Energy and Buildings 108 (2015) 223–235

Contents lists available at ScienceDirect

Energy and Buildings

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

Practical lognormal framework for household energy consumption modeling

P. Kuusela*, I. Norros, R. Weiss**, T. Sorasalmi

VTT Technical Research Centre of Finland Ltd., P.O. Box 1000, FI-020044 VTT, Finland

ARTICLE INFO

Article history:
Received 6 August 2015
Accepted 7 September 2015
Available online 12 September 2015

Keywords:
Energy consumption distribution
Mathematical modeling
Household
Lognormal distributions
Smart meter data
AMR

ABSTRACT

Substantial variability of household energy consumption is essential to recognize. In particular the variability occurs towards large magnitudes. This work focuses on modeling the distribution of household annual energy consumptions and promoting a practical modeling framework resulting to explicit models. Lognormal distributions describe well energy consumption of households that do not use electricity for heating. We model at city district level how consumption distributions depend on district housing characteristics. At individual household level, the joint distribution of the annual electricity consumption and key household characteristic (number of persons, floor area, number of bedrooms), is well approximated by multivariate lognormal distributions. This work utilizes numerous data sources: household annual electricity consumption data for city districts, city summary data on housing and people, data on sold apartments, and Irish Automatic Meter Reading (AMR) data with test survey information. We also demonstrate a process for adjusting the models when data relating individual household energy consumption with other characteristics are not available.

1. Introduction

For adequate understanding of energy consumption, it is important that the models capture features like variability and dynamics. Focusing on mean values alone leads to poor system dimensioning and performance, in particular when the system design involves not only new technologies in energy and telecommunications but also changes in energy usage and policy.

This paper focuses on the variation of annual energy consumption of households and its prediction on the basis of other characteristics of the households. Utilizing a number of data sources of different qualities, we model households' annual electricity consumption both at city district and individual household levels. We make systematic use of the observation that the variability of household energy consumption is surprisingly well described by lognormal distributions, even when measured at different time or population scales. Moreover, we show that some other important characteristics like the floor area and the household size (number of people in the household) can also be reasonably well modeled by lognormal variables, and even their dependences are well captured by correlations in log-transformed space so that multi-lognormal distributions can be applied. To the best of our knowledge, the framework of lognormal modeling of electricity consumption has not been addressed as coherently and explicitly before. Our approach is simple to use and it does not require any advanced algorithms.

This kind of consumption distribution models are valuable tools in energy system planning and dimensioning tasks. They can be applied in the calculation of probabilistic risk estimates. They can also be used as elements in energy system simulations, replacing the detailed simulation of individual consumers. Finally, the few and meaningful parameters of these models may become quantities whose long-term evolution over years is worth studying.

In this paper, all households are selected so that they do not use electricity for heating of space or water. First, for ten city districts of a major Finnish city we use datasets on AMR based household annual electricity consumptions to calculate district level distributions, and relate these distributions to simple district level average housing characteristics in terms of a regression model in a log-transformed space. Second, we consider an Irish smart meter trial data set that combines electricity consumption with other household information and develop a multivariate lognormal model relating the energy consumption with the household floor area, the number of bedrooms and the household size at individual

* Principal corresponding author.
** Corresponding author.
E-mail addresses: pirkko.kuusela@vtt.fi (P. Kuusela), robert.weiss@vtt.fi (R. Weiss).

http://dx.doi.org/10.1016/j.enbuild.2015.09.008
0378-7788/

household level. Although the apartment price level, probably indicating inhabitant wealth level of a district, was found to be a strong predictive factor in the first part of this study, this characteristic was not truly available in the Irish data. Third, we return to the Finnish city at individual household level. Because of data privacy reasons, we do not have any direct link to combine or relate individual energy consumption with the other individual housing characteristics. We solve the challenge by importing the missing parameters from the Irish household model, obtaining a good multi-lognormal model for the Finnish city households as well. This kind of transfer could in fact be a generic way to derive approximate multivariate consumption models for regions where the available data sets do not directly reveal all relevant dependencies, for privacy or other concerns.

The appearance of lognormal distributions in this context has been reported in several studies: the lognormality of annual electricity consumption in [1], the lognormality of hourly electricity consumption in [2,1], lognormal mixtures for daily electricity consumption in [3], and log–log plots of energy consumption vs. living area, together with lognormality of the former, in [4]. In general, many natural phenomena are multiplicative and generate lognormal distributions [5]. On the other hand, power-law behavior, i.e. linear relationships between the logarithms of input and output variables, has been identified in urban supply networks [6] and in most factors of aggregated residential electricity consumption in [7]. Power-law relations provide simple models with performance close to advanced methods, see [4].

Our work in district level electricity consumption has similarities to the work by Boulaire et al. [8] whose regression model uses logarithms of electricity consumption and household size. In contrast to our work, they handle variables solely through their means. A multilevel regression approach is utilized in [9] to model residential energy variations and to understand district variations by household features. Multiple linear regression has been utilized to relate energy and gas consumption to city building stock information [10]. For more detailed literature reviews, we refer to the recent papers [11,8,12], and the modeling review [13].

This paper is organized as follows. The basic lognormal modeling and model development framework is presented in Section 2. In Section 3, we work on the level of city districts in the Finnish city. Section 4 develops the household level model based on the Irish data, while Section 5 builds a corresponding model for the Finnish city household, adopting parameters from the Irish model. A summary of all data sets and models is provided in Section 6 and the concluding remarks in Section 7.

2. Lognormal modeling

2.1. Lognormal distributions in nutshell

If a random variable Z has normal (Gaussian) distribution with mean μ and variance σ^2, then $X = e^Z$ is said to have *lognormal distribution* with *scale parameter* μ and *shape parameter* σ. The mean and variance of X are given as

$$\mathrm{E}[X] = e^{\mu+\sigma^2/2}, \quad \mathrm{Var}[X] = e^{2\mu+\sigma^2}(e^{\sigma^2} - 1). \tag{1}$$

Lognormal distributions appear often with phenomena in the nature, when the quantity in question is positive, has a low mean value and a large variance [5]. The difference between normal and lognormal variability is in the action of independent forces which is additive in the normal case and multiplicative in the lognormal case. If X is lognormal then also aX^b is lognormal with a and b constants. For independent lognormal variables $X_1, \ldots, X_n$, the product $X_1 \cdots X_n$ is lognormal.

If a random vector $\boldsymbol{Z} = (Z_1, \ldots, Z_d)$ has a d-dimensional multinormal distribution with center at $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_d)$ and covariance matrix Γ, then the vector $\boldsymbol{X} = (e^{Z_1}, \ldots, e^{Z_d})$ is said to have a *multivariate lognormal (multi-lognormal) distribution* with parameters $\boldsymbol{\mu}$ and Γ. The rules governing the distributions of affine functions of $\boldsymbol{Z}$ imply corresponding rules for $\boldsymbol{X}$. Denote for an index set $J = \{j_1, \ldots, j_k\} \subset 1, \ldots, d$ $\Gamma(i, J) = (\Gamma(i, j_1), \ldots, \Gamma(i, j_k))$, and define respectively $\Gamma(J, i)$ and the submatrix $\Gamma(J) = \Gamma(J, J)$. Then, for any $i \in \{1, \ldots, d\} \setminus J$, the conditional distribution of Z_i given Z_J is Gaussian. If Z is centered, i.e. $\boldsymbol{\mu} = 0$, we have the relations

$$E[Z_i|Z_J] = \Gamma(i,J)\Gamma(J)^{-1}Z_J,$$

$$\mathrm{Var}[E[Z_i|Z_J]] = \Gamma(i,J)\Gamma(J)^{-1}\Gamma(J,i),$$

$$\mathrm{Var}[Z_i|Z_J] = \mathrm{Var}[Z_i] - \mathrm{Var}[E[Z_i|Z_J]],$$

and the non-centered versions are obtained in the obvious way. Similarly, the conditional distribution of X_i given X_J is lognormal with corresponding parameters, and we note the relations

$$E[X_i|X_J] = \exp\left(\mu_i + \Gamma(i,J)\Gamma(J)^{-1}(Z_J - \mu_J) + \frac{1}{2}(\sigma_i^2 - \Gamma(i,J)\Gamma(J)^{-1}\Gamma(J,i))\right), \tag{2}$$

$$\mathrm{Var}[X_i|X_J] = \exp(2(\mu_i + \Gamma(i,J)\Gamma(J)^{-1}(Z_J - \mu_J)) + \sigma_i^2 - \Gamma(i,J)\Gamma(J)^{-1}\Gamma(J,i)) \times (\exp(\sigma_i^2 - \Gamma(i,J)\Gamma(J)^{-1}\Gamma(J,i)) - 1) \tag{3}$$

2.2. Model development principles

All our approaches are related to lognormal distributions. Note that this framework only requires the estimation of simple bivariate relations and no higher factors regardless of the number of variables in question. When an appropriate multivariate data set is available, like in Section 4, we apply multivariate lognormal modeling to describe the dependent quantities. This amounts to estimating the mean $\boldsymbol{\mu}$ and covariance matrix Γ of the log-transformed data and to forming various types of estimates as conditional expectations using (2), or lognormal distributions with (2), (3) and (1).

In Section 3, we study the consumptions of households in a given district, but we lack data for matching any other information to these households. Instead, we gather the housing information from other data sources and relate electricity consumption to those on the district level.

We use the multiplicative properties of lognormal distributions as a guideline in running regression in the space of log-transformed variables and converting the results back to the original variables in a product form. Our regression variables are usually strongly correlated. The regression is performed by eliminating stepwise a variable whose coefficient has a high P-value in testing statistically the significance of the coefficient differing from zero. The final model is selected by balancing the number of variables and the goodness of fit. Field knowledge has been used to decide what variables to include in the model, instead of a feature selection from a large number of variables.

Table 1 illustrates how the variables and models are transformed between the original space and the log-transformed space.

2.3. Usage of distribution models

The large variability in household electricity consumption and in particular the variability towards large consumptions are essential