https://doi.org/10.17221/41/2019-SWR

Harmonisation of a large-scale historical database with the actual Czech soil classification system

Tereza Zádorová[1], Daniel Žížala[2], Vít Penížek[1], Aleš Vaněk[1]*

[1]Department of Soil Science and Soil Protection, Faculty of Agrobiology, Food and Natural Resources, Czech University of Life Sciences Prague, Prague, Czech Republic
[2]Research Institute for Soil and Water Conservation, Prague, Czech Republic
**Corresponding author: zadorova@af.czu.cz*

Citation: Zádorová T., Žížala D., Penížek V., Vaněk A. (2020): Harmonisation of a large-scale historical database with the actual Czech soil classification system. Soil & Water Res., 15: 101–115.

Abstract: The possibility of the adequate use of data and maps from historical soil surveys depends, to a large measure, on their harmonisation. Legacy data originating from a large-scale national mapping campaign, "Systematic soil survey of agricultural soils in Czechoslovakia (SSS, 1961–1971)", were harmonised and converted according to the actual system of soil classification and descriptions used in Czechia – the Czech taxonomic soil classification system (CTSCS). Applying the methods of taxonomic distance and quantitative analysis and reclassification of the selected soil properties, the conversion of two types of mapping soil units with different detailed soil information (General soil representative (GSR), and Basic soil representative (BSR)) to their counterparts in the CTSCS has been effectuated. The results proved the good potential of the used methods for the soil data harmonisation. The closeness of the concepts of the two classifications was shown when a number of soil classes had only one counterpart with a very low taxonomic distance. On the contrary, soils with variable soil properties were approximating several related units. The additional information on the soil skeleton content, texture, depth and parent material, available for the BSR units, showed the potential in the specification of some units, though the harmonisation of the soil texture turned out to problematic due to the different categorisation of soil particles. The validation of the results in the study region showed a good overall accuracy (75% for GSR, 76.1% for BSR) for both spatial soil units, when better performance has been observed in BSR. The conversion accuracy differed significantly in the individual soil units, and ranged from almost 100% in Fluvizems to 0% in Anthropozems. The extreme cases of a complete mis-classification can be attributed to inconsistencies originating in the historical database and maps. The study showed the potential of modern quantitative methods in the legacy data harmonisation and also the necessity of a critical approach to historical databases and maps.

Keywords: legacy data; soil classification; soil survey; soil mapping; taxonomic distance

The importance of legacy data has increased over the last decades, along with the development of international and global soil databases and map applications (Arrouyas et al. 2017). All over the world, the primary data (point soil profile observations) have been collected within national soil surveys and have been eventually derived into soil polygon maps and maps of soil properties. The historical soil data are available in the form corresponding to the period and place of their acquirement. This comprises differences in: the standards in the soil profile description, the laboratory methods, the used classification systems, the map scales and geo-referencing systems, the data density, the organisation and structure of the soil databases. The possibility, extent and quality of the harmonisation of legacy data with actual soil

Supported by the Ministry of Agriculture of the Czech Republic, Project No. QK1820389.

https://doi.org/10.17221/41/2019-SWR

description and classification systems determine the potential of their use in diverse soil applications on a national and international level. Harmonisation of the soil databases has been studied widely either for its use in global databases and maps (Arrouyas et al. 2017), in correlation with international soil classifications (Hughes et al. 2017, 2018; Láng et al. 2013) or in the actualisation of the soil data.

As has been shown, the correlation between two classification systems cannot be realised by a simple translation of the taxa's names (Krasilnikov et al. 2009; Michéli et al. 2016; Zádorová & Penížek 2011; Láng et al. 2016): such a conversion can be a source of significant errors. Moreover, the spatial distribution of the correlated units may differ from their counterpart in the other system (Láng et al. 2013).

Use of the taxonomic distance represents a possible approach for the correlation of two soil classifications. It clarifies the relationship among the soil taxa and promises an improvement in the correlation exactitude and accuracy. The taxonomical distance, as a method used to express the level of similarity and dissimilarity between the different soil taxa, had been introduced in 1960s' (Hole & Hironaka 1960) with local data. During the previous decade, the potential of this method has been further developed, firstly by Minasny and McBratney (2007) who used taxonomic distances in the spatial prediction and digital mapping of soil classes. Several studies have shown its effectiveness in studying the relationships between the soil classes within one classification system (Minasny et al. 2009), the correlation of the different classifications (Láng et al. 2013; Hughes et al. 2017) or the direct classification of the soil profiles (Láng et al. 2016). Two different approaches have been applied in the different studies: a conceptual approach based on the presence or absence of selected features or processes (dominant identifiers; Minasny & McBratney 2007) in the individual soil classes and a centroid-based approach using the mean values of the selected soil features typical for the individual soil classes. The actual quantitative data is entered to this model. Thus, a relevant dataset of soil profiles is needed in this approach (Láng et al. 2013, 2016).

Historical soil survey and soil classification in Czechia

A majority of the data acquired by traditional soil surveys originate in Czechia in the 1960s and 1970s. The extensive "Systematic soil survey of agricultural soils in Czechoslovakia (SSS)" was carried out between 1961–1971. The aim of the survey was to provide detailed and high-quality information on the soil cover, indispensable for a systematic increase of the soil quality and production potential (Kozák 2010). The whole area of the agricultural land in the former Czechoslovakia (7.2 M ha) had been surveyed (more than 700 000 point soil profile observations, 500 000 in Czechia) and mapped on a scale of 1 : 10 000. The map products cover: soil typological unit maps with geological units, soil texture maps, maps of recommended amelioration management. Other map outputs were represented by a series of 1 : 50 000 maps (a polygon soil map, a map of the soil texture and a map of the parent material). More than 2 million soil samples were analysed. The survey had been realised according to a unified methodology for sampling, description of soil morphology, laboratory analysing, classifying and mapping of soils (Damaška et al. 1967; Němeček et al. 1967; Sirový & Facek 1967).

The soils in the SSS were classified according to Genetic-agronomic soil classification (GAC, Němeček et al. 1967) that was developed for the survey. It is a complex hierarchical taxonomy built on a genetic base, but the classifying soils used a set of diagnostic features and their limits. It is comprised of four hierarchical classes (soil type, subtype, variety and erosion form) forming together the genetic soil representative (GSR). Moreover, the classification can be specified by adding lithological information on parent material, skeleton content, texture and soil depth (basic soil representative (BSR)). The GSR and BSR form the qualitative content of the polygon soil maps resulting from the soil survey. Three levels of soil profiles were established within the soil survey: basic pits (B-pits) with only the morphological description and the topsoil and subsoil sampling, selective pits (S-pits) with the morphological description and a full chemical and physical analysis for all the horizons and, finally, special profile pits (SP-pits) with a wider set of analysis compared to the S-pits. The list of soil types distinguished in the GAC and CTSCS and their brief characteristics are shown in Table 1.

The actual Czech soil classification – the Czech taxonomic soil classification system (CTSCS, Němeček et al. 2011) results from the aim to develop a modern soil classification predominantly based on the diagnostics of the soil features that will be comparable with international soil classifications upon its structure and nomenclature. The classification is hierarchical; the fundamental soil taxon is a soil type, classified according

https://doi.org/10.17221/41/2019-SWR

to diagnostic features. The soil types are, according to main paedogenetic process, united into reference soil groups that represent the most general soil taxon with the nomenclature comparable to the international soil classifications. GAC has been one of the fundamental sources in the development of CTSCS. Thus, the central concepts of the traditional soil taxa are similar. However, new soil classes have been introduced in the CTSCS and also the diagnostic features and their limits show differences in the two classifications. Another potential source of the error is the database itself (variable quality of outputs, loss of a part of the records, errors entered during the digitisation processes, low spatial accuracy of the profile records).

Objectives of the study

Based on the need of the relevant harmonisation of the database and the maps of the SSS, the objectives of the paper are to present a methodological approach for the conversion of the soil units at the soil representatives' level (GSR and BSR) of the SSS to the relevant classification units of the CTSCS in the following steps: (*i*) computation of the taxonomic distance of the GAC and CTSCS units, (*ii*) determination of the CTSCS soil classes corresponding to the GSR, (*iii*) harmonisation of the soil features defining the BSR, (*iv*) evaluation of the accuracy of the correlation using the representative soil profiles realised within the SSS.

METHODS

Correlation of the GSR with the CTSCS classes. At this level, we correlate an individual soil class (soil type, subtype, variety and erosion form) with no other specifying information. The conversion is based on the concept of the soil taxa in the two classifica-

Table 1. Description of the soil types distinguished in the Genetic-agronomic soil classification (GAC) and the Czech taxonomic soil classification system (CTSCS)

Description	GAC	CTSCS
Soils with deep, dark mollic horizon, include vertic and weakly leached soils	Chernozems	Chernozem
Soils with deep, dark A horizon and vertic properties		Smonice
Soils with migration of clay and humus, dark clay coatings	Hnedozems	Sedozem
Soils with clay migration, saturated		Hnedozem
Leached soils dominated by clay migration, with albic eluvial horizons and/or albeluvic tonguing	Illimerized soils	Luvizem
Redoximorphic soils	Stagnic soils	Pseudogley
Soils with brown horizons of soil weathering and mineral transformation	Brown soils	Kambizem
Soils with brown horizons of soil weathering and mineral transformation, heavy texture		Pelozem
Soils with reddish horizons of soil weathering, mineral transformation and initial podzolization		Kryptopodzol
Weakly developed soils, often skeletic, on consolidated calcareous parent material	Rendzinas	Rendzina
Weakly developed soils, often skeletic, on consolidated parent material rich $CaCO_3$		Pararendzina
Soils dominated by cheluviation and chilluviation	Podzolized soils	Podzol
Weakly developed sandy soils on non-consolidated sediments	Arenic soils	Regozem
Soils on fluvial sediments, often stratified with uneven humus distribution	Alluvial soils	Fluvizem
Soils on colluvial sediments, often stratified with uneven humus distribution	accumulated soil phase	Koluvizem
Soils with deep, dark mollic horizons, influenced by shallow groundwater, saturated	Meadow soils	Chernice
Weakly developed soils, skeletic, on consolidated parent material	undeveloped soils	Ranker
Weakly developed soils, very shallow		Litozem
Soils developed in predominant reducing conditions	Gleyic soils	Gley
Soils with deep organic layers	Organic soils	Organozem
Soils with high content of soluble salts	Solonchaks	Solonchak
Soils influenced by intensive agricultural management	anthropic subtypes	Kultizem
Man-made soils	Anthropogenic soils	Anthropozem

tions; the definitions, criteria and diagnostic features of each class are evaluated. The conceptual approach of the taxonomic distance based on the qualitative information has been applied for the correlation. The dominant identifiers are represented by the sets of the soil characteristics that developed due to the dominant soil processes and factors and are determinative for the particular soil classes. Each identifier was matched with the soil classes characterised in the CTSCS and GAC coded according to the probability of the presence of the features. The calculation was effectuated at the level of the soil types. The number of codes differs in the individual studies. Minasny et al. (2009) introduced codes 0 and 1 for the probable absence or presence of the attribute. Láng et al. (2013) added code 0.5 for the situation when the condition is possibly, but not necessarily present. In this study, we applied the approach of Láng et al. (2013) and used codes 0 (not present), 0.5 (possibly present) and 1 (present). The set of dominant identifiers and codes attributed to each class are shown in Tables 2 and 3. The taxonomic distance was assessed according to the calculation of Euclidean distance using the dist function in R environment (R Core Team 2018). Using the matrix of identifiers for the two classification systems, the Euclidean distance is calculated from following equation:

$$d_{ij} = \sqrt{(x_i - x_j)^{\mathrm{T}} - (x_i - x_j)}$$

where:

d_{ij} – an element of the distance matrix D ($c \times c$)

c – the number of the correlated soil classes

x – the vector of the identifiers' value

T – matrix transposition

Table 2. The dominant identifiers and codes attributed to each class of the Genetic-agronomic soil classification (GAC)

Identifier	Chernozem	Hnedozem	Illimerized soil	Stagnic soil	Rendzina	Brown soil	Undeveloped soil	Arenic soil	Podzolized soil	Alluvial soil	Meadow soil	Gley soil	Anthropogenic soil	Organic soil	Solonchak
Man-made	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0
Intensive agricultural use	0	0.5	0	0	0.5	0.5	0	0.5	0	0.5	0	0	0	0	0
Organic layers	0	0	0	0.5	0	0	0.5	0	0.5	0.5	0.5	0.5	0	1	0
Dark A. saturated	1	0.5	0	0	0.5	0.5	0.5	0.5	0	0.5	1	0.5	0.5	0	0
Texture contrast and clay coatings	0.5	1	1	0.5	0	0.5	0	0	0	0	0	0	0.5	0	0
Presence of eluvial horizon	0	0.5	1	0.5	0	0	0	0	1	0	0	0	0	0	0
Albeluvic glossae	0	0	0.5	0.5	0	0	0	0	0	0	0	0	0	0	0
Subsoil accumulation of Al-complexes	0	0	0.5	0	0	0.5	0	0	1	0	0	0	0	0	0
Redoximorphic features	0.5	0.5	0.5	1	0.5	0.5	0	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0
Groundwater affected	0.5	0.5	0.5	0.5	0	0.5	0	0.5	0.5	0.5	1	1	0.5	1	1
Vertic properties	0.5	0	0	0	0	0	0	0	0	0	0	0.5	0	0	0
Base saturated throughout	1	1	0.5	0.5	1	0.5	0.5	0.5	0	0.5	1	0.5	0.5	0.5	1
Base saturation under 30% in a part of soil profile	0	0	0.5	0.5	0	0.5	0.5	0.5	1	0.5	0	0.5	0.5	0.5	0
Strongly calcareous in majority of the soil profile	0	0	0	0	0.5	0	0	0	0	0.5	0	0	0	0	0
Secondary carbonates to 100 cm	1	0.5	0	0	0	0	0	0	0	0	0	0	0	0	0
Accumulation of soluble salts	0	0	0	0	0	0	0	0	0	0.5	0.5	0	0	0	1
Clay content more than 30% throughout	0.5	0.5	0.5	0.5	0.5	0.5	0	0	0	0.5	0.5	0.5	0.5	0	0.5
Sandy texture	0.5	0.5	0.5	0.5	0.5	0.5	0.5	1	0.5	0.5	0.5	0.5	0.5	0	0.5
Lithic contact in first 25 cm or high volume of coarse fragments	0	0	0.5	0.5	0.5	0.5	1	0.5	0.5	0.5	0	0.5	0.5	0.5	0
Moderately developed profile	0	0	0	0	0.5	0.5	1	1	0	0.5	0	0	0	0	0
From fluvial sediments	0	0	0	0	0	0.5	0	0	0	1	0.5	0.5	0	0	0.5
Accumulation of eroded material	0.5	0.5	0	0	0	0.5	0	0	0	0.5	0	0	0	0	0

https://doi.org/10.17221/41/2019-SWR

The value of d_{ij} then represents the taxonomic distance between two soil classes *i* and *j*.

Correlation of the BSR with soil types of the CTSCS. The level of the BSR includes, beside the information on the soil unit, the classes of the skeleton content, texture, depth and parent material. Harmonisation of the structure of these classes with their counterparts in the CTSCS is indispensable for the use of the BSR in the correlation.

Harmonisation of the parent material class. In both classifications, the parent materials are grouped into classes according to their genesis and specific

Table 3. The dominant identifiers and codes attributed to each class of the Czech taxonomic soil classification system (CTSCS)

Identifier	Llitozem	Ranker	Rendzina	Pararendzina	Regozem	Fluvizem	Koluvizem	Smonice	Chernozem	Chernice	Sedozem	Hnedozem	Luvizem	Kambizem	Pelozem	Kryptopodzol	Podzol	Pseudogley	Stagnogley	Gley	Solonchak	Organozem	Kultizem	Antropozem
Man-made	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
Intensive agricultural use	0	0	0.5	0.5	0.5	0.5	0	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0	0	0.5	0.5	0.5	0	0	1	0
Organic layers	0	0	0	0	0	0	0	0	0	0.5	0	0	0	0	0	0	0.5	0	0.5	0.5	0.5	1	0	0
Dark A saturated	0	0.5	0.5	0.5	0.5	0.5	0.5	1	1	1	0.5	0	0	0.5	0.5	0	0	0	0	0	0	0	0.5	0.5
Texture contrast and clay coatings	0	0	0	0	0	0	0.5	0	0.5	0	1	1	1	0.5	0	0	0	0.5	0	0	0	0	0.5	0.5
Presence of eluvial horizon	0	0	0	0	0	0	0	0	0	0	0.5	0.5	1	0	0	0	1	0.5	0	0	0	0	0	0
Albeluvic glossae	0	0	0	0	0	0	0	0	0	0	0	0	0.5	0	0	0	0	0.5	0	0	0	0	0	0
Subsoil accumulation of Al-complexes	0	0.5	0	0	0.5	0	0	0	0	0	0	0	0.5	0.5	0	1	1	0	0	0	0	0	0	0
Redoximorphic features	0	0	0	0.5	0.5	0.5	0.5	0.5	0.5	1	0.5	0.5	0.5	0.5	0.5	0.5	0.5	1	1	0.5	0.5	0	0.5	0.5
Groundwater affected	0	0	0	0	0.5	0.5	0.5	0	0	1	0	0	0	0.5	0	0.5	0.5	0.5	0.5	1	0.5	1	0.5	0.5
Vertic properties	0	0	0	0	0	0	0	1	0.5	0	0	0	0	0	0	0	0	0	0	0	0	0	0.5	0
Base saturated throughout	0.5	0.5	1	0.5	0.5	0.5	0.5	1	1	1	1	0.5	0.5	0.5	0.5	0	0	0.5	0.5	0.5	1	0.5	0.5	0.5
Base saturation under 30% in a part of soil profile	0.5	0.5	0	0	0.5	0.5	0.5	0	0	0	0	0	0.5	0.5	0	1	1	0.5	0.5	0.5	0	0.5	0.5	0.5
Strongly calcareous in majority of the soil profile	0.5	0	1	0.5	0.5	0.5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Secondary carbonates to 100 cm	0	0	0	0	0	0	0.5	0	1	0	0.5	0.5	0	0	0	0	0	0	0	0	0	0	0.5	0
Accumulation of soluble salts	0	0	0	0	0	0	0	0	0	0	0	0	0	0.5	0	0	0	0	0	0	1	0	0.5	0
Clay content more than 30% throughout	0.5	0.5	0.5	0.5	0.5	0.5	0.5	1	0.5	0.5	0	0.5	0	0.5	1	0	0	0.5	0.5	0.5	0.5	0	0.5	0.5
Sandy texture	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0	0.5	0	0	0	0.5	0.5	0	0.5	0.5	0	0	0.5	0	0	0.5	0.5
Lithic contact in first 25 cm or high volume of coarse fragments	1	1	1	1	0	0.5	0	0	0	0	0	0	0	0.5	0	0.5	0.5	0	0	0	0	0.5	0	0.5
Moderately developed profile	1	1	1	0.5	1	0.5	0.5	0	0	0	0	0	0	0.5	0	0	0	0	0	0	0	0	0	0
From fluvial sediments	0	0	0	0	0.5	1	0	0	0	0.5	0	0	0	0.5	0	0	0	0	0	0.5	0.5	0	0.5	0.5
Accumulation of eroded material	0	0	0	0	0	0.5	1	0	0.5	0	0	0.5	0	0.5	0	0	0	0	0	0	0	0	0	0

https://doi.org/10.17221/41/2019-SWR

properties (content of $CaCO_3$, mineralogy). The coding is not used in the soil diagnostics, thus, no harmonisation of the categories is needed.

Harmonisation of the texture classes. The texture classes give information needed for the classification of the soil subtypes and soil types determined by the extreme texture. The SSS database provides data of the texture in different formats and detail: categories of texture according to the field description (feel test) for the B pits, categories of texture according to the content of the particles of 0.01 mm for the BSR and a detailed particle size analysis in the S pits. The texture is categorised based on Novák's classification of the texture classes into 7 classes according to the content of the particles smaller than 0.01 mm in the SSS (Table 4). In the BSR, the texture class is given for the topsoil (plough layer) and the subsoil. In CTSCS, the USDA (United States Department of Agriculture) triangle is adapted for the soil texture classification.

The texture classes from the triangle are then grouped into five groups and, as such, are used for the soil classification (Table 5). Thus, we deal with two different types of categorisation: First, the different size of the particle categories, second, the different texture classes. The limits of each particle's category are as follows: SSS – clay (< than 0.001 mm), silt (0.001–0.05mm), sand (> than 0.05mm), CTSCS – clay (< than 0.002 mm), silt (0.001–0.05 mm), sand (> than 0.05 mm).

The harmonisation of the texture classes has been performed using the dataset of the selective profiles (S pits) that include the detailed texture analysis. More than 135 000 soil horizons from 35 916 soil profiles have been examined with the aim to evaluate the accuracy of the correlation between Nováks' texture classes and the CTSCS and the texture triangle classes. If sufficiently accurate, the texture can increase the potential of the correlation between the two classifications.

Firstly, the fractions of the clay (< 0.002 mm) have been derived. Instead of the general pedotransfer function used in the CTSCS (Němeček et al. 2011), the approximation of the texture curve was based on the analytical data of the S pits. Several parametric models describing the texture curve have been tested (Zádorová et al. 2018). The parameters have been identified according to the least squares method when the curve has been interspersed with known values from the S pits. The tested methods included various logarithmic functions, spline functions and logarithmic-exponential functions (according to Weibull 1951), Gompertz's model (Nemes et al. 1999) or Freudlund's model (Fredlund et al. 2000). Due to the significant variability of the curves, none of the models was robust enough to be applicable for all the profiles. Finally, we applied a simple logarithmic-linear transformation that uses the two closest fractions for the derivation of the unknown fraction. Due to the closeness of the searched fractions and the known values, we consider this method relevant. The method uses the data from each analysis, a unique function is then used for the transformation.

Table 4. The texture classes according to Nováks' classification (Genetic-agronomic soil classification, GAC)

Content of particles < 0.01 mm (%)	Category by Novák
0–10	sandy
10–20	loamy sand
20–30	sandy loam
30–45	loamy sand
45–60	clay loam
60–75	clayey
75–100	clay

Harmonisation of the skeleton classes. The class of the skeleton content provides information on the volume of the coarse fragments (gravel and stones) in the soil. In the SSS, five classes are distinguished (0: 0–5%, admixture: 5–10%, S1: 10–25%, S2: 25–50%, S3: > 50% of coarse fragments content). For the BSR, the skeleton class is given for the topsoil and the subsoil. No classes of the skeleton content are available for the CTSCS, they are given in volume percentages. However, the classification limits that can be extracted from the definitions of the individual classes are put in the same thresholds as in the SSS (5, 10, 25, 50 and 80%). Thus, the data from the SSS can be directly used for the soil correlation. The inaccuracies in the correlation can be, in the case of

Table 5. The texture classes used in the Czech taxonomic soil classification system (CTSCS)

USDA triangle category	CTSCS category
S, LS	light earth
SL	light medium earth
L, SiL, Si	medium earth
SCL, CL, SiCL	heavy earth
SC, SiC, C	very heavy earth

USDA – United States Department of Agriculture

https://doi.org/10.17221/41/2019-SWR

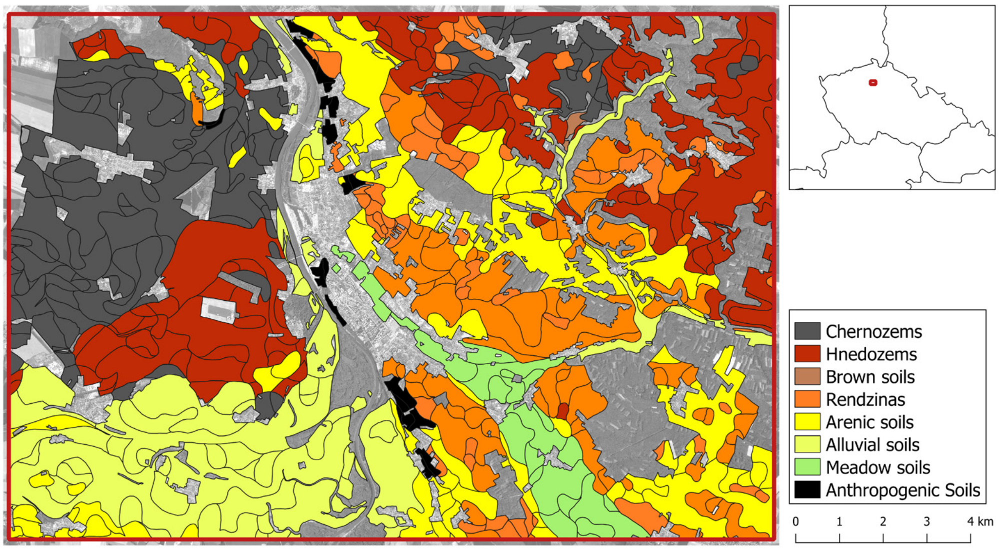

Figure 1. The pilot region delineation with the soil types in the Genetic-agronomic soil classification (GAC)

the skeleton content, caused not by the harmonisation itself, but by the vaguely defined criteria of the skeleton content in some soil classes defined in the CTSCS (soil types Rendzina, Pararendzina and some of the soil subtypes).

Harmonisation of the soil depth classes. The soil depth is classified into four classes in the SSS (shallow soil: 0–30 cm, moderately deep soil: 30–60 cm, deep soil: 60–120 cm, very deep soil: more than 120 cm). The soil depth is defined as the depth of solum to a layer containing more than 50% of coarse fragments or to a hard rock. The soil depth is not, with the exception of the Litozem soil type, used as a diagnostic criterion in the CTSCS.

Application of the correlation. According to the results of the taxonomic distance and the correlation of the additional soil criteria, the GSR and BSR have been transformed and a list of the most probable soil classes have been provided for each soil type and subtype in the GAC. The results have been visualised and validated in a pilot region (Figure 1), covering an area of 42 map sheets in the scale of 1 : 5000 (210 km^2). The area in the Mělník district has been selected because of a high pedodiversity of the mapped soil types, subtypes and parent material. Moreover, each individual soil type covers an area large enough to be statistically relevant and could be an object of validation. The soil cover is predominantly composed of luvic and mollic soils on loess, alluvial and semihydromorphic mollic soils on fluvial sediments and weakly developed soils on Cretaceous calcaric sediments and aeolian sands. Two types of maps have been produced - a map of the soil complexes showing the spatial distribution of the GSR and a map of the soil districts representing the distribution of the BSR.

Validation of the results. The accuracy of the correlation has been validated using a set of profiles described and analysed within the SSS in the study region. Overall, 1113 profiles with the morphological data (B pits) and 126 profiles with the morphological and analytical data (S pits) have been classified in the CTSCS. The profiles provide enough information for a relevant reclassification. Each of the profiles has been inspected and classified according to the decision trees for the conversion of the SSS profiles in the CTSCS (Figure S1 in the Electronic Supplementary Material (ESM)). The formal structure of the CTSCS prevents the use of a classification key and the gradual elimination of the not-fitting units. In the case of the CTSCS, we choose a taxon with the relevant diagnostic features from the whole set. This system is rather complicated in terms of an automated correlation and development of a conversion algorithm. A decision tree, that maintains the structure of the taxonomy to the greatest extent possible and enables an automated correlation, has been developed. Specific criteria typical for a limited number of units (e.g., a man-made soil, organic horizons, a high content of fragments, the shallowness of the profile) are placed first while the criteria characterising a larger group of soils (dark A, illimerisation, cambic horizon) are tested at the end of the tree. Two decision trees have been developed - one for the profiles

https://doi.org/10.17221/41/2019-SWR

with the morphological data only (B profiles), one for the profiles with the morphological and analytical data (S profiles).

The accuracy of the correlation has been evaluated using the overall accuracy method. The predictive maps of the converted GSR and BSR were overlaid by layers of the validation data (reclassified B and S pits). The data was compiled in a difference/error matrix. The accuracy assessment percentages were calculated using this matrix. The overall accuracy of the map is given by the percentage of the points whose predicted and real values correspond (Lunetta & Lyon 2004).

RESULTS AND DISCUSSION

Correlation of the GSR with CTSCS classes. Figure 2 shows the taxonomic distances (TD) between the GAC and CTSCS soil classes. The lower values mean the closer taxonomic relation. Generally, the results show a significant closeness of the concepts of the two classifications; a number of the soil classes have only one counterpart with a very low TD. The closest classes are Podzolised soil (GAC) and Podzol (CTSCS) with the distance 0.0; the reason of this absolute agreement is the concept of the soil types with the identical diagnostic features in both systems. Other soils with the taxonomic distance below 1.0 are Anthropogenic soils, Organic soils, Alluvial soils, Undeveloped soils, Brown soils, Rendzinas and Chernozems. Though some of these soils have the closest counterpart at a very low taxonomic distance, their other counterparts are also close. Rendzina can serve as an example: the closest unit lies at 0.7, but the taxonomic distance of the other six soil units range from 1.2–1.4. Analogically, Organic soil has its closest counterpart in Organozem (0.7), but it is very close to Gley and Stagnogley as well. Soils having only one relevant counterpart with a low taxonomic distance are the following units: Solonchak, Anthropogenic soils, Undeveloped soils, Meadow soils. In these cases, the soils have specific diagnostic criteria that are similar in both classifications. On the contrary, the soil units that show a high variability in the identifiable soil properties (humus content, base saturation, texture or hydromorphism) or have rather indistinctive diagnostic features, show low values of taxonomic distance with more than one unit from the correlated classification. Brown soil, Gley, Alluvial soil, Arenic soil, Rendzina, Stagnic soil and Hnedozem rank among these soils. Especially in these cases, the ambiguity and inaccuracy of the

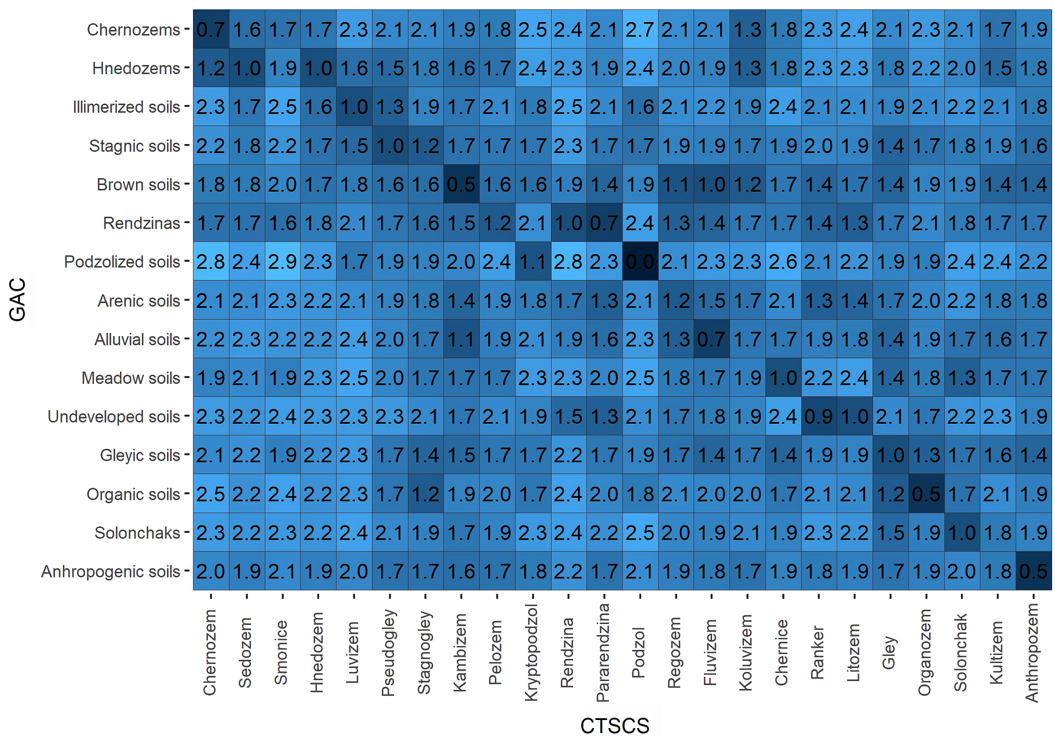

Figure 2. The taxonomic distances between the Genetic-agronomic soil classification (GAC) and the Czech taxonomic soil classification system (CTSCS) soil classes

https://doi.org/10.17221/41/2019-SWR

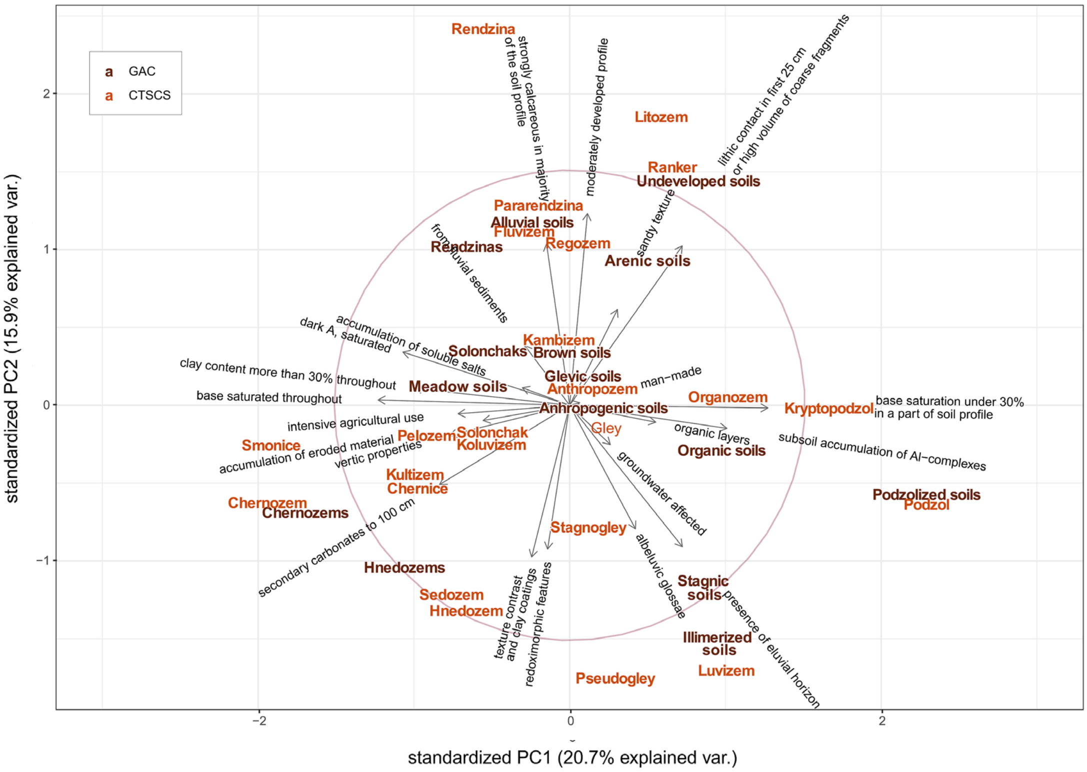

Figure 3. The PCA biplot

correlation must be taken into account and more of the close units should be considered a potential counterpart. Similar results are reported by Láng et al. (2016). They studied the possibilities of a profile classification using a TD and showed the analogical distribution of the TD efficiency with satisfactory results in the case of the soil units with strict and well-defined diagnostic features and less convincing for the soil groups similar to each other in various soil characteristics. Láng et al. (2013) used the TD for the correlation of the national soil classification with the World Reference Base (IUSS Working Group 2006). They found indistinctiveness in the brown soil correlation; pseudogleys, as in this study, were closest to Luvisols and Stagnosols.

The structure of the dataset and the relationship among the soil taxa and particular identifiers is shown in Figure 3. Despite the low significance (the principal components explain only ca 37% of the variance), the PCA biplot reveals some particularities in the dataset. One group of the soils is formed by young soils controlled by the weak development of the soil profile and high skeleton content. Saturated soils dominated by dark A horizons form another group. The close relationship between luvic and stagnic soils has also been shown while gleyic soils are located close to spodic and organic soils.

Table 6 shows the soil types in the GAC and their relevant counterparts in the CTSCS according to the taxonomic distance.

The whole GSR code refined by the soil subtype and erosion form enables a more accurate correlation. However, ambiguities are still possible. In the cases of the GSRs with a highly expressed ambiguity, an association of more soil types has been applied instead of an individual soil type. In Chernozems, the relevant counterparts resulting from the taxonomic distance illustrate a broad range of transitions towards the different soil types due to the processes of leaching, accumulation or the vertisol effect (Chernozem, Sedozem, Koluvizem, Smonice). However, in the case of the eroded form, Regozem must be considered a relevant unit, though its taxonomic distance is relatively high. This also applies in the eroded form of Hnedozem. The problematic classification of Rendzinas, divided into three equivalent units in the CTSCS (Rendzina, Pararendzina, Regozem), is related to its broader concept in the GAC, where it includes soils on both limestone and $CaCO_3$ rich materials (e.g., marlstone, calcareous sandstone) and with different amounts of coarse fragments.

Correlation of the BSR with the CTSCS classes. The BSR provides information on the class of the

https://doi.org/10.17221/41/2019-SWR

Table 6. The Czech taxonomic soil classification system (CTSCS) units with the closest taxonomic distance to the individual Genetic-agronomic soil classification (GAC) soil types

Soil type GAC	Closest units according to taxonomic distance
Chernozems	Chernozem, Koluvizem, Sedozem
Hnedozems	Hnedozem,Cernozem, Sedozem
Illimerized soils	Luvizem, Pseudogley, Hnedozem
Stagnic soils	Pseudogley, Gley, Luvizem
Brown soils	Kambizem, Regozem, Fluvizem
Rendzinas	Pararendzina, Rendzina, Regozem
Podzolized soils	Podzol, Kryptopodzol, Kambizem
Arenic soils	Regozem, Pararendzina, Kambizem
Alluvial soils	Fluvizem, Kambizem, Regozem
Meadow soils	Cernice, Gley, Solonchak
Undeveloped soils	Ranker, Litozem, Pararendzina
Gleyic soils	Gley, Ogamozem, Stagnogley
Organic soils	Organozem, Gley, Stagnogley
Solonchaks	Solonchak, Gley, Kambizem
Anhropogenic soils	Anthropozem, Kambizem

parent material, texture, skeleton content and soil depth. This information can give precision to the correlation of the particular soil units, namely at the level of the soil subtype. The class of parent material enables the distinguishing of the soil units derived from the Rendzina soil type - Rendzina and Pararendzina. Moreover, it can serve for the identification of the soil type Smonice determined by the specific clay sediments. At the subtype level, it determines the subtypes psefitic (gravel terraces) and fluvic (fluvial sediments).

Table 7 shows the distribution of the SSS texture classes into the CTSCS categories and qualifiers according to the texture analysis and harmonisation. The distribution indicates which classes can or cannot be converted with a high accuracy. The results show that the SSS category Clay can be attributed to CTSCS category 5 (very heavy earth) with a high certainty necessary for the identification of soil types Smonice and Pelozem. The category clayey soil can be attributed to categories 4 and 5 (very heavy and heavy earth) that together form the Pelic subtype. The classes of clay loam, loam and sandy loam cannot be relevantly attributed to any CTSCS category. The loamy sand category comes at a high probability under category 2 (light medium earth) delimiting the Arenic subtype in the Chernozem. The sandy soil category coincides with CTSCS category 1 (light earth) and the Arenic subtype at 92 %.

The skeleton content enables the identification of the soil types Ranker, Rendzina and Pararendzina. The relevance of the conversion is decreased by the absence of data for the whole profile (topsoil and subsoil) and namely by the vague criteria set for this feature in the case of Rendzinas and Pararendzinas in the CTSCS. The two soil types are categorised in the Leptosols reference group that is generally defined by a "significant amount of coarse fragments starting from the top", However, the quantitative limit of the skeleton content is not explicitly given in Rendzina and Pararendzina. The CTSCS defines Pararendzinas as "the soil that might be skeletic". The analysis performed in the study area showed that only 7 profiles from 197 classified as Rendzina in the GAC had a skeleton content of at least 25% in the subsoil, 13 had more than 10% and 81 more than 5%. According to the expert knowledge and the

Table 7. The number of the selective pits (S-pits) in the Systematic soil survey of agricultural soils in Czechoslovakia (SSS) and the Czech taxonomic soil classification system (CTSCS) texture classes and qualifiers according to the texture analysis and harmonisation

	USDA triangle categories	S	LS	SL	L	SiL	Si	SCL	CL	SiCL	SC	SiC	C
	CTSCS categories	light earth		light medium earth	medium earth			heavy earth			very heavy earth		
Nováks' categories	sandy	2 477	3 908	522	1	4	0	4	0	0	0	0	0
	loamy sand	4	2 008	17 026	844	299	0	7	0	0	1	0	0
	sandy loam	0	0	11 657	12 164	6 268	1	849	9	1	1	0	1
	loam	0	0	257	14 931	23 429	1	2 092	4 564	5 447	56	1	8
	clay loam	0	0	0	477	1 734	0	21	5 925	7 288	18	766	1 671
	clayey	0	0	0	2	15	0	0	68	1 442	0	2 531	2 402
	clay	0	0	0	0	1	0	0	0	29	0	851	1 435

https://doi.org/10.17221/41/2019-SWR

traditional concept of these two soil types, the limits have been assessed as 25% and more in the subsoil in Rendzina and at least 5% (category Admixture) in Pararendzina. The units with no occurrence of coarse fragments are classified as Regozem. In the BSR, the categories 0 and Admixture are unified in one skeleton class. These soils were classified as an association of Pararendzina/Regozem. The criteria in Ranker have been set at 50 % in the subsoil.

Application of the correlation. Figures 4 and 5 and Table 8 show the distribution of the correlated GSR and BSR classes in combination with the soil profiles classified according to their profile stratigraphy and soil properties. For the reason of good clarity, the figures and tables depict the correlation at the level of the soil type. The soil subtype level is shown in Tables S1 and S2 in the ESM.

The overall accuracy of the BSR and GSR conversion has been computed as a percentage of the soil profiles with an equal soil classification as the BSR and GSR units within which they spatially belong. The overall accuracy at the soil type is good when it goes beyond 70% in both cases. The slightly higher overall accuracy in the case of the BSR conversion (76.1%) is due to a more precise separation of the Rendzina soil unit using the skeleton content criteria. In fact, the performance is even more valuable if we take into account the fact that the number of the identified soil associations, covering the variable soils, decreased in the BSR conversion. This also applies for the overall accuracy at the subtype level. It decreases to ca. 65% and remains higher at the BSR units.

As for the conversion accuracy of the individual soil units, a high variability has been observed. Generally, the best conversion accuracy has been, as expected, reached in the soil units expressed as associations as they cover a more variable spectrum of the soil units. The best performance among the individual soil units has been achieved in Fluvizem, Kambizem (only 2 profiles), Pararendzina and Sedozem. In the case of Fluvizem, it is due to the equal setting of

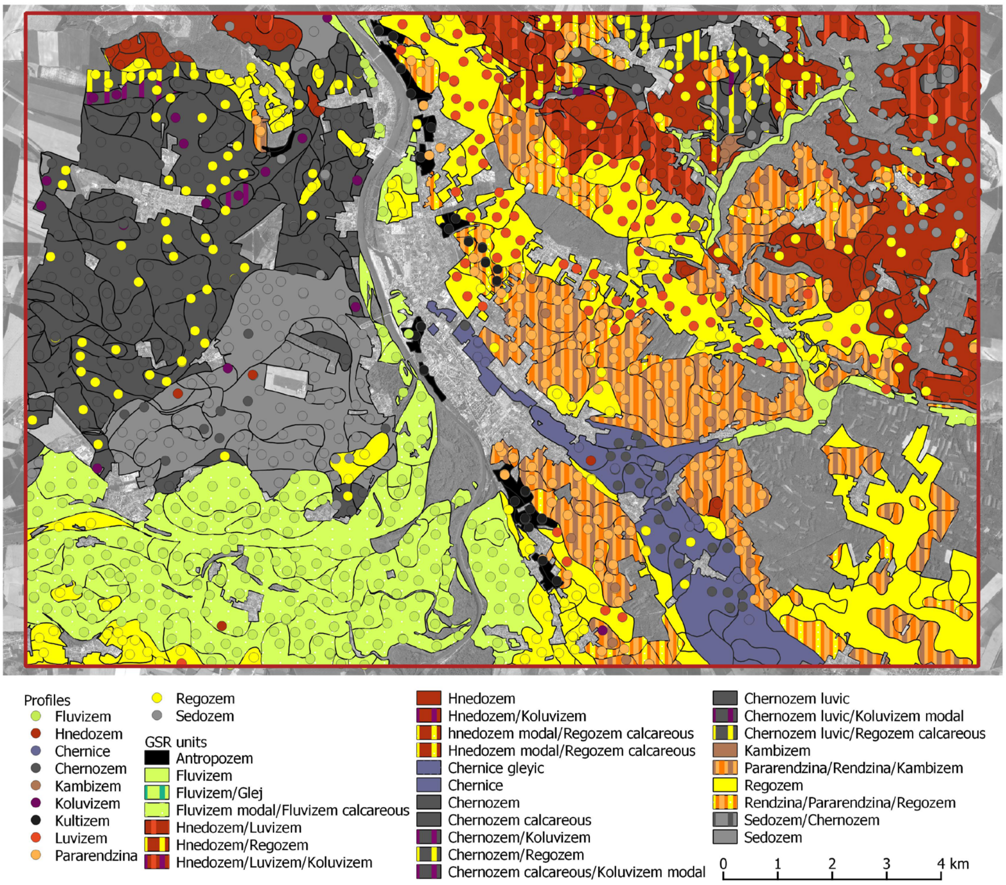

Figure 4. The general soil representatives (GSRs) converted in the Czech taxonomic soil classification system (CTSCS) nomenclature and the relevant soil profiles in the pilot region

https://doi.org/10.17221/41/2019-SWR

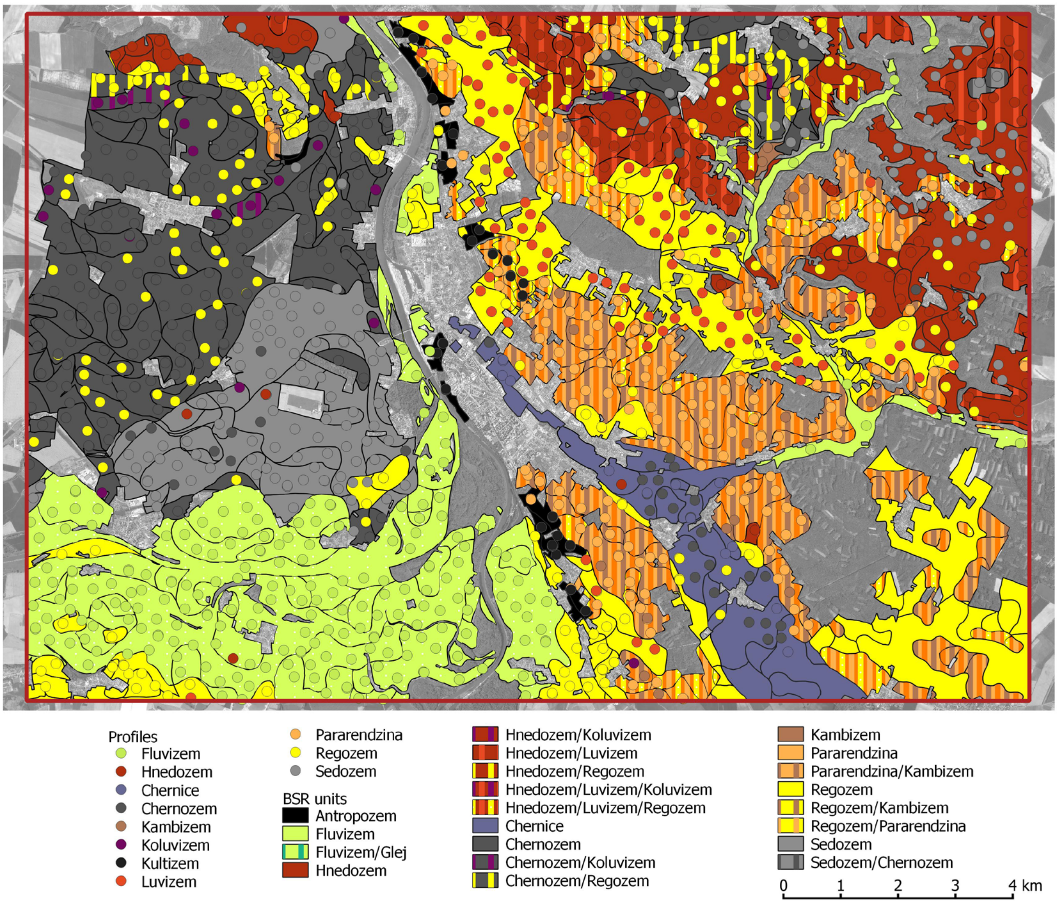

Figure 5. The basic soil representative (BSRs) converted in the Czech taxonomic soil classification system (CTSCS) nomenclature and the relevant soil profiles in the pilot region

the diagnostic properties and the limited profile information: we examined the gleyic properties as a potential indicator of the Gley soil unit; otherwise, the profile of the alluvial soil has been directly converted to Fluvizem. In the Sedozem BSR and GSR, almost 90% of the soil profiles corresponded with the classification. The mis-classified profiles were Chernozems luvic in the majority, thus, the closest soil group. Pararendzina, as an individual soil unit, has been distinguished only as BSR (the association of Rendzina/Pararendzina in the GSR) and the accuracy of its classification reached 91%. The results showed the reasonability of the Pararendzina/Regozem association for the BSR with a skeleton content expressed as 0/0: 75% of the belonging soil profiles have been classified as Pararendzina, the rest as Regozems. This also applies for the Pararendzina/Kambizem association. The accuracy of the Chernozem BSR and GSR conversion is lower, with 75% matching profiles. The profiles with a different classification included mainly Regozems and Koluvizems, the erosion derivatives of Chernozems, and Sedozems, that were classified in the BSRs of Chernozem luvic. The performance differs among the individual subtypes: Chernozem modal reached 73% while we observed lower accuracy in the other subtypes, 61% in Chernozem calcareous and 50% in Chernozems luvic (50% of the profiles not fulfilling the criteria of the SOC content or A horizon thickness were classified as Sedozem). The high percentage of profiles classified as Regozems correspond to the diagnostic properties set in the CTSCS, that in Chernozems include the rather strict criteria of the A horizon thickness, the SOC content and the colour. The Hnedozem BSR and GSR showed medium accuracy (51%), when the majority of the mis-classified profiles belong to Sedozem soil type. This is due to the identification of a specific feature - humus-clay coatings, that in the CTSCS implies the classification of Sedozem. The distinguishing of the Hnedozem/Luvizem association can be considered redundant as no Luvisol profile has

https://doi.org/10.17221/41/2019-SWR

occurred within the unit. The Chernice BSR and GSR showed low accuracy, when the majority of the relevant soil profiles fell within the Chernozem soil type. The results at the subtype level show that Chernozem gleyic is the most abundant soil in this unit. However, the very low TD between Meadow soil and Chernice promised more convincing results of this conversion. A major discrepancy has been observed in the Regozem BSR and GSR, when almost 43% of the profiles spatially belonging to this unit were classified as Luvisols (modal and stagnic). The majority of Regozems (namely arenic)

Table 8. The general soil representative (GSR) and the basic soil representative (BSR) converted in the Czech taxonomic soil classification system (CTSCS) nomenclature and the number of the relevant soil profiles according to their reclassification

Genetic soil representative	Soil profiles											
	Chernice	Chernozem	Fluvizem	Hnedozem	Kambizem	Koluvizem	Kultizem	Luvizem	Pararendzina	Regozem	Sedozem	conversion accuracy (%)
GSR												
Antropozem		2	1				6		19	6		0
Chernice	13	25		1						4		30.2
Chernozem		170				11				51	14	69.1
Chernozem /Koluvizem		1				4				1		83.3
Chernozem /Regozem		2								20	5	81.5
Sedozem		10		2		1					88	87.1
Sedozem /Chernozem											2	100
Fluvizem			214	1		1						99.1
Kambizem					2							100
Pararendzina /Rendzina /Kambizem					24	1		2	111	2		96.4
Rendzina /Pararendzina /Regozem				1			1		59	14		97.3
Regozem			3		2			80	3	75	2	45.5
Hnedozem		1		49	1					16	29	51.0
Hnedozem /Luvizem		1	1	55	1			1			1	93.3
Hnedozem /Regozem				11						13		100
BSR												
Antropozem		2	1				6		19	6		0
Chernice	13	25		1						4		30.2
Chernozem		166				10				33	14	74.4
Chernozem /Koluvizem		1				4				1		83.3
Chernozem /Regozem		2								20	5	81.5
Sedozem		10		2		1					88	87.1
Sedozem /Chernozem											2	100
Fluvizem			214	1		1						99.1
Kambizem					2							100
Pararendzina /Kambizem					24	1		2	111	2		96.4
Pararendzina							1		21	1		91.3
Regozem		4	3		2			80	3	93	2	49.7
Regozem /Pararendzina				1					38	13		98.1
Hnedozem		1		49	1					16	29	51
Hnedozem /Luvizem		1	1	55	1			1			1	93.3
Hnedozem /Regozem				11						13		100

https://doi.org/10.17221/41/2019-SWR

located on the parent material is characterised by a significant lithological discontinuity (aeolian sand over calcareous sandstone). In some cases, these soils have been classified as Regozems according to the upper material that markedly influences their agricultural functions. However, the diagnostic features described in the soil profiles (clay coatings, a decrease in the clay content below the Bt horizon) show the pedogenetic origin of the clay, urging us to classify the soil as a Luvisol. This conflict is evident in the database itself. The mapping units are determined as "arenic soils" while the majority of the respective profiles are classified as "illimerised soils". Thus, this poor performance can more likely be attributed to the problematic designation of the soil units in the SSS database than to the erroneous conversion. A zero overall accuracy has been reached in the Anthrosols BSR and GSR. The unit spatially intersects with the areas of vineyards with deeply trenched soils, mainly Pararendzinas. These profiles with a subsoil mixed down to 1 m by deep cultivation were classified as Kultizems – soils influenced by an intensive and long-term management. The soil type Anthropozems does not correspond to this kind of cultivation neither in the CTSCS nor in the GAC. Thus, we account this mis-classification again for a questionable determination of mapping units.

CONCLUSION

The methods of the taxonomic distance and harmonisation of the selected soil properties have been successfully applied for the purposes of the conversion of the large-scale national database into actual soil classification nomenclature. The results show a closeness of the concepts of the two classifications when a number of soil classes had only one counterpart with a very low taxonomic distance. The additional information on the soil skeleton content, texture, depth and parent material showed potential in the specification of some units. The harmonisation of the soil texture demonstrated that, due to the different categorisation of the soil particles, only some of the texture categories used in the SSS can be relevantly applied for the classification purposes.

The validation of the results in the study region showed the good overall accuracy for both spatial soil units, when the better accuracy has been observed in the BSRs. The usage of the soil associations have proven to be successful in the cases of the ambiguous assignment of the counterparts.

The major general conclusions are as follows: (*i*) the classification update is possible only to a certain level of accuracy that significantly varies among the different soil units, (*ii*) the conversion accuracy increases with the increased amount of the input data (both morphological and analytical) available for the classified entity, (*iii*) there is always some level of uncertainty that cannot be overcome because of missing information or different approaches to the soil profile description or classes settings used in the past, (*iv*) the level of uncertainty can result from a bias or the unclear definitions used in the classification system that disable the clear quantified application of the limits, (*v*) we recommend one to present polygon based maps as associations of the most probable units instead of the pure soil units.

REFERENCES

Arrouays D., Leenaars J.G B., Richer-de-Forges A.C. *et al.* (2017): Soil legacy data rescue via GlobalSoilMap and other international and national initiatives. GeoResJ, 14: 1–19.

Damaška J., Němeček J., Šimek J., Ryglevicz J., Mattauschová E., Haruda F. (1967): Survey of Agricultural Soils in Czechoslovakia. Vol. 2, Praha, MZVŽ. (in Czech)

Fredlund M.D., Fredlund D.G., Wilson G.W. (2000): An equation to represent grain size distribution. Canadian Geotechnical Journal, 37: 817– 827.

Hole F.D., Hironaka M. (1960): An experiment in ordination of some soil profiles. Soil Science Society of America Proceedings, 24: 309–312.

Hughes P., McBratney A.B., Huang J., Minasny B., Micheli E., Hempel J. (2017): Comparisons between USDA Soil Taxonomy and the Australian Soil Classification System I: Data harmonization, calculation of taxonomic distance and inter-taxa variation. Geoderma, 307: 198–209.

Hughes P., McBratney A.B., Minasny B., Huang J., Micheli E., Hempel J. (2018): Comparisons between USDA soil taxonomy and the Australian Soil Classification system II: Comparison of order, suborder and great group taxa. Geoderma, 322: 48–55.

IUSS Working Group WRB (2006): World Reference Base for Soil Resources 2006. Rome, FAO.

Krasilnikov P., Ibanez-Martí J.-J., Arnold R.W. (2009): The theoretical bases of soil classifications. In: Krasilnikov P., Ibanez-Martí J.-J., Arnold R.W., Shoba S. (eds.): Handbook of Soil Terminology, Correlation and Classification. London, Earthscan: 5–43.

Kozák J. (2010): Soil Atlas of the Czech Republic. 1st Ed. Praha, ČZU.

https://doi.org/10.17221/41/2019-SWR

Láng V., Fuchs M., Watner I., Michéli E. (2013): Soil taxonomic distance, a tool for correlation: as exemplified by the Hungarian brown forest soils and related WRB reference soil groups. Geoderma, 192: 269–276.

Láng V., Fuchs M., Szegi T., Csorba A., Michéli E. (2016): Deriving World Reference Base Reference Soil Groups from the prospective Global Soil Map product — A case study on major soil types of Africa. Geoderma, 263: 226–233.

Lunetta R.S., Lyon J.G.(2004): Remote Sensing and GIS Accuracy Assessment. Boca Raton, CRC Press.

Michéli E., Láng V., Owens P.R., McBratney A., Hempel J. (2016): Testing the pedometric evaluation of taxonomic units on soil taxonomy – A step in advancing towards a universal soil classification system. Geoderma, 264: 340–349.

Minasny B., McBratney A.B. (2007): Incorporating taxonomic distance into spatial prediction and digital mapping of soil classes. Geoderma, 142: 285–293.

Minasny B., McBratney A.B., Hartemink A.E. (2009): Global pedodiversity, taxonomic distance, and the world reference base. Geoderma, 155: 132–139.

Nemes A., Wosten J.H.M., Lilly A., Voshaar J.H.O. (1999): Evaluation of different procedures to interpolate particle-size distributions to achieve compatibility within soil databases. Geoderma, 90:187–202.

Němeček J., Damaška J., Hraško J., Bedrna Z., Zuska V., Tomášek M., Kalenda M. (1967): Survey of agricultural soils in Czechoslovakia. Vol. 1, Praha, MZVŽ. (in Czech)

Němeček J., Mühlhanselová M., Macků J., Vokoun J., Vavříček D., Novák P. (2011): Czech Taxonomic Soil Classification System. Praha, ČZU. (in Czech)

R Core Team (2018): R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna. Availale at https://www.R-project.org/.

Sirový V., Facek Z. (1967): Survey of Agricultural Soils in Czechoslovakia. Vol. 3, Praha, MZVŽ. (in Czech)

Weibull W. (1951): A statistical distribution function of wide applicability. Journal of Applied Mechanics, 18: 293–297.

Zádorová T., Penížek V. (2011): Problems in correlation of Czech national soil classification and World Reference Base 2006. Geoderma, 167–168: 54–60.

Zádorová T., Žížala D., Penížek V. (2018): Harmonization of the Database of Complex Soil Survey with the CTSCS and WRB 2014. Praha, VÚMOP. (in Czech)

Received: April 29, 2019
Accepted: September 2, 2019
Published online: October 21, 2019