DOCTORAL DISSERTATION THESES

VIKTÓRIA LABANCZ

GÖDÖLLŐ

2024

HUNGARIAN UNIVERSITY OF AGRICULTURE AND LIFE SCIENCES

THE EFFECT OF DIFFERENT SOIL PROPERTIES ON THE MACRO AND MICROAGGREGATE STABILITY OF SOILS

DOI: 10.54598/007100

Viktória Labancz

Gödöllő

2024

The name of the doctoral school: Doctoral School of Environmental Sciences

Field of study: Environmental Sciences

Head: **Dr. Erika Csákiné Michéli**

university professor, head of department,

Institute of Environmental Sciences

Hungarian University of Agriculture and Life Sciences

Supervisor: **Dr. Tamás András Szegi**

associate professor

Department of Soil Science

Institute of Environmental Sciences

Hungarian University of Agriculture and Life Sciences

Dr. András Szabolcs Makó

university professor

Department of Soil Physics and Water Management

HUN-REN Centre for Agricultural Research

1. Background and objectives of the work

Currently, there is no precise, objective, or universally applicable method for measuring and describing soil structure (HILLEL, 1998). However, it can be closely approximated with indicators that are suitable for describing structural conditions, such as aggregate stability. Aggregate stability provides information on how structural properties change over time due to various degradation effects (e.g., the mechanical impact of soil cultivation tools or the dispersion effect of rain or irrigation water) and indicates a soil's susceptibility to water or wind erosion (AMÉZKETA et al., 1999).

Comparing the literature reveals that different size aggregate fractions are stabilized by different mechanisms and binding agents (LE BISSONNAIS, 1996; AMÉZKETA 1999; TOTSCHE ET AL., 2018). Consequently, the testing procedures for macro- and microaggregate stability also differ. Among the methods for determining macroaggregate stability, procedures conducted in aqueous media, particularly various versions of wet sieving methods are widely used. The methods for examining the stability of microaggregates are fundamentally based on quantifying the clay and/or silt-sized particles of the soil. Their distribution is mostly determined using sieve-pipette methods or the faster laser diffraction method (LDM). LDM analysis, used for determining the partical size distribution (PSD) of soils and deriving dispersion indices for assessing microaggregate stability, has become a widely adopted testing method (MCCAVE et al., 1986; BLOTT & PYE 2006; RYZAK & BIEGANOWSKI 2011; YANG et al. 2015; FISHER et al. 2017), though standardization of the methods is still pending.

Nowadays, AI and machine learning methods offer several advantages, including the ability to uncover hidden correlations that traditional statistical methods cannot. Despite potential statistical errors (e.g., overfitting), these methods can be effectively used in developing pedotransfer functions (PTF) that estimate water retention and conductivity. Research results have shown that the estimation efficiency of PTF increases when structural indicators of the soil are included as input data (PACHEPSKY et al. 2001; RAWLS & PACHEPSKY 2002; MAKÓ et al. 2019). However, these soil structural characteristics are only included in a few specialized soil physical databases; in most cases, structural state can only be inferred from soil type or other measured soil properties. This raises the question: With the creation of a Hungarian soil structure database that includes information on soil structure in addition to many easily measurable soil parameters, can reliable PTFs be developed to characterize soil structural properties? And in a later step, can these estimated structural characteristics be used to refine PTFs for estimating hydro-physical soil properties?

To estimate aggregate stability indicators (and later incorporate this structural information into hydro-physical PTFs), we need to understand which abiotic and biotic factors (soil properties, external, and possibly indirect effects) most strongly determine the stability of structural elements within a given size category. The impact of these factors is often difficult to distinguish from each other; however, using machine learning methods—similar to hydro-physical PTFs—the order, strength, and direction of influencing factors can be well studied.

During the research, I sought to answer the following questions

LDM methodology development (preliminary experiment)

- How does the quality of the aqueous medium forming the soil suspension influence the particle size distribution measurement results obtained using the LDM, during the individual or combined application of different dispersion procedures?

Aggregate Stability Studies

- How applicable is the Random Forest analysis, a machine learning method, for estimating the macro- and microaggregate stability of soils, as tested on samples from the Hun-SSD, which represents Hungarian soil databases (AIIR, MARTHA)?

- Based on the ***Random Forest analysis***, which combinations of soil properties most significantly determine the ***macroaggregate stability*** of ***Hungarian soils***?

- Based on the ***Random Forest analysis***, which combinations of soil properties most significantly determine the ***microaggregate stability*** of ***Hungarian soils?***

- Considering the inaccuracies of the Random Forest predictive models and the review of existing literature, which (soil and environmental) properties should be considered for future inclusion in the study of influencing factors?

2. Material and method

The investigations involved two types of soil databases. The samples used for the preliminary experiments were provided by the soil sampling of the research TÁMOP- 4.2.1/B-09/ 1/ KONV-2010-0003 “KESZTHELY” database. The basis of the aggregate stability investigations is a new Hungarian Soil Structure Database, the “HunSSD” which contains data from the examined samples related to the NKFIH K119475.

2.1. Characterization of the soils included in the LDM methodology development

Table 1. Important physical and chemical characteristics of the soils included in the study

Sample code	**WRB soil classification** ***& name of the closest city***	**Symbol and depth of genetic horizons (cm)**	**Clay + Fe-oxihydrates (%) (<0.002 mm)**	**Silt (%) (0.002-0.05 mm)**	**Sand (%) (>0.05 mm)**	**SOM (%)**	**$CaCO_3$ (%)**	**CEC (mgEq/100 g)**	**Exchangeable Na^+ (mgEq/100 g)**	**Stability of macroaggregates (MaAS) (%)**
			SPM PSD* (solid part of the soil (100 %))							
M1	Vertic Stagnic Solonetz (Clayic) Karcag	**B** 5-30	51.09	45.90	0.88	2.00	0.13	40.85	20.63	20.84
M2	Hortic Terric Cambisol (Dystric, Siltic) Keszthely	**A** 0-30	21.09	33.13	44.28	1.45	0.05	11.84	0.14	53.40
M3	Hortic Terric Cambisol (Dystric, Siltic) Keszthely	**B** 30-50	22.90	33.87	42.29	0.93	0.00	12.38	0.13	38.47
M4	Cutanic Luvisol (Siltic) Várvölgy	**A** 0-20	15.27	29.35	54.05	1.33	0.00	10.36	0.12	87.57
M5	Cutanic Luvisol (Siltic) Várvölgy	**B** 20-50	22.30	26.56	50.49	0.65	0.00	12.78	0.15	38.38
M6	Vertic Gleyic Luvisol (Manganiferric, Siltic) Magyarszombatfa	**B** 20-50	38.96	25.93	34.61	0.49	0.00	16.78	0.17	44.41
M7	Vermic Calcic Chernozem (Anthric, Siltic) Kápolnásnyék	**A** 0-30	27.60	51.68	7.50	3.70	9.52	30.25	0.25	64.56
M8	Gleyic Vertisol (Clayic) Kisújszállás	**A** 0-30	53.88	41.19	1.05	3.89	0.00	35.69	0.29	59.14

*SPM PSD: Particle size distribution measured according conventional standardized sieve-pipette method (*ISO 11277:2009(E)*)

Eight soil samples from the "KESZTHELY" database (*Table 1.)* were included in the laser diffractometric preliminary experiment. The samples differed in their physical, chemical and mineralogical properties and were representative of the genetic horizon of the main Hungarian soil types. The basic tests (pH (H_2O), $CaCO_3$%, humus%) and CEC tests of the air-dried soil samples sifted through a 2 mm sieve were carried out according to the Hungarian soil testing methodology based on Buzás (1988) (Barna et al., 2015). The PSD was determined according to the international (ISO 11277:2009(E)) standard (with complete destruction of the aggregates). The macroaggregate stability measurements (MaAS (%)) were performed in an Eijkelkamp wet sieving apparatus based on the Kemper & Koch (1966) method.

2.2. Characterization of the soils involved in aggregate stability studies

The samples included in the tests came from 55 soil profiles (*Table 2.*) (Hungarian Soil Structure Database: HunSSD). When selecting the sampling locations, the main consideration was that the main soil types listed in the larger soil databases (AIIR, MARTHA) should be explored in a number corresponding to their proportion (within the planned 50-60 profile excavations). The basic tests of the soil samples were carried out according to the Hungarian standard soil testing methodology (MSZ-08.0206-2-78; BUZÁS, 1988; 1993), the results of the basic tests are summarized in *Table 2.*

Table 2. The main characteristics of the soils of the HunSSD database

	Mean	**Standard Dev.**	**Minimum**	**Maximum**
pH (H_2O)	7,52	1,19	4,15	9,88
$CaCO_3$ (%)	7,51	11,16	0,00	53,10
EC (µS)	326,35	502,88	22,00	5160,00
Humus (%)	2,06	2,25	0,10	21,80
Results of organic matter tests				
Hargitai Q	20,53	33,51	0,13	177,25
Hargitai K	7,57	11,46	0,05	64,39
C/N	9,12	1,80	5,26	20,06
Results of base exchange tests				
Exchangeable Ca (mmol*$100g^{-1}$)	19,81	10,50	0,50	57,50
Exchangeable Mg (mmol*$100g^{-1}$)	4,19	2,56	0,05	12,85
Exchangeable Na (mmol*$100g^{-1}$)	0,79	2,03	0,05	11,60

Exchangeable K ($mmol*100g^{-1}$)	0,55	0,56	0,05	2,62
Base cations ($mmol*100g^{-1}$)	25,31	11,71	0,65	66,51
CEC ($mmol*100g^{-1}$)	27,96	10,94	3,08	73,10
Acid cations ($mmol*100g^{-1}$)	2,73	2,54	0,00	17,34
Na (S%)	5,77	16,08	0,11	86,50
Ca (S%)	74,06	17,77	12,90	96,87
Mg (S%	18,42	11,88	0,33	82,80
K (S%)	2,56	2,83	0,14	15,70
Results of particle size distribution				
clay (%) (LDM)	30,30	8,68	3,40	47,89
silt (%) (LDM)	49,71	6,36	19,36	61,61
sand % (LDM)	20,00	11,30	3,92	71,62

The pH and conductivity (EC) were determined from a 1:2.5 soil:water suspension. The $CaCO_3$ content was determined using the Scheibler method (diluted hydrochloric acid method), and the humus content (H%) was determined using the Tyurin method (chromic acid oxidation). CEC tests were also performed. The amount of exchangeable calcium, magnesium, sodium and potassium ions was measured by ICP. The T-value was determined using a modified Mehlich method, $BaCl_2$. Humus quality tests were performed based on HARGITAI method (1988). The C/N ratio of the samples was also determined. The PSD of the soils was determined using with LDM (BIEGANOWSKI et al., 2018).

2.3. Methods

2.3.1. Test methods of PSD measurements

The sieve-pipette PSD measurement method according to the ISO 11277:2009 standard

In the case of the ISO 11277:2009 (ISO) standard, after the removal of elements larger than 2 mm, the next step was the removal of humic substances, carbonates and iron (oxy)hydroxides.

Humic substances were removed with a 30% hydrogen peroxide solution, carbonates were removed with 0.1 M hydrochloric acid. For the removal of iron (oxy)hydroxides, Na-dithionite was used in a Na-citrate-Na-bicarbonate buffer medium. For comparability, we performed removals on the entire sample material according to the proposed methodology. The amount of reagents used in each treatment was adjusted to the humus, carbonate and iron (oxy)hydroxide content in accordance with the ISO standard. According to the ISO method, the entire sand fraction above 0.05 mm was separated by sieving. To determine the pipetting depth, we used the average density of soils (2.65 g cm^{3}-). The clay, silt and sand content of the examined soils was expressed as a percentage of the dispersed soil.

The LDM PSD test method used for the experiments

The LDM PSD tests were performed with a Malvern Mastersizer 3000 device. We used the same measurement method during both the pre-experiments and the analysis of the structural database. An automatic wet preparation unit of the Hydro LV type was connected to the device. The stirring speed was 2750 rpm (BIEGANOWSKI 2018). The power of the ultrasound is 40W maximum (frequency: 40 kHz (nominal)), which operated at 100% power; the ultrasound was performed for 240 seconds.

During the measurements, the obscuration ranged from 5 to 20 % in accordance with the setting recommendations. In the case of measurements, 2 cm^3 of Calgon solution (a mixture of 33 g of Na-hexametaphosphate and 7 g of anhydrous Na-carbonate L-1 prepared according to the ISO 11277:2009(E) standard) was added to the tested air-dry samples as a dispersant, and then was mixed it gently on the watch glass with a glass rod. Then we added another 25 cm^3 of Calgon solution to the preparation unit. Each measurement took place for 5 minutes with the recording of 5 records, with a minimum of 3 repetitions from new measurements. The laser light intensity data were converted to PSD results based on the Mie theory, with the following settings: AI = 0.1; RI_{soil} = 1.52 and RI_{water} = 1.33. The size limit of the clay (<7 µm), silt (7-50 µm), and sand fraction (>50 µm) was determined based on the results of previous research on Hungarian soils (MAKÓ et al., 2019).

Preliminary experiments to study the effect of different aqueous media

In connection with the LDM preliminary experiments, we examined eight soil samples of the KESZTHELY database with different properties and studied the effect of aqueous media and dispersing methods based on the comparison of the measured particle size distribution curves, the calculated clay content, and the texture categories determined from the PSD (USDA 12 category triangle diagram).

The aqueous measurement media loaded into the Mastersizer 3000 Hydro LV preparation unit were: distilled water (DW), high-purity, commercially available deionized water (DIW), tap water (TV) (*Table 3.*).

Table 3. Properties of the aqueous environments used during LDM measurement

Aqueous medium	**pH**	**EC (µS/cm)**	**Na^+ ppm**	**K+ ppm**	**Ca^{2+} ppm**	**Mg^{2+} ppm**	**Total hardness mg/l CaO**
DW	6.4	2.2	0	0	0	0	0
DIW	6.1	1.5	0	0	0	0	0
TW	7.7	530	26	2,6	67	17	136

The LDM PSD measurements were performed in different combinations (according to a kind of matrix) MAKÓ, et al. (2017) and POLAKOWSKI, et al. (2021), including the three most common main variables: type of aqueous medium, physical and/or chemical dispersion (as pretreatment). We worked with four types of treatment: no treatment T1; only Calgon T2, only ultrasound T3, combined treatment T4. The soil samples were prepared in a traditional way: air-dried, ground, sifted through a 2 mm sieve, free of macroscopic plant remains.
From the examination of the fit of the particle distribution curves, it is possible to deduce the degree of dispersion between repetitions, the formation of chemical precipitates, possible artifacts, and bubbles. For this I used the own analysis software of the Malvern Mastersizer 3000 device (v4.10).
To compare the PSD results, we performed a GLM univariate analysis (UNIANOVA) to test the combined effect of the investigated factors (aqueous media, treatments, soil variables) on the clay, silt and sand content measured by LDM. One-Way ANOVA (Duncan test or Tamhane test depending on the homogeneity of variance) and Boxplot analysis (IBM SPSS Statistics 20 software) were performed to compare the clay, silt, and sand content of the data groups.
By displaying the PSD results determined in different aqueous media on the USDA 12 category texture triangle diagrams, I compared the texture classifications with each other, as well as with the PSD results determined by the pipette method according to the ISO 11277:2009 standard. Triangular diagrams were created using OriginPro Version 2021. OriginLab statistical software.

2.3.2. Methods of aggregate stability tests

Determination of macroaggregate stability using a wet sieving device

For the wet sieving tests, we used an Eijkelkamp device (Wet sieving method set, product code: 08.13). The device contains 8 sieves with a sieve diameter of 250 µm. The test was carried out with air-dried, ground, plant residue-free soil samples passed through a 2 mm sieve. This was followed by sieving the sieved samples to a size between 1-2 mm based on the method of Kemper & Rosenau (1986). In each case, 4 g of sample was measured on the sieves. In each case, the measurements were performed in eight repetitions. Contrary to the recommendation of the original method, the pre-moistening was carried out slowly by capillary means (Barna et al., 2017). The sieves containing the soil samples were placed in the device, immersed in the vessels of the device containing distilled water, and after switching on, the device lifted the sieves up and down in the liquid 34 times per minute, with a lifting height of 1.3 cm. After the 3 minutes, we replaced the dishes under the samples. A Na-pyrophosphate dispersing solution (concentration: 2 g L^{-1}) was poured into the new vessels, the sieves containing the samples were immersed in it, and the silting of the samples was continued as described above, this time for 8 minutes.

Both the distilled water and the contents of the collection vessel containing the dispersing solution were washed into a beaker without residue, then evaporated on an electric hotplate, dried in an oven at 105 °C, and then their mass was measured. The proportion of stable macroaggregates (MaAS (%)) of the soil samples was calculated based on the following equation:

$$MaAs\ (\%) = \frac{fd}{fnd + fd} * 100$$

where: *fnd* is the vessel containing distilled water (aggregates not stable in water), and *fd* is the evaporation residue of the vessel containing the dispersing solution (aggregates stable in water).

Determination of microaggregate stability values with LDM device

The microaggregate stability (MiAS (%)) of the soils was calculated based on the Vageler structure factor (Vageler, 1932), known from the literature and based on pipette measurements, from the ratio of clay fractions determined with dispersed and non-dispersed laser diffractometry (Horel et al. 2019).

$$MiAS\ (\%) = \frac{cd - cnd}{cd} * 100$$

where: *cd* is the dispersed clay fraction, *cnd* is the non-dispersed clay fraction.

The measurements required to determine the MiAS (%) were performed (similarly to the PSD tests) with a Mastersizer 3000, Hydro LV LDM device. The measurements required to determine the MiAS (%) were performed (similarly to the PSD tests) with a Mastersizer 3000, Hydro LV LDM device. We filled the preparation unit with deionized water (pH: 7.9; EC (µS/cm): 564). In accordance with the LDM PSD tests, converted the data coming into the detector into PSD results using the Mie theory. The PSD of the dispersed fraction was determined using the previously described particle size analysis determination method (Bieganowski et al., 2018) on the Mastersizer 3000, Hydro LV laser diffractometry device, with the addition of ultrasound and

Calgon solution dispersant. The PSD of the non-dispersed fraction was determined without the use of ultrasound or the addition of Calgon solution. During the PSD analysis of the non-dispersed fraction, the duration of the measurement was 30 minutes per repetition, during which time 30 records (particle size distribution curve) were recorded. For the calculation, due to better comparability with the dispersed PSD curves, I used the data measured at the time closest to 240 seconds from the PSD curves obtained without dispersion.

Phases of developing Random Forest models estimating aggregate stability

The various soil properties that potentially influence aggregate stability and the macro (MaAS (%)) and micro aggregate stability (MiAS (%)) results were sorted into a database after sorting, filtering and sorting the data (using Excel and SPSS software).

Correlation matrix analysis

The development of the estimating models was preceded by an analysis based on Pearson's correlation. This revealed the relationship between pairs of variables - between the MaAS(%) and MiAS(%) results of the HUN-SSD database and variables describing different soil properties. The value of the linear correlation coefficient was calculated between all variables, which I arranged in a matrix and displayed as a correlation matrix. At the same time, an examination of the significance of the correlation coefficients was also carried out. The strength of the relationship between two variables was assessed in a simplified manner according to Guilford's (1950) method.

Calculation of Random Forest models

To create the estimation models, I used the Random Forest statistical method (Wright & Ziegler, 2017) using R software (R Core Team, 2021). During the analysis, with the help of the algorithm, I established the relationship between the measured (real) and estimated MaAS (%) and MiAS (%) results, depending on which soil properties (independent variables) are used as input data for the calculation of the model. In the same way, it was determined which soil properties (independent variables) are decisive for the development of stability in the case of a given estimating model.

Performed the Random Forest tests on the attribute sets of the database consisting of soil samples from 258 sampled genetic levels of 55 excavated soil profiles representing domestic soil properties. When fitting each model, 500 decision trees were created, and the values estimated in terms of the reliability of the models were formed from the averages of these decision trees. Justified

cases during the creation of the individual models, also performed a recursive feature selection, prior to which I normalized the non-normally distributed data series (Na-(S%), EC, CaCO3 %, humus %, Q and K values of Hargitai humus quality) (ln log transformation). This selection method improved the estimation accuracy of the models in many cases.
Compared the statistical accuracy of the models estimating aggregate stability based on the square of the coefficient of determination (R^2) and the root mean square error (RMSE).

<u>Building Random Forest models</u>

The background database of Random Forest estimation models for both macro-aggregate stability (MaAS (%)) and micro-aggregate stability (MiAS (%)) can basically be divided into two groups (examples in *Tables 4-5*). The models 1 are based on a larger database (MaAS (%) - 177N; MiAS (%) - N224) that does not contain humus quality measurement data and is built up from tested soil samples from different genetic levels. The basis of the models 2 is a smaller database narrowed down from the database of model 1, but also containing humus quality data (MaAS (%) - N118; MiAS (%) - N133).

The structure of the models was expanded hierarchically. In the case of model groups 1 and 2, the data type expansion order is the same. During the development of the estimation models, the main aspect was to study how the accuracy of the estimation changes with the expansion of the range of soil test data, as well as which properties most influence the stability of different types of soil aggregates (MaAs(%), MiAS(%)).

Table 4. Random Forest models estimating macroaggregate stability from a database without humus quality variables

Aggregate stability	**Model**	**Database**	**Input data**	**N**
MaAS (%)	1/A	without humus quality variables	basic test data	177
MaAS (%)	$1/B_1$	without humus quality variables	basic test data+ topsoil/subsoil	177
MaAS (%)	$1/B_2$	without humus quality variables	basic test data + soil mid-depth	177
MaAS (%)	$1/B_3$	without humus quality variables	basic test data + land use	177
MaAS (%)	$1/B_4$	without humus quality variables	basic test data+topsoil/subsoil+ soil mid-depth + land use	177
MaAS (%)	$1/C_1$	without humus quality variables	basic test data+ base exchange	177
MaAS (%)	$1/C_2$	without humus quality variables	basic test data+ base exchange + topsoil/subsoil	177

Aggregate stability	Model	Database	Input data	N
MaAS (%)	$1/C_3$	without humus quality variables	basic test data+ base exchange + soil mid-depth	177
MaAS (%)	$1/C_4$	without humus quality variables	basic test data+ base exchange + land use	177
MaAS (%)	$1/C_{4f}$	without humus quality variables	basic test data+ base exchange + land use +feature selection	177
MaAS (%)	$1/C_5$	without humus quality variables	basic test data+ base exchange + topsoil/ subsoil+ soil mid-depth + land use t	177

Table 5. Random Forest models estimating microaggregate stability from a database containing humus quality variables

Aggregate stability	Model	Database	Input data	N
MaAS (%)	2/A	with humus quality variables	basic test data	118
MaAS (%)	2/B1	with humus quality variables	basic test data+ topsoil/subsoil	118
MaAS (%)	2/B2	with humus quality variables	basic test data +	118
MaAS (%)	2/B3	with humus quality variables	basic test data + land use	118
MaAS (%)	2/B4	with humus quality variables	basic test data+topsoil/subsoil+ center of soil profile+ land use	118
MaAS (%)	2/C1	with humus quality variables	basic test data+ base exchange	118
MaAS (%)	2/C2	with humus quality variables	basic test data+ base exchange + topsoil/subsoil	118
MaAS (%)	2/C3	with humus quality variables	basic test data+ base exchange + soil mid-depth	118
MaAS (%)	2/C4	with humus quality variables	basic test data+ base exchange + land use	118
MaAS (%)	2/C5	with humus quality variables	basic test data+ base exchange + land use +feature selection	118
MaAS (%)	2/D1	with humus quality variables	basic test data + humus quality	118
MaAS (%)	2/D2	with humus quality variables	basic test data + humus quality+ base exchange	118
MaAS (%)	2/D3	with humus quality variables	basic test data + humus quality+ base exchange + topsoil/subsoil	118
MaAS (%)	2/D4	with humus quality variables	basic test data + humus quality+ base exchange + soil mid-depth	118
MaAS (%)	2/D5	with humus quality variables	basic test data + humus quality+ base exchange + land use	118
MaAS (%)	2/D6	with humus quality variables	basic test data + humus quality+ base exchange + topsoil/subsoil+ soil mid-depth + land use	118

3. Results and discussion

3.1. Test results of preliminary experiments related to the measurement of PSD

The results of the effect of different aqueous media

The results show that the use of different waters has an effect on the measurement results, even though the information sheets of the Malvern Mastersizer, Fritsch, Beckman Coulter devices and the numerous publications do not mention the choice of the type and quality of the aqueous medium and its importance (e.g. Bieganowski et al., 2010).
Examining the particle distribution curves of the eight soil samples of the measured KESZTHELY database, it can generally be said that the parallel repetitions of the PSD measurements showed the highest standard deviation in TW medium. In the case of tap water measurements, we experienced the appearance of secondary peaks clearly caused by an artifact (probably calcium phosphate) in certain samples (*Fig.1.)*

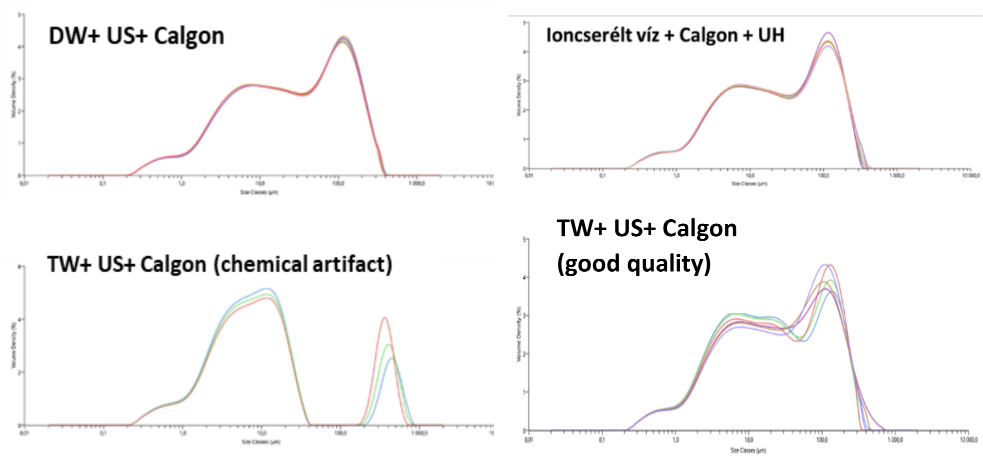

Figure 1. PSD examination of sample M2, a Cambisol soil sample, in distilled water, deionized water and tap water - with the combined use of Calgon solution and ultrasound

The GLM Univariate Analysis (UNIANOVA) confirmed the significant effect of the factors (pretreatments, aqueous medium quality, soil variables) and their combinations ($P<0.001$ in all cases) on the measured clay, silt and sand content. Comparing the amount of individual particle fractions (the results measured on all soil samples considered as a common group) per pretreatment and per liquid, it can be said that the comparisons made with One-way ANOVA tests showed that for all three aqueous media, T1 resulted in the significantly lowest clay content in all soil samples. (in our experience, the releasable clay content mostly indicates the success of the dispersion).

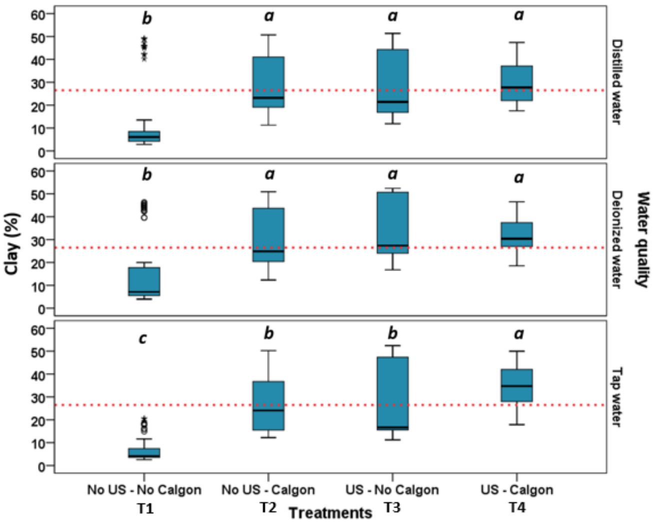

Figure 2. Changes in LDM clay content in different treatments and aqueous media (all soil samples considered together). Means denoted by the same letter did not significantly differ at $p < 0.05$ (One-way ANOVA). The dashed red lines indicate the median values of SPM PSD results of the samples.

In addition to DW and DIW used, the clay content did not differ significantly between T2 (use of Calgon solution only) and T4 (combined treatment), while in the case of TW, T4 significantly resulted in the highest clay content *(Figure 2*).

Based on the results presented on the 12-category USDA texture triangle diagrams (Figures 3-4), it can be concluded that the differences in the choice of the aqueous medium and the pretreatments resulted in differences in the texture classification of the soil samples. In the case of T1 (without the use of ultrasound and Calgon solution), the LDM PSD measurements showed a predominance of the coarser particle fractions when using all aqueous media (DW- distilled water, DIW- deionized water, TW- tap water). (*Figure 3*)

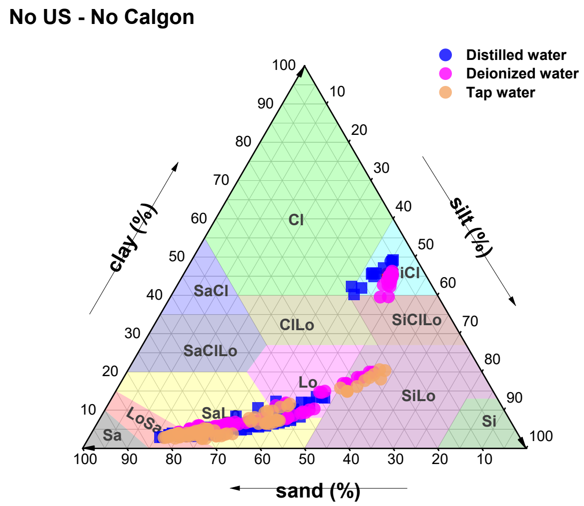

Figure 3. Texture classification of PSD results of T1 treatment in case of LDM measurement

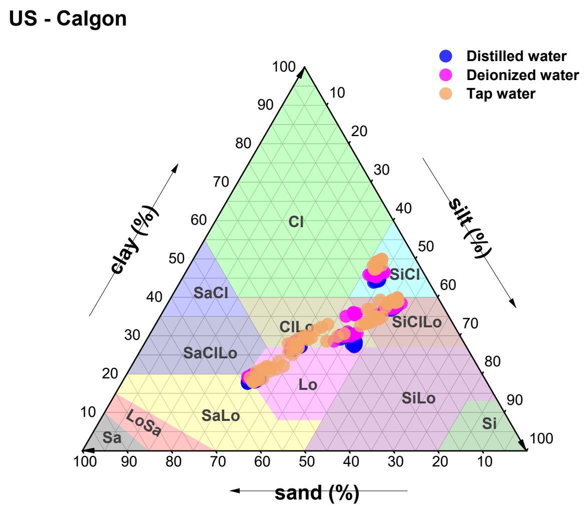

Figure 4. Texture classification of PSD results of T4 treatment in case of LDM measurement

The triangle diagrams representing the results of K4 (the combined application of ultrasound + Calgon solution) show the strongest dispersing effect (the measured clay content is usually the highest here). The scatter diagram of the obtained soil texture classifications is the most consistent in this case (*Figure 4*).

3.2. Macroaggregate stability results (HunSSD database)

The MaAS (%) database excluding humus quality variables

Based on the correlation matrix analysis, it can be said that there is a significant relationship between MaAS (%) and other soil properties, a medium correlation can only be seen in the case of humus content (r= 0.503***), its direction is positive. In the case of the investigated soils, macroaggregate stability increased parallel to the increase in humus content. A significant, sure but weak relationship can be seen with dust content (r= -0.212***), Na (S%) (r= -0.284***), Mg (S%) (r= -0.337***), pH (H2O) (r= -0.237***) and MaAs (%). In all cases, their direction was negative, so as the value of the variables increased, the value of the macroaggregate stability also decreased. A weak, negligible relationship was observed between the other variables included in the analysis (soil properties) and the macroaggregate stability.

Based on the correlation analysis, it appeared that the distribution of certain soil variables (Na (S%); EC (µS/cm); CaCO3 (%); humus (%)) differed from the normal distribution, therefore, prior to the Random Forest estimates, these variables were logarithmically transformed normalized.

The type 1 MaAS (%) estimation models (N= 177), which do not include humus quality variables, and were established from a larger database, show that, without examining the role of humus quality, which soil properties most influence the stability of the macroaggregate of soils.

The estimation accuracy of the majority of the type 1 MaAS% models proved to be very poor (R2 <0.5). Only 4 models showed acceptable estimation accuracy: 1/B3, 1/C4, 1/C4f and 1/C5 (*Figure 5.*). The same models also had the smallest RMSE value (*Figure 6.*). In the case of all four, it can be said that land use information was included as common input data, while none of the models with lower estimation accuracy used land use data.

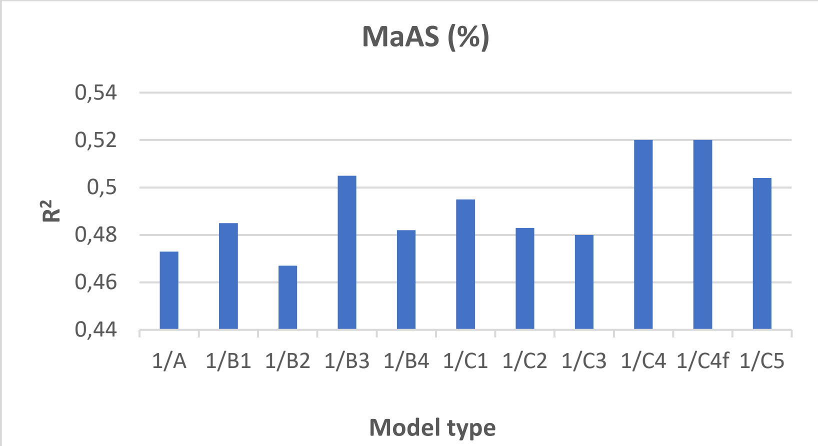

Figure 5. Statistical accuracy (R^2) of type 1 Random Forest models that estimate macroaggregate stability without humus quality variables

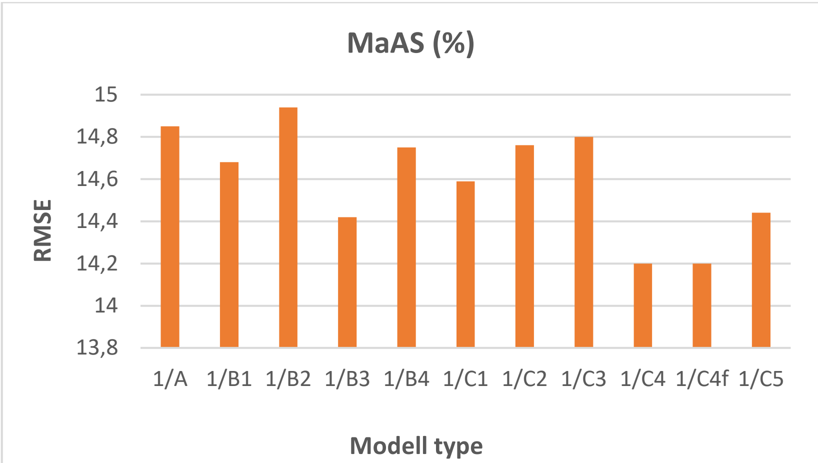

Figure 6. Error (RMSE) of type 1 models estimating macroaggregate stability without humus quality variables

MaAS (%) database including humus quality variables

The Random Forest models for estimating macroaggregate stability of the type 2 model group were created from a smaller database containing humus quality variables (N=118), so only the test results of those samples were included in the database that forms the basis of the 2 models, where the humus content exceeded 0.5%, as only these samples were tested for humus quality.
The correlation matrix analysis did not show a moderate or stronger relationship in any case. However, he revealed a certain but weak significant (***) connection in several cases. For silt (%) (r= -0.210), Mg (S%) (r= -0.239) and pH (H_2O) (r=-0.265), the relationship showed a negative direction, while humus

(%) (r = +0.396) and C/N (r= +0.216) soil variables, we established a positive relationship. No correlation was found with regard to other variables.

Based on the evaluation of the models estimating type 2 macroaggregate stability, it can be said that the coefficient of determination (R^2) of none of the models reached the value of 0.5, so the accuracy of the estimates cannot be considered acceptable. The RMSE values decreased in parallel with the increase in the R2 values. It can be concluded that, similarly to the type 1 models, the pH effect and the humus content had a decisive role in the formation of the macroaggregate stability in the case of the 2 models as well. The Random Forest model type 2 (2/D5f) MaAS (%), which proved to be the best, included as input: base test + humus quality + base exchange tests + land use data. In order to improve the performance indicators of the model, recursive feature selection was performed. The overall order of importance of the soil properties forming macroaggregate stability for the 2/D5f model was as follows: pH (H2O) > Hargitay K-value > humus (%) > land use > Hargitay Q-value > Mg (S%) > clay (%) > Na (S%) > EC (µS/cm) > Ca (S%).

3.3. Microaggregate stability results (HunSSD database)

MiAS (%) database excluding humus quality variables

Correlation matrix analysis between microaggregate stability and soil properties that did not include humus quality variables showed a significant moderate correlation or significant (***) relationship in three cases. These were: humus content (r=0.454), Ca (S%) (r=0.405), and a negative correlation was observed for Na (S%) (r=-0.629). In addition to the three variables, a certain, significant relationship was detected in only one case, in relation to Mg (S%), the direction of which was negative (r=-0.25). Based on the linear correlation analysis, the relationship between MiAS (%) and other soil variables was negligible.
The RMSE value) of type 1 models was related to the accuracy of the R^2(*Fig. 7-8.*). The estimation accuracy of the models can be categorized as acceptable (R^2=0.75-0.5) in all cases.

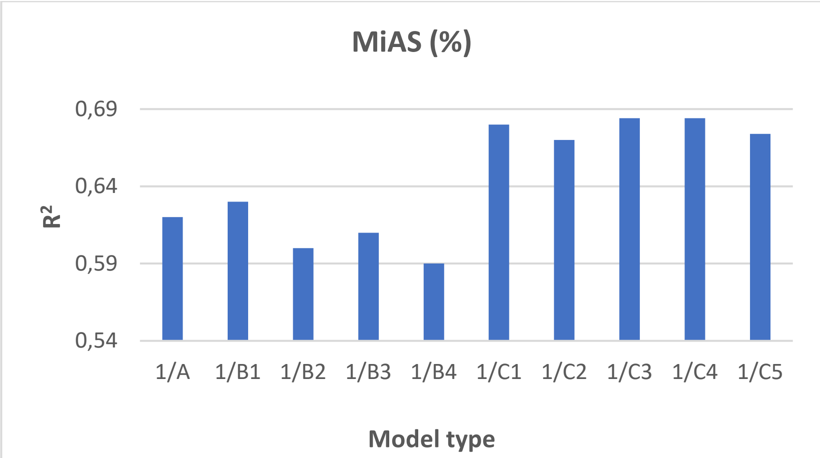

Figure 7. Statistical accuracy (R^2) of type 1 Random Forest models that estimate microaggregate stability without humus quality variables

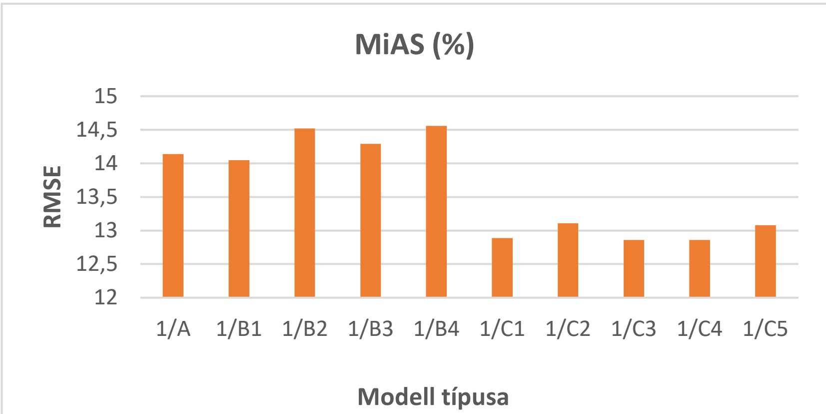

Figure 8. Error (RMSE) of type 1 models estimating microaggregate stability without humus quality variables

MiAS (%) database including humus quality variables

The linear correlation analysis revealed a significant (***) medium correlation in two cases, i.e. a significant relationship; for Na (S%) it was negative (r=-0.620), for Ca (S%) it was positive (r=0.472). The correlation matrix analysis revealed a significant (***) weak but reliable relationship in four additional cases. The humus content (r=0.240), Hargitai's (Q) value (r=0.289) and Hargitai's (K) value (r=0.231) are positive, while Mg (S%) (r =-0.313) showed a negative relationship. The linear correlation analysis did not reveal a relationship between MiAS (%) and other soil variables.
In the Random Forest analysis, among the 2/D model types, the 2/D5 model proved to be the most favorable estimation model. Therefore, examined its potential for further improvement using recursive feature selection; this model

was designated as 2/D5f. The algorithm did not recommend retaining land use information as input data, as the analysis indicated that it is not significant for the development of microaggregate stability.

4. Conclusions and recommendations

The preliminary experiments with LDM

A large percentage of literature reports on LDM measurements used to determine the PSD of soils report the use of distilled water or deionized water. The manufacturers of laser diffractometry devices make various recommendations regarding the selection of the aqueous medium. No manufacturer makes a separate recommendation for the use of water for testing soil samples. Based on the examination of the particle distribution curves, clay, silt, and sand contents, it can be said that the PSD determination procedure described in the general methodological chapter (Calgon solution + ultrasound) resulted in significant deviations when applied with tap water (TW), compared to distilled (DW) -, or using deionized water (DIW).

If we examined the effect of different treatments (T1, T2, T3, T4) carried out in each aqueous medium (DW, DIW, TW), it is difficult to explain the different dispersion order obtained in the case of soil samples with different properties. Presumably, the equilibrium states of the pairs of processes acting in parallel (dissolution-precipitation, adsorption-desorption, dispersion-flocculation, aggregation-deaggregation) determined the amount of clay released from the aggregates, the dust and sand content in the case of a chosen aqueous medium and treatment. There are only a small number of publications on the combined effects of the processes involved, rather on the effects of the individual sub-processes separately.

Based on the experience of preliminary tests, it can be said that the chemical properties of the aqueous medium (ion content and soluble salt content) can influence the degree of disaggregation of soil samples, dispersion and flocculation, ion exchange processes and the formation of artificial products. The optimal aqueous medium for more complete disaggregation may vary depending on the properties of the soil and the dispersion methods used.

Random Forest models for estimating aggregate stability

In the case of macroaggregate stability, a moderate correlation was only observed for the humus content, the direction of which was positive. This linear relationship is supported by numerous Hungarian and international publications (e.g. BALLENEGGER, 1933; TISDALL & OADES, 1982; CHENU, 2000)

Based on the correlation analyses, the microaggregate stability of the soils of the HunSSD database depended most significantly on Na (S%), but the direction of the relationship was negative, so with a high sodium content, low microaggregate stability was experienced. This is the same as described in all literature on the subject (TOTSCHE et al. 2018).

During the MaAS (%) and MiAS (%) estimations, the Random Forest analysis algorithm considered different soil properties important in a different order. By summarizing these and scaling them according to their importance (*Table 6.*) (based on their place in the order of importance of the analysis models and the frequency of the positions), it can be determined which of the examined soil properties are more likely to determine the macro- or micro-aggregate stability of the Hungarian soils of the HunSSD database. In addition, the table displays the relationship of individual soil properties (exclusively examined in the HunSSD database) with the stability of macro- and micro-aggregates (AMÉZKETA, 1999; TOTSCHE et al., 2018). as a summary of the information found in Hungarian and international publications.

Table 6. The relationship between macro- and micro-aggregate stability and different soil properties based on the summation of literature reports and in the case of soils from the HUNSSD database

Soil Properties (Variables)		**Macroaggregate Stability (>250µm)**	**HunSSD MaAS% (>250µm)**	**Microaggregate Stability (<250 µm)**	**HunSSD MiAS% (<250 µm)**
Electrolyte concentration and composition	pH (H_2O)	✓	✓✓✓	✓✓	✓✓✓
	EC (µS/cm)	×	✓✓	✓✓	✓✓
Texture	clay (%)	(✓)	✓✓	✓✓	✓✓
	silt (%)	×	✓✓	×	✓
	sand (%)	×	×	×	×
Calcium Carbonate	$CaCO_3$ (%)	(✓)	✓	✓✓	✓✓✓
Base Exchange Properties	Na (S%)	✓	✓	✓✓	✓✓✓
	Ca (S%)	✓✓	✓✓	✓	✓✓✓
	Mg (S%)	✓	✓✓✓	✓✓	✓✓
	K (S%)	✓	×	✓	✓
	T-S Value (mmol/100 g; acidic cations)	(×)	✓	(×)	✓
Organic matter	Humus (%)	✓✓✓	✓✓✓	✓✓	✓✓
	Hargitai Q		✓✓		✓✓✓
	Hargitai K		✓✓✓		✓
	C/N ratio	✓	✓	✓	×
Profile Position	Topsoil/subsoil	✓✓	✓	✓	✓
	Mid-depth (cm)		✓✓		✓✓
Land Use (arable, forest, grassland, orchard)			✓✓✓	(✓)	×

✓✓✓: strong effect; ✓✓: medium effect; ✓: weak effect; ×: no effect: ()= possible effect

The examination of other soil properties (e.g., iron content determination, clay mineral composition analysis, etc.) would incur additional costs and complicate the future application of PTFs in estimating soil macro- or microaggregate stability. However, numerous indirectly determinable soil properties and information can be found in the records of soil surveys conducted using either Hungarian genetic or international diagnostic soil classification systems (e.g., WRB). These valuable data, with appropriate logic, could potentially be indirectly incorporated into structure stability estimations.

New scientific results

Theses of laser diffractometric methodological preliminary experiments:

1. Depending on the ***physical, chemical and mineralogical properties*** of the soils, the ***different dispersion methods gave different LDM PSD results*** with the choice of ***different aqueous media***. In connection with the standardization efforts of future LDM measurements, attention should be drawn to the need to find a solution to more precisely determine the grain size distribution of soil samples with diverse properties (if possible with complete disaggregation, complete dispersion of elementary particles, elimination of artificial product formation) of the grain size distribution of soil samples in order to define it more precisely.
2. Since neither the manufacturers of the laser diffractometers nor the relevant literature deals with the quality of the aqueous medium of the soil suspension entering the measuring cells, no recommendations are made for its choice when measuring the ***LDM PSD of soils***. With methodological preliminary experiments of various soil samples, I proved that the ***use of tap water is not recommended*** in the case of the device (Mastersizer 3000, Hydro LV) and preparation methodology (US+ Calgon) used in our laboratory. Instead, it is advisable to use ion-exchanged water or distilled water, the possible exchange of which does not cause a verifiable difference in the measurement results in addition to the given methodology.

The theses related to aggregate stability studies:

3. Using the ***Random Forest algorithm***, determined that in the models estimating aggregate stability, built within a hierarchical system and representing the Hungarian soil databases (AIIR, MARTHA) within the Hungarian Soil Structural Database (HunSSD), the estimation of soil ***MiAS% proved to be more accurate*** (based on model performance indicators) than the estimation of MaAS%, considering the input test data used.

4. Using the above method, investigated which soil properties, and to what extent, enable the estimation of macroaggregate stability (MaAS (%)) in the

case of information from Hungarian soil databases with different levels of detail. The ***MaAS (%) model proved to be the most accurate*** when, in addition to the basic soil test data, we also have the results of the base exchange properties and land use information available. In this case, the most important soil properties (%) from the point of view of MaAS (%) were ***soil pH***, ***humus content*** and ***exchangeable*** $\boldsymbol{Mg^{2+}}$ content.

5. Using the above method - similarly to the previous ones - investigated which soil properties, to what extent, allow the estimation of microaggregate stability (MiAS (%)) in the case of information from Hungarian soil databases with different levels of detail, and to what extent these properties are important from the point of view of the estimation. The ***MiAS (%) model proved to be the most accurate*** when, in addition to the basic soil test data, we also have the results of the base exchange properties tests available. In this case, the most important soil properties in terms of MiAS (%) were, in order, the soil's ***exchangeable*** $\boldsymbol{Na^{+}}$ ***content***, **pH** and $\mathbf{CaCO_3}$ content. We could not verify the effect of land use in terms of the stability of microaggregates.

Major publications related to the thesis

Scientific articles in foreign language in foreign-printed reviewed journals with Impact Factor

Makó, A., Szabó, B., Rajkai, K., Szabó, J., Bakacsi, Zs., **Labancz, V.**, Hernádi, H., Barna, Gy.: Evaluation of soil texture determination using soil fraction data resulted from laser diffraction method. International Agrophysics 11 p. (2018). Q2; IF: 1,242

Scientific articles in foreign language in Hungarian-printed reviewed journals with Impact Factor

Labancz, V., Hernádi, H., Barna, Gy., Bakacsi, Zs., Szegi, T., Kocsis., M., Makó, A. 2024. The effect of different water types used for the measurement of soil particle size distribution with laser diffraction method. Hungarian Geographical Bulletin- 21 p.Q2; IF:1,71. *Submitted*

Szecsődi, O., Makó, A., **Labancz, V.,** Barna Gy., Gálos, B., Bidló, A., Horváth, A.: Using Different Approaches of Particle Size Analysis for Estimation of Water Retention Capacity of Soils: Example of Keszthely Mountains (Hungary) Acta Silvatica et Lignaria Hungarica: An International Journal in Forest, Wood and Environmental Sciences 17: 1 pp. 37-50., 14 p. (2021); Q4

Hungarian language in Hungarian publication

Labancz V., Barna Gy., Szegi T., Makó A.: A talajok aggregátum-stabilitásának vizsgálati lehetőségei I. Makroaggregátum-stabilitás Agrokémia és Talajtan 70: 1 pp. 87-109., 23 p. (2021); Q4

Makó A., Varga T., Hernádi H., **Labancz V**., Barna Gy.: Talajminták lézeres szemcseanalízisének módszertani tapasztalatai. Agrokémia és Talajtan, 66(1): 223–250 (2017); Q4

Makó A., Hernádi H., Barna Gy., Balázs R., Molnár S., **Labancz V**., Tóth B., Bakacsi Zs.: A talajok mechanikai összetétel vizsgálata pipettás ülepítéses módszerrel: a hazai és a nemzetközi szabvány szerinti eljárások összehasonlítása és konverziója. Agrokémia és Talajtan, 66(2): 295–315 (2017); Q4

Labancz, V., Makó, A., Bakacsi, Zs.: A klímaváltozás hatása a talajok vízgazdálkodására Mezőhír: Országos Agrárinformációs Szaklap 2020: 2 pp. 48-50., 3 p. (2020)

Conference publications in foreign language

Barna, Gy.; Bieganowski, A.; **Labancz, V**.; Ryżak, M.; Polakowski, C.; Sochan, A.; Bakacsi, Zs.; Rajkai, K.; Hernádi, H.; Molnár, S. et al. Comparative analysis of aggregate stability indices of typical Hungarian soil types: poster In: 22nd World Congress of Soil Science: poster book of abstracts Glasgow, Egyesült Királyság / Skócia: International Union of Soil Sciences (IUSS) (2022) Paper: P-525

Makó, A.; **Labancz, V.**; Bakacsi, Zs.; Hernádi, H.; Barna, Gy. Methodological comparison of particle size distribution data on a nationally representative soil database in Hungary In: Artur, Zdunek; Agata, Pacek-Bienek (szerk.) 13th International Conference on Agrophysics: Agriculture in changing climate: Book of abstracts Lublin, Lengyelország: Institute of Agrophysics, Polish Academy of Sciences (2021) 213 p. p. 147 Paper: P68. , 1 p.

Barna, Gy.; Tóth, T.; **Labancz, V.**; Bakacsi, Zs.; Rajkai, K.; Hernádi, H.; Makó, András Some methodological aspects of macro-aggregate stability measurements: poster In: Vaisvalavičius, Rimantas; Povilaitis, Virmantas (szerk.) 5th international symposium of soil physics: book of abstracts Kaunas, Litvánia: Vytauto Didziojo Universitetas (2022) 105 p. pp. 45-47., 3 p.

Barna Gy., Bakacsi Zs., **Labancz V**., Hernádi H., Makó A.: Methodological experiences of aggregate stability measurements. In: Celkova A. (szerk): Proceedings of Transport of water, chemicals and energy in the soil-plant-atmosphere system. 24th International Poster Day and Institute of Hydrology Open Day. Pozsony, Szlovákia, 2017.11.08. 29–35 (2017)

Labancz V., Šinkovičová M., Barna Gy., Szegi T., Tóth J., Kardos A. F., Herczeg E., Földényi R., Makó A.: Particle size distribution analysis of differently dispersed clayey soils measured by laser diffraction method. In: Celkova A (szerk): Proceedings of Transport of water, chemicals and energy in the soil-plant-atmosphere system. 24th International Poster Day and Institute of Hydrology Open Day. Pozsony, Szlovákia, 2017.11.08. 160–166 (2017)

Conference publications in Hungarian language

Labancz, V.; Sinkovicová, M.; Barna, Gy.; Szegi, T.; Tóth, J.; Kardos, F A.; Herczeg, E.; Földényi, R.; Makó, A. Különbözőképpen diszpergált, nagy agyagtartalmú talajminták lézeres szemcseanalízisének módszertani tapasztalatai In: Füleky, György (szerk.) XIV. Kárpát-medencei Környezettudományi Konferencia Gödöllő, Magyarország: MAG Mezőgazdaságért Alapítvány, (2018) pp. 170-174., 5 p.

References

AMÉZKETA, E., 1999. Soil aggregate stability: a review. J. Sustain. Agr. 14 (2–3). 83–151.

BALLENGER R. & DI GLÉRIA J., 1962. Talaj-és trágyavizsgálati módszerek. Mezőgazdasági Kiadó. Budapest.

BIEGANOWSKI, A., RYŻAK, M. & WITOKOWSKA-WALCZAK, B., 2010. Determination of soil aggregate disintegration dynamics using laser diffraction. *Clay Min.* 45. 23–34.

BIEGANOWSKI, A., RYŻAK, M., SOCHAN, A., BARNA, G., HERNÁDI, H., BECZEK, M., POLAKOWSKI, C. &MAKÓ, A. 2018. Laser Diffractometry in the Measurements of Soil and Sediment Particle Size Distribution Advances in Agronomy 151 pp. 215-279., 65 p.

BLOTT S.J. & PYE K., 2006. Particle size distribution analysis of sand-sized particles by laser diffraction: an experimental investigation of instrument sensitivity and the effects of particle shape. Sedimentology, 53, 671-685.

CHENU, C. & STOTZKY, G. 2002. Interactions Between Microorganisms and Soil Particles: An Overview, in Huang, P. M., Bollag, J.-M., Senesi, N. (eds.): Interactions Between Soil Particles and Microorganisms—Impact on the Terrestrial Ecosystems. John Wiley and Sons, Chichester, UK, pp. 3-40.

FISHER, P., AUMANN, C., CHIA, K., HALLORAN, N.O. & CHANDRA, S., 2017. Adequacy of laser diffraction for soil particle size analysis. PLoS One 1e20. https://doi.org/10.1371/journal.pone.0176510

LE BISSONNAIS, Y., 1996. Aggregate stability and assessment of soil crustability and erodibility: Theory and methodology. Eur J. Soil Sci. 47 (4). 425–437.

MAKÓ, A.,SZABÓ, B., HERNÁDI, H., FARKAS, CS.& MARTH, P.. 2010. Introduction of the Hungarian Detailed Soil Hydrophysical Database (MARTHA) and its use to test external pedotransfer functions. Agrokémia és Talajtan. 59. 10.1556/Agrokem.59.2010.1.4.

MAKÓ A., 2018. Új talajfizikai mérő- és becslőmódszerek kidolgozása vizes és nem-vizes folyadékfázist tartalmazó talajokra. MTA doktori értékezés. Budapest.

MAKÓ, A., SZABÓ, B., RAJKAI K., SZABÓ, J., BAKACSI, ZS., LABANCZ, V., HERNÁDI, H., & BARNA, GY., 2019. Evaluation of soil texture determination using soil fraction data resulting from laser diffraction method. International Agrophysics 33. 445-45

MCCAVE, I. N., BRYANT, R. J, COOK, H. F. & COUGHANOWR, C. A., 1986. Evaluation of a laser-diffraction size analyzer for use with natural sediments. Research Methods Papers. 561–564.

RYZAK, M. & BIEGANOWSKI, A., 2011. Methodological aspects of determining soil particle-size distribution using the laser diffraction method. J. Plant Nutr. Soil Sci. 174, 624e633. https://doi.org/10.1002/jpln.201000255.

STEFANOVITS P., FILEP GY. & FÜLEKY GY., 1999. TALAJTAN. MEZŐGAZDA KIADÓ, BUDAPEST.

TISDALL, J.M. & J.M. OADES. 1982. ORGANIC MATTER AND WATER-STABLE AGGREGATES IN SOIL. J. SOIL SCI. 33:141–163.

TOTSCHE, K.U., AMELUNG, W., GERZABEK, M.H., GUGGENBERGER, G., KLUMPP, E., KNIEF, C., LEHNDORFF, E., MIKUTTA, R., PETH, S., PRECHTEL, A., RAY, N. & KÖGEL-KNABNER, I., 2018. Microaggregates in soils. J. Plant Nutr. Soil Sc. 181 (1). 104–136.

VAGELER, P., 1932. Der Kationen- und Wasserhaushalt des Mineralbodens: Vom Stand-punkt der Physikalischen Chemie und Seine Bedeutung für die Land- und Forstwirtschaftliche Praxis. Springer, Verlag Berlin Heidelberg.

WRIGHT, M. N., & ZIEGLER, A. 2017. Ranger: A Fast Implementation of Random Forests for High Dimensional Data in C++ and R. Journal of Statistical Software, 77(1), 1-17.

YANG, X., ZHANG, Q., LI, X., JIA, X., WEI, X. & SHAO, M., 2015. Determination of soil texture by laser diffraction method. Soil Sci. Soc. Am. J. 79, 1556.