Geosciences and Engineering, Vol. 11, No. 1 (2023), pp. 56–67.
https://doi.org/10.33030/geosciences.2023.01.005

DIGITAL SOIL MAPPING APPROACH TO ESTIMATE SOIL PLASTICITY USING GEOREFERENCED TECHNICAL DATA OF TRACTORS

ENDRE DOBOS[1*], EMESE SZABÓ[2], SÁNDOR CSENKI[3], TSEGAY BEREKET MENGHIS [4], FERENC MOLNÁR[5], ANDRÁS DOBAI[6], MOHAMED RAJHI[7],
[1]**Faculty of Earth and Environmental Sciences and Engineering, University of Miskolc, Hungary; endre.dobos@uni-miskolc.hu;*
[2]*KITE Zrt., Nádudvar, Hungary; szaboemese@kite.hu*
[3]*Faculty of Earth and Environmental Sciences and Engineering, University of Miskolc, Hungary; sandor.csenki@uni-miskolc.hu,*
[4]*Faculty of Earth and Environmental Sciences and Engineering, University of Miskolc, Hungary; tsegay.bereket.menghis@student.uni-miskolc.hu*
[5]*Faculty of Earth and Environmental Sciences and Engineering, University of Miskolc, Hungary; ferenc.molnar@uni-miskolc.hu*
[6]*Faculty of Earth and Environmental Sciences and Engineering, University of Miskolc, Hungary; andras.dobai@uni-miskolc.hu*
[7]*Faculty of Earth and Environmental Sciences and Engineering, University of Miskolc, Hungary; mohamed.rajhi@uni-miskolc.hu*
[1]*https://orcid.gov/0000-0002-9798-6376*
[3]*https://orcid.gov/0009-0002-9603-3765*
[4]*https://orcid.gov/0009-0006-4263-1005*
[5]*https://orcid.gov/0000-0002-4872-1136*
[6]*https://orcid.gov/0000-0002-5267-0728*
[7]*https://orcid.gov/0009-0006-2842-6240*

Abstract: Soil and terrain properties are the most important environmental elements for characterization of within-field variability supporting the management zone definition. This study aims to develop a digital soil mapping approach to map texture related soil property, namely soil plasticity (Ka) using georeferenced machinery data derived from log file created along soil-tillage activities. The results indicated that speed and distance values have the highest correlation with Ka, having values around 0.4–0.5. The development of the regression model resulted in an adjusted R^2 value of 0.71, but the p-value was still around 0.2. These results indicate that the approach has a good potential, but statistics need to be improved with data preprocessing.

Keywords: *soil plasticity, precision agriculture, management zones*

1. INTRODUCTION

Precision farming is the most state-of-the-art technology in agriculture that has been introduced in the last few decades. Its major advantage is that all agrotechnological elements – like soil tillage, planting-seeding, fertilization, plant protection – can be adjusted to the current environmental and plant conditions [1]. The most important environmental factors that define the productivity are the soils and the terrain properties. Better soil productivity may support higher plant density and requires lower amount of fertilizer. Therefor the precise characterization of the soil-landscape conditions and their variability determine the effectiveness of the agricultural activity. Precision agriculture requires very high spatial accuracy to navigate the field equipment. Real-time kinematic positioning (RTK) is used to reach 1 cm scale accuracy to navigate between plant rows using this navigation system. The machinery always knows its position and adjusts the current treatment to the known conditions of the actual point. Therefore, the effective production requires an intensive survey of the static and dynamic soil conditions, accurate digital elevation data and up to date information on the plant condition and its spatial patterns and variability [2–7].

High resolution terrain data can be easily derived from the RTK navigation logger data. The RTK equipped machinery also collects accurate elevation data in a preadjusted density and is able to log it. This high density elevation point data can be interpolated or converted to any raster datasets of 1 meter scale raster resolution. This DEM can later be used to model the surface and subsurface water flow and nutrient translocation within the field, and it is also a very efficient predictor for digital soil mapping approaches [8–14]. Plant condition characterization is using proximal and remote sensing datasets and produce vegetation indices and other indirect information to describe the biomass and the crop "wellness" around the field. Accurate and high resolution soil data is difficult to capture. There are several indirect indicators of soil properties and their variability, like different kinds of geophysical measurements [15, 16]. These are quick methods, but not easy to validate. The most important soil property that these measurements are sensitive to is the soil moisture content. However, soil moisture depends both on the weather conditions and related soil properties that define the water holding capacity of the soils – namely texture, structure, porosity, CEC, humus content, etc. For that reason, direct relationship between the current soil moisture conditions and certain soil properties are difficult to define. Soil sampling and data validations are crucial to derive real soil property maps. On the other hand, EM38, soil conductivity and electrical resistance data is a very good predictor for soil mapping. Accurate soil mapping requires the opening, description and sampling of soil pits on the representative geomorphological, geological units of the fields and a good model that explains the landscape development of the quaternary periods. DEM and remotely sensed historical data explain a lot on the surface soil patterns. Understanding the real soil variability requires information on the vertical variability of soil moisture and nutrient contents in the rooting zone, namely the sequences of the underground genetic horizons and geological layers and the corresponding chemical, physical and biological properties. Field mapping is time consuming and expensive, requires expert knowledge that is difficult to schematized. Digital soil mapping activity that explains the short range variability within the field can minimize this activity and make soil mapping more effective and reintegrated into the agrotechnological planning.

Soil texture and particle size distribution analysis are one of the most basic and important soil properties, that is used for almost all soil related agrotechnological treatments, like soil cultivation, fertilization and irrigation [17–19]. Despite its importance, lab analysis is time consuming and one of the most expensive standardized measurements used by soil experts for agricultural and soil conservation consultancy. There are several easy to do measurements aimed to simplify the texture characterization, for example soil plasticity, which is strongly related to the texture. In Hungary, there is a widely used plasticity parameter, the Arany plasticity number (Ka), which is often directly translated to texture classes *(Table 1)*. The method is very simple. 100 g air-dried soil sample prepared for lab analysis is used for the measurement. Ion-exchanged water is continuously added to the soil with continuous stirring till reaching the end of the plastic stage or beginning of liquefaction, using the so-called "yarn test". This test is done by using the stirring stick. The plastic soil material is touched by the stick and pulled slowly upwards. In the plastic soil stage, the stirring stick pulls the soil material up to a point, when its breaks away and the thin soil material bends back following the gravitation like a dropped yarn. The amount of water in ml used to reach this stage defines the value of Ka. Due to the simplicity of this examination, this measurement is the most commonly used and understood property both by farmers and soil specialist.

Table 1. *The translation/correlation table between texture classes and the Ka-values* [19]

Texture class	**Ka-value**
Coarse sand	<25
Sand	25–30
Sandy loam	30–37
Loam	37–42
Clay-loam	42–50
Clay	50–60
Heavy clay	>60

The spatial distribution of the texture classes and the different particle size classes is very important in the definition of the management zones of precision farming, therefor a continuous map of Ka values covering the entire field in high resolution can be a very important covariant of management zone definition. This work aims to develop a digital soil mapping procedure to assess the Ka values using georeferenced machinery data like fuel consumption, speed and forward moved distance between regular time intervals derived from the soil tillage events. Physical soil properties, like texture, compaction, plasticity have a great impact on the pulling power for soil tillage. Heavy, clayey soils have much higher resistance against ploughing, disking, harrowing, so the fuel consumption is higher, while the speed and the forwarding distance is lower than that of the sandy soils. This work aims to make use of this simple relationship and tries to build a quantitative model for surveying Ka and support soil mapping.

2. Materials and methods

2.1. Pilot area

The pilot area is located in Eastern-Hungary, west of Tépe settlement *(Figure 1)*. The area belongs to the Körös-Berettyó geographic meso-region and the Berettyó-Kálló interfluve small region [20]. The size of the field is 62 hectares. It is located in a relatively low-lying, backwater region between higher elevated levees of historical water channels. Therefor the area in general is poorly drained, but some small abandoned drainage channels are crossing the area. According to the WRB classification [21], the most dominant soil type is Chernozem, but in the lower lying depressions of the field Gleysols, Solonetz, and Phaeozem soils occur as well.

The parent material is mainly alluvial, the texture is loam and clay-loam with smaller clayey spots. The average sand-silt-clay percentages are 40-30-30, but the silt may go up 55%, indicating the infusional loess, as one of the most important parent material along the alluvial clays. The Ka values range between 41 and 52.

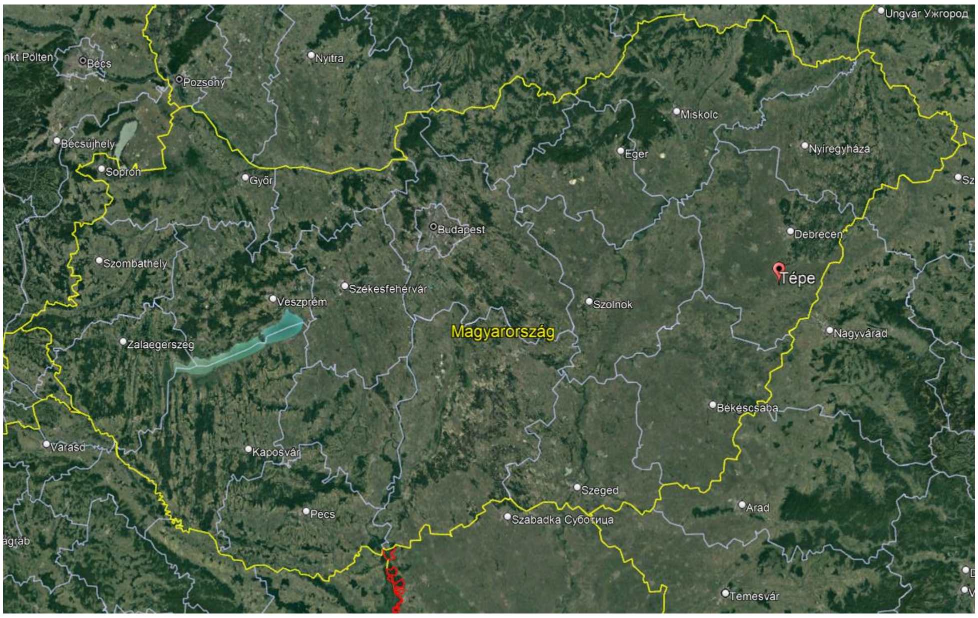

Figure 1

The location of the pilot area within Hungary

2.2. Data

Ka values from 11 sampling points were used to calibrate the model. The sampled points are shown on *Figure 2*. Four datasets from the target field were applied for the study. The dates and the tillage type are shown in *Table 2*. Three variables were extracted from the log files, the speed, the fuel consumption and the forwarding distance per time unit. The example dataset of the log file is shown on *Figure 2*.

Table 2

Dates and tillage types of the datasets used for the study

Date	**Tillage type**	**Naming code**
November 30, 2018.	Ploughing	18 plus variable name
June 12, 2019.	Strip till	1906 plus variable name
November 5, 2019.	Ploughing	1911 plus variable name
April 16, 2023	Primary soil tillage	23 plus variable name

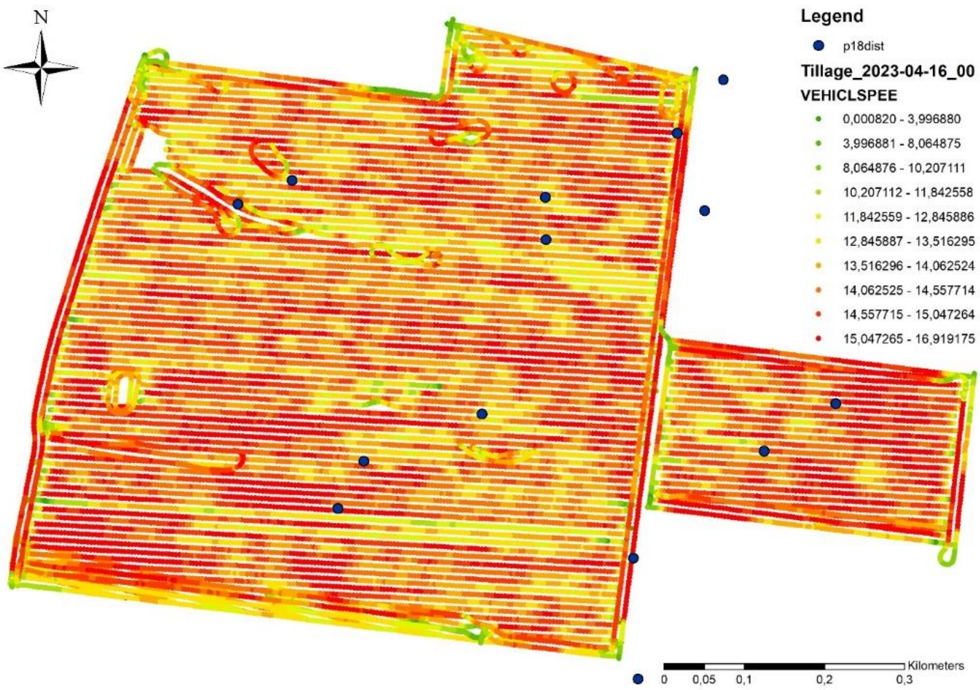

Figure 2

The log data of the field from 2023 April, showing the speed values on the logged positions. The blue points show the sampled sites

2.3. Data processing and Ka estimation procedure

Three variables from the four dates were extracted from the datasets and these twelve explanatory variables were complemented with the corresponding Ka value. In the first step, correlation between the Ak value and each of the selected 12 variables were calculated using the “korrel” function of the Microsoft Excel 2016 software package [22].

Table 3

The extracted database

Ak	23fuel	23speed	1911speed	1911fuel	1906speed	1906fuel	18speed	18fuel	1911dist	1906dist	23dist	18dist
42	0.019	14.057	7.049	0.014	5.256	0.002	7.012	0.011	2.076	1.459	3.941	2.016
42	0.019	13.436	8.159	0.013	10.366	0.009	8.146	0.013	2.264	2.883	3.747	2.263
46	0.011	14.985	7.454	0.011	5.941	0.020	7.045	0.019	2.069	1.971	4.040	2.033
48	0.019	15.304	6.308	0.014	9.035	0.006	7.857	0.012	1.751	2.506	4.254	2.181
48	0.019	14.712	8.876	0.014	10.710	0.007	8.332	0.014	2.462	2.978	4.094	2.316
48	0.019	14.345	8.837	0.014	10.230	0.008	8.654	0.014	2.454	2.837	3.998	2.402
52	0.019	13.406	8.939	0.014	10.375	0.009	8.581	0.013	2.480	2.885	3.725	2.384
52	0.019	14.931	7.585	0.014	10.342	0.008	8.212	0.013	2.107	2.868	4.150	2.280
48	0.019	13.692	8.505	0.014	9.885	0.006	8.762	0.014	2.362	2.748	3.799	2.432
41	0.019	14.413	8.677	0.014	10.003	0.007	8.043	0.014	2.410	2.781	3.995	2.233
48	0.019	14.028	8.365	0.013	10.198	0.008	8.110	0.013	2.323	2.835	3.898	2.254

The variables showing the highest correlation with Ka values were selected and used as inputs for multiple linear regression. Two models were developed. The first model used only the four predictors individually, not taking into consideration the potential predictor interactions using the model below:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 \tag{1}$$

In the second run, the model was extended to use the squared variables as potential part of the model according to the model below.

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_{1,1} x_1 * x_1 + \beta_{2,2} x_2 * x_2 + \beta_{3,3} x_3 * x_3 + \beta_{4,4} x_4 * x_4 \tag{2}$$

When running the models the following assumptions were made:

- Values of the response variable Ka vary according to a normal distribution with standard deviation σ for any values of the explanatory variables 18dist, 18speed, 1906dist and 1906speed. The quantity σ is an unknown parameter.
- Repeated values of response variable Ka are independent of one another.
- The relationship between the mean response of Ka and the explanatory variables are linear.

3. RESULTS AND DISCUSSION

3.1. The data analysis

The results of the correlation study is shown in *Table 4*.

Table 4

The result of the correlation calculation between Ka value and the predictor variables. The variables eith the highest correlation values are highlighted as gray lines

Variable name	Correlation value
23fuel	0.060
23speed	0.142
1911speed	0.126
1911fuel	0.231
1906speed	0.397
1906fuel	0.099
18speed	0.489
18fuel	0.007
1911dist	0.065
1906dist	0.422
23dist	0.144
18dist	0.503

The correlation values are ranging between 0.007 to 0.5. Interestingly, none of the fuel consumption variables for the four dates show any promising details, their correlation values are tending toward 0, or being very low. The speed and distance values have relatively higher correlation with the Ka values. Two dates had the highest correlation values – around 0.4–0.5 –, these are 2018 with ploughing and 2019 July with strip till In both cases the distance and the speed variables explained more of the Ka values. Therefore, these four were selected as inputs for the regression study. The kind of soil cultivation method has probably great impact on these values. Ploughing with higher pulling power need was expected to come among the best predictors, which was proved only in the 2018 case. The current soil moisture condition may also have a great impact on the pulling power need, dry soils have higher resistance against the tillage. Unfortunately, soil moisture data for cultivation events were not available, therefor this theory was not proved.

3.2. The multiple regression analysis results

A multiple regression analysis was performed using the Ka value as predicted, response variable, and the distance values from 2018 (18dist) and July 2019 (1906dist) and speed values from the same two days (18speed and 1906speed variables) were put as explanatory variables into the regression model 1. The regression equation was calculated as below:

$$\mathrm{Ak} = 4.8088 + 0.2292 \cdot 18\mathrm{dist} - 0.0128 \cdot 18\mathrm{speed} + 0.1037 \cdot 1906\mathrm{dist} - 0.0284 \cdot 1906\mathrm{speed} \quad (3)$$

The summary statistics are shown in *Table 5*.

Table 5

Statistics summary of the regression model 1

Predictor	**Coefficient**	**Estimate**	**Standard Error**	**t-statistic**	**p-value**
Constant	β0	4.8088	38.8002	0.1239	0.9054
18dist	β1	0.2292	1.6038	0.1429	0.891
18speed	β2	–0.0128	0.4444	–0.0288	0.9779
1906dist	β3	0.1037	0.1866	0.5561	0.5983
1906speed	β4	–0.0284	0.0643	–0.4414	0.6744
Summary of Overall fit R-Squared	0.375				
Adj. R-Squared	0.154				
Residual std error	4.043 on 6 degrees of freedom				
Overall F-stat	0.666 on 4 and 6 degrees of freedom				
Overall p-value	0.638				

The results of the first model show very limited potential for the Ka prediction. The p-values are extremely high, showing that the null hypothesis, namely that there is no real correlation between the predicted and predictor variables has very high probability. The same result is highlighted by the very low R^2 values as well. It is also interesting, that the speed and the distance, which should be highly correlated have a positive and negative signs in the same time. However, running model 2 has very different and very promising results *(Table 6.)*

The regression equation was calculated as below:

$$\begin{gathered}\mathrm{Ak}{=}{-}9806{,}85 + 19{,}74 * (18\mathrm{dist}) + 19{,}04 * (18\mathrm{speed}) \\ + 15{,}18 * (1906\mathrm{dist}) - 4{,}8 * (1906\mathrm{speed}) \\ - 0{,}019 * (18\mathrm{dist})^2 - 0{,}0131 * (18\mathrm{speed})^2 \\ - 0{,}04 * (1906\mathrm{dist})^2 + 0{,}0033 * (1906\mathrm{speed})^2\end{gathered} \quad (4)$$

Table 6

Statistics summary of the regression model 2

Predictor	**Coeff.**	**Estimate**	**Standard Error**	**t-statistic**	**p-value**
Constant	β0	–9806.8524	7545.8483	–1.2996	0.3234
18dist	β1	19.7438	116.0675	0.1701	0.8806
18speed	β2	19.0398	19.1116	0.9962	0.4241
1906dist	β3	15.1805	10.3176	1.4713	0.279
1906speed	β4	–4.8016	2.6326	–1.8239	0.2097
18dist∗18dist	β1,1	–0.019	0.2584	–0.0735	0.9481
18speed∗18speed	β2,2	–0.0131	0.0117	–1.1225	0.3783
1906dist∗1906dist	β3,3	–0.0403	0.0296	–1.3634	0.3059
1906speed∗1906speed	β4,4	0.0033	0.0022	1.5108	0.2699
Summary of Overall fit R-Squared	0.942				
Adj. R-Squared	0.714				
Residual std. error	2.011 on 2 degrees of freedom				
Overall F-statistics	4.373 on 8 and 2 degrees of freedom				
Overall p-value	0.209				

In model 2, where the squared formulas are included into the model as well, the model performance increased a lot. The R^2 value is almost 1, and even the adjusted R^2 value is exceeding the value of 0.7, which is already a promising result. Of course, the observation number of 11 for the entire area is quite limited, which may include some significant bias on the existing input data, which is clearly indicated by the relatively high p-value of 0.2. It means that statistically interpreting the results, it is not proved significantly that accurate Ka values can be estimated with the use of machinery log data. However, taking into consideration the limited number of observations and the potential of input data cleaning and processing in increasing the model performance – which was not the aim of this study, this methodology has a potential in estimating and mapping soil physical parameters.

4. SUMMARY AND CONCLUSION

This study aimed to develop and test a provisional methodology to assess soil physical properties using digital soil mapping tools and machinery data collected and logged through different RTK navigation based on soil tillage activities. Georeferenced data (like speed, fuel consumption and forwarding distance per time unit) from four different tillage activities (ploughing, strip tillage and primary soil tillage) were collected and used to predict their applicability for mapping. Correlations between a commonly used soil plasticity index (Ak) and these three variables from four different dates were calculated. Fuel consumption was found to be irrelevant, but the speed and the forwarding distance values had correlation between 0.4 and 0.5. A multiple linear regression was used to assess the Ak values using these four predictor variables. Ploughing and strip tilt showed the highest correlation, but data on soil moisture was not recorded to backup this statement. The regression study indicated – but statistically not proved – that this approach has a great potential to assess physical soil properties. This study only aimed to test the data, but further data processing studies and higher number of calibration data is still needed to prove the validity of the concept.

ACKNOWLEDGEMENT

The research work was published in the framework of Széchenyi 2020 entitled *Creation of innovation operative groups and support of the investment necessary for the implementation of the innovative project* (VP3-16.1.1-4.1.5-4.2.1-4.2.2-8.1.1-8.2.1-8.3.1-8.5.1-8.5.2-8.6.1-17) was made with the support of a tender. Application ID: 1866004120.

REFERENCES

[1] Franzen, D., Mulla, D. (2015). A history of precision agriculture. In: *Precision agriculture technology for crop farming*. pp. 1–20.

[2] Yari, A., Madramootoo, C. A., Woods, S. A., Adamchuk, V. I., Huang, H. H. (2017). Assessment of field spatial and temporal variabilities to delineate site-specific management zones for variable-rate irrigation. *Journal of Irrigation and Drainage Engineering*, 143, 9, 04017037.

[3] Saleh, A., Belal, A. (2014). Delineation of site-specific management zones by fuzzy clustering of soil and topographic attributes: A case study of East Nile Delta, Egypt. *IOP Conference Series: Earth and Environmental Science*, 18, 012046.

[4] Ruffo, M. L., Bollero, G. A., Bullock, D. S., Bullock, D. G. (2006). Site-specific production functions for variable rate corn nitrogen fertilization. *Precision Agriculture*, **7,** 5, pp. 327–342.

[5] Reyes, J., Wendroth, O., Matocha, C., Zhu, J. (2019). Delineating Site-Specific Management Zones and Evaluating Soil Water Temporal Dynamics in a Farmer's Field in Kentucky. *Vadose Zone Journal*, 18, pp. 1–19.

[6] Pedersen, S. M., Lind, K. M. (2017). Precision Agriculture – From Mapping to Site-Specific Application. In: Pedersen, S., Lind, K. (eds.). *Precision Agriculture: Technology and Economic Perspectives. Progress in Precision Agriculture*. Springer, 6330 Cham, Switzerland, pp. 1–20. https://doi.org/10.1007/978-3-319-68715-5

[7] Godwin, R., Miller, P. (2003). A Review of the Technologies for Mapping Within-field Variability. *Biosystems Engineering*, 84, pp. 393–407.

[8] Dobos, E. Micheli, E. Baumgardner, M. F. (1997). Digitális magassági modellek és a felszín-tagoltsági mutató használata kis méretarányú talajtérképezésben. *Agrokémia és Talajtan*, 46, 1–4, pp. 311–325.

[9] Dobos, E., Micheli, E., Baumgardner, M. F. (1998). Statistical Analysis of Advanced Very High Resolution Radiometer Data (AVHRR) Soil Relationship. Hungarian Contribution to the 16th International Congress of Soil Science. *Agrokémia és Talajtan*, 47, 1–4, pp. 49–62.

[10] Dobos, E., Micheli, E., Baumgardner, M. F., Biehl, L., Helt, T. (2000). Use of combined digital elevation model and satellite radiometric data for regional soil mapping. *Geoderma*, 97, pp. 367–391.

[11] Dobos, E., Normann, B., Worstell, B., Montanarella, L., Johannsen, C., Micheli, E. (2002). The use of DEM and satellite data for regional scale soil databases. *Agrokémia és Talajtan,* 51, 1–2, pp. 263–272.

[12] Kibirige D., Dobos, E. (2021). Soil Moisture Estimation Using Citizen Observatory Data, Microwave Satellite Imagery, and Environmental Covariates. *Water*, Vol. 13 (6), p. 837.

[13] Illés, G., Kovács, G., Bidló, A., Heil, B. (2006). Digital Soil and Landsite Mapping in Forest Management Planning. *Agrokémia és Talajtan*, 55, 1, pp. 99–108.

[14] Moore, I. D., Gessler, P., Nielsen, G. A. E., Peterson, G. (1993). Soil Attribute Prediction Using Terrain Analysis. *Soil Science Society of America Journal*, 57, pp. 443–452.

[15] Kitchen, N. R., Drummond, S. T., Lund, E. D., Sudduth, K. A., Buchleiter, G.W. (2003). Soil electrical conductivity and topography related to yield for three contrasting soil–crop systems. *Agronomy Journal*, 95, 3, pp. 483–495.

[16] Landrum, C., Castrignanò, A., Mueller, T., Zourarakis, D., Zhu, J., De Benedetto, D. (2015). An approach for delineating homogeneous within-field zones using proximal sensing and multivariate geostatistics. *Agricultural Water Management*, 147, pp. 144–153.

[17] Skopp, J. (1992). *Concept of soil physics*. University of Nebraska, Lincoln, NE.

[18] Burt Rebecca (eds.) (2004). *Soil Survey laboratory methods manual*. Soil Survey Investigation reports. No42. Version 4.0. USDA, November, 2004.

[19] Buzás, I. (1993). *Talaj- és agrokémiai vizsgálati módszerkönyv. I. A talaj fizikai, vízgazdálkodási és ásványtani vizsgálata*. INDA 4231 Kiadó. Budapest.

[20] Dövényi, Z. (szerk.) (2010). *Magyarország kistájainak katasztere. 2nd edition.* MTA Földrajztudományi Kutatóintézet, Budapest.

[21] IUSS Working Group WRB (2015). *World Reference Base for Soil Resources 2014, update 2015 International soil classification system for naming soils and creating legends for soil maps*. World Soil Resources Reports No. 106, FAO, Rome.

[22] Benesty, J., Chen, J., Huang, Y., Cohen I. (2009). Pearson Correlation Coefficient. In: *Noise Reduction in Speech Processing*. Springer Topics in Signal Processing. Vol 2. Springer, Berlin, Heidelberg.