Heat transfer and MLP neural network models to predict inside environment variables and energy lost in a semi-solar greenhouse

Morteza Taki a,*, Yahya Ajabshirchi a, Seyed Faramarz Ranjbar b, Abbas Rohani c, Mansour Matloobi d

a Department of Biosystems Engineering, Faculty of Agriculture, University of Tabriz, Iran
b Department of Mechanical Engineering, Mechanical Engineering Faculty, University of Tabriz, Iran
c Department of Biosystems Engineering, Faculty of Agriculture, Ferdowsi University of Mashhad, Iran
d Department of Horticultural Science, Faculty of Agriculture, University of Tabriz, Tabriz, Iran

ARTICLE INFO

Article history:
Received 5 September 2015
Received in revised form 17 October 2015
Accepted 3 November 2015
Available online 12 November 2015

Keywords:
Semi-solar greenhouse
Heat transfer
Energy lost
Intelligent system

ABSTRACT

The greenhouse environment is an uncertain nonlinear system which classical modeling methods have some problems to solve. There are many control methods, such as adaptive, feedback and intelligent control and they require a precise model. Therefore, many modeling methods have been proposed for this purpose, including physical transfer function and black-box modeling. The main goal of this paper is to compare some mathematical models (include dynamic and multiple linear regression (MLR)) with innovative method (Artificial Neural Network) and select the best one to predict inside air and roof temperature (T a and T o) and energy lost in a semi-solar greenhouse in Iran. For this purpose, 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). The environment factors influencing the T a and T o include outside air temperature (T xa), wind speed (v o), solar radiation on the roof (I o), inside soil temperature (T s) and inside air humidity (RH s), which were all collected as data samples. Then, the relationship between the factors, 4 main factors were extracted, and the relationship between the main factors and the original data was discussed by MLP and MLR models. Results showed that the Durbin–Watson statistic for MLR method to estimate T a and T o was 0.04 and 0.06 respectively, so this method cannot predict the output parameters correctly. Comparing MLP and dynamic models showed that the performance of MLP model was better according to small values of RMSE and MAPE and large value of EF indices. Statistical comparisons of the predicted data by neural network models and the actual data of the inside air and roof temperature showed that there is no significant difference between them. Also, the minimum value of the TSSE was 16.68 and 30.87 (°C) for T a and T o in ANN implementation. The performance of best model (MLP) to estimate the energy lost and exchange in a semi-solar greenhouse showed that MLP method is applicable to estimate the real data in greenhouse and then predict the energy lost and exchange. This method can be used online in greenhouses to decrease some cost related to application of sensors and some record instruments.

© 2015 Elsevier B.V. All rights reserved.

1. Introduction

The greenhouse environment is a very complex dynamic system covered with thin and transparent materials. This system satisfies the conditions for plant growth, but creates difficulties in controlling the greenhouse environment because of time delays and intensive disturbances from the surroundings, such as global radiation, wind speed and direction, and external air temperature and humidity. In the view of classical feedback control, such a system is poorly controlled if disturbance monitors and model based feed-forward control is not applied. For this reason, computer control technology for the greenhouse environment continues to require serious attention from researchers and engineers in various fields, although it has been studied since the 1980s [1]. Many modeling methods have been applied to control the environment of greenhouse, such as mechanism, transfer function and black-box modeling. The mechanism model provides a clear physical explanation of the greenhouse environment, such as the early static and dynamic model presented by Bot [2] or improved models presented...
The rate of Farhan Bagheri and Els A. Joudi [10] have used solar air, greenhouse roof and sky radiation for the purpose of greenhouse energy modeling, design and application. This paper is a part of this project. The main objective of this research is comparison between mathematical models (heat transfer and multiple regression analysis) and innovative way (ANN model) to predict the inside air and roof temperature in a semi-solar greenhouse according to the use of some inside and outside parameters, and then, select the best results. In the second part, we use the best results of the above models and apply them to calculate energy lost and exchange and compare the results with experimental data. This method can be used in automatic greenhouse in the future to decrease the cost for farmers and customers.

2. Materials and methods

2.1. Selection of the site and structure of sample greenhouse (semi-solar)

In this study, a semi-solar greenhouse was designed and constructed at the North-West of Iran in Azerbaijan Province. In solar greenhouse design, the engineers try to have the maximum solar radiation and decrease the heat lost. Warm and cold water aquifer layers 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 [34]. The solar greenhouse has some changes compared to a conventional greenhouse such as [35] 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, University of Tabriz with the cooperation of Science and Technology Vice-Presidency in Iran. Shape and orientation of the greenhouse, has been selected based on some common greenhouse shapes (Fig. 1) and according to the received maximum solar radiation. For this, meteorological data recorded by Iran
Meteorological Office for the period of 1992–2013 was used and after some analysis, this structure was selected. Also internal thermal screen (cloth type) and cement north wall were used to store and prevent heat lost during the cold period of 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 m² and 26.4 m³. The orientation of this greenhouse is East–West and perpendicular to the direction of the wind prevailing (Fig. 2).

2.2. Dynamic model

In order to simulate the inside air and roof temperatures, the system is divided into three elements, namely, the soil surface, internal air and roof cover. The present model consists of three first-order differential equations which were derived from energy balances for these elements. In this dynamic model, the following assumptions were made for the heat exchange between greenhouse components without crops:

- The greenhouse elements are considered as lumped systems.
- The air and top soil temperatures are uniform.
- No evaporation occurs from the soil.
- Radiation energy was neither absorbed nor emitted by the inside air.
- The greenhouse has dry soil and does not contain plants.

The first of these equations is the energy balance derived for the inside air which can be written as [35]:

$$\frac{dT_a}{dt} = \frac{Q_{a-s} - Q_{a-o} - Q_{a-ri} - Q_{nwi-nwo}}{\rho_a \times c_{p-a} \times V_a}$$

(1)

The energy balance equation derived for the roof area is the second equation and can be written as [35]:

$$\frac{dT_{ri}}{dt} = \frac{Q_{rd-ri} + Q_{a-ri} + Q_{s-ri} - Q_{ri-o} - Q_{ri-sk}}{\rho_r \times c_{p-r} \times V_r}$$

(2)
The third equation is the energy balance equation derived for the bare top soil of greenhouse and can be written as [35]:

$$\frac{dT_s}{dt} = \frac{Q_{rd-s} + Q_{a-s} - Q_{ss} - Q_{w-v}}{\rho_s \times c_p \times V_s}$$  \tag{3}

The energy transferred between greenhouse elements by convection and conduction is expressed as [4]:

$$Q_{a-s} = A_s \times \alpha_{a-s} (T_a - T_s)$$ \tag{4}

$$Q_{a-r} = A_r \times \alpha_{a-r} (T_a - T_r)$$ \tag{5}

$$Q_{r-d} = A_r \times \alpha_{r-d} (T_r - T_o)$$ \tag{6}

$$Q_{w-v} = A_{mv} \times \frac{\lambda_{mv}}{d_{mv}} (T_{w-v} - T_{mv})$$ \tag{7}

$$Q_{ss} = A_s \times \frac{\lambda_s}{d_s} (T_s - T_{ss})$$ \tag{8}

Empirical relations reported in the literature to estimate the heat transfer coefficients between the different surfaces in a greenhouse are as follows [34]:

$$\alpha_{a-r} = 3 |T_a - T_r|^{1/3}$$ \tag{9}

$$\alpha_{a-s} = 1.7 |T_a - T_s|^{1/3} \quad \forall T_s < T_a$$ \tag{10}

$$\alpha_{a-s} = 1.3 |T_a - T_s|^{0.25} \quad \forall T_s \geq T_a$$

$$\alpha_{r-d} = 2.8 + 1.2V_o \quad \forall V_o < 4$$  \tag{11}

$$\alpha_{r-i} = 2.5V_o^{0.8} \quad \forall V_o \geq 4$$

Infiltration through the greenhouse can be calculated as [36]:

$$Q_{a-o} = \rho_a \times c_{p-a} \times \phi_{a-o} (T_a - T_o)$$ \tag{12}

$$\phi_{a-o} = A_{rd}(8.3 \times 10^{-5} + 3.5 \times 10^{-5} \times f_a \times V_o)$$ \tag{13}

The solar radiation absorbed directly by the roof and soil surface ($Q_{rd-ni}$ and $Q_{rd-s}$) in Eqs. (3) and (2) is given as [34]:

$$Q_{rd-ni} = A_r \times \eta_{ri-ni} \times I_{roof}$$ \tag{14}

$$Q_{rd-s} = A_s \times \eta_{si-s} \times I_{inf}$$ \tag{15}

The net solar radiation heat exchanges between soil and roof ($Q_{ri-ni}$) and also between roof and sky ($Q_{ri-sk}$) are given as [35]:

$$Q_{ri-ni} = A_r \times E_{ri} \times E_{n-i} \times F_{ri-ni} \times \sigma (T_s^4 - T_{n-i}^4)$$ \tag{16}

$$Q_{ri-sk} = A_r \times E_{ri} \times E_{sk} \times F_{ri-sk} \times \sigma (T_s^4 - T_{sk}^4)$$ \tag{17}

The sky temperature suggested by Joudi and Farhan [11] is:

$$T_{sk} = 0.0552 (T_o)^{1.5}$$  \tag{18}

Fig. 3 shows all the above equations graphically (heat transfer between greenhouse components).

MATLAB software was used to solve the mathematical equations. The entire set of equations was solved at each 1 min time.
step using appropriate values of input parameters at the specific time step. When the solution converged, the computed data were taken as initial values for the next time step. The first values of $T_a$, $T_{ri}$ and $T_{pi}$ were measured. The input data for solution are given in Table 1. Simulation was done between 9:00 and 18:00 pm on 30/06/2015 in a semi-solar greenhouse located in University of Tabriz, Department of Biosystems Engineering.

### 2.3. Internal and external climate data

To measure the temperature and the relative humidity of the air, soil and roof inside and outside the greenhouse, the SHT11 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 micromachining (CMOSens® technology) ensures highest reliability and excellent long term stability. The device includes a capacitive polymer sensing element for relative humidity and a band-gap temperature sensor. Both are seamlessly coupled to a 14-bit 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 (Fig. 4). 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. We used these sensors in soil, on the roof (inside greenhouse) and in the air of greenhouse and outside to measure temperature and relative humidity.

On the greenhouse roof, we used a pyranometre type TES 1333. Its sensitivity is proportional to the cosine of the incidence angle of the radiation. It can measure the global radiation of spectral band solar in the 400–1110 nm range. Its measurement accuracy is approximately ±5%. Fig. 5 shows the place of SHT11 sensors and TES1333 pyranometre to collect the data in the semi-solar greenhouse.

---

### Table 1

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
<th>Parameter</th>
<th>Value</th>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\eta_{ls}$</td>
<td>0.0173</td>
<td>$\rho_s$</td>
<td>1.21</td>
<td>$A_s$</td>
<td>15.36</td>
</tr>
<tr>
<td>$E_s$</td>
<td>0.7</td>
<td>$c_{p-a}$</td>
<td>1000</td>
<td>$\lambda_s$</td>
<td>0.6</td>
</tr>
<tr>
<td>$E_{vok}$</td>
<td>0.8</td>
<td>$f_s$</td>
<td>1</td>
<td>$d_s$</td>
<td>0.65</td>
</tr>
<tr>
<td>$A_{vw}$</td>
<td>0.86</td>
<td>$A_v$</td>
<td>17.7</td>
<td>$V_v$</td>
<td>9.984</td>
</tr>
<tr>
<td>$\lambda_{vw}$</td>
<td>11.52</td>
<td>$V_i$</td>
<td>0.0708</td>
<td>$c_{p-a}$</td>
<td>800</td>
</tr>
<tr>
<td>$d_{vw}$</td>
<td>0.397</td>
<td>$V_s$</td>
<td>26.4</td>
<td>$\rho_s$</td>
<td>1400</td>
</tr>
<tr>
<td>$E_{s}$</td>
<td>0.25</td>
<td>$\rho_i$</td>
<td>2500</td>
<td>$F_{s-ri}$</td>
<td>0.8</td>
</tr>
<tr>
<td>$E_{s}$</td>
<td>840</td>
<td>$\tau$</td>
<td>$5.67051 \times 10^{-8}$</td>
<td>$\eta_{ls}$</td>
<td>0.86</td>
</tr>
</tbody>
</table>

---

![Fig. 3](image1.png) The greenhouse schematic and its heat exchange with surroundings.

![Fig. 4](image2.png) Performance of SHT11 sensor to record and convert the data to digit.
2.4. Artificial Neural Network

Prior to any ANN 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 ANN 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 [37]:

\[ x_n = \frac{x - x_{\text{min}}}{x_{\text{max}} - x_{\text{min}}} \times (r_{\text{max}} - r_{\text{min}}) + r_{\text{min}} \]  

where x is the original data, \( x_n \) the normalized input or output values, and \( x_{\text{max}} \) and \( x_{\text{min}} \) are the maximum and minimum values of the concerned variable, respectively. \( r_{\text{max}} \) and \( r_{\text{min}} \) 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.

2.4.1. The multilayer perceptron neural network

Among various ANN models, multilayer perceptron (MLP) has maximum 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 nondecreasing and differentiable. The most popular function is unipolar sigmoid [37]:

\[ f(\theta) = \frac{1}{1 + e^{-\theta}} \]  

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 \( q = 1, 2, \ldots, Q \)). 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 \( q = 1, 2, \ldots, Q \)). 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 [38]. In this study, Basic Back-propagation (BB) algorithm was employed.

2.4.2. Basic Back-propagation algorithm

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. We used an N-fold cross-validation method in which data are randomly divided into two sets: training set and cross-validation set [38]. All inputs we used to predict inside air and roof temperature are illustrated in Fig. 6. A computer code was also developed in MATLAB software for the feed-forward and back-propagation network.

2.4.3. Multiple linear regressions

In the present study, multi-linear regression analysis was carried out to simplify inside and roof temperature calculation. Multiple linear regression model is the relationship between two or more explanatory variables and a response variable by fitting a linear equation to observed data. Every value of the independent variable \( x \) is associated with a value of the dependent variable \( y \). The following form of the regression equation was used for inside air and roof temperature [39]:

\[ y = \alpha + \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \cdots + \beta_p x_p \]  

**Fig. 5.** SHT11 sensors place to collect inside and outside parameters.
where $y$ is the inside air and roof temperature, $x_i$ presents the value of parameter and $\hat{y}_i$ is the corresponding regression coefficient.

2.4.4. Performance evaluation criteria

To evaluate the performance of a model, some criteria have been defined in the literature. These criteria include: Total Sum of Squared Error (TSSE), Mean Square Error (MSE), Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), Mean Absolute Percentage Error (MAPE), Root Mean Square Percentage Error (RMSPE), coefficient of determination ($R^2$), Model Efficiency (EF), etc. Among them, MAPE, RMSE, EF, TSSE and $R^2$ are the most widely used performance evaluation criteria and may be used to compare the predicted and actual values which will be used in this study. They are defined as follows:

$$\text{MAPE} = \frac{1}{n} \sum_{j=1}^{n} \left| \frac{d_j - p_j}{d_j} \right| \times 100$$  \hspace{1cm} (22)

$$\text{RMSE} = \sqrt{\frac{\sum_{j=1}^{n} (d_j - p_j)^2}{n}}$$  \hspace{1cm} (23)

$$\text{EF} = \frac{\sum_{j=1}^{n} (d_j - \bar{d})^2 - \sum_{j=1}^{n} (p_j - \bar{d})^2}{\sum_{j=1}^{n} (d_j - \bar{d})^2}$$  \hspace{1cm} (24)

$$\text{TSSE} = \sum_{j=1}^{n} (d_j - p_j)^2$$  \hspace{1cm} (25)

$$R^2 = \left[ \frac{\sum_{j=1}^{n} (d_j - \bar{d})(p_j - \bar{p})}{\sum_{j=1}^{n} (d_j - \bar{d})^2 \sum_{j=1}^{n} (p_j - \bar{p})^2} \right]^2$$  \hspace{1cm} (26)

where $d_j$ is the $j$th component of the desired (actual) output for the $j$th pattern; $p_j$ is the component of the predicted (fitted) output produced by the network for the $j$th pattern; $\bar{d}$ and $\bar{p}$ are the averages of the whole desired (actual) and predicted output and $n$ is the number of variable outputs. A model with the smallest RMSE, MAPE and ESSE with largest EF and $R^2$ is considered to be the best [40].

3. Results and discussion

3.1. Selection the best group of variables to predict temperature by MLP and MLR models

The procedure applied in this development started by distinguishing the correlation coefficients between all inputs and outputs (Table 2). Studying the correlation coefficient among different characters makes it possible to decide more precisely about selected indirect selection indices and removing ineffective characters. As we can see, the inputs have a significant correlation to each other except outside wind speed ($v_o$). This fact is because these inputs have a very complex dynamic relation to each other. Every change in each of them can change the others and it is one of the problems in modeling the inside greenhouse environment by heat and mass transfer method. In the view of classical feedback control, such a system is poorly controlled if disturbance monitors and model based feed-forward control are not applied. All inputs have a positive correlation except inside air humidity (RHa). It is clear that when we increase humidity, temperature will decrease. Also, it will be seen for the relation between humidity and solar radiation (increase in solar radiation will change the temperature and humidity directly). Some researchers used external and internal variables for prediction inside environment parameters in greenhouse. He and Ma [33] proposed a back-propagation neural network (BPNN) based on principal component analysis (PCA) for modeling the internal greenhouse humidity in winter of North China. They collected the environment 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.

Selection of the best group of variables to predict $T_a$ and $T_{tn}$ by MLP and MLR models started with all inputs (i.e. each input separately), and then we removed the $v_o$ because it had the worst results and then the procedure continued with other inputs in two, three and four groups and finally the variables that had the best results were selected (Table 3). These inputs ($x_1$-$x_4$) were used to train and test MLP and estimate the output ($T_a$ and $T_{tn}$) by MLR models. The effect of selected inputs on outputs ($T_a$ and $T_{tn}$) is shown in Fig. 7.

3.2. MLR model result

The aim of regression analysis is to determine simple and sufficiently accurate models for predicting inside air and roof temperature from inside and outside variables ($x_1$-$x_4$). Minitab 17 software was used to run the models of multiple regression. The obtained regression equations for $T_a$ and $T_{tn}$ estimation based on four selected variables ($x_1$-$x_4$) are shown in Table 4. The coefficients of determination ($R^2$) values show that these equations can justify 98.15% and 97.86% of changes in $T_a$ and $T_{tn}$, respectively. The percent of contribution (PC) of the regression models’ parameters (i.e. $x_1$-$x_4$) and the models’ errors in predicting inside air and roof temperature were calculated by the sum of squares (SS) in Table 4. According to the results, 2.14% and 1.85% of total changes in the inside air and roof temperature regression models are because of errors which are acceptable. From the PC values of each geometric parameter, all inputs have the most to least impact in estimating estimated parameter. Although the contribution of $x_2$-$x_4$ is less than $x_1$, the p-values in Table 3 show that the effects of these variables are significant and therefore they
Table 3
Selection the best group of variables between \((x_1 = T_s, x_2 = I_r, x_3 = T_o, x_4 = RH_a, x_5 = v_0)\).

<table>
<thead>
<tr>
<th>Parameter</th>
<th>(T_a)</th>
<th>(T_s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>MLP</td>
<td>MLR</td>
</tr>
<tr>
<td>Criteria</td>
<td>RMSE</td>
<td>MAPE</td>
</tr>
<tr>
<td>(x_1)</td>
<td>1.31</td>
<td>1.57</td>
</tr>
<tr>
<td>(x_2)</td>
<td>2.07</td>
<td>4.16</td>
</tr>
<tr>
<td>(x_3)</td>
<td>2.28</td>
<td>2.85</td>
</tr>
<tr>
<td>(x_4)</td>
<td>3.36</td>
<td>4.11</td>
</tr>
<tr>
<td>(x_5)</td>
<td>5.15</td>
<td>6.29</td>
</tr>
<tr>
<td>(x_1, x_2)</td>
<td>0.26</td>
<td>0.30</td>
</tr>
<tr>
<td>(x_1, x_3)</td>
<td>1.01</td>
<td>1.10</td>
</tr>
<tr>
<td>(x_1, x_4)</td>
<td>0.71</td>
<td>0.79</td>
</tr>
<tr>
<td>(x_2, x_3)</td>
<td>0.21</td>
<td>0.27</td>
</tr>
<tr>
<td>(x_2, x_4)</td>
<td>0.41</td>
<td>0.44</td>
</tr>
<tr>
<td>(x_3, x_4)</td>
<td>1.50</td>
<td>1.74</td>
</tr>
<tr>
<td>(x_1, x_2, x_3)</td>
<td>0.38</td>
<td>0.44</td>
</tr>
<tr>
<td>(x_1, x_2, x_4)</td>
<td>0.20</td>
<td>0.25</td>
</tr>
<tr>
<td>(x_2, x_3, x_4)</td>
<td>0.45</td>
<td>0.53</td>
</tr>
<tr>
<td>(x_1, x_3, x_4)</td>
<td>0.64</td>
<td>0.71</td>
</tr>
<tr>
<td>(x_1, x_2, x_3, x_4)</td>
<td>0.17</td>
<td>0.23</td>
</tr>
</tbody>
</table>

Some statistical properties of the actual and predicted values by MLR models are shown in Table 5. 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 values are remarkable. Fig. 8 shows the distribution of the actual data and the predicted data by MLR models for $T_\text{a}$ and $T_\text{r}$.

A one-way between subjects ANOVA was conducted to compare the effect of $T_{\text{air}}-T_{\text{roof}}$ on inside air and roof temperature ($T_\text{a}$ and $T_\text{r}$). There was a significant effect of all inputs on outputs at the $p<0.01$ level (Table 6). In statistics, the Durbin–Watson statistic is a kind of test that is used to detect the presence of autocorrelation (a relationship between values separated from each other by a given time lag) in the residuals (prediction errors) from a regression analysis. The Durbin–Watson statistic ranges in value from 0 to 4. A value near 2 indicates nonautocorrelation; a value toward 0 indicates positive autocorrelation; a value toward 4 indicates negative autocorrelation [41]. Results of test of this statistic are shown in Table 6. The autocorrelation is illustrated between input variables and then we cannot use multiple regression models to predict $T_\text{a}$ and $T_\text{r}$. In economics, it is a kind of regression called “fake regression” [41]. In some researches in this area, researchers applied regression models to estimate inside variables in greenhouse but they did not care about the Durbin–Watson statistic and autocorrelation between variables. For example Dariouchy et al. [15], proposed MLP and MLR models to predict the inside air humidity and temperature in a tomato greenhouse in semi-arid area in Morocco. Model database was collected starting from the greenhouse climatic data. The External Moisture ($M_{\text{ext}}$), the Total Radiation ($R_l$), the Wind Direction ($D_{\text{w}}$), the Wind Velocity ($V_{\text{w}}$) and the External Temperature ($T_{\text{ext}}$) are retained like relevant entries of the time-series model. The results showed that the correlations between actual and predicted values by MLR model are 0.970 and 0.978 for inside air temperature and humidity respectively. The final results showed that the MLP model had a higher accuracy. They did not report the autocorrelation between variables in this research.

### 3.3. Dynamic model result

After recording the experimental temperatures, solar radiation and wind speed, we used Eqs. (1)–(18) and applied initial values and estimated $T_\text{a}$ and $T_\text{r}$. The air temperature was regarded as the average temperature of two points in center of greenhouse and roof temperature was the average temperature of two points on the center of glass roof. The comparison between simulated and experimental values of average air and roof temperatures is shown in Figs. 9 and 10. Some statistical properties of the actual and predicted values by dynamic models are shown in Table 7. Considering the average values of standard deviation and variance, it can be deduced that the values and the distribution of real and predicted...
Table 6
ANOVA analysis for regression models.

<table>
<thead>
<tr>
<th>DF</th>
<th>T1, MS</th>
<th>p-Value</th>
<th>T2, MS</th>
<th>p-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>x1</td>
<td>1</td>
<td>28407.4</td>
<td>0.00</td>
<td>29123.0</td>
</tr>
<tr>
<td>x2</td>
<td>1</td>
<td>1235.6</td>
<td>0.00</td>
<td>469.4</td>
</tr>
<tr>
<td>x3</td>
<td>1</td>
<td>124.0</td>
<td>0.00</td>
<td>309.8</td>
</tr>
<tr>
<td>x4</td>
<td>1</td>
<td>224.7</td>
<td>0.00</td>
<td>41.9</td>
</tr>
<tr>
<td>Error</td>
<td>428</td>
<td>655.5</td>
<td>97.86%</td>
<td>564.0</td>
</tr>
<tr>
<td>R²</td>
<td></td>
<td></td>
<td>97.84%</td>
<td></td>
</tr>
<tr>
<td>R² adj</td>
<td></td>
<td></td>
<td>97.86%</td>
<td></td>
</tr>
<tr>
<td>DW</td>
<td></td>
<td>0.06</td>
<td></td>
<td>0.04</td>
</tr>
</tbody>
</table>

Fig. 9. Relationship between the actual values (dv) and the predicted values by dynamic models (pv). (a: inside roof temperature and b: inside air temperature).

Fig. 10. Comparison of hourly prediction of inside roof and air temperature (a and b) by dynamic model and dished values.

data are not analogous and the differences of minimum values are remarkable. In dynamic method, there are a lot of kinds of error such as equation and material properties errors. Typically in dynamic method, researcher has more than 10% error in such as this work [11,12] and one of the important aims of this paper is to consider this fact and suggest the reviewers to change the classical methods into innovative tools. However, in this part, we used this method to have a good comparison with others. A lot of researches
have used this method since 1960–2015 in simulation and modeling the greenhouse environment. So we focus on other methods in this paper and investigate the least error and high accuracy to select the best model to evaluate the energy lost and environmental changes in semi-solar greenhouse.

### 3.4. The results of ANN modeling

Considering the pseudo-periodicity of the climatic variations in a semi-arid area, it was sufficient to work on 1 set data (24 h) to learn the BB-MLP model. The compiled database represents 1-day sets of parameter values inside the greenhouse. The training phase of 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), coefficient of determination ($R^2$) and Mean Absolute Percentage Error (MAPE). Therefore, required ANN model was developed in two phases: training (calibration) and testing (generalization or validation) phase. To select the best part of data for train and test, we tested four groups of data (Table 8) and selected the best one according to lowest RMSE, MAPE, TSSE and highest EF. So in the training phase, a larger part of database (80%) was used to train the network and the remaining part of the database (20%) was used in the testing phase.

Based on universal approximation theorem, a neural network with a single hidden layer and sufficiently a large number of neurons can well approximate any arbitrary continuous function [21]. 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 is improved 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 not enough, the network may not be able to learn. During 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. Fig. 11 demonstrates MSE versus the epoch number (number of learning runs) for the MLP models. As it can show, after 143 and 278 repetitions, the MSE for $T_a$ and $T_{ri}$ reached to $2.69e-5$ and $2.34e-5$ respectively and the MLP model learned to estimate the outputs in the lowest error.

From a 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 [40]. $p$ 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. The results are shown in Table 9. 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 ($p > 0.98$). 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 ($p > 0.20$). Finally in the training, validation and testing phases of ANN implementation, there is no significant difference between the actual data and the predicted data and minimum value of $R^2$ for the inside air and roof variables in all three phases of ANN is equal to 0.9991 and 0.9354 respectively. The $R^2$ of testing phase is less than the validation and training phases, because the imported data to the network are new at this phase. According to these results, the ANN ($x_1$–$x_4$) model could learn well the pattern of changes in $T_a$ and $T_{ri}$ using these inputs, and has a good ability to extend the flexibility. Thus, the prediction results of ANN ($x_1$–$x_4$) models are reliable. Fig. 12 shows the comparison of predicted data by ANN to actual data collected by SHT11 sensors in semi-solar greenhouse.

### 3.5. Select the best model to predict $T_a$ and $T_{ri}$ between MLR, ANN and dynamic method

Finally, the three proposed models for estimating the inside air and roof temperature were compared with each other to select the best one and propose it for other works. The results of this comparison are shown in Table 10. According to small values for RMSE, MAPE and TSSE and large value for EF and $R^2$ indices, the prediction performance of ANN ($x_1$–$x_4$) model is better than the other models. Since, the ANN models have better performance than the regression and dynamic models, the ANN modeling is acceptable. Fig. 13 shows the comparison between actual and predicted values by three models. The regression model has better results than dynamic but the autocorrelation makes these results fake. We suggest using two variables (i.e. outside air temperature and solar radiation on the roof) for future works and comparing the results with ANN model. One-way analysis of variance was done for four
Table 8
The percentage of four-group data and the statistical index of them to select the best part of data for training the network.

<table>
<thead>
<tr>
<th>Train</th>
<th>( T_a )</th>
<th>MAPE (%)</th>
<th>RMSE (°C)</th>
<th>EF (%)</th>
<th>TSSE (°C²)</th>
<th>( T_r )</th>
<th>MAPE (%)</th>
<th>RMSE (°C)</th>
<th>EF (%)</th>
<th>TSSE (°C²)</th>
</tr>
</thead>
<tbody>
<tr>
<td>70</td>
<td>0.30</td>
<td>0.25</td>
<td>0.99</td>
<td>35</td>
<td>0.32</td>
<td>0.28</td>
<td>0.99</td>
<td>43</td>
<td></td>
<td></td>
</tr>
<tr>
<td>50</td>
<td>0.37</td>
<td>0.32</td>
<td>0.99</td>
<td>56</td>
<td>0.37</td>
<td>0.32</td>
<td>0.99</td>
<td>56</td>
<td></td>
<td></td>
</tr>
<tr>
<td>25</td>
<td>0.40</td>
<td>0.34</td>
<td>0.99</td>
<td>66</td>
<td>0.50</td>
<td>0.42</td>
<td>0.99</td>
<td>97</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>0.54</td>
<td>0.45</td>
<td>0.99</td>
<td>112</td>
<td>0.76</td>
<td>0.60</td>
<td>0.99</td>
<td>196</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 9
The \( p \) values and \( R^2 \) between the actual and predicted data for the ANN models.

<table>
<thead>
<tr>
<th>Index</th>
<th>Training phase</th>
<th>Validation phase</th>
<th>Test phase</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>Variance</td>
<td>Distribution</td>
</tr>
<tr>
<td>( T_a )</td>
<td>( p )</td>
<td>0.9984</td>
<td>0.9976</td>
</tr>
<tr>
<td>( R^2 )</td>
<td>0.9992</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( T_r )</td>
<td>( p )</td>
<td>0.9997</td>
<td>0.9968</td>
</tr>
<tr>
<td>( R^2 )</td>
<td>0.9354</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Fig. 12. Comparison between the desired and predicted values by the ANN model for the inside roof and air temperature (a and b).

Fig. 13. Comparison between three models (i.e. ANN, MLR and dynamic) to estimate inside air and roof temperature in semi-solar greenhouse.
sets of data including the actual and the predicted data by ANN, MLR and dynamic models. The results of mean comparison by LSD method for these three sets of data are shown in Fig. 14 using box plots. The difference between the amounts of all four series of data is not significant at 1% level for $T_a$ but it is significant for $T_{ri}$. So there is no difference between ANN and MLR models statistically to predict $T_a$, but dynamic model is not suitable to estimate $T_{ri}$. These results were approved by some researchers [16,31,33]. Dariouchy et al. [15], developed MLP and MLR models to predict inside air and humidity in a semi-arid region in Morocco. The comparison between the obtained results and experimental ones indicates that the ANN method is suitable to predict the greenhouse climatic data.

### 3.6. Prediction of the energy lost in a semi-solar greenhouse

The final part of this paper is to use the results of the best selected model to estimate the energy lost and exchange in a semi-solar greenhouse. Energy lost includes $Q_{as}$, $Q_{awi-nwo}$, $Q_{ri-o}$ and $Q_{ri-sk}$ that $Q_{ri-o}$ and $Q_{ri-sk}$ related to inside roof temperature and energy exchange include $Q_{as}$, $Q_{awi}$, and $Q_{ri-o}$. The inside air and roof temperature affect these parameters. Fig. 15 shows the comparison of three models to predict energy lost and exchange during 9:00–18:00 pm on 30/06/2015. As we can see, the MLP model has the best results compared to actual. Prediction of inside roof temperature by dynamic method had some problems. So the amount of energy exchange between inside air and inside part of roof ($Q_{ri-o}$) and energy transfer between soil and roof ($Q_{ri-sk}$) were very different in comparison to other models. Fig. 16 shows the hourly energy lost and exchange by three models in semi-solar greenhouse. Energy exchange between air and soil ($Q_{as}$) increases until 12:00 pm and then starts to decrease. At noon, the average of inside air and roof temperature could be equal and then the energy exchange between them was zero, but after noon, the difference between these parameters made the energy exchange between them different. Overall, the MLP method was the best and more accurate model compared to others. Fig. 17 shows the input and energy lost for a sunny day (30/06/2015) from 9:00 am to 18:00 pm. The results show that the difference between MLP and actual data to predict energy lost is about 0.278 MJ (it is equivalent to 0.005 kg LPG (Propane n Butane)) [42]. So we can conclude that the physical greenhouse environment models have a high degree of complexity with lots of parameters that have to be determined by measurements or other sub-models and also have more errors. In contrast to physical models, black-box models do not suffer from the need to determine every parameter's value. These models can be used to estimate the inside environment changes and they can be very helpful for climate control purpose. So because of nonlinear system in greenhouse, time-invariant, and strong coupling, present applications of Artificial Neural Network (ANN) model for the simulation and prediction of greenhouse inside climate can be very useful and applicable. In this paper, we tried to show this fact that innovative methods are simple and more accurate than physical heat and mass transfer method to predict the environment changes. Furthermore, this method can be used to predict other changes in greenhouse such as final yield, evapotranspiration, humidity, cracking on the fruit, CO₂ emission and so on. For example, Kok et al. [43] and Seginer et al. [44] trained NN to imitate greenhouse models predicting the inside air temperature, ventilation and other environmental factors. Linker et al. [25] applied the NN greenhouse model to predict CO₂ concentration for optimizing CO₂ control and found that this method is very simple, accurate and user friendly. Landeras et al. [45] investigated the performance of ARIMA and ANNs in estimating the weekly ET in Basque Country, north Spain. They compared the ET values predicted by ARIMA and ANNs models with those assessed by FAO-PM56. They found out that the ANN models improved the performance of the

### Table 10

Performances of three methods for prediction of environment data in semi-solar greenhouse.

<table>
<thead>
<tr>
<th>Model</th>
<th>$T_a$</th>
<th>$T_{ri}$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>MAPE (%)</td>
<td>RMSE (°C)</td>
</tr>
<tr>
<td>MLP</td>
<td>0.29</td>
<td>0.23</td>
</tr>
<tr>
<td>MLR</td>
<td>1.49</td>
<td>1.18</td>
</tr>
<tr>
<td>Dynamic</td>
<td>4.14</td>
<td>2.74</td>
</tr>
</tbody>
</table>

*Fig. 14. Box plots for changes in the actual values of inside air and roof temperature versus the predicted values by different models.*

*Fig. 15. Comparison of the actual average of energy lost and exchange between the greenhouse air, soil and roof by MLP, MLR and Dynamic methods.*
ET prediction. Ladlani et al. [46] compared the efficiency of generalized regression neural networks (GRNNs) with that of radial basis function neural networks (RBFNNs) in estimating ET in Algeria. ET was calculated by FAO-PM56, and the calculated ET was then utilized for calibration and validation of GRNNs and RBFNN models. The results of two models were also compared with the results of Hargreaves (HG) and Hargreaves–Samany and Priestly Taylor (PT) methods. They concluded that ET data estimated by GRNNs model had a better correlation with the reference ET. The future research will focus on online MLP model in the semi-solar greenhouse to decrease the cost (sensor and other tools) and go toward building the automatic greenhouse, the first one in Iran.
4. Conclusion

This paper presents a comparison between mathematical and innovative models to select the best method to predict the inside air and roof temperature ($T_a$ and $T_r$) and energy lost in a semi-solar greenhouse starting from some external and internal climatic data in East Azerbaijan province, Iran. According to the results:

1. The relationship between input variables according to RMSE, MAPE and TSSE showed that inside soil temperature, outside air temperature, inside air humidity and solar radiation on the roof have the highest impact on $T_a$ and $T_r$.
2. The autocorrelation between input variables showed that multiple linear regression method cannot estimate the $T_a$ and $T_r$ correctly.
3. The results of mean comparison by LSD method for MLP and dynamic model showed that dynamic method has very bad results to calculate inside roof temperature and should improve the mathematic-experimental equations related to this parameter.
4. Comparison between neural network and dynamic model showed that ANN technique can be used as a reliable method to solve the nonlinear relation between inside environment variables and estimate $T_a$ and $T_r$ with high accuracy and without complex procedure. This method can calculate the energy lost and exchange in greenhouse very accurately and should develop to an online method in future solar greenhouse to decrease main cost and increase efficiency.
5. Using the same methodology can develop models to predict fuel consumption, CO$_2$ emission, and other agricultural production (yield) in the greenhouses. It is possible to use the same method to collect some suitable data related to the above subject and investigate on these aspects. Modeling fuel consumption, CO$_2$ emission, yield, and energy consumption based on social and technical parameters would open new doors to advances in agriculture and modeling.

Acknowledgments

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. Also the authors thank Dr. R. Van Ootheghem for helping to make basic proposal of this research. This study was supported by a grant from University of Tabriz, Iran. The authors are grateful for the support provided by University of Tabriz.

References