GRÓ Land Restoration Training Programme
Árleynir 22, 112 Reykjavik, Iceland

Final project 2022

CROP YIELD ESTIMATION USING SENTINEL-2 DATA: A CASE STUDY FROM UGTAALTSAIDAM SUB-PROVINCE, MONGOLIA

Erdenetuya Boldbaatar
Mongolian University of Life Sciences
17024 Zaisan, Khan-uul District, Ulaanbaatar Mongolia
ch_erdenetuya@muls.edu.mn

Supervisor
Michaela Hrabalikova
National Land Survey of Iceland
michaela.hrabalikova@lmi.is

ABSTRACT

Globally, remote sensing technology is widely used in cropland monitoring. Even in Mongolia, there is a need to use remote sensing technology to monitor agricultural fields located in a vast area. In this study, the relationship between NDVI, NDMI, and LAI calculated using the 2017 field measurement data and Sentinel 2 satellite data of Ugtaaltsaidam sum cultivation area of Central Province was evaluated. Exploratory data analysis and land cover classification methods were used to evaluate the descriptive statistics of the collected data using QGIS Semi-Automatic Plugin. In addition, Python was used in QGIS to calculate correlations. The three calculated indices had a positive but very weak correlation with crop yield. Since there was yield data but no crop type data, unsupervised classification of land cover was done and divided into four crop types based on which were obtained by the clustering (k-mean) of Sentinel 2 image. When calculating the correlation between each of these, the three indices were positively and weakly correlated for the Crop 1 category. However, the yield of Crop 2 category was positively and weakly correlated with NDVI and LAI, but not correlated with NDMI. However, other parameters of the soil had a weak negative correlation with the indices. The results were opposite to expectations based on the literature review. The results indicate that the crop yield recording methodology has to be improved by adding information about the method of crop yield estimation, crop type, and more information about the date of harvest. The methods developed and tested in this study will be of great importance for future research on crop yield in Mongolia and will hopefully enable enhanced data collection, classification, and evaluation.

Keywords: remote sensing, vegetation indices, correlation, crop yield, Mongolia

This paper should be cited as:

Boldbaatar E (2022) Crop yield estimation using sentinel-2 data: a case study from Ugtaaltsaidam sub-province, Mongolia. GRÓ Land Restoration Training Programme [Final project] https://www.grocentre.is/static/gro/publication/849/document/boldbaatar2022.pdf

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABBREVIATIONS

NDVI	Normalized Difference Vegetation Index
LAI	Leaf Area Index
NDMI	Normalized Difference Moisture Index
RS	Remote Sensing
ESA	European Space Agency
EDA	Earth System Dynamic
OSAVI	Optimized Soil-Adjusted Vegetation Index
MTVI	Modified Triangular Vegetation Index
EVI	Enhanced Vegetation Index
NIR	Near Infra-Red
RED	Red
SWIR	Short Wave Infra-Red

1. INTRODUCTION

The World Health Organisation's report from 2021 highlights the importance of food safety in upcoming years due to climate change. To support and facilitate the decision-making process for various applications in ensuring food security, reliable data on agriculture and crops are essential. One crucial step in assisting policy, land use allocation, and environmental concerns is crop estimates and monitoring (UN 2021; Muruganantham & Wibowo 2022; Thenkabail 2010).

The crop agriculture sector in Mongolia is relatively young, with a history of only 60 years (Batzorig 2011). Moreover, estimating crop yields, regionally and nationally, is a challenge not only for agricultural policy and management but also for the economy and society. Currently, the agricultural database of Mongolia does not have information on the yield for single crop plots but only general information for the region (Agricultural database 2022). In specific locations, but not everywhere, the harvest records are miswritten. For instance, the yield is not included or written as a sum for administrative units like in the Ugtaaltsaidam sub-province (Agricultural database 2022). In addition, in Mongolia, cropland agriculture is mainly carried out by private enterprises. Therefore, the size of the cultivated area, the type of crops, and the crop production are not included in the database consistently using the same methodology. It therefore becomes problematic for the government to monitor crop yields in a unified manner. Furthermore, crop yields are predicted only by traditional methods.

Crop estimates and forecasting may be performed using various techniques. However, satellite earth observation is a good option due to its coverage and data type. The modern methods of remote sensing technologies allow the exploration of large areas in short time periods which is a promising solution for the Mongolian crop sector.

Remote sensing technology has been used intensively in Mongolia since the year 2000. Nevertheless, most of the associated studies focus solely on the classification of cultivated and uncultivated land. For example, Batzorig (2011) used Landsat satellite data to classify cultivated and uncultivated areas in Tuv and Selenge Aimags by classification based on NDVI. Another application of remotely sensed data in agriculture explored the possibility of using a wide variety of satellites to monitor pastureland and cropland (Tsevengee 2016a). Tsevengee (2016b) also examined the possibility of defining the boundaries of cropland using the land cover classification method. Studies researching relationships between vegetation indices and Mongolia crop yields have, however, not yet been conducted.

The main goal of this research was therefore to assess the application of remotely sensed data for crop estimation in Mongolia. The study evaluated selected vegetation indices along with field records of crop yield. The objectives were to:

1. Calculate NDVI, LAI, and NDMI from Sentinel-2 satellite data
2. Conduct exploratory data analysis
3. Examine relationships between selected vegetation indexes and field data

1.1 The Mongolian crop agricultural sector

Mongolia has an area of 1.5 million km^2 and is sparsely populated with 1.8 person/km^2 (Mongolian Statistic Information Center 2022). Historically, pastoralism has been the dominant sector, and cropland has developed in the last 60 years. The crop agriculture sector in Mongolia is relatively young and small, covering 0.07 % of the total land and accounting for 0.7-1.29 % of the gross domestic product.

In 1959, Mongolia launched the first *Atar Campaign*, a process involving a large number of people, aimed at developing the country's economy and improving the food supply of the population by plowing the atar land and developing agriculture (Lonjid 1995). Then, during the second campaign in 1976, fallow cultivation began. The 1990s began the transition from a centrally planned economy to a market economy (Batzorig & Banzragch 2012) with a collapse in 2005. Therefore, the 3rd *Atar Campaign* in 2008 took place to recover and boost the agriculturalsector.

As a result of the last *Atar Campaign* in 2008, the cultivated area increased, but crop yields have remained stable (Agricultural database 2022). Figure 1 shows how the cropland area rapidly increased within the last three years from 526 thousand hectares to 675.1 thousand hectares. The fallow land for agricultural purposes decreased from 304.9 thousand hectares to 260.6 thousand hectares. Nevertheless, the crop yield remains almost the same.

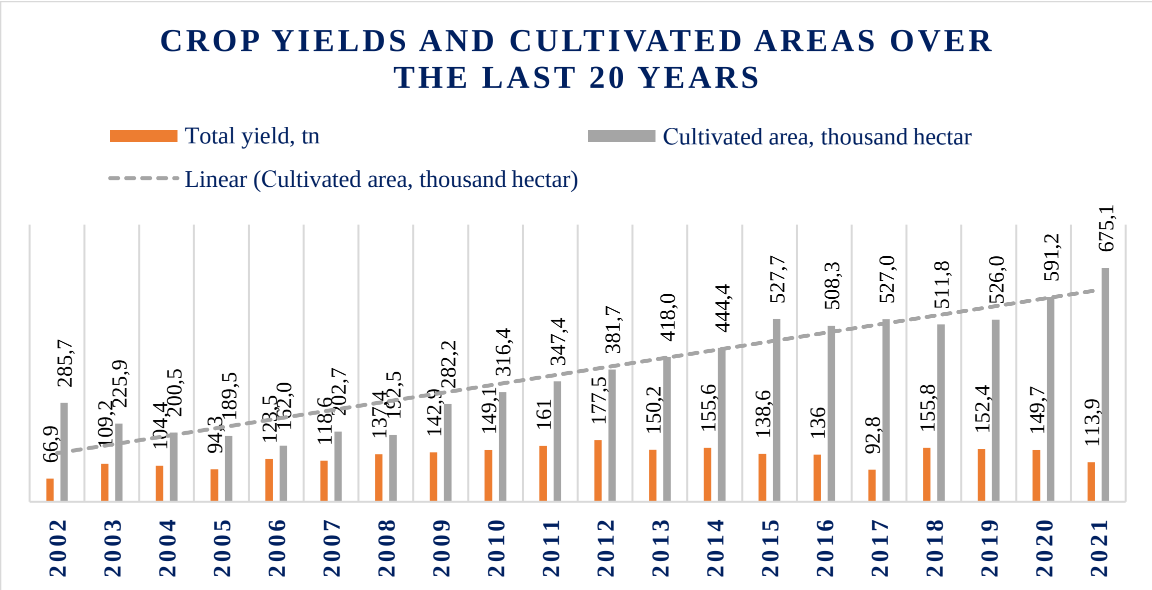


Figure 1. Changes in harvest and cultivated area over the past 20 years. (Source: Agricultural database 2022).

One of the goals of Mongolian policymakers is to increase self-sufficiency in the production of strategic crops, such as wheat, barley, potato, and vegetables. Sharav (1999) conducted an extensive study where he found that the Mongolian agriculture sector needs to increase production to secure food safety significantly. Between 2011 and 2020 there was a need to produce 250-300 thousand tons of wheat and 300-350 thousand tons of potatoes and vegetables according to the Agricultural Database (2022). This was impossible to reach with the level of agriculture at that time.

1.2 Application of remote sensing for agriculture

Although satellite observations of agricultural monitoring started with the launching of Landsat 1 in 1972, only in recent years have space agencies, national ministries of agriculture, and worldwide efforts revitalised agricultural remote sensing (EROS n.d.). The current abundance of high-resolution optical satellite data offers evaluation opportunities of various geophysical and biophysical factors to monitor agricultural systems. The optical imagery can be used for precipitation, temperature, evapotranspiration, soil moisture, vegetation health, crop monitoring and forecasting.

Over the last several decades, much research has been conducted on the use of remote sensing for yield prediction at various levels, from the field to the national level (Mamatkulov et al. 2021). As a result, these research findings are critical for individual farmers, national governments, and international organisations.

The spectral, temporal, and geographical resolutions of satellite pictures vary. The pictures that are most typically utilised for regional yield estimate studies still have a coarse resolution and a daily revisit period (Duncan et al. 2015). Low-resolution data sources (Table 1) include the Advanced Very High-Resolution Radiometer (AVHRR), Satellites Pour l'Observation de la Terre or Earth-observing Satellites VeGeTation (SPOT-VGT), MODerate Resolution Imaging Spectroradiometer (MODIS), and MEdium Resolution Imaging Spectroradiometer (MERIS) (Duncan et al. 2015).

The most significant limitation of low-resolution time-series images in crop monitoring assessment is that their spectral reflectance contains mixed pixel information of various surfaces and vegetation types (Durgun et al. 2020). This makes it challenging to interpret the signal and directly relate a spectral or temporal signature to a specific crop condition. This issue should progressively decrease when a time series of images with better spatial resolutions are established, for instance by Landsat-8, Landsat 9 and Sentinel-2 (Table 2). However, as Durgun et al. (2020) stated, images with high spatial resolutions are limited by our coping with much larger data volume and processing time.

Table 1. Satellites sensors for agricultural applications with low resolution. (Adapted from: NASA 2022, ESA 2022).

							Specific products					
Satellite	**Sensor**		**Spatial resolution**	**Spectral resolution**	**Temporal resolution**	**Coverage**	**land surface reflectance**	**Evapotrans-piration**	**Land surface temperature**	**Soil moisture**	**Vegetation greenness**	**Structure**
Terra	MODIS	1999 - present	250 m, 500 m, 1 km	36 bands 400 – 14400 nm	Daily, 8-day, 16-day, monthly, yearly	Global with gaps in the tropic	x	x	x		x	
Aqua	MODIS	2002 - present					x	x	x		x	
Suomi-NPP	VIIRS	2012 - present	375 m, 750 m	22 bands 400 – 12500 nm		Global	x		x		x	
NOAA-20	VIIRS						x		x		x	
Soil Moisture Active Passive	L-band radar		9 km	1.41 Ghz	3-hourly, daily					x		

Table 2. Satellites sensors for agricultural applications with high resolution. (Adapted from: NASA 2022, and ESA 2022).

							Specific products					
Satellite	**Sensor**		**Spatial resolution**	**Spectral resolution**	**Temporal resolution**	**Coverage**	**Land surface reflectance**	**Evapotrans-piration**	**Land surface temperature**	**Soil moisture**	**Vegetation greenness**	**Structure**
Landsat 8	OLI	2013 - present	30 m	9 bands 430 – 2290 nm	16- days	Global	x				x	
Landsat 9	OLI	2020 - present	30 m		16-days	Global	x				x	
Sentinel 2	MSI	Sentinel 2A: 2015 - present Sentinel 2B: 2017 - present	10 m (VNIR), 20 m (red edge, SWIR), & 60 m (atmospheric bands)	13 bands 440 – 2200 nm	5 days	Global	x				x	
Sentinel 1	C-band radar	2014 - present	~10 m	Dual polarisation: VV+VH or HH+HV	12-days	Global						x

Remote sensing is used extensively in statistical algorithms to estimate crop yields (Liang et al. 2015). Crop production may be predicted empirically using vegetation indices obtained from satellite reflectance. For example, Liang et al. (2015) used the hybrid inversion method to calculate the grain leaf area index. They concluded that the most sensitive indices were the optimized soil-adjusted vegetation index (OSAVI) r^2=0.928 and the modified triangular vegetation index (MTVI) r^2 = 0.910. Ali et al. (2019) calculated the correlation of grain yields using the NDVI, EVI, SAVI, GNDVI, GCI, and SR indices over an area of 11.07 hectares in northern Italy over five years. Grain yields were calculated using Landsat satellite data, with a 64-86 % correlation.

Kern et al. (2018) developed a multi-linear regression model using the example of Hungary to predict the 2000-2016 winter wheat, rapeseed, corn, and sunflower yields. The results of the study showed that remote sensing data only improved the overall prediction ability of the winter wheat model. The results were able to provide simple yet robust models for spatially specific yield predictions as well as near-future yield projections.

A similar approach was used by Ganshukh (2020), who used multi-spectral instrument analysis of Sentinel 1, Sentinel 2, and SAR data from time-series data to monitor wheat growth in the agricultural region of Mongolia to estimate NDVI and NDWI as the most appropriate indices.

Karlson et al. (2020) conducted a study to measure crop yield in 22 agricultural fields in central Burkina Faso with varying rainfall. The feasibility of using the Sentinel 2 time series was investigated, showing that this satellite data could explain 41-80 % of the variation in crop data measurements.

Hunt et al. (2019) studied within-field wheat yield variability over a one-year period using Sentinel 2 satellite data. A comparison was made with environmental data collected from a total of 39 wheat fields. The results showed that, by using data from the Sentinel 2 satellite with a resolution of 10 meters, it was possible to accurately map yield changes within the field.
Another sudy used Sentinel 2 satellite data with two different spatial resolutions of 10 and 20 meters to estimate wheat yield. Thus, to do the calculation 3 months before harvest was highly relevant (Franch et al. 2021).

A leaf area index calculated from Sentinel 2 satellite data was used in a simple wheat yield estimation algorithm and compared to 48 field measurements in 2014 and 2015 (Ma et al. 2022). It was concluded that the use of remote sensing data in crop land research facilitates the estimation of LAI, biomass and yield.

A feasibility assessment of using Sentinel 2 satellite data to monitor wheat field dynamics in 2018-2019 was conducted in the Ebro Delta growing area, Spain by Soriano Gonzalez et al. (2022). For this purpose, a time series was generated using four different vegetation indices (NDVI, NDWIMF, NDWIGAO, BSI) derived from 20-m resolution smoothed Sentinel 2 data. They concluded that, while the results do not guarantee continuous phenology monitoring for all fields, it is possible for wheat fields.

Tuvdendorj et al. (2022) conducted a study using nine types of remote sensing indices to predict yields. The aim was to find the best index for the development of the spring wheat yield calculation model using correlation and regression methods, and the most suitable indices were NDWI, VSDI, and NDVI. The study also found that the flowering period in late June and early July were suitable for harvest forecasts.

2. METHODS

2.1 Study area

Mongolia is an administrative unit with 21 provinces and one capital city. The provinces are divided into 330 sub-provinces.

The field site is Ugtaaltsaidam soum (sub-province; 105˚23´51.729″, 48˚11´07.252″) in the Tuv province in the central part of Mongolia. Figure 2 shows the spatial distribution of crop plots. Table 3 summarises the basic physio-geographical characteristics of the study site.

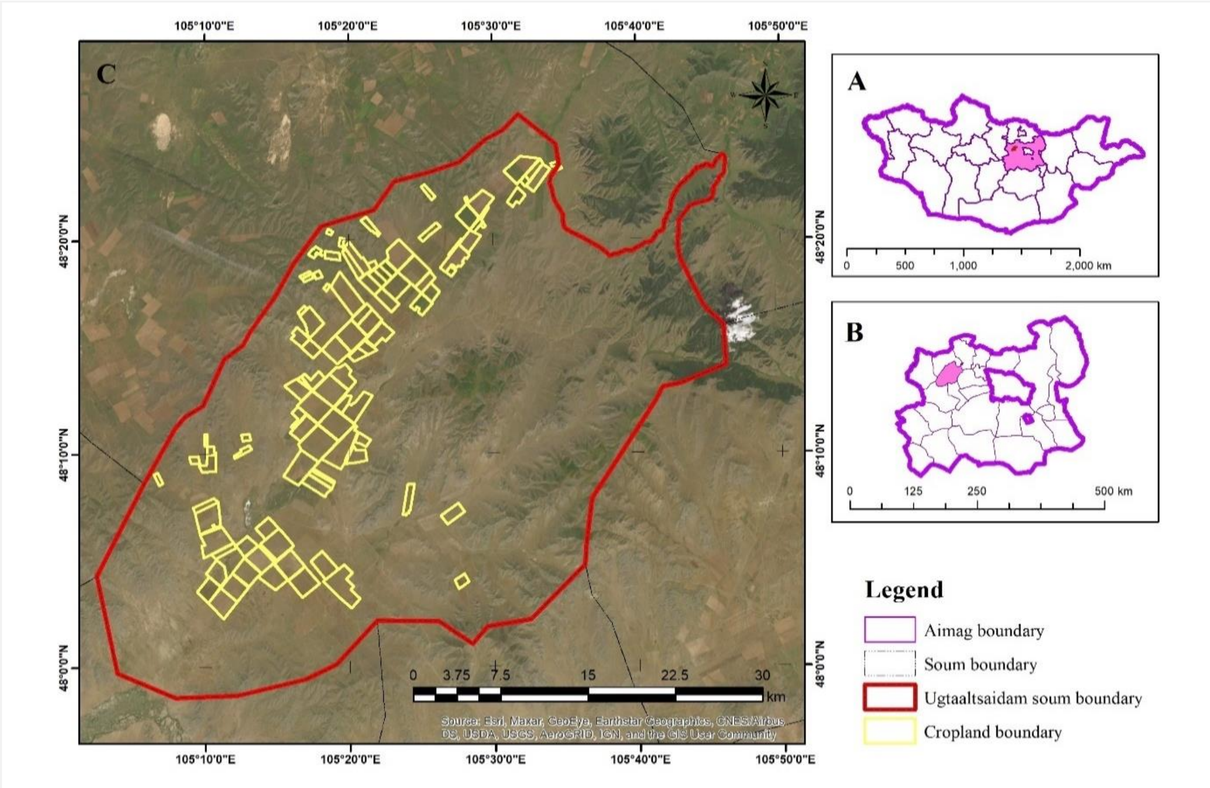


Figure 2. Study area: A) Tuv aimag, Mongolia, B) Ugtaaltsaidam soum, Tuv aimag, C) Cropland in Ugtaaltsaidam soum. (Source: Mongolian environmental database 2022; ESRI base maps 2022).

Table 3. Information about the study area.

Characteristic	Description
Area	The total area is 18.5 thousand hectares (Agricultural database 2022). It represents 12 per cent of the sub-province's total area.
Climate	Dry and cool summers (the average temperature in July is +17.2 degrees), harsh winters (the average temperature is -29.5 degrees in January). The average wind speed is 18 meters per second, and the total annual precipitation is 504.8 mm (Development 2022).
Soil	The typical dark soils and dark coloured mountain forests with mountain meadow-forest soils are prevalent (Mongolian national classification system) (Mongolian environmental database 2022).
Landscape	The fields are located on a flat surface with a slope of 0-6 degrees.
Natural zone	Located in the central zone between forest-steppe and steppe, it is suitable for livestock and agriculture (Mongolian environmental database 2022).
Cultivated plants	Potatoes, vegetables, wheat, and fodder crops are grown alternately using a cropping system (Minister of Economic Development 2022).
Growing season	This area's growing season is 100-110 days (number of days above +5 degrees) in summertime (Lonjid 1995).

2.2 Data and methodology

2.2.1 Field data collection

In a 2017 field survey, soil samples were taken from 99 crop plot parcels, and data on soil parameters were collected. In addition, data on the harvest of each year's crop was obtained. The data from this field study was collected as part of another research project. Soil samples were taken from each crop plot at 0-20 cm. Information about all methods used for soil analyses was unfortunately not available, only the results.

The following information was collected and used about soil parameters:

1. Soil classification: The area had brown soil, brown mountain soil, and brown meadow soil.
2. Soil texture, clay content (Ministry of Construction and Urban Development 2019). Clay is one of the important variables influencing the water uptake of plants, thus it indirectly influences the crop harvest. The ratio of clay and sand in the soil determines its soil texture. According to soil classification in Mongolia, if the soil texture is 0-20%, it is sand or sandy soil, 20-60% is loamy soil, and more than 60% is classified as clay soil.
3. Humus content: Humus content is also expressed as a percentage and determined by taking soil samples. In Mongolia, the content of humus is 5-10% for black soil, 3-5% for dark brown soil, 2-3% for brown soil, and 1.5-2% for light brown soil (Ministry of Construction and Urban Development 2019).
4. Potassium: Potassium supply was expressed in units of mg/kg (Ajindai 1999). It plays a major role in plant nutrition. Potassium supply is classified internally as very low (<10 mg/kg), low (10-15 mg/kg), medium (15-20 mg/kg), and high (<20 mg/kg).
5. pH: soil pH of 7 is neutral, less than 7 is acidic, and greater than 7 is alkaline.

Crop yield data for the 2017 harvest was also obtained. The unit of measurement was kg per hectare.

2.2.2 Imagery acquisition, processing, and software

Sentinel 2 satellite data is open source, and the spatial resolution is high-resolution (10 meters), so it was chosen for use in this study. Sentinel-2 consists of two satellites placed in sun-synchronous orbit at 180 degrees from each other. It is designed to monitor land surface conditions and changes, and its wide bandwidth (290 km) and revisit time will support the monitoring of changes in the Earth's surface. The Sentinel-2 satellites contain one Multi-Spectral Instrument (MSI) with 13 spectral channels in the visible/near-infrared (VNIR, 10 m spatial resolution) and short-wave infrared (SWIR, 20 m spatial resolution) spectral ranges. The 13 bands enables continued interoperability with SPOT-5 and Landsat-8 missions (ESA 2022).

QGIS (Semi-Automatic Classification Plugin) was used to develop survey maps. A Sentinel 2 satellite image dated August 1, 2017, T035541 was downloaded by writing a data retrieval script in Python in QGIS. The downloaded data was pre-processed with the help of the Semi-Automatic Classification Plugin. After that, the vegetation index (NDVI, LAI, SMI) was evaluated for crop yield estimation. The general workflow is showed in Figure 3.

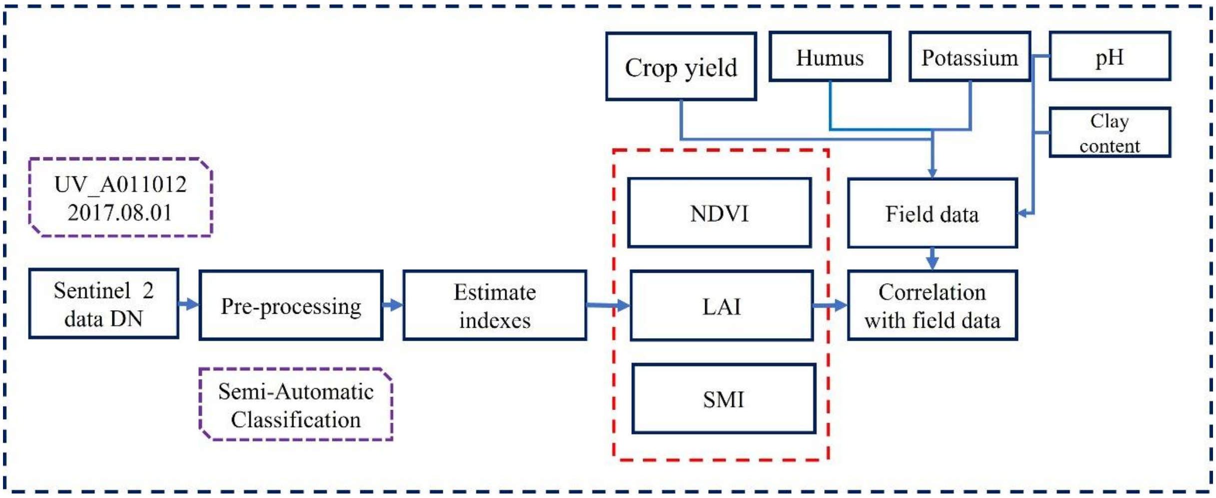


Figure 3. Study method flow chart.

2.2.3 Image classification and computing vegetation indices

Normalised Difference Vegetation Index

NDVI anomaly data are routinely used to determine crop and pasture vegetation conditions. Healthy plants reflect more infrared and green light and absorb more red and blue light.
Vegetation absorbs a lot of sunlight in the visible red and near-infrared. An attempt was made to assess the greenness of the plants with different indices using this property, and the NDVI was calculated. The NDVI is between -1 and +1. The closer to +1, the better the vegetation cover. It is expressed by the following general formula (NASA Earth Observatory 2000).

NDVI = (NIR-RED)/(NIR+RED) [1]

Red- Red
NIR- Near infrared

Leaf Area Index

The leaf area index was initially defined as the area of photosynthetic tissue per unit surface area (Watson 1947). This index is an essential indicator of the leaf cover and chlorophyll content during the growing season. Therefore, LAI is used in plant and evaporation studies, light absorption, crop setting, growth phases, and chemical element metabolism. The index has also been used in cropland and forestry studies. The most straightforward and most suitable substitute for the leaf index is the normalized difference vegetation index (NDVI), which is based on the red and near-infrared channels and determined by the following formula (Tewari et al. 2003).

LAI = 0.57*exp (2.33*NDVI) [2]

LAI- Leaf Area Index
NDVI- Normalized Difference Vegetation Index

Normalized difference moisture index

NDMI is a normalized difference moisture index that uses the near infrared and short wave infrared bands to represent moisture (Wilson & Sader 2002). The SWIR band reflects changes in plant water content and spongy mesophyll structure in the plant canopy, while NIR reflectance depends on leaf internal structure and leaf dry matter content but not water content. By combining NIR with SWIR, the variation due to leaf internal structure and dry matter content can be eliminated and the accuracy of obtaining plant water content can be improved. The amount of water in the leaf's internal structure largely controls the spectral reflectance in the SWIR interval of the electromagnetic spectrum. Thus, SWIR reflectance is negatively related to leaf water content. Briefly, NDMI is used to monitor changes in leaf water content and was proposed by Gao (1996).

$$SMI = (NIR-SWIR)/(NIR-SWIR) \quad [3]$$

NIR- Near infrared
SWIR- Short wave infrared
NDMI has a value of -1 to 1. A negative NDMI value indicates infertile soil, and a value of 0.2-0.4 indicates water stress (Alberto 2022).

<u>K-Means unsupervised classification</u>

Information on the type of crops planted in 2017 was not available for each crop plot. Therefore, the k-means unsupervised classification method was used to classify by crop type.

An unsupervised land cover classification is performed to obtain initial general information about the area under study (GISGeography 2022). Unsupervised classification creates clusters based on similar spectral features in an image. Then each cluster is classified without providing training samples. Unsupervised classification is performed following the steps of clustering and segmentation. K-Means unsupervised classification uses the minimum distance method to iteratively cluster pixels into the closest class. At each iteration, the classification values are rethought, and the pixels are refined and classified.

<u>Zonal statistical analysis</u>

After that, the indices calculated by the amount of sub-province were extracted for each crop plot using zonal statistics (ESRI 2022a). The zone statistics operation is an operation to calculate the statistics of the cell values of the raster (value raster) with different values by zone. When calculating statistics by zone, you can either calculate Zone Statistics or Zone Statistics by Table. Zonal statistical tables were used in this study. They truncate the pre-computed raster values at values with specific boundaries and generate statistics for each zone.

3. RESULTS

3.1 Calculating indices

Figure 4 shows calculated vegetation indices for the study area. As of August 1, 2017, NDVI was -0.59-+0.83, NDMI -0.79-+0.87, and LAI was 0.14-3.9.

As for NDVI, the minimum value of 0.095 indicated the presence of vegetation in those areas, and the maximum value of 0.48 indicated the area with the densest vegetation (Figure 5, Table 4). The mean value was 0.128, indicating that the average vegetation cover of the plots was low. The NDVI range was 0.386, which shows that there was a large difference in the index of each area. Considering the NDVI value of each area, 30 areas have a value of 0.1-0.12, 25 areas have a value of 0.12-0.14, and 13 areas have a value of 0.14-0.16, or 69% of the total area was between 0.1-0.16 as of August 2, 2017.

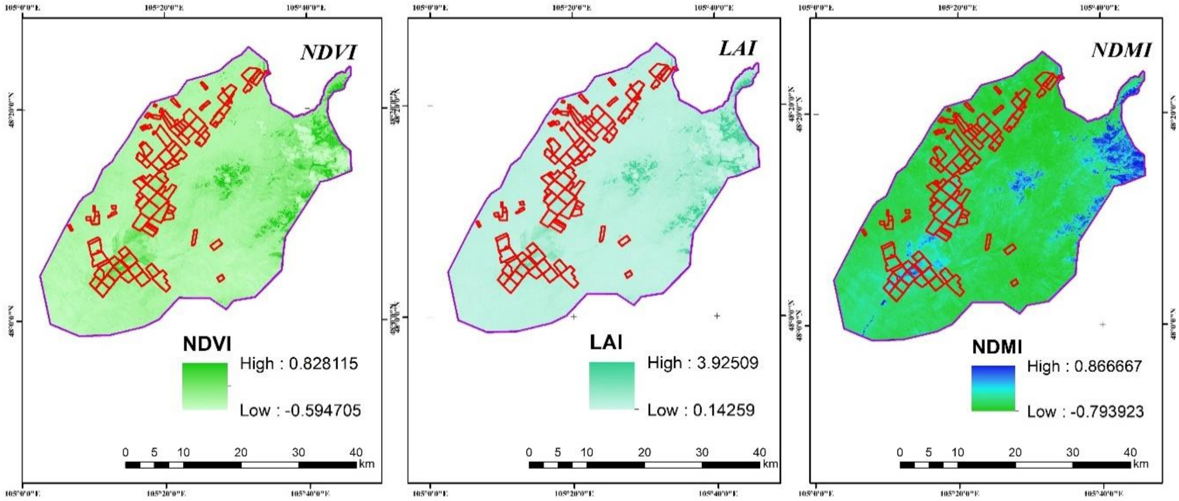


Figure 4. Indices calculated from Sentinel 2 satellite data. A) NDVI, B) LAI, C) NDMI.

NDMI calculation showed a value of -0.37-0.019, and it was classified as low, medium, and high-water stress with moderate canopy cover. NDMI was -0.3- -0.28 in 25 areas, -0.32- -0.3 in 13 areas, -0.28- -0.26 in 12 areas, and 50 percent of the total area was low canopy cover, dry or very low canopy cover, wet.

For the leaf area index, values of 0.711-1.749 were calculated. 64 percent of the total area had a value of 0.7-0.8 or lower. The range of LAI was 1.038, which showed that the vegetation cover of each field was significantly different.

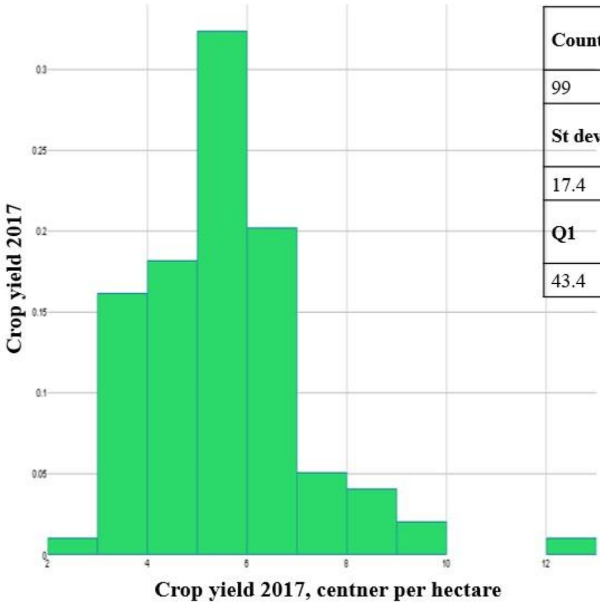
Figure 5. Scatter plot of NDVI, NDMI and LAI.

Table 4. Statistics of indices.

Index	Sum	Mean	Median	St dev	Minimum	Maximum	Range	Q1	Q3	IQR	Missing values
NDVI	14.899	0.152	0.128	0.062	0.095	0.481	0.386	0.115	0.162	0.047	0
NDMI	-27.074	-0.276	-0.285	0.058	-0.374	0.019	0.393	-0.311	-0.255	0.056	0
LAI	80.567	0.822	0.769	0.147	0.711	1.749	1.038	0.745	0.831	0.086	0

3.2 Exploratory Data Analysis of the field data set

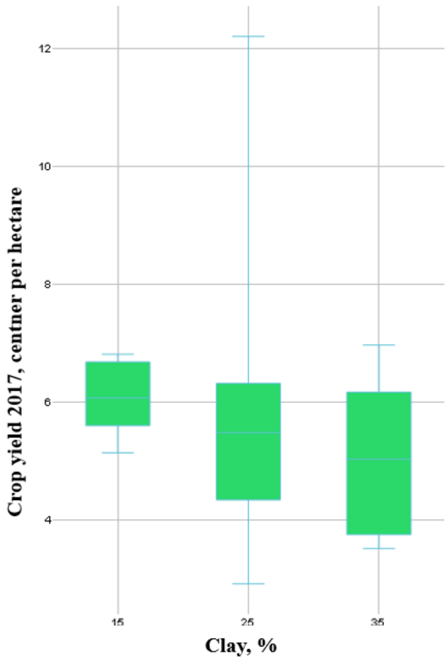
The distribution of crop yield records (Figure 6) was mostly symmetrical in a bell shape, indicating a normal distribution. There is an outlier, zero values that can be caused by missing values and one value of 122 kg/ha. With a standard deviation of 17.4 kg/ha and interquartile range (IQR) 19.6 indicating variation in the crop yield data. Most of the yield fall between 35 – 65 kg/ha, where 29 plots have yield in the range 55-65 kg/ha and 22 plots in the range 35-45 kg/ha.



Count	Sum	Mean	Median
99	5387.5	54.4	55.0
St dev	**Minimum**	**Maximum**	**Range**
17.4	0	122.0	122.0
Q1	**Q3**	**IQR**	**Missing values**
43.4	63.0	19.6	2

Figure 6. Distribution and summary statistics of crop yield data.

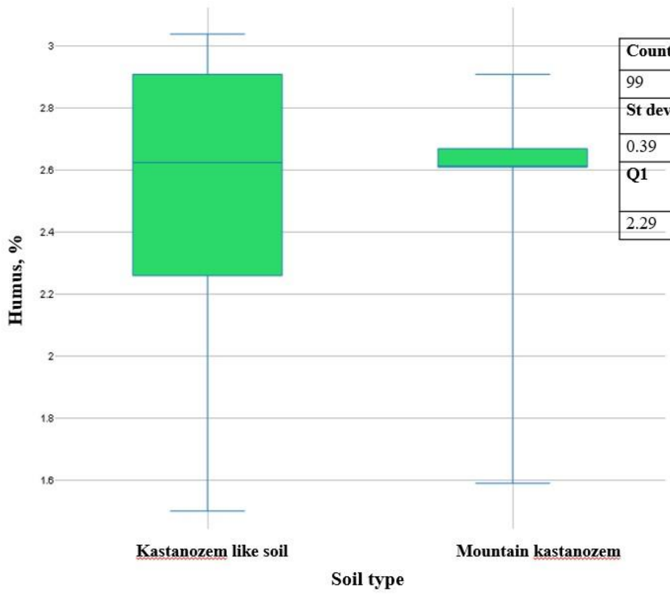
As an additional parameter to crop yield data, soil texture was analysed in relation to crop yield and plotted as a boxplot (Figure 7). The proportion of clay was 15-35%, i.e., sandy, light loamy, and medium loamy soil. The mean is 25% and the standard deviation is 4.67%. Most of the cultivated land in the study area is loamy or 35%.



Count	Sum	Mean	Median
99	2535	25.6061	25
St dev	**Minimum**	**Maximum**	**Range**
4.67492	15	35	20
Q1	**Q3**	**IQR**	**Missing values**
25	25	0	0

Figure 7. Boxplot of soil clay data and crop yield. The table on the right shows' descriptive statistics of soil texture.

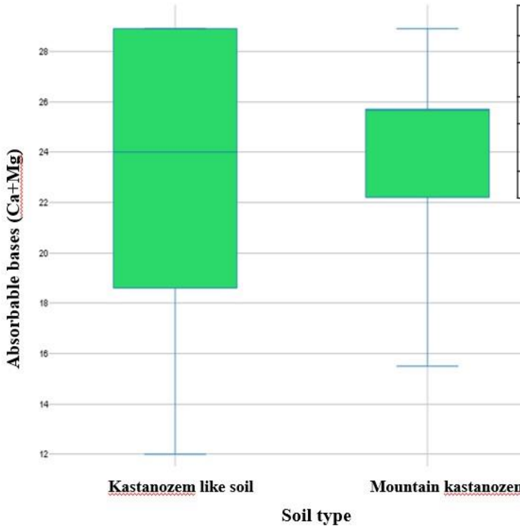
The humus content of the soil (Figure 8) ranged from 1.5 to 3.04%, with an average value of 1.61, indicating brown soil. A standard deviation of 0.39% and an interquartile range (IQR) of 0.51% indicate variation in soil humus content. Most data values are between 2.3 and 2.9%.



Count	Sum	Mean	Median
99	249.23	2.517	2.61
St dev	**Minimum**	**Maximum**	**Range**
0.39	1.5	3.04	1.54
Q1	**Q3**	**IQR**	**Missing values**
2.29	2.8	0.51	0

Figure 8. Boxplot of soil humus data and soil type. The table on the right shows' descriptive statistics of humus of soil.

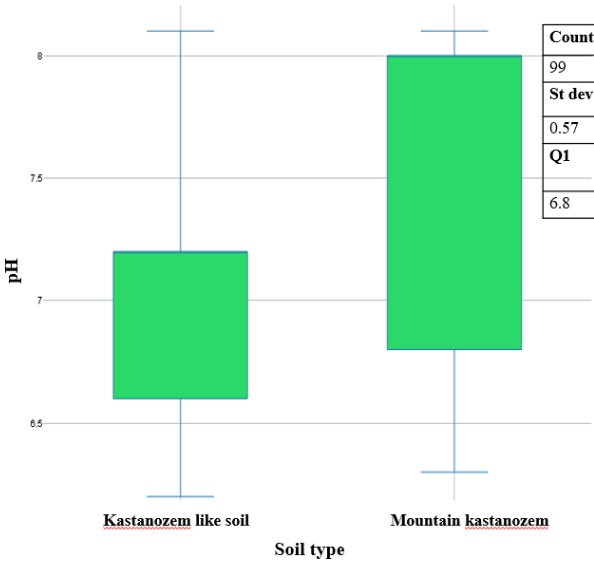
The total amount of potassium (Figure 9) was 12-28.9 mg/eq, which classified the fields as low, medium, and high potassium supply. A standard deviation of 4.53 mg/eq and an interquartile range (IQR) of 6.9 mg/eq indicate variability in the data. The mean value was 23 mg/eq, indicating a high potassium supply, with a maximum of 19–29 mg/eq.



Count	Sum	Mean	Median
99	2282.9	23.05	24.1
St dev	**Minimum**	**Maximum**	**Range**
4.53	12	28.9	16.9
Q1	**Q3**	**IQR**	**Missing values**
18.8	25.7	6.9	0

Figure 9. Boxplot showing the sum of the potassium data and soil type. The table on the right shows' descriptive statistics of the sum of the absorbed of soil.

Soil pH values ranged from 6.2 to 8.1, indicating a mix of acidic, alkaline, and neutral soils. The average value is 7.18%, and the maximum spread is 6.9-8.0%, indicating that alkaline soils are predominant. A standard deviation (St. dev.) of 0.57 % and an interquartile range (IQR) of 0.9 % indicate variability in soil pH data (Figure 10).



Count	Sum	Mean	Median
99	711	7.18	7.2
St dev	**Minimum**	**Maximum**	**Range**
0.57	6.2	8.1	1.9
Q1	**Q3**	**IQR**	**Missing values**
6.8	7.7	0.9	0

Figure 10. Boxplot of soil pH and soil type. The table on the right shows' descriptive statistics of pH of soil.

3.3 Crop type unsupervised classification

Unsupervised classification was performed because there was no data on crops planted in 2017 for 99 fields in the study area. In doing so, by using the Semi-automatic classification plugin of the QGIS program, a cluster of crop plots was created and classified into 4 types.

0.06-0.32 µm when looking at the electromagnetic wave plot by plotting class clusters as shown in Figure 11.

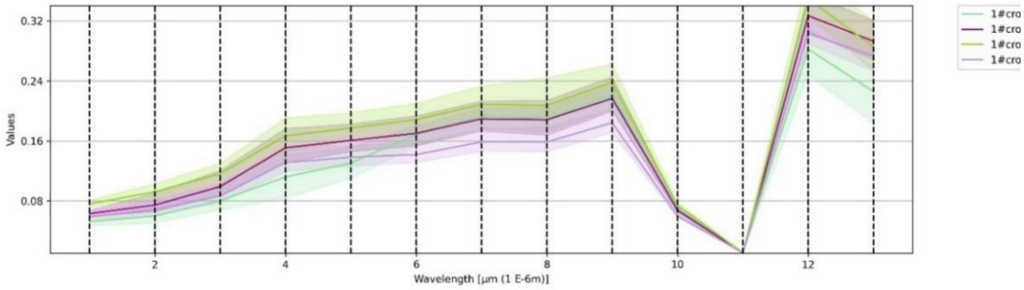


Figure 11. Wavelength graphic of cropland clusters.

Out of a total of 99 croplands, 22 were classified as crop 1, 15 as crop 2, 11 as crop 3, and 51 as crop 4 (Figure 12).

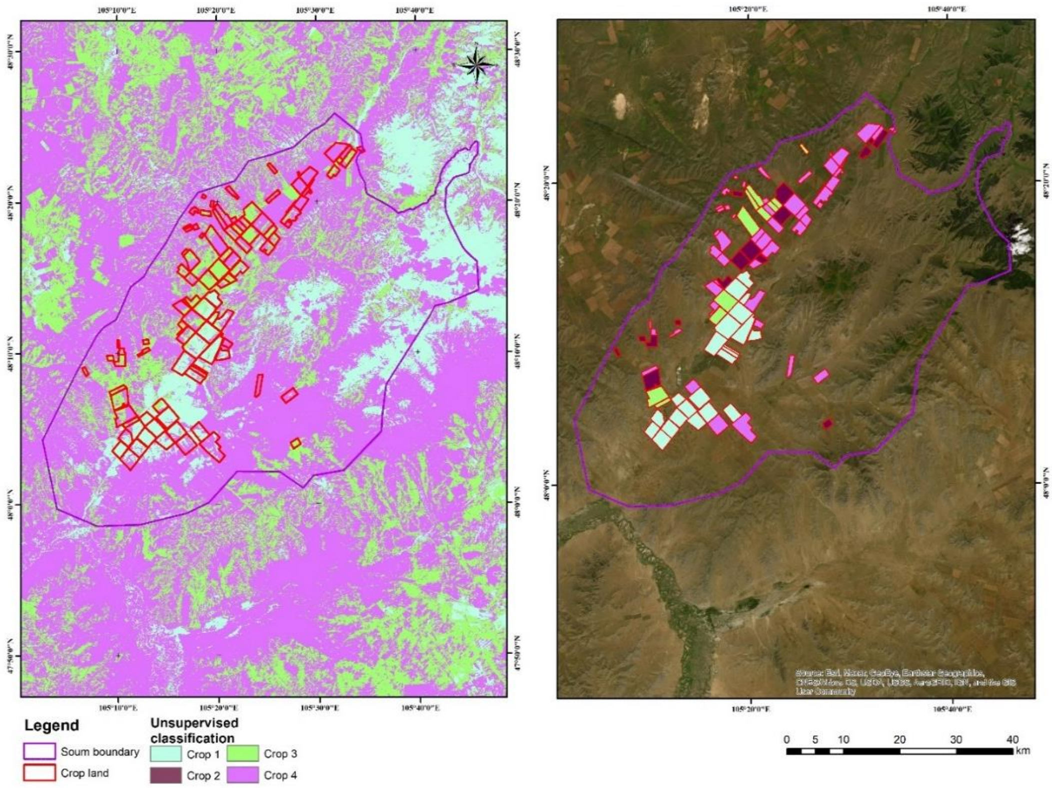


Figure 12. Unsupervised classification of cropland.

3.4 Relationship between vegetation indexes and field data

NDVI, NDMI, and LAI determined by the data of August 1, 2017, were found to be not correlated with crop yields (Table 5). When calculating the correlation between the indices determined by the Sentinel 2 satellite data and the field research data, it was found that the yield of the cultivated area was positively correlated (0.01-0.17), and other soil parameters were negatively correlated (-0.21- -0.06).

Table 5. Correlation between the indexes and field data.

	Crop yield	Humus	pH	Potassium	Clay
NDVI	0.165633	-0.19611	-0.09473	-0.0896	-0.06788
SMI	0.013968	-0.175	-0.03714	-0.09812	-0.11603
LAI	0.172571	-0.20783	-0.10368	-0.11485	-0.06567

Next, the correlation of the indices with the yield of each class classified by type of crop was calculated. In this way, the crop yield of crop 1 had a positive and low correlation with the indices (0.23-0.35). For class crop 2, NDVI and LAI had a positive and low correlation (0.44-0.45), and NDMI had a negative and no correlation (-0.01). For crop 3 and 4, NDVI and LAI had no correlation (0.01-0.02), and NDMI had no correlation (-0.15- -0.14) (Table 5, Figure 13-14).

Table 6. Correlation between the indexes and crop type's yield.

Index	Crop yield			
	Crop 1	Crop 2	Crop3	Crop 4
NDVI	0.317841	0.441252	0.020638	0.017236
NDMI	0.235447	-0.015662	-0.144209	-0.152957
LAI	0.353492	0.446997	0.016155	0.010362

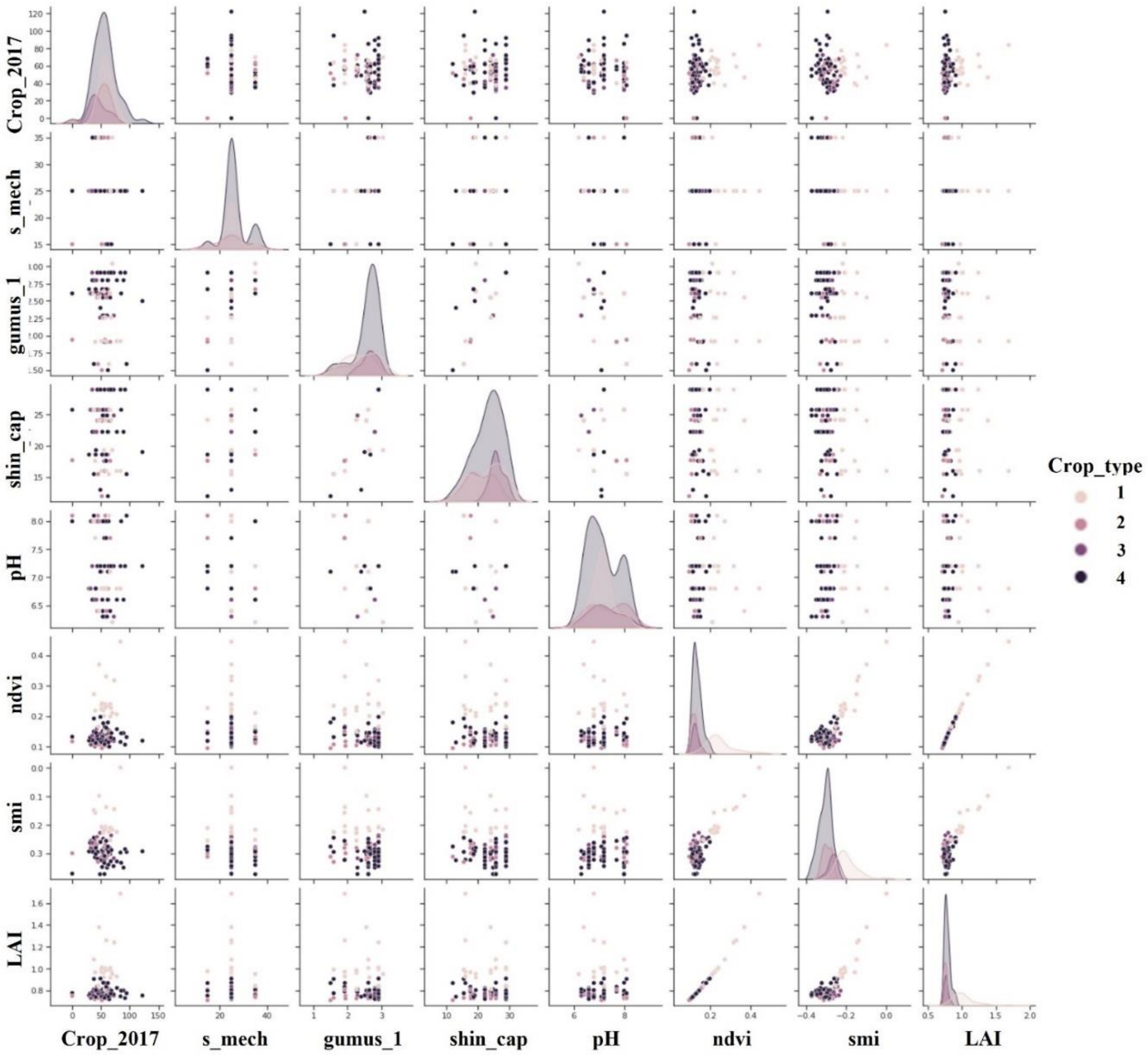


Figure 13. Scatterplot Matrix for crop yield (crop_2017), indexes (NDVI, LAI, smi for NDMI), and soil data (s_mech for clay content, humus, shin_cap for potassium, pH).

	crop_2017	s_mech	gumus_1	shin_cap	pH_1	ndvi	smi	LAI	crop_type
crop_2017	1.000000	-0.064046	-0.040496	-0.002628	-0.107901	0.165534	0.063269	0.169739	-0.016909
s_mech	-0.064046	1.000000	0.369750	0.150851	-0.117053	-0.056028	-0.111741	-0.053973	0.112580
gumus_1	-0.040496	0.369750	1.000000	0.734308	-0.253810	-0.190280	-0.158846	-0.192818	0.216781
shin_cap	-0.002628	0.150851	0.734308	1.000000	0.078426	-0.090697	-0.088846	-0.106260	0.090747
pH_1	-0.107901	-0.117053	-0.253810	0.078426	1.000000	-0.097002	-0.039137	-0.103149	-0.016294
ndvi	0.165534	-0.056028	-0.190280	-0.090697	-0.097002	1.000000	0.854865	0.989071	-0.608864
smi	0.063269	-0.111741	-0.158846	-0.088846	-0.039137	0.854865	1.000000	0.860480	-0.678463
LAI	0.169739	-0.053973	-0.192818	-0.106260	-0.103149	0.989071	0.860480	1.000000	-0.579698
crop_type	-0.016909	0.112580	0.216781	0.090747	-0.016294	-0.608864	-0.678463	-0.579698	1.000000

Figure 14. Correlation Matrix for crop yield (crop_2017), indexes (NDVI, LAI, smi for NDMI), and soil data (s_mech for clay content, humus, shin_cap for potassium, pH).

3.5 Data interpretation

This study correlated data from the 2017 field survey with indices determined by Sentinel 2 satellite data. Yield data for each plot was used, but it was not known which crop type it was. Therefore, the cultivated crops were classified into four classes by the unsupervised classification method. It can be seen from the scatter plot (Figure 13) that calculates the relationship between index and yield that there are two types of crops. Correlation calculations showed that crop yields and indices were not correlated (Figure 14).

The reason for not obtaining the expected results might be the following:

- Crop yield records methodology: Durgun et al. (2020) conducted a survey on winter wheat yield in 2020. In their study, crop yield was measured by weighing. However, the methodology used to collect field yield data for this study was unclear.

- Crop yield records: Another possible issue with field data in this study is the lack of sufficient harvest information like crop type or harvest day. For example, Soriano-González et al. (2022) highlighted the need for detailed harvest information. They recommended to record at least the variety of crops, and harvesting and planting time.
-
- Date of selected satellite data: The date of the selected satellite data may not have been appropriate. Because comparing the 2017 NDVI value of the surveyed area for four types of crops, some types of crops have high yields on July 15th and some types of crops have high yields on August 15th (Figure 15). This was revealed in GEE using Sentinel 2, Landsat 8 and MODIS data.

In addition to the reasons mentioned above, crop monitoring studies are often conducted using time series analysis. For example, Tuvdendorj et al. (2019) analyzed 18 years of data from June to August when determining the spring wheat yield using satellite data. Another example is Soriano-González et al. (2022) who calculated NDVI, NDWI, and BSI using Sentinel 2 A/B data for the entire 2-year period between 2018 and 2019 for monitoring wheat yields. Therefore, it may be possible to use time series analysis to extend this study further.

Figure 15. Average NDVI value for estimated crop types.

4. DISCUSSION

The main objective of this study was to evaluate the use of remote sensing data for crop estimation in Mongolia. A few remote sensing methodologies were studied to achieve this goal. However, the results of the study differed from the results of other researchers. NDVI, NDMI, and LAI determined were either not related to yield or the correlation coefficients were 0.16, 0.06, and 0.17.

For comparison, the results of the following researchers who conducted the same type of research were considered. Ali et al. (2019) calculated the correlation of grain yields using the NDVI, EVI, SAVI, GNDVI, GCI, and SR indices over an area of 11.07 hectares in northern Italy over 5 years. Grain yields were calculated using Landsat satellite data, with a 64-86 % correlation. Ma et al. (2022) calculated leaf area index from Sentinel 2 satellite data in a simple wheat yield estimation algorithm and compared it to 48 field measurements in 2014 and 2015. The results showed that the calculated values were 65-92 % correlated with the measured values. The highest correlation (r^2=0.55) was given by NDWI, VSDI, and NDVI among the nine indexes calculated in a study by Tuvdendorj et al. (2019).

In another example, Gansukh et al. (2020) monitored wheat growth using Sentinel 1 and Sentinel 2 satellite data. NDVI and NDWI were calculated between May 2019 and September 2019. The highest values of NDVI and NDWI were after August 15. Therefore, it would be wise to use all the data from the plant growth period or the data after August 15.

Field data could also be improved and re-collected to redo this research. In that case, the type of crops planted and the methodology for recording the yield should be clarified. Furthermore, it may be necessary to choose the date of the satellite data in accordance with the timing of crop growth.

5. CONCLUSIONS

The study was conducted using data from the Ugtaaltsaidam sum field, which is located in the central agricultural region of Mongolia. Crop yield and selected soil parameters were evaluated with NDVI, NDMI, and LAI. Vegetation indices were estimated from Sentinel 2 satellite data. Indices calculated from satellite data on August 1, 2017 showed no to very weak correlation with crop yields and soil parameters.

The research did not achieve the expected results. This may be due to the fact that the method of recording crop yields was unclear, the type of crop was unknown, and the data dates selected were inappropriate. Therefore, in order to further improve this research, following points should be considered:

- Identify the types of crops grown in the field
- Clarify the methodology for determining the yield of cultivated plants
- Determine the period of growth and maturity of cultivated plants
- Conduct time series analysis whenever possible
- Increase the indices to calculate the dependence

To conclude from the above, it is necessary for Mongolia to improve the cropland database and intensify research on cropland monitoring using remote sensing.

ACKNOWLEDGEMENTS

First of all, I would like to thank my colleagues at GRÓ-LRT for providing me with scientific knowledge about land degradation and its restoration, as well as learning new research methods and developing my personal and research skills.

I am very grateful to GRÓ-LRT director Dr. Sjöfn Vilhelmsdottir and deputy director Berglind Orradottir, who gave valuable suggestions, project manager Brita Berglund, who coordinated my research work and gave valuable advice, and operations manager Halldora Traustadottir, who helped me in every way to live in Iceland.

In addition, I am deeply grateful to Michaela Hrabalikova, who guided my research work, taught me many new research methods, supported me, and guided me in my future work.

I would also like to express my sincere gratitude to Isabel C. Barrio, Olafur Arnalds, Asa L. Aradottir, Brita Berglund, and Johann Thorsson for teaching me, broadening my knowledge, encouraging me, and giving me valuable advice during the program.

I want to express my appreciation to the administration of the school of Agroecology, Mongolian University of Life Sciences, my colleagues at the Land management department, and my supervisor, Dr. Associate Professor D. Tuvshinbayar, for giving me the opportunity to participate in this training.

Also, I wanted to express my gratitude to Dr. B. Batbileg and D. Byambasuren who always gave advice and shared information.

LITERATURE CITED

(EROS) Earth resources observation and science center (n.d.) 50 years of food security and spatial research. https://www.usgs.gov/news/featured-story/fifty-years-landsat-pioneer-promoting-food-security-space

Agricultural database (2022) Mongolian statistic information center. https://www.1212.mn/stat.aspx?LIST_ID=976_L10_2 (accessed 6 June 2022)

Ajindai C (1999) Тариалангийн технологийн үндэс [Fundamentals of agricultural technology]. Ulaanbaatar (in Mongolian)

Alberto (2022) Measuring moisture: Normalized Difference Moisture Index (NDMI) Sentinel-2 2022. https://geovisualization.net/2022/07/14/measuring-moisture-normalized-difference-moisture-index-ndmi-sentinel-2-2022/

Ali A, Martelli R, Lupia F, Barbanti L (2019) Assessing multiple years' spatial variability of crop yields using satellite vegetation indices. Remote Sensing 11

Batzorig E (2011) Монгол орны газар тариалангийн төв бүсийн кадастрын зураглал мониторинг хийх [Monitoring and mapping of agricultural regions of Mongolia using survey and cadastral mapping methods] (in Mongolian)

Batzorig E, Banzragch B (2012) Mapping and monitoring principal crop land cover/ use changes in Mongolia using remote sensing. International Geoscience and Remote Sensing Symposium (IGARSS):2261-2263

Development Ministry of Economy (2022) Орон нутгийн хөгжлийн индекс [Local Development Index]. http://ldi.nda.gov.mn/?province_id=18&region_id=377 (in Mongolian)

Duncan JMA, Dash J, Atkinson PM (2015) The potential of satellite-observed crop phenology to enhance yield gap assessments in smallholder landscapes. Frontiers in Environmental Science 3:1-16

Durgun YÖ, Gobin A, Duveiller G, Tychon B (2020) A study on trade-offs between spatial resolution and temporal sampling density for wheat yield estimation using both thermal and calendar time. International Journal of Applied Earth Observation and Geoinformation 86:101988

ESA (2022) No title. https://directory.eoportal.org/web/eoportal/satellite-missions/content/-/article/sentinel2

ESRI (2022)(a) ArcGIS pro. https://pro.arcgis.com/en/pro-app/latest/tool-reference/spatial-analyst/how-zonal-statistics-works.htm

ESRI (2022)(b) Base map.

Franch B, San Bautista A, Fita D, Rubio C, Tarrazó-Serrano D, Sánchez A, Skakun S, Vermote E, Becker-Reshef I, Uris A (2021) Within-field rice yield estimation based on sentinel-2 satellite data. Remote Sensing 13

Gansukh B, Batsaikhan B, Dorjsuren A, Jamsran C, Batsaikhan N (2020) Monitoring wheat crop growth parameters using time series sentinel-1 and sentinel-2 data for agricultural application in Mongolia. International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences - ISPRS Archives 43:989-994

Gao B (1996) NDWI-A normalized difference water index for remote sensing of vegetation liquid water from space. Remote Sensing of Environment 58:257-266

GISGeography (2022) Supervised and unsupervised classification in remote sensing. https://gisgeography.com/supervised-unsupervised-classification-arcgis/

Hunt ML, Blackburn GA, Carrasco L, Redhead JW, Rowland CS (2019) High resolution wheat yield mapping using Sentinel-2. Remote Sensing of Environment 233:1-37

Karlson M, Ostwald M, Bayala J, Bazié HR, Ouedraogo AS, Soro B, Sanou J, Reese H (2020) The potential of Sentinel-2 for crop production estimation in a smallholder agroforestry landscape, Burkina Faso. Frontiers in Environmental Science 8:1-13

Liang L, Di L, Zhang L, Deng M, Qin Z, Zhao S, Lin H (2015) Estimation of crop LAI using hyperspectral vegetation indices and a hybrid inversion method. Remote Sensing of Environment 165:123-134

Lonjid T (1995) Газар тариалан [Crop production]. Ulaanbaatar. (in Mongolian)

Ma C, Liu M, Ding F, Li C, Cui Y, Chen W, Wang Y (2022) Wheat growth monitoring and yield estimation based on remote sensing data assimilation into the SAFY crop growth model. Scientific Reports 12:1-16

Mamatkulov Z, Safarov E, Oymatov R, Abdurahmanov I, Rajapbaev M (2021) Application of GIS and RS in real time crop monitoring and yield forecasting: A case study of cotton fields in low and high productive farmlands. E3S Web of Conferences 227

Ministry of Construction and Urban Development (2019) Тариалангийн газрын төлөв байдал, чанарын улсын хянан баталгааны ажлын аргачилсан заавар [Methodical instructions for the state control and assurance of the condition and quality of agricultural land]. (in Mongolian)

Mongolian environmental database (2022) Газар зүйн мэдээллийн сан [Geoinformation database]. https://eic.mn/(in Mongolian)

Muruganantham P, Wibowo S (2022) A systematic literature review on crop yield prediction with deep learning and remote sensing. Remote Sensing
Observatory NASA Environmental (2000) Measuring vegetation.
https://earthobservatory.nasa.gov/features/MeasuringVegetation/measuring_vegetation_2.php

Soriano-González J, Angelats E, Martínez-Eixarch M, Alcaraz C (2022) Monitoring rice crop and yield estimation with Sentinel-2 data. Field Crops Research 281

Tewari S, Kulhavý J, Rock BN, Hadaš P (2003) Remote monitoring of forest response to changed soil moisture regime due to river regulation. Journal of Forest Science 49:429-438

Thenkabail PS (2010) Global croplands and their importance for water and food security in the twenty-first century: Towards an ever green revolution that combines a second green revolution with a blue revolution. Remote Sensing 2:2305-2312

Tsevengee E (2016)(a) Төрөл бүрийн хиймэл дагуулын мэдээг ашиглан Монгол орны бэлчээр, тариалангийн талбайн өөрчлөлтийг тодорхойлох [Determining the growth of pastures and arable land in Mongolia using various satellite datad]. Газрын харилцаа 2016 [Land communication 2016] (in Mongolian)

Tsevengee E (2016)(b) Тариалангийн талбайн зураглалын харьцуулсан үр дүн [Comparative results of field mapping]. Газрын харилцаа 2016 [Land communication 2016] (in Mongolian)

UN (2021) Pathways to zero hunger. https://www.un.org/zerohunger/content/challenge-hunger-can-beeliminated-our-lifetimes

Watson DJ (1947) Comparative physiological studies on the growth of field crops: I. variation in net assimilation rate and leaf between species and varieties, and within and between years. Annals of Botany 11:41-76

Wilson EH, Sader SA (2002) Detection of forest harvest type using multiple dates of Landsat TM imagery. Remote Sensing of Environment 80:385-396

World Health Organization (2021) World hunger is still not going down after three years and obesity is still growing. https://www.who.int/news/item/15-07-2019-world-hunger-is-still-not-going-down-after-three-years-and-obesity-is-stillgrowing- un-report