

Building Soil Spectral Library of the Czech Soils for Quantitative Digital Soil Mapping

LUKÁŠ BRODSKÝ, ALEŠ KLEMENT, VÍT PENÍŽEK, RADKA KODEŠOVÁ
and LUBOŠ BORŮVKA

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

Abstract: Spectral libraries are the data archives of spectral signatures measured on natural and/or man-made materials. Here, the objective is to build a soil spectral library of the Czech soils (SSL-CZ). Further on, the overall aim is to apply diffuse reflectance spectroscopy as a tool for digital soil mapping. An inevitable part of the library is a metadata database that stores the corresponding auxiliary information on the soils: type of material (soil, parent material), sample preparation, location of the sample with geographic coordinates, soil classification, morphological features, soil laboratory measurements – chemical, physical, and potential biological properties, geophysical features of and climatological information on the sample location. The metadata database consists of seven general tables (General, Spatial, Soil class, Environmental, Auxiliary, Analytical and Spectra) relationally linked together. The stored information allows for a wide range of analyses and for modelling developments of digital soil mapping applications. An example of partial least-square regression (PLSR) modelling for soil pH and clay content with 0.84 and 0.68 coefficients of determination is provided on the subset of the collected data. Currently, the SSL-CZ database contains more than 500 records in the first phase of development. Spectral reflectance signatures are stored in the range of 350 to 2500 nm with a step of 1 nm measured by ASD FieldSpec 3. The soil spectral library developed is fully compatible with Global Soil Spectral Library (Soil Spectroscopy Group).

Keywords: diffuse reflectance spectroscopy; digital soil mapping; soil; spectral library

Spectral libraries are data archives that consist of spectral signatures measured on selected natural and/or man-made materials. These libraries should also include the corresponding meta-information on the records. There are in principle two potential purposes: the use of the data in remote sensing as in-situ radiance/reflectance measurements for the calibration and/or end-members selection for further data processing (SELIGE *et al.* 2006), or the use of the measurements for rapid laboratory assessment by diffuse reflectance spectroscopy (DRS) of e.g. soil properties (VISCARRA ROSSEL *et al.* 2006; BROWN 2007). The focus of this paper is on the spectral library use as described in the latter case.

A number of publicly available spectral libraries exist that vary in many specifications. The most widely used spectral libraries for the remote sensing application are the United States Geological Survey (USGS) with the name splib06a (CLARK *et al.* 2007, available from <http://speclab.cr.usgs.gov>); Jet Propulsion Laboratory (JPL) spectral library by C.I. Grove, S.J. Hook, and E.D. Paylor II incorporated in ASTER compilation; Johns Hopkins University spectral library (JHU) also incorporated in ASTER compilation available from <http://speclib.jpl.nasa.gov> (BALDRIGE *et al.* 2009). The ASTER library collection includes the spectra of rocks, minerals, lunar soils, terrestrial

Supported by the Czech Science Foundation, registration number 526/09/1762 and by the Ministry of Education, Youth and Sports of the Czech Republic, Project No. MSM 6046070901.

soils, man-made materials, meteorites, vegetation, snow and ice covering the visible through thermal infrared wavelength region (0.4–15.4 μm) of the electromagnetic spectrum. The library consists of limited auxiliary information. The soil samples are represented by few examples and do not cover a wide variation of the existing soils and their properties. Further on, complimentary spectral libraries are available from other collections, for example: thermal emission spectral library of rock-forming minerals available from <http://speclib.asu.edu> (CHRISTENSEN *et al.* 2000) or Raman Spectra of Minerals (RAMAN 2006, available only online from <https://www.fis.unipr.it/phevix/ramandb.php>).

In the soil science, the Global Soil Spectral Library is currently being developed, initiated by Raphael Viscarra Rossel under soil spectroscopy group (VISCARRA ROSSEL 2008). At present, the library contains 5223 soil spectra signatures. The requirements for contributing spectra to the global library are defined as follows:

- soil spectra collected in the 350–2500 nm range every 1 nm,
- at least soil organic carbon and clay content data measured on the soil samples; and also any other soil chemical, physical, biological, and mineralogical data,
- noting which analytical techniques were used,
- coordinates (in WGS84 format) for each sample,
- soil classification for each sample, preferably using FAO-WRB (FAO 1998),
- future access to the soil samples for mid-IR scanning.

This paper provides the specifications of the first version of the Czech Soil Spectral Library, proposes further library evolution and provides the first example of the spectra use in a model calibration. The overall objective is to build a representative soil spectral library of the Czech soils for statistical inference models to allow the exploitation of rapid soil quantitative spectroscopy for digital soil mapping and monitoring the activities.

MATERIAL AND METHODS

Soil samples

The materials used in the library come first of all from various studies carried out by the Department of Soil Science and Soil Protection of the Czech University of Life Sciences Prague. Secondly, the

selection of the soil samples in the continuously updated library is based on statistical analysis of representative locations, initially limited to arable land. The aim is to describe the wide range of soil variations in the Czech Republic. The main soil properties examined are soil organic carbon, pH, carbonate content, texture (clay, silt, sand), cation exchange capacity (CEC), and soil chemical properties important for the crop production – available phosphorus, potassium, calcium and magnesium. Clay mineralogy is another important aspect that plays a role in the spectroscopy measurements.

Initially 11 topsoil representative locations in the Czech Republic (where detailed monitoring takes place at the research stations) and one subsoil and parent material were selected that are further used as the reference soil samples, Table 1. The same set of the soil samples had been previously used to study various pesticides sorption in the soils of the Czech Republic (KODEŠOVÁ *et al.* 2010). These reference samples represent various soil and climatic conditions of the Czech Republic. Eleven humic horizons of different soil types (four Chernozems, one Phaeozem, one Luvisol, one Arenosol, and four Cambisols) and two complementary parent materials (sand and loess) are included.

The other records in the library consist of the soil samples collected to describe fieldscale soil variation (e.g. Sedlčany – 3 ha, Lány – 14 ha, Chrástany – 4 ha, Brumovice – 100 ha). The number of samples varied field to field from 10 to 500 samples.

Further on, the selection of the soil samples based on statistical analysis of the representative locations was performed. The statistical analysis is based on the available GIS layers: soil classification map (KOZÁK *et al.* 1996), parent material map, and other environmental co-variables like climate regions. The stratification of the samples was based on the analysis of characteristic soil substrates within climatic regions. The soil substrates and climate are the main soil forming factors in the area of the Czech Republic. Their combination can better describe the soil variability than sample stratification based directly on the soil units, because some of the soil units (e.g. Cambisols) are very variable in their properties. Firstly, soil and substrate units of a small extent (< 2% of total soil cover), representing local peculiarities rather than being part of a typical soil cover, were excluded from the analysis (number of substrate classes decreased from 17 to 9). Secondly, class distribution of the samples was based on the combination of 9 substrate classes (reclassified from

Table 1. Reference soil samples basic characteristics (part of metadata tables) (in %)

Location	WRB_CLASS*	PARENT_MAT**	Cox	pH-KCl	CaCO ₃	Clay	Silt	Sand
Ivanovice na Hané	Haplic Chernozem	loess	1.8	6.3	0	11.4	75.5	13.2
Předbořice	Haplic Cambisol	syenit	1.7	5.0	0.2	4.8	36.2	59.0
Semice	Arenosol Epieutric	sand	0.7	5.7	0.16	3.5	12.8	83.7
Jince	Haplic Cambisol	quartzite	1.6	4.9	0	20.3	43.7	35.9
Vysoké nad Jizerou	Dystric Cambisol	gneiss	2.3	4.8	0.24	16.9	57.8	25.3
Čáslav	Greyic Phaeozem	loess	1.4	6.5	0.1	13.4	77.1	9.5
Humpolec	Haplic Cambisol	gneiss	1.6	4.4	0	9.9	34.7	55.4
Hněvčeves	Haplic Luvisol	loess	1.0	5.6	0	13.9	76.5	9.6
Písková Lhota	–	sand	0.02	8.1	1	3.3	3.0	93.7
Milčice	Stagnic Chernozem Siltic	marlite	2.9	7.4	28	15.8	54.6	29.5
Praha Suchdol – top	Haplic Chernozem	loess	2.0	7.2	7.8	19.4	56.3	24.4
Velké Chvalovice	Chernozem Arenic	gravel-sand	0.9	6.9	2.5	6.4	20.0	73.6
Praha Suchdol – sub	–	loess	0.4	7.4	11.6	24.5	51.8	23.7

*WRB_CLASS stands for the international standard taxonomic soil classification, the World Reference Base for soil resources

**PARENT_MAT stands for parent material of soil sample

the original classes to reduce the number of classes but to keep the main characteristics according to the studied properties) and 9 climatic regions, characterised by temperature and precipitation (Table 2). The number of samples in the substrate-climate classes corresponds to their spatial extent. The soil classes (groups) were statistically analysed under subsets of the combined substrate-climate regions. The number of samples for each soil type was assigned according to the spatial extent of the three most prevailing soil groups.

Cross-validation of this method was done by analysis of the spatial extents of the soil units in the soil map. The proportions of the representative main soil types from the soil map correspond to the proportions of the sampled soil types. The difference of the sampled soils using the substrates-climatic classes is only a few percent, which does not influence the characterisation of the soil cover on this scale.

Sample preparation and laboratory measurements

Soil samples from the topsoil of arable land dedicated to the spectral library were subjected to standard laboratory procedure. The soil samples were

air-dried, ground and sieved to 2-mm. The basic chemical and physical soil properties were obtained using standard laboratory procedures under a constant laboratory temperature of 20°C. The soil pH was measured using a 1:5 (w/v) soil to water ratio (pH_{H₂O}) and 1M KCl (pH_{KCl}) solutions (ISO 10390 1994) using an inoLab Level 1 pH-meter. CEC was measured using the method proposed by Bower and Hatcher (BOWER & HATCHER 1966 in KLUTE 1996). The exchange sites were saturated with Na using 1M sodium acetate adjusted to pH 7.0. Exchangeable Na was removed with 1M ammonium acetate. Sodium concentrations in the ammonium acetate solutions were determined by atomic adsorption spectrophotometry. The soil carbon content was measured using the dichromate redox titration method (SKJEMSTAD & BALDOCK 2008). The wet oxidation (K₂Cr₂O₇) was followed by the potentiometric titration with ferrous ammonium sulphate. CaCO₃ content was measured using the volumetric calcimeter method described by LOOPP and SUAREZ (1996). The carbonates reacted with HCl (in a sealed system) to form CO₂. The pycnometer method (FLINT & FLINT 2002) was used to measure the particle density. The particle size distribution (fractions of clay, silt, and sand) was obtained by the hydrometer method (GEE & OR 2002). Mehlich

Table 2. Stratification of 100 samples according to the substrate groups and climatic regions

Substrates groups	MW10	MW11	MW3	MW5	MW7	MW9	W2	W4	Total
Calcareous clays		1 Cambisol Clayic					1 Cambisol Clayic 2 Chernozem Gleyic 1 Vertisol		5
Fluvial (non-calcareous) sediments	2 Fluvisol 1 Gleysol	1 Fluvisol 1 Gleysol					3 Fluvisol 1 Gleysol 1 Chernozem Gleyic		10
Polygenetic (fluvio-delluvial) sediments		2 Stagnosol 2 Albeluvisol		2 Stagnosol 1 Albeluvisol 1 Gleysol	2 Stagnosol 1 Gleysol				11
Loess		3 Luvisol					12 Chernozem 2 Luvisol	3 Luvisol	20
Decalcified loess	3 Albeluvisol 2 Luvisol	2 Albeluvisol 1 Stagnosol				1 Albeluvisol 1 Stagnosol			10
Delluvial sediments from calcareous rocks		1 Cambisol	1 Cambisol				2 Cambisol 1 Leptosol		5
Delluvial sediments from calcareous sedimentary, igneous and metamorphic rocks	3 Cambisol	5 Cambisol			3 Cambisol				11
Delluvial sediments from dyctric sedimentary, igneous and metamorphic rocks			9 Cambisol	9 Cambisol	5 Cambisol				23
Gravel and gravel-sands on terraces		1 Cambisol					3 Regosol 2 Cambisol		5

MW – moderately warm regions; W – warm regions; for more details see Climatic regions according to Quitt's classification (Quitt 1971; Květoň & Žák 2007)

III extraction procedure was applied to measure the available P, K, Ca and Mg (MEHLICH 1984).

Soil spectra collection

The diffuse reflectance spectra in the wavelength range of 350–2500 nm were acquired in the laboratory with the ASD High Intensity Contact Probe, designed for contact measurements of solid raw materials, and FieldSpec-3 instrument. The Contact Probe used is equipped with a halogen bulb light source (6.5 W , with colour temperature $2901 \pm 10\%$ K). Spot size of the Contact Probe has the diameter of 22 mm. The instrument has spectral resolution 3 nm at 350–700 nm and 10 nm at 1400–2100 nm, while the sampling intervals account for 1.4 nm at 350–1050 nm range and 2 nm at 1000–2500 nm wavelengths. Spectralon as the diffuse white standard for the reflectance calibration was used.

The soil samples were poured into PP plastic dishes of 60 mm in diameter and 5 mm in depth. The upper surface of the sample was smoothed (Figure 1) with care taken not to introduce any preferred grain orientation. The Contact Probe glass protection in the chamber was regularly cleaned with soft paper tissue.

Reference samples

A set of 13 homogenised soil samples were used as standards to measure potential uncertainties. These samples were measured multiple times during the



Figure 1. Soil sample in dish prepared for spectra measurements

soil spectra acquisitions. The mean and standard deviations of Haplic Luvisol determined during the sample measurement are shown in Figure 2 as an example. The variation from the mean spectra ranged between 1.27 and 4.09% in 30 repeated spectra collection. The spectra collection showing larger deviations, and/or discontinuity between the three spectral regions of the instrument was rejected and after re-calibration new acquisition was performed.

Library organisation

The library has two main parts, the spectra files and the metadata (Figure 3). The metadata model consists of seven general tables relationally linked together. The first table, called GENERAL consists of the identification number, date of record, update and shortened name of location or field survey for simplified search purposes. The other linked tables contain spatial information SPATIAL, soil classification records SOIL_CLASS, environmental conditions of the sample location ENVIRONMENTAL, auxiliary information on soil samples AUXILIARY, analytical laboratory records on the samples ANALYTICAL and finally the spectra records that are only links to the files stored separately from the database SPECTRA. Each soil spectrum has a separate ASCII file with standard naming convention to be unique. The file name (e.g. culs.100148_vrch.txt) includes a unique integer ID record and suffix with shortened name or abbreviation of the location.

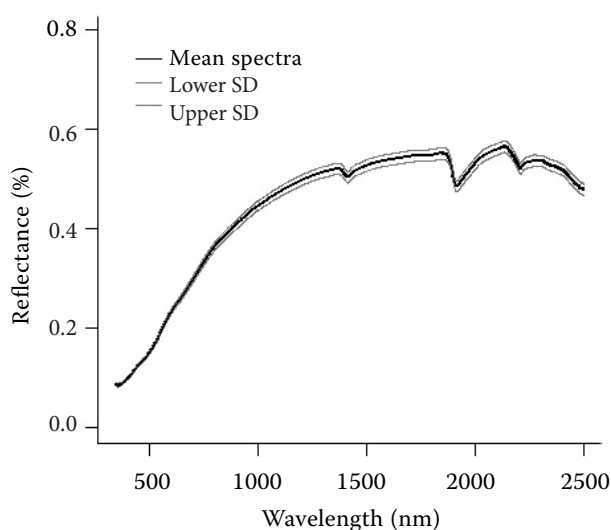


Figure 2. Example of the mean and standard deviations of Haplic Luvisol spectra

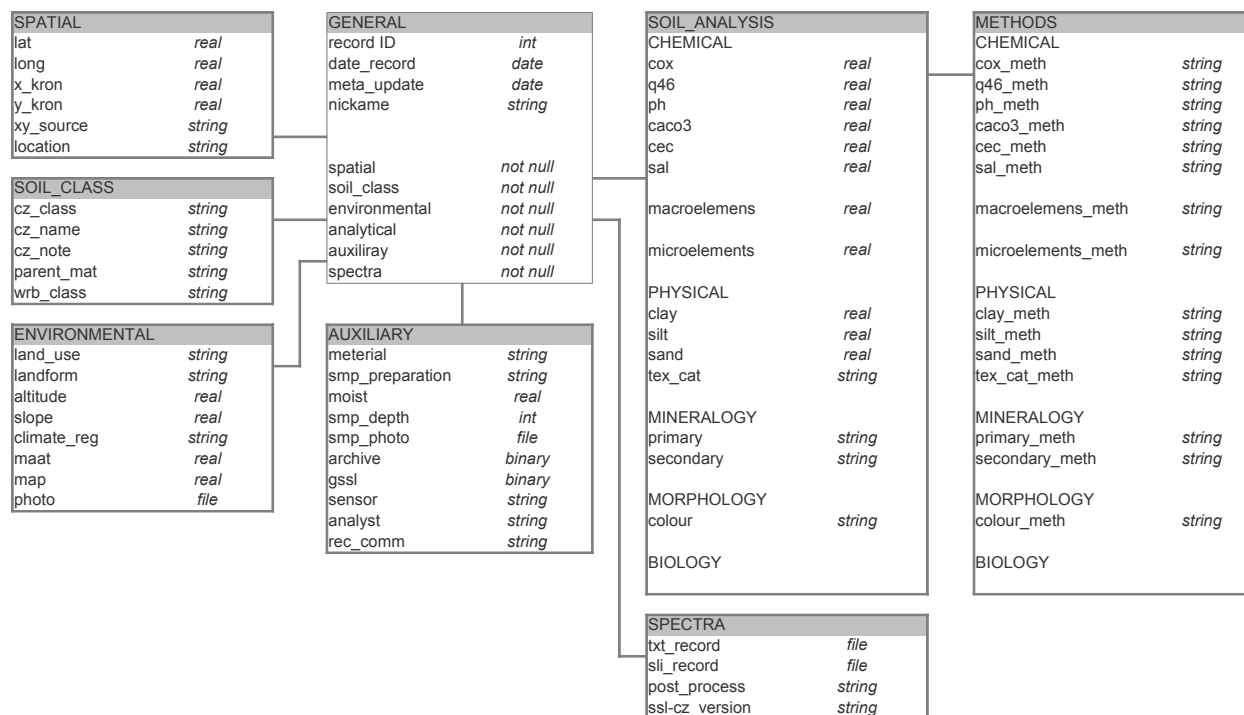


Figure 3. Database model of the soil spectral library metadata

RESULTS AND DISCUSSION

The samples of the library spectra measurements are shown in Figure 4. The spectra demonstrate the basic variety in the soil spectral shapes across the visible to shortwave infrared (0.4–2.5 μm) wavelength ranges. These are examples from 11 research stations. Figure 4 also presents examples of normalised spectra by continuum removal. Convex hull (EDDY 1977) boundary points were used to isolate the spectral continuum. This technique allows the comparison between the spectra by scaling the feature depths to unity. Continuum removal is commonly used as a component in the identification procedures of hyperspectral mapping because of its ability to standardise image reflectance and eliminate the effects of the sun and view geometry. In this case, the continuum removed line highlights the important soil absorption features, in particular Fe absorption bands, $\text{H}_2\text{O}/\text{OH}^-$ bands, clay minerals band, and carbonate band in one soil example.

Example modelling

Soil properties can be predicted in spectroscopy quantitatively and simultaneously by a multivariate calibration approach using partial least-square re-

gression (PLSR), multiple linear regression (MLR), (PCR) or artificial neural network procedure (ANN). PLSR was used in this first example derived from a subset of Czech Soil Spectral Library data (Chrášťany field and reference samples). Model calibration procedure was applied on a subset of 75 soil samples measured for soil pH and clay content, Figures 5 and 6. To evaluate the model fit, several indicators

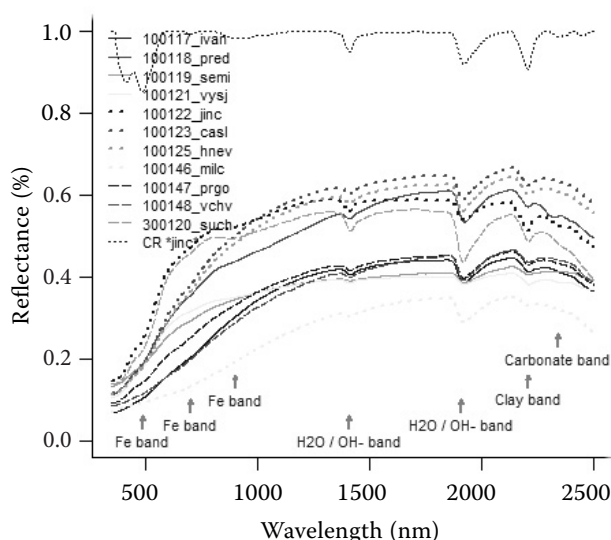


Figure 4. Examples of soil library reflectance spectra (11 reference soils)

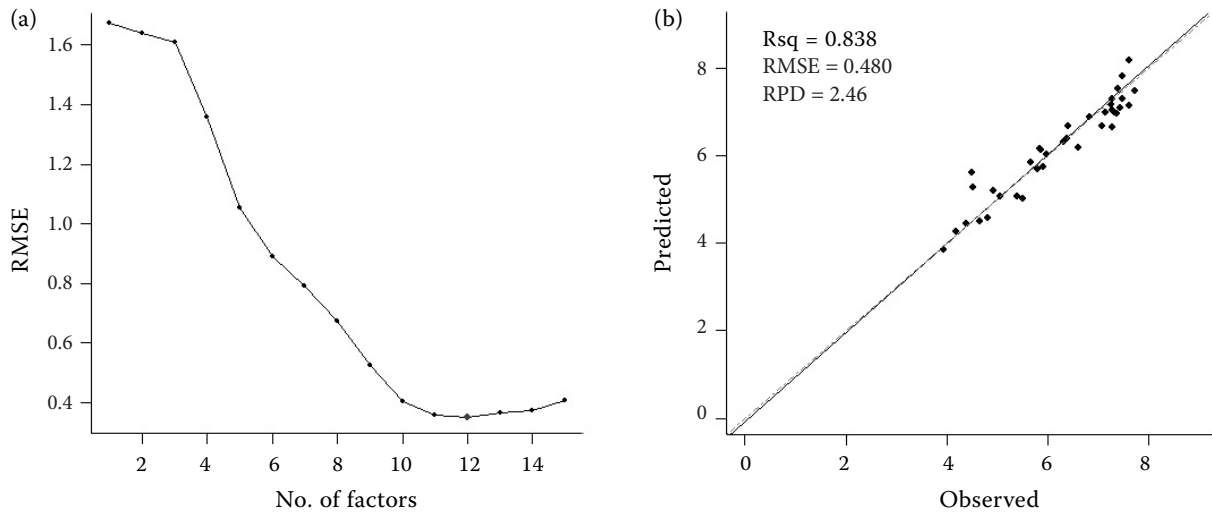


Figure 5. Partial least-squares regression modelling and prediction output for soil pH; (a) – cross-validated root mean square error (RMSE) of prediction against the number of factors; (b) – observed against the cross-validated PLSR predictions of soil pH with basic statistics

were employed. First, the coefficient of determination between the predicted and observed values provided a common measure of variance reduction. Second, RMSE was used to evaluate the prediction error in the model for each of the soil attributes. Leave-one-out cross-validation was used to determine the number of factors to retain in the calibration models (Figures 5 and 6). The model with the lowest RMSE was selected. The resulting R^2 values for the soil pH and clay content examples were 0.84 and 0.68. The models calibration achieved for this example field was fairly good and comparable with other

studies found in the literature (VISCARRA ROSSEL *et al.* 2006; BROWN 2007 and others). The result of the statistical model calibration demonstrates the potential of quantitative soil predictions using the VNIR spectroscopy and the PLSR decomposition that can be further employed in digital soil mapping.

CONCLUSIONS

This first version of the spectral library contains over 500 collected spectra. The version includes

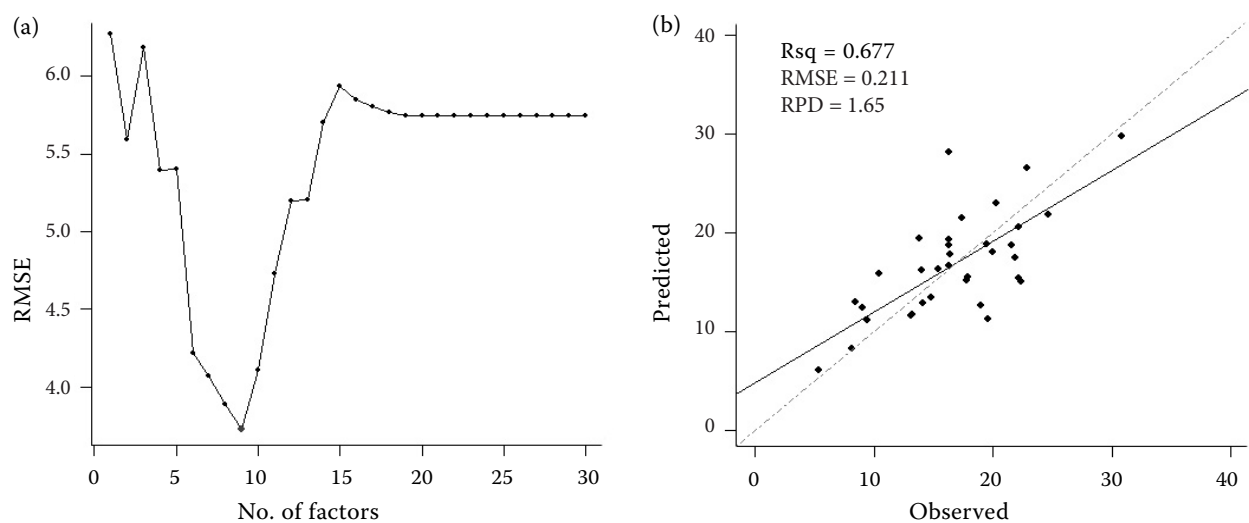


Figure 6. Partial least-squares regression modelling and prediction output for soil clay content; (a) – cross-validated root mean square error (RMSE) of prediction against the number of factors; (b) – observed against the cross-validated PLSR predictions of soil clay with basic statistics

the basic soil spectra variation of the Czech soils. The continually updated library provides initial collections of spectra covering the wavelength range 350–2500 nm that allows for the modelling of basic soil properties by statistical inference, as for instance the PLSR model in the example here, and is suitable for further soil spectroscopy exploration in the Czech Republic and practical application of digital soil mapping procedures.

References

- BALDRIDGE A.M., HOOK S.J., GROVE C.I., RIVERA G. (2009): The ASTER spectral library version 2.0. *Remote Sensing of Environment*, **113**: 711–715.
- BOWER C.A., HATCHER J.T. (1966): Simultaneous determination of surface area and cation-exchange capacity. *Soil Science Society of America Proceedings*, **30**: 525–527.
- BROWN D.J. (2007): Using a global VNIR soil-spectral library for local soil characterization and landscape modeling in a 2nd-order Uganda watershed. *Geoderma*, **140**: 444–453.
- CHRISTENSEN P.R., BANDFIELD J.L., HAMILTON V.E., HOWARD D.A., LANE M.D., PIATEK J.L., TIFF S.W., STEFANOV W.L. (2000): A thermal emission spectral library of rock-forming minerals. *Journal of Geophysical Research*, **105**: 9735–9739.
- CLARK R.N., SWAYZE G.A., WISE R., LIVO E., HOEFEN T., KOKALY R., SUTLEY S.J. (2007): USGS digital spectral library splib06a: U.S. Geological Survey, Digital Data Series 231.
- EDDY W.F. (1977): A new convex hull algorithm for planar sets. *ACM Transactions on Mathematical Software*, **3**: 398–403.
- FAO (1998): World Reference Base for Soil Resources. World Soil Resources Report 84. FAO, Rome.
- FLINT A.L., FLINT L.E. (2002): Particle density. In: DANE J.H., TOPP G.C. (eds): *Methods of Soil Analysis, Part 4 – Physical Methods*. Soil Science Society of America, Inc., Madison, 229–240.
- GEE G.W., OR D. (2002): Particle-size analysis. In: DANE J.H., TOPP G.C. (eds): *Methods of Soil Analysis, Part 4 – Physical Methods*. Soil Science Society of America, Inc., Madison, 255–294.
- ISO 10390 (1994): Standard of Soil Quality – Determination of pH. International Organization of Standardization, Geneva.
- KLUTE A. (1996): *Methods of Soil Analysis*. Agronomy Monograph No. 9, American Society of Agronomy, Madison.
- KODEŠOVÁ R., KOČÁREK M., KODEŠ V., DRÁBEK O., KOZÁK J., HEJTMÁNKOVÁ K. (2010): Pesticide adsorption in relation to soil properties and soil type distribution in regional scale. *Journal of Hazardous Materials*, **186**: 540–550.
- KOZÁK J., NĚMEČEK J., JETMAR M. (1996): The database of soil information system – PUGIS. *Rostlinná Výroba*, **42**: 529–534.
- KVĚTOŇ V., ŽÁK M. (2007): New climate atlas of Czechia. *Studia Geophysica et Geodaetica*, **51**: 345–349.
- LOOPEPPT H.R., SUAREZ L.D. (1996): Carbonate and gypsum. In: SPARKS D.L., PAGE A.L., HELMKE P.A., LOEPPERT R.H., SOLTANPOUR P.N., TABATABAI M.A., JOHNSTON C.T., SUMNER M.E. (eds): *Methods of Soil Analysis. Part 3 – Chemical Methods*. Soil Science Society of America, Inc., Madison, 437–474.
- MEHLICH A. (1984): Mehlich-3 soil test extractant: a modification of Mehlich-2 extractant. *Communications in Soil Science and Plant Analysis*, **15**: 1409–1416.
- QUITT E. (1971): Klimatické oblasti Československa. Academia, *Studia Geographica* 16, GÚ ČSAV, Brno. (in Czech)
- RAMAN (2006): Raman Spectra of Minerals. Available at <https://www.fis.unipr.it/phevix/ramandb.php> (accessed March 1, 2006)
- SELIGE T., BOHNER J., SCHMIDHALTER U. (2006): High resolution topsoil mapping using hyperspectral image and field data in multivariate regression modeling procedures. *Geoderma*, **136**: 235–244.
- SKJEMSTAD J.O., BALDOCK J.A. (2008): Total and organic carbon. In: CARTER M.R., GREGORICH E.G. (eds): *Soil Sampling and Method of Analysis*. Canadian Society of Soil Science, Taylor and Francis, Boca Raton, 225–237.
- VISCARRA ROSSEL R. (2008): The Soil Spectroscopy Group and the development of a global soil spectral library. *Pedomtron*, The Newsletter of the Pedometrics Commission of the IUSS, Issue 25.
- VISCARRA ROSSEL R.A., WALVOORT D.J.J., MCBRATNEY A.B., JANIK L.J., SKJEMSTAD J.O. (2006): Visible, near infrared, mid infrared or combined diffuse reflectance spectroscopy for simultaneous assessment of various soil properties. *Geoderma*, **131**: 59–75.

Received for publication April 21, 2011

Accepted after corrections September 13, 2011

Corresponding author:

Ing. LUKÁŠ BRODSKÝ, Ph.D., Česká zemědělská univerzita v Praze, Fakulta agrobiologie, potravinových a přírodních zdrojů, katedra pedologie a ochrany půdy, Kamýcká 129, 165 21 Praha 6-Suchbát, Česká republika
e-mail: brodsky@af.czu.cz
