Hungarian Geographical Bulletin 62 (2) (2013) 133–156.

Soil parent material delineation using MODIS and SRTM data

Endre DOBOS[1], Anna SERES[1], Péter VADNAI[1], Erika MICHÉLI[2], Márta FUCHS[2], Vince LÁNG[2], Réka Diána BERTÓTI[1] and Károly KOVÁCS[1]

Abstract

A digital mapping procedure was developed and tested to support the SOTER (SOil and TERrain digital database) database development. The SOTER mapping unit delineation is based on terrain and parent material information. Terrain information can be derived from SRTM data, but parent material information is often difficult to obtain. Legacy data are scarce and they are not always easily accessible. A procedure is needed to derive parent material information with limited legacy data support. The aim of this study was to develop a robust, remote sensing (MODIS and SRTM) based procedure to delineate soil parent material (PM) classes for small scale soil mapping applications. This quantitative procedure aimed to maintain the original mapping concept and to develop an analogue database to the "manually" created, existing SOTER databases. A simplified soil parent material classification was developed and tested for remote sensing based application. Multitemporal, visible, near infrared and thermal channels were used to compile the RS dataset which was later complemented with SRTM data and terrain derivatives as well. The suggested method can produce the first level delineations of the major PM units which can be further subdivided into smaller and more specific units as more data become available. The detail produced for the unconsolidated PM part of the pilot area is quite promising it shows all the major soil and landscape units that are important for soil mapping at the targeted - 1:1M -scale of SOTER. The consolidated part has much less detail, those areas still need legacy data for more detailed PM unit delineation. That procedure can be used anywhere in the world, where PM data are limited.

Keywords: digital soil mapping, SOTER, small scale soil mapping, digital terrain modeling, global soil database

[1] Institute of Geography, University of Miskolc, Hungary. E-mail ecodobos@uni-miskolc.hu
[2] Department of Soil Science and Agrochemistry, Szent István University, Gödöllő, Hungary, E-mail: micheli.erika@mkk.szie.hu

Introduction, rational and framework conditions for the soil parent material classification

SOTER database development

SOTER (Soil and TERrain Digital Database, ISRIC, 1993) is meant to replace the existing World Soil Map published by FAO. Its suggested scale is 1:1 M. The major limitation of the traditional SOTER products – as it has been identified so far - is the inconsistency between the national SOTER coverages. SOTER is designed to incorporate the existing data into a harmonized database. It is more like a correlation and a harmonization system than a mapping procedure. However, the spatial basis of correlation is the SOTER unit which should be consistent throughout the database, but this is not always the case. The delineations of the SOTER units vary from country to country. Therefore, the work to develop a more consistent mapping procedure – utilizing our state of the art in terrain modeling knowledge and the newly emerged SRTM data – to delineate the units started. Using SRTM as a common input data and a standardized digital elevation modeling toolset to define the terrain units guarantees the consistency between the databases of the different countries. A procedure to delineate SOTER terrain units was developed by Dobos, E. *et al.* (2005) and it was refined later by Dobos, E. *et al.* (2010a). Those physiographic units have to be further subdivided by the Parent Material (PM) information to define the final SOTER-units.

There are three potential situations which the SOTER database developers face when PM information has to be compiled:

– The first case assumes that there are existing and accessible legacy data for the whole area at an appropriate scale. It requires a harmonization effort of the input data sources as far as its thematic/attribute information is concerned, and a procedure to spatially incorporate, link the input PM polygons to the SOTER–units. That example was tested and a case study for incorporating the 1:1 M European Soil Database information into the SOTER database was demonstrated by Daroussin, J. *et al.* (2007). In that optimal case, a close to complete set of descriptive attributes can be loaded into the database.

– The second one is the limited data case when data are available for only a certain portion of the mapping area. It requires a harmonization effort, input data development efforts (digital mapping of PM info) and at the end a procedure to incorporate the info into the SOTER database. The latter one is the most typical case. The limited coverage PM info is used as training and calibration info for creating a full coverage for the whole area. That approach assumes that the areas with data represent the whole range of environmental, PM setups/variations. Hence, algorithms, rule systems can be developed to estimate the spatial distribution of the classes using environmental covariates like SRTM and MODIS data. Here, a limited set of PM information can be derived with varying accuracy depending on the environmental conditions and the quality and quantity of the training info.

– The third one is the no data situation when only general relationships and rules can be used to derive some delineations.

Parent material data are often limited or not accessible, which is already hindering the completion process of the SOTER database. PM classification – just like the soil one – varies from country to country as well. Therefore, a correlation system to harmonize the classes is needed. An international system has to have very general classes to be able to incorporate and correlate all national units. Both the thematic and the spatial correlations of the national PM data sets are difficult. Polygons coming from the national system inherit their own and often different way of delineations and interpretations of the PM classes. Importing those units introduces significant spatial inconsistency into the database. The only way to avoid that problem is to develop PM coverage in a controlled, quantitative way or at least increase the quantitative portion within the whole procedure. A quantitative procedure for PM delineations is needed to complete SOTER where no PM information is available and to increase the level of harmonization where legacy data can be incorporated into the database development.

Methods

Development of a simplified parent material classification for Digital Soil Mapping application

Information on the soil parent material is important to understand the current soil processes and properties better. One of the major limitations of the existing PM classification for soil scientist is the lack of expertise in geology. The majority of soil scientists have not got enough field knowledge to differentiate and classify certain rock types even if they occur in a pure stand. Classification of the rocks and parent materials is the responsibility of the geologists. There is a great diversity of PM definitions and interpretations in the national and international systems. Even the definition varies a lot.

There are general approaches like that of FAO, where the parent material is defined as the material from which the soil is presumably derived (FAO, 2006). Lawley (2009) defines the term as the geological deposits which immediately underlie the layers commonly known as 'topsoil' and 'subsoil' (Lawley, R. 2009). The most important source for a soil scientist is weathered, unconsolidated materials from which the soils are actually formed. That approach is shared by the Soil Science Society of America which defines the parent material as the unconsolidated and more or less chemically weathered mineral or organic matter from which a soil's solum is developed by pedogenic processes (SSSA, 2001; SSS & NRCS, 2007; Brady, N.C. and Weil, R.R. 1999). The US system and definition clearly defines that parent material is only an unconsolidated material in which soil horizons form and it is the only material which can be described in appropriate details by soil scientists.

Definitions are important, but the practice followed by the soil database developers is even more important. Many national systems provide information on the underlying consolidated rock rather than on the actual parent material, which are not necessarily the same. The most common example is a consolidated rock having a 2 meter thick loess cover on top. Soil is forming on the loess, but the identified PM is the consolidated rock type. That approach is misleading and it is also difficult to apply in the field. But most importantly, it does not provide appropriate information for the data users. Geological data are collected and kept by geologist. Soil science needs the description of the unconsolidated material from a soil formation point of view. A new, simplified system is suggested here to characterize the parent material, which can be applied for digital mapping procedures like digital terrain modeling and remote sensing.

The selected properties are important for the soil formation processes and can be easily described in the field with high level of reliability. The suggested system is shown in *Figures 1* and *2*. The first level of classification starts with the differentiation between the consolidated and unconsolidated parent materials.

Consolidated areas

Consolidated material is defined here as solid rock with its shallow weathering residuum with the typical mountain soil associations like bare rock/Leptosol/Cambisol. By genetics, it can be further classified as eluvial, colluvial or bare rock areas. Eluvial material is defined as in situ weathered material, weathering residuum, while colluvium is a loose deposit moving on slopes due to gravitational forces (in some cases water may play a role in the initiation of the movement). Expanding the content with the weathering residuum is basically an unavoidable compromise, because the existing soil maps with parent material information describe only the underlying rocks and they give no information on the properties of weathered materials. Information in the available legacy datasets exist only about the underlying rock. There is no reliable digital mapping procedure to derive geology/lithology information in the detail required by the existing PM classification of the SOTER methodology (ISRIC, 1993), so it is not possible to derive them using remote sensing. They often occur as mountains in the target area with relatively dense vegetation (no way for RS applications) with higher relief and elevation and a very high and large-scale diversity.

The consolidated areas are further subdivided into bare and none-bare rock. The non-bare rock areas usually have two subunits, eluvial and colluvial. However, the spatial mixing of those two is often too complex to differentiate between them at the target scale of 1 : 1 M (the only option is to give proportions within the geometric units). This is the detail in case of which quantitative procedure has to stop (Level of Genetics, *Figure 1*).

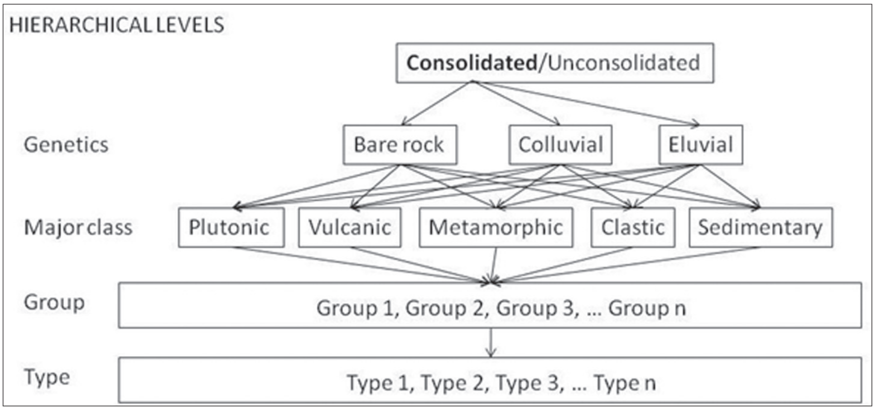

Fig. 1. Classification scheme for the Consolidated PM classes

Unconsolidated areas

An unconsolidated material is a loose inorganic/organic material which is – by nature – accumulated/deposited in a deeper stratum by wind, water or ice (aeolian, fluvial, estuarine, lacustrine, marine, glacial) or by mass movements (like the colluvial materials). The texture subclass of the unconsolidated areas can be delineated relatively easily with quantitative procedures using RS and DEM data with certain level of accuracy, but further separation of the lithology units is feasible only for areas having legacy data (*Figure 2*) when the classification can go further and the level of information defined by the revised lithology, classification can be filled into the database.

Based on the rationales described above, a new, revised classification system was developed (*Figures 1.* and *2.*). That system is designed to integrate the remote sensing techniques developed within the e-SOTER project where the consolidated/unconsolidated texture of unconsolidated materials, bare rock, alluvial, aeolian, marine and lacustrine material images were developed using RS and digital elevation modeling techniques. These layers can be used

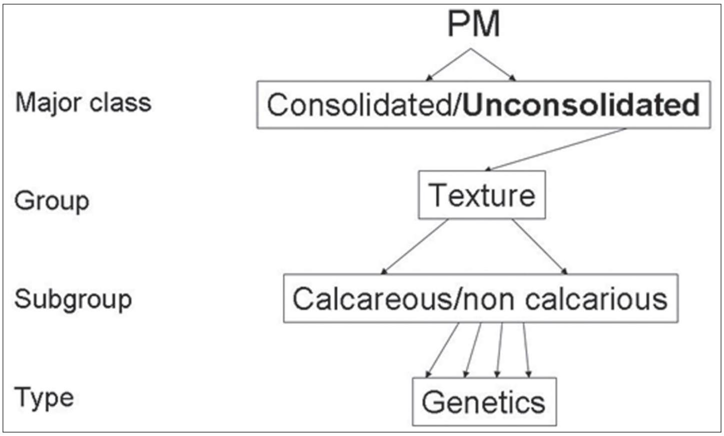

Fig. 2. Classification scheme for the Unconsolidated PM classes

as standalone information layers or they can be combined in any combinations depending on the data availability to refine the parent material information. That approach creates images with equivalent priorities, no hierarchy is kept in the system any more. The more input data and input layers, the better reliability and parent material map detail. At the end, the suggested system can provide the spatial detail needed for the SOTER database development.

The attribution of the units with legacy data happens with respect to the data availability. One of the major advantages of the system is the lack of hierarchy in the development process. In a hierarchical system, a classification tree is used to classify the phenomenon. The naming stops at the first level where data are missing. All pieces of information below which level are lost for the classification. Making the levels independent from each other helps to maintain each available information in the system and it fits a GIS database approach better. The system was tested using national databases and severe limitations were identified due to the hierarchy of the system.

Determination of soil parent material within the landform units based on low-resolution satellite imagery (MODIS) and DEM (SRTM) in combination with legacy soil parent material data

The study area

The pilot area is located in Central Europe and it covers the area of Austria, Hungary, Slovakia, Czech Republic, Southern Poland and a small part of Germany and Romania. The pilot window was chosen to cover the Central European pilot area of the e-SOTER project *(Figure 3).* The final area is much larger than the pilot area, it follows the tile borders of the SRTM (Farr, T.G. and Kolbrick, M. 2000) and MODIS images which include the whole e-SOTER pilot. Training data were available only for the pilot window and the territory of Hungary.

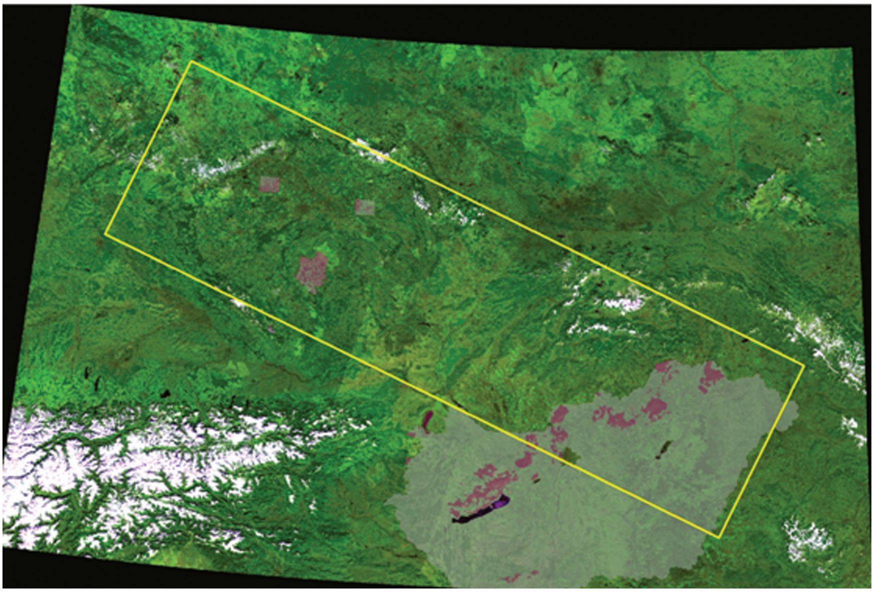

Fig. 3. The Central European window (yellow box) and the training areas for the consolidated PM image

The terrain and the soils of the area are quite variable. It includes some parts of the Alps, the Carpathian mountain range, the Czech–Moravian Mountains, the Pannonian Basin and the southern, hilly and flat region of Poland. The parent materials vary as well, all kinds of consolidated siliceous and carbonaceous rocks occur in the area together with Holocene alluvial and aeolian sediments and Pleistocene glacial and periglacial materials. The soils in the lowland are mainly Chernozems, Vertisols, Arenosols, Gleysols and Calcisols, while in the hilly and mountainous areas Luvisols, Cambisols, Stagnosols, Regosols and Leptosols are the dominant ones (FAO, IUSS Working Group WRB, 2007). Erdas Imagine 9.3 and ArcGIS 9.3.1 software were used for processing and classifying the data layers.

Covariates used to derive the thematic layers

In order to strengthen the performance of the classification, multi-temporal images of none-altered MODIS bands were compiled into an image of 55 layers representing the visible, NIR, MIR and thermal bands to capture the temporal environmental conditions and changes revealing to surface conditions. However, the 55 layers have a significant portion of overlapping information, redundant info in the images, hence a PCA was used to decrease the number of input images and decorrelate the information of the bands. The best 15 PCA components were maintained and incorporated into the final image.

MODIS-multi-temporal 8-day composites were used, 5 dates are evenly distributed over the vegetation period:

– MOD09A1: Band 1–7 (Layers 1–7), 500 m resolution,

– MOD11A2: Band 31–32 (Layers 9–10), LST (Land Surface Temperature) Day (Layer 1) and LST Night (Layer 5), 500 m resolution.

Surface temperature information like the thermal bands of the MODIS (Bands 31, 32) and the LST (Land Surface Temperature) products (night and day) having been derived from them were used as well. The daily temperate fluctuation is a function of the thermal capacity of the surface materials, actually a function of the parent material, texture, color and water content, basically the factors we are interested in. Therefore, a new normalized band combination was developed. The daily temperature differences were calculated with simply subtracting the LST night from the LST day and the values were multiplied with the ratio of the LST_(max for the whole area)/LST_(day) to reduce the effect of the climatic variation due to the differences in potential energy intake from the sun.

There were many attempts recorded in the literature to use band ratios to identify certain lithology classes or to highlight/enhance lithology differences in Landsat images (Knepper Jr., D.H. and Simpson, S.L. 1992; Vincent, R.K. 1997). Three band ratios, 6/1, 1/3, 7/6 were selected to represent lithological variations better. The band ratios were adopted to MODIS and they were derived for each of the 5 dates, resulting 15 new images added to the final image.

SRTM (Farr, T.G. and Kolbrick, M. 2000) data were used in combination with the MODIS derived layers as well. The basic parameters are the followings:

– Elevation (sinks are filled up to certain level),

– Slope percent,

– Relief Intensity,

– Potential Drainage Density (Dobos, E. *et al.* 2010a),

– Groundwater level (developed via the interpolation of the drainage network height and it is subtracted from the original elevation values),

– Topographic Wetness Index,

– Upland/Lowland,

– Convexity (not added to the basic image, used only for the colluvial image derivation).

The listed derivates have either been used in the SOTER methodology or they are believed to add significant information to make difference between the classified parameters. The SRTM images were spatially degraded to the level of MODIS resolution and an image of 43 layers containing 15 PCA layers, 8 SRTM derivatives, 5 normalized LST difference images and 15 band ratios.

Besides of the 43 layers described above, three further layers were added to the image to represent the climatic variability: the images of Easting and Northing to represent the geographic location. The distance from the sea was calculated as well. With the three further layers, an image of 46 layers was developed and used for the classification.

The PM classification procedure

The procedure is developed for situations with a limited set of PM variables available when only basic properties can be estimated with a relatively high level of reliability/accuracy. Those properties are the consolidated status, the texture classes and the genetic classes of fluvial, marine/estuarine, aeolian, colluvial, eluvial and bare rocks. Large-scale studies can be done for further refinement and more attributes, but they are site specific. No generally adaptable procedure exists.

PM is considered as unconsolidated material characterized by its texture, carbonate status and genetic classes (*Figure 2*). The three (four with the consolidated/unconsolidated) parameters require to develop four separate layers. The first layer is the consolidated/unconsolidated one, also being the first step in the procedure. It stratifies the area into two classes. The two main areas require different approaches and different classification steps.

Pre-stratification of the area to consolidated/unconsolidated classes

The first step in the process is the classification of the window into two classes (consolidated and unconsolidated). Maximum likelihood supervised classification algorithm using the combined, base image of 46 layers was applied to derive the image (Dobos, E. *et al.* 2010b). Several direct approaches were evaluated, but no one had an appropriate overall performance. There are only stochastic relationships between certain terrain parameters and the consolidatedness of the PM and the same is true for the RS images, especially in the temperate and tropical zones where the vegetation masks out the PM signal of the images. The stochastic relationships can be utilized in a supervised classification framework well.

Training data were limited to the window, as only 10 percent of the whole area was covered with legacy data. We used three small training areas for the Czech Republic and the Hungarian part of the window (*Figure 3*). The data sources had to be interpreted in the training areas for the classes defined.

Parent materials for the areas having consolidated rocks with shallow unconsolidated material /soil on the top are considered as consolidated and they are named according to the underlying materials such as granite, basalt, etc., regardless of the source of the materials, whether they are in situ materials or mixed ones with other aeolian or colluvial sediments. It is a traditional approach taken from the existing legacy databases. Anything having a bare rock, Leptosol/Cambisol soil association in a mountainous area is considered as a solid rock even if it does not appear directly on the surface. The training classes were merged that way, meaning that the in situ, weathered, relatively shallow eluvial or colluvial materials belong to the consolidated rock system and they are not considered to be independent unconsolidated strata.

Classification of the consolidated material areas

The classified consolidated areas are divided into three further major classes: Bare rock, Colluvial and Eluvial. The units can be further described with legacy data, but that one is the final stage for areas where legacy data are not available.

Bare rock

The classification of the bare rock was done using an NDVI (Normalized Difference Vegetation Index) (Kriegler, F.J. *et al.* 1969; Tucker, C.J. 1979) image generated from the peak season (early summer) of the vegetative period when strong vegetation cover is expected. Only areas having no soil and thus vegetative cover are expected to have very low NDVI value. A threshold of 0,8 was set to select the low NDVI areas as part of the bare rock class. That value was set by comparing the images with Landsat and other high-resolution images where the existence of the bare rock surface was evident. The threshold value is date and climatic zone dependent.

Colluvial areas

The colluvial areas were delineated applying the assumption that the colluvial materials accumulate in the lower sections of the slopes often converting the shape of the slopes into concave. The assumptions were translated to the following SRTM based decision rules:

- Curvature < 0.00
- Slope percent > 2.00

Classification of the unconsolidated materials

Based on the revised lithology classification, there are three property groups within the unconsolidated material:

- Texture:
  gravel, sand, loam, clay, diamicton (organic material);
- Carbonate status:
  calcareous and non-calcareous;
- Genetic:
  fluvial, aeolian, lacustrine, marine, estuarine, glacial till, glaciofluvial.

Out of those properties, the texture and the selected subgroups of the genetic classes were targeted to define as a minimum set of PM descriptions, namely the fluvial/lacustrine, aeolian and marine genetic classes.

Developing the texture class layer

The texture classification was done the same way as the consolidated/unconsolidated layer using the 43 layer combined image and a texture class training data set as inputs for the Maximum likelihood supervised classification. That step of the procedure requires a thorough knowledge of the area (for validation purposes) and the application options of the classification tools to achieve the best optimal results. No automatic approach can be developed, an expert user is needed.

Training data

Training data are the most critical part of the procedure. In an optimal case relatively high resolution training data are available with well-defined, none-overlapping classes. 1 : 100 K to 1 : 250 K data sources are commonly available in the developed World.

The data sources contain aggregated but still concrete classes (not associations). These data sources can be used as direct inputs for the supervised classification.

Development of the genetic layers

The SOTER procedure identifies several genetic PM classes, namely the fluvial, alluvial, lacustrine, glaciofluvial, marine and estuarine, aeolian, older terraces and glacial till plain. Those materials often have major differences in their origin and composition, but some of them look very similar from a geomorphological point of view. Taken the case of limited knowledge and information about the origin of the material, the only available information to characterize the PM is geomorphology derived from digital elevation models, like SRTM. Having only the geomorphology as separating tool, several classes with similar morphology but different origin had to be merged.

Terrain modeling alone is not enough to explain the origin of the PM, but it is often enough to delineate the spatial units by highlighting the changes in the surface morphology. The allocation of descriptive variables to the spatial delineations – name of the genetic classes – can be done using expert knowledge for the certain area.

The following combined classes and genetic layers were developed:

– Fluvial/alluvial/lacustrine/glaciofluvial:

– Plain, low slope and low relief intensity areas close to the groundwater level;

– Marine and Estuarine (does not exist in the pilot window, not used in this study):

– Follows the seashore line and it is characterized by 0–5 m elevation along the seashore;- Aeolian/older terraces/glacial till plain:

– Higher relief, higher above the groundwater level, not influenced by the recent fluvial activities.

Fluvial/alluvial/lacustrine/glaciofluvial class

This combined class has a plain, smooth surface, low relief intensity and small depth to the ground water/surface water level. Those assumptions were translated to terrain modeling language. The final solution for the delineation has only one criterion, namely the depth to the surface/groundwater level system. Ground/surface water level grid was simply extracted from the SRTM DEM and the difference is the depth to the ground water (DGW). The layer of groundwater level was developed using a simple modeling exercise.

First, a surface drainage/channel network was created by selecting the pixels as drainage pixels which have more than 50 pixel contributing area. The pixels were selected and the corresponding elevation pixel values were copied from the SRTM to those pixels. The pixels were interpolated to create a continuous surface. That image was extracted from the actual elevation to create the DGW image. That difference image was thresholded using a value of 3 to create the final image. Areas having values lower than the threshold were classified as low lying areas being endangered by potential annual floods by both the rivers and lakes, basically all kind of surface waters, therefore they are smooth and plain. That is why the fluvial and lacustrine sediments are combined in this classification. The glaciofluvial areas have similar appearance to the fluvial ones, only the source of the deposited material is different.

Aeolian/Older terraces/Glacial till moraines areas

Aeolian areas are the ones which are free from flooding impact and also from significant ground water impact. Therefore, the unconsolidated, non-fluvial areas – which are often dry and therefore the object of wind erosion and deposition – are the potential Aeolian ones reworked by the wind or covered by wind-blown sediments. Those areas have slightly higher relief intensity due to the longer acting dissecting processes or to the dune nature of the wind carved/built surfaces.

The most problematic genetic class is the Glacial till. Tills can be till plains which are relatively plain (low relief intensity), but lying higher than the fluvial areas.

They can be identified with selecting areas with higher than 3 m elevation above the water level and have a plain/low relief intensity surface. Glacial till moraines have much higher relief, sometimes similar to the Aeolian areas and they lie above the fluvial areas as well. That is why the Glacial till class was merged with the Aeolian. However, no area with significant amount of glacial tills occurs in the pilot window.

Finalizing the PM coverage

Input data:

- from satellite image classification
- consolidated-unconsolidated image,
- texture (consolidated, gravel, sand, loam, clay, peat, sapropel, diamicton, water);
- from SRTM DEM derivation
- alluvial/Aeolian,
- bare rock,
- marine/estuarine.

At the final stage of the PM development procedure the previously developed layers were integrated into a combined PM image.

Results and discussion

A novel approach of parent material delineation was developed and tested to support the SOTER completion process. SOTER is a polygon based database where the polygons are defined based on physiography/major landforms which are further subdivided by parent materials. One of the major limitations of the SOTER database development is the great variability of the input datasets having different systems of soil classification, mapping and delineations. The classes of different soil classification systems can likely to be correlated to reach a common platform, but the spatial units, namely, the polygons – are very difficult to change. Redrawing the units means a totally new polygon system, a new mapping campaign. That is why the existing international SOTER datasets have severe edge matching problems highlighting the thematic and spatial inconsistency of the datasets. It is an inherent property of the SOTER procedure, there is no way to avoid the use of the traditional SOTER mapping approach. That problem was identified and the need for a coherent developmental approach was expressed by the soil science community (Worstell, B. 2000).

Dobos, E. and Montanarella, L. (2004), Dobos, E. *et al.* (2005, 2010a,b) have developed a quantitative procedure to develop the terrain units of the SOTER database using GTOPO30 and later the SRTM database. That procedure was a feasible solution to solve the inconsistency problem, but the problem of lacking a coherent and consistent PM data layer to further subdivide the terrain units was still unsolved.

This study aimed to develop a quantitative method to develop the PM layer needed for the SOTER unit development using remote sensing and digital modeling tools. That system was designed to provide homogeneous spatial units of PM for areas lacking appropriate PM information. It does not require detailed, classified input data and the approach does not even aims to develop the classes. What it does is basically a stratification of the land surface based on very basic properties of the parent material and the surface landform types correlated strongly with the PM. The system is able to delineate the homogeneous PM units at a general level, but it cannot describe them at the level needed for the SOTER PM classification. It defines the PM polygons but it cannot fill up its record completely in the PM attribute table.

Besides of the remote sensing and digital terrain modeling approach, the major novelty of the system is its disaggregated nature. There is no hierarchy in the system, only property layers are created and combined to reach the appropriate level of delineation. The first layer is the consolidated/unconsolidated image (*Figure 4*).

The image was developed using a maximum likelihood classification of MODIS and SRTM data. Consolidated PM in the target area is mainly represented by the mountain. The mountains were lifted up and all uncon-

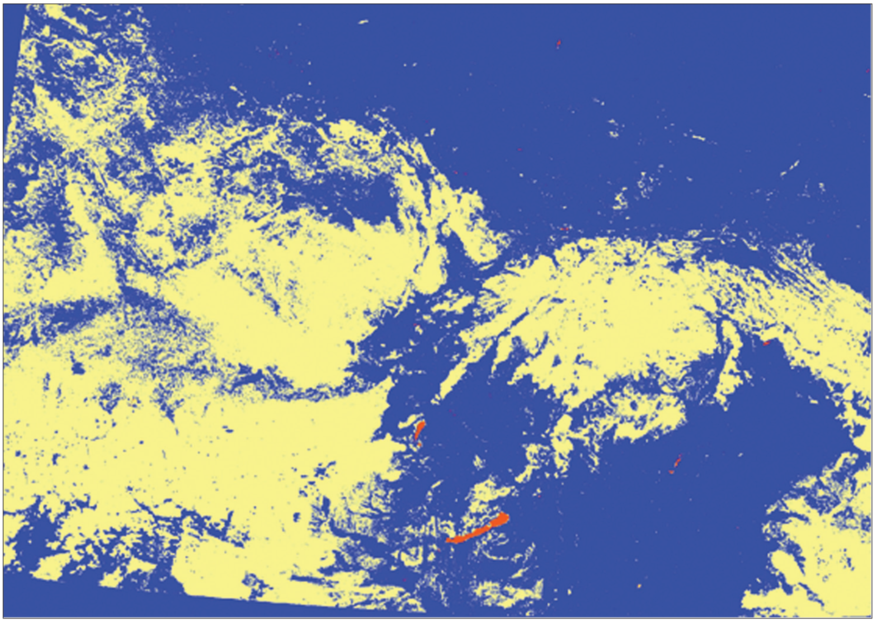

Fig. 4. The consolidated/unconsolidated image. Blue is the unconsolidated, while the yellow is the consolidated PM class. Orange color is surface water

solidated material were eroded away and resulted in the outcropping of the consolidated materials. The mountains are characterized by higher elevations, higher slope, higher relief intensity and they have mainly forest cover on them. Unconsolidated materials were accumulated in the basins and they filled them up to the current elevation. Accumulation happened mainly through fluvial and partly by aeolian processes resulted in a more or less flat surface for the alluvial areas or a little bit higher relief for the aeolian (reworked sand) sediments. The lower laying areas are under mainly agricultural use.

The two areas are so different from each other, both in the relief and in the land use that the combined image of MODIS and SRTM did a good job on the classification. The consolidated PM area is slightly over-classified, probably due to the forest cover. Areas having a dominant forest cover and higher relief were classified as consolidated materials. It leads to the over-representation of the consolidated class in the Southwestern part of Hungary. West from Lake Balaton and also in the hilly regions between the mountain ranges Pannonian sea sediments were accumulated and uplifted. However, in general, the consolidated-unconsolidated PM areas were classified with a relatively high performance and they can serve as a good basis for the further stratification steps.

Classification of the consolidated PM class

There were two more additional steps in the procedure for the consolidated areas. The first step was to classify the bare rock areas with no soil cover and vegetation on it while the second step was to separate the colluvial and eluvial areas. Classifying the bare rock areas is important, because those areas are excluded from the further soil classification process. Due to the 500 m resolution, only larger, contagious areas of bare rock were identified, whereas smaller spots were not recognized in the process. Significantly large areas occurred only in the High and Low Tatras and in the Alps.

After the exclusion of the bare rock regions, the remaining area were classified into two classes, namely the eluvial areas developed by in situ weathering and the colluvial one where unconsolidated sediments were deposited at the lower part of the hill slopes. Based on our expert knowledge, the slopes were divided into two parts.

The upper one has convex or linear slopes and it is characterized by erosion or in situ weathering. The lower section of the slope is the depositional surface where the colluvic materials are the dominant ones and which has a concave shape. Having only those two classes, one class defines directly the other one as well – pixels not classified as colluvial would be automatically classified as eluvial.

The classification was done using the curvature command of ArcGIS and the pixels forming a concave surface were identified as colluvial area, while the rest of the pixels were assigned to the eluvial class *(Figure 5*). The SRTM image based classification has a relatively high resolution (90 m) with many slope scale patterns. The combination or aggregation of the 90 m pixels into the 500 m resolution of the overall system was totally meaningless. Resampling the pixels to 500 m resolution resulted in a mixed and often balanced composition between the two classes, simply averaging out the two properties for the whole area. Therefore, the colluvial-eluvial classification was concluded not to be relevant at that scale where only the major landforms are taken into consideration and no slope scale patterns are considered.

Significant and relevant information to further subdivide the consolidated PM area – other than the bare rock regions – could not be derived from RS and SRTM images when legacy data were not available.

Classification of the unconsolidated PM class

The unconsolidated areas identified by the consolidated/unconsolidated image were further subdivided by the texture image in two steps. *Figure 6* shows the estimated texture classes of the pilot area.

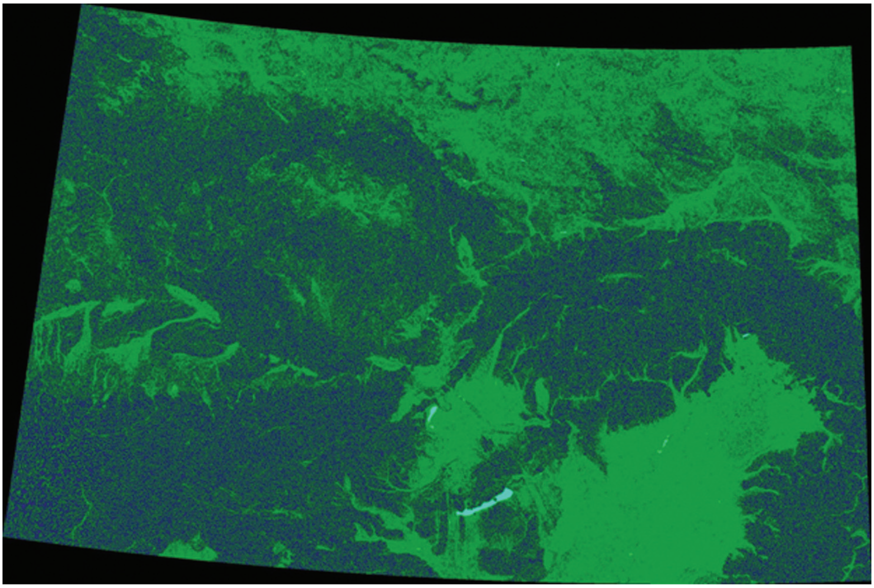

Fig. 5. Colluvial areas colored dark blue (curvature<0, slope%>2)

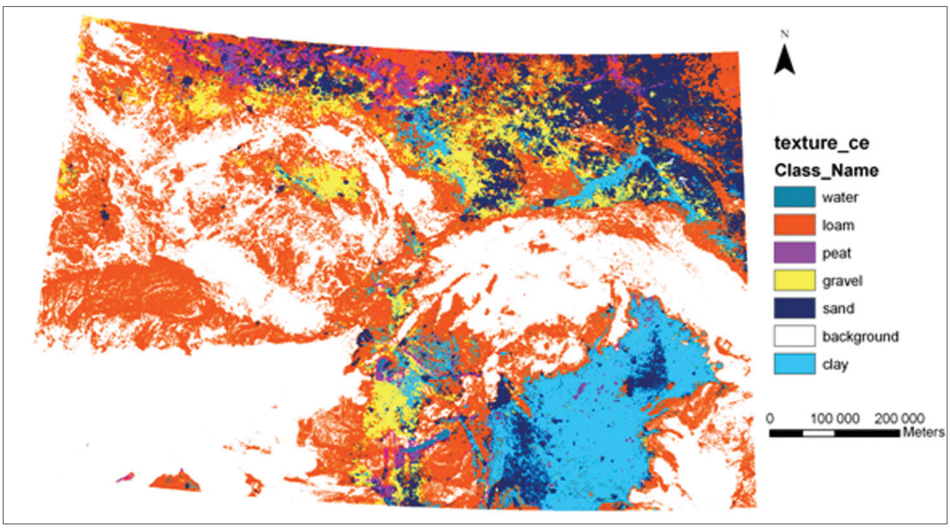

Fig. 6. The classified texture classes (background class means the consolidated areas where texture classes were not estimated)

The Hungarian part of the image was interpreted by a group of experts. That image corresponds well with the known picture of the soil class distribution of the area. The only problematic area was the Great Hungarian Plain where almost all of the clay-loam areas existing there disappeared from the image and were dissolved in the clay class. The silt-loam areas like the Transdanubian area with huge contiguous loess cover were classified correctly. There are two potential reasons for the regional misclassification. The first one is probably the limited capability of the covariates to distinguish between the clay-loam and clay classes. The two classes look very similar spectrally to each other and they occur on a very similar landscape.

The difference between the two classes is only 10 percent of clay increase, 30 to 40 percent clay required for the clay-loam class and over 40 percent clay is needed for the clay class. The majority of the low lying areas of the Hungarian plain are clay-loam. There are only smaller regions, like Taktaköz or a part of the Bodrogköz area where the clay content is high enough for the clay texture class.

The second reason for the low separability between the loam and clay classes is probably caused by the preprocessing steps of the training dataset. The training sites were point observations with known texture class or having only diagnostic horizons corresponding to certain texture classes, like Arenosols to sand or Vertisols to clay. There were 951 points in the whole pilot area where a decision on the existence or non-existence of the Vertisols could be made and 140 out of them were classified as Vertisols, all of them located on the Great Hungarian Plain. By definition, Vertisols need to have at least 30 percent clay. There is no Vertisol soil type in the Hungarian classification system, so the classification or the allocation of the profile of the Vertisol reference soil group of the WRB was done based on only the texture using a simplified rule. Soils having more than 30 percent clay and a thickness of 25 cm or more were classified automatically to the Vertisol reference group and all of them were used as training samples for the clay classes. However, it was a mistake, because the majority of them had less than 40 percent clay, not meeting the clay texture class requirement, but matching with the one for the Vertisols. Using th 140 sites as clay areas widened and distorted the clay class histogram, which later caused a significant over-classification of the class.

The texture of the unconsolidated sediments on the plains has a strong correlation with the origin of the parent material in Central Europe. The classes make a real and significant difference between the soil forming processes of the plain areas and they separate the different soil associations. Therefore, the texture was used to refine the stratification of the plain areas and divide them into two subclasses of fine and coarse textured parent materials. The soils developing on the sand and gravelly-sand (called gravel in the image of *Figure 6*) areas are totally different from the ones forming on loamy or clayey materials. Further differentiation within the fine texture class to loam and clay classes would have been advantageous, but the input texture image did not make a good separation between the classes. Areas having clay-loam texture – very common in the alluvial area of the Pannonian plain – were classified as clay. Therefore clayey soils are artificially overrepresented in the image of the target area. Luckily, the separation between the fine and coarse textured areas is much more reliable and it makes a good input for further classification.

Genetic classes, the delineation of the Fluvial/Alluvial areas

As the last step in the procedure, the basic geomorphological characteristics were used to classify the origin of the surface material. Two SRTM derivatives were used for that purpose, the relief intensity and the depth to groundwater (DGW). The relief intensity is the elevation range within a unit size area. Low relief means level surface, alluvial areas, while higher relief intensity alludes to Aeolian sedimentation or older, uplifted and more dissected areas.

The second terrain property, the depth to groundwater referred to the elevation of the area over the current surface drainage network. Small DGW values highlighted that the low lying areas, namely the recent alluvial areas had a higher chance of annual flooding. As it was concluded, the latter one – the DGW – alone was able to separate the alluvial areas, further differentiation using the relief intensity did not added significantly more spatial details. Therefore, the relief intensity factor was not even integrated into the final procedure. *Figure 7* shows the image of the alluvial areas.

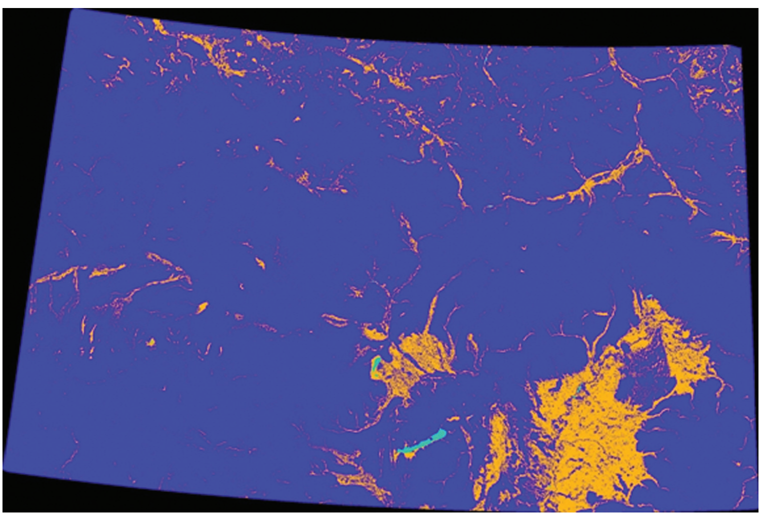

Fig. 7. Fluvial/Alluvial sediments (yellow) of the pilot area

All the current and recent floodplains are classified correctly. Unconsolidated areas not classified as alluvial refer to older terraces or Aeolian sediments. At the end the two major genetic-class groups of the unconsolidated areas, namely the Fluvial/alluvial/lacustrine/glaciofluvial and the Aeolian/Older_terrace/Glacial_till/Moraines areas could be separated by using the simple terrain derivative.

Developing the final parent material image

The final, combined image of the parent material is shown on *Figure 8*. It shows tremendous amount of spatial details representing information on consolidated- unconsolidated materials, bare rock areas, colluvial areas, texture classes and genetic classes of the unconsolidated areas. They are very basic properties of the surface material, relatively easy to map using digital mapping tools. Each of the layers represents important properties for the soil formation, but the real added value of the procedure is the combination of them into one combined image. The classes of the parent material image represent uniform areas where the geologic processes resulting in the current parent material for the soil formation were uniform as well. The procedure recognizes the differences and it defines

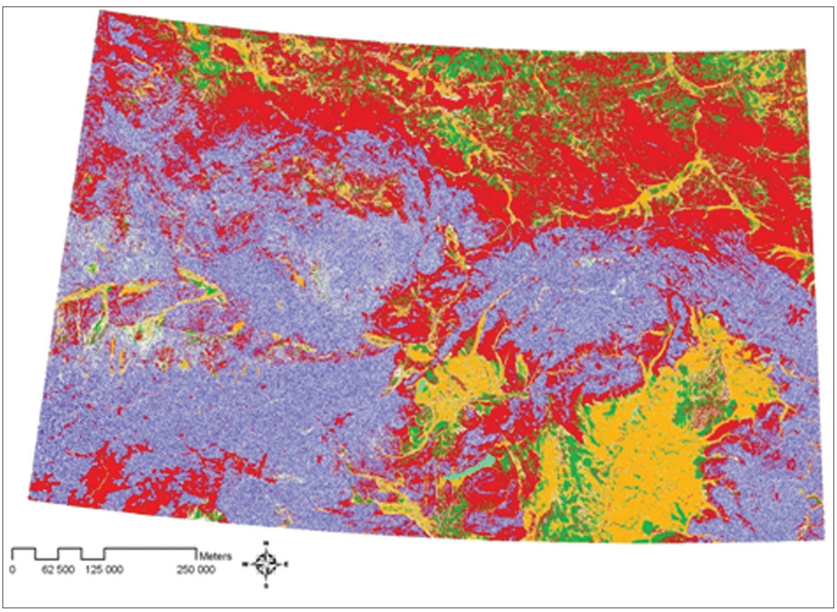

Fig. 8. The combined parent material image. No legend was allocated to the units, only the spatial patterns are used

the spatial units helping the delineation, but it does not explain the geologic origin in detail, only small segments of the whole information are given.

The traditional procedure needed to have PM input layer for the SOTER-unit delineation (ISRIC, 1993). The layers are always country specific and they follow the local traditions and local PM and soil classification systems. Therefore, all of the national layers are different with given, differently defined and not necessarily correlatable spatial units. The procedure follows the original SOTER idea of having broad melting pot classes which are capable of integrating the different national inputs. The procedure takes only the attributes of the input data by allocating, correlating, reclassifying them to the common property classes without importing their spatial definition, delineation. The geometry is defined by the remote sensing and terrain modeling based approach presented here. The resulting image has a consistent spatial delineation throughout the entire area.

The final image used only a simplified texture classification having only coarse and fine texture classes. The lowest class separability happened between the gravelly sand and sand and between the clay-loam and clay classes – as it was described earlier. Merging the overlapping and problematic classes together solved the majority of the misclassification problems while still kept the important information for soil formation. Separating the coarse textured areas – with huge amount of none or slowly weathering minerals and low buffer and CEC capacity – from the fine textured areas – with very different weathering processes and colloid characteristics – is important for any further soil characterization.

However, by merging the related classes, some important features are lost as well, especially in the fine texture class where the clay-loam areas – mainly the lower lying alluvial areas – and the silt-loam ones – representing the loess plateaus – are really different from each other. The first class is mainly covered by Fluvisols, Gleysols and salt effected soils, while the second one is Chernozem and Phaeozem dominated one. The differences are not always evident in *Figure 6*, but they are highlighted in the final image (*Figure 8*). The loess plateaus – like Hajdúság – and the sandy regions have higher elevation over the surface drainage network and therefore their areas are excluded from the alluvial-fluvial genetic class (*Figure 7).* After the combination of all of the input layers, the "missing classes" were reintroduced. The fine textured areas not classified as alluvial felt to the other class of Aeolian and appear in the final image solving the problem of texture image.

Please note that the image has no legend. The primary goal is to assist the SOTER unit delineation by providing the spatial delineations of the PM units. Of course, there are existing attributes allocated to each pixel, but their thematic accuracy – as it was described in the case of the texture class map – may have high uncertainty. The procedure separates the spatial units with relatively high certainty and the attribution of those units with descriptive data should be case dependent. Areas with high number of reliable training data can be described with the same input data layers, while in case of limited number or quality training data, only the spatial units should be used for further processing.

Conclusions

One of the major restricting factors for the SOTER database development is the lack of harmonized parent material information that could be used for the SOTER unit delineation. SOTER units are based on terrain and parent material information. Terrain information can be extracted from DEM using quantitative procedure. The terrain units have to be further subdivided by PM information which is often difficult to obtain. A procedure to produce that layer is crucially needed. The aim of this study was to develop a simplified procedure for PM unit delineation using easy to access RS and SRTM data and a limited set of PM information.

The original SOTER requires quite a complex and detailed information on lithology which is often difficult to acquire. Therefore, a simplified – property based and hierarchy free classification system focusing only on the most basic PM information – was developed to characterize the PM. There is only one hierarchical segment, the consolidated/unconsolidated material layer which is used as a first level stratification layer. After that, the consolidated and unconsolidated areas were treated differently throughout the process and were rejoined only at the end of the process. The segmentation of the consolidated PM area is difficult, only the bare rock areas can be identified with high reliability. No further division could be made without reliable legacy data. The most important factor, the lithology, is hidden under the forest cover and cannot be identified using neither RS nor digital terrain modeling tools. The unconsolidated material was further stratified by texture classes and by alluvial-aeolian processes. The two layers were combined to achieve the final PM units.

The resulting database is just a first approximation of the major PM units. However, the detail produced for the unconsolidated PM part of the pilot area is quite promising, it shows all the major soil and landscape units being important for soil mapping at the targeted – 1 : 1 M – scale of SOTER. The consolidated part has much less details, those areas still need legacy data for more detailed PM unit delineation.

The procedure can be used anywhere in the world where PM data are limited. It creates the first level delineationwhich can be further subdivided into smaller and more specific units as more data become available. Due to its quantitative approach of spatial unit definition, the resulting dataset is consistent and free of any artifacts and edge matching problems.

Acknowledgement: Our work was supported by FP7 project “Regional pilot platform as EU contribution to a Global Soil Observing System” Grant agreement no.: 211578, financed by the European Commission“; by the Hungarian National Scientific Research Foundation (OTKA, Grant No. K105167); by the "Validation of the Central European Soil database" Strategic Grant of the Visegrád Fund, No. 31210072, and by the BONUS-HU Grant No. OMFB-01251/2009.

REFERENCES

Brady, N.C. and Weil, R.R. 1999. *The nature and properties of soils*. 12nd edition. New Jersey, Upper Saddle River, Prentice Hall, 881 p.

Daroussin, J., King, D., Le Bas, C., Vrcaj, B., Dobos, E. and Montanarella, L. 2007. The soil geographic database of Eurasia at a scale 1 : 1.000.000: History and perspective in digital soil mapping. In: *Digital soil mapping: An introductory perspective. Development in Soil Science.* Vol 31. Eds. Lagacherie, P., McBratney, A. and Voltz, M. The Netherlands. Elsevier B.V. 55–67.

Dobos, E. and Montanarella, L. 2004. *The development of a quantitative procedure for soilscape delineation using digital elevation data for Europe*. Digital Soil Mapping workshop. Montpellier, France, Sept. 14–17.

Dobos, E., Daroussin, J. and Montanarella, L. 2010a. A quantitative procedure for building physiographic units for the European SOTER database. In *Digital Terrain Modelling. Development and Applications in a Policy Support Environment*. Eds. Peckham, R. and Jordán Gy. European Commission, Springer, 227–259.

Dobos, E., J. Daroussin, J. and. Montanarella, L. 2005. *An SRTM-based procedure to delineate SOTER Terrain Units on 1:1 and 1:5 million scales.* EUR 21571 EN, Office for Official Publications of the European Communities, Luxembourg, 55 p.

Dobos, E., Seres, A., Penizek, V., Dijkshoorn, K., Schuler, U. and Micheli, E. 2010b. *Soil parent material delineation using MODIS data.* Digital Soil Mapping Workshop. Rome, Italy, 24–26 May, 2010.

FAO, 2006. *Guidelines for soil description*. 4th edition. Rome, FAO, 97. p.

FAO, IUSS Working Group WRB, 2007. *World Reference Base for Soil Resources*. Rome, ISRIC.

Farr, T.G. and Kolbrick, M. 2000. *Shuttle Radar Topography Mission produces a wealth of data.* EOS, Trans. Am. Geophysical Union 81. 583–585.

ISRIC 1993. *Global and National Soils and Terrain Databases (SOTER).* Procedures Manual. Wageningen, UNEP-ISSS-ISRIC-FAO. ISRIC, International Soil Reference and Information Center.

Knepper, Jr., D.H. and Simpson, S.L. 1992. Remote Sensing in U.S. Geological Survey and Servicio Geologico de Bolivia, Geology and mineral resources of the altiplano and cordillera occidental, Bolivia. *U.S. Geological Survey Bulletin* 1975. 47–55.

Kriegler, F.J., Malila, W.A., Nalepka, R.F. and Richardson, W. 1969. *Preprocessing transformations and their effects on multispectral recognition.* Proceedings of the Sixth International Symposium on Remote Sensing of Environment, 97–131.

Lawley, R. 2009. *The Soil parent material database: A User Guide*. British Geological Survey Internal Report, OR/08/034, 47 p.

Soil Survey Staff, Natural Resources Conservation Service. 2007. *National Soil Survey Handbook.* Part 629 – Glossary of landform and geologic terms. Washington, USDA 600–655.

Tucker, C.J. 1979. Red and Photographic Infrared Linear Combinations for Monitoring Vegetation. *Remote Sensing of Environment* 8. (2):127–150.

Vincent, R.K., 1997. *Fundamentals of geological and environmental remote sensing.* Prentice Hall, Upper Saddle River, New Jersey, 366 p.

Worstell, B. 2000. *Development of soil terrain (SOTER) map units using digital elevation models (DEM) and ancillary digital data*. M.Sc. Thesis. Purdue University, Indiana, USA.