Performance of classic multiple factor analysis and model fitting in crop modeling

Jiang Zhaohui*, Zhang Jing, Yang Chunhe, Rao Yuan, Li Shaowen

(*School of Information and Computer, Anhui Agricultural University, Hefei 230036, China*)

Abstract: Multivariate statistical analysis and regression, which are typical methods for crop modeling, have direct influence on the accuracy of model, but the applications of these methods usually depend on experiences. In this research, the performances of some common methods of statistical analysis and regression model were compared and verified, in order to avoid the blindness in crop modeling. The monitoring data of growth environment and photosynthesis of tomato, pumpkin and cucumber were obtained by PTM-48A. For the object variable of CO_2 exchange rate, selectivity on the main environmental factors by correlation analysis and path analysis were quantitatively compared. The performances of four kinds of multivariate binomial regression equations were compared using a comprehensive aggregative indicator, and the effectiveness of modeling was verified with the selected optimized multivariate statistical analysis and regression equation. Results showed that path analysis was more comprehensive and effective than correlation to discrimination of the variables, especially the path analysis ruled out some suspected independent variables which were not really independent, and the pure quadratic was more suitable to crop modeling because of its simple structure and high accuracy when the data set was small. The conclusion of this research has a general applicability, and offers a useful reference and guide for the other study and application of crops' modeling.

Keywords: crop model, multivariate statistical analysis, path analysis, regression, comparison

DOI: 10.3965/j.ijabe.20160902.1023

Citation: Jiang Z H, Zhang J, Yang C H, Rao Y, Li S W. Performance of classic multiple factor analysis and model fitting in crop modeling. Int J Agric & Biol Eng, 2016; 9(2): 119－126.

1 Introduction

Crop model was one of the core technologies supporting the Precision Agriculture, and it was also the essential part of the intelligent production[1]. Crop model was to obtain accurately of growth environmental factors and physiological parameters, and to indicate quantitatively the relationship of them[2]. The typical crop models are DSSAT (Decision Support System of Agricultural Technology Transfer), APSIM (Agricultural Production Systems sIMulator), CERES (Crop and Environment Research Synthesis), and so on[3]. Because of the difference of growth condition, location and variety, every crop model has its own adaptation and target. Based on the specific conditions, the universal model should be improved to satisfy the actual demand. The improved process not only demands the complete agriculture knowledge, but also needs high-level mathematics and computer resources. So the application of typical crop model is not applied widely, and the small sample size dedicated crop model for the actual demand is common in practice.

To build up a crop model, firstly the independent variables and dependent variables, which have large

Received date: 2013-12-24 **Accepted date:** 2015-10-17

Biographies: Zhang Jing, Master student, Research interests: modeling and simulation of biological system, Email: 805276133@qq.com; **Yang Chunhe,** Master student, Research interests: agricultural information detection and processing, Email: 951265920@qq.com; **Rao Yuan,** PhD, Associate Professor, Research interests: agricultural Internet of Things, Email: ry9925@gmail.com; **Li Shaowen,** PhD, Professor, Research interests: agricultural informatization, Email: shwli@ahau.edu.cn.

***Corresponding author: Jiang Zhaohui,** PhD, Professor, Research interests: agricultural information detection and processing. Anhui Agricultural University. Mailing address: No.130 West Changjiang Road, Hefei 230036, China. Tel: +8613966675580, Email: jiangzh@ahau.edu.cn.

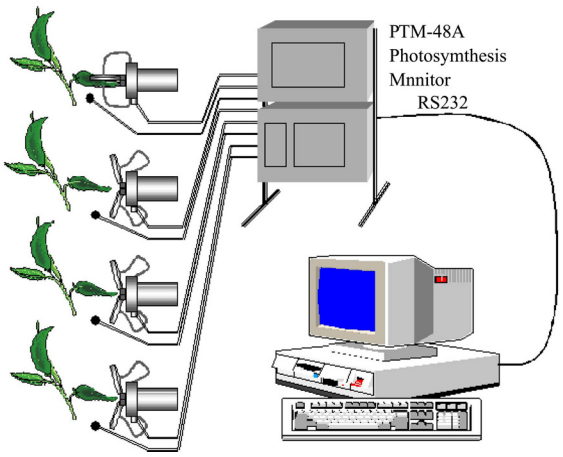
influence on the object variable of model, should be selected using methods of multivariate statistical analysis, using the professional data process tools like Excel, SPSS, and SAS. Then regression model or the other models should be used to build up the crop model. The practicality and accuracy of model should be verified, using the test data to predict the unknown variable. The common multivariate statistical analysis includes standard correlation analysis[4], principal components analysis[5], path analysis[6], and so on. Regression analysis is generally used in mathematical modeling. But the independent variables and form of equation are always chosen by experience. And empirical equations are often multitude and inconsistent[7]. The existing small sample size crop model is well-directed but has low adaptability. Although the model has high accuracy, the multivariate statistical analysis and regression model are selected according to the experience, instead of the quantification comparison results of the methods[8]. The empirical and blind selection becomes the hindrance of application and development of model[9], which is also the difficulty of agricultural informatics. The methods of multivariate statistical analysis and regression have been compared and verification in the previous study[10], but the experiment and expression of selecting the optimal method were not convincing enough.

In this research, methods of multivariate statistical analysis and regression were compared, using the preprocessed monitoring data of growth environment and photosynthesis of the tomato, pumpkin and cucumber. By the methods of correlation analysis and path analysis, the environmental characters impacting CO_2 exchange were picked out to put the analysis methods into comparison. The regression equations were found to compare the equations in respects of complexity and accuracy. The data of cucumber were used to verify the optimum multivariate statistical analysis and regression equation.

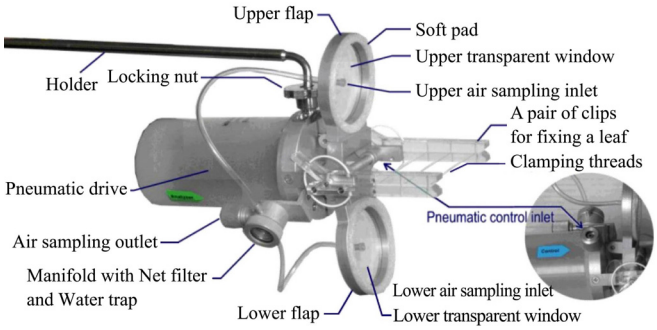
2 Materials and methods

2.1 Experimental data

In this research, three common outdoor vegetables, including tomato, pumpkin and cucumber, were continuously monitored for at least 24 hours, by using the PTM-48A photosynthesis monitor of B.F. Agritech Company[11,12], expecting to obtain the data of their photosynthetic physiological characteristic and growth environment. Figure 1 shows that the photosynthetic physiology and environmental factors online monitoring system includes five modules: the LC-4B leaf chambers, the multiple environmental factors sensors, the system console, the communication module and computer[13].



a. Schematic diagram of online monitoring system



b. Structure of LC-4B leaf chamber

Figure 1 Photosynthetic physiology and environmental factors online monitoring system

The system console controls the LC-4B leaf chamber open and close. If the LC-4B leaf chamber is closed, the system console will detect the LC-4B leaf chamber's concentration of carbon dioxide. After calculating, the system console obtains leaf's CO_2 exchange rate and transpiration rate. In addition, the system console is connected to the air temperature and humidity sensor, leaf's temperature and humidity sensor, soil moisture sensor, light radiation sensor and other environmental factor sensors. The data of real-time photosynthetic physiology and environmental parameters are stored in the system console. Through the RS232 serial port or

the wireless communication module, the system console transfers the data to a computer. The system console makes four LC-4B leaf chambers alternately opened and closed. Each closing time is 30 s, which is the detection time. Each opening time is one minute. During this time, the CO_2 absorber column will absorb the excess carbon dioxide in the console. When the four LC-4B leaf chambers' detections are completed, the system console will transfer the data to a computer. The monitoring time interval was uniform and set to 30 min, and the air flow was set to 0.84 LPM.

PTM-48A can obtained data of 31 variables for single monitor. The monitored data can be divided into two parts, the first part of which was the data of four leaves' physiological characteristic, and the rest was the ambient environmental parameters. There were eight environmental variables in all. We picked out the average value of four leaves' CO_2 exchange rates as the object variable, and the eight ambient environmental parameters as independent variables to study the influence of growth environment on the CO_2 exchange rate of crop. In order to decrease the amount of experimental work and study the physiological change of the whole crop, we took the average physiological value of four leaves as the physiological data of the whole crop. Eight environmental factors which were the independent variables, were summarized as follow: Air Temperature T_a (°C), Relative Humidity R_H (%RH), Radiation R_a (micromole/m^2·s), Absolute Humidity A_H (g/m^3), CO_2 concentration in the air C_{CO_2} (ppm), Vapour Pressure Deficit V_{pd} (kPa), Dew Point D_p (°C), Atmospheric Pressure P_a (mbar), defined as X_1-X_8. The CO_2 Exchange E (micromole/m^2·s) was defined as the dependent variable Y.

2.2 Analysis methods

The analysis and process of experimental data was as Figure 2.

At first, the monitoring data was preprocessed to convert the output table into the table which the processing software can read. Then Chauvenet's criterion and linear smoothing algorithm were used to check and process outlier. Forty-eight groups of data in 24 hours were chosen from the physiological and environmental data of the tomato and pumpkin as the experimental samples. The experimental samples were used for methods comparison, and the data of cucumber was for verification. The cucumber's data were divided into two parts, the first 48 groups of which were sampled data for modeling, and the rest were test data for predicting.

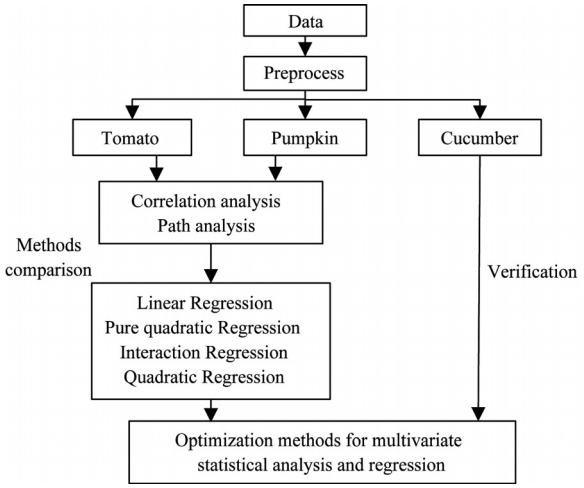


Figure 2 Flowchart of analysis and process

The preprocessed data were multivariate analyzed, using two methods, namely the correlation analysis and path analysis, for the purpose of getting the environmental factors that affect the rate of CO_2 exchange. In addition, the two methods were compared. In this research, the Person simple correlation coefficient[14,15] was adopted to indicate the relation between variables[16].

$$R_{ij} = \frac{C_{ij}}{\sqrt{C_{ii}C_{jj}}} \tag{1}$$

where, C_{ij}=cov(x_i, x_j) is the covariance of x_i and x_j. The effect of environmental factors on CO_2 exchange is expressed by the direct path coefficient ρ_i (i=1, 2, …, n), which is calculated as follow[17]:

$$\begin{bmatrix} \rho_1 \\ \rho_2 \\ \vdots \\ \rho_n \end{bmatrix} = \begin{bmatrix} c_{11} & c_{12} & c_{13} & \cdots & c_{1n} \\ c_{21} & c_{22} & c_{23} & \cdots & c_{2n} \\ \vdots & \vdots & \vdots & & \vdots \\ c_{n1} & c_{n2} & c_{n3} & \cdots & c_{nn} \end{bmatrix} \begin{bmatrix} r_{1y} \\ r_{2y} \\ \vdots \\ r_{ny} \end{bmatrix} \tag{2}$$

where, n is the number of independent variables; y is the dependent variable, r is their correlation matrix, and $c=r^{-1}$ was the invertible matrix of r. The indirect path coefficient $b_{ij} = r_{ij}\rho_j$ is the product of the direct path coefficient and correlation coefficient, indicating the

influence of independent variable x_i on dependent variable y through x_j.

The direct path coefficient ρ_i (i=1, 2, …, n) in Equation (2) was also the partial regression coefficient. And for N groups of data, the significance test statistics of path coefficient F was computed as Equation (3)[18].

$$F_i = \frac{\rho^2{}_i / c_{ii}}{\left(\sum_{i=1}^{n} \rho_i r_i\right) \Big/ (N-n-1)} \quad (3)$$

Regression analysis was carried out, regarding the analyzed environmental factors as independent variables and the CO_2 exchange as a dependent factor. After determining the independent and dependent variables, the suitable model should be selected. Then the coefficients could be calculated by the least square method, according to the historic statistics. The multivariate binomial regression includes linear, pure quadratic, interaction and quadratic.

Linear: $$y = \beta_0 + \beta_1 x_1 + \cdots + \beta_m x_m \quad (4)$$

Pure quadratic: $$y = \beta_0 + \beta_1 x_1 + \cdots + \beta_m x_m + \sum_{j=1}^{m} \beta_{jj} x_j^2 \quad (5)$$

Interaction:

$$y = \beta_0 + \beta_1 x_1 + \cdots + \beta_m x_m + \sum_{1 \le j < k \le m} \beta_{jk} x_j x_k \quad (6)$$

Quadratic: $$y = \beta_0 + \beta_1 x_1 + \cdots + \beta_m x_m + \sum_{1 \le j \le k \le m} \beta_{jk} x_j x_k \quad (7)$$

In order to pick out the optimal regression model, an aggregative indicator in Equation (8) was put forward, taking account of two aspects including accuracy and complexity comprehensively.

$$G = a_1 \tilde{RMSE} + a_2 \tilde{MAE} + a_3 \tilde{R^2} + a_4 \tilde{N_b} \quad (8)$$

where, $\tilde{RMSE}$, $\tilde{MAE}$, $\tilde{R^2}$ and $\tilde{N_b}$ are the data of $RMSE$, MAE, R^2 and N_b by normalization operation. $RMSE$, MAE, and R^2 are calculated as follows:

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{i=1}^{n} (\mathrm{OBS}_i - \mathrm{SIM}_i)^2}{n}} \quad (9)$$

$$MAE = \frac{\sum_{i=1}^{n} | OBS_i - SIM_i |}{n} \quad (10)$$

$$R^2 = \frac{SSR}{SST} = \frac{SSR}{SSE + SSR} = \frac{\sum_{i=1}^{n} ((\sum_{j=1}^{n} OBS_{ij} \Big/ n) - SIM_i)^2}{\sum_{i=1}^{n} (OBS_i - SIM_i)^2 + \sum_{i=1}^{n} ((\sum_{j=1}^{n} OBS_{ij} \Big/ n) - SIM_i)^2} \quad (11)$$

In Equations (9)-(11), n is the number of samples, OBS_i is the observed value, and SIM_i is the simulated value.

RMSE (Root Mean Square Error), is the standard error[19], and *MAE* (Mean Absolute Error), is the average of absolute errors. R^2 (R-square) is the coefficient of determination ranged from 0 to 1, and the nearer the R^2 approaches 1, the stronger explanatory ability was, and the better performance of the model was. N_b is the number of polynomial coefficients of each regression equation. *RMSE*, *MAE*, and R^2 can reflect the accuracy of model and complexity is smaller if N_b is smaller. To take both accuracy and complexity into consideration, set a_1, a_2, a_3, a_4 to $\frac{1}{6}$, $\frac{1}{6}$, $-\frac{1}{6}$, $\frac{1}{2}$. And if G is smaller, the model is better.

Using the optimal method of multivariate analysis, the modeling data of cucumber was analyzed and the regression model was developed. Practicability and accuracy was verified and the results were compared with that of other prediction models.

3 Results and discussion

3.1 Multivariate analysis of tomato and pumpkin

Eight environmental factors which were the independent variables, were summarized as follow: Air Temperature T_a, Relative Humidity R_H, Radiation R_a, Absolute Humidity A_H, CO_2 concentration in the air C_{CO_2}, Vapour Pressure Deficit Vpd, Dew Point D_p, Atmospheric Pressure P_a, defined as X_1-X_8. The CO_2 Exchange rate E is defined as the dependent variable Y. In Table 1, the Person correlation coefficients between CO_2 exchange and its impact factors were calculated from Equation (1). In Table 1, the upper halves of cells are the correlation coefficients of factors in tomato, and the lower halves are those in pumpkin.

Table 1 Correlation of CO_2 exchange and environmental factors in tomato (T) and pumpkin (P)

		X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8
X_2	T	−0.840							
	P	−0.756							
X_3	T	0.803	−0.918						
	P	0.794	−0.914						
X_4	T	−0.236	0.689	−0.564					
	P	0.896	−0.418	0.527					
X_5	T	−0.882	0.712	−0.658	0.070				
	P	−0.833	0.723	−0.619	−0.668				
X_6	T	0.861	−0.996	0.930	−0.674	−0.707			
	P	0.850	−0.972	0.918	0.583	−0.767			
X_7	T	−0.365	0.808	−0.728	0.932	0.237	−0.786		
	P	0.877	−0.349	0.474	0.977	−0.654	0.502		
X_8	T	−0.284	0.028	−0.013	−0.268	0.196	−0.047	−0.240	
	P	0.055	−0.256	0.444	−0.111	0.245	0.179	−0.091	
Y	T	0.868**	−0.844**	0.923**	−0.296*	−0.792**	0.851**	−0.525**	−0.136
	P	0.782**	−0.817**	0.942**	0.546**	−0.486**	0.819**	0.531**	0.573**

Note: * $p<0.05$; ** $p<0.01$.

As shown in Table 1, there are six environmental factors greatly affecting the E of tomato, including T_a, R_H, R_a, C_{CO_2}, V_{pd} and D_p. Besides of the five factors, A_H also expressed a highly significant correlation with CO_2 exchange. While all environmental factors showed greatly significant correlation with E in pumpkin. The variables such as T_a, R_H, A_H, V_{pd} and D_p were chosen as independent variables, but these variables were not really independent, and that might lead to unnecessary complication of the models. Results showed that correlation analysis could eliminate the irrelevant variables, but could not remove the variables with dependence relation.

Considering that the variables of the regression had to be really independent from each other, path analysis was carried out. Estimates of direct and indirect path coefficient were presented in Table 2. The elements in the last column were the significances of variables in tomato and pumpkin. The bold diagonal elements were direct path coefficients, and the rest were indirect path coefficients.

Table 2 Path analysis of CO_2 exchange and environmental factors in tomato (T) and pumpkin (P)

		X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	Sig.
X_1	T	**2.340**	0.292	0.696	−0.058	0.023	−3.067	0.649	−0.007	***
	P	**−1.598**	1.263	0.428	−0.575	−0.390	−0.275	1.925	0.004	
X_2	T	−1.966	**−0.348**	−0.796	0.171	−0.018	3.548	−1.434	0.001	
	P	1.208	**−1.670**	−0.493	0.268	0.338	0.314	−0.765	−0.017	**
X_3	T	1.879	0.320	**0.867**	−0.140	0.017	−3.311	1.292	−0.000	***
	P	−1.268	1.527	**0.540**	−0.338	−0.289	−0.297	1.040	0.029	***
X_4	T	−0.552	−0.240	−0.489	**0.248**	−0.002	2.401	−1.656	−0.007	
	P	−1.432	0.697	0.285	**−0.642**	−0.313	−0.188	2.146	−0.007	***
X_5	T	−2.065	−0.248	−0.571	0.017	**−0.026**	2.517	−0.422	0.005	
	P	1.330	−1.208	−0.334	0.429	**0.468**	0.248	−1.435	0.016	***
X_6	T	2.015	0.347	0.806	−0.167	0.018	**−3.561**	1.395	−0.001	***
	P	−1.358	1.623	0.495	−0.374	−0.359	**−0.324**	1.103	0.012	
X_7	T	−0.855	−0.281	−0.631	0.231	−0.006	2.798	**−1.776**	−0.006	***
	P	−1.401	0.582	0.256	−0.627	−0.306	−0.163	**2.196**	−0.006	
X_8	T	−0.664	−0.010	−0.011	−0.067	−0.005	0.168	0.427	**0.026**	
	P	−0.088	0.428	0.239	0.071	0.114	−0.058	−0.199	**0.065**	

Note: * $p<0.1$; ** $p<0.05$; *** $p<0.01$.

From the information in Table 2, the results of path analysis in tomato showed that T_a, R_a, V_{pd} and D_p had direct effect on E, and the significances were also great. While R_H, R_a, A_H, C_{CO2}, showed notability influence in pumpkin.

Therefore, path analysis could not only reflect the relation between independent variables and dependent variable, but also eliminate the variables with dependence relation. The path analysis could help to further identify the unwanted independent variables caused by couplings, measuring error, and accidental factors, especially when sample data was small.

3.2 Regression modeling

According to the results of multivariate analysis, impact factors were picked out as the independent variable. The multivariate regression equation was found using these parameters and E. The comparison of four models' performance on complexity and accuracy was shown in Table 3. Here the aggregative indicator calculated by Equation (8) was used to measure performance of each model.

From Table 3, the linear model was the simplest model, but its accuracy was the worst, because of its largest error. Although the accuracy of full quadratic model was the best, this model had the most coefficients. The accuracy of pure quadratic, interaction and full quadratic models were similar, and the pure quadratic model had the substantially lower complexity than others. Overall, pure quadratic was the most optimal model among the four models because that its aggregative indicator was the smallest.

Table 3 Comparison of four binomial regression models

			Linear	Pure quadratic	Interaction	Quadratic
Accuracy	*RMSE*	T	0.1169126	0.0709691	0.0703446	0.0715963
		P	0.195575	0.084806	0.087762	0.083788
	MAE	T	0.4092002	0.2265184	0.2250922	0.2162365
		P	1.601308	0.727856	0.714807	0.669218
	R^2	T	0.9629535	0.9876189	0.9884596	0.9893377
		P	0.913914	0.985319	0.985084	0.987874
Complexity	Coefficient number	T	5	9	11	15
		P	5	9	11	15
Aggregative Indicator G		T	0	−0.1994837	0.129129	0.026878
		P	0	−0.18751	0.117755	0

3.3 Modeling and verification

Since the experimental results of tomato and pumpkin showed that path analysis and pure quadratic were the optimal methods of analysis and modeling, we use the data of cucumber to verify the conclusion. The monitored data of cucumber were divided into two parts, the first 48 groups of which were sampled data for modeling, and the rest were test data. Using the modeling data of cucumber, we used path analysis to analyze the influence among the variables. The result was shown in Table 4.

Table 4 Path analysis of CO_2 exchange and environmental factors in cucumber

	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	Sig.
X_1	**6.627865**	−1.50186	0.655671	0.395317	−0.07721	−1.41846	−3.83509	0.039327	*
X_2	−6.01264	**1.655532**	−0.54668	−0.34477	0.113546	1.402925	3.043559	−0.07814	
X_3	5.817003	−1.21147	**0.747068**	0.350636	−0.04696	−1.22341	−3.51417	0.031735	***
X_4	6.536008	−1.42384	0.653445	**0.400873**	−0.07374	−1.38223	−3.86769	0.039918	
X_5	−3.6023	1.323275	−0.24695	−0.2081	**0.142055**	0.977725	1.338523	−0.1136	
X_6	6.420502	−1.58617	0.624183	0.378412	−0.09485	**−1.46427**	−3.49804	0.055153	**
X_7	6.393759	−1.26744	0.660375	0.390001	−0.04783	−1.28841	**−3.97551**	0.012519	*
X_8	1.870796	−0.92853	0.170165	0.114852	−0.11583	−0.57964	−0.3572	**0.139326**	

Note: * p<0.1; ** p<0.05; *** p<0.01.

From Table 4, R_a had great impact on CO_2 exchange, V_{pd} had impact, and both T_a and D_p. had little impact. Then the environmental characters related to CO_2 exchange, containing T_a, R_a, V_{pd}, and D_p were obtained by the method of path analysis. And the results were consistent with that in tomato. On the basis of that, we found four regression equations, and the comparison of four equations' performance on complexity and accuracy was in Table 5.

Table 5 Comparison of four binomial regression equations

		Linear	Pure quadratic	Interaction	Quadratic
Accuracy	RMSE	0.165699	0.128066	0.132527	0.104023
	MAE	0.62114	0.401531	0.424407	0.310404
	R^2	0.929638	0.961879	0.961271	0.978719
Complexity	Coefficient number	5	9	11	15
Aggregative indicator		0	−0.08667	−0.02451	0

The results in Table 5 show consistency with those in Table 3. The linear equation's accuracy was the worst. And full quadratic equation was so complex that it not suitable for modeling. Besides, the errors of pure quadratic and interaction equations were similar. But the pure quadratic equation was simpler than interaction equation. The aggregative indicator of pure quadratic equation was the smallest, and that indicated the model showed the best comprehensive performance.

Otherwise substituting the predicted data of cucumber into the fitting regression equations, the ordinate was CO_2 exchange in terms of time. The comparison of predicted results of four regression equations was shown in Figure 3.

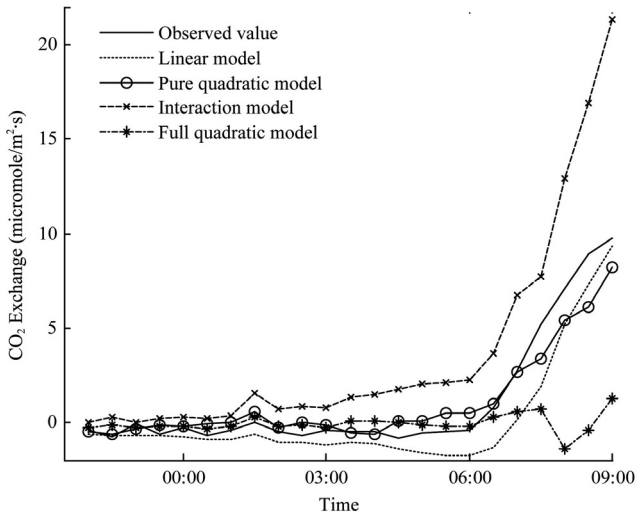


Figure 3 Comparison of predicted results of four multivariate regression models

The results shown in Figure 3 indicated that the error of pure quadratic model was obviously the smallest. And the next best model was linear model. In addition, the accuracies of interaction and full quadratic models were worse, compared with the linear and pure quadratic model. The error of interaction model increased rapidly after 6:00 am, and there was the fluctuation in the curve of full quadratic model after 3:00. The CO_2 exchange rate of photosynthesis was much larger than that of the respiration, especially if the radiation was enough. So the absolute error rose significantly in the morning, while the relative error might vary more smoothly. What more, the accuracy would usually be high, if the monitoring time of testing data was close towards that of the sample data. There were over 10 hours from the last 22:00 to 9:00. The period was so long that interaction and full quadratic models could not predict very accurately.

4 Conclusions

This research took tomato, pumpkin and cucumber as examples to study the optimal methods of multivariate statistical analysis and regression, to avoid the blindness in crop modeling. According to the analyzed results of tomato and pumpkin, path analysis was more suitable to for multivariate statistical analysis than correlation analysis, because it could not only express comprehensively the inner influence between variables, but also eliminate the variables with dependence relation. In the analysis process of regression model, an aggregative indicator was put forward, taking account of two aspects including accuracy and complexity comprehensively. From the comparison between four multivariate binomial regression models, the pure quadratic model had advantages over the others in comprehensive performance. The results of cucumber verified that path analysis and pure quadratic model were suitable for statistical analysis and regression in crop modeling. However, the accuracy of the model was not good enough during the last three hours. That showed the methods could only be applied for short-term prediction if the sample was little.

All in all, through the comparison of common statistical analysis and regression model, we verified the better performance of path analysis and pure quadratic model. The research could offer reference for crop modeling with small sample size, by the comparison of methods of multivariate analysis and regression. The results can not only be used in modeling of CO_2 exchange rate, but also extend to the modeling of other ecological and physiological characteristics in other crops.

Acknowledgement

This research was sponsored by Natural Science Foundation of Anhui Province (1508085MF110 & 1608085QF126), Key Programs for Science and Technology of Anhui Province (1501031102), Open Foundation of Key Laboratory in Application and Integration of Internet of Things (IOT) in Agriculture of

Ministry of Agriculture (2015-kf01), and International S&T Cooperation Project of Ministry of Agriculture (2015-Z44).

[References]

[1] Xie Z J, Cao W X, Luo W H. Application and prospects of crop growth simulation models in precision agriculture and intellectualized greenhouse in Shanghai. Acta Agriculturae Shanghai, 2001; 17(2): 17–21.

[2] He D J, He Y, Li M Z, Hong T S, Wang C H, Song S, et al. Research progress of information science-related problems in precision agriculture. Bulletin of National Natural Science Foundation of China, 2011; 1: 10–16.

[3] Nangia V, Ahmad M D, Du J T, Yan C R, Hoogenboom G, Mei X R, et al. Modeling the field-scale effects of conservation agriculture on land and water productivity of rainfed maize in the Yellow River Basin, China. Int J Agric & Biol Eng, 2010; 3(2): 5–17.

[4] Olaoye J O, Oni K C, Olaoye M M. Computer applications for selecting operating parameters of stationary grain crop thresher. Int J Agric & Biol Eng, 2010; 3(3): 8–18.

[5] Meng R F, Zhong J J, Zhang L F, Ye X Q, Liu D H. Ultrasonic concentration measurement of citrus pectin aqueous solutions using PC and PLS regression. Int J Agric & Biol Eng, 2012; 5(2): 76–81.

[6] Tiwari J K, Upadhyay D. Correlation and path-coefficient studies in tomato (*Lycopersicon esculentum Mill.*). Research Journal of Agricultural Sciences, 2011; 2(1): 63–68.

[7] Li Y Y. Application of BP neural network for forecasting parameters used in laser working computer simulation. Journal of Guangdong University of Technology, 2000; 17(3): 26–30. (in Chinese with English abstract)

[8] Luo Y, Guo W. Development and problems of crop models. Transactions of the CSAE, 2008; 24(5): 307–312. (in Chinese with English abstract)

[9] Zhang J, Jiang Z H, Wang C S, Yang C H. Modeling and prediction of CO_2 exchange response to environment for small sample size in cucumber. Computers and Electronics in Agriculture, 2014; 108: 39–45.

[10] Jiang Z H, Zhang J, Yang C H, Yao Y, Li S W. Comparison and verification of methods for multivariate statistical analysis and regression in crop modelling. 2015 International Conference on Electrical, Automation and Mechanical Engineering (EAME2015). 2015; pp.491–496. doi:10.2991/eame-15.2015.163

[11] Phyto-Sensor Group. http://www.phyto-sensor.com/

[12] BF Agritech - IT Systems. http://bf-ag.co.il/ITSystems.asp

[13] Jiang Z H, Wang C S, Zhang J, Yue Y, Li S W. Online monitoring and analysis of plant photosynthetic physiology and environmental factors. Applied Mechanics and Materials, 2013; 241: 75–80. doi: 10.4028/www.scientific.net/AMM.241-244. 75

[14] Aminrad Z, Zakariya S Z B S, Hadi A S, Sakari M. Relationship between awareness, knowledge and attitudes towards environmental education among secondary school students in Malaysia. World Applied Sciences Journal, 2013; 22(9): 1326–1333.

[15] Yao Y Z, Li J M, Zhang R, Sun S J, Chen K L. Greenhouse tomato transpiration and its affecting factors: Correlation analysis and model simulation. Chinese Journal of Applied Ecology, 2012; 23(7): 1869–1874. (in Chinese with English abstract)

[16] Moghadam F M, Ahmadi A, Keynia F. A new iris detection method based on cascaded neural network. Journal of Computer Sciences and Applications, 2013; 1(5): 80–84.

[17] Wang X S, Liu Z G, Liu H, Yang S J, Zhang X P, Meng Z J. Path Analysis and numerical simulation of MDS of tomato stem diameter. Transactions of the CSAM, 2012; 43(8): 187–192. (in Chinese with English abstract) doi: 10.6041/j.issn. 1000-1298. 2012. 08. 034.

[18] Ming D X. Path analysis-significance test. Journal of Sichuan Agricultural College, 1985; 3(1): 59–66. (in Chinese with English abstract)

[19] Jami M S, Husain I A F, Kabashi N A, Abdullah N. Multiple inputs artificial neural network model for the prediction of wastewater treatment plant performance. Australian Journal of Basic and Applied Sciences, 2012; 6(1): 62–69.