Modeling Safflower Seed Productivity in Dependence on Cultivation Technology by the Means of Multiple Linear Regression Model

Raisa Vozhehova[1], Mykhailo Fedorchuk[2], Serhii Kokovikhin[1], Pavlo Lykhovyd[1*], Vasyl Nesterchuk[1], Ivan Mrynskii[3], Olena Markovska[3]

[1] Institute of Irrigated Agriculture, Naddniprianske, 73483, Kherson, Ukraine

[2] Mykolaiv National Agrarian University, Heorhiia Honhadze 9 Street, 54000, Mykolaiv, Ukraine

[3] Kherson State Agrarian University, Stritenska 23 Street, 73006, Kherson, Ukraine

* Corresponding author's e-mail: pavel.likhovid@gmail.com

ABSTRACT

The results of the study devoted to the evaluation of reliability of the multiple linear regression model for safflower seed yields prediction were presented. Regression model reliability was assessed by the direct comparison of the modeled yields values with the true ones, which were obtained in the field trials with safflower during 2010–2012. The trials were dedicated to study of the effect of various cultivation technology treatments on the safflower seed productivity at the irrigated lands of the South of Ukraine. The agrotechnological factors, which were investigated in the experiments, include: A – soil tillage: A1 – disking at the depth of 14–16 cm; A2 – plowing at the depth of 20–22 cm; B – time of sowing: B1 – 3rd decade of March; B2 – 2nd decade of April; B3 – 3rd decade of April; C – inter-row spacing: C1 – 30 cm; C2- 45 cm; C3 – 60 cm; D – mineral fertilizers dose: D1 – N_0P_0; D2 – $N_{30}P_{30}$; D3 – $N_{60}P_{60}$; D4 – $N_{90}P_{90}$. Regression analysis allowed us to create a model of the crop productivity, which looks as follows: $Y = -1.3639 + 0.0213X_1 + 0.0017X_2 - 0.0121X_3 + 0.0045X_4$, where: Y is safflower seed yields, t ha^{-1}; X_1 – soil tillage depth, cm; X_2 – sum of the positive temperatures above 10°C; X_3 – inter-row spacing, cm; X_4 – mineral fertilizers dose, kg ha^{-1}. A direct comparison of the modeled safflower seed yield values with the true ones showed a very slight inaccuracy of the developed model. The maximum amplitude of the residuals averaged to 0.27 t ha^{-1}. Therefore, we conclude that multiple linear regression analysis can be successfully used in purposes of agricultural modeling.

Keywords: cultivation technology, prediction, statistical analysis, yields.

INTRODUCTION

Mathematical modeling is a widely used practice in almost all branches of modern science. One of the most popular mathematical methods of statistical analysis and development of simple models is linear regression, which finds application in solving diverse practical and theoretical tasks [Kutner et al. 2004; Neter et al. 1996]. Agricultural science is not an exception, and linear regression models are also successfully used for satisfying the needs in statistical data evaluation and forecasting [Mead 2017]. However, linear regression is nowadays considered to be an out-of-date and insufficiently accurate method of modeling the natural processes [Lykhovyd 2018]. Most scientists tend to use more modern and complicated methods of non-linear and spatial statistics, for example, artificial neural networks, multiple non-linear fuzzy regression analysis with improved calculations algorithm, etc. [Cheng, Lee 2001; Cross et al. 1995; Gelfand et al. 2010]. However, we should take into account that the above-mentioned methods often may not be available and understandable for everyone. Therefore, we decided to prove the efficiency of the linear regression analysis use in agricultural science on the example of modeling safflower seed productivity in dependence on the crop cultivation technology.

MATERIALS AND METHODS

Methodology of the field trials conduction

The field trials devoted to the investigation of safflower productivity in dependence on the cultivation technology treatments were carried out in the period from 2010 to 2012 at the experimental field of the Institute of Rice of the National Academy of Agrarian Sciences of Ukraine. The coordinates of the experimental field are: latitude 46°08′34″N, longitude 32°57′15″E, altitude is 8 m. The trials were carried out in accordance to the common recommendations on scientific work in agronomy [Ushkarenko et al. 2016] in four replications by using the randomized split plot design method. The study was devoted to the investigation of cultivation technology treatments on the safflower seed productivity, including:

- A – soil tillage: A1 – disking at the depth of 14–16 cm; A2 – plowing at the depth of 20–22 cm;
- B – time of sowing: B1 – 3rd decade of March; B2 – 2nd decade of April; B3 – 3rd decade of April;
- C – inter-row spacing: C1 – 30 cm; C2 – 45 cm; C3 – 60 cm;
- D – mineral fertilizers dose: D1 – N_0P_0; D2 – $N_{30}P_{30}$; D3 – $N_{60}P_{60}$; D4 – $N_{90}P_{90}$.

The cultivation technology of the crop was common for the irrigated conditions of the South of Ukraine excepting the studied factors. The previous crop was winter barley. Primary soil tillage was performed in accordance to the experimental design. Safflower cultivar Soniachnyi was sown by means of a seed drill at the depth of 5–6 cm. The inter-row spacing width was set according to the design of the trials. The crops were rolled instantly after sowing. Harrowing was performed before the sprouting stage of the crop, and then it was repeated at the stage of 2 leaves of the crop. Two inter-row cultivations were carried out on the plots with wide (60 cm) inter-row spacing. The irrigation of safflower in the trials was performed by using the frontal irrigation machine by maintaining the soil moisture at 75–80% level of the field water-holding capacity. The safflower seed yields were harvested by means of a self-propelled combine harvester "Sampo-130". The yields volumes were recorded at the standard moisture content in the seeds.

The climate of the zone, where the trials were carried out, is a coastal moderately continental one. It experiences a great influence of the nearly situated Black Sea. The weather conditions and meteorological indices were fixed at the local meteorological station installed directly on the experimental field of the Institute. The years of the study were characterized as follows: 2010 – extremely wet, 2011 – moderately dry, 2012 – extremely dry. The weather conditions during the studied period are represented in the Table 1.

Data processing

The multi-factor analysis of variance (ANOVA) of the crop yields data was performed by using the standard methodology within AgroStat add-on for Microsoft Excel software application [Kim 2014; Rosner 2006; Ushkarenko et al. 2016]. Statistical evaluation was performed for the reliability level of 95% ($p<0.05$). The safflower seed productivity was modeled by the results of the linear regression analysis, which was conducted by using the common calculations by the method of the least squares within Microsoft Excel software [Draper, Smith 2014; Gelfand et al. 2010; Seber, Lee 2012]. The model of safflower yields was developed as a common linear function $Y = b_0 + b_1X_1 + b_2X_2 + b_3X_3 + … + b_nX_n$. The accuracy and reliability of the developed regression model was checked by the direct comparison of the true crop yields values with the modeled ones.

RESULTS AND DISCUSSION

The mathematical processing of safflower yields data allowed us to determine the effect of different cultivation technology treatments on the crop yields, as well as define their coefficients of multiple and pair correlation, regression, and determination (Table 1). We should mention that we had calculated the sum of the positive temperatures for different times of sowing to enable expressing the above-mentioned factor in mathematical form for further statistical analysis. The statistical analysis proved the significant effect of the studied cultivation technology treatments on the crop yields (Table 2).

The results of the regression analysis showed the high strength of ties between the safflower seed productivity and cultivation technology

Table 1. Weather conditions during the period of the field trials with safflower

2010			2011			2012			Months
AT, °C	AH, %	PA, mm	AT, °C	AH, %	PA, mm	AT, °C	AH, %	PA, mm	
-3.0	85	33.0	-1.3	88	36.2	-0.3	87	62.9	Jan
2.0	84	0.0	-2.4	74	4.2	-6.7	85	18.3	Febr
3.4	79	14.6	2.6	78	17.3	2.7	79	29.0	Mar
10.7	70	11.0	9.9	73	38.9	12.8	77	12.7	Apr
17.6	66	77.3	16.7	77	47.3	20.8	74	58.2	May
22.5	65	69.5	22.2	71	68.8	23.8	65	12.3	Jun
24.7	63	44.9	25.2	71	11.0	26.5	58	13.3	Jul
26.1	62	44.0	23.2	65	12.3	24.4	62	9.0	Aug
17.7	68	64.1	19.3	68	7.1	20.1	70	0.1	Sept
7.8	76	36.2	10.7	77	19.3	15.8	77	17.7	Oct
10.5	86	43.0	3.1	76	6.0	7.7	85	8.4	Nov
1.6	88	68.9	4.4	89	27.5	0.4	86	13.5	Dec
11.8	74.3	506.5	11.1	75.6	295.9	12.3	75.4	255.4	Annual

Notes: AT – air temperature, AH – air humidity, PA – precipitation amounts.

Table 2. Safflower seed yields in t ha^{-1} depending on soil tillage, time of sowing, inter-row spacing and mineral fertilizers application doses (average for the studied period)

Inter-row spacing, cm (B)	Time of sowing (C)	Mineral fertilizers dose (D)				Factorial mean	
		N_0P_0	$N_{30}P_{30}$	$N_{60}P_{60}$	$N_{90}P_{90}$	C	B
Disking at the depth of 14–16 cm (A)							
30	early	1.48	1.74	1.96	1.88	1.77	1.60
	middle	1.29	1.65	1.85	1.89	1.67	
	late	1.03	1.31	1.58	1.53	1.36	
45	early	1.17	1.46	1.57	1.62	1.46	1.23
	middle	1.01	1.20	1.32	1.33	1.22	
	late	0.79	0.99	1.15	1.17	1.02	
60	early	1.01	1.26	1.37	1.40	1.26	1.05
	middle	0.86	1.03	1.13	1.15	1.04	
	late	0.67	0.82	0.96	0.98	0.86	
Factorial mean by D		1.04	1.27	1.43	1.44		
Plowing at the depth of 20–22 cm (A)							
30	early	1.63	1.89	2.02	2.11	1.91	1.73
	middle	1.48	1.73	1.89	1.92	1.76	
	late	1.22	1.48	1.71	1.72	1.53	
45	early	1.36	1.62	1.76	1.79	1.63	1.38
	middle	1.12	1.33	1.45	1.49	1.35	
	late	0.90	1.09	1.29	1.31	1.15	
60	early	1.13	1.39	1.50	1.53	1.39	1.16
	middle	0.93	1.14	1.25	1.27	1.15	
	late	0.73	0.91	1.05	1.09	0.94	
Factorial mean by D		1.17	1.40	1.55	1.58		
LSD$_{05}$ for the studied factors: A – 0.042; B – 0.027; C – 0.027; D – 0.061							

(Table 3). The coefficient of multiple correlation was 0.8277, and the coefficient of determination was 0.6851. However, the factors of soil tillage and inter-row spacing width had a very slight effect on the seed productivity of safflower, because their coefficients of determination were less than 0.1. Besides, it should be mentioned that inter-row spacing had negative value of the coefficient of correlation (-0.0566). This fact shows that an increase in the inter-row spacing width will have a negative impact on the crop productivity. The combination of positive temperatures (r – 0.6647)

Table 3. The results of regression analysis of the average safflower seed yields for the studied period depending on the cultivation technology treatments

Related X_i	R – multiple and r_i – pairs' coefficients of regression	D – general and d_i – partial coefficients of determination	b_0 and b_i – coefficients of regression
$X_1X_2X_3X_4$	0.8277	0.6851	-1.3639
X_1	0.1879	0.0353	0.0213
X_2	0.6647	0.4418	0.0017
X_3	-0.0566	0.0032	-0.0121
X_4	0.4525	0.2048	0.0045

Notes: X_1 – soil tillage depth, cm, X_2 – sum of the effective temperatures above 10°C, X_3 – inter-row spacing, cm, X_4 – mineral fertilizers dose, kg ha^{-1}.

and mineral fertilizers application (r – 0.4525) had the strongest effect on the crop productivity.

Using the results of regression analysis, the multiple linear regression model of safflower seed productivity depending on the studied inputs was created (1):

$$Y = -1.3639 + 0.0213X_1 + +0.0017X_2 - 0.0121X_3 + 0.0045X_4 \quad (1)$$

where: Y – safflower seed yields, t ha^{-1},
   X_1 – soil tillage depth, cm,
   X_2 – sum of the positive temperatures above 10°C,
   X_3 – inter-row spacing, cm,
   X_4 – mineral fertilizers dose, kg ha^{-1}.

The model shows that an increase in the depth of tillage on 1 cm, an increase of the sum of the positive temperatures on 1°C, and an increase of the application dose of mineral fertilizers on 1 kg ha^{-1} lead to respective increases of safflower seed yields on 21.3, 1.7, and 4.5 kg ha^{-1}. However, an increase of the inter-row spacing width on 1 cm causes a decrease in the crop productivity on 12.1 kg ha^{-1}.

The results of statistical data processing allowed us to determine the peculiarities in the influence of agrotechnological treatments on the safflower seed yields. It was established that the factor of the effective temperatures (X_2) with the share of 60.2% had the strongest influence on the crop productivity. The mineral fertilizers doses (X_4) can also be considered as a determinant factor for safflower yields, its share was 27.9%. At the same time, soil tillage (X_1) and inter-row spacing (X_3) are the factors of the least effect on the crop with their total share of only 5.2%, which is even lower than the share of occasional influence of other unaccounted in the study factors (6.7%).

A comparison of the true and modeled safflower yields showed sufficiently high relevance and accuracy of the developed linear regression model of the crop productivity (Table 4).

The amplitude of the seed productivity residuals averaged to -0.22...0.27 t ha^{-1}. This is a relatively small discrepancy between the crop yields. However, most of the trial variants obtained much lesser discrepancy between the modeled and true safflower productivity values (in the borders of 0.15 t ha^{-1}).

Therefore, the model showed quite a good reliability and accuracy, and it can be used for prediction of the crop yields in dependence on the cultivation technology treatments. We should mention that the model has limitations due to the fact that it was created for specific climatic and soil conditions. Thus, it might be successfully used only in modeling for the conditions of the South of Ukraine.

CONCLUSIONS

The developed multiple linear regression model of safflower seed yields depending on the cultivation technology treatments showed sufficient reliability and accuracy. The testing of the model gave us an opportunity to conclude that it is suitable to make an approximate forecast of the crop productivity in accordance to the cultivation technology parameters, such as soil tillage depth, inter-row spacing, and mineral fertilizers application doses, with taking into account terms of the crop sowing, which could be expressed in the sum of the effective temperatures needed for seed ripening.

Table 4. A comparison of the true and multiple linear regression predicted values of safflower seed yields depending on the cultivation technology treatments, t ha^{-1}

Soil tillage	Inter-row spacing	Time of sowing	Mineral fertilizers dose	True yields value	Modeled yield value	Residuals
Disking at the depth of 14–16 cm	30	early	N_0P_0	1.48	1.48	0.00
			$N_{30}P_{30}$	1.74	1.62	0.12
			$N_{60}P_{60}$	1.96	1.75	0.21
			$N_{90}P_{90}$	1.88	1.89	-0.01
		middle	N_0P_0	1.29	1.31	-0.02
			$N_{30}P_{30}$	1.65	1.45	0.20
			$N_{60}P_{60}$	1.85	1.58	0.27
			$N_{90}P_{90}$	1.89	1.72	0.17
		late	N_0P_0	1.03	1.14	-0.11
			$N_{30}P_{30}$	1.31	1.28	0.03
			$N_{60}P_{60}$	1.58	1.41	0.17
			$N_{90}P_{90}$	1.51	1.55	-0.04
	45	early	N_0P_0	1.17	1.30	-0.13
			$N_{30}P_{30}$	1.46	1.44	0.02
			$N_{60}P_{60}$	1.57	1.57	0.00
			$N_{90}P_{90}$	1.62	1.71	-0.09
		middle	N_0P_0	1.01	1.13	-0.12
			$N_{30}P_{30}$	1.20	1.27	-0.07
			$N_{60}P_{60}$	1.32	1.40	-0.08
			$N_{90}P_{90}$	1.33	1.54	-0.21
		late	N_0P_0	0.79	0.96	-0.17
			$N_{30}P_{30}$	0.99	1.10	-0.11
			$N_{60}P_{60}$	1.15	1.23	-0.08
			$N_{90}P_{90}$	1.17	1.37	-0.20
	60	early	N_0P_0	1.01	1.12	-0.11
			$N_{30}P_{30}$	1.26	1.25	0.01
			$N_{60}P_{60}$	1.37	1.39	-0.02
			$N_{90}P_{90}$	1.40	1.52	-0.12
		middle	N_0P_0	0.86	0.95	-0.09
			$N_{30}P_{30}$	1.03	1.08	-0.05
			$N_{60}P_{60}$	1.13	1.22	-0.09
			$N_{90}P_{90}$	1.15	1.35	-0.20
		late	N_0P_0	0.67	0.78	-0.11
			$N_{30}P_{30}$	0.82	0.91	-0.09
			$N_{60}P_{60}$	0.96	1.05	-0.19
			$N_{90}P_{90}$	0.98	1.18	-0.20
Plowing at the depth of 20–22 cm	30	early	N_0P_0	1.63	1.61	0.02
			$N_{30}P_{30}$	1.89	1.75	0.14
			$N_{60}P_{60}$	2.02	1.88	0.14
			$N_{90}P_{90}$	2.11	2.02	0.09
		middle	N_0P_0	1.48	1.44	0.04
			$N_{30}P_{30}$	1.73	1.58	0.15
			$N_{60}P_{60}$	1.89	1.71	0.18
			$N_{90}P_{90}$	1.92	1.85	0.07
Plowing at the depth of 20–22 cm	30	late	N_0P_0	1.22	1.27	-0.05
			$N_{30}P_{30}$	1.48	1.41	0.07
			$N_{60}P_{60}$	1.71	1.54	0.17
			$N_{90}P_{90}$	1.72	1.68	0.04
	45	early	N_0P_0	1.36	1.43	-0.07
			$N_{30}P_{30}$	1.62	1.56	0.06
			$N_{60}P_{60}$	1.76	1.70	0.06
			$N_{90}P_{90}$	1.79	1.83	-0.04
		middle	N_0P_0	1.12	1.26	-0.14
			$N_{30}P_{30}$	1.33	1.39	-0.06
			$N_{60}P_{60}$	1.45	1.53	-0.08
			$N_{90}P_{90}$	1.49	1.66	-0.17
		late	N_0P_0	0.90	1.09	-0.19
			$N_{30}P_{30}$	1.09	1.22	-0.13
			$N_{60}P_{60}$	1.29	1.36	-0.07
			$N_{90}P_{90}$	1.31	1.49	-0.18
	60	early	N_0P_0	1.13	1.25	-0.12
			$N_{30}P_{30}$	1.39	1.38	0.01
			$N_{60}P_{60}$	1.50	1.52	-0.02
			$N_{90}P_{90}$	1.53	1.65	-0.12
		middle	N_0P_0	0.93	1.08	-0.15
			$N_{30}P_{30}$	1.14	1.21	-0.07
			$N_{60}P_{60}$	1.25	1.35	-0.10
			$N_{90}P_{90}$	1.27	1.48	-0.21
		late	N_0P_0	0.73	0.91	-0.18
			$N_{30}P_{30}$	0.91	1.04	-0.15
			$N_{60}P_{60}$	1.05	1.18	-0.13
			$N_{90}P_{90}$	1.09	1.31	-0.22

REFERENCES

1. Cheng CB, Lee ES. 2001. Fuzzy regression with radial basis function network. Fuzzy Sets and Systems 119(2), 291–301.

2. Cross SS, Harrison RF, Kennedy RL. 1995. Introduction to neural networks. The Lancet 346(8982), 1075–1079.

3. Draper NR, Smith H. 2014. Applied regression analysis. John Wiley & Sons, New York City,

4. Gelfand AE, Diggle P, Guttorp P, Fuentes M. 2010. Handbook of spatial statistics. CRC Press,

5. Kim HY. 2014. Analysis of variance (ANOVA) comparing means of more than two groups. Restorative Dentistry & Endodontics 39(1), 74–77.

6. Kutner MH, Nachtsheim C, Neter J. 2004. Applied linear regression models. McGraw-Hill, Irwin,

7. Lykhovyd PV. 2018. Prediction of sweet corn yield depending on cultivation technology parameters by using linear regression and artificial neural network methods. Biosystems Diversity 26(1), 11–15.

8. Mead R. 2017. Statistical methods in agriculture and experimental biology. Chapman and Hall, CRC,

9. Montgomery DC, Peck EA, 2012. Vining GG. Introduction to linear regression analysis (Vol. 821). John Wiley & Sons, New York City,

10. Neter J, Kutner MH, Nachtsheim CJ, Wasserman W. 1996. Applied linear statistical models (Vol. 4). Chicago: Irwin, p. 318.

11. Rosner B. 2006. Fundamentals of biostatistics. Duxbury Press, Belmont CA.

12. Seber GA, Lee AJ. 2012. Linear regression analysis (Vol. 936). John Wiley & Sons, New York City.

13. Ushkarenko VO, Naidionova VO, Lazer PN. 2016, Scientific investigations in agronomy: the textbook. Kherson, Grin DS, pp. 314.