Jun 26th, 10:40 AM - 12:00 PM

Energy Consumption Prediction for Recreation Facilities using hybrid Neuro-Fuzzy Inference Systems

Paul Banda
paul.banda@student.rmit.edu.au

Muhammed Bhuiyan
RMIT, muhammed.bhuiyan@rmit.edu.au

Guomin Zhang
RMIT University, kevin.zhang@rmit.edu.au

Andy Song
RMIT University, andy.song@rmit.edu.au

9th International Congress on Environmental Modelling and Software
Fort Collins, Colorado, USA, Mazdak Arabi, Olaf David, Jack Carlson, Daniel P. Ames (Eds.)
https://scholarsarchive.byu.edu/iemssconference/2018/

Energy Consumption Prediction for a Recreation Facility using Hybrid Neuro-Fuzzy Inference Systems

Paul Banda[a], Muhammed A Bhuiyan[a] Goumin Zhang[a], Andy Song[b]
[a] School of Civil and infrastructure, RMIT University 124 La Trobe Street, Melbourne, Australia
[b] Computer Science & Information Tech, RMIT University 124 La Trobe Street, Melbourne, Australia
paul.banda@student.rmit.edu.au, muhammed.bhuiyan@rmit.edu.au, kevin.zhang@rmit.edu.au, andy.song.rmit.edu.au

Abstract: Leisure centers are growing in popularity as health and physical fitness awareness is becoming an integral part of human society. Leisure centers consume more energy compared to most office buildings but are less studied in the area of non-residential energy consumption prediction. This work presents an energy consumption prediction effort for a leisure center using a class of two ANFIS based adaptive networks: ANFIS GP and ANFIS SC and multi-linear regression. Climatic, periodicity and energy use data collected over a period of eight months were pre-processed, normalized and split into training and testing sets before being presented to the adaptive networks for neuro-fuzzy inferencing. The results were compared to those of the multi-linear regression models and showed that adaptive networks were superior in performance and there was only a small difference between the two ANFIS algorithms. The combination which gave the best results comprised temperature, the hour of the day and relative humidity with MAE, RMSE and R^2 values of 0.69 kWh, 0.70 kWh and 0.73 respectively, represented by ANFIS SC (0.5) model. This is a good predictive method offering an opportunity for better attainment of efficient energy management.

Keywords: ANFIS GP; ANFIS SC; energy consumption prediction; leisure center.

1 INTRODUCTION

Building energy efficiency is undoubtedly a subject of great importance in global sustainability as the increase of world population and economic growth keep adding pressure on energy supply. In addition, the world is calling for efficient and environmentally clean energy such as the Kyoto Protocol (Grubb et al., 1999). One key area in energy management is the building sector. According to Zuo and Zhao (2014), almost 39% of total energy consumption in the US is from buildings. In the same breath, China's energy usage by buildings is expected to reach a high 35% by 2030 (Chen et al., 2015). In Europe, buildings account for 40% energy usage which is equivalent to 36% CO_2 emissions, European Union Journal, Recast (2010). Recreation facilities cannot be ignored because of the increase in sports facilities in the past decade, which is a result of the increased awareness of health and fitness in modern lifestyle (Costa et al., 2011). Sports facilities account for 8% of the total building energy usage in Europe (EUROSTAT, 2008) and they are popular in Australia as well. Leisure centers in Australia offer an array of different activities under one roof such as, swimming pools (indoor and outdoor), physical fitness centers, spas and children's play park. This arrangement represents high and variant energy use profiles that present energy management challenges for building managers.

Widely used techniques found in the literature for building energy use prediction are broadly classified under engineering (white box), statistical, gray-box modeling and artificial intelligence based (black box) methods. Comprehensive discussions on the above techniques their advantages and disadvantages can be located in review papers (Foucquier et al. 2013; Fumo 2014).

The machine learning group of techniques provide a practical approach to building energy consumption prediction. By abstracting information through learning the structures in input and output data patterns from historical data, machine learning techniques can model building energy consumption. The approach has been a great area of interest in the research community, particularly with the boom in public web data sources and proliferation of smart meters. Simulation and ANN were integrated to optimize energy flows and HVAC control in sports and recreation buildings in (Costa et al., 2011). Artificial neural networks were calibrated using simulated data in Yuce et al., (2014) to predict and optimize energy consumption in an indoor swimming pool, wherein the Levenberg–Marquardt algorithm gave the best results among other trained ANN algorithms. Artuso and Santiangeli (2008) proposed a tool to provide a preliminary estimation of the power and energy required by the sports centers. In other typical building prediction works, ANFIS has been used in Li et al. (2011) to predict hourly energy use of a non-residential building and obtained the smallest CV of 2.66% compared to the back-propagation based neural network. (Jovanović et al. 2015) incorporated ANFIS in their neural network ensemble and their work mention the superiority of ANFIS in heating load energy prediction for non-residential buildings.

Most of the limited studies in the energy use prediction by leisure centers or sports and recreation facilities are mainly reliant on simulation-based results and also literature has limited studies on neuro-computing techniques' making use of real measured energy use data. Therefore, this study is part of the initial efforts to introduce neuro-computing techniques using measured real energy use data in investigating their performance in energy consumption prediction of leisure centers.

2 PROPOSED METHODOLOGY

2.1 Description of Building Case Studies

Don Tatnell Leisure Centre in Melbourne is one of the many Kingston municipality run buildings, which houses various indoor leisure activities under one roof including, a fitness center, a spa, indoor swimming pool, a formal pool and an occasional day-care center. The Don Tatnell leisure center is situated to the north – east corner of the site, with the main entrance facing east. It is constructed from a concrete slab base, with rendered masonry walls and covered with a pitched tin roof. Fitted with aluminum framed windows and doors. The building is located at the corner of Warren Rd & Brisbane Terrace Parkdale VIC, Australia. The longitude and latitude of the site are approximately 145.0924°E and 37.9911°S respectively. The center is open every day of the week, from 6am-9pm weekdays and 7am-6pm weekends.

2.2 Adaptive Network-based Fuzzy Inference System (ANFIS)

ANFIS is a fusion of two learning mechanisms, the linguistic learning ability of the fuzzy logic system and that of the classical neural network. An adaptive network is a network structure comprising nodes and directional links through which the nodes are connected. The parts or all of the nodes are adaptive, implying each output of these nodes depends on the parameters pertaining to this node and the learning rule specifies how these parameters should be changed to minimize a prescribed error measure, Jang (1993). Generally, both square (adjustable nodes) and circle (fixed nodes) symbols are used to represent different properties of adaptive learning, as seen in Figure 1.

In order to perform desired input-output mapping, adaptive learning parameters are updated based on gradient learning rules, Jang (1993). By utilizing a hybrid learning procedure, the proposed ANFIS constructs an input-output mapping of energy consumption (E) and the related factors (T, RH) based on both fuzzy if-then rules and stipulated input-output data pairs. Each or a combination of input parameters for example, temperature and relative humidity (T, RH) are subjected to the fuzzy inference process workflow which mainly entails: fuzzification of the input variables (Layer 1); application of the fuzzy operator in the antecedent (Layer 2); implication from the antecedent to the consequent (Layer 3); aggregation of the consequents across the rules (Layer 4) and finally defuzzification (Layer 5) Jang (1993). Figure 1 illustrates the typical ANFIS structure, in a layered, feed-forward network structure with T and RH as inputs and energy use (E) as the output, and for more details on ANFIS literature, readers are referred to Jang (1993). ANFIS is considered in this study because it is fast, adaptive and does not require a complex mathematical model, while at the same time the developed prediction tool can be implemented quickly, which is ideal for the complex task of electrical energy consumption prediction. In this study, two variations of ANFIS algorithm are investigated namely, ANFIS GP and ANFIS SC.

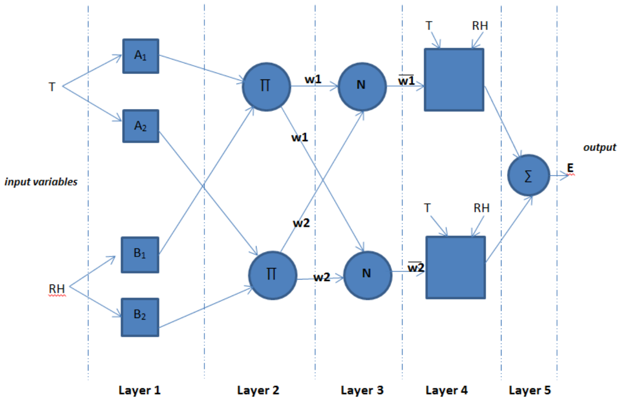

Figure 1. ANFIS structure.

2.3 Grid Partitioning and Subtractive Clustering (ANFIS)

ANFIS GP is a combination of Grid Partitioning and ANFIS. The input space is divided into a number of local fuzzy regions using axis paralleled partition by considering the number of membership functions (MFs) and their corresponding types for each dimension. Prior fuzzy sets and parameters are calculated using the least squares method based on the partition and MF types. During training, ANFIS identifies and learns the fuzzy rules and refines parameters. The ANFIS SC fuzzy inference system combines the Subtractive Clustering (mountain) method and ANFIS. The potential of data points in the feature space is measured resulting clustering of data points. The subtractive clustering notion is that, each data point defines the cluster center based on density of surrounding data points. The data point demonstrating highest potential is selected as the first cluster center, and the next cluster center is determined by revising potential of data points to cancel the effect of the preceding cluster center. The cluster radius is important in deciding the number of clusters, such that, a small number cluster radius gives many small clusters in the data space that result in many rules and vice versa. Once the number of fuzzy rules and MF is determined, linear squares estimate is used to determine corresponding output MF, thus generating a valid FIS.

2.4 Multiple-Linear Regression Model Development

Multiple linear regression, a data-driven technique, is a popular technique of mathematically modeling relationships between independent and dependent variable(s). In general, response variable Y may be related to n regressor variables.

$$Y = \beta_0 + \beta_1 * x_1 + \beta_2 * x_2 + \beta_3 * x_3 + \cdots .. + \beta_n * x_n \quad (1)$$

where β_0 is a constant and β_i, i = 1 ••• n are regression coefficients. This model describes a hyperplane in n-dimensional space of the regressor variables x_j. The parameter β_j represents the expected change in response Y per unit change in x_j when all the remaining independent variables x_i (i ≠ j) are held constant. In this study, the coefficients β_0, β_1, $\beta_{2,}$ β_n were determined through execution in the SPSS statistical software package.

3 MODEL DEVELOPMENT

3.1 General Description of Data Set

Don Tatnell Leisure energy monitoring system provided 15min resolution Electrical Energy Consumption (EEC) (kWh), while the outdoor weather conditions data (15min resolution) were obtained from the Bureau of Meteorology in Victoria. Inputs relating to periodicity namely, the hour of the day, the day of the week, the day of month and month of the year were derived from the 15min resolution data. Outdoor dry bulb temperature, solar radiation, relative humidity, wind and periodicity data were used as inputs while building's EEC values were regarded as the output. Training data comprised data from May 2017 to September 2017 while data from November to December 2017 was used for testing the calibrated model.

3.2 Implementation Details

This paper focuses on the prediction of EEC of a fitness center which contributes to the overall aggregate energy consumption at Don Tatnell Leisure center. The 15-minute resolution samples of eight potential input variables namely, month of the year, day of the week, hour of the day, temperature, solar radiation, wind velocity, relative humidity and one output variable (energy consumption) were considered in developing the proposed framework to predict the energy consumption. In both ANFIS and MLR models, May to October (6 months) data was assigned as the training set and Nov to December (2 months) input-output pairs was used for testing the performance of the model predictions. Both ANFIS and MLR models employ the same training and testing data sets for an appropriate performance comparison.

In this study, two algorithms of ANFIS are investigated namely, ANFIS GP and ANFIS SC together with their membership function types, in the prediction of EEC at the leisure center. Results on prediction accuracy from these neurocomputing techniques were compared to those of another counterpart traditional data-driven technique, MLR. The topology and learning procedure underlying ANFIS is investigated and determined.

3.3 Statistical description of Data

SPSS statistical package was used to run the bi-variate correlation of the attributes used for EEC prediction and the output from SPSS is given in Table 1.

Table 1. Potential predictor variables.

Category	**Dataset**	**Unit/Index**	**Correlation with (*EEC* kWh)**
Environment	Air temperature (T)	°C	0.739**
	Relative humidity (RH)	%	-0.406**
	Solar radiation (SR)	W/m^2	0.321**
	Wind (*u*)	Km/hr.	0.172**
Time indicator	Hour of day (H)	1-24	0.211**
	Day of month (DoM	1-31	0.146**
	Day of week (DoW)	1-7	-0.021**
	Month of year	5-12	0.388**

****** Pearson Correlation is significant at the 0.01 level (2-tailed).

As can be seen from Table 1, from all the potential input variables outdoor temperature stands out as the one parameter with a high linear correlation with EEC. Relative humidity, however, has a negative correlation to EEC meaning as relative humidity goes down EEC tends to be increasing. These two correlation output results follow the natural laws governing the relationships between the climate factors and EEC at this study site.

3.4 Data Transformation

Since the climatic variables used have different units of measurements (Table 1) major differences among values exists, thus it was considered pertinent to normalize the primary data. Normalization ensures that data is mostly uniformly distributed between inputs and outputs of the network which permits for faster, memory efficient and accurate forecasting of results by the network. Data in this study was normalized using the Min-max normalization technique represented in Equation (2) below:

$$X_n = \frac{X_i - X_{min}}{X_{max} - X_{min}} \qquad (2)$$

where X_n is the normalized value, X_i is the real value of the variable, X_{max} is the variable's maximum value, X_{min} is the variable's minimum value. All input and output data were normalized using Equation (2) above.

3.5 Metrics for Prediction Assessment

Determining with which criteria to assess the performance of the developed model is one major step since it has a bearing on optimal network topology selection. System errors that occur during learning and prediction accuracy are generally the widely adopted performance criteria. During prediction accuracy determination, ANFIS ability to estimate is determined by using data that has not been used in the training process. A plethora of model performance analysis methods exist in literature and described below are the ones adopted in this study.

The study used multi-criterion performance evaluation by considering determination coefficient (R^2), root mean square error (RMSE) and mean absolute error (MAE). These simple error analysis and linear regression statistical principles used to evaluate performance between the real energy use data and ANFIS models are represented by equations below;

$$R^2 = \frac{\left[\sum_{i=1}^{N}\left(E_{meas,i} - E_{meas\ mean,i}\right) \times \left(E_{pre,i} - E_{pre\ mean,i}\right)\right]^2}{\sum_{i=1}^{N}\left(E_{meas,i} - E_{meas\ mean,i}\right)^2 \times \sum_{i=1}^{N}\left(E_{pre,i} - E_{pre\ mean,i}\right)^2} \qquad (3)$$

$$\mathrm{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left[E_{meas,i} - E_{pre,i}\right]^2} \qquad (4)$$

$$\mathrm{MAE} = \frac{1}{N}\sum_{i=1}^{N}\left[E_{meas,i} - E_{pre,i}\right] \qquad (5)$$

where $E_{meas,i}$ is the measured energy consumption value at point i, $E_{pre,i}$ is the energy consumption prediction value at point i, $E_{meas\ mean,i}$ and $E_{pre\ mean,i}$ represents the average values of the corresponding variables, *N* is the number of data considered.

Higher R^2 values and lower RMSE and MAE values indicate better prediction accuracy by the model. The R^2 measures the extent of the linear relationship between two variables and values range from 0 to 1. A value closer to 1 indicated a good agreement between the measured and predicted values whereas values closer to 0 indicate inferior agreement hence indicates poor model performance. The MAE statistic measures the goodness of fit relating to moderate EEC values while RMSE considers the goodness of fit relative to high EEC values. In this study, the best model is chosen based on statistical parameters obtained during the testing phase.

4 MODELING STRATEGY

The selection criteria for inputs can largely be influenced by a general understanding of the physical system to be modeled. In the study, 15min resolution observations of outdoor temperature (T), mean relative humidity (RH), wind speed (U_2), mean solar radiation (SR), hour of the day (H), day of the month, day of the week, and month of the year, were used as inputs for estimation of electrical energy consumption. The correlation matrix between all the potential input variables is presented in Table 1. Following careful sensitivity analysis of the potential inputs together with correlation analysis experiments, only five models were established for discussion. These five models are the one-factor input vector model (outdoor air temperature), two-factor input vector models (outdoor air temperature, relative humidity; outdoor air temperature and the hour of day; the hour of the day and relative humidity) and lastly the three-factor input vector model (outdoor air temperature, relative humidity and hour of the day) respectively.

5 RESULTS AND DISCUSSION

In order to achieve the goal of energy prediction, the ANFIS models are evaluated and compared using a set of measured data. The data set contains highlights as a function of time the electric power consumption by the fitness center with a 15min sampling rate over a period of eight months.

5.1 Comments on Input Performance

Temperature tends to be the most significant single parameter of all the inputs considered, in influencing energy consumption at the fitness center. This confirms the results of the statistical correlation analysis shown in Table 1, in which temperature shows the highest correlation to energy consumption at the facility. However, during EEC prediction, temperature alone shows some relatively higher error rates having MAE, RMSE and lower R^2 values of 0.93 kWh, 1.33 kWh and 0.58 respectively. Hour of the day as a single input, as highlighted in Table 3, has a weak influence in predicting energy consumption, with high MAE and RMSE values of 1.43 kWh, RMSE 3.38 kWh and a low coefficient of determination value of 0.27 for the leading ANFIS SC model. A combination of the hour of day and temperature did not significantly improve EEC prediction, though the RMSE error value decreased slightly for the ANFIS SC algorithm. The combination which gave the best results comprised temperature, hour and relative humidity. This result shows that of all the factors considered at the study site for EEC prediction, temperature parameter needs to be given due consideration during data collection. The results further highlight the importance of the interaction of the climatic variables in influencing EEC at the site.

5.2 ANFIS GP and SC Model Performance

Generally, ANFIS models performed better than MLR models (Table 3) for all the models developed in this study. The study results show that for the problem of energy consumption prediction at Don Tatnell's fitness center, ANFIS SC algorithm is superior to its counterpart ANFIS GP, in EEC prediction at the fitness center according to the model performance measurement statistics. ANFIS SC three-factor input vector with a range of influence, squash factor and acceptance ratio of 0.5, 0.5 and 0.15 respectively is the superior model. This model comprising temperature, relative humidity and the hour of the day as inputs, is the superior model with the lowest MAE, RMSE and highest R^2 values of 0.69 kWh, 0.70 kWh and 0.73 respectively. The study observed that generally, the model did not have many problems in predicting lower consumption values of up to 4 kWh. The model, however, did not predict well values above this 4 kWh threshold. It is also observed that both ANFIS algorithms were sensitive to high EEC values as a result of abrupt increase in consumption, evidenced by the underestimation or overestimation of high EEC values, however, high consumption values with a gradual increase did not present prediction difficulties for both ANFIS algorithms. This phenomenon can be attributed to the relatively finer resolution of the EEC data set which does not allow for ample adjustment by the prediction model.

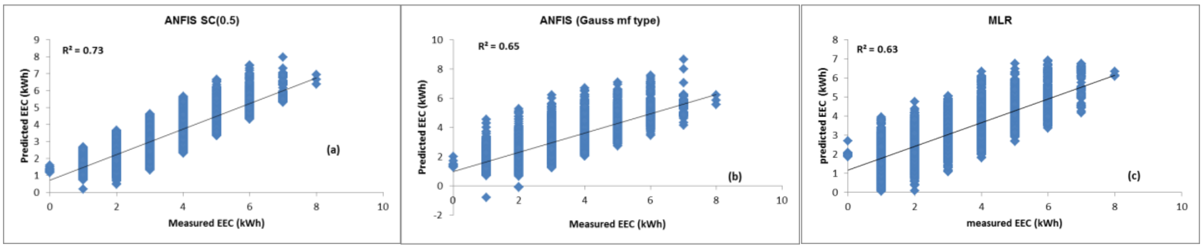

Figure 2. Scatterplots of observed versus predicted EEC values for the ANFIS and MLR models during the testing phase.

5.3 Overall Simulation Results of the Prediction Models

Efficient and reliable modeling of EEC is important for building facilities managers. Such a task is widely carried out based on conventional methods such as simulation engines and linear regression analysis. However, alternatives are offered by novel neurocomputing techniques such as ANNs, support vector machines, genetic programming, or ANFIS of which herein, the latter is used to model EEC at a fitness center.

Table 2. ANFIS computing parameters of the superior EEC prediction model ANFIS SC (0.5).

Type	Sugeno
Input structure	1 x 3
Output structure	1 x 1
Squash factor	0.5
Number of linear parameters	60
Number of nonlinear parameters	90
Total number of parameters	150
Number of checking data pairs	0
Number of fuzzy rules	12

As seen from Table 2, the total number of parameters for the superior ANFIS model developed is 150, and the number of linear parameters is 60, while the number of non-linear parameters is 90, with 12 fuzzy rules based on the three performance criteria. The initial parameters of this ANFIS model are identified using the subtractive clustering method having a clustering radius (R) of 0.5 that was optimally determined through a trial and error procedure. The clustering radius was varied between 0.4 and 1 with a step size of 0.1, in seeking to minimize the RMSE obtained on a representative testing set. The number of clusters was determined experimentally, by developing various models and analyzing the rules and their respective parameters. The final number of clusters (rules) is twelve for the superior model.

Table 3. Comparative analysis of the three model types in energy use prediction in the testing phase.

	Testing								
Input	ANFIS GP			ANFIS SC (0.5)			MLR		
	MAE	RMSE	R^2	MAE	RMSE	R^2	MAE	RMSE	R^2
T	0.92	1.33	0.58	0.97	1.48	0.50	3.07	10.89	0.58
H,T	0.92	1.35	0.59	0.90	1.33	0.58	3.07	10.92	0.59
H,RH,T	0.78	0.89	0.65	0.69	0.70	0.73	0.77	0.95	0.62

As shown in Table 3, the MLR models performance (for all input combination(s)) were relatively inferior compared to the ANFIS type models as shown by the former's high error values in the testing phase. The rather inferior performance by the linear MLR models can be attributed to the non-linearity and complexity of building related processes. Generally, ANFIS SC algorithm is the domineering method of all the models created in the testing phase, as it had the lowest error values. Figure 2a – c shows scatterplots of observed against calculated values of energy use for ANFIS GP, SC and the MLR models. Addition of relative humidity and hour of the day inputs resulted in a lowering of the MAE and RMSE values for all the models. ANFIS models have difficulties if the number of inputs increases as more rules will need to be generated and this tends to make the algorithm slow and perform poorly, therefore, work is underway to include more neurocomputing techniques that can incorporate more inputs.

6 CONCLUSIONS

Most studies in the energy use are mainly reliant on simulation engines, which require domain expertise, manual inputs and a time-consuming process. Limited work is reported using neuro-computing techniques and real energy usage data. This study is part of the initial efforts to address this, investigating their performance in energy consumption prediction of these afore-mentioned building types. The capability for estimating electrical energy consumption directly from observed climatic data and time indicators, using ANFIS for a leisure center was tested. Before commencement of ANFIS modeling, the eight-month period data set was divided into two sets namely training and testing subsets. According to the results, ANFIS outperformed the conventional MLR method based on criteria for model accuracy considered. During EEC prediction, both ANFIS algorithms were generally superior to MLR models in both the training and testing phases. A combination of temperature, relative humidity and hour of the day, has the greatest influence in modeling EEC at the fitness center. ANFIS can be an alternative solution for modeling electrical energy consumption at the leisure center. The results also highlight the importance of the interaction of the climatic variables, in influencing EEC at the site.

The fitness center studied in this paper represents a section of a large leisure center. Our further work is underway to predict the overall energy consumption of the entire leisure center. This work, therefore, acts a precursor to the overall bigger scale prediction work under consideration, were more neuro-computing models and larger data sets used together with data mining techniques from within the various sections in the leisure center will be studied.

ACKNOWLEDGMENTS

The authors would like to acknowledge the great contribution by Kingston City Council, Melbourne (Anthony Ziem) and Bilal Cemek (Ondokuz Mayis University) for providing all the data and technical support during the study.

REFERENCES

Artuso, P. and Santiangeli, A., 2008. Energy solutions for sports facilities. International journal of hydrogen energy, 33(12), pp.3182-3187.

Chen H, Lee W, Wang X. 2015. Energy assessment of office buildings in China using China building energy codes and LEED 2.2. Energy and Buildings.86:514-524.

Costa, A., Garay, R.S., Messervey, T. and Keane, M.M., 2011, November. Value of building simulation in sports facilities operation. In 12th Conference of International Building Performance Simulation Association.

EUROSTAT (2008) EUROSTAT: Environment and energy database

Foucquier A, Robert S, Suard F, Stéphan L, Jay A. 2013. State of the art in building modeling and energy performances prediction: a review. Renewable and Sustainable Energy Reviews. 23:272-288.

Fumo N. 2014. A review of the basics of building energy estimation. Renewable and Sustainable Energy Reviews.31:53-60.

Jang, J.S.R., "ANFIS: Adaptive-Network-based Fuzzy Inference Systems," IEEE Transactions on Systems, Man, and Cybernetics, Vol. 23, No. 3, pp. 665-685, May 1993.

Jovanović RŽ, Sretenović AA, Živković BD. 2015. The ensemble of various neural networks for prediction of heating energy consumption. Energy and Buildings.94:189-199.

Recast E. 2010. Directive 2010/31/EU of the European Parliament and of the Council of 19 May 2010 on the energy performance of buildings (recast). Official Journal of the European Union.18:2010.

Trnsys A. 2000. Transient System Simulation Program. The University of Wisconsin.

Yuce B, Li H, Rezgui Y, Petri I, Jayan B, Yang C. 2014. Utilizing artificial neural network to predict energy consumption and thermal comfort level: An indoor swimming pool case study. Energy and Buildings.80:45-56.

Zhao H, Magoulès F. 2012. A review of the prediction of building energy consumption. Renewable and Sustainable Energy Reviews.16:3586-3592

Zuo, J. and Zhao, Z.Y., 2014. Green building research–current status and future agenda: A review. Renewable and sustainable energy reviews, 30, pp.271-281.