Renewable and Sustainable Energy Reviews 74 (2017) 512–521

Contents lists available at ScienceDirect

Renewable and Sustainable Energy Reviews

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

Statistic linear parametric techniques for residential electric energy demand forecasting. A review and an implementation to Chile

Humberto Verdejo*, Almendra Awerkin, Cristhian Becker, Gabriel Olguin

Department of Electrical Engineering, University of Santiago, Santiago, Chile

ARTICLE INFO

Keywords:
Residential demand
Statistical modelling
Stochastic process
Times Series

ABSTRACT

In operational and planning studies in power electric distribution systems, one of the most important tasks is to quantify the evolution of the system. In particular, it is necessary to be able to measure the growth of electrical demand, with special attention to residential consumption. For that reason, it is fundamental to predict its future values. Considering the availability of real measure data, statistic parametric methods are widely used to describe and forecast those residential loads. This paper reviews the principal statistical linear parametric methods and implements four of them to analyse real measure data from Chilean systems. Additionally, those methods are compared among them and the performance of a non-tested continuous approach based on diffusion processes can be evaluated. In each case, the parametric adjustment and the validation methods are explained.

1. Introduction

At distribution level of an electric power system (EPS), residential energy demand is one of the principal matters in planning and control studies. It is necessary to be able to measure the growth of the grid and understand the consumption behaviour for residential customers [1].

There are different mechanism to address this task, some cases are engineering and statistical approaches. The differences between them are related to how to look at the problems. For example, engineering methods are often used to describe end-uses of the energy, and they are associated to physical models which base their assumptions on the natural knowledge of phenomena. Besides, they do not need any historical consumption information [2]. On the other hand, statistical models are appropriate when data measurements of the object under study are available.

As that is the case for electrical residential energy consumption, statistic methods can be used to develop a representative model, in order to describe the current and future behaviour of this type of load, focused on the real measurement of the variables.

There are a number of statistical techniques to obtain an expression for energy demand. Depending on the nature of equation, these techniques can be classified in: linear and non-linear, discrete and continuous. In addition, according to the adjustment methods used, it can be distinguished between parametric and non-parametric types.

As can be assumed, depending on the study orientation, and the type of variables used, one given tool will be more convenient than the rest. The principal merits and disadvantages of each statistical category mentioned before are summarized in Table 1.

It is important to mention that the classification for the demand models is not standard and it exclusively depend on the point of view from which the different techniques are seen [3,4,2].

Some research which classify and compare different methods can be found in [3,4,2,5]. For example, in [3], a number of approaches using previous observation of the electric demand so as to predict future behaviour, are discussed. This classification separates the models in 7 categories: Fourier series, neuronal networks (NN), Gaussian process, autoregressive, Fuzzy logic, Wavelets, and multiple regression. These groups are contrasted showing their benefits and inconveniences. Additionally, Fourier series and Gaussian process techniques are implemented and compared.

On the other hand, in [4,2]the methods are classified depending on the class of data, amount and the type of variables to be used in the modelling process. In other words, two big groups are distinguished: bottom-up and top-down models, which are sub classified into: deterministic, statistic, engineering, etc. The treated models are compared, in order to determine advantage, and disadvantages of each adjustment.

In [5], they investigated the Conditional Demand Analysis (CDA) technique, which is based on regression models. Furthermore, CDA is compared with the methods of Neuronal Networks (NN) and engineering models.

So far, developments in strategic topics such as: energy efficiency,

* Corresponding author.
E-mail address: humberto.verdejo@usach.cl (H. Verdejo).

http://dx.doi.org/10.1016/j.rser.2017.01.110
Received 21 January 2016; Received in revised form 7 November 2016; Accepted 16 January 2017
1364-0321/

Table 1
Modelling techniques comparison for residential energy demand.

Category	Advantage	Disadvantage
Linear	Easy implementation and interpretation. Estimation techniques well established. Scientific acceptance and a wide range of applicability.	In real world a few phenomena correspond with models assumptions, this leads sometimes, does not really useful results.
Non-linear	More general than linear methods, which gives a major flexibility at the moment to fit a data series.	The function that gives the optimum fit should be determined, this hinders the preparation analysis. Less amount of validations tools: for example, there is not exist a explicit calculus for the R^2 coefficient.
Discrete	Uses and gives more simplified information	Analysis and result too robust for short time intervals studies.
Continuous	Wide application field in the description of any phenomenon	Estimation, simulation and validation techniques more complex and sophisticated For good parametric estimation, a lot data and/or small time intervals between observation are required. Explicit analytic solutions do not always be able and numeric approximations must to be used exist
Parametric	Greater provision of information, due to certain probability distribution is assumed for the data.	To assume that the data come from of a specific probabilistic model, biased conclusions could be obtained if a wrong model is used.
Non-parametric	Less condition about the data should be assumed. Which is better in situations when the truly distribution is unknown or can not be approximated easily.	Limited software implementation. Oriented to hypothesis test, instead of effect estimations. Non-parametric estimations and confidence intervals extraction does not easy.

new technologies, renewable energy sources incorporation and smart grids, keep the residential consumption modelling on constant review. Then, the need of quantify and qualify the performance of the techniques under new scenarios appears. In this context, it is usual to find modern hybrid methods, which are presented in order to cover the new technological frameworks [6].

Some new hybrid methods that have appeared in the literature are based on stochastic differential equations (SDE) [7–9], but they have not been considered in the previous reviews, so one of the questions that this paper addresses is how SDE behaves in residential consumption, and compares it with other known methods.

Thus, in order to contribute to the analysis of electric power demand model and forecasting, when real data is available, this paper implements three of the common linear parametric techniques: multilinear and harmonic regression, and time series analysis, and a fourth hybrid technique based in SDE, with the aim of computing real performances and compare the non-tested SDE method with the rest of the methodologies. With the experimental results, which have been validated by theoretical techniques for the different models, a qualification of common and new approaches is established, in order to extract the advantage and underscore the uses of each method.

The benefit of this procedure is that with statistics methods, the information that can be extracted from the data is captured, and, on the other hand, the work presents an algorithm to validate the models, which gives a measure of quality of the fit, which is used to ensure that the forecast data is a good representation of the present and future values of the residential consumption.

The paper is organized as follows: in Section 2 the principal stochastic linear techniques found in literature are explained, in addition to a concise theoretical-practical presentation of the subject matter. In Section 3 some of the models studied in Section 2 are applied to a real case using real observation of a residential feeder of Santiago, Chile for the parametric estimation. The methodology used is described, including a proposal for the model, parametric adjustment and corresponding validation. In Section 4 the results for forecasting data are compared, and the advantages and disadvantages of each instrument are deduced. Finally, in Section 5 the conclusions of the paper are given.

2. Statistic linear models description

2.1. Regression

Regression models are among the most popular due to their simplicity and easy interpretation of the results obtained. Furthermore, they possess a great flexibility when assigning variables to the equations given to describe a certain behaviour [10].

There are several categories of regression: linear, multilinear, logistic, ridge, Least Absolute Shrinkage and Selection Operator (LASSO), ecological, Least Absolute Deviations (LAD), Bayesian, non-parametric, Wavelets, etc. Some of these techniques are more robust than others, like ridge and LASSO, which are similar to the linear type, but they restrict the regression coefficient values, in a way to make them more natural and to avoid a overfitting. The previous happens when there are more equations than unknown variables.

Another types, like logistic regression, depend on the nature and the variable response. The latter considers binary variable in the models.

As examples of application models adopting these tools, in [11] the use of a multilinear regression model can be found, with the objective of predicting future values of residential energy demand. Other variables related to the use of the energy and other factors concerning household such as: size, number of habitant, etc. are included.

In [12] it also used the multilinear regression methods so as to forecast the electric energy consumption, but with more general variables as those used in the previous case. An analysis of variance (ANOVA) is used as a model construction process.

The logistic regression method is used in [13], where it is applied to analyse the domestic electric consumption types. It is concluded that the focus can be used to estimate the added consumption of household electric demand.

Another variant of these models, that can be included in this review of linear techniques, corresponds to the harmonic regression models, due to the fact that the estimation method can be reduced to the linear case. This methodology is particularly useful to describe processes with marked seasonality [14], as in the case of residential energy consumption. In [15], for example, this technique is used to forecast the energy price values.

With regards to the estimators considered among the regression models, the most famous ones are the moment, maximum likelihood and least squares estimators, and depending on the type of regression to be applied, one technique will be more suitable than the others. For example, in logistic regression, the probabilistic distribution of the dependent variable does not allow a direct connection with the regression line, and the least squares method is not applicable. There are many variations of these methods, depending on the connection between variables, the assumption regarding a particular model, residual nature, among other things, will influence in used technique. More information about other methods or variations of those already known, can be found in [16–18].

The next and final step in regression analysis, is to determinate the goodness of fit in order to obtain appropriate conclusions. Perhaps, the most famous statistic to quantify the good adjustment is the determination coefficient R^2, which provides a measure of the dependence or