Energy and Buildings 110 (2016) 377–386

Contents lists available at ScienceDirect

Energy and Buildings

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

A statistical modeling approach to detect anomalies in energetic efficiency of buildings

Angel L. Zorita[a], Miguel A. Fernández-Temprano[b], Luis-Angel García-Escudero[b], Oscar Duque-Perez[a,*]

[a] Department of Electrical Engineering, University of Valladolid, 47011 Valladolid, Spain
[b] Department of Statistics and Operational Research, University of Valladolid, 47011 Valladolid, Spain

ARTICLE INFO

Article history:
Received 4 February 2015
Received in revised form
15 September 2015
Accepted 3 November 2015
Available online 12 November 2015

Keywords:
Energetic efficiency
Multivariate statistical techniques
Energy consumption
Efficiency index
Energetic behavior model

ABSTRACT

This paper proposes a methodology that provides a tool to support decisions in the monitoring of electrical energy consumption. The purpose is to help the manager of a set of non-residential buildings, such as, for example, factories, hospitals or university buildings, to objectively compare the energetic performance of all the installations under its responsibility and take measures when necessary. Firstly, an energy consumption model is developed considering the variables that have an influence on energy consumption in the building and processing them using the appropriate multivariate statistical techniques. Then, an efficiency index is defined which allows the manager to monitor the evolution of the installations and detecting deviations that later can be investigated, from an energetic efficiency point of view. A case study is presented where the developed methodology is applied to the public hospitals in the region of Castilla y León (Spain).

1. Introduction

Energetic efficiency is receiving a lot of attention during the last years. Due to its high importance, this issue is being considered at all possible levels. At the lowest level, any of us may be interested in saving in his/her bill by a more efficient energy usage [1–3]. At a higher level, a factory manager may try to improve energetic efficiency in the plants under his/her control to reduce the expenses generated, among others, by the electric lighting, heating and cooling systems [4–7]. At the highest level, governments [8] and international agencies [9] are enacting laws and establishing requirements to act on this matter not only to obtain economic benefits but also to take care of the ecological side of energy consumption. For example, in Europe, the European Parliament has stated directive 2010/31/UE [10] to promote energetic efficiency in the buildings in the EU taking into account external climate conditions and many other factors such as the requested ambient conditions inside the building.

The work we present in this paper is on the middle level of the scale (asset management). We provide the manager of a set of non-residential buildings, such as, for example, factories, hospitals or University buildings, with a methodology to take decisions leading to a better control of energy consumption.

In a first step, we will develop an energy consumption model for the buildings under our control. To develop this model, we will consider the variables that have an influence on energy consumption in the building (some of them related to the constructive characteristics of the building, other with the activities performed in the building and other variables dealing with the climatic characteristics of the place where the building is placed [11]), and process them using the appropriate multivariate statistical techniques. This statistical foundation of the methodology is one of its strongest points since, in that way, the subjectivity usually appearing in these studies is minimized since the decisions to be taken would not only be based in the opinion of the managers but will also be supported by analytical facts obtained by a global analysis of the buildings, not just by individual considerations. Obviously, in order to implement this methodology, we will need past records for the consumptions and any other variable considered.

It is relevant to notice that, as the whole set of buildings will be considered for the establishment of the model, the possible inefficiencies existing in any of the buildings from the start of the study will also be detected in the scheme that we will develop, as the

* Corresponding author at: Departamento Ingeniería Eléctrica, Escuela de Ingenierías Industriales, Paseo del Cauce 59, 47011 Valladolid, Spain.
E-mail addresses: zorita@eii.uva.es (A.L. Zorita), miguelaf@eio.uva.es (M.A. Fernández-Temprano), lagarcia@eio.uva.es (L.-A. García-Escudero), oscar.duque@eii.uva.es (O. Duque-Perez).

http://dx.doi.org/10.1016/j.enbuild.2015.11.005
0378-7788/

model will estimate the expected consumption of each building talking into account the characteristics specified in the variables included in the model.

Once the model is established, in a second step we will define an efficiency index that allows us to control the evolution of each of our buildings and to know if any of them is performing, from an energetic efficiency point of view, better or worse that it should, taking into account all its characteristics collected in the previous variables. This index will thus be useful for the manager to take any required decision and to do it supported by facts founded on a strong statistical methodology and with a very low level of subjectivity.

This work is part of a project developed by the Departments of Statistics and Operations Research and Electrical Engineering of the Universidad de Valladolid (Spain) for the Regional Health Management of Castilla y León (Spain). The main objective of the project is to provide the managers of the public hospitals with a set of tools to handle them efficiently from an energetic point of view. In fact, in a previous paper we already showed how to optimize the cost of the electrical contracted capacity for a hospital or any other large building or factory [7]. Since we have mainly dealt with data coming from these hospitals, we will use them to illustrate our methodology.

Section 2 in the paper describes the steps of the methodology used to establish the model and to define the appropriate efficiency index. Section 3 contains a case study where the methodology is used to obtain a model for the public hospitals in the region of Castilla y León (one of the largest regions of the European Union, covering an area of 94,223 km^2) and examples of the use of the proposed index are given. Finally, conclusions are exposed in Section 4.

2. Methodology

As explained in the introduction, in the proposed two-step methodology we will first establish an energetic behavior model for the set of buildings that we are considering. Obviously, the buildings in this set must have analogous characteristics and functionality. For example, (as we will do in the case study section) we can consider a set of hospitals or a set of factories producing similar items but clearly there is no purpose in trying to establish a single model covering both hospitals and factories as there is no homogeneity in their energetic consumption.

In the second step, we will use the model to define an energetic efficiency index that will allow us to monitor the evolution of the energetic consumption of the buildings. With this index, we will be able to check, for example, if there are buildings that are behaving worse than they should or to detect any trend meaning improvement or worsening in their energetic consumption. In this way, we will have a very useful tool for energy management.

2.1. Model establishment

In order to establish the model, we will mainly use regression techniques to relate the energetic consumption of the buildings to the variables that are supposed to be influential on that consumption. These regression techniques will also provide us with estimators of the variability of the consumptions which will be necessary for the definition of the energetic efficiency index in the second step as they allow detecting discrepancies with the expected consumption.

The response variable in the model is the electric energy consumed in the buildings. The explanatory variables will reflect the constructive characteristics and functions of the building that may have an influence on the energetic behavior of the building and the climatic characteristics of the site where the building is. In this way, three sets of variables will be considered. The classification of the variables in these different sets will help us to process the information correctly as for the statistical techniques used later it is recommended that the variables considered measure structurally similar characteristics in order to obtain results that can be interpreted more easily.

One of these sets will contain the constructive characteristics of the building such as the total surface or the heated or cooled surfaces that have an obvious role in our matter of interest. Another set of variables having an influence contains those taking into account the different activities developed in the building. For example, if the building is a hospital, variables of interest will include, among others, the number of beds or the number of operating rooms. If the building is a school or college, variables of interest will be the number of classrooms, number of computer rooms, number of students, timetable, etc., or, in the case of an office building, variables such as the number of employees, number of visitors, timetable, number of office rooms, etc. would be included.

The third set of variables may deal with the climate characteristics of the building site including, for example, mean, maximum or minimum temperatures.

The response variable and the last two sets of variables considered (activity variables and climatic variables) can be measured in different length time periods (weekly, monthly, quarterly,...). Our method can obviously deal with these periodicities without any trouble. Also, although it is less frequent that changes in the constructive characteristics of the buildings appear, these constructive changes can also be included in our analysis scheme as changes in the appropriate variables.

For each of the periods considered we will fit a regression model, with its own set of variables and coefficients, which will allow us to model the behavior of the set of buildings in that period and to evaluate the performance of each building according to its characteristics. We will then join the results of the different periods to check the evolution of the performance of each building.

2.1.1. Regression

Regression is a well-known statistical technique [12] used to model the behavior of a numeric response variable, Y, the variable to be predicted; from a set of explanatory variables having some influence on the value of the response ($X_1, \ldots, X_k$). This technique is very useful, and there are tools that allow selecting which of the explanatory variables have a statistically significant influence on the response variable. The equation of the model can be written as

$$Y = f(X_1, \ldots, X_k) + \varepsilon \qquad (1)$$

where $f(\cdot)$ is a linear function that can be written as $f(X_1, \ldots, X_k) = \beta_0 + \beta_1 X_1 + \cdots + \beta_k X_k$, ε is the random error usually modeled as a $N(0,\sigma)$ variable and σ is the unknown standard deviation of the random error. It is also common to write the model in matrix form as $f(X_1, \ldots, X_k) = X\beta$ where $\beta = (\beta_0, \beta_1, \ldots, \beta_k)'$ is the vector of unknown parameters, $X = (1, X_1, \ldots, X_k)$ is the $n \times (k+1)$ design matrix where n is the total number of buildings in the model, $X_1, \ldots, X_k$ are the values taken by the explanatory variables in these n buildings (as column vectors) and 1 is a column vectors of ones.

It is very easy to adjust a regression model as any statistical package, or even many spreadsheets can do it in a very easy way. However, not all researchers adjusting regression models check its statistical assumptions (linearity, homoscedasticity, independence of observations and normality) and the possible existence of anomalous points in the model. This is a fundamental point as, if the statistical assumptions of the model are not verified, the conclusions may be completely wrong. Residual analysis is the basic tool for checking these hypotheses and for detecting the existence of atypical points in the model. Occasionally, in order to ensure the validity of the hypotheses or to integrate the atypical values, the response variable or some of the explanatory variables will have