Mapping the Topsoil pH and Humus Quality of Forest Soils in the North Bohemian Jizerské hory Mts. Region with Ordinary, Universal, and Regression Kriging: Cross-Validation Comparison

Radim VAŠÁT, Lenka PAVLŮ, Luboš BORŮVKA, Ondřej DRÁBEK and Antonín NIKODEM

Department of Soil Science and Soil Protection, Faculty of Agrobiology, Food and Natural Resources, Czech University of Life Sciences Prague, Prague, Czech Republic

Abstract

Vašát R., Pavlů L., Borůvka L., Drábek O., Nikodem A. (2013): **Mapping the topsoil pH and humus quality of forest soils in the North Bohemian Jizerské hory Mts. region with ordinary, universal, and regression kriging: cross-validation comparison**. Soil & Water Res., **8**: 97–104.

North Bohemia belongs to one of the most heavily industrialized and polluted regions in Europe. The enormous acid deposition which culminated in the 1970s has largely contributed to the accelerated acidification process in the soils and consequently to the wide forest decline in North Bohemian mountains. In this paper we map the active topsoil pH and humus quality with ordinary, universal, and regression kriging and compare the accuracy of resulting maps with cross-validation. For the regression kriging we use two types of spatially exhaustive auxiliary information, first the altitude derived from digital elevation model and second the land cover classes derived from satellite imagery. The leave-one-out (cross-validation) statistics, i.e. mean error, root mean squared error, and mean squared deviation ratio, are taken for comparison since they are widely accepted as measurements of the accuracy of digital soil maps. The results show that the regression kriging is superior over other kriging methods in this case. Out of 97 sampling sites the regression kriging with land cover classes is the best predictor at 32 sites for pH and at 30 sites for humus quality, the regression kriging with altitude at 31 and 25 sites, the universal kriging at 21 and 23 sites, and the ordinary kriging at 13 and 18 sites. The highest number of best predictions for regression kriging implies that the topsoil pH and humus quality are driven approximately equally by land cover and altitude and little less by pure geographic position. Furthermore, the universal kriging maps show a northeast to southwest spatial trend of topsoil pH and a northwest to southeast spatial trend for humus quality.

Keywords: Black Triangle; digital soil mapping; geostatistics; map accuracy

North Bohemia as a part of the so-called Black Triangle (an area where Poland, Germany and Czech Republic come together) falls into one of the most heavily industrialized and polluted regions in Europe. In the 1970s and 1980s the electric power stations and district heating plants concentrated in this area were producing probably the highest quantity of SO_2 and NO_2 emissions in the world (Hruška & Ciencala 2003). Obviously, rising emissions strongly affected all parts of the ecosystem, including air, water, soil and living organisms. The enormous acid deposition largely contributed to the accelerated acidification process in the soils which resulted, among others, in a wide forest decline in North Bohemian mountains in 1980s (Akselsson *et al.* 2004). The soil pH decreasing, lowering of base saturation, Al mobilization, depletion of base cations, development of poor quality humus material, deceleration of the decomposition process, accumulation of raw organic material are only few of the main symptoms of this acidifying degradation process (Mládková *et al.* 2004; Borůvka *et al.* 2005; Nikodem *et al.* 2010).

To access the current state of acidification, digital soil maps might be of a great help. Besides using only point measurements as input for spatial prediction, one can possibly enhance the map accuracy involving available spatially exhaustive auxiliary information such as digital elevation model (DEM) and its derivatives, land cover classes, soil type classes, geological maps, etc. For this purpose so-called hybrid interpolation methods which combine two conceptually different approaches to modeling and mapping spatial variability were developed: (a) interpolation relying solely on point observations of the target variable and (b) interpolation based on regression of the target variable on spatially exhaustive auxiliary information. The power of this hybrid interpolation techniques was shown previously in several studies (e.g. Knotters *et al.* 1995; Bishop & McBratney 2001; Leopold *et al.* 2005; Lopez-Granados *et al.* 2005;), while its advantage over other interpolation methods is especially evident for contrasting landscapes (Zhu & Lin 2010). One of these hybrid interpolation methods is known as regression-kriging (RK) (Odeh *et al.* 1995; Hengl *et al.* 2004). It first uses regression on auxiliary information and then it uses simple kriging (SK) with known mean (equals to 0) to interpolate the residuals from the regression model. The comprehensive description of the regression kriging methods is offered in Hengl *et al.* (2007). To quantify the accuracy of the resulting maps, the cross-validation procedure, also known as leave-one-out, can be successfully applied in situations where no extra validation data are available. Consequently, the mean error (ME), the root mean squared error (RMSE), and the mean squared deviation ratio (MSDR) are computed since they are widely accepted as a measure of accuracy of resulting digital soil maps (Brus *et al.* 2011).

In this paper we map the spatial distribution of active topsoil pH and humus quality in the Jizerské hory Mts. region. We profit from available spatially exhaustive auxiliary information, i.e. from the altitude derived from DEM and the land cover classes derived from satellite imagery. The regression kriging (RK) predictions are compared with universal (UK) and ordinary kriging (OK). To quantify the accuracy of resulting maps and to compare different kriging methods the cross-validation statistics is computed.

Although the UK is mathematically equal to RK (or *vice versa*), many authors (Deutsch & Journel 1998; Wackernagel 1998; Papritz & Stein 1999) reserve the term universal kriging for the case where the trend is modeled as a function of coordinates. To make clear the difference between RK and UK as a difference in input parameters we also stick to this definition.

MATERIAL AND METHODS

Description of the study area and sampling strategy. The Jizerské hory Mts. region covering an area of about 368 km^2 is situated in the north of the Czech Republic, close to the border with Poland and Germany (Figure 1). The altitude ranges from 300 to 1124 m a.s.l. Annual precipitation attains to approximately 1700 mm and the average annual air temperature is 6°C. Like in the case of the whole Czech Republic, west winds predominate here.

To perform the survey, a total of ninety-seven Jizerské hory Mts. sites dislocated to cover up the whole area were visited within a sampling campaign in 2004 (Mládková *et al.* 2004; Borůvka *et al.* 2005). The allocation of all the 97 sampling sites is shown in Figure 2. Vegetation cover is formed mainly by Norway spruce monoculture (*Picea abies* [L.] Karst.) and partly by acidophilic European beech (*Fagus sylvatica* L.) accompanied by dominating *Calamagrostis arundinacea* and *Calamagrostis villosa* in the herbal layer. The highest

Figure 1. Location of the study area

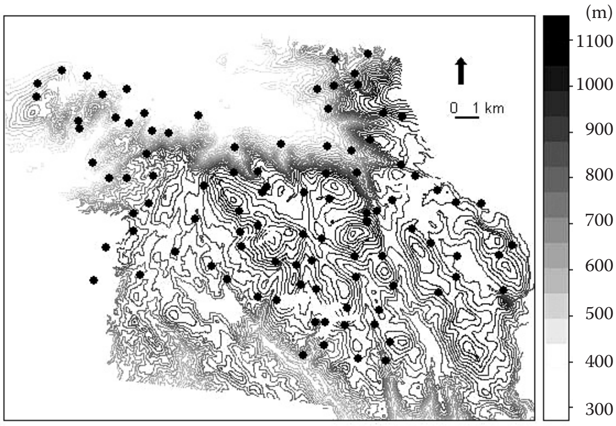

Figure 2. Allocation of 97 sampling sites with contours (grey lines)

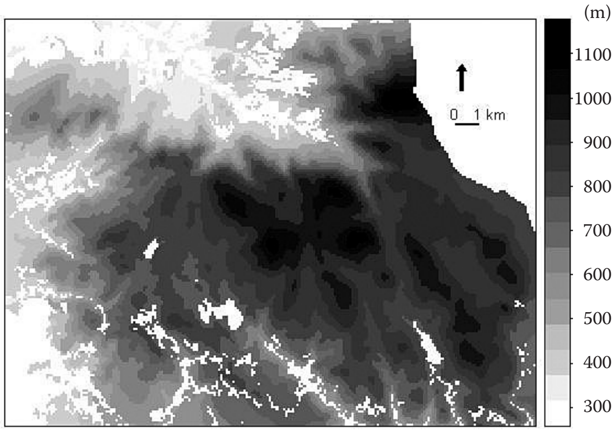

Figure 3. Altitudes of the Jizerské hory Mts. region in m a.s.l.

parts of the area (over 800 m a.s.l.) are to a large extent covered mainly with grass (*Calamagrostis villosa*) with a sparse occurrence of young spruce and rowan (*Sorbus aucuparia* [L.]) as a result of historical wide forest decline (Borůvka *et al*. 2005). More detailed information on the spatial coverage of different land cover types is offered in Figure 4. Soils were classified referring to the World Reference Base for Soil Resources (WRB 2006) as Podzols in most cases (Haplic or Entic) and Cambisols (mainly Dystric). All soils are developed on granite bedrock, and hence the effect of geology on soil characteristics spatial variation can be neglected (Borůvka *et al*. 2005).

The active topsoil pH was measured potentiometrically in the soil leachate (10 g of dried soil material with 20 cm^3 H_2O and 5 min shaking before measured). The humus quality was assessed by the ratio of absorbances of pyrophosphate soil extract at the wavelengths of 400 and 600 nm (A_{400}/A_{600}) (Pospíšil 1981).

Further in the study, two types of spatially exhaustive auxiliary information available for the study area, i.e. the altitude derived from DEM (Figure 3) and land cover classes derived from satellite imagery (Figure 4), are used. On the land cover map the spruce category represents areas with spruce monoculture (*Picea abies* [L.] Karst.) aging around 100 years which replaced the original vegetation. The beech category is represented by the natural beech wood (*Fagus sylvatica* L.) which has survived to these days on limited protected spots. The meadow category is represented by grasslands where haymaking is made regularly once a year. Areas mostly covered with peat where the water level attains close to or above the surface throughout the year fall in with the wetland category. The low-vegetation category occurs in areas of declined forest, former spruce monoculture areas mostly covered with grass (*Calamagrostis arundinacea* and *Calamagrostis villosa*) and partially with young vegetation of spruce and rowan (*Sorbus aucuparia* [L.]).

Ordinary, universal, and regression kriging prediction. All computations, graphs, and figures are done using the R software tool for statistical computing (R Developement Core Team 2011), which is a very rich and widely available environment (freeware). For kriging prediction we use the gstat package (Pebesma 2004), which is suitable for a wide range of geostatistical computations including variogram modeling, ordinary, universal, regression, kriging and co-kriging methods and their derivatives, kriging with external drift (KED), and many others. We also use packages maptools, sp and RColorBrewer. For the theory

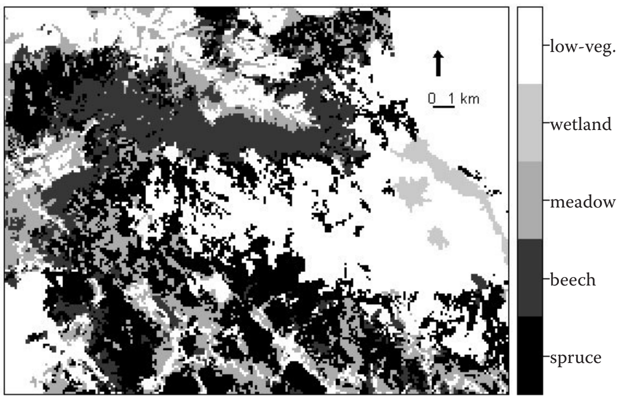

Figure 4. Land cover classes of the Jizera Mountains region (Landsat 7)

about the family of kriging methods we refer to Hengl *et al.* (2007).

Map accuracy, cross-validation, ME (bias), RMSE, and MSDR. To measure the accuracy of final maps we use cross-validation (also known as leave-one-out) procedure and consequently we compute root mean squared error – RMSE (Eq. 1), mean error – ME (Eq. 2) and mean squared deviation ratio – MSDR (Eq. 3):

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{i=1}^{n}(x_{1,i}-x_{2,i})^2}{n}} \qquad (1)$$

$$\mathrm{ME} = \frac{\sum_{i=1}^{n}(x_{1,i}-x_{2,i})}{n} \qquad (2)$$

$$\mathrm{MSDR} = \frac{\sum_{i=1}^{n}\left[\frac{(x_{1,i}-x_{2,i})^2}{\mathrm{var}_i}\right]}{n} \qquad (3)$$

where:
x_1 – prediction of the variable X
x_2 – measure of that variable
n – number of records
var – kriging variance

The smaller the RMSE and ME values, the more accurate the map. The MSDR says if the variance of measurement data is well reproduced with the kriging interpolation and ideally it equals to 1.

RESULTS AND DISCUSSION

Modeling the spatial auto-correlation (variograms and residual variograms)

In the case of topsoil pH the variogram and residual variograms indicate a moderately strong spatial auto-correlation (Figure 5), while there is not apparent trend, either with land cover classes (Figure 5a), altitude (Figure 5b) or with geographic position (Figure 5c). A bit stronger spatial auto-correlation is found for humus quality, but even here there is not apparent trend with any of auxiliary information since all variogram models have the basic parameters highly similar (Figure 6).

Spatial prediction (kriging)

In the case of topsoil pH the spatial patterns of regression kriging with land cover classes (RK-LC), regression kriging with altitude (RK-ALT), UK, and OK maps are very similar (Figure 7). This is not surprising since there is no apparent trend in the pH data either with direction (UK) or with altitude (RK-ALT) or land cover classes (RK-LC). Apparently, higher pH values occur at higher altitudes in the centre and in the southeast part

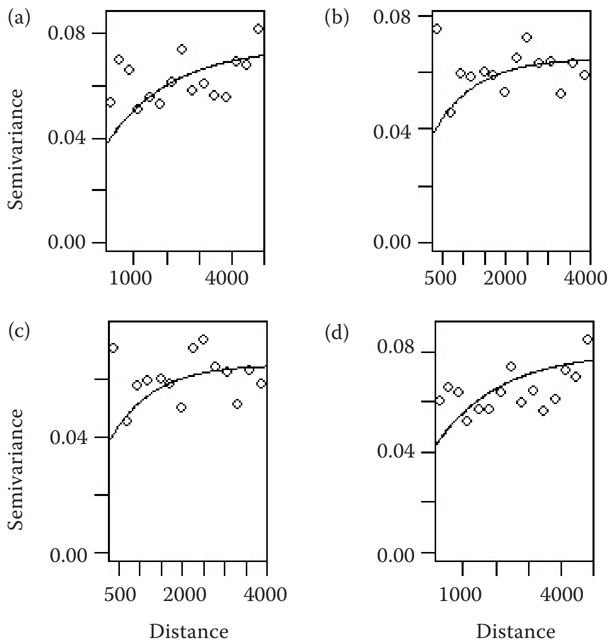

Figure 5. Variogram (d) and residual variograms with land cover classes (a), altitude (b), and geographic position (c) of topsoil pH with fitted models

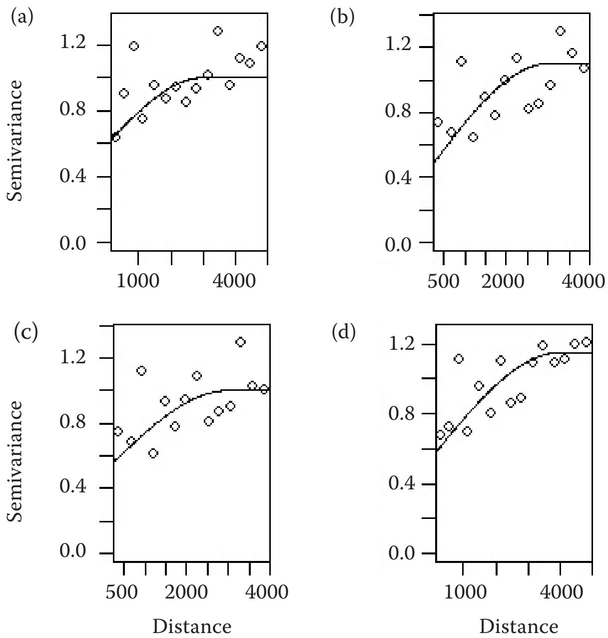

Figure 6. Variogram (d) and residual variograms with land cover classes (a), altitude (b) and geographic position (c) of topsoil humus quality with fitted models

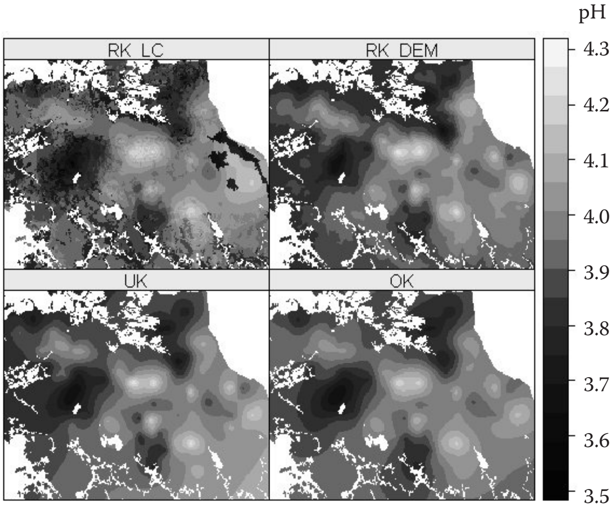

Figure 7. Kriging maps of topsoil pH: RK-LC – regression kriging with land cover classes; RK-ALT – regression kriging with altitude; UK – universal kriging; OK – ordinary kriging

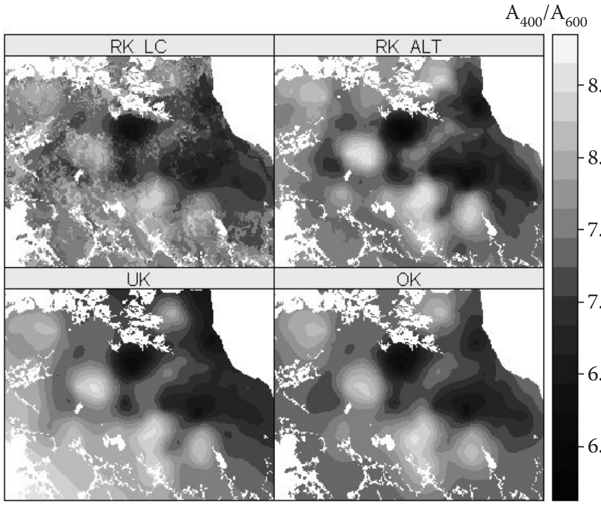

Figure 8. Kriging maps of humus quality: RK-LC – regression kriging with land cover classes; RK-ALT – regression kriging with altitude; UK – universal kriging; OK – ordinary kriging

of the area (here altitude ranges 700–1200 m), where the land cover is described as grassland with sparse occurrence of young spruce and rowan (former death forest). Lower pH values occur at lower altitudes in the northwest part of the area (here altitude ranges 300–700 m) with dominance of Norwey spruce monoculture. An exception is the natural beech wood in the south-west, where, despite of lower altitudes, the topsoil pH is significantly higher than that of the surroundings. These findings largely coincide with results of Mládková *et al.* (2004) and Borůvka *et al.* (2005) and support the theory of lower acid deposition at the highest altitudes caused by the limited grass canopy. Another interesting point is that the UK map offers a presumption of the northwest to southeast spatial trend (topsoil pH increases in this direction), which coincides with the direction of dominating winds. In summary, throughout the territory the topsoil is extremely acidic as shown by kriging predictions, which range 3.5–4.3. And indeed, the forested areas of the Czech Republic, Poland and the Netherlands belong to the most acidified in Europe and will require an exceptional time to recover (Reinds *et al.* 2009).

Focused on humus quality, the spatial pattern of all kriging maps is again largely similar (Figure 8). Soils with better quality of humus material (lower A_{400}/A_{600} values) occur in the centre and in the northeast part of the area at higher altitudes where the vegetation is formed mainly by grass (former dead forest) with sparse occurrence of young spruce and rowan. On the other hand, worse humus quality (higher A_{400}/A_{600} values) occurs in the west and south part of the area at lower altitudes where the vegetation cover is described as spruce monoculture and meadows with a small area of beech in the northwest. In general, the quality of humus material is very low across the whole area. For illustration, fertile soils like chernozems have the A_{400}/A_{600} value close to 1 (somewhere between 1.1 to 1.4), while poor soils like podzols usually exceed 4.6 (Pospíšil 1981).

Leave-one-out (cross-validation) statistics and the best prediction count

In the case of topsoil pH, the RMSE indicates that the UK and RK-ALT predictions are more accurate than predictions made with OK and RK-LC. However, the ME is lower for OK and RK-LC while it is slightly higher for UK and RK-ALT. The MSDR, which is supposed to ideally equal to 1, is the highest for RK-ALT and equals to 0.985. Although, it must be noted, the differences are very small in all cases (Table 1).

For humus quality lower RMSE values belong to UK and OK predictions (0.989 and 0.990 respectively), while a bit higher RMSE values are linked with RK-ALT (1.018) and RK-LC (1.031) predictions. Similarly,

Table 1. Root mean squared error (RMSE), mean error (ME), and the mean squared deviation ratio (MSDR) statistics for topsoil pH and humus quality kriging cross-validation

	Statistics	RK-LC	RK-ALT	UK	OK
pH	RMSE	0.242	0.240	0.240	0.242
	ME (bias)	–0.002	–0.003	–0.003	–0.002
	MSDR	0.976	0.985	0.980	0.923
Humus quality	RMSE	1.031	1.018	0.989	0.990
	ME (bias)	–0.002	–0.002	–0.001	–0.001
	MSDR	1.108	1.184	1.119	1.089

RK-LC – regression kriging with land cover classes; RK-ALT – regression kriging with altitude; UK – universal kriging; OK – ordinary kriging

the ME is lower for UK and OK (–0.001) and a bit higher for RK-ALT and RK-LC (–0.002) predictions. The MSDR is the best for OK and equals to 1.089. In sum, the cross-validation statistics is slightly better for UK and OK than for both RK, but the differences are again very small (Table 1).

In percentage terms in the case of topsoil pH the RK-LC is the best predictor at 33% of all sites, the RK-ALT at 32%, the UK at 22% and the OK at 13%. Similarly, in the case of humus quality the RK-LC is the best predictor at 31% of all sites, the RK-ALT at 26%, the UK at 24%, and the OK at 19%. In both cases the RK (either with land cover classes or with altitude) clearly offers a higher percentage of best predictions than UK and OK (Table 2). The ratio of best predictions indicates that the soil pH and humus quality are driven mostly by land cover and altitude and little less by the geographic position itself. Since significant differences in soil chemistry between beech and spruce land cover have been showed in many recent studies aimed either on this particular area (Tejnecký *et al.* 2013), or on the highly similar Železné hory Mts. (Oulehle & Hruška 2005; Oulehle *et al.* 2007), the superiority of RK-LC is not a surprise. This does not apply to RK-ALT, the second best predictor which incorporates the regression with altitude, when none relation to the altitude was found across this area

Table 2. Count of best predictions for different kriging methods in percentage terms

	RK-LC	RK-ALT	UK	OK
pH	33	32	22	13
Humus quality	31	26	24	19

RK-LC – regression kriging with land cover classes; RK-ALT – regression kriging with altitude; UK – universal kriging; OK – ordinary kriging

for a range of soil characteristics such as content of different Al species, organic carbon content, soil pH, exchangeable hydrogen cations, sorption characteristics, etc. (Borůvka *et al.* 2009).

Generally, according to a bunch of previous studies (e.g. Knotters *et al.* 1995; Bishop & McBratney 2001; Leopold *et al.* 2005; Lopez-Granados *et al.* 2005;), the hybrid interpolation methods, both modifications of RK, outperformed other two prediction techniques. As showed by Zhu and Lin (2010) the advantage of combined interpolation methods is especially evident for contrasting landscape, which is precisely the situation what this study is devoted to.

CONCLUSIONS

We conclude that higher pH values occur at higher altitudes (700–1200 m) in the centre and in the southeast part of the area where the land cover is described as grassland with sparse occurrence of young spruce and rowan (former death forest zone). Contrary, lower pH values occur at lower altitudes (300–700 m) in the northwest part of the area with dominant spruce monoculture cover and meadows. An exception is the original beech wood in the northwest part, where, despite lower altitudes, the topsoil pH is significantly higher than that of the surroundings. Soils with better quality of humus material occur in the centre and in the northeast part of the area at higher altitudes where the vegetation is formed mainly by grass. Worse humus quality is found in the west and south part of the area at lower altitudes where the vegetation cover is described as spruce monoculture and meadows.

The ordinary, universal and regression kriging maps of topsoil pH as well as maps of humus quality are very similar. There are negligible differences in the cross-validation statistics, but in percentage terms the regression kriging is superior over other kriging methods since the highest percentage of best predictions in total was made by this technique. The results imply that the topsoil pH and humus quality are driven approximately equally by land cover and altitude and little less by pure geographic position. Furthermore, the universal kriging maps show a northeast to southwest spatial trend of topsoil pH and a northwest to southeast spatial trend of humus quality.

Acknowledgments. Supported by the Ministry of Agriculture of the Czech Republic, Project No. QI92A216.

References

Akselsson C., Ardö J., Sverdrup H. (2004): Critical loads of acidity for forest soils and relationship to forest decline in the Northern Czech Republic. Environmental Monitoring and Assessment, **98**: 363–379.

Bishop T., McBratney A. (2001): A comparison of prediction methods for the creation of field-extent soil property maps. Geoderma, **103**: 149–160.

Borůvka L., Mládková L., Drábek O. (2005): Factors controlling spatial distribution of soil acidification and Al forms in forest soils. Journal of Inorganic Biochemistry, **99**: 1796–1806.

Borůvka L., Nikodem A., Drábek O., Vokurková P., Tejnecký V., Pavlů L. (2009): Assessment of soil aluminium pools along three mountainous elevation gradients. Journal of Inorganic Biochemistry, **103**: 1449–1458.

Brus D.J., Kempen B., Heuvelink G.B.M. (2011): Sampling for validation of digital soil maps. European Journal of Soil Science, **62**: 394–407.

Deutsch C., Journel A. (1998): GSLIB: Geostatistical Software and User's Guide. 2nd Ed. Oxford University Press, New York.

Hengl T., Heuvelink G., Stein A. (2004): A generic framework for spatial prediction of soil variables based on regressionkriging. Geoderma, **122**: 75–93.

Hengl T., Heuvelink G.B.M., Rossiter D.G. (2007): About regression-kriging: From equations to case studies. Computer & Geoscience, **33**: 1301–1315.

Hruška J., Ciencala E. (eds.) (2003): Long-Term Acidification and Nutrient Degradation of Forest Soils – Limiting Factors Forestry Today. Ministry of the Environment of the Czech Republic, Prague.

Knotters M., Brus D.J., Oude Voshaar J.H. (1995): A comparison of kriging, co-kriging and kriging combined with regression for spatial interpolation of horizon depth with censored observations. Geoderma, **67**: 227–246.

Leopold U., Heuvelink G.B., Tiktak A., Finke P.A., Schoumans O. (2005): Accounting for change of support in spatial accuracy assessment of modelled soil mineral phosphorous concentration. Geoderma, **130**: 368–386.

Lopez-Granados, F., M. Jurado-Exposito, J. Pena-Barragan, L. Garcia-Torres (2005): Using geostatistical and remote sensing approaches for mapping soil properties. European Journal of Agronomy, **23**: 279–289.

Mládková L., Borůvka L., Drábek O. (2004): Distribution of aluminium among its mobilizable forms in soils of the Jizera mountains region. Plant, Soil and Environment, **50**: 346–351.

Nikodem A., Kodešová R., Drábek O., Bubeníčková L., Borůvka L., Pavlů L., Tejnecký V. (2010): A numerical study of the impact of precipitation redistribution in a beech forest canopy on water and aluminum transport in a podzol. Vadose Zone Journal, **2**: 238–251.

Odeh I., McBratney A., Chittleborough D. (1995): Further results on prediction of soil properties from terrain attributes: heterotopic cokriging and regression-kriging. Geoderma, **67**: 215–226.

Oulehle F., Hruška J. (2005): Tree species (*Picea abies* and *Fagus sylvatica*) effects on soil water acidification and aluminium chemistry at sites subjected to long-term acidification in the Ore Mts., Czech Republic. Journal of Inorganic Biochemistry, **99**: 1822–1829.

Oulehle F., Hofmeister J., Hruška J. (2007): Modelling of the long-term effect of tree species (Norway spruce and European beech) on soil acidification in the Ore Mountains. Ecological Modelling, **204**: 359–371.

Papritz A., Stein A. (1999): Spatial prediction by linear kriging. In: Stein A., van der Meer F., Gorte B. (eds): Spatial Statistics for Remote Sensing. Kluwer Academic Publishers, Dodrecht, 83–113.

Pebesma E.J. (2004): Multivariable geostatistics in S: the gstat package. Computers & Geosciences, **30**: 683–691.

Pospíšil F. (1981): Group- and fractional composition of the humus of different soils. In: Transactions 5th Int. Soil Science Conf. Vol. 1. Research Institute for Soil Improvement, Prague, 135–138.

R Development Core Team (2011): R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna. Available at http://www.R-project.org

Reinds G. J., Posch M., Leemans R. (2009): Modelling recovery from soil acidification in European forests under climate change. Science of the Total Environment, **407**: 5663–5673.

Tejnecký V., Bradová M., Borůvka L., Němeček K., Šebek O., Nikodem A., Zenáhlíková J., Rejzek J., Drábek O. (2013): Profile distribution and temporal changes of sulphate and nitrate contents and related soil properties under beech and spruce forests. Science of the Total Environment, **442**: 165–171.

Wackernagel H. (1998): Multivariate Geostatistics: An Introduction with Applications. 2nd Ed. Springer, Berlin.

Zhu Q., Lin H.S. (2010): Comparing ordinary kriging and regression kriging for soil properties in contrasting landscapes. Pedosphere, **20**: 594–606.

Received for publication November 15, 2012
Accepted after corrections February 28, 2013

Corresponding author:

Ing. Radim Vašát, Česká zemědělská univerzita, Fakulta agrobiologie, potravinových a přírodních zdrojů, katedra pedologie a ochrany půd, Kamýcká 129, 165 21 Praha, Česká republika
e-mail: vasat@af.czu.cz