A Life Factor Approach to the Yield Prediction: a Comparison with a Technological Approach in Reliability and Accuracy

Pavlo Lykhovyd[1]

[1] Institute of Irrigated Agriculture, Naddniprianske, 73483, Kherson, Ukraine
   e-mail: pavel.likhovid@gmail.com

ABSTRACT

There are a number of various approaches to the development of yield predictive models in agriculture. One of the most popular ones is based on the yield modeling from the parameters of crop cultivation technology. However, there is another view on the yield prediction models, which is based on the use of life factors as yielding parameters. Our study is devoted to the comparison of a conventional technological approach to the yield prediction with a less prevalent approach of life factor based yield modeling. The testing of two approaches was performed by using the yielding data of sweet corn cultivated in the field trials under the drip-irrigated conditions of the Southern Ukraine, under the different technological treatments, viz. plowing depth, nutrition, and crop density. We developed two multiple linear regression models to compare their efficiency in the yielding predictions. One of the models used cultivation technology parameters as the inputs while the other used life factors as the inputs. Life factors were expressed in numeric values by using the following converter: total water consumption of the crop was used as the factor of water, the total sum of positive temperatures was used as the factor of heat, and the total sum of the main nutrients (NPK) available in the soil was used as the factor of nutrition. The results of the study proved an equal accuracy and reliability of the studied models of sweet corn yields, which is obvious from the values of RSQ. RSQ of the both studied regression models was 0.897. However, additional check of the modeling approaches applied in the feed-forward artificial neural network showed that the life factor based model with the RSQ value of 0.953 provided better yield predictions than the technologically based model with the RSQ value of 0.913. Therefore, we concluded that the life factor approach should be preferred to the technological approach in the development of yield predictive models for agriculture.

Keywords: artificial neural network, life factor, multiple linear regression, technological factor, yield modeling.

INTRODUCTION

Mathematical modeling is applied in various branches of modern scientific investigations. Agricultural sciences are not an exception. Particularly, different mathematical approaches are implemented in simulation of the processes, which take place in agricultural biosystems, and in making predictions. For example, special attention is paid to the yield predictive models. These models are useful for better understanding of the processes related to the crop productivity formation in particular environmental and agro-industrial conditions of its cultivation. Of course, these models also help scientists to select the most efficient cultivation treatments among the vast diversity of currently used in agricultural practice. Besides,

mathematical models lay in the basis of various decision support systems in agriculture; thus, they have to be reliable enough to provide producers with accurate results of simulations to help them take right steps and not fail [Jame, Cutforth 1996]. A number of scientific studies concerning different methods of mathematical modeling used for crop productivity predictions depending on different technological treatments have been conducted. For example, there are empirical mathematical models of crops productivity depending on fertilization [Cerrato, Blackner 1990], water use [Rennie, de Jong 1989], weed management [Doyle 1991], complex of technological treatments [Lavrenko et al. 2015], normalized difference vegetation index values [Kogan et al. 2013], climate and weather conditions [Balaghi

et al. 2008; Dixon et al. 1994], etc., and the studies related to the mathematical crop productivity predictions continue.

There are different approaches to the development of empirical yielding models in agriculture. The most popular one is based on the technological approach, when the output of the model (yield) is determined by the number of inputs, which are represented by the agrotechnological parameters. However, this approach is considered not to be a quite correct. In order to predict the productivity of any crop one should firstly take into account those factors, which have direct effects on the yielding capacity. These factors are not of technological nature. They are so-called factors of life, which directly determine the crops productivity. The technological factors have a mediated effect on the yield through the changing of availability and quantity of the life factors such as air, heat, water, light, and nutrition. Taking the above-mentioned statement into account, it must be more reasonable to develop the predictive models of crops productivity by using these life factors as inputs of the mathematical models.

Different mathematical models provide different yield predictions, which often might be in great discrepancy with each other [Lobell, Burke 2010]. The choice of the right type of mathematical model is very important, because it results in the accuracy and reliability of the productivity prediction, which is necessary to obtain positive results in the application of the modeling advice in real life situations [Cerrato, Blackner 1990]. Therefore, the goal of our study was to determine whether the technologically based approach to the yield prediction is better than the proposed one based on life factors.

MATERIALS AND METHODS

Testing and comparison of two approaches to the development of yield predictive models was performed by using the average yielding data of sweet corn ears obtained in the field trials during the period of 2014–2016. The trials were conducted at the drip-irrigated lands of the Southern Ukraine (the coordinates of the experimental plots were: latitude 46°43′42″N, longitude 32°17′38″E, 42 m above the sea level) by using the common methodology of agricultural investigations under the irrigated conditions

[Ushkarenko et al. 2014]. The field experiments were devoted to the improvement of sweet corn cultivation technology, and foresaw the study of the effects on the crop yields of such technological factors as: plowing depth (20–22, 28–30 cm), NP amounts applied (0, 60, 120 kg ha^{-1} of the active substance of mineral fertilizer), quantity of plants per ha (35000, 50000, 65000, 80000 plants ha^{-1}). The cultivation technology was common for growing sweet corn under the irrigated conditions of the Southern Ukraine excluding the realization of studied factors. The cultivar used in the trials was Ukrainian Brusnytsia (simple sweet corn of su type). After harvesting of the previous crop (winter wheat) soil disking and plowing was carried out. Further, in the spring period, harrowing was performed for several times, and two cultivator tillages (the last one was combined with application of the herbicide acetochlor, 900 g L^{-1} in the dose of 2.0 L ha^{-1}). The pre-sowing cultivator tillage was conducted at the depth of 5–6 cm. Sowing was carried out by the means of a seed drill with the inter-row spacing of 70 cm. The time of sowing varied by the years of study: 1st of May in 2014, 22nd of May in 2015, 21st of May in 2016. The crops were rolled instantly after sowing. Further care for the crops was in formation of the desired plants density, chemical treatments against weeds (with the complex herbicide containing foramsulfuron, 31.5 g L^{-1}, iodosulfuron, 1.0 g L^{-1}, tienecarbazon-methyl, 10 g L^{-1}, cyprosulfamide (antidote), 15 g L^{-1}) and insects (with chlorantraniliprole, 200 g L^{-1} in dose of 0.15 L ha^{-1}). The crop was drip irrigated to maintain the soil moisture of the root zone at the level of 80% of water-holding capacity. The average volume of irrigation water applied was 150 mm, and varied by the years of the study: 170 mm in 2014, 120 mm in 2015, 160 mm in 2016. Harvesting of sweet corn ears was conducted by hands at the stage of their technical ripeness. The time of harvesting varied by the years of study and variants of trials: from 15th to 25th of July in 2014; from 31st of July to 10th of August in 2015; from 1st to 7th of August in 2016.

The technological factors, which were used as the input in the models, were expressed by the numeric values of their quantity, namely: the absolute depth for plowing, amounts of the active substance applied for fertilizers, and quantity of sweet corn plants per ha during the vegetation, respectively.

Life factors, which were also used as the inputs in another model, obtained their expression by the following transformations:

1) Heat quantity was expressed in the sum of positive temperatures during the vegetative period of the crop with accordance to the duration of the latter. Weather conditions of the trials are given in the Table 1.

2) Water quantity was described by the total water consumption of the crop including the rainfall, irrigation water, and soil water used by the crops. The values were calculated with the method of Kostiakov water balance. Soil water amounts used by the crops were determined by the difference between the soil moisture at the beginning and in the end of the crop vegetation period by using the balance-drier method [Ushkarenko 1994; Ushkarenko et al. 2014].

3) Nutrition was expressed as the sum of the main nutrients (NPK) available in the soil layer of 30 cm in kg ha[-1] taking into account natural soil fertility, bulk density, and the doses of applied mineral fertilizers. The soil contained 35 mg kg[-1] of nitrogen (by Kornfield), 32 mg kg[-1] of phosphorus (by Machygin), and 430 mg kg[-1] of potassium (by Machygin) [Arinushkina 1970; Shkonde 1971]. These values were transformed into kg ha[-1] by taking into account the mass of 1 ha of the 0–30 cm soil layer [Smirnov, Muravin 1984]. The mass of 1 ha of the soil was calculated with accordance to the bulk density, which was different under the two variants of tillage and averaged to 1.270 t m[-3] under the plowing at 20–22 cm; 1.235 t m[-3] under the plowing at 28–30 cm. Bulk density of the soil was determined with the method of Kachynskii [Ushkarenko et al. 2014]. The amounts of available for plants in the first year nitrogen and phosphorus applied with mineral fertilizers were added to the amounts of the soil nutrients (Table 3).

The yielding data of sweet corn ears without husks were processed by using the common methodology of multiple linear regression analysis

Table 1. Weather conditions in the years of the field trials with sweet corn

Month	Decade	Air temperaturę, ºC	Relative air humidity, %	Precipitation, mm
2014				
May	I	13.7	75	33.0
	II	17.8	75	5.2
	III	22.2	61	0.0
June	I	22.4	64	13.3
	II	20.0	58	28.6
	III	20.0	64	22.5
July	I	23.5	53	0.0
	II	25.5	56	9.4
	III	26.1	49	10.0
2015				
May	III	19.6	69	70.7
June	I	21.3	61	7.1
	II	21.3	67	3.4
	III	20.0	73	27.8
July	I	22.8	74	84.9
	II	21.0	66	19.7
	III	26.0	67	0.0
August	I	26.0	49	0.0
2016				
May	III	18.5	77	20.7
June	I	17.8	70	16.2
	II	21.9	75	12.8
	III	26.5	62	14.0
July	I	22.4	61	21.6
	II	25.8	59	0.0
	III	25.0	54	24.7
August	I	26.0	55	0.6

Table 2. A comparison of the developed multiple linear regression models of sweet corn yields

Inputs of the model						True yields of sweet corn, t ha^{-1}	Outputs of the model			
Technological approach			Life factors approach				Technological approach		Life factors approach	
I_1 (plowing depth, cm)	I_2 (applied NP, kg ha^{-1})	I_3 (plants per ha)	I_1 (sum of T, ºC)	I_2 (amount of water used, mm)	I_3 (available NPK, kg ha^{-1})		Yields, t ha^{-1}	Residuals, t ha^{-1}	Yields, t ha^{-1}	Residuals, t ha^{-1}
20	0	35000	1567.4	258.3	1054	2.67	3.01	0.34	2.66	-0.01
20	0	50000	1584.7	261.6	1054	2.85	3.41	0.56	3.18	0.33
20	0	65000	1601.7	265.7	1054	3.01	3.81	0.80	3.90	0.89
20	0	80000	1636.2	266.8	1054	2.96	4.20	1.24	3.64	0.68
20	60	35000	1653.4	262.4	1105	5.56	5.63	0.07	4.76	-0.80
20	60	50000	1670.5	268.1	1105	6.31	6.02	-0.29	5.85	-0.46
20	60	65000	1713.9	271.4	1105	7.67	6.42	-1.25	5.98	-1.69
20	60	80000	1731.3	274.0	1105	6.80	6.82	0.02	6.33	-0.47
20	120	35000	1687.9	267.1	1156	7.53	8.24	0.71	7.77	0.24
20	120	50000	1713.9	270.7	1156	8.81	8.64	-0.17	8.24	-0.57
20	120	65000	1731.3	277.1	1156	10.93	9.04	-1.89	9.49	-1.44
20	120	80000	1774.7	277.6	1156	9.58	9.44	-0.15	8.95	-0.63
28	0	35000	1593.4	259.5	1053	3.00	2.23	-0.77	2.51	-0.49
28	0	50000	1602.0	262.3	1053	3.34	2.63	-0.71	3.04	-0.30
28	0	65000	1619.0	266.2	1053	3.57	3.03	-0.54	3.71	0.14
28	0	80000	1644.9	267.5	1053	3.37	3.43	0.06	3.63	0.26
28	60	35000	1662.0	263.3	1104	4.89	4.85	-0.04	4.79	-0.10
28	60	50000	1679.2	268.6	1104	5.55	5.25	-0.30	5.79	0.24
28	60	65000	1722.6	271.7	1104	6.25	5.64	-0.61	5.87	-0.38
28	60	80000	1748.6	274.8	1104	5.64	6.04	0.40	6.21	0.57
28	120	35000	1705.2	267.6	1155	6.23	7.46	1.23	7.58	1.35
28	120	50000	1731.3	271.4	1155	7.36	7.86	0.50	8.09	0.73
28	120	65000	1757.3	277.5	1155	8.59	8.26	-0.33	9.15	0.56
28	120	80000	1792.0	278.6	1155	7.56	8.66	1.10	8.88	1.32

(the method of the least squares) by the means of the built-in analysis tool of LibreOffice Calc 6 [Draper, Smith 2014]. As a result, we obtained the regression coefficients and slopes of the models, which enabled to express the yielding model in the form of a mathematical equation:

$$O = b_0 + b_1 I_1 + b_2 I_2 + b_3 I_3.$$

Feed-forward artificial neural network models were developed by using TiberiusXL add-in in the environment of Microsoft Excel 2019. TiberiusXL uses a supervised learning algorithm. The networks contained 3 inputs, 5 hidden neural nodes, and single output (yield). The learning process was performed in 1000 epochs with the learning rate of 1.00. TiberiusXL does not provide users with a model, which could be represented by the mathematical expression like in regression models. The results of the modeling process are given in the form of predicted output values and errors.

The accuracy of the developed models was determined by the values of RSQ and the amplitude of the residuals. The RSQ was calculated by using the formula 1:

$$RSQ = 1 - \frac{V(y \mid x)}{V(y)} \qquad (1)$$

where: $V(y \mid x)$ is the dispersion of the dependent argument [Devore 2011].

RESULTS

The calculations performed for creation of linear regression models allowed expressing sweet corn ears yield as the following function:
1) In the case of technological modeling:

$$O = 4.027 - 0.0972 I_1 + 0.0436 I_2 + 0.0265 I_3$$

2) In the case of life factors based modeling:

$$O = -84.998 - 0.0151 I_1 + 0.2371 I_2 + 0.0475 I_3$$

Table 3. A comparison of the developed artificial neural network models of sweet corn yields

Inputs of the model						True yields of sweet corn, t ha^{-1}	Outputs of the model			
Technological approach			Life factors approach				Technological approach		Life factors approach	
I_1 (plowing depth, cm)	I_2 (applied NP, kg ha^{-1})	I_3 (plants per ha)	I_1 (sum of T, °C)	I_2 (amount of water used, mm)	I_3 (available NPK, kg ha^{-1})		Yields, t ha^{-1}	Residuals, t ha^{-1}	Yields, t ha^{-1}	Residuals, t ha^{-1}
20	0	35000	1567.4	258.3	1054	2.67	2.72	0.05	2.73	0.06
20	0	50000	1584.7	261.6	1054	2.85	2.38	-0.47	2.73	-0.12
20	0	65000	1601.7	265.7	1054	3.01	3.19	0.18	2.97	-0.04
20	0	80000	1636.2	266.8	1054	2.96	3.48	0.52	3.11	0.15
20	60	35000	1653.4	262.4	1105	5.56	5.76	0.20	5.47	-0.09
20	60	50000	1670.5	268.1	1105	6.31	5.98	-0.33	5.76	-0.55
20	60	65000	1713.9	271.4	1105	7.67	7.55	-0.12	7.14	-0.53
20	60	80000	1731.3	274.0	1105	6.80	7.45	0.65	6.57	-0.23
20	120	35000	1687.9	267.1	1156	7.53	8.53	1.00	7.61	0.08
20	120	50000	1713.9	270.7	1156	8.81	9.22	0.41	9.22	0.41
20	120	65000	1731.3	277.1	1156	10.93	10.60	-0.33	11.06	0.13
20	120	80000	1774.7	277.6	1156	9.58	10.14	0.56	8.78	-0.80
28	0	35000	1593.4	259.5	1053	3.00	1.69	-1.31	2.59	-0.41
28	0	50000	1602.0	262.3	1053	3.34	1.22	-2.12	2.67	-0.67
28	0	65000	1619.0	266.2	1053	3.57	2.61	-0.96	3.08	-0.49
28	0	80000	1644.9	267.5	1053	3.37	3.29	-0.08	3.31	-0.06
28	60	35000	1662.0	263.3	1104	4.89	4.38	-0.51	4.63	-0.26
28	60	50000	1679.2	268.6	1104	5.55	4.28	-1.27	5.63	0.08
28	60	65000	1722.6	271.7	1104	6.25	5.85	-0.40	6.97	0.72
28	60	80000	1748.6	274.8	1104	5.64	6.06	0.42	6.63	0.99
28	120	35000	1705.2	267.6	1155	6.23	8.34	2.11	6.84	0.61
28	120	50000	1731.3	271.4	1155	7.36	8.63	1.27	8.72	1.36
28	120	65000	1757.3	277.5	1155	8.59	9.44	0.85	9.10	0.51
28	120	80000	1792.0	278.6	1155	7.56	9.10	1.54	8.73	1.17

The results of sweet corn yields prediction by using the above-mentioned linear regression functions are represented in the Table 2.

The RSQ values of the both regression models were the same – 0.897. This fact indicates comparable accuracy in the yield prediction of the both developed models independently of the used approach to their creation. However, it should be mentioned that the amplitude of the residuals was lower at the life factor approach model (3.04 t ha^{-1}) comparatively to the technological approach (3.13 t ha^{-1}).

The results of sweet corn ears yields prediction by the artificial neural network are represented in the Table 3.

The RSQ value of the technological model was 0.913, and RSQ of the life factors based one was considerably higher – 0.953. Besides, the amplitude of the residuals was lower under the use of life factors (2.16 t ha^{-1}) than technological parameters (4.23 t ha^{-1}) as the inputs. This fact allowed concluding that the life factors based approach to the development of yield predictive models should be preferred to the common technological approach.

DISCUSSION

Some scientists consider technological models more suitable and flexible in terms of usage in agricultural practice [Mikheiev 2005].

The results of the study proved that neural network models provide considerably better accuracy of predictions [Kaul et al. 2005; Choubin et al. 2016; Lykhovyd 2018]. Besides, these models can operate with non-numeric information that is very helpful in some cases. For example, it is very difficult to express different types of soil tillage in the form of figures (plowing, disking, chisel tillage, etc.). If developing the regression model is necessary, it becomes problematic in this

case . Such issue does not appear if dealing with artificial neural networks.

However, every separate mathematical model is very limited to be used as the only instrument for the yield prediction. The future is in the development and implementation of complex crop models taking into account a number of technological, climatic, soil, genetic factors for possible simulation of the processes and interrelations, which take place in agricultural field ecosystems [Jame, Cutforth 1996]. Besides, we also see the future of crop modeling systems in a combination of empirical models (regression, neural network, statistically based) with simulation models used in modern decision support systems like DSSAT [Jones et al. 2003], EPIC [Williams et al. 1989], CERES [Godwin, Vlek 1985], heuristic and optimization approaches to productivity prediction [Mikheiev 2005]. However, it should be taken into account that too complicated mathematical systems tend to be less efficient and useful in real life after some point of complexity, which determines a significant decrease of their usability and accuracy [Mikheiev 2005].

In our study, we have developed two different models on the basis of the experimental data. Although statistical and empirical models could be considered as the most reliable and accurate under certain conditions, the study is limited. We need to create a number of different models using not only different approaches to their development but using different types of the input information in general, for example, including empirical, expert and normative types of models for yielding prediction of different crops;, then, we have to compare their accuracy to provide agricultural engineers and specialists with the substantiated recommendations on use of mathematical apparatus in such predictive models.

CONCLUSIONS

The results of the study enabled to draw the following conclusions:

1. Mathematical modeling of the yield performed by using the linear regression analysis provided no advantage to any of the studied approaches. Both technological and life factor models granted the same level of reliability, which is proved by the same value of RSQ that was 0.897.

2. Artificial neural network models are significantly more precise and reliable than regression ones. At the same time, the life factor predictive model provided distinctly higher accuracy of sweet corn yield prediction with RSQ of 0.953 in comparison to RSQ of 0.913 of the technologically based network model. Therefore, life factor approach should be preferable in the yield predictive models.

3. An artificial neural network has an advantage over a linear regression model in the accuracy. However, it is more complicated in the development and implementation. Besides, the network model has a drawback of impossibility to express it in the form of a mathematical equation, which makes it less versatile for application in other environment. It is almost impossible to transit the network model in other software or perform the prediction without using a computer power. Linear regression models do not have this drawback because they are expressed in the form of a mathematical formula, which could be used under various computational environments.

4. No separate approach to the yield models can be presented as the only method of obtaining reliable productivity prediction. The future is in the combined models, which encompass all advantages of empirical and simulation approaches.

REFERENCES

1. Arinushkina E.V. 1970. Guide on the chemical analysis of soils. Moscow. Moscow State University.

2. Balaghi R., Tychon B., Eerens H., Jlibene M. 2008. Empirical regression models using NDVI, rainfall and temperature data for the early prediction of wheat grain yields in Morocco. International Journal of Applied Earth Observation and Geoinformation, 10(4), 438–452.

3. Cerrato M.E., Blackmer A.M. 1990. Comparison of models for describing; corn yield response to nitrogen fertilizer. Agronomy Journal, 82(1), 138–143.

4. Choubin B., Khalighi-Sigaroodi S., Malekian A., Kişi Ö. 2016. Multiple linear regression, multilayer perceptron network and adaptive neuro-fuzzy inference system for forecasting precipitation based on large-scale climate signals. Hydrological Sciences Journal, 61(6), 1001–1009.

5. Devore J.L. 2011. Probability and Statistics for Engineering and the Sciences. Boston. Cengage learning.

6. Dixon B.L., Hollinger S.E., Garcia P., Tirupattur V. 1994. Estimating corn yield response models to predict impacts of climate change. Journal of Agri-

cultural and resource economics, 19(1), 58–68.

7. Doyle C.J. 1991. Mathematical models in weed management. Crop Protection, 10(6), 432–444.

8. Draper N.R., Smith H. 2014. Applied regression analysis. New York City. John Wiley & Sons.

9. Godwin D.C., Vlek P.L.G. 1985. Simulation of nitrogen dynamics in wheat cropping systems. Wheat Growth and Modelling. Boston. Springer, 311–332.

10. Jame Y.W., Cutforth H.W. 1996. Crop growth models for decision support systems. Canadian Journal of Plant Science, 76(1), 9–19.

11. Jones J.W., Hoogenboom G., Porter C.H., Boote K.J., Batchelor W.D., Hunt L.A., Wilkens P.W., Singh U., Gijsman A.J., Ritchie J.T. 2003. The DSSAT cropping system model. European journal of agronomy, 18(3–4), 235–265.

12. Kaul M., Hill R.L., Walthall C. 2005. Artificial neural networks for corn and soybean yield prediction. Agricultural Systems, 85(1), 1–18.

13. Kogan F., Kussul N.N., Adamenko T.I., Skakun S.V., Kravchenko A.N., Krivobok A.A., Shelestov A.Yu., Kolotii A.V., Kussul O.M., Lavrenyuk A.N. 2013. Winter wheat yield forecasting: A comparative analysis of results of regression and biophysical models. Journal of Automation and Information Sciences, 45(6), 68–81.

14. Lavrenko S.O., Lavrenko N.N., Pichura V.I. 2015. Neural network modeling of chickpea grain yield on ameliorated soils. Nauchnyi zhurnal Rossiiskogo NII problem melioratsii, 2(18), 16–30.

15. Lobell D.B., Burke M.B. 2010. On the use of statistical models to predict crop yield responses to climate change. Agricultural and Forest Meteorology, 150(11), 1443–1452.

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

17. Mikheiev E.K. 2005. Informational systems in agronomy. Part I: Technological decision support systems on the level of planning. Kherson.

18. Rennie D.A., de Jong E. 1989. InnovationAcres: MaintainingProductivityand Soil Quality. AgriScience. December, 5–6.

19. Ushkarenko V.O. 1994. Irrigated agriculture: Textbook. Kyiv. Urozhaj.

20. Ushkarenko V.O., Kokovikhin S.V., Holoborodko S.P., Vozhehova R.A. 2014. Methodology of the field experiment (Irrigated agriculture): Textbook. Kherson. Hrin DS.

21. Shkonde E.I. 1971. On using the methodology of Kornfield for estimation of needs for Nitrogen fertilization of soils. Chemistry in Agriculture, 12, 50–60.

22. Smirnov P.M., Muravin E.A. 1984. Agrochemistry. 2nd Ed. Moscow. Kolos.

23. Williams J.R., Jones C.A., Kiniry J.R., Spanel D.A. 1989. The EPIC crop growth model. Transactions of the ASAE, 32(2), 497–0511.