Kurt Heil

School of Life Sciences, Forschungsdepartment Life Science Systems, Chair of Plant Nutrition, Technical University of Munich, Emil-Ramann-Str. 2, D-85350 Freising, Germany;

School of Life Sciences, Forschungsdepartment Life Science Systems, Chair of Organic Agriculture and Agronomy, Technical University Munich, Liesel-Beckmann-Str. 2, D-85354 Freising, Germany;

Correspondence: kheil@tum.de

A paper from the Proceedings of the
16th International Conference on Precision Agriculture
21-24 July 2024
Manhattan, Kansas, United States

Abstract.
In the current study, we used intensively collected information from soil profile analyses at the Dürnast site (Germany, 30 km in the north of Munich) for calculation and validation.
Based on the soil units of the overview soil map (1: 25000) and maps of topography, erosion (2 × 2 m each), and soil estimation (1: 5000), a sequence of sequential calculations was performed to derive soil units:
The application of existing soil maps, modeling of topography and erosion paths lead to plausible results. Larger spatial units are well identifiable, but smaller areas with different soils compared to the surrounding area can only be detected from small-scale measurements such as soil survey with EC_a and / or biomass survey (crop yield from harvester, drone or satellite imagery). However, the area-wide survey of vertical C_{org}, N_t and texture content requires a point-by-point assessment. Supervised classification random forest provided mainly topography as well as EC_a values as significant predictors for soil C_{org}, N_t and texture.

Keywords.
Digital soil mapping, Electical conductivity, Machine learning, Soil maps, C, N and texture

Detailed derivation of spatial soil attributes using soil sensor data, terrain analysis and soil maps with random forest classification

1. Introduction

Detailed knowledge of the spatial distribution of soils is critical for improved management and modeling in agriculture and forestry. However, information from existing soil maps is often not accurate enough and soil units are too large.

2. Materials and methods

2.1. Site description

The study area is approximately 2 ha and is located in Freising, 30 km north of Munich, Germany (4477221.13 E, 5362908.78 N), in a hilly, Tertiary landscape. According to the German Soil Survey (Bodenkundliche Kartieranleitung, 2005), Eutric Cambisol (Siltic, Aric)**,** Eutric Stagnic Cambisol (Siltic, Aric) and Eutric Cambisol (Loamic, Aric).

2.2. Geophysical survey, topographical parameters and soil sampling

The EM38-MK2 (Geonics) was mounted on a sledge, covered with a plastic cap and pulled by a tractor. A Trimble AG132 DGPS system (Trimble Navigation Ltd., Sunnyvale, CA) with an accuracy of 1 m or less was used to georeference the EC_a (vertical 1.0 m and 0.5 m, horizontal 1.0 m and 0.5 m). Besides the height, different primary and secondary complex relief attribute parameters were calculated with the software package System for Automated Geoscientific Analyses (SAGA, produced by Scilands GmbH Gottingen, www.scilands.de). From the soil evaluation the parameters, “soil water household” and “condition level” are used.
Sixty-four soil core samples were collected at depths 0-25 cm, 25-50 cm and 50-75 cm at different locations based on the EC_a-maps.

2.3. Random forest approach

The random forest regression (RFM) calculation is a non-parametric technique (Breiman 2001) as a continuation of the Classification and Regression Trees (CART) program with the target to improve the prediction performance of the model.
The RFMs were carried out using the package random forest in R (R Development Core Team, 2007). After Liaw and Wiener (2002) three parameters must be defined: the number of trees (ntree), the number of variables used per tree (mtry) as well as the minimum amount of data per terminal node (nodesize).

3. Results

3.1. Derivation C, N and texture

Results with the relative importance of the environmental covariates are shown in Table 1, in which the R^2 and the RMSD can be observed. It is striking that the R^2 values of C_{org} and N_t increase significantly with increasing depth. This is certainly partly due to the larger number of predictors in the subsoil. The most important cause, however, is probably the higher values in the subsoil. The results show that the best explanatory variables for the C_{org}- and also N_t-contents modelling are the EC_a-readings in combination with more area-related relief values (catchment area, valley depth, curvature). The dominating factors are catchment and elevation. The derivations of soil texture deliver clear results (not shown). Clay and silt are modeled in an excellent way with R^2 higher than 0.74 and RMSD mainly lower than 3%. In contrast to these results the sand models indicate no usable calculations. The main reason here is the low variance of the sand content. This prevents relationships with the covariates. The highest importance show the EC_a readings in the clay and silt models. The validation dataset (calculated with the same predictors as calibration) deliver similar gradations with sufficient models.

Table: 1 Models for the derivation of C, N and texture with Random forest

Depth	Target variable	Predictors	Sig.	Rel. importance	RMSD	Adj. R^2 sig.	RMSD	Adj. R^2 sig.
					Calibration		Validation	
0-25 cm	C [%]^3	1/Elevation	***	43.4	0.07	0.451***	0.16	0.22
		Valley depth	**	39.2				
		LS-factor	**	17.4				
	N [%]^3	1/Catchment area	***	25.6	0.006	0.445***	0.019	0.24*
		EC_a(h-05^3)	***	44.6				
		Valley depth	*	29.8				
25-50 cm	C [%]^3	1/Catchment area	***	43.6	0.12	0.71***	0.276	0.55***
		Catchment area^3	***	24.8				
		EC_a(v-10^2)	**	18.9				
		Profilecurvature	*	12.6				
	N [%]	Elevation^2	***	51.9	0.01	0.681***	0.053	0.56***
		EC_a(√h-05)	***	31.7				
		Profilecurvature^2	*	6.1				
		Valley depth	*	10.2				
50-75 cm	C [%]^2	1/Catchment area	***	54.87	0.19	0.962***	0.44	0.76***
		EC_a(v-10)	***	24.94				
		Profilecurvature^3	***	12.4				
		Plancurvature^3	**	7.8				
	N [%]^2	1/Catchment area	***	58.4	0.012	0.961***	0.09	0.76***
		1/Elevation	***	21.5				
		Plancurvature^3	***	11.6				
		Log10(Slope)	**	8.4				

4. Conclusion

On this rather small area, the covariates used have led to very good results (except for sand) for the selected target parameters. Soil evaluation and soil maps are not necessary for a prediction here. This is now the base for the aggregation of soil properties that lead to the separation of soil science and crop production sub-areas. If larger areas are surveyed, additional covariates are included in the calculations.

5. References

Arbeitsgruppe Bodenkunde, (2005). Bodenkundliche Kartieranleitung. 5. Verbesserte und erweiterte Auflage (*Herausgegeben von der Bundesanstalt für Geowissenschaften und Rohstoffe in Zusammenarbeit mit den Staatlichen Geologischen Diensten der Bundesrepublik Deutschland). Schweizerbart'sche Verlagsbuchhandlung, Hannover.*

Breiman, L. Technical report for Version 3. (2014 2001). Available in: http//oz.berkeley.edu/users/breiman/randomforest2001.pdf (Accessed at: 12/28/14).

Heil, K., Schmidhalter, U., (2015). Comparison of the EM38 and EM38-MK2 electromagnetic induction-based sensors for spatial soil analysis at field scale. *Comput. Electron. Agric*. 110, 267–280.

Liaw, A., Wiener, M., (2002). Classification and regression by randomForest. *R News* 2 (3), 18–22