

Building Energy Modeling Using Artificial Neural Networks

Maya Arida, Nabil Nassif, Rand Talib and Taher Abu-Lebdeh

*Department of Civil, Architectural and Environmental Engineering,
North Carolina A&T State University, Greensboro, USA*

Article history

Received: 07-05-2017

Revised: 31-05-2017

Accepted: 06-06-2017

Corresponding Author:

Nabil Nassif

Department of Civil,
Architectural and
Environmental Engineering,
North Carolina A&T State
University, Greensboro, USA
E-mail: nnassif@ncat.edu

Abstract: Accurate modeling of total building energy is now vital to reduce energy consumption. This is especially true for buildings since they are considered as the largest energy consumer in the United States. This paper investigates modeling methods for building energy-systems using non-linear auto-regression artificial neural networks. The proposed model can forecast the whole building energy consumptions given the four input variables: Dry-bulb and wet-bulb outdoor air temperatures, hours of day and type of days. In addition, the paper presents optimization process that uses genetic algorithm to determine the best model structure by minimizing the model errors. Statistical indexes such as the root mean-square error RMSE and the coefficient of variance CV of RMSE are used to measure the model accuracy. The data was collected from existing buildings and from simulations. The collected data was used to test and train the proposed models as well as in the optimization process. Various neural network structures were tested using different inputs and feedback delays. The results show that the proposed model can accurately predict the energy consumptions. The CV values were within a range of 1.7 and 7.7%. It also proves that the model can be used for saving estimation applications and different energy efficiency.

Keywords: Building Energy Model, Neural Network, HVAC System, Regression Model

Introduction

In 1972, sustainable development concept existed and extended throughout the years where it becomes one of the most important solution to save natural resources and energy. However, since energy use is continuously in the rise, it may result in the consciousness of the global warming. Thus, the need for energy efficient design has become increasingly urgent. In the United States, buildings consume roughly 48% of the nation's total Energy consumption (EIA), from which heating and cooling systems consume about 55% of this energy, while the other 45% used by lights and appliances of existing buildings. If this level of energy consumption continues, buildings would be considered as the top consumer of global energy by 2025. Therefore, it's important to understand the parameters that effect the energy use in the building, specially the HVAC systems and how to reduce such energy use by designing a building that will not only provide higher quality of living for the inhabitants but also has the potential to reduce energy use.

Several key factors may cause the increase in energy use and they can range from weather, size of building and use of building. HVAC systems can contribute to half of the energy consumption in developed countries. Therefore, finding a new and advanced modeling methods to predict the energy consumption become urgent. Further, the incorporation of Artificial Neural Networks (ANNs) in the computing process is inspiring and can be an invaluable resource to heating and air conditioning system architects. These methods are being regarded as the next main progress in the computing industry. This paper investigates modeling methods for building energy-systems using non-linear auto-regression artificial neural networks. ANNs by nature are pattern and experience based learning programs. Most importantly, this modeling tool will provide researchers and designers a powerful, simple method in addressing their needs of creating more energy-efficient HVAC systems.

Recent modern buildings are usually equipped with power meter dashboard and Building Auto-motion System (BAS) which provides large amount of data and the ability of collecting such data. However, because of the need of embedded calculations means and centralized

explanation, these buildings still do not operate optimally. Thus, there is a true need to analyze how new computational techniques can be used to generate the required data and to utilize the benefits of the available online data. Additionally, it is needed to operate many intelligent applications such as modeling, optimization, energy efficiency and energy assessment and fault detection and diagnosis (ASHRAE, 2011; Wang *et al.*, 2010; Seem, 2007; Nassif, 2013).

This paper presents optimization process using genetic algorithm to determine the best model structure by minimizing the model errors. To determine the optimal model structural (Model parameters) that produces the minimum error between the actual and simulated data during the test period, the genetic algorithm GA is used. The problem variables (or ANN model parameters) are: (1) Input time delays, (2) feedback time delays and (3) number of neurons (hidden layer size). The objective function could be the Mean-Square Error (MSE), root mean-square error RMSE, or coefficient of variance CV. The constraints cover the lower and upper limits of design variables are the maximum and minimum size of number of neurons and time delays.

Preprocessing and Data Analysis

Data Collection

In this study, data was collected from two buildings. For each study case, the building energy consumption was collected every 15 min by smart meter and converged into a spreadsheet. Also, the weather data which was recorded every hour uploaded in a different form of spreadsheet called (.csv) files which contains potential conditions that will have an impact at the output of the energy consumption. This includes time of the recorded weather (local airport). After the hourly data was collected for each month, it was organized into a single spreadsheet by assembling all files together consecutively from March 2014 to May 2016. Energy consumption data was organized for each month in a form of spreadsheet with the hourly data.

Preprocessing

This work utilized two different types of data:

- Synthetic data which was collected through eQuest software. It uses the actual organized data for the dry and wet temperature, type of the day and hour of the day for one of the building. Three different trends were noted: Occupied period which is 8 am and 5 pm, unoccupied period (between 5 pm and 7 am) and holidays and HVAC system start and stop period. HVAC system turns on at 7 am, 1 h before occupied period and it turns off at 4 pm, 1 h before the end of occupied period, but the ventilation system stays on between 4 to 5 pm.

Figure 1 shows the hourly energy consumptions as a function of outdoor air temperatures for synthetic data

- Actual data which was collected using dashboard power meter system for two of the buildings. The data was organized into spreadsheet and used to generate the model. Figure 2 shows the hourly energy consumptions as a function of outdoor air temperatures for the actual data.

Modeling

As aforementioned, the need for accurate dynamic models is vital in developing the energy solution tools. Depending on the types of functions and the required precision, the models can vary from simple to very sophisticated and detailed calculations (Nassif *et al.*, 2008). Moreover, it is of practical importance to develop simple, yet accurate and reliable models to better capture the real dynamic behavior of the subsystems and overall system over the entire operating range. In this study, non-linear time series auto-regression artificial neural networks are used. Artificial neural networks are computational models that are inspired by the natural neurons of the brain.

Natural neurons collect signals by synapses located on the membrane of a neuron. When the received signals are good enough, the neuron is activated and emits a signal through to the axon. That signal may be sent to different synapse and also may trigger more neurons for activation. The strength of the neurons' interconnections is called the adaptive weights. These are numerical parameters that are tuned by a learning algorithm. The higher the weight of an artificial neuron, the stronger the input will be. Weights can also be a negative value, so signals may have inhibited by a negative weight. Depending on the weights, the computation of the neuron in a network will be different. Adjusting the weights can produce the outputs needed for specific inputs.

Figure 3 shows the schematic of non-linear time series auto-regression artificial neural networks used in this research work.

It should be noted that the model inputs are outdoor air dry-bulb temperature, wet-bulb temperature, type of day as (weekend/weekday) and the hours of day. The model output is building level energy consumption. In the process, the type of day variable is either "zero" or "one" corresponding to weekdays or weekend and holidays. The hour of day varies from 1 to 24 h. A large set of various neuron networks structures with various number of neurons and time delays n are tested in order to determine the best and simple structures yielding adequate accuracy in terms of mean-square error MSE, root mean-square error RMSE, or coefficient of variances CV.

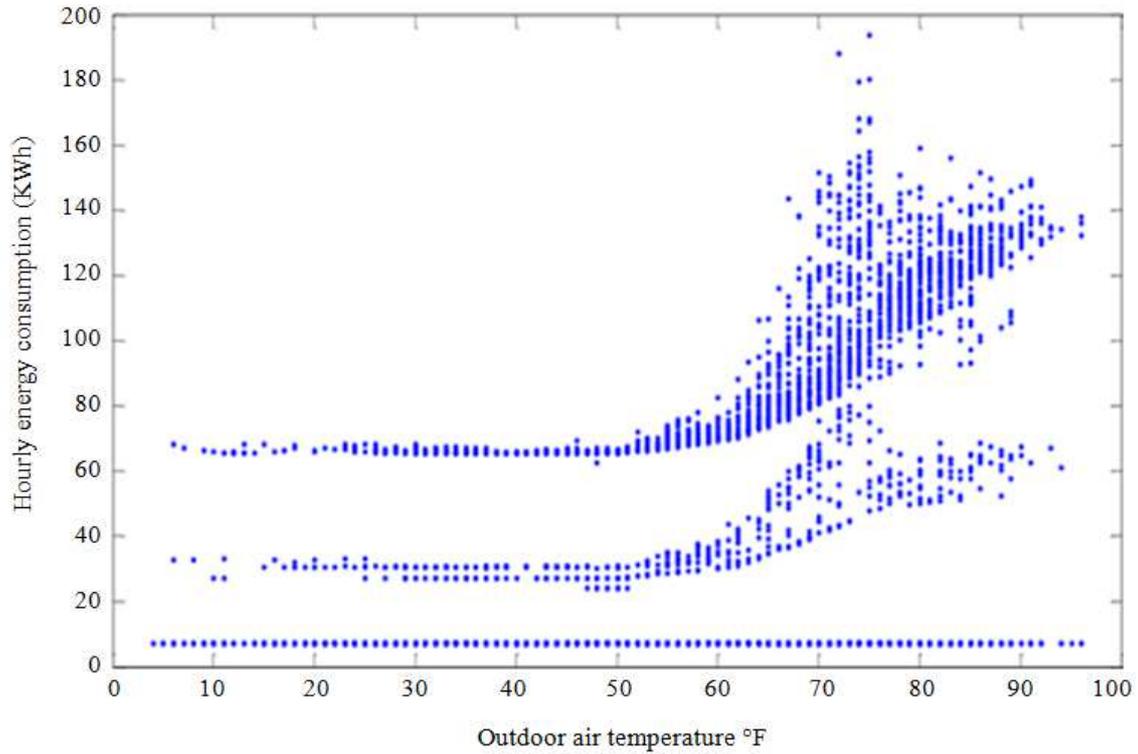


Fig. 1. Hourly energy consumptions as a function of outdoor air temperatures for the simulated building

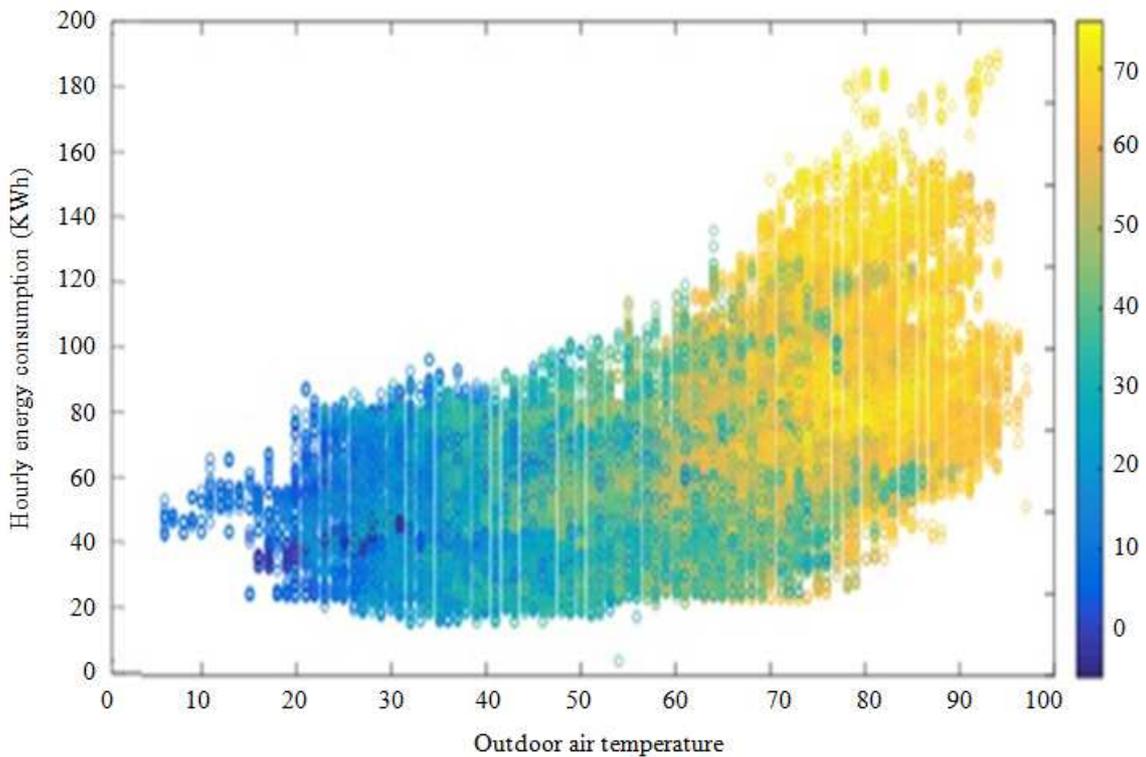


Fig. 2. Hourly energy consumptions as a function of outdoor air temperatures for the actual data

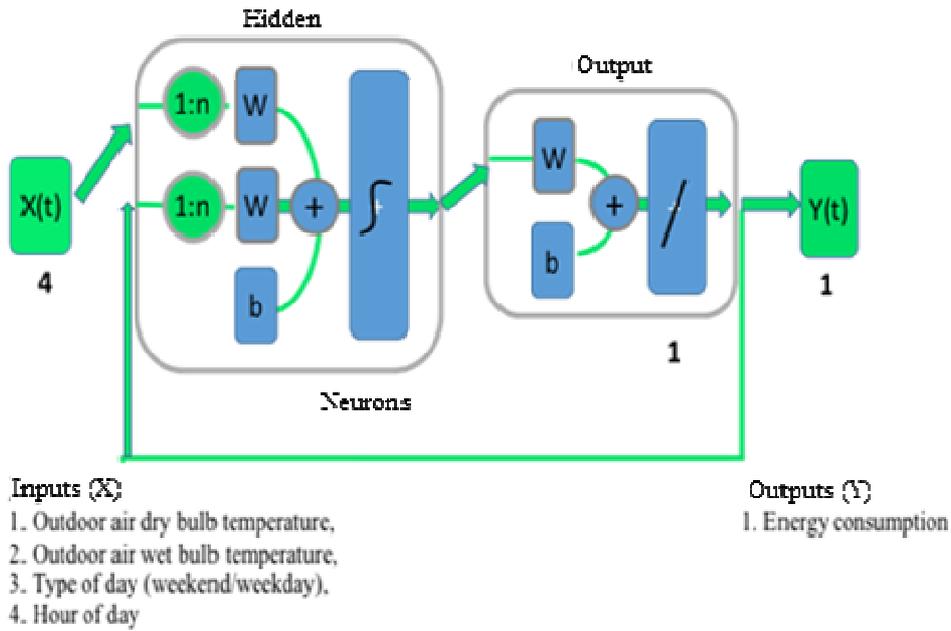


Fig. 3. A schematic of non-linear time series auto-regression artificial neural networks

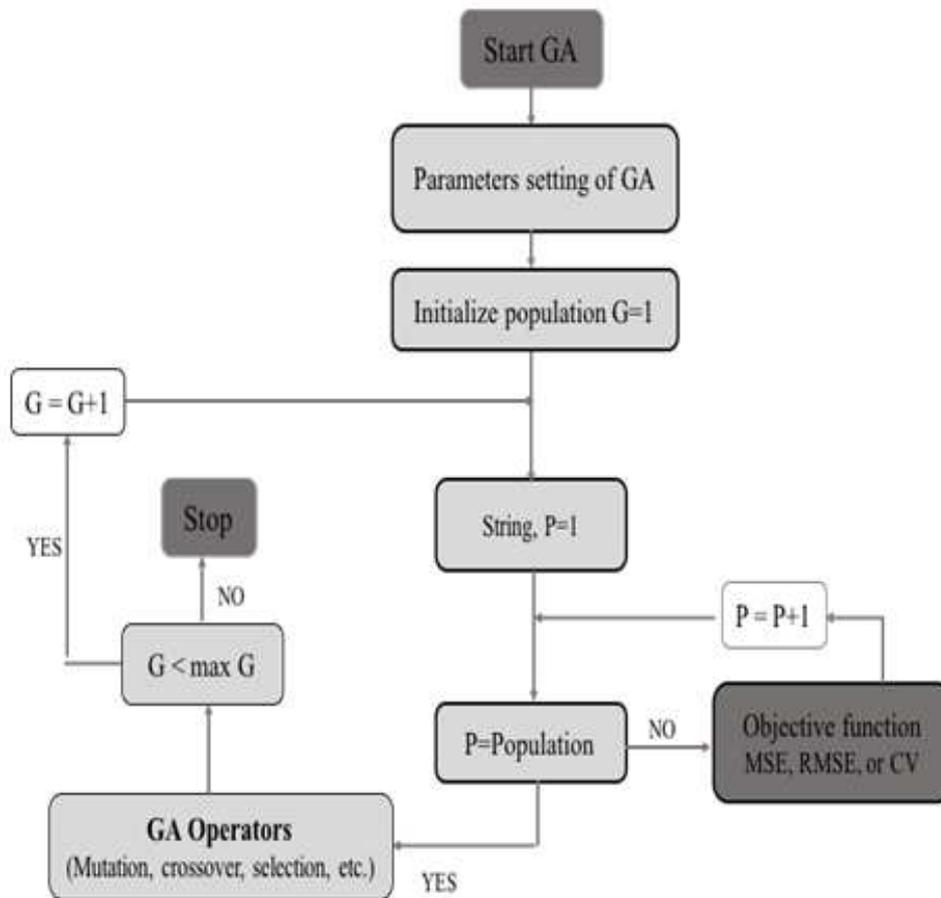


Fig. 4. Flow chart of genetic algorithm GA for ANN model parameters optimization

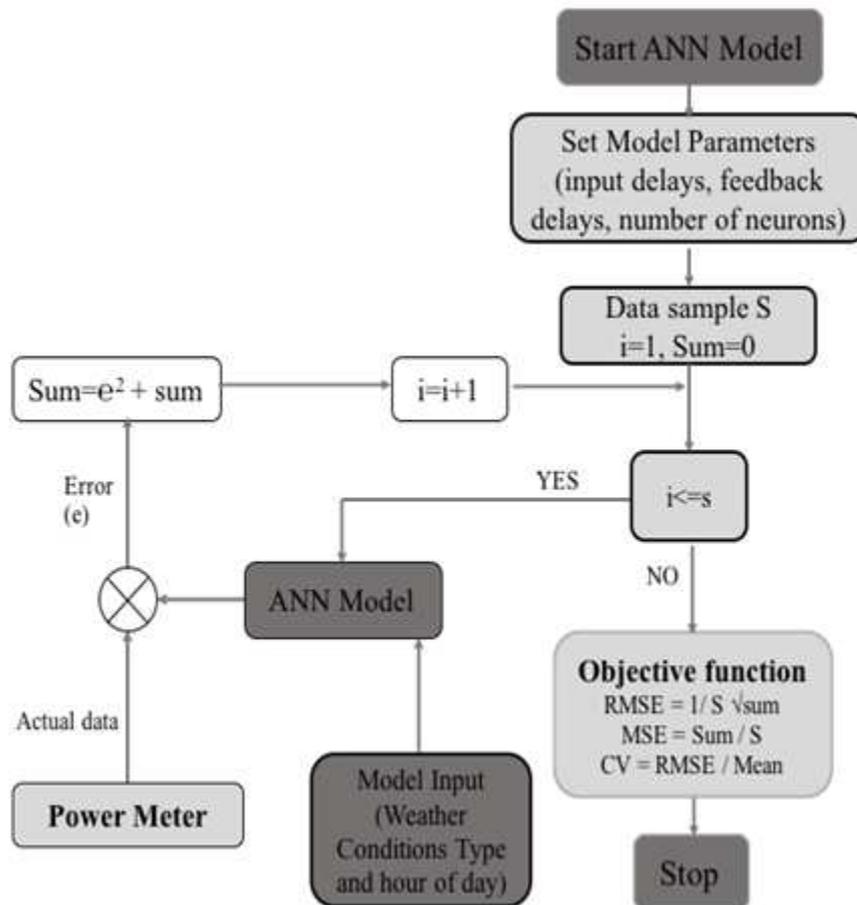


Fig. 5. Flow chart of objective function calculations for a sample S of data

Optimization

To determine optimal model structural (Model parameters) that produces the minimum error between the actual and simulated data, the genetic algorithm GA is used. The problem variables (or ANN model parameters) are: (1) Input time delays, (2) feedback time delays and (3) number of neurons (hidden layer size). The objective function is the Mean-Square Error (MSE), Root Mean-Square Error (RMSE), or Coefficient of Variance (CV). The constraints cover the lower and upper limits of design variables, such as the maximum and minimum size of number of neurons and time delays. Figure 4 shows the flow chart of genetic algorithm GA for ANN model parameters optimization.

The genetic algorithm GA is a method for solving both constrained and unconstrained optimization problems that is based on natural selection, process that drives biological evolution. The GA is extensively used for HVAC system optimization (Nassif, 2014). The genetic algorithm repeatedly modifies a population of individual solutions. At each step, the genetic algorithm selects individuals at random from the current population

to be parents and uses them to produce the children for the next generation. Over successive generations, the population "evolves" toward an optimal solution. As shown in Fig. 4, the GA starts with a random generation of the initial population (initial solution) and ends with the optimal solutions including the optimal variables. It is worth noted that the problem variables (input time delays, feedback time delays and number of neurons) represent an individual solution in the population. The performance or objective function (MSE, RMSE, or CV) of each individual of the first generation is estimated. The second generation is generated using operations on individuals such as selection, crossover and mutation, in which individuals with higher performance (fitness) have a greater chance to survive. The performance of each new individual is again evaluated. The process is repeated until the maximum number of generations is reached.

Figure 5 shows the flow chart of how objective function is calculated for each individual solution. For each individual solution (specific model parameters), the ANN model simulates energy consumption for all sample data "S" and compares the results with the actual

values collected from the power meters. The Mean-Square Error (MSE), root mean-square error, or Coefficient of Variance (CV) (i.e., objective function) is determined and sent back to GA.

Implementing and Results

Parametric Study

The first part of the results, which depends on parametric study, was obtained by using MATLAB software. In this part of the analysis, the synthetic data were divided into two sets: (1) Training set which contains data collected during January first to August thirty first and (2) testing set from September first to December thirty first. The Artificial neural network model was built using four inputs: (1) Outdoor air dry-bulb, (2) web-bulb temperatures, (3) type of day (weekend/weekday/holidays) and (4) the hour of day. The output from (eQuest) process also entered to be compared with output from the ANN

Synthetic Data

To predict the best model structure, various ANN structures were tested with various input delays ranging between 1-3 and feedback delays between 1-3. The number of neurons between 5-100 with an increment of five were selected in order to better understand the relationship between the investigated variables and to evaluate its forecasting aptitude by calculating the MSE and CV for each structure.

After obtaining all results, including all the CV for testing and training, the results were transferred to excel sheet for organization and comparison. The design chart includes three numbers of CV for the same delay time but in three different feedback delays time. The charts were constructed for both the testing and training results separately as shown below. Moreover, the data which was compiled together in two charts shows that the performance of the energy consumption in terms of CV is better in ANN number less than 50. Higher value will result in higher CV. This is due to the over fitting in neurons number more than 50. Figure 6 shows the values of CVs for different number of neurons and input delays for testing and training period. The test was carried out for input time delays varying from 1 to 3 and number of neurons varying from 5 to 100 with an increment of five. Feedback time delay = 1.

Values of the minimum MSE and minimum CV for the optimal solution are tabulated in Table 1. These values correspond to number of neurons of 45 at time delays of one and feedback delays of one for both the training and testing period.

Actual Data

The actual data collected by the power meter for each building included: Outdoor air dry bulb and wet bulb temperatures, type of day (weekend/weekday/holidays)

and the hour of day. To predict the best model for the actual data, various ANN structures were tested with various input delays ranging between 2-6 and feedback delays between 2-6. The number of neurons was between 10-60 with an increment of ten were selected to better understand the relationship between the investigated variables and to evaluate the forecasting aptitude by calculating the MSE and CV for each structure. After obtaining all results, CVs were determined for testing and training. The results were then transferred to excel sheet for organization and comparison. The design chart included three numbers of CV for the same delay time and feedback delays time. The charts were constructed for both the testing and the training results separately as shown below. Figure 7 and 8 show the resulted coefficient of variance CVs for the first building (Blufordlibrary) where the number of neurons varying from 10 to 60 with an increment of ten, input delay = 2, 4, 6 and feedback delays equal to the input delays for training and testing period.

Values of the minimum MSE and minimum CV for the optimal solution at 40 neurons, time delays of six and feedback delays of six for both training and testing period are tabulated in the Table 2.

The process was repeated to cover the second existing buildings (Corbett Building). As in building 1, the actual data was divided into training set of five months (from Feb to June 2014) and testing set of three months (July, August and September 2014). After obtaining the required results and determining the CV for testing and training, the results transferred to excel sheet for organization and comparison. Again, the design chart included three numbers of CV for the same delay time and feedback delays time for both the testing and training as shown below.

Figure 9 and 10 show the resulted coefficient of variance CVs for the second building (Corbett Building) for number of neurons varying from 10 to 60 with an increment of ten, input delay = 2, 4, 6 and feedback delays equal to the input delays for training and testing period.

Table 3 illustrates the minimum MSE and minimum CV for the optimal solution at 10 number of neurons, at time delays of six and feedback delays of six for both training and testing period.

Optimization

Secondly, the optimal solution was determined using the optimization algorithm GA as described earlier. To get the optimization result, the Genetic algorithm tool was run through MATLAB after defining the objective function as: the mean-square error MSE, root mean-square error RMSE, or coefficient of variance CV. Again, the problem variables include: Input time delays; feedback time delays and number of neurons. Table 4 shows the parameter used for synthetic and actual data.

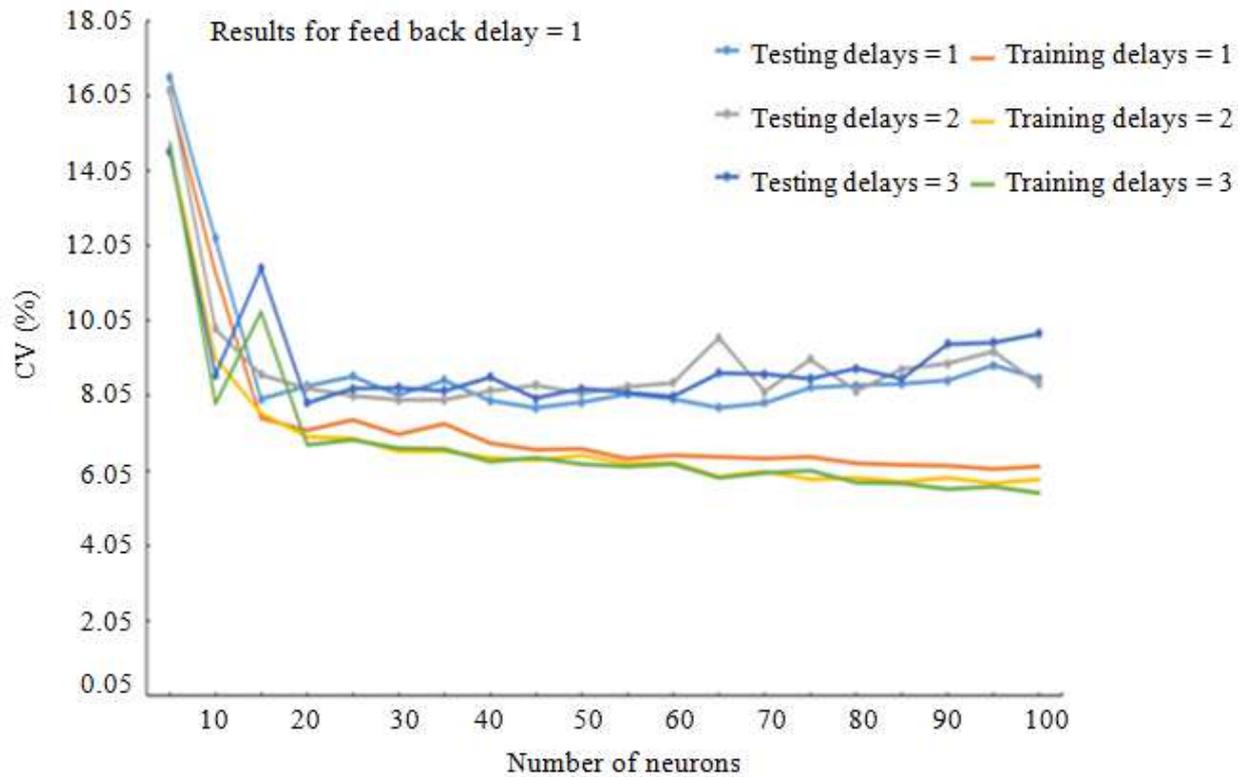


Fig. 6. Coefficient of Variance (CV) for different number of neurons and different input delays for testing and training data (feedback delays = 1)

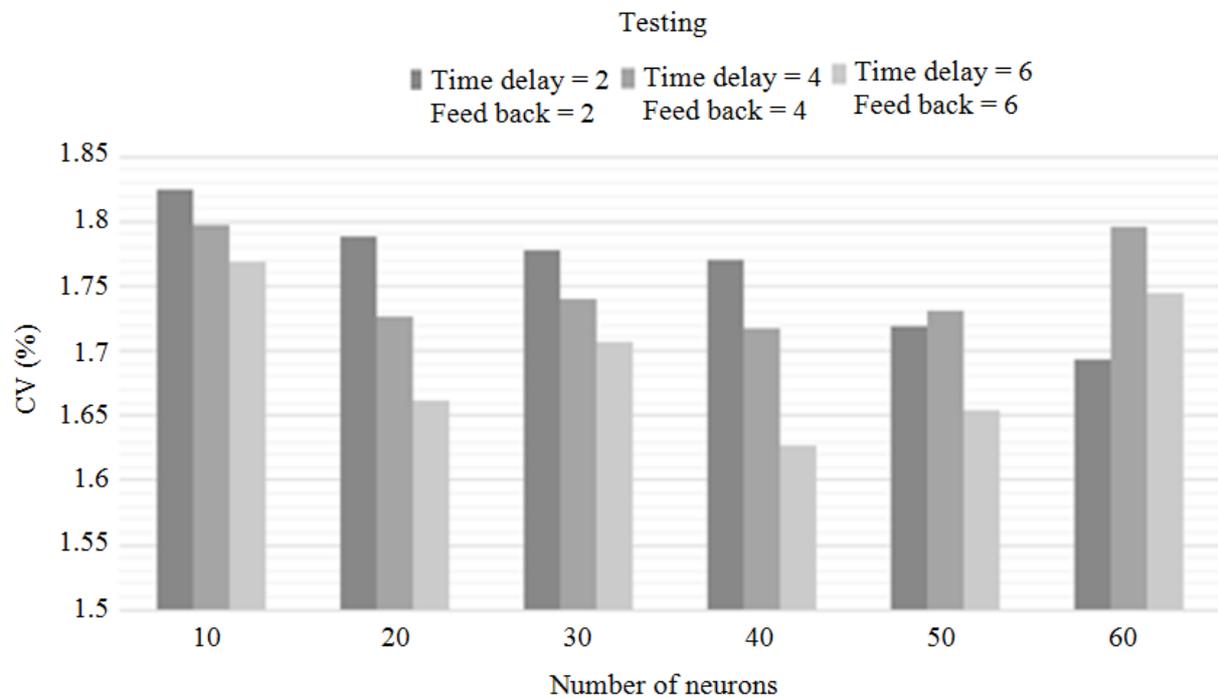


Fig. 7. Coefficient of variance for (Blufordlibrary) for the same input delay and feedback delays = 2, 4, 6 for testing period

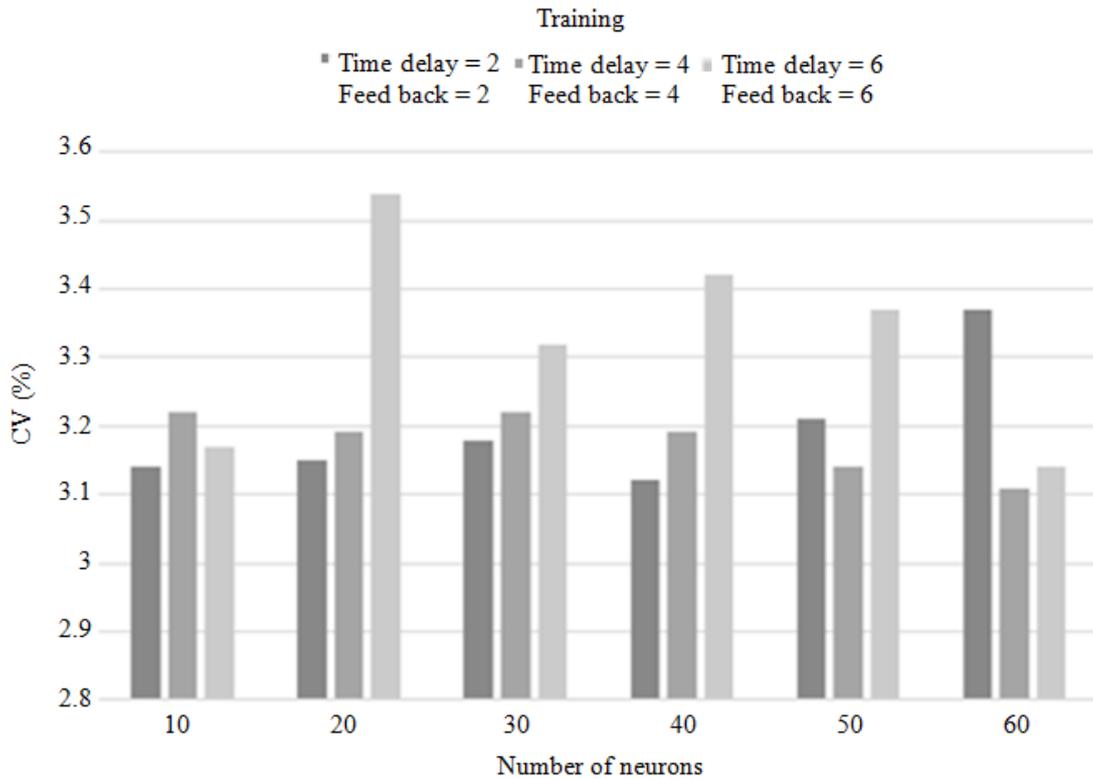


Fig. 8. The resulted coefficient of variance for (Bluford library) for the same input delay and feedback delays = 2, 4, 6 for training period

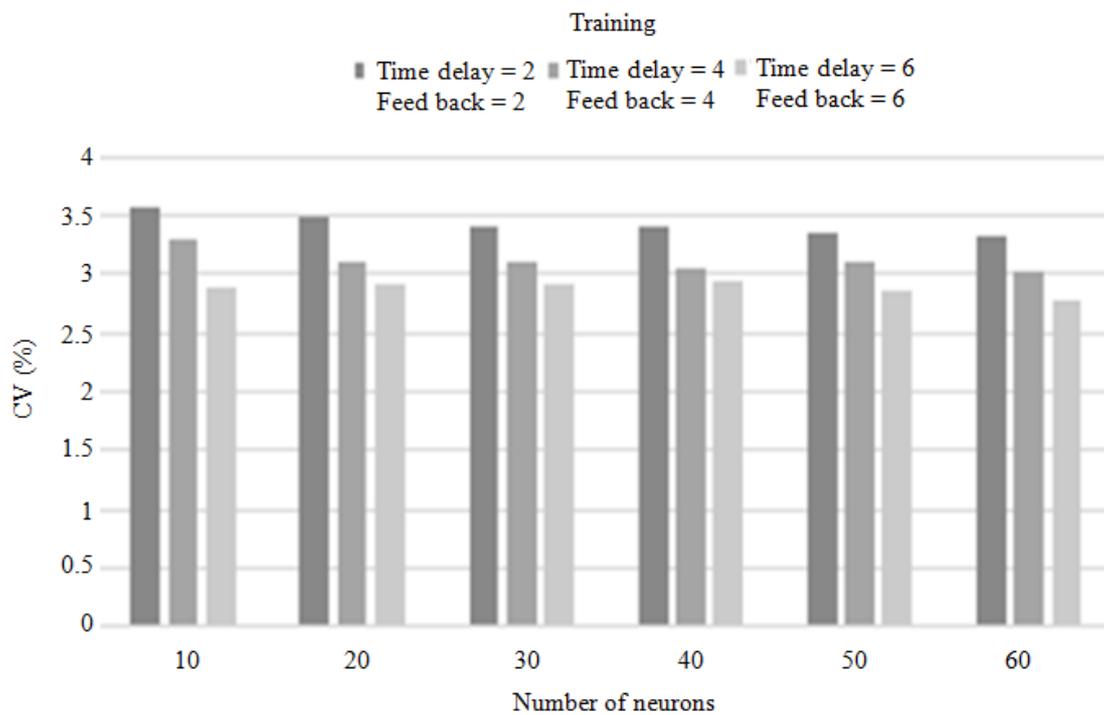


Fig. 9. Coefficient of variance for (Corbett Building) for the same input delay and feedback delays = 2, 4, 6 for testing period

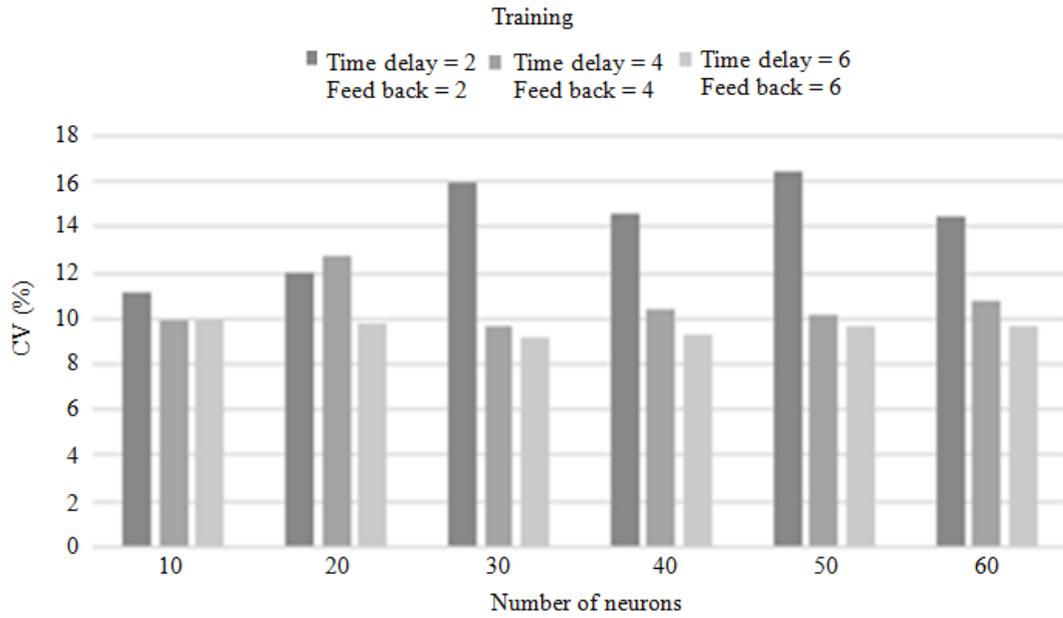


Fig. 10. The resulted coefficient of variance for (Corbett Building) for the same input delay and feedback delays = 2, 4, 6 for training period

Table 1. Result of the optimal 45 neurons at input delay of one and feedback of one

Number of neurons at ID = 1 FD = 1	Training		Testing	
	MSE	CV	MSE	CV
45	12.65	6.62	13.38	7.72

Table 2. Result of the optimal 40 neurons at input delay of six and feedback of six

Number of neurons at ID = 6 FD = 6	Training		Testing	
	MSE	CV	MSE	CV
40	12.2	3.42	27.7	1.62

Table 3. Result of the optimal 60 neurons at input delay of six and feedback of six

Number of neurons at ID = 6 FD = 6	Training		Testing	
	MSE	CV	MSE	CV
60	1423200	9.67	817500	2.79

Table 4. Constraint parameters used for synthetic and actual data to run the optimization

Objective function problem variables	Mean-Square Error (MSE) Root Mean-Square Error (RMSE) Coefficient of Variance (CV)		
	Number of neurons	Input time delays	Feedback time delays
Synthetic data	5-100	1-3	1-3
Actual data	10-60	2-6	2-6

Table 5. Optimal solutions obtained by the optimization

	Optimal solution		
	Number of neurons	Input time delays	Feedback time delays
Synthetic data	45	1	1
Actual data/building 1	40	6	6
Actual data/building 2	60	6	6

Synthetic Data

The optimization was run twice for different parameters based on the data; first it runs for synthetic data for the three variables (number neurons, input and feedback time delays). As illustrated in able 5, the optimization run for number of neurons between 5 to 100, input delay ranges between 1 and 3, feedback delay equal to the input delay. After obtaining the solution and examining the proper operation of the optimization algorithm GA, the optimal solution is compared with the one obtained from the parametric studies above. The optimal solution by GA matches exactly the one obtained by the parametric studies, which is the number of neurons of 45 and the feedback and input time delays of one.

Actual Data

Second, the optimization of the actual data was carried out once for each building considering three variables (number neurons and input and feedback time delays). As shown in Table 5, the optimization runs for number of neurons between 10 and 60, input delay ranges between 2 and 6, feed back delay equal to the input delay. Since the actual data was recorded in 15 min increments, the input time was selected to range from 2-6 to match the input time delays used for the synthetic data of 1-3.

After obtaining the solution and examining the proper operation of the optimization algorithm GA, the optimal solution is compared with the one obtained from the parametric studies above. The optimal solution by GA matches exactly the one obtained by the parametric studies, which is the number of neurons of 40 and the feedback and input time delays of six for the first building and the number of neurons of 60 and the feedback and input time delays of six for the second building. Table 5 illustrates the optimal solution after running the optimization for three times: First for the synthetic data, then for the actual data for each building separately.

In this study, the energy consumption network models were carried out for various configurations. Each network configuration contained at least two to three training algorithms that accurately predicted the data. Unfortunately, there were also many cases where algorithms either contained anomalous data, or failed to model the network to acceptable level.

The best architecture for modeling the energy consumption can be determined by comparing the optimal solution (i.e., algorithm and neurons that produces best performance) of each network structure. The best performing algorithms in each network configuration contained 5 or less hidden neurons, validating the assumption that the networks were originally over-fitted.

Conclusion

This paper presented modeling methodologies for building energy systems using non-linear auto-regression artificial neural networks. The model predicts whole building energy consumptions as a function of four input variables: Dry-bulb and wet-bulb outdoor air temperatures, hours of day and type of days. Data from simulations and actual buildings are used to model both the training and testing cases.

- The study was constructed using data from simulations as well as actual data collected from two buildings located at North Carolina A&T State University. The computational software MATLAB was utilized to develop the ANN models, using the integrated Building Dashboard system
- Different neural network structures were tested along with various input delays to determine the one yielding the best results in term of CV. When a neural network is trained with a given set of data, it builds a predictive model based on the data. This model reflects a minimization in error when the network's out prediction is compared with the known outcome. Each learning algorithm has its own set of error-correction rules for reducing this error
- The type of day variable is either "zero" or "one" corresponding to weekdays or weekend and holidays. The hour of day varies from 1 to 24 h. A large set of various neuron networks structures with various number of neurons and time delays are tested in order to determine the best and simple structures yielding adequate accuracy in term of Mean-Square Error (MSE), Root Mean-Square Error (RMSE), or Coefficient of Variances (CV) of RMSE
- To determine optimal model structure (Model parameters) with minimum error between the actual and simulated data during the test period, the Genetic Algorithm (GA) is used. The problem variables (or ANN model parameters) are (1) input time delays, (2) feedback time delays and (3) number of neurons (hidden layer size). The objective function could be the Mean-Square Error (MSE), Root Mean-Square Error (RMSE), or Coefficient of Variance (CV). The constraints cover the lower and upper limits of design variables, such as the maximum and minimum size of number of neurons and time delays
- The optimization method to improve the selection process of the best model parameters are developed and tested. Further, the optimal solution was compared with the one obtained from the parametric studies. The model with neurons of between 40 and 60 and with a time

delay of one hour holds the best results. The testing results show that the coefficient of variance is 7.7% for the simulated building, 1.7% for Building #1, 2.8% for Building#2

- The optimal solution by GA matches exactly the one obtained by the parametric studies, in which the number of neurons was 45 for the simulation and 40 and 60 for the actual data. The feedback and input time delays was one for the simulation and six for the actual data
- The model can provide accurate prediction of the building energy consumptions and that can be used for many applications such as energy building assessment, fault detection and diagnosis, energy saving estimation and saving measurement and verifications
- This study has shown that ANN's can be used as powerful tools in modeling the performance of building energy consumption. From the results analyzed in this work, neural networks are capable of modeling the systems to superior levels of prediction accuracy. They are scalable, easy to configure and highly adaptable to the information they are given. Researchers and designers should consider the viability of introducing ANNs to model the performance of building energy consumption
- While traditional modeling techniques require continuous updating as these systems grow in complexity, the flexible nature of neural networks makes them a very attractive alternative for these applications

Acknowledgement

The authors gratefully acknowledge the support of the U.S. Department of Commerce, National Institute of Standards and Technology for funding this project.

Funding Information

This work was performed under the following financial assistant award 70NANB14D1319 from the U.S. Department of Commerce, National Institute of Standards and Technology.

Author's Contributions

Maya Arida, Nabil Nassif and Rand Talib: Contributed in the design concept, data-analysis and contributed to the writing of the manuscript.

Taher Abu-Lebdeh: Assisted in reviewing the data analysis and in reviewing and improving the manuscript.

Ethics

The authors declare that there are no ethical issues that could arise after the publication of this study.

References

- ASHRAE, 2011. ASHRAE Handbook: Applications. 1st Edn., American Society of Heating Refrigeration and Air Conditioning Engineers, Atlanta.
- eQuest. Quickenergy simulation tool. eQuest Version 3.65. <http://www.doe2.com/equest/>
- EIA. Energy information administration. USA. www.eia.gov
- Nassif, N., 2013. Single and multivariate regression models for estimating monthly energy consumptions in schools in hot and humid climates. *Energy Eng.*, 110: 33-54.
- Nassif, N., S. Moujaes and M. Zaheeruddin, 2008. Self-tuning dynamic models of HVAC system components. *Energy Build.*, 40: 1709-1720.
- Nassif, N., 2014. Modeling and optimization of HVAC systems using artificial neural network and genetic algorithm.
- Seem, J.E., 2007. Using intelligent data analysis to detect abnormal energy consumption in buildings. *Energy Build.*, 39: 52-58.
- Wang, S.W., Q. Zhou and F. Xiao. 2010. A system-level fault detection and diagnosis strategy for HVAC systems involving sensor faults. *Energy Build.*, 42: 477-490.