DOI: 10.5433/1679-0359.2020v41n6p2587

Modeling and analysis of greenhouse environmental factors in north China based on path analysis and stepwise regression

Modelagem e análise de fatores ambientais de estufa no norte da China com base na análise de caminho e regressão stepwise

Xu Zhang[1*]; Zhiyong Zhao[2]; Lijun Gao[2]; Biwu Ren[3]

Highlights:
This study uses statistical methods to analyze environmental factors in agricultural greenhouses.
The path analysis method yields an in-depth understanding of the direct and indirect effects of greenhouse environmental factors.
The model's accuracy is validated using stepwise regression modeling.

Abstract

To explore the relationship between environmental factors in a greenhouse on sunny/cloudy days, an environmental factor model was developed using path analysis and stepwise regression analysis. The environmental factors studied include greenhouse air temperature (GAT), greenhouse air humidity (GAH), soil temperature (ST), soil humidity (SH), greenhouse radiation (GR), and carbon dioxide concentration (CDC). The results showed that on a sunny day, the models can describe the GAT and GAH well (R^2=0.957, 0.936), and the model's tested determination coefficient was above 0.87. However, due to the delay and other main control factors of ST and SH, the models' determination coefficient was poor (R^2=0.587, 0.625). However, there was a fifth-order polynomial fitting relationship between ST and SH (R^2=0.817). On a cloudy day, the coupling effect between dependent variables and environmental factors was well described (R^2=0.97). The model test results for GAT and ST were better (R^2=0.997, 0.981), and the GAH and SH model test results were also good (R^2=0.789,0.882). In summary, the established coupling model of greenhouse environmental factors was suitable for simple greenhouse environment prediction, allowing greenhouse managers to easily predict greenhouse environmental change trends and reduce the cost of testing, laying a foundation for the subsequent establishment of a simpler, more accurate greenhouse factor model.
Key words: Greenhouse. Environmental factors. Path analysis. Stepwise regression analysis.

Resumo

Para explorar a relação entre fatores ambientais em casa de vegetação em dias ensolarados / nublados, foi obtido o modelo de fatores ambientais, utilizando análise de caminho e análise de regressão passo a passo. Os principais fatores ambientais incluem temperatura do ar da estufa (GAT), umidade do ar da estufa (GAH), temperatura do solo (ST), umidade do solo (SMC), radiação do efeito estufa (GR),

[1] Researcher, Dr., Mechanical and Electrical Engineering, Inner Mongolia Agricultural University, IMAU, Hohhot, China. E-mail: 1099499366@qq.com
[2] Students, Mechanical and Electrical Engineering, Inner Mongolia Agricultural University, IMAU, Hohhot, China. E-mail: 401787180@qq.com; 760530470@qq.com
[3] Prof. Dr., Solar Thermal Research Institute, STRI, Baotou, China. E-mail: reng_1965@126.com
[*] Author for correspondence

Received: May 01, 2020 - Approved: July 18, 2020

concentração de dióxido de carbono (CDC). Os resultados mostraram que: Em um dia ensolarado, os modelos puderam descrever o poço GAT e GAH (R^2=0.957, 0.936). O coeficiente de determinação do teste do modelo foi superior a 0.87. No entanto, devido ao atraso e outros fatores de controle principais de ST e SH, o coeficiente de determinação dos modelos foi ruim (R^2=0.587, 0.625). No entanto, verificou-se que havia uma relação de ajuste polinomial de quinta ordem entre ST vs SH (R^2=0.817). Em um dia nublado, o efeito de acoplamento entre variáveis dependentes e fatores ambientais foi bem descrito (R^2=0.97), o teste do modelo GAT e ST foi melhor (R^2=0.997,0.981), o teste GAH e SH também foi bom (R^2=0.789,0.882). Em resumo, o modelo de acoplamento dos fatores ambientais da estufa estabelecido foi adequado para a previsão simples do ambiente da estufa, facilitando para os gerentes da estufa prever a tendência das mudanças ambientais da estufa e reduzir o custo do teste, além de estabelecer as bases para o estabelecimento subsequente de um modelo de fator de efeito estufa mais preciso e simplificado.
Palavras-chave: Casa de vegetação. Fatores ambientais. Análise de trajetória. Análise de regressão gradual.

Introduction

The greenhouse is an important part of agricultural production in North China. However, in most greenhouse production, environmental factor interaction model theory is weakly applied, mainly because the models are complex and difficult to apply in practice. After years of research, foreign scholars have established greenhouse environmental factor models in various forms, such as the GDGCM (Mashonjowa, Ronsse, Mubvuma, Milford, & Pieters, 2013; Chayangira, 2012), KASPER (Zwart, 1996; Speetjens, Hemming, Wang, & Tsay, 2012), MIC GREEN (Singh, Singh, Lubana, & Singh, 2006), SimFreC (Dimokas, Katsoulas, Tchamitchiam, & Kittas, 2008), and SIMICROC models (Briceño-Medina, Ávila-Marroquín, & Jaimez-Arellano, 2011). Domestic scholars have established a solar greenhouse microclimate mathematical model by using the basic theories of thermodynamics, heat transfer and architectural lighting (Li, Wu, & Yu, 1994; Chen & Wang, 1996). By analyzing greenhouse radiation, ventilation, heat exchange, and other basic processes, a dynamic model of greenhouse air humidity was established (Jiang, Qin, & Shi, 2013). Some scholars selected the model of peak-fitting function to model the soil temperature and soil moisture (Ta, Wu, Ma, Chen, & Zhu, 2015; Sai, Ma, & Ta, 2019).

At present, many important achievements have been made in greenhouse simulation research; however, there are few studies of the relationship between environmental factors in a solar greenhouse without heating equipment in cold and dry areas. In this study, a typical solar greenhouse in a cold and dry area in northern China was taken as the research object. Using path analysis and stepwise regression analysis, this study established an environmental factor model for the greenhouse in different winter weather conditions. By fitting the model predicted and measured values, the model's feasibility was proven, providing a basis for subsequent greenhouse research and scientific management.

Materials and Methods

Test area overview

The experimental greenhouse was located in Baotou, China (40.657°N,109.84°E). The greenhouse faced south, with a length of 70 m from east to west, a span of 8.5 m from north to south, and a ridge height of 4.75. The 3.5 m high north wall was made of brick and earth. From February 1 to 25, 2019, data were collected at the center of the greenhouse. Soil temperature (humidity) was taken at the surface layer. The sunny day on February 4 and the cloudy day on February 9 were selected for data analysis, and the sunny day on February 6 and the cloudy day on February 18 were selected for testing the model.

Analysis method

Path analysis

Path analysis is used to further study the quantitative relationship between dependent variables and independent variables based on correlation analysis and regression analysis by using the path coefficient analysis method (Luo, & Cheng, 2011; Gong, Zhang, & Chao, 2011; Zhou, Ma, & Zhang, 2005). The formula is as follows:

$$\begin{bmatrix}\rho_1\\\rho_2\\\vdots\\\rho_p\end{bmatrix}=\begin{bmatrix}C_{11} & C_{12} & C_{13} & \cdots & C_{1p}\\C_{21} & C_{22} & C_{23} & \cdots & C_{2p}\\\vdots & \vdots & \vdots & \cdots & \vdots\\C_{p1} & C_{p2} & C_{p3} & \cdots & C_{pp}\end{bmatrix}\times\begin{bmatrix}r_{1y}\\r_{2y}\\\vdots\\r_{py}\end{bmatrix} \quad (1)$$

where is the direct path coefficient, is the inverse of the correlation matrix, and is the correlation coefficient. The direct path coefficient (DPC),, is often expressed as ; the indirect path coefficient (IPC), , can be calculated by the correlation coefficient and the direct path coefficient. .

Stepwise regression analysis

Stepwise regression analysis introduces variables into the model one by one. Then, a certain independent variable is added or eliminated from all available independent variables until the optimal regression equation is established. The F test is used for selection; when >F, we introduced the independent variable ; when≤F, we did not introduce the independent variable .

$$F_{1i}=\frac{V_i(x_1,x_2,\ldots x_l)}{SS_S(x_1,x_2,\ldots x_l,x_i)/(n-l-2)} \quad (2)$$

where n is the sample size, and l is the number of “introduced” independent variables. is the sum of the squares of the residual dispersion, and is the contribution of the added independent variable to the regression sum of squares.

Determination coefficient

The determination coefficient R^2 was selected as the evaluation standard for the model and the model test.

$$R^2=\frac{\sum_{i=1}^{N}(y_{pi}-\bar{y}_m)^2}{\sum_{i=1}^{N}(y_{mi}-\bar{y}_m)^2} \quad (3)$$

where is the measured value of model parameters, is the estimated value of model parameters, and is the average of the measured values.

Results and Discussion

To explore the relationship between factors, sunny and cloudy days were selected for analysis (a solar greenhouse receives solar radiation). To simplify the following statement, the environmental factor naming was simplified: X1 was for RA, X2 was for GAT, X3 was for GAH, X4 was for ST, X5 was for SH, X6 was for CDC.

Path coefficient analysis

In the path coefficient analysis, when DPC > TIPC = total indirect path coefficient), this factor had a direct effect on the dependent variable; when DPC < TIPC, the indirect effect was dominant, showing that the factor’s effect on the dependent variable was synergistic with other factors, and the effect of the single factor on the dependent variable was weak. DPC and IPC were used to determine the relationship factor, which played a decisive role on the dependent variable.

X2 as the dependent variable (Table 1): On a sunny day, the main factors affecting the change in X2 were X1, X3, and X6, among which X1 and X6 were mainly reflected in the direct effect. X1 was a positive effect, X6 was a negative effect, and X3 was mainly an indirect effect. On a cloudy day, the main factors affecting the change in X2 were X1, X3, X4, and X6, and the influence of each factor on X2 was indirect.

X3 as the dependent variable (Table 1): On a sunny day, the main factors affecting the change in X3 were X1, X2, and X5, among which X2 was mainly reflected in direct action, with a negative effect. The effect of X1 and X5 on X3 is indirect. On a cloudy day, the main factors affecting the change in X3 were X1, X2, X4, and X5, among which X2 and X5 were mainly reflected in direct action with negative effects. X1 and X4 were indirect, positive effects.

Table 1
Path analysis of environmental factors on sunny/cloudy day in the solar greenhouse

Item	Dependent variable	Environmental factor	DPC	IPC						
				X1	X2	X3	X4	X5	X6	TIPC
Sunny day	X2	X1	0.46	—	—	0.35	—	—	0.10	0.45
		X3	-0.38	-0.4	—	—	—	—	-0.14	-0.54
		X6	-0.38	-0.11	—	-0.14	—	—	—	-0.25
	X3	X1	-0.38	—	-0.50	—	—	-0.06	—	-0.56
		X2	-0.57	-0.33	—	—	—	-0.01	—	-0.34
		X5	-0.27	-0.18	-0.10	—	—	—	—	-0.28
	X4	X5	0.61	—	—	—	—	—	0.12	0.12
		X6	-0.27	—	—	—	—	-0.27		0.27
	X5	X2	0.73	—	—	-0.83	0.01	—	—	-0.732
		X3	0.90	—	-0.67	—	0.11	—	—	-0.56
		X4	0.52	—	0.01	0.20	—	—	—	0.21
Cloudy day	X2	X1	0.26	—	—	0.37	0.06	—	0.22	0.66
		X3	-0.36	-0.20	—	—	-0.12	—	-0.30	-0.62
		X4	0.16	0.08	—	0.35	—	—	0.27	0.58
		X6	-0.32	-0.18	—	-0.34	-0.13	—	—	-0.65
	X3	X1	0.17	—	-0.99	—	0.22	-0.17	—	-0.94
		X2	-1.22	0.14	—	—	0.48	-0.37	—	0.25
		X4	0.60	0.06	-0.99	—	—	-0.46	—	-1.39
		X5	-0.48	0.06	-0.98	—	0.57	—	—	-0.35
	X4	X1	-0.32	—	1.19	-0.75	—	0.25	—	0.69
		X2	1.46	-0.26	—	-0.96	—	0.57	—	-0.65
		X3	0.98	0.25	-1.44	—	—	-0.58	—	-1.77
		X5	0.71	-0.11	1.18	-0.81	—	—	—	0.26
	X5	X2	-0.94	—	—	1.00	0.75	—	—	1.75
		X3	-1.02	—	0.93	—	-0.73	—	—	0.2
		X4	0.92	—	-0.77	0.81	—	—	—	0.04

X4 as the dependent variable (Table 1): On a sunny day, the main factors affecting the change in X4 were X5 and X6, among which X5 was mainly reflected in direct action with a positive effect. X6 was mainly reflected in indirect action with a negative effect. On a cloudy day, the main factors affecting the change in X4 were X1, X2, X3, and X5, among which X2 and X5 were mainly reflected in direct action with a positive effect. X1 and X3 were indirect effects.

X5 as the dependent variable (Table 1): On a sunny day, the main factors affecting the change in X5 were X2, X3, and X4, among which X3 and X4 were mainly reflected in direct action, and X2 had an indirect effect. On the cloudy day, the path analysis results were similar to the sunny day.

Stepwise regression analysis of environmental factors in the solar greenhouse

On a sunny day, with X2 and X3 as the dependent variable (Table 2), the stepwise regression equation's determination coefficient was better (R^2=0.957, 0.936); however, when X4 and X5 were the dependent variables, the stepwise regression equation's determination coefficient was relatively poor (R^2=0.587, 0.625). When X4 was the dependent variable, and X1, X2 or X3 were added into the stepwise regression, the regression equation's determination coefficient was smaller than that of the determination coefficient containing only X5 and X6. However, when only X5 and X6 were included, the regression equation's determination coefficient remained poor. Therefore, on a sunny day, other factors and delays affected the changes in X4 and X5. Studies have shown that ST had a certain amount of delay on sunny days, with ST at its lowest at 11:00 and the surface temperature at its maximum at 17:30-18:00 (Sai et al., 2019). In January during the winter, the test study of Hohhot city solar greenhouse found that the maximum ST appeared at 15:00, and SH also reached its peak at that time (Ta et al., 2015).

Table 2
Stepwise regression analysis of environmental factors in the solar greenhouse

Item	Dependent variable (Y)	Stepwise regression	F Value	Statistical significance (P)	Determination coefficient R2
Sunny day	X2	Y=(7.6E-3)*X1-0.256*X3-0.01*X6+37.403	74.58	＜0.01**	0.957
	X3	Y=-(9.95e-3)*X1-0.83*X2+3.87*X5-35.22	48.52	＜0.01**	0.936
	X4	Y=3.29*X5-(2.4E-3)*X6-95.80	7.81	＜0.01**	0.587
	X5	Y=0.045*X2+0.038*X3+0.096*X4+27.85	5.54	＜0.01**	0.625
Cloudy day	X2	Y=(1.038E-2)*X1-0.345*X3+0.872*X4-(3.883E-3)*X6+30.485	443.29	＜0.01**	0.996
	X3	Y=(6.938E-3)*X1-1.252*X2+3.551*X4-14.281*X5+546.47	135.37	＜0.01**	0.984
	X4	Y=-(2.19E-3)*X1+0.253*X2+0.164*X3+3.506*X5-126.063	81.77	＜0.01**	0.976
	X5	Y=-(3.27E-2)*X2-(3.427E-2)*X3+0.185*X4+30.581	400.42	＜0.01**	0.971

On a cloudy day, each regression equation's determination coefficient was better (R^2=0.996, 0.984, 0.976, 0.971), and the regression model reached an extremely significant level, indicating that the coupling effect of greenhouse environmental factors on a cloudy day was stronger than for a sunny day.

Regression model validation

Data on a sunny day (February 6) and a cloudy day (February 18) were selected to verify the regression model. The dependent variable value was calculated by the regression model and defined as the predicted value. The predicted value and the experimental value were fitted, and the predicted results were judged by the data distribution on the 1:1 line.

On a sunny day, the predicted results for X2 and X3 by the stepwise regression model were better (R^2=0.882, 0.874), showing that the model has high prediction accuracy (Figure 1-A, Figure 1-B). However, the X4 and X5 model test results were worse (Figure 1-C, Figure 1-D). By analyzing the change trend of the predicted and measured values of X4 and X5 (Figure 1-E, Figure 1-F), although the deviation between the predicted and measured values was small, the predicted value and the measured value did not have the same change trend. Once again, on sunny days, the ST and SH could not be accurately predicted by the main environmental factors of the greenhouse in this study.

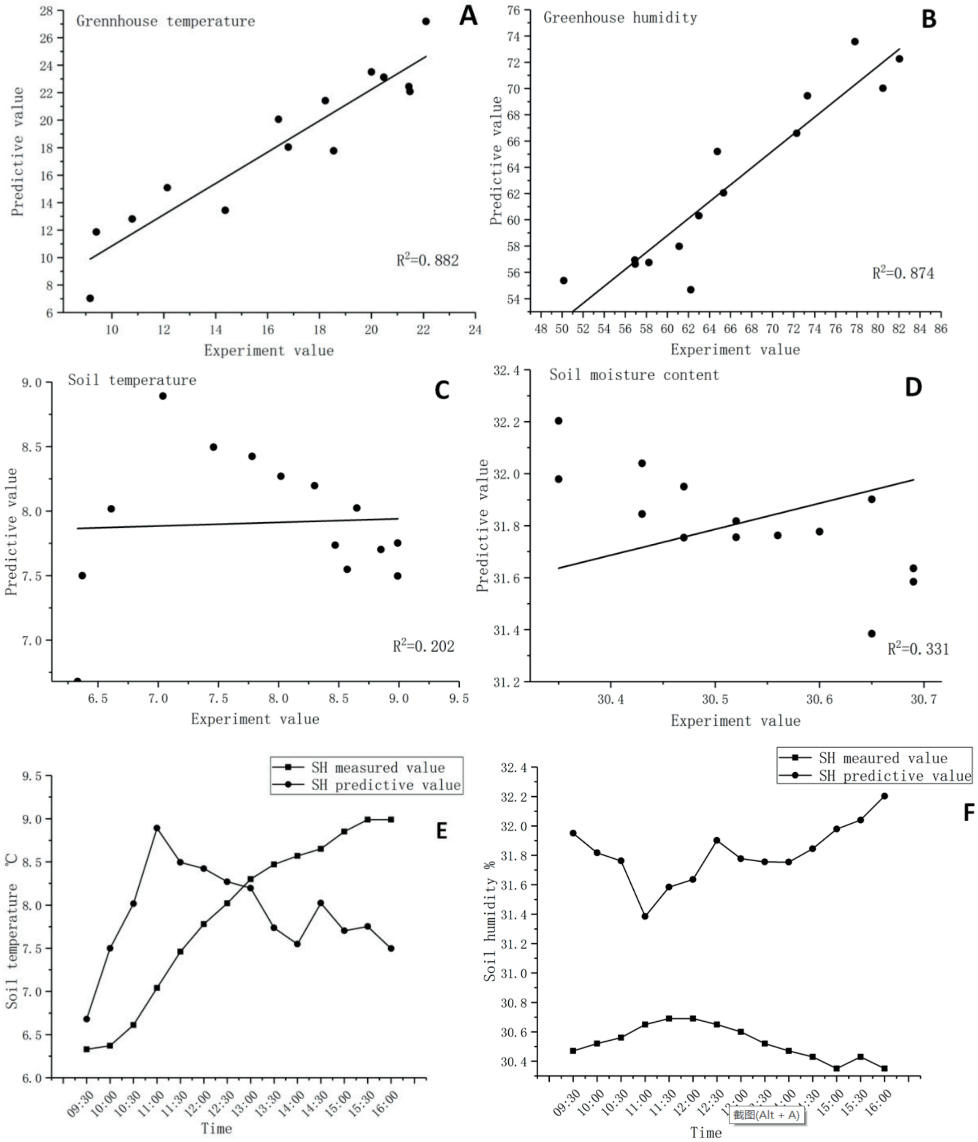

Figure 1. Fitting relationships between the predicted and measured values on a sunny day.
A: The predicted value and the measured value of the GAT were fitted;
B: The predicted value and the measured value of the GAH were fitted;
C: The predicted value and the measured value of the ST were fitted;
D: The predicted value and the measured value of the SH were fitted;
E: Change trend of ST test value and measured value;
F: Change trend of SH test value and measured value.

On a cloudy day, the prediction results for X2 and X4 from the stepwise regression model were superior (R^2=0.997, 0.981) (Figure 2-A, Figure 2-C). The prediction results for X3 and X5 from the stepwise regression model were good (R^2=0.789, 0.882) (Figure 2-B, Figure 2-D). These results indicate that the stepwise regression model is more accurate in describing dependent variables, for which the models can be applied.

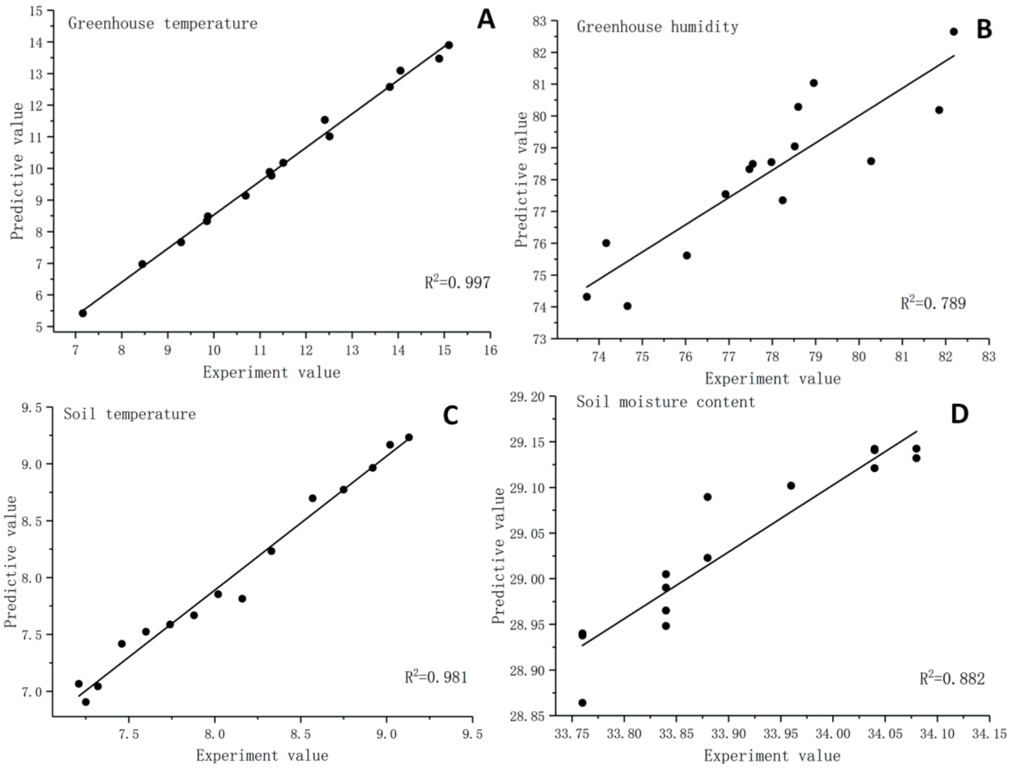

Figure 2. Fitting relationship between the predicted and measured values on a cloudy day
A: The predicted value and the measured value of the GAT were fitted;
B: The predicted value and the measured value of the GAH were fitted
C: The predicted value and the measured value of the ST were fitted;
D: The predicted value and the measured value of the SH were fitted.

Analysis of the relationship between soil temperature and soil humidity

Studies have shown that there is a good linear relationship between soil surface temperature and humidity in a winter greenhouse (Saiyin et al., 2019). Based on the analysis of X4 and X5, the fifth-order polynomial regression model with X5 as the dependent variable and X4 as the independent variable had a better fitting effect (R^2=0.817) (Figure 3-A). The regression model was as follows:

$$y = 174.077 * x - 46.584 * x^2 + 6.185 * x^3 - 0.407 * x^4 + 0.011 * x^5 - 226.898 \quad (4)$$

The data from February 6 (a sunny day) were selected to verify the model. Since the predicted and experimental values had the same time variable, the independent variable was presented in the form of a natural number (Figure 3-B). The predicted and experimental values had the same variation trend, and the maximum humidity error was less than 1%, proving that the established model is accurate.

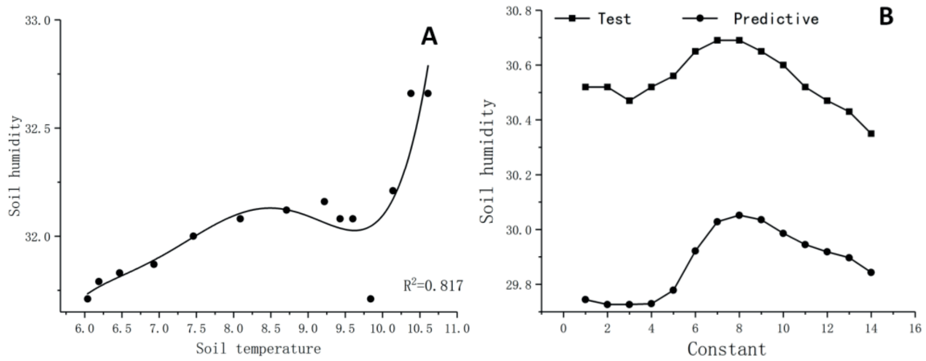

Figure 3. Regression fitting and testing of soil temperature and soil moisture on a sunny day.
A: Fitting effect of the fifth-order polynomial of ST vs SH;
B: Change trend of predicted and measured values of SH (with ST as the independent variable).

Shortcomings of this paper

Solar radiation is short in winter in northern China (7-8 hours). In this study, data was recorded every 30 min. In the follow-up study, to improve the model's accuracy, the data collection time could be shortened to record every 10 min, increasing the number of data samples. In follow-up studies, transpiration, water absorption by the root system, leaf heat dissipation and heat release from soil microbial reactions were considered to establish a more comprehensive stepwise regression model of soil temperature and soil moisture on a sunny day. In the regression model, the coefficients and constants of each variable have a certain error range. For a more accurate model, more data verification should be performed to reduce the influence of the error of each coefficient and constant, to make the model more universal and applicable.

Conclusions

This paper used the path analysis and stepwise regression analysis methods to model greenhouse environmental factors. The model verification showed that the model had a certain degree of credibility, which proved the feasibility of the analysis method in this study. This study provides a method for modeling greenhouse environmental factors in different regions of the world, thereby helping greenhouse operators predict the greenhouse environment and reduce test equipment investment.

Acknowledgments

This work was supported by Prof. Ren, and I would like to thank American Journal Experts (AJE) organization for checking the English writing of this paper.

References

Briceño-Medina, L. Y., Ávila-Marroquín, M. V., & Jaimez-Arellano, R. E. (2011). SIMICROC: Greenhouse microclimate simulation model. *Agrociencia*, *45*(7), 801-813. Recovered from https://xueshu.baidu. com/usercenter/paper/show?paperid=3911f580b4afc1949620fab9c7a08e02&site=xueshu_se&hitarticle=1

Chayangira, J. (2012). Effects of different ventilation strategies on the microclimate and transpiration of a rose crop in a greenhouse. *Greener Journal of Science, Engineering and Technological Research, 3*(5), 166-180. doi: 10.1029/2005TC001829

Chen, Q. Y., & Wang, Z. F. (1996). Dynamic simulation of the thermal environment in the energy-saving solar greenhouse. *Journal of China Agricultural University,*(1), 67-72. doi: CNKI:SUN:ZWZY.0.2019 96-05-013

Dimokas, G., Katsoulas, N, Tchamitchian, M., & Kittas, C. (2008). Validation of a climate simulator (Sim Gre C) inmediterranean greenhouses during winter period [J]. *Proceedings of the 4th International Conference on Information and Communication Technologies in Bio and Earth Sciences*, 9, 173-181. Recovered from www.researchgate.net/publication/251880917

Gong, X., Zhang, R. L., & Chao, H. X. (2011). Diurnal variation characteristics and influencing factors of photosynthesis and transpiration in 4 coconut cultivars. *Journal of Tropical Crops, 32*(2), 221-224. doi: CNKI:SUN: RDZX.0.2011-02-008

Jiang, Y. X., Qin, L. L., & Shi, C. (2013). Modeling of the modern greenhouse humidity system based on mechanism model. *Journal of Jiangnan University (Natural Science Edition), 12*(5), 535-540. doi: 10.3969/j.issn.1671-7147.2013.05.006

Li, Y. Z., Wu, D. R., & Yu, Z. (1994). Simulation and experimentalstudyofsolargreenhousemicroclimate[J] *Journal of Agricultural Engineering, 1*(19):130-136. doi: CNKI:SUN:NYGU.0.1994-01-019

Luo, Y. Z., & Cheng, Z. Y. (2011). Effects of water stress on water potential, transpiration rate, and stomatal conductance of alfalfa leaves. *Journal of Grassland Science, 19*(2), 215-221. doi: 10.11733/j.issn.1007-0435.2011.02.006

Mashonjowa, E., Ronsse, F., Mubvuma, M., Milford, J. R., & Pieters, J. G. (2013). Estimation of leaf wetness duration for greenhouse roses using a dynamic greenhouse climate model in Zimbabwe [J]. *Computers and Electronics in Ag-Riculture*, 95, 70-81.doi: 10.1016/j.compag.2013.04.007

Sai, Y. Z. G. T., Ma, W. J., & Ta, N. (2019). Effect of soil temperature on soil moisture content in a solar greenhouse, *Journal of China Agricultural University*, *24*(3), 102-108. doi: 10.11841/j.issn.1007-4333. 2019.03.13

Singh, G., Singh, P. P., Lubana, P. P. S., & Singh, K. G. (2006). Formulation and validation of a mathematical model of the microclimate of a greenhouse[J]. *Renewable Energy, 31*(10): 1541-1560. doi: 10.1016/j. renene. 2005.07.011

Speetjens, B., Hemming, S., Wang, D., & Tsay, J. R.(2012). *Design of a vegetable greenhouse system for subtropical conditions in Taiwan.* 1-60. Recovered from http://edepot.wur.nl/238571

Ta, N., Wu, S. L., Ma, W. J., Chen, B., & Zhu, Y. K. (2015). The peak-fitting method for soil temperature variation in a solar greenhouse under different water content is proposed. *Transactions of the Chinese Society of Agricultural Engineering, 30*(20), 204-210. doi: 10.3969/j.issn.1002-6819.2014.20.025

Zhou, X. X., Ma, Y. Q., & Zhang, Z. Q. (2005). Effects of environmental factors on photosynthetic and transpiration characteristics of pepper. *Journal of Plant Resources and Environment, 14*(4), 15-20. doi: CNKI:SUN: ZWZY.0.2005-04-003

Zwart, H. F. D. (1996). *Analyzing energy-saving options in greenhouse cultivation using a simulation model.* Doctoral thesis, Wageningen University, Wageningen, Netherlands. Recovered from https://www.r esearchgate.net/publication/40189453