

Szent István University

Ph.D. School of Environmental Sciences

Prediction of soil properties and determination of soil classification units
by means of reflectance spectroscopy

Thesis of Ph.D. dissertation

Csorba Ádám

Gödöllő

2017

Ph.D. School: Ph.D. School of Environmental Sciences

Discipline: Environmental sciences

Head of School: Csákiné Dr. Michéli Erika, D.Sc.
Professor
Szent István University
Faculty of Agricultural and Environmental Sciences
Institute of Environmental Sciences
Department of Soil Science and Agrochemistry

Scientific supervisor Csákiné Dr. Michéli Erika
Professor
Szent István University
Faculty of Agricultural and Environmental Sciences
Institute of Environmental Sciences
Department of Soil Science and Agrochemistry

Approval of Head of School

Approval of Scientific supervisor

1. Introduction and objectives

The significance of the agricultural and other environmental functions of soils have been highlighted by environmental changes, new demands for energy, the need for sustainable development and the recognition of global problems. Due to the above mentioned facts nowadays there is a serious need for reliable, accessible and good quality soil data and information. This growing interest in the soil has revealed the spatial, temporal and financial limits of conventional soil surveys, sampling and analytical methods. Nowadays a global trend is forming which supports the development and use of technologies allowing rapid, cheap and environment friendly data collection and analysis (VISCARRA ROSSEL and CHEN, 2011). Reflectance spectroscopy, based on reflectance measurements in the visible (VIS) and near-infrared (NIR) regions of the electromagnetic spectrum (350 – 2500 nm), is able to satisfy these requirements.

Based on the results of studies dealing with the possibilities of soil spectral characterization the reflectance measurements in the VIS – NIR are suitable to characterize the soils in a qualitative and quantitative way as well (VISCARRA ROSSEL et al., 2006). The application of the technology in the soil classification procedure is not a standard yet, although in the modern soil classification systems the integration of standard measurements and numeric data in the determination of diagnostic horizons and properties defined by strict threshold limits is essential (MICHÉLI, 2011). Because the spectral measurements provide information simultaneously on soil parameters and properties playing key role in defining of certain diagnostic soil classification units the technology has high potential in supporting the soil classification efforts.

Based on the above mentioned requirements, the objectives of this research are the followings:

1. The development of multivariate mathematical-statistical chemometric models based on the spectral reflectance of samples collected in the Gödöllő hilly region and reference soil data determined by conventional laboratory techniques. By the application of these models the soil organic carbon (SOC), calcium carbonate (CaCO_3) content and pH of soils can be estimated based on their spectral reflectance.
2. The investigation of the depth distribution of spectral properties and soil parameters based on selected soil profiles.
3. The application of the spectral reflectance data to define diagnostic classification units, and to compare the depth distribution of generated spectral variables and soil parameters determined by conventional laboratory methods.
4. The application of spectroscopic measurements in diagnostic soil classification. The application of spectra-based taxonomic distance measurements to determine the taxonomic relationships of the studied soils. The comparison of the soil groups defined by the distance metrics and the units defined by the World Reference Base for Soil Resources (IUSS Working Group WRB, 2015).

2. Materials and methods

2.1 Methods of spectral based soil parameter prediction

The objective of this part of the study was to test the suitability of VIS – NIR spectral measurements to predict soil parameters. The conduction of the research was justified by the need of the determination of SOC, CaCO₃ content and pH for a large number of soil samples collected during a soil survey which aimed the mapping of the carbon stock of the Gödöllő hilly region soils. For the multivariate calibration beside the spectral measurements conduction of reference laboratory analysis is also required. Based on laboratory reference SOC and CaCO₃ content and pH measurements on selected calibration samples representing the spectral variability of the entire dataset, and on the spectral reflectance of those samples the development of mathematical-statistical chemometric models can be established. The developed models can be used to predict these soil properties based on the spectral reflectance of samples with unknown composition. The main reason to use of chemometric methods in soil science is to replace (partially or even entirely) conventional laboratory methods with a simple and fast technique. This will enable the cheaper and faster determination of properties of given number of samples, or the analysis of more samples during given time period and financial circumstances.

2.1.1 Study area, soil sampling

The study area included typical agricultural and forested areas in the Gödöllő hilly region. Following the Land Degradation Surveillance Framework (LDSF) (VÅGEN et al., 2015) of Africa Soil Information System (AfSIS) project 160 sampling points were defined. Finally 154 sampling points were visited. The samples were taken from 0-20, 20-40, 40-60, 60-80, 80-100 cm depth intervals. The samples were air dried, sieved through a 2 mm sieve before further analyses.

2.1.2 Spectral and laboratory reference analysis

For the laboratory spectroscopic measurements the Analytical Spectral Devices (ASD) FieldSpec 3 MAX portable spectroradiometer in the Hungarian Institute of Agricultural Engineering of the National Agricultural Research and Innovation Centre was used. The instrument records the spectra in the spectral region between 350 and 2500 nm on 2151 spectral bands. The spectral reflectance of the collected 770 soil samples was acquired using the High Intensity Contact Probe attachment. By using this probe the direct contact with the samples, thus the minimization of the environmental effects negatively affecting the quality of measures was ensured. By one measurement the probe provides spectral information from an area of 10 mm diameter.

For laboratory reference measurements ~10 % (80) of the samples were chosen by random method. Further samples were chosen based on the visual interpretation of the spectra and on the results of the chemometric analysis. Finally SOC, CaCO₃ content and pH determination was performed on 135, 140 and 160 samples, respectively.

2.1.3 Spectroscopic transformations

As first step the transformation of spectral reflectance to spectral absorbance was performed using the following equation:

$$Absorbance = \log\left(\frac{1}{Reflectance}\right)$$

As next step Multiplicative Scatter Correction (MSC, GELADI et al., 1985) and Standard Normal Variate (SNV, BARNES et al., 1989) transformation were performed. Both methods have key role in reducing the effect of wavelength dependent scattering phenomena. On the SNV corrected spectra first and second order derivatives were calculated using the Savitzky-Golay method (SAVITZKY and GOLAY, 1964), which transformations reduce noise, resolve and enhance the overlapping absorption features.

2.1.4 Chemometric analysis

Chemometric methods based on multivariate mathematical-statistical techniques are suitable to quantify the statistical relationship between independent (X-variables, eg.: spectral reflectance of soil samples) and dependent (Y-variables, eg.: reference soil parameters) (DUCKWORTH, 1998). The developed chemometric models enable the estimation of soil parameters from the spectral reflectance.

2.1.5 The applied methods – Partial Least Squares Regression

In this study Partial Least Squares Regression (PLSR) with leave-one-out cross validation was used to calibrate the spectral data with the reference laboratory soil data. The PLSR algorithm selects successive orthogonal factors that maximize the covariance between predictor (spectra) and response (laboratory) data. The PLSR is one of the most widely used chemometric methods, which is the most suitable when the number of variables exceed the number of samples, and there is a high multicollinearity between the variables.

2.2 The methodology of the spectral property based soil classification

In this part of the study the application of VIS – NIR reflectance spectroscopy in soil characterization and classification is demonstrated and evaluated. The theoretical basis of the research is that a wide number of soil constituents - which play key role in soil classification - have characteristic spectral fingerprints in the VIS – NIR region. Thus the spectral measurements provide qualitative and quantitative information on those soil properties whose depth function distribution is the basis of soil horizon delineation. Furthermore the taxonomic relationship of soils is determined by the similarity or dissimilarity in the distribution of these soil parameters along the profile. Because the spectral data incorporates information on these soil properties, by the visualization of the depth-function spectral variability one can get a picture on the distribution of main soil parameters, and by using numeric methods the taxonomic relationship can be expressed. Figure 1. shows the schematic representation of the workflow.

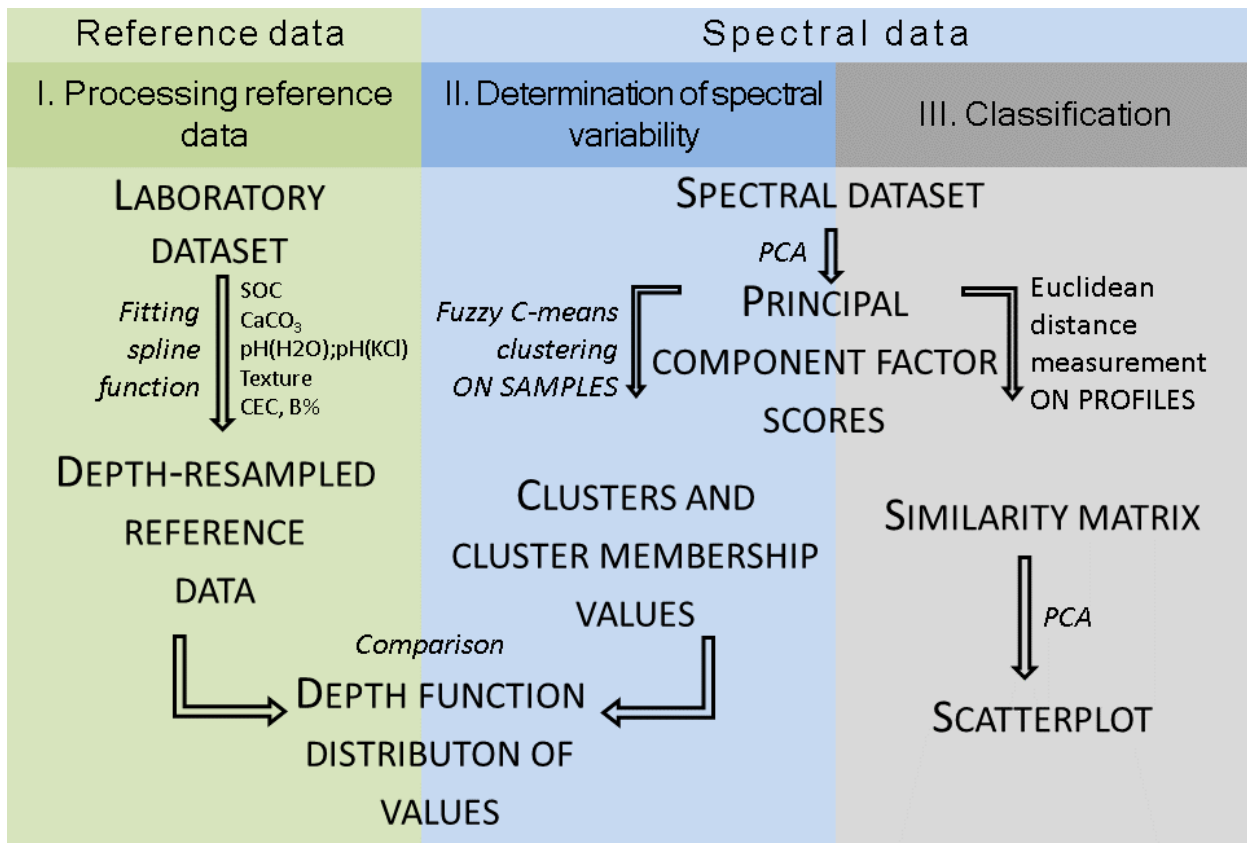


Figure 1.: The schematic workflow of the spectral property based soil classification

2.2.1 The selected soil profiles and soil sampling

During the study thirteen soil profiles from different locations in Hungary were evaluated. The emphasis were put on Chernozems, Kastanozems and Luvisols. The exact locations of the profiles were chosen based on previous knowledge on their properties. During the field work each pit was prepared to a depth of 1.5 m. The profile description was conducted by using the FAO guideline (FAO, 2006). Using the field descriptions and the auxiliary laboratory data the soils were classified to the reference soil group (RSG) level according to the WRB classification (IUSS WORKING GROUP, WRB, 2015). For detailed characterization the principal and supplementary qualifiers were also determined.

The soil sampling scheme was designed to be able to characterize the spectral variability of the soil profiles as detailed as it is possible. Soils were sampled by 5 cm depth intervals to 1 m depth, and by 10 cm intervals between 1 and 1.5 m. This sampling scheme resulted in 25 samples per profile that were air dried, crushed and sieved through a 2 mm sieve. For reference laboratory measurements samples from genetic horizons were collected from each soil profile.

2.2.2 Laboratory reference analyses and depth standardization

Samples collected from the genetic horizons were analyzed for SOC, CaCO₃ content, texture, pH, cation exchange capacity, exchangeable bases and base saturation. Considering that the sampling intervals differ in case of each soil profile mass preserving spline function was fitted on the genetic horizon based reference data. This procedure allowed the estimation of the soil properties for the fixed depth intervals from the

input horizon data. For resampling procedure the depth intervals were identical as described for the spectral measurements.

2.2.3 Spectral measurements and spectral transformations

The VIS – NIR reflectance spectra of the 325 soil samples from the 13 profiles were acquired using the Analytical Spectral Devices (ASD) FieldSpec 3 MAX portable spectroradiometer with a Contact Probe attachment. For further analysis the spectra were transformed to units of absorbance ($\log(1/\text{reflectance})$) and 1st derivatives were calculated using Savitzky-Golay method.

2.2.4 Derivation of spectral variables

To reduce the high dimensionality of the dataset, and to derive spectral variables which express the underlying information in the spectral data, and can be plotted against the depth principal component analysis (PCA) was performed on the spectral database as the first step. The number of the principal components (PCs) to be involved into the further analysis was determined based on the expressed variance of the principal components. PCs expressing less than 2 % of the total variance were ignored. The PC factor scores calculated for each sample were used as profile descriptor variables that represent the depth function spectral variability of the profiles.

2.2.5 Testing the “profile description ability” of the spectral dataset

In order to get picture on the “profile description ability” of the spectral dataset Fuzzy C-means clustering was performed on the matrix containing the PC score values. The method allows each data point (sample) to belong several clusters with a degree of membership to each one. The depth function spectral variability was examined by plotting the resulting Fuzzy C membership values against the depth.

2.2.6 Spectral property based taxonomic distance measurements

The similarity and dissimilarity of the soil profiles were determined by Euclidean distance measurement which is a widely used method in determining taxonomic relationships (MINASNY et al., 2009). The calculation was applied to the matrix containing the principal component factor scores per soil sample arranged per soil profile. To visualize the similarity and dissimilarity of the soils principal component analysis was performed on the similarity matrix containing the Euclidean distance values.

3. Results and discussion

3.1 The results of the chemometric analysis

3.1.1 Descriptive statistics of reference soil parameters

Table 1. and Figure 2. show the descriptive statistics of the reference SOC, CaCO₃ and pH values.

Table 1.: The descriptive statistics of the reference soil parameters

	Mean	Max	Min	SD	Median	Q1	Q3
SOC (%)	1.6	5.9	0.1	1.1	1.4	0.7	2.2
CaCO ₃ (%)	14,7	31.8	0.6	9.6	15	5.4	23.1
pH _(H₂O)	7.4	8.4	5.1	0.8	7.7	7.0	8.0
pH _(KCl)	6.7	7.7	3.8	1.1	7.0	6.3	7.4

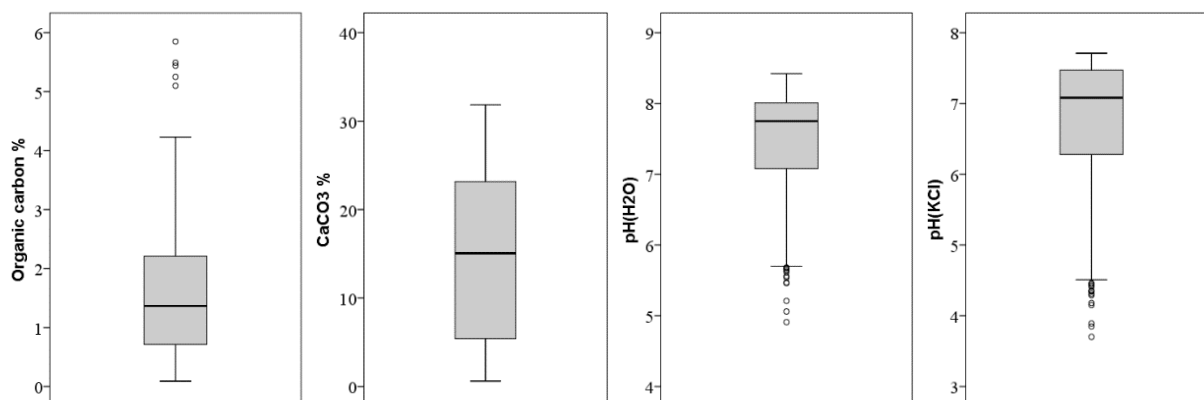


Figure 2.: Statistical parameters of reference soil parameters indicated by boxplots

3.1.2 Descriptive statistics of spectral measurements

The descriptive statistics of the spectral data for the whole dataset and for the samples chosen for calibration are compared on Figure 3. The minimum, maximum, mean, Q1 and Q3 values for the whole dataset are indicated with continuous lines, and for the calibration samples with dotted lines. For pH calibration the same samples were chosen. Based on the plots the chosen samples are representative to the whole dataset.

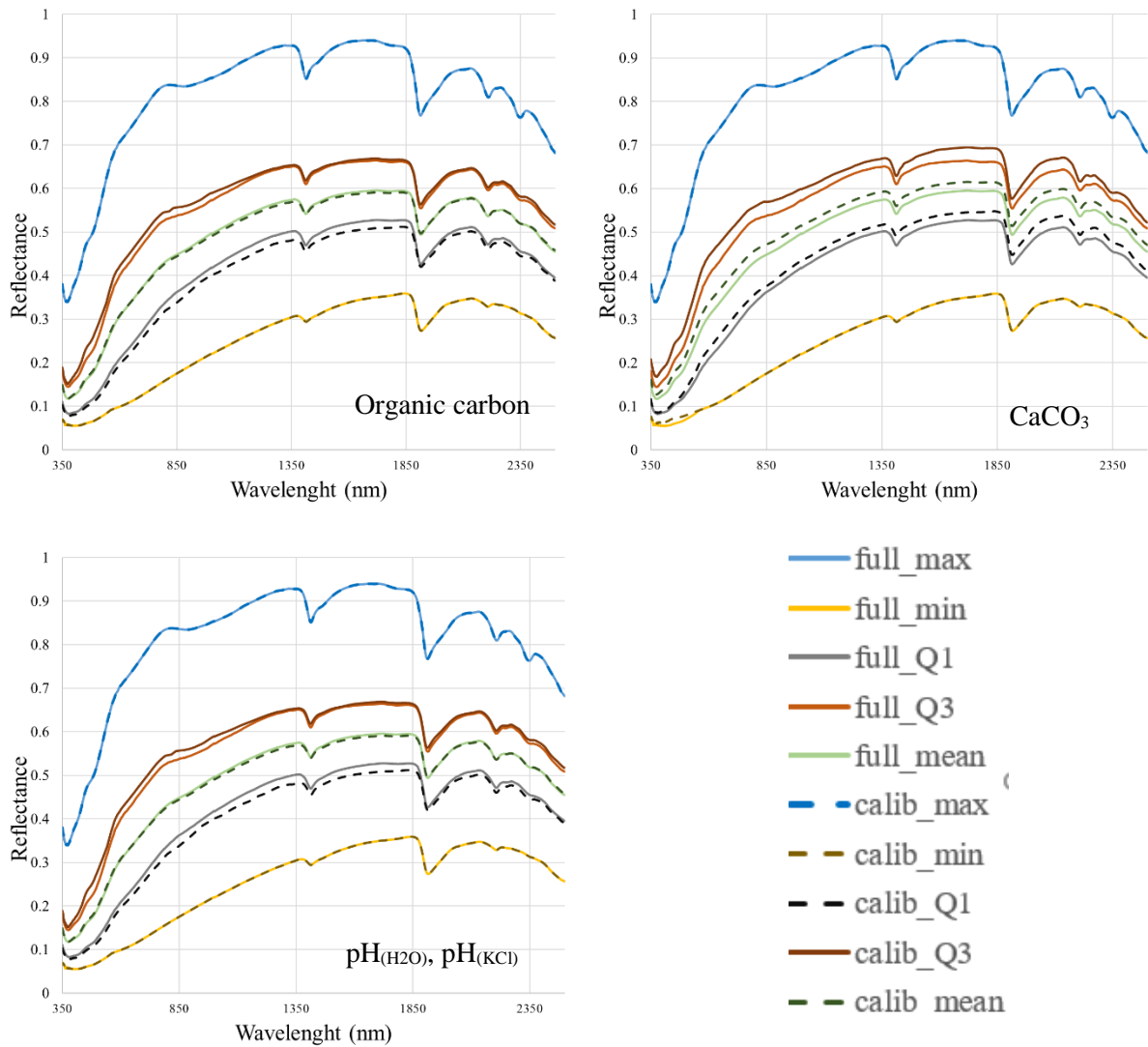


Figure 3.: The descriptive statistics of the spectral data for the whole dataset (continuous lines) and for the samples chosen for calibration (dotted lines). Samples chosen for a.) SOC b.) CaCO_3 , c.) $\text{pH}_{(\text{H}_2\text{O})}$ and $\text{pH}_{(\text{KCl})}$ calibration

3.1.3 The result of PLSR modeling

Figure 4. shows the scatterplot of laboratory measured versus the PLSR predicted values of SOC, CaCO_3 content, $\text{pH}_{(\text{H}_2\text{O})}$ and $\text{pH}_{(\text{KCl})}$. In case of each parameter during the calibration step higher value of R^2 and lower value of RMSE were calculated than during the validation step. This is due to that after the calibration step the model tests itself with the same dataset which was used during the calibration. But during the validation step the result of the leave-one-out cross validation can be seen, which approximates the real performance of the developed model.

The high value of R^2 in case of SOC and CaCO_3 is can be explained by the strong spectral activity of both soil constituent in the VIS – NIR region. The RMSE values ($\text{RMSE}_{\text{SOC}} = 0.39$, $\text{RMSE}_{\text{CaCO}_3} = 3.31$) are higher than most of the values published in literature (VISCARRA ROSSEL et al., 2006). The higher RMSE values are mostly due to the application of a portable spectroradiometer. Using a portable device for spectral measurements it is needed to taking into consideration of a wide number of factors which

negatively affect the data acquisition. To improve the performance of the models the development of a common and standard measurement protocol is needed (BEN DOR et al., 2010).

The prediction of pH proved to be very reliable ($R^2_{\text{pH}(\text{H}_2\text{O})} = 0.87$, $\text{RMSE}_{\text{pH}(\text{H}_2\text{O})} = 0.29$; $R^2_{\text{pH}(\text{KCl})} = 0.92$, $\text{RMSE}_{\text{pH}(\text{KCl})} = 0.3$). However the VIS – NIR spectra reflect the physical/chemical/mineralogical composition of soils it is possible to estimate soil properties which are driven by different soil constituents and do not have direct effect on the spectra. Thus, the estimation of pH cannot be done directly but indirectly through the amount of soil constituents which do have spectral fingerprint and show good correlation with the pH.

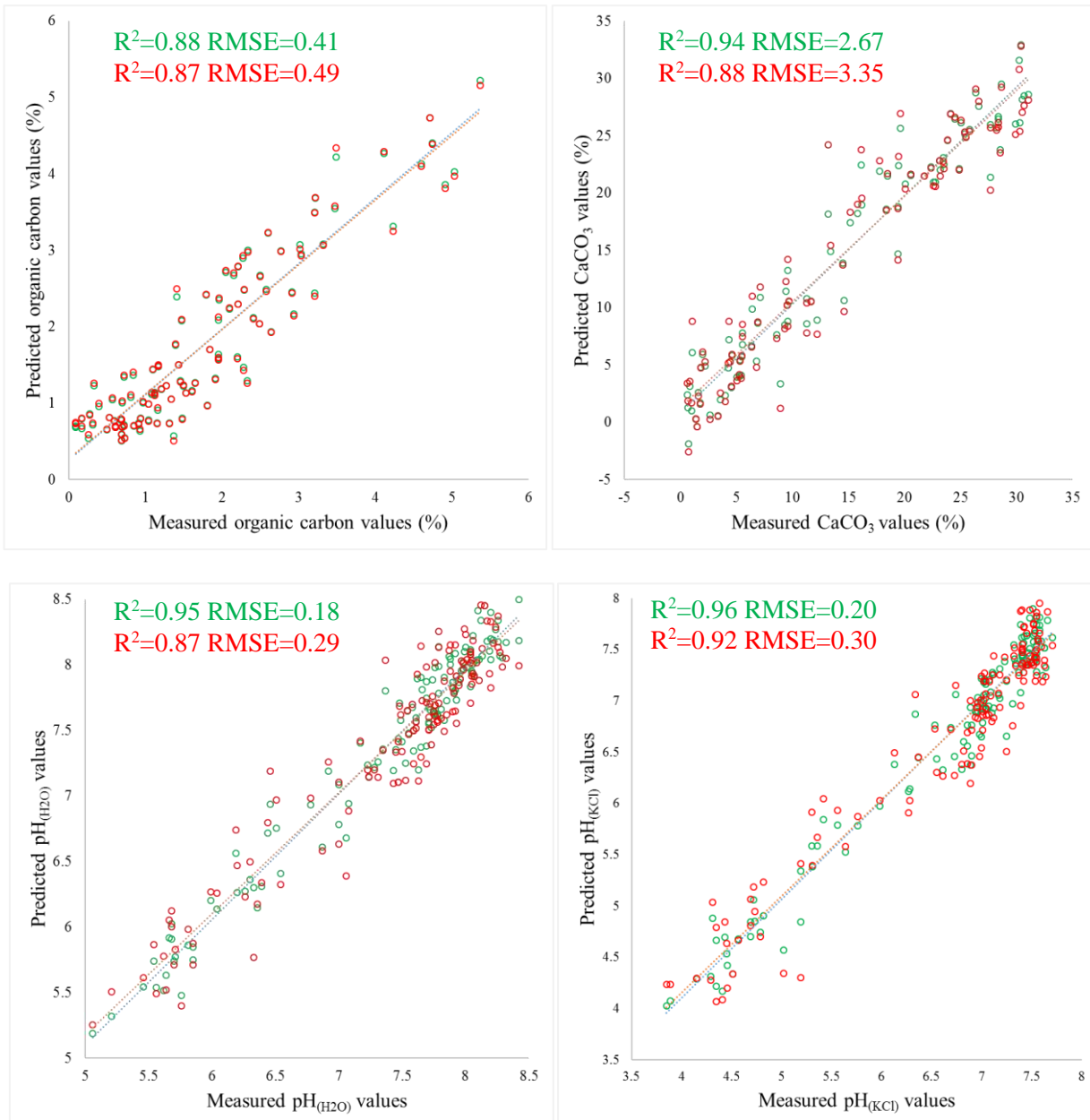


Figure 4.: Measured versus predicted values of organic carbon, CaCO_3 content, $\text{pH}_{(\text{H}_2\text{O})}$ and $\text{pH}_{(\text{KCl})}$.

3.2 The results of spectral property based soil classification

3.2.1 Results of field work and classification

During the field work eight soils were classified as Chernozems, two as Kastanozems and three as Luvisols according to the WRB 2015.

Table 2.: The diagnostic horizons/properties/materials, WRB Reference Soil Groups and principal and supplementary qualifiers determined for the 13 investigated soil profiles.

ID	Diagnostic horizons	Diagnostic horizons, materials	WRB classification IUSS WORKING GROUP (2015)	Hungarian classification STEFANOVITS (1972), JASSÓ (1989)
CH1	Calcic (90 -); Chernic (0-65)	Protocalcic properties	Endocalcic Vermic CHERNOZEM (Pantoloamic, Pachic)	Leached chernozems (180)
CH2	Calcic (50 -); Chernic (0-50)	Protocalcic properties	Katocalcic CHERNOZEM (Pantoloamic)	Typical Pseudomyceliar chernozems (191)
CH3	Calcic (80-); Chernic (0-80)	Protocalcic properties;	Endocalcic Vermic CHERNOZEM (Pantoloamic, Pachic)	Leached chernozems (180)
CH4	Calcic (80-); Chernic (0-100)	Protocalcic properties;	Endocalcic Vermic CHERNOZEM (Episiltic, Katoloamic, Pachic)	Typical Pseudomyceliar chernozems (191)
CH5	Calcic (30-); Chernic (0-70)	Protocalcic properties; Gleyic properties	Katocalcic Bathyglyeyic Vermic CHERNOZEM (Pantoloamic, Pachic)	Typical Pseudomyceliar chernozems (191)
CH6	Calcic (70-); Chernic (0-70)	Protocalcic properties; Gleyic properties	Endocalcic Endogleyic CHERNOZEM (Pantoloamic, Pachic)	Typical Pseudomyceliar chernozems (191)
CH7	Calcic (75-); Chernic (0-75)	Protocalcic properties; Gleyic properties	Endocalcic Bathyglyeyic CHERNOZEM (Pantoloamic, Pachic)	Typical Pseudomyceliar chernozems (191)
CH8	Calcic (50-); Chernic (0-75)	Protocalcic properties	Katocalcic CHERNOZEM (Pantoloamic, Pachic)	Typical Pseudomyceliar chernozems (191)
KZ1	Argic (75-90); Calcic (90 -); Mollic (0-30)	Protocalcic properties	Endocalcic Endoluvic KASTANOZEM (Pantoloamic)	Non-calcareous chernozem brown forest soil (161)
KZ2	Argic (65-90); Calcic (90-); Mollic (0-25)	Protocalcic properties	Endocalcic Endoluvic KASTANOZEM (Pantoloamic)	Non-calcareous chernozem brown forest soil (161)
LU1	Argic (30-95); Calcic (95-)	Protocalcic properties	Endocalcic LUVISOL (Epiloamic, Amphiclayic Cutanic, Differentic)	Non-podzolic brown forest soil with clay illuviation (112)
LU2	Argic (50-95); Calcic (95-)	Protocalcic properties	Endocalcic LUVISOL (Epiloamic, Amphiclayic, Cutanic, Differentic)	Non-podzolic brown forest soil with clay illuviation (112)
LU3	Argic (0-150)	Stagnic properties	Nudiargic Amphistagnic LUVISOL (Pantoclayic Cutanic Epidystric Pantoclayic)	Stagnant brown forest soil with clay illuviation (122)

3.2.2 Result of the PCA

The principal component analysis performed on the spectral database resulted in eight factors that express 98 % of the total variance in the dataset. PCs explaining less than 2% of the total variance were excluded from the analysis.

3.2.3 Result of classification of factor scores using Fuzzy C-means clustering

Figure 5. shows the distribution of the samples along the first three principal components that express the 92,3 % of the total variance. The color coding and the symbols refer to the determined WRB diagnostic horizon from where the samples were collected (Figure 5.a) and to the Fuzzy C-means cluster the samples belong to (Figure 5.b). Based on the visual inspection of the 3D scatterplots the majority of samples clustered to the Cluster A were part of a Calcic horizon, samples clustered to the Cluster B were part of a Mollic horizon and samples of the Cluster C were part of an Argic horizon. Samples belonging to a Calcic and a Mollic horizon were classified either in Cluster A or B.

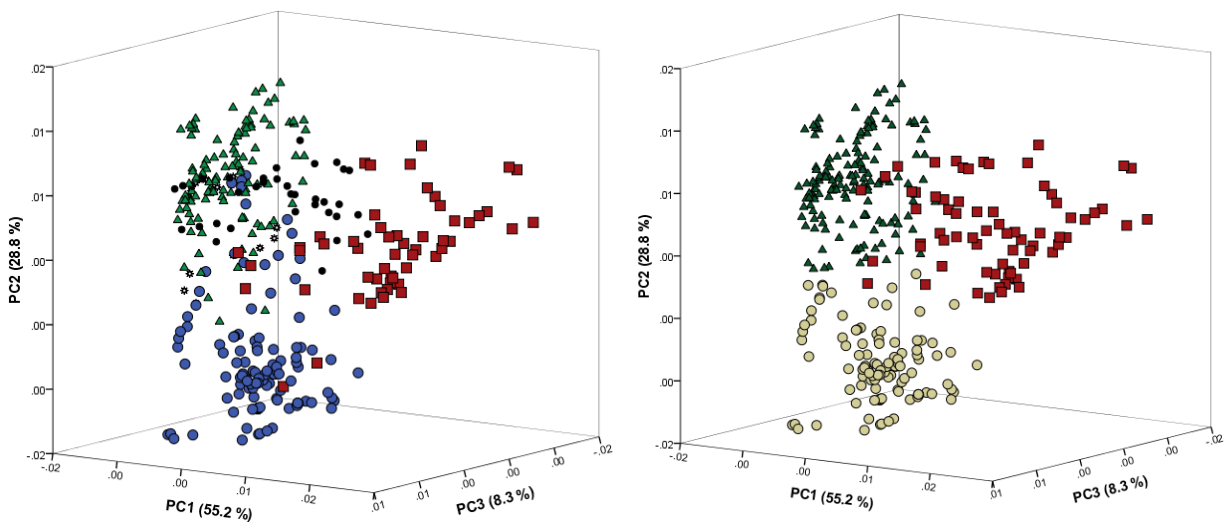


Figure 5.: a.) The 3D scatterplots showing the distribution of samples along first three principal components. A.) The coloring and symbols refer to the WRB diagnostic horizons

Calcic: ● Argic: ■ Mollic: ▲ Mollic - Calcic: ○ Non-diagnostic horizon: ●

b.) The coloring and symbols refer the Fuzzy C-means clusters

Cluster A: ○ Cluster B: ■ Cluster C: ▲

Figure 6. shows the spectral reflectance curves of samples per class. However the Fuzzy C-means clustering was performed on the first derivative of the absorbance spectra the original reflectance spectra per class are indicated because of the easier interpretation.

The spectra belonging to the Cluster A show the highest overall reflectance values compared to the Clusters B and C (Figure 6/a). The presence of the broad and shallow absorption features near 450 nm, 550 nm and 850 nm indicates the presence of iron-oxides, hydroxides and oxihydroxides (eg. hematite, goethite) (SCHWERTMANN, 1993). The absorption feature near 2340 nm characteristic for the different kinds of

carbonates is the most expressed in this cluster indicating the highest concentration of carbonate minerals (LAGACHERIE et al., 2008).

The spectra of the Cluster B have the lowest overall reflectance values compared to the Clusters A and B (Figure 6/b). The most convex shaped spectra belong to this cluster. The convexity of the spectra in wavelength region shorter than 1000 nm indicates the high organic matter content and its darkening effect. The increasing concavity of the curves indicates the decreasing organic matter content (STEVENS et al., 2010). The spectra belonging to the Cluster B have overall lack of or less expressed presence of broad and shallow absorption features on the shorter wavelengths (< 1300 nm) and the narrow features on longer (>2000 nm) wavelengths that indicate the high organic matter content and the mimicking of weaker features even in case of spectra with high overall reflectance (O'ROURKE and HOLDEN, 2011).

The spectra belonging to the Cluster C show the most expressed absorption features in the region of iron-oxides indicating the presence of a large amount of these components that give a deep yellow and red color to the soil (Figure 6/c). The deep absorption feature near 1400 and 1900 nm indicate the presence of H₂O bearing-minerals in high concentration. The first overtones of the OH stretch occur at about 1400 nm and the combinations of the H-O-H bend with the OH stretches are found near 1900 nm. The absorption feature near 2200 nm is the most expressed in the spectra belonging to this cluster indicating the highest amount of minerals that bear Al-OH groups in their lattice structure, particularly clay minerals (CHABRILLAT et al., 2003). The less expressed absorption near 2340 nm is due to the lack or very low concentration of carbonates.

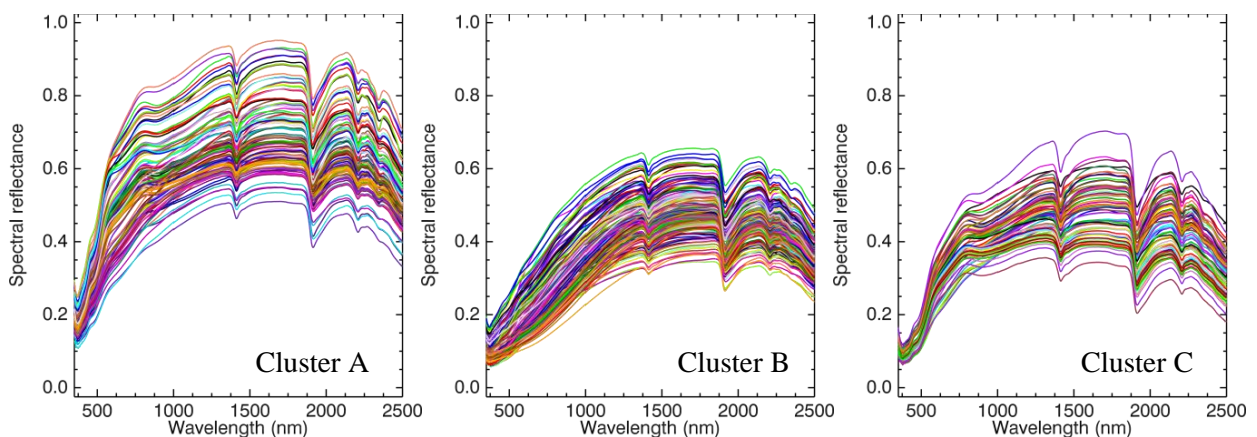


Figure 6.: The spectral reflectance of samples according to the cluster they belong to.

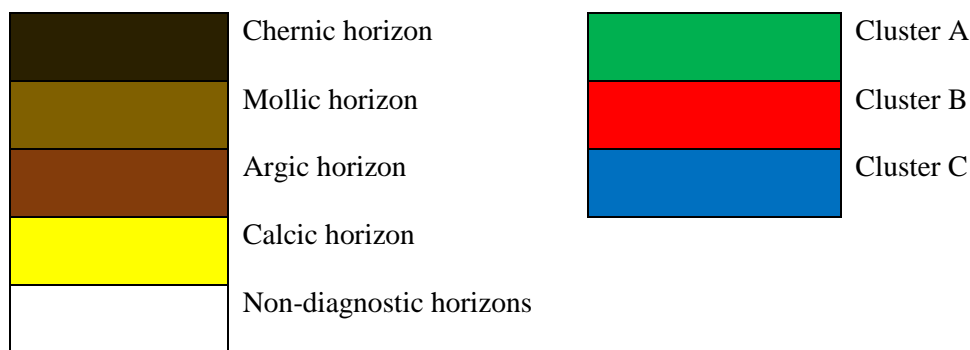
On the Figure 7. the distribution of the cluster membership values and the “spline-resampled” OC, CaCO₃ and clay content values versus the depth are shown. The cluster membership values of the Cluster A show similar depth-function pattern as the “spline estimated” CaCO₃ values. In case of soils showing high or moderate level of carbonate leaching (Luvisols: LU1, LU2; Kastanozems: KZ1, KZ2; Chernozems CH1, CH2) the membership values increase rapidly without or with very thin transition zone. In case of soils where the evidence of leaching process is less pronounced the membership values increase gradually. The first cluster appears always in the lower parts of the profiles which depth intervals are in good agreement with the Calcic horizon determined during the field and laboratory work.

The membership values of the Cluster B show similar depth function pattern as the “spline estimated” organic carbon values. The Cluster B appears always in the uppermost part of the profiles which depth intervals are in good agreement with the Mollic and Chernic horizons determined during the field work.

The explanation of the depth function distribution of the membership values of the Cluster C needs a different approach. Its depth function distribution shows similarity with the clay content only in cases where the soil was affected by considerable clay illuviation (LU1, LU2, KZ1, KZ2, CH1). The Cluster C appears almost always as a subsurface cluster and its depth and thickness is in good agreement with the Argic horizon determined in the field. The only exception is the LU3 profile where the whole profile belongs to the Cluster C. This profile is the most affected by erosion processes and it is eroded to the Argic horizon. Additionally, because of strong leaching processes no Calcic horizon is present.

On Figure 7. the comparison of distribution of Fuzzy C-means clusters, cluster membership values and reference soil parameters along the profile is shown. The comparison is exemplified by three soil profiles (CH2, KZ1 and LU2) and three soil parameters (SOC, CaCO₃ and clay content). Next to the photo of the profile the diagnostic horizons (*Diag*) and the Fuzzy C-means clusters (*Fuzzy klasz*) are shown.

Legend:

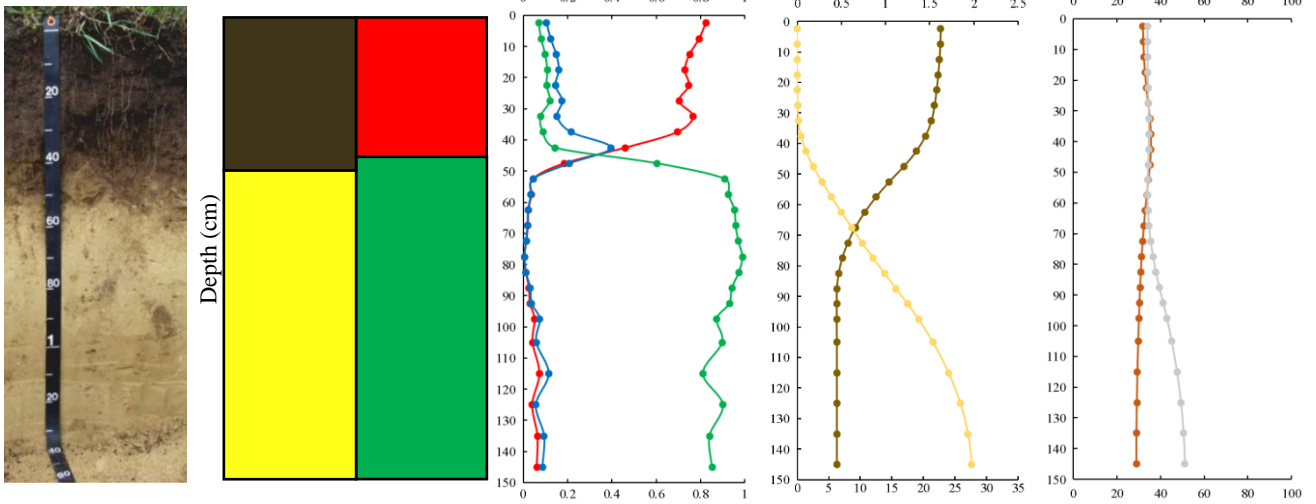


In *Clust.memb* block the Fuzzy C cluster membership values calculated for each sample are plotted against the depth. The cluster A, B and C are indicated with green, red and blue colors, respectively. On both horizontal axes the membership values can be read.

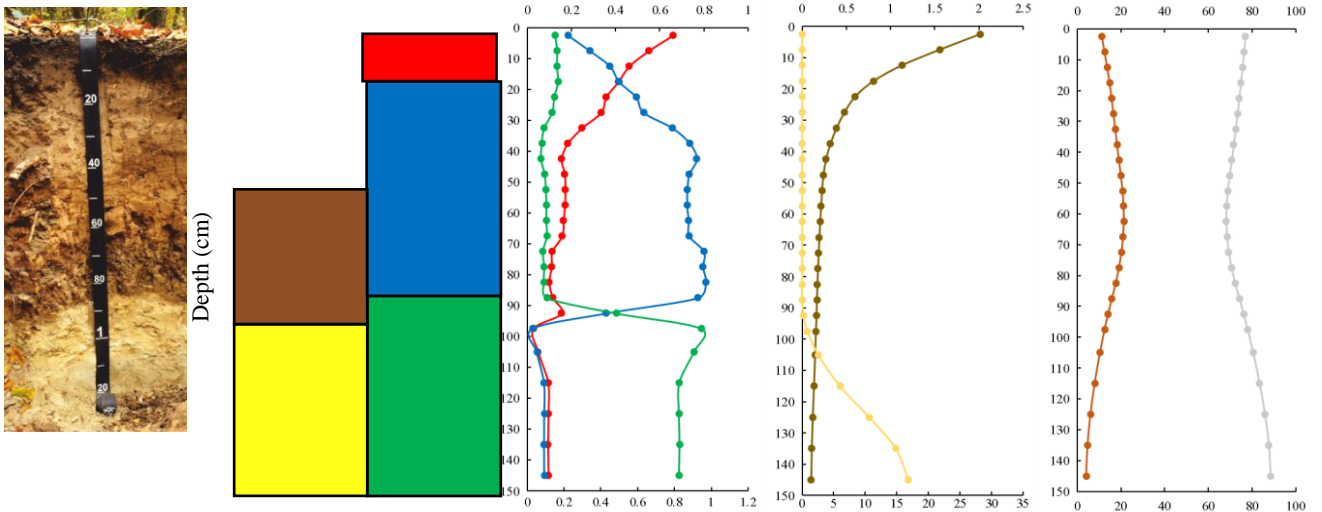
In the blocks of the reference soil parameters their distribution along the depth is shown. There are two parameters per block. On the principal horizontal (upper) axis the values of the first, on the secondary horizontal (lower) axis the values of the second parameter can be read.

Figure 7.: The depth distribution of the cluster membership values and the reference “spline-resampled” OC, CaCO₃ and clay content values

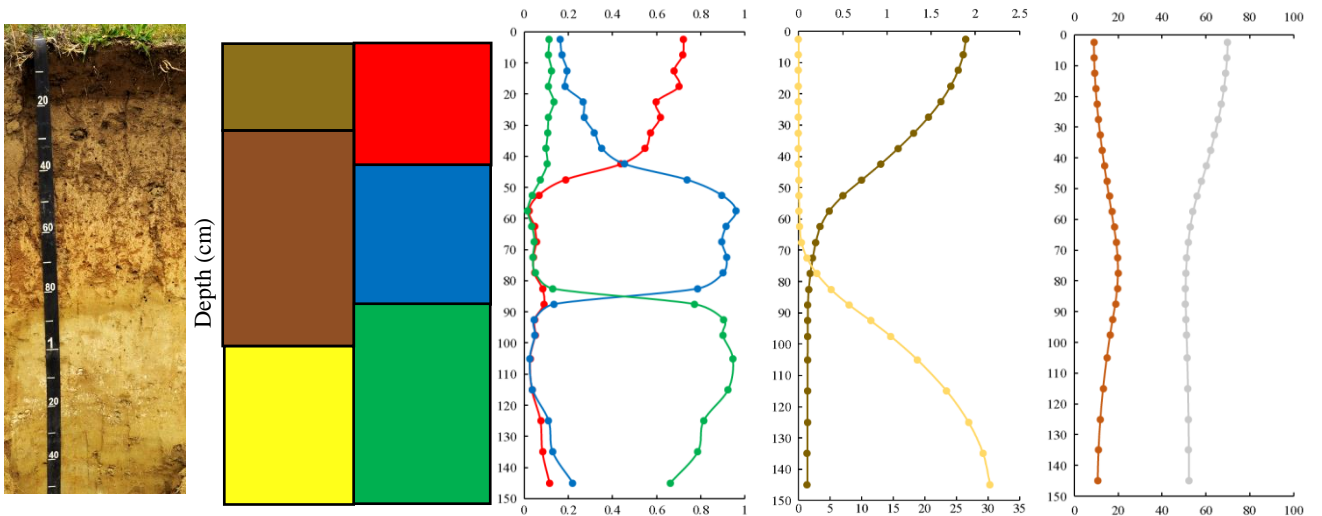
CH2



LU2



KZ1



3.2.4 Results of taxonomic distance measurements

To present the result of the similarity measurement Euclidean distance values were visualized with colored similarity matrix (Table 3.), where the “taxonomic” relationships are represented using a color ramp that indicates the difference between the studied soils. Green color represents greater similarity while the even reddish colors represent lesser similarities.

Table 3.: The similarity matrix of the 13 investigated soil profiles based on the depth function distribution of PC factor score value.

	CH1	CH2	CH3	CH4	CH5	CH8	CH6	CH7	KZ1	KZ2	LU1	LU2	LU3
CH1	0.000	0.052	0.038	0.043	0.058	0.047	0.064	0.074	0.054	0.054	0.069	0.077	0.092
CH2	0.052	0.000	0.059	0.052	0.069	0.040	0.057	0.072	0.062	0.055	0.077	0.093	0.090
CH3	0.038	0.059	0.000	0.024	0.040	0.046	0.054	0.055	0.075	0.070	0.088	0.103	0.108
CH4	0.043	0.052	0.024	0.000	0.027	0.036	0.037	0.050	0.078	0.071	0.091	0.106	0.114
CH5	0.058	0.069	0.040	0.027	0.000	0.048	0.046	0.059	0.094	0.087	0.106	0.121	0.129
CH8	0.047	0.040	0.046	0.036	0.048	0.000	0.036	0.053	0.082	0.074	0.097	0.111	0.118
CH6	0.064	0.057	0.054	0.037	0.046	0.036	0.000	0.053	0.093	0.083	0.105	0.122	0.132
CH7	0.074	0.072	0.055	0.050	0.059	0.053	0.053	0.000	0.106	0.097	0.117	0.137	0.143
KZ1	0.054	0.062	0.075	0.078	0.094	0.082	0.093	0.106	0.000	0.022	0.040	0.049	0.066
KZ2	0.054	0.055	0.070	0.071	0.087	0.074	0.083	0.097	0.022	0.000	0.043	0.061	0.074
LU1	0.069	0.077	0.088	0.091	0.106	0.097	0.105	0.117	0.040	0.043	0.000	0.054	0.067
LU2	0.077	0.093	0.103	0.106	0.121	0.111	0.122	0.137	0.049	0.061	0.054	0.000	0.070
LU3	0.092	0.090	0.108	0.114	0.129	0.118	0.132	0.143	0.066	0.074	0.067	0.070	0.000

Based on the similarity matrix (Table 4.) and the scatter plot (Figure 8.) the Chernozems and the Kastanozems/Luvisols can be discriminated obviously. Chernozems that show the evidence of groundwater (Gleyic colour patterns), and have very dark and thick organic carbon rich Chernic horizon were grouped together (CH7, CH5, CH6). Chernozems that lack the morphological features of groundwater but still have thick (> 70 cm) Chernic horizon , furthermore secondary CaCO₃ accumulation at various depths proved to be similar based on their depth function spectral properties (CH3, CH4, CH8). Chernozems showing thinner Chernic horizon (40-60 cm) and considerable level of secondary CaCO₃ accumulation are close together on the scatterplot (CH1, CH2). Kastanozems showing the evidence of Mollic horizon, furthermore Argic and Calcic are clearly separated (KZ1, KZ2). The three Luvisols do not form a clear group. However they show a transition starting from the soil having a thin but organic carbon rich surface horizon, and a secondary CaCO₃ accumulation in 100 cm depth (LU1) through a soil having similar properties but lower level of surface OC accumulation (LU2) to a soil where strong leaching and erosion processes are responsible for lack of CaCO₃ and a OC rich horizon (LU3).

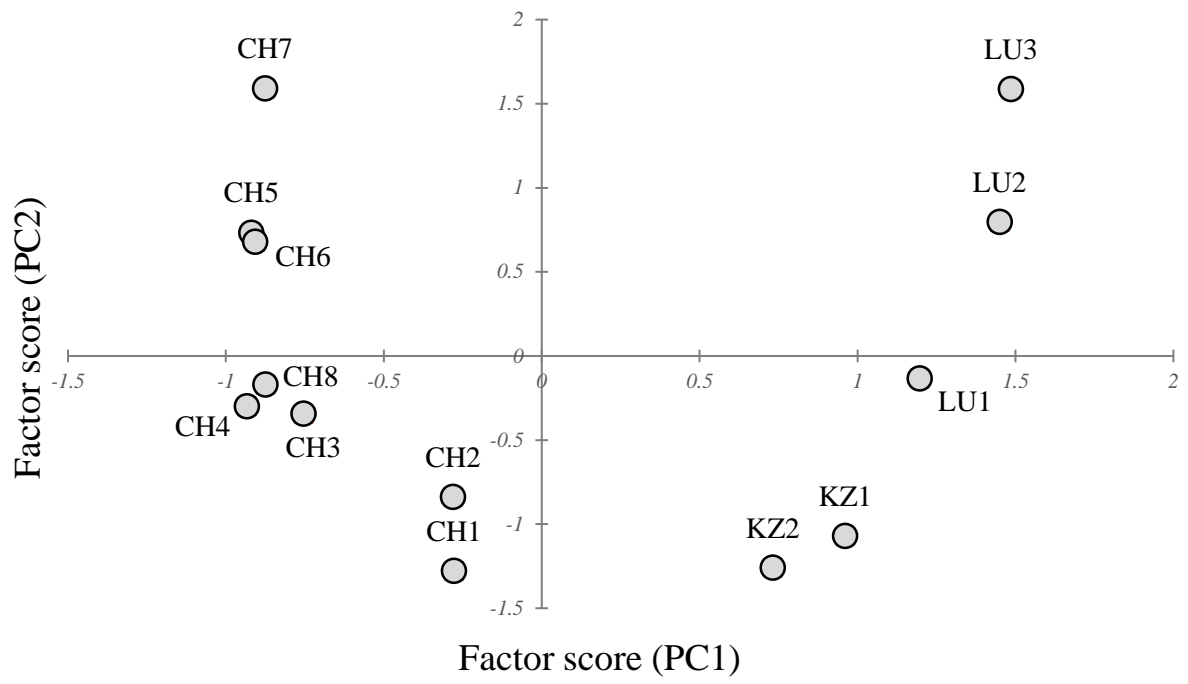


Figure 8.: Scatterplot of PC1 and PC2 factor scores resulting from the PCA on the similarity matrix.

4. Conclusions

The results of this research confirm that the reflectance measurement in the VIS – NIR region of the electromagnetic spectrum might have a key role in soil analysis and classification.

In spectral-based soil investigation the availability of good quality soil data (in this case the spectral reflectance) is also essential. In case of the use of a portable device the development and adherence of a common, standardized and detailed measurement protocol are even pronounced.

The selection of the calibration samples for the multivariate analysis is the next critical step towards the development of reliable chemometric models. For reference analysis the careful selection of calibration samples which are representative regarding to the whole population has a top priority. Besides using random methods, and visual interpretation of spectral data the application of methods which take into consideration of the statistical parameters of the whole population during the selection of the representative samples is recommended (eg.: Kennard-Stone algorithm).

During the multivariate chemometric analysis paying attention to the definition of samples which do not fit into the model is crucial. If the samples proved to be outliers because of measurement error, the measurement has to be performed again. But if the sample appears as an outlier because it belongs to a different soil population, the involvement of samples having similar composition into the calibration set, or using non-linear modeling methods is recommended.

By investigating the depth function spectral properties of soils it has been concluded that it is possible to derive variables which express the spectral variability along the soil profile and by plotting them against the depth show correspondence with the distribution of soil parameters using simple multivariate methods. Using the spectral data the delineation of horizons showing good accordance with determined diagnostic horizons is also possible, thus using spectral reflectance as input data in taxonomic studies can be very effective.

During the study 13 soils belonging to three Reference Soil Groups of the WRB were investigated which represent a narrow soil population. For further evaluation of the possibilities of using spectral measurements in determining taxonomic relationships it is necessary to involve soils which are affected by different soil forming factors.

The final conclusion of this work is that the widespread application of VIS – NIR reflectance spectroscopy might open extraordinary opportunities in practical soil science. But the requirement for success is the development of a spectral library which represents the pedological diversity of Hungary. With the establishment and continuous expansion of such a database the prediction of soil parameters will become more effective, the soil classification procedure will be more data driven and based on standard measurements it will become even objective.

5. Summary of scientific results

1. A spectral library representing the spectral variability of the Gödöllő hilly region soils has been developed.
2. By using multivariate chemometric methods, spectral models with known statistical measures were developed which can be used to predict soil organic carbon calcium carbonate content and pH of soils.
3. Based on the investigation of the depth spectral variability of soils the spectral reflectance data contain information on physical and chemical properties necessary to characterize the soils and to determine soil horizons.
4. The taxonomic distance measurement based on the depth spectral properties can distinguish the soil groups that are in good match with the WRB reference soil groups.

7. Related publications

1. Peer-reviewed research articles

1.1. With impact factor (according to WEB OF SCIENCE), in English

1.1.1. International publisher

NOCITA, M., STEVENS, A., VAN WESEMAEL, B., AITKENHEAD, M., BACHMANN, M., BARTHÈS, B., BEN DOR, E., BROWN, D. J., CLAIROTTE, M., CSORBA, A., DARDENNE, P., DEMATTÊ, J. A. M., GENOT, V., GUERRERO, C., KNADEL, M., MONTANARELLA, L., NOON, C., RAMIREZ-LOPEZ, L., ROBERTSON, J., SAKAI, H., SORIANO-DISLA, J. M., SHEPHARD, K. D., STENBERG, B., TOWETT, E. K., VARGAS, R., WETTERLIND, J. (2015): Chapter Four - Soil Spectroscopy: An Alternative to Wet Chemistry for Soil Monitoring. *Advances in Agronomy*. 132, 139 – 159 pp. (Impact Factor: 5.02 (2014))

1.2. Without impact factor, in English

1.2.1. Hungarian publisher

CSORBA, Á., LÁNG, V., FENYVESI, L., MICHÉLI, E. (2013): Prediction of soil parameters based on reflectance spectroscopy and multivariate modelling methods. *Növénytermelés*. 62 (Suppl) 213 - 216 pp.

1.3. Without impact factor, in Hungarian

CSORBA, Á., LÁNG, V., FENYVESI, L., MICHÉLI, E. (2012): Reflektancia spektroszkópia alkalmazása talajok szerveszén- és CaCO₃-tartalmának becslésében. *Agrokémia és Talajtan*. 6/2, 277 - 290 pp.

5. Conference proceedings without ISBN, ISSN or other certification

5.1. Full text, peer reviewed in foreign language

CSORBA, A., LÁNG, V., FENYVESI, L., MICHÉLI, E. (2013). Prediction of soil pH, soil organic carbon and calcium carbonate content based on reflectance measurements. In: 3rd Global Workshop on Proximal Soil Sensing. Konferencia helye, ideje: Potsdam, Németország, 2013.05.26 Potsdam: 243 - 247 pp.

CSORBA, A., LANG, V., MICHELI, E., FENYVESI, L., (2012): Prediction of Soil Organic Carbon and Calcium Carbonate Content Based on Reflectance Spectroscopy and Multivariate Modelling Methods. *International Conference of Agricultural Engineering, CIRG-AgEng 2012, Valencia, Spanyolország, 2012.07.08-12, Papers Book*.

LÁNG, V., CSORBA, Á., MICHÉLI, E.,: Prediction of Selected Soil Parameters Based On Reflectance Measurements: a Case Study In a Hungarian Catchment Area, ASA, CSSA, SSSA Annual Meeting, San Antonio, 2011

8. References

- BARNES, R. J., DHANOA, M. S., LISTER, S. J. (1989): Standard normal variate transformation and detrending of near-infrared diffuse reflectance spectra. *Applied Spectroscopy*. 43, 772 – 777 pp.
- DUCKWORTH, J. H., 1998. Spectroscopic quantitative analysis. In: *Applied Spectroscopy* (Eds.: WORKMAN, J. & SPRINGSTEEN, A.) Academic Press. San Diego, California. 93–163 pp.
- GELADI, P., MACDOUGALL, D., MARTENS, H. (1985): Linearization and scatter-correction for near-infrared reflectance of meat. *Applied Spectroscopy*. 39, 491 – 500 pp.
- IUSS WORKING GROUP WRB (2015): World Reference Base for Soil Resources. World Soil Resources Reports No. 103. FAO. Rome.
- JASSÓ, F. (szerk.)(1989): *Melioráció-öntözés és talajvédelem. Útmutató a nagyméretarányú országos talajtérképezés végrehajtásához* '89 melléklet, Agroinformációs vállalat, Budapest.
- MICHÉLI, E. (2011): A talajképző folyamatok megjelenése a diagnosztikai szemléletű talajosztályozásban. *Agrokémia és Talajtan*, 60/1. 17-32 pp.
- MINASNY, B., MCBRATNEY, A. B., HARTEMINK, A. E. (2009): Global pedodiversity, taxonomic distance, and the World Reference Base. *Geoderma*. 155. 132 – 139 pp.
- O'ROURKE, S. M., HOLDEN, N. M. (2011): Optical sensing and chemometric analysis of soil organic carbon – a cost effective alternative to conventional laboratory methods? *Soil Use Manage*, 27, 143 – 155 pp.
- SAVITZKY, A., GOLAY, M. (1964): Smoothing and differentiation of data by simplified least squares procedures. *Analytical Chemistry*. 36, 1627 – 1639 pp.
- SCHWERTMANN, U. (1993): Relations between iron oxides, soil color and soil formation. In: *Soil Color, Special Publication*. (Eds.: BIGHAM, J. M. & CIOLKOSZ, E. J.) 51–70. Soil Science Society of America. Madison, Wisconsin.
- STEFANOVITS, P. (1972): *Talajtan*. Mezőgazda Kiadó, Budapest
- STEVENS, A., UDELHOVEN, T., DENIS, A., TYCHON, B., LIOY, R., HOFFMANN, L., WESMAEL, B. (2010): Measuring soil organic carbon in croplands at regional scale using airborne imaging spectroscopy. *Geoderma*, 158, 32 – 45 pp.
- VÅGEN, T-G., WINOWIECKI, L.A., TAMENE DESTA, L., TONDOH J.E. (2015): *The Land Degradation Surveillance Framework (LDSF) – Field Guide v 4.1*. World Agroforestry Centre, Nairobi, Kenya 14 pp.
- VISCARRA ROSSEL, R. A., CATTLE, S. R., ORTEGA, A., FOUAD, Y. (2009): In situ measurements of soil colour, mineral composition and clay content by vis-NIR spectroscopy. *Geoderma*, 150: 253-266 pp.
- 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 assesment of various soil properties. *Geoderma*, 131, 59 – 75 pp.
- VISCARRA ROSSEL, R.A., CHEN, C. (2011): Digitally mapping the information content of visible-near infrared spectra of surficial Australian soils. *Remote Sensing of Environment*. 115/6, 1443 – 1455 pp.

