Application of Neural Networks and multiple regression models in greenhouse climate estimation

Morteza Taki[1*], Yahya Ajabshirchi[1], Seyed Faramarz Ranjbar[2], Mansour Matloobi[3]

(1. Department of Biosystems Engineering, Faculty of Agriculture, University of Tabriz.

2. Department of Mechanical Engineering, Mechanical Engineering Faculty, University of Tabriz.

3. Department of Horticultural Science, Faculty of Agriculture, University of Tabriz, Tabriz, Iran)

Abstract: Artificial Neural Networks (ANNs) are biologically inspired computer programs designed to simulate the way in which the human brain processes information. After a comprehensive literature survey on the application of ANNs in greenhouses, this work describes the results of using ANNs to predict the roof temperature, inside air humidity, soil temperature and inside soil humidity (Tri, RHia, Tis, RHis), in a semi-solar greenhouse according to use some inside and outside parameters in the institute of renewable energy in East Azerbaijan province, Iran. For this purpose, a semi-solar greenhouse was designed and constructed for the first time in Iran. The model database selected beside on the main and important factors influence the four above variables inside the greenhouse. Neural estimation models were constructed with (Vo, Tia, Toa, Ir, Tis, RHia, Tri) as the inputs and (Tri, RHis, Tis, RHia) as the outputs. Optimal parameters for the network were selected via a trial and error procedure on the available data. Results showed that MLP (Multilayer Perceptron) algorithm with 4-6-1(4 inputs in first layer, 6 neurons in hidden layer and an output) and 4-9-1(4 inputs in first layer, 9 neurons in hidden layer and an output) topologies can predict inside soil and air humidity and inside roof and soil temperature with a low error (RMSE=0.25 ℃, 0.30%, 1.06 ℃ and 0.25% for Tri, RHis, Tis and RHia), respectively. Also the results showed that regression model has a low error to predict Tri (RMSE=0.71 ℃) and high error to estimate Tis (2.71 ℃), respectively. In overall, the error for regression model to predict all 4 parameters (Tri, RHis, Tis, RHia) was about 2 times higher than MLP method. It is concluded that ANN represents a promising tool for predicting inside climate in a greenhouse and will be useful in automatic greenhouses. For practical application, however, the farmers should use metrological and experimental data for 12 months of the year to decrease the prediction error.

Keywords: Artificial Neural Networks, semi-solar greenhouse, multiple linear regression model, Iran

Citation: Taki, M., Y. Ajabshirchi, S. F. Ranjbar, and M. Matloobi. 2016. Application of Neural Networks and multiple regression models in greenhouse climate estimation. Agricultural Engineering International: CIGR Journal, 18(3):29-43.

1 Introduction

Artificial neural networks (ANNs) are one of the widely used forecasting models that have enjoyed fruitful applications in forecasting social, economic, engineering, foreign exchange, stock problems, etc. Several distinguishing features of artificial neural networks make them valuable and attractive for a forecasting task. First, as opposed to the traditional model-based methods, artificial neural networks are data-driven self-adaptive methods in that there are few a priori assumptions about the models for problems under study. Second, artificial neural networks can generalize. After learning the data presented to them (a sample), ANNs can often correctly infer the unseen part of a population even if the sample data contain noisy information. Third, ANNs are universal functional approximators. It has been shown that a network can approximate any continuous function to any desired accuracy. Finally, artificial neural networks can solve linear and nonlinear functions (Khashei and Bijari, 2010). Because neural networks can be used to model nonlinear systems, it has been applied to greenhouse environment modeling (Ferreira et al., 2002; Seginer, 1997; Caponetto et al., 2000; Morimoto and Hashimoto, 2000; Wang et al., 2009).

The ANNs models are powerful prediction tools for the relation between the external climatic data and those inside the greenhouse parameters (Dreyfus et al., 2004;

Received date: 2016-01-18 **Accepted date:** 2016-07-28

***Corresponding author: Morteza Taki,** Department of Biosystems Engineering, Faculty of Agriculture, University of Tabriz. Iran. E-mail: mortezataaki@gmail.com.Tel: (+98) 9139204415

Haykin, 1994; Dariouchy et al., 2009). The results will support decision making of the farmer to avoid heat overload during the summer season (most frequent in this area) or heat shortage during the hibernal season. Besides, they assist the farmer planners to undertake the necessary measures to face the bad predictions (Hornik et al., 1989; Acosta and Tosini, 2001). Several studies were made on the greenhouses, for example, to control the greenhouse climate (Bakker et al., 1994; Linker et al., 1998; Peŕez-Alonso et al, 2012), physical modeling of greenhouse climate (Bot, 1991),estimate and prediction of greenhouse climatic parameters (Fernandez and Bailey, 1992; Jolliet, 1994; Boaventura et al., 1997 and 2000; Coelho et al., 2002).

Segineret al, 1994, presented neural network models to control greenhouse climate in Israel. The ANNs model showed that this is a useful method for following tasks: as a model for optimal environmental control, as a screening tool in preparation for developing physical models and this model does not require explicit evaluation of transfer coefficients and need no model formulation. The main disadvantage is that they cannot be used for design purposes.

Seginer (1997) reviewed some artificial neural networks applications for greenhouse environmental control. He concluded that the ANNs greenhouse modeling only refers to existing structures. These models cannot be used to design new greenhouses, since they lack explicit expressions for the various components and transfer coefficients. Changes in equipment will also require model modification. However, one could contemplate a situation where a manufacturer of greenhouses dose not only supply a turn-key facility but also its NN model. This model could later be fine-tuned to local conditions and requirements, based on data collected on location. NN models can also be useful as controllers, since they may be taught various control rules. Two examples are the mimicking of a model-based optimal (feed-forward) controller and a human optimizer (expert grower), who uses some feedback information from the state of the crop.

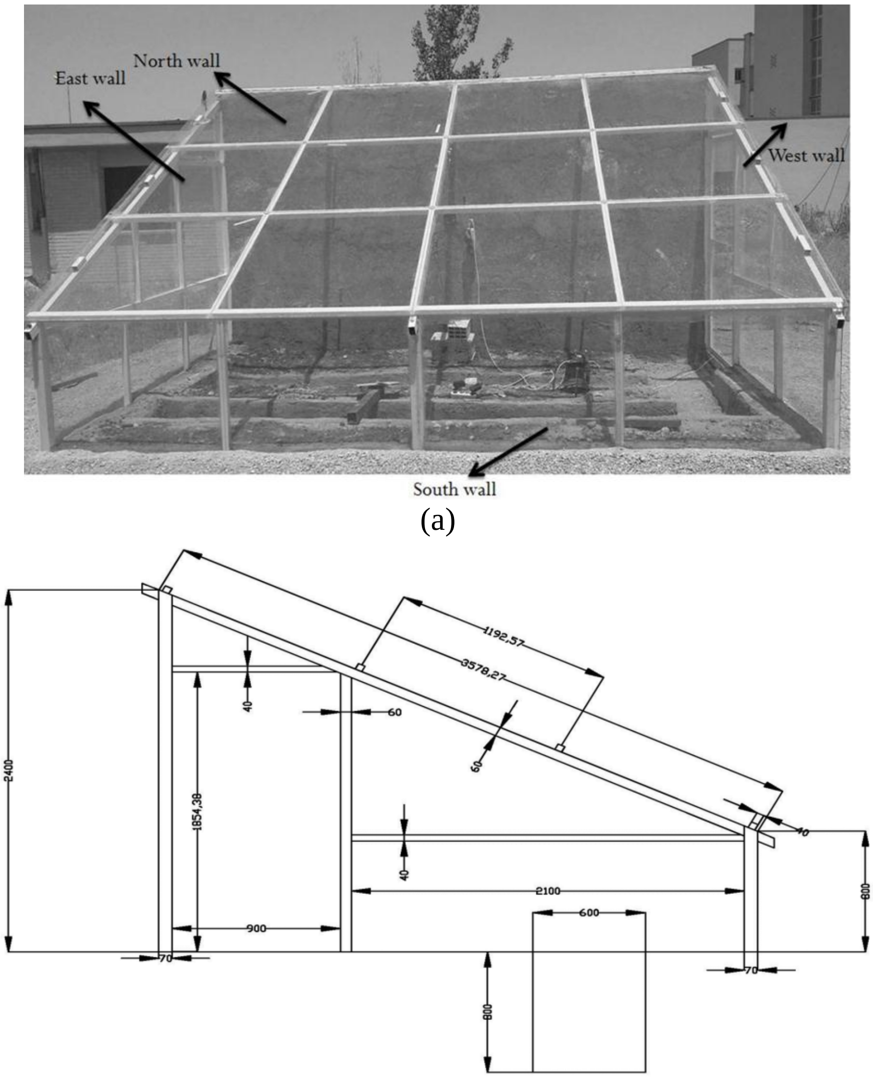
Linker and Seginer, 2004, made a comparison between the performance of three types of models trained with several seasonal sub-sets of data: (1) black-box (BB) sigmoid neural network (NN) trained only with in situ data, (2) hybrid physical-RBF (radial basis function) model, and (3) sigmoid neural network trained with a combination of in situ data and synthetic data generated with a physical model (termed ‘prior-K sigmoid model’) to predict the greenhouse air temperature in Israel. Results showed that The BB sigmoid model gives the best predictions within the training domain, but performs very badly outside it. On the other hand, the hybrid and prior-K sigmoid models produce useful predictions over the whole operating domain, although they are slightly less accurate within the training domain.

At this moment, we are not aware of any study about the use of a solar greenhouse in Iran because of cheap fossil fuel and the need of high investments. So we decided to start a big project in Tabriz University (East Azarbaijan Province of Iran) on solar greenhouse modeling, design and application in two Ph. D studies. This paper is a part of this project. The main objective of this research is to develop a model to predict the roof temperature, inside air humidity, soil temperature and humidity in a semi-solar greenhouse according to use some inside and outside parameters and then, select the best and simple one using artificial neural network. For this purpose, data were recorded from a semi-solar greenhouse located in Tabriz University, Department of Renewable Energy.

2 Materials and methods

2.1 Semi-solar greenhouse

In this study, for the first time, a semi-solar greenhouse was designed and constructed at the North-West of Iran in Azerbaijan Province (geographical location of 38°10′ N and 46°18′ E with elevation of 1364 m above the sea level). In the solar greenhouse design, the heat insulation and the transmission of solar radiation are maximized. A warm and a cold water aquifer layer are used to store and retrieve the surplus solar energy. At times of heat demand, the greenhouse can be heated with little energy input with a heat pump and warm aquifer water. At times of heat surplus, the greenhouse can be cooled with a heat exchanger and cold aquifer water, while energy is harvested for use at times of heat demand (Van Ooteghem, 2007). The solar greenhouse has some differences compared to a conventional greenhouse such as (Van Straten, 2011): improved insulation value and improved light transmission cover, ventilation with heat recovery, aquifer (an aquifer is a formation of water-bearing sand material in the soil that can contain and transmit water. Wells can be drilled into the aquifers and water can be pumped into and out of the water layers), heat extraction, heat pump, boiler, carbon dioxide supply and gas motor or electric drive. In this research we started a new project in Department of Biosystems Engineering, Tabriz University. Shape and orientation of this greenhouse, was selected between some common greenhouse shapes and according to receive maximum solar radiation during the whole year. For this, meteorological data recorded by Iran Meteorological Office for the period of 1992–2013, were used and after some analysis, this structure was selected. Also internal thermal screen and cement north wall was used to store and prevent heat loss during the cold period of the year. So we called this structure, 'semi-solar' greenhouse. It is covered with glass (4 mm thickness). It occupies a surface of approximately 15.36 m^2 and a volume of 26.4 m^3. The orientation of this greenhouse is East–West and perpendicular to the direction of the prevailing wind (Figure1).



(a)

(b)

Figure 1 (A) Front view and (B) Design picture of semi-solar greenhouse in Tabriz University, Iran.

2.2 Record the inside and outside inputs

To measure the temperature and the relative humidity of the air, soil and roof inside and outside the greenhouse, the SHT 11 sensors were used. The SHT11 is a single chip relative humidity and temperature multi sensor module comprising a calibrated digital output. Application of industrial CMOS processes with patented micro-machining (CMOSens® technology) ensures a high reliability and long term stability. The device includes a capacitive polymer sensing element for relative humidity and a bandgap temperature sensor. Both are seamlessly coupled to a 14bit analog to digital converter and a serial interface circuit on the same chip. This results in superior signal quality, a fast response time and insensitivity to external disturbances (EMC) at a very competitive price. The accuracy of the measurement of temperature is ±0.4% at 20°C and the precision measurement of the moisture is ±3% for a clear sky (company information). Figure 2 shows the diagram of this sensor. At 1 m height above the ground outside the greenhouse, we used a pyranometre type TES 1333. Its sensitivity is proportional to the cosine of the incidence angle of the radiation. It is a measure of global radiation of the spectral band solar in the 400–1110 nm range. Its measurement accuracy is approximately ±5% (company information).

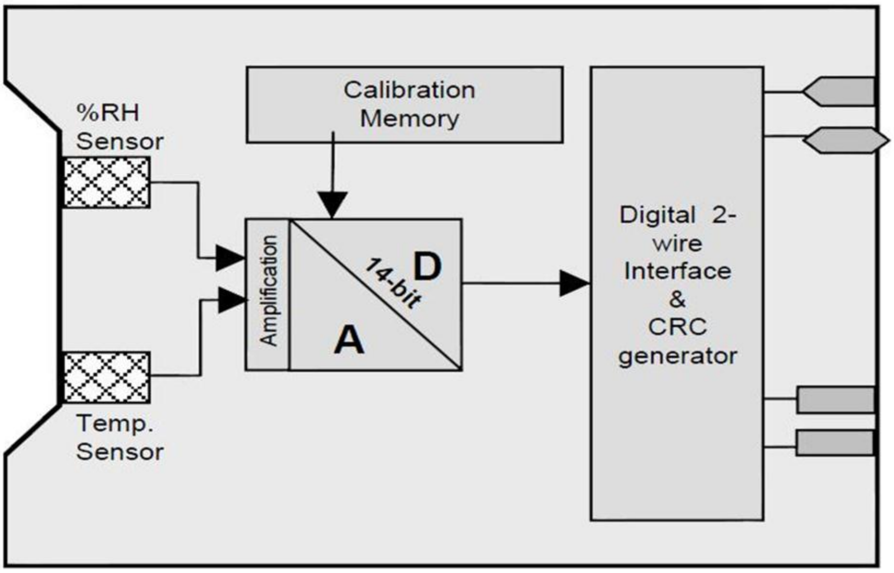


Figure 2 SHT 11 sensor diagram

2.3 Artificial Neural Network

Prior to any ANNs training process with the trend free data, the data must be normalized over the range of [0, 1]. This is necessary for the neurons' transfer functions, because a sigmoid function is calculated and consequently these can only be performed over a limited range of values. If the data used with an ANNs are not scaled to an appropriate range, the network will not converge on training or it will not produce meaningful results. The most commonly employed method of normalization involves mapping the data linearly over a specified range, whereby each value of a variable x is transformed as follows (Equation 1):

$$x_n = \frac{x - x_{min}}{x_{max} - x_{min}} ?\ r_{max} - r_{min}) + r_{min} \tag{1}$$

where x is the original data, x_n the normalized input or output values, x_{max} and x_{min}, are the maximum and minimum values of the concerned variable, respectively. r_{max} and r_{min} correspond to the desired values of the transformed variable range. A range of 0.1–0.9 is appropriate for the transformation of the variable onto the sensitive range of the sigmoid transfer function (Taki et al, 2016b). Among various ANNs models, Multilayer Perceptron (MLP) has high practical importance. MLP is a feed-forward layered network with one input layer, one output layer, and some hidden layers.

Every node computes a weighted sum of its inputs and passes the sum through a soft nonlinearity. The soft nonlinearity or activity function of neurons should be non-decreasing and differentiable. The most popular function is unipolar sigmoid (Equation 2) (Taki et al., 2012):

$$f(\theta)=\frac{1}{1+e^{-\theta}} \tag{2}$$

The network is in charge of vector mapping, i.e. by inserting the input vector, x^q the network will answer through the vector z^q in its output (for q=1,2,..., Q). The aim is to adapt the parameters of the network in order to bring the actual output z^q close to corresponding desired output d^q (for q=1,2,..., Q). The most popular method of MLP training is the Back-Propagation (BP) algorithm, and in literatures there exist many variants of this algorithm. This algorithm is based on minimization of a suitable error cost function (Taki et al, 2016b). In this study, Basic Back-propagation (BB) algorithm was employed.

MLPs are normally trained with Back Propagation (BP) algorithm. It is a general method for iteratively solving for weights and biases. The knowledge obtained during the training phase is not stored as equations or in a knowledge base but is distributed throughout the network in the form of connection weights between neurons. BP uses a Gradient Descent (GD) technique that is very stable when a small learning rate is used but has slow convergence properties. Several methods for speeding up BPs have been used, including adding a momentum term or using a variable learning rate. GD with a momentum (GDM) algorithm that is an improvement to the straight GD rule in the sense that a momentum term is used to avoid local minima, speeding up learning and stabilizing convergence, is used (Taki et al., 2012). Multiple layers of neurons with non-linear transfer functions allow the network to learn nonlinear and linear relationships between input and output parameters. Several MLP network architectures with one, two, three and four hidden layers have been trained and evaluated aiming at finding the one that could result in the best overall performance. In this work, the learning rules of Gradient Descent Momentum (GDM) and Levenberg-Marquardt (LM) were considered. No transfer function for the first layer was used. For the hidden layers the sigmoid functions were used, and for the output layer a linear transfer function was applied as desired for estimating problems. A computer code was also developed in MATLAB software for the feed forward and back propagation network.

We used an N-fold cross validation method that in this method data are randomly divided into two sets; training set (70% of all data) and cross validation set (the remaining 30% of all data) (Taki et al., 2012). The neural network model is formed for output (roof temperature, inside air humidity, soil temperature and humidity) by using four models of MLP according to (Figure 3 shows the place of these sensors):

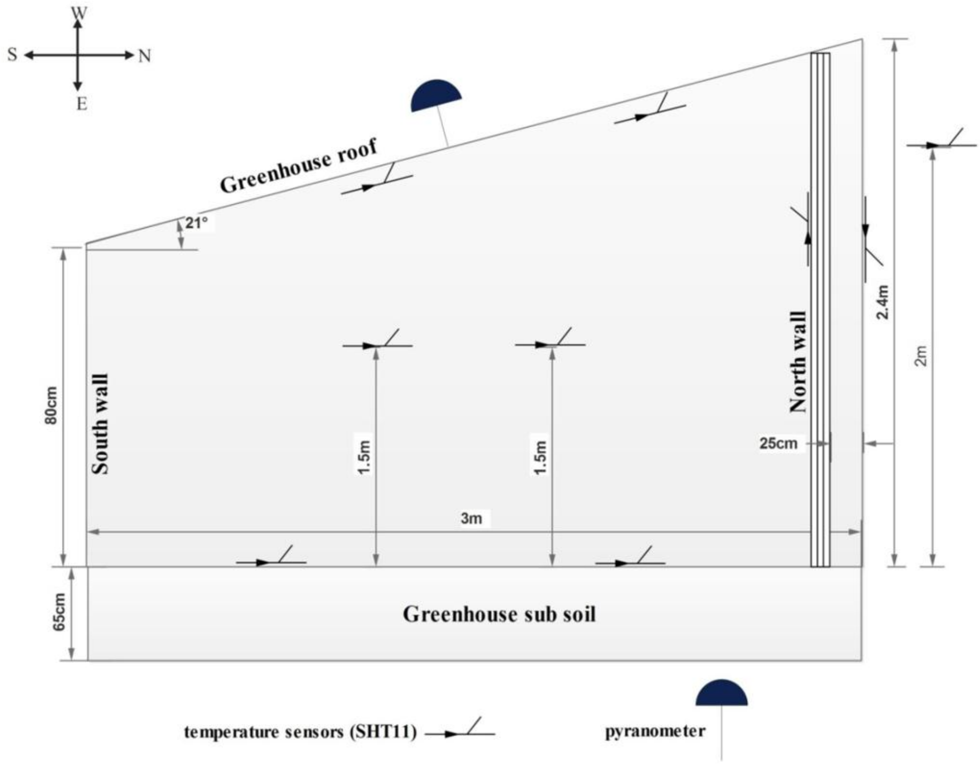


Figure 3Locations of SHT11 sensors to collect inside and outside T/RH values

I. Four inputs for predicting roof temperature (inside air temperature (T_{ia}), solar radiation on the roof (I_r), wind speed (V_o), outside air temperature (T_{oa}).

II. Four inputs for predicting soil humidity (inside soil temperature (T_{is}), inside air humidity (RH_{ia}), solar radiation on the roof, inside air temperature.

III. Four inputs for predicting soil temperature (inside air temperature, solar radiation on the roof, inside roof temperature (T_{ri}) and inside air humidity).

IV. Four inputs for predicting inside air humidity (inside air temperature, inside roof temperature, outside air temperature, solar radiation on the roof).

2.4 Regression model

For the regression model, we used all the inputs that were used in the ANN model (inputs and outputs mentioned in the former paragraph). A stepwise multiple regression method was applied to choose the pertinent independent variables influencing the dependent variable. Furthermore, in order to verify the validity of multiple regression models, a chi-square test was carried out using the predicted and experimental data. Minitab 17 software was used for data analysis.

To evaluate the performance of a model some criteria have been used. These criteria include: Root Mean Squared Error (RMSE), coefficient of determination (R^2) and Model Efficiency (EF). They are defined as following Equation 3, Equation 4 and Equation 5:

$$RMSE = \sqrt{\frac{\sum_{j=1}^{m}(d_j - p_j)^2}{m}} \tag{3}$$

$$R^2 = \frac{\left[\sum_{j=1}^{n}(d_j - \bar{d})(p_j - \bar{p})^2\right]^2}{\sum_{j=1}^{n}(d_j - \bar{d})^2 \times \sum_{j=1}^{n}(p_j - \bar{p})^2} \tag{4}$$

$$EF = \frac{\sum_{j=1}^{n}(d_j - \bar{d})^2 - \sum_{j=1}^{n}(p_j - d_j)^2}{\sum_{j=1}^{n}(d_j - \bar{d})^2} \tag{5}$$

Where d_j is the ith component of the desired (actual) output for the jth pattern; $p_j$is the component of the predicted (fitted) output produced by the network for the jth pattern; $\overline{d}$ and $\overline{p}$ are the average of the whole desired (actual) and predicted output and n is the number of variable outputs. A model with the small RMSE and the high EF and R^2 is considered to be the best (Taki et al, 2016a; Willmott et al., 1985).

3 Results and discussion

3.1ANNs model

The proper physiological functioning of the plant requires a controlled climate especially in the summer when the increase of temperature can harm the growth or the plants (Fatnassi et al., 2002). The prediction of the internal temperature of the greenhouse can bring a substantial help for controlling the distribution of the shelters. Considering the pseudo-periodicity of the climatic variations in a semi-arid area, it was sufficient to work on 1 set data (24 hours) to learn our artificial network model. The compiled database represents 1 day sets of parameters values inside the greenhouse. The training phase of the ANN model was terminated when the error on the testing data bases were minimal. The training process goal is to reach an optimal solution based on some performance measurements such as root mean squared error (RMSE), R square (R^2) and mean absolute percentage error (MAPE). Therefore, the required ANN model was developed in two phases: training (calibration) phase and testing (generalization or validation) phase. In the training phase, a large part of the database (70%) was used to train the network and the remaining part of the database was used for testing.

Based on universal approximation theorem, a neural network with a single hidden layer and with a sufficient large number of neurons can well approximate any arbitrary continuous function (Haykin, 1994). Therefore, the ANNs designed in this study are equipped with a single hidden layer. Determination of the number of neurons in the hidden layer is rather an art than science, because it may vary depending on the specific problem under study. In this study, the optimal number of neurons in the hidden layer was selected using a trial-and-error method. The process was repeated several times. It is observed that the performance of BB-MLP may improve as the number of hidden neurons increased. However, too many neurons in the hidden layer may cause over-fitting problems, which results in good network learning and data memorization, but lack of ability to generalize. On the other hand, if the number of neurons in the hidden layer is too low, the network may not be able to learn.

During the training step, the network used the training data set. Training was continued until a steady state was reached. The BB algorithm was utilized for model training. Some statistical properties of the sample data used for training process and the prediction values associated with different training algorithms are shown in Table 1. Considering the average values of standard deviation and variance, it can be deduced that the values and the distribution of real and predicted data are analogous. But, the differences of minimum and maximum values are remarkable. This is probably due to the fact that the extreme values were not well represented in the training data set, because these were only one or two points.

Table1 Statistical variables of desired and predicted values in training phase (MLP).

ANN Models with structure		Average, % and °C	Variance $(\%)^2$ and $(°C)^2$	Standard Deviation, % and °C	Skewness (-)	Kurtosis (-)
$T_{ri}=f(V_o,T_{ia},T_{oa},I_r)$	Desired	55.462	116.620	10.799	-1.148	0.017
	Predicted	55.481	116.907	10.812	-1.150	0.018
$RH_{is}=f(T_{is},I_r,T_{ia},RH_{ia})$	Desired	10.159	29.342	5.416	1.014	-0.321
	Predicted	10.216	29.682	5.448	1.019	-0.326
$T_{is}=f(T_{ia},I_r,T_{ri},RH_{ia})$	Desired	72.056	201.994	14.212	-0.980	-0.404
	Predicted	72.010	204.192	14.289	-1.008	-0.337
$RH_{ia}=f(I_r,T_{ri},T_{oa},T_{ia})$	Desired	17.757	20.395	4.516	0.537	-1.176
	Predicted	17.744	20.120	4.485	0.533	-1.199

Testing sets are usually used to select the best performing network model. There is no systematic method to find a suitable structure of a neural network. The test is done in a heuristic way according to the following steps. Initially, the networks with only one hidden layer were built by successively adding two additional neurons on this one. This technique which has of course the advantage of decreasing the number of tests is founded on the intuitive idea that the addition of two neurons instead of only one in general does not generate too large differences in performances between two architectures consecutive. Secondly, for the networks with two hidden layers, the triangular structures were considered, for which the number of neurons on a layer is higher than the following layer (Dariouchy et al, 2009). The results of training and testing were shown in Figure 4.

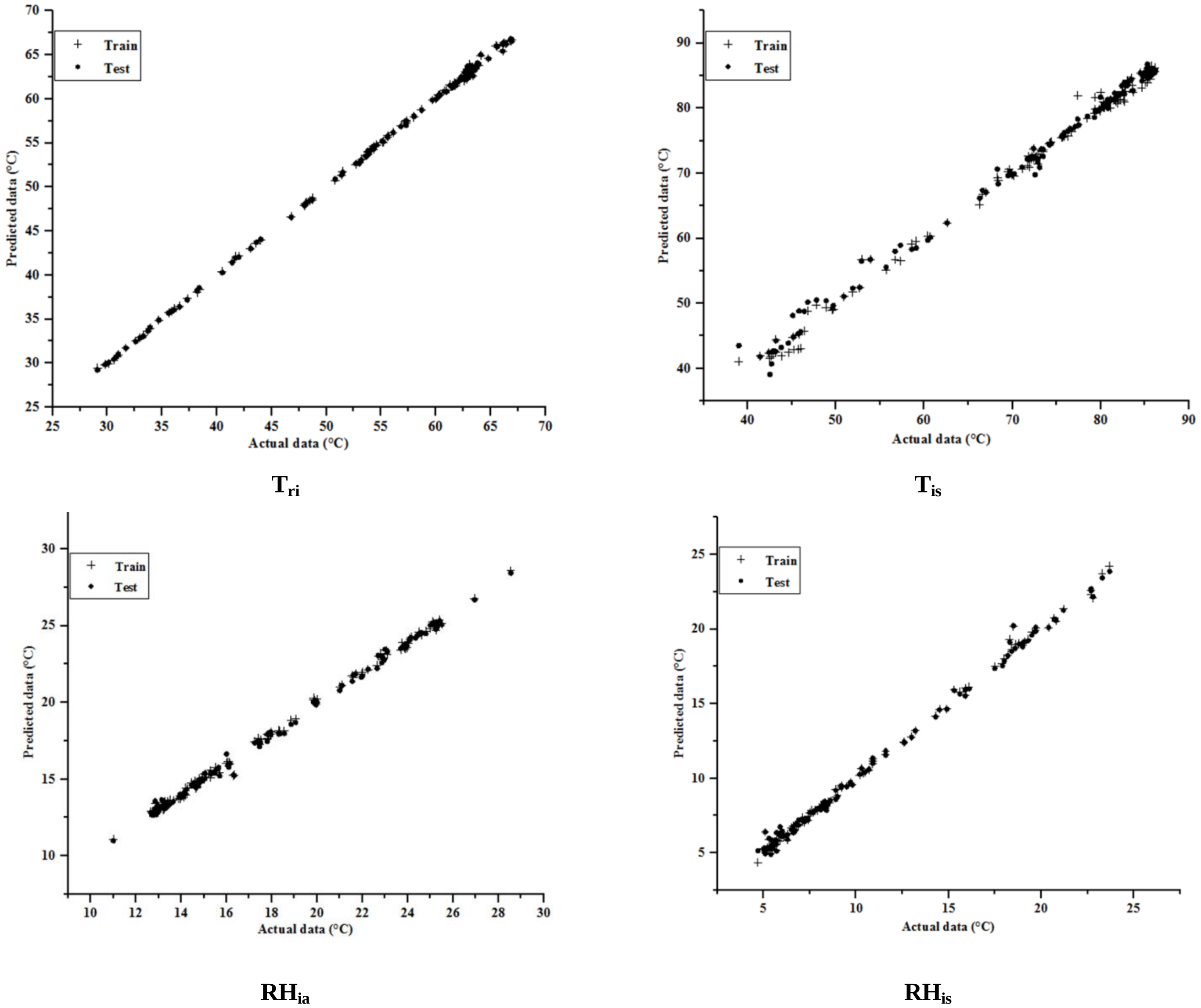


Figure 4.Comparison between the desired and predicted values by the ANNs model for the roof temperature (T_{ri}), soil temperature (T_{is}), inside air humidity (RH_{ia}) and inside soil humidity (RH_{is}) during the test and training phase

In this section, we used the selected structure with the previously adjusted weights (training step, Table 1). The objective of this step was to test the network generalization property and to evaluate the competence of the trained network. Therefore, the network was evaluated by data, outside the training set. Table 2 shows some statistical properties of the data used in the test phase and the corresponding prediction values associated with different training algorithms.

Table2 Statistical variables of desired and predicted values (test phase)

ANN Models with structure		Average, % or °C	Variance	Standard Deviation	Skewness (-)	Kurtosis (-)
$T_{ri}=f(V_o,T_{ia},T_{oa},I_r)$	Desired	55.455	116.518	10.794	-1.149	0.019
	Predicted	55.473	116.952	10.814	-1.149	0.017
$RH_{is}=f(T_{is},I_r,T_{ia},RH_{ia})$	Desired	10.141	29.120	5.396	1.010	-0.320
	Predicted	10.216	29.249	5.408	1.023	-0.319
$T_{is}=f(T_{ia},I_r,T_{ri},RH_{ia})$	Desired	70.120	200.120	14.146	-0.954	-0.391
	Predicted	72.256	196.763	14.027	-0.981	-0.355
$RH_{ia}=f(I_r,T_{ri},T_{oa},T_{ia})$	Desired	17.893	21.126	4.596	0.581	-1.038
	Predicted	17.852	20.829	4.563	0.600	-1.000

Table 3 shows the effect of number of neurons in the hidden layer on the performance of BB-MLP model. Considering Table 1, a BB-MLP model with nine neurons in the hidden layer seems to be appropriate for (T_{ri}) and (T_{is}) and with six neurons for (RH_{is}) and (RH_{ia}). These topologies can be more versatile for future applications.

Table3 Performance variation of a one-layer MLP with different number of neurons in the hidden layer

Parameters		Number of neurons in the hidden layer						
	Criterion	4	5	6	7	8	9	10
$T_{ri}=f(V_o,T_{ia},T_{oa},I_r)$	R^2 (-)	0.9994	0.9993	0.9991	0.9992	0.9993	**0.9994**	0.9990
	RMSE (°C)	0.2579	0.2729	0.3142	0.3339	0.2678	**0.2526**	0.3143
$RH_{is}=f(T_{is},I_r,T_{ia},RH_{ia})$	R^2 (-)	0.9965	0.9962	**0.9968**	0.9959	0.9967	0.9955	0.9950
	RMSE (°C)	0.3216	0.3220	**0.3056**	0.3444	0.3252	0.3850	0.3891
$T_{is}=f(T_{ia},I_r,T_{ri},RH_{ia})$	R^2 (-)	0.9912	0.9909	0.9929	0.9931	0.9926	**0.9945**	0.9914
	RMSE (°C)	1.3498	1.3549	1.2102	1.1945	1.2400	**1.0679**	1.3385
$RH_{ia}=f(I_r,T_{ri},T_{oa},T_{ia})$	R^2 (-)	0.9913	0.9955	**0.9971**	0.9966	0.9950	0.9962	0.9915
	RMSE (°C)	0.4194	0.3028	**0.2502**	0.2610	0.3123	0.2826	0.4213

3.2 Statistical analysis

From statistical point of view, both desired and predicted test data have been analyzed to determine whether there are statistically significant differences between them. The null hypothesis assumes that statistical parameters of both series are equal. *p*value was used to check each hypothesis. Its threshold value was 0.05. If *p* value is greater than the threshold, the null hypothesis is then fulfilled. To check the differences between the data series, different tests were performed and *p* value was calculated for each case (Taki et al, 2016a). The results are shown in Table 4. The so called *t-test* was used to compare the means of both series. It was also assumed that the variance of both samples could be considered equal. The obtained *p* values were greater than the threshold; hence the null hypothesis cannot be rejected in all cases ($p > 0.99$). The variance was analyzed using the *F-test*. Here, a normal distribution of samples was assumed. Again, the *p* values confirm the null hypothesis in all cases ($p > 0.97$). Finally, the *Kolmogorov–Smirnov test* also confirmed the null hypothesis. From statistical point of view, again, the *p* values confirm the null hypothesis in all cases ($p > 0.97$).

Table 4Statistical comparisons of desired and predicted test data and the corresponding *p* values (The high *p* values show that there are no differences between actual and predicted values by ANNs)

Parameters	Analysis type		
	Comparisons of means	Comparisons of variances	Comparisons of distribution
$T_{ri}=f(V_o,T_{ia},T_{oa},I_r)$	0.9941	0.9890	0.9950
$RH_{is}=f(T_{is},I_r,T_{ia},RH_{ia})$	0.9920	0.9781	0.9760
$T_{is}=f(T_{ia},I_r,T_{ri},RH_{ia})$	0.9916	0.9920	0.9770
$RH_{ia}=f(I_r,T_{ri},T_{oa},T_{ia})$	0.9918	0.9701	0.9740

3.3 Sensitivity analysis

According to the obtained results in Table 5, the share of each input item of developed MLP model on desired outputs can be seen clearly. Sensitivity analysis provides insight into the usefulness of individual variables. With this kind of analysis it is possible to judge what variables are the most significant (with sensitivity value close to 1) and the least significant (with sensitivity value close to 0) during generation of the satisfactory MLP. It is evident that solar radiation on the roof (30%), inside soil temperature (29%), solar radiation on the soil (38%) and inside air temperature (35%) had the high sensitivity on (T_{ri}), (RH_{is}), (T_{is}) and (RH_{ia}), respectively.

Table 5 Sensitivity analysis of various inputs on outputs of ANNs models

	Inputs							
	T_{ia}	I_r	V_o	T_{oa}	RH_{ia}	I_s	T_{ri}	T_{is}
$T_{ri}=f(V_o,T_{ia},T_{oa},I_r)$	0.19	0.33	0.20	0.28	-	-	-	-
$RH_{is}=f(T_{is},I_r,T_{ia},RH_{ia})$	0.17				0.27	0.27		0.29
$T_{is}=f(T_{ia},I_r,T_{ri},RH_{ia})$	0.25				0.21	0.38	0.16	
$RH_{ia}=f(I_r,T_{ri},T_{oa},T_{ia})$	0.35	0.25	0.14	0.12			0.14	

Note: T_{ia}: Inside air temperature, I_r: Roof radiation, V_o: Outside wind speed, T_{oa}: Outside air temperature, RH_{ia}: Inside air humidity, I_s: Inside soil radiation, T_{ir}: Inside roof temperature, T_{is}: Inside soil temperature

3.4 Comparison between multiple regression and MLP models

Any relationship, linear or nonlinear, can be learned and approximated by an ANNs such as a three-layer MLP with sufficiently large number of neurons in the hidden layer. Another advantage of ANNs is its capability of modeling the data of multiple inputs and multiple outputs. In contrast, the conventional regression techniques can only be used to learn the relationship between a single output and one or more inputs but cannot be used to model the data of multiple inputs and multiple outputs. The results of a multiple linear regression analysis between (T_{ri}), (RH_{is}), (T_{is}) and (RH_{ia}) and series of independent variables (V_o, T_{ia}, T_{oa}, I_r, T_{is}, RH_{ia}, T_{ri}) are presented in Table 6. All main factors in these models had a significant effect at 5% probability level. Also Figure 5 shows the contribution of inputs on final output. Also see Table 6 please.

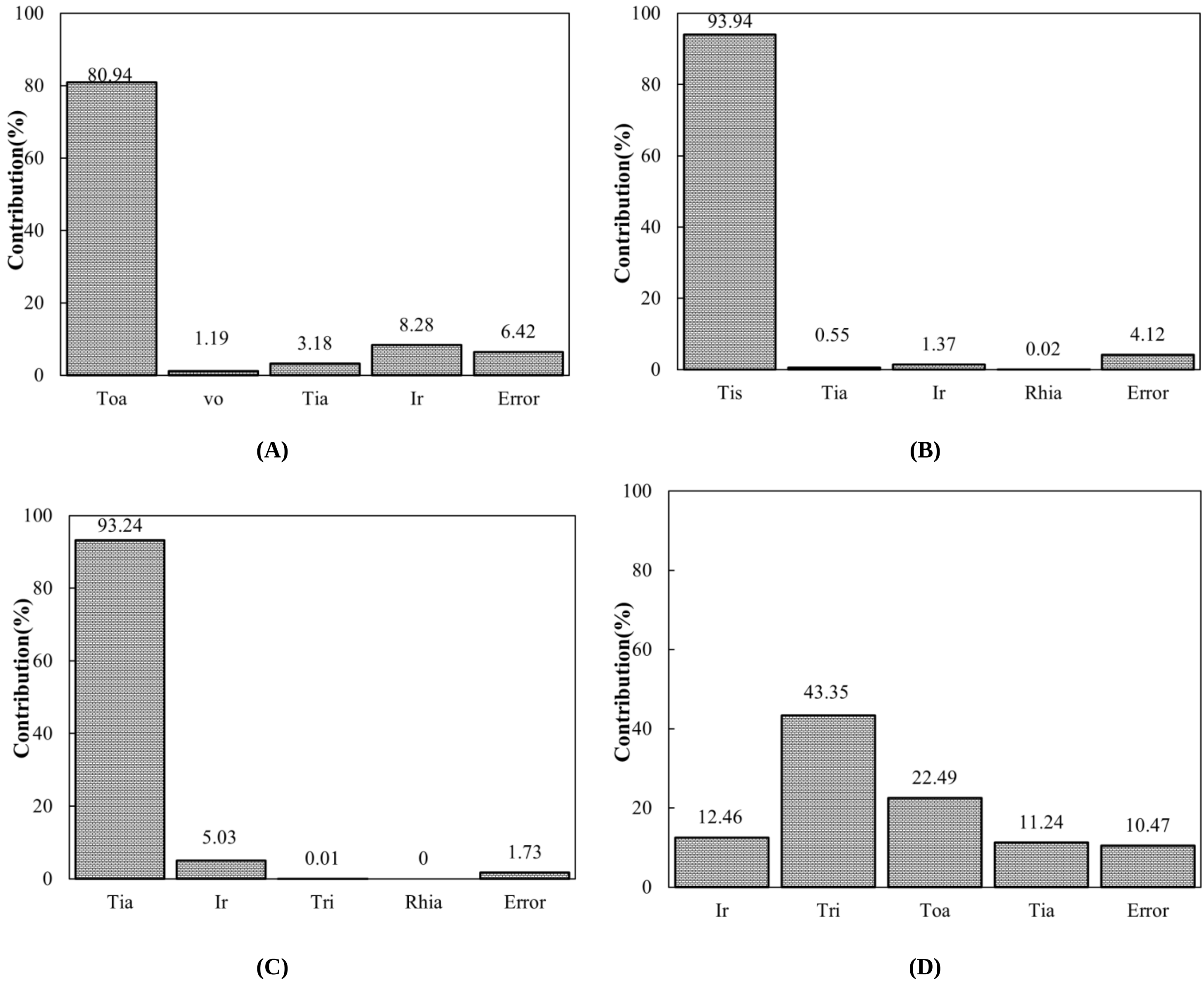


Figure 5 Contribution of each input on final output to (A): roof temperature, (B): inside soil humidity, (C): inside soil temperature, (D): inside air humidity in a semi-solar greenhouse (T_{ia}: Inside air temperature, I_r: Rood radiation, T_{ri}: Inside roof temperature, Rh_{ia}: Inside air humidity, T_{oa}: Outside air temperature, V_o: Outside wind speed, T_{is}: Inside soil temperature)

Table 6 Results of multiple regression analysis to predict (T_{ri}), (RH_{is}), (T_{is}) and (RH_{ia})

Model	Regression models*	R^2
$T_{ri}=f(V_o,T_{ia},T_{oa},I_r)$	T_{ri}=3.39+(0.505譚$_{oa}$)-(3.705譾$_{o}$)+(0.4236譚$_{ia}$)+(0.019028譏$_{r}$)	0.9358
$RH_{is}=f(T_{is},I_r,T_{ia},RH_{ia})$	RH_{is}=38.46-(0.6050譚$_{is}$)+(0.1532譚$_{ia}$)+(0.008472譏$_{r}$)+(0.0443識H_{ia})	0.9588
$T_{is}=f(T_{ia},I_r,T_{ri},RH_{ia})$	T_{is}=2.1-0.4817+(1.0819 $_{ia}$ $_{r}$ T_{ri})-(0.0032識H_{ia})	0.9827
$RH_{ia}=f(I_r,T_{ri},T_{oa},T_{ia})$	RH_{ia}=46.601+(0.008829譏$_{r}$)+(0.2341譚$_{ri}$)+(0.0089譚$_{oa}$)-(0.8105譚$_{ia}$)	0.8953

Note: *Minimum probability threshold $p \leq 0.05$

The plots of predicted (T_{ri}), (RH_{is}), (T_{is}) and (RH_{ia}) against measured values are depicted in Figure6. The results reveal a very good agreement between the predicted and the measured values ($R^2>0.9$). Also, these figures reveal that the (T_{ri}), (RH_{is}), (T_{is}) and (RH_{ia}) predictions from regression model were not as good as fit to measured data in comparison to MLP model prediction. Comparisons of measured versus predicted values for MLP model resulted in a least squares linear regression lines with slopes almost equal to regression model, while the MLP model resulted in lines with y-intercepts much lower than the regression model.

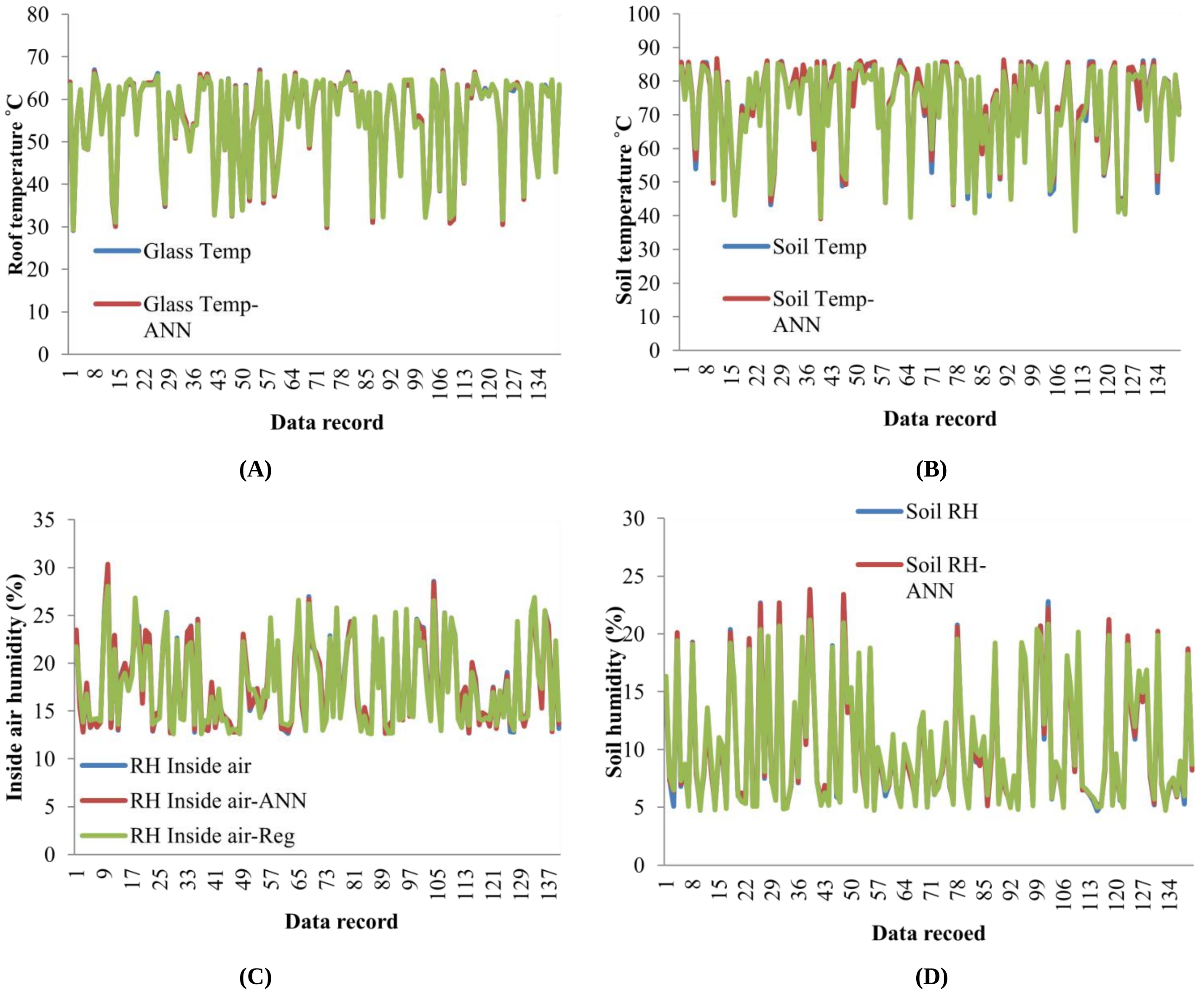


Figure 6Comparison between predicted values by the ANNs and regression model for the roof temperature (A), soil temperature (B), inside air humidity (C) and inside soil humidity (D) parameters on the testing base.

Comparing the results generated using MLP network with those generated by the regression model (Table 7), it can be concluded that MLP model has a higher capability of producing accurate predictions in comparison to regression model, because the MLP model had lower values of RMSE and higher values of EF and R^2 in comparison to regression model. He and Ma (2010), proposed a back propagation neural network (BPNN) based on principal component analysis (PCA) for modeling the internal greenhouse humidity in the winter of North China. They collected the environmental factors influencing the inside humidity including outside air temperature and humidity, wind speed, solar radiation, inside air temperature, open angle of top vent and side vent and open ration of sunshade curtain. Through PCA of these data samples, 4 main factors were extracted, and the relationship between the main factors and the original data was discussed. Taking the principal component values as the input of BPNN, the model showed a good performance. A comparison was made between the performances of the BPNN based on PCA and the stepwise regression method with 20 data samples which had not been used to establish the NN model, and the prediction of stepwise regression method was less accurate than the BPNN based on PCA.

Table7 Performances of two methods in prediction of climate data in semi-solar greenhouse

Parameters	Model type	Performance criterion		
		RMSE, °C	EF,%	R^2
$T_{ri}=f(V_o,T_{ia},T_{oa},I_r)$	MLP	**0.2526**	**0.9994**	**0.9994**
	Regression	0.7137	0.9955	0.9358
$RH_{is}=f(T_{is},I_r,T_{ia},RH_{ia})$	MLP	**0.3056**	**0.9967**	**0.9968**
	Regression	0.7691	0.9796	0.9588
$T_{is}=f(T_{ia},I_r,T_{ir},RH_{ia})$	MLP	**1.0679**	**0.9970**	**0.9945**
	Regression	2.7184	0.9500	0.9827
$RH_{ia}=f(I_r,T_{ri},T_{oa},T_{ia})$	MLP	**0.2502**	**0.9943**	**0.9945**
	Regression	1.0232	0.9631	0.8953

Linker et al (1998) applied ANN model to control the CO_2 balance in a small greenhouse in Israel. Neural network greenhouse models trained using data collected over two summer months in a small greenhouse. The models reduced to minimum size, by predicting separately the temperature and CO_2 concentration, and by eliminating any unessential input. The results showed that ANN models not only fitted the data well, they also seemed qualitatively correct, and produce reasonable optimization results. Wang et al (2009) used Online Sparse Least-Squares Support Vector Machines Regression (OS LSSVMR) to predict some environmental variables in a greenhouse. They used a simplified greenhouse model, in which only greenhouse internal and external air temperatures were considered. Results showed a promising performance in the greenhouse environment with potential improvements, if a more complete data setup is used.

4 Conclusions

This article focused on the comparison between the Artificial Neural Networks (ANNs) and multiple linear regression to predict some internal climate data in a semi-solar greenhouse located in Iran. To show the applicability and superiority of the proposed approach, the measured data of inside soil and roof temperature, inside air and soil humidity were used. To improve the output, the data were first preprocessed. MLP network and multiple linear regression was used and applied with (V_o, T_{ia}, T_{oa}, I_r, T_{is}, RH_{ia}, T_{ri}) as the inputs variable. The network was trained using BB learning algorithm. Statistical comparisons of measured and predicted test data were applied to the selected ANN. From statistical analysis, both measured and predicted test data are similar (with p values greater than 0.9). After testing all possible networks, it has been demonstrated that MLP network with 4-9-1(4 inputs in first layer, 6 neurons in hidden layer and an output) and 4-9-1(4 inputs in first layer, 9 neurons in hidden layer and an output) and LM algorithm had the best output to predict (T_{ri}), (T_{is}) and (RH_{is}) and (RH_{ia}), respectively. It is also found that neural network is particularly suitable for learning nonlinear functional relationships which are not known or cannot be specified. The RMSE for MLP to predict (T_{ri}, T_{is}, RH_{is}, RH_{ia}) was between (0.25°C-1.06°C) and for regression model was between (0.71°C-2.71°C). Because the ANNs do not assume any fixed form of dependence between the output and input values, unlike the regression methods, it seems to be more successful in the application under consideration. It could be concluded that a neural network provides a practical solution to the problem of estimating internal climate data in a greenhouse in a fast, yet accurate and objective way. One of the important problems is that for any situation we need to have a very complete database, i.e. for 12 months of the year, we should have all inputs variables to predict the inside environmental factor very accurate. So, future studies should focus on the applicable structure of ANN and complete this method for artificial greenhouses.

Acknowledgement

The authors would like to thank the editor in chief and the anonymous referees for their valuable suggestions and useful comments that improved the paper content substantially. This study was supported by a grant from Tabriz University, Iran. The authors are grateful for the support provided by the University.

References

Acosta, G., and M.Tosini. 2001. A firmware digital neural network for climate prediction applications, in: Proceedings of the 2001 IEEE International Symposium on Intelligent Control ISIC'01, pp. 127–131.

Bakker, J.C., G.P.A.Bot, H.Chella, and N.J.Van de Braak. 1994. Greenhouse Climate Control, WageningenPers Publishers, Wageningen.

Boaventura, J., C. Couto, and A.E.B. Ruano, 1997. Real-time parameter estimation of dynamic temperature models for greenhouse environmental control, *Control Engineering Practice,*10 (5): 1473–1481.

Boaventura, J., C. Couto, and A.E.B.Ruano. 2000. A greenhouse climate multivariable predictive controller, *Acta Horticulturae,* ISHS 534(31): 269–276.

Bot, G.P.A. 1991. Physical modelling of greenhouse climate, Proceeding of the IFAC/ISHS Workshop, pp. 7–12.

Caponetto, R., L. Fortuna, G. Nunnari, L. Occhipinti, and M.G. Xibilia. 2000. Soft computing for greenhouse climate control. *IEEE Trans. Fuzzy Syst*., 8(6): 753–760.

Coelho, J.P., J. Boaventura, and B.P.M. Oliveira. 2002. Solar radiation prediction methods applied to improve greenhouse climate control. World Congress of Computers in Agriculture and Natural Resources, 13–15 March, pp. 154–161.

Dariouchy, A., E.Aassif, K.Lekouch, L.Bouirden, and G.Maze. 2009. Prediction of the intern parameters tomato greenhouse in a semi-arid area using a time-series model of artificial neural networks. *Measurement*, 42 (3): 456–463.

Dreyfus, G., J.M. Martinez, M. Samuelides, M.B. Gorden, F. Badran, S. Thiria, and L. Hérault. 2004. Réseaux de neurons méthodologie et applications, second ed., ISBN: 2-212-11464-8.

Fatnassi, H., T. Boulard, H. Demrati, L. Bouirden, and G. Sappe. 2002. Ventilation performances of a large canarian, type greenhouse equipped with insect-proof nets. *Biosystems Engineering*, 82 (1): 97–105.

Fernandez, J. E., and B. J. Bailey.1992. Measurement and prediction of greenhouse ventilation rates, Agric. *Forest Meteorol*., 58(3-4):229–245.

Ferreira, P.M., E.A. Faria, and A. E. Ruano. 2002. Neural network models in greenhouse air temperature prediction. *Neurocomputing*, 43 (1-4):51–75.

Haykin, S. 1994. Neural Networks, MacMillan, New York.

He, F., and C. Ma. 2010. Modeling greenhouse air humidity by means of artificial neural network and principal component analysis. *Computers and Electronics in Agriculture*, 71S, S19–S23.

Hornik, H., M. Stinchcombe, and H. White. 1989. Multilayer feedforward networks are universal approximators, *Neural Netw*, 2(5): 359–366.

Jolliet, O. 1994. A model for predicting and optimizing humidity and transpiration in greenhouse, *Journal of Agricultural Engineering Research*, 57(1): 23–37.

Khashei, M., and M.Bijari. 2010. An artificial neural network (p, d, q) model for time series forecasting. *Expert Systems with Applications*, 37(1): 479–489

Linker, R., and I. Seginer. 2004. Greenhouse temperature modeling: a comparison between sigmoid neural networks and hybrid models. *Mathematics and Computers in Simulation*, 65(1-2): 19–29.

Linker, R., I. Seginer, and P. O. Gutman.1998. Optimal CO_2 control in a greenhouse modeled with neural networks. *Computers and Electronics in Agriculture*, 19(3): 289–310

Morimoto, T., and Y. Hashimoto. 2000. An intelligent control for greenhouse automation, oriented by the concepts of SPA and SFA—an application to a post-harvest process. *Computers and Electronics in Agriculture*, 29 (1-2):3–20.

Pérez-Alonso, J., M. Pérez-García, M. Pasamontes-Romera, and A.J. Callejón-Ferre. 2012. Performance analysis and neural modelling of a greenhouse integrated photovoltaic system. *Renewable and Sustainable Energy Reviews*, 16(7): 4675–4685.

Seginer, I. 1997. Some artificial neural network applications to greenhouse environmental control. *Computers and Electronics in Agriculture*, 18(2–3):167–186.

Seginer, I., T. H. Boulard, and B. J. Bailey. 1994. Neural network models of the greenhouse climate. *Journal of Agricultural Engineering Research*, 59(3): 203–216.

Taki, M., Y. Ajabshirchi, S. F. Ranjbar, A. Rohani, and M. Matlooi, 2016a Heat transfer and MLP neural network models to predict inside environment and energy lost in a semi-solar greenhouse. *Energy and Buildings,* 110(1): 314-329.

Taki, M., Y. Ajabshirchi, S.F. Ranjbar, and. A. Rohani.2016b. Modeling and experimental validation of heat transfer and energy consumption in an innovative greenhouse structure. Info Proc Agri, http://dx.doi.org/10.1016/j.inpa. 2016.06.002

Taki, M., M. Mahmoudi, H. Ghasemi-Mobtaker, and H. Rahbari. 2012. Energy consumption and modeling of output energy with multilayer feed-forward neural network for corn silage in Iran. *Agric Eng Int: the CIGR Journal*, 14(4): 93－101.

Van Ooteghem, R.J.C. 2007.Optimal Control Design for a Solar Greenhouse, Systems and Control.Wageningen: Wageningen University.

Van Straten, G., G.Van Willigenburg, E.Van Henten, and R.Van Ootghem. 2011. Optimal control of greenhouse cultivation. CRC press, Taylor and Francis, New York.

Wang, D., M.Wang, and X.Qiao. 2009. Support vector machines regression and modeling of greenhouse environment. *Computers and Electronics in Agriculture,* 66(1): 46–52.

Willmott, C.J., S. G. Ackleson, R. E. Davis, J. J. Feddema, K. M. Klink, D. R. Legates, J. O'Donnell, and C. M. Rowe. 1985. Statistics for the evaluation and comparison of models. *J. Geophys,* Res 90(C5), 8995–9005.