SOIL SCIENCE ANNUAL
2024, 75(2), 189540

https://doi.org/10.37501/soilsa/189540

Assessment of soil impact on pre- and post-harvest NDVI extrema by machine learning

Artur Łopatka[1]*, Piotr Koza[1], Beata Suszek-Łopatka[1], Grzegorz Siebielec[1], Jan Jadczyszyn[1]

[1] Institute of Soil Science and Plant Cultivation – State Research Institute, Department of Soil Science Erosion and Land Protection, Czartoryskich 8, 24-100 Puławy, Poland

* Corresponding author: MSc, Artur Łopatka, artur@iung.pulawy.pl, ORCID iD: https://orcid.org/0000-0002-6977-4464

Received: 2023-11-22
Accepted: 2024-06-01
Published online: 2024-06-01
Associated editor: Łukasz Uzarowicz

Keywords:

Soil Agricultural Map
Remote Sensing
NDVI
Crop harvest
Machine Learning
Sentinel-2

Abstract

It was observed that the difference in the maximum and minimum NDVI values at a time close to harvest (mxNDVI and mnNDVI, respectively), referred to as the haNDVI index (harvest amplitude of NDVI), correlates with agricultural soil quality and the share of sowings. The NDVI becomes saturated when the values of the Leaf Area Index (LAI) significantly exceed one so spatial variation in haNDVI is mainly due to the minimum post-harvest NDVI (mnNDVI). To explain the variability of mnNDVI values three hypotheses were formulated: i) impact of crop selection, ii) field size impact, and iii) impact of soil. To determine which of these hypotheses had the highest impact on the variation in the mnNDVI, the developed machine learning models of this indicator were subjected to a test removing individual explanatory variables from them. Removing a variable does not cause a significant increase in model error if a variable does not contribute useful information to the model. This test showed that the mnNDVI index depends almost exclusively on the crop indicator which was the median of mnNDVI for crops, not directly from soil variables such as the agricultural quality of soil or soil moisture. According to this, the hypothesis of direct impact of soil was rejected. The explanation for the observed correlation of haNDVI with soil quality is the agricultural practice of choosing crops with low mnNDVI (cereals, rapeseed) at better soil conditions and crops with high mnNDVI (fodder crops, grassland) for worse soil conditions.

1. Introduction

Meeting the challenge of sustainable development in agriculture creates a need for continuous monitoring of soil conditions and environmental changes in rural areas. Indicators based on data provided by farmers and indicators based on measurements (e.g. Faber, 2007; Harasim, 2013) are available with a delay compared to the current state and have limitations related to the aggregation method (e.g., the indicator calculated for municipalities is difficult to assign to natural units such as river catchments). Moreover, they are often subject to methodological changes, which limits the interpretation of long-time series and ambiguities. An alternative to traditional indicators are remote sensing data, particularly satellite images. These indicators are base for surveilling the natural environment (weather patterns, flooding, drought, crop productivity, atmospheric pollution, landscape diversity, and cyanobacterial blooms). Their advantages are up-to-date, simultaneity and long range, low cost of development, and the possibility of processing archival data when new analytical methods are available and when changes in the methodology occur.

The Normalized Difference Vegetation Index (NDVI) has been proven to be particularly useful for remote sensing of plant conditions and development (Rouse et. al., 1974; Tucker 1979). NDVI is proportional to the fraction of the soil surface covered by the photosynthetically active surface of plants and is a component of numerous indicators used to assess yield levels, agricultural drought, fertilization needs, plant diseases, etc. Agricultural crops typically exhibit a clear minimum NDVI value in the post-harvest period (e.g., Sicre et al., 2016), while forests, orchards, and grasslands, except for short periods of mowing and grazing, maintain a relatively constant NDVI that only decreases in autumn. The NDVI values for crops, especially after harvest, are related to the reflection of radiation from the soil surface and crop residue. The maximum NDVI values recorded before harvest were related to the reflection of radiation from the canopy at maximum density.

It was observed (Łopatka and Koza, 2020) that the difference in the maximum and minimum NDVI values close to the harvest period (mxNDVI and mnNDVI, respectively), referred to as the haNDVI index (harvest amplitude of NDVI), correlates with agricultural soil quality measured by the WWRPP index (Witek, 1993) (r=0.60), NPK fertilization intensity (r=0.69), and the share of sowings in the total area (r=0.86) for communes in low-resolution satellite images (MODIS; pixel 1 km). For high-resolution images (Sentinel-2; pixel 20 m), a preliminary visual analysis of the haNDVI index values within the fields of the Agricultural Experimental Station (AES) of IUNG "Kępa-Puławy" and their surroundings indicated (Łopatka and Koza, 2020) that they are higher for soils with higher agricultural quality and significantly lower on similar soils outside the AES boundaries where fertilization is much lower.

Because the NDVI becomes saturated when the values of the Leaf Area Index (LAI) significantly exceed one (Jones and Vaughan, 2010), it is less sensitive to differences in leaf area when plants are in the full development phase at dates close to the harvest period. Therefore, the observed differences in the haNDVI = mxNDVI – mnNDVI index at the farm scale cannot be explain by differences in the level of maximum NDVI before harvest (mxNDVI). Consequently, the spatial variation in haNDVI is mainly due to the minimum post-harvest NDVI (mnNDVI). This conclusion is confirmed by analysis of the mxNDVI and mnNDVI maps for Poland (Fig. 1), where the mnNDVI diversity in agricultural areas is higher than the mxNDVI diversity. Furthermore, the mnNDVI map, which shows the NDVI at a minimum after the harvest, clearly reflects Poland's soil quality and planting share. The variability of NDVI before harvest is similar to that of the difference between annual precipitation and evapotranspiration. Therefore, in areas with high agricultural soil quality and high fertilization, as well as a high share of sowings, a low level of mnNDVI is observed.

The problem to be solved is to identify the reasons for the variation in the level of NDVI at minimum after harvest and its correlation with the crop production indicators. Before undertaking the research, three hypotheses were formulate:

1. Agronomical – in areas with intensive production (located mainly in lowland areas with a spatial concentration of soils of higher agricultural quality), the crop structure is dominated by cereals, which, unlike fodder plants, dry out and lose chlorophyll before the harvest period. Moreover, when fields are flat, the grain is mown low (modern combines automatically adjust header height).
2. Structural – in the case of remote sensing images with low resolution (raster pixels with a side longer than 10 m), in addition to the NDVI of the analyzed crop, NDVI values from the field boundaries (field margins, mid-field trees, etc.) were also read. That means that after harvest, the NDVI recorded is higher than NDVI inside agricultural fields, especially in areas where the plots are very narrow. In areas

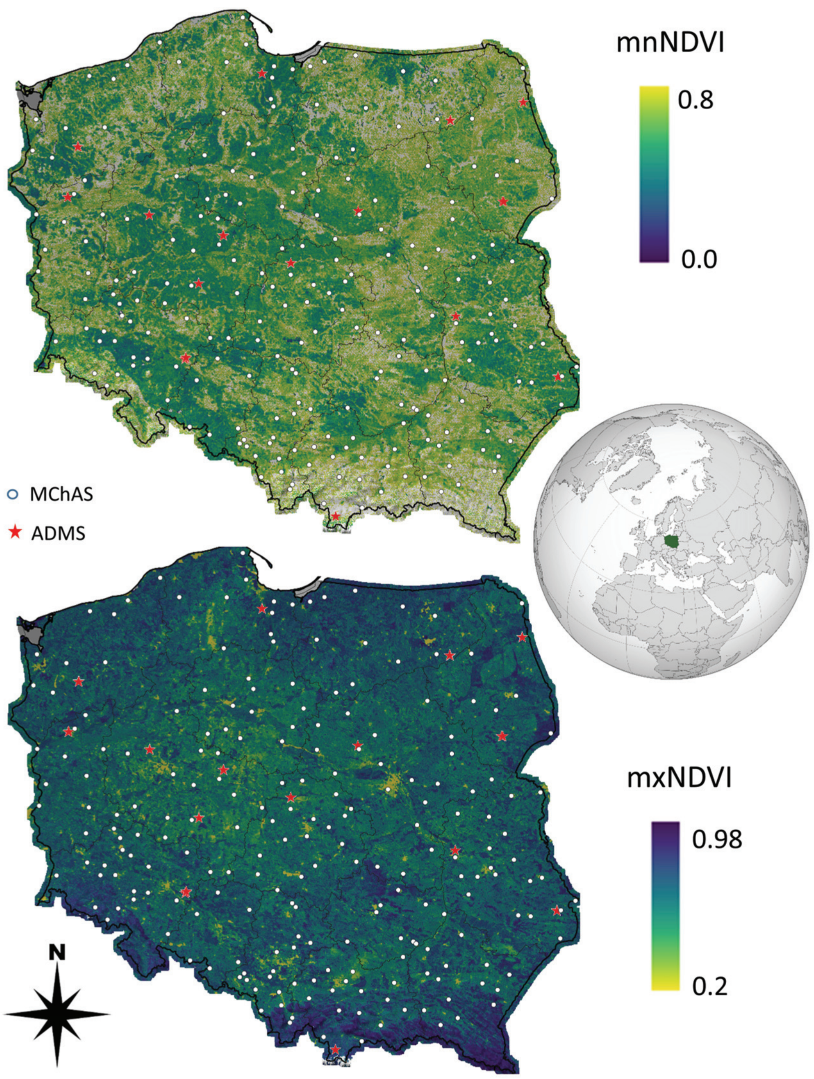

Fig. 1. Locations of measurements of the medians mnNDVI and mxNDVI for individual crops (MChAS) and soil moisture content (ADMS) against the background of long-term mxNDVI and mnNDVI in Poland

with small and narrow field plots, the production intensity (use of fertilizers, pesticides, and retardants) was low for economic reasons. Given these circumstances, nutrient deficiencies and illnesses can emerge, resulting in a decrease in the highest pre-harvest NDVI level and an increase in the probability of crops being abandoned in the field due to weed overgrowth, disease, pests, or lodging.

3. Soil – after harvest, the soil is partially visible from under the crop residues; the higher its agricultural quality (higher humidity, higher humus content), the darker the soil and the lower its NDVI.

Understanding the reasons for the correlation between mnNDVI and production intensity may contribute to the development of remote sensing of agricultural soil quality or agricultural intensity level. In particular, if the third soil hypothesis is confirmed, mnNDVI will be easy to obtain measure for agricultural soil quality. This means a simplification of the current methodology for remote sensing of soil properties, which most often require analyzing rare images of fields without vegetation cover, which limits the study of large areas.

2. Materials and methods

2.1. Hypothesis verification

To verify each hypothesis, it was tested whether, in the mnNDVI model, the variables with the greatest contribution to the reduction of model error are: i) an index of land roughness and information on the cultivated crops with the agronomical hypothesis; ii) an area of agricultural plots in the case of the structural hypothesis; and iii) relative soil moisture and water capacity of the soil arable layer and agricultural soil quality in the case of the soil hypothesis.

To determine which of the initial hypotheses had the highest impact on the variation in the mnNDVI, the developed models of this indicator were subjected to a test to remove individual variables from them. Removing should not cause a significant increase in model error if a variable does not contribute useful information to the model. After removing the model variables that contain information that is important for the analyzed phenomenon and do not duplicate the information in other variables, the model's ability to predict should decrease, resulting in error growth. The model error was measured as the mean RMSE (the square root of the mean of the squared model residuals) in the 10-fold cross-validation procedure. Cross-validation folds were selected from the full number of observations using random sampling.

Because the choice of crop type is determined to a large extent by the agricultural quality of the soil, we decided to observe the elimination of model variables while constantly present in the model variable corresponding to the crop impact on NDVI. If, for example, the agricultural soil quality variable turns out to be significant in such a model, it will mean that the soil is actually visible from under crop residues, and the observed impact of soil quality on NDVI is not only the result of correlation between crop type and soil quality.

2.2. Data preparation

At the research planning stage, because of time-consuming process of downloading and pre-processing satellite images, faster processing of images from the MODIS sensor with spatial resolutions of 1 km, 500 m, or 250 m was planned. However, during the project, the possibility of processing satellite images by cloud computing, without downloading and storing them on local hard drives was noticed. Such opportunities are provided by among others, the Google Earth Engine (GEE) platform (Goerlick et al., 2017; Zhao et al., 2021) used in the task. As a result, it was possible to quickly process Sentinel-2 imagery with a spatial resolution of 10 m (in the NIR and RED bands necessary to calculate the NDVI index). Such an image pixel does not contain a mixture of arable fields, balks, and trees; therefore, testing the structural hypothesis was limited to the more interesting problem of the direct influence of the area of agricultural plots.

On the GEE platform, a code has been run to select clear Sentinel-2 image pixels above selected monitoring locations (using cloudBitMask and cirrusBitMask parameters in the QA60 band and clouds, cirrus clouds, and shadow parameters in the SCL band). Images recorded between 2018 and 2022 were filtered from geometrically and atmospherically corrected images the 'COPERNICUS/S2_SR_HARMONIZED' collection. The NDVI values obtained from images are, at best, in the absence of cloud cover, recorded on Poland area approximately every one-two days according to Sentinel-2 revisit time (5 days) shortened by effect overlapping swaths from adjacent orbits. To get more continuous time series, NDVI daily values were aggregated into decade periods. The NDVI time series were not smoothed because the use of typical methods such as the Savitzky-Golay filter in the case of cultivated vegetation carries the risk of reducing the NDVI extremes examined in this work. In particular, this applies to the post-harvest minimum which follows the sharp decline in NDVI observed in the case of fodder crops and grass vegetation. The decade NDVI patterns developed in this way were combined with the database with information on the crop structure of individual agricultural plots in selected locations.

The soil moisture data came from locations where soil moisture measurements were conducted as a part of the Agricultural Drought Monitoring System (ADMS 2023). To determine the crop dependent average values of mnNDVI and mxNDVI, the soil measurement points of the Monitoring of the Chemistry of Arable Soils (MChAS 2023) were selected. The selected locations cover the entire country area, although, with the ADMS, there was only one mountain location (Fig. 1). The aggregated decadal NDVI values were read from a single pixel 10 by 10 m in which a given monitoring point was located.

The locations from both monitoring sites were evaluated to eliminate potential non-crop sources for the reading in a 10x10 meters pixel. The orthophoto maps that were currently available on the Geoportal.gov.pl website were used for the review. Locations where roads, trees or shadows were observed within 15 meters of the precise measurement position, were excluded. Out of the 52 ADMS locations with continuous observations from 2018 to 2022 and a nearby meteorological station, 19 were excluded from the analysis. In case of MChAS locations, 76 out of

216 places were removed. Models of mnNDVI and mxNDVI was constructed on N=141 ADMS data records (33 ADMS locations × 5 years with removed years without full data set).

2.3. Calculation of the crop impact index on mnNDVI and mxNDVI

The variable responsible for the influence of the crop type on mnNDVI (or mxNDVI) was assessed by converting a non-parametric variable, which is the crop type, into a parametric variable, the median value of nmNDVI (or mxNDVI) measured for this crop in numerous other locations.

To get the variable of crop impact on the mnNDVi and mxNDVI indicators, their values were calculated from the NDVI time series recorded in the subsequent years 2018–2022 in selected 140 MChAS locations. To process the 700 NDVI time series (5 years × 140 locations), a code dedicated to this task was developed in R. Medians of mnNDVI and mxNDVI for individual crops were calculated from the time series recorded at least five times, which minimizes the impact of outlier observations (e.g., related to the discrepancy between the farmer's crop declaration in the ARMA database and the actual situation).

This task was performed for two definitions of mnNDVI and mxNDVI, differing in the choice of extreme dates defining the time in which the minimum and maximum were searched. The first definition was used in a previous study on the haNDVI index (Łopatka and Koza, 2020) and required the maximum NDVI in the period from May 15 to September 30 and the minimum in the period from June 15 to October 30. This translates for the minimum into the decades from 17 to 31, and for the maximum to the decades from 14 to 28. It was also observed that in some situations, the minimum values for maize and sugar beet are registered before the harvest, which is contrary to the general idea of the indicator, and the maximum values for winter wheat and rapeseed are registered in the period after the harvest on the next winter crop emerging in the same year (Fig. 2). To reduce this problem, we selected a better range of extreme dates, which were performed by trial and error. Satisfactory results (Fig. 3) were obtained for the minimum, assuming the period between the 19th and 34th decades, and for the maximum between the 10th and 24th decades. Therefore, the crop impact indexes on nmNDVI and mxNDVI were obtained as the median of nmNDVI and mxNDVI for individual crops with the optimal selection of dates.

As seen in Figure 4, the variability of mxNDVI between crops was small compared to the variability of mnNDVI. It can be seen that crops with the most frequently used high inputs have a lower mnNDVI than extensively cultivated plants. The exception applies only to sugar beets and is probably related to the frequent leaving of beet leaves in the field after harvest.

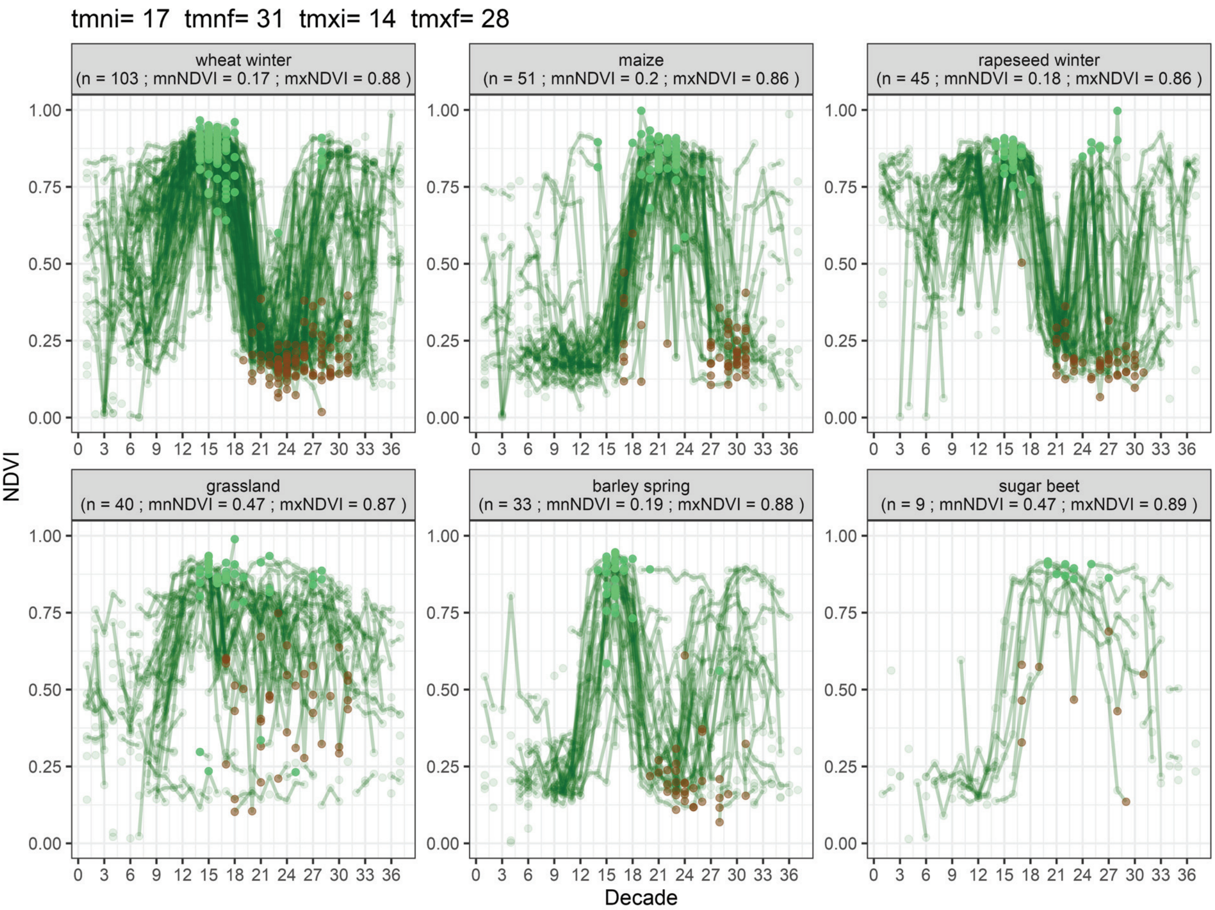

Fig. 2. NDVI time series with mnNDVI highlighted in brown and mxNDVI in green for the original time frame definition

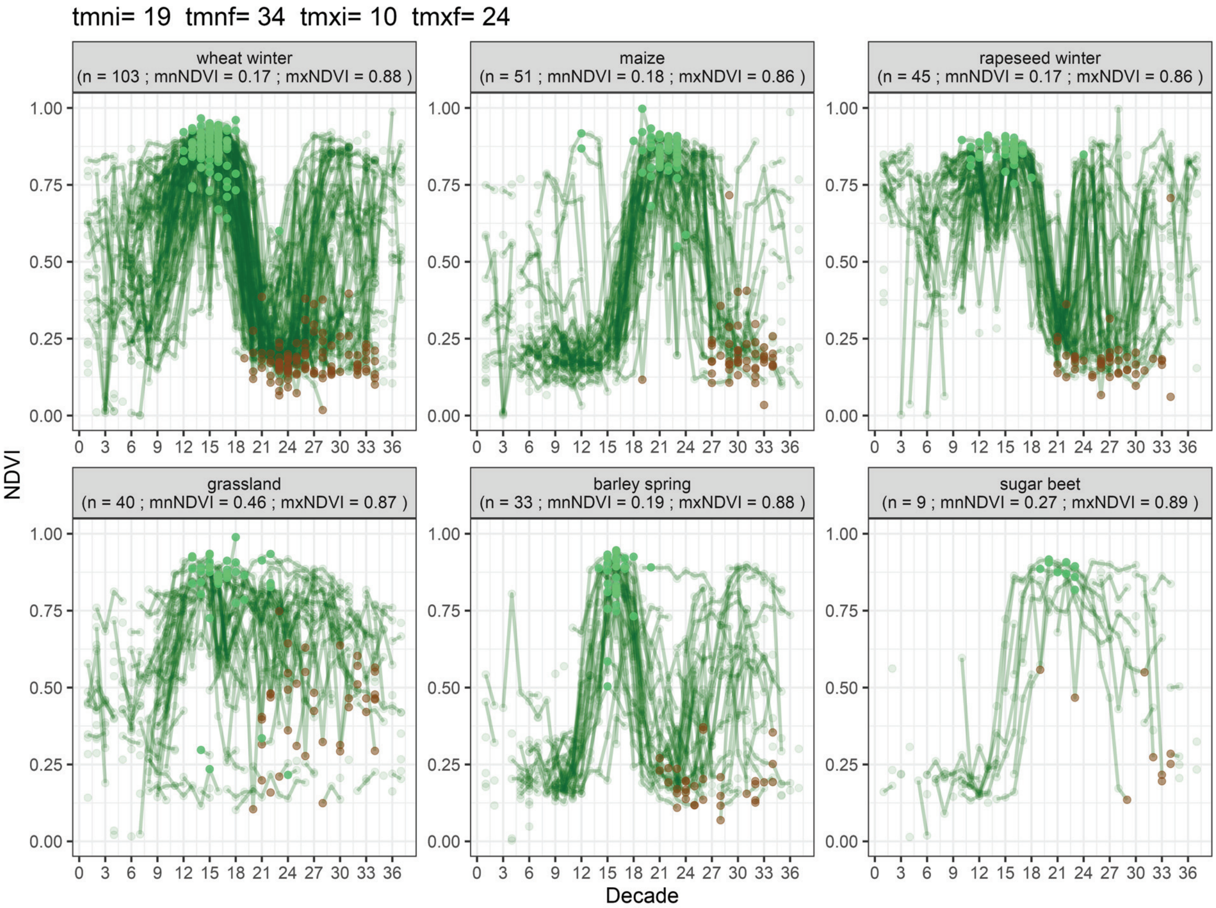

Fig. 3. NDVI time series with mnNDVI highlighted in brown and mxNDVI in green for the optimal time frame definition

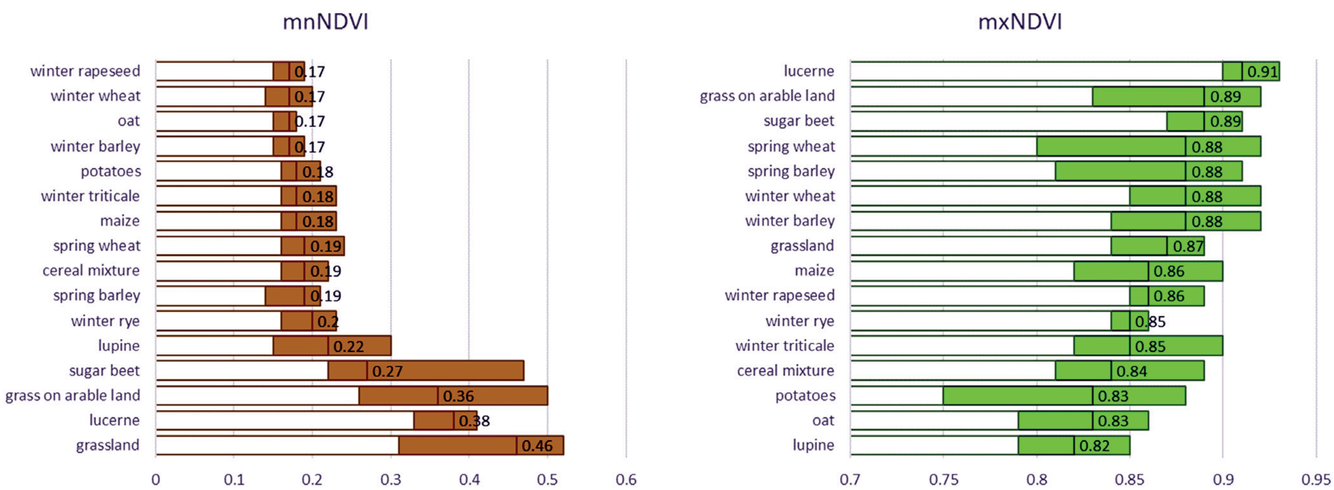

Fig. 4. Median mnNDVI and mxNDVI for individual crops from MChGO locations in 2018–22. The color bar shows the range between the 1st and 3rd quartiles. The line on the bar and the number is the median value

2.4. Variables and models

The explanatory variables were selected using the expert method, trying to take into account all the most important factors affecting the variability of NDVI. As explanatory variables were selected all drivers that were expected as the most important factors affecting the variability of NDVI. Soil moisture, besides its direct impact on albedo (Angström 1925; Liu et al., 2002; Lobell, Asner 2002), may, as a result of larger and long-term deficits, also directly affect the amount of chlorophyll in plant leaves and the value of the NDVI. Effects similar to drought but not necessarily related to water deficit in the soil may be the result of stress related to high air temperature. To record such effects, we include also a variable which is the number of days in a year in which the maximum daily air temperature exceeded 30^{0}C (Schauberger et al. 2017).

The most important information about the variables used in the analysis is included in Table 1, and their Spearman correlation coefficients are summarized in the matrix in Figure 5. The explained variables mnNDVI and mxNDVI revealed, as expected, a significant and strong correlation with the Cmn and Cmx variables, which are the medians of these variables for crops in the MChAS locations. In the case of explanatory variables, the strongest correlations were recorded for the following pairs: i) agricultural quality of soil (SQI) and the natural logarithm of agricultural plot area (lnPA), which indicates that plots on good soils are larger than on poor ones, and ii) number of hot days (HD) and relative soil moisture (SMI), which is an expected result due to the connection between high air temperature and high evaporation.

Due to the observed strong right-skewness of the distributions for TRI and PA, these variables were logarithmized for the purposes of linear regression and LASSO analyses. It is worth noting that the RF model, as a non-parametric method, does not require attention to the distribution of explanatory variables.

Because of the nonlinearity and lack of recognition of the detailed mechanisms of the studied processes, models of mnNDVI and mxNDVI in addition to the Ordinary Least Squares (OLS) linear regression used here as the reference method were constructed using two popular methods of supervised machine learning.

In the first method, the Least Absolute Shrinkage and Selection Operator (LASSO) linear regression analysis and variable selection (Tibshirani 1996) were performed simultaneously, which limits the effect of overfitting the model to the training data. The second selected method, Random Forest (RF), which is a set of decision trees, apart from reducing the effect of overfitting, allows for the reflection of non-linear dependencies in the model (Breiman 2002). Random Forest is a commonly selected first-choice method in machine learning because of the reasonably good prediction results achieved with the default model parameters.

According to this we build three models (linear regression, LASSO, RF) for each of the two explained variables. Estimation, calibration, and cross-validation were performed in the R Statistical Software (v4.3.0; R Core Team 2023) using scripts developed by the author's team, using the glmnet package (Friedman et al. 2010) for LASSO regression calculations, the randomForest package (Liaw, Wiener 2002) for the RF model, and the Caret package (Kuhn 2008) for cross-validation of all models.

Table 1
Explained and explanatory model variables

variable	volatility range				data source and preparation
	Minimum	Median	Mean	Maximum	
mxNDVI maximal NDVI before harvest (-)	0.57	0.86	0.84	0.93	10 m Sentinel2 data filtered from clouds and averaged in GEE; separate layer for each year
mnNDVI minimal NDVI after harvest (-)	–0.12	0.18	0.21	0.76	10 m Sentinel2 data filtered from clouds and averaged in GEE; separate layer for each year
Cmx crop impact index on mxNDVI	0.83	0.87	0.87	0.89	median mxNDVI values for individual crops based on NDVI time series at MChAS points (details of calculation in chapter 2.3)
Cmn crop impact index on mnNDVI	0.17	0.18	0.20	0.46	median mnNDVI values for individual crops based on NDVI time series at MChAS points (details of calculation in chapter 2.3)
SQI Soil Quality Index (-)	1	70	60.7	94	Soil Agricultural Map 1:25000 and table of agricultural value of soil complexes (Witek 1993)
SM95th 95th percentile of soil moisture (%) in the top 25 cm of soil	13.1	19.4	20.0	33.9	averaged profile probe measurements (ADMS IUNG, 2023); constant in time
SMI relative soil moisture of vegetation period on depth 0–25 cm (%)	29	64	64	90	averaged profile probe measurements (ADMS IUNG, 2023); separate values for each year
HD number of Hot Days (with maximum temperature above 30 Celsius degree)	0	14	13.5	27	own calculations based on meteorological measurements (ADMS IUNG, 2023); separate values for each year
TRI Terrain Ruggedness Index (m) (Riley et al., 1999)	0.14	0.71	1.41	20.7	own calculations based on the elevation data (Geoportal, 2023) aggregated to 20 m resolution; temporally constant
PA agricultural parcel area (ha)	0.2	3.93	9.14	43.5	averaged area from agricultural land payment system (ARMA, 2023); vector layer; temporally constant

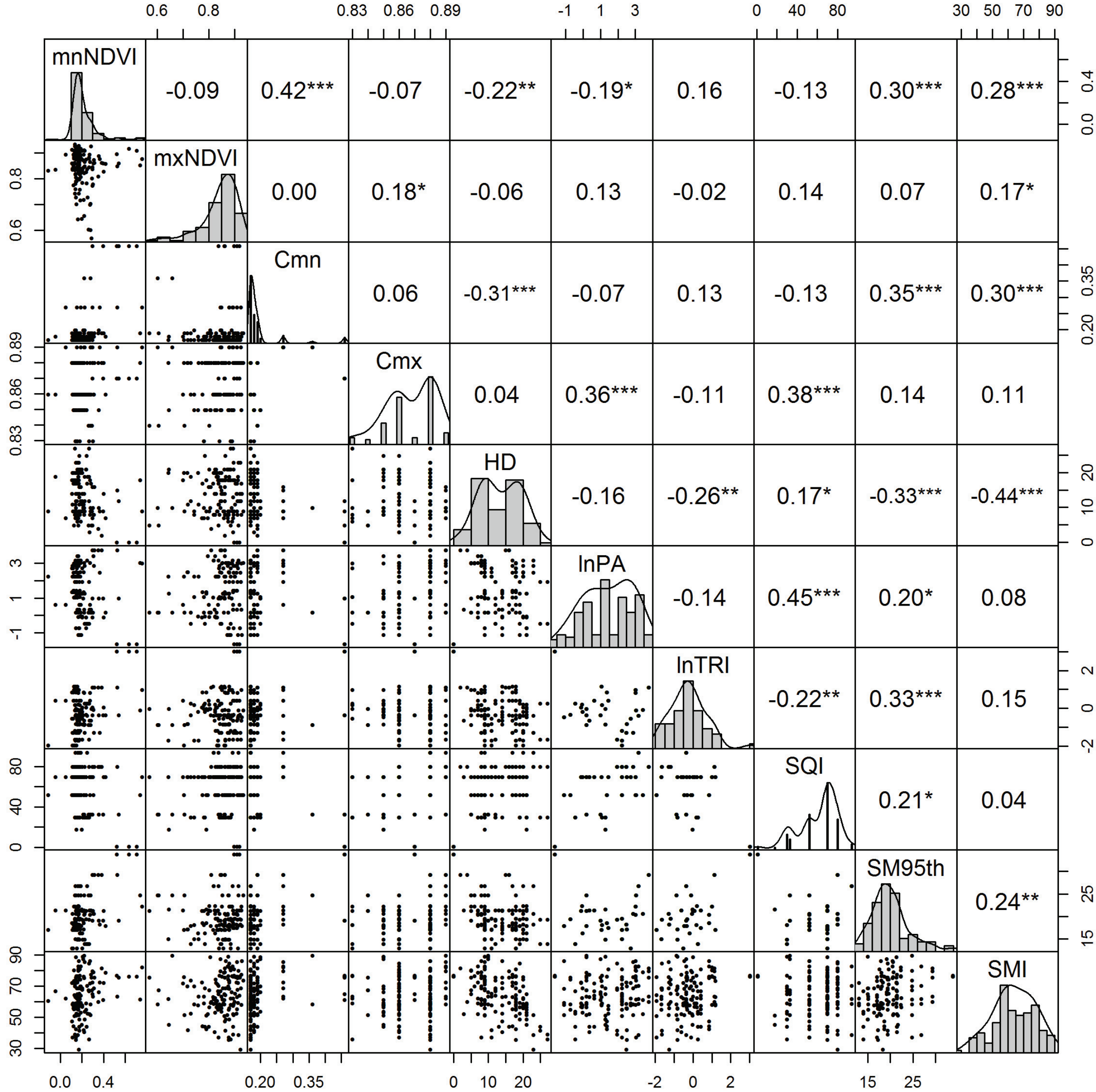

Fig. 5. Spearman correlation coefficients between variables used in the analyses

3. Results

The results obtained for both regression models (Table 2) indicated that the explanatory variables that were most often significant in linear regression or remained after selection in LASSO regression were the C variable related to cultivation, soil water capacity SM95th, and the logarithm of the terrain roughness index lnTRI. However, the positive sign of the estimated parameter at field water capacity SM95th is inconsistent with the remote sensing soil hypothesis of exposed soil, which should be darker if it has a higher water capacity and leads to a decrease in NDVI. The obtained models explain about 42% of the variability of mnNDVI and from 1 to 7% of the variability of mxNDVI. Moreover, stringent cross-validation tests show that only models for the NDVI minimum allow its prediction in locations other than those used to construct the model.

The obtained models were then tested for their sensitivity to the removal of variables to show which variables have significant information value for the models. In such a case, an increase in the $RMSE_{cv}$ error determined in the cross-validation test indicates that the variable is important for model predictions, and the lack of change indicates that the variable is redundant. In the case of mnNDVI, it was observed that only the removal of the Cmn variable caused a significant increase in model errors (Figure 6). In the case of mxNDVI, it was not observed that the removal of any variable (including Cmx) resulted in worse predictions (Figure 7). This means that this study failed to find an answer as to what are the variables that determine the level of maximum NDVI before harvest.

It was also observed that after removing Cmn from the model, the terms responsible for soil properties SQI and SM95th became significant in the regression for mnNDVI, as expected

Table 2
Parameters of regression and RF models

		mnNDVI LR	mnNDVI Lasso	mnNDVI RF	mxNDVI LR	mxNDVI Lasso	mxNDVI RF
regression coefficients	constant	–0.113	–0.062	not relevant	0.366	0.801	not relevant
	C	0.765***	0.752		0.523	0.045	
	lnPA	0	0		0.004	0	
	HD	0.001	0		0	0	
	lnTRI	0.018	0.016		0.016*	0	
	SQI	–0.001	0		0.001	0	
	SM95th	0.007**	0.006		–0.002	0	
	SMI	0.001	0		0	0	
model error measures	RMSE	0.093	0.094	0.056	0.068	0.070	0.030
	MAE	0.062	0.062	0.035	0.052	0.051	0.022
	R^2	0.426	0.421	0.793	0.076	0.007	0.816
	$RMSE_{cv}$	0.093	0.092	0.098	0.070	0.069	0.057
	MAE_{cv}	0.066	0.065	0.064	0.055	0.052	0.042
	R^2_{cv}	0.024	0.064	–0.137	–0.162	–0.104	0.207

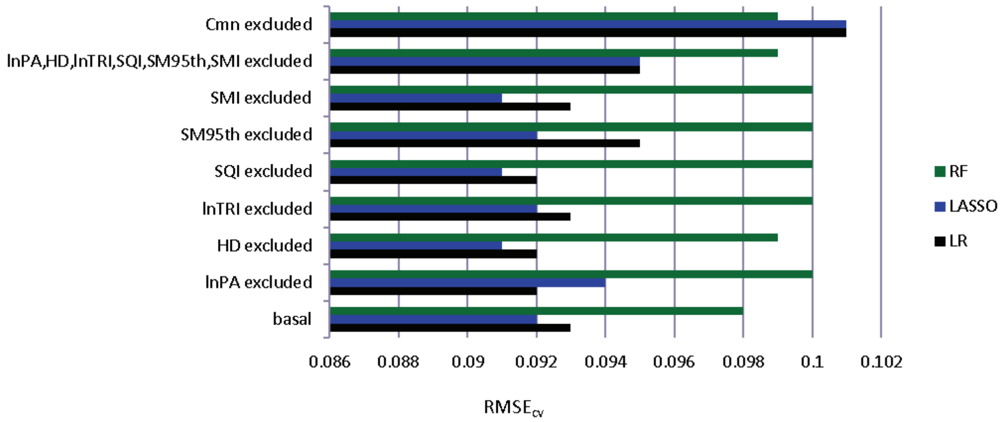

Fig. 6. Effect of variable removal on mnNDVI model prediction errors

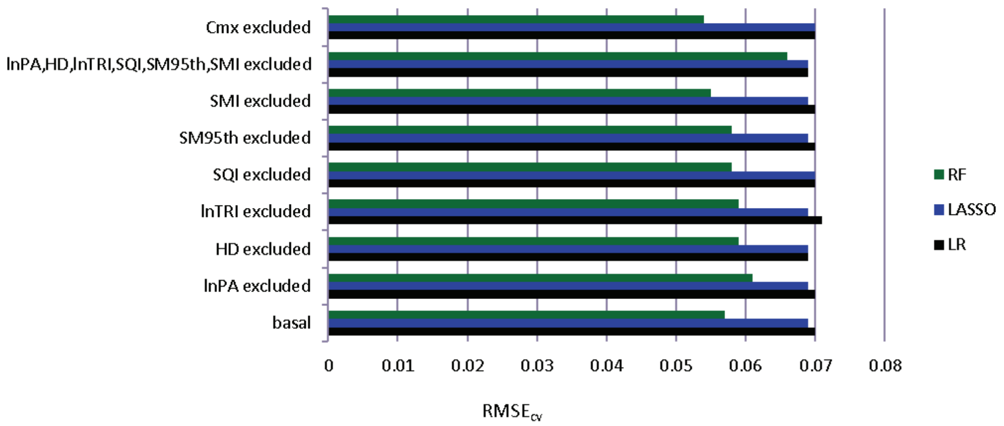

Fig. 7. Effect of variable removal on mxNDVI model prediction errors.

(Table 3). It follows that the observed relationship between mnNDVI and soil quality comes only from the relationship between the choice of crop and soil conditions. Therefore, the soil hypothesis about the influence of soil reflection recording was not confirmed.

4. Discussion

The obtained results indicated that the variability of the maximum NDVI before the harvest (mxNDVI) could not be explained, while the variability of the minimum NDVI after the

Table 3
Parameters of regression and RF models after removing the crop-related variable C.

		mnNDVI LR	mnNDVI Lasso	mnNDVI RF	mxNDVI LR	mxNDVI Lasso	mxNDVI RF
regression coefficients	constant	–0.016	0	not relevant	0.800***	0.836	not relevant
	-	–	–		–	–	
	lnPA	–0.004	–0.003		0.006	0	
	HD	0	0		0	0	
	lnTRI	0.014	0		0.016*	0	
	SQI	–0.001**	0.014		0.001	0	
	SM95th	0.012***	–0.001		–0.001	0	
	SMI	0.001	0.011		0	0	
model error measures	RMSE	0.101	0.101	0.058	0.068	0.070	0.029
	MAE	0.068	0.067	0.035	0.052	0.051	0.021
	R^2	0.333	0.332	0.777	0.066	0.010	0.833
	$RMSE_{cv}$	0.101	0.101	0.099	0.070	0.070	0.054
	MAE_{cv}	0.072	0.070	0.064	0.054	0.052	0.041
	R^2_{cv}	–0.260	–0.233	–0.087	–0.131	–0.093	0.297

harvest (mnNDVI) was explained by the variability of the crop. The choice of crop is related to soil conditions, which was also confirmed in the current study. Hence, the rejection of the soil hypothesis does not limit the use of the haNDVI = mxNDVI – mnNDVI index for remote sensing of soil conditions.

The rejection of the soil hypothesis indicates that blocking the reflection of radiation from the soil by a dense cover of crop residues plays an important role. Moreover, analysis of the data in Figure 4 indicates that only grasses, fodder plants and sugar beet have higher NDVI levels at the minimum after harvest, which is because they are harvested when their leaves are green, unlike cereals.

NDVI is rarely studied as an explained variable, but often as a variable explaining yields (e.g. Panek and Gozdowski, 2020), percentage canopy cover (Tenreiro et al., 2021) or soil properties (e.g. Jędrejek et al., 2023). To the authors' knowledge, there have also been few studies on the variability and dependence of the extremes of the NDVI index on other factors. These include the analysis of the NDVI amplitude on a global scale, where satellite images were not filtered from cloud cover and snow cover (Julien and Sobrino, 2008), and the global analysis of the use of the difference in maximum and minimum NDVI to distinguish between annual and perennial vegetation and the detection of monsoon forest and double cropping (Liu, 2017). Recognizing the relationship between mnNDVI and crops (cereals vs fodder) and rejecting the hypothesis about the importance of direct soil detection is therefore an original contribution to understanding the correlation between the amplitude of the NDVI index and the agricultural quality of soils.

In larger areas (continental scale) where climatic differences are decisive for crop phenology, the recognition of the time window for determining NDVI extremes should be improved. It can be based on the detection of, for example, decreasing trend of NDVI at harvest (Gao et al., 2020). There is also a need for a better understanding of the factors that determine the variability of maximum NDVI before harvest.

5. Conclusions

The tested mnNDVI index depends almost exclusively on the crop; thus, the agronomic hypothesis is confirmed, and the soil hypothesis about the possibility of the influence of the recorded soil reflection from among crop residues is rejected. However, it should be remembered that the structure of crops depends on the soil; therefore, mnNDVI and haNDVI have the potential to be used in remote sensing and machine learning of soil conditions. The mxNDVI did not significantly depend on any of the potential explanatory variables indicated in this study. The reason for the lack of impact of drought on NDVI values before harvest remains unexplained.

Acknowledgments

The project was funded by statutory activities IUNG-PIB (task 2.03).

References

ADMS IUNG, 2023. Agricultural Drought Monitoring System, Soil moisture monitoring. https://susza.iung.pulawy.pl/mwg/
Angström, A., 1925. The albedo of various surfaces of ground. Geografiska Annaler 7, 323–327. https://doi.org/10.2307/519495
ARMA, 2023. https://geoportal.arimr.gov.pl/
Breiman L., 2001. Random Forests. Machine Learning, 45, 5–32. https://doi.org/10.1023/A:1010933404324
Faber, A., 2007. Przegląd wskaźników rolnośrodowiskowych zalecanych do stosowania w ocenie zrównoważonego gospodarowania w rolnictwie. [In:] Harasim, A. (Eds.) Sprawdzenie przydatności wskaźników do oceny zrównoważonego gospodarowania zasobami środowiska rolniczego w wybranych gospodarstwach, gminach i województwach. „Studia i Raporty IUNG-PIB" 5, 9–24. (in Polish) https://doi.org/10.26114/sir.iung.2007.05.01
Friedman, J., Tibshirani, R., Hastie, T., 2010. Regularization Paths for Generalized Linear Models via Coordinate Descent. Journal of Statistical Software 33(1), 1–22. https://doi.org/10.18637/jss.v033.i01

Gao, F., Anderson, M.C., Hively, W.D., 2020. Detecting cover crop end-of-season using VENµS and Sentinel-2 Satellite imagery. Remote Sensing 12, 3524. https://doi.org/10.3390/rs12213524

Geoportal, 2023. https://mapy.geoportal.gov.pl/

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

Harasim, A., 2013. Metoda oceny zrównoważonego rolnictwa na poziomie gospodarstwa rolnego. Studia i Raporty IUNG-PIB 32(6), 25–75. (in Polish) https://doi.org/10.26114/sir.iung.2013.32.02

Jędrejek, A., Jadczyszyn, J., Pudełko, R., 2023. Increasing accuracy of the soil-agricultural map by Sentinel-2 images analysis—Case study of maize cultivation under drought conditions. Remote Sensing 15, 1281. https://doi.org/10.3390/rs15051281

Jones, H.G., Vaughan, R.A., 2010. Remote sensing of vegetation – Principles, techniques, and applications. Oxford University Press, New York, 353 pp., ISBN: 9780199207794.

Julien, Y., Sobrino, J.A., 2008. NDVI seasonal amplitude and its variability. International Journal of Remote Sensing 29(17–18), 4887–4888. https://doi.org/10.1080/01431160802036607

Kuhn, M., 2008. Building Predictive Models in R Using the caret Package. Journal of Statistical Software 28(5), 1–26. https://doi.org/10.18637/jss.v028.i05

Liaw, A., Wiener, M., 2002. Classification and Regression by random Forest. R News, 2(3), 18–22. https://CRAN.R-project.org/doc/Rnews/

Liu, R., 2017. Compositing the Minimum NDVI for MODIS Data. [In:] IEEE Transactions on Geoscience and Remote Sensing 55(3), 1396–1406. https://doi.org/10.1109/TGRS.2016.2623746

Liu, W., Baret, F., Gu, X.F., Tong, Q., Zheng, L., Zhang, B., 2002. Relating soil surface moisture to reflectance. Remote Sensing of Environment 81, 238–246. https://doi.org/10.1016/S0034-4257(01)00347-9

Lobell, D., Asner, G., 2002. Moisture Effects on Soil Reflectance. Soil Science Society of America Journal 66, 722–727. https://doi.org/10.2136/sssaj2002.7220

MChAS 2023, Monitoring Chemizmu Gleb Ornych Polski, GIOŚ. https://www.gios.gov.pl/chemizm_gleb/

Łopatka, A., Koza, P., 2020. Crop production intensity and haNDVI indicator – amplitude of NDVI related to harvest. Polish Journal of Agronomy 42, 24–33. https://doi.org/10.26114/pja.iung.411.2020.42.03

Panek, E., Gozdowski, D., 2020. Analysis of relationship between cereal yield and NDVI for selected regions of Central Europe based on MODIS satellite data. Remote Sensing Applications: Society and Environment 17, 100286. https://doi.org/10.1016/j.rsase.2019.100286

R Core Team, 2023. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/

Riley, S., Degloria, S., Elliot, S.D., 1999. A terrain ruggedness index that quantifies topographic heterogeneity. International Journal of Science 5, 23–27.

Rouse, J.W, Haas, R.H., Scheel, J.A., and Deering, D.W., 1974. Monitoring Vegetation Systems in the Great Plains with ERTS. Proceedings, 3rd Earth Resource Technology Satellite (ERTS) Symposium, vol. 1, p. 48–62. https://ntrs.nasa.gov/citations/19740022614

Sicre, M.C., Inglada, J., Fieuzal, R., Baup, F., Valero, S., Cros, J., Huc, M., Demarez, V., 2016. Early detection of summer crops using high spatial resolution optical image time series. Remote Sensing 8, 591. https://doi.org/10.3390/rs8070591

Schauberger, B., Archontoulis, S., Arneth, A. et al., 2017. Consistent negative response of US crops to high temperatures in observations and crop models. Nature Communications 8, 13931. https://doi.org/10.1038/ncomms13931

Tenreiro, T.R., García-Vila, M., Gómez, J.A., Jiménez-Berni, J.A., Fereres, E., 2021. Using NDVI for the assessment of canopy cover in agricultural crops within modelling research. Computers and Electronics in Agriculture 182, 106038. https://doi.org/10.1016/j.compag.2021.106038

Tibshirani, R., 1996. Regression Shrinkage and Selection via the lasso. Journal of the Royal Statistical Society. Series B (methodological), 58(1), 267–88. https://doi.org/10.1111/j.2517-6161.1996.tb02080.x

Tucker C.J., 1979. Red and photographic infrared linear combinations for monitoring vegetation. Remote Sensing and Environment 8, 127–150. https://doi.org/10.1016/0034-4257(79)90013-0

Witek, T., 1993. Waloryzacja rolniczej przestrzeni produkcyjnej Polski według gmin. IUNG Puławy, Seria (A) 56. (in Polish)

Zhao, Q., Yu, L., Li, X., Peng, D., Zhang, Y., Gong, P., 2021. Progress and trends in the application of Google Earth and Google Earth Engine. Remote Sensing 13, 3778. https://doi.org/10.3390/rs13183778

Ocena wpływu gleby na ekstrema NDVI przed i po zbiorach za pomocą uczenia maszynowego

Słowa kluczowe

Mapa glebowo-rolnicza
Teledetekcja
NDVI
Zbiory plonów
Uczenie maszynowe
Sentinel-2

Streszczenie

Zaobserwowano, że różnica maksymalnych i minimalnych wartości NDVI w okresie zbliżonym do żniw (odpowiednio mxNDVI i mnNDVI), określana jako wskaźnik haNDVI (amplituda NDVI związana ze żniwami), koreluje z rolniczą jakością gleby i udziałem zasiewów. NDVI ulega nasyceniu, gdy wartości wskaźnika powierzchni liści (LAI) znacznie przekraczają jeden, więc za przestrzenne zróżnicowanie haNDVI odpowiadają głównie wartości minimalnego NDVI po zbiorach (mnNDVI). Aby wyjaśnić zmienność wartości mnNDVI, sformułowano trzy hipotezy o: i) wpływie doboru upraw, ii) wpływie wielkości pola oraz iii) wpływie gleby. Aby określić, która z tych hipotez miała największy wpływ na zmienność mnNDVI, opracowane modele uczenia maszynowego tego wskaźnika poddano testowi usuwającemu z nich poszczególne zmienne objaśniające. Usunięcie zmiennej nie powoduje istotnego wzrostu błędu modelu, jeśli zmienna nie wnosi do modelu użytecznych informacji. Test ten wykazał, że wskaźnik mnNDVI zależy prawie wyłącznie od wskaźnika uprawy, który był medianą mnNDVI dla upraw, a nie bezpośrednio od zmiennych glebowych, takich jak rolnicza jakość gleby czy wilgotność gleby. W związku z tym odrzucono hipotezę o bezpośrednim wpływie gleby. Wyjaśnieniem obserwowanej korelacji haNDVI z jakością gleby jest praktyka rolnicza polegająca na wyborze upraw o niskim mnNDVI (zboża, rzepak) przy lepszych warunkach glebowych oraz upraw o wysokim mnNDVI (rośliny pastewne, użytki zielone) przy gorszych warunkach glebowych.