

# Modelling the Electricity Consumption in a Manufacturing Company through Probability Distribution

Mária Michalková<sup>[0000-0001-7488-8514]</sup>, Ivana Pobočíková<sup>[0000-0003-4357-260X]</sup>, and Zuzana Sedliačková<sup>[0000-0002-4083-6262]</sup>

University of Žilina, Faculty of Mechanical Engineering, Department of Applied Mathematics, Univerzitná 1, Žilina 010 26, Slovakia

**Abstract.** The paper deals with model of electricity consumption in a company providing hot and cold forming. As modelled data, there were used hourly records of electricity consumed in the manufacturing processes. The data covered working hours over a year. The probability distributions, chosen for modelling, were normal distribution, gamma distribution, logistic distribution, Weibull distribution and Rayleigh distribution. The goodness of fit of these distributions was judged based on the values of the information criteria (*AIC*, *AICc* and *BIC*), coefficient of determination ( $R^2$ ) and the root mean square error (*RMSE*). According to these criteria, the best fit was achieved by the two-parameter Weibull distribution.

## 1 Introduction

The aim of the paper is to model the electricity consumption in a manufacturing company that provides hot and cold bending. Despite the growing electricity demand from residential sector, commercial and public services, and transportation (partially caused by the spreading interest in electric cars), the industrial sector still possesses the leading role as main electricity end-user worldwide [1]. The electricity consumption by industries over the globe has increased by about 0.9% annually between years 2010 and 2018 [2]. Rapid growth has been observed in energy-intensive industrial subsectors where also processes of metal forming belong. With lessening supplies of fossil fuels and only slowly growing incorporation of renewable electricity production sources, such as wind, solar, thermal and biomass, we might witness possible electricity shortage in the future. In order to prevent such scenario, companies need to transform their processes to be more energy-efficient. In literature, there have been proposed several possibilities – rescheduling the processes in order to optimize the peak demand patterns [3-6], supplementing part of the energy supply by renewable energy sources [7-10], applying modern procedures and machines in manufacturing [11-13], etc. Such solutions do not need to be limited to industry, but are applicable also in households. However, optimization of electric energy consumption is conditioned by appropriate electricity consumption model.

The electricity consumption can be considered as a variable with stochastic character. Such assumption is supported by consumption variation in time and place as well as by being influenced by a number of internal and external factors. Due to this fact, several statistical approaches were adopted to modelling the electricity consumption or electricity demand. Particular method of modelling depends on the structure of available data and on the purpose of the analysis. There have been developed energy consumption models from the level of countries [14-17] to level of individual units, such as households, companies, machines [18-20]. Models of electric energy consumption include models based on analysis and forecasting of time series, probability distributions, regression analysis or Markov chains. For review on modelling methods, see for example [21].

In this paper, we model electricity consumption in manufacturing processes recorded over the period of one year. Five probability distributions are fitted to data: normal, Weibull, Rayleigh, logistic and gamma distribution. The goodness of fit of each probability distribution is judged by the Kolmogorov-Smirnov test. The most appropriate probability distribution is chosen based on values of information criteria – Akaike’s information criterion, corrected Akaike’s information criterion and Bayesian information criterion; coefficient of determination and the root mean square error.

The paper is organized as follows: in section 2, the modelled data are described alongside with brief characteristics of probability distributions chosen to fit the data. Moreover, the estimation method for probability distribution parameters is defined and the model selection criteria and goodness of fit test are outlined. Results of the fitting procedure are discussed in section 3. Summarization of the paper is given in section 4.

## 2 Data and methodology

In the following section, we briefly characterize data along with probability distributions that are fitted to the data. As a method for estimation of parameters of each probability distribution, the maximum likelihood method is chosen. Further in the section, the criteria for goodness of fit are listed.

### 2.1 Characteristics of data

Data modelled in this paper represent electricity consumed in manufacturing processes in a company providing hot and cold bending. The data were collected from four production halls and were recorded hourly. They cover the consumption of electric energy (in [kW]) during the working hours throughout one year. According to this, the sample includes 3984 observations. Descriptive statistics of the data are given in Table 1.

As we may see from the coefficient of skewness, the data are positively skewed. This might indicate that the probability distribution fitting the data will be asymmetrical. Such assumption is not surprising since consumptions of low and moderate amount are more common; on the other hand, high consumptions occur seldom.

**Table 1.** Descriptive statistics of modelled data.

Min	Max	Mean	Variance	Quartiles			Skewness	Kurtosis
				1 <sup>st</sup>	2 <sup>nd</sup>	3 <sup>rd</sup>		
31.82	1419.24	501.80	71 640.07	291.11	469.95	679.92	0.48	-0.37

## 2.2. Probability distributions

As authors in [22] noted, a variety of probability distributions can be used to fit the electricity consumption data. In [23], authors fitted two-parameter Weibull and two-parameter lognormal distribution to electricity consumptions in selected Swedish households. Electricity consumption in university campus in Nigeria was modelled by normal, two-parameter Weibull and Gumbel distribution [24]. The electricity loads have been modelled by several probability distributions: in [25], Weibull distribution was fitted to energy consumption of customers in Northern Ireland; lognormal distribution was applied to electricity consumption data in [26]. In [27], a number of distributions – Weibull, normal, inverse-normal, Rayleigh, gamma, lognormal – were fitted to consumption data. Similarly in [28] where data of electricity load in two substations of Venezuelan power system were fitted by normal, Weibull, gamma, Rayleigh and exponential probability distributions, respectively. Probability distributions chosen for fitting to data considered in this paper are summarized in Table 2.

**Table 2.** Characteristics of chosen probability distributions.

Probability distribution	Probability density function $f(x)$ Cumulative distribution function $F(x)$	
Normal $N(\mu, \sigma^2)$	$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$ $F(x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^x \exp\left(-\frac{(t-\mu)^2}{2\sigma^2}\right) dt$	$x \in R,$ $\mu \in R,$ $\sigma > 0$
Weibull $W(a, b)$	$f(x) = \frac{a}{b^a} x^{a-1} \exp\left(-\left(\frac{x}{b}\right)^a\right)$ $F(x) = 1 - \exp\left(-\left(\frac{x}{b}\right)^a\right)$	$x > 0,$ $a > 0,$ $b > 0$
Gamma $Gamma(a, b)$	$f(x) = \frac{1}{b^a \Gamma(a)} x^{a-1} \exp\left(-\frac{x}{b}\right)$ $F(x) = \frac{1}{b^a \Gamma(a)} \int_0^x t^{a-1} \exp\left(-\frac{t}{b}\right) dt$	$x \in [0, \infty),$ $a > 0,$ $b > 0$
Logistic $Logist(\mu, \sigma)$	$f(x) = \frac{\exp\left(\frac{x-\mu}{\sigma}\right)}{\sigma \left[1 + \exp\left(\frac{x-\mu}{\sigma}\right)\right]^2}$ $F(x) = \frac{\exp\left(\frac{x-\mu}{\sigma}\right)}{1 + \exp\left(\frac{x-\mu}{\sigma}\right)}$	$x \in R,$ $\mu \in R,$ $\sigma > 0$
Rayleigh $Ra(b)$	$f(x) = \frac{x}{b^2} \exp\left(-\frac{x^2}{2b^2}\right)$ $F(x) = 1 - \exp\left(-\frac{x^2}{2b^2}\right)$	$x > 0,$ $b > 0$
	$a$ – shape parameter; $b$ – scale parameter; $\mu$ – location parameter; $\sigma$ – scale parameter	

### 2.3 Parameter estimation method

Among the methods for estimation the parameters of probability distributions, the maximum likelihood method (MLM) has been chosen since it is considered as the most common one.

Let  $X_1, X_2, \dots, X_n$  be a random sample from a population with probability density function  $f(\mathbf{x}; \boldsymbol{\theta})$  where  $\boldsymbol{\theta} = (\theta_1, \theta_2, \dots, \theta_p)$  is a vector of parameters,  $\boldsymbol{\theta} \in \Omega$ , and  $\Omega$  is the parameter space. The likelihood function  $L(\mathbf{x}, \boldsymbol{\theta})$  is a function of vector parameter  $\boldsymbol{\theta}$  defined as follows

$$L(\mathbf{x}, \boldsymbol{\theta}) = \prod_{i=1}^n f(x_i, \boldsymbol{\theta}) \tag{1}$$

with  $\mathbf{x} = (x_1, x_2, \dots, x_n)$  being the realization of random sample. If there exists a value  $\hat{\boldsymbol{\theta}} \in \Omega$  such that for all  $\boldsymbol{\theta} \in \Omega$

$$L(\mathbf{x}, \hat{\boldsymbol{\theta}}) \geq L(\mathbf{x}, \boldsymbol{\theta}) \tag{2}$$

then  $\hat{\boldsymbol{\theta}}$  is called the maximum likelihood estimate of parameter  $\boldsymbol{\theta}$ .

When estimating parameters of particular probability distribution via MLM, we need to find values that maximise the likelihood function (1). In order to achieve that, the system of equations in the form

$$\frac{\partial L(\mathbf{x}, \boldsymbol{\theta})}{\partial \theta_i} = 0, \quad i = 1, 2, \dots, p \tag{3}$$

need to be solved. However, it is more convenient to work with the logarithm of likelihood function

$$\ln L(\mathbf{x}, \boldsymbol{\theta}) = \sum_{i=1}^n \ln f(x_i, \boldsymbol{\theta}) \tag{4}$$

called loglikelihood function. Substituting (4) in (3), we obtain a system of likelihood equations for finding the maximum likelihood estimates of parameters. The likelihood equations for each considered probability distribution are given in Table 3.

We can see that shape parameter of Weibull and gamma distribution, respectively, as well as both parameters of logistic distribution cannot be calculated directly. In such cases, the iterative methods are required to find these particular estimates.

**Table 3.** Maximum likelihood estimates of parameters for chosen probability distributions.

Probability distribution	MLM estimates
Normal $N(\mu, \sigma^2)$	$\hat{\mu} = \frac{1}{n} \sum_{i=1}^n x_i,$ $\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n \left( x_i - \frac{1}{n} \sum_{j=1}^n x_j \right)^2$
Weibull $W(a, b)$	$\frac{1}{a} + \frac{1}{n} \sum_{i=1}^n \ln x_i - \frac{\sum_{i=1}^n x_i^a \ln x_i}{\sum_{i=1}^n x_i^a} = 0,$ $\hat{b} = \left( \frac{1}{n} \sum_{i=1}^n x_i^{\hat{a}} \right)^{\frac{1}{\hat{a}}}$
Gamma $Gamma(a, b)$	$\frac{\partial \ln \Gamma(a)}{\partial a} - \ln a + \ln \bar{x} - \frac{1}{n} \sum_{i=1}^n \ln x_i = 0,$ $\hat{b} = \frac{\bar{x}}{\hat{a}},$ $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$
Logistic $Logist(\mu, \sigma)$	$1 - \frac{2}{n} \sum_{i=1}^n \frac{\exp\left(\frac{x_i - \mu}{\sigma}\right)}{1 + \exp\left(\frac{x_i - \mu}{\sigma}\right)} = 0,$ $\sigma - \mu + \frac{1}{n} \sum_{i=1}^n x_i - \frac{2}{n} \sum_{i=1}^n \frac{(x_i - \mu) \exp\left(\frac{x_i - \mu}{\sigma}\right)}{1 + \exp\left(\frac{x_i - \mu}{\sigma}\right)} = 0$
Rayleigh $Ra(b)$	$\hat{b} = \sqrt{\frac{1}{2n} \sum_{i=1}^n x_i^2}$

### 2.4 Model selection criteria

In order to choose the probability distribution that fits the data the best, we apply information criteria, coefficient of determination and root mean square error (Table 4).

**Table 4.** Model selection criteria.

Criterion	Formula
Akaike’s information criterion [29]	$AIC = -2 \ln L(x_1, x_2, \dots, x_n; \hat{\theta}) + 2m$
Akaike’s information criterion corrected	$AICc = AIC + \frac{2m(m + 1)}{n - m + 1}$
Bayesian information criterion [30]	$BIC = -2 \ln L(x_1, x_2, \dots, x_n; \hat{\theta}) + m \ln n$
Coefficient of determination	$R^2 = \frac{\sum_{i=1}^n [\hat{F}(x_i) - \bar{F}]^2}{\sum_{i=1}^n [\hat{F}(x_i) - \bar{F}]^2 + \sum_{i=1}^n [F_n(x_i) - \hat{F}(x_i)]^2}$
Root mean square error	$RMSE = \left\{ \frac{1}{n} \sum_{i=1}^n [F_n(x_i) - \hat{F}(x_i)]^2 \right\}^{\frac{1}{2}}$

Here  $X_1, X_2, \dots, X_n$  represent a random sample from a population with cumulative distribution function  $F(x)$  and  $x_1, x_2, \dots, x_n$  are respective observations. Term  $L(x_1, x_2, \dots, x_n; \hat{\theta})$  is the maximum value of likelihood function (1) for estimated model,  $m$  is number of estimated parameters,  $n$  is the sample size. Further, for coefficient of determination  $R^2$  and  $RMSE$ , the term  $\hat{F}(x)$  represents the estimated cumulative distribution function with  $\bar{F}$  being the average value defined as

$$\bar{F} = \frac{1}{n} \sum_{i=1}^n \hat{F}(x_i). \tag{5}$$

Function  $F_n(x)$  is the empirical distribution function

$$F_n(x) = \frac{1}{n} \sum_{i=1}^n I(x_{(i)} \leq x) \tag{6}$$

where  $I(x_{(i)} \leq x) = 1$  if inequality  $x_{(i)} \leq x$  holds; otherwise 0. Values  $x_{(1)}, x_{(2)}, \dots, x_{(n)}$  are observations in ascending order.

When choosing the best fitting model among the considered ones according to given criteria, we judge by the values of these criteria as follows – we select the model with lowest values of  $AIC$ ,  $AICc$ ,  $BIC$  and  $RMSE$  and with  $R^2$  closest to 1.

However, these criteria provide only partial information. Their main disadvantage is that the value of information criterion alone does not express whether the probability distribution fits the data properly or not. Therefore, it is optimal to combine the information criteria with goodness-of-fit tests. Among the existing tests, we opt for Kolmogorov-Smirnov goodness-of-fit test (KS test). The test determines whether the data follow given probability distribution with cumulative distribution function  $\hat{F}(x)$ (null hypothesis), against the alternative that they do not come from such distribution. The null hypothesis is rejected on significance level  $\alpha$  when the test statistic

$$D_n = \max_{1 \leq i \leq n} \left\{ \left| \frac{i}{n} - \hat{F}(x_{(i)}) \right|, \left| \hat{F}(x_{(i)}) - \frac{i-1}{n} \right| \right\} \tag{7}$$

is greater than the critical value of KS test. Usually, the  $p$ -value is used for comparison with null hypothesis being rejected at significance level  $\alpha$  when  $p$  – value  $< \alpha$ .

### 3 Results and Discussion

Parameter estimates obtained by MLM, along with the information criteria, coefficient of determination and the root mean square error are summarized in Table 5. All the calculations were done in software Matlab R2019b.

**Table 5.** Parameter estimates and values of model selection criteria for chosen probability distributions.

Probability distribution	Parameter estimates	Model selection criteria				
		<i>AIC</i>	<i>AICc</i>	<i>BIC</i>	$R^2$	<i>RMSE</i>
Normal $N(\mu, \sigma^2)$	$\hat{\mu} = 501.801$ $\hat{\sigma}^2 = 71640.27$	55 847.871	55 847.874	55 860.451	0.989	0.031
Weibull $W(a, b)$	$\hat{a} = 565.796$ $\hat{b} = 1.952$	55 430.672	55 430.675	55 443.252	0.999	0.010
Gamma $\text{Gamma}(a, b)$	$\hat{a} = 2.915$ $\hat{b} = 172.170$	55 602.996	55 602.999	55 615.576	0.994	0.023
Logistic $\text{Logist}(\mu, \sigma)$	$\hat{\mu} = 487.249$ $\hat{\sigma} = 155.805$	56 008.041	56 008.044	56 020.621	0.992	0.027
Rayleigh $Ra(b)$	$\hat{b} = 402.136$	55 432.418	55 432.419	55 438.708	0.998	0.013

According to values of model selection criteria, considered probability distributions are ranked from best to worst as follows:

1. Weibull
2. Rayleigh
3. Gamma
4. Normal/Logistic
5. Logistic/Normal

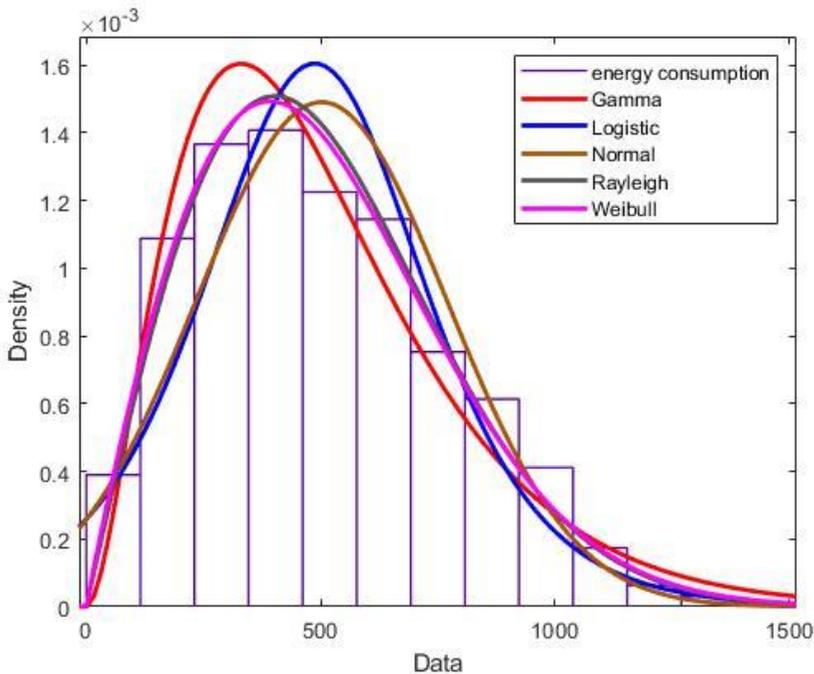
In two last places, there are placed two probability distributions because normal distribution is better than logistic with respect to information criteria; on the other hand, logistic distribution achieves better values of  $R^2$  and *RMSE*. However, these distributions are less suitable for modelling the electricity consumption as both of them extend to negative values.

Results of KS test are summarized in Table 6. Critical value of the test is 0.0215 on significance level  $\alpha = 0.05$ . Obviously, Weibull distribution significantly fits the electricity consumption data on given significance level  $\alpha$ . However, in case of Rayleigh distribution, we cannot unequivocally decide whether to reject or not the null hypothesis because the value of test statistic is very close to critical value ( $p$ -value very close to significance level). In order to make decision upon goodness of fit, we apply Anderson-Darling goodness-of-fit test on significance level  $\alpha = 0.05$ . As  $p$ -value of the test is 0.0045, we may conclude that Rayleigh distribution does not fit the data significantly. Other three probability distributions do not significantly fit the modelled data based on the results of KS test.

**Table 6.** Test statistics and  $p$ -value of KS test for chosen probability distributions.

Probability distribution	Test statistics	$p$ -value
Normal $N(\mu, \sigma^2)$	0.0549	7.1432e-11
Weibull $W(a, b)$	0.0170	0.1985
Gamma $\Gamma(a, b)$	0.0438	4.4276e-7
Logistic $Logist(\mu, \sigma)$	0.0510	1.8459e-9
Rayleigh $Ra(b)$	0.0214	0.0516

Fit of all considered probability distributions to energy consumption data is visualised in Figure 1.



**Fig. 1.** Probability distributions fitted to electric energy consumption data.

## 4 Conclusions

In the paper, data of the electricity consumption in manufacturing processes of hot and cold bending were modelled. The data covered hourly consumption in working hours throughout one year. They were modelled by fitting probability distributions. As tested distributions, following were selected: two-parameter Weibull distribution, two-parameter gamma distribution, normal distribution, logistic distribution and one parameter Rayleigh distribution. As criteria for selecting the best fitting model, information criteria – Akaike’s information criterion, Akaike’s information criterion corrected, Bayesian information criterion, coefficient of determination and root mean square error – were considered. In order to assess the goodness of fit of each distribution to modelled data, the Kolmogorov-

Smirnov test was realized. According to model selection criteria, Weibull distribution provides the best fit, closely followed by Rayleigh distribution, with gamma distribution being the third best. However, only Weibull distribution passed the KS test, i.e. this distribution significantly fitted the data. Thus, we may conclude that electricity consumption data from manufacturing processes in considered company follows Weibull distribution with shape parameter  $\hat{a} = 565.796$  and scale parameter  $\hat{b} = 1.952$ .

## Acknowledgements

This paper was supported by the Slovak Grant Agency KEGA through the project No. 027ŽU-4/2020 Innovation of learning text and implementation of new didactical means to improve the quality of Mathematics II at the university of the technical type, and project No. 040ŽU-4/2018 Implementation of progressive methods for education of university level mathematics for engineers.

## References

1. International Energy Agency: Electricity Information 2019, [https://www.iea.org/reports/electricity-information-2019\\_-](https://www.iea.org/reports/electricity-information-2019_-) access 01.07.2020.
2. P. Levi, T. Vass, H. Mandova, A. Gouy Tracking Industry <https://www.iea.org/reports/tracking-industry-2020> (2020) - access 24.07.2020.
3. A.A.G. Bruzzone, D. Anghinolfi, M. Paolucci, F. Tonelli Energy-aware scheduling for improving manufacturing process sustainability: A mathematical model for flexible flow shops. *CIRP Annals – Manufacturing Technology*, 61(1), 459–62 (2012).
4. K. Fang, N. Uhan, F. Zhao, J.W. Sutherland A new approach to scheduling in manufacturing for power consumption and carbon footprint reduction. *J Manuf Syst* 30(4), 234–40 (2011).
5. S. Ashok Peak-load management in steel plants. *Appl Energy*, 83(5), 413–24 (2006).
6. F. Shrouf, J. Ordieres-Meré, A. García-Sánchez, M. Ortega-Mier Optimizing the production scheduling of a single machine to minimize total energy consumption costs. *J Clean Prod* 67, 197–207 (2014).
7. International Renewable Energy Agency: Renewable Energy Options for the Industry Sector: Global and Regional Potential until 2030, [https://irena.org/-/media/Files/IRENA/Agency/Articles/2016/Nov/IRENA\\_RE\\_Potential\\_for\\_Industry\\_BP\\_2015.pdf?la=en&hash=1214D8FDBD507297FC61073DACE78F8F31927663](https://irena.org/-/media/Files/IRENA/Agency/Articles/2016/Nov/IRENA_RE_Potential_for_Industry_BP_2015.pdf?la=en&hash=1214D8FDBD507297FC61073DACE78F8F31927663) - access 28.07.2020.
8. N. Preston, A. Maroufmashat, H. Riaz, S. Barbouti, U. Mukherjee, P. Tang, J. Wang, E. Hagi, A. Elkamel, M. Fowler How can the integration of renewable energy and power-to-gas benefit industrial facilities? From techno-economic, policy, and environmental assessment. *Int J Hydrog Energy* 45(51), 26559–26573 (2020).
9. P. Ďurčanský, R. Nosek Modelling and application of Stirling engine with renewable sources in electricity production. *Structure and Environment*, 10(4), 386–89 (2018).
10. C. McMillan, R. Boardman, M. McKellar, P. Sabharwall, M. Ruth, S. Bragg-Sitton, Generation and Use of Thermal Energy in the U.S. Industrial Sector and Opportunities to Reduce its Carbon Emissions. Technical report, Joint Institute for Strategic Energy Analysis, <https://www.nrel.gov/docs/fy17osti/66763.pdf> - access 28.07.2020.
11. J. Paralikas, K. Salonitis, G. Chryssoulouris Energy efficiency of cold roll forming process. *Int J AdvManufTechnol*, 66, 1271–84 (2013).

12. P. Albertelli, M. Strano Tube bending machine modelling for assessing the energy savings of electric drives technology. *J Clean Prod* 154, 83–93 (2017).
13. N. Diaz, E. Redelsheimer, D. Dornfeld, Energy consumption characterization and reduction strategies for milling machine tool use. In: Hesselbach J., Herrmann C. (Eds) 18th CIRP International Conference on Life Cycle Engineering. Globalized Solutions for Sustainability in Manufacturing, pp. 263–67. Springer, Berlin, Heidelberg (2011).
14. J. Campillo, F. Wallin, D. Torstensson, I. Vassileva Energy demand model design for forecasting electricity consumption and simulating demand response scenarios in Sweden. In: *IntConf on Applied Energy ICAE 2012*, Suzhou, China (2012).
15. C. Peñasco, P. del Río, D. Romero-Jordán Gas and electricity demand in Spanish manufacturing industries: analysis using homogenous and heterogenous estimators. *Utilities Policy* 45, 45–60 (2017).
16. G. Bölük, A.A. Koç Electricity demand of manufacturing sector in Turkey: A translog cost approach. *Energy Economics* 32(3), 609–15 (2010).
17. F. Boulaire, A. Higgins, G. Foliente, C. McNamara Statistical modelling of district-level residential electricity use in NSW, Australia. *Sustain Sci*, 9(1), 77–88 (2014).
18. J. Munkhammar, J. Rydén, J. Widén On a probability distribution model combining household power consumption, electric home-charging and photovoltaic power production. *Appl Energy* 142, 135–43 (2015).
19. J.V. Paatero, P.D. Lund A model for generating household electricity load profiles. *Int J Energy Res* 30(5), 273–90 (2006).
20. C.A. Babu, S. Ashok Peak load management in electrolytic process industries. *IEEE Transactions on Power Systems* 23(2), 399–405 (2008).
21. L. Suganthi, A.A. Samuel Energy models for demand forecasting – A review. *Renew SustEnerg Rev* 16(2), 1223–40 (2012).
22. R. Singh, B.C. Pal, R.A. Jabr Statistical representation of distribution system loads using Gaussian mixture model. *IEEE Trans Power Syst* 25(1), 29–37 (2010).
23. J. Munkhammar, J. Rydén, J. Widén Characterizing probability density distributions for household electricity load profiles from high-resolution electricity use data. *Appl Energy* 135, 382–90 (2014).
24. J.O. Okeniyi, A.A. Atayero, S.I. Popoola, E.T. Okeniyi, G.M. Alalade Smart campus: Data on energy generation costs from distributed generation systems of electrical energy in a Nigerian University, *Data in Brief* 17, 1082–90 (2018).
25. G.W. Irwin, W. Monteith, W.C. Beattie Statistical electricity demand modelling from consumer billing data. *IEE Proceedings C (Generation, Transmission and Distribution)* 133(6), 328–35 (1986).
26. A. Seppälä Statistical distribution of customer load profiles. In: *Proceedings 1995 IntConf on Energy Management and Power Delivery EMPD '95 (Singapore)*, 2, 696–701. IEEE (1995).
27. E. Carpaneto, G. Chicco Probability distributions of the aggregated residential load. In: *IntConf on Probabilistic Methods Applied to Power Systems (Stockholm, Sweden)*, pp. 1–6. IEEE (2006).
28. F. Gonzales-Longatt, J. Rueda, I. Erlich, W. Villa, D. Bogdanov Mean variance mapping optimization for the identification of Gaussian mixture model: test case. In: *6-th IEEE IntConf Intelligent Systems (Sofia)*, pp. 158–63. IEEE(2012).
29. H. Akaike A new look at the statistical model identification. *IEEE Transactions on Automatic Control* 19(6), 716–23 (1974).
30. G. Schwarz Estimating the dimension of a model. *Annals of Statistics* 6(2), 461–64 (1978).