



## Application of remote sensing methods for statistical estimation of organic matter in soils

Vadym Belenok<sup>1\*</sup>, Liliia Hebryn-Baidy<sup>1,5</sup>, Natalyya Bielousova<sup>1</sup>, Halyna Zavarika<sup>2</sup>, Sergiy Kryachok<sup>3</sup>, Dmytro Liashenko<sup>4</sup>, Tetiana Malik<sup>4</sup>

1. National Aviation University, Aerospace Geodesy and Land Management Department, UA-03058, 1, Kosmonavta Komarova Ave, Kyiv, Ukraine

2. Sievierodonetsk Volodymyr Dahl East Ukrainian National University, International Economics and Tourism Department, UA-01042, 17, John Paul II Street, Kyiv, Ukraine

3. Chernihiv Polytechnic National University, Scientific Institute of Architecture, Design and Geodesy, Department of Geodesy, Cartography and Land Management, UA-14035, 95, Shevchenko Street, Chernihiv, Ukraine

4. Taras Shevchenko National University of Kyiv, Institute of Geology, Geoinformatics Department, UA-03022, 90, Vasylykivska str., Kyiv, Ukraine

5. University of Cambridge, Scott Polar Research Institute, Lensfield Road, Cambridge CB2 1ER, Cambridge, United Kingdom

\*Corresponding author: belenok.vadim@nau.edu.ua

### ABSTRACT

Maintaining soil fertility and optimizing land management strategies on a global scale necessitates a comprehensive understanding of soil's physicochemical properties. This study proposes the deployment of remote sensing techniques as a more efficient, accurate, and cost-effective alternative for exploring the properties of chernozem soils and for predicting soil organic matter (SOM) in specific regions of Kyiv, Ukraine. The spectral properties of the chernozem soils were examined by Landsat 5 TM and 8 OLI satellite imagery. A mosaic of the mean spectral reflectance values for the study period (1986-2015) was created on the Google Earth Engine. The construction of the predictive model employed reflectance values across six bands, a range of vegetation indices (RNDSI, NDWI, NDBI, BAEI, MSAVI, and NDVI), topographic data (slope), and climatic parameters (temperature, precipitation, soil moisture) as regressors. The optimal model was subsequently determined using the Forward stepwise method, which is predicated on F-test and Akaike criteria. This model yielded notable metrics of Multiple R = 0.651, R<sup>2</sup> = 0.424, and Adjusted R<sup>2</sup> = 0.397. The most significant predictors within the model were determined to be RED, NIR, slope, and soil moisture. Upon validation, the model displayed no evidence of heteroscedasticity, multicollinearity, or potential outliers, thereby underscoring its reliability and robustness. The proposed model serves as a valuable tool, offering both reliability and precision in predicting SOM in specific regions of Ukraine.

*Keywords: spectral brightness, Landsat, multiple linear regression, vegetation index, soil organic matter*

## Aplicación de métodos de teledetección para la estimación estadística del material orgánico en suelos

### RESUMEN

Mantener la fertilidad del suelo y optimizar estrategias de manejo de tierras a una escala global demanda un conocimiento amplio de las propiedades fisicoquímicas del suelo. Este estudio propone la implementación de técnicas de teledetección, por su eficiencia, precisión y economía, para la exploración de las propiedades de los suelos chernozem (tierra negra) y para la predicción de Material orgánico del suelo (MOS) en algunas regiones específicas de Ucrania. Las propiedades espectrales de los suelos chernozem se examinaron con imágenes de los satélites Landsat 5TM y 8 OLI. Luego se creó un mosaico de los valores de reflectividad espectral media para el período de estudio (1986-2015) con la herramienta Google Earth Engine. Para la construcción del modelo predictivo se tomaron valores de reflectividad de seis bandas, un amplio rango de índices de vegetación (RNDSI, NDWI, NDBI, BAEI, MSAVI y NDVI), de información topográfica (pendiente), y parámetros climáticos (temperatura, precipitación, humedad del suelo) como regresores. Luego se determinó el modelo óptimo a través del método de selección por pasos hacia adelante, el cual se basó en los criterios de información F-Test y Akaike. Este modelo presentó métricas notables en los valores Multiple R = 0.651, R<sup>2</sup> = 0.424, y Adjusted R<sup>2</sup> = 0.397. Los predictores más significativos del modelo fueron RED, NIR, pendiente y humedad del suelo. Tras la validación, el modelo no presentó evidencias de heterocedasticidad, multicolinealidad, u otros posibles valores atípicos, lo que garantiza su fiabilidad y robustez. Este modelo propuesto sirve como una herramienta útil que ofrece fiabilidad y precisión en la predicción de composición del material orgánico del suelo en regiones específicas de Ucrania.

*Palabras clave: brillo espectral; Landsat; regresión lineal múltiple; índice de vegetación; material orgánico del suelo.*

Manuscript received: 02/01/2022

Accepted for publication: 03/10/2023

#### How to cite item:

Belenok, V., Hebryn-Baidy, L., Bielousova, N., Zavarika, H., Kryachok, S., Liashenko, D., & Malik, T. (2023). Application of remote sensing methods for statistical estimation of organic matter in soils. *Earth Sciences Research Journal*, 27(3), 299-312. <https://doi.org/10.15446/esrj.v27n3.100324>

## 1. Introduction

The significance of soil organic matter (SOM) research cannot be overstated, being at the forefront of solutions to pressing global issues. As a substantial carbon repository, SOM is indispensable to strategies aimed at mitigating climate change, with research on its dynamics providing invaluable insights for enhancing carbon sequestration efforts (Baveye et al., 2020; Navarro-Pedreño, Almendro-Candel and Zorpas, 2021). Moreover, SOM bolsters soil fertility, a critical factor in augmenting agricultural productivity, thereby addressing concerns of food security in the face of an escalating global population (Khangura et al., 2023; Parikh and James, 2012; Das et al., 2021; Chatohin and Lindin, 2001). Beyond that, SOM serves as a vital energy source for soil microorganisms, fostering soil biodiversity (Bach et al., 2020; Wang et al., 2022; Zhang et al., 2021). Recent advancements in remote sensing technologies offer transformative possibilities in the field of SOM research, allowing for large-scale, non-invasive, and temporal assessments of soil conditions (Omia et al., 2023; David, 2013; Das et al., 2021; Bouzekri et al., 2015). The ability of satellites to rapidly cover extensive and often inaccessible geographical regions presents an unrivaled advantage over traditional, labor-intensive soil sampling methods. Combined with ground-based measurements, satellite data offers improved accuracy of SOM estimates, providing a comprehensive analysis of soil fertility indicators (Belenok et al., 2021; Huang et al., 2018; Weiss, Jacob and Duveillerc, 2020; Chen et al., 2006; Zhang et al., 2021; Yuzugullu et al., 2020). Satellite imagery, particularly multispectral and hyperspectral images, are essential tools in assessing SOM due to their ability to capture reflectance and absorption properties of different soil constituents (Luo et al., 2023; Gopp et al., 2017; Gopp et al., 2019; Khangura et al., 2023). Organic matter influences reflectance in the visible (VIS) and near-infrared (NIR) regions, making these spectral bands key in SOM evaluation (Reis et al., 2021; Bouasria et al., 2020; Yu et al., 2021; Yuzugullu et al., 2020; Zhang et al., 2021).

The spectral characteristics of bare soil are shaped by several factors, most notably the content of organic matter, often referred to as humus or SOM, and the level of soil moisture. Specifically, increases in either SOM content or soil moisture are associated with reduced spectral brightness coefficients, a pattern particularly pronounced in the red band of the spectral range (Zhang et al., 2021; Wang et al., 2022; Kravtsova, 2005; Shikhov et al., 2020; Prudnikova and Savin, 2021). Dry soils exhibit a common trait: a monotonic escalation in spectral brightness coefficients as the wavelength extends from 0.4 to 2 microns. As the SOM content rises, the spectral brightness of the soil correspondingly diminishes (Rukhovich et al., 2021). Further, the soils structure can influence its spectral brightness. Unstructured soils, for example, reflect approximately 10-15% lighter than well-structured soils, underscoring the intricate interplay of soil properties on its spectral characteristics (Kravtsova, 2005; Jing et al., 2020; Wang et al., 2022; Yang et al., 2021; Zhang et al., 2021).

The evaluation of SOM using remote sensing techniques utilizes specific spectral indices such as the Ratio Normalized Difference Soil Index (RNDSI) or the Normalized Difference Vegetation Index (NDVI). These indices help in amplifying the SOM signal while simultaneously suppressing the effects of other variables (Yang et al., 2021; Deng et al., 2015; Montero et al., 2023). While the primary application of NDVI lies in vegetation analysis, it can indirectly infer SOM levels as high vegetation areas often correlate with increased SOM content. Many techniques underscore the use of the NDVI as an identifier for bare soil surfaces (Ding et al., 2016; Tian et al., 2021; Lu, Liu and Liu, 2022). Guidelines have been established, which suggest scales of NDVI values for reflectance in the RED (0.25) and NIR (0.3) ranges for open ground (the average NDVI value is thus is equal to 0.025) (Koroleva et al., 2017; Kulyanitsa et al., 2017; Luo et al., 2021). According to the study (Gasmi et al., 2021) suggested using an NDVI threshold value of 0.22 to separate bare soils from other land types. Further, ground cover studies utilize mathematical models built upon parameters like fractional vegetation coverage, nitrogen reflectance index, yellow leaf index, bare soil index, and slope, achieving an accuracy of approximately 90% (Das et al., 2021). Land degradation hot spots and soil salinity are studied using spectral characteristics like surface albedo, the Salinized Land Degradation Index (SDI), the Soil Surface Moisture Index (LSM), Enhanced Vegetation Index (EVI) (Rukhovich et al., 2021; Higginbottom and Symeonakis, 2014; Jiang et al., 2008; Yuzugullu et al., 2020; Gopp et al., 2019). To minimize the effect of some vegetation spectral indices on soil brightness, based on the wavelengths of the RED and NIR ranges bands,

it is proposed to use indices like the Soil-Adjusted Vegetation Index (SAVI) (Huete, 1988) and the Modified Soil Adjusted Vegetation Index (MSAVI) (Qi et al., 1994), as well as the NPV-Soil Separation Index (NSSI), developed to distinguish non-photosynthetic vegetation soils using two NIR bands (Tian et al., 2021). In the study (Chen et al., 2006), authors proposed the utilization four vegetation indices - NDVI, the Normalized Difference Water Index (NDWI) (Gao, 1996), the Difference Built-up Index (NDBI), the Normalized Difference Bareness Index (NDBaI) - for differentiating bare soils from other land use and land cover (LULC) types, setting threshold values for each. These indices are based on the spectral properties inherent to all LULC categories (Chen et al., 2006; Li and Chen, 2014).

A multitude of studies have established a correlation between SOM content in soils with pixel values of multispectral satellite data, often employing multiple linear regression (MLR) or alternative mathematical methodologies (Lu, Liu and Liu, 2022; Luo et al., 2021; Ahmed and Iqbal, 2014; Bouasria et al., 2020). Some authors (Gasmi et al., 2021; Demattê et al., 2007) have successfully used MLR to predict clay content. In a distinct approach (Mirzaee et al., 2016) authors applied geostatistical methods, specifically different types of the kriging method, to predict SOM content in soils. In the study (Jing et al., 2020) authors examined the impact of time gaps between field sampling and the acquisition of Landsat TM/OLI satellite data on soil nutrient predictions, using both MLR and artificial neural network methodologies. According to study (Ahmed and Iqbal, 2014; Fiorio and Demattê, 2009), authors employed MLR to correlate soil surface variables with spectral data. These studies demonstrate the diversity of approaches and tools used in exploring the link between SOM and spectral data.

The technique of this study consists in the correct separation of the bare soil cover from other LULC in order to establish an excellent linear relationship between the spectral indices and the actual soil parameters. It is this indicator and the calculated vegetation indices that will be the basis for forecasting the fertility of agricultural land in the Kyiv region.

The aim of the study is to establish a statistical relationship between SOM and the mean spectral reflectance extracted from satellite images. This process based on using MLR and further assessing the accuracy and significance of this relationship for the chernozems of Kyiv region. The establishment of such a relationship will make it possible to forecast the SOM content in the soil for further assessment of soil fertility.

## 2. Materials and Methods

### *Characteristics of the study area*

Located in the Dnipro River basin in the northern part of Ukraine, the Kyiv region encompasses an area of 28.1 thousand square kilometers, excluding the Kyiv. This region constitutes approximately 4.7% of Ukraine's total land area, with the inclusion of Kyiv bringing the total to 28.9 thousand square kilometers. From an administrative perspective, the region is divided into 25 districts, along with 13 cities of regional subordination. In addition, it includes 30 cities of district subordination, classified as urban-type settlements, and 1,182 rural settlements. As of the start of 2021, the population of the Kyiv region stood at approximately 1,788,530 thousand individuals (Website of the Kyiv Regional State Administration, 2021).

The Kyiv region's lands are uniquely positioned within a transitional soil-climatic zone. This zone serves as an interface between Polissya, characterized by soddy-podzolic soils, and the Forest-steppe region, known for its transitional chernozem soils. This distinctive location contributes to the diversity of soil types in the Kyiv region and influences its ecological and agricultural potential (State Department of Environmental Protection in Kyiv region, 2021).

According to the Main Directorate of the State Land Agency, the total land area within the administrative boundaries of the Kyiv region is 2816.2 thousand hectares. The region's agricultural land spans 1658.9 thousand hectares, representing approximately 58.9% of the region's total land area. Built-up areas account for 137.4 thousand hectares, which equates to 4.9% of the total. Industrial lands - 22.9 thousand hectares (0.8%), while lands designated for transport and communications represent 29.7 thousand hectares (0.9%). Land allotted to law enforcement agencies covers 26.3 thousand hectares (0.9%). The region designates 56.0 thousand hectares for environmental protection. Recreational areas cover a total of 1.8 thousand hectares, with 0.4 thousand

hectares intended for general recreation and an additional 1.4 thousand hectares set aside specifically for recreational facilities. Finally, 1.2 thousand hectares are reserved for historical and cultural purposes (State Department of Environmental Protection in Kyiv region, 2020; Main Department of the State Land Agency in Kyiv region, 2012).

Figure 1 illustrates the territory of the Kyiv region with an emphasis on 15 highlighted districts (excluding city councils). For these highlighted districts, a statistical relationship will be established between the humus content (SOM) in the soil and the Landsat 8 OLI satellite data. The creation Figure 1 employed geoinformation technologies, following the methods delineated in the study (Belenok et al., 2021; Liashenko et al., 2020) in ArcGIS Pro software.



Figure 1. Study area

The Kyiv region boasts a diverse soil cover, with chernozems, constituting approximately 50% of the region's arable land, being the most prevalent. Other soil types found across a significant portion of the region include podzolized soils. This diverse mix of soil types underscores the rich and varied terrain of the Kyiv region (Figure 2). The degree of plowing of the territory exceeds 60% (State Institution "Soils protection Institution of Ukraine", 2016). In the Kyiv region, an estimated 1353.7 thousand hectares of land are degraded, representing 48.1% of the region's total land area and 81.4% of its agricultural land. This level of degradation is among the highest in the Forest-Steppe and Steppe regions of Ukraine. One of the key contributing factors to these negative processes is the excessive pressure exerted on the region's land resources, which includes intensive agricultural activity and widespread plowing of the territory (State Service of Ukraine for Geodesy, Cartography and Cadastre, 2020).

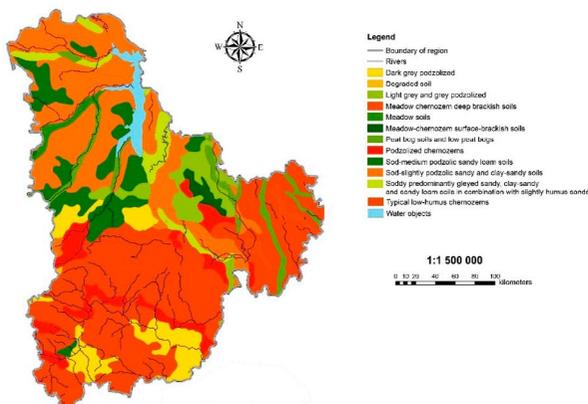


Figure 2. Soil of Kyiv region, constructed by the authors based on the data described in (Atlas of natural conditions and natural resources of the Ukrainian SSR, 1978)

The overall condition of the soils in the Kyiv region is currently deemed satisfactory. Nevertheless, in order to maintain and enhance this status, it is imperative to implement consistent, reliable, and timely soil cover monitoring (State Institution "Soils protection Institution of Ukraine", 2020).

It is important to note that since February 24, 2022, a full-scale war has been unfolding in Ukraine. Consequently, the current statistical data concerning land resources may be subject to change depending on the impact of the ongoing conflict. Therefore, when interpreting the data, the influence of these extraordinary circumstances should be taken into consideration.

### 3. Materials and data

#### *Conducting ground surveys and obtaining factual data on the percentage of SOM*

Agrochemical research is conducted in Ukraine every five years. To date, ten complete rounds of such studies have been completed, and the 11<sup>th</sup> round is currently underway with data still being gathered. The State Institution 'Soils Protection Institute of Ukraine' carries out agrochemical studies on soils and produces accompanying reports (State Institution "Soils protection Institution of Ukraine", 2020). In our study, we utilized data on the humus content, also known as Soil Organic Matter (SOM), from the 5<sup>th</sup> to the 10<sup>th</sup> rounds of investigation. It is pertinent to note that the 5<sup>th</sup> round spanned the years 1986-1990, the 6<sup>th</sup> round covered 1991-1995, the 7<sup>th</sup> round included 1996-2000, the 8<sup>th</sup> round ranged from 2001-2005, the 9<sup>th</sup> round incorporated 2006-2010, and the 10<sup>th</sup> round covered 2011-2015. Thus, the SOM percentage represents the average value over five years for each round, broken down by district. During the 5<sup>th</sup> round, 1094.8 thousand hectares of agricultural land were surveyed, with an average SOM percentage of 2.7% for the entire Kyiv region. In the 6<sup>th</sup> round, the surveyed area contracted to 770.8 thousand hectares, and the average SOM decreased to 2.6%. For the 7<sup>th</sup> round, the surveyed area expanded to 939.3 thousand hectares while the average SOM remained at 2.6%. In the 8<sup>th</sup> round, the surveyed area began to shrink again, reaching 843.2 thousand hectares, with a SOM of 2.87%. This decrease continued into the 9<sup>th</sup> round (793.2 thousand hectares, SOM – 2.9%) and the 10<sup>th</sup> round (761.4 thousand hectares, SOM – 2.98%) (Romanova et al., 2018; Yacuk and Balyk, 2013). As a result, the SOM data for each district were extracted from statistical reports provided by the State Institution 'Soils Protection Institute of Ukraine.' This information can also be found in the published article by (Romanova et al., 2018).

The State Institution 'Soils Protection Institute of Ukraine' collects soil samples from March to May. Following this, laboratory tests are conducted and the SOM data is calculated as an average value spanning five years. Consequently, we selected satellite images for the months of March through May from 1986 to 2015. It is important to note that many of the images from May display vegetation. This article proposes a method for estimating the predicted SOM indicator based on spectral characteristics. To ensure the reliability of the study, we decided to focus on a single soil type, specifically chernozems, which predominate in the southern part of the region, as depicted in Figure 2. The 15 selected districts from the Kyiv region, along with their corresponding humus content (SOM) in the soils and the soil types, were identified by the authors and are presented in Figure 3.

Drawing from empirical data (Romanova et al., 2018), a significant increase in SOM over time was reported, with marked alterations especially observable in the southern part of the region. Despite the ongoing intensification of agricultural land use in the region, it is evident that the SOM data can be enhanced through the application of various organic fertilizers. Specifically, within the Kyiv region, the incorporation of by-products and residual vegetable matter from agricultural crops is identified as the primary mechanism for replenishing SOM. Romanova's research provides further validation, confirming the annual increase in organic contributions. The second factor contributing to the rise in SOM within the region is the strategic removal, initiated in 2001, of highly eroded and unproductive lands with low SOM content from active cultivation. The area of the investigated lands decreased by 30.5% (from 1986 to 2015). SOM content in soils of Kyiv region according to (Romanova et al., 2018) by survey rounds (1986-2015) in 15 selected districts shown in Table A1 (Appendix A).

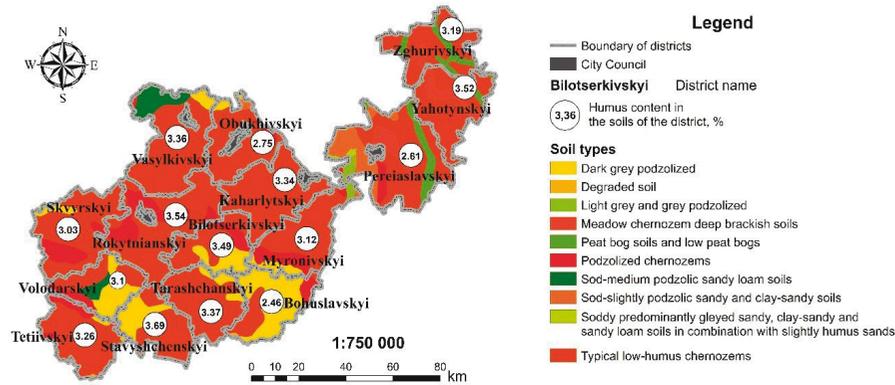


Figure 3. Humus content (SOM) in selected 15 districts of Kyiv region (10<sup>th</sup> round)

Satellite data

Landsat data

In this study, we utilized data from the Landsat 5 TM (“USGS Landsat 5 Level 2, Collection 2, Tier 1” collection in GEE) and Landsat 8 OLI (“USGS Landsat 8 Level 2, Collection 2, Tier 1” collection in GEE) satellites, both of which are publicly accessible as remote sensing materials. As the ground-based measurements for the SOM data were collected between 1986 and 2015, we ensured that the satellite data also spanned this time interval. Among the Landsat satellite system, three missions, specifically Landsat 5, 7, and 8, were operational during this period. However, it is worth noting that data acquired after May 31, 2003, from these missions may contain data gaps (Landsat Missions, 2021).

For the 5<sup>th</sup> through 9<sup>th</sup> rounds (1986-2010), we exclusively used data from Landsat 5 TM. The 10<sup>th</sup> round, which spanned the interval from 2011 to 2015, was covered by data from both Landsat 5 TM (which operated until 2012) and Landsat 8 OLI (which has been operational since 2013). When data from the Landsat 5 TM were loaded for what were considered the most favorable periods - March and May of 2011 and 2012, the resulting images were found to be of poor quality. This could possibly be due to the deterioration of the detectors as the Landsat 5 TM satellite mission approached its end. As a result, the decision was made to use data from the Landsat 8 OLI satellite. The Landsat 8 OLI data offers a spatial resolution of 30 meters in the visible, near-infrared (NIR), and shortwave infrared (SWIR 1 and SWIR 2) ranges, and it has a 16-day temporal resolution (Landsat Missions, 2021).

The number of images that were used for conducting the research in each of the rounds are displayed in Table 1.

Table 1. The number of Landsat images used.

No round of investigation	Investigation time	Satellite	Bands	Number of images
5 <sup>th</sup>	1986-1990	Landsat 5 TM	1-5, 7	160
6 <sup>th</sup>	1991-1995	Landsat 5 TM	1-5, 7	117
7 <sup>th</sup>	1996-2000	Landsat 5 TM	1-5, 7	126
8 <sup>th</sup>	2001-2005	Landsat 5 TM	1-5, 7	72
9 <sup>th</sup>	2006-2010	Landsat 5 TM	1-5, 7	109
10 <sup>th</sup>	2011-2015	Landsat 8 OLI	2-7	102

Satellite elevation and climate data

Elevation and climate data from GEE were utilized to construct a regression model in this study. The ‘NASA SRTM Digital Elevation 30 m’ collection served as the source for terrain slope data, expressed in degrees. Additionally, the ‘TerraClimate: Monthly Climate and Climatic Water Balance for Global Terrestrial Surfaces, University of Idaho’ collection provided data on precipitation accumulation, measured in millimeters, as well as soil moisture, also measured in millimeters, which was derived using a one-dimensional

soil water balance model. Temperature data (°C) were also sourced from this collection.

Satellite Data Preprocessing in GEE

Landsat 5 TM and Landsat 8 OLI data can be freely downloaded from GEE (Google Earth Engine, 2021). Hence, the study employs the GEE cloud computing platform, offering powerful tools for processing remote sensing data and conducting geoinformation analysis. Using GEE mitigates challenges associated with loading, storing, and processing satellite data at different times. This issues commonly encountered when dealing with large volumes of spatial data (Mateo-García et al., 2018; Shetty Ai., Umesh and Shetty A., 2021). To access GEE over the internet, a Google account is required for coding operations. The data catalog houses a vast repository of publicly available geospatial datasets. These incorporate observations from a variety of satellite and aerial imaging systems spanning both optical and non-optical wavelength ranges. The catalog also includes environmental variables, weather and climate data, land cover, topographic information, and socio-economic data. All this data is preprocessed and made available in a ready-to-use format (Gorelick et al., 2017). The GEE platform comprises two principal components that function in tandem - the Google Earth Engine Explorer, which is used for viewing datasets, and the Google Earth Engine Playground. The Google Earth Engine Playground application employs a JavaScript API to download and analyze extensive satellite images, as well as to execute complex geostatistical and geospatial operations (Shetty Ai., Umesh and Shetty A., 2021).

In GEE Landsat 5 TM and Landsat 8 OLI data were downloaded from Collection 2 Level-2 already with surface reflectance values. That is, they are atmospheric corrected. Landsat 5 TM and Landsat 8 OLI surface reflectance data (Level 2 Science Product - L2SP) were generated using the Land Surface Reflectance Code algorithm. This algorithm converts pixel values to Top of Atmosphere (TOA) Reflectance and Brightness Temperature values using calibration parameters from metadata. Then the TOA Reflectance values are subjected to atmospheric correction using additional data (Landsat 8 Collection 2 (C2) Level 2 Science Product (L2SP) (Guide, 2020; Vermote et al., 2016)

The subsequent step in processing Landsat data in GEE involves masking clouds and their shadows. Pixels obscured by cloud cover and cloud shadows do not provide useful information about the Earth’s surface. Therefore, their inclusion in data processing would lead to deliberately inaccurate results. GEE masks cloud pixels according to the C function of the mask (CFmask) algorithm using data from the QA\_PIXEL (Pixel Quality Assessment) band. The CFmask algorithm can be found, for example, in the study (Joshi, Wynne and Thomas, 2019; Zekoll et al., 2021).

Finally, employing a date filter, mosaic Landsat images were created for the months of March through May for rounds 5<sup>th</sup> through 10<sup>th</sup>, encapsulating 15 months across five years for each round. These mosaic Landsat images consist of images with atmospheric corrected pixel values (surface reflectance) that have been averaged over all scenes of this satellite covering the 15 districts of the Kyiv region for the designated time interval, with clouds and their shadows masked.

## 4. Methods

### Snow masking

Images from March and May were considered for this study because during this time, the snow in the area would have already melted and the primary growing season for most plants would not have begun yet. Extensive vegetation on the soils can lead to reflections that distort the soil's spectral curve. To account for any areas where snow might still be present, a trait common in March and April for this latitude, the Normalized Difference Snow Index (NDSI) was utilized. We employed the NDSI, as described in (Riggs, Hall, and Salomonson, 1994) to mask areas that might still be covered with snow. This specific threshold was established after examining the satellite data for the study area. The calculation of the NDSI was conducted using the following formula:

$$\text{NDSI} = \frac{\text{Green} - \text{SWIR1}}{\text{Green} + \text{SWIR1}}, \quad (1)$$

where *Green* and *SWIR 1* are the pixel values in Landsat's bands (2 and 5 for Landsat 5 TM and 3 and 6 for Landsat 8 OLI, respectively).

The threshold  $\text{NDSI} < 0$  was utilized for snow masking, with positive NDSI values indicating areas covered with snow. The mean spectral reflectance (MSR) value was obtained for pixels unaffected by snow, clouds, and cloud shadows. Subsequently, the resulting mosaic was cropped to the precise area of the polygonal feature representing the Kyiv region. Finally, the processed mosaic was uploaded to the GEE platform from the Humanitarian Data Exchange (The Humanitarian Data Exchange, 2021).

### Determination of bare soils from remote sensing data

According to (Romanova et al., 2018) soil samples were collected from bare soils in the Kyiv region for further determination of SOM. Consequently, the subsequent step involved developing a method to isolate bare soils using the mosaic of Landsat satellite data. Figure 2 indicates that chernozem soils, which are the focus of this article's spectral analysis, are predominantly located in the south and southeast of the Kyiv region. Therefore, for the upcoming

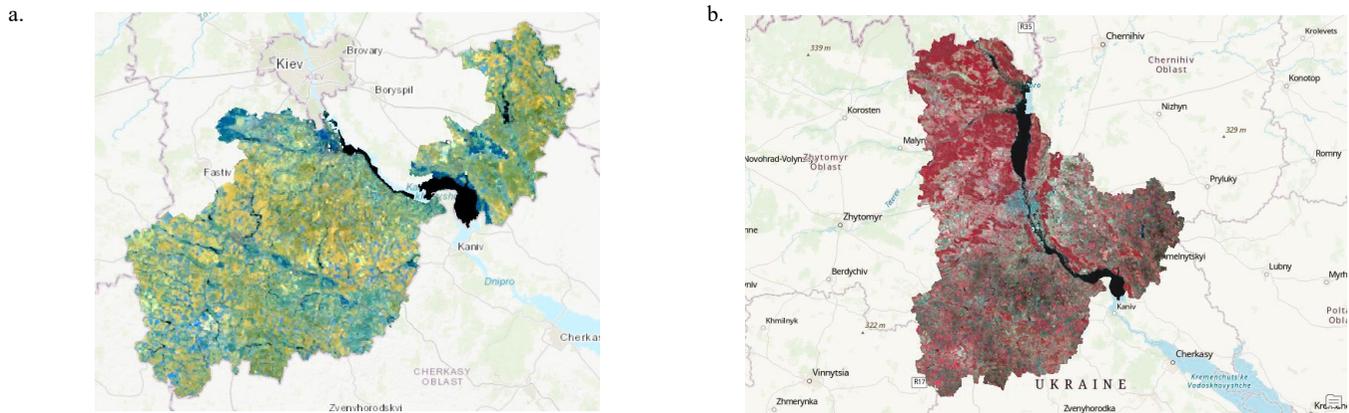
investigation, the corresponding 15 districts of the Kyiv region, as depicted in Figure 1 and Figure 3 were specifically chosen. The polygon data for these districts were also obtained from the Humanitarian Data Exchange (The Humanitarian Data Exchange, 2021).

The conventional approach for identifying different types of land use, known as LULC (Land Use/Land Cover), involves satellite data classification. Both supervised and unsupervised classification methods are commonly employed, as described in the scientific literature (Mekuriaw, Cherinet and Tsegaye, 2021; Olthof and Fraser, 2014). Nevertheless, the classification method is rather cumbersome and can be effectively used to distinguish all types of LULCs typical for a given territory. In our study, we need to select only bare soils, and we are not interested in selecting other LULC types. Therefore, it is advisable to use a technique based on the reflective spectral properties of bare soils, making it possible to decipher this particular type of LULC. These data for the first seven bands of Landsat 8 OLI in GEE are shown in Figure 4.

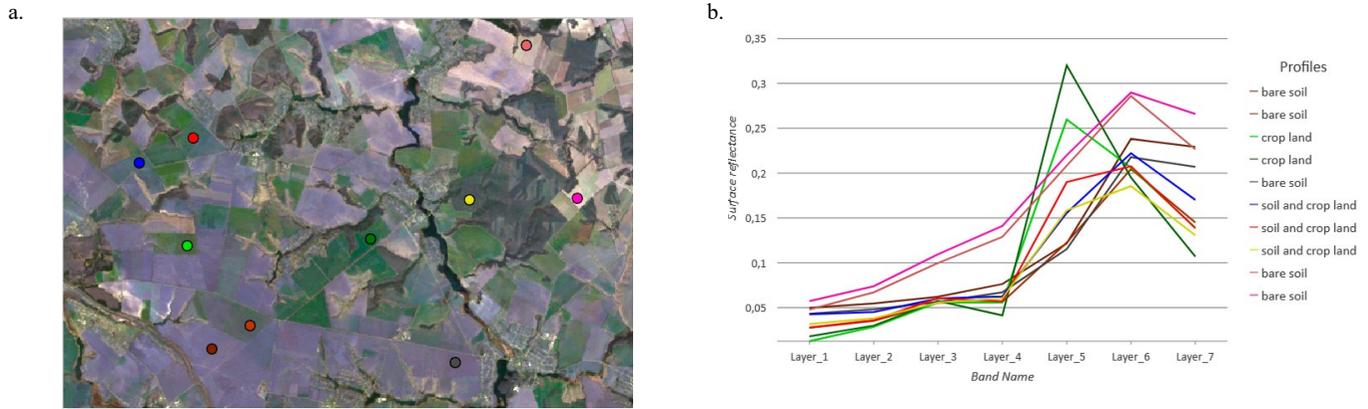
The spectral reflectance properties of soils in the study area were examined by analyzing spectral profiles and using different combinations of bands from the Landsat 8 OLI satellite. (Figure 5).

By analyzing Figure 5a and Figure 5b, it is possible to determine the color representation of bare soils in composite R-G-B = 4-3-2 and decipher them. It is important to consider the soil property described in (Shikhov et al., 2020), as mentioned earlier, which states that soils of different genetic affiliations exhibit a monotonic increase in reflection coefficients as wavelength increases in the range from 0.4 to 2  $\mu\text{m}$ . In our study, this corresponds to the B1-B6 band of Landsat 8 OLI. By comparing the spectral profiles of different bare soils (Figure 5a) and overlaying a thematic map of soils in the Kyiv region (Figure 2) onto Figure 5a it can be concluded that the last two spectral profiles in the "Profiles" list do not correspond to chernozem soils. Therefore, for further analysis, they should be masked.

Based on the findings from previous studies regarding the spectral reflectance properties of chernozems and the obtained spectral profiles, the authors chose to utilize specific threshold values for vegetation indices to isolate bare soils chernozems, as presented in Table 2. Additionally, the Built-Up Area Extraction Index (BAEI) was employed to separate bare soils from built-up areas.



**Figure 4.** Landsat 8 OLI satellite data mosaic plotted in GEE for the month's March-May for round 10<sup>th</sup> (2011-2015): a - for the territory of the Kyiv region (False Color Composite R-G-B = 5-4-3), b - for the territory of the selected 15 districts of the Kyiv region (Composite R-G-B = 7-6-5)



**Figure 5.** The determination of the LULC type is indicated by points (a), and the corresponding spectral profiles (b) are provided. A fragment of the Landsat 8 OLI image for the study area is presented as a composite with the bands R-G-B represented by 4-3-2.

**Table 2.** Threshold values of vegetation indices for bare soils detection.

Vegetation indices	Threshold values	
	From	To
NDVI	0.1	0.35
NDBI	0	0.45
NDWI	-0.5	0
MSAVI <sub>2</sub>	0	0.3
BAEI	< 1.3	
NDSI	< 0	

The vegetation indices were calculated using the formulas provided below.

NDVI differentiates green vegetation from other surfaces, and elevated NDVI values indicate a high presence of leaf biomass, dense canopy coverage, or a large leaf area (Jasinski, 1990; Tucker, 1979; Wilson and Sader, 2002):

$$NDVI = \frac{NIR - Red}{NIR + Red} \quad (2)$$

NDBI (Normalized Difference Built-Up Index) is utilized to emphasize built-up areas while mitigating variations in surface illumination and atmospheric effects (Zha, Gao and Ni, 2003):

$$NDBI = \frac{Band_6 - Band_4}{Band_6 + Band_4} \quad (3)$$

Normalized Difference Water Index (NDWI) was introduced as a novel index for remote sensing applications, specifically for detecting vegetation liquid water and assessing the spectral reflectance properties of soils (Gao, 1996):

$$NDWI = \frac{\rho_{0.86} - \rho_{1.24}}{\rho_{0.86} + \rho_{1.24}} = \frac{NIR - SWIR1}{NIR + SWIR1} \quad (4)$$

where  $\rho$  is surface reflectance value at the wavelength specified in  $\mu m$ ;

MSAVI<sub>2</sub> (Modified Soil Adjusted Vegetation Index) expands the dynamic range of the vegetation signal, thereby reducing the impact of soil background and enhancing the sensitivity of vegetation. This sensitivity is determined by the ratio of the vegetation signal to the soil noise (Qi et al., 1994):

$$MSAVI_2 = \frac{2 \cdot NIR + 1 - \sqrt{(2 \cdot NIR + 1)^2 - 8 \cdot (NIR - Red)}}{2} \quad (5)$$

Built-Up Area Extraction Index (BAEI) is calculated by the equation (Bouzekri et al., 2015):

$$BAEI = \frac{Red + 0.3}{Green + SWIR1} \quad (6)$$

Vegetation indices were computed based on the average spectral reflectance of Landsat 5 TM and Landsat 8 OLI, following the formulas -. Using the predetermined threshold values of vegetation indices (Table 2) and the polygons representing the territories of city councils acquired from (The Humanitarian Data Exchange, 2021), a binary raster mask was generated. In this mask, a value of 1 indicates bare soils chernozems, while 0 represents other types of LULC. The generated mask was then applied to the previously obtained MSR values. As a result, images of bare soils were obtained, where the pixels retain the MSR values.

### 5. Results

To establish the statistical relationship between the MSR value obtained from Landsat and SOM, it was decided to employ a multiple linear regression (MLR) model. Since each of the 15 selected districts in the Kyiv region is characterized by a single the SOM value, this value should correspond to a single mean spectral reflectance value from Landsat for each band. Consequently, within GEE, zonal statistics were utilized to calculate the average MSR values for the seven selected Landsat bands, the average values of vegetation indices for each area, as well as additional data on relief and climate factors such as slope, precipitation, soil moisture, and temperature.

Indeed, MLR enables the assessment of the individual impact of each factor and their collective influence on the target variable, in this case, the SOM data. MLR allows for the determination of the degree of influence that each factor exerts on the productive trait and provides insights into the combined effect of these factors on SOM. By analyzing the MLR model coefficients and statistical significance, we can understand the relative importance and contribution of each factor in explaining the variations in SOM. In the MLR model, SOM was considered as the dependent variable. As predictors or regressors in the MLR analysis, we utilized spectral bands with wavelengths common to Landsat 5 TM and Landsat 8 OLI, which are associated with the spectral reflectance properties of soils. Additionally, vegetation indices capable of detecting bare soils were employed as predictors. Furthermore, topographic data, specifically slope, and climatic data including temperature, precipitation, and soil moisture, were also included as predictors in the model (Table 3).

**Table 3.** Predictors of MLR model

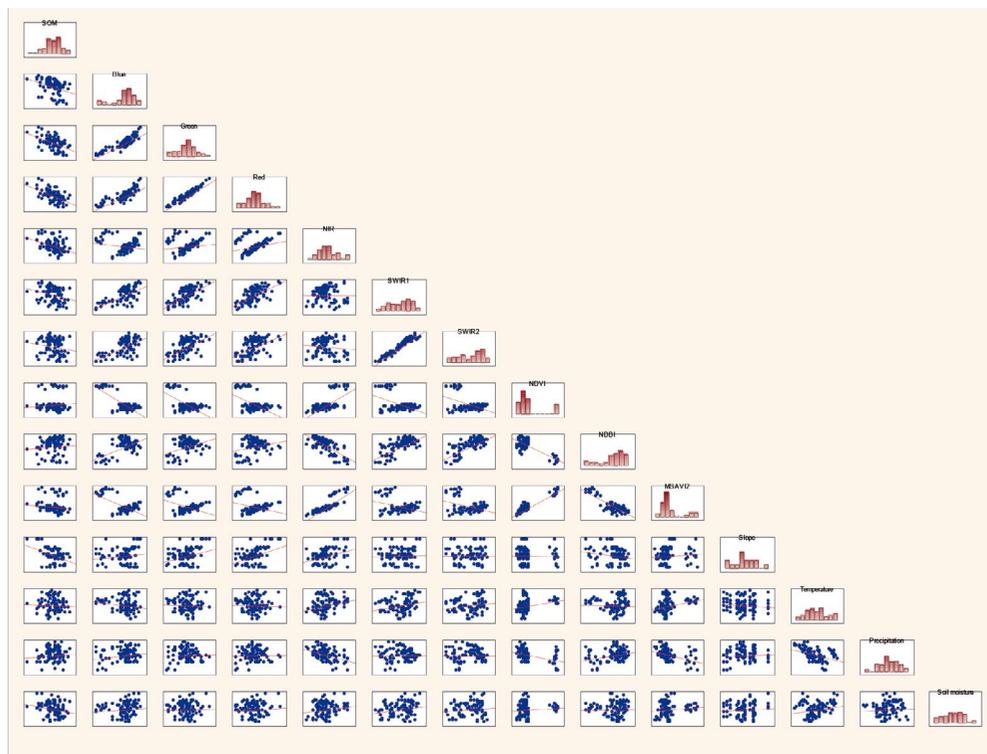
<b>Predictors, <math>x_i</math></b>	$x_1$	$x_2$	$x_3$	$x_4$	$x_5$	$x_6$	$x_7$
Values, MSR	Blue	Green	Red	NIR	SWIR 1	SWIR 2	NDVI
<b>Predictors, <math>x_i</math></b>	$x_8$	$x_9$	$x_{10}$	$x_{11}$	$x_{12}$	$x_{13}$	
Value, MSR	NDBI	MSAVI2	Slope	Temperature	Precipitation	Soil moisture	

The  $\overline{\text{MSR}}$  value in Table 3 indicates the MSR obtained from Landsat 5 TM and Landsat 8 OLI for each of the six selected bands, band combinations, and vegetation indices. These MSR values were calculated within the 15 districts of interest. In each of the six rounds of ground surveys (5<sup>th</sup>-10<sup>th</sup> rounds), there is one averaged value of SOM. Considering that ground surveys were conducted in 15 districts of the Kyiv region, each MLR regressor model consists of 90 values (6 rounds  $\times$  15 districts = 90 values). These input data for the MLR calculations are presented in Table A2 in Appendix A. Afterwards,

a Pearson's correlation analysis was conducted to examine the relationships between the variables used in the MLR model. The Pearson's correlation coefficient (Pearson's  $r$ ) was employed to evaluate the strength and direction of the linear association between the variables (Mendenhall and Sincich, 2013). The results of the Pearson correlation analysis between  $\overline{\text{MSR}}$  and SOM values are presented in Table 4. Marked \* correlations are significant at  $p < 0.05$ . For clarity, the data from Table 4 are shown on Figure 6.

**Table 4.** Correlation matrix model variables

SOM	*	*	*	*	*				*				*
-0.37	Blue	*	*		*	*	*	*	*				
-0.51	0.90	Green	*		*	*	*	*	*	*			
-0.56	0.84	0.96	Red	*	*	*	*	*	*	*			
-0.40	-0.20	0.13	0.26	NIR			*	*	*	*	*	*	*
-0.28	0.67	0.68	0.62	-0.01	SWIR 1	*	*	*	*		*		*
-0.18	0.61	0.59	0.50	-0.11	0.97	SWIR 2	*	*	*				
0.08	-0.80	-0.60	-0.52	0.69	-0.46	-0.46	NDVI	*	*		*	*	*
0.08	0.61	0.38	0.24	-0.72	0.70	0.76	-0.81	NDBI	*			*	*
-0.10	-0.65	-0.38	-0.27	0.86	-0.34	-0.38	0.96	-0.84	MSAVI2		*	*	*
-0.41	0.15	0.32	0.42	0.33	0.04	-0.01	-0.03	-0.21	0.10	Slope			
-0.04	-0.18	-0.05	-0.03	0.29	0.26	0.19	0.29	-0.03	0.31	-0.11	Temp.	*	*
0.11	0.20	0.05	0.02	-0.42	0.00	-0.01	-0.39	0.30	-0.43	0.07	-0.43	Precip.	
-0.23	0.12	0.03	0.08	0.18	0.23	0.13	0.10	0.03	0.14	0.08	0.24	0.07	Soil mois.



**Figure 6.** Scatterplot matrix for correlation between model variables

The analysis of Table 4 reveals a notable and statistically significant negative relationship between SOM and  $\overline{MSR}$  in the Green and Red bands ( $r \approx -0.5 \div -0.6$ ). There is also a slightly weaker negative relationship between SOM and  $\overline{MSR}$  in the Blue and NIR bands

( $r \approx -0.4$ ). Additionally, an inverse relationship is observed between SOM and  $\overline{MSR}$  in the SWIR 1 band ( $r \approx -0.28$ ). Regarding the climatic data and topography, the analysis indicates a significant negative relationship between SOM and Slope, as well as between SOM and Soil moisture. However, the correlations between SOM and the other  $\overline{MSR}$  variables are relatively weak and not statistically significant at  $p < 0.05$ .

At the initial stage, a total of 13 predictors for the MLR model were selected. Subsequently, employing the method of least squares, the MLR coefficients and other regression parameters were calculated. In our analysis, it was employed two methods to determine the most significant variables: the Forward stepwise method and the Backward stepwise method. The Forward stepwise method involves iteratively adding or removing independent variables from the model at each regression step based on the results of the F test. Variables are added to the model if they contribute significantly to improving the regression model's fit. This process continues until a better regression model is achieved. On the other hand, the Backward stepwise method removes independent variables from the regression equation one at a time, again based on the F test results. Variables are eliminated from the model if their inclusion does not significantly improve the fit of the regression model. This process continues until a better regression model is obtained. The essence of the F test in testing the significance of MLR is to compare two variances: model-explained

( $MSR$ ) and unexplained ( $MSE$ ) (Mendenhall and Sincich, 2013). If  $MSR \approx MSE$ , then the regression is insignificant (the constructed model does not allow explaining the behavior of the predicted value depending on the the predictors' values). If  $\frac{MSR}{MSE} > F_{\alpha}$ , then the regression is significant. Two hypotheses are considered:  $H_0^{(1)}$  – regression is not significant,  $H_1^{(1)}$  – regression is significant, which is what it is used for F test:

$$F = \frac{MSR}{MSE} \quad (7)$$

Next, the significance is checked  $F$  (value is calculated  $F_0$ ) through the F-distribution (Fisher-Snedecor distribution) for the obtained values  $F$ ,  $DFR$ ,  $DFE$ . If,  $F_0 > \alpha$  where  $\alpha = 0.05$  is significance level, then the hypothesis  $H_0^{(1)}$ , is accepted, otherwise, the hypothesis  $H_1^{(1)}$  is accepted.

For a detailed step-by-step procedure of the Forward stepwise and Backward stepwise methods, please refer to Appendix B. As a result of these methods, we obtained two models, and their respective parameters are presented in Table 5.

In order to determine the best model among the two obtained, we utilized the Akaike Information Criterion (AIC) as proposed by (Akaike, 1973). The AIC serves as a heuristic measure that aims to balance the reduction in the number of model parameters and the quality of model fit. According to this criterion, the model with the smaller AIC value should be selected, irrespective of its absolute value.

**Table 5.** The resulting regression models obtained through the Forward stepwise and Backward stepwise methods

Model of Forward stepwise method (Model I)						
Multiple $R = 0.651$ , $R^2 = 0.424$ , Adjusted $R^2 = 0.397$ , $F(4,85) = 15.667$ , $p\text{-value} = 1.2E-9 < \alpha = 0.05$ , Std. Error of estimate, $\sigma_y = 0.25655$						
	Std. regression coefficient, $\beta^*$	Std.Err. of $\beta^*$	Accurate estimates of model parameters, $\beta$	Std.Err. of $\beta$	t Stat	p-value
Intercept			6.345	0.455	13.956	1.055E-23
Red	-0.437	0.092	-22.055	4.630	-4.763	7.760E-6
NIR	-0.210	0.089	-4.883	2.072	-2.357	0.021
Soil moisture	-0.151	0.084	-0.004	0.002	-1.802	0.075
Slope	-0.143	0.094	-0.072	0.047	-1.523	0.131
Model of Backward stepwise method (Model II)						
Multiple $R = 0.631$ , $R^2 = 0.398$ , Adjusted $R^2 = 0.384$ , $F(2,87) = 28.763$ , $p = 2.6E-10 < \alpha = 0.05$ , Std. Error of estimate, $\sigma_y = 0.25932$						
Intercept			1.946	0.241	8.080	3.40E-12
NDVI	2.286	0.305	17.517	2.337	7.496	5.166E-11
MSAVI2	-2.296	0.305	-32.701	4.345	-7.527	4.482E-11

The obtained values for the AIC criterion are as follows:

$$AIC_1 = -240.023$$

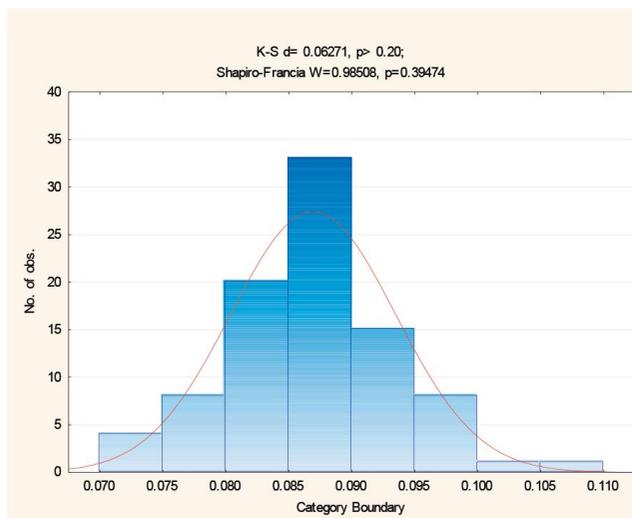
$$AIC_1 = -239.994$$

Based on the selection process, Model I derived from the Forward stepwise method was chosen for predicting SOM. To assess the normality of the values of the model regressors and the residuals, we conducted the Shapiro-Francia test and the Kolmogorov-Smirnov test. The Shapiro-Francia test, which is an extension of the Shapiro-Wilk test suitable for samples larger than 50, was used for this purpose. These tests help us determine if the values of the model

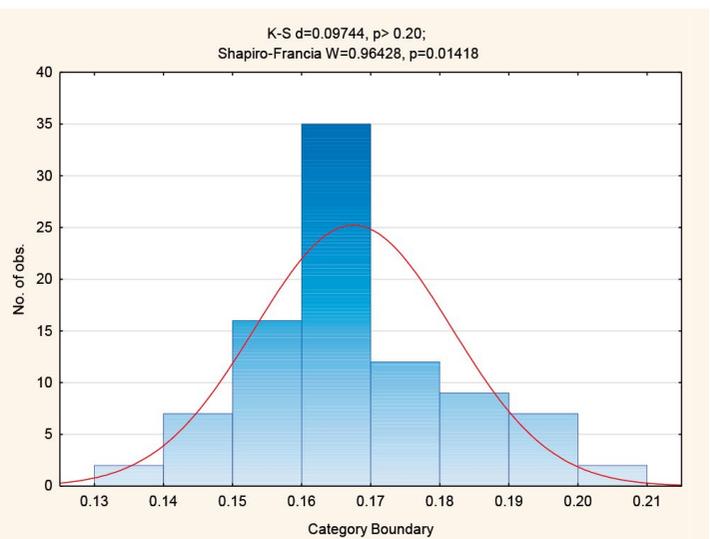
regressors and the residuals follow a normal distribution (Shapiro and Francia, 1972). The results of the regressor test are shown in the following Figure 7.

The values of all four regressors in Model I are normally distributed at a significance level of  $\alpha = 0.001$ , as  $\forall p\text{-value}: p \geq \alpha$ . As a limitation of our model at this stage, it should be noted that the NIR and Slope regressors will not be normally distributed if we consider a significance level of  $\alpha = 0.01$ . During the assessment of the residuals' normality using the Shapiro-Francia test and the Kolmogorov-Smirnov test, the results depicted in Figure 8 were obtained. It can be observed that the residuals are normally distributed at a significance level of  $\alpha = 0.001$  ( $p = 0.951 > 0.001$ ). The effect size KS - D, which measures the difference between the sample distribution and the normal distribution, is very small, specifically 0.04269. This suggests that the deviation from normality is negligible.

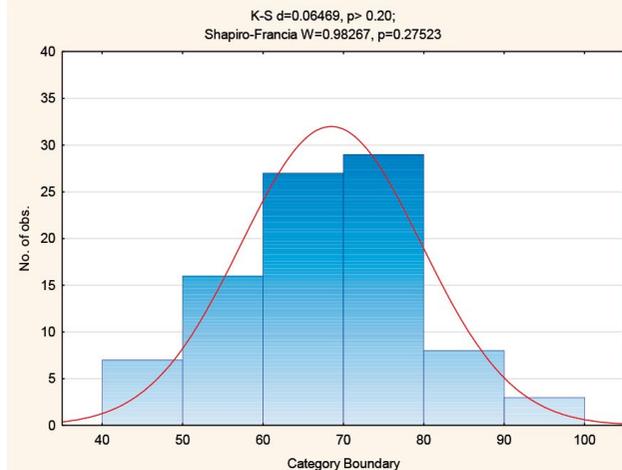
a - Red



b - NIR



c - Soil moisture



d - Slope

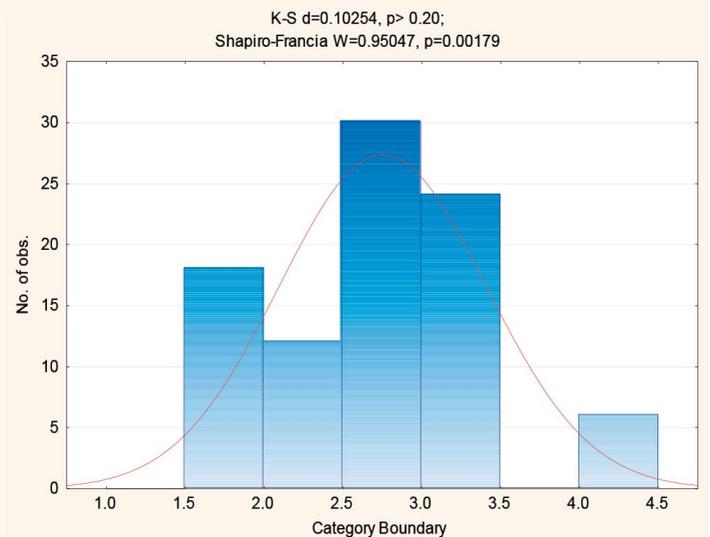


Figure 7. Normality check histograms of regressors values

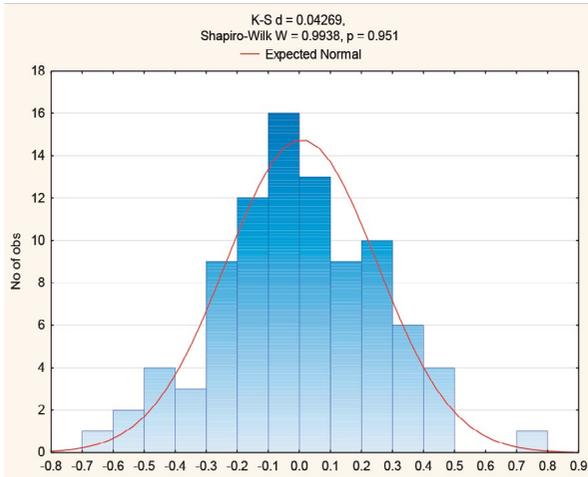


Figure 8. Normality check histograms of residuals

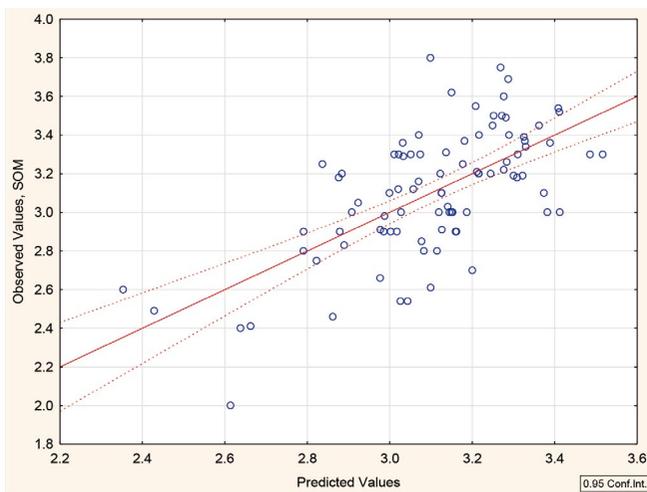


Figure 9. Predicted vs. Observed Values

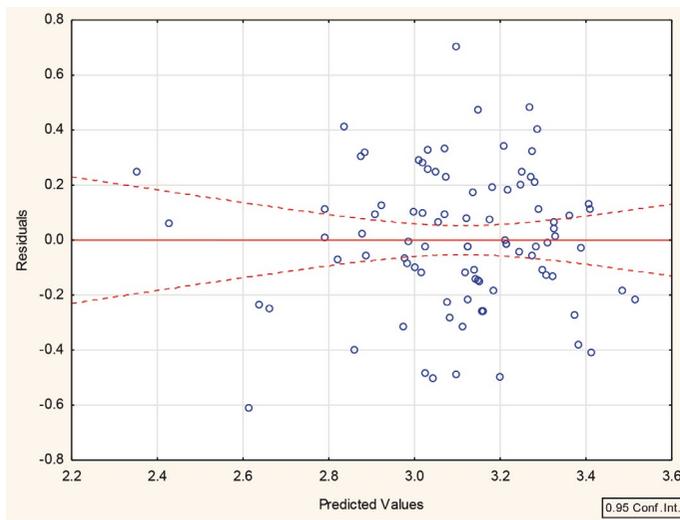


Figure 10. Predicted vs. Residual Scores

To test our model on heteroscedasticity (the hypothesis  $H_0$  – homoscedasticity, error variances are equal) we used The Breusch-Pagan test (Breusch and Pagan, 1979). The test result showed that there was no heteroscedasticity:

$$X_{calc}^2 = 1.226, X_{table}^2(p = 0.95; df = 9.488, X_{calc}^2 < X_{table}^2)$$

To test our model for multicollinearity, we used the Variance Inflation Factor (VIF) (Mendenhall and Sincich, 2013). At first glance, there is a suspicion of multicollinearity in the model, since according to the Table 5 p-value of t Stat for Soil moisture and Slope greater than  $\alpha = 0.05$  (0.075 and 0.131, respectively). But the VIF calculation showed that there is no multicollinearity:

$$VIF_{Red} = 1.24, VIF_{NIR} = 1.18, VIF_{Slope} = 1.30, VIF_{Soil\ moisture} = 1.03$$

According to (Mendenhall and Sincich, 2013) a severe multicollinearity problem exists if the largest of VIF is greater than 10.

To identify potential outliers, we calculated Cook's distance. All obtained values of Cook's distance are approximately equal and are in the range from 0 to 0.16, which indicates the absence of potential outliers (Mendenhall and Sincich, 2013).

Finally, we have constructed graphs that illustrate the relationship between the Predicted vs. Observed Values (Figure 9) and Predicted vs. Residual Scores (Figure 10). Confidence bands were constructed at a probability level of 0.95.

In Figure 9 and Figure 10 the confidence bands line (red dashed curve) shows the confidence level (0.95), which is the probability that the “true” fitted line (in the population) falls between the bands.

## Discussion

The findings from the constructed MLR model for predicting SOM in the chernozem soils of the Forest-Steppe zone of Ukraine provide significant insights into the region's soil fertility. The application of a combination of reflectance values in six bands of Landsat 5 TM and Landsat 8 OLI satellites, vegetation indices (NDSI, NDWI, NDBI, BAEI, MSAVI, and NDVI), topographic data (slope), and climatic data (temperature, precipitation, soil moisture) as regressors has contributed to the development of an effective predictive model. The Forward stepwise method (Mendenhall and Sincich, 2013) was instrumental in achieving the optimal model, as it allowed for the selection of the most impactful regressors based on F test and Akaike criteria (Akaike, 1973). The model yielded a Multiple R = 0.651 and an  $R^2 = 0.424$  and the most significant predictors of an effective predictive model are RED, NIR, slope and soil moisture. These metrics suggest that the model can explain approximately 42.4% of the variation in the SOM in the soils of the examined regions. The adjusted  $R^2$  of 0.397 further supports this and confirms that the model maintains a decent level of predictive power even when adjusting for the number of predictors in the model. Fairly similar results were obtained in the study (Yu et al., 2021) where the accuracy of the model was improved by adding the same environmental factors and spectral combination. However, the estimation effect of back propagation neural network (BPNN) model (with  $R^2$  being 0.947) was better than that MLR model (with  $R^2$  being 0.410). Similar to the findings of this study (Forkuor et al., 2017; Wang and Ge, 2012) also recorded elevation, slope, RED and NIR as the most important variable influencing soil indicators. According to the study (Luo et al., 2023) adding the multiyear average temperature effectively improves the SOM mapping performance. The authors used Sentinel images (10 m) and the random forest algorithm, which obviously affects the accuracy of the model. Some studies (Lu, Liu and Liu, 2022) suggest that the environmental variables (mainly clay index) have the greatest influence on the prediction accuracy of SOM. Using time series 7 MOD09A1 images (Zhang et al., 2021) noted that the pixel dates of the training samples and precipitation data were the main factors controlling the model performance. The simultaneous combination of dead fuel index (DFI), NDVI and environmental factors improves the SOM mapping performance (Yang et al., 2021) especially in infertile soils. Undoubtedly, the accuracy of SOM mapping performance will be affected by the spatial resolution of the images. In our study, the period 1986-2015 was selected, which is decisive for why we used the Landsat system.

Given that the RED and NIR bands emerged as the most significant spectral predictors in our analysis, it would be valuable to compare our findings with the results of Ukrainian scientists who have studied the same type of soils.

According to the study (Chornyy and Abramov, 2016; Achasov and Bidolah, 2008; Chatochin and Lindin, 2001; Sakhatsky, 2008; Truskavetsky, 2006; Byndych, 2021) the humus content (SOM) is most linearly related to RED, NIR and GREEN band. In the study (Abramov, 2012) author built a parabolic relationship between spectral index RED/NIR and the humus content (SOM). Long-term studies of soils in Transcarpathia made it possible to conclude that the best relationship was traced between the humus content (SOM) and the spectral index NIR/RED (Hebryn-Baidy, 2017). According to study (Truskavetsky et al., 2015) there is a linear relationship between SWIR1,2 bands and the humus content (SOM). The fact that in the listed studies authors used exclusively spectral information is interesting. In our study, we used environmental factors in addition to spectral data. It should be noted that the mentioned authors did not specify the correlation between the bands. According to our research, for the selected area, there existed a high positive correlation between the GREEN and RED bands ( $R = +0.96$ ). Therefore, including both of them simultaneously in the model would yield an incorrect result. This increased the accuracy of our research to some extent. Some studies prove that there is a relationship between SOM and other bands. Based on the MLR analysis, the authors (Ahmed and Iqbal, 2014) establish a significant relationship ( $R^2 = 0.55$ ) between SOM and spectral reflectance in the BLUE, Thermal and Mid-Infrared bands. Whereas in our study, the BLUE band showed worse significance, and others band were not considered. In the study (Demattê et al., 2007; Fiorio and Demattê, 2009; Reis et al., 2021), authors demonstrate that SOM, clay and sand are the most important attributes that can be predicted by MLR. It is important to note that the validation procedure compared the content of each soil attribute determined by the spectral models with laboratory traditional analysis. In our study, the accuracy of the obtained models was evaluated exclusively on the basis of mathematical calculations. A noteworthy approach is shown in the study (Bouasria et al., 2020), authors increase the resolution of the image (from 30 m to 15 m) by pansharpening method. Thus, the significance of the relationship between SOM and spectral reflectance in the multispectral range increases.

It should be noted that the research period and soil conditions vary depending on the natural and geographical location of study area. The authors also note that the soil should be in an air-dry state, the surface roughness should be minimal, and there should be no vegetation on the soil surface (Gopp et al., 2017; Gopp et al., 2019; David, 2013; Wang and Ge, 2012; Wang et al., 2022; Savin et al., 2021). These conditions were met in our study by using NDSI, NDWI, NDVI, NDBI, and BAEI masks and climatic data. Stable relationships between SOM and spectral reflectance exist only at medium and high content of SOM (1-5%) (Sullivan et al., 2005; Reis et al., 2021). Indicators of SOM in our study fluctuate within the limits from 2.69% to 3.69%. It is also worth noting that the sample size can affect the accuracy of the model. However, this factor is worth a separate study. Additional factors not considered in this study, such as other soil components, the impact of human activity, and biodiversity changes, could further refine the model's predictive capacity. Also, the model was built and tested only in the Kyiv region of Ukraine, primarily on chernozem soils, which may limit its applicability to other soil types and regions. In conclusion, the model serves as a useful tool for understanding and predicting SOM in chernozem soils. With the continuous advancements in satellite technology and increasing availability of detailed environmental data, refining this model for higher accuracy and broader applicability is a promising avenue for future research.

## Conclusion

This article focuses on constructing a MLR model to predict the SOM in chernozem soils within the Forest-Steppe zone of Ukraine. The statistical data used for analysis consisted of percent values of SOM obtained from six rounds of field studies conducted between 1986 and 2015. The model development was carried out for 15 districts located predominantly on chernozem soils within the Kyiv region, Ukraine.

As regressors in the model, reflectance values from six bands of Landsat 5 TM and Landsat 8 OLI satellites, vegetation indices for bare soil detection, topographic data (slope), and climate data (temperature, precipitation, soil moisture) were employed. The optimal model was obtained through the Forward stepwise method, based on F-test and Akaike criteria. The model exhibited the following metrics: Multiple  $R = 0.651$ ,  $R^2 = 0.424$ , and Adjusted  $R^2 = 0.397$ . Validation of the model demonstrated the absence

of heteroscedasticity, multicollinearity, and potential outliers, indicating its reliability. The most influential predictors in the effective predictive model were identified as RED, NIR, slope, and soil moisture. The developed model can be utilized for predicting the SOM content in chernozem soils based on remote sensing data. Future research efforts could be directed towards exploring the efficacy of nonlinear regression models for predicting SOM in chernozem soils and evaluating the model's performance on other soil types.

## Acknowledgments

The authors would like to express their gratitude to the Editor-in-chief of the journal and the reviewers for their valuable comments. The corrections and suggestions provided by the editor and reviewers have significantly enhanced the scientific value of the article.

The author, Dr. Liliia Hebryn-Baidy, express gratitude to the British Academy and the Council for At-Risk Academics for support to this research through the Researchers at Risk Research Support Grants. Sincere gratitude to the Department of Geography at the University of Cambridge and the Scott Polar Research Institute for their support throughout the research process.

## References

- Abramov, D. (2012). Dynamics of the humus state of dark chestnut soils using remote sensing data. *Scientific works. Ecology*, 206(194), 69-73. (In Ukrainian).
- Achasov, A., & Bidolah, B. (2008). Use of space and terrestrial digital photography to determine the humus content of soils. *Soil science*, 3, 280-286. (In Russian).
- Ahmed, Z., & Iqbal, J. (2014). Evaluation of Landsat TM5 multispectral data for automated mapping of surface soil texture and organic matter in GIS. *European Journal of Remote Sensing*, 47(1), 557-573. <https://doi.org/10.5721/EuJRS20144731>
- Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In: B. Petrov and F. Csake. (Eds). *2nd International Symposium on Informational Theory*. Budapest, Akademiai Kiado.
- Alkharabsheh, H. M., Seleiman, M. F., Battaglia, M. L., Shami, A., Jalal, R. S., Alhammad, B. A., Almutairi, K. F. & Al-Saif, A. M. (2021). Biochar and Its Broad Impacts in Soil Quality and Fertility, Nutrient Leaching and Crop Productivity: A Review. *Agronomy*, 11(5), 993. <https://doi.org/10.3390/agronomy11050993>
- Allohverdi, T., Mohanty, A. K., Roy, P., & Misra, M. (2021). A Review on Current Status of Biochar Uses in Agriculture. *Molecules*, 26(18), 5584. <https://doi.org/10.3390/molecules26185584>
- Atlas of natural conditions and natural resources of the Ukrainian SSR (1978). *Main Directorate of Geodesy and Cartography*. Moscow. (In Russian).
- Bach, E. M., Ramirez, K. S., Fraser, T. D. & Wall, D. H. (2020). Soil Biodiversity Integrates Solutions for a Sustainable Future. *Sustainability*, 12, 2662. <https://doi.org/10.3390/su12072662>
- Baveye, P. C., Schnee, L. S., Boivin P., Laba M., & Radulovich, R. (2020). Soil Organic Matter Research and Climate Change: Merely Re-storing Carbon Versus Restoring Soil Functions. *Frontiers in Environmental Science*, 8, 2296-665. <https://doi.org/10.3389/fenvs.2020.579904>
- Belenok, V., Hebryn-Baidy, L., Bielousova, N., Gladilin, V., Kryachok, S., Tereshchenko, A., Alpert, S., & Bodnar, S. (2023). Machine learning based combinatorial analysis for land use and land cover assessment in Kyiv City (Ukraine). *Journal of Applied Remote Sensing*, 17(1), 014506. <https://doi.org/10.1117/1.JRS.17.014506>
- Belenok, V., Hebryn-Baidy, L., Bielousova, N., Zavarika, H., Sakal, O., & Kovalenko, A. (2022). Geoinformation Mapping of Anthropogenically Transformed Landscapes of Bila Tserkva (Ukraine). *Acta Scientiarum Polonorum. Formatio Circumiectus*, 21(1), 69-84. <https://doi.org/10.15576/ASPFC/2022.21.1.69>
- Belenok, V., Noszczyk, T., Hebryn-Baidy, L., & Kryachok, S. (2021). Investigating anthropogenically transformed landscapes with remote sensing. *Remote Sensing Applications: Society and Environment*, 24, 100635. <https://doi.org/10.1016/j.rsase.2021.100635>

- Bouasria, A., Ibno, N. K., Rahimi, A., & Mostafa, E. (2020). *Soil organic matter estimation by using Landsat-8 pansharpened image and machine learning*. 4rth International Conference On Intelligent Computing in Data Sciences (ICDS). <https://doi.org/10.1109/ICDS50568.2020.9268725>
- Bouzekri, S., Lasbet, A. A., & Lachehab, A. (2015). A new spectral index for extraction of built-up area using Landsat-8 data. *Journal of the Indian Society of Remote Sensing*, 43(4), 867–873. <https://doi.org/10.1007/s12524-015-0460-6>
- Breusch, T. S., & Pagan, A. R. (1979). A simple test for heteroscedasticity and random coefficient variation. *Econometrica*, 47, 1287–1294.
- Byndych, T. Y. (2021). *Diagnostic and parameterization of lateral soil heterogeneity based on multispectral space scanning data*. [Thesis for a Doctor of Science degree], Kharkiv. (In Ukrainian).
- Chatohin, A., & Lindin, M. (2001). Conjugate study of Donbass chernozems by land and remote sensing methods. *Soil science*, 9, 1037-1044. (In Ukrainian).
- Chen, X. L., Zhao, H. M., Li, P. X., & Yin, Z. Y. (2006). Remote sensing image-based analysis of the relationship between urban heat island and land use/cover changes. *Remote Sensing of Environment*, 104(2), 133-146. <https://doi.org/10.1016/j.rse.2005.11.016>
- Chornyy, S., & Abramov, D. (2016). The monitoring of southern chernozem soil humus content with using multispectral satellite images Landsat: Spatial and temporal aspects. *Fundamental and Applied Soil Science*, 17(1-2), 22-30. <http://dx.doi.org/10.15421/041602>
- Das, S., Deb, P., Bora, P. K., & Katre, P. (2021). Comparison of RUSLE and MMF Soil Loss Models and Evaluation of Catchment Scale Best Management Practices for a Mountainous Watershed in India. *Sustainability (Switzerland)*, 13(1), 232, 1-22. <https://doi.org/10.3390/su13010232>
- David, J. M. (2013). Twenty five years of remote sensing in precision agriculture: Key advances and remaining knowledge gaps. *Biosystems Engineering*, 114(4), 358-371. <https://doi.org/10.1016/j.biosystemseng.2012.08.009>
- Dematté, J. A. M., Galdos, M. V., Guimarães, R. V., Genú, A. M., Nanni, M. R., & Zullo, J. (2007). Quantification of tropical soil attributes from ETM/LANDSAT-7 data. *International Journal of Remote Sensing*, 28(17), 3813-3829. <https://doi.org/10.1080/01431160601121469>
- Deng, Y., Wu, C., Li, M., & Chen, R. (2015). RNDSI: A ratio normalized difference soil index for remote sensing of urban/suburban environments. *International Journal of Applied Earth Observation and Geoinformation*, 39, 40-48. <https://doi.org/10.1016/j.jag.2015.02.010>
- Ding, Y., Zheng, X., Zhao, K., Xin, X., & Liu, H. (2016). Quantifying the Impact of NDVIsoil Determination Methods and NDVIsoil Variability on the Estimation of Fractional Vegetation Cover in Northeast China. *Remote Sensing*, 8(1), 29. <https://doi.org/10.3390/rs8010029>
- EarthExplorer (2021). USGS. <https://earthexplorer.usgs.gov> (last accessed 16.12.2021).
- Fiorio, P. R., & Dematté, J. A. M. (2009). Orbital and laboratory spectral data to optimize soil analysis. *Scientia Agricola*, 66(2), 250-257. <https://doi.org/10.1590/s0103-90162009000200015>
- Forkuor, G., Hounkpatin, O., Welp, G., & Thiel, M. (2017). High Resolution Mapping of Soil Properties Using Remote Sensing Variables in South-Western Burkina Faso: A Comparison of Machine Learning and Multiple Linear Regression Models. *PLoS ONE*, 12(1), e0170478. <https://doi.org/10.1371/journal.pone.0170478>
- Gao, B. C. (1996). NDWI - A normalized difference water index for remote sensing of vegetation liquid water from space. *Remote Sensing of Environment*, 58(3), 257-266. [https://doi.org/10.1016/S0034-4257\(96\)00067-3](https://doi.org/10.1016/S0034-4257(96)00067-3)
- Gasmi, A., Gomez, C., Lagacherie, P., Zouari, H., Laamrani, A., & Chehbouni, A. (2021). Mean spectral reflectance from bare soil pixels along a Landsat-TM time series to increase both the prediction accuracy of soil clay content and mapping coverage. *Geoderma*, 388, 114864. <https://doi.org/10.1016/j.geoderma.2020.114864>
- Gaudutis, A., Jotautienė, E., Miėdažys, R., Bivainis, V., & Jasinskas, A. (2023). Sustainable Use of Biochar, Poultry and Cattle Manure for the Production of Organic Granular Fertilizers. *Agronomy*, 13(5), 1426. <https://doi.org/10.3390/agronomy13051426>
- Google Earth Engine, 2021. <https://earthengine.google.com/> (last accessed 30.11.2021).
- Gopp, N. V., Nechaeva, T. V., Savenkov, O. A., Smirnova, N. V., & Smirnov, V. V. (2019). Effect of Slope Mesorelief on the Spatial Variability of Soil Properties and Vegetation Index Based on Remote Sensing Data. *Izvestiya, Atmospheric, Oceanic Physics*, 55, 1329-1337. <https://doi.org/10.1134/S0001433819090202>
- Gopp, N. V., Nechaeva, T. V., Savenkov, O. A., Smirnova, N. V. & Smirnov, V. V. (2017). Indicative capacity of NDVI in predictive mapping of the properties of plow horizons of soils on slopes in the south of Western Siberia. *Eurasian Soil Science*, 50, 1332-1343. <https://doi.org/10.1134/S1064229317110060>
- Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., & Moore, R. (2017). Google Earth Engine: Planetary-scale geospatial analysis for everyone. *Remote Sensing of Environment*, 202, 18-27. <https://doi.org/10.1016/j.rse.2017.06.031>
- Hebryn-Baidy, L. (2017). Application of remote sensing methods to evaluation of soil fertility indicators of Zakarpattia lands. Interdepartmental scientific and technical collection. *Geodesy, cartography and aerial photography*, 85(85), 42-52. <https://doi.org/10.23939/istgcgcap2017.01.042>
- Higginbottom, T. P., & Symeonakis, E. (2014). Assessing Land Degradation and Desertification Using Vegetation Index Data: Current Frameworks and Future Directions. *Remote Sensing*, 6, 9552-9575. <https://doi.org/10.3390/rs6109552>
- Huang, Y., Chen, Z. X., Yu, T., Huang, X. Z., & Gu, X. F. (2018). Agricultural remote sensing big data: Management and applications. *Journal of Integrative Agriculture*, 17, 1915-1931. [https://doi.org/10.1016/S2095-3119\(17\)61859-8](https://doi.org/10.1016/S2095-3119(17)61859-8)
- Huete, A. R. (1988). A soil-adjusted vegetation index (SAVI). *Remote Sensing of Environment*, 25, 295–309. [https://doi.org/10.1016/0034-4257\(88\)90106-X](https://doi.org/10.1016/0034-4257(88)90106-X)
- Jasinski, M. F. (1990). Sensitivity of the normalized difference vegetation index to subpixel canopy cover, soil albedo, and pixel scale. *Remote Sensing of Environment*, 32, 169-187. [https://doi.org/10.1016/0034-4257\(90\)90016-F](https://doi.org/10.1016/0034-4257(90)90016-F)
- Jiang, Z., Huete, A., Didan, K., & Miura, T. (2008). Development of a two-band enhanced vegetation index without a blue band. *Remote Sensing of Environment*, 112, 3833-3845. <https://doi.org/10.1016/j.rse.2008.06.006>
- Jing, Z., Dongli, J., Dong, D., Jinjie, M., Hongwei, L., & Yaonan, B. (2020). Temporal paradox in soil potassium estimations using spaceborne multispectral imagery. *Catena*, 194, 104771. <https://doi.org/10.1016/j.catena.2020.104771>
- Joshi, P. P., Wynne, R. H. & Thomas, V. A. (2019). Cloud detection algorithm using SVM with SWIR2 and tasseled cap applied to Landsat 8. *International Journal of Applied Earth Observation and Geoinformation*, 82, 101898. <https://doi.org/10.1016/j.jag.2019.101898>
- Khangura, R., Ferris, D., Wagg, C., & Bowyer, J. (2023). Regenerative Agriculture - A Literature Review on the Practices and Mechanisms Used to Improve Soil Health. *Sustainability*, 15(3), 2338. <https://doi.org/10.3390/su15032338>
- Khellouk, R., Barakat, A., Jazouli, A. E., Lionboui, H., & Benabdellouahab, T. (2021). Assessment of water cloud model based on SAR and optical satellite data for surface soil moisture retrievals over agricultural area. *Eurasian Journal of Soil Science*, 10(3), 243-250. <https://doi.org/10.18393/ejss.926813>
- Koroleva, P. V., Rukhovich, D. I., Rukhovich, A. D., Rukhovich, D. D., Kulyanitsa, A. L., Trubnikov, A. V., Kalinina, N. V., & Simakova, M. S. (2017). Location of Bare Soil Surface and Soil Line on the RED-NIR Spectral Plane. *Eurasian Soil Science*, 50, 1375-1385. <https://doi.org/10.1134/S10642293171100040>

- Kravtsova, V. (2005). *Space methods of soil research*. Aspect Press, Moscow. 190 p. (In Russian).
- Kulyanitsa, A. L., Rukhovich, A. D., Rukhovich, D. D., Koroleva, P. V., Rukhovich, D. I., & Simakova, M. S. (2017). The application of the piecewise linear approximation to the spectral neighborhood of soil line for the analysis of the quality of normalization of remote sensing materials. *Eurasian Soil Science*, 50, 387-395. <https://doi.org/10.1134/S1064229317040044>
- Landsat 8 Collection 2 (C2) Level 2 Science Product (L2SP) Guide (2020). U.S. Geological Survey. <https://www.usgs.gov/media/files/landsat-8-collection-2-level-2-science-product-guide> (last accessed 19.08.2021).
- Landsat Missions. (2021). USGS. <https://www.usgs.gov/landsat-missions> (last accessed 16.12.2021).
- Li, S., & Chen, X. (2014). A new bare-soil index for rapid mapping developing areas using Landsat 8 data. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences - ISPRS Archives*, 40(4), 139-144. <https://doi.org/10.5194/isprsarchives-XL-4-139-2014>
- Liashenko, D., Belenok, V., Spitsa, R., Pavlyuk, D., & Boiko, O. (2020). *Landslide GIS modelling with QGIS software*. In: XIV International Scientific Conference on Monitoring of Geological Processes and Ecological Condition of the Environment. European Association of Geoscientists & Engineers, Kyiv, Ukraine. <https://doi.org/10.3997/2214-4609.202056069>
- Lu, M., Liu, Y., & Liu, G. (2022). Precise prediction of soil organic matter in soils planted with a variety of crops through hybrid methods. *Computers and Electronics in Agriculture*, 200, 107246. <https://doi.org/10.1016/j.compag.2022.107246>
- Luo, J., Zhang, W., Zhang, X., & Liu, H. (2022). Mapping soil organic matter content using Sentinel-2 synthetic images at different time intervals in Northeast China. *International Journal of Digital Earth*, 16(1), 1094-1107. <https://doi.org/10.1080/17538947.2023.2192005>
- Main Department of the State Land Agency in Kyiv region. (2012). Program of land use and protection in Kyiv region for the period 2012-2020. (In Ukrainian) <http://www.Kyivoblzem.org/> (last accessed 05.11.2021).
- Mateo-García, G., Gómez-Chova, L., Amorós-López, J., Muñoz-Marí, J., & Camps-Valls, G. (2018). Multitemporal cloud masking in the Google Earth Engine. *Remote Sensing*, 10(7), 1079. <https://doi.org/10.3390/rs10071079>
- Mekuriaw, A., Cherinet, M., & Tsegaye, L. (2021). Assessing the impact of land cover changes on selected ecosystem services in upper Suha watershed, Gojjam, Ethiopia. *International Journal of River Basin Management*, 19(4), 459-467. <https://doi.org/10.1080/15715124.2019.1704767>
- Mendenhall, W., & Sincich, T. (2013). *A Second Course in Statistics: Regression Analysis*. 7<sup>th</sup> Edition. Pearson New International Edition. 749 pp.
- Mirzaee, S., Ghorbani-Dashtaki, S., Mohammadi, J., Asadi, H., & Asadzadeh, F. (2016). Spatial variability of soil organic matter using remote sensing data. *Catena*, 145, 118-127. <https://doi.org/10.1016/j.catena.2016.05.023>
- Montandon, L., & Small, E. (2008). The impact of soil reflectance on the quantification of the green vegetation fraction from NDVI. *Remote Sensing of Environment*, 112, 1835-1845. <https://doi.org/10.1016/j.rse.2007.09.007>
- Montero, D., Aybar, C., Mahecha, M. D., Martinuzzi, F., Söchting, M., & Wieneke, S. (2023). A standardized catalogue of spectral indices to advance the use of remote sensing in Earth system research. *Scientific Data*, 10(197), 6254. <https://doi.org/10.1038/s41597-023-02096-0>
- Navarro-Pedreño, J., Almendro-Candel, M. B., & Zorpas, A. A. (2021). The Increase of Soil Organic Matter Reduces Global Warming, Myth or Reality? *Sci*, 3(1), 18. <https://doi.org/10.3390/sci3010018>
- Olthof, I., & Fraser, R. H. (2014). Detecting landscape changes in high latitude environments using Landsat trend analysis: 2. classification. *Remote Sensing*, 6(11), 11558-11578. <https://doi.org/10.3390/rs6111558>
- Omia, E., Bae, H., Park, E., Kim, M. S., Baek, I., Kabenge, I., & Cho, B. K. (2023). Remote Sensing in Field Crop Monitoring: A Comprehensive Review of Sensor Systems, Data Analyses and Recent Advances. *Remote Sensing*, 15(2), 354. <https://doi.org/10.3390/rs15020354>
- Parikh, S. J., & James, B. R. (2012). Soil: The Foundation of Agriculture. *Nature Education Knowledge*, 3(10), 2. <https://www.nature.com/scitable/knowledge/library/soil-the-foundation-of-agriculture-84224268/> (last accessed 05.11.2021).
- Prudnikova, E., & Savin, I. (2021). Some Peculiarities of Arable Soil Organic Matter Detection Using Optical Remote Sensing Data. *Remote Sensing*, 13, 2313. <https://doi.org/10.3390/rs13122313>
- Qi, J., Chehbouni, A., Huete, A. R., Kerr, Y. H., & Sorooshian, S. (1994). A modified soil adjusted vegetation index. *Remote Sensing of Environment*, 48(2), 119-126. [https://doi.org/10.1016/0034-4257\(94\)90134-1](https://doi.org/10.1016/0034-4257(94)90134-1)
- Reis, A. S., Rodrigues, M., Santos, G. L. A. A., Oliveira, K. M., Furlanetto, R. H., Crusiol, L. G. T., Cezar, E., & Nanni, M. R. (2021). Detection of soil organic matter using hyperspectral imaging sensor combined with multivariate regression modeling procedures. *Remote Sensing Applications: Society and Environment*, 22, 100492. <https://doi.org/10.1016/j.rsase.2021.100492>
- Riggs, G. A., Hall, D. K., & Salomonson, V. V. (1994). *A snow index for the Landsat Thematic Mapper and Moderate Resolution Imaging Spectroradiometer*. Proceedings of IGARSS '94 - 1994 IEEE International Geoscience and Remote Sensing Symposium, 8-12 Aug., 1994, Pasadena, CA, USA.
- Romanova, S., Hrichenko, O., Venglinskiy, M., & Yarmolenko, E. (2018). Humus state of soils of Kyiv region. *Agroecological journal*, 3, 34-40. (In Ukrainian). <https://doi.org/10.33730/2077-4893.3.2018.148057>
- Rukhovich, D. I., Koroleva, P. V., Rukhovich, D. D., & Kalinina, N. V. (2021). The Use of Deep Machine Learning for the Automated Selection of Remote Sensing Data for the Determination of Areas of Arable Land Degradation Processes Distribution. *Remote Sensing*, 13, 155. <https://doi.org/10.3390/rs13010155>
- Sakhatsky, O. (2008). Experience in using satellite data to assess soil status to solve natural resource problems. *Reports of the National Academy of Sciences of Ukraine*, 3, 109-115. (In Ukrainian).
- Savin, I., Prudnikova, E., Chendev, Y., Bek, A., Kucher, D., & Dokukin, P. (2021). Detection of Changes in Arable Chernozemic Soil Health Based on Landsat TM Archive Data. *Remote Sensing*, 13, 2411. <https://doi.org/10.3390/rs13122411>
- Shapiro, S. S. & Francia, R. S. (1972). An approximate analysis of variance test for normality. *Journal of the American Statistical Association*, 67(337), 215-216. <https://doi.org/10.1080/01621459.1972.10481232>
- Shetty, A., Umesh, P., & Shetty, A. (2021). An exploratory analysis of urbanization effects on climatic variables: a study using Google Earth Engine. *Modeling Earth Systems and Environment*, 8(1), 1363-1378. <https://doi.org/10.1007/s40808-021-01157-w>
- Shikhov, A., Gerasimov, A., Ponomarchuk, A., & Perminova, E. (2020). *Thematic deciphering and interpretation of cosmic signs of middle and high spaciouness*. State National Research University, Perm. (In Russian).
- Srinivasan, R., Singh, S. K., Nayak, D. C., Hegde, R., & Ramesh, M. (2019). Estimation of soil loss by USLE model using remote sensing and GIS Techniques - A case study of Coastal Odisha, India. *Eurasian Journal of Soil Science*, 8(4), 321-328. <https://doi.org/10.18393/ejss.598120>
- State Department of Environmental Protection in Kyiv region. (2020). Regional report on the state of the environment in Kyiv region in 2020. (In Ukrainian) <http://eco-Kyiv.com.ua> (last accessed 05. 11. 2021).
- State Department of Environmental Protection in Kyiv region. (2021). (In Ukrainian) <https://ecology-Kyivoblast.com.ua/page/stan-dovkillya-kyiyvskoyi-oblasti> (last accessed 05. 11. 2021).
- State Institution "Soils protection Institute of Ukraine". (2016). Report on the implementation of design and technological and research work in 2015 (final). (In Ukrainian).
- State Institution "Soils protection Institute of Ukraine". (2020). Periodic report of the condition of soils on the land of agriculture of UKRAINE according to the results of the 10th round (2011-2015) agrochemical survey of lands. (In Ukrainian) <http://www.iogu.gov.ua/wecompress.com.pdf> (last accessed 05.11.2021).

- State Institution "Soils protection Institute of Ukraine". (2020). Agrochemical survey of land. (In Ukrainian) <https://www.ioгу.gov.ua/pasportizaciya/ahrohimichne-obstezhennya-silskohospodarskyh-uhid/> (last accessed 05. 11. 2021).
- State Service of Ukraine for Geodesy, Cartography and Cadastre. (2020). Distribution of land fund of Kyiv region. (In Ukrainian) <http://land.gov.ua/> (last accessed 05.11.2021).
- Sullivan, D. G., Shaw, J. N., Rickman, D., Mask, P. L., & Luvall, J. C. (2005). Using remote sensing data to evaluate surface soil properties in Alabama ultisols. *Soil Science*, 170(12), 954–968. <https://doi.org/10.1097/01.ss.0000187350.39611.d7>
- Taplin, R. H. (1999). Robust F-tests for linear models. *Canadian Journal of Statistics*, 27(2), 361–371. <https://doi.org/10.2307/3315645>
- The Humanitarian Data Exchange. (2021). <https://data.humdata.org> (last accessed 16.12.2021).
- Tian, J., Su, S., Tian, Q., Zhan, W., Xi, Y., & Wang, N. (2021). A novel spectral index for estimating fractional cover of non-photosynthetic vegetation using near-infrared bands of Sentinel satellite. *International Journal of Applied Earth Observation and Geoinformation*, 101, 102361. <https://doi.org/10.1016/j.jag.2021.102361>
- Truskavetsky, S. (2006). *The use of multispectral space scanning and geoinformation systems in the study of the soil cover of Polissya of Ukraine*. [PhD Thesis]. (In Ukrainian).
- Truskavetsky, S., Byndych, T., Sherstyuk, A., & Viatkin, K. (2015). Studying the Condition of Soil Protection Agrolandscape in Ukraine Using Remote Sensing Methods. *Journal of Agricultural Science and Technology A* 5, 235-240. <https://doi.org/10.17265/2161-6256/2015.04.001>
- Tucker, C. J. (1979). Red and photographic infrared linear combinations for monitoring vegetation. *Remote Sensing of Environment*, 8, 127-150. [https://doi.org/10.1016/0034-4257\(79\)90013-0](https://doi.org/10.1016/0034-4257(79)90013-0)
- Vermote, E. F., Justice, C. O., Claverie, M., & Franch, B. (2016). Analysis of the performance of the Landsat 8/OLI land surface reflectance product. *Remote Sensing of Environment*, 185, 46-56. <https://doi.org/10.1016/j.rse.2016.04.008>
- Wang, X., & Ge, L. (2012). *Evaluation of filters for ENVISAT ASAR speckle suppression in pasture area*. Proceedings of the ISPRS Annals of the XXII ISPRS Congress-Photogrammetry, Remote Sensing and Spatial Information Sciences. Melbourne, pp. 341–346.
- Wang, Y., Luo, C., Zhang, W., Meng, X., Liu, Q., Zhang, Y., & Liu, H. (2022). Remote Sensing Prediction Model of Cultivated Land Soil Organic Matter Considering the Best Time Window. *Sustainability*, 15(1), 1-15. <https://doi.org/10.3390/su15010469>
- Website of the Kyiv regional State Administration. (2021). (In Ukrainian) <http://www.Kyiv.obl.gov.ua> (last accessed 05. 11. 2021).
- Weiss, M., Jacob, F., & Duveiller, G. (2020). Remote sensing for agricultural applications: A meta-review. *Remote Sensing of Environment*, 236, 111402. <https://doi.org/10.1016/j.rse.2019.111402>
- Wilson, E. H., & Sader, S. A. (2002). Detection of forest harvest type using multiple dates of Landsat TM imagery. *Remote Sensing of Environment*, 80(3), 385-396. [https://doi.org/10.1016/S0034-4257\(01\)00318-2](https://doi.org/10.1016/S0034-4257(01)00318-2)
- Wu, C., Wu, J., Luo, Y., Zhang, L., & DeGloria, S. (2009). Spatial Prediction of Soil Organic Matter Content Using Cokriging with Remotely Sensed Data. *Soil Science Society of America Journal*, 73(4), 1202-1208. <https://doi.org/10.2136/sssaj2008.0045>
- Yacuk, I., & Balyk, S. (2013). Methods of agrochemical certification of agricultural lands, 99. (In Ukrainian).
- Yang, J., Li, X., Wu, B., Wu, J., Sun, B., Yan, C., & Gao, Z. (2021). High Spatial Resolution Topsoil Organic Matter Content Mapping Across Desertified Land in Northern China. *Frontiers in Environmental Science*, 9, 668912. <https://doi.org/10.3389/fenvs.2021.668912>
- Yu, Q., Yao, T., Lu, H., Feng, W., Xue, Y., & Liu, B. (2021). Improving estimation of soil organic matter content by combining Landsat 8 OLI images and environmental data: A case study in the river valley of the southern Qinghai-Tibet Plateau. *Computers and Electronics in Agriculture*, 185, 106144. <https://doi.org/10.1016/j.compag.2021.106144>
- Yu, T., Jiapaer, G., Bao, A., Zheng, G., Jiang, L., Yuan, Y., & Huang, X. (2020). Using Synthetic Remote Sensing Indicators to Monitor the Land Degradation in a Salinized Area. *Remote Sensing*, 13, 2851. <https://doi.org/10.3390/rs13152851>
- Zatserkovny, V. I., Babych V. Y., Belenok, V. Y., Frolov, H. O., & Hebryn-Baidy, L. V. (2019). *Black sea level change monitoring using altimetry data and geoinformation technologies*. 18th International Conference on Geoinformatics – Theoretical and Applied Aspects, 13-16 May 2019, Kyiv, Ukraine. 15785. <https://doi.org/10.3997/2214-4609.201902060>
- Zekoll, V., Main-Knorn, M., Alonso, K., Louis, J., Frantz, D., Richter, R., & Pflug, B. (2021). Comparison of masking algorithms for Sentinel-2 imagery. *Remote Sensing*, 13(1), 137: 1-21. <https://doi.org/10.3390/rs13010137>
- Zha, Y., Gao, J., & Ni, S. (2003). Use of normalized difference built-up index in automatically mapping urban areas from TM imagery. *International Journal of Remote Sensing*, 24(3), 583-594. <https://doi.org/10.1080/01431160304987>
- Zhang, M., Liu, H., Zhang, M., Yang, H., Jin, Y., Han, Y., Tang, H., Zhang, X., & Zhang, X. (2021). Mapping Soil Organic Matter and Analyzing the Prediction Accuracy of Typical Cropland Soil Types on the Northern Songnen Plain. *Remote Sensing*, 13(24), 5162. <https://doi.org/10.3390/rs13245162>