Energy and Buildings 49 (2012) 380–387

Contents lists available at SciVerse ScienceDirect

Energy and Buildings

journal homepage: www.elsevier.com/locate/enbuild

Multiple regression models to predict the annual energy consumption in the Spanish banking sector

Alfonso Aranda*, Germán Ferreira, M.D. Mainar-Toledo, Sabina Scarpellini, Eva Llera Sastresa

CIRCE – Centre of Research for Energy Resources and Consumption – University of Zaragoza, Mariano Esquillor Gómez, 15 – 50018 Zaragoza, Spain

ARTICLE INFO

Article history:
Received 6 December 2011
Received in revised form 7 February 2012
Accepted 24 February 2012

Keywords:
Banking sector
Multiple regression model
Energy consumption prediction

ABSTRACT

This paper presents a regression analysis of energy consumption in the banking sector. In our case study, the target area is the Spanish banking sector, for which we divide the available data into a prediction and a validation subset. Power models were developed using test data from 55 banks. From the analysis, three models were obtained; where the first proposed model can be used to predict the energy consumption of the whole banking sector, while the rest of the models estimate the energy consumption for branches with low winter climate severity (Model 2) and high winter climate severity (Model 3). Models 2 and 3 differ from the first model in that they need independent variables measured in situ. As a result, the uncertainty of the response variable in the function of the independent variables is reduced by 56.8% for the first model and by 65.2% and 68.5% for the second and third proposed models, respectively. The validation of the first model, which is the model with the lowest determination coefficient, shows that this model is appropriate for predicting the energy consumption of bank branches with good energy consumption performance and detecting inefficiencies in bank branches with poor energy consumption performance.

1. Introduction

Energy consumption and demand trends show that for the 27 members of the European Union (EU 27) greater energy efficiency will be required by 2020. The most recent projections, using 2009 as the reference, show that primary energy consumption will decrease to 70,170,768 TJ (1676 Mtoe) by 2020 (6,950,088 TJ [166 Mtoe] less than the reference year 2005), which is 8,457,336 TJ (202 Mtoe) more than the 2020 objective of achieving a 20% reduction in energy consumption [1]. The total annual energy consumption in the EU 27 increased continuously until 2007, with a total of 48,470,583.6 TJ (1157.7 Mtoe), and the service sector was responsible for 39.5% of the energy consumption, with 11,915,632.8 TJ (284.6 Mtoe) for households, 1,332,480.257 TJ (27.8 Mtoe) for agriculture and 6,079,233.6 TJ (145.2 Mtoe) for services, etc. After the economic downturn that started in 2008, there was a reduction in economic activity and, consequently, energy consumption, but there was also a slowdown on the progression towards energy efficiency [2]. The economic recovery period will see an increase in equipment refurbishing that will allow for progress in energy efficiency to take place.

In 2007, the service sector accounted for nearly 10% of the total final energy consumption in Spain – 408,799.152 GJ (9764 ktoe) – [3], with the largest consumers being in the office sector, and an increasing trend is expected in the coming years. Greenhouse gases (GHG) emitted by the commercial and institutional sector were equivalent to approximately 8.2 million tonnes of CO_2 [4]. Consequently, energy savings in this sector offer the best means of reducing the energy demand. Several European countries, led by the European Commission, are currently interested in improvements in energy efficiency in buildings and in reducing carbon dioxide equivalent emissions (CO_2 eq). As result of Directive 2002/91/EC in the last few years, intense development is taking place in Spain with the intention of reducing carbon dioxide emissions. Since the enactment in November 2007 of the Royal Decree 47/2007, of 19th January, approving the Basic Procedure to certify energy efficiency in new-construction buildings and in certain retrofitting buildings (in which more than 1000 m^2 or more than the 25% of the building envelope is refurbished), buildings that undergo this Royal Decree must be qualified in terms of energy efficiency at the project level for the work to be completed. Directive 2010/31/EU regarding the energy performance of buildings (recast) from 19 May 2010 repeals the current directive that requires the certification of buildings that are for rent or sale. Consequently, a new Royal Decree is expected in the near future. Additionally, recent initiatives focusing on both the improvement of energy efficiency as well as the provision of renewable energy sources in the building sector by Energy Service Companies (ESCOs) are expected to have a positive effect on these issues in the short term [5].

* Corresponding author.
E-mail address: alaranda@unizar.es (A. Aranda).

0378-7788/$ – see front matter © 2012 Elsevier B.V. All rights reserved.
doi:10.1016/j.enbuild.2012.02.040

Nomenclature

Latin alphabet

AGE	age
ANOVA	analysis of variance
ATMs	automated teller machines
CO_2 eq	carbon dioxide emissions equivalent
CI	confidence interval
Cp	Mallows' statistic
CTE	building technical code, Código Técnico de la Edificación
E	mathematical expectation
EIF	energy inefficiency factor
ENERCON	annual energy consumption
ESCOs	energy services companies
F-test	*F*-test statistic
GHG	greenhouse gases
GLASUR	glazed surface in the façade
H_0	null hypothesis
HEIGHT	office height
HVAC	heating, ventilation and air conditioning
NCASH	number of automated teller machines
NEMP	number of employees
p	number of independent variables plus one, which corresponds to the intercept
p-value	*p*-value statistic
r	Pearson's correlation coefficient
R^2	determination coefficient
R^2 (adj)	adjusted determination coefficient
s	number of coefficients
S	residuals standard deviation
SE	standard error
SURF	office surface area
SUSEV	summer climatic severity
t-test	Student's *t*-test statistic
Var	variance
VIF	variance inflation factor
WCSEV	winter climatic severity
x_i, x_j	independent variable, predictor, predictive variable
X	independent variable matrix
Y	dependent variable, response variable

Greek symbols

α	confidence level
β	estimates of the regression coefficients
ε, ε_a, ε_b	random errors or perturbations
σ^2	population variance

In spite of significant trends and interest of the research community in the energy performance improvement of the service sector, there is a lack of information regarding the specific service sectors that have dedicated research efforts to study their energy consumption in more detail [6–13]. G.N. Spyropoulos and C.A. Balaras reported that among the different office subcategories, banks and other financial offices are the most energy intensive in the US, but similar data for European buildings has not been published for the banking sector [14].

Modelling techniques that predict the energy performance of buildings have been used, including multiple regression methods, artificial neural networks, decision trees or Fourier series models [15–19]. In this paper a regression model has been selected to find a compromise between the simplicity of the evaluation method and the accuracy in the result without requiring a considerable amount of input data and simulation energy [17]. Multiple regression is used frequently in research; the present work aims to identify explicative variables to develop a model in which the chosen variables influence the response and the variables that do not contribute relevant information are rejected.

The objectives of this study are to develop a regression model that determines how efficient or inefficient a bank branch is in terms of energy consumption, depending on its construction characteristics, climatic area and energy performance, by predicting its annual energy consumption. Furthermore the energy requirements for heating and cooling demand of bank branches are supplied only by electricity. The mathematical model permits researchers to predict the energy consumption without widespread analysis, and the model is validated and used to detect energy-inefficient bank branches in Spain and to propose energy saving measures that could reduce energy consumption. The results provide relevant information on the energy performance of the Spanish banking sector and contribute new data for the energy performance of the service sector.

2. Methodology

An inference analysis was developed to obtain three multiple regression models for the prediction of the annual energy consumption in the Spanish banking sector. The aim was to first obtain a model that will serve as a pre-diagnostic tool for energy performance in the bank branches analysed, and the model was based on easy-to-obtain variables, precluding the necessity of a walkthrough audit. Additionally, alternative models were calculated to obtain a more accurate energy performance based on independent variables that represent features that need an in situ measurement or more detailed information than is collected in a walkthrough energy audit finding a balance between accuracy and feasibility in obtaining the predictors.

Relevant independent variables that define the energy consumption were selected to develop models that were obtained by means of regression models, which were validated and discussed for future predictions and reproducibility [20]. Regression coefficients were estimated using the least squares method. This method estimates the regression coefficients by minimising the sum of the squares of the deviations to the proposed regression model [21]. A regression equation is as shown:

$$\hat{Y} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_p x_p \quad (1)$$

where $\hat{Y}$ is the fitted value and $\beta_0, \beta_1, \ldots,$ and β_p are the estimations of the regression parameters.

The real value for *Y* is:

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_p x_p + \varepsilon \quad (2)$$

where ε is the random error [22].

β_0, β_1, ..., β_p describe the expected change in the predicted variable *Y* in response to a unitary change in x_i when the rest of predictors remain constant [23].

It is not recommended to predict the response variable for a set of values for predictors that are out of the range of data used for the regression equation obtained, which would lead to an extrapolation error [24]. The graphical and regression analysis were performed using Minitab [25] and SPSS [26].

2.1. Sampling

Fifty-five bank branches were selected from the 12 total climatic areas across Spain. Using the Minitab software, the sample size was obtained according to the 1 sample *Z*-test method outlined by Douglas C. Montgomery [22] for a standard deviation of 7224 for the population analysed, for a confidence level of 90% and an acceptable

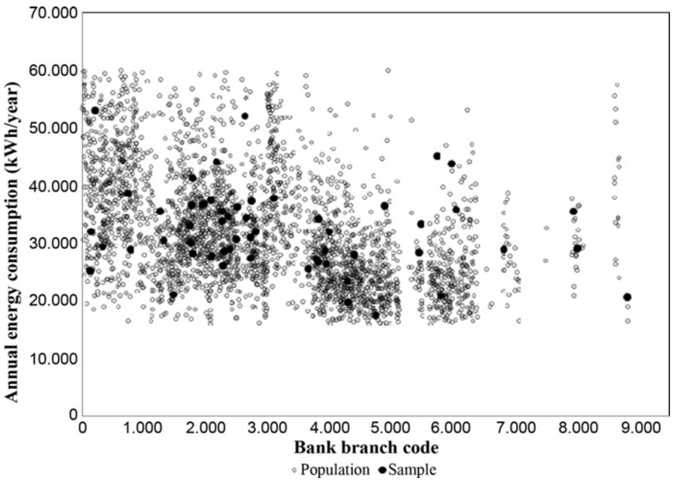

Fig. 1. Scatter plot for the annual energy consumption (kWh/year).

energy consumption error of 10% (3207 kWh/year) of the population mean (32,075 kWh/year) which is similar to the sample mean (32,100 kWh/year). The bank branches were equally represented in terms of climate area, number of employees in the branch location and the annual energy consumption (ranges from 16,000 to 60,000 kWh/year, which corresponds with the value interval that characterise the total population being studied, see Fig. 1).

For the analysis, a two-phase audit was carried out from June 2010 to May 2011 to obtain the energy performance along the four seasons, through the following steps: (a) preliminary data gathered by questionnaire sent via to the company responsible for the maintenance and (b) an in situ walkthrough audit with data gathering and measurement. An energy efficiency analysis of both energy consuming systems such as lighting, HVAC, office equipment and automated teller machines (ATMs) and employers behaviour was performed. The questionnaires consisted of questions about contracted power, total annual energy consumption (ENERCON), number of employees (NEMP), office surface area (SURF) and storage room surface area. The rest of data were gathered during the walkthrough audit. Measurements done were: lighting level by means of a luxmeter, lighting system energy demand by a power analyser, bank branch energy consumption by a power analyser, ATMs energy consumption by a power analyser, glazed surface through a distance meter, ceiling height through a distance meter and a quality measurement of energy losses through an IR thermographic camera. The rest of variables such as equipment power and electrical characteristics were obtained from the registers facilitated by the maintenance person in charge.

All variables are quantitative where continuous and discrete type are included. The data collected corresponding to the variables being studied are shown in Table 1, which summarises their name and measurement units. The dependent variable to consider and predict is ENERCON which is the total annual sum of the lighting, heating and air conditioning, equipment energy consumption and other consumption, such as ATMs and office equipment for the reference year 2010. The winter and summer climate severities (WCSEV and SUSEV) quantify the climatic conditions of every climatic area in the winter and summer, respectively. These variables are the result of a combination of the degree days and the solar radiation for the location under study. They are calculated from the Basic Document HE of Energy Saving, section HE-1 (Building Technical Code, Código Técnico de la Edificación, CTE).[1] The WCSEV variable is divided into five ranges of severity codified from A to E, where A is the lowest and E the highest severity. For this study the data considered is the mean value of every range (see Table 2). The SUSEV variable is divided into four ranges, 1–4, from lowest to highest severity, and data are analysed in the same manner as WCSEV, as shown in Table 3.

Table 1
Sample variables.

Variable	Name	Measurement unit
Annual energy consumption	ENERCON	kWh/year
Winter climatic severity	WCSEV	Dimensionless
Summer climatic severity	SUSEV	Dimensionless
Office surface area	SURF	m^2
Number of employees	NEMP	
Office height	HEIGHT	m
Number of ATMs	NCASH	
Energy inefficiency factor	EIF	Dimensionless
Age	AGE	year
Glazed surface in the façade	GLASUR	m^2
HVAC installed power	HVAC	kW

Regarding the office geometrical configuration, the SURF variable does not include the storage room, bunker or any other spaces that are not air conditioned. The office height (HEIGHT) is the height of the room in the air conditioned space, the energy inefficiency factor (EIF) is a function of lighting, heating and air conditioning energy-saving measures that are not applied in the bank branches. Therefore the bank branch that has not implemented any energy-saving measure has an EIF value of 1; otherwise the energy-saving measure percentage is calculated taking into account the bank branch annual energy consumption before and after applying the saving measures. The final value is deducted from the 1 value corresponding to an inefficient bank branch EIF; the age (AGE) corresponds to the period between the construction phase of new buildings or from the last refurbishment of existing buildings to the year 2010.

2.2. Sample analysis

Variables were analysed using the scatter plot tool to detect linear, quadratic, logarithmic and other relationships between variables [22,27] as well as possible outliers. In this phase, the variables were also analysed such that trends could be found, as the sample was to be divided into subsamples that explain the energy consumption better.

Subsequently, cluster analysis was used to forecast predictive variables [28] in order to create groups of variables that were not known initially, and the number of these variables was reduced by grouping them into new families of variables with common features, given the high similarity in the correlation coefficient [29]. This method follows a hierarchical agglomeration process that starts with all the variables separately, and in the first step, the two nearest variables are grouped together, and in the following step the next variable or variables are combined with the last two into a different cluster. This process continues until all the clusters are grouped into one cluster. Additionally, analysing the predictive variables through a correlation analysis is important to avoid dependencies between these variables, which may not result in a correct prediction model.

2.3. Multiple regression analysis

As previously mentioned, the multiple linear regression analysis was adopted to estimate alternative models for predicting the

[1] The Código Técnico de la Edificación (CTE) is the regulatory framework that establishes the basic requirements of safety and habitability to be fulfilled by buildings established by Law 38/1999 of November 5, Ordenación de la Edificación, Management of Construction Planning (LOE).

Table 2
Winter climate severity (WCSEV) according to the CTE and considered for the study.

A	B	C	D	E
WCSEV ≤ 0.3	0.3 < WCSEV ≤ 0.6	0.6 < WCSEV ≤ 0.95	0.95 < WCSEV ≤ 1.3	WCSEV > 1.3
WCSEV = 0.15	WCSEV = 0.45	WCSEV = 0.775	WCSEV = 1.125	WCSEV = 1.45

energy consumption by the banking sector in Spain. Linear models have to fulfil a series of basic hypotheses regarding the random error of the model to achieve a regression coefficient estimation with appropriate statistical features. The expectation is that the random perturbation should be null ($E(\varepsilon) = 0$), the perturbations should be homoscedastic ($\mathrm{Var}(\varepsilon) = \sigma^2$ = constant) and all the perturbations should be uncorrelated ($r(\varepsilon_a, \varepsilon_b) = 0$). For the distributional hypothesis, the perturbation has to follow a normal distribution, which is an essential feature if the purposed model is to be predictive. The explicative variables should be linearly independent, and the amount of data has to be larger than the number of explicative variables in the model. The residues will allow for the testing of the structural and statistical hypothesis because the perturbation is not observable. To determine the correct model, it is necessary to have at our disposal a measure of comparison between the models by means of hypothesis proofs regarding the model parameters of interest and the adjustment measures, such as correlation coefficients, coefficient of determination, confidence and prediction intervals [30] as well as outliers, because of outliers may influence on the regression model obtained through the least squares method. The process of obtaining a regression model is as follows:

a) Coefficient estimation: assuming the existence of $(XTX)^{-1}$ where $\beta = (XTX)^{-1}XT$ where matrix X corresponds to $\{x_1, x_2, \ldots, x_p\}$, XT denotes the transpose matrix of X and β corresponds to $\{\beta_0, \beta_1, \beta_2, \ldots, \beta_p\}$.
b) Inference of the coefficients: for the contrast hypothesis, H_0: $\beta_i = 0$, the p-value of each contrast should be lower than 0.05, and in this case, if the null hypothesis is rejected, all the coefficients are significantly different from 0.
c) Assessment of the confidence intervals for the different regression coefficients: $\mathrm{IC}_{\beta i,1-\alpha/2s}$, where s is the number of coefficients when the confidence interval is evaluated for several predictors.
d) Analysis of variance (ANOVA).
e) An F test for the regression: this permits contrast if there is a linear relationship between Y and the predictors by means of the contrast H_0: $\beta_i = 0$.
f) Coefficient of determination: R^2.
g) Estimation and prediction of the mean response.

The selection of the correct model was obtained using the Minitab tool 'Best Subsets' and 'Stepwise'. Values of the Mallows' Cp statistic for each regression model should be evaluated by function p, where p is the number of independent variables plus one, which corresponds to the intercept. The selected regression model should have a low Cp value and should be close to the number of independent variables to be estimated in the model to avoid a considerable difference between the expected value and the real value of the dependent variable Y [31]. Additionally, the selected model should have the lowest possible residual deviation (S) combined with the highest R^2 (adj) value.

Table 3
Summer climate severity (SUSEV) according to the CTE and considered for the study.

1	2	3	4
SUSEV ≤ 0.6	0.6 < SUSEV ≤ 0.9	0.9 < SUSEV ≤ 1.25	SUSEV > 1.25
SUSEV = 0.3	SUSEV = 0.75	SUSEV = 1.075	SUSEV = 1.4

2.4. Model validation

The validation of the first model was carried out by analysing 4732 new random bank branches and comparing the difference between the results obtained with the model regression as a function of the variables SUSEV, number of ATMs (NCASH), SURF and the actual energy consumption of the branches. The branches were distributed across Spain and its 12 climatic areas based on the input data requirements for the regression analysis. The energy consumption, along with the rest of independent variables, was obtained by sending questionnaires to 10,000 bank branches, of which 5772 responded; some responses had to be filtered because of missing information or irrational values given for the independent variables used as inputs in the prediction of the regression model or for values of the dependent variables. After filtering, 4732 bank branches were analysed for the validation of the regression model. These data were inserted into the regression model to obtain the estimated energy consumption to compare with the real energy consumption and to calculate the deviation, which reflects the difference between the real annual energy consumption in kWh/year for the reference year 2010 and the statistical average of the annual energy consumption for a bank branch with the same characteristics and that is located in the same climatic area. When the difference is positive, energy saving potential exists.

The branches with a deviation higher than the 10% set by the accepted random error were examined further, and 6 bank branches were randomly selected to be tested in more detail to clarify the causes of the energy deviation. This was performed with a walkthrough audit, in which the structural characteristics, the lighting and heating, ventilation and air conditioning (HVAC) systems, as well as other energy consuming devices, such as office equipment and ATMs, were checked.

The audit results were used to indicate the energy inefficiency or energy performances of each bank branch. Along with the bank branches a set of energy efficiency measures and good practices were selected for each location to reduce their annual energy consumption and increase their energy efficiency.

3. Results and discussion

3.1. Sample analysis results

Non-linear relationships between variables have not been detected, in consequence the posterior results are based on lineal dependences. In this way the WCSEV and SUSEV variables are highly negatively correlated (Table 4). The Pearson's correlation coefficient, r, measures the linear association between two variables (the extent to which a variable changes with another). The r value is always between −1 and 1 [30]. However, a correlation does not imply a cause–effect relationship; for instance, a strong

Table 4
Correlation analysis between the independent variables.

	WCSEV	SURF	NCASH	ENERCON
SUSEV	−0.664 (0.002)			
ENERCON		0.529 (0.000)	0.333 (0.013)	
NEMP		0.401 (0.002)		0.280 (0.038)

Cells contain Pearson's correlation (P-value).

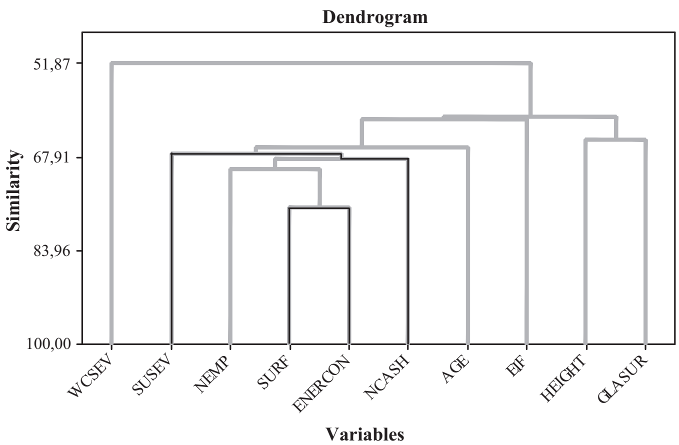

Fig. 2. Dendrogram of the variable cluster analysis.

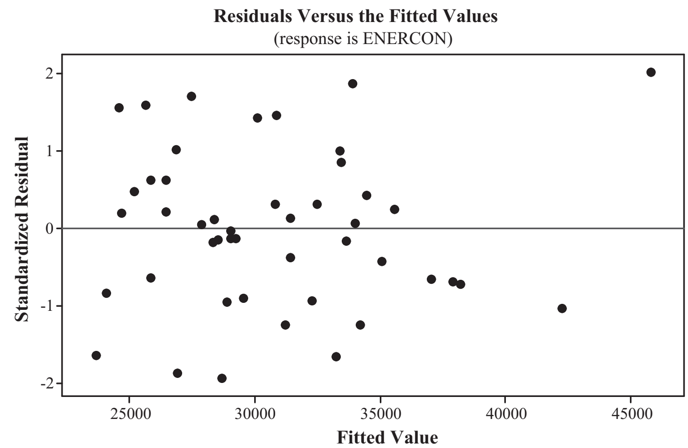

Fig. 3. Scatter plot of the residuals versus the fitted value for Model 1.

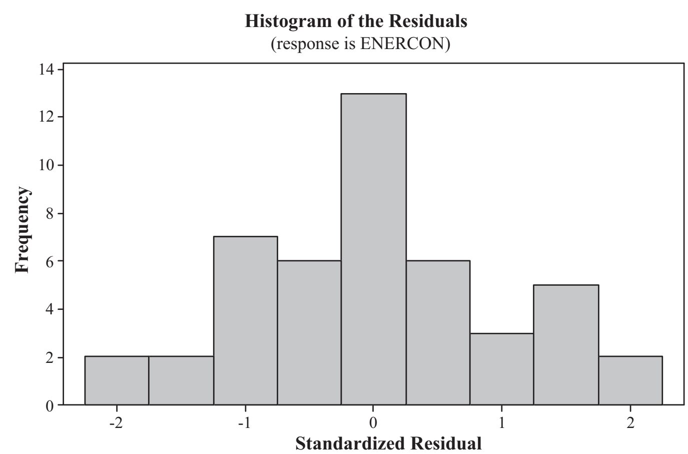

Fig. 4. Histogram of the residuals for Model 1.

correlation between two variables could be influenced by a third variable that was not considered [32]. The variables NEMP and SURF correlate with the ENERCON variable, and this correlation seems to be a function of the variables NEMP, SURF and NCASH, as seen in Table 4.

For the cluster analysis, the results of which can be seen in Fig. 2, it could be concluded that there exists a correlation between the variables NEMP, NCASH, SUSEV, SURF, AGE and ENERCON. The introduction of the NEMP variable in the model could have led to an uncorrected model because of the high correlation of this variable with the rest of the independent variables, as seen in Table 5.

3.2. Multiple regression analysis results

The best subsets and stepwise regression resulting model was:

$$\text{ENERCON} = 14,702 - 4175\ \text{SUSEV} + 62.5\ \text{SURF} + 7837\ \text{NCASH} \quad (3)$$

The regression statistics of Eq. (3) are given in Table 6. This first model achieves an equilibrium between a high R^2 value and low S value, and the corresponding independent variables are highly significant with respect to the model. The variables contained in the model are SUSEV, SURF and NCASH with a R^2 (adj) of 56.8% and an S value of 4062.38. The standard error (SE) for each regression coefficient should be smaller than the coefficient value because if it were greater, there would be a large confidence interval, which indicates a low significance for the inclusion of the corresponding variable in the regression model. Additionally, if the interval contains 0, the model would become unstable. The p-values from the Student's t-test give the same result: the most significant variable is NCASH with a p-value of 0, and the rest of variables have a p-value less than the 0.05 level of confidence for the alternative hypothesis of including these variables in the regression model. It can be seen in Table 6 that the variance inflation factor (VIF) is included, where a higher value indicates a higher multiple collinearity between the independent variables. For some authors, the value 4 corresponds to the highest limited value from which collinearity is a problem. In this model, VIF is less than 4 for each variable, indicating the lack of collinearity between the independent variables. The results of the ANOVA analysis are also included in Table 6. The adequacy of having included the set of explanatory variables is considered, and where the p-value is 0, the linear regression model is adequate and is in concordance with the p-value and Student's t-tests for each explanatory variable. An F-test value of 20.71 is found to be concordant with the p-value and contrasts with the alternative hypothesis for the significance contribution of at least one explicative variable in the regression model.

The residual graphics are taken into account when analysing the homoscedasticity of each model, and they can be seen in Figs. 3–5. The probability graph shows that the observations are within the confidence interval limit of 95%, and from this result, it can be seen that the residuals follow a normal distribution, and their average is 0, as seen in Fig. 6.

Some independent variables have been removed from the model because they are not significant and do not meet the requirement of being predictive in the regression model if they are included.

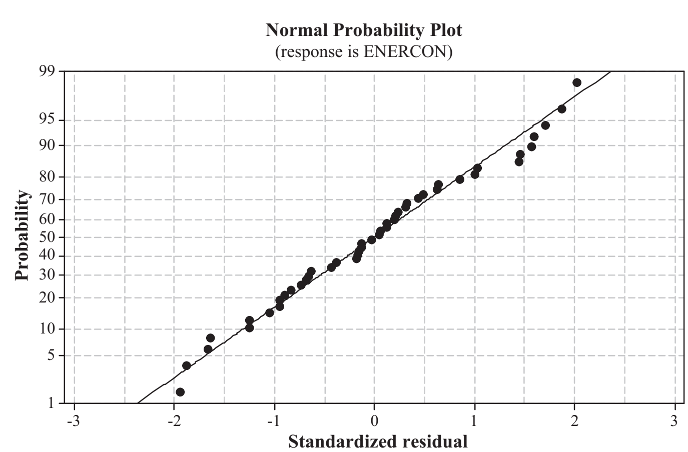

Fig. 5. Normal probability plot of the standardised residuals for Model 1.

Table 5
Correlation analysis between the dependent and independent variables.

	SUSEV	NEMP	SURF	AGE	NCASH
NEMP	0.347 (0.010)				
SURF	0.184 (0.179)	0.401 (0.002)			
AGE	0.267 (0.048)	0.323 (0.016)	0.257 (0.058)		
NCASH	0.309 (0.022)	0.366 (0.006)	0.174 (0.204)	0.011 (0.938)	
ENERCON	0.043 (0.756)	0.280 (0.038)	0.529 (0.000)	−0.026 (0.852)	0.333 (0.013)

Cells contain Pearson's correlation (*P*-value).

Table 6
Regression statistics for Model 1.

Model	Intercept	WCSEV	SUSEV	SURF	NEMP	HEIGHT	NCASH	EIF	AGE	GLASUR	HVAC	R^2 (%)	R^2 (adj) (%)	*S*	*F*	DW
1	14,702		−4175	62.5			7837					59.7	56.8	4062.38	20.71	1.58
SE coefficient	2738		1555	19.37			1256									
P	0.000		0.010	0.002			0.000									
VIF			1.220	1.104			1.272									

Observing Eq. (3) (Model 1), it can be deduced that the NCASH variable is an important contribution to the annual energy consumption of the bank branches; more ATMs result in a higher energy consumption, which is related to the total electricity consumption. Apparently the equation would contradict common sense, taken into account that the ENERCON variable is the sum of the annual energy consumption, this model implies that the SUSEV variable has greater importance in areas where their summer climate conditions are more severe, which also (due to the characteristics of Spain) coincides mostly with climatic zones with milder winters. In these climatic areas the total annual energy consumption is less than the climatic zones with more severe winters, the latter is the reason why the coefficient for SUSEV is negative. In addition, bank branches working hours are reduced in three hours and a quarter in summer time. The annual energy consumption is influenced by the surface area to be heated or cooled, if it increases, the HVAC consumption increases. This model reduces the uncertainty of the response variable as a function of the independent variables by 56.8%.

In addition to this first result, it is necessary to obtain models that describe the energy performance with more accuracy, and for this purpose, the total sample has been divided into several samples. After analysing different trends in winter and summer climate areas as well as energy inefficiency factors, it was decided to split the sample into two samples in order to observe the two different behaviours depending on winter severity value and see the different trends of these two subsamples. The first sample is composed of A, B and C winter climatic severities, and the second sample is composed of C, D and E winter climatic severities. The C climatic area is included in both samples because its energy performance could be intermediate. The proportion of bank branches in climate zones covering all areas of winter weather is not homogeneous, the heterogeneity is caused by the geography of Spain, which causes that the winter severity influence does not appear in the equation. There are a greater number of branches in winter weather severity C, D and E. Two regression models have been obtained:

$$\text{ENERCON} = 22,900 + 5155 \quad \text{NCASH} + 55.3 \quad \text{GLASUR} \tag{4}$$

$$\text{ENERCON} = -35,308 + 137 \quad \text{SURF} + 5853 \quad \text{NCASH} + 41,156 \quad \text{EIF} \tag{5}$$

The results from Eq. (4) (samples A, B and C) and Eq. (5) (samples D, E and C) are included in Table 7. Eqs. (4) and (5) reduce the uncertainty of the response variable by 65.2% and 68.5%, respectively, in the function of the independent variables. Uncertainty for this models are accepted considering that other factors influence in the energy consumption that have not been controlled in this study and the data acquisition is based on a study for a relatively small sample [19].

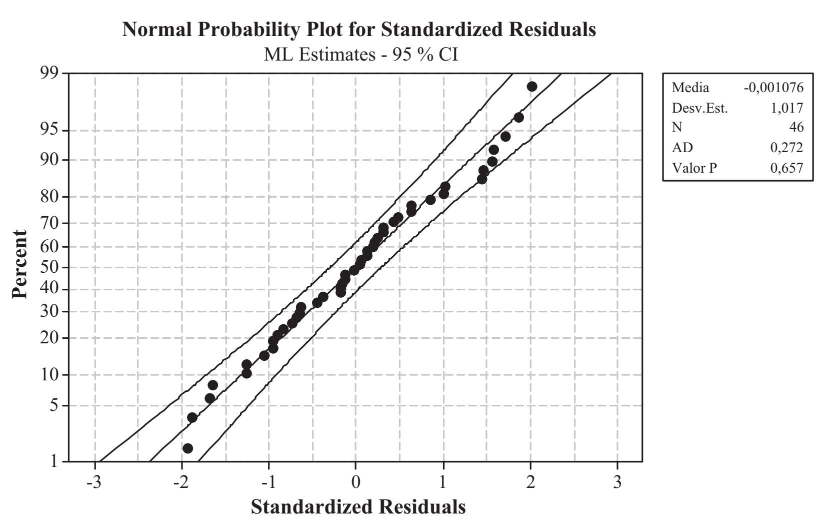

Fig. 6. Normal probability plot of standardised residuals with a 95% CI limits for Model 1.

Table 7
Regression statistics for Models 2 and 3.

Model	Intercept	WCSEV	SUSEV	SURF	NEMP	HEIGHT	NCASH	EIF	AGE	GLASUR	HVAC	R^2 (%)	R^2 (adj) (%)	*S*	*F*	DW
2	22,990						5155			55.3		71.1	68.5	2176	27.06	2.41
SE coefficient	1364						761.8			18.70						
P	0.000						0.000			0.007						
VIF							1.000			1.000						
3	−35,308			137			5853	41,156				69.7	65.2	5311	15.36	2.78
SE coefficient	22,814			22.42			1707	24,044								
P	0.137			0.000			0.003	0.102								
VIF				1.000			1.000	1.000								

Table 8
Analysed bank branches for the validation of Model 1.

Branch number	Real annual consumption 2010 (kWh/year)	Predicted annual consumption (kWh/year)	Error
592	104,333	48,331	40.00%
601	78,946	57,731	27.00%
813	72,240	47,581	35.00%
888	64,558	37,907	40.00%
887	57,424	38,244	33.00%
681	55,911	43,456	40.00%

It is important to highlight that the regression models are only valid over the range of values that determine the independent variables considered, and once out of this range of values, there is no certainty that successful results will be obtained.

3.3. Model validation

The application of the first regression model to the 4732 bank branches is plotted in Fig. 7. It can be seen there are some bank branches with great differences between the adjusted value and the real value of annual energy. Initially, it is easy to pinpoint a bank branch that has higher consumption by comparing similar characteristics. These banks are the object of a more detailed study. The real and adjusted energy consumption for the 6 randomly selected branches, mentioned above, used for the validation are included in Table 8. The proposed energy-saving measures have been applied toward lighting and HVAC because these represent 70–80% of the total energy consumption in a bank branch.

The deviation between real and adjusted energy consumption for these 6 branches, which the model allows to predict, was mainly caused by bank branches with improper energy accounting, significant inefficiencies in energy behaviour and awareness, obsolete equipment with low efficiency, energy overconsumption due to lack of regulation and sectoring, high storage surface area with continuous use of lighting and air conditioning systems even when not necessary for working hours. Therefore, it is highly recommended to check them as first step to correct the energy consumption deviation.

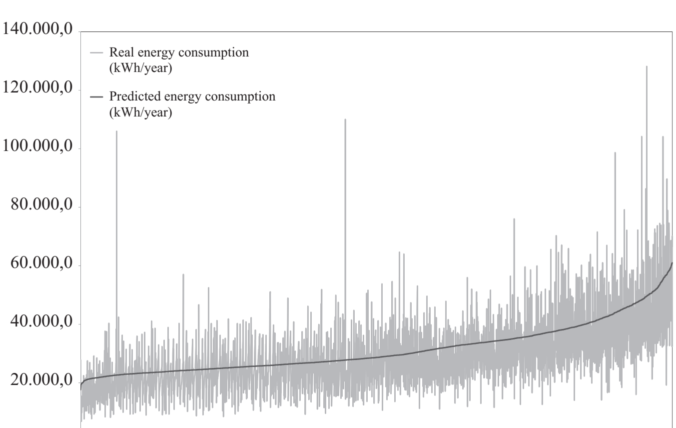

Fig. 7. Predicted annual energy consumption versus real annual energy consumption (kWh/year).

The medium and high cost energy efficiency measures proposed to improve the bank branches' efficiencies are summarised and classified by the application system as follows:

- Lighting system: lighting system replacement based on fluorescents (T8), the substitution of halogen lamps with fluorescent systems and incorporating presence detection systems for storing stays.
- Conditioning systems: the measures address equipment replacement by systems that permit the sectoring and regulating the energy demand as well as obsolete thermostat replacement and relocation to the correct locations.

4. Conclusions

Multiple linear regression analysis was used to estimate mathematical models to predict the energy consumption for the banking sector in Spain. From the analysis of the samples, it was observed that there is correlation between the variables NEMP, NCASH, SUSEV, SURF, AGE and ENERCON. It was found that the inclusion of the NEMP in the model could result in an uncorrected model because of NEMP high correlation with the rest of the independent variables.

In this study, three regression models were used to predict the energy consumption of a bank branch as a function of its construction characteristics, climatic area, and energy performance by predicting the annual energy consumption of banks. The first model presented allows for the detection inefficient bank branches in Spain and can be used as a tool to propose energy-saving measures to reduce the bank branch's energy consumption. The uncertainty of the response variable is reduced by 56.8% as a function of the independent variables using this model.

Finally, Models 2 and 3 reduce the uncertainty of the response variable as a function of the independent variables by 65.2% and 68.5%, respectively. Uncertainty for these models are accepted considering that other factors which have influence in the energy consumption have not been controlled in this study and the data acquisition is based on a study for a relatively small sample [19]. In spite of the fact that Models 2 and 3 were not validated, it is expected that they would achieve a better adjustment using the real energy consumption. To validate it, more study needs to be carried out since important information that these models require are lacking in our data set; yet, based on our experience with gathering these data, procuring such detailed information for individual buildings is a non-trivial task that requires a significant investment of time.

References

[1] E. Commission, Key Figures (2011), http://ec.europa.eu/energy/observatory/eu_27_info/doc/key_figures.pdf.
[2] D.L.M.P. Capros, N. Tasios, A. De Vita, N. Kouvaritakis, (E3M-Lab) EU Energy Trends to 2030 – Update 2009, European Commission, Belgium, 2010.
[3] IDAE, Energy Efficiency Policies and Measures in Spain, in: T.y.C. Gobierno de España, Ministerio de Industria (Eds.), Monitoring of Energy Efficiency in EU 27, Norway and Croatia, Intelligent Energy Europe, Odysee-Mure, 2009.
[4] E. Commission, EU Energy in Figures 2010. CO_2 Emissions by Sector. Extended time series (2010), http://ec.europa.eu/energy/publications/doc/statistics/ext_co2_emissions_by_sector.pdf.
[5] Mure-Odyssee, Energy Efficiency Profile: Spain (2011), http://www.odyssee-indicators.org/publications/country_profiles_PDF/esp.pdf.
[6] S. Chen, N. Li, J. Guan, Research on statistical methodology to investigate energy consumption in public buildings sector in China, Energy Conversion and Management 49 (8) (2008) 2152–2159.
[7] I. Ward, A. Ogbonna, H. Altan, Sector review of UK higher education energy consumption, Energy Policy 36 (8) (2008) 2939–2949.
[8] S. Becken, C. Frampton, D. Simmons, Energy consumption patterns in the accommodation sector—the New Zealand case, Ecological Economics 39 (3) (2001) 371–386.
[9] T.N.T. Lam, K.K.W. Wan, S.L. Wong, J.C. Lam, Impact of climate change on commercial sector air conditioning energy consumption in subtropical Hong Kong, Applied Energy 87 (7) (2010) 2321–2327.
[10] A. Salem Szklo, J. Borghetti Soares, M.c. Tiomno Tolmasquim, Energy consumption indicators and CHP technical potential in the Brazilian hospital sector, Energy Conversion and Management 45 (13–14) (2004) 2075–2091.
[11] O.A. Nekrasova, An energy consumption forecast for the Soviet residential and commercial sector, Energy Economics 13 (1) (1991) 10–18.
[12] J.O. Jaber, S.D. Probert, Purchased-energy consumptions in Jordan's commercial and public-service sector, Applied Energy 71 (1) (2002) 31–43.
[13] N. Mairet, F. Decellas, Determinants of energy demand in the French service sector: a decomposition analysis, Energy Policy 37 (7) (2009) 2734–2744.
[14] G.N. Spyropoulos, C.A. Balaras, Energy consumption and the potential of energy savings in Hellenic office buildings used as bank branches—a case study, Energy and Buildings 43 (4) (2011) 770–778.
[15] S.E. Chidiac, E.J.C. Catania, E. Morofsky, S. Foo, A screening methodology for implementing cost effective energy retrofit measures in Canadian office buildings, Energy and Buildings 43 (2–3) (2011) 614–620.
[16] J.C. Lam, K.K.W. Wan, D. Liu, C.L. Tsang, Multiple regression models for energy use in air-conditioned office buildings in different climates, Energy Conversion and Management 51 (12) (2010) 2692–2697.
[17] T. Catalina, J. Virgone, E. Blanco, Development and validation of regression models to predict monthly heating demand for residential buildings, Energy and Buildings 40 (10) (2008) 1825–1832.
[18] G.K.F. Tso, K.K.W. Yau, Predicting electricity energy consumption: a comparison of regression analysis, decision tree and neural networks, Energy 32 (9) (2007) 1761–1768.
[19] A.F. Tzikopoulos, M.C. Karatza, J.A. Paravantis, Modeling energy efficiency of bioclimatic buildings, Energy and Buildings 37 (5) (2005) 529–544.
[20] A.S.H.S. Chatterjee, B. Price, Regression Analysis by Example, third ed., John Wiley, New York, 2000.
[21] D. Peña, Regresión y Diseño de Experimentos, Alianza Universidad Textos, Madrid, 2002.
[22] G.C.R.D.C. Montgomery, Probabilidad y Estadística Aplicadas a la Ingeniería, second ed., Limusa Wiley, Mexico D.F., 2004.
[23] H.S.N.R. Draper, Applied Regression Analysis, third ed., Wiley, New York, 1998.
[24] M.H.K.J. Neter, C.J. Nachtsheim, W. Wasserman, Applied Linear Statistical Models, fourth ed., WBC/McGraw-Hill, Chicago, 1996.
[25] M. Inc., Minitab 15.1.30.0. The Fully Functional Demo Version (2007).
[26] S. Inc., SPSS Statistics 17.0. Release 17.0.0 (2008).
[27] L.M.A.P. Grima, J. Tort-Martorell Llabrés, Estadística Práctica con Minitab, Prentice Hall, Universitat Politecnica de Catalunya, 2004.
[28] Q. Shao, Y. Wu, A consistent procedure for determining the number of clusters in regression clustering, Journal of Statistical Planning and Inference 135 (2) (2005) 461–476.
[29] C. Smyth, D. Coomans, Y. Everingham, T. Hancock, Auto-associative multivariate regression trees for cluster analysis, Chemometrics and Intelligent Laboratory Systems 80 (1) (2006) 120–129.
[30] J. Gravier, V. Vignal, S. Bissey-Breton, J. Farre, The use of linear regression methods and Pearson's correlation matrix to identify mechanical–physical–chemical parameters controlling the micro-electrochemical behaviour of machined copper, Corrosion Science 50 (10) (2008) 2885–2894.
[31] S.J. Steel, D.W. Uys, Influential data cases when the Cp criterion is used for variable selection in multiple linear regression, Computational Statistics & Data Analysis 50 (7) (2006) 1840–1854.
[32] R. Kan, X. Wang, On the distribution of the sample autocorrelation coefficients, Journal of Econometrics 154 (2) (2010) 101–121.