DOI: 10.32347/2412-9933.2024.57.67-71
UDC 631.165:528.88

Huang Mingxin
PhD student in the Department of Information Systems and Technologies,
https://orcid.org/0000-0002-2365-9815
Shevchenko National University of Kyiv

DEVELOPMENT OF A YIELD MONITORING MODEL BASED ON ANALYSIS OF SURVEYS AND IMAGES OF THE FIELD

Abstract. *The study is devoted to the construction of a mathematical model of the yield of agricultural crops, which includes three components: trend, seasonal, and random. The developed model shows the dependence of yield on phenological indicators, the quality of land resources, management efficiency, and other random factors. The trend and seasonal components of the yield model do not depend on random factors and can therefore be used to predict yield. It is proposed to consider the trend component of the model as a piecewise linear function, and the seasonal component of the model as a linear harmonic regression. A method for analyzing multispectral images with consideration of geoinformation data has been developed to assess phenological indicators. This method includes determining the threshold value using Otsu method to find the density of the agricultural crop in the field. Data on crop density, supplemented with geodata about the plot boundaries, are used to calculate the yield. A comparison of yield forecasts for three crops in the Chernihiv region was made using observations of the phenological indicators of crops throughout the year and over 3 months. It was found that yield is significantly determined by plant development in the first months after germination. The comparison of yield forecasts was made with data from the State Statistics Service of Ukraine and forecasts made using the WOFOST simulation model. It was established that the average relative error of yield prediction using the developed model is 2.96% when using observations of the phenological indicators of crops throughout the year and 4.51% when observing over 3 months. This accuracy is sufficient and comparable to the average accuracy of predictions based on the WOFOST model, which is 3.62%).*

Keywords: yield monitoring; GIS; project management; biomonitoring; information management; critical infrastructure

Introduction

Population growth and the demand for food quality are the reasons for the intensification of the struggle for limited land resources. The desire to maximize profit leads to the degradation of land resources and the loss of biodiversity. Global climate changes also have a negative impact. Therefore, the rational use of land to preserve and increase its productivity and to support life in ecosystems is a current task.

The use of geographic information systems is an important tool for managing sown areas, analyzing yields, and predicting future yields. Currently, the concept based on Site Specific Crop Management (SSCM) is popular. An important task of SSCM is to increase yield, i.e., to ensure the quality and quantity of agricultural production obtained. SSCM is implemented based on GPS technology, which determines the location of objects and the condition of plants in agricultural areas through a satellite network. The RTK (Real Time Kinematic) technology allows for high-precision object location in the field. After collecting information about soil conditions, SSCM technology allows for the management of fertilizers, chemicals, herbicides, pesticides, and other substances. The application of this technology is relevant not only for Ukraine but also for China, where in some regions farmlands yield several harvests a year. This requires prompt management of the ripening process of agricultural crops. Equipment using a GPS module allows for the quick processing of large areas and minimizes human impact.

The density and health of agricultural crops can be monitored by analyzing field surface images obtained at regular intervals. Such time series of images contain valuable information about chlorophyll activity, plant activity, possible presence of diseases or pests. The analysis of multispectral field images allows for the identification of necessary indicators that determine crop growth, ripening, and so on.

In the study [1], a hypothesis is described suggesting that the yield of agricultural crops can be directly determined by the heterogeneity in field images. The processing of digital field images provides valuable information on the state of the agricultural crop, allowing for an assessment of plant health and the prediction of yield, timing, quantity, and quality of future production.

Images obtained from remote sensing of the Earth's surface or by other means are converted into time series, which must be analyzed to consequently make informed decisions about managing these areas. Specialized software is necessary for processing time series of digital images for decision-making in agriculture, capable of handling both one-dimensional and multi-dimensional time series data at a high level [2]. Another limitation of modern geoinformation systems and software packages for image processing and pattern recognition is that these systems do not provide all the necessary functions for time series processing and require data storage arrays for image access capabilities, i.e., they do not offer optimization for efficient monitoring of the state of agricultural crop areas [3].

The study [4] describes the capabilities of the BFAST method (Breaks For Additive Seasonal and Trend) to detect long-term phenological changes in time series of images. This method combines change detection techniques in time series behavior with decomposition methods that break down series into components defining trend changes, seasonal changes, and random components. This method is based solely on the integral indicators of the area and does not take into account their geographical distribution.

The purpose of the article

The aim of the study is to develop a model for monitoring the yield and growth activity of agricultural crops based on the analysis of time series obtained from the processing of multispectral images of the field surface, obtained at certain time intervals. This will increase the efficiency and justification of decisions in agriculture.

To achieve the goal, the following tasks were set:

1. Develop a mathematical model to represent the relationship between phenological indicators and the yield of agricultural crops. In developing the model, describe other factors affecting yield.
2. Develop a method for analyzing multispectral images with consideration of geoinformation data.
3. Apply the developed yield monitoring model for an agricultural crop based on the analysis of geodata and plot images to predict the yield of agricultural crops.
4. Conduct verification of the developed model and methods.

Presenting main material

For the model construction, let's consider the assumption on which the research is based: there is a causal relationship between the quality of land resources, management decisions, phenological indicator values, and the yield of agricultural crops. The main phenological indicator used in the analysis of sowing area images is the Normalized Difference Vegetation Index (NDVI). It is a quantitative indicator of photosynthetically active biomass used for the quantitative assessment of vegetation cover, calculated by the formula:

$$NDVI = \frac{N-R}{N+R}, \qquad (1)$$

where R is the intensity of light reflected from the surface area in the visible red spectrum (630-690 nm), and N is the intensity of light reflected in the infrared spectrum (760-900 nm).

The assumption implies that there is a functional dependence between yield and NDVI. That is, the yield V can be expressed as some function g:

$$V = g(\beta, Z, M, e, t), \qquad (2)$$

where β is the NDVI value, Z is the indicators of land resource quality, M is the management efficiency indicators, e are other factors also affecting the yield, and t is time. Under other factors, weather conditions, natural disasters, epidemics, and random processes that cannot be predicted and are not subject to influence are implied.

As a consequence of the assumption, it follows that multispectral field images can be used to predict yield.

Let's build a model for monitoring the yield of an agricultural crop based on the analysis of geodata and plot images. For this, consider that according to the BFAST method, the additive decomposition model of time series forecasting of phenological indicators is:

$$\beta_t = T_t + S_t + e_t, \qquad (3)$$

where β_t are the time series data, i.e., the phenological indicator values recorded at time t, T_t is the trend component, S_t is the seasonal component, and e_t are the random components, with t being the number of observations, i.e., the number of elements in the time series.

Considering (2) and (3), the model for monitoring the yield of an agricultural crop based on the analysis of geodata and plot images will look like this:

$$V = T(\beta, Z, M, t) + S(\beta, Z, M, t) + e(t), \qquad (4)$$

where T defines the trend component, S defines the seasonal component, and e are the random components. The trend and seasonal components do not depend on random factors and therefore can be predicted using corresponding methods.

To predict the yield of an agricultural crop using the developed model, it is necessary to have a knowledge base with reference images corresponding to various growth indicators of defined agricultural plants. By comparing the current image with images from the knowledge base to establish similarities, conclusions can be made about possible deviations from normal growth and maturation of the crop. In case of negative deviation, necessary decisions are required to correct the situation. The developed model can also be used to predict the amount of harvest. However, this forecast must be adjusted considering other factors affecting plant growth and maturation.

The input data for the crop yield assessment are images of the plot, which can be represented as a certain matrix of size M×N. Agricultural crops do not grow uniformly in the field. Therefore, it is necessary to consider the concept of crop distribution density as the ratio of the area where the crop grows to the total field area. A portion of the image pixels corresponds to the part of the field where the crop grows. Other pixels correspond to parts of the field where no crop grows. Let's introduce a threshold function:

$$\delta(x,y) = \begin{cases} 1, \ if \ \beta(x,y) \geq B \\ 0, \ if \ \beta(x,y) < B' \end{cases} \tag{5}$$

where x and y are the coordinates of the pixel in the image, β(x, y) is the NDVI value for the image pixel with corresponding coordinates, and B is some threshold value.

The threshold function defines the operation of threshold segmentation of the image, which converts the input color image into a black and white image. For the binarization of monochrome images, a simple yet effective method is Otsu method [5]. This method is used for automatic finding of the binarization threshold value based on the analysis of the histogram shape of intensity frequencies.

For the task of monitoring crop yield in a specific area, the most important task is to establish the boundaries of the area and link the available data to it. Also important are the a priori data about the soils and meteorological data. Having the boundaries of the area allows for the introduction of a membership function:

$$\omega(x,y) = \begin{cases} 1, \text{ if pixel } (x,y) \text{ is inside,} \\ 0, \text{ if pixel } (x,y) is \ outside, \end{cases} \tag{6}$$

which, when predicting yield, allows for consideration only of the part of the image that belongs to the area. Then, the yield estimation is calculated taking into account the membership function using the formula:

$$V = \sum_{x=0}^{X} \sum_{y=0}^{Y} \omega(x,y) p(x,y). \tag{7}$$

Timely information about the field's condition (presence of pests, plant diseases, soil condition) allows for the creation of yield maps. These maps enable the assessment of future benefits from growing agricultural crops, as there can be significant differences in yield within the same field.

Let's represent the trend component of the crop yield monitoring model (4) as a piecewise linear function:

$$T(\beta, Z, M, t) = a_i + b_i t, \tag{8}$$

where ak and bk are control points of observation.

Let's represent the seasonal component of the crop yield monitoring model (4) as a linear harmonic regression:

$$S(\beta, Z, M, t) = \sum_{k=1}^{K} \chi_k \left(\cos\gamma_k \sin\left(\frac{2\pi kt}{\lambda}\right) + \right.$$
$$\left. + \sin\gamma_k \cos\left(\frac{2\pi kt}{\lambda}\right)\right), \tag{9}$$

where the amplitude χk and phase γk are unknown, and the frequency λ is known. For Ukraine, the frequency is typically 1 year. For China, a frequency of 6 months is characteristic.

In the use of the crop yield monitoring model based on the analysis of geodata and plot images for yield prediction in this study, the random component is not considered and is assumed to be equal to 0. Considering (20) and (21), the crop yield monitoring model (4) can be written as:

$$V = a_i + b_i t + \sum_{k=1}^{K} \chi_k \left(\cos\gamma_k \sin\left(\frac{2\pi kt}{\lambda}\right) + \right.$$
$$\left. + \sin\gamma_k \cos\left(\frac{2\pi kt}{\lambda}\right)\right). \tag{10}$$

Let's consider the algorithm for using the crop yield monitoring model based on the analysis of geodata and plot images for yield prediction:

1. Use the available GIS data about the plot boundaries to calculate the membership function (6).

Process a series of plot images in different spectra. For each image:

2.1. find the NDVI value for each pixel using formula (1);

2.2. construct the frequency histogram of NDVI intensity and find the threshold value using Otsu method;

2.3. using the found value to calculate the threshold function (5), find the density value of the crop distribution in the area;

2.4. calculate the yield estimates considering the membership function using formula (7). Obtain a time series of yield estimates.

3. Using the time series of yield estimates, find the coefficients of the trend component of the crop yield monitoring model (8) using linear regression.

4. Using the time series of yield estimates, find the amplitudes and phases of the seasonal component of the crop yield monitoring model (9).

5. Find the predicted yield values using formula (10) by substituting t equal to the harvest time.

Conversation

For the verification of the crop yield monitoring model based on the analysis of geodata and plot images, a comparison of the yield forecasts for three crops: winter wheat, corn, and barley collected as of October 1, 2019, in the Chernihiv region was conducted (figure). The reason for choosing the 2018-2019 period for prediction is the lack of data on plantings for 2020-2022 on the Public Cadastre Map [6]. The forecasting was done in three ways.

The first method of prediction involves using the developed crop yield monitoring model based on the analysis of geodata and images. All available multispectral images obtained from the SENTINEL-2B satellite from October 1, 2018, to October 1, 2019, were used as input data [7].

Figure – NDVI calculation example

The second prediction method uses the developed model with all available multispectral images obtained from the SENTINEL-2B satellite for the period from March 1 to June 1, 2019, as input data.

The third prediction method is based on the use of the WOFOST simulation model [8], with calculations performed using the WOFOST Control Centre 2.1 software.

The accurate yield value is considered to be the data from the State Statistics Service of Ukraine about the yield by regions for 2019.

Data from the State Statistics Service of Ukraine on yield by regions and multispectral images obtained from the SENTINEL-2B satellite for 2011-2018 were used to find the parameters of the crop yield monitoring model based on the analysis of geodata and images.

The prediction results are presented in Table. The results demonstrate that the proposed crop yield monitoring model based on the analysis of geodata and plot images provides sufficiently accurate forecasts. It can also be concluded that yield is significantly determined by plant development in the first months after emergence, making monitoring of plant conditions during this period most important.

A significant limitation of the proposed model is the low resolution of images. For instance, the EOS AM-1 research satellite operated by NASA photographs the Earth's surface with a resolution from 250 m to 1 km. More modern satellites, like SENTINEL-2B, photograph the Earth's surface with a resolution from 20 m. However, this resolution is not sufficient to distinguish individual plants. The low resolution leads to errors where weeds and other vegetation can be perceived as the crop. To reduce the error, the proposed model should be used for monitoring the yield of agricultural crops of sufficiently large areas.

*Table – **Yield prediction for crops in the Chernihiv region for 2019***

Prediction Method	Yield, t per hectar			Relative Error, %	
	Winter Wheat	Corn	Barley	avg	max
Yield according to the SSSU	48,2	80,0	41,2	0	0
WOFOST	50,3	82,1	42,8	3,62	4,35
Monitoring model, images for 12 months	46,8	78,9	39,3	2,96	4,61
Monitoring model, images for 3 months	46,2	79,1	38,7	4,51	6,06

Additionally, the proposed crop yield monitoring model based on the analysis of geodata and images does not sufficiently take into account management decisions regarding fertilization, weeding, etc.

Conclusions

1. A mathematical model has been constructed to illustrate the relationship between phenological indicators and the yield of agricultural crops. The yield model includes three components: trend, seasonal, and random. The developed model demonstrates the dependency of yield on phenological indicators, the quality of land resources, management efficiency, and other random factors. The trend and seasonal components of the model do not depend on random factors and can therefore be predicted using appropriate methods.

2. A method for analyzing multispectral images with consideration of geoinformation data has been developed. This method includes determining the NDVI threshold value to separate the part of the field where the crop grows from the part where it does not grow using the Otsu method.

3. An algorithm for applying the developed crop yield monitoring model based on the analysis of geodata and plot images has been proposed for predicting the yield of agricultural crops.

4. The developed model and methods have been verified. A comparison of yield forecasts for three crops was conducted using observations of phenological indicators of crops throughout the year and over 3 months. It was established that yield is significantly determined by plant development in the first months after emergence. The yield forecasts were compared with data from the State Statistics Service and the WOFOST simulation model. It was found that the average relative error of yield prediction using the developed model is 2.96%.

References

1. DeJoia, A. & Duncan, M. (2015). What is “Precision Agriculture and why is itimportant”. Retrived at: https://soilsmatter.wordpress.com/2015/02/27/what-is-precision-agriculture-and-why- is-it-important/.
2. Rasinmäki, J. (2003). Modelling spatio-temporal environmental data. *Environ. Model. Softw*, 18 (10), 877–886.
3. Babu, A. J., Thirumalaivasan, D. & Venugopal, K. (2006). STAO: a component architecture for raster and time series modeling. Environ. *Model. Softw*, 21 (5), 653–664.
4. Verbesselt, J., Hyndman, R., Zeileis, A. & Culvenor D. (2010). Phenological Change Detection while Accounting for Abrupt and Gradual Trends. *Satellite Image Time Series Remote Sensing of Environment*, 114 (12), 2970–2980.
5. Otsu, N. (1979). A threshold selection method from gray-level histograms. *IEEE Transactions on Systems, Man, and Cybernetics*, 9 (1), 62–66. http://doi.org/10.1109/TSMC.1979.4310076.
6. Public cadastral map. (2024). Retrieved at: https://map.land.gov.ua.
7. Copernicus: Sentinel-2 – The Optical Imaging Mission for Land Services. (2024). Retrieved at: https://directory.eoportal.org/web/eoportal/satellite-missions/c-missions/copernicus-sentinel-2.
8. WOFOST – WOrld FOod STudies. (2024). Retrieved at: https://www.wur.nl/en/Research-Results/Research-Institutes/Environmental-Research/Facilities-Tools/Software-models-and-databases/WOFOST.htm.

The editorial board received the article on March 17, 2024

Хуан Мінсінь
Аспірант кафедри інформаційних систем та технологій,
https://orcid.org/0000-0002-2365-9815
Київський національний університет імені Тараса Шевченка, Київ

РОЗРОБКА МОДЕЛІ МОНІТОРИНГУ УРОЖАЙНОСТІ СІЛЬСЬКОГОСПОДАРСЬКИХ КУЛЬТУР НА ОСНОВІ АНАЛІЗУ ГЕОДАНИХ ТА ЗОБРАЖЕНЬ ДІЛЯНКИ

Анотація. *Дослідження присвячено побудові математичної моделі врожайності сільськогосподарських культур яка включає три складові: трендову, сезонну та випадкову. Розроблена модель засвідчує залежність врожайності від фенологічних показників, якості земельних ресурсів, ефективності управління та інших випадкових факторів. Трендова та сезонна складова моделі врожайності не залежать від випадкових факторів, а тому можуть бути використані для прогнозування урожайності. Запропоновано трендову складову моделі розглядати як лінійно-кускову функцію, а сезонну складові моделі – як лінійну гармонічну регресії. Для оцінки фенологічних показників розроблено метод аналізу мультиспектральних зображень з урахуваннями геоінформаційних даних. Цей метод включає визначення порогового значення методом Оцу для знаходження щільності сільськогосподарської культури на полі. Дані про щільність культури, доповнені геоданими про межі ділянки, використовуються для обчислення урожаю. Здійснено порівняння прогнозів урожайності трьох культур для посівів Чернігівської області при використанні спостережень за фенологічними показниками посівів протягом всього року та протягом трьох місяців. Встановлено, що врожайність значною мірою визначається розвитком рослин у перші місяці після сходів. Порівняння прогнозів врожайності здійснено з даними Державної служби статистики України та прогнозами, зробленими на основі імітаційної моделі WOFOST. Встановлено, що середня відносна похибка прогнозування врожайності за допомогою розробленої моделі становить 2,96% при використанні спостережень за фенологічними показниками посівів протягом всього року та 4,51% при спостереженні протягом трьох місяців. Така точність є достатньою і співставною із середньою точність прогнозування на основі моделі WOFOST, яка становить 3,62%.*

Ключові слова: моніторинг врожайності; GIS; управління проєктами; біомоніторинг; інформаційний менеджмент; критична інфраструктура

Link to publication

APA Huang, Mingxin. (2024). Development of a yield monitoring model based on analysis of surveys and images of the field. *Management of Development of Complex Systems*, 57, 67–71, dx.doi.org\10.32347/2412-9933.2024.57.67-71.

ДСТУ Хуан Мінсінь. Розробка моделі моніторингу урожайності сільськогосподарських культур на основі аналізу геоданих та зображень ділянки. *Управління розвитком складних систем.* Київ, 2024. № 57. С. 67 – 71, dx.doi.org\10.32347/2412-9933.2024.57.67-71.