Estimating Hungarian Household Energy Consumption Using Artificial Neural Networks

András Szűts

Doctoral School of Applied Informatics and Applied Mathematics
Óbuda University
Bécsi út 96/b, H-1034 Budapest, Hungary
szuts.andras@phd.uni-obuda.hu

István Krómer

Kandó Kálmán Faculty of Electrical Engineering, Óbuda University
Bécsi út 96/b, H-1034 Budapest, Hungary
kromer.istvan@kvk.uni-obuda.hu

Abstract: In the European Union and Hungary more than one-third of the total energy consumption comes from households. Therefore, during both the planning of energy efficient investments and the design of energy production and consumption, one of the most important factors is estimating the rate of consumption. The determination of the exact consumption of households has not been achieved because the rates of consumption are calculated solely on the basis of technical parameters, and the results of these calculations are not sufficiently accurate or reliable, especially considering the uncertainties arising from consumer habits. To resolve the issue, we have created a database structure and a calculation model that helps to estimate a household's annual energy consumption based on factors that we have defined. Obviously, this does not mean that calculation based on technical parameters is now unnecessary, but applying the two methods together can significantly increase the accuracy of the estimation of household energy consumption at the individual, regional or national level.

Keywords: households; energy consumption; neural networks

1   Introduction

A current issue in the estimation and calculation of household energy consumption, is the lack of correct data and unified databases. Information about consumers, as well as, the information and databases related to technical solutions are virtually absent, or unavailable in a standardized form and, therefore, useless

for comparative analysis. In the United States, under the management of the National Renewable Energy Laboratory (NREL), a unified database in both software and manual on-demand format was created that contains data of various building materials; it provides the necessary information for the proper functioning of the energy simulation programs. The purpose for creating such a database is for use as a unified national repository, during the energy efficient renovation of residential buildings. It provides an appropriate basis for the cost-benefit analysis of investments.

During the planning of such investments, knowledge of environmental factors and the technical parameters of building materials are not sufficient to determine the energy consumption of households. In our opinion, a knowledge of the consumer and their consumption behavior is at least as important, if not more important, than the technical parameters. Thus, in order to calculate the expected energy consumption, it is important to create a database based on real consumer data and an inference system built upon it. In this article we attempt to provide a solution to eliminate the above mentioned information deficit related to consumers and to utilize this information captured. It is important to clarify that in our research the energy consumption of a household is understood as the total energy consumed. Therefore, we do not make such exceptions in the calculation that exist in regulation 7/2006 TNM, commonly used in Hungary, according to which, in the case of residential buildings, the energy demands of lighting does not need to be included in the calculation of the energy performance of the building, even though it may also be a significant part of a household's energy demand.

It must be noted that to manage the problem outlined above, with a single model, applicable world-wide, is only possible in a theoretical sense, because the factors that significantly impact energy consumption will vary across different geographic and social regions. Therefore, this article looks for a solution considering only Hungarian conditions. As the method is elaborated, it will be clear that the algorithm can be used in other countries or regions by simply gathering the appropriate data concerning the characteristic features for that area and society.

2   The Structure of Household Energy Consumption

According to our research, in order to ensure the effectiveness of investments in energy efficiency it is essential to know the structure of household energy consumption, since this is the way to identify the areas of energy saving potential and to achieve the desired decrease in consumption. Most of the households in Hungary generally use primary (e.g. gas, wood) and secondary (e.g. electricity) energy sources. In extreme cases, the use of only one kind of energy source is possible; however, it is relatively rare in Hungary due to the widespread

distribution of the natural gas service. The energy of the primary and secondary energy sources can be divided into three categories according to their use [2]:

- Energy consumption for heating, cooling and ventilation systems which help to ensure adequate air comfort in living spaces;

- Energy required for the production of domestic hot water;

- Energy required to operate household electrical appliances (e.g. fridge, stove) and lighting.

The above listed factors are obviously not completely independent from each other, since the lighting or any household electrical appliance can certainly produce some degree of heat energy, which may thus influence the energy needs of heating and cooling systems [2]. Thus, the areas with potential for energy savings can be grouped according to the previous list, which can be even further detailed by breaking it down to the energy sources used.

According to G. Swan et al. [2], the energy demand for heating and cooling basically depends on climatic conditions and the technical characteristics of the building, while the second two groups of the list are linked to consumer behavior, the household appliances, the demographic conditions and the number of the household residents. Based on these considerations, he finds the methods for determining the energy needs for indoor comfort based on engineering calculations sufficient; but he recommends statistical methods to determine the energy needs for hot water, household appliances and lighting. However, in contrast with the positions of the studies of G. Swan et al. [2], we believe we can benefit from statistics when considering all three factors. Our reasons are detailed herein.

Looking at households and energy consumption, the uncertainties of consumer behavior are present at virtually all levels, since it does not matter how effective the heating system of a house is if the residents have unrealistic expectations about achievable air comfort in the interior space or they improperly use the systems that are designed for major energy savings. Thinking through the above-mentioned example and similar cases, it is easy to see the importance of the uncertainties that derive from consumer behavior regarding the heating and cooling energy demand and that, according to our knowledge, can be solved by using statistical methods.

In regard to domestic hot water, we agree with the author of the study [2]. As regards household appliances and lighting, it is important to mention that these are an extremely important part of this research, since according to current practice in Hungary, this factor is not taken into consideration at all during the examination of household energy consumption. This importance is confirmed by a 2012 European Union report [3] that established that the achievable saving in this area is 12% of the total achievable potential savings in the residential sector. Since in our study we consider the total energy consumed by households, this factor is also part of the calculation model established.

3   Selection of the Proper Method

It seems to be obvious that some sort of IT-based solution is necessary to address the problems of the information gap that emerged during our research. According to research from international sources, multiple models and multiple methods can be used to estimate the energy consumption of a consumer unit (in this case a household). The basic properties of the different models are summarized in the following table based on Zhao and Magoulès [4].

Table 1
Models to calculate the energy consumption of buildings, and their attributes [4]

Methods	Model complexity	Running speed	Inputs needed	Accuracy
Elaborate Engine	Fairly high	Low	Detailed	Fairly high
Simplified Engine	High	High	Simplified	High
ANN	High	High	Historical data	High
SVM	Fairly high	Low	Historical data	Fairly high

Currently the most common method in Hungary for determining the energy consumption of buildings is calculation according to regulation 7 / 2006th TNM, which, with a little exaggeration, corresponds to the simplified simulation method detailed in Table 1 above. The calculation can be done relatively quickly, even in the case of limited data availability. However, given the results and limited accuracy, we believe there is a need to develop other methods. The greatest drawback with the method required by Hungarian law is that the calculation only considers the technical parameters, and thus fully excludes the impact of consumer behavior on energy consumption. If we consider the findings of the above section, examining the structure of household energy consumption, the consumers have an impact at least as important as the technical factors.

The full and detailed simulation models are also widely used methods. However, due to the time and expense of their use, they are only utilized for research and projects with considerable financial support. Their computational demand is very high, and therefore the running speed is too slow, so their use is not practical in the case of a large information demand. Artificial neural network-based (ANN) techniques are widely used in various calculations of energy consumption and for the approximation of energy consumption functions. Despite the fact that we are talking about a relatively complex model, thanks to today's IT systems with high processing speeds, it is possible to use IT solutions for solving the problems we have highlighted after forming the appropriate database and modeling. The SVM (support vector machine) model, like the artificial neural network based models, relies on a statistical information-based database, but the computational complexity and speed are also a problem in the large-scale application of the system.

After considering the various computation methods, our choice for the estimation of household energy consumption based on non-technical factors is an artificial neural network based inference system, the support for which is confirmed by Aydinalp-Koksal and Ugursal [5], according to whom, such models are suitable for considering technical parameters, as well as, social and economic conditions.

4 Determination of Energy Consumption Affecting Factors

The first step in building the necessary database for operating an artificial neural network based inference system is to identify the non-technical factors that affect household consumption. Research was carried out on the relevant international and national literature, and it was determined that Aydinalp et al. [6] best outline the factors affecting a household's energy consumption, beyond the technical parameters. These factors are as follows:

- Size of the settlement
- Form of the housing (owner, renter)
- Number of children living in the housing
- Number of the adults living in the housing
- Income level of the household
- Type of the housing/flat
- Size of the housing
- Number of the people staying at home during day-time hours

Adapting their conclusions to local Hungarian conditions and taking into consideration the availability of the statistical information required for the database to operate the inference system, the factors listed above were modified as follows:

- Size of the settlement
- Form of the housing (owner, renter)
- Number of people living in the housing
- Number of children living in the housing
- Number of household members actively working/employed
- Highest education level of the head of the household
- Income group of the household,
- Size of the actual living space of the property
- Type of the housing/flat

The factors specified in the list above are the backbone of the database and the research, according to which data collection should be performed, either to the desired geographical area or even to the whole country. Of course it is also necessary to include in the model the total energy consumption of the household,

to which the items of the above list are contributing factors. In case we do not want to use our system to determine the overall energy consumption, it is possible, for example, to group by energy sources or even by energy consumption purposes, as needed to solve the given task. The resulting database, with the adequate amount of sample data, provides the basic data for the operation of the neural network based inference system.

5   Data Source

To develop the database needed for the proper operation of the neural network, we basically relied on two sources. The primary source was the 2011 census conducted by the Central Statistical Office (CSO), which needed to be supplemented by a secondary data source, data from a 2011 survey on household consumption, also prepared by the CSO. The census presumably shows a realistic picture of domestic real estate holdings, family statistics, demographic data and other conditions, as participation and response was mandatory for the entire Hungarian population. Consequently, the backbone of the database was created based on the processing of these data, which define seven out of the earlier established nine factors for our work.

The other household consumption survey was unfortunately done on a smaller sample, but since in Hungary no comprehensive and detailed survey on income and consumption data has been done, it is the only source of statistics that includes feasible data for this research. The simultaneous use of two sources of statistics might appear strange, but there are plenty of examples in the international literature, for similar reasons, as we have listed. Also, on the topic of energy consumption, Aydinalp Koksal and Ugursal [5] are developing their own database based on two sets of statistics in an examination of the energy consumption of Canadian households.

Returning to the already mentioned income and consumption statistics, the sometimes lack of detailed information causes some uncertainty in the data accuracy, but this has no effect on the model or on the operation of the developed system, since in the future, a new, accurate, private or institutional survey for data input is possible in order to enhance the size of the database and thus the computational precision of the model. However, regarding the income and consumption conditions, it should be noted that in Hungary the rate of hidden income is measurable, and the presence of a black or gray economy can distort the statistics, as can the illegal purchase of energy. The preparation of a comprehensive energy consumption and income survey that can be combined with a real estate portfolio would be useful but would certainly raise concerns about the preservation of anonymity and data protection.

6 The Structure and Use of the Database

The process of the development of the database structure can be tracked on the flowchart shown in Fig. 1. Here we can observe the details for defining the various factors and the relationship between them, which is only needed to prepare the starting, training database necessary for the artificial neural network based inference system. In the future, a representative survey could significantly help to populate the database with real data, and the database could later also be expanded with the results from investments in energy efficiency. Since the preparation of such a survey is beyond both the time frame and financial resources of this current research, at present we rely solely on statistics. The spread of smart metering could help create a useful database, as connecting the consumption data with the household characteristics would supply us with detailed energy consumption data. However, this also raises data protection and legal issues.
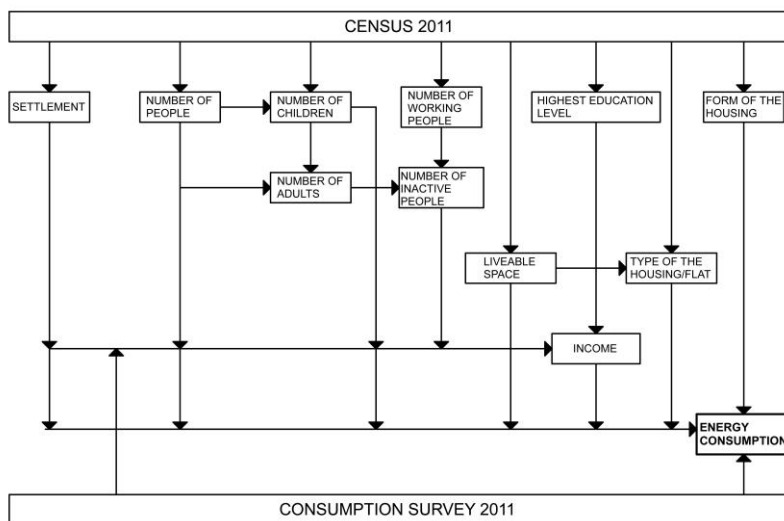

Figure 1
The block diagram of the database structure

On the basis of the above-described method, we have created a database of 1,000 samples, a sample size that in our experience satisfies the margin of error requirements in the training of artificial neural networks. In order to train and test the inference system, we need to divide the database into three parts. The largest part, 70%, is used to train the net, while 15% each is designated for checking and testing. Thus, from the database of 1,000 samples, we randomly assign the sample data in the given proportion to the defined groups. The greater the number of samples we have in our database for training, the greater will be the accuracy of the results and the fewer will be the number of errors that we can expect at the preparation of the model. Nevertheless, a database of 1,000 samples provides good accuracy, as we will demonstrate later.

7　The Neural Network Model

To create the artificial neural network based model, we used the Matlab software package, "Neural Network Toolbox", due to the availability of the devices. With the neural network module creator in the software package, we created a multilayer feed-forward network. The input layer consists of sensory units that are connected to a hidden layer of sigmoid transfer function neurons, which connects to an output layer consisting of linear transfer function neurons. [7] The sigmoid transfer function neurons are required because of the nonlinear nature of the problem; in the case of using linear transfer function neurons, we could also use a single-layer network, although only for linear approximation [7]. The sigmoid function used is defined in the software as follows:

$$tansig(x) = \frac{2}{1 + e^{-2x}} - 1$$

According to our research, this type of neural network model is capable of solving multi-dimensional problems if there are sufficient hidden neurons and a consistent data set. The scheme of the neural network used with 10 neurons in the hidden layer we can see in Fig. 2.
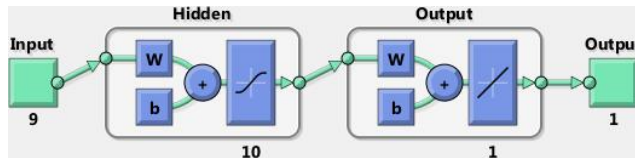

Figure 2

The scheme of the neural network used with 10 neurons in the hidden layer

The network is trained with the software according to the Levenberg-Marquardt backpropagation method that is based on the database discussed in the previous paragraph. The training mechanism is shown in Fig. 3. The number of iterations during the application of the algorithm depends on the degree of the gradient vector, the mean square error, or if these two values are not adequate, on the number of the pre-set maximum cycle number. The Levenberg-Marquardt algorithm can be called into the program with the "trainlm" command, which approximates the weight of the connections of neurons in the network as follows:

$$\mathbf{x}_{k+1} = \mathbf{x}_k - [\mathbf{J}^T \cdot \mathbf{J} + \mu \cdot \mathbf{I}]^{-1} \cdot \mathbf{J}^T \cdot \mathbf{e}$$

where J is the Jacobian matrix that contains the first derivatives of the network errors by weights, μ is the scalar, I is the identity matrix, and e is the vector of network errors.
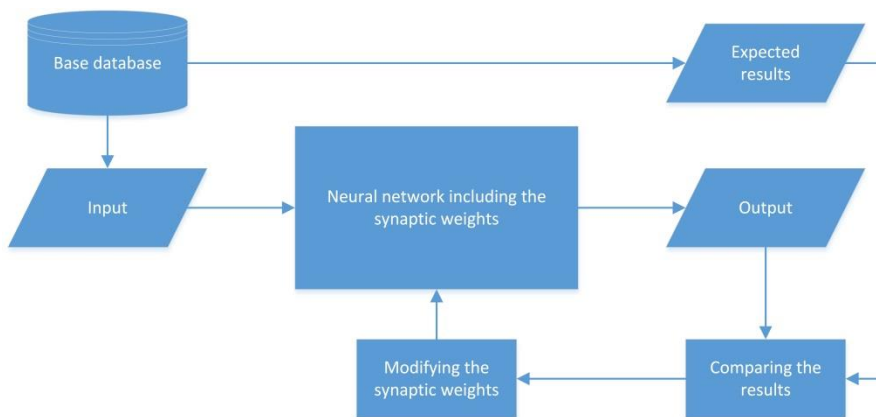
Figure 3

Neural network training with the Levenberg-Marquardt backpropagation method

The model was adjusted in a number of ways, in order to find the ideal number of neurons in the hidden layer. During modeling, this step is primarily required to maintain computational speed within reasonable limits. It should also be kept in mind that a "too complicated" model for a given problem not only decreases the running speed, but it can also significantly degenerate the accuracy of the estimations. Table 2 presents the results after the analysis of three significantly different models, thus supporting our assumptions as to the number of hidden neurons. During model testing the best results were obtained when using 10 hidden neurons, so the same structure is used hereafter.

Table 2

The performance of different models according to the number of hidden neurons

The number of hidden neurons	MSE - Training	MSE - Validation	MSE - Testing	Average difference
5				
Training cycle 1.	19,1280	17,1726	19,0161	
Training cycle 2.	19,5039	22,9476	18,0702	
Training cycle 3.	17,4450	29,9349	19,1352	
Training cycle 4.	16,7433	22,1691	23,4732	
Training cycle 5.	21,9828	19,6401	23,8509	
Mean	18,9606	22,3729	20,7091	**8,5085**
10				
Training cycle 1.	15,1974	20,9568	25,2963	
Training cycle 2.	18,1158	21,2136	16,7463	
Training cycle 3.	17,1003	22,0410	17,7402	
Training cycle 4.	17,3442	19,9203	17,8749	
Training cycle 5.	19,0449	18,7782	14,5440	
Mean	17,3605	20,5820	18,4403	**8,0301**

20				
Training cycle 1.	17,0109	15,3666	22,7337	
Training cycle 2.	16,2405	19,0206	24,9339	
Training cycle 3.	14,0937	26,0877	25,3026	
Training cycle 4.	17,4423	23,5560	18,0054	
Training cycle 5.	16,7802	18,5667	22,2432	
Mean	16,3135	20,5195	22,6438	**20,7830**

MSE - Mean Squaded Error

8 Results

We have created the artificial neural network-based model so that it can be embedded into the Matlab Simulink module. In our case, it is not necessary to create the Simulink model in order to estimate the household consumption because, as Fig. 4 illustrates, the operation requires only one input and one output variable. In further research, the option of embedding is an important aspect, since our neural network inference system can be part of a more comprehensive energy model.
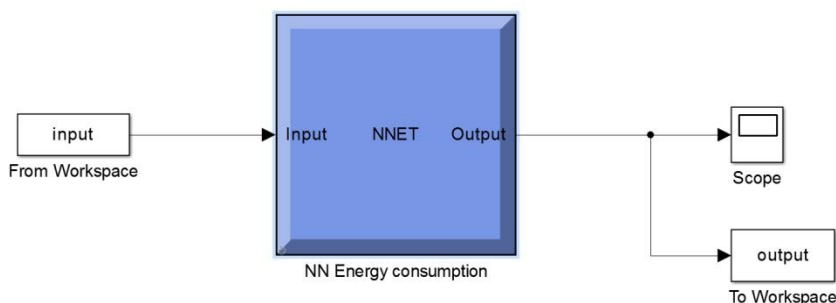

Figure 4
The embedded neural network model

Table 5 and Fig. 3 illustrate how the system works. We defined fictitious households in the sample database and then ran the inference system according to established criteria. We obtained the values presented in the last column of Table 3 and which we show as a bar chart in Fig. 5 for better visibility. We can see from the sample how the different types of settlements, the highest education level of the head of the household, the number of household members and other factors can influence changes in the annual energy consumption. Let us examine a random example: a household in Budapest in a 25 m^2 apartment, containing one working person with a university degree, and an energy consumption of 10,763 kWh of energy on an annual basis, which is considered realistic.

Table 3

The sample database to test the model and results

Number of sample	Settlement *1 = capital* *2 = county town* *3 = other towns* *4 = village*	Form of the housing *0 = renter* *1 = owner*	Number of people	Number of children	Number of working people
1	4	1	2	1	1
2	4	1	3	1	1
3	4	1	4	2	1
4	4	1	5	2	1
5	4	1	6	2	1
6	3	1	1	0	1
7	3	1	2	0	1
8	3	1	3	1	1
9	3	1	4	2	1
10	3	1	5	2	1
11	2	0	1	0	1
12	2	1	2	0	1
13	2	1	3	1	1
14	2	1	4	2	1
15	2	1	4	2	1
16	1	1	1	0	1
17	1	1	2	0	1
18	1	1	3	0	1
19	1	1	4	1	1
20	1	0	5	2	1

Number of sample	Highest education level *0 = nothing* *1 = primary school* *2 = secondary school* *3 = college graduate*	Income decile group	Liveable space *[m²]*	Type of the housing *0 = house* *1 = flat*	Energy Consumption Results *[kWh/household* year]*
1	0	3	70	0	14831
2	0	4	70	0	24063
3	0	5	70	0	32188
4	0	6	90	0	33747
5	0	7	105	0	34447
6	1	3	45	1	8197
7	1	4	55	1	13876

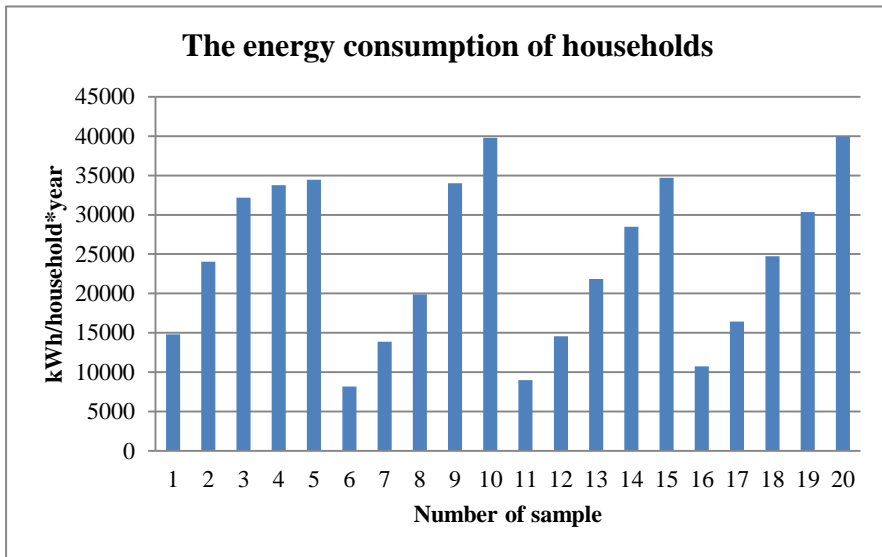|					
8	1	5	55	1	19910
9	1	6	70	0	34026
10	1	7	70	0	39764
11	2	4	35	1	8988
12	2	5	35	1	14576
13	2	6	45	1	21842
14	2	7	45	1	28490
15	2	7	70	0	34716
16	3	6	25	1	10763
17	3	7	35	1	16442
18	3	8	45	1	24717
19	3	9	55	1	30343
20	3	10	70	1	39933

Figure 5
The results belong to the sample database

Conclusions

In summary, we can conclude that the method described in this work may be suitable for estimating the annual energy consumption of households based on statistical data, helping the goals of energy-efficiency and energy-saving programs. The developed model and the defined component factors or principals can be the basis for an accurate estimation of a household's energy consumption in other cases besides the one described; for example, the model can be used in the design phase for a new building to estimate the future energy consumption of the

households, though this is possible only if the prospective owners are known. Based on the obtained consumption results, we can produce a detailed energy cost analysis and identify the available savings potential for different parts of the household, though this would require taking into account technical characteristics as well. For practical application, further research is needed.

Acknowledgement

References

[1]   National Renewable Energy Laboratory: National Residential Efficiency Measures Database, http://www.nrel.gov/ap/retrofits/about.cfm, download time: 2013.11.14.

[2]   Lukas G. Swan, V. Ismet Ugursal, Ian Beausoleil-Morrison: Occupant-related Household Energy Consumption in Canada: Estimation Using a Bottom-Up Neural-Network Technique, Energy and Buildings, Vol. 43, pp. 326-337, 2011

[3]   Federal Ministry for the Environment, Nature Conservation and Nuclear Safety, Fraunhofer Institute for Systems and Innovation Research ISI: Policy Report, Contribution of Energy Efficiency Measures to Climate Protection within the European Union until 2050, 2012

[4]   Hai-xiang Zhao, Frédéric Magoulès: A Review on the Prediction of Building Energy Consumption, Renewable and Sustainable Energy Reviews, Vol. 16, pp. 3586-3592, 2012

[5]   Merih Aydinalp-Koksal, V. Ismet Ugursal: Comparison of Neural Network, Conditional Demand Analysis, and Engineering Approaches for Modeling End-Use Energy Consumption in the Residential Sector, Applied Energy, Vol. 85, pp. 271-296, 2008

[6]   Merih Aydinalp, V. Ismet Ugursal, Alan S. Fung: Modeling of the Space and Domestic Hot-Water Heating Energy-Consumption in the Residential Sector Using Neural Networks, Applied Energy, Vol. 79, pp. 159-178, 2004

[7]   Arash Kialashaki, John R. Reisel: Modeling of the Energy Demand of the Residential Sectorin the United States Using Regression Models and Artificial Neural Networks, Applied Energy, Vol. 108, pp. 271-280, 2013

[8]   D. Kolokotsa, C. Diakaki, E. Grigoroudis, G. Stavrakakis and K. Kalaitzakis: Decision Support Methodologies on the Energy Efficiency and Energy Management in Buildings, Advances in Building Energy Research, Vol. 3, pp. 121-146, 2009

[9] Kangji Li, Hongye Su, Jian Chu: Forecasting Building Energy Consumption Using Neural Networks and Hybrid Neuro-Fuzzy System: A Comparative Study, Energy and Buildings, Vol. 43, pp. 2893-2899, 2011

[10] Alberto Hernandez Neto, Flávio Augusto Sanzovo Fiorelli: Comparison between Detailed Model Simulation and Artificial Neural Network for Forecasting Building Energy Consumption, Energy and Buildings, Vol. 40, pp. 2169-2176, 2008

[11] Lukas G. Swan, V. Ismet Ugursala: Modeling of End-Use Energy Consumption in the Residential Sector: A Review of Modeling Techniques, Renewable and Sustainable Energy Reviews, Vol. 13, pp. 1819-1835, 2009

[12] Merih Aydinalp, V. Ismet Ugursal, Alan S. Fung: Modeling of the Appliance, Lighting, and Spacecooling Energy Consumptions in the Residential Sector Using Neural Networks, Applied Energy, Vol. 71, pp. 87-110, 2002

[13] Central Statistical Office (CSO) "Census 2011", http://www.ksh.hu/nepszamlalas/?langcode=hu, 2013 (in Hungarian: "Népszámlálás 2011")

[14] Central Statistical Office (CSO) "The Level and Structure of Household Consumption, 2011", Statistical Mirror, Vol. 108, 2012 (in Hungarian: "A háztartások fogyasztásának színvonala és szerkezete, 2011")

[15] Satute 7/2006. TMN (V.24): Regarding the Definition of Building Energy Characteristics, 2006 (in Hungarian: Az épületek energetikai jellemzőinek a meghatározásáról)