Svitlana KUZNICHENKO[1], Yurii GUNCHENKO[2], Iryna BUCHYNSKA[1]

KLASYFIKACJA ROŚLINNOŚCI Z WYKORZYSTANIEM ZDJĘĆ SATELITARNYCH I TECHNOLOGII GIS

Streszczenie: W artykule przedstawiono podejście do oceny dynamiki zmian stanu ziemi dla celów rolniczych z wykorzystaniem technologii teledetekcji i GIS. Wyjaśniono wybór metody obliczenia indeksu wegetatywnego NDVI i klasyfikacji zdjęć kosmicznych przy wykorzystaniu oprogramowania z otwartym kodem. Zaproponowano program dla zabezpieczenia i algorytm wyliczeniowy dla klasyfikacji danych teledetekcyjnych, celem których jest wyznaczenie obszarów z różną gęstością roślinności, co pozwala robić monitorowanie wzrostu kultur rolniczych roślinności i osobliwości sezonowej wegetacji oraz dokonać obliczeń prognozowanych zbiorów.

Słowa kluczowe: technologie geoinformacyjne, zdjęcia satelitarne, klasyfikacja, pokrycie terenu, uprawa, wskaźniki roślinności

VEGETATION CLASSIFICATION USING SATELLITE IMAGERY AND GIS TECHNOLOGIES

Summary: The paper presents an approach to assessing the dynamics of changes in the state of agricultural lands using remote sensing data and GIS technologies. The choice of the method for calculating the vegetation index NDVI and the classification of space images using open source software has been substantiated. Software and a computational algorithm for the classification of remote sensing data in order to determine areas with different vegetation densities are proposed; it allows monitoring the agricultural crops and the characteristics of seasonal vegetation and calculating quantitative estimates of the predicted yield.

Keywords: geoinformation technologies, satellite images, classification, crop, vegetation indices

1. Introduction

Vegetation is the most dynamic component of natural systems. The complexity of the organization and a wide variety of structural elements of the vegetation cover make it difficult to distinguish discrete parts in it as classification units. However, the

[1] Odessa State Environmental University, Dept. of Information Technologies, Odessa, Ukraine, Ph.D, e-mail: skuznichenko@gmail.com; and Ph.D student, email: buchinskayaira@gmail.com:
[2] Odessa I. I. Mechnikov National University, Prof. D.Sc., email: gunchenko@onu.edu.ua

systematization of objects is a necessary stage of any research, and the choice of classification principles is determined by specific goals and objectives.

Currently, due to the increasing anthropogenic influence, the vegetation cover is rapidly changing; mapping is an important aspect of the study of vegetation cover, an effective method of its spatial inventory, assessment of dynamics and biodiversity.

This is especially important for monitoring and analyzing the state of agricultural land and crops, as well as forecasting yield. In addition, this kind of research helps to identify the causes of occurrence, spread, as well as timely elimination of land degradation.

An effective and modern way of monitoring changes in the state of agricultural lands is the use of historical high-resolution remote sensing data and GIS technologies. The choice of methods for geo-information analysis of these data at different times can make it possible to trace the changes that have taken place with the object of study during the designated period. Based on the results of this analysis, to obtain quantitative assessments of the processes and develop measures to reduce the influence of negative factors on the process of agricultural land degradation.

The purpose of this work is to develop automated approaches to the classification of vegetation cover based on GIS, remote sensing data and ground-based research materials. The study used publicly available high-resolution satellite images and open source geographic information systems.

The object of the study is the land resources of Odesa region (3331.4 thousand hectares), which are characterized by an extremely high level of development. The largest is the share of agricultural land - 2659.2 thousand hectares, of which arable land – 2075.5 thousand hectares. In the structure of land, agricultural land is occupied by 79.8%, including arable lands – 62.3% (Table 1) T. The area of degraded lands in the Odesa region is 33.0 thousand hectares [1].

Table 1. The structure of the fund of agricultural lands of the Odesa region

Main types of agriculture	Arable land	Fallows	Perennial plantations	Hayfields and pastures
Total thous. ha.	2074,4	27,4	87,3	403,3
% from the total area	62,3	0,8	2,6	12,1

According to the Center for Assessment of Social and Environmental Risks [2], the problem of soil degradation as a result of their intensive use is attributed to the one of the main environmental risks in the Odesa region.

2. Materials and Methods

2.1. Earth remote sensing data used in the research

Most monitoring tasks use high-resolution satellite images: LANDSAT – 7 (ETM +), LANDSAT – 4, 5 (TM), LANDSAT – 1, 3 (MSS) TERRA (ASTER), HYPERION (EO-1).

The images taken by different satellites differ in the following main parameters:

- spatial resolution: a square on the earth's surface corresponds to 1 pixel of the image raster (varies from 0.3x0.3 m to 1000x1000 m);

- the number of channels and the spectral range of each channel: these characteristics directly affect what properties of the earth's surface can be detected from the images (Table 2);
- temporal resolution - the approximate interval at which the image is repeated in a given area (per day);
- the total area of one shot (the lower the resolution, the larger the size of the shot itself);
- the time interval in which the satellite imagery was carried out by the sensor;
- purchase cost (original photograph and derivative products).

Table 2. Characteristics of satellite imageries

Sensor	ASTER	Landsat (MSS, TM, ETM+, OLI)	MODIS
Satellite	Terra	Landsat1…Landsat8	Terra, Aqua
Years of work	since 2000	since 1972	since2002
Number of spectral channels	15	4(L1)…11(L8)	36
Total spectral range (μm)	0.52 – 11.65	0.43 – 12.51 (по L8)	0.4 – 14.4
Area of one imagery (sq.km)	60 x 60	185 x 185	1130 x 1130
Spatial resolution (meters in 1 pixel)	15 – 90	15 – 100	250 – 1000
Time interval (day)	16 and less	16	0.5 – 1

Each satellite sensor takes a surface survey in spectral bands, each of which has its own clearly defined interval of the electromagnetic spectrum in which the image is taken. The number of spectral channels and the spectral range of each channel is different in different sensors. The result of satellite photography of the earth's surface in one channel is called a “scene of image.” Each image has as many scenes as there are spectral channels in the sensor. In other words, all scenes of the same time taken in different spectral channels of the sensor.

The satellite's sensor registers the reflected and released electromagnetic radiation from the earth's surface and stores this measurement for each pixel in the image in a special unit called "gray level" or Digital Number (DN).

The gray levels of modern sensors are divided into 64, 256 and 65536. The gray level of a pixel is the color of a pixel stretched a digital number from pure black (DN = 0) to pure white (DN = the maximum possible value, that is, 64, 256, or 65536). The DN levels depends on the number of binary digits (bits) - 0 and 1, required to store the maximum value of the parameter in digital form. For example, with 16 bits, the digital number is $2^{16} = 65536$ levels (so many levels have images of the Landsat 8 satellite, unlike earlier Landsat satellites having 256 levels).

The gray level (DN) of each pixel of the satellite image raster is the average measurement of the energy reflected and released from that territory on the planet's surface corresponding to a given pixel: that is, a square, the side of which in modern images varies from 0.3 to 1000 meters. Actually, the sensor captures the gray level in pixels for each such square, based on the average reflected and released energy of all types of surfaces that make up the square.

Data from Landsat satellites 5, 7 and 8 (USGS, USA) were used in the research. They are distributed free of charge [3,4] and have sufficient spatial resolution.

2.2. Vegetation indices

To identify the characteristics of the earth's surface by remote sensing methods, a graph of the spectral curve of the object is used. The spectral curve graph shows the percentage of reflected and emitted radiation for a given object over the entire electromagnetic spectrum. Deciphering the image and identifying the characteristics of objects on the earth's surface is based on comparing the already known spectral curves with the spectral curves of each pixel in the image. The shape of the spectral curves of various objects can differ significantly, and the principle of object recognition and the identification of their characteristics from satellite images is based on these differences. The plot of the spectral curves for the vegetation cover are shown in Fig. 1. Knowledge of the relationship between the structure and state of vegetation with its spectral reflected abilities allows using space images for mapping and identifying vegetation types and their stress state [5-7].

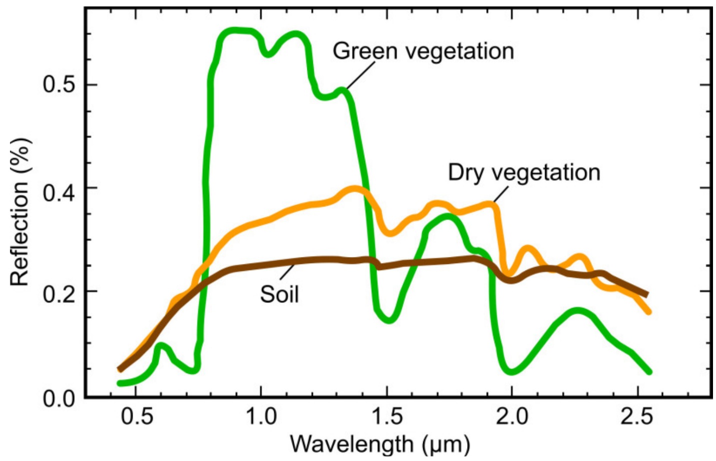

Figure 1. Graphs of spectral curves of dry vegetation, green (healthy) vegetation and soil

To work with spectral information, vegetation indices are often used - indicators calculated as a result of operations with different spectral ranges (channels) of remote sensing data, and are related to the parameters of vegetation in a given image pixel.
There are three most common vegetation indices: NDVI, PVI, and SAVI.
NDVI is a Normalized Difference Vegetation Index, a quantitative indicator of the amount of photosynthetic active biomass. One of the most common and used indices for solving problems that apply quantitative estimates of vegetation cover. NDVI is moderately sensitive to changes in soil and atmospheric background, except in cases with poor vegetation [5-7].
NDVI is calculated using the formula [7]:

$$NDVI = \frac{(NIR - Red)}{(NIR + Red)}, \quad (1)$$

where NIR – reflection in the near infrared region of the spectrum; RED – reflection in the red (visible) region of the spectrum.

Estimating the NDVI value from the available images, we can confidently speak about the structure of the study area and changes in the state of vegetation. NDVI can be calculated based on any high, medium or low-resolution images, have spectral channels in the red (0.55-0.75 μm) and infrared (0.75-1.0 μm) ranges. The NDVI values for different types of objects are given in Table 3.

Table 3. NDVI values for different object types

Type of object	Reflection in Red Region of the Spectrum	Reflection in Infrared Region of the Spectrum	NDVI values
Dense vegetation	0.1	0.5	0.7
Sparse vegetation	0.1	0.3	0.5
Open ground	0.25	0.3	0.025
Clouds	0.25	0.25	0
Snow and ice	0.375	0.35	-0.05
Water	0.02	0.01	-0.25
Artificial materials (concrete, asphalt)	0.3	0.1	-0.5

The main disadvantages of using NDVI are:

- inability to use data that has not passed the stage of radiometric correction (calibration);
- errors introduced by weather conditions, heavy cloudiness and fog;
- the need for most tasks to compare the results obtained with pre-collected data from test sites (standards), which should take into account seasonal environmental and climatic indicators, both of the image itself and of test sites at the time of data collection;
- the possibility of using surveying only for the growing season for the studied region.

2.3. Pre-processing of satellite imageries

Preliminary processing of a satellite imagery consists in carrying out geometric, radiometric and atmospheric corrections.

Geometric correction removes geometric distortions associated with the fact that the imagery of the Earth's surface by a device directed exactly downward contains an undistorted imagery only in the center of the image; when shifting to the edges, the distances between points on the image and the corresponding distances on Earth are increasingly different. Correction of such distortions is carried out in the process of photogrammetry.

Radiometric correction includes two types of corrections:

- radiometric correction of distortions arising from the fault of the sensor optics;
- radiometric correction of distortions arising from the angle of incidence of sunlight and landforms (elimination of defects in sunspots and shadows).

Atmospheric correction - removes various distortions introduced by the presence of atmosphere [8].

Calibration coefficients required for making corrections are supplied with the original package of the image; However, there is an opportunity not to deal with the correction and even the main processing of imageries on their own, but to use ready-made "secondary" images, if they are supplied for the selected sensor and the quality of their data is suitable for the tasks being solved.

Sensor products are supplied to the end user in two forms [8]:

- original image with all scenes of spectral channels. According to Table 3, these are equal to Level0 (the original version of the image) and Level1 (the same initial version with applied but not applied correction factors)
- derivative products - a snapshot processed by operators with a well-defined algorithm to detect certain characteristics. These are Level2, Level3, Level4 for Table 4.

Table 4. Levels of remote sensing data processing (by NASA system)

Level	Description
Level 0	Reconstructed, unprocessed instrument and payload data at full resolution, with any and all communications artifacts (e.g., synchronization frames, communications headers, duplicate data) removed.
Level 1A	Reconstructed, unprocessed instrument data at full resolution, time-referenced, and annotated with ancillary information, including radiometric and geometric calibration coefficients and georeferencing parameters (e.g., platform ephemeris) computed and appended but not applied to Level 0 data.
Level 1B	Level 1A data that have been processed to sensor units (not all instruments have Level 1B source data).
Level 2	Derived geophysical variables at the same resolution and location as Level 1 source data.
Level 3	Variables mapped on uniform space-time grid scales, usually with some completeness and consistency.
Level 4	Model output or results from analyses of lower-level data (e.g., variables derived from multiple measurements).

Derivative products do not require any special processing from the end user; they can be downloaded and used immediately to analyze the data they represent. To carry out the study, we will use Landsat 8 Level2 satellite images without performing any special preprocessing.

2.4. Classification of satellite imageries

When classifying imageries, all pixels (their spectral brightness) are represented as vectors in the space of spectral features. When analyzing the quantitative relationships of the spectral brightness of various objects, the points are divided into classes. Image classification is divided into supervised (with training) and unsupervised classification (without training) [9].

Classification with training assumes the presence of a standard with the brightness of which the brightness of each pixel is compared. As a result, having several predefined standards, we get many objects, divided into classes. This classification works only if the objects that are reflected in the image are known in advance, the classes are clearly visible and their number is small.

Consider some of the methods that can be used in learning classification [9].

Minimum distance method.The brightness value of pixels are considered as vectors in the space of spectral features. Between these values and the values of the vectors of the reference areas, the spectral distance is calculated as the root of the sum of the squares of the difference between the vectors of the pixel and the reference (in other words, the Euclidean distance between them). All pixels are divided into classes depending on whether the distance between them and the reference is greater than the specified one or not. So, if the distance is less, then the class is defined, the pixel can be attributed to the ideal.

The Mahalanobis distance method is very similar to the first method, only it is not the Euclidean distance between vectors that is measured, but the Mahalanobis distance, which takes into account the variance of the brightness values of the reference.

The Mahalanobis distance (*Dm*) from the multidimensional vector $x = (x_1, x_2, x_3, ..., x_N)^T$ to the set with the mean vector $\mu = (\mu_1, \mu_2, \mu_3, ..., \mu_N)^T$ and the covariance matrix *S* is determined as:

$$D_m = \sqrt{(x-\mu)^T S^{-1} (x-\mu)}\ , \quad (2)$$

Spectral angle mapper algorithm. The maximum value of the spectral angle (the angle between the reference vector and the vector of a given pixel) is set.

The spectral angle is found, and, as with the Euclidean distance, if the angle is less than the specified one, then the pixel falls into the class of the reference with which it is compared.

Parallelepiped classification is a nonparametric method that is implemented by setting boundaries for the spectral brightness values for each class. The bounds of the parallelepiped for each class can be determined by the minimum and maximum values of the matrix in the given class or, alternatively, by a certain number of standard deviations on either side of the class mean, determined from the training sample.

The rule of thumb is to check if the point that represents a pixel in feature space is inside each of the parallelepiped. The parallelepiped classification is fast and easy, but errors can occur, especially when a pixel lies within more than one parallelepiped or outside all parallelepipeds. These two situations, obviously, can arise, since the distribution of template vectors in space is often quite complicated [10].

In the process of parallelepiped classification:

- pre-create reference areas;
- the value of the brightness of pixels is considered as a vector fij in the space of spectral features, i and j is the value of the brightness of a pixel in different spectral channels;
- the spectral distance between the reference vectors and the vectors of the brightness values of all pixels of the image is calculated. The distance between the two vectors (r) is calculated as:

$$r = \sqrt{\sum_k \left(f_{kij} - f_{kmn}\right)^2}\ , \quad (3)$$

where *k* is the number of the spectral channel.

Classification without training is based on a fully automatic distribution of points into classes based on the statistics of the distribution of pixel brightness values. This type of classification is used if at first it is not known how many objects are present in the image, the number of objects is large, as a result, the machine itself issues the resulting classes, and we already determine which objects to match them.

The ISODATA (Iterative Self-Organizing Data Analysis Technique Algorithm) is based on cluster analysis using the successive approximation method. After considering the brightness of pixels as vectors in the space of spectral features, the nearest ones are determined into one class. For each spectral zone, the statistical parameters of the brightness distribution are calculated. All pixels are divided into some n number of equal ranges, within each of which there is an average value. For each pixel in the range, the spectral distance to the mean is calculated. All pixels, the distance between which is the smallest, are defined in one cluster. This is how the first iteration goes. During the second iteration and subsequent iterations, the real average values for each cluster are calculated. Each new iteration refines the boundary of future classes.

The K-Means method is similar to the previous method, with the only exception that the initial average values are specified (this is possible only if the objects in the image are well read).

In this work, to classify territories by vegetation density based on NDVI values, the parallelepiped classification was used. The task of known values of the boundaries of parallelepipeds was carried out on the basis of the classification adopted for Ukraine (no vegetation (-1-0), average vegetation (0-0.15), vegetation medium density (0.15-0.3), dense vegetation (0.3-0.45), very dense vegetation (0.45-0.6), high density vegetation (0.6-1)). The algorithm for data analysis is shown in Fig.2.

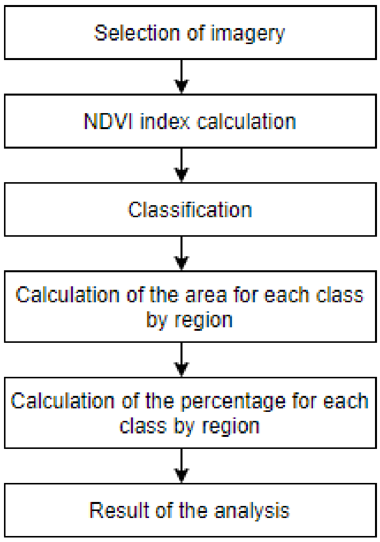

Figure 2. Algorithm for data analysis

The study used the QGIS geographic information system, which is distributed under the GNU General Public License v2 [11,12]. Landsat 7, 8 satellite images downloaded free of charge from the USGS (United States Geological Survey) geodatabase [3] for the vegetation period (April to October) 2011 and 2015-2018. The images already had atmospheric correction, that is, there was no need to carry out it separately.

3. Results and Discussion

3.1. Calculation of vegetation indices

At the first stage of image classification, it is necessary to calculate vegetation indices, to find ranges of different types of vegetation. In the second stage: apply the segmentation method to the converted images, taking into account the calculated ranges of vegetation indices, then carry out vectorization and calculate the accuracy of the definition of vegetation objects, create a map-scheme.
The diagram of the methodology of segmentation of vegetation presented in Fig. 3.
The NDVI index can be calculated in QGIS using a Raster calculator. NIR and RED channels were used to calculate the index, the Landsat satellite has the fourth and third channels, respectively (the Landsat 8 satellite has the fifth and fourth channels).
The calculation of the index for each pixel of the satellite image in the red and near infrared spectral zones allows you to obtain a derived image - the NDVI map. For green vegetation, the reflection in the red region is always less than in the near infrared, due to the absorption of light by chlorophyll, so the NDVI value for vegetation cannot be less than zero. As a result, the values for the NDVI index are distributed as follows:

$$S(x,y) = \begin{cases} 1, & \text{if } 0.54 < f(x,y) \leq 0.85 \\ 2, & \text{if } 0.29 < f(x,y) \leq 0.54 \\ 3, & \text{if } 0.15 < f(x,y) \leq 0.29 \\ 0, & \text{else} \end{cases} \tag{4}$$

where $S(x,y)$ is the desired imagery;
1 – forester, dense vegetation;
2 – swamp, swamps, average vegetation density;
3 – farm, grass, poor vegetation;
0 – water; no vegetation.

3.2. Classification of satellite imageries

The images were segmented using the Semi-Automatic Classification plugin in the QGIS environment [13]. The segmentation algorithm diagram is shown in Fig.4. The result of the algorithm operation is the separation of all pixels in the image according to the specified limit value of the spectral index and the creation of an initial raster file of the GeoTIFF format containing information about the types of pixels in the form of a binary image. Then we turn the resulting raster file into polygons of the vector layer with the saved projection; and a text file that contains information about the coordinates and values of the pixels of each class, which will allow us to identify areas of different types of vegetation.
After loading the space image of the territory, it is necessary to create directories, where ROI files (regions of interest) and signature files will be saved later. Next, we need to generate ROI files and signatures for further classification. For objects with a large area, such as lake mirror, we use the polygonal capture button (Fig. 5). Then we need to set the selected object ID and click on "SAVE ROI". After recording the ROI the signature list will automatically change.
After carrying out the all steps, the signature list will be formed (Fig.6), in which the signatures have their own auto color.

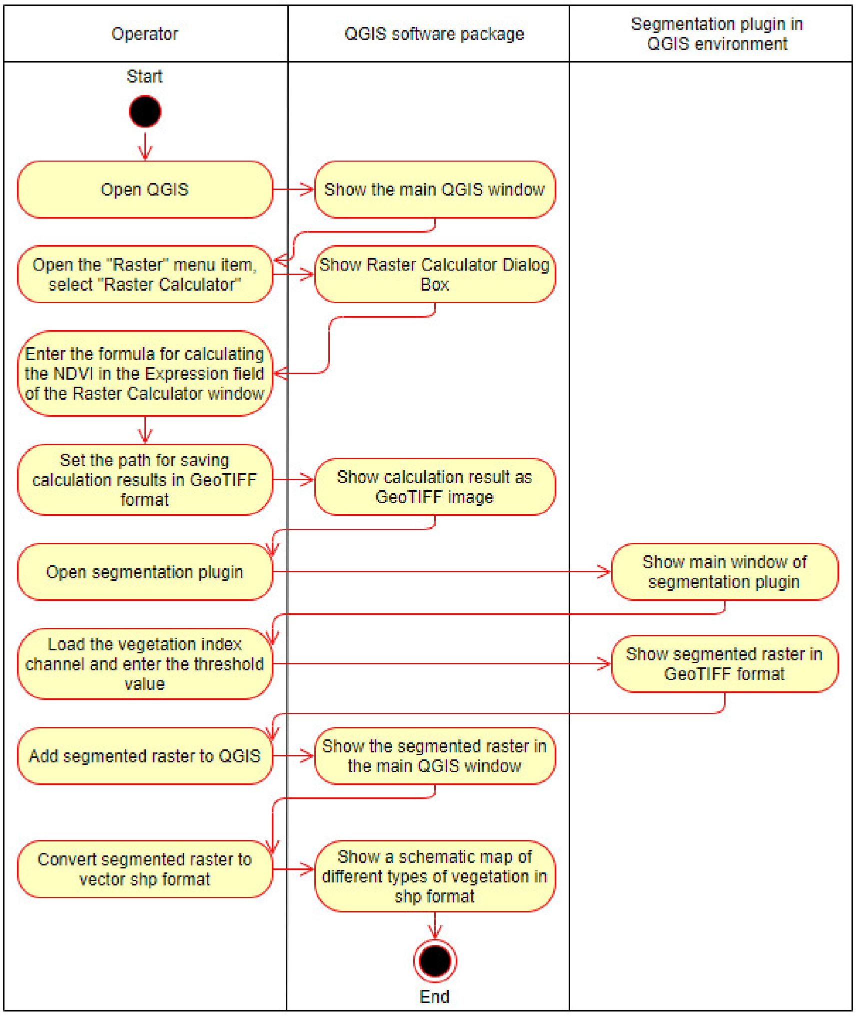

Figure 3. Process diagram of the satellite imagery segmentation

At the next stage, we select the classification method "Spectral Angle Mapping", "Size = 500" with which a raster map is generated, which displays the classification results (Fig.7).

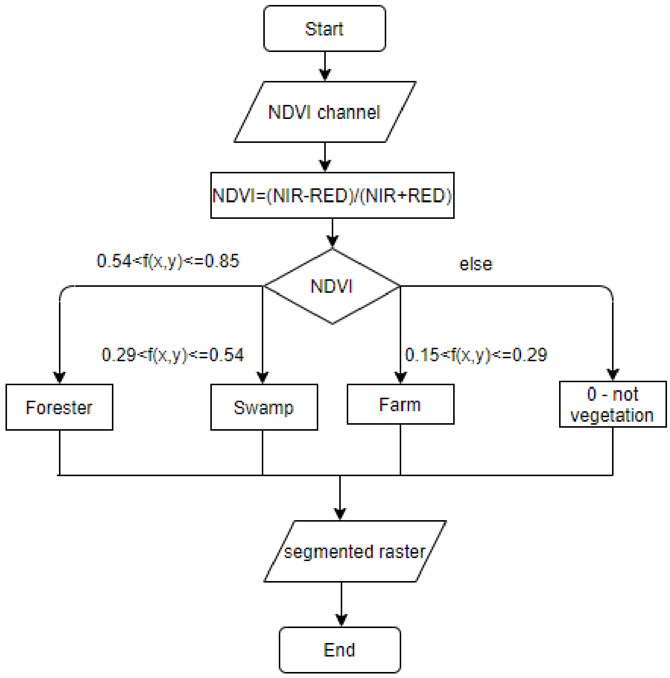

Figure 4. Diagram of the image segmentation algorithm

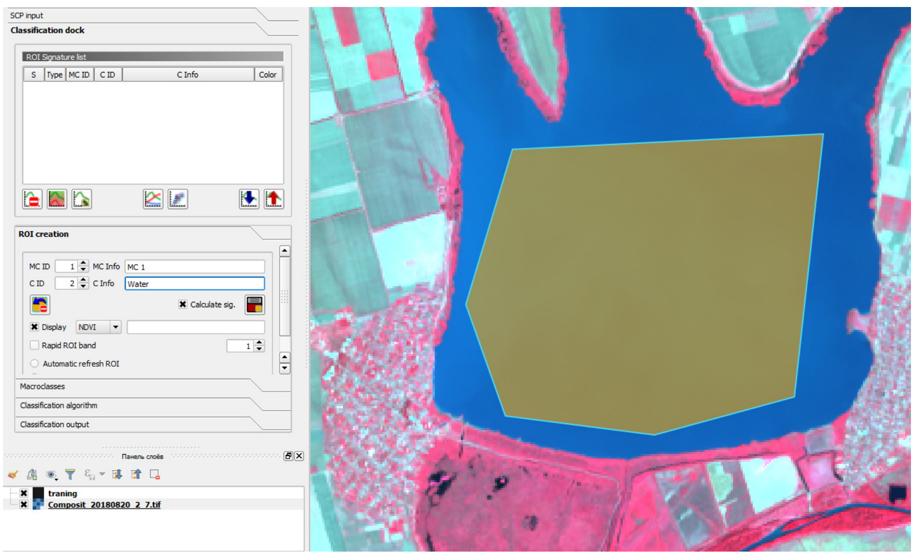

Figure 5. Capturing pixels for water bodies

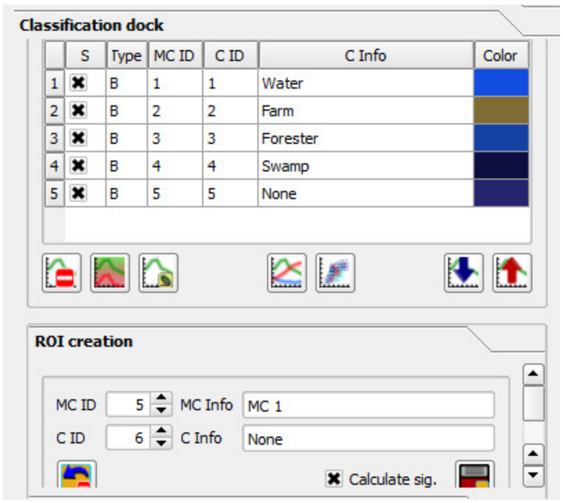

Figure 6 Signature list

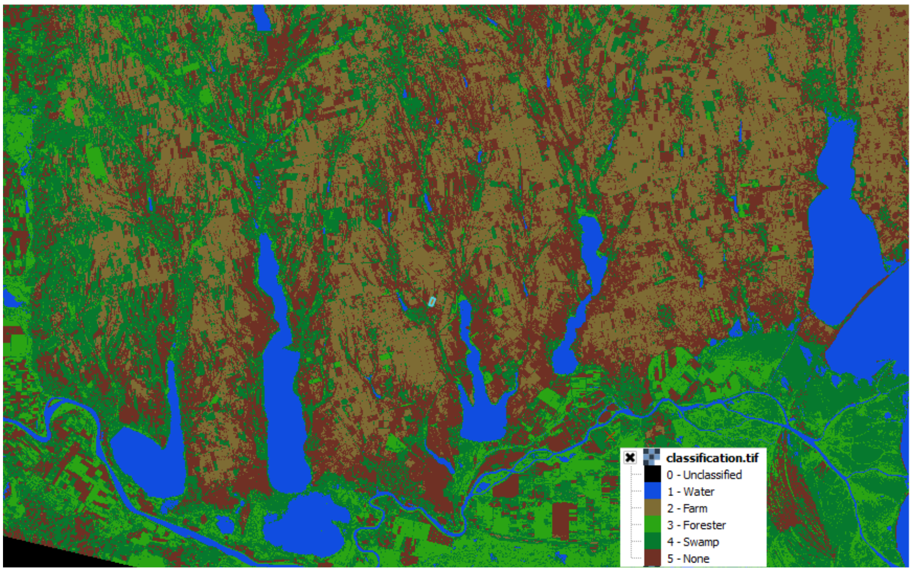

Figure 7. The result of classification with training

The map shows that most of the study area is farmland. It is difficult to classify forests and swamps because they have approximately the same vegetation index. According to the resulting raster classification map, one can build a vector map of vegetation by vectorising it.

The created vegetation index map have made it possible to assess the yield of rice paddy fields in the Odesa region. The task of assessing the yield of rice crops is usually difficult due to the noise effect created by water on flooded checks. The construction of the NDVI index allows one to predict the real productivity of crops and to produce their quantitative estimates. Fig.8 shows NDVI maps for the Kiliyskiy district near

Vilkovo, where rice fields are located. The maps can be used to trace the dynamics of changes in the state of these lands for the vegetation period of 2018.

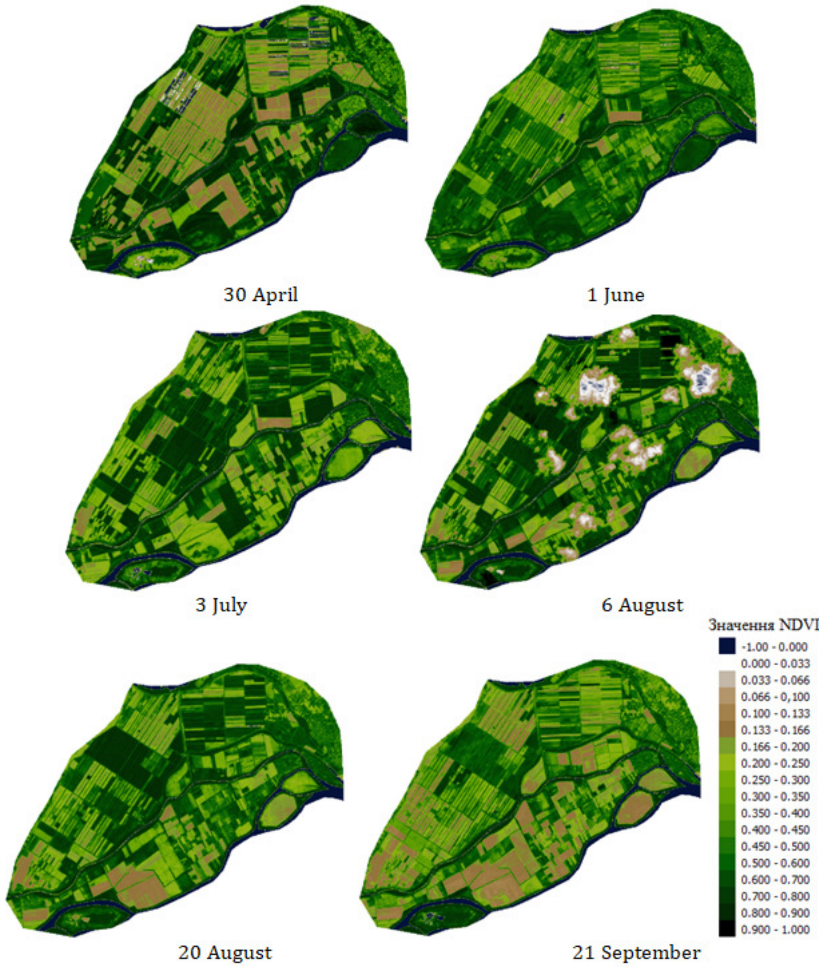

Figure 8. NDVI maps of rice paddies for the vegetation period

4. Conclusions

The above approach to assessing the state of agricultural land is based on the calculation of the NDVI index, which is a quantitative indicator to measure of photosynthetic active biomass. It makes possible to find the areas of sites with different density of vegetation, to identify degraded lands, to trace the dynamics of changes in vegetation cover. Satellite images of different spectral channels of the Landsat satellite with primary atmospheric processing were used as the initial data for calculating NDVI. Archival imageries allowed you to trace changes in vegetation cover over time.

To classify satellite images by density and vegetation type, the parallelepiped classification alghoritm was used, which refers to a learning classification method. Classification and calculation were performed in the open source GIS QGIS using a special plugin Semi-Automatic Classification.

REFERENCES

1. Regional report on the state of the environment in Odessa region in 2016. Department of Ecology and Natural Resources. Available online: *http://ecology.odessa.gov.ua/files/ecology_portal/reg_onal_dopov_d_2016.pdf* (accessed on 21 October 2020).
2. Website of the Center for Social and Environmental Risk Assessment. Available online: *https://coser.com.ua* (accessed on 21 October 2020).
3. USGS. U.S. Geological Survey. Available online: *https://www.usgs.gov/* (accessed on 21 October 2020).
4. Landsat 8 Data Users Handbook. USGS. Available online: *https://landsat.usgs.gov/sites/default/files/documents/Landsat8DataUsersHandbook.pdf* (accessed on 21 October 2020).
5. REED B.C., BROWN J.F., VANDER ZEE D., LOVELAND T.R., MERCHANT J.W., OHLEN, D.O.: Measuring phenological variability from satellite imagery. J. Veg. Sci. (1994)5, 703–714. https://doi.org/10.2307/3235884
6. ROUSE J.W., HAAS R.H., SCHELL J.A., DEERING D.W.: Monitoring Vegetation Systems in the Great Plains with ERTS. Third ERTS Symposium, NASA SP-351, 1973, 309-317.
7. SELLERS P.J.: Canopy Reflectance, Photosynthesis and Transpiration. International Journal of Remote Sensing, 6(1985), 1335-1372.
8. CHAVEZ P.S.: Image-Based Atmospheric Corrections–Revisited and Improved Photogrammetric Engineering and Remote Sensing, [Falls Church, Va.]. Am. Soc. Photogramm. (1996)62, 1025–1036.
9. SAHAR A. EL_RAHMAN: Performance of Spectral Angle Mapper and Parallelepiped Classifiers in Agriculture Hyperspectral Image, International Journal of Advanced Computer Science and Applications(ijacsa), 7(2016)5, http://dx.doi.org/10.14569/IJACSA.2016.070509
10. ALAWADI F., AMOS C., BYFIELD V., PETROV P.: The application of hyperspectral image techniques on MODIS data for the detection of oil spills in the RSA, Proc. SPIE 7110, Remote Sensing for Environmental Monitoring, GIS Applications, and Geology VIII, 71100Q (10 October 2008); https://doi.org/10.1117/12.799374
11. QGIS API. QGIS API Documentation. Available online: *http://www.qgis.org/api/* (accessed on 21 October 2020)
12. GNU Operating System. GNU General Public License. Available online: *https://www.gnu.org/licens es/gpl-3.0.en.html* (accessed on 21 October 2020).
13. CONGEDO L.: Semi-Automatic Classification Plugin Documentation. Available online: *https://fromgist ors.blogspot.com/p/semi-automatic-classification-plugin.html* (accessed on 21 October 2020).