МЕЛІОРАЦІЯ, ЗЕМЛЕРОБСТВО, РОСЛИННИЦТВО

UDC 528.8:519.2:633.1
DOI https://doi.org/10.32848/agrar.innov.2020.1.1

FORECASTING THE YIELDS OF SPRING ROW CROPS BY THE REMOTE SENSING DATA

VOZHEHOVA R.A. – Doctor of Agricultural Sciences, Professor,
Corresponding Member of the National Academy of Agrarian Sciences of Ukraine
https://orcid.org/0000-0002-3895-5633
Institute of Irrigated Agriculture of the National Academy of Agrarian Sciences of Ukraine
MALIARCHUK M.P. – Doctor of Agricultural Sciences
https://orcid.org/0000-0002-0150-6121
Institute of Irrigated Agriculture of the National Academy of Agrarian Sciences of Ukraine
BILIAIEVA I.M. – Doctor of Agricultural Sciences
https://orcid.org/0000-0003-0688-4209
Institute of Irrigated Agriculture of the National Academy of Agrarian Sciences of Ukraine
LYKHOVYD P.V. – Candidate (Ph.D.) of Agricultural Sciences
https://orcid.org/0000-0002-0314-7644
Institute of Irrigated Agriculture of the National Academy of Agrarian Sciences of Ukraine
MALIARCHUK A.S. – Candidate (Ph.D.) of Agricultural Sciences
https://orcid.org/0000-0001-7119-269X
Institute of Irrigated Agriculture of the National Academy of Agrarian Sciences of Ukraine

Problem statement. Remote sensing is a developing branch of modern science and technology, which is used in GIS for mapping, land surface management, water and environmental management, ecological monitoring, modeling and forecasting of natural ecosystems conditions, etc. Besides, it is an important part of modern systems of precise agriculture, where remote sensing provides the data that is further integrated with various decision support systems to improve agricultural management [12; 15; 27]. The technique provides great opportunities for fast and precise evaluation of crops vegetation conditions for making reasonable amendments to cultivation technology and obtain the maximum crops productivity [16; 22]. To perform mentioned functions, a number of vegetation indices, which are calculated by the remote sensing imagery, are used.

One of the most widely spread vegetation indices is Normalized Difference Vegetation Index (NDVI), which was first introduced by Rouse et al. (1974) [29]. The index was the first, which was derived from the satellite imagery data and applied to distinguish vegetation cover and get information about its conditions. It is calculated by the Eq. (1):

$$NDVI = \frac{(a_{nir} - a_{vis})}{(a_{nir} + a_{vis})} \quad (1)$$

where: a_{nir} is the reflective infrared range of the spectrum, a_{vis} is the visible red range of the spectrum [3].

Analysis of the last studies and publications. NDVI applications are not limited just to detection of vegetation and description of its conditions. Moreover, it is an indirect indicator that testifies about potential photosynthetic activity of flora, and, as a result, it could be used to obtain information on the potential productivity of crops. The connection between NDVI and volumes of absorbed photosynthetic active radiation (PAR) is direct and linear [10]. The strong connection between NDVI and PAR makes it possible to find out a relationship between NDVI and yields, as the latter depends directly on the volumes of PAR, which is efficiently used by crops [26, 31].

Purpose. The goal of the study was to determine the connection between the values of NDVI and yields of the major spring row crops for early predictions of their productivity.

Materials and methods. Finding the connection between NDVI values and yields was performed using polynomial regression analysis under by the Cramer's rule [11]. True values of spring row crops (corn, sorghum and soybean) were the inputs and corresponding values of NDVI obtained from the Sentinel-2 and Sentinel-1 combined imagery at critical stages of the crops growth, namely: V2 (second trifoliate) and R2 (full bloom) for soybean [19]; S3 (growing point differentiation) and S6 (half bloom) for sorghum [28]; VT (tasselling) and R1 (silking) for corn [23]. True yields of the studied crops were obtained during the harvesting of the studied crops at the experimental field of the Institute of Irrigated Agriculture of NAAS in 2017–2018. The yields were calculated for the standard moisture content in grain (14 % for corn, 13.5 % for sorghum and 12 % for soybean). Coefficient of yields variation (CV) was calculated as a ratio of the standard deviation (SD) to the mean [6]. The yields were connected to the corresponding NDVI and the data were processed using Microsoft Excel at the probability level of 95 % ($p<0.05$) with further approximation and calculation of mean absolute percentage errors (MAPE) for yield predictions [4].

Results. Analyzing NDVI values at the stages of corn growth we observed the regulation that the values of the index at the stages of the crop growth (VT and R1) were similar (Figure 1).

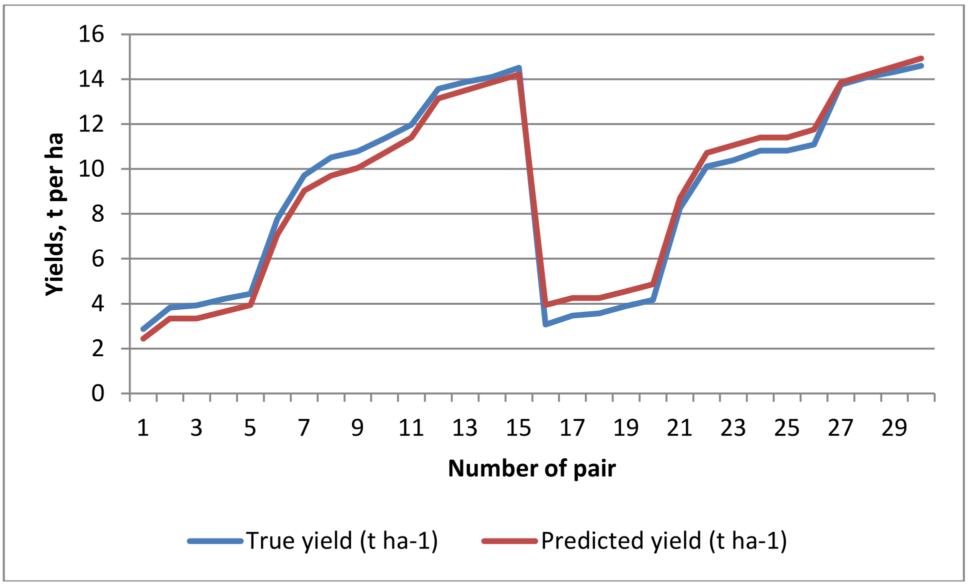

Figure 1. True and predicted by the NDVI-based model yields of corn

Regression analysis determined the connection of NDVI and corn yields by the rule of thumb, which is very high and positive: coefficient of correlation R=0.9906, and coefficient of determination R^2 =0.9813 [21]. Quadratic Eq. (2) describes the relationship between the index values and yield of the crop:

$$y = 8.571 \times x^2 + 22.755 \times x - 8.035 \qquad (2)$$

where: *y* is the yield of maize in t ha^{-1}, and *x* is the value of NDVI at VT or R1 stage.

Approximation of the regression model and calculation of MAPE, that is less than 10 %, proved high accuracy and reliability of the prediction model [20].

As for other studied crops, NDVI corresponding to different stages of their growth differed providing unequal accuracy for the yield predictions.

Thus, the least accuracy of the regression model for the yield forecasting was for sorghum (Figure 2).

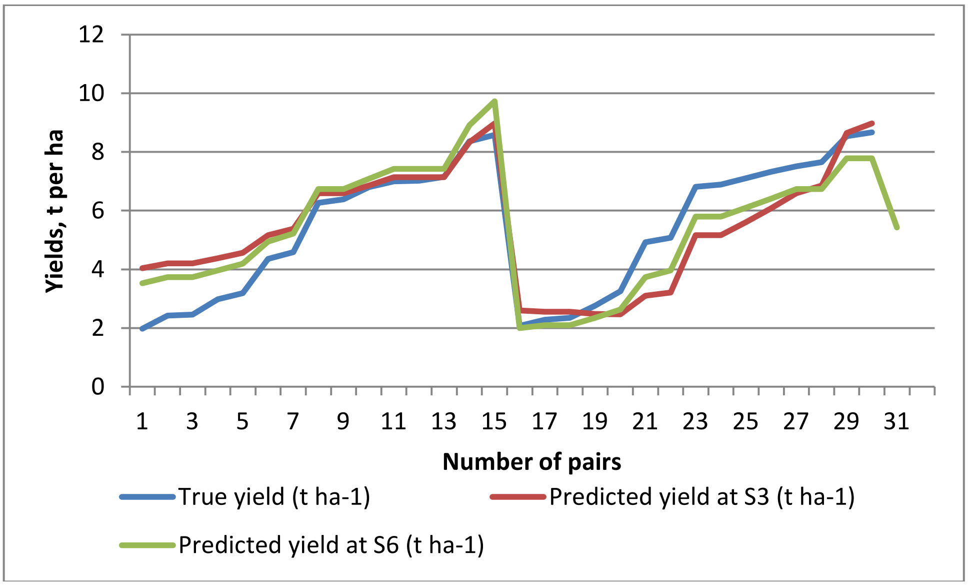

Figure 2. True and predicted by the NDVI-based model yields of sorghum

The most inequality in the yields assessment was observed at S3 stage, where it averaged to 22.01 %. The model at this stage provides reasonable forecasting [20]. The regression Eq. for S3 stage is as follows (3):

$$y = 42.311 \times x^2 - 30.065 \times x + 7.833 \qquad (3)$$

where: *y* is the yield of sorghum in t ha^{-1}, and *x* is the value of NDVI at S3 stage.

The coefficient of correlation R for the model is 0.8809, the coefficient of determination is 0.7760, showing a high positive correlation according to the rule of thumb [21].

Sorghum yield prediction using NDVI at S6 stage has greater accuracy with an average MAPE of 17.62 % that is a good forecasting [20]. The regression Eq. for the model is as follows (4):

$$y = 52.193 \times x^2 - 42.126 \times x + 10.014 \quad (4)$$

where: y is the yield of sorghum in t ha^{-1}, and x is the value of NDVI at S6 stage.

The coefficient of correlation for the model R is 0.9298, and the coefficient of determination is 0.8645, that testifies about a high positive correlation according to the rule of thumb [21].

Regression analysis of NDVI and soybean yields testified that the highest level of correspondence between the inputs and outputs was at V2 stage of the crop, when MAPE averaged to 3.75% testifying about very high accuracy of the forecast [20] (Figure 3).

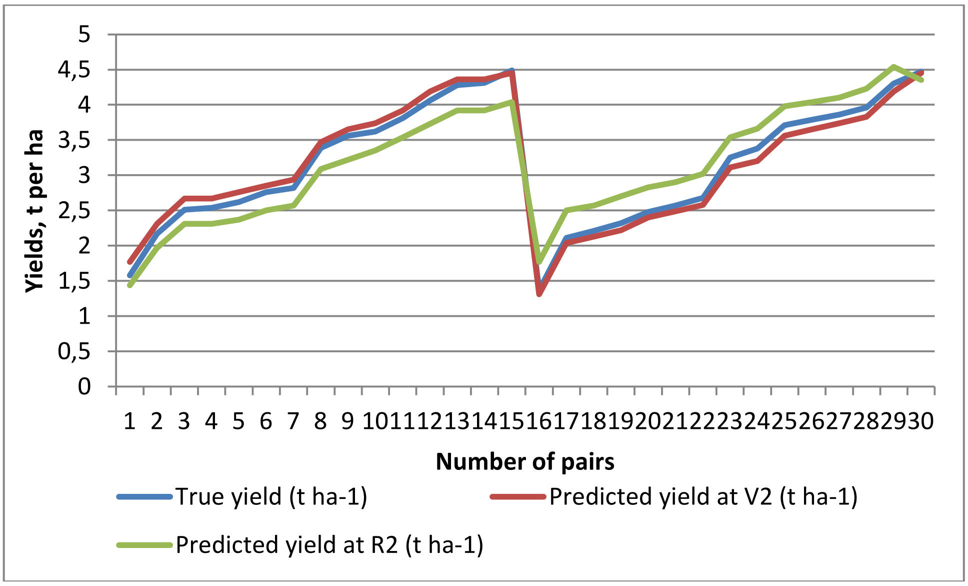

Figure 3. True and predicted by the NDVI-based model yields of soybean

The forecasting model could be expressed as Eq. (5):

$$y = -0.221 \times x^2 + 9.220 \times x - 2.338 \quad (5)$$

where: y is the yield of soybean in t ha^{-1}, and x is the value of NDVI at V2 stage.

The coefficient of correlation R for this model is 0.9914, the coefficient of determination is 0.9829, which is an extremely high positive correlation [21].

The polynomial regression model for soybean yield at R2 stage is less accurate with MAPE that is 10.16 %, however, this value also certifies about the possibility of precise productivity prediction for the crop [20]. The model for the yield prediction is as follows in the Eq. (6):

$$y = -0.221 \times x^2 + 9.220 \times x - 2.338 \quad (6)$$

where: y is the yield of soybean in t ha^{-1}, and x is the value of NDVI at R2 stage.

The coefficient of correlation R for the model is 0.9377, the coefficient of determination is 0.8793, testifying about a very high positive correlation [21].

Our results testify that it is possible to forecast the crop yields by NDVI with a reasonably high accuracy exceeding 90 % for corn and soybean, while the model accuracy for sorghum is just about 80 %. Lower accuracy in the sorghum yield prediction could be put on higher variation in the input data set used in the study: CV for S3 stage was the highest among the studied crops and reached 23 %, while we observed the tendency to an increase of the forecasting model performance under lower fluctuations of the input index (the closest prediction was obtained at the lowest CV – 16 % at V2 stage of soybean).

Another study on soybean yield prediction by the NDVI has approved that there is a strong non-linear relationship between the crop productivity and NDVI with the value of adjusted R^2 of 0.721 under the implementation of flexible Fourier transform model [35]. Very close to our results concerning soybean yield prediction were obtained in the work of Bolton & Friedl (2013), where the accuracy of soybean yield prediction by the MODIS NDVI data had a reasonable accuracy with the coefficient of determination averaging to 0.69 [2]. Another recent study discovered that NDVI values have a positive correlation with corn and soybean yields and are good for the yield prediction [13]. NDVI has also been proved to be efficient for a large-scale corn yield prediction by the means of regression models on the basis of long-term data providing reliable results 6–8 weeks before the harvesting period [24]. Regression analysis of corn yield and NDVI time series discovered a strong dependence of the harvest on the NDVI at pre-silking period allowing to predict yield losses due to unfavourable conditions in this period [34]. There is a study on the very high reliability of an empirical model "corn yield – NDVI at flowering stage" that provided just 3 % difference from the true yields [9]. The study devoted to the determination of corn yields depending on the NDVI at different stages of the crop growth showed that the best yield prediction performance was obtained under the implementation of R2 stage NDVI inputs [18], while our study showed that the model performance is best at R1 stage. Some scientists claimed about strong dependence of the "NDVI – corn yield" prediction model on the plant density [5], while we did not take this factor into account in our study. As for sorghum, there are a few findings related to the yield prediction based on the NDVI. There is a report stated about high accuracy (MAPE<20 %) of sorghum biomass prediction using NDVI data 6 months before harvesting [32]. A comprehensive large-scale study, which was performed in the US with different crops, including sorghum, corn, soybean, on the establishment of the connections between the yields and NDVI showed positive correlation between these indices for all the studied crops testifying about the possibility for the use of remote sensing data in yield prediction [13]. Another big study devoted to the determination of corn, soybean and sorghum yields through the multivariate regression analysis of satellite imagery and computed vegetation indices testified on the reasonable correlation between the indices and yields (the coefficient of correlation figures were 0.86, 0.74 and 0.65 for corn, soybean and sorghum, respectively) [25]. These results agree with our study, namely, that the least relationship with the coefficient of determination of 0.42 was recorded for sorghum.

Besides pure NDVI-based models, scientists developed combined models using additional indices related to crop productivity, namely, PAR, leaf area index (LAI), enhanced vegetation index (EVI), etc. [1; 7; 17]. This is reasonable in many cases when it is difficult to obtain reasonable prediction performance using vegetation index as the only input, because introduction of additional indices usually significantly improves the accuracy. Besides, using better computation techniques can also be valuable step for the enhancement of yield forecasting [30; 33]. However, sometimes complicated computations, for example, such as using artificial neural networks (ANN), do not guarantee the performance, which is significantly better than of regression analysis: the ANN NDVI-based model of sugarcane yield prediction had the coefficient of determination equal to 0.61 that cannot be considered as a good forecast [8].

Conclusions. Statistical analysis of the yields of three spring row crops, namely, corn, sorghum and soybean, in the connection to NDVI obtained from the Sentinel-2 imagery at the critical stages of the studied crops growth testified about a high positive correlation between the index and the yields. Polynomial regression NDVI-based models for early yield prediction are good for early yield prediction at the probability level of 95 % ($p<0.05$). The values of MAPE for the best prediction models are: 8.75 % for corn, 17.62 % for sorghum, 3.75 % for soybean. Thus, the NDVI could be used as a tool for early yield prediction both for scientific and practical needs.

Notwithstanding the fact that a huge number of studies are devoted to yield simulation by remote sensing indices it is necessary to obtain more knowledge on the technique of their implementation in the systems of precision agriculture for giving reasonable advice to agricultural producers.

REFERENCES:

1. Báez-González, A.D., Chen, P.Y., Tiscareño-López, M., & Srinivasan, R. (2002). Using satellite and field data with crop growth modeling to monitor and estimate corn yield in Mexico. *Crop Science*, 42, 1943–1949.

2. Bolton, D.K., & Friedl, M.A. (2013). Forecasting crop yield using remotely sensed vegetation indices and crop phenology metrics. *Agricultural and Forest Meteorology*, 173, 74–84.

3. Carlson, T.N., & Ripley, D.A. (1997). On the relation between NDVI, fractional vegetation cover, and leaf area index. *Remote sensing of Environment*, 62, 241–252.

4. De Myttenaere, A., Golden, B., Le Grand, B., & Rossi, F. (2016). Mean absolute percentage error for regression models. *Neurocomputing*, 192, 38–48.

5. de Oliveira, M.F., Ormond, A.T.S., de Freitas Noronha, R.H., dos Santos, A.F., Zerbato, C., & Furlani, C.E.A. (2019). Prediction Models of Corn Yield by NDVI in Function of the Spacing Arrangement. *Journal of Agricultural Science*, 11, 493–500.

6. Everitt, B.S., & Skrondal, A. (2010). *The Cambridge Dictionary of Statistics*. Cambridge Univ. Press, Cambridge, UK.

7. Fang, H., Liang, S., & Hoogenboom, G. (2011). Integration of MODIS LAI and vegetation index products with the CSM–CERES–Maize model for corn yield estimation. *International Journal of Remote Sensing*, 32, 1039–1065.

8. Fernandes, J.L., Ebecken, N.F.F., & Esquerdo, J.C.D.M. (2017). Sugarcane yield prediction in Brazil using NDVI time series and neural networks ensemble. *International Journal of Remote Sensing*, 38(16), 4631–4644.

9. Fernandez-Ordoñez, Y.M., & Soria-Ruiz, J. (2017). Maize crop yield estimation with remote sensing and empirical models. In: *2017 IEEE International Geoscience and Remote Sensing Symposium (IGARSS).* IEEE, pp. 3035–3038.

10. Gamon, J.A., Field, C.B., Goulden, M.L., Griffin, K.L., Hartley, A.E., Joel, G., Penuelas, J., & Valentini, R. (1995). Relationships between NDVI, canopy structure, and photosynthesis in three Californian vegetation types. *Ecological Applications*, 5, 28–41.

11. Gong, Z., Aldeen, M., & Elsner, L. (2002). A note on a generalized Cramer's rule. *Linear Algebra and its Applications*, 340, 253–254.

12. Herold, M., Scepan, J., & Clarke, K.C. (2002). The use of remote sensing and landscape metrics to describe structures and changes in urban land uses. *Environment and Planning A*, 34, 1443–1458.

13. Johnson, D.M. (2014). An assessment of pre-and within-season remotely sensed variables for forecasting corn and soybean yields in the United States. *Remote Sensing of Environment*, 141, 116–128.

14. Johnson, D.M. (2016). A comprehensive assessment of the correlations between field crop yields and commonly used MODIS products. *International Journal of Applied Earth Observation and Geoinformation*, 52, 65–81.

15. Kustas, W.P., & Norman, J.M. (1996). Use of remote sensing for evapotranspiration monitoring over land surfaces. *Hydrological Sciences Journal*, 41, 495–516.

16. Liaghat, S., & Balasundram, S.K. (2010). A review: The role of remote sensing in precision agriculture. *American Journal of Agricultural and Biological Sciences*, 5, 50–55.

17. Lykhovyd, P.V. (2020). Sweet corn yield simulation using normalized difference vegetation index and leaf area index. *Journal of Ecological Engineering*, 21, 228–236.

18. Maresma, A., Chamberlain, L., Tagarakis, A., Kharel, T., Godwin, G., Czymmek, K.J., Shields, E., & Ketterings, Q.M. (2020). Accuracy of NDVI-derived corn yield predictions is impacted by time of sensing. *Computers and Electronics in Agriculture*, 169, 105236.

19. McWilliams, D.A., Berglund, D.R., & Endres, G.J. (1999). *Soybean growth and management quick guide.*

20. Moreno, J.J.M., Pol, A.P., Abad, A.S., & Blasco, B.C. (2013). Using the R-MAPE index as a resistant measure of forecast accuracy. *Psicothema*, 25, 500–506.

21. Mukaka, M. (2012). Statistics Corner: A guide to appropriate use of. *Malawi Medical Journal*, 24, 69–71.

22. Mulla, D.J. (2013). Twenty five years of remote sensing in precision agriculture: Key advances and remaining knowledge gaps. *Biosystems Engineering*, 114, 358–371.

23. Nafziger, E. (2013). *Corn. Illinois Agronomy Handbook*. Crop Science Extension and Outreach, Urbana IL, USA.

24. Nagy, A., Fehér, J., & Tamás, J. (2018). Wheat and maize yield forecasting for the Tisza river catchment using MODIS NDVI time series and reported crop statistics. *Computers and Electronics in Agriculture*, 151, 41–49.

25. Petersen, L.K. (2018). Real-time prediction of crop yields from MODIS relative vegetation health: A continent-wide analysis of Africa. *Remote Sensing*, 10, 1726.

26. Raines, C.A. (2011). Increasing photosynthetic carbon assimilation in C3 plants to improve crop yield: current and future strategies. *Plant Physiology*, 155, 36–42.

27. Rogan, J., & Chen, D. (2004). Remote sensing technology for mapping and monitoring land-cover and land-use change. *Progress in Planning*, 61, 301–325.

28. Roozeboom, K.L., & Prasad, P.V. (2019). Sorghum growth and development In: *Sorghum: State of the Art and Future Perspectives*. American Society of Agronomy and Crop Science Society of America, Inc.

29. Rouse, J.W., Haas, R.H., Schell, J.A., & Deering, D.W. (1974). Monitoring vegetation systems in the Great Plains with ERTS. *NASA Special Publication*, 351, 309.

30. Stas, M., Van Orshoven, J., Dong, Q., Heremans, S., & Zhang, B. 2016. A comparison of machine learning algorithms for regional wheat yield prediction using NDVI time series of SPOT-VGT. In *2016 Fifth International Conference on Agro-Geoinformatics (Agro-Geoinformatics)* (pp. 1–5). IEEE.

31. Zhu, X.G., Long, S.P., & Ort, D.R. (2010). Improving photosynthetic efficiency for greater yield. *Annual review of Plant Biology*, 61, 235–261.

32. Zinke-Wehlmann, C., De Franceschi, P., Catellani, M., & Dall'Agata, M. (2019). Early within-season yield prediction and disease detection using Sentinel satellite imageries and machine learning technologies in biomass sorghum. In *Software Technology: Methods and Tools: 51st International Conference*, TOOLS 2019, Innopolis, Russia, October 15–17, 2019, Proceedings. Springer Nature, pp. 227–234.

33. Tiwari, P., & Shukla, P. (2020). Artificial neural network-based crop yield prediction using NDVI, SPI, VCI feature vectors. In *Information and Communication Technology for Sustainable Development* (pp. 585–594). Springer, Singapore.

34. Wang, R., Cherkauer, K., & Bowling, L. (2016). Corn response to climate stress detected with satellite-based NDVI time series. *Remote Sensing*, 8, 269.

35. Xu, C., & Katchova, A.L. (2019). Predicting soybean yield with NDVI using a flexible Fourier transform model. *Journal of Agricultural and Applied Economics*, 51, 402–416.

Vozhehova R.A., Maliarchuk M.P., Biliaieva I.M., Lykhovyd P.V., Maliarchuk A.S. Forecasting the yields of spring row crops by the remote sensing data

Purpose: to develop statistical models to forecast the yields of major spring row crops, namely, corn, sorghum and soybean, depending on the data of remote sensing – normalized difference vegetation index (NDVI), recorded at the critical stages of the crops growth. **Methods.** We used analytical, statistical, GIS-technologies methods to conduct the study. Remote sensing data for the NDVI computation was obtained from the satellite Sentinel-2 imagery. Regression analysis of a polynomial type was applied to work out forecasting models on the basis of true yielding data, which were recorded during the harvesting of the studied crops in the period of 2017–2018 at the experimental field of the Institute of Irrigated Agriculture of NAAS. **Results.** Statistical processing of the data revealed that regression models are suitable for accurate forecasting of the crops' yields. The best performance of the regression models was under the use of NDVI values, which were recorded at the stage of tasselling (VT) and silking (R1) for corn (the coefficient of determination is 0.9813), at the stage of second trifoliate (V2) for soybean (the coefficient of determination is 0.9829), and at the stage of half bloom (S6) for sorghum (the coefficient of determination is 0.8645). NDVI assumption in other studied stages of the crops growth led to a decrease in the accuracy of the forecasting models. **Conclusions.** NDVI is a convenient and flexible, easy-in-use tool for early yield prediction of major spring row crops. Further investigations in this field and enhancement of the performance of the developed models through the introduction of additional data and use of better computation techniques is needed to improve the quality of yield predictions.

Key words: NDVI, regression analysis, precise agriculture, corn, sorghum, soybean.

Вожегова Р.А., Малярчук М.П., Біляєва І.М., Лиховид П.В., Малярчук А.С. Прогнозування врожайності ярих просапних культур за даними дистанційного зондування

Мета: розробити статистичні моделі прогнозування врожайності основних ярих просапних культур, а саме: кукурудзи, сорго та сої залежно від даних супутникового зондування – нормалізованого диференційного вегетаційного індексу (NDVI), отриманого в критичні фази розвитку культур. **Методи.** Аналітичний, статистичний, ГІС-технологічний методи були застосовані для виконання дослідження. Дані супутникового зондування для розрахунку NDVI було отримано за зображеннями із супутника Sentinel-2. Поліномінальний регресійний аналіз було застосовано в розробленні прогностичних моделей із використанням фактичних величин урожайності, які було отримано під час збирання досліджуваних культур у період 2017–2018 рр. на дослідному полі Інституту зрошуваного землеробства НААН. **Результати.** Статистична обробка даних показала, що регресійні моделі добре підходять для точного прогнозування врожайності культур. Найкраща точність регресійних моделей була за використання величин NDVI, отриманих у фазу викидання волоті (VT) та цвітіння качанів (R1) у кукурудзи (коефіцієнт детермінації 0.9813), у фазу другого трійчастого листка (V2) у сої (коефіцієнт детермінації 0.9829) та в першу половину фази цвітіння (S6) сорго (коефіцієнт детермінації 0.8645). Величини NDVI, отримані в інші фази росту культур, приводили до зниження точності моделей. **Висновки.** NDVI є зручним і гнучким, простим у використанні інструментом раннього прогнозування врожайності основних ярих просапних культур. Подальші дослідження в цій і підвищення точності розроблених моделей шляхом введення додаткових даних і використання поліпшеної техніки обчислень необхідні для поліпшення якості прогнозування врожайності.

Ключові слова: NDVI, регресійний аналіз, точне землеробство, кукурудза, сорго, соя.