REVIEW OF MONITORING AND FORECASTING TOOLS OF THE CROP YIELD

Abstract. A theoretical basis for research was created to solve the problem of creating an information technology for monitoring the yield of agricultural crops based on the analysis of multispectral images obtained by remote sensing. The geoinformation system created on the basis of this technology should monitor and yield the analysis of satellite time series of images to identify quantitative and qualitative indicators of yield, possible diseases of plants, and the like. The problem described in the work is relevant in conditions of environmental uncertainty. It is revealed that as a result of the food crisis, the consequence of which is a constant increase in the cost of food products, agriculture is increasingly functioning in conditions of uncertainty and risk, which requires the use of special research methods.

Keywords: geoinformation technologies; agriculture; monitoring crop yields

Introduction

The use of geoinformation systems is an integral tool for managing sown areas, analyzing yields and forecasting future yields. By 2050, according to the Food and Agriculture Organization, the world’s population may grow to 9.6 billion [1]. This is a big burden on agriculture, because productivity in the coming years should increase significantly. Moreover, agriculture now uses most of the water supply, and arable land is not enough. That is why the urgency of the development of geoinformation technology to support decision-making in the field of agriculture is not in doubt [2].

The image of the crop area can be presented in the form of time series, which allows applying the appropriate methods for their prediction [3 – 9]. The articles [10 – 12] describe the methods and technologies for processing time series of digital images for decision-making in agriculture.

The management of crop yields is a part of the managerial task of agribusiness, which requires the use of new conceptions of project and program management [13 – 17]. For the analysis of yields, it is necessary to take into account external factors of influence, in particular, the air quality in the growing area of agricultural crops [18].

The purpose of the article

The purpose of the article is to conduct a review of software products, services, devices, as well as methods for processing multispectral images, monitoring and forecasting of the crops’ youth.

Presenting main materials

As is well known, the results obtained due to remote sensing of the earth’s surface or other image of the crop area are transformed into time series, which need to be analyzed and, consequently, to make a rational decision on the management of these areas. This requires specialized software that can process both data of one dimensionality and multidimensional time series at a high level [19]. Another disadvantage of modern geographic information systems and software packages for image processing and pattern recognition is that such systems do not provide all the necessary functions for handling time series, and also require the storage of data arrays for access to images, that is, they do not provide for optimization for effective monitoring. I will plant sown areas of agricultural crops [18].

In the 90's of the last century, WinDisp software was developed, which reflected and implemented the key needs of managers in agriculture. In the early 2000s, the support of this software was stopped, because the tools that were implemented in it, did not meet the challenges of time and outdated [1]. To analyze the agricultural land image for yield monitoring, you can use specialized mathematical and statistical packages and tools for MatLab [20], R environment [21], as well as image processing software GRASS [22], IDRIDI [23], ENVI [24]. You can also use MySQL or PostgreSQL spatial databases to store information. However, the processing of agricultural land images requires special skills in programming and analysis, that is, the involvement of relevant professionals.

Necessary tools for analyzing information obtained after ground sensing, such as the NDVI vegetation index [25], surface humidity, anomalies, etc., provide a number of online platforms: MARS [26], the US Geological Survey Platform [27], the GLAM Platform [28], CropExplorer [29]. However, these platforms don’t have the necessary function for the processing of agricultural landscapes, which corresponds to the needs for management of sown areas and forecasting yields.

© Mingxin Huang
Some other online services also provide services related to image processing, but they are mostly used for specific tasks [30 – 32]. For example, the TIMESAT [33] service is quite effective, but its lack is the lack of a proper graphical interface. Another TimeStats service has built-in intelligent data analysis and a sufficient amount of archive data, but it requires the involvement of data analysis specialists and remote sensing experts. Means for the purpose of rheological modulation [34; 35] provide efficient work with time series of images, but their capabilities do not meet the needs for processing large numbers of images.

A system that is designed to monitor crop yields should take into account the large number of different growth indicators, perform statistical analysis and forecasting of yields based on monitoring results:

1. Meteorological indicators (weather monitoring).
2. Agrometeorological indicators (crop monitoring).
3. Phenological indicators (phenological monitoring).

In order to ensure effective monitoring at all levels, high-quality aerospace images that can be stored and analyzed in time series should be provided for analysis.

Crop monitoring is a system for observing and measuring crop growth, taking into account meteorological, agrometeorological, phenological and other indicators based on the analysis of time series images obtained as a result of photographing crop areas, in order to evaluate and predict the potential yield of a crop. It is also envisaged to take into account uneven yields at different sections of the crop area, which can be visualized using yield maps, in which the various colors indicate the volume of the resulting products after harvesting. It should also be borne in mind that crop monitoring should provide for a two-dimensional assessment: the quantity of products potentially obtained and the quality of the product on the appropriate scale.

Phenological indicators of plant cultivation take into account the relationship between seasonal changes, weather conditions and climate change affecting the maturity period, the occurrence of diseases, the timing of the start of certain agricultural work, and so on. The study of phenological indicators is a necessary element of quality management of sown areas.

Most studies concerning the allocation of phenological indicators in the analysis of images of sown areas by remote sensing of the earth's surface are in the calculation of normalized vegetative index (NDVI). This is a simple quantitative indicator of photosynthetic active biomass used to quantify vegetation cover:

$$\text{NDVI} = \frac{N - R}{N + R},$$

where \( R \) = 630-690 nm, visible red area spectrum; \( N \) = 760-900 nm, reflective infrared spectrum area.

According to this formula, the vegetation density (NDVI) at a certain point in the image is equal to the difference in the intensity of the reflected light in the visible and infrared range divided by the sum of their intensities [36 – 38]. The general concept for detecting phenological indicators is based on the definition of critical points in the NDVI trajectory. Let

$$\beta = \{\beta_1, \beta_2, \ldots, \beta_n\},$$

\( \beta \) – discrete time series of NDVI ratings fixed at certain time intervals (day, week, two weeks, etc.); \( \beta_i \) – estimation of NDVI at the i-th moment of time.

Then, by constructing this time series on the graph, one can investigate how the vegetation index changes in dynamics, which allows real-time tracking of abrupt changes in the vegetation level and adopting an appropriate solution to ensure effective crop management.

Consider the BFAST (Breaks For Additive Seasonal and Trend) method, which can help determine long-term phenological changes in time series of images. This method combines the methods of detecting changes in the behavior of time series with the methods of decomposing rows into components that determine tendency changes, seasonal changes, and random components [39]. The article [40] describes the possibilities of the BFAST method to detect long-term phenological changes with the help of a harmonious seasonal model. According to this method, the additive decomposition model is that the time series of the image has the form:

$$Y_t = T_t + S_t + e_t,$$

where \( Y_t \) – time series data that are recorded at the time \( t \), \( T_t \) – trend component, \( S_t \) – seasonal component, \( e_t \) – residual, random components, \( t = 1, n \), \( n \) – the number of observations or the number of elements in the time series of the image. Residual components are variations of the time series that are diverging from trend or seasonal components:

$$T_t = a_t + t \cdot b_t,$$

where \( r_{t-1} \leq t \leq r_j, i = 1, m \) – monitoring points.

The seasonal component changes in a similar way. Changes can occur from one segment to another between control points. Also, for the seasonal component, control points can be determined differently. In [40], a harmonic model for describing the seasonal component is defined:

$$S_t = \sum_{k=1}^{l} \alpha_{jk} \sin \left( \frac{2\pi k t}{\lambda} + \beta_{jk} \right),$$

where unknown is the amplitude \( \alpha_{jk} \), phase \( \beta_{jk} \), and the frequency is known \( \lambda \).

Also, to determine the seasonal component, you can set the model of linear harmonic regression:
\[ S_t = \sum_{k=1}^{K} \left( \gamma_{jk} \sin \left( \frac{2\pi kt}{\lambda} \right) + \chi_{jk} \cos \left( \frac{2\pi kt}{\lambda} \right) \right), \]

where \( \gamma_{jk} = \alpha_{jk} \cos \beta_{jk}, \quad \chi_{jk} = \alpha_{jk} \sin \beta_{jk}, \)

- coefficients of the model. The amplitude can be defined as

\[ \alpha_{jk} = \sqrt{\gamma_{jk}^2 + \chi_{jk}^2}, \]

and the phase for the frequency \( \frac{\lambda}{k} \) is set to

\[ \beta_{jk} = \frac{1}{\tan} \left( \frac{\chi_{jk}}{\gamma_{jk}} \right). \]

The described model has the following advantages, unlike the usual seasonal model:

1. The model is less susceptible to short-term changes and noise.
2. To calculate the parameters of the model of multiple regression, many observations are not required.
3. Parameters are relatively easy to calculate.

In addition to identifying phenological changes, one can consider methods that allow meteorological and agrometeorological changes to be identified. Their content does not change in general. It is necessary to allocate time series of data (meteorological and agrometeorological indicators) which characterize the research area, make its decomposition, allocate seasonal and trend components, and also estimate the residual component. After this procedure, it is possible to calculate the forecast as a seasonal component with a certain confidence interval \( \hat{S}_t \), and directly to the time series of data \( \hat{Y}_t \). The forecast data are valuable information for the planning and organization of agricultural work, in particular, in the monitoring of crop yields. In Figure 1 Two pictures of the same field are displayed at intervals of 10 days. Changes in the image are visualized without the use of software, only by viewing the image. However, for effective crop management, such undesirable changes need to be identified before they are visible to the visually in order to take the necessary action and not to lose the crop.

In order to distinguish between phenological, meteorological and agrometeorological indicators, it is necessary to construct the so-called Satellite Image Time Series (SITS) satellite image. Such time series are satellite images of a certain area, executed at fixed moments of time. It is necessary to construct it:

1. To take pictures of the territory with a certain period of time.
2. Perform recognition for these photos using special image recognition methods.
3. Perform a visualization of the map of the territory, where you can display (preferably in color) changes that are related to the investigated feature (Figure 2).

The collected information is analyzed and the conclusion is reached regarding the estimation of agricultural crop yields at a given site.

![Figure 2 – Satellite time series of image of the territory and displaying the color of deviations from the norm by two indicators](image)

It should be noted that an arbitrary image can be defined as a function \( z(x, y) \), where \( x \) defines the coordinate abscissa, and \( y \) – coordinate ordinates on the plane. The value of the function is intensity.

In the case of a black and white image, the value of the function \( z(x, y) \) is a gray level. A digital image is a function \( z(x, y) \), the arguments of which accept a finite number of discrete values. If you divide the image into a finite number of pixels, then the position of each pixel will be determined by the coordinates \( x, y \).

As a rule, black and white images for monitoring yields are not used, since the color contains very valuable information on the basis of which it is possible to determine the NDVI indicator or to identify possible plant diseases.

Color models mostly have three views:

1. RGB color mixing model.
2. CMYK model used in typography.
3. YIQ, YUV, YCrCb models used in video systems.

Figure 1 – Satellite images of the same field are made at intervals of 10 days
One way of comparing images with each other is the method of perceptual hashing. One of the simplest hash functions is the average of low frequencies. In images, high frequencies provide detail, and low frequencies display the structure. A large, detailed photo contains many high frequencies. In a very small image, there are no details, so it consists entirely of low frequencies. The hash algorithm consists of these steps:

1. Reduce the size of the image. When reducing the image, the high frequencies disappear. It is suggested to reduce the image to at least 8 by 8 pixels. Even if the image is not square, you still need to compress it to a square of the minimum size.

2. The next step is to convert the image to shades of gray without color.

3. Calculation of the average value of all shades of all image pixels.

4. If the pixel tint is darker than the average tint, then it gets 0, that is, black. If the tint of the image pixel is lighter than the average tint, then it gets 1, that is, white. Thus the image is converted to black and white.

5. Constructing hash, that is, converting bits of an image into one value. In the case of an image 8 by 8, it will be a 64-bit value based on the number of pixels or bits.

The final hash will not change if the image is scaled, compressed or stretched. To compare images, you can quantify the similarity using, for example, Heming distance.

Another method is the perceptual image hashing, which uses the discrete cosine-transformation of the image. Discrete cosine-transformation (DCT) is an orthogonal transform, a cosine-transform variant for a real number vector. It is used in compression algorithms for loss information, such as MPEG and JPEG. This transformation is closely related to the discrete Fourier transform and is a homomorphism of its vector space. Mathematically, the transformation can be accomplished by multiplying the vector into a transformation matrix. In this case, the matrix of the inverse transformation to the multiplier is equal to the transposed matrix [41].

The algorithm of perceptual hashing consists of these steps:

1. Reduce the size of the image. Perceptual hash is calculated on the basis of a small image, for example 32 to 32.

2. Remove the color and perform a discrete cosine transform that cuts the image into a set of frequencies and vectors.

3. In the upper left corner of the image obtained after the cosine transform, only lower frequencies will be stored, so it is suggested to fix the part of the image for further analysis: the upper left corner is 8 to 8.

4. Calculate the average value in the reduced block.

5. For the image, the cosine of the transformation is assigned 0 or 1, depending on whether the darker or lighter shade is.

6. After that, you need to build one meaning of the perceptive hash and compare it with the hashes of other images.

There are varieties of perceptible hash algorithms that also increase performance. For example, you can crop the image before reducing the size. In this case, the loss of information around the main part of the image does not play a special role.

The disadvantage of the hashing method for the task is the need to maximize the consideration of high frequencies. Also, the disadvantage is that the change in the pixel color of the image will not significantly affect the hash design, and this change can signal the negative processes occurring in the development of plants.

Another approach that may be useful in analyzing the fields of agricultural crops is the image segmentation. Segmentation involves separating the image into segments according to certain rules. Segmentation methods, taking into account high frequencies, allow you to detect changes that occurred on the site, new objects that appeared on it, and so on. Each pixel of the image corresponds to a certain label, and it is considered that pixels with the same label are visualized equally.

There are two basic approaches to segmentation. The first one is based on image splitting by brightness levels. The sharp change in brightness when moving from one pixel to another can indicate the identity of the object. The second class of methods is to partition the image in the area according to predetermined criteria.

One of the missions of the research satellite EOS AM-1, which operates under the guidance of NASA, is taking photos of the earth’s surface in 36 range ranges from the length from 0.4 mkm to 14.4 mkm and an extension from 250 m to 1 km. This mission is called MODIS (Moderate Resolution Imaging Spectroradiometer) and is generally designed to study processes (radiation, meteorological, etc. on the surface of the earth and the oceans). Also, this tool is used in agriculture, in particular for calculating vegetation indices and soil moisture levels.

The MODIS sensor has 36 spectral bands, seven of which are intended for study of vegetation and terrestrial surface: blue (459-479 nm), green (545-565 nm), red (620-670 nm), infrared (NIR1: 841-875 nm ; NIR2: 1230-1250 nm), and short-wave infrared (SWIR1: 1628-1652 nm, SWIR2: 2105-2155 nm). The MODIS Land Science team provides a set of standard MODIS products for users, including the MODIS Surface Reflectance 8-day Composite (MOD09A1). According to satellite data, vegetation indices are calculated according to the formulas:

\[
NDVI = \frac{R_{\text{NIR}} - R_{\text{RED}}}{R_{\text{NIR}} + R_{\text{RED}}},
\]

where \( NDVI \) – Normalized Difference Vegetation Index, \( R_{\text{NIR}} \) – infrared band, \( R_{\text{RED}} \) – infrared band.
where

\[ LSWI = \frac{R_{NIR} - R_{SWIR}}{R_{NIR} + R_{SWIR}}, \]

\[ EVI = \frac{5}{2} \frac{R_{NIR} - R_{RED}}{1 + R_{NIR} + 6R_{RED} - \frac{15}{2}R_{BLUE}}, \]

where \( LSWI \) – Land Surface Water Index; \( R_{NIR} \) – infrared band; \( R_{SWIR} \) – shortwave infrared band; \( R_{RED} \) – red band; \( R_{BLUE} \) – blue band.

So far, vegetation indices, and especially the NDVI index, are the most informative indicator of automated crop evaluation. However, its use for this purpose has a number of features. First, all studies on the application of the NDVI index for yield assessment usually relate to local areas, a defined list of crops, and were carried out at the regional level, which does not cover sufficient representation of plants, regions and geographical features. Often these studies are used to analyze low-resolution images. Therefore, the results of studies in which the only indispensable index in the calculation of agricultural crop yields is the NDVI index is not entirely correct. Because the quality of the images is low, weeds and other vegetation can be perceived as cultivated plants, which cannot be filtered out from the original image. The results obtained are then adjusted to take into account the final harvest.

In general, most studies indicate a high correlation between the NDVI index and the final harvest. However, in these studies, small plots were analyzed, and many of them used photographs derived from terrestrial platforms.

Manufacturers offer many devices to monitor yields of agricultural products. One of them is the Grain yield monitor (GYM). This is a device with sensors to calculate the yield of grain placed on the combine. GYM is part of high-precision agricultural production, which provides agricultural producers with tools to reduce costs, increase yields and increase efficiency. GYM is designed to measure the mass flow of grain yield, moisture and speed to determine the total amount of grain harvested. In most cases, it is now combined with a global positioning system for recording yields and other field information variables. This allows you to create a grain yield map that provides information on spatial variability and supports managerial solutions for producers.

The yield monitor is a device that records the data that determines the yield of the grain. Modern yield monitors provide operators with a user interface that displays grain yield, grain moisture, as well as a spatial color coding map that displays the yield of grain of certain parts of the field.

Grain yield maps can be displayed on the monitor or through spatial data management software such as SMS or Apex. Crop maps are used in managerial decisions such as fertilizer rates and sowing rates, etc. [44]. Crop maps are also used to make decisions about best management practices in terms of comparing varieties of crops, types of fertilizers and standards of their application, as well as the use of pesticides. These and other methods of precise farming can be written as spatial maps and superimposed on grain yield maps for further analysis and decision making.

The density and health of crops can be monitored by analyzing surface field images that are obtained at a certain interval of time. Such time series of images contain valuable information on the activity of chlorophyll, plant activity, the possible presence of diseases or pests. The analysis of multispectral field images allows us to identify the necessary indicators that determine the growth of a crop, its maturity, and others like that. Therefore, we can assume that NDVI vegetation indices, or other similar indices, can be used to monitor the yield and crop yields in the local field.

It should be noted that productivity is also influenced by other factors: soil quality, type and efficiency of management, environmental situation and weather conditions, etc.

**Conclusions and recommendations for further research**

According to the results of the study of modern means of processing time series of images for monitoring of crop yields, it was found that none of the services or software does not combine all the necessary package of possibilities for processing and analysis of data that are necessary for the management of sown areas: finding anomalies, identifying phenological changes, crop evaluation, etc. Therefore, the important task is to create specialized software that would allow you to download and work with archives of large-scale images and implemented, built-in methods of intelligent data processing and image recognition.

The urgent task is to develop methods for analyzing such digital images for the establishment and forecasting of yield levels, the identification of areas of over-drying of the soil, the content of weeds, the incidence of plants, etc. Integration of these methods into geoinformation systems will create a multifunctional decision support system in agriculture.

Investigating the connection between remote sensing and yield data is relevant, since there has not been much research in this area. Especially important research in this direction is due to significant environmental uncertainty, which complicates decision-making on management of sown areas.

As a result of the analysis of scientific researches in the direction of establishing the relationship between yields and the results of calculating vegetation indices and other indicators of the growth rate of agricultural products, the urgent task is to develop management systems that combine remote sensing data and currently used software for decision-making on the management of sown areas.
plants, it has been found that most of them are performed either on local sites, taking into account a large number of controlled parameters, or in large plots of a significant degree of generalization, since some fragments of the image in this case is difficult to study in detail. It should be noted that many factors influence the growth of agricultural crops, including quality of soil, quality of management, ecological situation, etc.

Prospective study objectives are:
1. Establish the relationship between the values of vegetation indices, such as NDVI, and the yield of crops.
2. Describe the factors of influence on yield, in particular, the control factor, ecological factor, soil quality factor, etc.
3. Estimate the possibility of using multispectral images for forecasting crop yields.
4. To propose information technology for monitoring of crop yields based on the geographic information system.

References

Ключові слова: геоінформаційні технології; сільське господарство; моніторинг врожайності

Анотація. Створено теоретичну основу дослідження для вирішення задачі створення інформаційної технології моніторингу врожайності сільськогосподарських культур на основі аналізу мультиспектральних зображень, отриманих шляхом дистанційного зондування. Створена на основі цієї технології геоінформаційна система повинна моніторити та прогнозувати врожайності аналізуючи супутникові похилі ряди зображень для виявлення кількісних та якісних показників врожайності, можливих захворювань рослин тощо. Описана в роботі задача є актуальною в умовах екологічної невизначеності. Виявлено, що через продовольчу кризу, наслідком якої є постійне зростання вартості продовольчих товарів, сільське господарство все більше функціонує в умовах невизначеності та ризику, що потребує застосування спеціальних методів дослідження.

Ключові слова: геоінформаційні технології; сільське господарство; моніторинг врожайності

Link to publication
