International Journal of Agricultural Technology 2022Vol. 18(3):1033-1044
Available online http://www.ijat-aatsea.com
ISSN 2630-0192 (Online)

Beans (*Phaseolus vulgaris* L.) yields forecast using normalized difference vegetation index

Lavrenko, S.[2], Lykhovyd, P.[1*], Lavrenko, N.[3], Ushkarenko, V.[2] and Maksymov, M.[2]

[1]Institute of Irrigated Agriculture of NAAS, Naddniprianske, Kherson, Ukraine; [2]Faculty of Agriculture, Kherson State Agrarian and Economic University, Kherson, Ukraine; [3]Faculty of Architecture and Construction, Kherson State Agrarian and Economic University, Kherson, Ukraine.

Lavrenko, S., Lykhovyd, P., Lavrenko, N., Ushkarenko, V. and Maksymov, M. (2022). Beans (*Phaseolus vulgaris* L.) yields forecast using normalized difference vegetation index. International Journal of Agricultural Technology 18(3):1033-1044.

Abstract One of the main tasks of modern agrarian science is to enhance the productivity using available natural, labor, and technical resources. Introduction of remote sensing achievements in precision agriculture systems could be very helpful for early correction of cultivation technology and yield prediction. Based on the results at the irrigated field of Agricultural Cooperative Farm “Radianska Zemlia” (Paryshevo, Kherson oblast, 46.706631 N, 32.274669 E) with common beans (*Phaseolus vulgaris* L.), the polynomial regression (PR) model for the crop yield prediction using the highest values of normalized difference vegetation index (NDVI) within the crop growing season, which were recorded at the stage of blossoming – pods formation (V8 – R2), is developed. The model provided good fitting (RSQ value was 0.8069) with great accuracy of predictive performance (MAPE value was 7.12%). Using the model and the model-based gradual yields scale could be useful both for the operative adjustment of the crop cultivation technology and beans yield prediction. Additionally, an artificial neural network-based (ANN) model was developed for providing somewhat better fitting and predictive performance: RSQ value was 0.8988, MAPE value was 7.08%. The combined equation based on the results of polynomial regression model adjustment with superstition of the artificial neural network forecasting results (PRANN model) provided better fitting accuracy than the original polynomial model (RSQ value was 0.8080) along with the best forecast precision (MAPE value was 6.83%). The combined PRANN model is the best option for crop yields modeling and forecasting based on the values of NDVI.

Keywords: Neural network, *Phaseolus vulgaris* L., Polynomial regression, Prediction, Productivity

Introduction

The highest productivity at the least expense is the main goal of crop producers. Modern cultivation technologies used in industrial and well-

[*] **Corresponding Author:** Lykhovyd, P.; **Email:** pavel.likhovid@gmail.com

developed agrarian countries allow obtaining levels of yields, which would be assumed as fantastic just 30-50 years ago (Tsusaka and Otsuka, 2013; Fischer *et al.*, 2014). However, until now there are more questions than answers regarding regularities of crop formation and the influence of various technological operations on it. In this relation, a lot of farmers and scientists struggle with the problem of early detection of unfavorable events in the field and early crop prediction to react timely on threats to the crop formation.

Yield prediction is a valuable part of agricultural science and production as it is an important and irreplaceable component for guarantying food security (Al-Gaadi *et al.,* 2016). Crop growth and development monitoring systems are of great interest and importance for sustainable development of agrarian sector of the economy of Ukraine. At its earliest stages, the science of yield prediction grounded mainly on the results of field experiments that provided researchers with the information required for determination of the regularities related to crop formation in various environmental and agrotechnological conditions. Although field experiment remains the leading and most reliable method of acquiring scientific knowledge in agriculture, novel approaches are also involved to resolve the problems of accurate crop yield estimation. Amazing opportunities for the purpose are opened owing to remote sensing implementation.

Remote sensing technologies development allowed indirect studying vegetation conditions through the computation of vegetation indices. The most popular and widely used one is normalized difference vegetation index or NDVI, provided by Rouse *et al.* (1974). The use of NDVI gives agricultural scientists and practitioners an opportunity for indirect assessment of crops' conditions throughout their growing season even without visiting fields. The technology, if applied properly, is extremely helpful in monitoring of vast crop arrays in the dynamics, provides support in making reasonable decision regarding cultivation technology options based on the status of plants in the field (Liaghat and Balasundram, 2010). Remote sensing provides a possibility of full-time supervision for the crops' conditions and their reaction to agrotechnological intrusions. It is a prospective tool for yield prediction either, which has been confirmed by numerous studies with almost all cultivated crops.

Nowadays, remote sensing data is used for crops prediction through versatile mathematical models, among which regression models remain predominant due to their simplicity and high practicality (Lykhovyd, 2020). But it has been proved that regression models on their own because of the appearance of new, much more accurate and robust mathematical modeling methods, should be replaced. Currently, agricultural science, along with simple linear, polynomial, multiple regression data analysis, applies such approaches as fuzzy regression models, artificial neural networks (Verma *et al.,* 2018;

Khaki and Wang, 2019), combined models of multiple regression and artificial neural networks, etc. (Gopal and Bhargavi, 2019). The latter approach seems to be most mathematically complicated but provides the highest reliability and accuracy making it the most prospective for further introduction to yield prediction.

In Ukraine, common beans (*Phaseolus vulgaris* L.) is an important leguminous crop, which is cultivated in almost all the regions of the country. It is a valuable source of plant protein and essential amino acids, which are an irreplaceable constituent of healthy human nutrition. Recently, an interest to beans cultivation in Ukraine has significantly increased, and leading agricultural producers are keen to obtain the maximum practical knowledge on the crop cultivation (Poedinceva *et al.,* 2020). Much has been done to study and improve the cultivation technology of common beans but there were almost no studies aimed to forecast the crop yields using NDVI (Ushkarenko *et al.,* 2018). There is no doubt that the knowledge on beans yield forecasting is of a great interest and practical value for the crop producers.

The goal of our study was to develop and tested the model for beans yields estimation using NDVI values applying combined mathematical approach of polynomial regression and artificial neural network. Thence, the gradual yielding scale of the crop estimated yield depending on the NDVI was to create and propose to agricultural producers.

Materials and methods

The field experiment with common beans was conducted in 2016 on the irrigated lands of Agricultural Cooperative Farm "Radianska Zemlia" (Paryshevo, Kherson oblast, Ukraine, 46.706631 N, 32.274669 E). The study aimed to improve the crop cultivation technology and embraced the investigation of such factors as tillage depth (two gradations), fertilization (three gradations), inter-row spacing (four gradations). The study was set in four replications using partially randomized split plot design. Therefore, there were 96 experimental plots representing the experimental field. The accounted area of one experimental plot was 150 m^2. The yield from each plot was harvested using combined harvester machine and weighed. Further the gross yield was recalculated to the standard grain moisture of 14% and 100% pure condition.

The values of NDVI were obtained using OneSoil GIS online service, providing the computation of the spatial vegetation index from Sentinel-2 and Sentinel-1 combined images. Each pixel of the image represents 5m ×5m square (Dunaieva *et al*., 2018). Mean NDVI values from 6 pixels were used to represent each variant of the field experiment. The highest values of NDVI through the growing season of common beans were used to create forecasting models. The highest NDVI was recorded at the stage of blossoming – pods formation (V8 – R2) or 50 – 55 days prior to harvesting.

Relative standard deviation (RSD) was calculated using the formula provided in the statistical dictionary by Everitt and Skrondal (2010), while standard deviation (SD) was estimated according to Bland and Altman (1996).

The combined approach to common beans yield forecast was applied. Firstly, the data set (containing the pairs of "NDVI-beans yield") was processed by the standard algorithm of polynomial regression (PR) analysis (the polynomic function of the second grade) using the means of Microsoft Excel 365 add-in BioStat v7 (Neter *et al*. 1996). Thence, the same data set was processed using artificial neural network with back propagation learning algorithm in Tiberius XL software (Brierley and Batty, 1999). The neural network initializing and learning parameters was set up as follows: 5 hidden neurons in the layer, redisplay rate – 100, the number of epochs – 1000, learning rate – 0.50. The results of the beans yield forecast, obtained using the artificial neural network (ANN), was further involved through superposition with the forecast results of the polynomial model to additional linear regression post-processing in BioStat v7 to create the combined polynomial regression and artificial neural network model (PRANN model) through the computation of adjustment coefficient.

Forecasting accuracy evaluation of the created mathematical models were performed using the mean absolute percentage error (MAPE), while the quality of fitting was assessed using RSQ values (Moreno *et al*., 2013; Khair *et al*., 2017). Graphical approximation of the developed models has been performed using the tools of Microsoft Excel 365.

The gradual scale of beans productivity depending on NDVI at the crop's stage V8 – R2 was created using the combined PRANN model taking into consideration the dispersion of the predicted values, the dispersion is assumed to be equal to the absolute error.

Results

Generalization of the recorded experimental data which aggregated from average values for each replication of the field trial is presented in the Table 1. There were 24 analytical pairs of "NDVI-beans yield" altogether created for further processing using previously mentioned mathematical algorithms.

The results of polynomial regression analysis are presented in the Tables 2, 3, and 4. Based on the results of this statistical analysis, PR model for beans productivity depending on the values of NDVI was recorded at the V8 – R2 stage of the crop's development which was created as (1):

$$Y = 4.4278 - 16.7325 \times NDVI + 22.8703 \times NDVI^2 \quad (1)$$

where Y is beans yield, t ha^{-1}; $NDVI$ – value of the index at the V8 – R2 stage of beans, pts.

Table 1. Analytical pairs of the experimental data used for beans yield prediction

Beans yield (t ha^{-1})	**NDVI at the V8 – R2 stage (pts)**
1.39	0.45
1.71	0.50
2.32	0.48
1.96	0.55
1.93	0.55
2.24	0.58
2.96	0.61
2.54	0.60
2.08	0.55
2.31	0.57
3.11	0.65
2.84	0.57
1.39	0.45
1.77	0.52
2.41	0.57
2.03	0.55
1.96	0.53
2.31	0.58
3.02	0.62
2.60	0.60
2.09	0.55
2.37	0.60
3.16	0.65
2.90	0.60
RSD = 0.22; SD = 0.51 t ha^{-1}	RSD = 0.10; SD = 0.05 pts

Table 2. Regression statistics for the PR model of beans yield prediction

Statistical index	**Value**
Correlation coefficient R	0.8983
Coefficient of determination (RSQ)	0.8069
RSQ adjusted	0.7885
SD	0.2346

Table 3. Analysis of variance for the PR model of beans yield prediction

Parameter	d.f.	SS	MS	F	P-value
Regression	2	4.83	2.42	43.88	3.17×10^{-8}
Residuals	21	1.16	0.06		
Total	23	5.99			

Table 4. Regression coefficients and additional statistics for the PR model of beans yield prediction

Input relation	Coefficient	SE	LCL	UCL	t-statistics	P-value
Constant	4.4278	4.1417	-4.1853	13.0409	1.0691	0.2972
NDVI	-16.7325	15.1589	-48.2572	14.7921	-1.1038	0.2822
NDVI^2	22.8703	13.7908	-5.8093	51.5498	1.6584	0.1121

The ANN model of beans productivity developed using Tiberius XL, which provided no possibilities to derive any equation and interpret the track of computations required to achieve the result, which is the major drawback of pure ANN models with their "black box nature". However, the fitting graph (Figure 1) and modeled values of beans yield allowed to compare the quality of fitting of both PR and ANN models, as well as to evaluate their accuracy using MAPE (Table 5). Also, the calculated R, and RSQ values for the ANN model which amounted to 0.9480 and 0.8988, respectively which significantly outperforming the PR model.

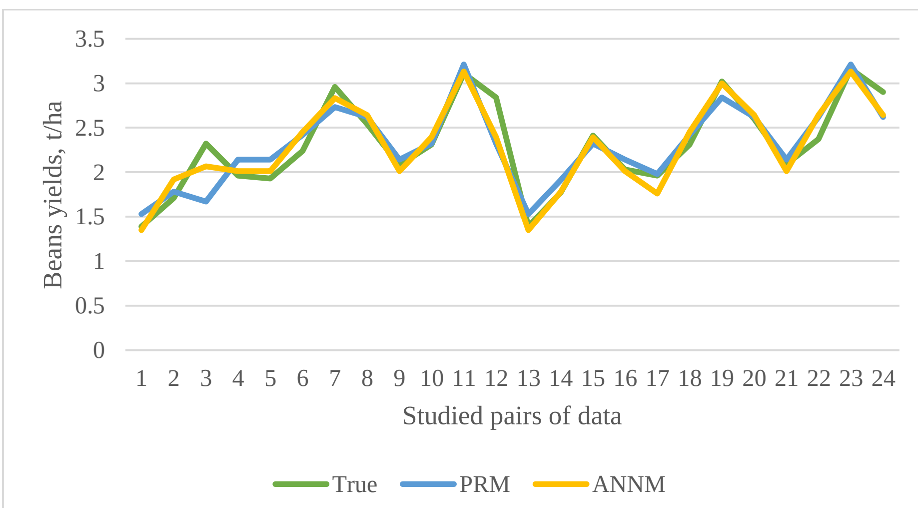

Figure 1. Approximation and fitting quality of the developed PR and ANN models of beans yields depending on NDVI at the stage V8 – R2 of the crop

Table 5. True and forecasted by PR and ANN models beans yields and MAPE of the models

True yields, t ha^{-1}	PR forecasted yields, t ha^{-1}	ANN forecasted yields, t ha^{-1}
1.39	1.53	1.35
1.71	1.78	1.92
2.32	1.67	2.07
1.96	2.14	2.01
1.93	2.14	2.01
2.24	2.42	2.45
2.96	2.73	2.83
2.54	2.62	2.64
2.08	2.14	2.01
2.31	2.32	2.39
3.11	3.21	3.13
2.84	2.32	2.39
1.39	1.53	1.35
1.77	1.91	1.78
2.41	2.32	2.39
2.03	2.14	2.01
1.96	1.98	1.76
2.31	2.42	2.45
3.02	2.84	3.00
2.60	2.62	2.64
2.09	2.14	2.01
2.37	2.62	2.64
3.16	3.21	3.13
2.90	2.62	2.64
MAPE, %	7.12	7.08

Further adjustment of the PR model using the results of the ANN model prediction through the linear regression coefficient calculation was performed. The results of the adjustment are presented in the Tables 6, 7, and 8.

Table 6. Regression statistics for the adjustment of the PR model of beans yield prediction using the ANN model through linear regression analysis

Statistical index	Value
Correlation coefficient R	0.9983
Coefficient of determination (RSQ)	0.9965
RSQ adjusted	0.9965
SD	0.1408

Table 7. Analysis of variance for the adjustment of the PR model of beans yield prediction using the ANN model through linear regression analysis

Parameter	d.f.	SS	MS	F	P-value
Regression	1	131.39	131.39	6629.80	0
Residuals	23	0.46	0.02		
Total	24	131.85			

Table 8. Regression coefficients and additional statistics for the PR model of beans yield prediction using the ANN model through linear regression analysis

Input relation	Coefficient	SE	LCL	UCL	t-statistics	P-value
Constant	0					
PR model	0.9956	0.0122	0.9703	1.0209	81.4236	0

After the adjustment, the PRANN combined model for beans yield prediction is as follows (2):

$$Y = 4.4083 - 16.6589 \times NDVI + 22.7697 \times NDVI^2 \quad (2)$$

where *Y* is beans yield, t ha^{-1}; *NDVI* – value of the index at the V8 – R2 stage of beans, pts.

The results of the PRANN model approximation and accuracy evaluation are presented in the Figure 2 and Table 9. The computations testify about intermediate fitting quality (RSQ value was 0.8080) of the PRANN model, comparing to the PR and ANN ones, and its best accuracy with MAPE of 6.83%.

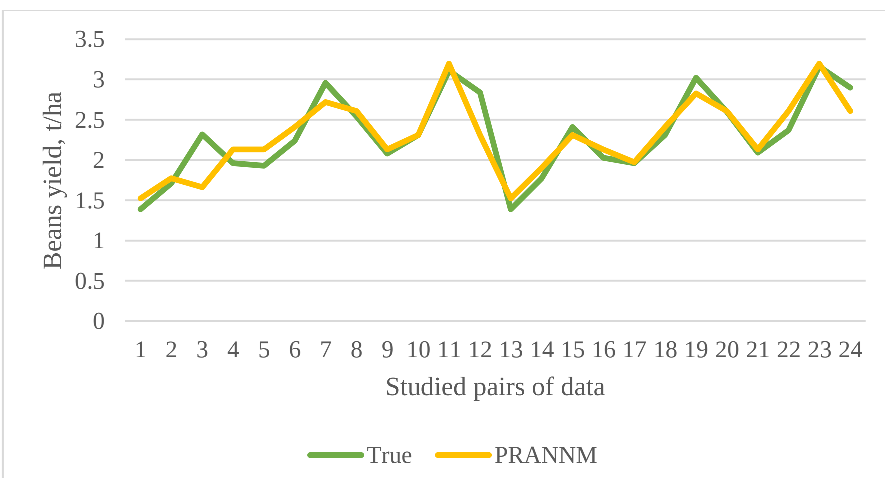

Figure 2. Approximation and fitting quality of the developed PRANN model of beans yields depending on NDVI at the stage V8 – R2 of the crop

Table 9. True and forecasted by the PRANN model beans yields and MAPE of the model

True yields, t ha^{-1}	PRANN forecasted yields, t ha^{-1}
1.39	1.53
1.71	1.78
2.32	1.67
1.96	2.14
1.93	2.14
2.24	2.42
2.96	2.73
2.54	2.62
2.08	2.14
2.31	2.32
3.11	3.21
2.84	2.32
1.39	1.53
1.77	1.91
2.41	2.32
2.03	2.14
1.96	1.98
2.31	2.42
3.02	2.84
2.60	2.62
2.09	2.14
2.37	2.62
3.16	3.21
2.90	2.62
MAPE, %	6.83

Table 10. The gradual scale for beans yield estimation using NDVI values at the V8 – R2 stage of the crop

NDVI, pts	Estimated yield, t ha^{-1}		
	min	avg	max
0.30	1.36	1.46	1.56
0.35	1.27	1.37	1.46
0.40	1.29	1.39	1.48
0.45	1.42	1.52	1.63
0.50	1.65	1.77	1.89
0.55	1.99	2.13	2.28
0.60	2.43	2.61	2.79
0.65	2.98	3.20	3.42
0.70	3.64	3.90	4.17
0.75	4.40	4.72	5.04
0.80	5.27	5.65	6.04
0.85	6.25	6.70	7.16

Based on the results of the PRANN model, the gradual scale for beans yield estimation using NDVI values at the V8 – R2 stage of the crop was developed (Table 10). The dispersion of the probable yield was taken into account, providing the minimum (min), average (avg), and the maximum values of possible yield.

The gradual scale started from 0.35 NDVI value because the values below 0.30 were not representative for the late stages of the crop development according to OneSoil platform guidelines. The developed gradual scale would be helpful for the crop producers allowing them to estimate possible yields with high reliability and almost no time and labor expenses.

Discussion

The results of beans yield modeling testify about high fitting quality and accuracy of all the developed models, including the PR, ANN, and combined PRANN ones. The difference in performance between the models is quite slight, although we must admit that the least MAPE is provided by the PRANN model, while the best fitting quality is acquired to the ANN one. All the models showed great accuracy according to Moreno *et al.* (2013). Gradation of the models by MAPE values. But considering the purpose of modeling, we suggest that the model with the least MAPE, in our case – the PRANN, must be applied for beans yield estimation to obtain the most reliable results. The worst fitting quality and accuracy among the developed models was in the PR one. This fact agrees with the claims of most other studies conducted to compare the performance of regression and neural networks in yield prediction of various crops, for example, sweet corn (Lykhovyd, 2018), corn and soybeans (Li *et al.*, 2007), potato (Abrougui *et al.*, 2019), barley (Ayoubi and Sahrawat, 2011). While neutral networks used in yield prediction by vegetation indices is not a novel approach, the adjustment of regression models based on the results of neural network forecasting provides new insights on the enhancement of regression models performance (Panda *et al*., 2010).

Examination of the studies by other scientific groups revealed that there were very few ones devoted to the yield estimation of beans using NDVI values. For example, Gonzalez-Gonzalez (2018) successfully applied NDVI to estimate yields of beans on a large scale in Mexico lands, having received RSQ value of the model amounted to 0.70. The study by Nemeskéri *et al*. (2018) is in absolute agreement with our results, claiming that NDVI at the stage of flowering – pod formation of beans is closely correlated with the crop yield, and it could be used for preliminary productivity estimation. The researchers testify that the highest NDVI (0.85) corresponded to the yield of beans fluctuating within 6 to 9.5 t ha^{-1} depending on the varietal traits, while in our conditions the yield of the crop at the NDVI value of 0.85 is estimated as slightly lower – 6.25 to 7.16 t ha^{-1}. However, it is evident that the predicted yield fluctuation was much less than in the study by Nemeskéri *et al.* (2018). Quite similar results supported the statement of strong correlation between NDVI at the early pod formation stage of beans and the crop yield were obtained by Sankaran *et al.* (2019), who claimed about RSQ values fluctuating within 0.18-0.53.

Moreover, it is revealed the first to propose crop producers a gradual scale based on the results of mathematical modeling for early beans yield estimation. The developed scale would increase the interest of farmers in usage of remote sensing in their practice and enhance crop’s productivity. It recommended that gradually yielding scale for practical use for all the growers of common beans are concerned.

Acknowledgements

The authors would like to express gratitude to OneSoil team for the provision of NDVI images required for the study conduction. Besides, the authors thank Dr Rafiq Islam (Ohio State University, USA) for his valuable support in publication of this manuscript.

References

Abrougui, K., Gabsi, K., Mercatoris, B., Khemis, C., Amami, R. and Chehaibi, S. (2019). Prediction of organic potato yield using tillage systems and soil properties by artificial neural network (ANN) and multiple linear regressions (MLR). Soil and Tillage Research, 190:202-208.

Al-Gaadi, K. A., Hassaballa, A. A., Tola, E., Kayad, A. G., Madugundu, R., Alblewi, B. and Assiri, F. (2016). Prediction of potato crop yield using precision agriculture techniques. PloS one, 11:e0162219.

Ayoubi, S. and Sahrawat, K. L. (2011). Comparing multivariate regression and artificial neural network to predict barley production from soil characteristics in northern Iran. Archives of Agronomy and Soil Science, 57:549-565.

Bland, J. M. and Altman, D. G. (1996). Statistics notes: measurement error. BMJ, 312:1654.

Brierley, P. and Batty, B. (1999). Data mining with neural networks - an applied example in understanding electricity consumption patterns. In: Knowledge Discovery and Data Mining (Professional Applications of Computing Series), Institution of Engineering & Technology, pp.240-303.

Dunaieva, I., Mirschel, W., Popovych, V., Pashtetsky, V., Golovastova, E., Vecherkov, V., Melnichuk, A., Terleev, V., Nikonorov, A., Ginevsky, R., Lazarev, V. and Topaj, A. (2018). GIS services for agriculture monitoring and forecasting: development concept. International Scientific Conference Energy Management of Municipal Transportation Facilities and Transport 2018, Cham, Springer, pp.236-246.

Everitt, B. S. and Skrondal, A. (2010). The Cambridge dictionary of statistics. Cambridge, Cambridge University Press, 480 p.

Fischer, R. A., Byerlee, D. and Edmeades, G. (2014). Crop yields and global food security. ACIAR: Canberra, ACT, pp.8-11.

Gonzalez-Gonzalez, M. A. (2018). Spatial modeling of rainfed bean production in the semiarid central Mexico: climate surface and vegetation remote sensing determinants. (Ph.D. Thesis). The University of Arizona, US.
Gopal, P. M. and Bhargavi, R. (2019). A novel approach for efficient crop yield prediction. Computers and Electronics in Agriculture, 165:104968.
Khair, U., Fahmi, H., Al Hakim, S. and Rahim, R. (2017). Forecasting error calculation with mean absolute deviation and mean absolute percentage error. Journal of Physics: Conference Series, 930:012002.
Khaki, S. and Wang, L. (2019). Crop yield prediction using deep neural networks. Frontiers in plant science, 10:621.
Li, A., Liang, S., Wang, A. and Qin, J. (2007). Estimating crop yield from multi-temporal satellite data using multivariate regression and neural network techniques. Photogrammetric Engineering & Remote Sensing, 73:1149-1157.
Liaghat, S. and Balasundram, S. K. (2010). A review: The role of remote sensing in precision agriculture. American Journal of Agricultural and Biological Sciences, 5:50-55.
Lykhovyd, P. (2020). Sweet corn yield simulation using normalized difference vegetation index and leaf area index. Journal of Ecological Engineering, 21:228-236.
Lykhovyd, P. V. (2018). Prediction of sweet corn yield depending on cultivation technology parameters by using linear regression and artificial neural network methods. Biosystems Diversity, 26:11-15.
Moreno, J. J. M., Pol, A. P., Abad, A. S. and Blasco, B. C. (2013). Using the R-MAPE index as a resistant measure of forecast accuracy. Psicothema, 25:500-506.
Nemeskéri, E., Molnár, K. and Helyes, L. (2018). Relationships of spectral traits with yield and nutritional quality of snap beans (*Phaseolus vulgaris* L.) in dry seasons. Archives of Agronomy and Soil Science, 64:1222-1239.
Neter, J., Wasserman, W. and Kutner, M. H. (1996). Applied Linear Statistical Models. Vol. 4. Irwin, Chicago, 318 p.
Panda, S. S., Ames, D. P. and Panigrahi, S. (2010). Application of vegetation indices for agricultural crop yield prediction using neural network techniques. Remote Sensing, 2:673-696.
Poedinceva, A. A., Turenko, V. P., Stankevych, S. V., Matsyura, A. V., Bilyk, M. O., Zabrodina, I. V., Zhukova, L. V., Goryainova, V. V., Bezugla, O. N., Zayarna, O. Yu. and Batova, O. M. (2020). A review of protection measures against the principal bean diseases in Ukraine. Ukrainian Journal of Ecology, 10:236-240.
Rouse, J. W., Haas, R. H., Schell, J. A. and Deering, D. W. (1974). Monitoring vegetation systems in the Great Plains with ERTS. NASA special publication, 351:309.
Sankaran, S., Quirós, J. J. and Miklas, P. N. (2019). Unmanned aerial system and satellite-based high resolution imagery for high-throughput phenotyping in dry bean. Computers and Electronics in Agriculture, 165:104965.
Tsusaka, T. W. and Otsuka, K. (2013). The changing effects of agro-climate on cereal crop yields during the green revolution in India, 1972 to 2002. Journal of Sustainable Development, 6:11.
Ushkarenko, V. O., Lavrenko, S. O., Lykhovyd, P. V., Lavrenko, N. M. and Maksymov, D. O. (2018). Yield components of haricot beans (*Phaseolus vulgaris* L.) depending on cultivation technology elements at the irrigated lands of the Steppe zone. Modern Phytomorphology, 12:73-79.
Verma, A., Jatain, A. and Bajaj, S. (2018). Crop yield prediction of wheat using fuzzy c means clustering and neural network. International Journal of Applied Engineering Research, 13:9816-9821.

(Received: 18 November 2021, accepted: 10 April 2022)