A paper from the Proceedings of the
15th International Conference on Precision Agriculture
June 26-29, 2022
Minneapolis, Minnesota, United States

COMPARISON AND VALIDATION OF DIFFERENT SOIL SURVEY TECHNIQUES TO SUPPORT A PRECISION AGRICULTURAL SYSTEM

Lang V.[1], Tóth G.[1], Dafnaki D.[1], Csenki S.[1]

[1] Discovery Center Nonprofit Kft., Hungary, Gödöllő, Harsfa street 1.

Abstract

The data need of precision agriculture has resulted in an intensive increase in the number of modern soil survey equipment and methods available for farmers and consultants. In many cases these survey methods cannot provide accurate information under the used environmental conditions. On a 36 hectare experimental field, several methods have been compared to identify the ones which can support the PA system the best. The methods included contact and non contact soil scanning, yield mapping, high accuracy field soil survey using soil pits, grid based soil sampling, remote sensing with satellites and UAVs. Soil classification diagnostic features were collected from >130 individual points to assess the usability of modern soil classification systems for precision agriculture purposes.

The accuracy of the tested methods was validated on the soil map, and with the use of several years of harvester based yield maps. We have found that although all the methods have identified certain features on the field, many of these had no or minimal effect on the yield potential of the field zones, thus the usage of these tools or techniques have no or negative effect on the cropping system and on the environment. Modern soil classification diagnostic units, although in many cases neglected for PA systems, can provide a more detailed baseline information for decision support, through the aggregated soil chemical, physical and morphological information contained in one class. With modern digital soil mapping technologies and a custom made sampling methodology these tools are comparable in speed and provide deeper knowledge on the field than the ‘modern’ quick survey techniques. Since precision agriculture is an intensively expanding

market and is supposed to be one of the main tools for sustainable agriculture a broad validation of these tools should be performed to identify the soil and climatic conditions where these can be properly used with providing the required precision and reliability.

Keywords.
Management zone, World Reference Base, Soil classification, NDVI, Soil scanners

Introduction

The data need of precision agriculture has resulted in an intensive increase in the number of modern soil survey equipment and methods available for farmers and consultants. Often the cost and fast data acquisition has more importance than the accuracy and reliability of the collected soil or soil related information. In many cases these survey methods can hardly provide accurate information under the used environmental conditions.

Several techniques have been adopted recently to provide soil information for site specific nutrient management. Besides the most common grid sampling, management zone delineation based on different data sources have been widely used among farmers and consultants. This paper compares these survey techniques from a yield potential perspective.

As far as the grid sampling is concerned, a regular grid of points is defined and sampled. Thus, we subdivide the sampling problem by areas. The main weak spot of systematic sampling is the absence of an unbiased estimator of the variance [9]. Nevertheless, an intensive grid sampling approach reveals more variability than sampling approaches based on zones representing larger field [16]. However, a sampling approach like this cannot be practically adopted by producers when one takes into account sampling and analysis costs, labor and likely responses of crops to fertilization. On-farm scientific research has shown that it is not possible to measure small-scale nutrient variability cost-effectively using current grid sampling methods [16, 17, 25, 8]. It is, also, noteworthy that working on an equal-area projection is a good choice to ensure that each point represents the same area [9].

On the other hand, Management Zones Sampling improves the conventional sampling method by using information that can be collected using precision agriculture technologies. It reduces the number of samples and sampling costs while maintaining acceptable information about nutrient variability within a field [16, 17]. However, a management zone scheme will represent small-scale soil test variability less accurately than an intensive grid sampling scheme, because there will be fewer samples. Such a sampling scheme will be especially effective when soil type and nutrient removal by crops are major factors in determining nutrient variability across large areas [16]. Also, expertise and subjective judgement are prerequisites for the success of the method [9, 17, 8].

The main steps for the designation of Management Zones are the following: (i) Mapping of properties involved in the analysis by using measurements from proximal and remote sensing and interpolation of the unsampled points, (ii) Subjecting the data to a clustering procedure based on an algorithm, which aims to depict the natural structure of the data and in most cases either minimizes the sum of the within-cluster variances, uses the data density, or is based on distance connectivity between pairs of observation, (iii) Finding the optimal number of classes in order to achieve a balance between spatial variation in soil properties and a manageable number of zones that are spatially well distributed, (iv) Delineating final zones using a GIS system, since the boundaries between clusters can shape the potential zones and (v) Assessing the efficiency of the classification by several criteria (e.g. variance reduction of the management zones compared to the within-field variance, accuracy, cost–benefit analysis etc.) [5, 13, 20].

Some of the most widely used precision agriculture technologies that enable soil scientists to obtain data and create management zones are yield mapping, proximal sensing, remote sensing, traditional soil surveys all of them with known advantages and disadvantages.

Yield Mapping

The essential data for the creation of a yield map are the actual yield within the field as it is harvested and the position of the combine harvester at a specific time. This means that yield is spatially related and, thus, yield mapping is used for managerial purposes. There are two main errors that have been identified in many yield maps; the lag time between detachment and sensing of the grain, and the unknown width of crop entering the header. Furthermore, it is hard to achieve practical implications of management if the yield map contains many perturbances. In that case, smoothing of the data is required, so that only the important highs and lows are disclosed [4].

Proximal Sensing and Soil Scanners

Proximal sensing has been successfully used to acquire quantitative and qualitative soil information, as it includes field-based sensors that are in contact with or close to the soil. These sensors can be more time- and cost- efficient than conventional laboratory analyses and they are becoming smaller, faster, more accurate, more energy efficient, wireless and more intelligent along with the progress of technology [24].

EC Scanners

Electro conductivity of the soil is an indirect index, which correlates well with several physical and chemical properties of soil such as size and texture of soil particles, soil organic matter content, cation exchange capacity, water content in the soil, soil depth above the clay or stone layer, salinity and soil temperature. Experts highlight the high quality of the method, the low production cost and the low environmental footprint [14].

EM Scanners

Electromagnetic induction (EMI) scanning is a rapid, non-invasive method for obtaining soil ECa (apparent soil electrical conductivity) information, especially on soil moisture content and soil texture. However, there is not always a perfect 100% agreement between the spatial location of ECa boundaries and the principal soil units of the intensive soil map [12]. This is because soil ECa is a function of soil texture and soil water content, but also of other factors, including soil bulk density and organic matter content. It is a fact that only when the costs are spread over a number of years would they be affordable for commercial farming. In fields where there are clear contrasts in soil texture, EMI surveying techniques can provide a cost-effective method as an addition to traditional soil survey practices [24, 12].

Remote Sensing

Remote sensing techniques use radiation and are now delivering data on land surface and subsurface characterization at increasingly higher spectral and spatial resolutions [23]. For large scale surveys airborne methods are required, whereas medium scale surveys may be aided by both airborne and spaceborne methods. Lastly, small-scale surveys are served most by the use of satellite-data. Of equal importance for the selection of the most appropriate remote sensing means is the purpose of the study, the specific characteristics of objects at the earth's surface and the climatic conditions. [19] UAV-based remote sensing is less accurate and less feasible than proximal sensing, but it is cost-effective, fast in producing, manageable for the local staff and has good geometric accuracy [22]. It is now possible to produce accurate maps of within-field yield variation at 10 m resolution using Sentinel-2 data. Still, the potential of Earth Observation has been limited by image costs and limited repeat frequency [11].

Satellite Images

The use of satellite images in soil science has mostly been driven by the growth of a wide range of medium resolution space-borne sensors on stable and fixed orbits. Cost and availability remain important issues as imagery from high-resolution sensors is not only more expensive, but also less likely to cover the area in consideration [22, 23]. Several spaceborne platforms enable soil scientists to do the quantitative measurement of ground features with a relatively lower cost per unit area of coverage. So, satellite images are especially useful to study large areas [18]. Also, higher resolution data require great storage and computing capacity. Some significant resources that were brought out before 2000 have adequate repeat coverage, large scene size and low cost of entry. It is also noteworthy that images from Landsat 8 and Sentinel-2 are free to use [23].

Digital Elevation Model and derivatives

The Digital Elevation Model (DEM) is a 3D representation of the bare soil surface, without depicting any objects that may exist on it [26]. It is a quantitative data source and, as a result, it has some important advantages like consistency and homogeneity, let alone the fact that data generalization and edge-matching problems are significantly reduced. The DEM-derived variables are elevation, slope, aspect, curvature, potential drainage density etc.. In some cases these attributes show high correlation with the soil classes [7]. This method is used for mapping large areas quickly and cost-effectively. However, the success of the method depends on the quality of the input imagery and the meticulousness during pre-processing [3]. Last but not least, the lower altitude at which they fly, allows much higher spatial resolutions than satellite imagery [22, 23].

Traditional soil survey and soil classification

The soil landscape can be methodically described by the WRB diagnostic horizons, which correspond to a standard depth and, consequently, the high intensity soil survey makes the results of soil genesis quantifiable. Relating the diagnostic horizons in their thickness to the upper 100 soil centimeters is a good option to replace traditional soil classification and identify the soil profile [15]. On the other hand, classifying depending on the morphology and formation of soil profiles, doesn't allow the examination of other properties, like slope, nor the achievement of high accuracy and speed [1, 6]. Also, the issue of inevitable subjectivity occurs when different soil scientists map similar areas [6].

Random Forest

A Random Forest (RF) classifier produces multiple independent decision trees, using a randomly selected subset of training samples and variables to separate the data into more homogenous units, for the purpose of a single prediction [2, 11]. Two essential parameters in this method are the number of trees (ntree) and the number of variables available for selection in each split (mtry) [21]. The Random Forest algorithm presents clear advantages over other models for mapping. To begin with, it can increase the amount of data available for training, and consequently its estimative potential, because keeping back data for validation is less important for this model. Moreover, RF is able to take advantage of relationships between explanatory variables in order to control confounding factors. The number of variables that are needed for a precise estimation can be decreased, due to its ability to deal with multi-variate relationships between data of different types and resolutions.

Materials and Methods

The research was performed in Somogy county, Hungary. The area is described by loess blanketed or quicksand-covered sediments with milder slopes (5-15°) and a relative relief intensity of 20 - 50 m/4 km2. The 36 hectare field in which the different soil survey techniques were tested is a very heterogenous field described by different soil types due to the high relief intensity and the resulting erosion.

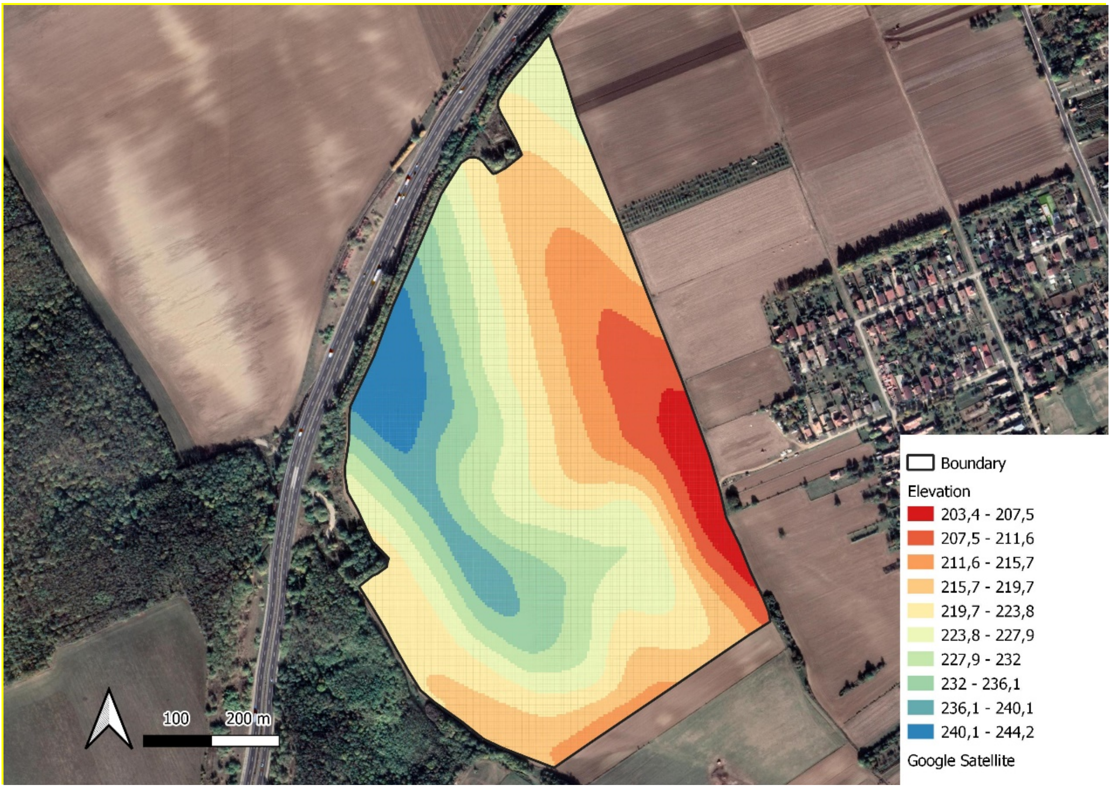

Figure 1. Digital Elevation Model of the study area (UAV based)

In the region vineyards were dominating the landscape. The land use resulted in a very intensive erosion process before the vineyards were removed and most of the region was transformed into croplands. This long lasting process is very well described by the dominant soil types of the field. The ridges, shoulders, and steep upper slopes are dominated by Calcisols, where the topsoil have been eroded and the calcium carbonate rich loess and sand based parent material is exposed on the surface. On the more stable, and higher slope positions Luvisols or argic properties are dominating with very well developed reddish clay accumulation horizons. The lower, valley positions are mainly described by very deep humus rich horizons and layers, due to accumulation of eroded surface horizon materials. These soils can be as deep as 2 meters with colluvic material layered on top of the original, buried soil (Figure 2.).

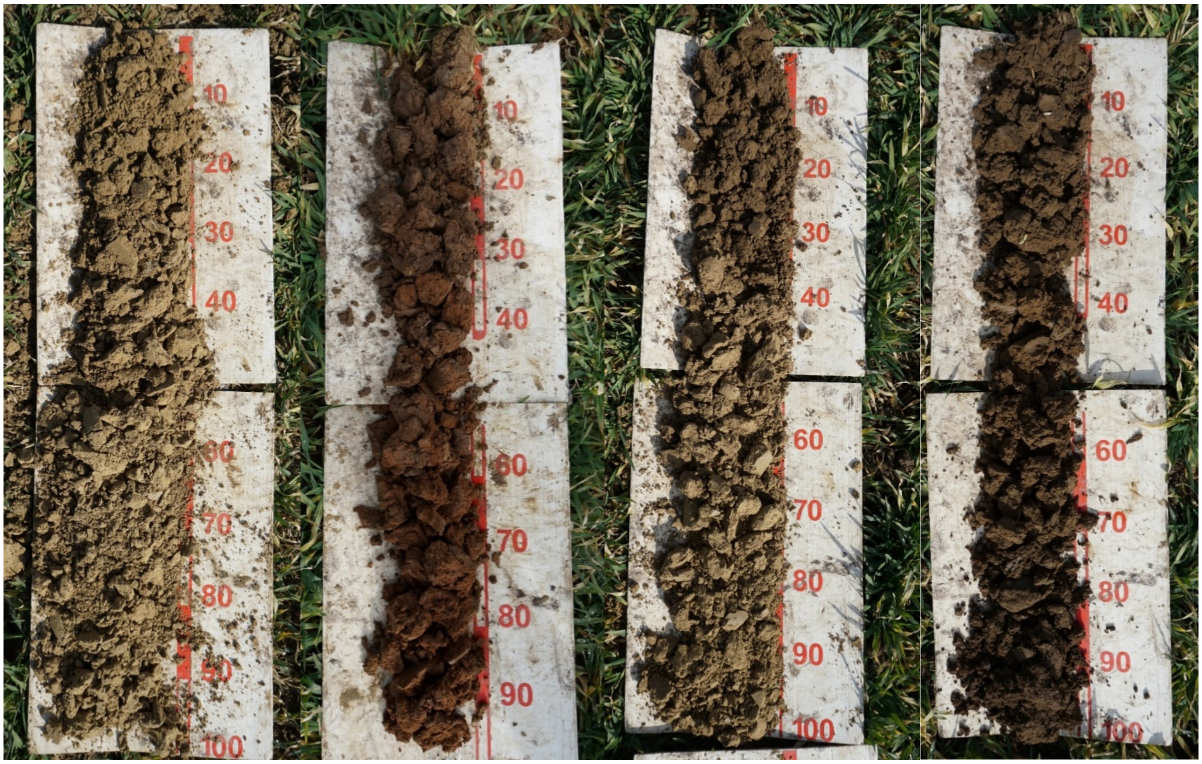

Figure 2. Dominant soil types of the study area (left to right: Calcisols, Luvisols, Kastanozems, Phaeozems (colluvic))

This heterogeneity very well describes, the larger region, where these highly eroded soils occur, thus the results of the evaluation of the tested methods can be extrapolated to larger regions, although cannot be used as a basis for tool selection under principally different soil and landscape conditions.

The selected soil survey methods were performed within a very small time period to minimize the error resulted by different soil moisture conditions. No relevant precipitation was observed between the surveys. Most of the survey methods were performed by the same research team, with the exception of the contact and non-contact soil scanning methods, which were performed by professional consultants. The main goal of the research was to evaluate the survey methodologies in relation with the soil properties, and yield, but as a connecting study the different agronomical created by the subcontractors were also evaluated to investigate the potential differences, resulted by the different tool selection.

The tested methods included contact electrical conductivity soil scanning, non-contact soil electromagnetic induction scanning, several years of yield mapping, high accuracy field soil survey using soil pits and auguring (soil classification), grid based soil sampling, remote sensing based vegetation indexes, digital elevation model derivatives, and a handheld soil scanner (Table 1.).

Table 1. Description of survey methods used in the study

Yield Mapping	Yield mapping was performed through a John Deere harvester in the years of 2017-2020.
EC Scanners	A contact EC scanner was used in this study, the scanning was performed with a swath width of 30 meters
EM Scanners	2 Non contact EM scanning were performed during different time periods, the same branded scanner was used, but the scanning was performed by different operators. Swath width was 30 meters for both surveys.
UAV-based DEM	A DJI Phantom 4 RTK was used to create a detailed DEM model of the area. The survey was performed after barley seeding, but before germination.
Satellite Images	ESA Sentinel 2. images were used to derive NDVI and MSBI indexes. NDVI indexes were derived from several periods, and were averaged after normalization of each image.
Intensive soil survey	An intensive soil survey was performed, with soil profile descriptions in every ¼ hectare to a a depth of 1 meter. For the profile descriptions the World Reference Base for Soil Resources (WRB, 2015) was used. Calcic, Cambic, Mollic, Ochric, Argic horizons were identified and described, along with colluvic soil material and surface soil horizon Munsell color, which was translated into RGB triplets for further analysis. Besides the field description laboratory analysis was performed for the following properties: pH (KCl), Organic matter, Plasticity, $CaCO_3$ content, Soil Electrical Conductivity, Total soluble salts, P_2O_5 content, K_2O content

To derive continuous property maps for the described survey methods, classification units and soil laboratory analysis Random Forest based predictions were used on a 5x5 meter grid, where the DEM based covariates and the satellite based MSBI index were used as covariates. These property maps were overlain by the yield maps, which were normalized and averaged through the 4 studied years to generate a yield potential variable, which was the main target of the random forest classification to identify the most relevant properties to describe the yield potential of the field. Random Forest classification was performed with and without using the DEM derivatives.

Results

The minimum, maximum and mean values along with the prediction variances of the Random Forest based property maps can be found in Table 2, along with some example property maps in Figure 3, respectively.

Table 2. Minimum, maximum, mean values of predicted soil property maps and the Prediction Variance of the Random Forest model

Property	Min	Max	Mean	Prediction Variance
R1 (EM scanning)	0	36,10766	21,29266	6,639361
R2 (EM scanning)	0	44,59372	25,50693	5,606096
R3 (EM scanning)	0	48,33708	24,12783	5,976343
R4 (EM scanning)	0	51,72342	28,84053	4,209277
pH (KCl)	5,4306	7,5745	7,044848	0,206307
Plasticity	31,32	39,85	35,4296	4,365624
Total soluble salts	0,02	0,0513	0,026935	8,25E-05
CaCO3 (%)	0,03	27,83	7,398576	24,22984
Organic matter (%)	0,9332	1,8898	1,477206	0,041335
P2O5 (mg/kg)	69,112	244,454	147,0772	1813,574
K2O (mg/kg)	120,9	297,81	196,0539	764,5057
Mollic (binary)	0	0,55	0,006439	0,005891
Ochric (binary)	0,07	1	0,837518	0,071142
Argic (binary)	0	1	0,352202	0,134097
Calcic (binary)	0,5	150	67,92322	1726,686
Cambic (binary)	0,1	1	0,916811	0,045467
Colluvic (binary)	0	0,95	0,127658	0,05465
red (RGB triplet)	91,9	201,52	113,4571	483,0083
green (RGB triplet)	64,57	190,15	85,38896	595,7352
blue (RGB triplet)	12,2	179,63	49,5157	869,4197
Shallow EC (contact scanner)	3,88537	161,4382	73,15394	101,3694
Deep EC (contact scanner)	134,5879	365,0972	271,1483	235,5286

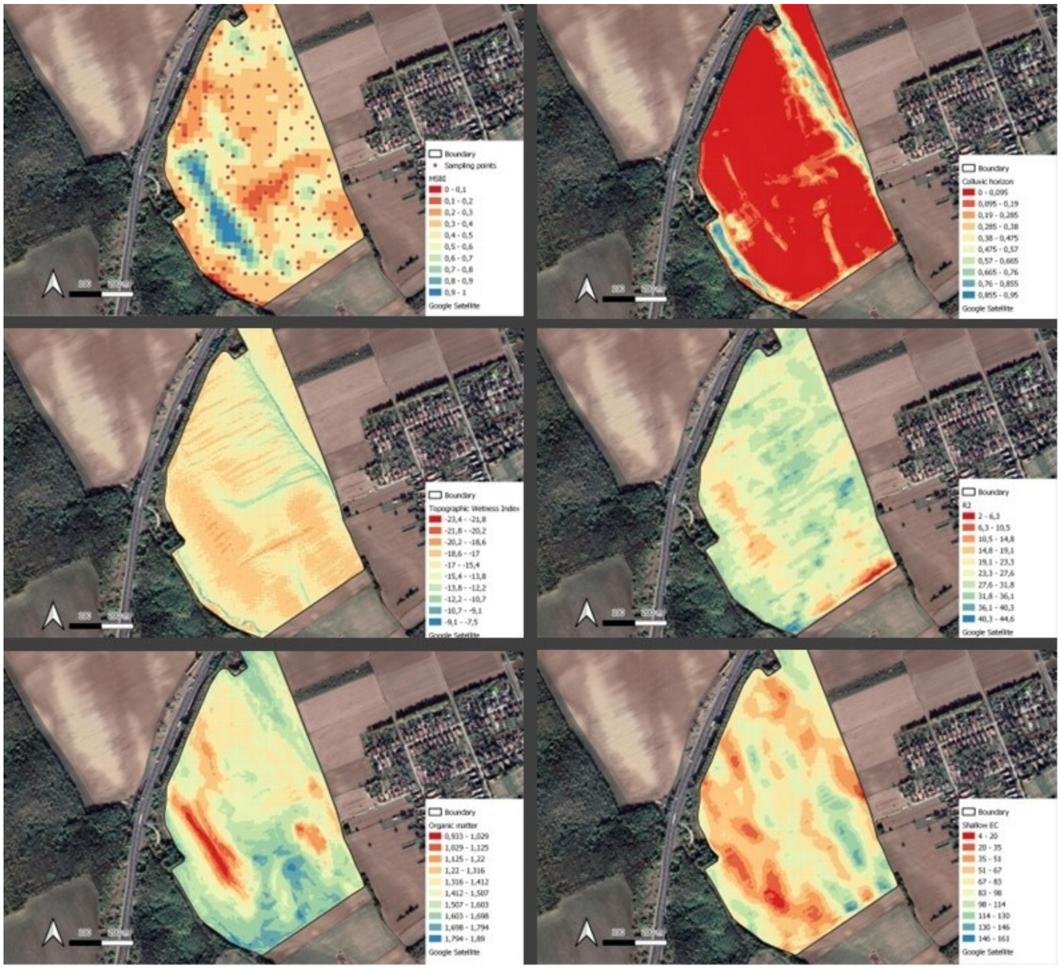

Figure 3. Property maps of the study area (left to right and top to bottom: MSBI index with soil sampling points, probability of colluvic material (0-1), TWI, R2 value (EM scanning), Organic matter (%), Shallow EC (contact EC scanner))

The classification of the Random Forest regression model indicated that the DEM derived topographic properties had a large effect on the yield potential, since out of the 10 main splitting properties 4 were from this group (slope, TWI and Channel Network, Relative slope position). Soil classification or field survey related information were also among the top relevant properties with colluvic material, Calcic horizon and topsoil color, leaving K_2O content and $CaCO_3$ content the only laboratory measured data appearing as main classifier. Interestingly, either the soil scanner nor the satellite based properties appear among the top classifiers.

In the classification scenario when Dem derivatives were left out of the training dataset, the top classifiers mentioned above, were completed pH, Ochric horizon, NDVI and MSBI. These results very well resonate with the results of the linear correlation where the above mentioned properties appear as the main correlating (negative or positive) properties.

Summary

At this stage of the analysis we have found that although all the methods have identified certain features on the field, many of these had no or minimal effect on the yield potential of the field zones, thus the usage of these tools or techniques have no or negative effect on the cropping system and on the environment. Modern soil classification diagnostic units, although in many cases neglected for PA systems, can provide a more detailed baseline information for decision support, through the aggregated soil chemical, physical and morphological information contained in one class. With modern digital soil mapping technologies and a custom made sampling methodology these tools are comparable in speed and provide deeper knowledge on the field than the 'modern' quick survey techniques. Since precision agriculture is an intensively expanding market and is supposed to be one of the main tools for sustainable agriculture a broad validation of these tools should be performed to identify the soil and climatic conditions where these can be properly used with providing the required precision and reliability.

Acknowledgements

This work was supported by the Establishment of EIP-AGRI innovation operational groups and supporting the investment needed to implement 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), under the project with the grant agreement No. 1862420362.

References

Avery B.W., 1973. Soil Classification in the soil survey of England and Wales, 336

Belgiu M., Dragut L., 2016. Random forest in remote sensing: A review of applications and future directions, 24, 27

Betts H. D., DeRose R. C., 1999. Digital elevation models as a tool for monitoring and measuring gully erosion, 91, 94, 99

Blackmore B. S., Marshall C. J., 1996. Yield Mapping; Errors and Algorithms, 403-404, 411, 413

Cambouris A. N., Nolin M. C., Zebarth B. J., Laverdière M. R., 2006. Management Zones Delineated by Electrical Conductivity to Characterize Spatial and Temporal Variations in Potato Yield and in Soil Properties, 390

D'Avello T. P., Indorante S. J., Bathgate J. D., Fitch B. C., McCauley W. M., Williams D. R., Wilson M. A., 2015. Application of SollM for a High-Intensity Soil Survey: A Southern Illinois Example, 3, 5

Dobos E., Bliss N., Worstell B., Montanarella L., Johannsen C., Micheli E., 2002. The use of DEM and satellite data for regional scale soil databases, 649-1, 649-2, 649-7, 649-9

Fleming K.L. and Westfall D.G., 2015. Evaluating Management Zone Technology and Grid Soil Sampling for Variable Rate Nitrogen Application, 179-180, 184

Gallego J. F., 2005. Stratified sampling of satellite images with a systematic grid of points, 369-370, 371-373, 375-376

Gallo B. C. B. , Demattê J. A. M., Rizzo R., Safanelli J. L., Mendes W. S., Lepsch I. F., Sato M. V., Romero D. J., Lacerda M. P. C., 2018. Multi-Temporal Satellite Images on Topsoil Attribute Quantification and the Relationship with Soil Classes and Geology, 2-3, 6, 13

Hunt M. L., Blackburn G. A., Carrasco L., Redhead J. W., Rowland C. S., 2019. High resolution wheat yield mapping using Sentinel-2, 1-4, 10-11

James I.T., Waine T.W., Bradley R.I., Taylor J.C., Godwin R.J., 2003. Determination of Soil Type Boundaries using Electromagnetic Induction Scanning Techniques, 421, 428, 429

Javadi S. H., Guerrero Castillo A. P., Mounem Mouazen A., 2022. Clustering and Smoothing Pipeline for Management Zone Delineation Using Proximal and Remote Sensing, 2, 6-7

Jurišić M., Plaščak I., Barač Ž., Zimmer D., 2016. Application of the principle of electroconductivity and the use of EC scanner for determining the properties of the soil, 1617-1621

Ließ M., Glaser B., Huwe B., 2012. Making use of the World Reference Base diagnostic horizons for the systematic descriptions of the soil continuum — Application to the tropical mountain soil-landscape of southern Ecuador, 20, 27-28

Mallarino A., Wittry D., 2001. Management Zones Soil Sampling: A better alternative to grid and soil type sampling, 160-163

Mallarino A., Wittry D., 2004. Efficacy of Grid and Zone Soil Sampling Approaches for Site-Specific Assessment of Phosphorus, Potassium, pH, and Organic Matter, 132-133, 141-142

Mudgal Y., 2020.. A Review of Microwave Remote Sensing Platforms and Sensors, 1-3

Mulders M.A., 1987. Remote sensing in soil science, 8-10

Nawar S., Corstanje R., Halcro G., Mulla D., Mouazen A. M., 2017. Delineation of Soil Management Zones for Variable-Rate Fertilization: A Review, 181, 189-191

Pouladi N., Bjørn Møllera A., Tabatabaia S., Greve M. H., 2019. Mapping soil organic matter contents at field level with Cubist, Random Forest and kriging, 86, 88

Rokhmana C. A., 2015. The potential of UAV-based remote sensing for supporting precision agriculture in Indonesia, 245-246, 248-250, 252-253

Smith M.J. and Pain C.F., 2009. Applications of remote sensing in geomorphology, 572, 574, 578-579

Viscarra Rossel R.A., Adamchuk V.I., Sudduth K.A., McKenzie N.J., Lobsey C., 2011. Proximal Soil Sensing: An

Effective Approach for Soil Measurements in Space and Time, 245-246
Wollenhaupt N.C. and Wolkowski R.P., 1994. Grid Soil Sampling, 6-7, 9
Wulf H., Mulder T., Michael E. S., Keller A., Jörg C. P., 2015. Remote Sensing of Soils, 2-3, 7-8, 11-12, 18, 23