

ASSESSMENT OF SOIL CHARACTERISTICS USING A THREE-BAND AGRICULTURAL DIGITAL CAMERA

AGNIESZKA GLINKO ¹, CEZARY KAŻMIEROWSKI ², JAN PIEKARCZYK ²,
SŁAWOMIR KRÓLEWICZ ²

¹ Department of Nature, Institute of Technology and Life Sciences, National Research Institute, Falenty, Poland

² Department of Environmental Remote Sensing and Soil Science, Institute of Physical Geography and Environmental Planning, Adam Mickiewicz University, Poznań, Poland

Manuscript received: June 8, 2022

Revised version: July 18, 2022

GLINKO A., KAŻMIEROWSKI C., PIEKARCZYK J., KRÓLEWICZ S., 2022. Assessment of soil characteristics using three bands agricultural digital camera. *Quaestiones Geographicae* 41(3), Bogucki Wydawnictwo Naukowe, Poznań, pp. 127–140. 4 figs, 4 tables.

ABSTRACT: Remote sensing techniques based on soil spectral characteristics are the key to future land management; however, they still require field measurement and an agrochemical laboratory for the calibration of the soil property model. Visible and near-infrared diffuse reflectance spectroscopy has proven to be a rapid and effective method. This study aimed to assess the suitability of multispectral data acquired with the agricultural digital camera in determining soil properties. This 3.2-Mpx camera captures images in three spectral bands – green, red and near-infrared. First, the reference data were collected, which consist of 151 samples that were later examined in the laboratory to specify the granulometric composition and to quantify some chemical elements. Second, additional soil properties such as cation exchange capacity, organic carbon and soil pH were measured. Finally, the agricultural digital camera photograph was taken for every soil sample. Reflectance values in three available spectra bands were used to calculate the spectra indices. The relationships between the collected data were calculated using the independent validation regression model such as Cubist and cross-validation model like partial least square in R Studio. Additionally, different types of data normalisation [multiplicative scatter correction, standard normal variate, min–max normalisation, conversion into absorbance] were used. The results proved that the agricultural digital camera is suitable for soil property assessment of sand and silt, pH, K, Cu, Pb, Mn, F, cation exchange capacity and organic carbon content. Coefficient of determination varied from 0.563 (for K) to 0.986 (for soil organic carbon). Higher values were obtained with the Cubist regression model than with partial least squares.

KEY WORDS: remote sensing, soil, spectroscopy, cubist, R Studio

Corresponding author: Agnieszka Glinko, agaglinko@interia.pl

Introduction

Historically, our perception of soil and its characteristics has required comprehensive laboratory analysis. Conventional measurement techniques aiming to assess the relationship between the physical and chemical properties of

soil components often overlook their complex interaction. It is important to develop and improve the existing methods of measuring soil parameters to describe the entire soil system as accurately as possible (Viscarra Rossel et al. 2006). Spectroscopy makes it possible to deviate from the traditional techniques of laboratory

measurement of soil parameters by determining the relationship between electromagnetic radiation and an object in its natural environment. Spectroscopic measurements have shown enormous potential for calibration, prediction and data modelling in soil science (Milton 1987). Historically, research has been conducted to determine the possible method of testing soils in various ranges of electromagnetic radiation. One of the best-studied is where diffuse reflectance spectroscopy (DRS) has been used, *inter alia*, in the mid-visible and near-infrared (MVNIR) ranges. This method enables faster, more economically and non-chemically extracted soil measurement procedures (Raupach 1991).

Visible and near-infrared (VNIR) spectroscopy in soil research enables the simultaneous measurement of several parameters without prior laborious precise preparation of samples. In laboratory conditions, hyperspectral spectrophotometers with very high spectral resolution in the VNIR range are used to measure soil samples. However, soil properties are also estimated with a lower spectral resolution using satellite and airborne multispectral sensors. Imaging data from these sensors are recorded in only a few bands of the VNIR range and can be used to estimate the content of soil organic carbon (SOC) (Croft et al. 2012, de Paul Obade, Lal 2013) and clay (Nanni, Dematte 2006, Demattè, Fiorio 2009). However, better results can be obtained by combining satellite data with hyperspectral measurements (Peng et al. 2015). Other studies show that attempts are being made to use airborne multispectral data to improve the quality of soil maps (Wetterlind et al. 2008). The Cubist model is often used to estimate SOC. In this case, it is also advisable to use spectral indices as variables in addition to raw reflectance (Peng et al. 2015).

Unmanned aerial vehicle (UAV) - mounted multispectral VNIR sensors are very often used to observe agricultural crops in precision farming applications. However, there are also multispectral cameras that can be used for ground or laboratory imaging. An example of a multispectral data acquisition device is the agricultural digital camera (ADC). This sensor is specifically designed to capture three spectral channels that are most sensitive to changes in plant biomass, *i.e.* green, red and near-infrared. This fact makes

ADC suitable for estimating the size of biomass and yield (Swain et al. 2010), assessment of nitrogen content at various stages of plant development (Saberioon et al. 2012), calculation of vegetation indices (Liu et al. 2012) and even for discrimination of crop cultivars (Avola et al. 2019). ADC can also be used in field research as a part of UAV (Candiago et al. 2015, Vega et al. 2015, Matese et al. 2017).

Another approach is given by multispectral satellite sensors. Since 1972, Landsat satellites gather images that can be useful in environmental studies. For example, sensor thematic mapper (TM) on-board Landsat 5 was used to detect bare soil (Dematte et al. 2009). In 2015, the European Space Agency (ESA) began to deliver free of cost, good spatial resolution (10 m) Earth images. Sensors on-board optical Sentinel-2 satellites are equipped with 12 spectral bands, which can be useful for clay content mapping (Gasmi et al. 2022). There are other examples of clay content mapping using other multispectral sensors such as that on-board the ASTER satellite (Gasmi et al. 2019). These studies have proved that multispectral satellite sensors should be considered in soil research more often.

The usefulness of the acquired image data largely depends on the way it is processed and analysed. Many statistical methods are used to obtain reliable soil information from multispectral images, such as multiple linear regression (MLR) analysis, principal component regression (PCR) and partial least squares (PLS) regression. Application of the latter method to hyperspectral data allows to determine several soil parameters with high values of correlation coefficient and low errors, including grain size composition, pH, cation exchange capacity (CEC) or some chemical elements (Mammadov et al. 2020, Vestergaard et al. 2021). Recently, machine learning algorithms based on random forests and Cubist development models have been used to study the relationship between spectral data and soil characteristics. The Cubist model is often used to estimate SOC; in such cases, it is also advisable to use spectral indices as variables in addition to raw reflectance (Peng et al. 2015).

A proposed new approach for estimating soil parameters is to use for this purpose multispectral images obtained from ADC. The possibility of determining the condition of the soil substrate

based on such data is not well researched or described. This study aimed to determine whether it would be possible to estimate soil parameters using a sensor that guarantees measurements only in three spectral channels (green, red and near-infrared). More precisely, which soil parameters, with what method of data analysis and with what accuracy, can be estimated based on images taken with a multispectral camera in laboratory conditions.

Materials and methods

Study area

The research was conducted within two arable fields located in Pokrzywno (Wielkopolskie Voivodeship, Poznań Poviát). This region has a temperate transitional climate characterised by a small number of frosty days and low rainfall. The average annual temperature is 8.5°C and the annual rainfall is about 500–550 mm (WIOS 2013). It is an area with unfavourable water balance, exposed to periodic droughts. Soils classified as Luvisols and Phaeozems, according to the IUSS Working Group WRB (2015), dominate this study area (Fig. 1).

Sampling and laboratory analyses

A total of 151 samples were collected from both research fields. They were tested in the soil science laboratory of the Adam Mickiewicz University in Poznań. All samples were prepared for testing by drying, grinding in a ceramic mortar and sieving through a 2-mm mesh sieve. The soil texture was determined by the hydrometer method according to the standard PN-R-04032 (Polish Committee for Standardisation PKN 1998). SOC was determined using oxidation by $K_2Cr_2O_7$ with H_2SO_4 for 30 min on the digestion block at 150°C and titration of oxidant residues by $FeSO_4$ (Nelson, Sommers 1996). Total nitrogen was modified using the Kjeldahl method (International Standard ISO 12261 1995). The soil pH was measured in 1:1 soil solution ratio in water and 1M KCL (PN-ISO 10390 1997). The form of nutrients available to plants (K, Mg, Ca, Zn, Cu, Pb, Cd, Mn and Fe) was determined by the modified Mehlich 3 method (Mehlich 1984). CEC was determined by successive barium and magnesium chloride solution extraction and flame atomic absorption spectroscopy (International Standard ISO 11260 1994). Calcium carbonate content was determined twice using Scheibler volumetric (International Standard ISO 10693 2002) and

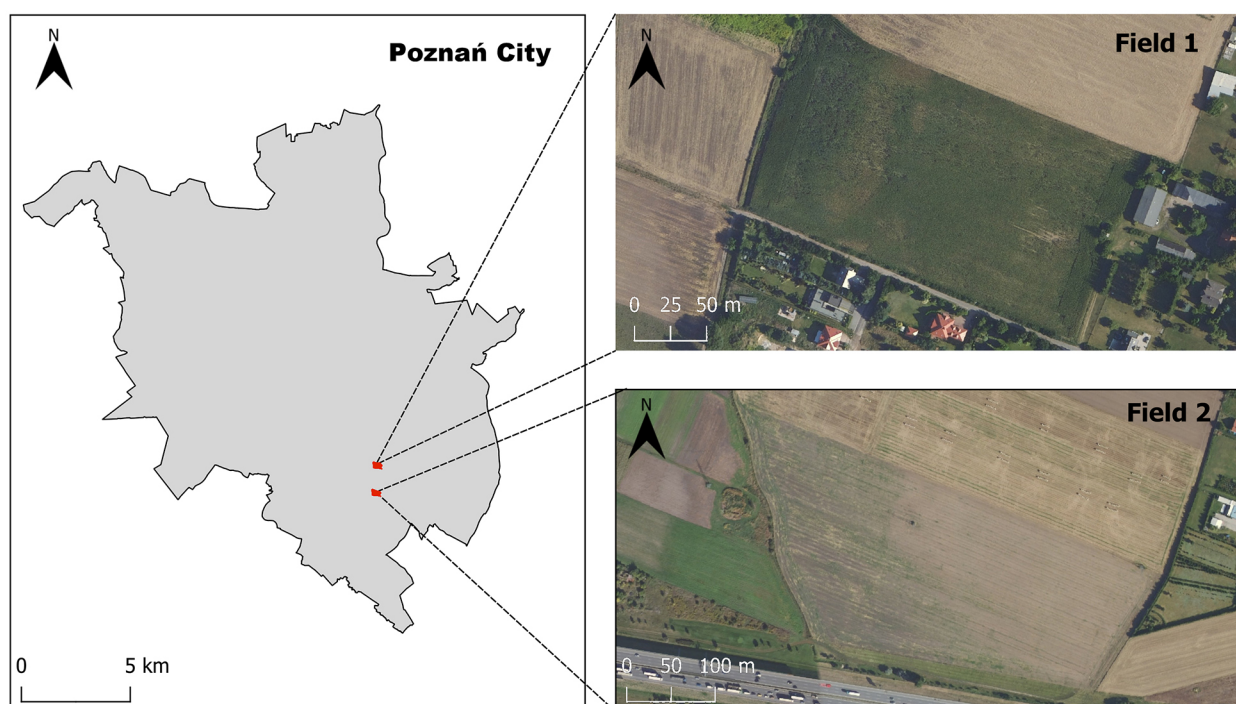


Fig. 1. Map of the study area.

titration methods (FAO 2021). As a result of the analyses, the soil particle size composition was determined, and the amount of organic carbon, nitrogen, the ratio of carbon content to nitrogen content, soil reaction, percentage of calcium carbonate, total CEC and the content of the elements potassium, magnesium, calcium, zinc, copper, lead, cadmium, manganese, iron and phosphorus was also determined.

Multispectral data

Multispectral data of soil samples were taken in the laboratory environment using the ADC by Tetracam. The specific design with optical Bayer filter mask in the complementary metal-oxide-semiconductor (CMOS) sensor allows to obtain three images at a resolution of 2048×1536 px (3.2 Mpx) (Swain et al. 2010). The images correspond to three Landsat Thematic Mapper 2, 3 and 4 spectral bands: green (520–600 nm), red (630–690 nm) and near-infrared (760–920 nm) (Lan et al. 2010) and the estimated ground pixel resolution is $0.000707 \text{ m px}^{-1}$ (Swain et al. 2010).

For the purposes of this study, photographs of 151 soil samples were taken under laboratory conditions. The ADC was placed on a tripod at a height of 70 cm above the test object and at an angle of 90° . In addition, a 400-W halogen lamp was used to illuminate the surface of the soil, which was set at a distance of 80 cm and at an angle of 45° . Then the images were processed into tiff

format in the Pixel Wrench 2 program dedicated by the manufacturer. The next step was to transform the original digital numbers to reflectance using the TNT Mips software.

Spectral indices

The evolution of remote sensing techniques caused the development of methods of evaluating remote measurements, processing and extracting as much information as possible from the collected data. Attempts to interpret the reflectance data from different available ranges of electromagnetic radiation have led to the development of a large number of indicators and their derivatives. A large group of spectral indices relate to the vegetation and soil substrate, which are calculated as the ratio of reflectance in two or more spectral channels of the selected device, sometimes with additional parameters (Bannari et al. 1995).

In this study, vegetation indices such as normalized difference vegetation index (NDVI), radar vegetation index (RVI, Bannari et al. 1995, Martínez M. 2017) and infrared percentage vegetation index (IPVI, Gunathilaka 2021) were used. All of them presented ratios between reflected radiation in red and infrared ranges. In addition, the indexes with green spectra such as IPVI (GNDVI, Candiago et al. 2015), green-red vegetation index (GRVI, Motohka et al. 2010) and modified GNDVI normal (Crippen 1990) were

Table 1. Summary of spectral indices used in the study.

Spectral index	Abbreviation	Formula
Normalized Difference Vegetation Index	NDVI	$\frac{NIR - RED}{NIR + RED}$
Green Normalized Difference Vegetation Index	GNDVI	$\frac{NIR - GREEN}{NIR + GREEN}$
Green Normalized Difference Vegetation Index normal	GNDVInormal	$\frac{NIR - GREEN}{NIR + GREEN} + 1$
Infrared Percentage Vegetation Index	IPVI	$\frac{NIR}{NIR + RED}$
Soil-adjusted Vegetation Index	SAVI*	$\frac{NIR - RED}{NIR + RED + L} \times (1 + L)$
Radar Vegetation Index	RVI	$\frac{NIR}{RED}$
Green-Red Vegetation Index	GRVI	$\frac{NIR}{GREEN}$

*SAVI25 - L=0.25, SAVI50 - L=0.50, SAVI75 - L=0.75.

added to the dataset. Finally, three variants of the soil-adjusted vegetation index (SAVI, Bannari et al. 1995) were calculated, considering the soil parameter in three values of 0.25, 0.5 and 0.75. Summarises the spectral indices used in the spectral data processing (Table 1).

Data normalisation

Raw spectra are subject to fluctuations and noise disturbance. For this reason, methods of standardising spectra data are often used. They consist in reducing the undesirable effects in the set of spectral measurements (Gholizadeh et al. 2015). In this study, we used methods of multiplicative scatter correction (MSC), standard normal variate (SNV), conversion of reflectance data into absorbance and scaling with minimum and maximum values. One of the most popular methods of data standardisation is MSC (Rinnan et al. 2009). This method relies on adjusting to each spectral measurement an ideal reference spectrum estimated based on additive and multiplicative correction factors (Rinnan et al. 2009). Another frequently used method is the SNV, which consists of common centring and scaling by subtracting the mean values and normalising with the standard deviation for each reflection spectrum (Vestergaard et al. 2021). MSC and SNV were introduced by the Prospectr 0.2.4 package implemented in the R-4.1.3 software for Windows. When working with various types of data, it can be noticed that occasionally, some measured results may significantly differ from others and thus disturb the work of the computational model (Gholizadeh et al. 2015). To eliminate this effect, the min-max scaling can be used (available in the R software in the Caret package). This type of data normalisation is based on scaling all data so that they fall in the range from zero to one. This reduces the value of the standard deviation and also the effect of outliers in the dataset.

Additionally, the spectral data were converted to the form of absorbance according to the following formula (Wenjun et al. 2014):

$$ABS = -\log R,$$

where R is the reflectance in a given spectral channel.

Regression modelling

Regression models are used to establish the relationship between variables y (dependent, explained variable) and x (independent, explanatory variable). Such analysis makes it possible to explain how the value of the explained variable developed under the influence of the explanatory variable. In the case of a greater number of variables, it is useful to use statistical programs that calculate relationships between the variables by selecting the best variables to determine the relationship within the given parameter. Regression models are often used to estimate the content of soil components. Among them are Cubist (Peng et al. 2015) and PLS (Vestergaard et al. 2021). Cubist, which is modelled on the M5 Quinlana model (Quinlan Bassier 1992), is a tool that allows to create a decision tree based on a given number of rules. Each rule represents a linear regression model to which the given variables are fitted. If the value of the variable matches the rule, its predicted value is calculated (Minasny, McBratney 2008). The final regression model is simplified to reduce the absolute error value (Yi Peng et al. 2015). Partial least squares (PLS), which was created by H.Wold in 1966, is a linear regression method recommended in the case of a large number of explanatory variables and when there is a high probability that they can be correlated with each other. The reduced set of x variables is used to create the regression model (Wold et al. 2001).

Variable importance

Variable importance in the projection (VIP) is useful for determining which predictor variables are best explained by explanatory variables. VIP determines the variables and the extent to which they contribute to the construction of a given regression model (Chong, Jun 2005, Xu et al. 2021).

VIP values can be obtained through dedicated software.

Accuracy assessment

There are many ways to determine how well an outcome estimate is guaranteed by a given regression model. One of them is to calculate the coefficient of determination R^2 . This measure

shows how well the variance of the predicted spectral values coincides with the values measured in the laboratory. R^2 is calculated according to the following formula (Ng et al. 2022):

$$R^2 = 1 - \frac{\sum_i^n (x_i - y_i)^2}{\sum_i^n (x_i - \bar{x}_i)^2},$$

where x_i i y_i are the values observed and predicted under i , and \bar{x}_i is the mean value of the observed (Ng et al. 2022).

Another factor is the root mean squared error (RMSE), which informs about the difference between the values estimated in the model. RMSE takes values equal to or greater than zero, with zero being a statistically perfect match of estimated values to those observed (Peng et al. 2015):

$$RMSE = \sqrt{\frac{\sum_i^n (x_i - y_i)^2}{n}},$$

where x_i i y_i are the values observed and predicted under i , and n is the number of samples (Yi Peng et al. 2015). Rel-RMSE – the derivative of RMSE is the relative mean square error calculated as the ratio of the RMSE value to the mean value of a given variable (Yiping Peng et al. 2019). Regression point displacement (RPD) is another value that is used to evaluate the quality of the model. It is the ratio between the standard deviation of the variable and the RMSE (Vestergaard et al. 2021). At present, the RPIQ measure is considered more frequently when assessing the model estimation. It considers only the interquartile range of the results. It is calculated as the quotient of the values measured in the interquartile space and the RMSE (Vestergaard et al. 2021).

Saeys et al. (2005) proposed to establish the criteria for the classification of the model according to the following values of R^2 and RPD: if R^2 is <0.5 or RPD is <1.5 , this indicates a poor model estimation and means that the calculated values cannot be used. If R^2 is in the range of 0.5–0.65 and RPD is between 1.5 and 2, then it is possible to distinguish between high and low values. R^2 from 0.66 to 0.81 or RPD from 2 to 2.5 shows that the model enables approximate quantitative predictions. When the value of the coefficient of determination is in the range of 0.82–0.9 or RPD is between 2.5 and ≥ 3 , then the model can be

assumed to be good. Finally, if R^2 is >0.91 combined with value RPD >3 , the model is perfect. As a rule, a good prediction model should have the highest R^2 i RPD (or RPIQ) values and as little RMSE as possible (or rel-RMSE) (Wenjun et al. 2014).

Results and discussion

Laboratory analyses

The results of laboratory analyses are presented in Table 2. The summary includes statistics on mean values, median, maximum and minimum values, and standard deviation. As shown in Table 2, mean values for all data are in the range of 0.12–1504.29 mg · kg⁻¹. Calcium has the most varied values. The standard deviation for this element is 2014.76 mg · kg⁻¹. Other variables with high standard deviation are phosphorus, iron, magnesium, potassium and manganese. All other parameters have second derivative (SD) values <10.00 ; cadmium has the lowest standard deviation value of 0.12 mg · kg⁻¹.

Soil spectra

Table 3 shows a summary of reflectance data obtained by ADC and the results of all used data normalisation methods. The mean reflectance values for each spectral band are 0.47 for green, 0.41 for red and 0.81 for near-infrared. After MSC data normalisation, the mean red and green band values were slightly changed, while the NIR value was the same in both. The SNV method changed spectral data completely with mean red and green values changed to negative. Application of min-max normalisation effected in NIR spectra became smaller than the red band value. The same was observed for the absorbance values. The standard deviation varied from 0.32 in SNV green band to 0.02 in the NIR MSC band. The MSC method had lower standard deviation values for each band.

Figure 2 presents the relationship between analysed soil characteristics and ADC spectral data in addition to the calculated spectral indices. Correlation values differ from -1.0 (marked as blue on the graph) to 1.0 (red). Most of the soil parameters have a strong negative correlation

Table 2. Summary of soil laboratory analyses.

Parameter	Unit	Min	Mean	Median	Max	SD
Sand	%	61.00	78.10	79.00	85.00	4.12
Silt	%	12.00	16.87	16.00	33.00	3.63
Clay	%	1.00	4.97	5.00	8.00	1.69
SOC	%	0.57	1.56	1.00	5.79	1.32
N	%	0.01	0.14	0.08	0.82	0.15
C/N	-	6.40	14.34	11.80	63.00	9.10
pH_{H2O}	-	4.00	5.67	5.15	7.67	1.13
pH_{KCL}	-	3.48	5.18	4.65	7.25	1.24
CaCO3 vol	%	0.00	0.87	0.00	12.00	2.03
CaCO3 titr	%	0.00	1.15	0.00	13.10	2.44
CEC	cmol kg⁻¹	5.46	9.98	7.99	27.28	5.16
K	mg kg⁻¹	29.30	147.88	147.60	409.50	41.51
Mg	mg kg ⁻¹	16.30	70.59	47.70	314.80	61.70
Ca	mg kg ⁻¹	32.60	1504.29	322.70	7424.60	2014.76
Zn	mg kg ⁻¹	4.40	9.06	8.50	27.40	3.70
Cu	mg kg⁻¹	1.00	1.75	1.60	3.50	0.56
Pb	mg kg⁻¹	0.50	5.36	5.00	13.90	1.90
Cd	mg kg ⁻¹	0.00	0.12	0.09	0.55	0.12
Mn	mg kg⁻¹	19.40	73.22	76.60	101.30	17.93
F	mg kg⁻¹	83.10	233.97	248.60	385.10	60.25
P	mg kg ⁻¹	23.30	176.88	153.60	275.60	63.21

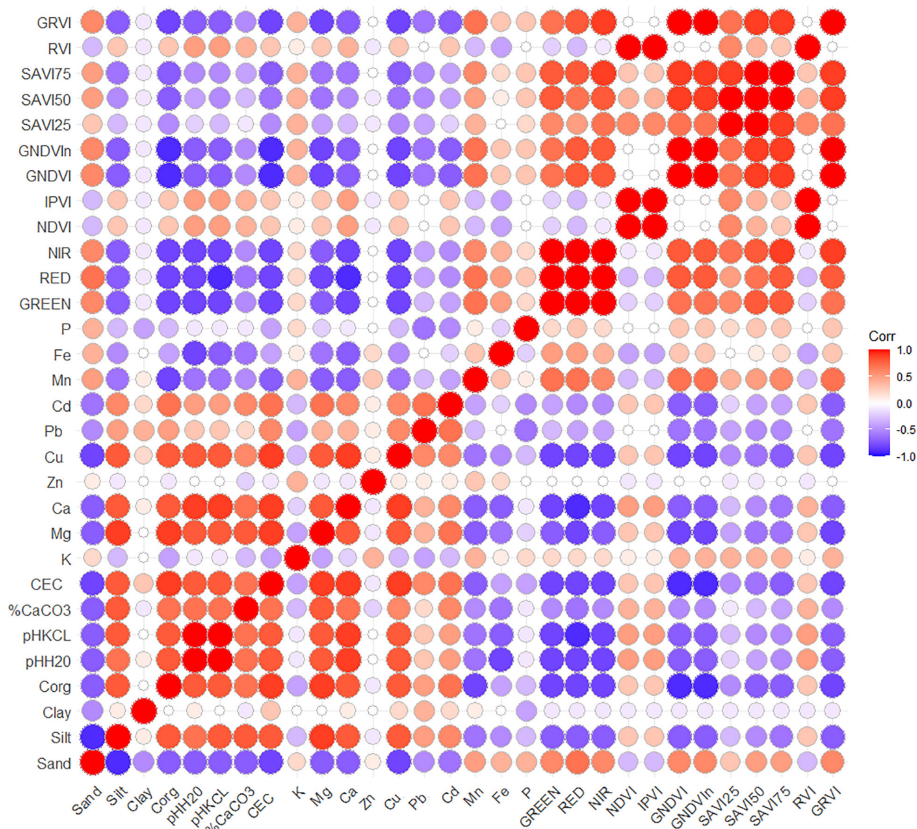


Fig. 2. Correlation between soil parameters and spectral indices.

Table 3. Summary of soil spectra.

	RAW		MSC		SNV		max-min NORM		ABS	
	MEAN	SD	MEAN	SD	MEAN	SD	MEAN	SD	MEAN	SD
GREEN	0.47	0.12	0.50	0.11	-0.34	0.32	0.44	0.19	0.35	0.13
RED	0.41	0.15	0.38	0.09	-0.74	0.17	0.52	0.23	0.43	0.21
NIR	0.81	0.29	0.81	0.02	1.08	0.17	0.33	0.15	0.13	0.20

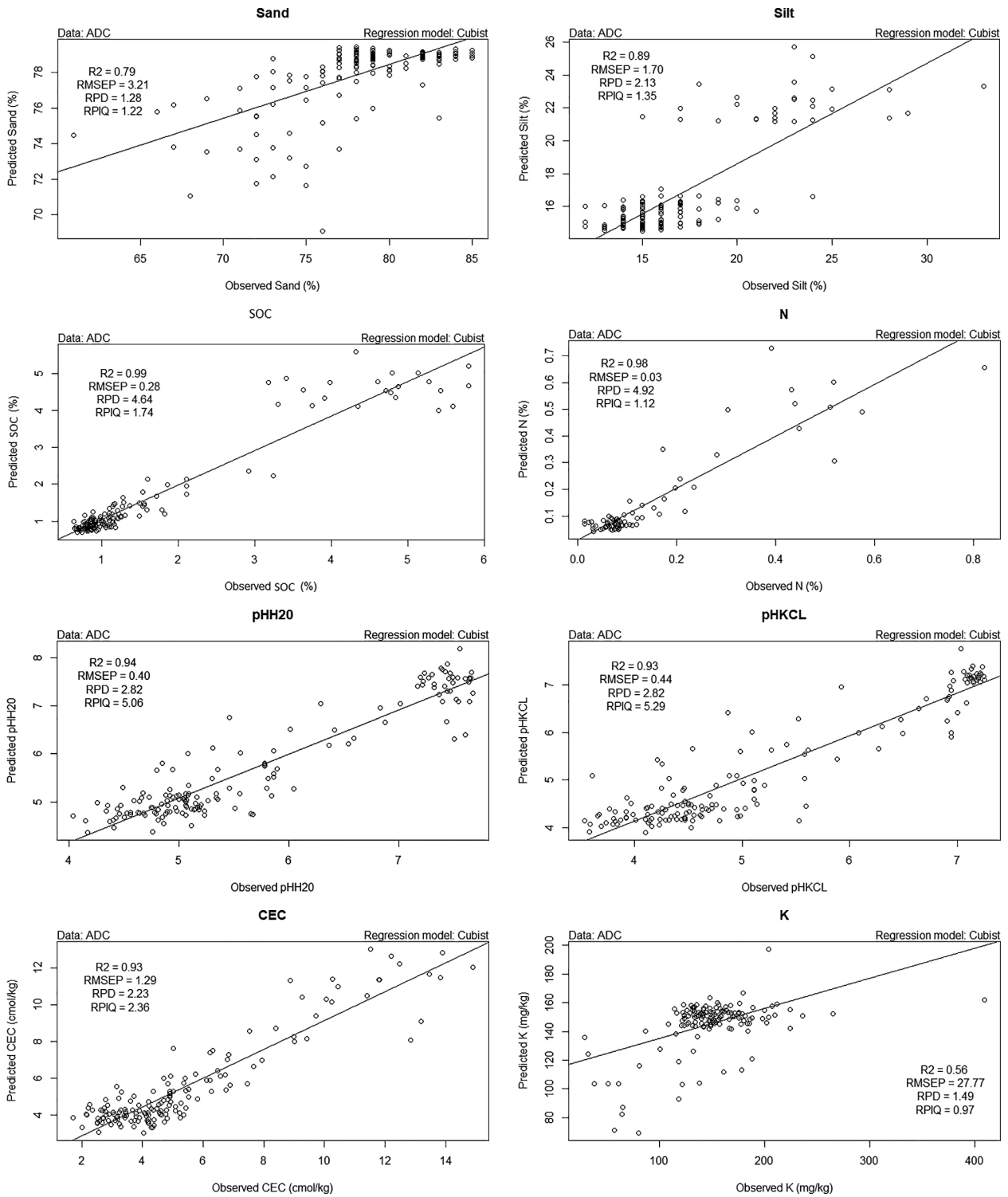


Fig. 3. Summary of graphs for soil parameters that achieved the desired values of R2 and rel-RMSE.

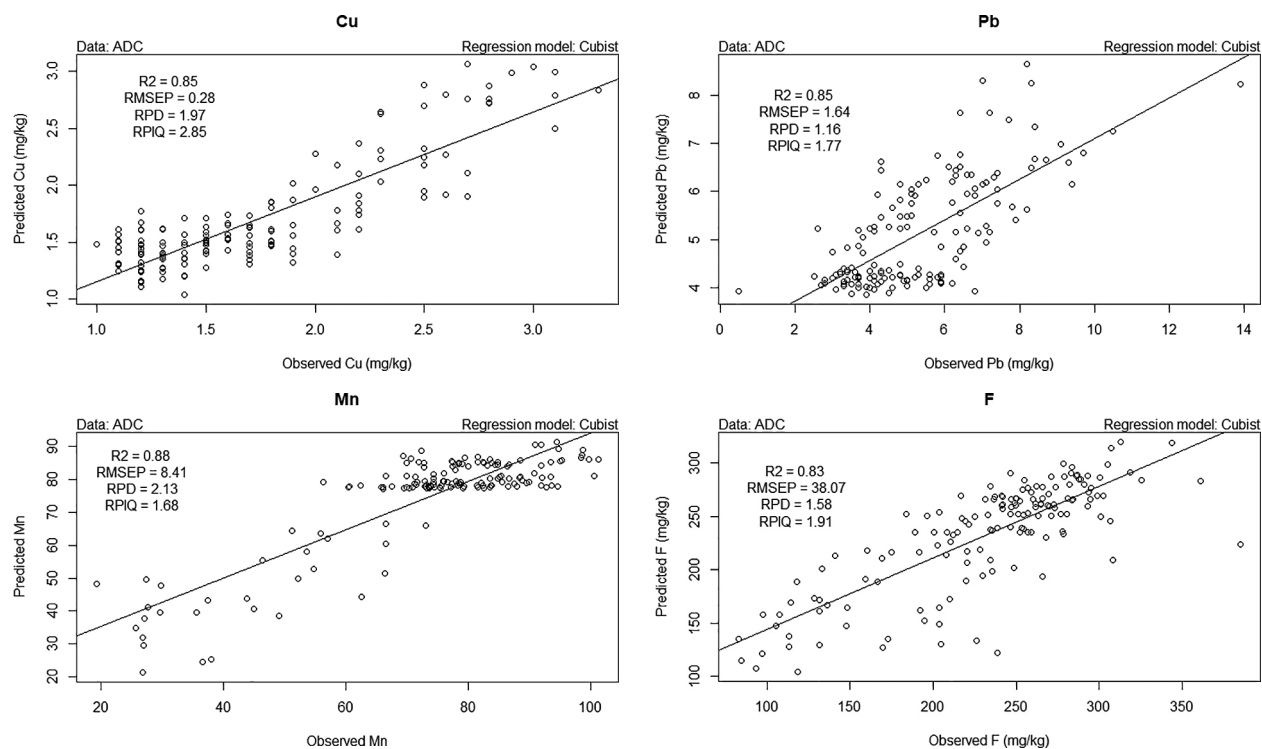


Fig. 3. cont.

with spectral data. Only the percentage of sand and some chemical elements, such as Mn, Fe and P, have positive correlation values. Almost every soil parameter is correlated with some of the spectral bands or indices, with the exception of clay and zinc.

Comparison of model prediction

Both regression models were calculated for 21 variables describing soil parameters and 12 variables corresponding to the average reflectance in three bands of the ADC device and the spectral

Table 4. Measures of goodness-of-fit for soil characteristic estimations.

Parameter	Unit	MEAN	SD	Model	pre-processing	R2	RMSE	Rel RMSE	RPD	RPIQ
Sand	%	78.10	4.12	Cubist	MSC	0.793	3 211	0.04	1.28	1.22
Silt	%	16.87	3.63	Cubist	MSC	0.886	1 704	0.10	2.13	1.35
Clay	%	4.97	1.69	PLS	ABS	0.043	1 634	0.33	1.03	1.22
SOC	%	1.56	1.32	Cubist	SNV	0.986	0.284	0.18	4.64	1.74
N	%	0.14	0.15	Cubist	RAW	0.980	0.031	0.21	4.92	1.12
C/N	-	14.34	9.01	PLS	SNV	0.018	9 100	0.63	1	0.54
pHH₂O	-	5.67	1.13	Cubist	ABS	0.937	0.403	0.07	2.82	5.06
pHKCL	-	5.18	1.24	Cubist	ABS	0.934	0.438	0.08	2.82	5.28
CaCO ₃ vol	%	0.87	2.03	Cubist	min-max Norm	0.780	0.673	0.77	3.01	0.48
CaCO ₃ titr	%	1.15	2.44	Cubist	min-max Norm	0.871	0.694	0.61	3.52	0.84
CEC	cmol kg⁻¹	9.98	5.16	Cubist	ABS	0.928	1 278	0.24	2.23	2.36
K	mg kg⁻¹	147.88	41.51	Cubist	RAW	0.563	27 770	0.19	1.49	0.97
Mg	mg kg ⁻¹	70.59	61.70	Cubist	SNV	0.951	23 463	0.33	2.63	1.78
Ca	mg kg ⁻¹	1504.29	2014.76	Cubist	min-max Norm	0.924	735 515	0.49	2.74	3.86
Zn	mg kg ⁻¹	9.06	3.07	PLS	SNV	0.068	3 548	0.39	1.04	1.15
Cu	mg kg⁻¹	1.75	0.56	Cubist	ABS	0.853	0.284	0.16	1.97	2.85
Pb	mg kg⁻¹	5.36	1.09	Cubist	ABS	0.848	1 638	0.31	1.16	1.77
Cd	mg kg ⁻¹	0.12	0.12	Cubist	min-max Norm	0.836	0.052	0.42	2.26	3.13
Mn	mg kg⁻¹	73.22	17.93	Cubist	min-max Norm	0.867	8 419	0.11	2.13	1.68
F	mg kg⁻¹	233.97	60.25	Cubist	RAW	0.832	38 073	0.16	1.58	1.91
P	mg kg ⁻¹	176.88	63.21	Cubist	ABS	0.463	67 680	0.38	0.93	2.12

indices calculated on their basis. Additionally, the model was calculated each time for each variant of standardised spectral variables. The data for the Cubist model were divided randomly

into a training set of 80% of all data, and a test set which received the remaining 20% of the data. For the PLS model, cross validation type was used which divides the data into segments. The

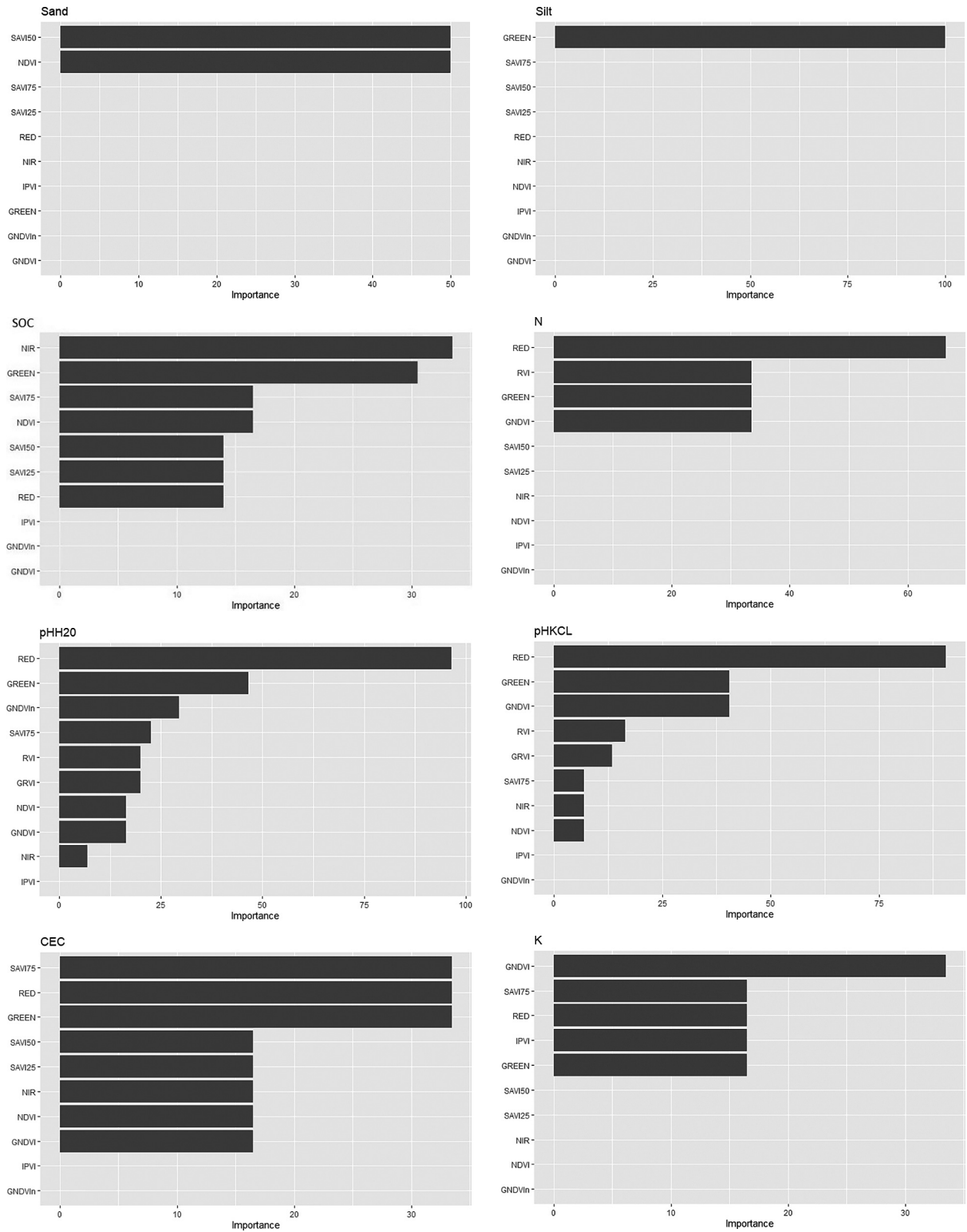


Fig. 4. Summary of graphs showing the most important variables used to estimate soil properties.

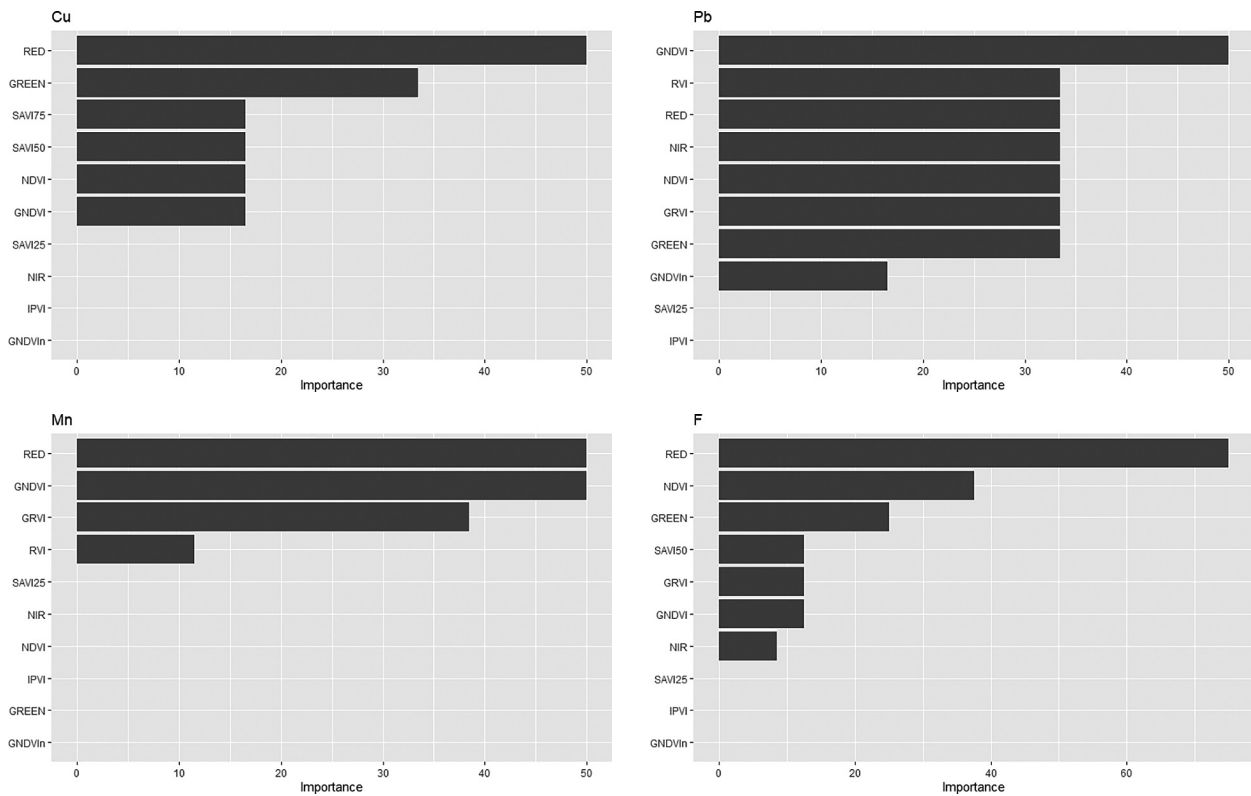


Fig. 4. cont.

number of segments was set to 10. Regression models, Cubist and PLS, were downloaded in the R software by dedicated packages Cubist, PLS. For the Cubist model, the chosen parameters were the number of committees set to 1 and the number of rules set to 3. The obtained values of the predicted soil parameters were compared with those obtained by laboratory measurements based on the values of the correlation coefficient, root mean square of errors, relative root mean square of errors, RPD and the ratio of yield to interquartile distance (RPIQ) (Table 4).

Yi Peng et al. (2015) used the Cubist model on 328 soil samples to improve SOC modelling at the regional scale. The reference data were a combination of two satellite images and laboratory Vis-NIR measurements. The obtained results were $R^2 = 0.69$, RMSE = 2.8, RPD = 1.6 and RPIQ = 0.8. Ng et al. (2022) conducted research that aimed to estimate the available nutrition of many soils using memory-based learning (MBL) algorithm and Cubist regression model. The validation statistics for the prediction of Mehlich III extractable elements using mid-infrared (MIR) spectroscopy are for Ca at $R^2 = 0.91$ and RPIQ = 2.11; for Mg $R^2 = 0.82$ and RPIQ = 2.39; for K $R^2 = 0.56$ and RPIQ

= 2.59; for Mn $R^2 = 0.65$ and RPIQ = 2.07; for P $R^2 = 0.5$ and RPIQ = 1.47; and for Zn $R^2 = 0.59$ and RPIQ = 1.59. The result of all elements, except zinc, is similar to or lower than the values obtained in our study. Other studies focused on different methods for soil features estimation. For example, Gholizadeh et al. (2015) focused on using support vector machine regression to establish a relationship between reflectance spectra in the visible near-infrared region and concentrations of Mn, Cu, Cd, Zn and Pb in soil. They used the first and second derivatives (FD and SD), SNV, MSC and continuum removal (CR) to normalise the data. The accuracy was defined by R^2 and RMSE. FD turned out to be the best normalisation method giving the highest R^2 . For Cu it was 0.78, for Mn 0.6, for Cd 0.8, for Pb 0.68 and for Zn 0.77.

The next step was to answer the question of which soil parameters can be estimated based on multispectral data obtained with ADC and with what accuracy? For that purpose, the threshold values were established for $R^2 \geq 0.5$ and rel-RMSE ≤ 0.31 . These criteria were met by 12 results (marked in bold in Table 4), which were the contents of sand, silt, SOC, nitrogen, potassium, copper, zinc, lead, manganese and iron, soil pH

($\text{pH}_{\text{H}_2\text{O}}$, pH_{KCL}) and CEC. All were obtained using Cubist regression. In Table 4, apart from the regression model, it was also specified which data normalisation method for each parameter guaranteed the highest possible R^2 value and the lowest possible rel-RMSE value. Normalisation by converting data to absorbance (ABS) most often gave the best results. The MSC method was selected only twice.

According to the model evaluation criteria (Saeys et al. 2005), models for sand, clay, C/N, Pb, Zn and P were considered as not suitable for prediction. The distinction between high and low values was guaranteed by models for K and F. Models for the percentage content of silt, CEC, Cu, Mn, Cd and the first variant of calculating the calcium carbonate content allow for approximate quantitative predictions. According to the given criteria, we can consider $\text{pH}_{\text{H}_2\text{O}}$, pH_{KCL} , Mg, Ca and the second method of determining the percentage of calcium carbonate as a good model. Finally, SOC and N were considered as perfect models.

The results, shown in the form of graphs (Fig. 3), present the ratio of the obtained values to the predicted values. The x -axis shows the observed values, i.e. those measured in the laboratory, and the y -axis values are estimated by the regression model.

Variable importance

For the 12 best-estimated parameters, additional graphs (Fig. 4) were created to illustrate which variables were considered as the most important for building the regression model for each of them. The y -axis contains a list of variables in the order from the one that had the greatest importance for building a given model. The x -axis shows the range of values from 0% to 100%, describing the weight of the variable. Cubist regression model is designed to select only certain parameters for prediction. For this reason, having access to information about VIP can help interpret and understand the given results (Fig. 4).

Conclusions

Based on the conducted research, it can be concluded that multispectral data are sufficient to determine the condition of the soil substrate.

Although only the reflection values in the green, red and near-infrared bands were used in the study, it is possible to estimate 12 out of the 21 described soil parameters with the use of appropriate data normalisation and regression model.

Although ADC was not dedicated to soil research, it can partially replace the classic spectroscopy.

Based on the VIP charts, it can be concluded that the use of spectral indices as additional explanatory variables is the correct assumption. Indicators played a large role in creating regression models for many soil parameters. Spectral indices such as GNDVI and NDVI were the most frequently used. The least frequently used indices were IPVI, SAVI25 and GNDVI normal.

The ease of use and portability of ADC makes it ideal for data acquisition in the field. For this reason, it is worth considering conducting similar studies based on images taken directly in the field. It is important to determine in what lighting conditions, at what angle of camera setting and for what types of soil it would be possible to best estimate the soil parameters.

Acknowledgments

The research has been funded by the Institute of Physical Geography and Environmental Planning at Adam Mickiewicz University. The authors would specially like to thank the anonymous reviewers for their helpful comments and feedbacks for this manuscript.

Author's contribution

Agnieszka Glinko performed all the statistical operations (data normalization, regression modelling, accuracy assessment) and wrote the manuscript. Cezary Kaźmierowski carried all laboratory analyses. Jan Piekarczyk prepared spectral measurements. Cezary Kaźmierowski, Jan Piekarczyk and Sławomir Królewicz designed the study, gathered all soil samples and contributed to the manuscript writing.

References

- Avola G., Gennaro, S.F. di Cantini C., Riggi E., Muratore F., Tornambè C., Matese A., 2019. Remotely sensed vegeta-

- tion indices to discriminate field-grown olive cultivars. *Remote Sensing* 11. DOI 10.3390/rs11101242.
- Bannari A., Morin D., Bonn F., Huete A.R., 1995. A review of vegetation indices. *Remote Sensing Reviews* 13: 95–120. DOI 10.1080/02757259509532298.
- Candiago S., Remondino F., Giglio M. de Dubbini M., Gattelli M., 2015. Evaluating multispectral images and vegetation indices for precision farming applications from UAV images. *Remote Sensing* 7: 4026–4047, MDPI AG. DOI 10.3390/rs70404026.
- Chong I.G., Jun C.H., 2005. Performance of some variable selection methods when multicollinearity is present. *Chemometrics and Intelligent Laboratory Systems* 78: 103–112. DOI 10.1016/j.chemolab.2004.12.011.
- Crippen R.E., 1990. Calculating the vegetation index faster. *Remote Sensing of Environment* 34: 71–73.
- Croft H., Kuhn N.J., Anderson K., 2012. On the use of remote sensing techniques for monitoring spatio-temporal soil organic carbon dynamics in agricultural systems. *Catena* 94: 64–74.
- Demattè J.A.M., Fiorio P.R., 2009. Orbital and laboratory spectral data to optimize soil analysis. *Scientia Agricola* 66(2): 250–257.
- Dematte J.A.M., Huete A.R., Ferreira Jr. L.G., Nanni M.R., Alves M.C., Fiorio P.R., 2009. Methodology for bare soil detection and discrimination by landsat TM image. *The Open Remote Sensing Journal* 2(1): 24–35. DOI 10.2174/1875413900902010024.
- De Paul Obade V., Lal R., 2013. Assessing land cover and soil quality by remote sensing and geographical information systems (GIS). *Catena* 104: 77–92.
- FAO [Food and Agriculture Organisation], 2021. Standard operating procedure for soil calcium carbonate equivalent - Titrimetric method. Online: www.fao.org/publications/card/en/c/CA8621EN/ (accessed July 17, 2022).
- Gasmi A., Gomez C., Chehbouni A., Dhiba D., Elfil H., 2022. Satellite multi-sensor data fusion for soil clay mapping based on the spectral index and spectral bands approaches. *Remote Sensing* 14(5). DOI 10.3390/rs14051103.
- Gasmi A., Gomez C., Lagacherie P., Zouari H., 2019. Surface soil clay content mapping at large scales using multispectral (VNIR-SWIR) ASTER data. *International Journal of Remote Sensing* 40(4): 1506–1533. DOI 10.1080/01431161.2018.1528018.
- Gholizadeh A., Boruvka L., Saberioon M.M., Kozák J., Vašát R., Nemeček K., 2015. Comparing different data pre-processing methods for monitoring soil heavy metals based on soil spectral features. *Soil and Water Research* 10: 218–227, Czech Academy of Agricultural Sciences. DOI 10.17221/113/2015-SWR.
- Gunathilaka M.D.K.L., 2021. Modelling the behavior of DVI and IPVI vegetation indices using multi-temporal remotely sensed data. *International Journal of Environment, Engineering & Education* 3(1): 9–16.
- International Standard (ISO) 11260, 1994. *Soil quality - Determination of effective cation exchange capacity and base saturation level using barium chloride solution*.
- International Standard (ISO) 10693, 2002. *Soil quality - Determination of carbonate content - Volumetric method*.
- IUSS Working Group WRB, 2015. World reference base for soil resources 2014, update 2015. International soil classification system for naming soils and creating legends for soil maps. World Soil Resources Reports No. 106. FAO, Rome.
- Lan Y., Thomson S.J., Huang Y., Hoffmann W.C., Zhang H., 2010, October. Current status and future directions of precision aerial application for site-specific crop management in the USA. *Computers and Electronics in Agriculture* 74: 34–38. DOI 10.1016/j.compag.2010.07.001.
- Liu W., Chen S., Qin X., Baumann F., Scholten T., Zhou Z., Sun W., et al., 2012. Storage, patterns, and control of soil organic carbon and nitrogen in the northeastern margin of the Qinghai-Tibetan Plateau. *Environmental Research Letters* 7 035401. DOI 10.1088/1748-9326/7/3/035401.
- Mammadov E., Denk M., Riedel F., Lewinska K., Kaźmierowski C., Glaesser C., 2020. Visible and near-infrared reflectance spectroscopy for assessment of soil properties in the Caucasus Mountains, Azerbaijan. *Communications in Soil Science and Plant Analysis* 51: 2111–2136. DOI 10.1080/00103624.2020.1820027.
- Martínez M.L.J., 2017. Relación entre el estado nutricional de los cultivos, las mediciones espectrales y las imágenes Sentinel 2. *Agroñomía Colombiana* 35: 205–215. DOI 10.15446/agron.colomb.v35n2.62857.
- Matese A., Gennaro S.F. di Berton A., 2017. Assessment of a canopy height model (CHM) in a vineyard using UAV-based multispectral imaging. *International Journal of Remote Sensing* 38: 2150–2160. DOI 10.1080/01431161.2016.1226002.
- Mehlich A., 1984. Mehlich 3 Soil Test Extractant: A Modification of Mehlich 2 Extractant. *Communications in Soil Science and Plant Analysis* 15: 1409–1416. DOI 10.1080/00103628409367568.
- Milton E.J., 1987. Review article: Principles of field spectroscopy. *International Journal of Remote Sensing* 8: 1807–1827. DOI 10.1080/01431168708954818.
- Minasny B., McBratney A.B., 2008. Regression rules as a tool for predicting soil properties from infrared reflectance spectroscopy. *Chemometrics and Intelligent Laboratory Systems* 94: 72–79. DOI 10.1016/j.chemolab.2008.06.003.
- Motohka T., Nasahara K.N., Oguma H., Tsuchida S., 2010. Applicability of Green-Red Vegetation Index for remote sensing of vegetation phenology. *Remote Sensing* 2: 2369–2387. DOI 10.3390/rs2102369.
- Nanni M.R., Dematte J.A.M., 2006. Spectral Reflectance Methodology in Comparison to Traditional Soil Analysis. *Soil Science Society of America Journal* 70: 393–407. DOI 10.2136/sssaj2003.0285.
- Nelson D.W., Sommers L.E., 1996. Total carbon, organic carbon, and organic matter. In: Sparks D.L., et al. (eds.), *Methods of soil analysis. Part 3. Chemical methods*. SSSA Book Series No. 5, SSSA and ASA, Madison, WI: 961–1010.
- Ng W., Minasny B., Jeon S.H., McBratney A., 2022. Mid-infrared spectroscopy for accurate measurement of an extensive set of soil properties for assessing soil functions. *Soil Security* 6: 100043. DOI 10.1016/j.soisec.2022.100043.
- Nguyen T.T., Janik L.J., Raupach B.M., 1991. Diffuse Reflectance Infrared Fourier Transform (DRIFT) spectroscopy in soil studies. *Australian Journal of Soil Research* 29: 49–67.
- Peng Y., Xiong X., Adhikari K., Knadel M., Grunwald S., Greve M.H., 2015. Modeling soil organic carbon at regional scale by combining multi-spectral images with laboratory spectra. *PLoS ONE* 10. DOI 10.1371/journal.pone.0142295.
- Peng, Y., Zhao L., Hu Y., Wang G., Wang L., Liu Z., 2019. Prediction of soil nutrient contents using visible and near-infrared reflectance spectroscopy. *ISPRS International Journal of Geo-Information* 8. DOI 10.3390/ijgi8100437.
- PN-ISO-10390, 1997. *Soil quality-determination of pH*. Polish Committee for Standardization, Warsaw.

- Quinlan J.R., 1992. Learning with continuous classes. *Proceedings of the 5th Australian joint Conference on Artificial Intelligence*, 16–18 November 1992, Hobart: 343–348.
- Rinnan Å., Berg F. van den Engelsen S.B., 2009, November. Review of the most common pre-processing techniques for near-infrared spectra. *Trends in Analytical Chemistry* 28(10): 1201–1222. DOI 10.1016/j.trac.2009.07.007.
- Saberioon M., Amin M., Gholizadeh A., 2012. Estimation of nitrogen of rice in different growth stages using Tetra-cam agriculture digital camera. *The Philippine Agricultural Scientist* 96(1): 116–121.
- Saeyns W., Mouazen A.M., Ramon H., 2005. Potential for onsite and online analysis of pig manure using visible and near infrared reflectance spectroscopy. *Biosystems Engineering* 91: 393–402. DOI 10.1016/j.biosystemseng.2005.05.001.
- Swain K.C., Thomson S.J., Jayasuriya H.P.W., 2010. Adoption of an unmanned helicopter for low-altitude remote sensing to estimate yield and total biomass of a rice crop. *Transactions of the ASABE* 53: 21–27.
- Vega F.A., Ramírez F.C., Saiz M.P., Rosúa F.O., 2015. Multi-temporal imaging using an unmanned aerial vehicle for monitoring a sunflower crop. *Biosystems Engineering* 132: 19–27. DOI 10.1016/j.biosystemseng.2015.01.008.
- Vestergaard R.J., Vasava H.B., Aspinall D., Chen S., Gillespie A., Adamchuk V., Biswas A., 2021. Evaluation of optimized preprocessing and modeling algorithms for prediction of soil properties using vis-nir spectroscopy. *Sensors* 21, 6745: 2–18. DOI 10.3390/s21206745.
- Viscarra Rossel R.A., Walvoort D.J.J., McBratney A.B., Janik L.J., Skjemstad J.O., 2006. Visible, near infrared, mid infrared or combined diffuse reflectance spectroscopy for simultaneous assessment of various soil properties. *Geoderma* 131: 59–75. DOI 10.1016/j.geoderma.2005.03.007.
- Wenjun J., Zhou S., Jingyi H., Shuo L., 2014. In situ measurement of some soil properties in paddy soil using visible and near-infrared spectroscopy. *PLoS ONE* 9(8), e105708. DOI 10.1371/journal.pone.0105708.
- Wetterlind J., Stenberg B., Söderström M., 2008. The use of near infrared (NIR) spectroscopy to improve soil mapping at the farm scale. *Precision Agriculture* 9: 57–69. DOI 10.1007/s11119-007-9051-z.
- Wojewódzki Inspektorat Ochrony Środowiska w Poznaniu 2013. Raport o stanie środowiska w Wielkopolsce w roku 2012. Biblioteka Monitoringu Środowiska, Poznań.
- Wold S., Sjostrom M., Eriksson L., Sweden S., 2001. PLS-regression: A basic tool of chemometrics. *Chemometrics and Intelligent Laboratory Systems* 58(2): 109–130. DOI 10.1016/S0169-7439(01)00155-1.
- Xu S., Wang M., Shi X., Yu Q., Zhang Z., 2021. Integrating hyperspectral imaging with machine learning techniques for the high-resolution mapping of soil nitrogen fractions in soil profiles. *Science of the Total Environment* 754, 142135. DOI 10.1016/j.scitotenv.2020.142135.