electric vehicles to minimize renewable power curtailment in polygeneration prosumer buildings. In: 2020 fifteenth international conference on ecological vehicles and renewable energies (EVER). 2020, p. 1–8. http://dx.doi.org/10.1109/EVER48776.2020.9243112.
[18] De la Puente-Gil Á, González-Martínez A, Borge-Diez D, Blanes-Peiró JJ, De Simón-Martín M. Electrical consumption profile clusterization: Spanish Castilla y León regional health services building stock as a case study. Environments 2018;5(12):12. http://dx.doi.org/10.3390/environments5120133.
[19] Kim Y, Son H-G, Kim S. Short term electricity load forecasting for institutional buildings. Energy Rep 2019;5:1270–80. http://dx.doi.org/10.1016/j.egyr.2019.08.086.
[20] Chung W. Using the fuzzy linear regression method to benchmark the energy efficiency of commercial buildings. Appl Energy 2012;95:45–9. http://dx.doi.org/10.1016/j.apenergy.2012.01.061.
[21] Hsu D. Comparison of integrated clustering methods for accurate and stable prediction of building energy consumption data. Appl Energy 2015;160:153–63. http://dx.doi.org/10.1016/j.apenergy.2015.08.126.
[22] Ignatiadis D, Henri G, Rajagopal R. Forecasting residential monthly electricity consumption using smart meter data. In: 2020 fifteenth international conference on ecological vehicles and renewable energies. 2019, p. 1–6. http://dx.doi.org/10.1109/NAPS46351.2019.9000285.
[23] Lazzari F, et al. User behaviour models to forecast electricity consumption of residential customers based on smart metering data. Energy Rep 2022;8:3680–91. http://dx.doi.org/10.1016/j.egyr.2022.02.260.
[24] Kim MK, Kim Y-S, Srebric J. Predictions of electricity consumption in a campus building using occupant rates and weather elements with sensitivity analysis: Artificial neural network vs. linear regression. Sustain Cities Soc 2020;62:102385. http://dx.doi.org/10.1016/j.scs.2020.102385.
[25] Ruiz LGB, Cuéllar MP, Calvo-Flores MD, Jiménez MDCP. An application of non-linear autoregressive neural networks to predict energy consumption in public buildings. Energies 2016;9(9):684. http://dx.doi.org/10.3390/en9090684.
[26] Weisberg S. Applied linear regression. John Wiley & Sons; 2005.
[27] Bianco V, Manca O, Nardini S. Electricity consumption forecasting in Italy using linear regression models. Energy 2009;34(9):1413–21. http://dx.doi.org/10.1016/j.energy.2009.06.034.
[28] Mirasgedis S, et al. Models for mid-term electricity demand forecasting incorporating weather influences. Energy 2006;31(2):208–27. http://dx.doi.org/10.1016/j.energy.2005.02.016.
[29] Leung PCM, Lee EWM. Estimation of electrical power consumption in subway station design by intelligent approach. Appl Energy 2013;101:634–43. http://dx.doi.org/10.1016/j.apenergy.2012.07.017.
[30] Sekhar Roy S, Roy R, Balas VE. Estimating heating load in buildings using multivariate adaptive regression splines, extreme learning machine, a hybrid model of MARS and ELM. Renew Sustain Energy Rev 2018;82:4256–68. http://dx.doi.org/10.1016/j.rser.2017.05.249.
[31] Sarah Guido AC. Introduction to machine learning with python: a guide for data scientists.
[32] von Storch H, Zwiers FW. Statistical analysis in climate research. Cambridge University Press; 1999.