Estimating Inside Air Temperature of a Glasshouse Using Statistical Models

ARISTIDIS MATSOUKIS[1] and KONSTANTINOS CHRONOPOULOS[2]*

[1]Department of Crop Science, Agricultural University of Athens,
Iera Odos 75, Athens 11855, Greece
[2]Department of Biotechnology, Agricultural University of Athens,
Iera Odos 75, Athens 11855, Greece

ABSTRACT

The efficiency of applying linear regression (LR) and artificial neural network (ANN) models to estimate inside air temperature (T) of a glasshouse ($37^0 48´20´´$N, $23^0 57´48´´$E), Lavreotiki, was investigated in the present work. The T data from an urban meteorological station (MS) at $37^0 58´55´´$N, $23^0 32´14´´$E, Athens, Attica, Greece, about 30 Km away from the glasshouse, were used as predictor variable, taking into account the actual time of measurement (ATM) and two hours earlier (ATM-2), depending on the case. Air temperature data were monitored in each examined area (glasshouse and MS) for four successive months (July-October) and averages on a two-hour basis were used for the aforementioned estimation. Results showed that ANN were better than LR models, considering their better performance as shown in the scatterplots of the distribution of observed versus estimated inside T data of the glasshouse, in terms of both higher coefficient of determination (R^2) and lower mean absolute error (MAE). The best ANN model (highest R^2 and lowest MAE) was achieved by using as predictor variables the T at ATM and the T at ATM-2 from MS. The findings of our study may be a first step towards the estimation of inside T of a glasshouse in Greece, from outside T data of a remote MS. Thus, the operation of the glasshouse could be improved noticeably.

Keywords: Air temperature, Estimation, Glasshouse, Linear regression,
Neural network models

INTRODUCTION

The growth of plants inside glasshouses is often necessary in order to create marketable plant products out of season. Glasshouses are entirely or partially closed constructions, in which there is manual or automatic control and regulation of the values of meteorological parameters for fulfilling the requirements of the cultivated plant species[1]. One of the key factors that impacts glasshouse plant production is inside air temperature (T) [2]. The estimation of this parameter has been reported to be of high importance to help growers to manage crop production and designers to improve the ventilation and heating systems[3].

Therefore, from both a theoretical and a practical point of view, a lot of attempts have been made to estimate the inside T of a glasshouse, presenting a variety of statistical models, for example, linear auto regressive models with external input (ARX) and auto regressive moving average models with external input[4], ARX combined with neural network architectures[5], partial least-square regression and back-propagation neural network[6], with higher or lower degree of accuracy.

One robust computational technique, the artificial neural network (ANN) model[7], can be used successfully for the estimation of inside T of a glasshouse, as it has been confirmed by several

studies, e.g. Salazar et al.[8] and Alipour and Loghavi[9]. This model coincides with a great potential of complex, non-linear and time-varying input-output mapping[10]. The use of ANNs for the estimation of inside T of glasshouses is limited, especially when using the outside T of a remote meteorological station (MS) as a predictor variable. To our knowledge, only some preliminary tests (unpublished data) have been made by the authors of the present work.

Our present work aims to investigate the hypothesis of satisfactory performance of ANNs, regarding the estimation of inside T of a glasshouse, based on outside T data of a remote MS. Moreover, the performance of selected linear regression (LR) models was evaluated and compared to that of the ANNs. In both cases, we used the scatterplots of the distribution of observed versus estimated inside T data, along with the coefficient of determination (R^2) and the mean absolute error (MAE) to evaluate the results.

MATERIALS AND METHODS

The field experiment was conducted in two places. The first place was a MS ($37^0 58' 55''$N, $23^0 32' 14''$E) in the highly populated urban region of the municipality of Athens, and the second one was a glasshouse ($37^0 48' 20''$N, $23^0 57' 48''$E) in the municipality of Lavreotiki, in the prefecture of Attica of southeast continental Greece. There was one examined site (S1) at the MS and one examined site (S2) inside the glasshouse, about 30 Km away from S1. The S2 site was located in a non-shaded plot, where there was an ornamental plant cultivation according to Matsoukis et al[11]. In brief, this plot had an open vertical side, with the other three vertical sides being covered by white opaque plastic sheets of polypropylene (model Velliflor of Vellis A.E. company, Greece). The same type of sheet was used to cover the ground surface for the prevention of weed emergence.

Air temperature data at S2 were monitored every 10 min, by a sensor (model 809 L 0-100, Wilh. Lambrecht, GmbH, Germany; accuracy ±0.3° at 0°) and recorded by a datalogger (model 903; Wilh. Lambrecht, GmbH, Germany). The sensor was placed at a height equal to the top of the plant canopies, for four successive months (July-October).

Air temperature averages were calculated on a two-hour basis. Simultaneously, and on the same time basis, T averages were calculated by the recorded T data of the MS[12], for the same period. The T averages of both glasshouse (S2) and MS (S1) were used for the estimation of the inside T of the glasshouse, with the aid of simple linear (SLR) and multiple linear regression (MLR) models, as well as ANN models, as determined by many preliminary tests. Finally, to estimate the inside T at S2, based on the data from S1, four models were distinguished from the others, in terms of higher R^2 and lower MAE. These models were named A, B, C and D. In model A, a SLR analysis was used, while in model B, a MLR analysis was adopted. Regarding models C and D, custom multilayer perceptrons (MLPs), which belong to the most commonly used ANN architectures[13], were used.

More specifically, the SLR analysis is defined by the equation:

$$y = \alpha + \beta x \qquad (1)$$

where y is the dependent variable, x the independent variable, and α, β the Y-axis intercept and the slope, respectively. In model A, the inside T of the glasshouse (dependent variable) was estimated, using the T of the MS, as independent variable, for the actual time of measurement (ATM). The MLR analysis is defined by the equation:

$$y = \alpha + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_n x_n \qquad (2)$$

where x_1, x_2,…,x_n: independent variables, α: Y-axis intercept and β_1, β_2,…,β_n: regression coefficients. Each regression coefficient represents the contribution of the respective independent variable to the prediction of the dependent variable. In model B, the T at site S2 (dependent variable) was estimated, using the T of S1 (first independent variable), at actual time of measurement (ATM), and the T of S1 two hours earlier (ATM-2), as second independent variable.

One of the most commonly used ANN models for T estimation, MLP, was chosen to be used for the present study. A major consideration when using MLPs for model building is the determination of the optimal architecture of the network, that is, the

number of inputs, number of layers and number of nodes per layer. To solve this problem, a trial-and-error method was used, the most common strategy to test many various alternative models, in order to keep the best performing network. It was found that for model C the best architecture was 1-6-1, that is, an input layer of 1 unit, (T of S1 at ATM) a hidden layer of 6 units and an output layer of 1 unit (estimated T of S2). In a similar way, for model D the best architecture was found to be 2-7-1, that is, an input layer of 2 units (T of S1 at ATM and T of S1 at ATM-2), a hidden layer of 7 units and an output layer of 1 unit (estimated T of S2, in this case). For both ANNs, the connections between the layers were feedforwarded and their weights and thresholds were determined by the training procedure of the neural network. The training set consisted of half of the data, the selection set of a quarter of the data and the test set of the remaining quarter of the data, randomly assigned[14]. After the proper training of the networks, we took into consideration only the test data set to determine the testing parameters (R^2, MAE) and compare the estimation models.

In order to evaluate the performance of the results obtained by LR and MLP models, two widely used criteria were used; the R^2 between observed and estimated T values at site S2 and the MAE of the estimated T values. The MAE is the average of the absolute errors after each model was applied and it is, side by side with R^2, a way to examine the overall efficiency of the models. The more efficient the model is, the higher the R^2 is and the lower the MAE is, which is the desirable goal. Examining the statistical significance of the parameters of LRs and MLPs, special attention needs to be given in the output P value. This value plays a major role in determining the parameters, which will be used in the various models so as to eliminate deficiencies. In the present study, it was ensured that the results were significant at $P \leq 0.05$.

RESULTS AND DISCUSSION

The results of the application of A, B, C and D models, in terms of scatterplots of the distribution of observed versus estimated T data for the site S2, are shown in Figures 1 and 2. The model with the worst performance was the simplest one, model A, derived from SLR analysis, due to the lowest R^2 and highest MAE (Figure 1a), compared to the other applied models. The model B, based on MLR analysis, showed a better performance than the previous one (A), and this improvement was justified by the higher R^2 and the lower MAE (Figure 1b). The additional input of the T of S1 at the time ATM-2 was the critical factor for this improvement. Similar results have been reported by Chronopoulos et al.[15] with regard to the estimation of T in a canyon in a National Forest in Greece. The model C, based on MLP with one input parameter (T of S1 at ATM), as it can be seen in Figure 2a, showed a slightly
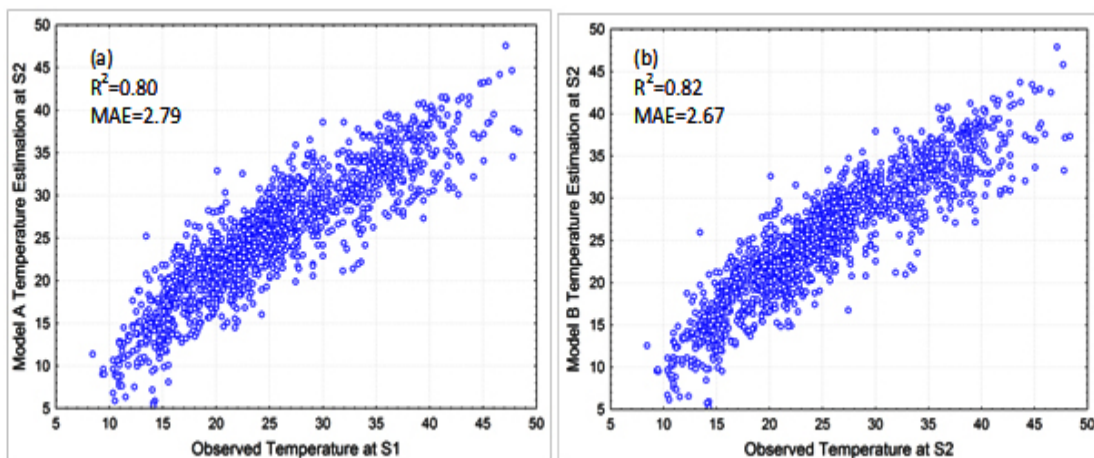

Fig. 1: Scatterplots of observed versus estimated air temperature (°) data for models A (a) and B (b). R^2 : determination coefficient, MAE: mean absolute error
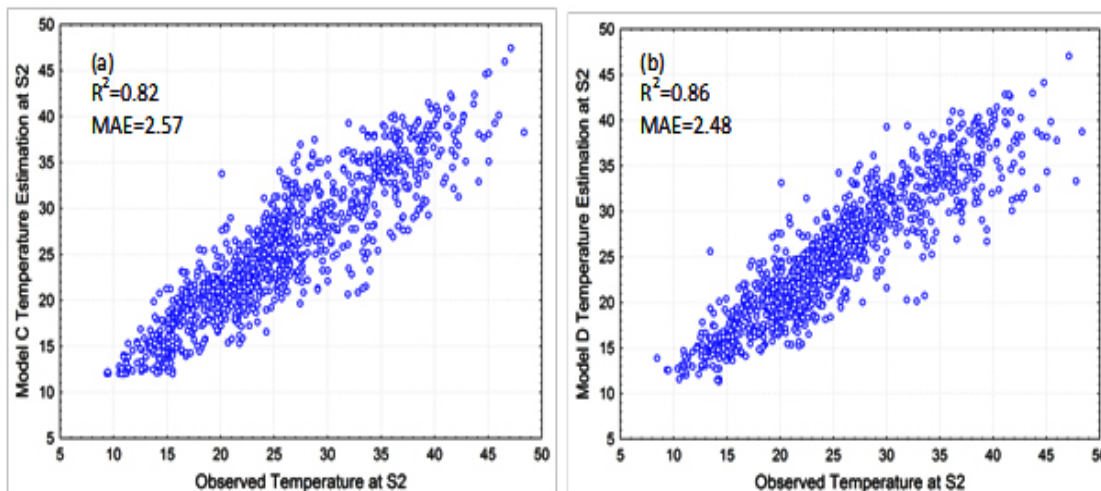
Fig. 2: Scatterplots of observed versus estimated air temperature (°) data for models C (a) and D (b). R^2 : determination coefficient, MAE: mean absolute error

better performance than model B, due to the lower MAE, although R^2 remained unchanged. The more sophisticated model, D, based on MLP method, in which the input parameters were both T at ATM and T at ATM-2 of S1, produced better results than any other model we tested, in terms of the highest R^2 and lowest MAE (Figure 2b). It has been reported that the introduction of ATM as an input in ANN models, produces better results in an urban area, concerning the estimation of T.[13]

The ability of the ANNs to take into account the nonlinear characteristics of the T data, produced better results than the LRs models, especially when we used combined input of the same data with a time lag of two hours. This clear improvement when using time lag can possibly be explained by the distance of the two sites, combined with the influence of the terrain profile. It should be noted that the terrain profile results in different sunrise and sunset times which vary more than one hour at the two examined sites.

In conclusion, the analysis of the results after the application of SLR, MLR and MLP models, with or without time lag, clearly showed that the MLP models were better than the SLR and MLR models, considering their better performance based on the scatterplots of the distribution of observed versus estimated inside T data for the glasshouse, the higher R^2 and the lower MAE. From the two examined MLP models, model D, in which the input parameters were T at ATM and T at ATM-2, had better performance based on the aforementioned scatterplots, the highest R^2 and the lowest MAE. Therefore, the examined MLP model with time lag of minus two hours could be beneficial, as a first step, for the estimation of inside T of a glasshouse in Greece, using the outside T of a remote MS. This estimation could be a valuable tool for more efficient operation of the glasshouse.

REFERENCES

1.   Chronopoulou-Sereli, A., Tsiros, I., Kamoutsis, A., Matsoukis, A., Droulia, F., Charalampopoulos, I. and Chronopoulos, K. General and Specific Topics of Bioclimatology, Applications-Exercises, *Ziti Publications,* *Thessaloniki, Greece*, 190 p (2012).

2.   Mesmoudi, K., Soudani, A. and Bournet, P. E. Determination of the inside air temperature of a greenhouse with tomato crop, under hot and arid climates. *Journal of Applied Sciences*

in Environmental Sanitation, **5**(2): 117-129 (2010).

3.   Wang, S. and Boulard, T. Predicting the microclimate in a naturally ventilated plastic house in a mediterranean climate. *Journal of Agricultural Engineering Research*, **75**(1): 27-38 (2000).

4.   Uchida Frausto, H., Pieters, J. G. and Deltour, J. M. Modelling greenhouse temperature by means of auto regressive models. *Biosystems Engineering*, **84**(2): 147-157 (2003).

5.   Uchida Frausto, H. and Pieters, J. G. Modelling greenhouse temperature using system identification by means of neural networks. *Neurocomputing*, **56**: 423-428 (2004).

6.   Li, X. F., Qin, L. L., Ma, G. Q. and Wu, G. Modeling greenhouse temperature by means of PLSR and BPNN. *Proceedings of the 35th Chinese Control Conference*: 2196-2200 (2016).

7.   Shank, D. B., Hoogenboom, G. and McClendon, R. W. Dewpoint temperature prediction using artificial neural networks. *Journal of Applied Meteorology and Climatology*, **47**(6): 1757-1769 (2008).

8.   Salazar, R., López, I. and Rojano, A. A neural network model to predict temperature and relative humidity in a greenhouse. *Acta Horticulturae*, **801**: 539-546 (2008).

9.   Alipour, M. and Loghavi, M. Development and evaluation of a comprehensive greenhouse climate control system using artificial neural network. *Universal Journal of Control and Automation*, **1**(1): 10-14 (2013).

10. Dibike, Y. B. and Coulibaly, P. Temporal neural networks for downscaling climate variability and extremes. *Neural Networks*, **19**: 135-144 (2006).

11. Matsoukis, A., Chronopoulou-Sereli, A. and Chronopoulos, I. Bioclimatic conditions in relation to shading in a glasshouse: The case study of *Lantana camara* L. cultivation in summer. *Current Agriculture Research Journal*, **4**(1): 47-53 (2016).

12. Chronopoulou-Sereli, A. Meteorological data of Votanikos area of Athens of the time period 1994 - 1998. Laboratory of Physics and Agricultural Meteorology, Agricultural University of Athens, Athens, Greece, (1999).

13. Kamoutsis, A. P., Matsoukis, A. S. and Chronopoulos, K. I. Air temperature estimation by using artificial neural network models in the greater Athens area, Greece. *ISRN Meteorology*: 1-7 (2013).

14. Chronopoulos, K., Kamoutsis, A., Matsoukis, A. and Manoli, A., An artificial neural network model application for the estimation of thermal comfort conditions in mountainous regions, Greece. *Atmósfera*, **25**(2): 171-181 (2012).

15. Chronopoulos, K., Tsiros, I., Dimopoulos, I., and Alvertos, N., An application of artificial neural network models to estimate air temperature data in areas with sparse network of meteorological stations. *Journal of Environmental Science and Health*, *Part A*, **43**(14): 1752-1757 (2008).