Forecasting Oil Crops Yields on the Regional Scale Using Normalized Difference Vegetation Index

Pavlo Lykhovyd[1*]

[1] Institute of Irrigated Agriculture of NAAS, Naddniprianske, 73483 Kherson, Ukraine
[*] Corresponding author's email: pavel.likhovid@gmail.com

ABSTRACT

Early prediction of crop yields on large cropland areas is of a great importance for operational planning in the agrarian sector of economy and ensuring food security. Large-scale forecasts became possible owing to the introduction of remote sensing technologies in the systems of precision agriculture, providing the information on crops conditions both on a certain field and large croplands. The study on the forecasting of major oil crop yields, namely, sunflower (Helianthus annuus L), winter rape (Brássica nápus) and soybean (Glycine max), on the regional level in Kherson oblast of Ukraine was conducted using historical yielding data and monthly MODIS Terrain NDVI smoothed time series imagery with 250 m resolution of the period from 2012 to 2019. The statistical data on the crop yields were linked to the corresponding values of monthly NDVI to determine the type of inter-relationship and work out the regression models for the oil crops yield prediction based on the remotely sensed vegetation index. The highest correlation between the yields of the oil crops and NDVI with the best prediction accuracy were obtained by using the index values at the period of April for winter rape, July for sunflower, and August for soybean. The developed regression models have reasonable accuracy with the mean absolute percentage errors of predictions reaching 25.23 percent for sunflower, 18.28 percent for winter rape, and 13.24 percent for soybean. The models are easy in use and might be recommended for introduction in theory and practice of precision agriculture.

Keywords: regression analysis, remote sensing, soybean, sunflower, winter rape.

INTRODUCTION

Early large-scale yield forecasting is of great important for modern agriculture to provide rational agricultural policy and import-export strategy, as well as guarantee food security. Recent advances in remote sensing, which is the science and technique targeted on obtaining the information about the on-land objects from the satellite imagery without direct contact [Sabins, 1987], provide extensive opportunities for monitoring and forecasting crop yields on the territories of various dimensions, for example, districts, field arrays of a whole farm, regions, states, etc. Remote sensing advancements are widely implemented in precision agriculture [Seelan et al., 2003]. The most prospective and convenient instrument for yield predictions by remote sensing data is the application of vegetation indices, calculated based on satellite imagery in special geoinformation system (GIS) software. Normalized difference vegetation index (NDVI) seems to fit best for the needs of yield forecasting because of its accessibility and simplicity of usage. Some services provide ready to use imagery with the previously calculated values of the index, so that the user has no need to deepen into the calculations and raster analysis.

Normalized difference vegetation index, which was proposed by Rouse et al. (1974), is a ratio between the difference of near-infrared and red spectra reflections from the surface and their sum. In turn, enhanced vegetation index or EVI, which was introduced later by Huete et al. (2002), has some benefits related to consideration of the distortions caused by aerosols and soil surface reflectance. Both these indices have been successfully used by scientists for estimation of crop yields. For example, both NDVI and EVI were applied in rice yield predictions by Son et al.

(2014). Mulianga et al. (2013) used the MODIS NDVI data in the study on the sugarcane yields estimation on large territories. Kouadio et al. (2014) applied the MODIS NDVI and EVI data to estimate the yields of spring wheat on the eco-district scale. Huang et al. (2013) applied the time series of NDVI values in their regression model of rice yield prediction. Vozhehova et al. (2020) and Lykhovyd (2020) applied the NDVI-based regression models for spring row crops yields prediction on the field scale. Therefore, there is several studies conducted in different regions of the Earth, which were devoted to the problem of grain crops yields estimation by the remotely sensed data. However, there is a lack of such studies for major oil crops, namely, sunflower, rape, and soybean [Lühs & Friedt, 1994; Sharma et al., 2012]. Thus, the goal of the study was to develop and test the model for major oil crops yield estimation on the regional level for Kherson oblast of Ukraine based on the values of MODIS NDVI smoothed time series data.

MATERIALS AND METHODS

The study on the estimation of major oil crops yields was carried out in Kherson oblast, which is located in the South of Ukraine (Figure 1), using the data of MODIS Terrain NDVI 16¬Day smoothed time series (with a resolution of 250 m) for the period from 2012 to 2019. The imagery was obtained using the services of the University of Natural Resources and Life Sciences in Vienna. The imagery was inserted for further processing and raster analysis in QGIS 3.10 software. Besides, all the screens were cut by the mask of vegetation cover of Kherson oblast, which was obtained at the NEXTGIS DATA service. This step has been taken to avoid the possible distortions and errors through accounting the areas, which were free from vegetation at the moment conducting the study. Further, 16-Day time series was converted into monthly series.

In order to develop the regression models for oil crops yields prediction the data on their yields in the studied region were obtained at the State Statistical Service of Ukraine. A regression analysis was performed at $p<0.05$ using the BioStat v7 software and involved the computation of Pearson correlation coefficient (R), coefficient of determination (R2) (raw, adjusted, predicted), slope, and mean absolute percentage error for the developed models during their approximation [Anscombe, 1973; Cook & Weisberg, 1982; Neter et al., 1996; Pedhazur, 1997; Stevens, 2002; Huber, 2004; Belsley et al., 2005]. Using the results of linear regression analysis, the models for each oil crop yield estimation were created.
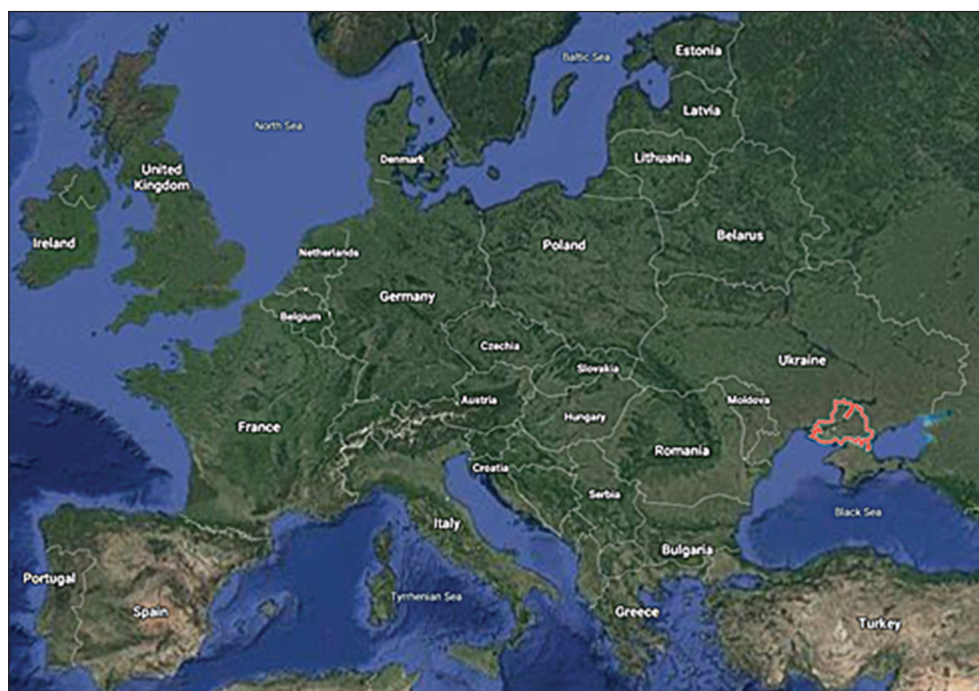

Figure 1. Kherson oblast (marked with red boarders) on the map of Europe

Table 1. The mean values of NDVI for vegetation cover of Kherson oblast during the active growing season of winter rape, sunflower and soybeans

Year	Mean						
	IV	V	VI	VII	VIII	IX	X
2012	0.39	0.46	0.50	0.51	0.50	0.49	0.47
2013	0.47	0.50	0.51	0.50	0.49	0.48	0.47
2014	0.49	0.52	0.52	0.49	0.46	0.43	0.42
2015	0.50	0.55	0.57	0.55	0.51	0.46	0.42
2016	0.51	0.56	0.57	0.56	0.53	0.49	0.46
2017	0.49	0.53	0.54	0.52	0.49	0.47	0.47
2018	0.50	0.53	0.53	0.53	0.51	0.48	0.45
2019	0.52	0.56	0.56	0.54	0.51	0.50	0.51

RESULTS AND DISCUSSION

Using the results of NDVI computation for the studied region, the mean values of the index by the months of the studied crops vegetation were obtained and generalized in the Table 1. The yields of the studied oil crops in Kherson oblast during the period from 2012 to 2019 are generalized in the Table 2.

The preliminary assessment of the relationships between the yields of the studied oil crops and NDVI values by the months were performed through the calculation of linear Pearson correlation coefficient and revealed that the highest correlation and, therefore, reasonability for regression modeling, is recorded in April for winter rape, in July for sunflower, and in August – for soybean (Table 3). Thus, the yields of winter rape, sunflower and soybean can be estimated 30–50 days in advance of the harvesting period, which in the region commonly starts at the end of June for the rape, at September for sunflower, and at the end of September – October for soybean.

Table 2. The mean values of yields of the oil crops in Kherson oblast (according to State Statistical Service), t ha^{-1}

Year	Crops		
	Winter rape	Sunflower	Soybean
2012	1.02	0.93	2.75
2013	1.80	1.44	3.60
2014	1.59	0.87	1.84
2015	1.84	1.62	3.30
2016	2.36	2.70	3.67
2017	2.45	1.34	3.05
2018	2.36	1.64	3.22
2019	2.20	1.79	3.40

Regression analysis, performed at $p<0.05$, calculated the values of slopes for the models of oil crops yields. The constant of the models was taken as "0". The regression statistics for the models of winter rape, sunflower and soybean yields is in the Table 4.

The regression statistics prove that the obtained forecasting models are quite reliable with high Predicted R^2 values exceeding 0.85, and MAPE within 13–25 percent. Such predictions could be considered as good and reasonable for winter rape and soybean, but the forecast accuracy for sunflower cannot be considered reliable enough [Caraka et al., 2019]. However, other classifiers claim that the forecasts with MAPE ranging from 10 to 20 percent are good, and from 20 to 50 percent are reasonable [Moreno et al., 2013]. Thus, it is still possible to estimate the sunflower yields on the regional scale in advance by means of the developed regression model. All in all, the best predictive model with the highest R^2 and the least MAPE was developed for soybean. The models for yield estimation for each studied crop are presented in the Table 5. A graphical approximation of the models for every studied crop is depicted in the Figure 2.

The studies previously conducted by Bu et al. (2017) proved the possibility of implementing satellite imagery in sunflower yield prediction

Table 3. Pearson correlation coefficient for the relationship between the yields of oil crops and monthly NDVI values

Crop	Month				
	April	May	June	July	August
Winter rape	0.8235	0.7851	0.6331	–	–
Sunflower	0.5859	0.7115	0.7544	0.8448	0.8303
Soybean	0.2906	0.3754	0.4662	0.6644	0.8102

Table 4. Regression statistics for the models of oil crops yields depending on the NDVI values

Criteria	Sunflower yield	Winter rape yield	Soybean yield
Mean square error (MSE)	0.2632	0.1339	0.2410
Coefficient of determination (R^2)	0.9135	0.9709	0.9788
Adjusted R^2	0.9135	0.9709	0.9788
Predicted R^2	0.8855	0.9629	0.9730
Mean absolute percentage error (MAPE). percent	25.33	18.28	13.24
Slope	2.9677	4.0723	6.2328

Table 5. Regression models of oil crops yields depending on the NDVI values

Crop	Yield estimation models
Winter rape	4.0723×NDVI±18.28%
Sunflower	2.9677×NDVI±25.33%
Soybean	6.2328×NDVI±13.24%

but they reported some limitations connected with obtaining qualitative images on time. Another study dealt with sunflower yield forecasting based on the satellite-derived vegetation indices, reported that the best prediction was obtained by using NDVI three months in advance of harvesting sunflower seeds – R^2 of the prediction model reached 0.74 [Narin & Abdikan, 2020], that is slightly less than in our study (0.89). As for winter rape, Dominguez et al. (2015) proved that Landsat satellite imagery in the case of appropriate pre-processing, might be successfully used

for winter rape growth and yield forecasting, that agrees with our results. The regression analysis of the relationship between soybean yield and NDVI at different stages of the crop growth also confirmed the possibility for precise yield forecasting using the index, with a coefficient of determination of yielding model reaching 0.80 at $p<0.01$ [Ma et al., 2001]. Some scientists used Flexible Fourier Transform Model to predict the soybean yields by NDVI values with higher accuracy [Xu & Katchova, 2019]. Thus, there are plenty of scientific studies reporting the results, which agree with ours, namely that it is possible to provide early estimation of oil crops yields by the values of remote sensing-derived vegetation indices. The models, proposed in our study, have comparatively high accuracy, and might be recommended for introduction in theoretical and practical precision agriculture.
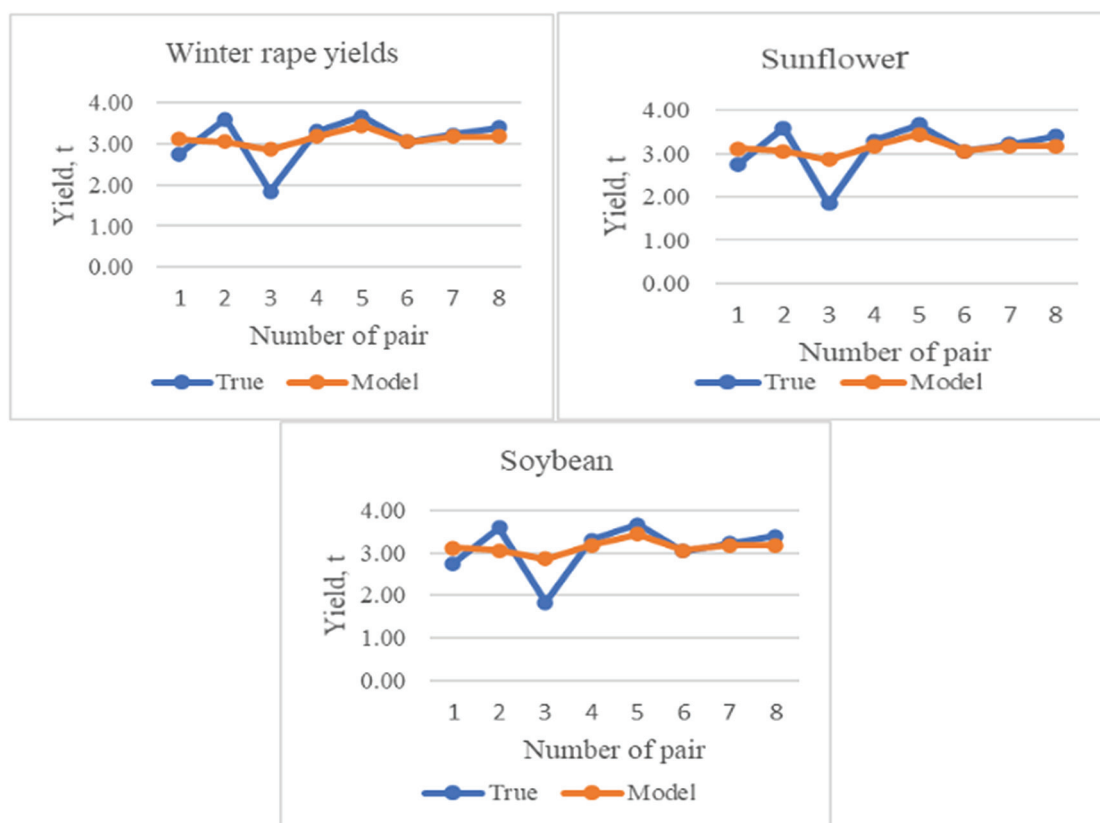
Figure 2. Approximation of the NDVI-based regression models of major oil crops yields in Kherson

CONCLUSIONS

Remote sensing provides an opportunity for fast, convenient, and precise early estimation of yields of major oil crops, namely, winter rape, sunflower and soybean, on large arrays. The regression models developed for Kherson oblast, Southern Ukraine, testify about reasonability of such an approach to yield prediction and crop growth modeling, providing the forecasts of the crops' yields with R2>0.88 at p<0.05.

Acknowledgements

REFERENCES

1. Anscombe, F.J. 1973. Graphs in statistical analysis. The American Statistician, 27(1), 17–21.

2. Belsley, D.A., Kuh, E., & Welsch, R.E. 2005. Regression Diagnostics: Identifying Influential Data and Sources of Collinearity. John Wiley & Sons, pp. 571.

3. Bu, H., Sharma, L.K., Denton, A., & Franzen, D.W. 2017. Comparison of satellite imagery and ground-based active optical sensors as yield predictors in sugar beet, spring wheat, corn, and sunflower. Agronomy Journal, 109(1), 299–308.

4. Caraka, R.E., Bakar, S.A., Tahmid, M., Yasin, H., & Kurniawan, I.D. 2019. Neurocomputing fundamental climate analysis. Telkomnika, 17(4), 1818–1827.

5. Cook, R.D., Weisberg, S. 1982. Residuals and Influence in Regression. Chapman and Hall. New York. USA.

6. Domínguez, J.A., Kumhálová, J., & Novák, P. 2015. Winter oilseed rape and winter wheat growth prediction using remote sensing methods. Plant, Soil and Environment, 61(9), 410–416.

7. Huang, J., Wang, X., Li, X., Tian, H., & Pan, Z. 2013. Remotely sensed rice yield prediction using multi-temporal NDVI data derived from NOAA's-AVHRR. PloS One, 8(8), e70816.

8. Huber, P.J. 2004. Robust Statistics. John Wiley & Sons, pp. 523.

9. Huete, A., Didan, K., Miura, T., Rodriguez, E.P., Gao, X., & Ferreira, L.G. 2002. Overview of the radiometric and biophysical performance of the MODIS vegetation indices. Remote Sensing of Environment, 83(1–2), 195–213.

10. Kouadio, L., Newlands, N.K., Davidson, A., Zhang, Y., & Chipanshi, A. 2014. Assessing the performance of MODIS NDVI and EVI for seasonal crop yield forecasting at the ecodistrict scale. Remote Sensing, 6(10), 10193–10214.

11. Lühs, W., & Friedt, W. 1994. The major oil crops. In: Designer oil crops: Breeding, processing and biotechnology., 5–71.

12. Lykhovyd, P.V. 2020. Sweet corn yield simulation using normalized difference vegetation index and leaf area index. Journal of Ecological Engineering, 21(3), 228–236.

13. Ma, B.L., Dwyer, L.M., Costa, C., Cober, E.R., & Morrison, M.J. 2001. Early prediction of soybean yield from canopy reflectance measurements. Agronomy Journal, 93(6), 1227–1234.

14. 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(4), 500–506.

15. Mulianga, B., Bégué, A., Simoes, M., Todoroff, P. 2013. Forecasting regional sugarcane yield based on time integral and spatial aggregation of MODIS NDVI. Remote Sensing, 5(5), 2184–2199.

16. Narin, O.G., & Abdikan, S. 2020. Monitoring of Phenological Stage and Yield Estimation of Sunflower Plant Using Sentinel-2 Satellite Images. Geocarto International, (just-accepted), 1–12.

17. Neter, J., Wasserman, W., & Kutner, M.H. 1996. Applied Linear Statistical Models. Irwin, Chicago.

18. Pedhazur, E.J. 1997. Multiple Regression in Behavioral Research. 3rd Ed. Orlando. FL. Harcourt Brace.

19. 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.

20. Sabins Jr, F.F. 1987. Remote sensing – principles and interpretation. WH Freeman and company.

21. Seelan, S.K., Laguette, S., Casady, G.M., & Seielstad, G.A. 2003. Remote sensing applications for precision agriculture: A learning community approach. Remote Sensing of Environment, 88(1–2), 157–169.

22. Sharma, M., Gupta, S.K., & Mondal, A.K. 2012. Production and trade of major world oil crops. In: Technological Innovations in Major World Oil Crops, Vol. 1, pp. 1–15. Springer, New York, NY.

23. Son, N.T., Chen, C.F., Chen, C.R., Minh, V.Q., & Trung, N.H. 2014. A comparative analysis of multitemporal MODIS EVI and NDVI data for large-scale rice yield estimation. Agricultural and Forest Meteorology, 197, 52–64. https://doi.org/10.1016/j.agrformet.2014.06.007

24. Stevens, J.P. 2002. Applied Multivariate Statistics for the Social Sciences. 4th Ed. NJ: LEA. Mahwah.

25. Vozhehova, R., Maliarchuk, M., Biliaieva, I., Lykhovyd P., Maliarchuk, A., Tomnytskyi, A. 2020. Spring row crops productivity prediction using normalized difference vegetation index. Journal of Ecological Engineering, 21(6), 176–182.

26. Xu, C., & Katchova, A.L. 2019. Predicting Soybean Yield with NDVI using a Flexible Fourier Transform Model. Journal of Agricultural and Applied Economics, 51(3), 402–416.