Geospatial Intelligence and Remote Sensing for Climate-Smart and Sustainable Agriculture

Pavlo Lykhovyd

DeepScience

Geospatial Intelligence and Remote Sensing for Climate-Smart and Sustainable Agriculture

Pavlo Lykhovyd

Institute of Climate-Smart Agriculture of NAAS

DeepScience

Published, marketed, and distributed by:

Deep Science Publishing, 2025
USA | UK | India | Turkey
Reg. No. MH-33-0523625
www.deepscienceresearch.com
editor@deepscienceresearch.com
WhatsApp: +91 7977171947

ISBN: 978-93-7185-994-3

E-ISBN: 978-93-7185-854-0

https://doi.org/10.70593/978-93-7185-854-0

Citation: Lykhovyd, P., (2025). *Geospatial Intelligence and Remote Sensing for Climate-Smart and Sustainable Agriculture*. Deep Science Publishing. https://doi.org/10.70593/978-93-7185-854-0

Preface

In the context of global climate change, related threats for food security, and the need for rational land and water resources usage in order to provide environmental sustainability, agricultural sector of Ukrainian economy faces serious challenges. The problems are particularly aggravated in the steppe zone, where unstable weather conditions, lack of natural moisture supply, and soil degradation require the introduction of novel approaches to agricultural production. In this regard, the use of geographic information systems (GIS), remote sensing technologies, and precision farming is essential. These technologies not only allow for dynamic and rapid monitoring of the state of agricultural phytocenoses but also enable the development of scientifically sound strategies for agricultural production management, optimization of resource use, and increased sustainable yields.

This monograph summarizes the results of a five-year scientific work. It consists of sections that sequentially cover the stages of research: from theoretical foundations to practical recommendations. The first chapter presents a review of relevant scientific literature and its detailed scientific analysis. The second chapter is devoted to the methodology of scientific work. The third chapter provides details on the developed mathematical models and methods of agroecological zoning of agricultural lands based on remote sensing data. The fourth chapter is devoted to the possibility of remote identification of agricultural crops and the possibilities of crop mapping based on the dynamic characteristics of vegetation indices. The fifth chapter presents information about the main databases and software products developed using the results of the performed study. The conclusions and recommendations are based solely on the results of scientific work that has been tested and implemented in agricultural farms and higher education establishments.

The results of the research have significant scientific and practical value. They create the basis for the systematic implementation of digital agriculture, contribute to improving the economic efficiency of agricultural production, optimizing the use of resources, and adapting to climate change. In the long term, this will significantly strengthen Ukraine's food security and ensure the sustainable development of the national agricultural sector.

Pavlo Lykhovyd

Table of Contents

Deep Science Publishing, 2025
https://doi.org/10.70593/978-93-7185-854-0

Chapter 1: Geographical information technologies as a cornerstone of modern agriculture

Nowadays, agriculture faces numerous challenges: from the need to increase productivity to satisfy the growing demand for food to simultaneous diminishing negative impacts on the environment and climate. To respond to these challenges, agricultural sector addresses to modern innovative technologies, in particular geographic information systems (GIS) and related digital technologies, which have become one of the cornerstones of modern precision farming. In its current state, GIS is not limited to digital maps and geospatial databases, but it is a powerful analytical tool that allows one to collect, store, analyze, predict and visualize spatial data of various formats and thematic scope, including those directly related to agricultural productivity. The use of GIS technologies in agriculture is multifaceted and covers almost every stage of the plant products manufacturing cycle.

1 Land management

One of the key aspects is mapping landscapes, soil and water resources. Using GIS tools that integrate data from GPS receivers, soil sensors, satellite images, detailed field maps are created, which, depending on the spectral channels of the images, allow various assessments of the condition of land to be made. Thus, remote sensing combined with GIS allows creating maps of soil fertility heterogeneity, determining the boundaries of water bodies, analyzing soil reclamation indicators, its physical, chemical, and mechanical properties, crops mapping, and determining the actual irrigation areas, etc. Based on aerospace data integrated into GIS, farmers can create maps of agrotechnological operations and apply differentiated fertilization, which allows them to optimize costs and prevent soil pollution; perform remote control for pests, diseases, weeds and optimize chemical protection of crops; perform harvesting campaign monitoring, etc. For scientific and practical purposes, satellite monitoring is widely used for cadastral surveys, updating boundaries and dynamic control of agricultural land use, mapping crop rotations and assessing the structure of cultivated areas, etc. [1].

Remote sensing data integrated into modern GIS significantly improves the economic and environmental efficiency of dynamic crop monitoring. Satellite or aerial photography (unmanned aerial vehicles (UAVs), drones) is used to obtain multispectral images, which are analyzed to calculate specific vegetation indices. Depending on the intended purpose, vegetation indices are estimated based on different spectral channels. One of the most popular indices in agricultural science and practice is the NDVI (Normalized Difference Vegetation Index), which is a universal indicator of the density and overall vigor of vegetation cover, allowing the early detection of problem areas affected by pests, diseases, or moisture deficiency. The NDMI (Normalized Difference Moisture Index) is actively used to assess water stress and provide agricultural crops irrigation scheduling. The EVI (Enhanced Vegetation Index) is a version of NDVI optimized for atmospheric distortion and therefore sometimes serves as a substitute for the latter in scientific research and farming. More specific indices such as SMI (Soil Moisture Index) are used to assess moisture accumulation in soils, NDWI (Normalized Difference Water Index) – for assessing water stress, irrigation efficiency, and establishing the boundaries of water bodies; NRI (Nitrogen Reflectance Index) – for assessing plant nitrogen availability; SAVI (Soil Adjusted Vegetation Index) – for a more accurate assessment of vegetation cover with soil correction, etc. In total, over 100 different vegetation indices are available now, both highly specific and more general in nature. However, only a limited number of vegetation indices found practical use, and most complex indices have their own scientific and theoretical significance [2, 3].

2 Irrigation and plant protection management

It is extremely important to ensure the rational use of available water resources, especially in the context of global warming, increasing freshwater scarcity, pollution of natural water bodies due to irrational anthropogenic activity, and increased pressure on water bodies due to growing demands for drinking water quality and quantity. GIS-controlled irrigation systems allow for differentiated irrigation based on optimized water use algorithms. Instead of uniform irrigation of the entire field, GIS systems analyze soil moisture maps and specific vegetation indices, which are used as the markers of water stress, supplying water only to those areas that need it most. This not only saves water resources, but also prevents waterlogging, which can be harmful to crops, and significantly increases the ecological and economic efficiency of irrigated agriculture [4].

Similarly, GIS is used for targeted spraying of crops as part of an integrated plant protection system. Modern aerial photography, combined with machine learning tools, allows for the accurate identification of areas with high levels of weed infestation or specific types of pests and diseases. This enables the creation of virtual field maps with clear geographical coordinates of problem areas. Using maps of the spreading phytopathogens, created based on data from drones, self-propelled sprayers are configured so that the system applies the appropriate control measures only to the affected areas. This significantly reduces the use of pesticides, which reduces the environmental impact and lowers economic costs [5].

3 Driving agricultural machinery

Integrating GIS with GPS navigation systems provides unique opportunities to automate the process of driving agricultural machinery. Autopilots on tractors, sprayers, and self-propelled harvesters ensure the best routes and provide the highest efficiency of agricultural aggregates movement in the field, minimizing overlaps or gaps when performing agrotechnological operations. As a result, farmers save fuel, labor and time and optimize the use of available technological resources. GIS, in combination with management decision support systems and artificial intelligence, allows for the creation of the most efficient routes for machinery, considering the macro and micro relief, soil features and configuration of the plots, helping to avoid soil erosion, over compaction, irrational resource usage and save time on operation conduction [6].

4 Monitoring and prediction

The use of GIS allows not only to analyze the conditions of agricultural land in the moment, but also to predict future trends. For example, analysis of historical data on crop yields, weather and soil conditions helps to build predictive models to forecast the likely yield in the next season. This helps farmers to make more informed and reasonable decisions about the selection of crops in specific agro-ecological conditions, plan investments in equipment and resources, and perform product marketing planning. At the state level, this creates the conditions for a rational agricultural and food policy and ensures sustainable development and food security.

In modern agriculture, GIS is becoming a key tool for increasing production efficiency, reducing risks, and ensuring resilience to the negative effects of climate change. However, it is impossible to unlock the whole potential of GIS without proper integration with remote sensing data, which allows fetching real-time information about the on-surface objects and their conditions.

GIS monitoring involves regular dynamic observation of the croplands and soil surface conditions, rapid detection of problems such as soil erosion or pest spreading and helps with assessment of the impact of climate factors on agricultural productivity. Prediction, in its turn, is based on the models that integrate historical data with current observations.

For example, WOFOST or SVAT models integrated into GIS allow simulating crop development dynamics based on data on precipitation, temperature, and soil conditions [7]. Such forecasts help farmers optimize sowing and harvesting times, reducing losses from adverse conditions. There are other forecasting algorithms, such as universal mathematical methods of moving average, triple exponential smoothing, autoregressive models, etc.

Remote sensing (RS) is a technology for obtaining information about the Earth's surface and on-land objects without direct contact, using satellites, unmanned aerial vehicles (UAVs), or aviation sensors. RS data includes spectral images in various ranges (from visible to infrared), which allow for the assessment of vegetation, soil moisture, and other parameters.

The integration of RS with GIS transforms raw data into useful information. GIS allows data layers to be superimposed: satellite images are combined with cropland or soil maps, meteorological data, and historical yield records. This integration results in multi-layered maps that visualize spatial variations and trends across multiple domains of agricultural production. This approach is particularly useful for large farms where manual monitoring is impossible because of scale, as well as for scientific work with a wide range of parameters and their interrelationships, which is difficult to handle manually. Also, GIS plus remote sensing combo is frequently used for large-scale research, where it is impossible to cover the whole area with on-land surveys.

The advantages of remote sensing include timeliness (data is updated daily or weekly), objectivity (no human factor), and coverage of large areas. For example, the European Space Agency's satellites Sentinel-2 fetch images with a resolution of 10 meters, allowing for detailed monitoring of even individual plots within a research field.

Integrating GIS and remote sensing starts with data collection: satellite images are processed to remove noise (e.g., cloud cover) and for geographic and radiometric correction, after which they are imported into GIS programs such as ArcGIS or QGIS. Here, the data is georeferenced – linked to coordinates on the map. Next, analysis algorithms are applied: crop classification, calculation of vegetation indices, and subsequent yield modeling based on the values of selected input parameters.

At the present stage, various machine learning algorithms (both traditional and modern, depending on the research needs, the set and nature of input data, computing power, and specialist competencies) integrated with GIS are most often used for forecasting. For example, artificial neural networks are trained on historical remote sensing data to forecast crop yields with an error of less than 10%. For example, in Kazakhstan, researchers have developed an algorithm for predicting sunflower yields using NDVI data (an index calculated from Sentinel-2 satellite images), achieving an error rate of 0.67% to 10.7%. This demonstrates how the integration of remote sensing, GIS, and machine learning improves the accuracy of crop yield forecasts and opens new prospects for highly efficient agricultural production planning [8].

One striking example of the effective use of the tandem of remote sensing, GIS, and machine learning is the classification and mapping of agricultural crops carried out in India (Karnataka) using NDVI from the Landsat 8 satellite in the ArcGIS system [9]. Another example is the use of UAVs for high-resolution monitoring. In Iraq, in the province of Nineveh, the integration of data from Landsat, Sentinel, and drones made it possible to create suitability maps for wheat with a classification accuracy of 90%. NDVI correlated with yield at R^2=0.85, which helped identify problems with nitrogen in the soil [10]. In the United States, according to information provided in USDA reports, more than 50 million acres of agricultural land are monitored automatically using remote sensing data to assess and dynamically control soil conditions.

Yield forecasting is a key area for the implementation of GIS and remote sensing. In China, for example, remote sensing data based on SVAT model algorithms was used to forecast yields and calculate the water use efficiency of winter wheat and corn [11]. Remote sensing data, combined with regression models and mapping tools, were effectively used in pre-season forecasting of sugarcane productivity and water use efficiency for this crop [12]. Ukrainian scientists have been successfully integrating remote sensing data into mathematical models for forecasting the yield of major agricultural crops for over 10 years. Basically, domestic scientists use vegetation indices to build regression models of productivity, such as the model of spring barley productivity depending on the NDVI value developed at the Pogorily Ukrainian Research Institute of Plant Industry [13]. However, recently, increased attention has been paid to expanding the range of vegetation indices used and strengthening the mathematical algorithms applied in such forecasts. A striking example of modern forecasting of winter wheat yield using remote sensing data in the form of NDVI and Chlorophyll Index (CI) vegetation indices based on linear approximation algorithms is the work carried out at the National University of Life and Environmental Sciences of Ukraine [14].

The effectiveness of GIS and RS integration is justified by several factors. First, it is highly accurate: recent studies show that forecast errors are within 10-15%. This surpasses traditional methods based on sample surveys. Second, it is economic feasibility: reduction of labor, time, and material resources expenditures. Third, it is sustainability: technologies help to effectively combat climate change, monitor and predict adverse climatic events (droughts, floods, storms), and prevent their devastating impact on agroecosystems. In the EU and the US, GIS and RS are technologies that guarantee food security and rational use of natural resources. Fourth, it is accessibility: open-source software for working with GIS, such as QGIS, reduces costs, making technologies accessible to everyone.

5 Benefits and challenges

The benefits of implementing GIS technologies are obvious:

- Increased cropland productivity and resource payout: Optimizing resources usage allows achieving maximum crop productivity under the lowest possible cost.
- Reduced costs: Precision GIS-based farming helps to cut the expenditure on fertilizers, plant protection products, fuel, and water.
- Reduced environmental footprint: Targeted use of chemicals and water minimizes soil and water pollution, providing for environmental sustainability.
- Improved management: Ability to make reasonable data-driven decisions rather than leaning upon intuition.

However, there are also certain challenges:

- High cost: The implementation of GIS systems, sensors, and GPS equipment requires significant initial investment.
- Technical complexity: The use of these technologies requires appropriate knowledge and skills from staff.
- Data availability: The quality and availability of data (e.g., satellite imagery) may be limited in some regions.

Given the above, GIS technologies are now an integral part of precise climate-smart and environmentally friendly farming. They are transforming agricultural production from traditional and intuitive to high-tech and scientifically based. Despite the initial difficulties and investments, the long-term benefits in terms of increased efficiency, environmental friendliness, and economic benefits make GIS systems an essential tool for every modern agricultural enterprise striving for sustainable development and competitiveness in the global market [15].

Deep Science Publishing, 2025
https://doi.org/10.70593/978-93-7185-854-0

Chapter 2: Methodology of research work

The research methodology based on Ukrainian national and internationally recognized methods for the studies in the field of GIS and information technologies (IT). The following methods and methodological approaches were used during the research work:

- Retrospective historical analysis of research data.
- Analytical and synthetic methods for processing and summarizing literature search and research results.
- Modern methods of conducting field experiments with major agricultural crops, which include vegetation and phenological observations, phytopathological surveys, agrometeorological observations, measurement of plant assimilation apparatus parameters, selection and analysis of soil samples in a certified laboratory of the Institute of Climate-Smart Agriculture of the National Academy of Agrarian Sciences, recording of yields of main and secondary products, etc.
- Mathematical and statistical methods for constructing mathematical and statistical models, establishing links between crop production processes, agrotechnology parameters, geographical components, and remote sensing data.
- GIS methods for integrating field research data, satellite monitoring, and geographical components of crop productivity models.
- Programming methods for working with cloud platforms such as Google Earth Engine and for developing software and databases.
- Cartographic and graphic packages Adobe Illustrator, MapChart for building maps; GIS-software QGIS to perform vegetation indices calculation based on satellite imagery.

1 Introduction to vegetation indices NDVI and LAI

Vegetation index is calculated based on the ratio of values retrieved from different spectral bands of remote sensing imagery. This index indirectly relates to vegetation parameters through estimation of spectral reflectance values for each pixel of spatial image. The implementation fields of vegetation indices are determined by the reflectance features. Usually, calculation of vegetation indices is performed using two most stable areas of the spectral reflectance curve of plants.

The Normalized Difference Vegetation Index (NDVI) is one of the most popular tools in precision agriculture for crop conditions monitoring. This index provides a quantitative assessment of vegetation density and vigor, allowing agronomists to evaluate germination, observe growth and development, control presence of weeds and infestation with pests and diseases, as well as predict crop productivity. The NDVI is based on the analysis of spectral characteristics of vegetation obtained from satellite images. Plants with high photosynthetic activity actively absorb electromagnetic radiation in the visible red range of the spectrum (0.62–0.75 μm), where chlorophyll has maximum absorption of solar radiation. At the same time, they intensively reflect radiation in the near-infrared range (0.75–1.3 μm), since the cellular structure of leaves maximally reflects energy in this zone.

Healthy plants show a low reflectance coefficient in the red spectral band and a high one in the near-infrared. The ratio of these coefficients allows to distinguish vegetation cover from other natural objects, such as soil or water. Based on the calculated ratio, the areas with abnormalities, which require proper agrotechnological intrusion, are identified.

However, NDVI has a list of limitations. The index is moderately sensitive to soil albedo and atmospheric conditions. Also, it may fail in the areas with sparse vegetation. In addition, in dense vegetation, the index may show so-called saturation effect, where dense canopy prevents adequate estimation of real green biomass conditions. This can lead to an underestimation or overestimation of the actual crop density and health.

There are over 150 different vegetation indices developed for precision farming, but the NDVI remains the most popular. Its spread is due to its simplicity and ease-in-use and interpretation, making it a versatile tool for monitoring vegetation status for farmers and scientists of different scholar levels.

Another important feature is spatial resolution of the imagery used to estimate NDVI. It is important to understand that the advantage in spatial resolution does not resolve issues of the index sensitivity to atmospheric distortions and soil albedo. It is the key for proper index utilization in different scenarios, such as local (field-level), regional or country-level or even global research scale. High resolution images (e.g., 1-10 meters) are perfect for detailed in-field analysis, while low resolution images (500-4000 meters) should be used in large-scale studies.

In contrast, other vegetation indices use different spectral channels, e.g. red-edge. They often have a lower resolution of 20 meters and higher, therefore, they are unsuitable or limited suitable for detailed in-field studies. These indices are designed to address specific tasks, such as detecting plant stress in the early stages or estimating nitrogen content, but their lower spatial detail may limit their application for accurate analysis of small areas. Thus, the advantage of NDVI lies in its versatility and effective combination with available high-quality data, which provides both high detail and a wide range of applications. The NDVI calculation in QGIS 3.22 (Quantum GIS) for Windows is performed using the QGIS raster calculator with pre-loaded data. The standard formula is used for this as follows:

$$NDVI = \frac{(NIR - Red)}{(NIR + Red)} \tag{2.1}$$

where NIR is the absorption and reflection of infrared rays; Red is the absorption and reflection of red rays [16].

Essentially, each vegetation index is a specific combination of measurements taken by sensors across two or more channels with different wavelengths of reflective properties (water content, chlorophyll, pigments, etc.), which characterize the state of vegetation in a certain way. As sensors improve, Earth observation satellites feed new data into remote sensing developments, improving existing analysis methods.

The Leaf Area Index (LAI) was developed to analyze the leaf surface of our planet in order to calculate the area of leaves in each area. It is calculated as the ratio of the one-sided (illuminated) leaf area to the soil surface it occupies. The LAI leaf area index is an important indicator in studies examining the state of agricultural crops, forest plantations, the environment, and climatic conditions.

LAI is defined as a ratio of plant canopy area to the soil surface area. The value obtained from the satellite corresponds to the total green LAI (embracing vegetation layers and undergrowth, which can represent a very significant contribution, especially in case of forests). In plain English, LAI describes the thickness of the vegetation cover. The LAI is estimated for both individual crops or entire regions and plant groups. It is calculated using following equation (2.2):

$$LAI = \frac{leaf\ area}{soil\ area} \tag{2.2}$$

The leaf area index was introduced for the NASA MODIS sensor to refine NDVI data. Unlike the latter, it takes into account topography, and the spectral bands used for its calculations undergo atmospheric correction. If the LAI index = 3, the leaves can cover the occupied surface in three layers. The LAI vegetation index is considered high at ≥ 3.5; however, its values increase in the presence of clouds and bright objects [17].

To compare and monitor the condition of grain corn crops cultivated on the Institute's experimental plots in Kherson region, not only conventional methods were used, but also satellite monitoring. The study used data from three satellites: Landsat-8, Sentinel-2, and MODIS, which have several differences.

The advantages of Landsat-8 are characterized by indirect characteristics, since its spatial resolution (the size of the smallest detail that the sensor can detect on the ground) ranges from 15 to 60 meters per pixel, depending on the range, which guarantees the highest image quality and the most accurate results compared to other satellites, However, the imaging frequency is 16 days, which does not meet the need for real-time monitoring of crop conditions.

The Sentinel-2 satellite had slightly better characteristics, with a spatial resolution of 10 to 60 meters per pixel, depending on the range and mode, and an imaging frequency of every 10 days.

MODIS sensors provide images with a spatial resolution of 250–1000 meters, depending on the range, which causes a large error on small experimental plots, although the advantages include a fairly high imaging frequency of 1–2 days and the ability to select the best quality samples. As of October 1, 2022, the satellite no longer provides new images; only a set of data for historical reference remains available, and users have been advised to switch to EVIIRS.

Due to different specifications, sensor types, and settings, the NDVI calculation formula for each satellite has different names for the calculation bands. To calculate based on Landsat-8 satellite image data, the formula takes the following form (2.3):

$$NDVI = \frac{(band5 - band4)}{(band5 + band4)}, \quad (2.3)$$

where band5 is the absorption and reflection of infrared rays; band4 is the absorption and reflection of red rays.

To calculate the data from Sentinel-2 satellite images, the formula takes the following form (2.4):

$$NDVI = \frac{(band8 - band4)}{(band8 + band4)}, \quad (2.4)$$

where band8 is the absorption and reflection of infrared rays; band4 is the absorption and reflection of red rays.

To calculate the MODIS satellite image data, the formula takes the following form (2.5):

$$NDVI = \frac{(band2 - band1)}{(band2 + band1)}, \quad (2.5)$$

where band2 is the absorption and reflection of infrared rays; band1 is the absorption and reflection of red rays.

Usually, NDVI calculation results vary from -1 to 1. The values below zero correspond to areas without objects different from plant vegetation; bare soil values are usually within 0.1-0.2; and active vegetation cover will always have positive values from 0.2 to 1. Healthy, dense vegetation cover's NDVI is usually >0.5, while weak vegetation is likely to be in the range of 0.2-0.5. However, this is an empirical rule, as individual features must always be considered to know exactly what the NDVI values mean.

The choice between Landsat-8, Sentinel-2, MODIS, and VIIRS for calculating the NDVI index depends on the specific requirements and research objectives.

High resolution: Landsat-8 has higher resolution (typically 30 meters) compared to MODIS and VIIRS, providing more detailed images. This is particularly important for small area studies or studies, which are concentrated on specific objects in detail. Wider spectral range: Landsat-8 has a wider spectral range compared to MODIS and VIIRS, fetching more potentially useful information about the studied objects. It is useful for studying different types of plants and vegetation cover. Landsat-8 has regular missions and short lag between images, allowing dynamic tracking vegetation dynamics. There are many programs and tools for processing Landsat-8 data that allow for the analysis of NDVI and other vegetation indices. However, MODIS and VIIRS have their own advantages, such as higher data update frequency and larger coverage area. They are useful for monitoring large areas and global changes in vegetation.

The choice of Landsat-8 for the study was based on circumstances such as the size of the study area, which did not exceed 1.6 hectares, the study period, since 15-16 days are sufficient to track the development of corn, the availability of data, and the number of resources for processing and analyzing this data. An example of an image processed in the EOSDA Crop Monitoring system from this satellite is shown in the figure below (Figure 2.1) [18].

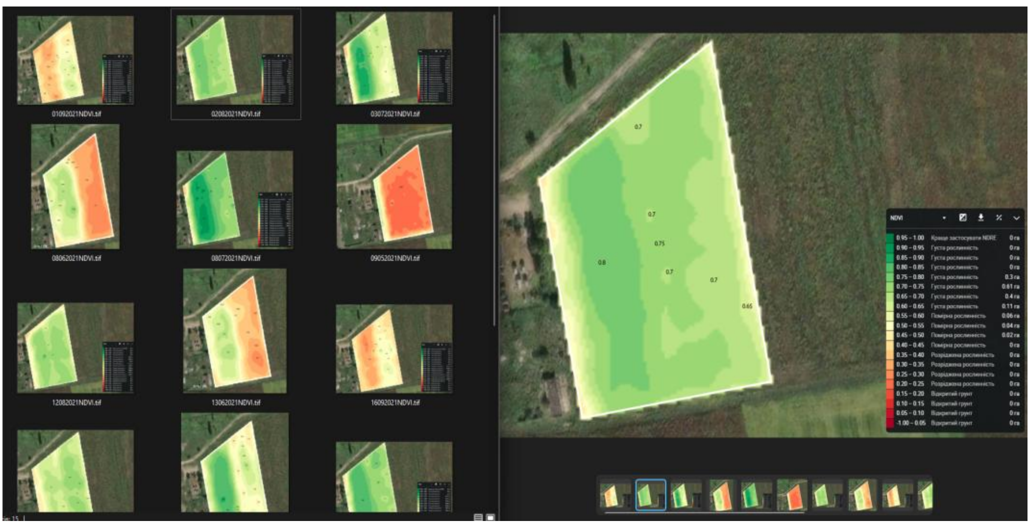


Fig. 2.1 Processed and analyzed images from the Landsat-8 satellite for 2021

2 Mathematical statistics methods

The scientific work was accompanied by a series of mathematical calculations.

Pearson's linear correlation coefficient indicates the strength of the relationship between the variables under study [19]. The calculations were performed using formula (2.6):

$$R_{XY} = \frac{\sum_1^N (x_i - \bar{x})(y_i - \bar{y})}{(N-1)s_X s_Y} \tag{2.6}$$

where s_X, s_Y – standard deviations in samples X and Y, respectively; N – number of data in the sample; x_i, y_i – values of the studied variables in pairs; (“x,”) ̅“y” ̅– arithmetic means for samples X and Y, respectively.

The interpretation of the closeness of the relationship between the studied parameters was performed according to Evans (1996) [20] (Table 2.1). The direction of the relationship is determined by the sign of the correlation coefficient: “+” (direct) and “–” (inverse).

Table 2.1 Degrees of closeness of correlation between parameters based on Pearson's correlation coefficient

Absolute value of |*R*|	Closeness of relationship
0,80-1,00	Very strong
0,60-0,79	Strong
0,40-0,59	Moderate
0,20-0,39	Weak
0,00-0,19	Very weak or no relationship

The coefficient of variation is an important statistical parameter in life science studies, which indicates the degree of variation of the feature of the studied object. It can be expressed both as a numerical value (from 0 to 1) and as a percentage. The calculation was performed using formula (2.7) [21]. It is generally accepted that a coefficient of variation of <10% corresponds to a low level of dispersion of the trait, 10-25% to an average level, and >25% to a high level.

$$CV = \frac{SD}{x} \tag{2.7}$$

where SD is the standard deviation; x is the arithmetic mean for the sample data under study.

The standard deviation (standard error) is a numerical expression of the uncertainty of the characteristics of the studied object and gives an idea of the range of its fluctuation within the upper and lower limits. The calculation was performed using the formula (2.8) proposed by Bland & Altman (1996) [22]:

$$SD = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(x_i - x)^2} \quad (2.8)$$

where N is the number of data points in the sample; x_i is the value of the indicator; x is the arithmetic mean for the sample under study.

Linear regression (simple or multiple) allows establishing a relationship between the parameters under study and describing it mathematically.

Regression analysis of data involves calculating regression statistics parameters, regression coefficients, and variance [23]. Additionally, statistical criteria are used for a comparative assessment of models.

Regression statistics are primarily correlation coefficients R and determination coefficients R^2 (ordinary, adjusted, and predicted), as well as the root mean square error. The coefficient of determination reflects the fitting quality of the model and is an indirect marker of its accuracy and reliability. It is calculated using formula (2.9):

$$R^2 = 1 - \frac{SS_{error}}{SS_{total}} \quad (2.9)$$

The adjusted (often referred to as normalized) coefficient of determination considers the amount of input data that explains changes in the model and is calculated using equation (2.10):

$$R^2_{adj} = 1 - (1 - R^2)\frac{N-1}{N-k-1} \quad (2.10)$$

where k is the number of predictors; N is the total number of data pairs in the sample.

The predicted coefficient of determination is calculated to assess the predictive value of the model, taking into account the predicted sum of squares of residuals d (PRESS) according to formulas 2.11 and 2.12:

$$R^2_{pred} = 1 - \frac{PRESS}{SS_{Total}} \quad (2.11)$$

$$PRESS = \sum d_i^2 \quad (2.12)$$

The mean square error (MSE) indicates the magnitude of the deviation of the modeled value of the studied indicator from the actual value and was calculated using formula (2.13):

$$MSE = \frac{1}{n}\sum_{i=1}^{n}\left(Y_i - \hat{Y}_i\right)^2 \quad (2.13)$$

where n is the number of data points in the input sample; Y_i-$\hat{Y}_i$ is the difference between the actual and predicted values of the target function.

The equation of the linear regression model is as follows (2.14):

$$Y = a_1x_1 + a_2x_2 + \cdots + a_kx_k + c + e \quad (2.14)$$

where Y is the dependent variable; ai are the regression coefficients for each independent variable of the corresponding independent arguments x_i; c is a constant (not a mandatory component of the model); e is the error (residuals).

Polynomial regression describes the relationship between variables using an nth-degree polynomial. Such models are nonlinear, which allows them to better describe the characteristics of processes in agroecosystems. At the same time, polynomials are not purely nonlinear models. Technically, polynomials are an advanced version of multiple linear regression [24]. Model selection and construction are performed using the least squares method. The polynomial equation has the form (2.15):

$$Y = \beta_0 + \beta_1 x + \beta_2 x^2 + \cdots + \beta_n x^n + \varepsilon \quad (2.15)$$

where β – regression coefficients; x – independent variable; n – degree of polynomial; ε – random error with zero mean.

Regression coefficients are calculated using formulas (2.16) and (2.17):

$$\beta_0 = \frac{1}{M}\sum_i y_i - \frac{\beta_1}{M}\sum_i x_i \quad (2.16)$$

$$\beta_i = \frac{M\sum_i x_i y_i - \sum_i x_i \sum_i y_i}{M\sum_i x_i^2 - (\sum_i x_i)^2} \quad (2.17)$$

where M is the number of data points in the sample; x is the independent variable; y is the dependent variable.

The accuracy of the model was assessed based on the mean absolute and relative errors. The MAPE is calculated using formula (2.18) [25]:

$$MAPE = \frac{100\%}{n}\sum_{t=1}^{n}\left|\frac{A_t - F_t}{A_t}\right| \quad (2.18)$$

where A_t – actual value of the indicator; F_t – predicted (simulated) value of the indicator.

MAE is calculated using equations (2.19, 2.20) [26]:

$$MAE = \frac{\sum_{i=1}^{n}|e_i|}{n} \quad (2.19)$$

$$|e_i| = |y_i - x_i| \quad (2.20)$$

where y_i is the predicted (simulated) value of the indicator; x_i is the actual value of the indicator.

Additional statistical criteria can be calculated for comparative evaluation of models.

The Darbin-Watson (DW) criterion is a statistical marker of autocorrelation in a regression model (first-order autocorrelation, ρ1), calculated using equation (2.21):

$$DW = \frac{\sum_{t=2}^{n}(\epsilon_t - \epsilon_{t-1})^2}{\sum_{t=1}^{n}\epsilon_t^2} \approx 2(1-\rho_1) \quad (2.21)$$

When there is no autocorrelation, the DW criterion approaches 2.0; when there is positive autocorrelation, DW is close to zero, and when there is negative autocorrelation, it is close to 4 [27].

The Akaike information criterion (AIC) is a statistical indicator for assessing the error of a non-selective forecast and is an indicator of the quality of a statistical model. The criterion is calculated using equation (2.22) [28]:

$$AIC = 2k - 2\ln(\hat{L}) \quad (2.22)$$

where k is the number of model parameters; L is the maximized value of the likelihood function for a given model.

For better assessment of the models accuracy when small data set was used, the adjusted Akaike criterion (AIC_c) is calculated considering the sample size (n) using formula (2.23) [29]:

$$AIC_c = AIC + \frac{2k^2 + 2k}{n-k-1} \quad (2.23)$$

Bayes' information criterion (BIC) or Schwarz's criterion – a smaller BIC corresponds to a better model. This criterion is closely related to the Akaike criterion. It is calculated using formula (2.24) [30]:

$$BIC = \ln(n)k - 2\ln(\hat{L}) \quad (2.24)$$

In regression models, the Akaike statistical criterion proves to be asymptotically optimal if it is necessary to select a model with minimal mean square error, while the Bayesian criterion better identifies the "true model."

The Hannan-Quinn criterion (HQC) is an additional metrics for choosing the best model. It is calculated using equation (2.25) [31]:

$$HQC = -2L_{max} + 2k\ln(\ln(n)) \quad (2.25)$$

It is one of the most stable and reliable criteria, which is not asymptotically efficient, though it has not gained much practical implementation.

Time series forecasting using the Holt-Winters triple exponential smoothing method was performed using an AAA-type algorithm with a confidence interval of 95% or $p<0.05$ [32]. Seasonality parameters – 6 years. The calculation algorithm is given below (2.26-2.33):

$$u_i = \alpha\left(\frac{y_i}{s_{i-c}}\right) + (1-\alpha)(u_{i-1} + v_{i-1}) \tag{2.26}$$

$$v_i = \beta(u_i - u_{i-1}) + (1-\beta)v_{i-1} \tag{2.27}$$

$$s_i = \gamma\left(\frac{y_i}{u_i}\right) + (1-\gamma)s_{i-c} \tag{2.28}$$

$$y_i = (u_{i-1} + v_{i-1})s_{i-c} \tag{2.29}$$

$$y_{i+h} = (u_i + hv_i)s_{i+h-ch'} \tag{2.30}$$

$$h' = INT\left(\frac{h-1}{c}\right) + 1 \tag{2.31}$$

$$v_c = \frac{1}{c}\sum_{j=1}^{c} y_j \tag{2.32}$$

$$s_i = y_i/u_c \tag{2.33}$$

where $0 < \alpha \leq 1$; $0 \leq \beta \leq 1$; $0 \leq \gamma \leq 1 - \alpha$; u_i is baseline; v_i is trend (or slope); s_i is seasonality; $1 \leq i \leq c$.

The mathematical essence of the moving average method is to calculate the arithmetic mean or weighted mean of a set of consecutive values of a time series for a certain fixed period of time (window). For a simple moving average (SMA), the forecast for the next period is the average value of the previous ‘n’ observations (2.34):

SMA_t = (X_{t-1} + X_{t-2} + ... + X_{t-n}) / n (2.34)

where SMA_t – simple moving average value at time t (forecast for the next period); X_{t-i} – actual value of the time series at time t-i; n – window size (number of previous periods used to calculate the average).

Thus, mathematically, the moving average method is a linear combination of previous values of the time series, where the coefficients of this combination are equal weights. This process is repeated, “sliding” along the time series to obtain a forecast for each subsequent period.

The main idea of the analysis of variance (ANOVA) is to break down the total variance of a random variable into independent components that can characterize the influence of a specific factor. The mathematical algorithm of ANOVA is described in [33, 34].

The source of variation is the components of the model (terms) into which the sum of squares is divided. The total sum of squares is divided into the between-groups sum of squares (effect) and the within-groups sum of squares (error).

The sum of squares (SS) is the sum of squares caused by a component. Intergroup SS variance is the sum of squares caused by the influence of the factor under study. Intragroup SS variance (partial, residual, random) characterizes a random variable. This part of the variation is caused by the influence of unaccounted factors and does not depend on the factor under study. The number of degrees of freedom (df) or the term of the component with a total number of input data pairs N (2.35-2.37):

$$df_{general} = N - 1 \qquad (2.35)$$

$$df_{effects} = \text{Number of groups} - 1 \qquad (2.36)$$

$$df_{error} = df_{general} - df_{effect} \qquad (2.37)$$

Mean square (MS) – mathematical expectation of the sum of squares, averaged value of the corresponding SS (2.38).

$$MS = \frac{SS}{df} \qquad (2.38)$$

The F-test for assessing reliability according to Fisher's criterion under the null hypothesis is distributed as (2.39):

$$F = \frac{MS_{effect}}{MS_{error}} \qquad (2.39)$$

If the significance level is less than the critical α, the null hypothesis is rejected, and it can be concluded that not all mean values for the factor are equivalent.

3 Introduction to supplementary vegetation indices

Besides the NDVI vegetation index, which is the most common in the scientific and practical activities of agronomists, we investigated additional vegetation indices. The EVI vegetation index is calculated using formula (2.40) [35]:

$$EVI = 2{,}5\frac{NIR - Red}{NIR + 6Red - 7{,}5Blue + 1} \qquad (2.40)$$

where NIR – reflection in the near-infrared range of the spectrum; Red – values for the red spectrum; Blue – values for the blue spectrum.

The normalized difference moisture index (NDMI), which is a marker of water stress in cultivated plants and the overall moisture supply of crops, was calculated using formula (2.41) [36]:

$$\text{NDMI} = (\text{NIR} - \text{SWIR}) / (\text{NIR} + \text{SWIR}) \qquad (2.41)$$

where NIR – reflection in the near-infrared range of the spectrum; SWIR – reflection in the short-wave infrared range of the spectrum.

The Vegetation Health Index (VHI) is an indicator used to assess general health conditions of vegetation cover. It is estimated combining the Vegetation Condition Index (VCI) and the Temperature Condition Index (TCI). VHI is a useful tool for monitoring droughts, crop yields prediction, and general assessment of the overall conditions of agricultural ecosystems [37]. VHI is calculated using the following formula (2.42):

$$VHI = \alpha * VCI + (1 - \alpha) * TCI \quad (2.42)$$

where VCI reflects the relative state of vegetation compared to historical data for the same period of the year. It is calculated based on the normalized difference vegetation index (VCI = 100 * (NDVI - NDVI_min) / (NDVI_max - NDVI_min)); TCI reflects the thermal state of the earth's surface and temperature-related plant stress. It is calculated based on land surface temperature data (TCI = 100 * (LST_max - LST) / (LST_max - LST_min)); α (alpha) is a coefficient that determines the contribution of VCI and TCI to VHI. α is typically set to 0.5, which means equal weighting of both indices, but can be changed depending on the specific application and environmental conditions.

The Agricultural Stress Index (ASI) is an indicator developed by the Food and Agriculture Organization of the United Nations (FAO) to quickly identify agricultural land with a high probability of water stress (drought). It uses satellite data to monitor vegetation and land surface temperature, integrating them to assess the impact of drought on agriculture. ASI allows for early warning of potential crop losses and threats to food security. ASI is the result of combining VHI temporal integration with crop coefficients and spatial assessment of the percentage of affected arable land with low VHI [38].

4 Software development technological stack

It is worth dwelling on the technology stack used to perform software work and develop online applications and databases.

The interactive electronic map was built based on remote sensing data obtained through JavaScript queries from Google Earth Engine, as well as crop yield data obtained from the Statistical Yearbooks of Ukraine.

To create the map, a database file in .csv format was created and integrated into the web application. The web application was developed using HTML5, CSS3, and JavaScript technological stack. The database on the soils and climate of Ukraine is built on a similar technology stack. The Agroland Classifier program is based on a pure HTML engine with minimal use of styles and scripts.

To create databases for agroecological zoning and features of Ukraine’s soil and climate conditions, Microsoft Access was used alongside web development technologies (HTML, CSS, JavaScript).

5 Experimental data collection

It is worth dwelling separately on the methodology for obtaining field experimental data, which was used in the research work.

For local models, NDVI values were retrieved from the OneSoil cloud platform, which calculates the vegetation index based on combined Sentinel-2 and Sentinel-1 images. Each pixel of the image corresponds to an area of 5 m × 5 m.

5.1 Common beans

A field experiment with common beans was conducted in 2016 on irrigated land belonging to the Agricultural Cooperative RadZemlia (village of Parishovo, Kherson region, 46.706631 N, 32.274669 E). The study was aimed at improving crop cultivation technology and included the study of factors such as soil cultivation depth (2 gradations), fertilizer application (3 gradations), and row spacing (4 gradations). The study was conducted in four replications using a partially randomized split-plot design. Thus, 96 experimental plots were laid out on the experimental field. The recorded area of the experimental plot was 150 m². The harvest from each plot was collected using a self-propelled grain harvester and weighed. The gross yield was then converted to a standard grain moisture content of 14% and 100% purity. The average NDVI values were used to represent each variant of the field experiment. The maximum NDVI values during the vegetation period of common beans were used to create prediction models. The maximum NDVI values were recorded in the "flowering – pod formation" phase or 50-55 days prior to harvest.

5.2 Grain corn, grain sorghum, winter wheat and soybeans

The yield of grain crops, including corn, sorghum, winter wheat and soybeans, for the period 2017-2018 (for winter wheat – 2017-2019) was recorded in the field trials of the Institute of Climate-Smart Agriculture of NAAS. The experimental fields were located on the Ingulets irrigated array within a square with coordinates: 46°44'36.5"N 32°42'07.0"E; 46°44'39.5"N 32°42'32.0"E; 46°44'33.3"N 32°42'33.7"E; 46°44'30.3"N 32°42'08.5"E.

Field experiments were carried out in an irrigated 4-field grain-row crop rotation. The soil of the plots was represented by dark-chestnut middle-loamy soil with humus horizon depth 0.40 m, and humus content in the arable layer of 2.15-2.30%. The equilibrium bulk density of the soil is 1.39-1.42 g/cm^3. Five basic tillage systems in the crop rotation (factor A) and three fertilization systems (factor B) were studied according to the following scheme.

Factor A (basic tillage): variable-depth moldboard plowing at 14-16 cm for winter wheat and 25-27 cm for soybeans; variable-depth plowless tillage at 14-16 cm for winter wheat and 25-27 cm for soybeans; shallow disk tillage at 12-14 cm for all crops of crop rotation;

differentiated tillage with one deep loosening at 38-40 cm; differentiated tillage with one shallow plowing at 18-20 cm.

Factor B (fertilization system): organic system with the use of plant residues; organo-mineral system with the application of $N_{82.5}P_{60}$ per 1 ha of the crop rotation area + plant residues; organo-mineral system with the application of $N_{120}P_{60}$ per 1 ha of the crop rotation area + plant residues.

The experiments were carried out in four replications using split plot design, the accounted area of the plots of the lowest grade was 50 m^2.

All elements of agrotechnology of the studied crops in the experiments were generally accepted for irrigated conditions of Southern Ukraine, except for the factors set for study. Soil moisture in the 0-50 cm layer was maintained at 70-75% of the field capacity by overhead sprinkling irrigation. Phenological observations were carried out according to the BBCH scale [39]. Yield was determined by direct harvesting of the plots upon reaching full grain ripeness. The obtained yield was converted to standard moisture: 14% for winter wheat and grain corn; 13.5% for grain sorghum; 12% for soybeans.

5.3 Sweet corn

Sweet corn yields and the time of phenological phases onset of the crop for the period of 2016 were recorded according to the results of field research in the fields of the private farm RadZemlia, where the influence of the depth of tillage (factor A), doses of mineral fertilizers (factor B) and plant density (factor C) on the crop productivity was studied. Field experiments were carried out on irrigated lands of the Ingulets system. Experimental scheme: factor A – moldboard plowing at 20-22 and 28-30 cm; factor B – doses of mineral fertilizers N_0P_0, $N_{60}P_{60}$, $N_{120}P_{120}$; factor C – plant density 35, 50, 65, 80 thousand per ha. The soil of the experimental plots was represented by dark-chestnut slightly saline soil with an average bulk density of 1.35 g/cm^3, humus profile of 0.45-0.55 m, humus content of 2.50% in the surface active layer. Cultivation technology in the trials was generally accepted for sweet corn grown on the irrigated lands of the South of Ukraine, except for the factors under study. Drip irrigation was used to keep up soil moisture in the root zone at 80% of the field capacity. The experiment was conducted in four replications using the split-plot design method, the area of the experimental plot was 30 m^2. Field experiments were accompanied by phenological observations using generally accepted methods and the BBCH scale. Harvesting and weighing the harvest of marketable corn cobs were carried out manually upon reaching technical ripeness.

The yield of sweet corn and the dates of phenological phases of the crop in the period 2019-2020 were recorded in the field research in the Kulyk State Enterprise, where crop hybrids were tested and the effect of the plant density at different row spacings and growth regulators on its productivity was studied. Field experiments were carried out on the irrigated dark-chestnut soil (Chornobaivka village, Kherson region, Bilozersky district). Experimental scheme was as follows: factor A – sowing scheme – 45×30 cm, 45×20, 70×20, 70×10 cm; factor B – hybrid composition (Deyneris, NBM-2020, Champion); factor C – treatment with the Regoplant growth regulator (50 mL/ha in the phase of 7-8 leaves of the crop). Irrigation – overhead sprinkling to maintain soil moisture at 80% of the field capacity in the 0-50 cm layer during the crop growth period. Mineral fertilizers at the dose of $N_{100}P_{100}K_{100}S_{40}$ were applied directly to the rows in the 3-5 leaf phase of the crop. Otherwise, the agricultural technology of growing sweet corn corresponded to that generally accepted in the south of Ukraine. The studies were accompanied by the necessary phenological observations. Harvesting and its weight accounting were carried out in the phase of technical (milky wax) ripeness of the kernels manually with subsequent weighing. The studies were carried out in four replications using the split-plot design, the area of the accounting plot was 20 m^2.

The experimental plots were linked to the corresponding NDVI values by geotagging on the OneSoil platform, which allows obtaining the average NDVI value from images from the Sentinel-1 and Sentinel-2 satellites (resolution 10-250 m). The OneSoil AI cloud platform uses a special algorithm for processing combined satellite images from Sentinel-2 (for plant recognition) and Sentinel-1 (for correcting distortions and reducing the error in the vegetation index estimation). NDVI data is received on the regular basis at 8-16-day intervals (occasionally the intervals are changing depending on meteorological conditions and satellite images availability).

5.4 Regional and national-scale studies

The images of the annual VHI for croplands of the 1st season were obtained for calculations and analysis from the Global Information and Early Warning System for Food and Agriculture (GIEWS) atlas map, presented through the ArcGIS Online platform. The croplands of the 1st season represent the growing season of most crops grown in Ukraine, as defined by the FAO as the period from March to August. The season 2 is not representative for active agricultural crops growth in Ukraine, as the period September – February is a cold period of a year.

The images were analyzed in the Pixel Color Counter program to obtain quantitative characteristics of the representation of each VHI class. The final score was calculated using the following formula (2.43):

$$VHI = a*na + b*nb + c*nc + d*nd + e*ne + f*nf + g*ng + h*nh + i*ni, \qquad (2.43)$$

where *a, b, c, d, e, f, g, h, i* – corresponding VHI classes, namely, 0,075, 0,20, 0,30, 0,40, 0,50, 0,60, 0,70, 0,80, 0,85; $n_{a\text{-}i}$ – corresponding share of the pixels in the screen that correspond to the certain class.

The ASI and WMVHI values were obtained from data provided by FAO. The relationship between VHI, ASI and WMVHI is presented in Figure 2.2 for the year 2000.

Figure 2.2 Relationship between the studied drought indicators according to remote sensing data of Kherson region in 2000 (according to ArcGIS Online platform)

Monthly regional NDVI values for the arable lands of Kherson region were computed in QGIS software using the raster analysis toolkit from raw satellite images (MODIS Terra, resolution 250 m, smoothed time series), retrieved from the satellite monitoring service of the University of Natural Resources and Life Sciences (Vienna, Austria); partially NDVI images were analyzed using the Global Agricultural Monitoring System GIMMS system.

Regarding the study of the relationship between the yields of the major agricultural crops, cultivated in Ukraine, depending on NDVI, NDMI, VHI, LST and PET data. The study was conducted for arable lands of Ukraine during 2015-2023. The yields of 12 crops, such as spring wheat, winter wheat, spring barley, winter barley, rapeseed, winter rye, oats, peas, sunflower, grain corn, soybeans and sugar beet, were retrieved from the Statistical Yearbooks of Ukraine, expressed in t/ha and summarized. Remote sensing data were obtained using Google Earth Engine capacities through JavaScript requests. The data were generalized monthly for each year of the study in the context of the country regions. The average annual data on NDVI, NDMI, VHI, LST and PET values were subjected to correlation analysis with the construction of corresponding correlation matrix. The .json files, which stored data on the indicators, were converted to .csv format and processed in Python 3.13 to build mathematical models for the yield prediction by the algorithms of linear regression [40], Random Forest regression [41] and Gradient Boosting regression [42]. If applicable, mathematical equations were presented.

The average annual values of the indices were used to build the models. In case of gaps in data for individual years, these years were omitted for the corresponding region of Ukraine. The general methodological workflow of the study is presented in Figure 2.3.

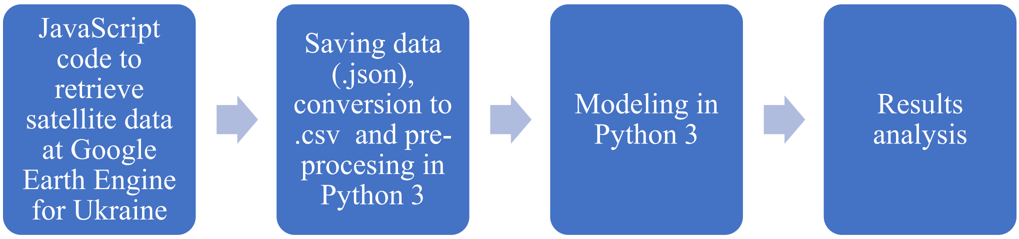


Fig. 2.3 Methodological process of research on the use of machine learning and remote sensing data for prediction of the yields of major agricultural crops in Ukraine

The study of the relationship between the NDVI value and the yield of potatoes, vegetables, fruits and berry crops in the Kherson region for 2005-2021 was performed using crop productivity data from the Statistical Yearbooks of Ukraine and the average monthly value of the vegetation index, estimated from the MODIS Terra time series imagery (resolution 250 m). The images were retrieved from the University of Natural Resources and Natural Sciences (Vienna, Austria). Regression analysis was performed using the ordinary least squares procedure, and the construction of an artificial neural network with the backpropagation of error algorithm was performed in Tiberius software. The artificial neural network had 5 layers of neurons, the learning rate was set to 0.80, and the training lasted for 1000 epochs.

Deep Science Publishing, 2025
https://doi.org/10.70593/978-93-7185-854-0

Chapter 3: Agroecological zoning of Ukraine using remote sensing data

The use of remote sensing (RS) data for agroecological zoning of agricultural crops is a very promising scientific and practical direction. Interest in the use of aerospace imaging for monitoring agricultural lands is growing. This is due to climate change, which affects the growing conditions and productivity of crops, threatening food security. Satellite monitoring data allows you to quickly and accurately assess the state of agrophytocenoses in large areas. This helps to determine how well the environment meets the needs of specific crops. Due to its accuracy, efficiency and the possibility of dynamic tracking, aerospace monitoring is one of the most promising tools for classifying land and increasing the efficiency of crop production. The goal of our scientific work was to classify the territory of Ukraine based on remote sensing-derived data to determine the optimal zones for the cultivation of strategic late spring crops (corn for grain, soybeans and sunflower) in non-irrigated conditions.

1 Agroecological zoning using NDVI

The study was performed using NDVI time series data obtained from pre-processed images provided by AgroMonitoring cloud platform. NDVI is calculated based on composite cloud-free images of Landsat-8 and Sentinel-2 satellites. Monthly average index values were estimated for the period of active vegetation "May–November". The study was performed in 2018 on fixed polygons representing the studied crops: grain corn, soybeans, sunflower. Ten randomly selected and fixed non-irrigated polygons were selected for each region of Ukraine to represent one of the studied crops. The full dataset consisted of 750 polygons; the data are characterized by a normal distribution.

According to the average values of seasonal NDVI, generalized for each geographical region, the most optimal zones of Ukraine for agricultural crops cultivation in non-irrigated conditions were established. It should be noted that the general geographical zoning of Ukraine was used as the basis for current agro-ecological zoning. In this case, 5 main zones are distinguished: South (Kherson, Mykolaiv, Odesa, Zaporizhia, Dnipropetrovsk regions and Crimea); Center (Kirovohrad, Poltava, Cherkasy, Vinnytsia regions); East (Luhansk, Donetsk, Kharkiv regions); North (Sumy, Chernihiv, Kyiv, Zhytomyr regions); West (Transcarpathian, Lviv, Volyn, Rivne, Ternopil, Ivano-Frankivsk, Khmelnytskyi, Chernivtsi regions). It should be emphasized that the conditions of Crimea were assessed only for the steppe zone part of the region, while the mountainous and coastal zones were excluded from the assessment, because they have their specific soil-climate conditions and are not representative of agricultural lands.

The NDVI dynamics during the growing season of the studied crops is presented in Fig. 3.1-3.3. Analysis of the experimental data pointed out that the peak NDVI values are reached in August for sunflower regardless of the cultivation zone (average NDVI = 0.60); in August for grain corn (average NDVI = 0.56), but with a slight dependence on the cultivation zone (in the west and north the peak is reached in September); in September for soybeans (average NDVI = 0.53), but with significant fluctuations across the country territory (peak values are reached in July in the east, in August in the south and center, in September in the west and north). The most uniform seasonal dynamics of NDVI is recorded for sunflower, while soybeans are characterized by the greatest variation across the cultivation zones. As for the highest average seasonal NDVI value for each crop, it was 0.33-0.34 for sunflower cultivated in the north and west; 0.35-0.36 for grain corn and soybeans, which were cultivated in the north and west of Ukraine. The lowest NDVI values were recorded in the east of the country for sunflower (0.26) and grain corn (0.25), but the lowest NDVI value for soybeans (0.27) was fixed in the south. This is mainly because of the great deficit of natural humidification in these regions. Therefore, it is impractical to cultivate these crops in the south and east of Ukraine in the non-irrigated conditions; the most favorable for their cultivation are the western and northern regions.

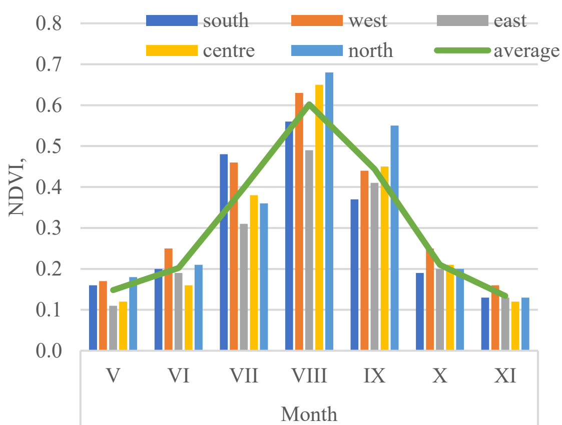


Fig. 3.1. Characteristics of NDVI dynamics during the growing season of sunflower depending on the cultivation zone

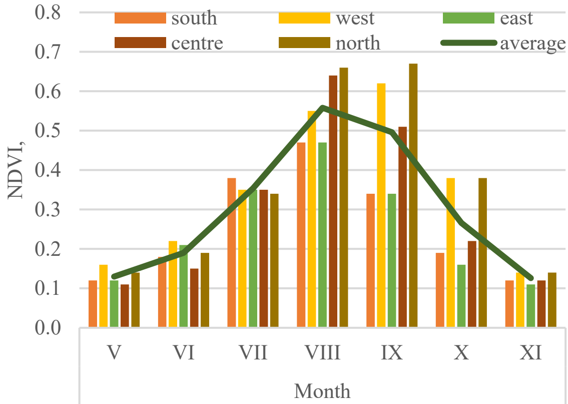


Fig. 3.2. Characteristics of NDVI dynamics during the growing season of grain corn depending on the cultivation zone

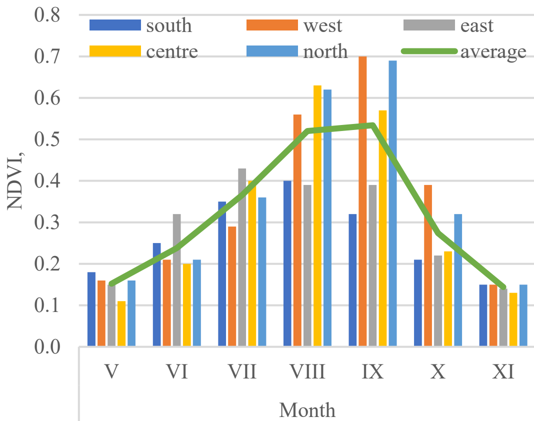


Fig. 3.3. Characteristics of NDVI dynamics during the growing season of soybean depending on the cultivation zone

The methodological scheme for the crops agroecological zoning by NDVI values is shown on Fig. 3.4. The proposed scheme is universal.

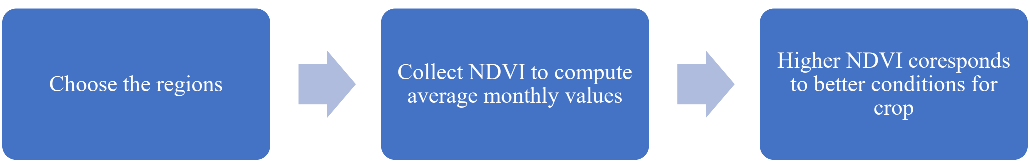


Fig. 3.4 Methodological scheme of crops agroecological zoning based on satellite NDVI data

However, its disadvantage is a certain one-sidedness, since several important factors, such as soil moisture regimes, nutrient regime, are estimated indirectly and generalized through the NDVI index, while a combined assessment considering specific satellite indices can significantly improve the accuracy and reliability of agroecological zoning. More details on current study one can learn from this scientific paper [43].

2 Comprehensive agroecological zoning using integrative AEZI values

As it was mentioned in the previous section, it is advisable to use integrative approach to perform more robust and reliable agroecological zoning, including such important parameters as moisture and key nutrients availability. In this regard, complex Agroecological Zoning Index (AEZI) is proposed as a complex indicator to perform remote sensing-based mapping of the agricultural lands' suitability for certain crops cultivation.

In the proposed approach, the assessment of moisture availability is carried out using the NDWI (Normalized Difference Water Index) value. This index can indirectly indicate the moisture content in the biomass of plants. There are developments in assessing the intensity of drought on agricultural lands using the NDWI value.

To assess nutrient availability, the nitrogen reflectance index NRI (Nitrogen Reflectance Index) is used, the value of which can indirectly determine the degree of nitrogen storage for plants, which is the basis for the formation of biomass and yield of most crops. The scientific and practical application of this index is limited. However, its value has been proven in assessing the potential productivity of crops, particularly those that consume a lot of nitrogen for crop formation, where it is not inferior in accuracy of yield modeling to the more widely used NDVI index.

To perform agroecological zoning of territories, a graduated scale of each of the indices is necessary, which would reflect their relationship to real environmental conditions and the state of crops. The scales of interpretation of NDWI, NDVI and NRI used during the development of the complex index are given in Tables 3.1, 3.2 and 3.3.

Table 3.1 Interpretation of the NDWI value from the point of view of water supply

NDWI values	Water supply
0–0.30	Extreme drought
0.31–0.40	Drought
0.41–0.50	Moderate drought, low water supply
0.51–0.60	Moderately low water supply
0.61–0.70	Moderate water supply
0.71–0.80	High water supply
Понад 0.81	Flooding or extremely high overwatering

Table 3.2 Interpretation of NDVI values from the perspective of general plants condition

NDVI values	Vegetation cover conditions
0–0.20	Bare soil
0.21–0.30	Sparse vegetation cover, weak plants, initial growth
0.31–0.40	Fragile green biomass
0.41–0.50	Moderate green biomass
0.51–0.60	Well-developed green biomass
0.61–0.70	Highly developed green biomass
0.71–0.80	Dense and well-developed vegetation cover
Понад 0.80	Very dense and developed vegetation cover

Table 3.3 Interpretation of NRI values form the perspective of nitrogen availability for plants

NRI values	Nitrogen supply
0–0.30	Low
0.31–0.50	Moderate
0.51–0.70	High
> 0.71	Very high

The AEZI is estimated using the following formula (3.1):

$$AEZI = \frac{NDWI + NDVI + NRI}{3} \times 100\% \qquad (3.1)$$

where NDWI, NDVI and NRI are the absolute averages for the growing season of the values of the corresponding satellite indices, calculated using a smoothed time series.

The interpretation of the complex agro-ecological index AEZI for zoning agricultural lands (both universal and specific) is given in Table 3.4. In the future, it is possible to adjust the main classes of agricultural land suitability according to the value of this complex index.

Table 3.4 Interpretation of AEZI values for zoning of agricultural land

AEZI values	Suitability for crops cultivation
0–20%	Unsuitable
21–30%	Marginal
31–40%	Conditionally suitable
41–60%	Suitable
61–75%	Optimal
> 75%	Ideal

Practical application of AEZI is possible for both general (universal) and specific zoning of agricultural lands according to their compliance with the biological requirements of a specific crop. At the same time, when performing universal zoning, satellite index values are used for the growing season as a whole, and for specific zoning – for the growing season of a certain crop. It is important to use satellite images of one satellite (recommended – Landsat-8, or combined images from Sentinel-1/2), with minimal (up to 10%) cloudiness and without distortions. As an example of practical application of AEZI, a universal zoning of the conditions that developed in the experimental fields of the Institute of Climate-Smart Agriculture of NAAS, located in Kherson region in 2023 during the growing season (March – October), are presented. Data on the value of satellite indices were calculated from Landsat-8 and Sentinel-2 images adapted from the AgroMonitoring service (Table 3.5).

Table 3.5 Characteristics of agroecological conditions in the experimental fields of the Institute of Climate-Smart Agriculture of NAAS during the growing season of 2023 according to aerospace monitoring data

Month	NDVI	NDWI	NRI	AEZI	Conclusion
March	0.19	0	0.19	12.7%	Unsuitable
April	0.08	0.15	0.07	10.0%	Unsuitable
May	0.48	0.23	0.31	34.0%	Conditionally suitable
June	0.41	0.15	0.40	32.0%	Conditionally suitable
July	0.37	0.15	0.40	30.7%	Conditionally suitable
August	0.43	0.04	0.42	29.7%	Marginal
September	0.37	0.04	0.40	27.0%	Marginal
October	0.31	0.10	0.41	27.3%	Marginal
Average	0.33	0.11	0.33	25.7%	Marginal

According to the results of AEZI analysis, the experimental fields of the Institute of Climate-Smart Agriculture of NAAS, which in 2023 were almost uncultivated for several years due to military activities, are of low suitability or even unsuitable for crops cultivation. This confirms the objectivity of the assessment of agricultural lands conditions according to the AEZI index. The decrease in the suitability of the Institute's lands was mainly contributed by the lack of irrigation, since the NDWI value was the minimum of all the studied indices, while nitrogen reserves in the surface soil layer were at an average level for most of the growing season. This can be confirmed by the analysis of the conditions in 2021 for the same test site of the Institute of Climate-Smart Agriculture of NAAS, when all the necessary agrotechnical and land reclamation measures were properly carried out (Table 3.6). It was established that the main part of the active vegetation of cultivated plants (April – August) is classified as suitable for crop production.

An agroecological assessment of the Steppe zone of Ukraine (Figure 3.5) was also carried out for the suitability for growing major agricultural crops such as grain corn, sunflower, soybeans, wheat, barley, rapeseed, alfalfa, sugar beets and rice.

Table 3.6 Characteristics of agroecological conditions in the experimental fields of the Institute of Climate-Smart Agriculture of NAAS during the growing season of 2021 according to aerospace monitoring data

Month	NDVI	NDWI	NRI	AEZI	Conclusion
March	0.28	0.12	0.39	26.3%	Marginal
April	0.49	0.20	0.45	38.0%	Conditionally suitable
May	0.59	0.40	0.34	44.3%	Suitable
June	0.55	0.24	0.37	38.7%	Conditionally suitable
July	0.42	0.22	0.38	34.0%	Conditionally suitable
August	0.33	0.06	0.38	25.7%	Marginal
September	0.27	0.09	0.36	24.0%	Marginal
October	0.25	0.13	0.37	25.0%	Marginal
Average	0.40	0.18	0.38	32.0%	Conditionally suitable

The AEZI was estimated using combined images from Sentinel-1/2 and Landsat-8 satellites (resolution was 250 m, cloud-free images only retrieved from AgroMonitoring cloud platform). Spatial indices were estimated for six randomly chosen and fixed fields representing the studied crops, viz., grain corn, sunflower, soybeans, rapeseed, wheat, barley, alfalfa, sugar beet and rice. In total, 594 fields were analyzed in 7 regions of the steppe zone (Crimea, Kherson, Mykolaiv, Zaporizhzhia, Dnipropetrovsk, Kirovohrad, Odesa regions). Sugar beet is not cultivated in Crimea, and rice is grown only in Kherson and Odesa regions. The dataset embraced 2018-2019 season. Based on the results of the AEZI calculations, the database "Agroecological zoning of the steppe zone of Ukraine v.1.00" was created, and agroecological mapping was performed.

The results of the AEZI calculation for the steppe zone of Ukraine pointed out to unfavorable conditions for late spring crops, while early spring and winter cereal crops, as well as rapeseed, could be grown in the zone even under non-irrigated conditions provided rational agricultural techniques. The best correspondence of ecological conditions in the non-irrigated steppe zone to crops biological demands was found for rapeseed and wheat, and the least suitability was fixed for grain corn (Table 3.7).

To visualize the results, agroecological mapping of agricultural lands was performed (Figure 3.5). The coastal areas of the steppe zone are the most optimal for rapeseed cultivation without irrigation.

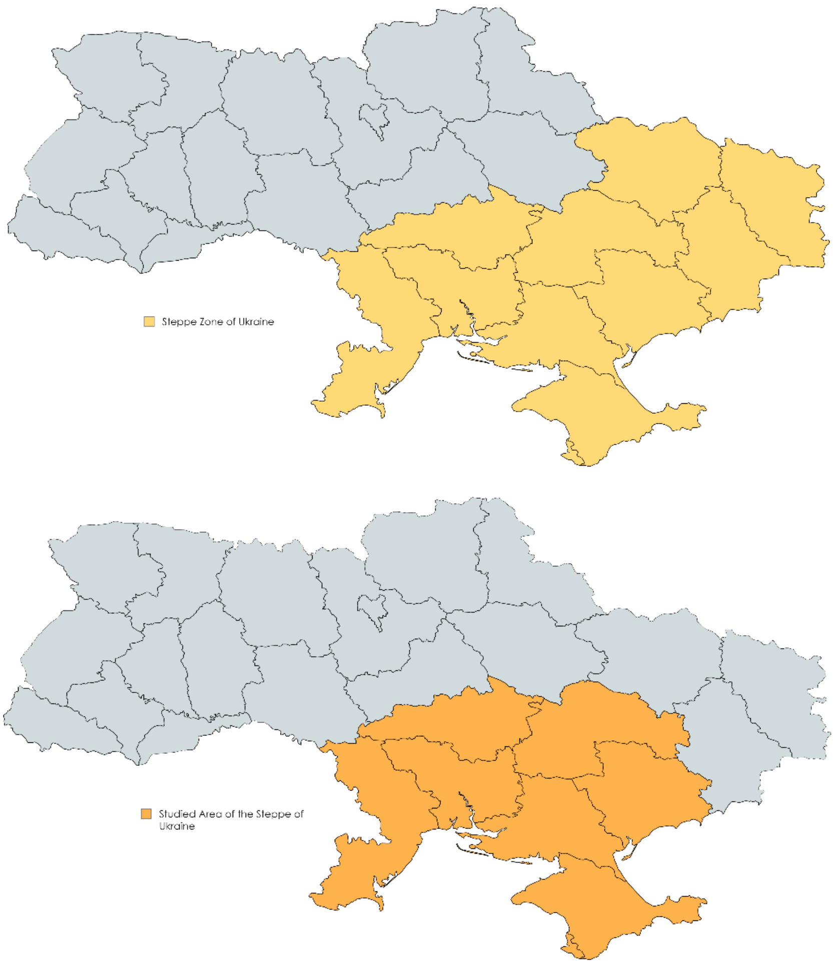


Fig. 3.5. Steppe zone of Ukraine (top – as a whole, bottom – the territory for which agroecological zoning has been carried out)

Table 3.7 Characteristics of agroecological conditions in the non-irrigated Steppe of Ukraine during the growing season 2018-2019 according to aerospace monitoring data

Region	**AEZI**					
	corn	**soybeans**	**sunflower**	**wheat**	**barley**	**rapeseed**
Crimea	21.0%	22.8%	25.1%	36.3%	35.5%	39.3%
Kherson	21.5%	22.0%	26.0%	40.1%	38.9%	46.2%
Mykolaiv	24.3%	23.0%	26.6%	39.7%	36.5%	42.1%
Odesa	24.5%	24.3%	30.0%	37.2%	38.3%	41.6%
Zaporizhzhia	22.0%	21.0%	27.4%	34.1%	34.5%	44.1%
Dnipro	25.1%	25.1%	27.4%	36.6%	37.7%	40.2%
Kirovohrad	24.3%	27.4%	25.0%	38.5%	38.5%	39.6%
Average	23.2%	23.7%	26.8%	37.5%	37.1%	41.9%

corn, soybeans, sunflower | wheat, barley | rapeseed

Fig. 3.6 Mapping the suitability of non-irrigated agricultural lands of the Steppe zone of Ukraine for major agricultural crops cultivation (color scheme: yellow – unsuitable; light green – conditionally suitable; green – suitable)

As for irrigated croplands, the situation is changing dramatically, mainly owing to better moisture availability for crops. There are no unfavorable regions for any of the studied crops under irrigated conditions. The irrigated steppe zone is suitable for wheat, barley and rapeseed cultivation, while it is less favorable for sunflower and grain corn (Table 3.8). As for rice, this crop is grown only in Kherson and Odesa regions, and the AEZI for the studied period was 47.3 and 50.5%, respectively (suitable lands).

Table 3.8 Characteristics of agroecological conditions in the irrigated Steppe zone of Ukraine during the growing season 2018-2019 according to aerospace monitoring data

Region	AEZI							
	corn	soybeans	sunflower	wheat	barley	rapeseed	alfalfa	sugar beets
Crimea	36.7%	48.1%	32.5%	50.6%	47.6%	51.6%	43.6%	N/A
Kherson	41.3%	41.0%	41.0%	51.0%	43.9%	50.8%	45.3%	44.4%
Mykolaiv	33.2%	39.0%	41.9%	41.7%	44.5%	47.2%	41.0%	48.3%
Odesa	40.0%	37.1%	36.1%	52.0%	55.2%	47.8%	46.6%	41.5%
Zaporizhzhia	39.0%	40.8%	41.3%	50.7%	56.1%	47.7%	45.8%	41.4%
Dnipro	46.8%	42.9%	40.1%	47.7%	44.2%	44.4%	41.0%	44.3%
Kirovohrad	50.7%	46.8%	42.3%	47.4%	46.3%	45.9%	43.3%	44.0%
Average	41.1%	42.2%	39.3%	48.7%	48.3%	47.9%	43.8%	44.0%

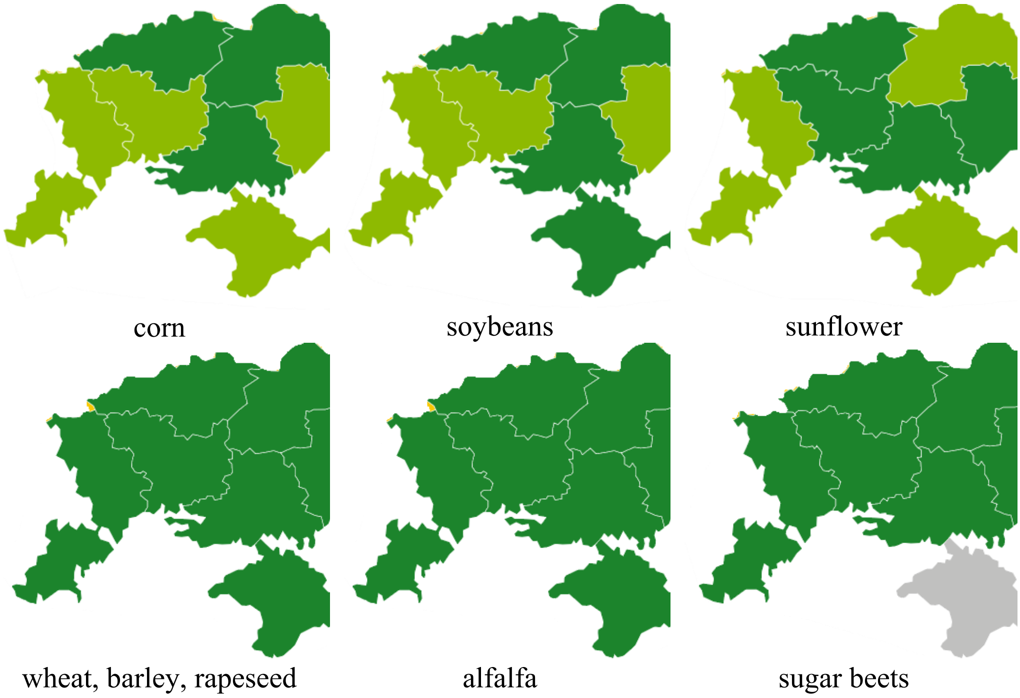


Fig. 3.7 Mapping the suitability of irrigated agricultural lands of the Steppe Zone of Ukraine for major agricultural crops cultivation (color scheme: light green – conditionally suitable; green – suitable; gray – data for the region are missing)

To visualize the results presented in Table 3.8, an agroecological mapping of the irrigated steppe zone was performed, which is demonstrated in Figure 3.7. It is noticeable that the best conditions for growing the studied crops were in the Kherson and Kirovohrad regions.

Additionally, the key features of the AEZI computation methodology and related agroecological zoning studies could be learned in detail by referring to following scientific papers [44-45].

Deep Science Publishing, 2025
https://doi.org/10.70593/978-93-7185-854-0

Chapter 4: Classification of agricultural lands using remote sensing data

NDVI is an effective tool for crop mapping and monitoring, as it allows for a quantitative assessment of the vegetation cover conditions. Studies have shown that each crop has its specific NDVI patterns, which allows for the identification of crop types and monitoring their development, including determining the optimal time for harvesting. Combining NDVI with other data and mathematical models significantly expands its capabilities.

The main goal of the study is to develop a methodology for recognizing and mapping crops using remote sensing NDVI data.

The study was carried out using a time series of NDVI calculated on the basis of composite cloud-free and atmospherically undistorted images from Landsat-8 and Sentinel-2 satellites with a resolution of 10-250 m. Monthly average index values were calculated for the period from May to November for grain corn, sunflower and soybeans. The study was conducted within 2018-2019 on the fixed fields, each field representing one of the studied crop. Ten non-irrigated polygons for each region of Ukraine were selected to represent every crop. In total, ten polygons per crop per region of Ukraine, therefore, thirty polygons of the studied crops were subjected to analysis in each region and 250 polygons per crop. The full dataset consists of 750 polygons; the data are characterized by a normal distribution.

1 Discriminant analysis for crops recognition

Based on the annual intra-seasonal patterns of NDVI values, regularities of the index change by growth and development phases of the studied crops were established and its suitability for their identification and potential application in crop mapping systems was determined. Mathematical data processing and construction of a classification model were performed using the multiclass linear discriminant analysis (MLDA) and canonical discriminant analysis (CDA) procedures. Calculations were performed in the statistical package BioStat v.7. The calculated canonical coefficients and constants were used to construct canonical discriminant functions for each crop. The functions were additionally analyzed for their accuracy in identifying the studied crops. All statistical calculations were performed for P <0.05.

The discriminant analysis of a generalized sample of 750 plots (Table 4.1) revealed the main patterns of NDVI changes during the growing season of the studied crops and the possibility of its use in the systems of automated crop identification and crop mapping systems.

Table 4.1 Probability of a priori classification for each group of studied spring row crops depending on their individual NDVI structures during the growing season

Group	N	p
Sunflower	250	0.33
Grain corn	250	0.33
Soybeans	250	0.33
Total	750	1.00

Calculation of Wilks' lambda (λ) using Bartlett and Rao model approximation algorithms allowed us to reject the null hypothesis of the absence of a statistically significant relationship between the seasonal dynamics of NDVI and the characteristics of the growth processes of the studied crops (Tables 4.2 and 4.3).

Table 4.2 Wilks' Lambda (Bartlett's) for Discriminant Analysis

Canonical function	λ	χ^2	df	p-value
Canonical function 1-2	0.7138	250.860	14	0
Canonical function 2	0.9285	55.212	6	4.2011×10^{-10}

Table 4.3 Wilks' Lambda (after Rao) for discriminant analysis

Statistical	Value
λ	0.7138
F-criterion	19.4390
p-value	0
F critical	1.6984
Null hypothesis	rejected

The analysis of canonical coefficients (Table 4.4) certified that the main role in the crops identification is played by the August NDVI (canonical function 1) and September NDVI (canonical function 2) values. Thus, the final phenophases of their development, namely the pre-harvest period, are of great importance for the recognition of late spring crops. The analysis of canonical variables shown in Fig. 4.1 indicates that canonical function 1 is much more decisive and reliable than canonical function 2. Thus, August NDVI values are the main predictor for the automated recognition of late spring crops.

Table 4.4 Canonical coefficients and the full canonical structure of the discriminant function

Variables	Canonical coefficients		Standardized canonical coefficients		Full canonical structure	
	Canonical function 1	Canonical function 2	Canonical function 1	Canonical function 2	Canonical function 1	Canonical function 2
May	-4.2620	4.5721	-0.2239	0.2402	-0.0436	0.5798
June	1.1546	4.9345	0.0934	0.3991	0.0250	0.6755
July	-1.0560	1.8225	-0.1197	0.2067	-0.5038	0.2071
August	-6.5555	-1.9442	-0.9070	-0.2690	-0.4935	-0.1969
September	4.0306	4.6711	0.7266	0.8421	0.5080	0.1872
October	2.7626	-6.1681	0.3289	-0.7343	0.5321	0.0110
November	-7.9549	12.4817	-0.3823	0.5999	-0.1092	0.6068

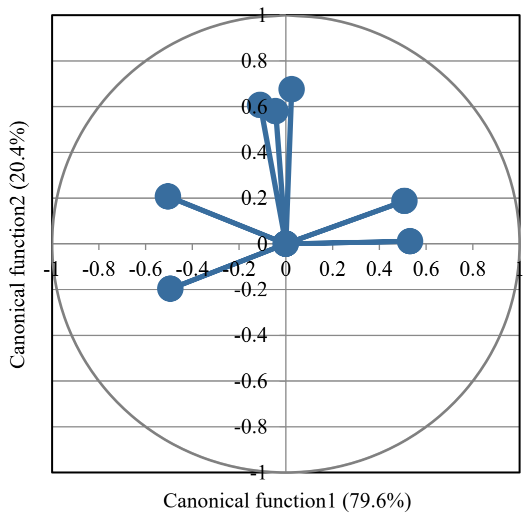


Fig. 4.1 Plot of canonical independent variables used to identify the study crops based on their individual NDVI patterns

Table 4.5 shows the classification matrix based on the results of MLDA and CDA. This matrix is used to evaluate prediction accuracy in classification analysis, using the predicted inputs as a rule to distinguish between classes.

Table 4.5 Classification matrix and recognition accuracy of corn, soybean and sunflower crops by canonical function

Group/Prediction	Grain corn	Soybeans	Sunflower	Total	Correct guesses
Grain corn	117	68	65	250	46.8%
Soybeans	66	131	53	250	52.4%
Sunflower	43	31	176	250	70.4%
Total	226	230	294	750	56.5%

The maximum classification accuracy was recorded for sunflower (70.4%), while for soybeans (52.4%) and grain corn (46.8%) it is low. The overall prediction precision for the studied crops was 56.5%. It was found that grain corn is frequently confused with soybeans (in 26.4-27.2% of cases). Sunflower has a unique seasonal NDVI dynamics, which is very different from both soybeans (incorrect guesses in 12.4% of cases) and grain corn (incorrect guesses in 17.2% of cases). Soybeans and grain corn have similar seasonal dynamics of biomass accumulation, and therefore the NDVI time series. This complicates the recognition of these two crops using pure remote sensing data [46].

In addition to late spring crops, winter cereals (wheat and barley) and rapeseed are strategic for Ukraine. The study of the possibility of using multiclass linear and canonical discriminant analysis (MLDA and CDA) for their recognition and subsequent mapping based on NDVI dynamics data was performed for 2018; the vegetation index was calculated using combined Landsat-8 and Sentinel-2 satellite images for 70 randomly selected and fixed fields of winter wheat, winter barley, and winter rapeseed (210 fields in total) located in the Steppe zone of Ukraine (geographically, Kherson, Mykolaiv, Odesa, Zaporizhzhia, Dnipropetrovsk, and Kirovohrad regions). Data on the NDVI value was generalized into a monthly time series for the period April – July. Data for March were not used, since, according to previous studies, they are unrepresentative of the active vegetation of winter crops. Algorithms for mathematical and statistical processing of experimental data were like in the previous study with late spring crops. Statistical calculations were performed in the BioStat v.7 at a confidence level of 95%. A general outlook on the NDVI in the fields of the studied crops during the period of study is represented by Table 4.6.

Table 4.6 The value of the mean monthly NDVI in winter crops

Crop	Parameter	Month			
		April	May	June	July
rapeseed	Mean±SD	0.42±0.08	0.48±0.10	0.59±0.08	0.20±0.07
	CV, %	19.61	21.10	12.69	34.30
wheat	Mean±SD	0.35±0.08	0.45±0.08	0.57±0.11	0.15±0.05
	CV, %	22.38	18.54	19.21	32.42
barley	Mean±SD	0.38±0.09	0.42±0.10	0.59±0.11	0.15±0.05
	CV, %	23.59	24.98	19.11	33.68

Winter crops have similar NDVI dynamics, which complicates the task of their recognition. On average, the maximum values of the vegetation index are recorded for winter rapeseed crops, and the minimum values for winter wheat.

The eigenvalues of the canonical functions for winter crops recognition are given in Table 4.7. The canonical correlation coefficient for the first function is significantly higher, therefore, this function suits better for identifying winter crops. The graphical model of the distribution of canonical changes by functions 1 and 2 additionally indicates a significantly higher weight of the first canonical function over the second (Fig. 4.2).

Table 4.7 Eigenvalues of canonical functions for recognizing winter crop crops by NDVI value

Canonical function	Value	Share	Cumulative	Canonical R
1	1.5582	0.8290	0.8290	0.78
2	0.3215	0.1710	0.1710	0.49

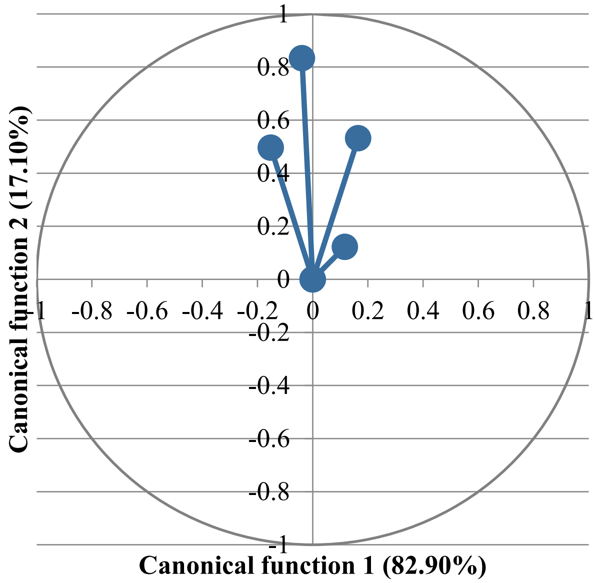


Fig. 4.2 Plot of canonical independent variables used to identify the study crops based on their individual NDVI patterns

The mathematical statistics of Wilks' lambda (according to the Bartlett and Rao approximation), which is a measure of the effectiveness of the developed discriminant model, are given in Tables 4.8 and 4.9. The calculation of the Pillai trace (Table 4.10) allows us to definitively state that the null hypothesis of recognizing winter crop crops according to the time series data of the normalized differential vegetation index is rejected.

Table 4.8 Wilks' Lambda (Bartlett approximation) of canonical functions for recognizing winter crop crops by NDVI value

Canonical function	λ	χ^2	df	P-value
1–2	0.2958	250.31	8	0
2	0.7567	57.28	3	2.235×10^{-12}

Table 4.9 Wilks' Lambda (Rao approximation) of canonical functions for recognizing winter crops by NDVI value

λ	F actual (df1=8, df2=410)	P-value	F critical	Null hypothesis
0.2958	42.77	0	1.96	rejected

Table 4.10 Pillai trace of canonical functions for recognizing winter crops by NDVI value

Pillai trace	F actual (df1=8, df2=410)	P-value	F critical	Null hypothesis
0.8524	38.07	0	1.96	rejected

The combined variance-covariance matrix, which gives an idea of the coupled compatible variability of the classification results from the input parameters of the model, and the full classification matrix are given in Table 4.11. Based on the calculations, classification functions for the studied winter crops were developed, and are given in Table 4.12.

Table 4.11 Combined variation-covariance matrix and full matrix of winter crops classification by NDVI value

		Combined variation-covariance matrix		
Variables	**April**	**May**	**June**	**July**
April	0.0070	0.0079	0.0041	-0.0001
May	0.0079	0.0094	0.0049	-0.0003
June	0.0041	0.0049	0.0102	0.0010
July	-0.0001	-0.0003	0.0010	0.0033
		Full matrix of winter cropd classification		
April	0.0076	0.0083	0.0043	0.0004
May	0.0083	0.0100	0.0049	0.0004
June	0.0043	0.0049	0.0102	0.0012
July	0.0004	0.0004	0.0012	0.0039

Table 4.12 Function for classifying winter crops by NDVI value

Group	April	May	June	July	Constant
Barley	84.1441	-49.1650	44.7771	29.8650	-21.2175
Rapeseed	10.1102	25.2633	36.7560	53.0527	-24.5530
Wheat	-85.8271	101.3789	37.9921	39.4493	-21.3046

To assess the correctness of the crop recognition, a classification matrix was calculated (Table 4.13). Taking into account the results of the statistical evaluation, the best identification accuracy was provided for winter wheat (over 75%). A somewhat lower accuracy was recorded for winter barley (almost 73%), while winter rapeseed was identified the worst – the prediction accuracy reached only 55.7%. In our opinion, this can be explained by the distortion of the NDVI value on winter rapeseed crops during the period of active flowering, when the bright yellow color of the crop's flowers prevents an adequate assessment of the vegetation index value. The results of the study proved the possibility of classifying winter cereals using the time series data on the remote sensing NDVI.

Table 4.13 Matrix for classifying winter crops by NDVI value

Group/Prediction	Barley	Rapeseed	Wheat	Total	Correct guesses
Barley	**51**	19	0	70	72.9%
Rapeseed	24	**39**	7	70	55.7%
Wheat	2	15	**53**	70	75.7%
Total	77	73	60	210	**68.1%**
N of correct = 143					

2 Discriminant analysis for distinguishing irrigated and rainfed croplands

Satellite monitoring is a promising technology for identification, dynamic monitoring, operational management, and control of agricultural land use. The aerospace monitoring data opens new opportunities for remote identification of irrigated lands, which is important for assessing the real scale of irrigation applied, monitoring of irrigated ecosystems, tracing their geographic localization, mapping, as well as optimizing water resources use. Given the high relevance of this subject and the lack of scientific developments devoted to this problem, the goal of research was to develop a methodology for remote sensing based identification of the irrigated and rainfed croplands in the Steppe zone of Ukraine, first for such major crops as winter wheat, grain corn, soybeans, and sunflower, based on mathematical discriminant analysis.

The analysis of the potential use of NDVI time series for distinguishing between irrigated and non-irrigated areas of major agricultural crops cultivated in Southern Ukraine was conducted using combined satellite imagery from Landsat-8 and Sentinel-2, with a spatial resolution of 250 m. The dataset was refined by excluding distorted or low-quality images, as well as those with high cloud cover. For each studied crop – winter wheat, grain corn, soybeans, and sunflower – NDVI time series were generated for randomly selected and subsequently fixed fields, and aggregated into monthly intervals for the vegetation period from May to October.

A total of 50 fields per crop (split to 25 irrigated and 25 non-irrigated, respectively), mainly located in Kherson and Mykolaiv regions, were subjected to analysis during the 2018 growing season. The resulting dataset contained 200 fields and 1,200 NDVI observations. Statistical analysis was performed using multiclass linear discriminant analysis (MLDA) and canonical discriminant analysis (CDA) algorithms implemented in the BioStat v.7 software package, with significance accepted at $P < 0.05$. Based on the results of the mathematical processing, a canonical discriminant function was constructed to classify irrigated and non-irrigated lands, both in general and separately for each of the investigated crops.

As a result of mathematical calculations, the canonical correlation coefficient was estimated as 0.81, which indicates the high quality of the discriminant model. The calculation of such statistical indicators as Wilks' lambda (0.34) and Pillai's trace (0.66) allowed us to confidently reject the null hypothesis. Additional evidence in favor of the model for classifying irrigated and non-irrigated lands is the value of the criterion $\chi^2 = 207.97$ with the number of degrees of freedom 6 (the control value of the criterion is $\chi^2 = 12.60 < \chi^2$), which allows us to completely reject the null hypothesis.

The combined variance-covariance matrix, as well as the full classification matrix, are given in Table 4.14. The canonical coefficients and the canonical structure of the classification function are given in Table 4.15.

Table 4.14 Combined variation-covariance matrix and full classification matrix of irrigated and rainfed lands based on NDVI data

Variable (NDVI)	May	June	July	August	September	October
Combined variation-covariance matrix						
May	0.0248	0.0323	-0.0020	-0.0270	-0.0223	-0.0042
June	0.0323	0.0474	-0.0021	-0.0399	-0.0317	-0.0059
July	-0.0020	-0.0021	0.0103	0.0029	-0.0034	-0.0017
August	-0.0270	-0.0399	0.0029	0.0443	0.0307	0.0052
September	-0.0223	-0.0317	-0.0034	0.0307	0.0385	0.0097
October	-0.0042	-0.0059	-0.0017	0.0052	0.0097	0.0058
Full classification matrix						
May	0.0252	0.0323	-0.0015	-0.0246	-0.0199	-0.0039
June	0.0323	0.0473	-0.0019	-0.0389	-0.0307	-0.0058
July	-0.0015	-0.0019	0.0108	0.0055	-0.0008	-0.0013
August	-0.0246	-0.0389	0.0055	0.0555	0.0422	0.0069
September	-0.0199	-0.0307	-0.0008	0.0422	0.0501	0.0115
October	-0.0039	-0.0058	-0.0013	0.0069	0.0115	0.0061

Table 4.15 Canonical coefficients and canonical structure of the classification function for irrigated and rainfed lands recognition based on NDVI data

Variables (NDVI)	Canonical coefficients		
	Raw	Standardized	Full structure
May	5.6108	0.8843	0.1639
June	4.3191	0.9406	0.0456
July	3.7303	0.3784	0.2913
August	5.1812	1.0900	0.5610
September	5.7194	1.1223	0.5988
October	-2.6224	-0.2004	0.2581

The classification function for irrigated and non-irrigated lands of the Ukrainian Steppe is given in Table 4.16. The classification matrix (resulting) is presented in Table 4.17.

Table 4.16 Function to classify irrigated and rainfed lands based on NDVI data

Group	V	VI	VII	VIII	IX	X	constant
irrigated	28.1418	61.3133	52.9820	54.9718	34.2783	21.5056	-46.7715
rainfed	12.7304	49.4499	42.7358	40.7402	18.5686	28.7086	-25.9281

Table 4.17 Matrix for classifying irrigated and rainfed lands based on NDVI data

Group / Prediction	Irrigated	Rainfed	Total	Correct guesses
irrigated	**91**	9	100	91.0%
rainfed	8	**92**	100	92.0%
total	99	101	200	**91.5%**

N correct = 183

Thus, it was established that according to the monthly NDVI time series data, it is possible to classify irrigated and non-irrigated croplands for the main agricultural crops of the Steppe of Ukraine with very high (91.5%) accuracy. We present the developed classification functions (Table 4.18), as well as the results of the classification matrix of irrigated and rainfed crops for each studied crop individually (Table 4.19).

Table 4.18 Classification functions for recognition of irrigated and rainfed crops based on NDVI data

Group	V	VI	VII	VIII	IX	X	constant
				Winter wheat			
irrigated	-9.7021	169.2668	-58.1099	69.8045	121.5592	219.9100	-71.8640
rainfed	-26.1876	154.4067	-66.0404	68.3065	114.6071	230.0972	-52.5441
				Sunflower			
irrigated	109.0168	23.6479	124.0220	111.3787	111.4712	29.3895	-94.4985
rainfed	91.7896	52.7838	100.7603	89.8808	88.7875	26.2191	-66.6667
				Grain corn			
irrigated	33.7654	95.2327	78.4457	72.7962	36.1976	46.9554	-65.5002
rainfed	63.4961	78.7181	58.5888	50.8028	15.9501	39.4445	-36.8342
				Soybeans			
irrigated	69.8560	74.8651	109.4211	94.6726	53.6681	-25.5538	-79.6537
rainfed	82.9034	47.4025	81.1902	49.3747	25.5037	-4.4578	-34.8158

It was found that the canonical functions provide the maximum (98.0%) classification accuracy for corn and soybeans, while the minimum (86.0%) for winter wheat. At the same time, certain intergroup differences were noted. For example, irrigated and non-irrigated sunflower is identified with the same accuracy; as for winter wheat and grain corn, the classification function provided the highest accuracy in recognizing non-irrigated lands, while for soybeans, on the contrary, for irrigated lands. The highest correlation coefficient for the canonical function was established for soybeans, while the smallest was for winter wheat. These features must be taken into account when applying the developed discriminant classification functions for scientific and practical purposes.

Table 4.19 Classification matrices for irrigated and rainfed croplands based on NDVI data

Group / Prediction	Irrigated	Rainfed	Total	Correct guesses
		Winter wheat		
irrigated	**20**	5	25	80,0%
rainfed	2	**23**	25	92,0%
total	22	28	50	**86,0%**
N correct = 43; R = 0,80				
		Sunflower		
irrigated	**24**	1	25	96,0%
rainfed	1	**24**	25	96,0%
total	25	25	50	**96,0%**
N correct = 48; R = 0,81				
		Grain corn		
irrigated	**24**	1	25	96,0%
rainfed	0	**25**	25	100,0%
total	24	26	50	**98,0%**
N correct = 49; R = 0,88				
		Soybeans		
irrigated	**25**	0	25	100,0%
rainfed	1	**24**	25	96,0%
total	26	24	50	**98,0%**
N correct = 49; R = 0,94				

The results of the study demonstrated the feasibility of using NDVI time series in combination with canonical discriminant functions for high-precision remote classification of irrigated and rainfed croplands in southern Ukraine. These findings highlight new prospects for the expanded integration of aerospace remote sensing data into national agricultural science and practice. Furthermore, the obtained results will serve as a foundation for developing an automated crop identification system [47].

3 Logistic regression for croplands recognition. Agroland Classifier application

The practical application of discriminant analysis and other mathematical classification methods requires users to possess specific analytical skills and methodological understanding. Therefore, it is insufficient merely to design an algorithm or develop a classification methodology; it is equally essential to provide a user-friendly and intuitive tool that enables the effective and accurate application of these analytical techniques.

As the discriminant analysis algorithms have been described in detail earlier, they are not revisited here. With respect to binary logistic regression, this approach was selected for crop-group differentiation because it yielded a notably higher overall classification accuracy – averaging 75.4%. Given the substantial similarity in seasonal NDVI profiles among certain crops, particularly soybeans and grain corn, as well as winter wheat and barley, and the correspondingly low recognition accuracy within these pairs, it was deemed appropriate to merge them into unified classification groups. Consequently, classification was performed for the following pairs: non-irrigated sunflower vs. non-irrigated soybean/grain corn, and winter wheat/barley vs. winter rapeseed.

Binary logistic regression was calculated using the sigmoid activation function algorithm, which has the form (4.1):

$$p(x)=1/(1+e^{(-(b_0+\cdots+b_n))}) \qquad (4.1)$$

where p(x) is a probability of a particular classification group; e is the Euler number; b is a regression coefficient and/or intercept value of the model.

The results of the statistical analysis of the logistic classification of late spring crops showed its reliability according to the Hosmer tests (Hosmer = 694.59; df = 623; P = 0.02<0.05) and the χ^2 criterion (χ^2 = 196.40; df = 5; P = 1.67×10^{-40}<0.05), which allowed us to reject the null hypothesis. The ROC curve, which allows us to assess the quality of binary classification and reflects the ratio between the proportion of truly correctly classified objects to the total number of incorrectly classified objects, is presented in Fig. 4.3. As for the classification of winter crops, the null hypothesis was rejected only according to the χ^2 criterion (χ^2 = 52.53; df = 3; P = 2.31×10^{-11}<0.05). The Hosmer test called the reliability of the model into question. However, the high percentage of correct definitions (74.3%) allows us to propose the developed logistic function for performing classification analysis. The corresponding ROC curve is presented in Fig. 4.4.

However, the classification would be incomplete without further recognition within the groups "grain corn – soybeans" and "wheat – barley". Despite the great similarity of the seasonal NDVI dynamics, an auxiliary (second-grade) logistic classification models were developed, allowing a complete classification analysis. The results of the statistical analysis of the logistic classification within the group "grain corn – soybeans" certified its reliability according to the χ^2 criterion (χ^2 = 60.71; df = 5; P=8.65×10^{-12}<0.05), which allowed to reject the null hypothesis. According to the results of the Hosmer test, the null hypothesis was not rejected. However, taking into account the sufficiently high level of true predictions (65.6%), it could be that the developed model may perform reasonably enough for crops recognition. The corresponding ROC curve is presented in Fig. 4.5. As for the classification of wheat and barley, the null hypothesis was rejected both by the χ^2 criterion (χ^2 = 132.36; df = 2; P=1.81×10^{-29}<0.05) and by the Hosmer test (Hosmer = 2258746; df = 15; P=0<0.05). The high percentage of correct guesses (95.7%)

allows to propose the developed logistic function for performing a second-grade classification analysis. The corresponding ROC curve is presented in Fig. 4.6.

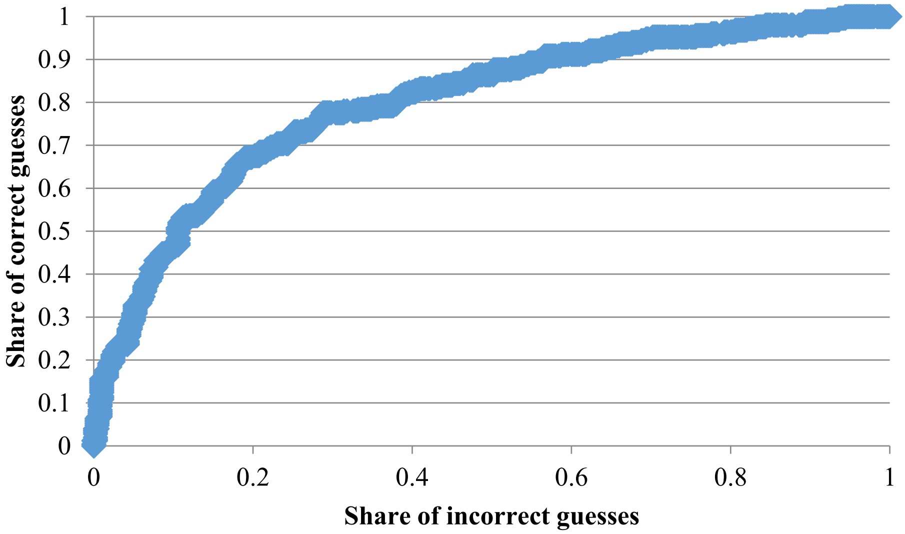


Fig. 4.3 ROC curve of logistic classification of late spring crops

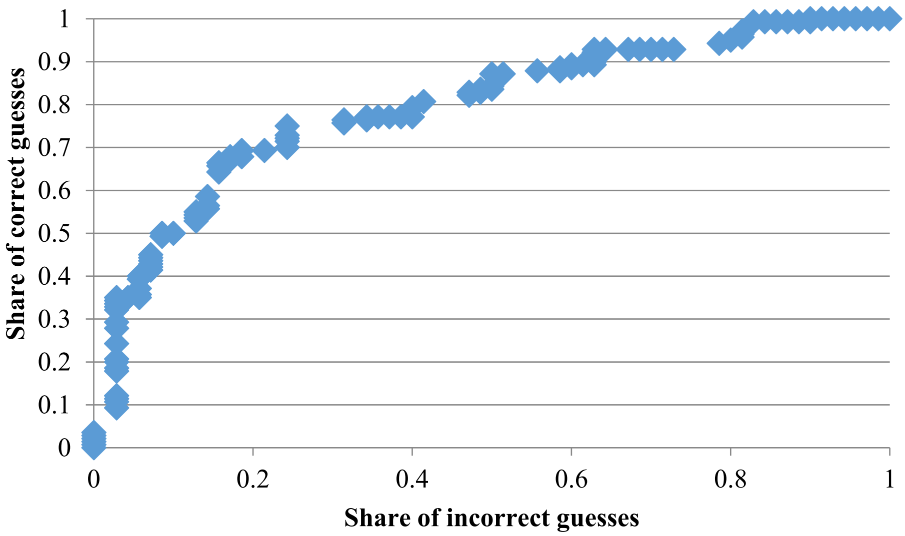


Fig. 4.4 ROC curve of logistic classification of winter crop crops

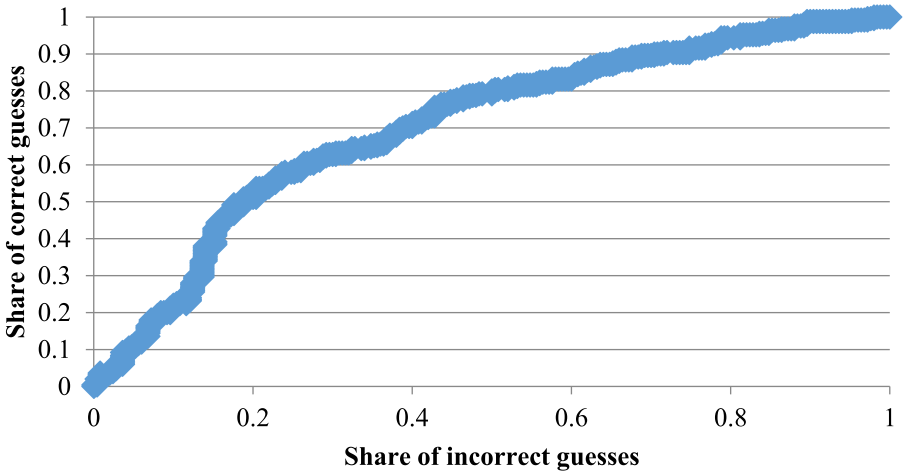


Fig. 4.5 ROC curve of logistic classification of corn and soybean crops

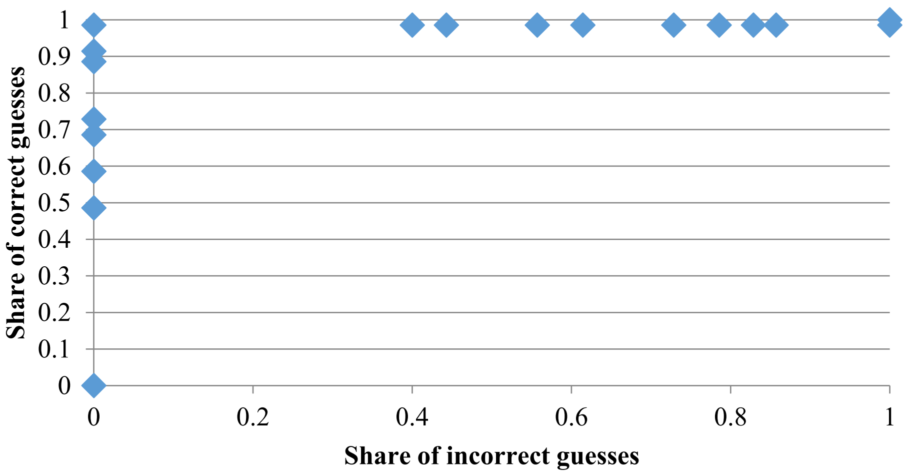


Fig. 4.6 ROC curve of logistic classification of winter wheat and barley crops

Based on the results of generalized mathematical analyses using discriminant and logistic regression algorithms, the Agroland Classifier application was developed to enable automated classification of irrigated and non-irrigated lands, as well as identification of the studied agricultural crop types. The application was built using common web development stack (HTML, CSS, and JavaScript) and features a user-friendly, intuitive interface in English (global version), with an additional Ukrainian-language version designed for domestic users. The software is cross-platform and can operate on devices running virtually any operating system; its functionality has been successfully tested on Windows 11, Android 11, and iPadOS 17.4.

To facilitate user comprehension of the operational principles, the application includes detailed instructions and provides direct links to primary research materials published in peer-reviewed scientific journals. These resources offer users an opportunity to explore the methodological foundations and classification algorithms upon which the program is based. The algorithm for performing the classification of croplands in the Agroland Classifier application is as follows:

1) Classification of Irrigated and Non-Irrigated Lands:
The classification of reclaimed lands into irrigated and non-irrigated categories is implemented in the Irrigated Lands Classifier section of the Agroland Classifier application. Classification can be performed using either the multiclass linear and canonical discriminant analysis algorithms or the logistic regression algorithm. Average monthly NDVI values for the period from May to October are entered into the corresponding data input cells.

When recognition is performed using discriminant analysis, classification decisions are based on two canonical functions. Ideally, both functions yield the same result; however, in cases where their outputs differ, the user should rely on the second canonical function (Decision 2) to make the final determination. The logistic regression algorithm may also be used either as a verification tool or as an independent classification method.

2) Classification of Major Agricultural Crops:
Crop classification is performed in the Crops Classifier section of the application. This module employs the two-stage logistic regression algorithm described previously to differentiate between major crop groups. For late spring row crops, the first stage distinguishes sunflower from the soybean/corn group, and the second stage uses the same NDVI series to differentiate between soybean and corn. For winter crops, the first stage separates rapeseed from the wheat/barley group, and the second stage distinguishes between wheat and barley within that group.

Given the higher classification accuracy of the logistic regression approach compared with multiclass linear and canonical discriminant analysis (71.1% for row crops and 85.0% for winter crops versus 56.5% and 68.1%, respectively), the latter algorithms were excluded from the final version of the application.

3) Operational Modes and Statistical Significance of Input Variables:
The application can be used in both online and offline modes, as well as in a corporate mode that allows users to share calculation and classification results with colleagues. Some input fields in the logistic regression models are intentionally disabled. This is due to the fact that not all monthly NDVI data for the growing season are statistically significant for classification purposes. The significance of each month's NDVI value was assessed using the P-value criterion, which reflects the probability of obtaining test results at least as extreme as the observed result under the null hypothesis.

Monthly data with regression coefficient β values having $P > 0.05$ were deemed statistically insignificant and thus excluded from the model, as their inclusion would reduce classification confidence below the 95% threshold.

The Agroland Classifier application is provided free of charge upon request and is protected by a copyright certificate. The online version of the application is available at: https://ssccust1.spreadsheethosting.com/1/47/af5f049e3fe9b1/Agroland%20Classifier/Agroland%20Classifier.htm. The interface of the application is represented in Fig. 4.7.

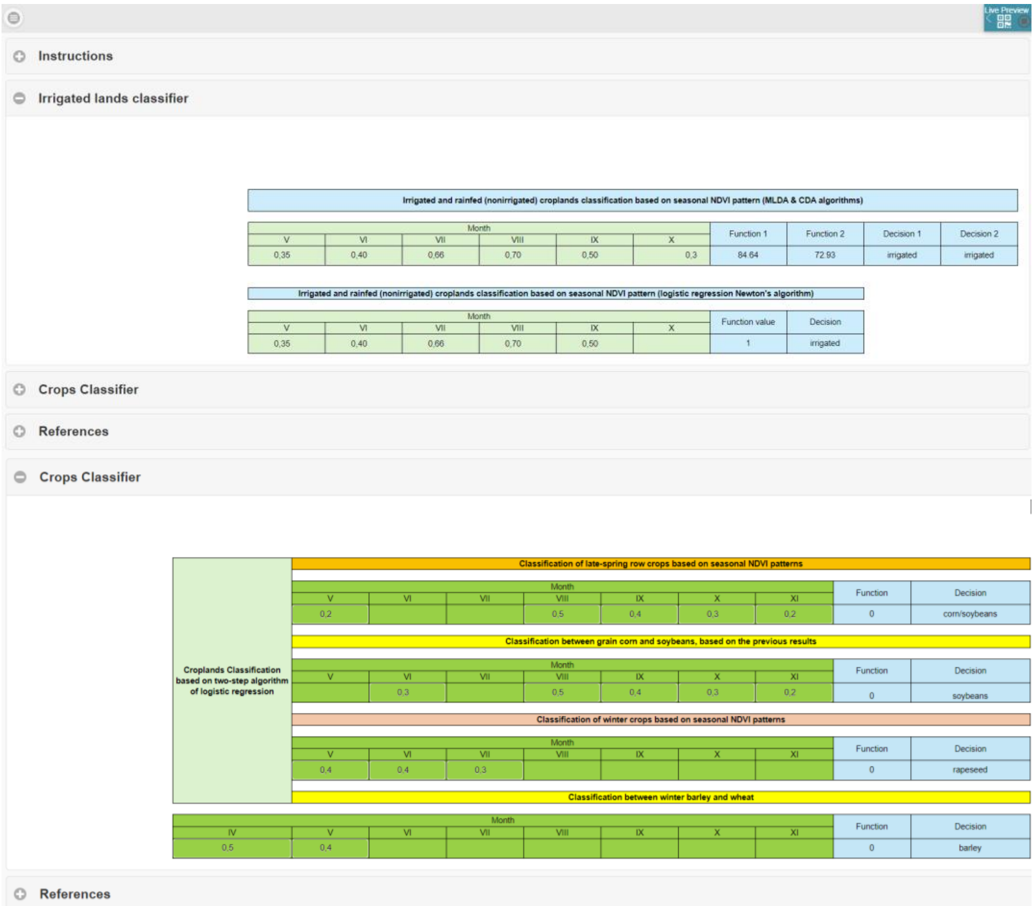


Fig 4.7 Agroland Classifier application interface

4 Validation and performance assessment of the Agroland Classifier application

Validation testing was conducted using an independent dataset comprising 100 fields (50 irrigated and 50 non-irrigated) located in the Kherson and Mykolaiv regions of Ukraine. The testing confirmed high classification accuracy in distinguishing irrigated from non-irrigated lands and revealed notable differences between the algorithms. The canonical discriminant function demonstrated more consistent results for both irrigated (88% accuracy) and non-irrigated lands (84% accuracy), while logistic regression performed less effectively for irrigated fields (78% accuracy) but more accurately for non-irrigated ones (96% accuracy).

Therefore, Agroland Classifier can be recommended for both scientific research and practical applications in the semi-automated identification of irrigated and non-irrigated lands and for monitoring agricultural water use.

An assessment of classification accuracy for 600 crop fields (100 fields per crop: wheat, barley, winter rapeseed, corn, soybean, and sunflower) showed that the highest accuracy (82%) was achieved for wheat, while the lowest (50%) was recorded for soybean. Overall, the observed error range corresponded to a good to reasonable forecast reliability, confirming the applicability of the tool for both theoretical and practical agronomic uses.

Further research will focus on improving the classification models through machine learning using additional labeled and unlabeled datasets (unsupervised learning introduction) to enhance recognition accuracy.

Deep Science Publishing, 2025
https://doi.org/10.70593/978-93-7185-854-0

Chapter 5: Databases of crops productivity and agroecological conditions

1 Agroecological zoning database

The use of remote sensing data is promising for addressing the challenges of croplands agroecological zoning by their suitability for crops cultivation. Satellite imagery acquired across multiple spectral bands provides valuable detailed information on numerous properties of on-land components of agricultural ecosystems, which are reflected in the values of corresponding vegetation indices.

The combination of several vegetation indices enables a comprehensive assessment of agroecosystem conditions and allows deep evaluation of the degree to which environmental parameters within a given area correspond to the biological requirements of the crops, which are due to be cultivated there. An example of such an integrative approach is the Agroecological Zoning Index (AEZI), which is derived from the combined analysis of three vegetation indices, viz., the Normalized Difference Vegetation Index (NDVI), which reflects the overall vegetation cover condition; the Normalized Difference Water Index (NDWI), which indicates water stress levels; and the Nitrogen Reflection Index (NRI), which characterizes the nitrogen availability for plants – the principal nutrient for crops growth, development and productivity formation.

Through this approach, nearly every limiting factor affecting plant development is considered. The value of AEZI provides an integrated measure of environmental suitability for crop cultivation, expressed as a percentage (0–100%), with the component indices normalized within a range of 0–1. Agroecological suitability is classified by the AEZI values according to the following gradual scale:

- **0–20%** – unsuitable;
- **21–30%** – poorly suitable (marginal);
- **31–40%** – conditionally suitable;
- **41–60%** – suitable;
- **61–75%** – optimal;

- **>75%** – ideal conditions.

Based on AEZI values calculated for strategic crops of the Steppe zone of Ukraine, a comprehensive database of agroecological zoning for the South of the country was developed for the 2018-2019 period.

The database was developed within the framework of Microsoft Access, providing a convenient and versatile platform for AEZI storage, management, and analysis. The main elements of the database interface are presented in Fig. 5.1 and Fig. 5.2. The interface is designed to be simple and intuitive, facilitating ease of use for both researchers and practitioners of different expertise level. The database is available in two language versions – English and Ukrainian – to accommodate both international and domestic users.

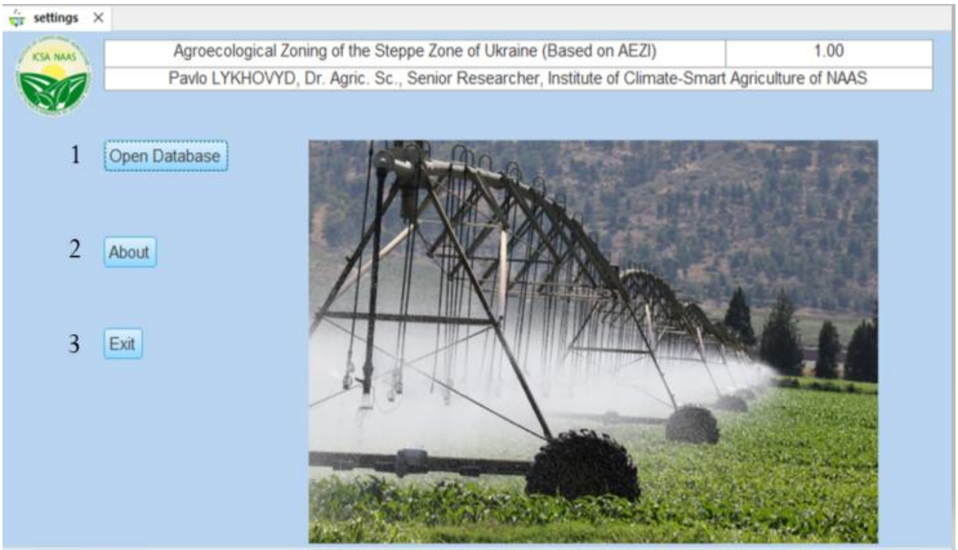


Fig. 5.1 Start window of the database Agroecological Zoning of the Steppe Zone of Ukraine (1 – open button; 2 – about button; 3 – exit button)

The database contains information on the agroecological zoning of seven southern regions of Ukraine within the Steppe zone, namely: Crimea, Kherson, Mykolaiv, Zaporizhzhia, Odesa, Dnipropetrovsk, and Kirovohrad regions. Zoning was performed based on the results of the calculated Agroecological Zoning Index (AEZI) for the 2018-2019 period for the strategic crops of the region, including grain corn, soybeans, sunflower, wheat, barley, rapeseed, alfalfa, sugar beet, and rice. Both irrigated and non-irrigated growing conditions were considered in the assessment.

The reference section of the database provides methodological background information on the principles of agroecological zoning applied to the Steppe territories of Ukraine, which constitute the conceptual foundation of the system. It also includes References section containing recommended citation formats, enabling users to properly reference the database in their scientific publications.

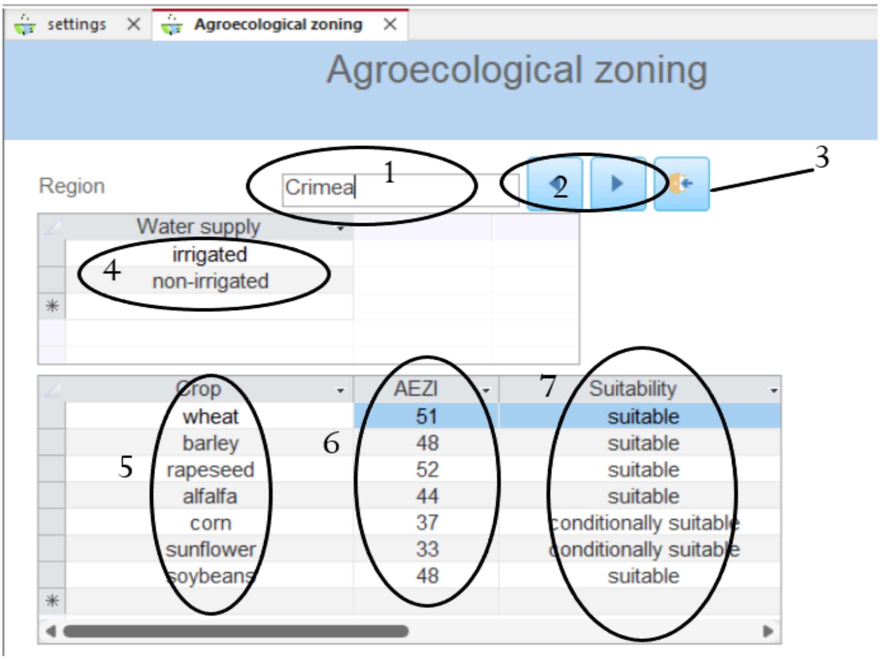


Fig. 5.2 Main window of the database Agroecological Zoning of the Steppe Zone of Ukraine (1 – region; 2 – region switcher; 3 – exit button; 4 – irrigated/rainfed toggle; 5 – crops; 6 – AEZI values; 7 – suitability estimation)

Future development of the database will focus on expansion through the inclusion of data from subsequent time periods, thereby enhancing its analytical potential and supporting long-term monitoring of agroecological conditions across southern Ukraine.

2 Soil and climate of Ukraine database

Comprehensive soil and climate databases are important tools for both researchers and agricultural practitioners. By providing standardized and integrated information about key environmental drivers, which are impactful for agricultural production, such databases enhance the quality of scientific research, support informed decision-making, and facilitate the adoption of innovative technologies by farmers.

These databases are essential for promoting evidence-based decisions for sustainable land, water, and crop management. They also contribute to the development of adaptive cultivation technologies under global warming conditions and serve as valuable resources for precision climate-smart agriculture systems, sustainable climate adaptation programs, and the management of natural resources.

Comprehensive databases typically integrate data from different sources, applying standardized parameters and performing data harmonization to ensure comparability and consistency. However, their establishment and long-term maintenance require specific technical expertise and considerable financial investment – costs that are often not fully covered by institutional or governmental budgets. As a result, their development frequently depends on the initiative and commitment of motivated researchers and practitioners.

Thus, comprehensive soil and climate databases represent a cornerstone of modern agriculture. They provide the foundation for high-quality research and effective transfer of innovative agricultural technologies, fostering the development of resilient and sustainable agricultural systems amid climate instability and biodiversity loss.

In Ukraine, several specialized databases have been created to support agroecological research. Among them is the database called “Agroecological Zoning of the Steppe Zone of Ukraine”, which was referred to in the Chapter 5.1, and the Database of Ecological Properties of Soils of Ukraine, which compiles information on soil quality and environmental attributes. Despite these achievements, Ukraine still lacks a universal, interactive soil and climate database that would integrate long-term observational data, provide open access to agroclimatic indicators, and support advanced scientific analysis through machine learning algorithms. Ukrainian researchers themselves emphasize the need for systematic data integration and coordination to facilitate the creation of such a comprehensive national database.

The Soil and Climate Database of Ukraine was developed using open-access agrometeorological data on air temperature, precipitation, wind speed, and relative humidity provided by the Ukrainian Hydrometeorological Institute. The dataset summarizes observations for the period 1946-2024 on a regional and monthly basis and is supplemented with data from the Central Geophysical Observatory of Ukraine.

Soil surface temperature data were obtained via JavaScript queries executed in Google Earth Engine (GEE) using the MODIS/061/MOD11A2 aerospace monitoring dataset, while potential evapotranspiration was derived from the IDAHO_EPSCOR/TERRACLIMATE dataset, with a spatial resolution of 1 km.

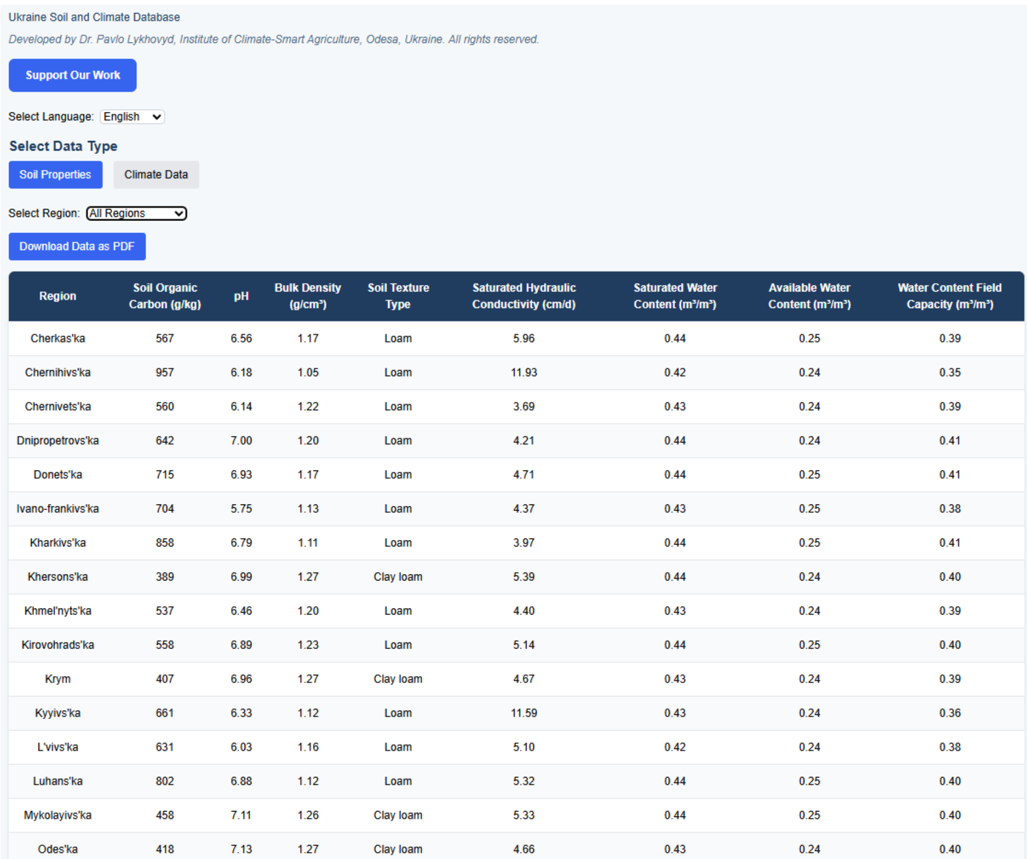
Meteorological indices – including the Standardized Precipitation Evapotranspiration Index (SPEI) and the Palmer Drought Severity Index (PDSI) – were calculated according to internationally recognized methodologies. The interpretation of index values follows current conventions used for agroclimatic classification and ecosystem gradation.

Soil data for Ukraine were obtained through the Google Earth Engine platform from the OpenLandMap databases (including soil type, organic matter content, soil pH, and bulk density, with a resolution of 250 m) and from HiHydroSoil v2.0, which provides data on soil hydro-physical properties. Regional data separation was performed using the FAO/GAUL/2015/level1 administrative boundary mask. As soil properties are relatively stable over time, they are presented in the database as static parameters, without temporal dynamics.

The database was developed using common web stack (HTML, CSS, JavaScript + React), which ensures efficient data access, an interactive graphical interface, and integrated analytical tools. The Soil and Climate Database of Ukraine is accessible at: https://ukr-soil-clim-database.web.app/.

The developed database serves as a convenient and flexible tool for obtaining primary soil and climatic data, as well as for performing their analysis in agronomic research.

The user interface of the application begins with a start page containing the database title, information about the developer, and a drop-down menu for language selection (Ukrainian or English). This page also presents tabular data on the soil properties of Ukraine, which can be filtered by region and downloaded (Fig. 5.3).



Region	Soil Organic Carbon (g/kg)	pH	Bulk Density (g/cm^3)	Soil Texture Type	Saturated Hydraulic Conductivity (cm/d)	Saturated Water Content (m^3/m^3)	Available Water Content (m^3/m^3)	Water Content Field Capacity (m^3/m^3)
Cherkas'ka	567	6.56	1.17	Loam	5.96	0.44	0.25	0.39
Chernihivs'ka	957	6.18	1.05	Loam	11.93	0.42	0.24	0.35
Chernivets'ka	560	6.14	1.22	Loam	3.69	0.43	0.24	0.39
Dnipropetrovs'ka	642	7.00	1.20	Loam	4.21	0.44	0.24	0.41
Donets'ka	715	6.93	1.17	Loam	4.71	0.44	0.25	0.41
Ivano-frankivs'ka	704	5.75	1.13	Loam	4.37	0.43	0.25	0.38
Kharkivs'ka	858	6.79	1.11	Loam	3.97	0.44	0.25	0.41
Khersons'ka	389	6.99	1.27	Clay loam	5.39	0.44	0.24	0.40
Khmel'nyts'ka	537	6.46	1.20	Loam	4.40	0.43	0.24	0.39
Kirovohrads'ka	558	6.89	1.23	Loam	5.14	0.44	0.25	0.40
Krym	407	6.96	1.27	Clay loam	4.67	0.43	0.24	0.39
Kyyivs'ka	661	6.33	1.12	Loam	11.59	0.43	0.24	0.36
L'vivs'ka	631	6.03	1.16	Loam	5.10	0.42	0.24	0.38
Luhans'ka	802	6.88	1.12	Loam	5.32	0.44	0.25	0.40
Mykolayivs'ka	458	7.11	1.26	Clay loam	5.33	0.44	0.25	0.40
Odes'ka	418	7.13	1.27	Clay loam	4.66	0.43	0.24	0.40

Fig. 5.3 Soil database interface

The interface includes buttons for switching between different data categories. By default, soil data are displayed. When climatic data are selected, an additional window opens, allowing users to specify the time span for analysis, as well as to visualize the corresponding climatic trends. Both tabular data and graphical outputs can be exported and downloaded for further use (Fig. 5.4).

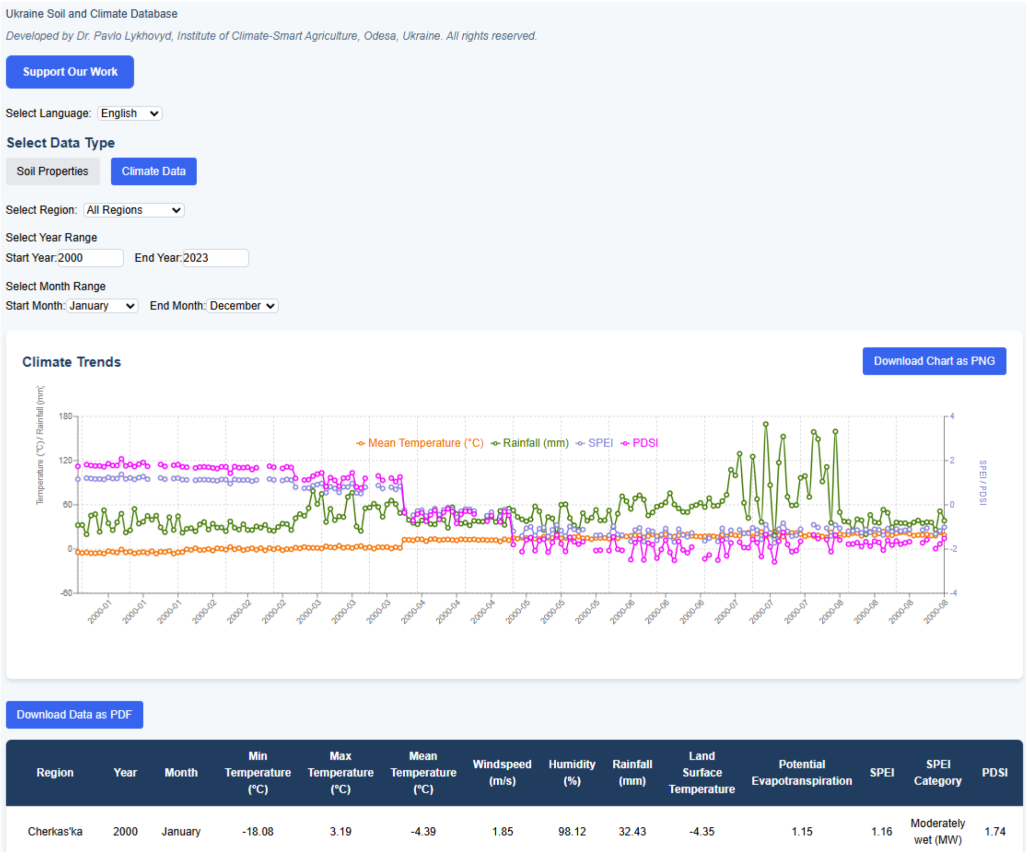


Fig. 5.4 Climate database interface

An important advantage of the database is its accessibility and scalability across different devices. Users can export data tables in PDF format and graphs in PNG format for offline analysis. Interactive charts allow detailed examination of data values when hovering the cursor over the elements of plots.

Compared with existing resources, the developed database represents a unique interactive platform for Ukraine. For the first time, it provides open access to an extensive range of agroclimatic indicators covering a long-term period (1946-2024) and offers built-in analytical capabilities. This significantly reduces the time required for data acquisition and processing, thereby supporting the development of adaptive agricultural technologies.

Future improvements will focus on expanding functionality through the inclusion of additional indicators and datasets, further enhancing its value for agronomic and environmental research.

3 Ukraine crop production map

The development of interactive maps to depict crop productivity through remote sensing data represents a highly relevant and rapidly evolving trend in modern agricultural science, particularly in the context of recent advances in GIS. Such maps provide a valuable source of operational, dynamic, and spatially detailed information, which is critically important for crop monitoring, yield prediction, land resource management, and food security insurance.

The development of modern information products relies on the integration of field-based data with satellite monitoring and coding. This approach enables tracking of crop growth and development, yield prediction, and informed decision-making across all levels – from individual farms to national agricultural policies. Moreover, these tools facilitate the analysis of climate change and technological impacts on agrophytocenoses, as well as the assessment of water and land resource availability and utilization, thereby contributing to effective agricultural planning and environmental risk reduction.

Recent scientific advances have enhanced the functionality of such systems across the multisource remote sensing data (e.g., MODIS, Landsat-8, Sentinel-2) with machine learning and deep learning algorithms implementation. These methods substantially improve the accuracy, resolution, and scalability of yield predictions, enabling the generation of reliable, near real-time data over large territories. Cloud platforms, particularly Google Earth Engine, play a pivotal role by facilitating large-scale data processing, reducing computational complexity, and streamlining multisource data retrieval and further re-integration in external systems.

The relevance of developing interactive productivity maps is determined by several key factors:

- Food security insurance – supports the rational management of agricultural resources under changing climatic conditions.
- Prediction accuracy improvement – integration of satellite-derived vegetation indices (NDVI, EVI, NDMI, etc.) with actual yield data significantly enhances reliability of predictions.
- Sustainable agricultural planning – provides a scientific basis for regional and national strategies aimed at achieving long-term sustainability in agricultural production.

The development and implementation of an interactive crop productivity map for Ukraine represents a strategic task aimed at enhancing agricultural efficiency, supporting adaptation to climate variability, and promoting the principles of sustainable development within the agricultural sector.

The goal of the study was to develop an interactive electronic map of productivity for strategic crops in Ukraine by integrating satellite-derived vegetation indices with mathematical statistics and machine learning methods, thereby enabling in-depth analysis of agroecosystem productivity.

To achieve this objective, a comprehensive interactive web application, the Ukraine Crop Production Map was developed. This resource combines cartographic visualization with extensive analytical functionality, allowing users not only to visualize crop productivity but also to access reference information, perform mathematical and statistical analyses, construct regression models, and generate short- and long-term forecasts based on remote sensing data.

The interactive map was implemented using satellite imagery processed via a custom Google Earth Engine (GEE) script. Vegetation indices were calculated from the following sources: NDMI from MODIS/006/MOD09A1, and NDVI, NDWI, EVI, and NRI from MODIS/061/MOD09GA, all at a spatial resolution of 500 m. Index calculations were performed within the script according to internationally recognized methodologies. The AEZI was computed using the author's method, integrating NDVI, NDWI, and NRI values. Yield data for the principal crops were obtained from official statistical sources, specifically the statistical yearbooks of Ukraine.

The technical implementation of the web application was carried out using a standard web technology stack, including HTML5, CSS3, and JavaScript, complemented by libraries for data processing and visualization such as D3.js, Leaflet.js, Chart.js, math.js, and PapaParse.js.

The application provides a range of tools for mathematical and statistical analysis, including:

- Correlation matrix construction, enabling assessment of the relationships between vegetation indices and crop productivity using the Pearson correlation coefficient.
- Trend analysis based on the non-parametric Mann-Kendall test, allowing the identification of statistically significant trends in time series data.
- Regression modeling, including pairwise and multiple linear regression methods to evaluate the determinism and predictive relationships among variables.
- Time series forecasting using moving averages and exponential smoothing within the Holt-Winters algorithm.

The web application interface is demonstrated in Fig. 5.5. The home page provides access to the core functionality, including an interactive map of Ukraine that visualizes both spatial and temporal data on the productivity of principal crops for the period 2005–2023, as well as satellite-derived vegetation and agroecological indices, including NDVI, EVI, NDMI, NRI, and AEZI.

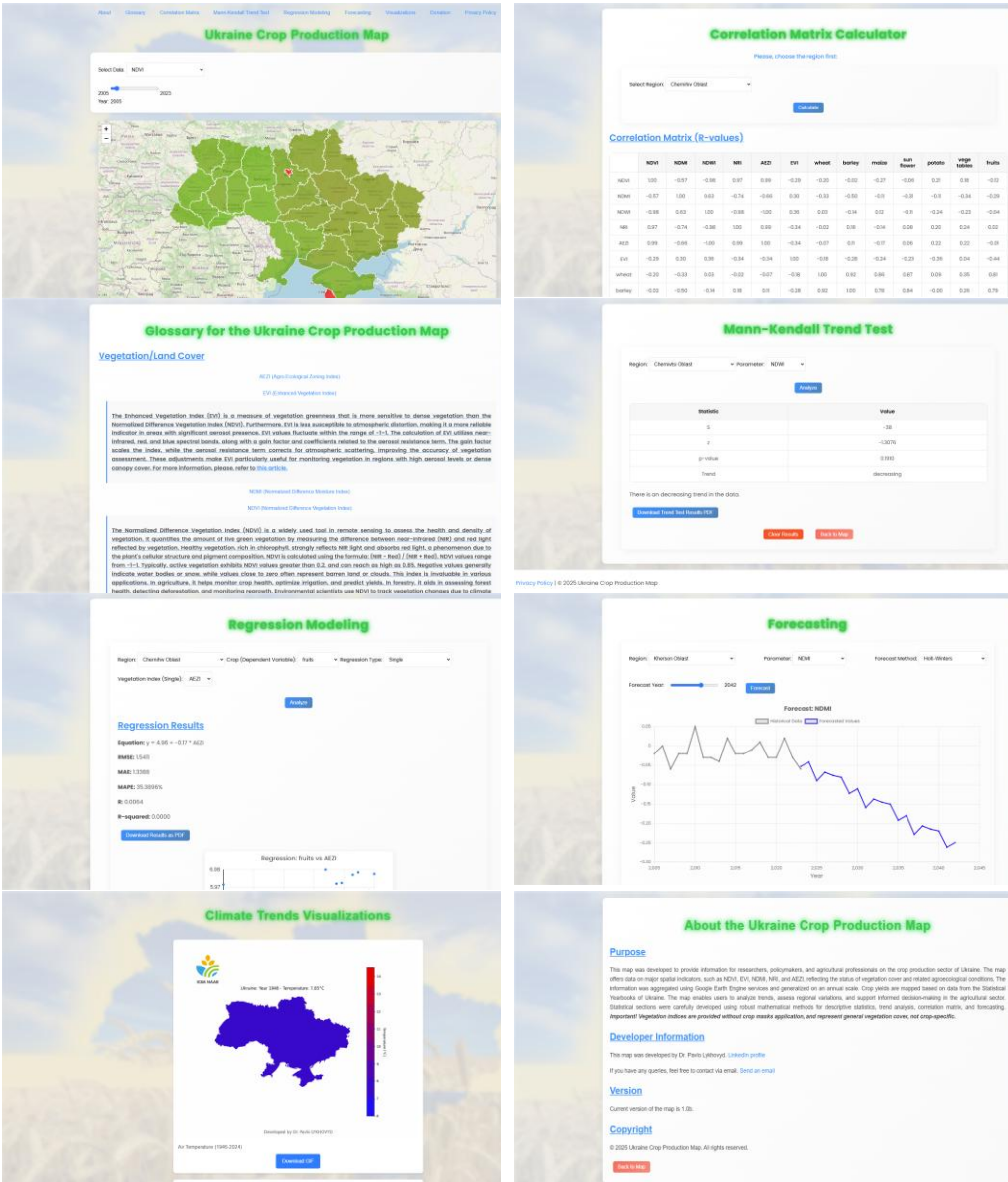


Fig. 5.5 Interface and main functionality of the Ukraine Crop Production Map

Users can select the required index from the Select Data menu, and a slider is provided to choose the corresponding year. By clicking on the map, detailed regional data for the selected parameter are displayed, including the value for the chosen year, the average value for the entire period, minimum and maximum values, standard deviation, and other statistical metrics. A translator widget in the upper-right corner allows users to switch languages, translating the interface content from English to other supported languages seamlessly.

The About section provides information on the application, its current version, the developer, relevant developer profiles, and contact information. The Glossary section offers an interactive terminological dictionary to facilitate understanding of the program's features and functionality.

The Correlation Matrix section enables calculation of correlation matrices for the selected region or for the country on the whole. Users can evaluate the relationships between yield indicators and satellite-derived indices, with outputs including the Pearson correlation coefficient and the coefficient of determination.

The Mann-Kendall Trend Test section allows users to assess trends in selected indicators using the non-parametric Mann-Kendall method, providing relevant statistical measures for trend detection.

The Regression Modeling section supports linear regression analyses, including simple and multiple regression approaches. Users can build crop productivity models and assess their reliability using metrics such as root mean square error (RMSE), mean absolute error (MAE), mean absolute percentage error (MAPE), Pearson correlation coefficient, and coefficient of determination. An approximation curve is automatically generated for visual assessment.

The Forecasting section enables time series prediction for selected indicators using moving averages or the Holt-Winters method. The application generates both a forecast graph and a corresponding table of numerical forecast values for further analysis.

The developed web application represents a unique and innovative platform for Ukraine, integrating satellite and agroclimatic data with crop yields and robust methods of statistical analysis. It employs modern mathematical, statistical, and machine learning methods to assess and predict the state and productivity of agroecosystems.

Owing to its extensive functionality, the application constitutes a full-featured scientific and analytical tool of the new generation. Its practical value lies in the provision of an accessible interface for analyzing large and complex datasets, thereby significantly facilitating evidence-based data-driven decision-making. This tool is relevant for climatology, agriculture, agroecology, and for supporting the development of agrarian policies aimed at resource-efficient agricultural production.

The interactive electronic Ukraine Crop Production Map is more than a visualization tool; it is a multifunctional platform for data analysis and interpretation. It fosters interdisciplinary interaction between science, practice, and management, opening new opportunities in several areas:

- Scientific research: supports analytical studies and the evaluation of agroecosystems.
- Education: serves as an effective tool for training students and professionals in agricultural sciences.
- Policy development: informs the design of strategies to ensure food security.
- Practical application: equips farmers with tools to adapt to climate change and enhance agricultural sustainability.

Overall, the developed platform represents a significant advancement toward the sustainable development of the national agricultural sector, combining scientific rigor with practical usability to support informed decision-making and long-term agricultural resilience [48].

Deep Science Publishing, 2025
https://doi.org/10.70593/978-93-7185-854-0

Chapter 6: Conclusions, recommendations and prospects

The results of this research demonstrate the high efficiency and promising potential of using remote sensing data, particularly vegetation and vegetation-related indices (e.g., NDVI, NDWI, NRI, AEZI, etc.), for assessing, recognizing, and mapping the fields of major agricultural crops in Ukraine.

The study facilitated the development of a series of mathematical models and functions (linear, canonical, discriminant, and logistic) for:

- Classification and identification of crop types;
- Recognition of irrigated and non-irrigated lands with high accuracy.

It was established that climate change significantly impacts agricultural production, and NDVI monitoring provides an effective means of assessing these impacts and identifying optimal conditions for cultivating specific crops across different zones of Ukraine. At the same time, although NDVI is a valuable tool, certain cases (e.g., sweet corn) may benefit from the use of traditional indices, such as the Leaf Area Index (LAI), which can offer higher predictive accuracy for productivity. Combining satellite data with on-land surveys is another way of increasing precision and getting better modeling results.

The results of the presented scientific research possess substantial scientific, theoretical, and practical value, as evidenced by its implementation in both higher education institutions, such as Kherson State Agrarian and Economic University, and agricultural production (e.g., in the field practice of the private farm “Vostok” located in Kherson region). Developed software, including the interactive electronic map Ukraine Crop Production Map, the Agroland Classifier application for crop and meliorated land classification, and databases such as the Agroecological Zoning of the Steppe Zone of Ukraine and the Ukraine Soil and Climate Database, make these results widely accessible to researchers, educators, and practitioners.

Based on the findings and identified regularities, the following recommendations are proposed to advance this research direction and enhance its practical value:

- Expand the range of studied crops and indices: Continue developing predictive models for less-studied crops and integrate new or combined indices to improve forecasting accuracy.
- Deepen climate change analysis: Examine the effects of sharp weather fluctuations on crop yields, particularly for early-ripening and late-ripening varieties, to guide adaptive agricultural practices.
- Enhance analytical tools and software: Further develop interactive maps, applications, and databases, incorporating new functionalities and datasets for more accurate and user-friendly agroecological zoning and monitoring.
- Focus on comprehensive territorial assessment: Conduct integrated analyses incorporating agrochemical, climatic, organizational, economic, and social indicators to generate holistic spatial models for effective agricultural management and normalization of anthropogenic impact.
- Implement advanced machine learning methods: Apply advanced robust algorithms for processing big data to achieve more precise analyses.
- Promote practical implementation: Strengthen collaboration with agricultural producers to integrate developed models and software directly into field operations, supporting rational resource use and enhancing food security.
- Support scientific and educational integration: Collaborate with higher education and research institutions to expand educational programs and implement original methodologies in research and teaching, thereby improving training quality and the competitiveness of agronomy graduates.

References

[1] Morozov VV, Lysogorov KS, Shaporybska NM. Geoinformation systems in agrosphere. Kherson: KhDU; 2007.

[2] Bilynsky JJ, Knysh BP. Analysis of features and substantiation of vegetation indices. Bulletin of Vinnytsia Politechnic University. 2021;(2):7-14.

[3] Solymosi KJ, Kövér G, Romvári R. The progression of vegetation indices: a short overview. Acta Agraria Kaposváriensis. 2019;23(1):75-90.

[4] Polishchuk VV. Method of effectiveness assessment of information system implementation for operational planning irrigation GIS-Irrigation. Bulletin of National University of Water Economy and Environmental Use. 2015;(3):431-8.

[5] Ganzelmeier H. GIS-aided application of plant protection products. Outlooks on Pest Management. 2005;16(6):273.

[6] Weidong Z, Chun W, Jing H. Development of agriculture machinery aided guidance system based on GPS and GIS. In: 2010 World Automation Congress; 2010 Sep 26-30; Kobe, Japan. IEEE; 2010. p. 313-7.

[7] De Wit A, Boogaard H, Fumagalli D, Janssen S, Knapen R, van Kraalingen D, et al. 25 years of the WOFOST cropping systems model. Agricultural Systems. 2019;168:154-67.

[8] Sadenova M, Beisekenov N, Varbanov PS. Assessing the effectiveness of RS, GIS, and AI data integration in analysing agriculture performance to enable sustainable land management. Discover Sustainability. 2024;5(1):478.

[9] Bharathkumar L, Mohammed-Aslam MA. Crop pattern mapping of tumkur taluk using NDVI technique: a remote sensing and GIS approach. Aquatic Procedia. 2015;4:1397-404.

[10] Abbas AA. Integration of remote sensing and GIS for precision agriculture: a case study. International Journal of Structural Design and Engineering. 2024;5(2):15-24.

[11] Mo X, Liu S, Lin Z, Xu Y, Xiang Y, McVicar TR. Prediction of crop yield, water consumption and water use efficiency with a SVAT-crop growth model using remotely sensed data on the North China Plain. Ecological Modelling. 2005;183(2-3):301-22.

[12] Bhoutika K, Das DP, Kumar A, Pandey A. Application of remote sensing and GIS in crop yield forecasting and water productivity. In: Geospatial technologies for land and water resources management. Cham: Springer International Publishing; 2021. p. 207-22.

[13] Kravchuk V, Novohatskyi M, Serdiuchenko N, Saidak R. Modeling yields of spring cereals using remote sensing data. Technical and technological aspects of development and testing new machinery and technologies for agriculture of Ukraine. 2013;17(2):4-16.

[14] Pasichnyk NA. The method of forecasting yields on the basis of remote sensing data of high resolution on the example of wheat has been suggested. Balanced Nature Using. 2020;(2):129-34.

[15] Khanal S, Kc K, Fulton JP, Shearer S, Ozkan E. Remote sensing in agriculture – accomplishments, limitations, and opportunities. Remote Sensing. 2020;12(22):3783.

[16] D'Allestro P, Parente C. GIS application for NDVI calculation using Landsat 8 OLI images. International Journal of Applied Engineering Research. 2015;10(21):42099-102.

[17] Fang H, Baret F, Plummer S, Schaepman-Strub G. An overview of global leaf area index (LAI): Methods, products, validation, and applications. Reviews of Geophysics. 2019;57(3):739-99.

[18] Oymatov R, Musaev I, Bakhriev M, Aminova G. Monitoring agricultural land areas using GIS-online program EOS DA: case study of Andijan region. In: E3S Web of Conferences. Vol. 401. EDP Sciences; 2023. p. 02005.

[19] Hinkle DE, Wiersma W, Jurs SG. Correlation: a measure of relationship. Applied Statistics for the Behavioral Sciences. 1998;4(1):105-31.

[20] Evans JD. Straightforward statistics for the behavioral sciences. Thomson Brooks/Cole Publishing Co.; 1996.

[21] Lee DK, In J, Lee S. Standard deviation and standard error of the mean. Korean Journal of Anesthesiology. 2015;68(3):220.

[22] Bland JM, Altman DG. Statistics notes: measurement error. BMJ. 1996;312(7047):1654.

[23] Belsley DA, Kuh E, Welsch RE. Regression diagnostics: identifying influential data and sources of collinearity. John Wiley & Sons; 2005.

[24] Chang YW, Hsieh CJ, Chang KW, Ringgaard M, Lin CJ. Training and testing low-degree polynomial data mappings via linear SVM. Journal of Machine Learning Research. 2010;11(4):1471-490.

[25] De Myttenaere A, Golden B, Le Grand B, Rossi F. Mean absolute percentage error for regression models. Neurocomputing. 2016;192:38-48.

[26] Willmott CJ, Matsuura K. Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. Climate Research. 2005;30(1):79-82.

[27] Porter DC, Gujarati DN. Basic econometrics. New York: McGraw-Hill Irwin; 2009.

[28] Akaike H. A new look at the statistical model identification. IEEE Transactions on Automatic Control. 1974;19(6):716-23.

[29] Konishi S, Kitagawa G. Information criteria and statistical modeling. Springer; 2008.

[30] Findley DF. Counterexamples to parsimony and BIC. Annals of the Institute of Statistical Mathematics. 1991;43(3):505-14.

[31] Hannan EJ, Quinn BG. The determination of the order of an autoregression. Journal of the Royal Statistical Society: Series B (Methodological). 1979;41(2):190-5.

[32] Winters PR. Forecasting sales by exponentially weighted moving averages. Management Science. 1960;6(3):324-42.

[33] Wickens TD, Keppel G. Design and analysis: a researcher's handbook. Upper Saddle River, NJ: Pearson Prentice-Hall; 2004.

[34] Kirk RE. Experimental design: procedures for the behavioral sciences. Sage Publications; 2012.

[35] Білинський ЙЙ, Книш БП. Аналіз характеристик та обґрунтування індексів рослинності. Вісник Вінницького Політехнічного Інституту. 2021;(2):7-14.

[36] Lykhovyd PV, Sharii VO. Normalised difference moisture index in water stress assessment of maize crops. Agrology. 2024;7(1):21-6.

[37] Zeng J, Zhou T, Qu Y, Bento VA, Qi J, Xu Y, et al. An improved global vegetation health index dataset in detecting vegetation drought. Scientific Data. 2023;10(1):338.

[38] Rojas O. Agricultural extreme drought assessment at global level using the FAO-Agricultural Stress Index System (ASIS). Weather and Climate Extremes. 2020;27:100184.

[39] Ушкаренко VO, Вожегова RA, Голобородько SP, Коковіхін SV. Методика польового досліду (зрошуване землеробство). ОлдіПЛЮС, Херсон; 2020.

[40] Maulud D, Abdulazeez AM. A review on linear regression comprehensive in machine learning. Journal of Applied Science and Technology Trends. 2020;1(2):140-7.

[41] Schonlau M, Zou RY. The random forest algorithm for statistical learning. The Stata Journal. 2020;20(1):3-29.

[42] Friedman JH. Greedy function approximation: a gradient boosting machine. Annals of Statistics. 2001;29:1189-232.

[43] Lykhovyd PV. Using normalised difference vegetation index in classification and agroecological zoning of spring row crops. Biosystems Diversity. 2023;31(4):506-12.

[44] Lykhovyd P, Vozhehova R, Hranovska L, Bidnyna I, Shablia O, Kokoiko V, et al. Implementation of agroecological zoning index to map the suitability of agricultural lands of the Steppe zone of Ukraine for the cultivation of major crops. Modern Phytomorphology. 2024;18:41-6.

[45] Lykhovyd P, Averchev O, Chaban V, Nikitenko M, Haievskyi S, Bidnyna I, et al. Agroecological zoning of Ukraine using remote sensing and unsupervised clustering. Modern Phytomorphology. 2025;19:420-5.

[46] Lykhovyd P. Using normalised difference vegetation index in classification and agroecological zoning of spring row crops. Biosystems Diversity. 2023;31(4):506-12.

[47] Lykhovyd P, Vozhehova R, Averchev O. Using remote sensing normalised difference vegetation index to recognise irrigated croplands via Agroland Classifier application. Visnyk of V N Karazin Kharkiv National University, series "Geology. Geography. Ecology". 2024;(61):223-33.

[48] Lykhovyd PV, Vozhehova RA, Hranovska LM, Chaban VO, Bidnyna IO. Ukraine Crop Production Map – an interactive map of crop productivity and analytical evaluation of agroecological conditions of Ukraine. Tavrian Scientific Herald. 2025;143(1):172-8.