CHEMICAL ENGINEERING TRANSACTIONS

VOL. 94, 2022

A publication of
AIDIC
The Italian Association
of Chemical Engineering
Online at www.cetjournal.it

Guest Editors:Petar S. Varbanov, Yee Van Fan, Jiří J. Klemeš, Sandro Nižetić

ISBN978-88-95608-93-8;**ISSN** 2283-9216

DOI: 10.3303/CET2294003

Modelling of Forecasting Crop Yields Based on Earth Remote Sensing Data and Remote Sensing Methods

Marzhan Anuarbekovna Sadenova*, Nail Alikuly Beisekenov, Yerbolat Tleubekovich Ualiyev, Natalya Anatolyevna Kulenova, Petar Sabev Varbanov

Priority Department Centre «Veritas» D. Serikbayev East Kazakhstan technical university, 19 Serikbayev str. 070000, Ust-Kamenogorsk, Kazakhstan
MSadenova@ektu.kz

In this work, the authors proposed a method of determining the yield of spring wheat based on the analysis of the dynamics of spectral parameters of its growth and development, determined by multispectral satellite images. It was found that by processing the satellite images of the fields in selected spectral regions, it is possible to estimate with a high degree of reliability the productivity of plants, biomass value, photosynthesis intensity and other parameters. By means of mathematical processing of the statistical data array of phosphorus, potassium and nitrogen content in the soil according to the Remote Sensing (RS) data in comparison with the actual data obtained after agrochemical analysis of soil samples, the total value of the average error (the average absolute error ranging from 24 to 36 % for the analysed period) was calculated. Using remote sensing data and Convolutional Neural Networks (CNN), the forecast of spring wheat yield in the conditions of soil and climatic zone of Eastern Kazakhstan was carried out. The results obtained with the predictive model are close to the actual yield results of the previous year (the error smaller than 9 %), indicating the relationship between yield and agrochemical analysis of the soil.

1. Introduction

Satellite methods of yield forecasting are the most promising among other methods due to their objectivity, efficiency, and the possibility to cover large territories. The use of the results of yield forecasting based on satellite data along with the results obtained by other methods allows for improving the quality of forecasts and their advance (Rembold et al., 2013). Over the past decade, several methods and approaches to crop yield forecasting have been formed, the main among which are: the method of trend analysis and cyclicality in yield dynamics; the method based on identifying the year-analogue; modelling of plant biomass growth; the method based on the analysis of synoptic processes; regression method using satellite data. The approach based on regression analysis, if a sufficiently extended series of quality satellite data is available, can provide good results (Bouras et al., 2021). Many factors affect crop yields, such as crop genotype, environment, and management practices. The spatially and temporally changing environment has a huge impact on the fluctuations in crop yields from year to year and from place to place. Under such conditions, accurate forecasting of yields is very useful for world production and food. Based on accurate forecasts, import and export decisions can be made in a timely manner. Farmers can use yield forecasts to make informed management and financial decisions. It is possible to predict the performance of new hybrids in new and untested sites (Khaki et al., 2019). However, successful yield prediction is very difficult due to many complex factors. For example, genotype and environmental factors often interact with each other, making yield prediction more difficult. Environmental factors such as weather components often have complex non-linear effects that are difficult to accurately estimate.

Currently, works on the development of this approach and analysis of the possibilities of its application for yield forecasting of various crops are being carried out quite actively. The need for research in this area is due to unreliable and/or distorted information provided by land users on current crop yields as well as the condition of agricultural lands. Traditional cartographic materials in the form of land management schemes and household planning are often poorly informative or inaccurate due to the use of outdated, irrelevant information bases.

Paper Received: 25 April 2022; Revised: 30 May 2022; Accepted: 03 June 2022
Please cite this article as: Sadenova M.A., Beisekenov N.A., Ualiyev Y.T., Kulenova N.A., Varbanov P.S., 2022, Modelling of Forecasting Crop Yields Based on Earth Remote Sensing Data and Remote Sensing Methods, Chemical Engineering Transactions, 94, 19-24 DOI:10.3303/CET2294003

State statistics and accounting of agricultural land, in this case, become biased, making it difficult to assess the effectiveness and control of the yield and use of agricultural land. From the results of studies given in (Beisekenov et al., 2021a), annual yield fluctuations are quite accurately predicted by spectral indices during the growing season, during which all the vital processes of plants are maximally manifested. According to van Klompenburg et al. (2020), the widely used deep learning algorithms for crop yield prediction are Convolutional Neural Networks (CNN). The results of many works show that combining data from several sources is superior to models based on only one data set. This method will be able to greatly facilitate the assessment of soil conditions by analysing land remote sensing data, especially useful for countries with large areas of cultivated land, where access to soil samples is difficult. Based on this analysis, Convolutional Neural Networks were adopted as the basic model, with further improvements. The novelty of this study lies in the fact that such studies have not previously been given on the example of spring wheat in East Kazakhstan. The purpose of this work is to assess the yield and current condition of agricultural soils through reliable satellite land assessment based on the use of remote sensing data.

2. Object and methodology of research

The studied site is located at the coordinates 82.7541 °E and 50.0809 °N and occupies an area of 126 ha on the territory of the farm "Oilseeds Experimental Farm" (OEF) of Eastern Kazakhstan. According to the taxonomy of soils, the chernozem type of soil is the dominant soil unit in the territory of this site. The geographical location of the selected representative profiles is shown in Figure 1.

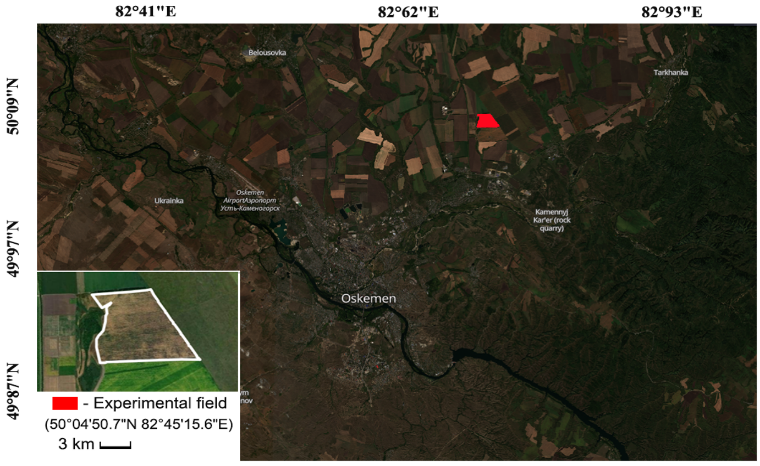

Figure 1: Study area and experimental field (*all data was obtained from Sentinel-2 satellite images*)

Figure 2 illustrates the methodology for the analysis and calculation of agrochemical indicators, the correlation of data obtained by the method of soil selection and also using land remote sensing data.

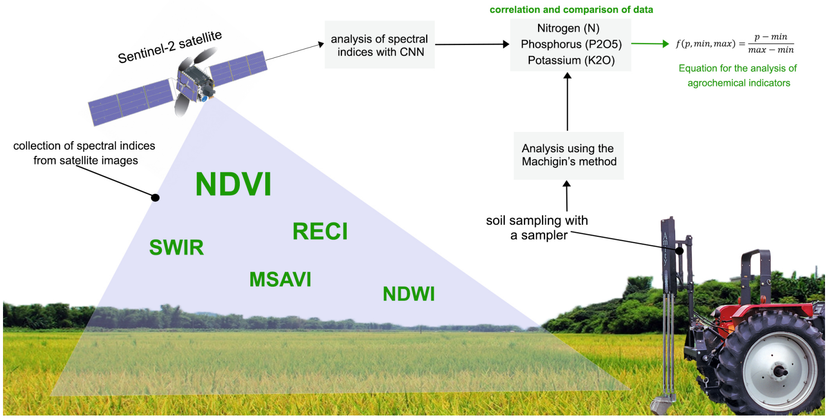

Figure 2: Methodology for conducting the experiment for the analysis of agrochemical indicators

In this study, agrochemical indicators of soil were evaluated by soil sampling in the field. Soil sampling was carried out on September 29, 2021, by envelope method in 5 points of the experimental plot presented in Figure 3. The staging was made under the agricultural crop: spring wheat. The selected soil was analysed using Machigin's method modified by TSINAO (GOST 26205-91, 2020) for exchangeable potassium, mobile phosphorus and total nitrogen. All three chemical elements are very important for productive growth and soil enrichment. Nitrogen is vital to plants for the proper development of the root system. All metabolic processes occurring in the plant body, from chlorophyll synthesis to vitamin assimilation, are activated by nitrogen. Lack of nitrogen can lead to incomplete yields or even death of the crop.

Phosphorus, along with Nitrogen and Potassium, is an essential element, but for successful growth and development, as well as fruiting, the right balance of these substances is important. First of all, it is to maintain the metabolism and strength of the growing plant. Also, phosphorus is involved in the regulation of plant respiration. Phosphorus is important for the normal development of roots and reproductive organs of crops. Potassium, along with nitrogen and phosphorus, are among the major plant nutrients. This important component helps regulate plant water balance through roots (osmotic gradient) and leaf stomata functioning. Potassium promotes starch and sugar accumulation in fruits and increases plant resistance to fungal and microbial diseases and insect damage. Adequate levels of potassium increase carbohydrate production and transport in the plant. Potassium plays an important role in increasing plant resistance to low temperatures, salinity, drought and disease. Based on the QGIS survey data, an agrochemical cartogram was developed.

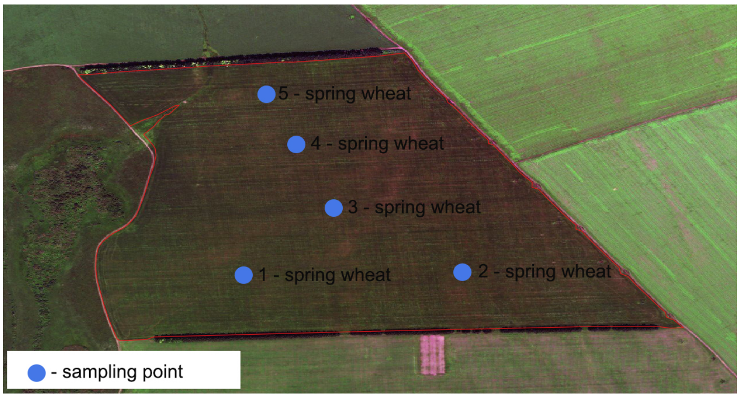

Figure 3: Soil sampling points at the experimental site (*all data was obtained from Sentinel-2 satellite images*)

A similar cartogram was constructed using remote sensing data. Sixteen-day composite images with 250 m resolution from the Sentnel-2 satellite were used as remote sensing data. Agrochemical indices on the experimental plot were evaluated using spectral indices: NDVI - vegetative development of plants; NDWI - normalised difference water index, SWIR - short-wave infrared range; MSAVI - modified soil vegetation index; RECI - chlorophyll index, as well as the elevation map. The main idea of the technique is an original approach to the analysis of agrochemical elements using land remote sensing data using multispectral images from the Sentinel-2 satellite using CNN Eq.(1).

$$f(p, min, max) = \frac{p - min}{max - min} \tag{1}$$

where f is the normalisation function; p is the value of the spectral index (space image); min – minimum pixel value of 0; max - maximum pixel value of 255. The input layer takes into account the two-dimensional image topology and consists of several maps (matrices), consisting of 3 maps, where each map corresponds to an image with a specific channel (red, blue and green). The input data for each specific pixel value is normalised to a range of 0 to 1.

3 Results and discussion

Agrochemical indicators of soil are an important characteristic for differentiated fertiliser application. If a single fertiliser rate is used in a field, some areas may have an excess of nutrients in other areas, a deficit. Differentiated fertiliser application takes into account the heterogeneity of the soil and chooses the exact fertiliser rates for all areas of the field. According to the results of agrochemical analysis on selected soils, all three elements: exchangeable potassium, mobile phosphorus and total nitrogen, showed excessive content of substances in the soil relative to the norm. Cartogram based on actual agrochemical data obtained during soil sampling is presented in Figure 4. Figure 5 shows spectral indices characterising the state of vegetation in the experimental field calculated and presented by the authors of the article based on the processing of images received from Sentinel-2 satellite, which are in the public domain. Colour gradation from light blue to dark green on the analysed cartograms reflects the content of a quantitative indicator of one or another agrochemical indicator.

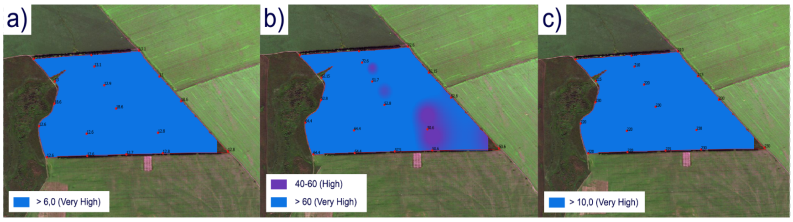

Figure 4: Cartogram of the content of agrochemical indicators in the soil on selected samples of the experimental field (all data were obtained from Sentinel-2 satellite images)

Figure 4 shows a) - a cartogram of the content of mobile phosphorus in the soil (mg/100 g); b) - a cartogram of exchangeable potassium in the soil (mg/100 g); c) - a cartogram of total nitrogen in the soil (mg/100 g).

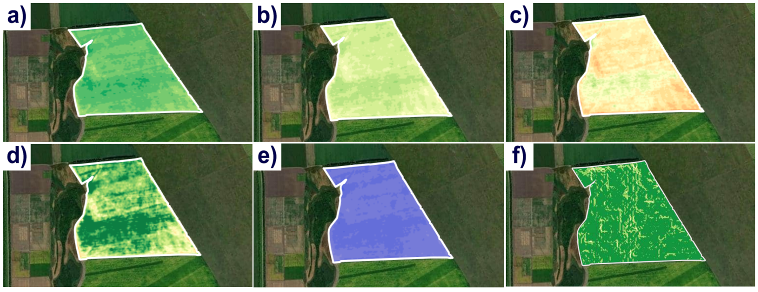

Figure 5: Spectral indices map of the test site (all data were obtained from Sentinel-2 satellite images) a)-NDVI; b) -NDRE; c) MSAVI; d) - RECI; e) -NDWI; f) - elevation map

Table 1: Percentage comparison of the ratio of agrochemical indicators of soil composition obtained by RSE and results obtained by chemical soil analysis

Collection points	Nitrogen (N), mg/100 g of soil		deviation %	Phosphorus (P2O5), mg/100 g of soil		deviation %	Potassium (K2O), mg/100 g of soil		deviation %
	Measured	RSE		Measured	RSE		Measured	RSE	
1	210	142	47.89	13.1	8.5	54.12	72.6	60.6	19.8
2	220	134	64.18	12.9	10.4	24.04	51.7	46.8	10.47
3	230	148	55.41	18.6	12.6	47.62	52.8	45.2	16.81
4	220	195	12.82	12.6	9.8	28.57	64.4	50.7	27.02
5	230	160	43.75	12.8	7.8	64.1	50.6	43.2	17.13

These spectral indices quantitatively depend on canopy structure, optical properties of plant elements such as the angle of the sun, atmospheric conditions, and angular configuration in general, as well as biophysical parameters characterising the projective coverage of the soil surface, calculated on the basis of remote sensing data. By statistical processing of the data obtained by direct experiment with soil sampling and analysis of satellite images, the correlation and deviations of data stacks, with similar evaluation criteria, as well as units of values, are calculated. This method is used in modelling and forecasting spring wheat yield. Based on the results of the calculation and analysis of agrochemical indicators using multispectral images from Sentinel-2 satellite, a large number of spectral indicators were extracted with the selection of required parameters. An artificial neural network architecture based on Convolutional Neural Networks was developed to calculate the agrochemical data. As input data, we used spring wheat yield data, agrochemical data for the previous year, as well as meteorological data for the experimental site. According to the results of agrochemical analysis according to RS data, all three elements: exchangeable potassium, mobile phosphorus and total nitrogen, showed a similar excess in the soil relative to the norm. According to the results of agrochemical indicators using RS data, a comparative analysis with the actual results of agrochemical analysis was carried out. The result of this comparison is presented in Table 1. According to the results of the comparative analysis presented in Table 1, the total value of the average error of the result obtained with the help of remote sensing relative to the actual data was, on average, 43 %. Despite such a high deviation, the value of the average error from the actual data is quite acceptable and fairly well reflects the excessive amount of elements in the soil. From the data in Table 1, it is clearly visible how the indicator of total nitrogen obtained from remote sensing, compared to the other elements, significantly deviated from the actual figures by more than 55 %, and this case is justified by the data of agrochemical analysis for the last year used as an input parameter. The availability of more input data for the machine learning model would increase the probability of a more accurate analysis of agrochemical indicators using RS data.

The developed method was created as an alternative way of obtaining agrochemical indicators for farmers in case of a hard-to-reach site and also allows saving money due to more accurate determination of soil composition and quantity of fertilisers. As a result, this method eliminates the need for annual soil sampling on agricultural lands. In order to obtain a reliable yield forecast for the next year, a model based on a machine learning CNN was built based on spectral indices, actual yields of previous years, meteorological data, and agrochemical indicators. In the first phase of the study, a large number of spectral indices were extracted from the obtained images with the selection of necessary parameters. Data on agrochemical indicators, weather, actual crops and spectral data were collected. A CNN based method was used as a machine learning algorithm. This algorithm showed good yield prediction results in a paper by Van Klompenburg et al. (2020). For the convolutional neural network, the number of output features in each dimension can be calculated using equation 2 below:

$$n_{out} = \left\lfloor \frac{n_{in} + 2p - k}{s} \right\rfloor + 1 \qquad (2)$$

where n_{in} - number of input functions; n_{out} - number of output functions; k - convolution kernel size; p - convolution size; s - convolution step size. As input functions into the equation are used: weather data, yield data, spectral indices, and agrochemical indicators. The coefficient of determination (R2) was used to evaluate the performance of the developed model in this study; this algorithm showed some of the best results in the work of Beisekenov et al. (2021b). The coefficient algorithm (R2) reflects the degree of a linear relationship between the observed and predicted yields of spring wheat (3 equation).

$$R^2 = \frac{\left(\sum_{i=1}^{n}(O_i - \bar{O})\,(F_i - \bar{F})\right)^2}{\sum_{i=1}^{n}(O_i - \bar{O})^2 \sum_{i=1}^{n}(F_i - \bar{F})^2} \qquad (3)$$

An overview of the method of yield forecasting from multispectral images using convolutional neural networks is presented as a flowchart in Figure 6.

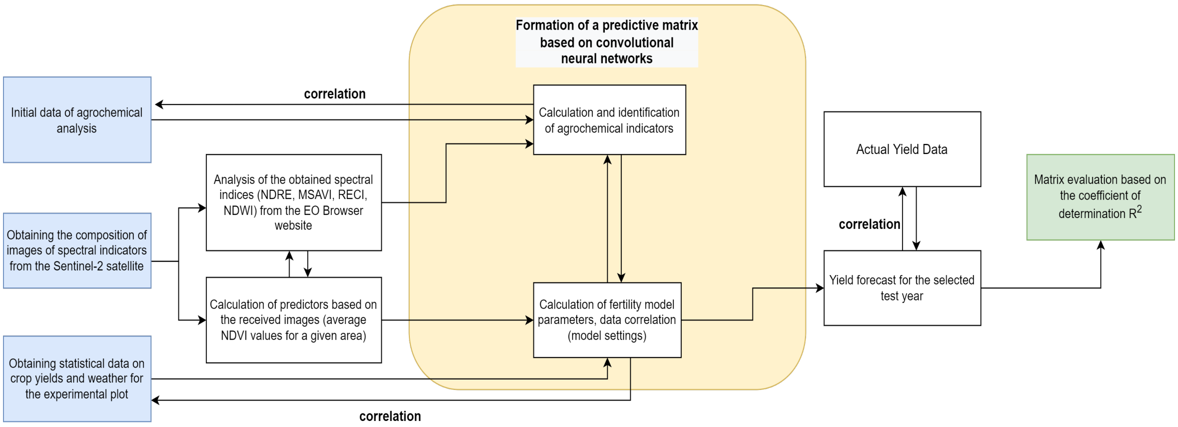

Figure 6: Method of yield forecasting from multispectral satellite images using a multivariate linear regression method

As a result of applying the developed mathematical model for predicting the yield of grain crops, the value of the average absolute error was 0.59 - 8.9 %/y. in the simulated period. The developed algorithm for determining the yield showed fairly good accuracy on a grain crop - spring wheat. The result was evaluated using a matrix assessment based on the algorithm of the coefficient of determination R^2 method.

4. Conclusion

Based on the results of computational experiments, the possibility of predicting the yield from multispectral satellite images using real statistical data on the yield of spring wheat in East Kazakhstan has been demonstrated. An analysis of the accuracy of yield forecasting was carried out, which showed fairly good estimates using remote sensing data. In further studies, it is planned to modify the constructed models to improve the accuracy and generalization in predicting yields within the framework of the formulated research areas. Compared with the data obtained by taking soil samples and using the analysis of remote sensing data, the total value of the average error of the obtained data on the content of exchangeable potassium, mobile phosphorus and total nitrogen in the soil was about 24-36 %. The resulting indicator of deviation compared to the actual indicators is within acceptable values, which indicates the feasibility of using this method. As a result of the study, the developed mathematical model for predicting grain crop yield showed the value of the average absolute error of 0.59 - 8.9 %/y. in the simulated period by using the calculation of machine learning algorithm: the coefficient of determination R^2. The results of the study showed that the use of remote sensing methods makes it possible to carry out a qualitative analysis of the ecological state and assessment of agricultural lands. The developed method allows farmers to save money by more accurately determining the composition and the amount of fertilisers in the soil. Future work should focus on developing a method for targeting and assessing the savings, as well as a mechanism for updating and improving the model with new data. The results are especially useful for countries with large areas of cultivated land where access to soil samples is difficult.

Acknowledgements

This research has been supported by the Project 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., Kulenova N. A., Mukhtarkanovna M.A., 2021a. 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. 2021 16th International Conference on Electronics Computer and Computation (ICECCO), 1-6, DOI: 10.1109/ICECCO53203.2021.9663758.

Beisekenov N.A., Sadenova M.A., Varbanov P.S., 2021b. Mathematical Optimization as A Tool for the Development of "Smart" Agriculture in Kazakhstan, Chemical Engineering Transactions, 88, 1219-1224.

Bouras Eh., Jarlan L., Er-Raki S., Balaghi R., Amazirh A., Richard B., Khabba S., 2021. Cereal Yield Forecasting with Satellite Drought-Based Indices, Weather Data and Regional Climate Indices Using Machine Learning in Morocco. Remote Sensing. 13(16)3101, DOI: 10.3390/rs13163101.

GOST 26205-91, 2020. Determination of mobile phosphorus and potassium in carbonate soils by the Machigin method in the modification of TSINAO: GOST 26205-91, Moscow, 2020 (in Russian).

Khaki S., Wang L., Archontouli S.V., 2020. A CNN-RNN framework for crop yield prediction. Frontiers in Plant Science, 10, 1750.

Rembold F, Atzberger C, Savin I, Rojas O., 2013. Using Low Resolution Satellite Imagery for Yield Prediction and Yield Anomaly Detection. Remote Sensing, 5(4), 1704-1733, DOI: 10.3390/rs5041704.

Van Klompenburg T., Kassahun A., Catal C., 2020. Crop yield prediction using machine learning. A systematic literature review. Computers and Electronics in Agriculture. 177.105709.