ANALIZA I PREDVIĐANJE POTROŠNJE ENERGIJE POSLOVNE ZGRADE KORIŠĆENJEM VIŠESTRUKO LINEARNO REGRESIONOG MODELA, METODE POTPORNIH VEKTORA I NEURONSKE MREŽE

MULTIPLE LINEAR REGRESSION, SUPPORT VECTOR MACHINES AND NEURAL NETWORKS FOR PREDICTION OF COMMERCIAL BUILDING ENERGY CONSUMPTION

Aleksandra SRETENOVIĆ,
Mašinski fakultet, Univerzitet u Beogradu, asretenovic@mas.bg.ac.rs,
Branislav ŽIVKOVIĆ,
Mašinski fakultet, Univerzitet u Beogradu, bzivkovic@mas.bg.ac.rs,
Radiša JOVANOVIĆ
Mašinski fakultet, Univerzitet u Beogradu, rjovanovic@mas.bg.ac.rs,

S obzirom na stalni porast interesovanja za povećanje energetske efikasnosti u zgradarstvu, neophodno je primenjivati i unapređivati postojeće i razvijati nove metode za predviđanje i analizu potrošnje zgrada. Na modelu tipične poslovne zgrade u Beogradu ispitivan je uticaj različitih karakteristika zgrade. Simulacija potrošnje energije na časovnoj bazi urađena je korišćenjem programa HAP (Hourly Analysis Program). Za dalju analizu izabrana su tri faktora koja u najvećoj meri utiču na potrošnju energije za hlađenje: specifična instalisana snaga osvetljenja, udeo prozora u spoljašnjem zidu i koeficijent propustljivosti Sunčevog zračenja kroz prozore. Analiza je vršena za različite vrednosti ova tri parametra. Za kreiranje i testiranje višeparametarskog modela korišćeno je 245 simulacija. Predložen je višestruko linearni model koji može da se koristi za određivanje potrošnje energije za hlađenje, a koji kao ulazne veličine koristi pomenuta tri parametra. U cilju predviđanja potrošnje, razvijeni su modeli primenom metode potpornih vektora (support vector machine) i veštačkih neuronskih mreža i izvršeno je poređenje rezultata sa višestruko linearnim modelom. Pokazano je da modeli zasnovani na metodi potpornih vektora i neuronskim mrežama postižu veću tačnost predvidjanja u odnosu na linearni višeparametarski model.

Ključne reči: predviđanje potrošnje energije; linearna regresija; metoda potpornih vektora; neuronska mreža;

Considering the constant growth of interest in energy efficiency in the building sector, it is necessary to apply and improve existing and also to develop new methods for prediction and analysis of building energy consumption. In this paper cooling consumption of the model of a typical commercial building in Belgrade is analyzed. Detailed energy simulation is done using software HAP (Hourly Analysis Program). The influence of various building characteristics is investigated, and for creating building consumption database, three variables that most largely affect the cooling consumption are chosen: specific lighting power, window area and window shade coefficient. Those three parameters are varied and 245 simulations in total are used for creating and testing the prediction models. The multiple linear model is created and the obtained equation is used for cooling consumption evaluation taking these three building parameters as input. The artificial neural network and support vector machine (SVM) models are also developed for prediction and their results are compared with linear regression model. It has been shown that the statistical methods, such are neural networks and support vector machines can achieve much higher accuracy in prediction than the linear regression model, gaining almost perfect match with simulated values (mean absolute percentage error for testing the SVM model 0,26%)

Key words: cooling consumption prediction; multiple linear regression; support vector machines, neural networks

I. Introduction

Due to the fact that in Europe buildings account for 40% of total energy use and 36% of total CO_2 emission [1] estimation or prediction of building energy consumption is lately topic of greatest interest. This research field involves various scientific domains. Excellent review of these methods with different approaches to their categorization can be found in recent literature [2]. Existing methods for building energy use calculation can be divided into three categories: "white box", "gray box" and "black box" methods. The "white box" approach uses physical techniques to model the building behavior. A lot of different software tools have been developed for this purpose, such as DOE-2, EnergyPlus, TRNSYS, BLAST, ESP-r, HAP, IES APACHE. They are based on solving of equations describing heat transfer. Using these tools requires detailed knowledge of the numerous building parameters (constructions, systems) and occupants' behavior. It is also necessary that all physical mechanisms are described with high accuracy, but they are usually based on assumptions and described by empirical equations, which makes models less accurate. Considering that creating physical model of building requires extensive knowledge of various mechanisms occurring inside and outside of the building envelope, in recent years, significant attention has been given to statistical methods that do not require any physical information. Using statistical or data-driven approach [3] building energy behavior is analyzed, while defining relationships with one or more different driving forces or parameters. It is required that the input and output variables are known and measured, and the development of the inverse model consists in determination of a mathematical description of the relationship between the independent variables and the dependent one. One of the main advantages of the so called "black box" models is that there is no need for detailed knowledge about physical characteristics of building, but it requires significant amount of measured data, which is usually difficult to collect. Moreover, it is usually complicated to interpret results obtained by statistical approach in physical term. Creating these models involve various statistical techniques, such as linear multiple regression, neural networks and support vector machines. Statistical regression models simply correlate the energy consumption with the influencing variables. Bauer and Scartezzini [4], Catalina et al. [5] and Katipamula et al. [6] showed that the regression models can be successfully used for estimating building energy consumption. Artificial neural networks (ANN) are the most used artificial intelligence models for the building energy consumption prediction. In [7] detailed model simulation (in EnergyPlus) was compared with a simple model based on ANN. In [8] actual recorded input and output data that influence long-term energy consumption were used in the training, validation and testing process. The produced ANN results were compared with the results obtained by a linear regression method, a support vector machine method and with real energy consumption records showing much better prediction accuracy. In [9] and [10] the authors proposed improving of neural network prediction by creating various ensembles. A review of the different neural network models used for building energy prediction can be found in [11]. Support vector machine is gaining popularity in recent years, since it has been proven in literature that the models can achieve high accuracy in various prediction tasks, such as flood stage forecasting [12], temperature-frequency correlation of a bridge [13], etc. Dong et al. in [14] applied support vector algorithm to predict building energy consumption in tropical region.

The main idea of this paper is to investigate the application of statistical models for energy use prediction, while using building characteristics as input variables. The database for statistical models is created using detailed hourly simulations. Considering that the cooling energy consumption is, due to the transient heat transfer and numerous influencing parameters, complicated to calculate, it is chosen for the analysis.

II. Artificial Neural Networks (ANN)

Artificial neural network (ANN) method is a computational intelligence technique, based on the information processing system of the human brain and which may be used as an alternative method in engineering analysis and prediction. ANN works as a black-box model, thus, it is not necessary to have detailed information about the system. Instead, they learn the relationship between input and output variables by means of historical data, similar to the way a nonlinear regression might perform [15]. The FFNN architecture consists of an input layer, an output layer, and one or more hidden layers built of processors called neurons, which are fully interconnected with neurons in the subsequent layer using adaptable weighted connections. (Figure 1). The nonlinear activation functions in the hidden layer neurons enable the neural network to be a universal approximator. The process of training network is the adjustment of the weights, so that the network can produce the desired response to the given inputs. Different training algorithms could be applied to minimize the error function, but the most widely used are the backpropagation algorithm and the algorithms derived from it. In this study, a multilayer feedforward network with single hidden layer and backpropagation learning algorithm (BPNN) is used. This learning algorithm has two phases. In the first stage

(forward pass), synaptic weights of the network are all fixed and actual response of the network is obtained based on input training dataset presented to the network input layer. The network output is compared with the desired values, the error signal is propagated backwards, and the weights are modified consequently (backward pass) [16].
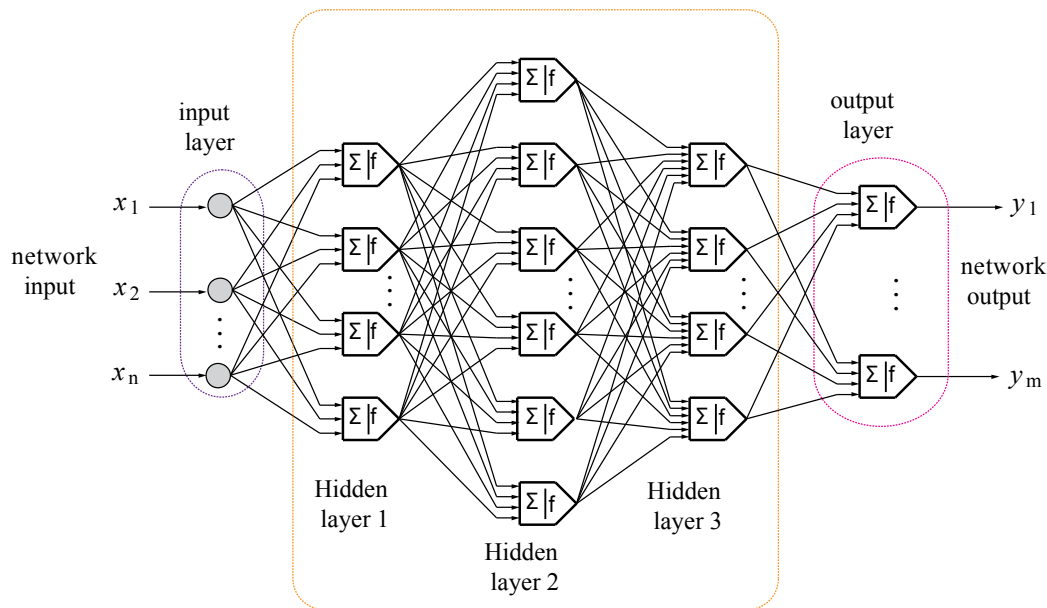

Figure1. Artificial Neural Network

III. Support Vector Machines (SVM)

Support Vector Machine (SVM) is the universal approximator of any multivariate function to any desired degree of accuracy [17], same (or similar) as the Neural Network (NN). There is a difference in the development between these two most common used statistical techniques. SVM was developed from theory, and later it was implemented in practice and experiments, while the NN followed the path in reverse order: from application and extensive experiments to theory. The foundations of SVM have been developed by Vapnik and co-workers in 1964/65 [18], but until recently it was widespread the opinion that it was not suitable for practical application. Today, SVMs show better (or comparable) results than NNs and other statistical models. Traditional NN approaches have suffered difficulties with generalization, producing models that can overfit the data [19]. This is a consequence of the optimization algorithms used for parameter selection and the statistical measures used to select the optimal model. SVMs are the so-called nonparametric model, which means that their "learning" (training) is the crucial issue. The parameters are not predefined and their number depends on the training data used. In other words, parameters that define the capacity of the model are data-driven in such a way as to match the model capacity to data complexity [17]. This is a basic paradigm of the structural risk minimization (SRM) introduced by Vapnik et al. which has been shown to be superior to traditional Empirical Risk Minimization (ERM) principle, employed by conventional neural networks. SRM of the generalization error consisting of the sum of the training error and a confidence level based on Vapnik–Chernoverkis (VC) dimension, as opposed to ERM that minimizes the error only on the training data. It is this difference which equips SVM with a greater ability to generalize, which is the goal in statistical learning. Another key characteristic of SVM is that training of SVM is equivalent to solving a linearly constrained quadratic programming problem so that the solution of SVM is always unique and globally optimal, unlike other network's training which requires non-linear optimization with the danger of getting stuck into local minima. In SVM, the solution to the problem is only dependent on a subset of training data points which are referred to as support vectors. Using only support vectors, the same solution can be obtained as using all the training data points. SVMs were developed to solve the classification problem, but recently they have been extended to the domain of regression problems, which is often referred as Support Vector Regression (SVR) [20].

Consider a set of training points $\{(x_1, y_1), \ldots (x_l, y_l)\}$, where $x_i \in \Re^d$ is a feature vector and $y_i \in \Re$ is the target output. The support vector regression for the linear case finds a linear function that can best approximate the actual output vector y, with a tolerance ε, and is as flat as possible. The regression function can be expressed as:

$$f(w,b) = w \cdot x + b \tag{1}$$

where w and b are the parameter vectors of the function. The regression problem can be expressed as the following convex optimization problem:

$$\min_{w,b,\xi,\xi^*} \quad \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{l}\left(\xi_i + \xi_i^*\right)$$
$$\text{subject to} \quad \begin{cases} y_i - w \cdot x_i - b \le \varepsilon + \xi_i \\ w \cdot x_i + b - y_i \le \varepsilon + \xi_i^* \\ \xi_i, \xi_i^* \ge 0 \end{cases} \tag{2}$$

where ξ_i and ξ_i^* are slack variables that specify the upper and the lower training errors subject to an error tolerance ε, and C is the positive constant, which determines the trade-off between the flatness of function and the empirical error. Minimizing the first term is equivalent to minimizing the confidence interval of the learning machine and minimizing the second term corresponds to minimizing the empirical risk. By introducing a dual set of Lagrange numbers optimization problem can be solved more easily in dual form, which is not detailed herein (more details can be found in [20]). Another advantage of the optimization problem in its dual form is that the input vectors are multiplied as dot product, which enables the use of the „kernel trick" for extending SVM to nonlinear case. The idea is to map the input data in a higher dimensional feature set, and perform linear regression in the feature set.
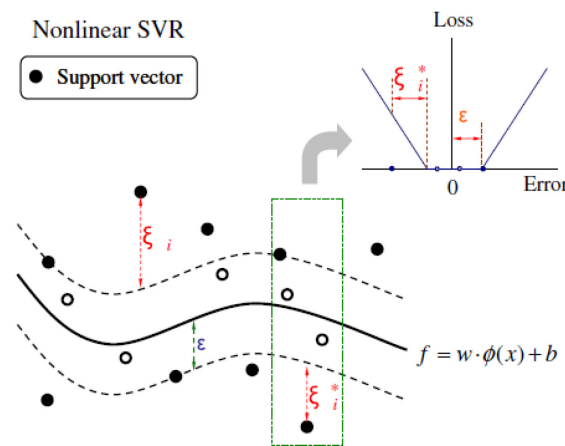

Figure 2. The non-linear Support Vector Regression [12]

The nonlinear Support Vector Regression is illustrated in Figure 2. The region enclosed by the tube is called ε -insensitive zone. The values with excess positive and negative deviations are denoted with ξ_i and, ξ_i^*, respectively and they are called the „slack variables". The optimization criterion penalizes those data points whose values of y lie more than ε distance away from the fitted function $f(x)$. If the predicted value is within the tube, the loss is zero, while if the predicted point is outside of the tube, the loss is magnitude of the difference between the predicted value and the radius ε of the tube.

Let x_i be mapped into a feature space by a non-linear function $\phi(x)$, than the function becomes:

$$f(w,b) = w \cdot \phi(x) + b \tag{3}$$

Similarly the non-linear regression problem can be expressed as the following optimization problem:

$$\min_{w,b,\xi,\xi^*} \quad \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{l}\left(\xi_i + \xi_i^*\right)$$
$$\text{subject to} \quad \begin{cases} y_i - w \cdot \phi(x_i) - b \le \varepsilon + \xi_i \\ w \cdot \phi(x_i) + b - y_i \le \varepsilon + \xi_i^* \\ \xi_i, \xi_i^* \ge 0 \end{cases} \tag{3}$$

By the use of kernels, all necessary computations can be performed directly in input space, without having to compute the mapping. Some commonly used kernels are: linear, polynomial, sigmoid and radial basis func-

tion (RBF) kernel. The RBF kernel nonlinearly maps samples into a higher dimensional space and it can handle the case when the relationship between dependent and independent variables is non-linear. Therefore it is one of the most widely used for various regression problems and it can be written as:

$$K(x_i, x_j) = \exp\left(-\gamma \left\| x_i - x_j \right\|^2\right) \tag{4}$$

where γ is the kernel parameter. Support vector algorithm for nonlinear case can be written in its dual form:

$$\min \quad -\frac{1}{2}\sum_{i,j=1}^{l}\left(\alpha_i - \alpha_i^*\right)\left(\alpha_j - \alpha_j^*\right)K\left(x_i, x_j\right) - \varepsilon \sum_{i=1}^{l}\left(\alpha_i + \alpha_i^*\right) + \sum_{i=1}^{l} y_i\left(\alpha_i - \alpha_i^*\right)$$

$$\text{subject to} \quad \begin{cases} \sum_{i=1}^{l}\left(\alpha_i - \alpha_i^*\right) = 0 \\ 0 \le \alpha_i, \alpha_i^* \le C \end{cases} \tag{5}$$

Only a number of coefficients α_i and α_i^* will be different from zero, and the datapoints associated to them are called the support vectors. Finally the kernel function allows the decision function of nonlinear SVR to be expressed as follows:

$$f(x) = \sum_{i=1}^{l}\left(\alpha_i - \alpha_i^*\right)K\left(x_i, x_j\right) + b \tag{6}$$

The parameters that define the nonlinear SVR are the cost constant C, the radius of the insensitive tube ε, and the kernel parameters (in case of RBF kernel γ). These parameters are mutually dependent, so changing the value of one parameter influences other parameters. The parameter C controls the smoothness or flatness of the approximation function. A greater C corresponds to greater penalty of errors and makes the learning machine more complex [21] . Smaller C may cause the errors to be excessively tolerated, which can lead to learning machine with poor approximation. The parameter ε also affects smoothness and dominates the number of support vectors (smaller ε lead to more support vectors and more complex machine). Determining optimal parameters for nonlinear SVR is often a heuristic trial-and-error process, while there are attempts in literature to define optimal range of values depending of the size and noise of the training data [22].

IV. Building model

Model of a typical office building located in Belgrade, which consists of ground floor and four floors, is selected for the case study. Building dimensions are 50 x 20 m, while the floor height is 2,7 m (Figure 3). Total building floor area is 5.000 m². Building is divided into 9 zones (one in each corner, one on each orientation and a central zone). The simulated HVAC system is 4-pipe fan-coil, with central preparation of fresh air in AHUs. Since the central zones mainly demand for cooling, due to the internal gains and no outside walls or windows, the separate AHU prepares the fresh air for the central zones, while the main AHU covers the ventilation needs of all other zones. For the heat source natural gas boiler is selected, while the cold water is provided from air-cooled chiller.
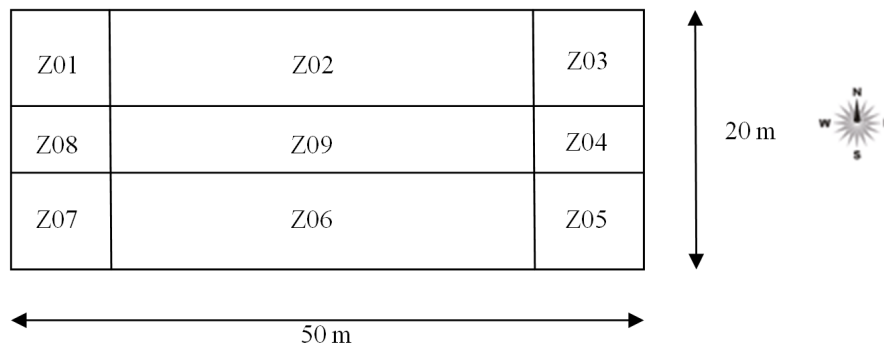
Figure 3. Model building layout

For building energy simulation Hourly Analysis Program (HAP), developed by Carrier, is used. HAP [23] is a software tool which assists engineers in designing HVAC systems for buildings. It estimates design

cooling and heating loads for commercial buildings in order to determine the required size for HVAC system components. Load profiles for cooling are computed for one design day in each month using design weather conditions, design day operating schedules and the ASHRAE-endorsed Transfer Function load calculation method, while for heating loads are computed for a single heating design condition. HAP also estimates annual energy use and energy costs for energy consuming systems in a building by simulating building operation for each of the 8,760 hours in a year. During this process simulation weather data (Typical Meteorological Year – TMY) is used to evaluate how the building's HVAC systems react to real sequences of weather over the course of a year. The Transfer Function load method provides accurate estimates of building loads considering the transient nature of heat transfer processes in the building. Yearly cooling consumption is chosen for further analysis. Energy use for cooling comprises of electricity consumption of: fans in AHUs, fans in FCUs, chiller, primary and secondary water pump.

V. One parameter study

In order to evaluate the influence of the specific input variable on yearly cooling consumption, as the first step, one parameter study is performed. The results are shown in Figures 4 to 9. It can be seen that windows area and window shade coefficient show approximately linear correlation with cooling consumption, while the specific lighting power, ventilation rate, thickness of wall and roof insulation and window U-value have mostly quadratic correlation.
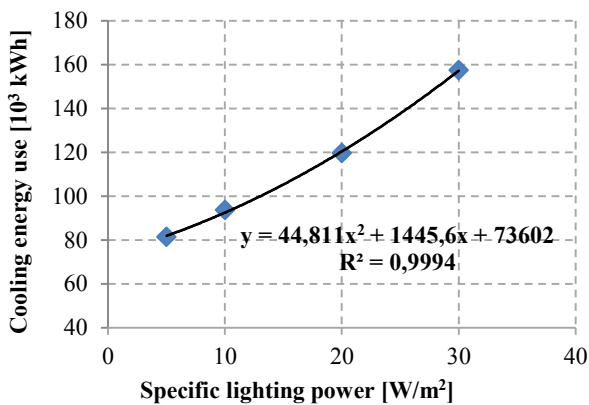

Figure 4. Correlation of cooling consumption and spec. lighting power
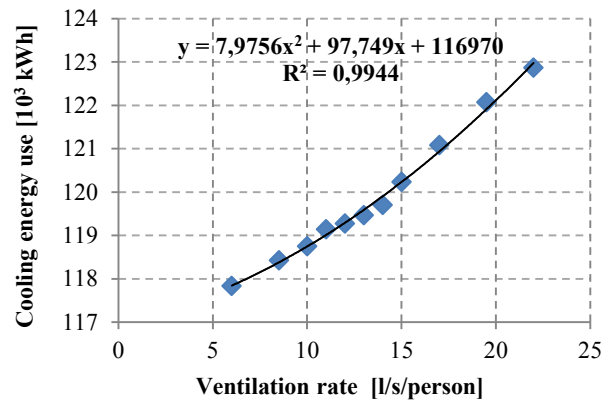
$y = 44{,}811x^2 + 1445{,}6x + 73602$
$R^2 = 0{,}9994$

Figure 5.Correlation of cooling consumption and ventilation rate

$y = 7{,}9756x^2 + 97{,}749x + 116970$
$R^2 = 0{,}9944$
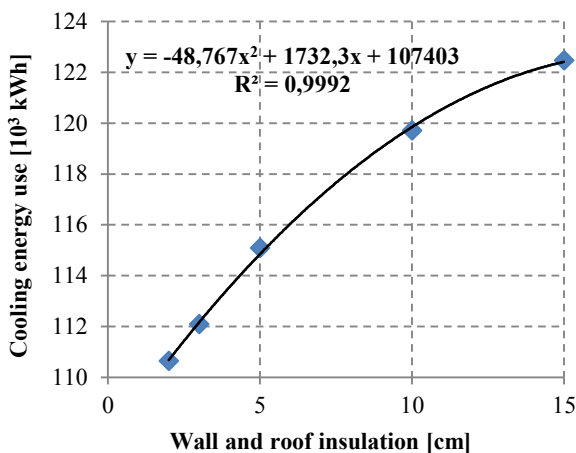
Figure 6. Correlation of cooling consumption and thickness of wall and roof insulation

$y = -48{,}767x^2 + 1732{,}3x + 107403$
$R^2 = 0{,}9992$
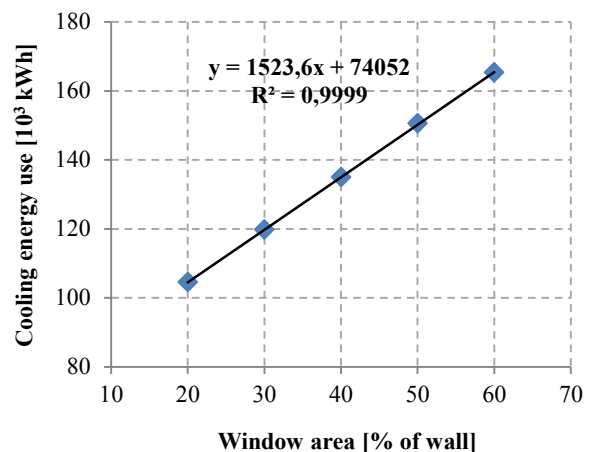
Figure 7.Correlation of cooling consumption and window area

$y = 1523{,}6x + 74052$
$R^2 = 0{,}9999$

After analysis of individual parameters, three variables that most largely affect cooling energy consumption are selected for creating consumption database. They are varied in the following way:

- window area – percent of outside wall (20%, 30%, 40%, 50% and 60%)
- specific installed lighting power (5, 10, 15, 20, 25, 30 and 35 W/m²)
- window shade coefficient (0,15, 0,3, 0,4, 0,5, 0,6, 0,7 and 0,9)

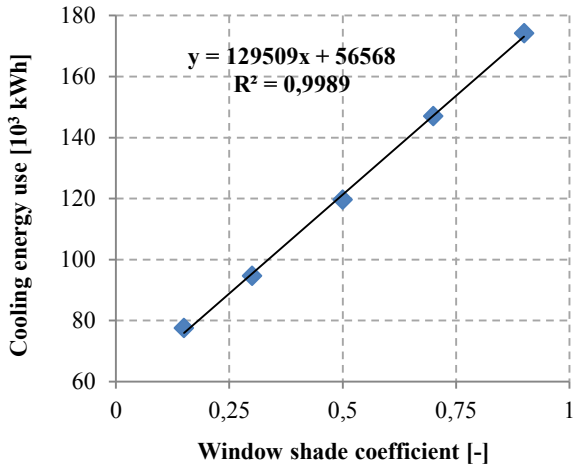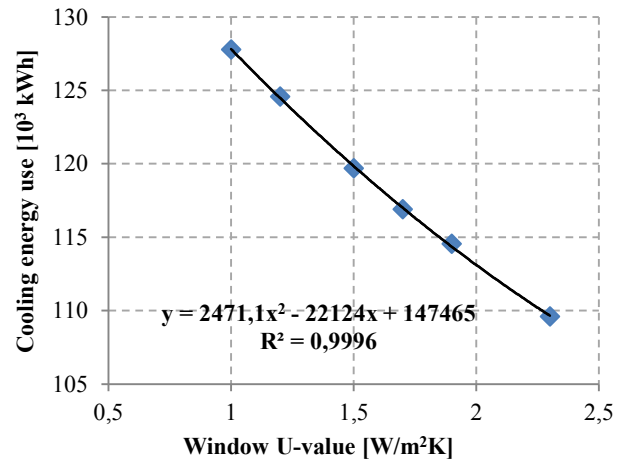*Figure 8.Correlation of cooling consumption and window shade coeff.* *Figure 9.Correlation of cooling consumption and window U-value*

In total 245 simulations are performed, of which 200 simulations are used for training the statistical models, and 45 for testing. The prediction accuracy is measured by the coefficient of determination (R^2), root mean square error (RMSE) and mean absolute percentage error (MAPE), used in the ASHRAE context, defined as follows:

$$\text{RMSE}=\sqrt{\frac{1}{N}\sum_{k=1}^{N}(y_k-\hat{y}_k)^2} \quad , \quad \text{MAPE}=\frac{1}{N}\sqrt{\sum_{k=1}^{N}\frac{|y_k-\hat{y}_k|}{y_k}}\cdot 100 \tag{7}$$

where y_k and $\hat{y}_k$ are the target and output value for the k-th pattern respectively and N is the total number of patterns.

VI. Results and discussion

A. Multiple linear model

Basic multiple linear model (LM) can be described as:

$$Y = a_0 + a_1 X_1 + a_2 X_2 + a_3 X_3 \tag{8}$$

where Y is the yearly cooling consumption, X_1 is window area [% of outside wall], X_2 is the specific installed lighting power [W/m^2], X_3 window shade coefficient [-] and a_0, a_1, a_2 and a_3 are the regression coefficients. Based on 200 simulations used for training the multiple linear regression model, the obtained relation is:

$$Y = -64.722,31 + 1.575,85 \cdot X_1 + 12.684,45 \cdot X_2 + 172,02 \cdot X_3 \tag{9}$$

The same equation is used to evaluate the cooling consumption for the 45 input combinations that are left for testing, and the obtained results are compared with the simulation results. The comparison between the modeled cooling consumption in HAP and predicted values based on the LM is shown in Figure 9. It can be seen that the linear model tends to predict smaller consumption than the software simulation, while achieving MAPE of 9,15% in the test period (Table 1).

B. Neural Network model

Neural network (NN) model is developed using Neural Network Toolbox available in Matlab. The model used in this study is a three-layer feedforward neural network composed of one input layer, one output layer and one hidden layer. The activation functions used for the hidden and output layers were the hyperbolic tangent (tansig) and linear (purelin) functions, respectively. In this study the Levenberg–Marquardt (LM) training algorithm, which is a variant of feedforward back-propagation algorithm is used. For the NN model the number of neurons in the hidden layer was found by a trial and error procedure. The best results are achieved with 12 neurons in hidden layer.

Scaling features before applying NN (and also SVM) is very important. The main reason is to avoid that variables with greater numerical range dominate those with smaller range. Another advantage is to avoid

the numerical difficulties during the calculation. Therefore, all variables are scaled to range [0,1]. It is important that both training and testing data are scaled in the same way.
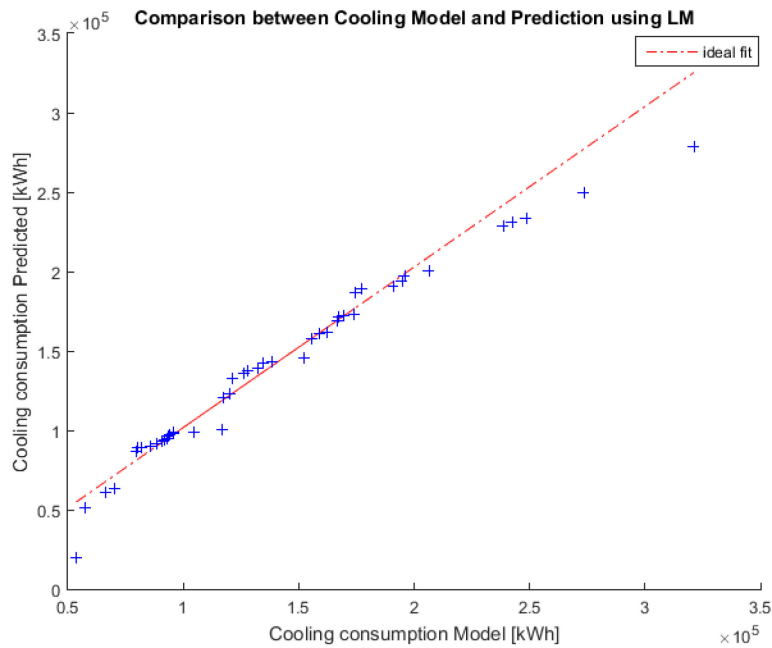

Figure 10. The comparison between modeled cooling consumption and prediction of the multiple linear model
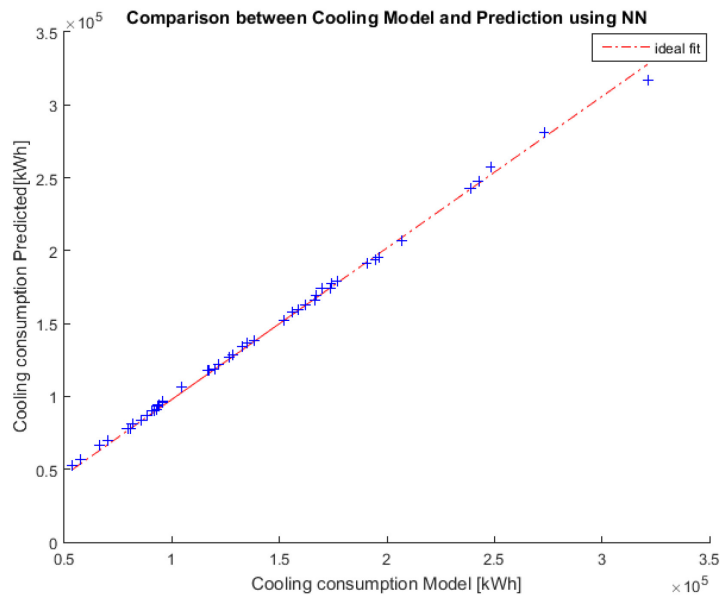

Figure 11. comparison between modeled cooling consumption and prediction of the neural network model

C. Support Vector Machine model

For building Support Vector Machine (SVM) model, library for Support Vector- libsvm for Matlab is used [24]. It has been indicated in literature that the centralized feature of the radial basis function enables it to effectively model the regression process. Since many papers on the use of SVR demonstrated the favorable performance of radial basis function, for this study it has been chosen for kernel. The parameters that define the selected nonlinear SVR are the cost constant C, the radius of the insensitive tube ε and the RBF kernel parameter γ. In this paper, grid search method is applied to find the optimal parameters. One of the advantages of SVM is the possibility to use the grid search method in order to overcome potential shortcomings of the trial and error method. When using neural networks for prediction, more parameters needs to be known a priori, such as training algorithms, number of layers, number of neurons in hidden layers, etc. The optimal parameters are usually chosen by trial-and-error method. In libsvm library, the "gridregression.py", which conducts grid search over the combination of the parameters, in defined range, with specified step is available. It calculates the training error for all combinations of parameters and picks out the combination

achieving the smallest error. In order to avoid the possible over-fitting, the case when the model is "too well" adjusted for training instances, but behaves badly on test data, the cross-validation is used. The obtained optimal SVR parameters are (C, ε, γ) = (32, 0,0009765625, 1).
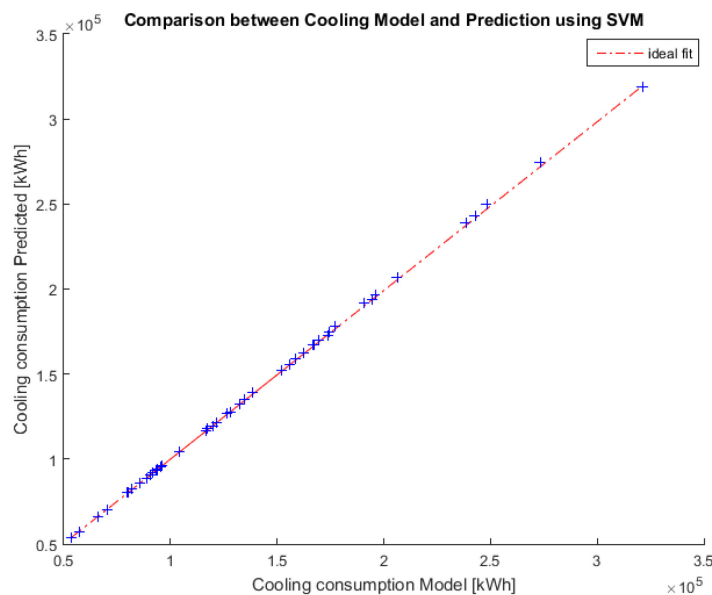

Figure 12. Comparison between modeled cooling consumption and prediction of the support vector machine model

In Figure 12, almost perfect match of modeled cooling consumption and the prediction using SVM is shown. The summary of all prediction indices of the proposed models is presented in Table 1.

Table 1. Prediction indices for proposed models

Model type	R^2 [-]		RMSE [kWh]		MAPE [%]	
	training	test	training	test	training	test
Multiple linear regression	0,9285	0,9408	14.558	14.801	9,4918	9,1534
Feedforward neural network	0,9988	0,9986	2.176	2.619	1,2947	1,2358
Support Vector Regression	0,9994	0,9999	1.306	593	0,2926	0,2605

While the multiple linear regression models shows the coefficient of determination on test data of 0,9408, and achieves the error of 9,15%, the NN and SVM models show much better results. Such low error of 1,24% for neural network prediction and even 0,26% for support vector regression is not surprising, considering that the database is created using modeled values, that come as a result of simulation. Software energy simulations are done by solving mathematical equations, so it is expected that the error would be higher when using the actual measured data. The idea of this paper is not to ideally fit the simulation data, but to show the strength and the possibilities of using statistical methods in energy use prediction.

In Figure 13 prediction of three models are compared.

VII. Conclusion

The main idea of this paper is to investigate the possibility of using statistical ("black box") methods for building energy use prediction. Detailed energy simulation of a typical office building in Belgrade is done using software HAP (Hourly Analysis Program). The influence of various building characteristics is investigated, and for creating building consumption database, three variables that most largely affect the cooling consumption are chosen: specific lighting power, window area and window shade coefficient. Those three parameters are varied and 245 simulations in total are used for creating and testing the models for yearly cooling consumption prediction. The multiple linear model is created and the obtained equation is used for cooling consumption evaluation using these three building parameters as input. The artificial neural network and support vector machine (SVM) models are also developed for prediction and their results are compared with linear regression model. It has been shown that the statistical methods, such are neural networks and

support vector machines can achieve much higher accuracy in prediction than the multiple linear model, gaining almost perfect match with simulated values (mean absolute percentage error for testing the SVM model 0,26%, and for NN 1,24%). Once created, these models can be used to predict the influence of changing one parameter on cooling energy consumption, without necessity to simulate the building in software. Later use of the created NN or SVM models does not require any knowledge about energy simulations, so it can be used as a tool for techno-economic analysis of energy efficiency measures by energy managers. Also, the database could be expanded with other input variables in order to become more widely used.
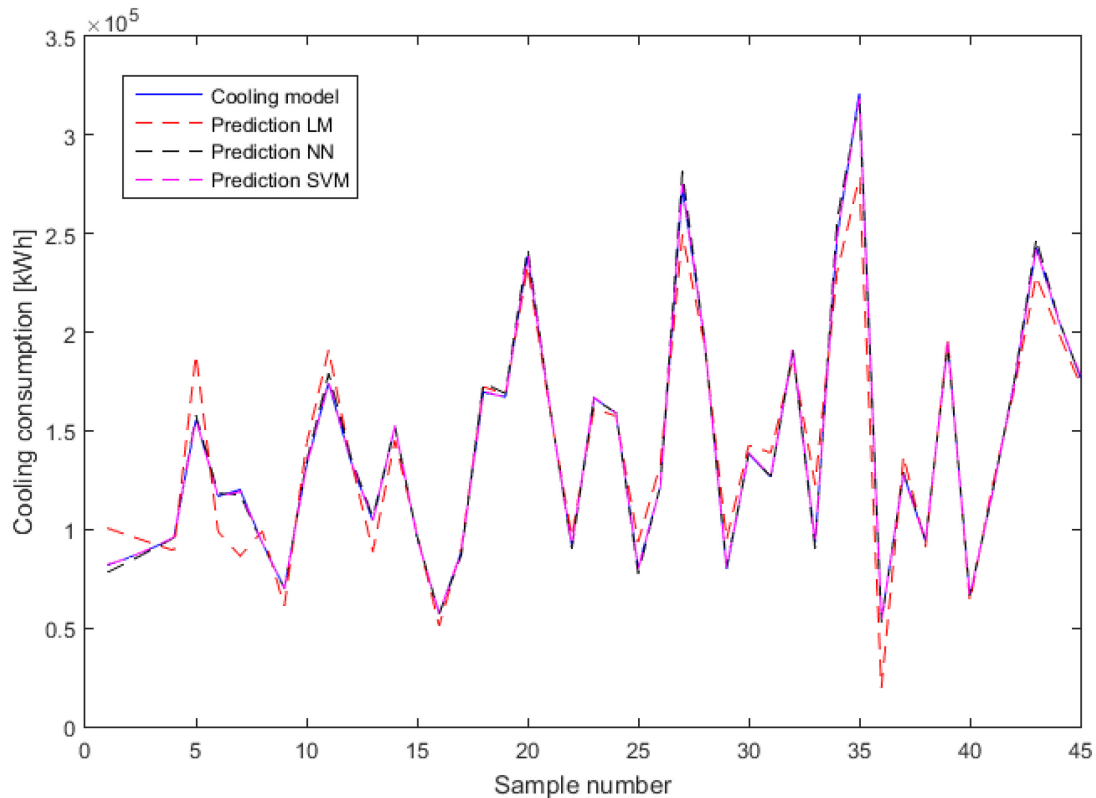

Figure 13. Comparison of the modeled cooling consumption simulation and various models

VIII. Literature

[1] ***, Council., E. P. a. Directive 2010/31/EU of the European Parliament and of the Council of 19 May 2010 on the energy performance of buildings. *Official Journal of the European Union 2010*, L153(2010), pp.13-35.

[2] **Zhao, H.-x. and Magoulès**, F. A review on the prediction of building energy consumption. *Renewable and Sustainable Energy Reviews*, 16, 6 (2012), pp. 3586-3592.

[3] **Kusiak, A., Li, M. and Zhang, Z.** A data-driven approach for steam load prediction in buildings. *Applied Energy*, 87, 3 (2010), pp. 925-933.

[4] **Bauer, M. and Scartezzini, J.-L.** A simplified correlation method accounting for heating and cooling loads in energy-efficient buildings. *Energy and Buildings*, 27, 2 (1998), pp.147-154.

[5] **Catalina, T., Virgone, J. and Blanco, E.** Development and validation of regression models to predict monthly heating demand for residential buildings. *Energy and buildings*, 40, 10 (2008), pp.1825-1832.

[6] **Katipamula, S., Reddy, T. and Claridge, D.** Multivariate regression modeling. *Journal of Solar Energy Engineering*, 120, 3 (1998), pp.177-184.

[7] **Neto, A. H. and Fiorelli, F. A. S.** Comparison between detailed model simulation and artificial neural network for forecasting building energy consumption. *Energy and Buildings*, 40, 12 (2008), 2169-2176.

[8] **Ekonomou, L.** Greek long-term energy consumption prediction using artificial neural networks. *Energy*, 35, 2 (2010), pp.512-517.

[9] **Jovanovic, R. and Sretenovic, A.** Various multistage ensembles for prediction of heating energy consumption, *Modeling Identification and Control,*36, 2 (2015), pp.119-132.

[10] **Jovanović, R. Ž., Sretenović, A. A. and Živković, B. D.** Ensemble of various neural networks for prediction of heating energy consumption. *Energy and Buildings*, 94, (2015), pp.189-199.

[11] **Kumar, R., Aggarwal, R. and Sharma, J.** Energy analysis of a building using artificial neural network: A review. *Energy and Buildings*, 65, (2013), pp.352-358.

[12] **Yu, P.-S., Chen, S.-T. and Chang, I.-F.** Support vector regression for real-time flood stage forecasting. *Journal of Hydrology*, 328, 3 (2006), pp.704-716.

[13] **Hua, X., Ni, Y., Ko, J. and Wong, K.** Modeling of temperature–frequency correlation using combined principal component analysis and support vector regression technique. *Journal of Computing in Civil Engineering*, 21, 2 (2007), pp.122-135.

[14] **Dong, B., Cao, C. and Lee, S. E.** Applying support vector machines to predict building energy consumption in tropical region. *Energy and Buildings*, 37, 5 (2005), pp.545-553.

[15] **Karatasou, S., Santamouris, M. and Geros, V.** Modeling and predicting building's energy use with artificial neural networks: Methods and results. *Energy and Buildings*, 38, 8 (2006), pp.949-958.

[16] **Jovanović, R. Ž., Sretenović, A. A. and Živković, B. D**. Multistage ensemble of feedforward neural networks for prediction of heating energy consumption. *Thermal Science*, 00 (2015).

[17] **Wang, L.** *Support Vector Machines: theory and applications*. Springer Science & Business Media, 2005.

[18] **Vapnik, V.** *The nature of statistical learning theory*. Springer Science & Business Media, 2013.

[19] **Gunn, S. R.** Support vector machines for classification and regression. *ISIS technical report*, 14 (1998).

[20] **Smola, A. J. and Schölkopf, B.** A tutorial on support vector regression. *Statistics and computing*, 14, 3 (2004), pp.199-222.

[21] **Li, Q., Meng, Q., Cai, J., Yoshino, H. and Mochida, A.** Applying support vector machine to predict hourly cooling load in the building. *Applied Energy*, 86, 10 (2009), pp.2249-2256.

[22] **Cherkassky, V. and Ma, Y.** Practical selection of SVM parameters and noise estimation for SVM regression. *Neural networks*, 17, 1 (2004), pp.113-126.

[23] ***, *Carrier Hourly Analysis Program,*
https://www.carrier.com/commercial/en/us/software/hvac-system-design/hourly-analysis-program/

[24] **Chang, C.-C. and Lin, C.-J.** LIBSVM: A library for support vector machines. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 2, 3 (2011)