ASSESSMENT OF PREDICTIVE MODELS FOR THE ESTIMATION OF HEAT CONSUMPTION IN KINDERGARTENS

by

Nebojša M. JURIŠEVIĆ**, *Dušan R. GORDIĆ, and Arso M. VUKIĆEVIĆ

Faculty of Engineering, University of Kragujevac, Kragujevac, Serbia

Original scientific paper
https://doi.org/10.2298/TSCI201026084J

The service sector remains the only economic sector that has recorded an increase (3.8%) in energy consumption during the last decade, and it is projected to grow more than 50% in the following decades. Among the public buildings, educational are especially important since they have high abundance, great retrofit potential in terms of energy savings and impact in promoting a culture of energy efficiency. Since predictive models have shown high potential in optimizing usage of energy in buildings, this study aimed to assess their application for both finding the most influential factors on heat consumption in public kindergarten and heat consumption prediction. Two linear (simple and multiple linear regression) and two non-linear (decision tree and artificial neural network) predictive models were utilized to estimate monthly heat consumption in 11 public kindergartens in the city of Kragujevac, Serbia. Top-performing and most complex to develop was the artificial neural network predictive model. Contrary to that, simple linear regression was the least precise but the most simple to develop. It was found that multiple linear regression and decision tree were relatively simple to develop and interpret, where in particular the multiple linear regression provided relatively satisfying results with a good balance of precision and usability. It was concluded that the selection of proper predictive methods depends on data availability, and technical abilities of those who utilize and create them, often offering the choice between simplicity and precision.

Key words: predictive models, data-driven approaches, kindergarten buildings, public buildings energy management

Introduction

Final energy consumption decreased by 7.1% in the *EU28* during the period 2005-2016 [1]. This decrease has reflected the reduction in energy consumption in all economic sectors except in the sector of services that have increased by 3.8% [1]. On the other side, worldwide energy demand for the building sector is expected to grow more than 50% from 2010-2050 [2]. Thus, it may be found that future strategies dealing with the reduction of CO_2 emissions should pay particular attention the constantly growing sector of services, including educational buildings due to the following three reasons.

First, educational buildings have relatively high abundance and great energy use intensity. Accordingly, their energy consumption presents a significant share of total energy consumption in public buildings. For an example, educational buildings are responsible for 61% of energy and water costs created in the public sector in the city of Kragujevac, Serbia [3], and 60% in the city of Ingolstadt (Germany) [4]. Final energy consumption in educational and cul-

* Corresponding author, e-mail: jurisevic@kg.ac.rs

tural institutions in the city of Banjaluka (Bosnia and Herzegovina) represent 72.5% of the total share of final energy consumption in all public buildings in the city [5], while that share in the city of Ohrid (Northern Macedonia) has the value of 53.17% [6]. Although many municipal energy reports do not include official figures dealing with energy consumption per building type, all of them emphasize that educational buildings have a great influence on municipal public energy demand [7].

Second, educational buildings have relatively high retrofit potential in terms of electricity [8] and overall energy savings [9]. As a result of the renovation of 16 educational buildings in central Serbia, their energy intensity has been reduced by approximately 40% [10]. According to the World Bank, renovated public buildings in Serbia decreased energy consumption by approximately 60% [11]. Similarly, demonstrated potential for energy savings in Bulgaria has been 40% [12], in Russia more than 30% [12], while potential savings in terms of thermal energy in educational buildings in central Italy are estimated to 38% [13]. Hu [14], according to the Severn, Md., USA, primary school case study, concluded that only exterior walls thermal improvement could contribute to 22% of the school's energy savings. A combination of different energy-saving measures in the C climatic zone in Greece results in energy savings of about 28.78% [15]. As an addition, there are many building renovation projects or pilot projects that are examples of good practice for local governments and communities. Renovation of Music Academy in Dresden, Germany, resulted in a 67% reduction of energy intensity [16]. Thermo-modernisation of two kindergartens and seven primary school buildings located in south-eastern Poland contributed to 46.8% of energy savings [17]. Caste study of two educational buildings in the city of Lecce, Puglia Region, Southern Italy, indicates that cost-optimal transition from traditional to nearly zero energy buildings contributes to 85% of energy savings [18]. Studies that deal with nearly zero energy buildings consider implementation of RES as essential, and often share collected data for evaluation purposes [19].

Third, improving the energy performance of public buildings is important for the promotion of a culture of energy efficiency [13]. In particular, educational institutions play an important role in creating awareness of sustainable development among the youth. Teaching staff and parents are indirectly included in the process. Thus, the approach of early education could result in multiple long-term benefits [20].

Kindergarten buildings have the highest values of energy intensity among educational buildings [3, 4, 21, 22]. This is the consequence of special building design that in particular refers to large windows and high ceilings. Besides, ambient conditions in these buildings have to be maintained with particular attention. Contrary to other educational buildings, kindergartens are open during the whole year. Usually, they most often present the first choice for energy retrofit and energy-saving programs among educational buildings.

Numerous researches utilized data-driven approaches to predict energy consumption in different types of buildings. Capozzolia *et al*. [23] utilized multiple linear regression (MLR) method to develop a model for the estimation of heating energy consumption in school buildings. For that purpose, the authors selected nine inputs (describing building physics, heating system, and climatic conditions) to be used on the sample of 80 school buildings in northern Italy. The model resulted in coefficient of determination, R^2, value of 0.86, meaning that was able to describe 86% of analyzed data. In order to introduce the heating energy demand estimation model for the practice of real estate management on the example of educational buildings in the city of Stuttgart, Germany, Elisabeth Beusker, Christian Stoy *et al*. [24] applied MLR model with five inputs. The relatively low R^2 value of their model (0.6) can be explained with the fact that different types of buildings were included in the model (*e.g.* elementary school with gymnasium, a secondary school with swimming pool*, etc.*). Similar results can be obtained for other sectors where building characteristics and operational activities significantly vary. To find a compromise between the simplicity of the evaluation method and the accuracy of energy consumption in the banking sector, Aranda *et al.* [25] developed three linear regression models with seven inputs on the sample of 55 bank buildings in Spain. The model resulted in R^2 values of 0.57, 0.65, and 0.69. Higher R^2 values can be expected when models are dealing with monthly electricity consumption in buildings because there are fewer uncertainties influencing electricity consumption than the uncertainties influencing heat consumption. Amber *et al.* [26] applied regression method with five inputs to create a reliable forecast of electricity consumption is the university campus in London South Bank, University of London, UK. The model resulted in a relatively satisfying R^2 value of 0.89. Similar to that, MLR analysis of energy (gas and electricity) consumption in a supermarket in northern England [27] resulted in R^2 of 0.95 for electricity and 0.86 for gas.

Five input decision tree (DT) prediction models dealing with heating energy consumption in northern Italy resulted in the same R^2 values as the MLR model (0.86) [23]. The same method was applied to predict the energy consumption of 67 residential buildings in Japan, with an R^2 value of 0.92. The ANN methods proved as the most precise. The principal component analysis – adaptive neuro-fuzzy inference system method used for the prediction of the energy performance of residential buildings resulted in mean absolute percentage error (MAPE) of just 1.56% [28]. Besides the ANN, support vector machines (SVM) are exquisite at determining the non-linear obstacles, executing them very appropriate to forecast energy in the building environment [29]. Moreover, the SVM proved to be more reliable than the three different ANN methods in the particular study predicting heating energy consumption in 35 campus buildings in Gloshaugen, Norway [30]. Some studies concluded that a data-driven approach could provide reliable predictions as the physical-based ones [31, 32]. Without any doubt, physical-based software, EnergyPlus, TRNSYS, DesignBuilder, SEFARIA software (Revit plug-in)*, etc.* are a powerful tool to rely on, especially *in the hands* of experts. However, public buildings energy management often relies on the experience of non-experts in the engineering field, or at least on the professionals who do not have enough experience to operate complex simulation software. On that behalf, data-driven approaches can be useful for two reasons: first, once developed they are relatively simple to use by non-experts as they are based on just 5-10 data inputs and second, they provide satisficing predicting results. Besides that, the approach provides universal patterns suitable for the application on a large number of buildings, contrary to other approaches relying on *one building – one calculus* approach.

To summarize, data-driven approaches became a popular management tool while energy consumption forecasting becomes a critical and necessary input for planning and controlling energy usage in the building sector [33]. They can help the city and regional planners predict the energy burdens that could result from alternative urban growth patterns and global warming scenarios [34]. However, the literature review indicated that there is a lack of research focused on the prediction of heat energy consumption in kindergartens. Accordingly, this study aimed to assess and compare the performances of various predictive models for the estimation of heat consumption in kindergarten buildings. As an addition, parameters that influence heat consumption were accessed for all considered models. Finally, all the data and predictive models developed in this study are available on the public repository*.

---

* The repository (DOI: 10.5281/zenodo.3889752) includes: 1) Excel files with data, SLR and MLR, 2) Developed Weka models and input *.arff files for DT and 3) Matlab script and *.mat files for running the developed EAANN model.

Materials and methods

Materials

Data used in this study were collected from 11 (out of 15) public kindergartens in the city of Kragujevac, Serbia. The 440 staff members provide services for around 3800 children annually. Four of the buildings were not included in the analysis because they do not have individual heat consumption metering, two buildings connected to district heating (DH) network and two buildings were under renovation process. The basic details and geographical locations of the analysed buildings are presented in fig. 1.

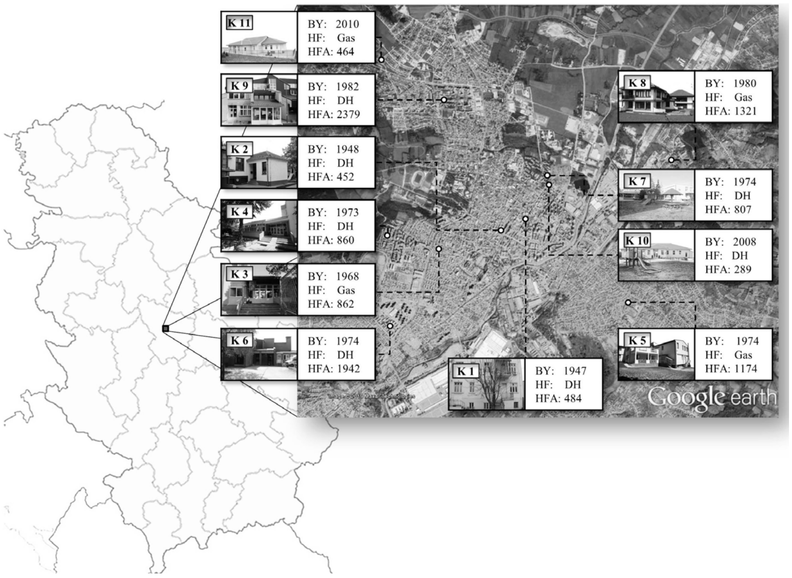

Figure 1. Location and basic details of analysed buildings: *K – Kindergarten, BY – Built year, HF – heating fuel/heating source, and HFA – Heated floor area*

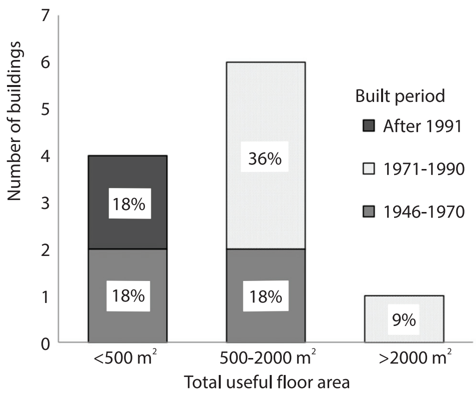

Figure 2. Characteristics of analysed public kindergartens in the city of Kragujevac

The classification of the buildings according to the total useful floor area as well as different built periods is presented in fig. 2.

Although there are some operational differences, a heating day with guaranteed room temperatures for all the buildings connected to the DH system begins at 6:30 a. m. and lasts until 9:30 p. m. during working days, and from 7:00 a. m. until 10:00 p. m. during the weekends. Moreover, there is no automatic control in substations of kindergarten buildings connected to the DH system and the buildings are heated during non-occupancy periods in the same manner asduring the occupancy periods. On the other side, buildings that use natural gas for heating have two heating regimes: a daily heating regime that starts at 5:00 a. m. and lasts until 9:00 p. m. during every day of the heating season providing room temperatures 22-24 °C, and night regime that lasts from 9:00 p. m. until 5:00 a. m. keeping temperatures on 17 °C.

Energy use intensity of kindergartens in some European cities (including Kragujevac) is present in fig. 3. Presented cities are listed according to their geographic latitudes (from north to south) so energy intensity comparison is easier to interpret.

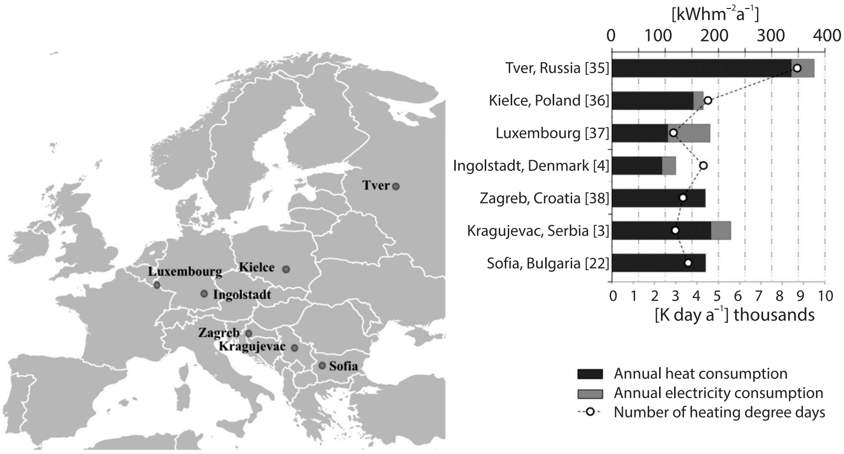

Figure 3. Comparison of energy consumption indicators in kindergartens in Kragujevac, Serbia with neighbouring cities and cities from other climatic zones [3, 4, 22, 35-38]

Public kindergartens in the city of Kragujevac have the highest energy use intensity in the public buildings sector. Moreover, analyzed kindergartens consume more heat than expected compared to the same type buildings in other countries. Because of that, these buildings are particularly interesting for municipal energy management monitoring.

Predictive models should be able to facilitate energy managers to calculate budget spending. Moreover, models should be able to contribute to the timely detection of system malfunctions, detection of incorrect metering, and, if that is possible, predict changes in energy consumption in the case of building thermal envelope renovation.

Energy audit and data collection

Data related to heat consumption and climate were collected for five years, starting from 2015 and ending in 2019. All the data referring to the heat consumption of the analyzed buildings were provided by the city utility companies for gas and heat distribution. Additional information related to buildings' characteristics and organizational structure were collected in co-ordination with representatives of the preschool buildings and according to building technical documentation, while meteorological data were provided by the Republic Hydrometeorological Service of Serbia. The pattern of natural ventilation was established through the process of surveying employees. Unfortunately, class registers were only partially available for the period of the last 2-3 years, therefore, it was not possible to include building occupation *i.e.* occupants behavior as the influential factor in the study.

Data preprocessing

Output data in this study is heating energy consumption in buildings, expressed in [kWh]. To remove poor-quality monthly outputs caused by missing data or monitoring issues, the data cleaning process was conducted manually. The practice of local DH utility company is to add energy consumed in the last half month of the ending heating season (April) to the consumption of the first half month of the following heating season (October). Thus, their combined values are measured in October reading. Because of that, measured April and October heat consumption cannot be considered as proper monthly data. Consequently, these values have been removed from the input/output list for all the analyzed buildings connected to the DH system for all heating seasons. The rest of the data included in the study analysis present input parameters, which can be classified into three categories.

The first category of data represents variable parameter (monthly weather data) delivered by the Republic Hydrometeorological Institute of Serbia. The average daily temperature values are processed to present the number of monthly heating degree days (HDD).

The second category of data consists of constant values that represent building physical details collected from building technical documentation and building energy audits. These parameters can be additionally classified into two subcategories:

– original, the category of *raw* data where one input represents one value, and
– derived, where the input values are derived from two original inputs.

Derived values were created to avoid multicollinearity among two important independent inputs by combining them into the one expression. The value derived in that way represents the combined influence of two original factors on the output value. This way the MLR model will be able to predict possible changes in energy consumption caused by building renovation *i.e.* changes of building envelope and fenestration details.

Original constant inputs are: building built year, type of built (traditional masonry or modern prefabricated), heating source (DH or natural gas), number of buildings (NB), heated floor area (HFA), heated building volume, external walls gross surface, external walls net surface (EWNS), gross fenestration area (GFA), ceiling surface, external walls average *U*-value (EWA-U), average fenestration *U*-value (AF-U), average ceiling *U*-value, gross building envelope surface, net building envelope surface (excluding fenestration area), roof type (flat or pitched), and number of building floors. Derived constant inputs are: products of external wall net surface and average external walls *U*-values (EWNS-U), the product of gross fenestration areas and average fenestration *U*-values (GFA-U), the relation of fenestration and floor area, and aspect ratio (AR); the ratio of heat transfer surface and gross heated volume, and building size (BS); greater or smaller than 500 m^2.

The third category of data refers to the employees' survey results. The aim of the survey was to define the pattern of natural classroom ventilation. It was concluded that all the analyzed buildings have more or less the same time pattern of single-sided natural ventilation, therefore, just the area of windows used for natural classroom ventilations (AWNCV) was considered as a parameter of influence.

Data stratification on train and test sets

To develop and objectively assessed predictive models, overall data acquired in this study were split into training and independent test set. Data stratification was performed by the random splitting of the overall data set ($N = 287$) into the training (~70% samples) and test set (~30% samples), ensuring that both sets kept characteristics of the original set. Statistical characteristics of data sets used in this study are given in tab. 1.

Methods

Four different methodologies were considered in this study: simple linear regression (SLR), MLR, DT, and ANN.

Table 1. Characteristics of training and test set

Characteristics	Unit	Overall (N = 287)	Training (N = 201)	Test (N = 86)	p-value
NB	[–]	1.08 ±0.28	1.09 ±0.29	1.07 ±0.26	0.79
HFA	[m^2]	1039 ±638	1081 ±666	940 ±559	0.21
HDD	[K·day·month^{-1}]	413 ±132	413 ±134	413 ±126	0.98
EWNS·U	[WK^{-1}]	1099 ±1095	1077 ±1025	1148 ±1247	0.29
GFA-U	[WK^{-1}]	958 ±662	1001 ±684	860 ±600	0.16
AWNCV	[m^2]	41 ±24	41 ±24	41 ±24	0.62
RT	[–]	233 (81%)	164 (82%)	69 (80%)	0.86
EWNS	[m^2]	755 ±617	772 ±606	716 ±644	0.13
EWA-U	[Wm^{-2}K^{-1}]	0.79 ±0.49	0.78 ±0.48	0.83 ±0.52	0.44
GFA	[m^2]	315 ±220	332 ±227	277 ±197	0.07
AF-U	[Wm^{-2}K^{-1}]	3.15 ±0.65	3.12 ±0.66	3.2 ±0.62	0.39
AR	[m^{-1}]	0.55 ±0.16	0.55 ±0.16	0.57 ±0.16	0.65

Particularly, they were selected as representative since include algorithms that are: linear (SLR, MLR) and non-linear (DT, ANN) and require manual statistical analysis and selection of features, to those that are self-configuring.

These variabilities among the considered algorithms enable investigation of various aspects and effects of using predictive modelling in building energy management (*i.e.* usability, complexity), besides traditional accuracy metrics.

Linear regressions were used in many studies dealing with the prediction of energy consumption in public, commercial, and residential buildings. The method is proven as relatively efficient providing a good balance of precision and simplicity (ease of use). To compare the results of two linear methods, one relatively advanced and other relatively simple, this study utilizes MLR and SLR model. The SLR model was developed using only one explanatory variable, the number of HDD. In addition this, to enable the algorithm to be applicable to the buildings of all sizes, output values were expressed as general energy consumption per floor area [kWhm^{-2}month^{-1}]. The SLR prediction formula can be expressed:

$$\frac{E}{HFA} = kHDD + m \tag{1}$$

where E stands for the predicted value of heating energy, k [–] – the coefficient determined by the regression of training data set, HDD – the number of monthly heating degree days (input variable), and m [–] the intercept determined by the regression of training data set. In order to express the monthly energy consumption of a building, one should just multiply the estimated general specific heat consumption with the value of heated floor area of a building. Developed SLR model has the eq. (2):

$$\frac{E}{HFA} = 0.0527HDD + 7.5498 \tag{2}$$

On the other side, the MLR model utilizes more input variables, tab. 2, selected from the Spearman correlation matrix, forming the model:

$$E = m + k_1x_1 + k_2x_2 + \ldots + k_nx_n \tag{3}$$

The Spearman correlation [39] is conducted to determine the linear correlation among potential influential factors and output values. The x_6, the most influential and least mutually correlated factors for MLR were selected according to input-output correlation. Additionally, selected parameters were tested on multicollinearity by variance inflation factors (VIF), tab. 2. Depending on literature, input parameters can be considered as collinear for VIF values greater than 4 [25] or 10 [23]. As all the analyzed inputs have the VIF values lower than the lowest recommended threshold proscribed in literature, it can be concluded that there is no multicollinearity and that selected inputs can be utilized in the MLR model.

Table 2. The MLR descriptive statistics

Parameter	Input parameter	Unit	Coefficients		Std. error	*t* stat	*p*-value	VIF
m	Intercept	[–]	*m*	–29186.00	3001.04	–9.73	0.00	–
x_1	*HDD*	[K day month^{-1}]	k_1	6386.92	1640.03	3.89	0.00	1.03
x_2	*NB*	[–]	k_2	6506.47	2028.73	3.21	0.00	1.26
x_3	*BS*	[–]	k_3	57.70	3.64	15.84	0.00	2.44
x_4	*EWNS-U*	[WK^{-1}]	k_4	1.54	0.55	2.79	0.01	1.37
x_5	*GFA-U*	[m^2]	k_5	21.59	1.04	20.80	0.00	3.31
x_6	*AWNCV*	[m^2]	k_6	46.88	23.16	2.02	0.04	2.79

A relatively intuitive prediction formula that is not difficult to interpret and utilize could be considered as the method strength. Shortcomings of the method are related to the fact that MLR is a liner approach, and that multicollinearity among independent variables restricts the utilization of some of the variables that could influence the output. Also, variables whose values do not have a linear correlation with the output values are not considered as significant, although they could be influential non-linear factors.

The third considered algorithm was DT. In contrast to SLR and MLR, it is non-linear and based on a flowchart-like tree structure to segregate a set of data into various predefined classes, thereby providing the description, categorization, and generalization of given datasets [40]. The strength of the method is the fact that the presence of multicollinearity is not of concern for the DT [41], therefore, mutually correlated variables can be included in the model regardless of their similarities. The selection of proper influential parameters in the present study was done by using the wrapper attribute evaluator algorithm [42] applied to the *j*48 classifier in the WEKA software. The wrapper works by excluding one by one input to examine input influence on the prediction result. Finally, the four most influential inputs were selected, HDD, HFA, and type of built (TB), *GFA*, tab. 3.

Since all the inputs are ordinal values, the predicted outputs have ranked value as well. The range of adopted numeric prediction values, shown in column 6, presents the average value of the class span from column 1.

The EAANN was proposed for the estimation of heat consumption in kindergartens as a solution that can fully automatically: configure hyperparameters of the ANN [43] and perform feature selection for the developed ANN [44].

This was achieved by setting the objective function of genetic algorithm (GA) to maximize the correlation coefficient, R^2, of ANN, while simultaneously selecting optimal features subset and the following list of the ANN parameters: number of neurons in a hidden layer, type of activation functions in the layers, learning algorithm, number of learning epochs, learning rate, and momentum. By using these parameters, the GA objective function iteratively config-

Table 3. Most significant DT inputs by class

Class/category [–]	Input/output parameters					
	Heat consumption [MWh month^{-1}]	*HDD* [K day month^{-1}]	*HFA* [m^2]	*TB* [–]	*GFA* [m^2]	Output average by class [MWh month^{-1}]
1	<10	<200	<500	<1947	<100	5
2	10-20	200-300	500-1000	1947-1970	100-200	15
3	20-30	300-400	1000-1500	1970-1980	200-300	25
4	30-40	400-500	1500-2000	1980-1991	300-400	35
5	40-50	500-600	2000>	1991>	400-500	45
6	50-60	600-700			500-600	55
7	60-70	700>			600-700	65
8	70-80					75
9	80>					85

ures and evaluates the ANN performances, which are improving with every GA generation until it converged or reached a maximum number of generations, fig. 4.

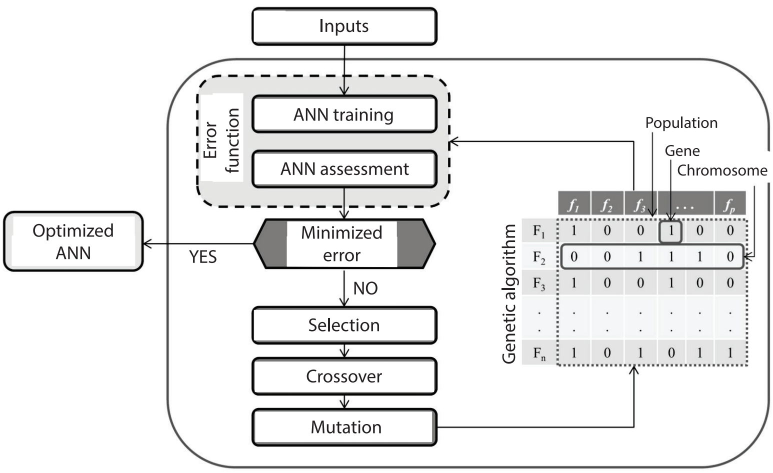

Figure 4. Procedure for the optimization of ANN using GA

Feature selection and configuration of the ANN hyperparameters were done simultaneously using the GA. The total number of *n* parameters was subjected for the GA optimization. Particularly, in each evaluation of the GA objective function, the ANN was developed using F_1 to F_7 chromosome, while F_8 to F_n were used as features selectors, so that the parameters F_8 to F_n with value 1 indicated features that should be selected while parameters with value 0 indicated features to be neglected during the training.

The major advantage of EAANN is the fact that it reduces needs for users' expertise in machine learning and statistics needed for the feature selection, while it ensures reaching high predictive performances. Regarding the GA settings, the number of generations was set to 50 and the size of the population was set to 200. The EAANN model was developed by using the MATLAB 2017b (MathWorks, Natick, Mass, USA). In the present study, automatically selected input parameters for the EAANN model were: HDD, EWA-U, GFA, AF-U, and roof type (flat/pitched).

Results and discussion

Performances of predictive models are expressed by measuring their R^2, which assesses the strength of the linear relationship between the real and predicted values, and MAPE, as an additional measure of prediction accuracy. The R^2 values that are close to 1, and MAPE values that are close to 0 indicating better forecasting results. A comparison of the SLR and MLR model predictions and the corresponding ground truth values are shown in fig. 5. In terms of R^2 values achieved on the training data set, the SLR model, as expected, showed smaller precision than the MLR (0.84 *vs*. 0.89, respectively), while the MAPE was the same for both of the models (33%). On the test data set SLR model resulted in a significant drop in R^2 (by 9%) and MAPE (by 6%) compared to the values achieved on the training data set. Contrary to that, the MLR model resulted in just 1% lower R^2 value, while MLR MAPE was 2% better than the value achieved on the training data set.

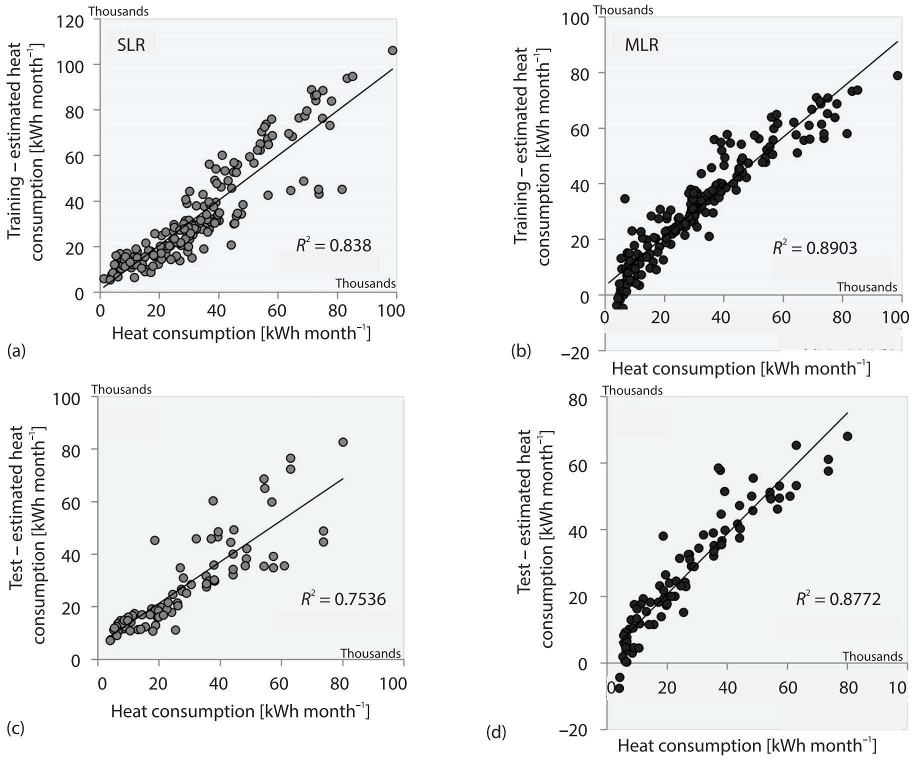

Figure 5. Correlations between ground truth and SLR and MLR predicted heat consumption

A comparison between DT and ANN model predictions and the corresponding ground truth values are shown in fig. 6. The ANN prediction model resulted in R^2 values of 0.96 and 0.92 on the training and data test, respectively, which is significantly better than the result achieved by DT (0.92 and 0.84). Moreover, the ANN model has MAPE of 14% on both data sets, which is more than 10% better than DT. More detailed performances obtained on the training and test data set are shown in tab. 4. To make more objective conclusions, the discussion was based on the results obtained on the independent test set. For analyzing prediction accuracy, the authors assessed predictive models concerning three ranges of monthly heat consumption (low < 10 MWh per month, medium 10-40 MWh per month, high > 40 MWh per month).

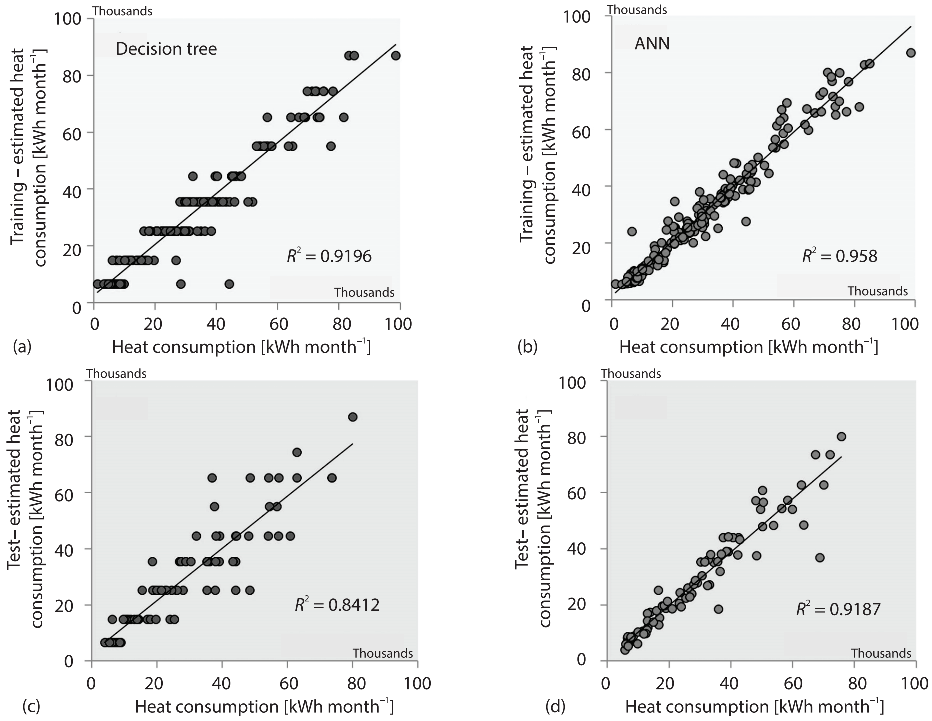

Figure 6. Correlations between ground truth and DT and ANN predicted heat consumption

Table 4. Performances of the developed predictive models on training and test set

Method	Consumption [MWh·month^{-1}]	<10	10-40	40>
	Data set	*MAPE*		
SLR	Training	91%	19%	21%
	Test	81%	22%	22%
MLR	Training	118%	16%	11%
	Test	70%	21%	11%
DT	Training	51%	15%	10%
	Test	28%	24%	16%
ANN	Training	38%	9%	8%
	Test	16%	14%	9%

Among linear methods, MLR provides better overall prediction results than the SLR. It is interesting however that both models show relatively great MAPE values for the consumptions lower than 10 MWh per month. Although in this range MLR has lower MAPE than SLR (70% vs 81%), it can be seen in fig. 5 that around 10% of MLR predictions have negative values of predicted consumption. Those errors are caused by negative intercept value that is determined for the cases where the number of HDD is predominantly higher than

250 K·day per month (90% of analyzed data). Consequently, the model is not suitable to provide satisfying results when the number of HDD is lower than mentioned (at the beginnings and endings of heating periods), and for all the predictions in the range of low heat consumptions. In the range of medium consumption, both models show similar precision. The MLR provides 11% better accuracy than the SLR in the range of great heat consumption. Among non-linear methods, and in general, the *ANN* model provides the most accurate predictions with overall R^2 and MAPE values. Although the DT model resulted in less accurate predictions (R^2 = 0.84 and *MAPE* = 24%), it utilizes only four input parameters comparing to nine utilized by the ANN. The ANN has more than 10% better predictions than the DT in the range of low and medium consumption (28% *vs*. 16%, and 24% *vs*. 12%), while in the range of high consumptions the ANN resulted with the MAPE of just 9%, compared to the DT with 16%.

Conclusion

This study utilizes linear and non-linear approaches with four different methodologies (SLR, MLR, DT, and ANN) to determine the most influential factors on heat consumption in public kindergartens and to develop heat consumption prediction models in these buildings. Prediction models were developed to forecast heat consumption according to selected input parameters and to predict a possible change in heat consumption caused by building renovation (or changes on building thermal envelope). In addition, the paper presents a short overview of heat energy consumption in public kindergartens in different countries with the aim to summarize energy-related data in the field (which could be useful for further studies and analysis of this particular educational sector). Referring to the models, a linear approach was represented with simple and MLR. Strengths of the linear methods were their simplicity, and ease of use, while downsides were restrictions referring to multicollinearity and relatively low prediction accuracy in the low heat consumption range. With the increase of the heat consumption range, model accuracy was increasing. On the other hand, non-linear methods provide better overall accuracy in terms of MAPE, although they do not provide interpretable formula as linear methods do. The DT model, in particular, showed more than 40% better precision in the low range of heat consumptions compared to the linear models. Precision in the medium range of heat consumption was more or less similar to those achieved by the SLR and MLR, with less than 2% difference of MAPE in favor of DT. Contrary to that the MLR model performed 5% better then the DT for the range of consumption that is higher than 40 MWh per month. The best performing model in all consumption ranges was the ANN. The model resulted in less than 16% in MAPE in all consumption ranges and 15% of MAPE in overall prediction. Contrary to its accuracy, the downsides of the ANN model refer to its complexity and the fact that the model is black-box based. Therefore, the model is hard to interpret and develop. As all of the utilized models have a range of strengths and downsides that are rather complementary than opposed, it is up to potential users to choose between better performances (ANN) or usability (MLR, DT).

Acknowledgment

This paper represents the results of research on the project that has been financed by the Ministry of Education, Science, and Technological Development of the Republic of Serbia.

Nomenclature

AF-U – average fenestration *U*-value, [Wm^{-2}K^{-1}]
AR – aspect ratio, [m^{-1}]
AWNCV – area of WINDOWS bused for natural classroom ventilations, [m^2]
BS – building size, [–]
BY – built year, [–]
DH – district heating
DT – decision tree

E – predicted heating energy consumption, [kWh·month^{-1}]
EAANN – evolutionary assembled artificial neural network
EWA-U – external walls average *U*-value, [Wm^{-2}K^{-1}]
EWNS – external walls net surface, [m^2]
EWNS-U – products of external wall net surface and average external walls *U*-values, [WK^{-1}]
GA – genetic algorithm
GFA – gross fenestration area, [m^2]
GFA-U – product of gross fenestration areas and average fenestration *U*-values, [WK^{-1}]
HDD – heating degree day, [K·day·month^{-1}]
HFA – heated floor area, [m^2]
MAPE – mean absolute percentage error
MLR – multiple linear regression
N – number of samples, [–]
NB – number of buildings, [–]
SLR – simple linear regression
SVM – support vector machines
TB – type of built, [–]
VIF – variance inflation factor

References

[1] ***, Final Energy Consumption by Sector and Fuel — European Environment Agency, https://www.eea.europa.eu/data-and-maps/indicators/final-energy-consumption-by-sector-9/assessment-4
[2] Yousefi, Y., *et al*., Energy-Efficiency In Educational Buildings in Iran: Analysis and Measures, *Proceedings*, Building Simulation Conference, Hyderabed, India, 2015, pp. 169-174
[3] Gordić, D., *et al.*, *Energy Efficiency Program for the City of Kragujevac for the Period 2018-2020* (in Serbian), University of Kragujevac, Kragujevac, Serbia, 2018
[4] ***, City of Ingolastadt, Energy Report 2017 (in German), Ingolstadt, Germany, 2017
[5] Tadić, D., *et al*., *Energy Efficiency Action Plan (APE) for the City of Banja Luka for The Period 2016-2019* (in Serbian), Banja Luka, Republic of Srpska, Bosnia and Herzegovina, 2019
[6] ***, Municipality of Ohrid, Energy Efficiency Program for The Municipality of Ohrid 2012-2014 (in Macedonian), Ohrid, Northern Macedonia, 2012
[7] Kipping, A., Trømborg, E., Modelling Hourly Consumption of Electricity and District Heat in Non-Residential Buildings, *Energy*, *123* (2017), C, pp. 473-486
[8] Josijevic, M. M., *et al*., A Method to Estimate Savings of LED Lighting Instalation in Public Buildings the Case Study of Secondary Schools in Serbia, *Thermal Science*, *21* (2017), 6, pp. 2931-2943
[9] Cukovic Ignjatovic, N., *et al*., Improving Energy Efficiency of Kindergartens in Serbia: Challenges and Potentials, *Thermal Science*, *24* (2020), 6A, pp. 3521-3532
[10] Bećirović, S. P., Vasić, M., Methodology and Results of Serbian Energy-Efficiency Refurbishment Project, *Energy Build*., *62* (2013), 2013, pp. 258-267
[11] Aiyer, M. G. S.-R., Report Number: ICRR14381, ICR Review, 2015
[12] ***, Energy Charter Secretariat, Energy Efficiency in the Publc Sector/Policies and Programs in ECT Member Countries, 2010
[13] Dall'O, G., Sarto, L., Potential and Limits to Improve Energy Efficiency in Space Heating in Existing School Buildings in Northern Italy, Energy Build., *67* (2013), 2013, pp. 298-308
[14] Hu, M., Assessment of Effective Energy Retrofit Strategies and Related Impact on Indoor Environmental Quality: A Case Study of an Elementary School in the State of Maryland, *Journal Green Build*., *12* (2017), 2, pp. 38-56
[15] Dimoudi, A., Kostarela, P., Energy Monitoring and Conservation Potential in School Buildings in the C′ Climatic Zone of Greece, *Renew. Energy*, *34* (2009), 1, pp. 289-296
[16] Grossmann, D., *et al*., A Methodology for Estimating Rebound Effects in Non-Residential Public Service Buildings: Case Study of Four Buildings in Germany, *Energy Build*., *111* (2016), Jan., pp. 455-467
[17] Zyczynska, A., *et al*., Energy Effects of Retrofitting the Educational Facilities Located in South-Eastern Poland Anna, *Energies*, *13* (2020), 10, pp. 1-16
[18] Congedo, P. M., *et al*., Efficient Solutions and Cost-Optimal Analysis for Existing School Buildings, *Energies*, *9* (2016), 10, pp. 1-24
[19] Zaca, I., *et al.*, Data of Cost-Optimal Solutions and Retrofit Design Methods for School Renovation in a Warm Climate, *Data Br*., *9* (2016), pp. 846-849
[20] Davis, J., *et al*., The Impact and Potential of Water Education in Early Childhood Care and Education Settings: A Report of the Rous Water Early Childhood Water Aware Centre Program, Brisbane, Australia, 2008
[21] Gordić, D., *et al*., Benchmarking of Heat Energy Consumption in Public Buildings in the City of Kragujevac, *Proceedings*, Energetics2018, Ohrid, North Macedonia, 2018, pp. 139-146

[22] Ivanov, B., Collecting Data from Energy Certification Monitor Performance Indicators for New and Existing Buildings (DATAMINE) – Monitoring of Energy Performance of Municipal Buildings in Bulgaria, Sofia, Bulgaria, 2007
[23] Capozzoli, A., *et al*., Estimation Models Of Heating Energy Consumption In Schools For Local Authorities Planning, *Energy Build.*, *105* (2015), Dec., pp. 302-313
[24] Beusker, E., *et al*., Estimation Model and Benchmarks for Heating Energy Consumption of Schools and Sport Facilities in Germany, *Build. Environ*., *49* (2012), 1, pp. 324-335
[25] Aranda, A., *et al.*, Multiple Regression Models to Predict the Annual Energy Consumption in the Spanish Banking Sector, *Energy Build*., *49* (2012), June, pp. 380-387
[26] Amber, K. P., *et al*., Electricity Consumption Forecasting Models for Administration Buildings of the UK Higher Education Sector, *Energy Build*., *90* (2015), Mar., pp. 127-136
[27] Braun, M. R., *et al*., Using Regression Analysis to Predict the Future Energy Consumption of a Supermarket in the UK, *Appl. Energy*, *130* (2014), Oct., pp. 305-313
[28] Nilashi, M., *et al*., A Soft Computing Method for the Prediction of Energy Performance of Residential Buildings, *Meas. J. Int. Meas. Confed*., *109* (2017), May, pp. 268-280
[29] Ahmad, T., *et al*., A Comprehensive Overview on the Data Driven and Large-Scale Based Approaches for Forecasting of Building Energy Demand: A Review, *Energy Build*., *165* (2018), Apr., pp. 301-320
[30] Aleksandra, S., *et al*., Support Vector Machine for the Prediction of Heating Energy Use, Thermal Science, *22* (2018), 4, pp. 1171-1181
[31] Catalina, T., et al., Development and Validation of Regression Models to Predict Monthly Heating Demand for Residential Buildings, *Energy Build*., *40* (2008), 10, pp. 1825-1832
[32] Amiri, S. S., *et al*., Using Multiple Regression Analysis to Develop Energy Consumption Indicators for Commercial Buildings in the U. S., *Energy Build*., *109* (2015), Dec., pp. 209-216
[33] Lu, X., *et al*., Modelling and Forecasting Energy Consumption for Heterogeneous Buildings using a Physical-Statistical Approach, *Appl. Energy*, *144* (2015), Apr., pp. 261-275
[34] Robinson, C., *et al*., Machine Learning Approaches for Estimating Commercial Building Energy Consumption, *Appl. Energy*, *208* (2017), Sept., pp. 889-904
[35] Vatin, N., *et al*., An Energy Audit of Kindergartens to Improve their Energy Efficiency, Adv. Civ. Eng. Build. Mater. IV – Sel. Peer Rev. Pap. from 2014, *Proceedings*, 4th Int. Conf. Civ. Eng. Build. Mater. CEBM 2014, Hong Kong, China, 2015, April 2016
[36] Pawelec, P., *et al*., Updating Assumptions for the Supply Plan Bydgoszcz into Heat and Electricity and Gaseosus Fuels by 2025, (in Polish), 2017, 1
[37] Thewes, A., *et al*., Field Study on the Energy Consumption of School Buildings in Luxembourg, *Energy Build*., *68* (2014), PARTA, pp. 460-470
[38] Čulig-Tokić, D., *et al*., Comparative Analysis of the District Heating Systems of Two Towns in Croatia And Denmark, *Energy*, *92* (2015), July, pp. 435-443
[39] Spearman, Spearman, C., The Proof and Measurement of Association between Two Things Author (S): C. Spearman Source: The American Journal of Psychology, Vol. 15, No . 1 (Jan ., 1904), pp . 72-101 Published By : University of Illinois Press Stable URL : Http://Www.Jstor.O, Am. J. Psychol., 15 (1904),1, pp. 72-101
[40] Yu, Z., *et al*., A Decision Tree Method for Building Energy Demand Modelling, *Energy Build*., *42* (2010), 10, pp. 1637-1646
[41] Piramuthu, S., Input Data for Decision Trees, *Expert Syst. Appl*., *34* (2008), 2, pp. 1220-1226
[42] Wu, B., *et al*., Simple Profile Rectifications Go a Long Way Statistically Exploring and Alleviating the Effects of Sampling Errors for Program Optimizations, Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics), 7920 LNCS (2013), 97, pp. 654-678
[43] Vukicevic, A. M., *et al*., Evolutionary Assembled Neural Networks for Making Medical Decisions with Minimal Regret: Application for Predicting Advanced Bladder Cancer Outcome, *Expert Syst. Appl*., *41* (2014), 18, pp. 8092-8100
[44] Vukicevic, A. M., *et al.*, Automated Development of Artificial Neural Networks for Clinical Purposes: Application for Predicting the Outcome of Choledocholithiasis Surgery, *Comput. Biol. Med*., *75* (2016), Aug., pp. 80-89

Paper submitted: October 26, 2020
Paper revised: December 12, 2020
Paper accepted: February 17, 2021