Energy & Buildings 164 (2018) 99–108

Contents lists available at ScienceDirect

Energy & Buildings

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

Statistical methodology to assess changes in the electrical consumption profile of buildings

Xavier Serrano-Guerrero [a,*], Guillermo Escrivá-Escrivá [b], Carlos Roldán-Blay [b]

[a] Grupo de Investigación en Energías, Universidad Politécnica Salesiana, Elia Liut y Calle Vieja, 010103, Cuenca, Ecuador
[b] Institute for Energy Engineering, Universitat Politècnica de València, Camino de Vera, s/n, Edificio 8E, escalera F, 2° piso, 46022, Valencia, Spain

ARTICLE INFO

Article history:
Received 6 October 2017
Revised 5 December 2017
Accepted 23 December 2017
Available online 12 January 2018

Keywords:
Consumption profiles
Changes in consumption
Electrical consumption patterns
Building management

ABSTRACT

Many efforts have been made to define patterns, predict, and forecast energy use. However, changes in energy consumption may be studied in detail using various methodologies. This work presents a statistical methodology to assess changes in a facility consumption profile. Consumption patterns are obtained from a historical database of a predefined time interval, according to the type of day (day of the week, working or non-working), and an index that assesses change in the electrical consumption profile is proposed. Assessing these changes enables associating these values with possible events in a facility, which can serve to generate alarms in an energy management system, and reduce costs and maintenance periods. Additionally, a multi-criteria interpretation of the applied test table is presented that offers explanations and identifies possible causes of anomalous consumption.

1. Introduction

Several studies have been made to reduce the demand and use of energy in buildings around the world. Energy efficiency has been studied from the point of view of construction materials, heat ventilation air conditioning systems (HVAC), lighting systems [1], use of renewable energy, changes of devices for others of greater efficiency [2], demand management [3], integration of distributed generation with clean energy [4], and so on. One of the most important efforts related to improving energy efficiency in buildings is the concept 'zero energy buildings', which proposes that buildings become self-sufficient in renewable energy to avoid emitting CO_2 to the environment [5].

An analysis of the energy behaviour (pattern) is crucial to propose actions for improving energy efficiency. Models for energy consumption have been developed from a statistical perspective [6], including multiple regression analysis and principal components analysis [7]. Researchers have statistically analysed energy patterns [8–11], where it is possible to cluster similar consumptions for different purposes, such as monitoring, classifying consumers, or analysing power depending on the type of day. In addition, using these methodologies, it is possible to predict power consumption for consumers, traders, utilities, or generators.

The terms prediction and forecasting differ in their meanings in the literature [3,12–15]. A prediction is the output of a statistical model – even if the data is historic. Forecasting refers to predictions of future values. Methods for predicting building energy consumption are summarised in [12] and they are grouped as: simplified designed engineering methods; statistical methods; artificial intelligence (especially neural networks); support vector machines (SVM); and support vector regression (SVR). SVR is usually used to classify and solve regression problems [13]. The selection of variables and their features is maybe as important as the statistical method applied in the data analysis. However, the application of key variables is scarce in energy consumption analysis and modelling [16]. There are many factors influencing energy consumption and so analysis can be complex. In [17], it is indicated how the selection of subassemblies of feature selection characteristics impacts on the performance of learning machines when a statistical learning method is applied.

In the literature many authors state that numerous efforts have been made to define patterns, predict, and forecast energy use; however, the identification and quantification of changes in energy consumption patterns have not been evaluated in depth. This work presents a statistical assessment for identifying changes in consumption methodology (SAICC methodology) to detect changes in electrical consumption for an analysed period. The analysed day is compared with a pattern previously obtained from a historical database of a predefined time interval. An index of change (IoC) is calculated to catalogue the change in consumption of the analysed

* Corresponding author.
E-mail address: jserranog@ups.edu.ec (X. Serrano-Guerrero).

https://doi.org/10.1016/j.enbuild.2017.12.059
0378-7788/

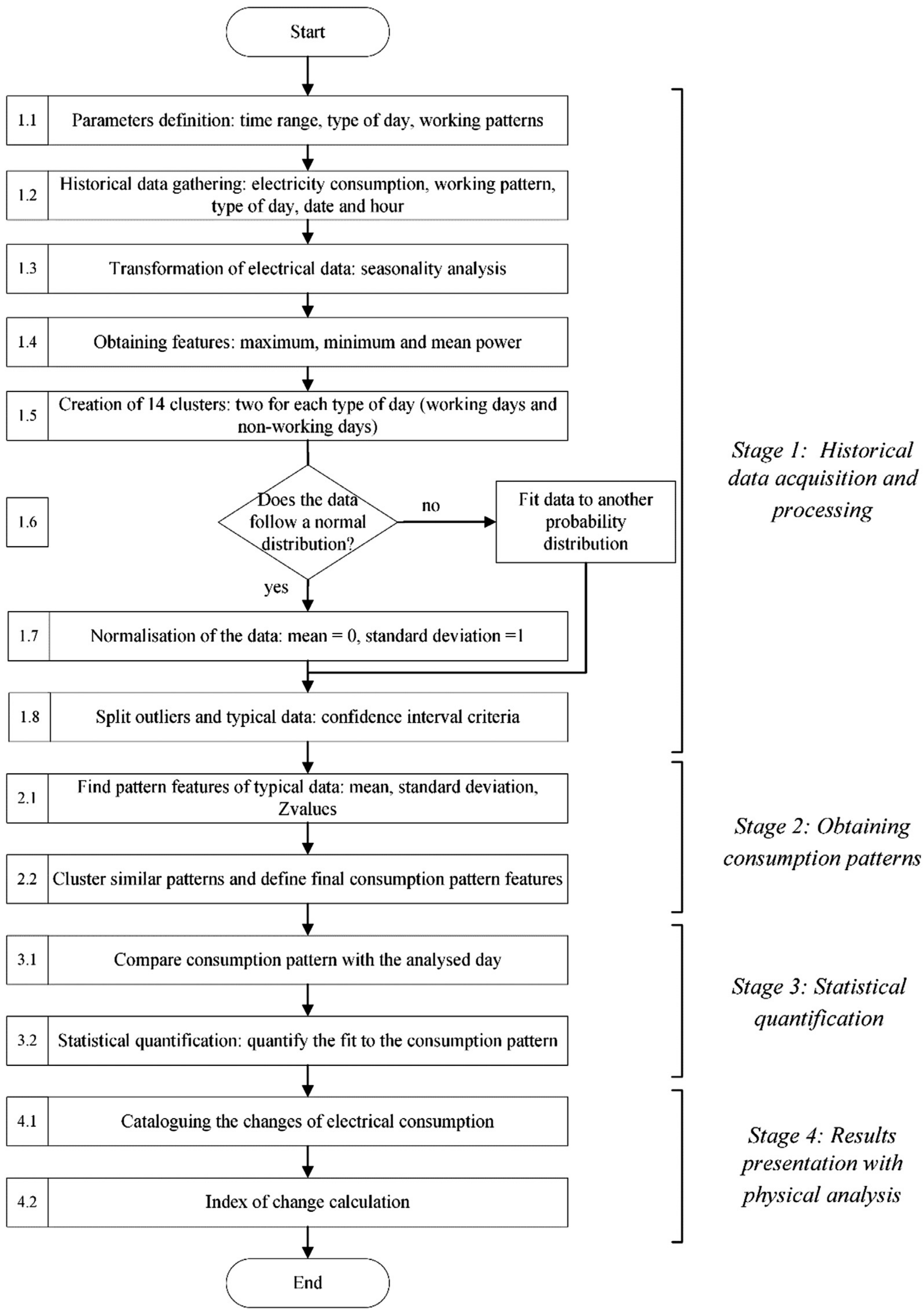

Fig. 1. Flowchart of SAICC methodology.

day and using a multi-criteria interpretation table, the cause of the abnormal consumption is inferred.

Assessing changes in consumption profiles enables associating these values with possible events of abnormal consumption in a facility. This may be used to generate alarms, reduce costs in maintenance, and respond quickly to anomalous consumption.

This paper is organised as follows. Section 2 explains the proposed methodology. Section 3 presents the application and validation of the proposed statistical methodology. Finally, some conclusions are drawn in Section 4.

2. Proposed statistical methodology

A statistical methodology (SAICC methodology) is proposed to analyse and evaluate changes in the electrical consumption in a facility.

Fig. 1 shows the flow diagram of SAICC methodology with four stages. Stage 1 acquires the data and selects data that will be used in Stage 2 to obtain consumption patterns. Stage 3 compares consumption patterns with the analysed day; and finally, Stage 4 catalogues the detected changes and calculates the IoC. The following sections describe this methodology in detail.

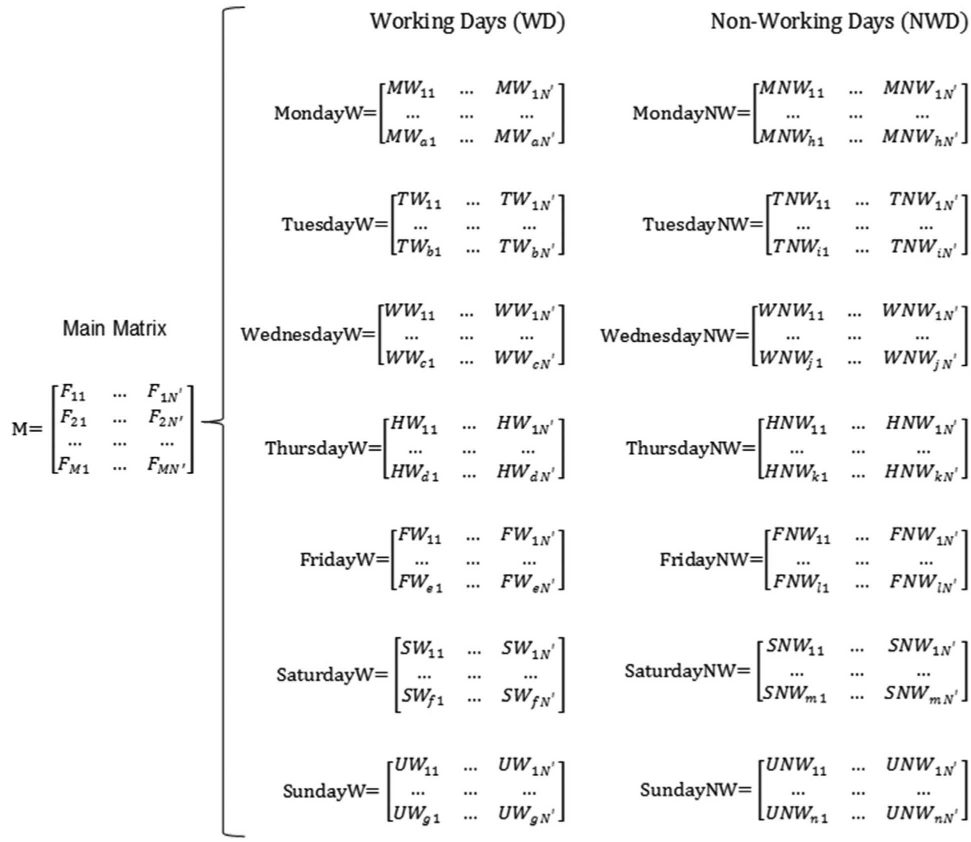

Fig. 2. Disaggregated matrices.

2.1. Stage 1: historical data acquisition and processing

2.1.1. Parameters definition

The analysis parameters to apply the methodology for obtaining patterns are initially established. The time range (number of days), type of day (days of the week), and working patterns – working days (WDs) or non-working day (NWDs) – are defined.

2.1.2. Historical data gathering

Historical data used in the methodology is organised in a matrix (MxN). Where M indicates the number of days in the time range for obtaining the consumption pattern and N indicates the number of parameters considered in each day.

For this study, the maximum analysis interval is a year – since it is considered that the consumption patterns change over time, and data is obtained from power meters installed at the Universitat Politècnica de València (UPV) using the Derd System [3], and so $M = 365$.

The parameters considered are electrical energy consumption in 15-minute intervals, date and hour, type of day (day of the week) and working pattern (WD or NWD), and so $N = 99$.

2.1.3. Transformation of electrical data

The electrical energy consumption in a building varies over time for many reasons, including external temperature changes, work operations, holidays, and so on. When making an annual data analysis, if the analysis interval is greater than three months, it is necessary to transform the data so that the seasonality of consumption does not reflect anomalous data [9].

For seasonality analysis, a power data vector of values of WDs is obtained and a base load value is calculated. The base load value is obtained with the value of the 1st percentile of the whole analysed period. When analysing the data, the authors determined that less than 1% of outliers exist in the left-tail of the probability distribution, so outliers related to extraordinary events, such as power outages, are ignored using this percentile. The base load value is subtracted from the original power data vector of the WDs, since the base load is produced in non-working hours (NWH) and does not influence seasonality analysis. A moving average method is then applied to the data vector to obtain seasonality indexes that are used to obtain the deseasonalised data vector [18]. Finally, the base load value is added to this vector. NWD are not considered in seasonality analysis because the energy consumption remains invariable when the building is empty.

2.1.4. Obtaining features

In this step, three columns with features are added to each day data. In this work, the maximum, mean, and minimum powers are added for each day. Daily energy consumption is not considered as a feature because it is directly related to the mean power (correlation equal to 1). Consequently, the data matrix has a dimension of $M \times N'$, where $N' = 102$.

2.1.5. Creation of 14 clusters

The matrix $M \times N'$ is segmented into 14 disaggregated matrices, two (WD or NWD) for each day of the week (Fig. 2). Data is separated: firstly, to process it more efficiently using only the data necessary in each operation; and secondly, because then the data is easily clustered if the consumption pattern of one type of day is similar to another. This provides greater robustness to the method.

2.1.6. Normal distribution analysis

It is necessary to prove if the data has a normal distribution in the disaggregated matrices, for each hour (column). To prove a normal distribution of the data, the Chi-square goodness of fit test

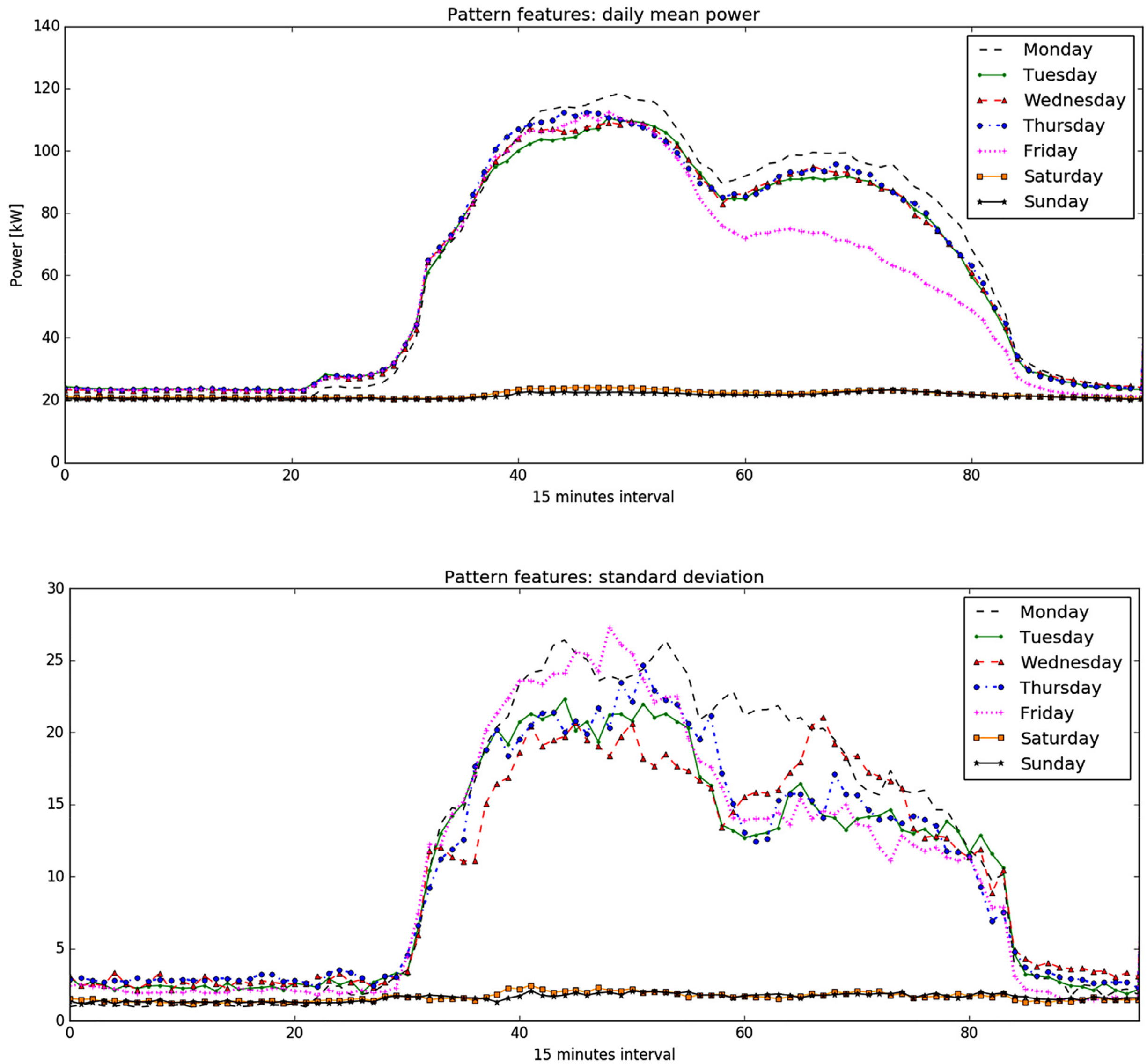

Fig. 3. Finding pattern features, mean power, and standard deviation (96 values).

is used [19].

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

Where:

- χ^2 is the value of the aleatory variable whose sampling distribution is approximated to the chi square distribution with k-1 degree of freedom.
- k is the number of class intervals.
- o_i and e_i represent the observed and expected frequencies, respectively, for the i value.

In this study, the tests made lead to the conclusion that 'the null hypothesis where the data follows a normal distribution is not rejected' . Therefore, for following stages, it is considered that data follows a normal distribution.

2.1.7. Normalisation of the data

The disaggregated matrices are standardised, once it is proved that data follows a normal distribution, making the mean equal to zero ($\mu = 0$), and the standard deviation is equal to one ($\sigma = 1$).

The normalisation procedure is made column by column. If the number of rows of the matrix is more than 30 (number of days considered), the statistic of the standard deviation is approximated to the real standard deviation [19]. The Z value for each value of data (per column) is computed (expression 2):

$$Z_{rc} = \frac{x_{rc} - u_c}{\sigma_c} \quad (2)$$

Where:

x_{rc} is the value of the variable (in the disaggregated matrix) of the row r and column c, u_c is the mean of the values in column c, and finally σ_c is the standard deviation of the values in column c.

The maximum values of Z in each column are the Z_{max} vector, while the minimum values of Z in each column is the Z_{min} vector.

2.1.8. Split outliers and typical data

A value is considered an outlier when it is out of the 95% confidence interval in a normal distribution function. This procedure is made for each column of each disaggregated matrix. A day is classified as atypical if one or more of the N´ values is out of the confidence interval.

Table 1
Results of clustering process.

N°	Cluster
1	Monday
2	Tuesday- Wednesday -Thursday
3	Friday
4	Saturday-Sunday

The probability that a value is inside the confidence interval is expressed as:

$$P(Z_{\alpha_1} < Z_{rc} < Z_{\alpha_2}) = 1 - \alpha_1 - \alpha_2 \tag{3}$$

Where $Z_{\alpha 1}$ and $Z_{\alpha 2}$, are the lower and upper limits of the confidence interval respectively, while α_1 and α_2 represent the areas of the left and right tail of the normal distribution respectively.

Finally, this step splits the data in two populations: one for the typical data (without outliers); and the second for the non-typical data (with outliers).

2.2. Stage 2: obtaining consumption patterns

2.2.1. Find pattern features

The population of typical data represents the consumption pattern and is represented in a 'pattern matrix' for each disaggregated matrix. The features considered for each pattern are: mean; standard deviation of the energy consumption data; and Z_{min} and Z_{max} (in Z dimension values).

Fig. 3 shows mean and standard deviation vectors of the typical values for each day of the week in a period of one year. In this case, Monday to Friday were considered WD and Saturdays and Sundays NWD.

2.2.2. Cluster similar patterns

The objective of this step is to group similar patterns to the 'analysed day', since the proposed method works better if more data is available. Euclidian distance has been used to cluster data. The Euclidian distance between a vector A and B with the same quantity of elements n is defined as follows:

$$Distance(A, B) = \sqrt{\sum_{i=1}^{n} (a_i - b_i)^2} \tag{4}$$

Where $n = 96$ in this study and each value (a_i and b_i) represents the mean power in an interval of 15 minutes.

Final consumption pattern is called 'final pattern matrix', and is formed by the set of days that define the features of the pattern. To obtain this matrix, the Euclidean distance is calculated between the vectors of the mean and standard deviation of each initial pattern of each day of the week. If the distance is less than an arbitrary threshold, the patterns are considered similar. The final pattern matrix is obtained by grouping similar patterns, and pattern features are again calculated.

In this work, the formation of clusters from Monday to Friday is determined by the mean power and not by standard deviation. Table 1 shows the results of the clustering process.

2.3. Stage 3: statistical quantification

2.3.1. Consumption pattern vs analysed day

Defined variables for the final pattern matrix serve as references for analysing the electrical consumption of a specific day. The analysed day is defined as a row vector that contains the N' columns previously defined.

The values of the analysed day are transformed using the equation:

$$Z_i = \frac{x_i - u_i}{\sigma_i} \tag{5}$$

Where Z_i is the Z value, i is the index column, x_i is the feature value in the position i. Meanwhile, u_i is the mean, and σ_i is the standard deviation of values of column i of the 'final pattern matrix' .

Once the data is transformed using Eq. (5), it can be represented in the *Z domain*.

The confidence interval is variable because the $Z_{min,i}$ and $Z_{max,i}$ values are different in each of the 99 values. Clearly, if Z_i is inside the confidence interval ($Z_{min,i} < Z_i < Z_{max,i}$), it is not considered as an outlier. The confidence interval can be considered as the probability that one value is a typical value.

The hypothesis test is used to accept or reject the existence of outliers in the data. The null hypothesis is defined as: 'The power consumption in the analysed interval is not an outlier'. The alternative hypothesis is defined as: 'The power consumption in the analysed interval is an outlier'.

It is important to emphasise that the hypothesis test is not approximated with a fixed probability of the type I error (reject the null hypothesis when it is true); in other words, it is not establishing an 'α' fixed level of significance [19].

2.3.2. Statistical quantification

The objective in this step is to statistically quantify if the analysed day fits the consumption pattern. The *Z value*, confidence interval, type I error, type II error (failure to reject the null hypothesis when it is false [19]), and outlier index (0 if the value is typical and 1 if an outlier is present) are calculated for each variable for the analysed day. Table 2 summarises the analysed variables.

2.4. Stage 4: results presentation with physical analysis

2.4.1. Cataloguing the changes of electrical consumption

The SAICC methodology tries to identify if a change in consumption occurred and additionally offers a possible cause of the change. For this, 12 tests are applied. These tests give relevant information about electrical consumption, such as the proportion of outliers, time of occurrence, duration, consecutive measured data with a constant value (used to detect measurement errors), and finally, three tests evaluate the behaviour of the maximum, mean, and minimum powers. Table 3 details each test.

Finally, the 12 tests are evaluated to present the conclusions. An algorithm was developed to obtain a multi-criteria interpretation of applied test table (MCIAT), where possible results are considered and 569 different cases were established. MCIAT gives clues to possible causes of abnormal energy consumption. Table 4 shows four of these cases as an example.

2.4.2. Index of change calculation

An index of change (IoC) is defined, to assess the consumption changes with respect to the calculated pattern. Firstly, I_{change} is defined

$$I_{change} = P_1 T_{1c} I_{C1} + P_2 I_2 I_{C2} + P_3 I_3 I_{C3} + P_4 I_4 I_{C4} \tag{6}$$

Where:

P_1, P_2, P_3 y P_4 are the weights of each term.

T_{1c} is the total number of outliers in the day expressed as $T_1/100$.

I_{C1} is the median of the values of the confidence interval of the 96 values of mean power values.

I_{C2}, I_{C3} and I_{C4} are the confidence intervals of the mean, maximum, and minimum powers of the pattern, respectively.

Table 2
Analysed variables.

Variable nº	Variable type	Results
1–96	Mean power consumption in an interval of 15 min	Z_i, confidence interval, type I and II errors, outlier index for the 99 variables
97	Mean power of the day	
98	Maximum power of the day	
99	Minimum power of the day	

Table 3
Tests applied in SAICC methodology.

Test nº	Description	Test value
1	Total number of outliers in the day, from 1 to 96 (TNO)	T1 = TNO/96 *100%
2	Number of outliers in NWH (NONWH)	T2 = NONWH/40 *100%
3	Number of outliers [0:00–7:00] (NO_0–7)	T3 = NO_0–7 / 28 *100%
4	Number of outliers [21:00–24:00] (NO_21–24)	T4 = NO_21–24/ 12 *100%
5	Number of outliers in WH [7:00–21:00] (NOWH)	T5 = NOWH / 56 *100%
6	Number of outliers [7:00–14:00] (NO_7–14)	NO_7–14 / 28 *100%
7	Number of outliers [16:00–21:00] (NO_16–21)	NO_16–21 / 24 *100%
8	Consecutive equal values CEV	(CEV/4) [hours]
9	Consecutive outliers number CON	(CON/4) [hours]
10	Mean power	Higher, lower, within range, and confidence interval
11	Maximum power	Higher, lower, within range, and confidence interval
12	Minimum power	Higher, lower, within range, and confidence interval

Table 4
Multi-criteria interpretation of applied test table (MCIAT).

Test	Case 1	Case 2	Case 3	Case 4	…
T1	x	0%	> 0%	> 0%	…
T2	x	x	< 10%	100%	…
T3	x	x	x	x	…
T4	x	x	x	x	…
T5	x	x	< 50%	> 0 and < 100%	…
T2/T5	x	x	x	> 1.35	…
T3/T4	x	x	x	x	…
T8	> 2	< 2	< 2	< 2	…
T9	x	0	< 1	x	…
T10	x	Within range	Within range	Higher	…
T11	x	Within range	Higher	Within range	…
T12	x	Within range	Within range	Higher	…
					…
Conclusion	Measurement error. More than two hours with the same measurement value	Day with normal power consumption	Less than 10% of anomalous data, both in WH and NWH. However, the maximum power consumption is higher than the maximum consumption of the pattern. Consumption exceeds peak hours	Consumption is anomalous in all NWH. Possible measurement error or a connected permanent load. Energy consumption is greater than the pattern. A load has probably been connected, or there is a measurement error. Minimum power consumption is greater than the pattern. A load has possibly been connected.	

Note: 'x' means that the result of the test can be any value.

I_2, I_3 and I_4 are the outlier indices of the mean, maximum, and minimum powers of the pattern, respectively.

In this work, P1 and P2 are equal to 1/3 because their terms have a greater incidence in consumption changes than the other terms. P3 and P4 are equal to 1/6.

I_{change} is normalised as it is always lower than 1 by definition. Hence, the maximum index I_{cmax} is calculated.

$$I_{cmax} = P_1\, I_{C1} + I_{C2}\, P_2 + I_{C3}\, P_3 + I_{C4}\, P_4 \tag{7}$$

Eq. (7) is a modification of Eq. (6), where,

$T_{1c} = 1$ considers that all the values are outliers.

$I_2 = 1,\ I_3 = 1,\ I_4 = 1$ considers that the mean, maximum, and minimum powers are out of range (energy consumption is considered very abnormal).

I_{cmax} can be 1 only when all the data is considered as outliers and when mean, maximum, and minimum powers are out of range.

Finally, the IoC is defined as:

$$IoC = \frac{I_{change}}{I_{cmax}} \tag{10}$$

IoC can vary between 0 and 1. However, it is very difficult to reach 1 because several conditions must be fulfilled as seen above. An IoC equal to 0 shows that the power consumption of the analysed day perfectly fits the calculated pattern and there is no anomalous data. The experience acquired in this work shows that the IoC values can be catalogued using Table 5.

3. Application and validation of SAICC methodology

3.1. SAICC methodology application in a UPV building

An example of application of SAICC methodology is presented using the electrical consumption values of Building 5E at the UPV

Table 5
Catalogue of day consumptions using IoC values.

IoC value	Change in consumption	Interpretation
0	None	Daily power consumption fits the pattern
(0–0.03]	Negligible	Very little anomalous data. This does not indicate significant changes in consumption
(0.03–0.15]	Noticeable	A number of important anomalous values (but less than 50%) are observed that have not shown changes in the maximum, minimum, or mean power. A load possibly remained connected during NWH
(0.15–0.3]	Obvious	There are many non-typical values throughout the day, such as a permanent connected load, which do not show great differences in power consumption compared to the pattern, but there may be out of range values for maximum or minimum powers
(0.3–0.55]	Large	Anomalous consumption during much of the day that causes energy consumption outside the range of the pattern
(0.55–0.75]	Very Large	Presence of non-typical values during most of the day that cause energy consumption outside the range of the pattern. In addition, the maximum or minimum power is also out of range
(0.75–1]	Extreme	Anomalous values of power consumption cause all the parameters to be out of range. Energy consumption is very non-typical

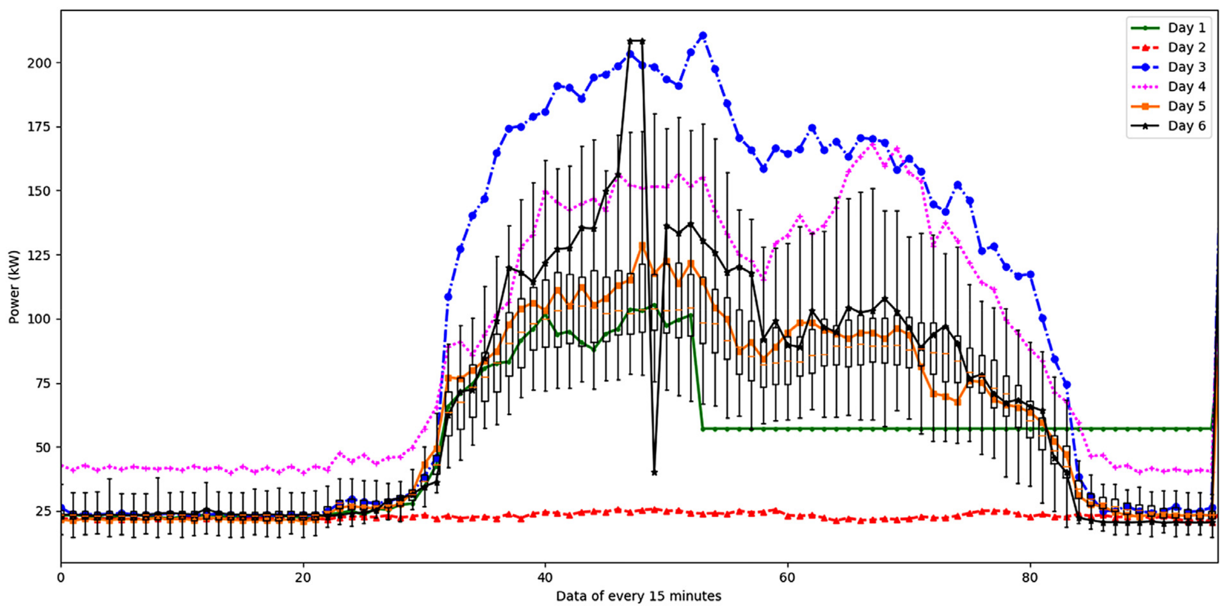

Fig. 4. Consumption pattern and analysed days 1–6 consumption profiles.

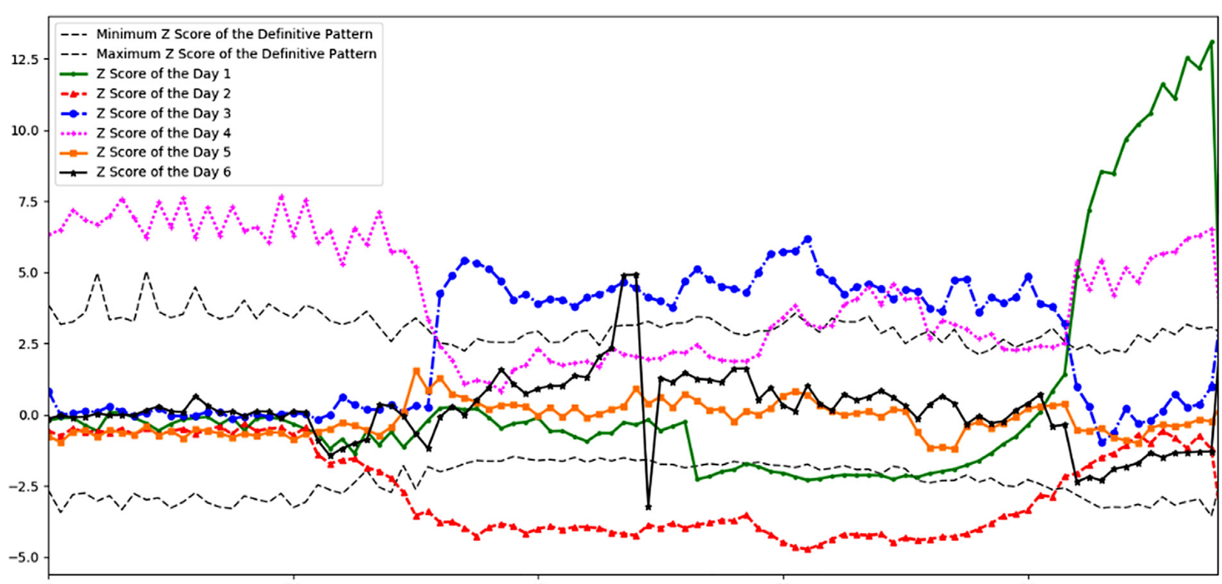

Fig. 5. Normalised data in Z domain for analysed days 1–6 and power features.

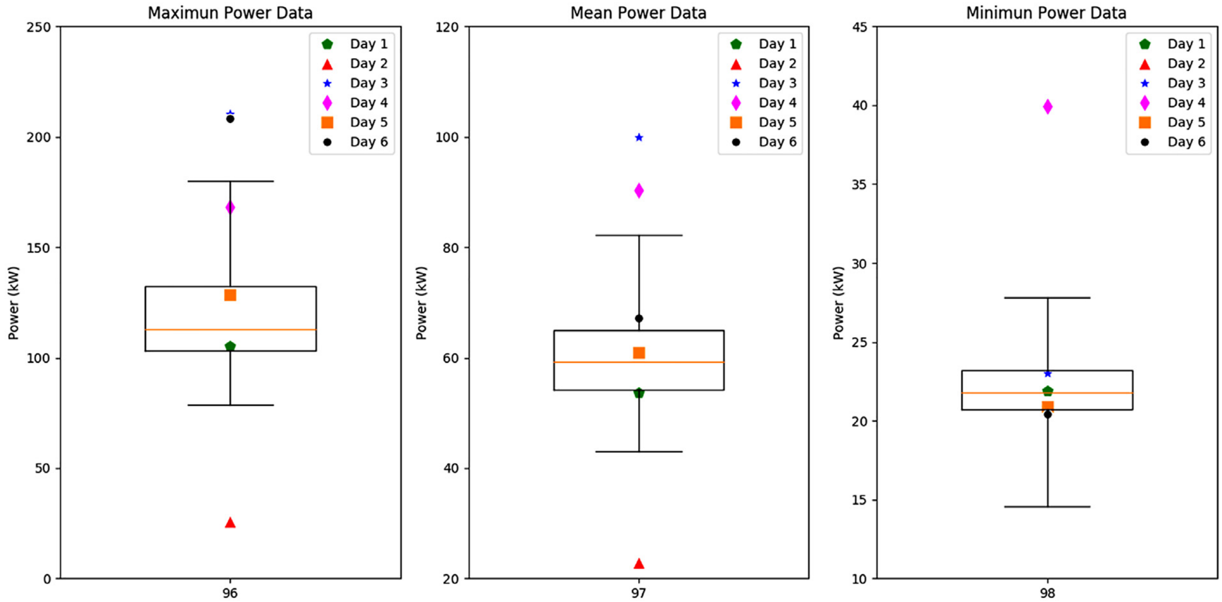

Fig. 6. Boxplot of maximum, mean, and minimum power of analysed days.

Table 6
Results of statistical quantification for analysed day 4.

Value Nº	Hour	Z_{min}	Z_{max}	Z_i	Confidence interval	Outlier index	Error I	Error II
1	0:00–0:15	-2.6530	3.8672	6.3155	0.9960	1	0.0040	0.0000
2	*0:15–0:30*	-3.4302	3.1707	6.4862	0.9989	1	0.0011	0.0000
3	*0:30–0:45*	-2.8047	3.2553	7.1816	0.9969	1	0.0031	0.0000
...	...	...	...	...	...	...	...	...
96	*23:45–24:00*	-3.5578	3.0664	6.5122	0.9987	1	0.0013	0.0000
97	–	-1.8279	2.8311	2.2791	0.9639	0	0.0361	0.0250
98	–	-2.1338	2.7363	3.7363	0.9805	1	0.0195	0.0000
99	–	-3.1755	2.6981	8.0570	0.9958	1	0.0042	0.0000

for a working Thursday. Six days of this type are analysed and the results are presented.

In stage 1 of the SAICC methodology, the initial parameters are defined. The time range is established at 365 days (consumption data from July 2015 to June 2016), and the seasonality analysis is then made.

In stage 2, the final consumption pattern of the selected type of day (a working Thursday) is obtained after finding the features and clustering process. Consumption pattern (boxplots) and consumption of the analysed days (96 power values) are shown in Fig. 4. As mentioned, data is transformed to *Z domain* to show the information more clearly (Fig. 5). Analysed days 1–4 and 6 correspond to days with detected outliers. Analysed day 5 corresponds to a Thursday without outliers. As can be seen, the proposed methodology is a simple statistical technique, with low computational load that automatically finds patterns of energy consumption for each type of day.

Additionally, maximum, mean, and minimum power features of six analysed days (defined in Table 2) are shown in Fig. 6 using boxplots.

In stage 3, a statistical quantification to test how analysed days fit the pattern is performed. This analysis is a powerful tool for energy managers, and enables them to automatically analyse large amounts of data. As an example, results for analysed day 4 are presented in Table 6.

In stage 4 of the SAICC methodology, 12 tests are considered to perform a physical analysis, as explained in point 2.4.1. The IoC is then calculated to assess changes in power consumption. Table 7 shows results of the tests applied to analysed day 4.

Additionally, results of 12 tests applied to the six analysed days are summarised in Table 8.

Conclusions that are automatically generated by SAICC methodology using the MCIAT are presented in Table 9. Additionally, a diagnosis of the six analysed days by a technical expert on power consumption behaviour is shown.

3.2. Validation of SAICC methodology

To validate the proposed methodology, the patterns obtained following a year of data collection were immediately tested during a six-month period. In this testing period, 30 days were catalogued as anomalous. A comparison between results given by the SAICC methodology and the diagnosis of a technical expert who manages the analysed building was made. In 19 of the 30 cases the proposed method and the expert give similar results, in 9 cases the proposed method gave more information and clues as to the causes of the anomalous consumption than the expert, and in 2 cases the expert offered a better explanation for changes in power consumption than the methodology. It is important to consider the time effort that the expert made in analysing the information and possible human errors (i.e. choosing the incorrect pattern for the analysed day).

In summary, the proposed SAICC methodology has the capacity to automatically assess and infer the causes of changes in power consumption profiles for an analysed day using an IoC and MCIAT. The analysed day is compared with a pattern obtained from a historical database. This methodology in real time can help reduce costs and energy consumption, as well as quickly detecting

Table 7
Results of 12 tests applied to analysed day 4.

Test	Value	Test result
Test 1: Total outliers	60	62.5% of change of daily consumption
Test 2. Outlier in NWH	40	100% of consumption change in NWH
Test 3. Outliers [0–7]	28	100% of consumption change at dawn
Test 4: Outliers [21–24]	12	100% of consumption change in NWH during the night
Test 5: Outliers [7-21]	20	35.71% of consumption change in WH
Test 6: Outliers (7–14)	4	14.3% of consumption change in WH during the morning
Test 7: Outliers (16–21)	12	50% of consumption change in WH during the afternoon and early evening
Test 8: Consecutive equal values	0	0 hours of consecutive equal values
Test 9: Consecutive outliers	32	8 hours of continuous anomalous consumption
Test 10: Mean power	Higher	Energy consumption is greater than energy consumption in the pattern. Confidence interval: 0.9805
Test 11: Maximum power	In range	Maximum power demanded is inside the range compared to the consumption pattern. Confidence interval: 0.9639
Test 12: Minimum power	Higher	Minimum power demanded is greater than minimum power in the pattern. Confidence interval: 0.9958

Table 8
Summary of test results applied to analysed days 1–6.

Day	T1 (%)	T2 (%)	T3 (%)	T4 (%)	T5 (%)	T2/ T5	T3/ T4	T8 (h)	T9 (h)	I2	I3	I4
1	31	30	0	100	32	0.94	0	10.75	4.5	0	0	0
2	58	0	0	0	96	0	–	0.5	13.5	1	1	0
3	56	0	0	0	93	0	–	0	13	1	1	0
4	62.5	100	100	100	36	2.77	1	0	8	1	0	1
5	0	0	0	0	0	–	–	0	0	0	0	0
6	3.1	0	0	0	5.4	0	–	0	0.75	0	1	0

* T1, T2, T3, T4 T5, T8 and T9 correspond to the tests described in Table 3; and I2, I3 and I4 are described in Eq. (6).

Table 9
Conclusions of power consumption of six analysed days.

Analysed day	SAICC methodology conclusions	Expert diagnosis	Best interpretation	IoC
1	Measurement error. More than two hours with the same measurement value	Measurement error due to having equal consecutive values	Both give same information	0.11
2	Energy consumed is less than the pattern consumption. It is probable that work activity decreased. Failure, measurement error, maintenance, or circuit disconnection may have occurred	The day was not a working day (it was a holiday)	Both give similar information	0.69
3	Energy consumption is greater than energy consumption in the pattern. It is probable that the work activity has increased The maximum power consumed is greater than in the pattern. It is probable that labour activity has increased	Higher consumption during the day. Day with extreme temperatures	Both give similar information	0.67
4	Consumption is anomalous for all NWH. Measurement error or a connected permanent load Energy consumption is greater than in the pattern. Something has probably been connected or there is a fault Minimum power is greater than in the pattern. Something probably has been connected, or a fault has occurred	Something connected at night. High ambient temperature	The proposed method gives additional information	0.71
5	Very normal day – fits the consumption pattern	Day with normal energy consumption	Both give same information	0
6	There is less than 10% of anomalous data; however, maximum power is higher than in the consumption pattern. Additional consumption extends to peak hours. Analysed consumption has 0% of anomalous data in NWH and 5.36% in WH, which does not indicate representative changes in power consumption	Measurement problems	Expert defines the problem in a better way	0.18

anomalies and failures. SAICC methodology may help improve energy consumption and help establish if energy saving policies are effective.

4. Conclusions

The proposed statistical assessment for identifying changes in consumption methodology (SAICC methodology) is a simple technique to automatically find patterns of energy consumption. SAICC methodology finds differences between the consumption patterns and the consumption of an analysed day.

In the methodology, the conclusions of the analysis are offered after 12 different tests are performed. An algorithm is developed to obtain a multi-criteria interpretation of applied test table (MCIAT) and 569 alternatives are established. MCIAT give clues about the possible causes of abnormal energy consumption.

A new index of change (IoC) is presented that assesses changes in power consumption for an analysed day according to a pattern obtained from similar consumptions.

The use of this methodology in real time can help reduce costs and energy, as well as determine consumption patterns and detect anomalies and failures. It is a time saving tool for managers and technicians in charge of the facilities of a building. Additionally, SAICC methodology may help identify actions to improve energy consumption, or assess changes in power consumption and establish if implemented energy saving policies are effective.

The authors believe that the method may be improved by integrating it within the energy management system of a building in operation.

Acknowledgments

This research work has been made possible by the support of Universidad Politécnica Salesiana Sede Cuenca and the GV/2015/068 grants for emerging research group projects.

References

[1] X. Guan, Z. Xu, Q.S. Jia, Energy-efficient buildings facilitated by microgrid, IEEE Trans. Smart Grid 1 (3) (2010) 243–252.
[2] G. Escrivá-Escrivá, Basic actions to improve energy efficiency in commercial buildings in operation, Energy Build 43 (11) (2011) 3106–3111.
[3] G. Escrivá-Escrivá, Nuevas Herramientas Para Facilitar La Respuesta Activa De Consumidores En Mercados Eléctricos liberalizados: Implementación y Retribución, Doctoral thesis, Universitat Politècnica de València, 2009.
[4] X. Serrano-Guerrero, G. Escrivá-Escrivá, Simulation model for energy integration of distributed resources in buildings, IEEE Lat. Am. Trans. 13 (1) (2015) 166–171.
[5] A.J. Marszal, P. Heiselberg, J.S. Bourrelle, E. Musall, K. Voss, I. Sartori, A. Napolitano, Zero energy building - a review of definitions and calculation methodologies, Energy Build 43 (4) (2011) 971–979.
[6] IAEA International Atomic Energy Agency, "Modelo para el Análisis de la Demanda de Energía (MAED-2)," pp. 7–16, 2007.
[7] T. Reddy, D. Claridge, Using synthetic data to evaluate multiple regression and principal component analyses for statistical modeling of daily building energy consumption, Energy Build 21 (1) (1994) 35–44.
[8] J. Kwac, J. Flora, R. Rajagopal, Household energy consumption segmentation using hourly data, IEEE Trans. Smart Grid 5 (1) (2014) 420–430.
[9] J.E. Seem, Pattern recognition algorithm for determining days of the week with similar energy consumption profiles, Energy Build 37 (2) (2005) 127–139.
[10] H. Xiao, Q. Wei, Y. Jiang, The reality and statistical distribution of energy consumption in office buildings in China, Energy Build 50 (2012) 259–265.
[11] Z. Ma, H. Li, Q. Sun, C. Wang, A. Yan, F. Starfelt, Statistical analysis of energy consumption patterns on the heat demand of buildings in district heating systems, Energy Build 85 (2014) 664–672.
[12] H.X. Zhao, F. Magoulès, A review on the prediction of building energy consumption, Renewable Sustain. Energy Rev. 16 (6) (2012) 3586–3592.
[13] A. Foucquier, S. Robert, F. Suard, L. Stéphan, A. Jay, State of the art in building modelling and energy performances prediction: A review, Renewable Sustain. Energy Rev. 23 (2013) 272–288.
[14] G. Escrivá-Escrivá, C. Álvarez-Bel, C. Roldán-Blay, M. Alcázar-Ortega, New artificial neural network prediction method based on buildings' end-uses for energy consumption forecast, Energy and Build 43 (2011) 3106–3111.
[15] H. Li, S. Guo, C. Li, J. Sun, A hybrid annual power load forecasting model based on generalized regression neural network with fruit fly optimization algorithm, Knowl.-Based Syst 37 (2013) 378–387.
[16] D. Hsu, Identifying key variables and interactions in statistical models of building energy consumption using regularization, Energy 83 (2015) 144–155.
[17] H. Zhao, F. Magoulès, Feature selection for predicting building energy consumption based on statistical learning method, J. Algorithm Comput. Technol. 6 (1) (2012) 59–78.
[18] D.A. Lind, W.G. Marchal, S.A. Wathen, Statistical Techniques in Business & Economics, McGraw-Hill/Irwin, 2012.
[19] R.E. Walpole, R.H. Myers, S. Myers, K. Ye, Probabilidad y Estadística Para Ingeniería y Ciencias, Pearson Educación, 2007.