UTILIZATION OF SOIL-PLANT INTERRELATIONS THROUGH THE USE OF MULTIPLE REGRESSION AND ARTIFICIAL NEURAL NETWORK IN ORDER TO PREDICT SOIL PROPERTIES IN HUNGARIAN SOLONETZIC GRASSLANDS

Tibor TÓTH[1] – Marcel G. SCHAAP[2] – Zsolt MOLNÁR[3]

[1] Research Institute for Soil Science and Agricultural Chemistry of the Hungarian Academy of Sciences, Budapest, Hermann O. 15, tibor@rissac.hu
[2] Department of Soil, Water and Environmental Science, University of Arizona, Tucson, Arizona, USA
[3] Ecological and Botanical Research Institute of the Hungarian Academy of Sciences, Vácrátót, Botanical Garden

Abstract: Soil and plant interrelations are strong enough in semi-natural solonetzic grasslands to permit the use of plant cover as predictor variable for soil salinity, sodicity and alkalinity. Four data sets were analysed which covered 4-7 plant association types, with sample sizes ranging from 20 to 120 and quadrat sizes 0.16 to 20 m^2; and correlation coefficients (R) of the multiple regression equations established between plant cover (independent or predictor variables) and soil (dependent or predicted variables) usually ranged from 0.65 to 0.80. Utilization of neural networks improved the prediction further and provided typically R values of 0.8. Plant cover observations consequently can be used to improve the precision of numerical maps of soil properties on solonetz soils and to delineate risk areas more precisely faster at a lower cost.

Keywords: plant cover, solonetz soil, salinity, alkalinity, pH, sodicity, salt-affected soil

Introduction

The semi-natural vegetation of solonetz soils shows close correlation with the soil properties (Rajkai et al., 1988, Zalatnai and Körmöczi, 2004, Ristolainen et al., 2006, Tóth et al., 2006). Since it is much easier to recognize the vegetation type *in situ* than to determine chemical properties of soil in the laboratory, use of that correlation for predicting soil properties is desired. Such use of the plant - soil interrelationship can be applied effectively in soil survey, as suggested early by Ballenegger, 1929. Isoline maps compiled on the basis of numerical values are nowadays considered to be more informative than the polygon-maps, especially in the case of soils whose fertility is limited by a few, known factors. The empirical-stochastic relation between soil properties (dependent or predicted variable) and plant cover percents (independent or predictor variables) in a location can be established and utilized for the quantitative estimation of the soil property by regression analysis (Webster, 1989). The initiatives of European Union for protecting the soils by the Soil Thematic Strategy emphasizes the prevention of soil salinization as one of six soil threats. For efficient numerical mapping ideal tools are salinity maps with equal intervals. Our previous research showed that plant categories are suitable for distinguishing not very high levels of soil salinity both in the deepest and highest elevation zones. Our hypothesis was that plant species cover is suitable predictor of salinity in the most saline, intermediate elevation zones. Therefore, our objective was to establish multiple regression equations between soil properties and plant cover percents and to interpret the precision provided by the equations.

Materials and methods

We evaluated published data of Bodrogközy (1965) and Magyar (1928) and two data sets of ours. All the observations have been made in the north-eastern part of the Great Hungarian Plain, in the regions of Hortobágy and Nagykunság

In the calculation of regression equations, those plant species were selected as independent variables which were dominant in the coenological quadrats. As dependent variable of the regression equation, ecologically important variables, which are related to salinization, sodification and alkalization, were selected. All plant cover data, that is all independent variables have been transformed with logit (ln x/(100-x)), but the transformation is not indicated in the equations. The method of variable selection in the multiple regression analysis was stepwise selection.

Results and discussion

In our study site at Kisszeg, Hortobágy the capacity of the vegetation (associations as described by Molnár and Borhidi, 2003) to predict distinct ranges of was very much limited compared to the results of Tóth and Kertész, 1996, who worked in a more salt-affected area. Our special focus now is on the information provided by the cover of indivial plant species.

Table 1 summarizes the result of the regression equations and neural networks. In this table only those plants are indicated which were selected to be included in the multiple regression equations by the stepwise selection method.

The studied cases covered 4-7 solonetzic grassland associations. The number of quadrats fell between 20 and 120, the size of quadrat changed between 0.16 and 20 m^2. The depth of sampling ranged from 0-3 to 30-40 cm. The plant cover was expressed as rating from 1 to 5 (Bodrogközy, 1965) and with percents in the case of other data sets. All four cases covered the plant associations of the microerosional complex "padkásszik", (Tóth and Rajkai, 1994), that is the plant associations **Puccinellietum limosae - Camphorosmetum annuae - Artemisio-Festucetum pseudovinae**, therefore the most efficient predictor plant covers were the same in each case. These were the plants, which tolerate the highest surface soil salt concentration and pH, the *Camphorosma annua* (CAM). and *Puccinellia limosa* (PUC). The total number of plants occurring in the quadrats (SNO) and the total plant cover (TOCO) were also often included in the regression equations.

The value of the multiple correlation coefficients fell mostly in the range of 0.65-0.80, and this indicated that the equations can be used for the estimation of soil properties, contrary to that the distribution of the variables is often non-normal. For the study in which the largest number of observations was available, the "Padkásszik" case study, there was a further analysis with combined artificial neural network and bootstrap as described by Efron and Tibshirani (1993), since neural networks can find complex relations between the input data and output parameters.

Conclusions

When isoline map of properties of solonetz soils is prepared, the number of soil sampling points required to meet a prefixed precision can be reduced if the information provided by plant cover data is utilized. In some points the soil sampling must be

Table 1
Multiple regression equations between soil properties and plant cover percents in solonetzic grasslands in the north-eastern part of the Great Hungarian Plain. The correlation coefficients (R) in italic and bold indicate correlation with neural network model for the „Padkásszik" dataset. Minimum (m) and maximum (M) values of soil properties of the depth layers are shown in italic. Plant assocition names are shown in bold.

Source of data	Bodrogközy,1965	Magyar, 1928	„Nagykunság"	„Padkásszik"
Number of quadrats	29	34	20	120
Maximum distance (km)	10	30	30	0.06
Area of quadrat (m^2)	20	20	0.16	0.25
Soil depths (cm)	0-40 by 10	0-40 by 20	0-3	0-5 & 10-15
Agrosti-Beckmannietum	4			
Agrosti-Alopecuretum	2		3	3
Pholiuro-Plantaginetum	2			
Puccinellietum	5	+	2	30
Camphorosmetum	3	+	2	4
Artemisio-Festucetum	8	+	6	83
Achilleo-Festucetum	5	+	7	
Agrosti-Eleochari-Alopecuretum		+		
Predicted property (Dependent variable)	Plants in the regression equations (Independent variables)			
Salt % minimum/Max	*m0.02M0.7*	*m0M1.1*	*m0.01M0.7*	*m0.01M0.65*
Salt % at surface	CAM.AGR	PUC	TOCO	CAM.SNO.PUC.REL
R	0.63	0.62	0.62	0.69 ***(0.78)***
Salt % min/Maximum	*m0.05M1.1*	*m0.04M1.7*		*m0.09M1.3*
Salt % in 0-20 cm	CAM	CAM		CAM.SNO.REL.PUC
R	0.69	0.46		0.77 ***(0.83)***
Soda% min/Max	*m0M0.15*			
Soda % at surface	CAM			
R	0.76			
Soda% min/Max	*m0M0.3*	*m0M0.23*		
Soda % in 0-20 cm	TOCO.CAM	CAM		
R	0.83	0.65		
Na meq/100 g min/Max				*m2M23*
Na at surface				T.AR.CAM MOS.RAU
R				0.70 ***(0.8)***
Na meq/100 g min/Max				*m5M33*
Na in 10-15 cm				CAM.RAU SNO.PUC
R				0.69 ***(0.77)***
pH min/Max			*m5.2M8.5*	*m6.3M9.98*
pH at surface			TOCO.AR. POL.ACH	PUC.CAM.NOS. SNO.TOCO.REL
R			0.89	0.73 ***(0.78)***
pH min/Max				*m8.51M10.6*
pH in 10-15 cm				CAM.REL. SNO.PUC.F
R				0.75 ***(0.72)***

coupled with plant coenological survey. If the soil properties determined in the sampling points and the cover percents of the plant species show strongly significant multiple regression equations, then the values of soil properties can be estimated in the remaining points with multiple regression equations.
The calculated multiple regression equations and neural network algorithms proved to be not comprehensive. It means that the values of regression coefficients probably depend on the range of plant associations covered, season of year, location, etc., therefore the equations intended for mapping of soil properties should be calculated for every mapping situation separately (Tóth and Rajkai, 1994). Two favourable conditions were noted that improved the correlation coefficient of the multiple regression equations, such as the utilization of comprehensive survey (this way covering wide range of soil salinity as shown in Table 1) as in the case of Bodrogközy, (1965) and the previous averaging of the plant cover percents ("Nagykunság" data set). Since the three basic chemical characteristics of solonetz soils, that is salinity, sodicity and alkalinity are very closely correlated, there was very similar precision obtained during their prediction. Based on the proven precision of the algorithms, the soil-plant interrelations can be made operative and the techniques shown can be incorporated in the routine soil mapping and risk area identification of solonetzic areas.

Acknowledgements

Authors acknowledge the financial support provided by the project NKFP6-00013/2005 and the Sixth Framework Project No.4420 "RAMSOIL"

References

Ballenegger, R. 1929. Les méthodes de la cartographie des sols alcalins (salins).pp. 9-12. in Transactions of the Alkali-Subcommision of the International Society of Soil Science.

Bodrogközy, Gy. 1965. Ecology of the halophilic vegetation of the Pannonicum.. II. Correlation between alkali ("szik") plant communities and genetic soil classification in the Northern Hortobágy. Acta Bot. Hun. 11,1-51.

Efron, B. and Tibshirani, R. .J. 1993. An introduction to the bootstrap. Chapman & Hall.

Magyar, P. 1928. Data to the plant coenological and geobotanical characterization of Hortobágy. Erdészeti. Kisérletek. **30:**210-226. (in Hungarian)

Molnár, Zs. and Borhidi, A. 2003. Hungarian alkali vegetation: Origins, landscape history, syntaxonomy, conservation. Phytocoenologia. **33:**377-408

Rajkai, K., Marchand, D. and Oertli, J. J. 1988. Study of the spatial variability of soil properties on alkali soils. pp 150-155. Proceedings of the international symposium on solonetz soils problems properties and utilization. Osijek, 1988. June 15-21.

Ristolainen, A., T. Tóth, Cs. Farkas. 2006. Measurement of soil electrical properties for the characterization of the conditions of food chain element transport in soils. Part I. Instrumental comparison. Cereal Research Communications. **34**:159-162.

Tóth, T. and Kertész, M. 1996. Application of soil-vegetation correlation to optimal resolution mapping of solonetzic rangeland. Arid Soil Research and Rehabilitation. **10:**1-12.

Tóth, T. and Rajkai, K. 1994. Soil and plant correlations in a solonetzic grassland. Soil Science **157:**253-262.

Tóth, T., A. Ristolainen, V. Nagy, D. Kovács, Cs. Farkas. 2006. Measurement of soil electrical properties for the characterization of the conditions of food chain element transport in soils. Part II. Classification of management units. Cereal Research Communications. **34**: 163-166.

Webster, R. 1989. Is regression what you really want? Soil use and management. **5**:47-53.

Zalatnai, M. and Körmöczi, L. 2004. Fine-scale pattern of the boundary zones in alkaline grassland communities. Community Ecology. **5:**235-246.