

Forecasting Crop Yields Based on Earth Remote Sensing Methods

Marzhan Anuarbekovna Sadenova*, Nail Alikuly Beisekenov, Petar Sabev Varbanov

Center of Excellence «Veritas», D. Serikbayev East Kazakhstan technical university, 19 Serikbayev str. 070000 Ust-Kamenogorsk, Kazakhstan
 MSadenova@ektu.kz

Observations of the dynamics of crop development using remote sensing data showed that in the spectral characteristics space, each crop species forms a compact cluster (a set of homogeneous photometric points) at a certain time and stage of development. It was found that derivation of images from satellite data by processing according to special algorithms in selected spectral regions allows studying plant productivity, biomass, photosynthesis intensity, and other parameters. In the present work, the authors developed a preliminary version of the algorithm for forecasting crop yield on the example of sunflower, which showed good accuracy. The method allows for early forecasting. Approximation of the dynamic curves corresponding to the values of the seven-day composite NDVI (Normalized Difference Vegetation Index) indices has been proposed using the Gaussian function and the Levenberg-Marquardt algorithm. The use of the approximating function for predicting the annual maximum NDVI on the arable land mask showed the magnitude of the average absolute prediction error depending on the predicted week ranging from 0.67 to 10.7 % per year in the evaluated period, which is acceptable accuracy for seasonal forecasts.

1. Introduction

The problem of reliable identification of agricultural plants recognition from space images is actively researched worldwide. Among the main tasks are the availability of accessible and reliable methods of crop identification, providing equipment with suitable imaging capability, and the use of flexible, mobile software systems for statistical and mathematical processing of the obtained data. To date, databases on the spectral reflectance of crops have been created. Soil quality is modelled at the level of compost processing (Lim et al., 2021). However, it is known that spectral properties of agricultural crops dynamically change during the growing season and depend on relief features and other natural conditions (Beisekenov et al, 2021a).

The presence of significant areas of agricultural production in Kazakhstan, as well as the variety of natural and climatic conditions of growing crops in the south, north and east of the country, set the task and are the basis for the development of science-based methods of identification of agricultural crops on the Earth remote sensing materials. An important indicator of agricultural production efficiency is crop yield, which is necessary for planning and regulating agricultural markets. Many methods of forecasting crop yields have been developed – such as (Makarov et al., 2020). However, they are not always adequate to the objectives of operational management of cropping patterns due to the low cognitive ability of the proposed methods for short- and medium-term forecasting. Among the methods based on estimates of physical environmental parameters, the most common are statistical methods and mechanistic models of plant growth (Jin et al, 2016). To develop an adequate mathematical model of the multifactorial process, it is important to rank the criteria under study. A new ranking system has been developed in (Beisekenov et al, 2021b), for the identification of a multitude of significant parameters, taking into account the specifics of the region. Climate change affects crop yields and disrupts the global food system (Rajakal et al, 2021). Therefore, the approach to solving the problem of yield forecasting has to build a fertility model based on space data. It should apply optimisation methods using a set of predictors derived from computer analysis of multispectral images from space.

2. Objectives and research methods

Based on the performed review, it was found that it is reasonable to use a model of crop biomass dynamics containing a small number of free parameters, taking into account the main meteorological factors, for yield forecasting. The calculation of biomass dynamics is made on the basis of meteorological observations and the NDVI index. The influence of solar radiation flux and soil moisture is taken into account not on the basis of direct measurements, but using equations with empirical constants whose values are determined for specific territories. The applicability of this approach was successfully demonstrated earlier in the Republic of Belarus when forecasting crop yields using available statistical data and satellite measurements of NDVI index (Moderate Resolution Imaging Spectroradiometer), as well as Era-Interim reanalysis (Lysenko et al., 2019). Earlier, the authors (Beisekenov et al, 2021c) developed a preliminary version of the algorithm for determining crop yield on the example of spring wheat, which showed fairly good accuracy. The method makes early predictions possible. Approximation of dynamic curves corresponding to the values of seven-day NDVI index composites was carried out using the Gauss function and Levenberg-Marquardt algorithm. The use of an approximating function for predicting the annual maximum NDVI by arable land mask in the farm "Yernar" of Glubokoe district of East Kazakhstan region yielded good results. The average absolute forecast error depending on the forecast week was from 0.4 to 9.7 % per year in the evaluated period. The use of European remote sensing satellite data and the developed in-house Python software enable the quick computation of the forecast and the possibility of correcting farmers' plans. In this paper, the study was carried out on the example of sunflower crops on the plot of the Limited Liability Partnership (LLP) "Experimental farming of oilseeds" (EFO), located in Eastern Kazakhstan. The total area of the three fields with sunflowers (Figure 1) is 264.8 ha, including field "a" of 118.5 ha, field "b" of 54.2 ha and field "c" of 92.1 ha. In the EO Browser web application, the plots of "Experimental farming of oilseeds" LLP were considered. The images of this crop were digitized in the NEXTGIS program and saved to a KML file. The file was then opened in the EO Browser web application. Sentinel-2 images were used because they have 10 m resolution and are updated every 3-5 days. The NDVI vegetation index was plotted on 13.07.2021. As a result of field mask processing in Figure 1 "a" there is quite a noticeable difference in colour coverage of NDVI index compared to fields 1 "b" and 1 "c". Probably this difference in reflectivity of sunflower fields is connected with peculiarities of land surface relief among other things.



Figure 1. The mask of the sunflower fields of "Experimental farming of oilseeds" LLP (Sentinel-2 satellite images from 13.07.2021) with the areas: "a" - 118.5 ha, "b" field - 54.2 ha and "c" field - 92.1 ha.

We used algorithms for processing satellite data obtained using the Leaflet and Sentinel Hub API; normalized NDVI values per calendar week were obtained, calculated as the average for 5 y of observations. Seven-day (weekly) NDVI index values (cloudless composites SPOT-5,6,7, KOMSPAT-3, Sentinel-2, Landsat 8) for a calendar year, calculated using the plough mask, were considered. A total of 10 time series of normalized indices corresponding to 2017-2021 were generated. The maximum NDVI value was reached during the 26th-30th calendar week, corresponding to mid-July to early August. Determination of the maximum NDVI for the early yield forecast was performed by finding the parameters of the approximating nonlinear function corresponding to the distribution of the normalized NDVI values of the previous years. Calculations were performed in the Python programming environment using the PyCharm add-on science package (Paszke et al., 2019), The 2021 sunflower yield data of "Experimental farming of oilseeds" LLP were obtained using the farm agronomists' indicator database. Annual values of gross sunflower harvest and harvested acreage were used to calculate yields. In the case of heterogeneity of the time series, the initial data were filtered by statistical methods. Figure 2 shows the statistics of changes in the NDVI of crops in the experimental plot under the letter "a", in the period from 2020 to 2021, which shows that, depending on different conditions, soil conditions, crops and other indicators of crop production vary markedly.



Figure 2. Statistics of crop NDVI indicator change in experimental plot "a" during the period from 2020 to 2021

3. Methodologies and Results

Analysis of the time series of the NDVI index by cropland mask showed that the change in values of the index during the year corresponds to a normal distribution, and, accordingly, we can use the Gaussian function to approximate the series:

$$F(i)NDVI_{max} \exp \frac{5-(i-b)^2}{2c^2} \quad (1)$$

where i is the sequence number of weeks, and b and c are the unknown parameters of the Gaussians. The solution of such a problem is usually carried out by the nonlinear least-squares method, in particular, the Levenberg-Marquardt algorithm has recently been used to solve such problems (Gavin, 2020). Table 1 shows the results of the Levenberg-Marquardt algorithm for the period from 2017-2021.

Table 1. Data obtained from the results of the Levenberg-Marquardt algorithm from the period from 2017-2021

Years	2017	2018	2019	2020	2021
Coefficients b	$b = 29.15 \pm 0.17$	$b = 30.24 \pm 0.13$	$b = 28.16 \pm 0.17$	$b = 31.06 \pm 0.16$	$b = 30.11 \pm 0.11$
	$c = 5.43 \pm 0.11$	$c = 6.61 \pm 0.15$	$c = 7.54 \pm 0.21$	$c = 6.78 \pm 0.21$	$c = 6.67 \pm 0.20$

Variable "b" in the table implies data by weeks of the calendar year, variable "c" implies data of vegetation index NDVI. As can be seen from Figure 3, built on the basis of actual data, it is quite consistent with the graph of the approximating function (Evergreen et al., 2016).

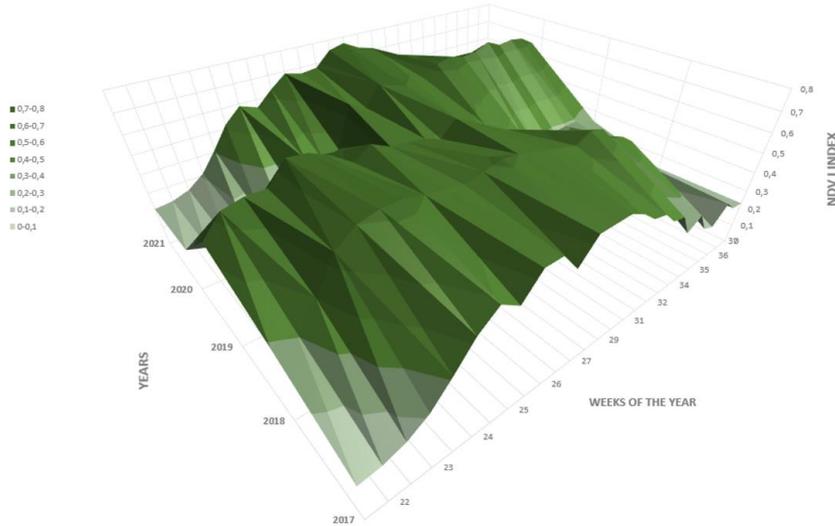


Figure 3. Plots of accumulation and Gaussian approximation of weekly NDVI index compositional values (14-43 weeks of the calendar year) for arable land in a particular field in 2017-2021

To predict maximum NDVI the actual values of $NDVI_i$, where i corresponds to the number of the calendar week, and by smoothed value when $NDVI_i$ is calculated as the average of vegetation indices ($i, i - 1, \dots, i - 3$) of calendar weeks... Maximum NDVI value was determined by the formula following from formula (1):

$$NDVI_{max} = \frac{NDVI_i}{\exp\frac{-(i-b)^2}{2c^2}} \tag{2}$$

As follows from the graphs, it makes sense to start forecasting the maximum NDVI from calendar weeks 18-20, which corresponds to the middle of May, and for an early forecast, it is more reasonable to use the value of NDVI of the last week of observations as in Eq.(2). Starting from calendar week 26, the smoothed forecast error in 2017-2021 did not exceed 7 %. To further assess the capabilities of the method, the average absolute forecast error (in %) was calculated. It was found that for a more accurate determination of the maximum from the 22-nd week, it is reasonable to use the smoothed value of the NDVI indicator. The obtained results are valid for 264.8 ha of sunflower crops and a gross yield of 15 cwt/ha in 2021 for "Experimental farming of oilseeds".

To build the forecast model, data on sown areas were selected and the gross harvest of the crops for the period 2017-2021 was evaluated. At the next stage, two regression models were built, where the average wheat yield (c/ha) was considered as the dependent variable, and maximum NDVI and the integral meteorological indices, the hydrothermal coefficient and the biological climate efficiency were considered as independent predictors. In the second model, only the maximum NDVI was considered as an independent variable. Next, for the regression model with one independent variable based on 2017-2021 data, the prediction error was estimated. In general, the developed approach to forecasting crop yields can be represented by the diagram in Figure 4.

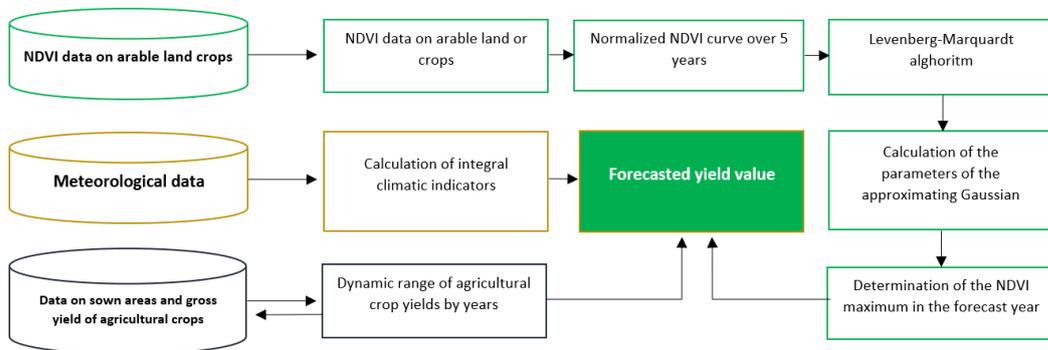


Figure 4. General scheme of crop yield forecasting at the regional level

The results of the predictive model were used to determine the maximum NDVI for 2022. The forecast model included weather data, 5-year arable NDVI data, and gross yield for 2021. All actual and simulated arable NDVI

values in the test region for the period 2017-2022 are presented in the table below. Based on the results of the Levenberg-Marquardt algorithm and the forecast model predicted using machine learning and auxiliary libraries of the Python programming language, the following result is obtained: in 2022: $b = 31.04 \pm 0.15$, $c = 8.25 \pm 0.32$. The 5-year average absolute forecast error in the 28-31 calendar week was 2.1-4.5 %, in the 24-27 week from 4.2 to 7.1 %, and in the 21-24 week from 4.9-8.5 %. Methodological provisions for yield forecasting in practice were used. Using data on the actual yield, meteorological data and NDVI data on arable land for 5 years of sunflower cultivation in the "Experimental farming of oilseeds" for 2017-2021 All calculations were performed using the methodology developed by the authors of sunflower yield forecasting for 2021-2022 The obtained data of the forecast model will be used for further comparison with actual yield data for 2022 in the same period to assess the quality of the model as a yield forecasting tool. Having developed the entire mathematical model in Jupyter Notebook, we obtained the result shown in Figure 5.

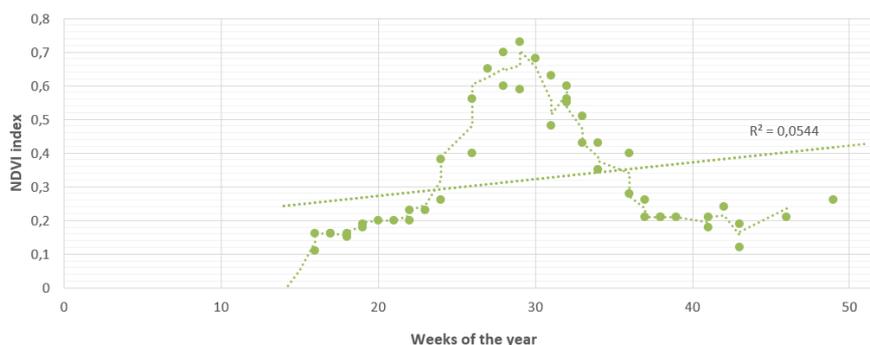


Figure 5. Scatter diagram of the results of yield forecasting for 2022 using Python software packages

Based on the results of building a predictive model based on all the data obtained using machine learning and auxiliary libraries of the Python programming language, the result was: in 2022: $p = 17.4 \pm 1.21$ c/ha. In order to assess the quality of the yield prediction model, the obtained result was compared with the result for 2021, relative to this year, the predicted result revealed a yield increase of 16 %. Using the approximating function to predict the maximum NDVI forecast mask of arable land in the example of LLP "Experimental farming of oilseeds" in East Kazakhstan is consistent with the results obtained earlier by the authors for the wheat crop. The average absolute error of the forecast ranges from 0.67 to 10.7 % per annum in the simulated period. This average absolute forecast error is calculated on the basis of results for previous years.

4. Conclusions

To date, satellite methods are the most promising among other methods of yield forecasting due to their objective and rapid results, as well as the possibility of covering large areas. Assimilation of satellite remote sensing data with various mathematical models significantly reveals the possibilities of assessing and forecasting the state of soil-vegetation systems of agro landscapes. Thus, the use of remote sensing data and developed Python software modules contribute to the rapid formation of forecasts and the possibility of adapting farmers' plans. Based on the Levenberg-Marquardt algorithm, the Python programming language with the Jupyter Notebook development environment was chosen because of the availability of a set of analytical visualization tools. The authors developed a variant of the algorithm for determining crop yields using sunflower as an example, which showed good accuracy. The method makes it possible to make early predictions. Approximation of dynamic curves corresponding to the values of seven-day composites of NDVI indices is proposed using the Gaussian function and the Levenberg- Marquardt algorithm. The use of the approximating function for forecasting the annual maximum of NDVI by arable land mask in East Kazakhstan showed good results: the average absolute forecast error depending on the forecast week was from 0.67 to 10.7 %/y, in the simulated period.

Acknowledgements

This research has been supported by Project IRN BR10865102 "Development of technologies for remote sensing of the earth (RSE) to improve agricultural management", funded by The Ministry of Agriculture of the Republic of Kazakhstan.

References

- Beisekenov N.A., Sadenova M.A., Varbanov P.S., 2021a. Mathematical Optimization as A Tool for the Development of 'Smart' Agriculture in Kazakhstan. *Chemical Engineering Transactions*, 88, 1219-1224.
- Beisekenov N.A., Sadenova M.A., Kulenova N.A., Mukhtarkanovna M.A., 2021b. Development of a preliminary version of a model for machine learning in predicting yield on the example of wheat in the conditions of East Kazakhstan, 16-th International Conference on Electronics Computer and Computation (ICECCO), 1-6, doi: 10.1109/ICECCO53203.2021.9663758.
- Beisekenov N.A., Anuarbekov T.B., Sadenova M.A., Varbanov P.S., Klemeš J.J., Wang J., 2021c. Machine Learning Model Identification for Forecasting of Soya Crop Yields in Kazakhstan. 6th International Conference on Smart and Sustainable Technologies (SpliTech), 2021, 1-6, doi: 10.23919/SpliTech52315.2021.9566376.
- Evergreen S.D.H., 2016. *Effective data visualization: The right chart for the right data*. Sage Publications, Los Angeles, United States, ISBN 9781506303055.
- Gavin H.P., 2020. The Levenberg–Marquardt method for nonlinear least squares curve-fitting problems. *Dep. Civ. Environ. Eng., Duke Univ.* <<https://people.duke.edu/hpgavin/ce281/lm.pdf>>, accessed 23.09.2019.
- Jin X., Kumar L., Li Z., Xu X., Yang G., Wang J., 2016. Estimation of Winter Wheat Biomass and Yield by Combining the AquaCrop Model and Field Hyperspectral Data. *Remote Sensing*, 8(12), 972, doi: 10.3390/rs8120972.
- Lim L.Y., Lee C.T., Bong C.P.C., Lim J.S., Ong P.Y., Klemeš J.J., 2021. Selection of Parameters for Soil Quality Following Compost Application: A Ranking Method, *Chemical Engineering Transactions*, 83, 505-510.
- Lysenko S.A., 2019. Forecasting of the yield of agricultural crops based on satellite monitoring of the carbon dynamics of the terrestrial ecosystems. *Issledovaniye Zemli iz kosmosa*, 2019(4), 48–59. (in Russian)
- Makarov V.Z., Gusev V.A., Shlapak P.A., Reshetarova D.A., 2020. Choice of the optimal method for recognition of agricultural crops from high-resolution space images (on the example of the Saratov Trans-Volga region), *Izvestiya Saratov, new series of Earth Sciences* 20, 3-4.
- Paszke A., Gross S., Massa F., Lerer A., Bradbury J., Chanan G., et al., 2019. PyTorch: An Imperative Style, High-Performance Deep Learning Library. In: *Advances in Neural Information Processing Systems* 32. Curran Associates, Inc.; 2019, 8024–35.
- Rajakal J.P., Andiappan V., Wan Y.K., 2021. Mathematical Approach to Forecast Oil Palm Plantation Yield under Climate Change Uncertainties, *Chemical Engineering Transactions*, 83, 115-120, doi:10.3303/CET2183020.