Hungarian Geographical Bulletin 59 (2) (2010) pp. 181–205.

A quantitative procedure for building physiographic units supporting a global SOTER database

Endre Dobos[1]–Joël Daroussin[2] and Luca Montanarella[3]

Abstract

Until recently, manual methods were used for delineating SOILSCAPE. The use of digital data sources, such as digital elevation models (DEM) and satellite data can speed up the completion of digital soil databases and improve the overall quality, consistency, and reliability of the database. Our approach uses DEMs for SOILSCAPE delineation based on the terrain classification system of "traditional manual" and Edwin Hammond's landform classification methods, published in 1954.

In this study, the goal was to use quantitative methods to derive terrain classes that match the criteria of the "Georeferenced Soil Database for Europe" Manual of procedure and to create a DEM-derived polygon (soilscape) system for Europe. Four terrain attributes were used to define the SOILSCAPE: hypsometry (~elevation and relief intensity), slope percentage (SP), relief intensity (RI), and dissection (PDD). The SRTM30 (Shuttle Radar Topographic Mission) database was used as a base DEM and for the derivation of the SP, RI, and PDD layers.

We concluded that no major modification is required for the procedures to incorporate information that is derived quantitatively from digital data sources. The resulting database will have all the advantages of quantitatively derived databases, including consistency, homogeneity, and reduced data generalization and edge-matching problems.

Keywords: digital soil mapping, SRTM, digital terrain modelling, small scale soil database, DEM, morphometric terrain analysis

Introduction

Soil database is needed for global scale yield forecasting, modelling and research. However the only available soil map with a global coverage is the 1:5

[1] University of Miskolc, Institute of Geography, Department of Physical Geography and Environmental Sciences, Miskolc-Egyetemváros, 3515, Hungary. E-mail: ecodobos@uni-miskolc.hu

[2] INRA-UR SOLS Centre de Recherche d'Orléans UR0272 Unité de Science du Sol 2163 avenue de la pomme de pin CS 40001 Ardon 45075 Orléans cedex 2, INRA. E-mail: Joel.Daroussin@orleans.inra.fr

[3] European Commission, Joint Research Centre, Institute of Environment and Sustainability, Soil and Waste Unit, TP 280, 21020 Ispra (VA), Italy. E-mail: luca.montanarella@jrc.it

million scale FAO (Food and Agriculture Organization, United Nations) soil map of the world, which has been compiled from data collected up to the late seventies. Since the completion of the FAO soil map much new data have been documented and new approaches of mapping and database development have been developed. The lack of a standardized, compatible, reliable soils database at appropriate scale is a major constraint to global environmental and agricultural modelling. Therefore, the SOTER (World SOil and TERrain Digital Database) project was initiated by the International Society of Soil Science (ISSS) in 1986 (ISSS, 1986). Initially SOTER was intended to have a global coverage at 1:1 million scale (Batjes, N.H. 1990; ISRIC, 1993), which goal was later degraded to 1:5M due to the lack of financial means. Other international organizations, such as the United Nations Environmental Programme (UNEP), FAO of the United Nations and the International Soil Reference and Information Centre (ISRIC) joined this project and supported the idea of having a global scale soil and terrain database useful for a series of applications. An international committee was appointed to develop a "universal map legend system" and to define a minimum necessary set of soil and terrain attributes suitable for compilation of a small-scale soil resources map. The database can provide information for a wide range of applications such as "crop suitability, soil degradation, forest productivity, global soil change, irrigation suitability, agro-ecological zonation, and risk of drought" (ISRIC, 1993). It is not feasible to delineate soil polygons on regional-to-global scale soil databases. Only homogeneous terrain units defined by their physiographic and parent material information can be defined to represent homogeneous units of the soil forming environment. Soil information appears only on the attribute level as assigned soil associations.

The central and eastern part of Europe is completed based on the SOVEUR project (Batjes, N.H. and Bridges, E.M. 1997). The database is currently operational and has been used for assessing different land and soil degradation processes acting in the area. However, it still shows the most typical limitations of data inconsistency listed below. The database was compiled from national databases provided and translated to the "SOTER language" by the national soil survey institutions. Differences in the interpretation of soil parameters are evident from the acidification map *(Figure 1).* Almost all political borders are visible on the thematic map. Besides the variation in resolution and quality of the incorporated data, this artefact is most likely due to the differences in interpretation of numerous soil terms, and to the misplacement of the national soil variability on the global variability range. More appropriate quantification procedure for soil property characterization is needed to solve this problem.

The work has been started to prepare the SOTER database for the western part of Europe as well. The European Commission has agreed to

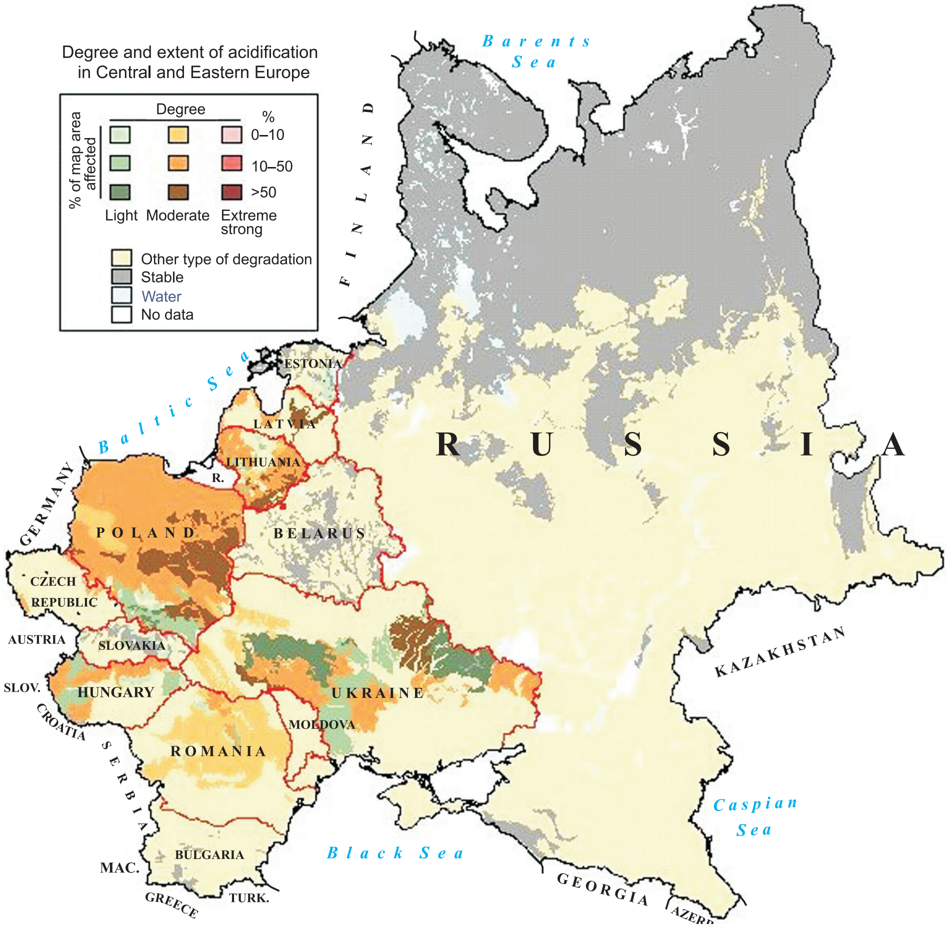


Fig. 1. Degree and extent of acidification risk in Central and Eastern Europe derived from the SOVEUR database

complete the SOTER database for the EU countries at a scale of 1:5M. The soil information for the EU SOTER map is taken from the SGDBE1M with an expert knowledge and decision rule controlled procedure developed by INRA Orleans (King, D. *et al.* 2002). A preliminary version of the database has been completed by aggregating the existing SGDBE1M polygons. Unfortunately, the SGDBE1M polygon delineation does not follow the SOTER methodology. Thus the SOTER Unit delineation based on the aggregation of SGDBE1M polygons does not necessarily match SOTER criteria.

Emerging digital technology and high resolution digital terrain information, such as the digital elevation model (DEM) obtained by the Shuttle

Radar Topography Mission (SRTM) (Farr, T. G. and Kolbrick, M. 2000), represent a great resource and potential for developing a quantitative procedure to replace the existing SOTER procedure. The main advantage of a quantitative procedure lies in the spatial and thematic consistency of the final product.

The existing SOTER procedure needs to be modified slightly, partly due to its natural evolution and partly because of the quantification of the criteria that were originally defined qualitatively. Dobos, E. and Montanarella, L. (2004) conducted preliminary studies on the use of digital terrain and remotely sensed data for the SOTER unit delineation. The results were promising and a program was launched to update and modify the SOTER procedure to incorporate digital elevation data in the delineation of the Terrain Units. This work is done jointly by the European Commission Joint Research Centre (EC JRC) – Soil and Waste Unit, ISRIC and FAO. A SOTER Procedure Modification (SPM) workshop was organized in Ispra, Italy, 21–22 of October 2004, to implement the changes that were agreed upon since the last published version of the Manual and define the needs for future research to update the procedure and incorporate the newly emerging tools and data sources such as DEM data. The decision was taken that Europe would be the first pilot area for the global soil database with EU SOTER as the database to test the new procedure.

This paper focuses on the development of the quantitative DEM-based procedure to delineate SOTER Terrain Units at both 1:1M and 1:5M scales. Up to now, no accepted procedure on the characterization of terrain units has been developed. This work has aimed to set up a DEM based procedure to delineate homogeneous terrain units that match the ones derived by following the traditional SOTER manual. In other words, the original SOTER procedure was translated to the digital manner using DEM as a major input for the terrain unit delineation.

Materials and methods

The study area

The study area covers the countries of the European Union. However, in this paper as well as for the procedure development, a smaller pilot area was selected, namely the eastern half of the Carpathian basin *(Figure 2).*

The data

The Shuttle Radar Topography Mission (SRTM) global elevation data covers almost 80 percent of the globe, almost all terrestrial land surfaces. Its cover-

Fig. 2. The study areas. The big window (1) comprises the full extent of pilot area, while the windows (1) (Bükk Mountains) and No. 3 (Great Hungarian Plain, Alföld) are the smaller areas to show specific characteristics of the procedure.

age extends between 60° north and 56° south latitudes. SRTM is a joint project between the National Aeronautics and Space Administration (NASA) and the Department of Defence's National Geospatial-Intelligence Agency (NGA) to produce a near global digital elevation data coverage at a relatively high spatial resolution. Farr, T. G. and Kolbrick, M. (2000) describe the data capturing and processing procedure. The data is handled and distributed by the United States Geological Survey and can be downloaded from their website [1].

Before the SRTM data could be used, the input DEM had to be hydrologically corrected. Sinks that are due to errors in the data and the micro-scale natural sinks, such as the sinkholes –which appear as noise at that scale – had to be filled, while the meso- and macro-scale natural variability of the area were kept. According to our experience in using SRTM data, a limit of 20 meters has to be set as a maximum sink depth to be filled. This limit is suggested in this study as well.

Methods

The classifications for the four terrain layers are described in the four sections below. *Figure 3* shows the flowchart of the analysis. ArcInfo® geographical information system software and its GRID raster analysis module were used to achieve the work.

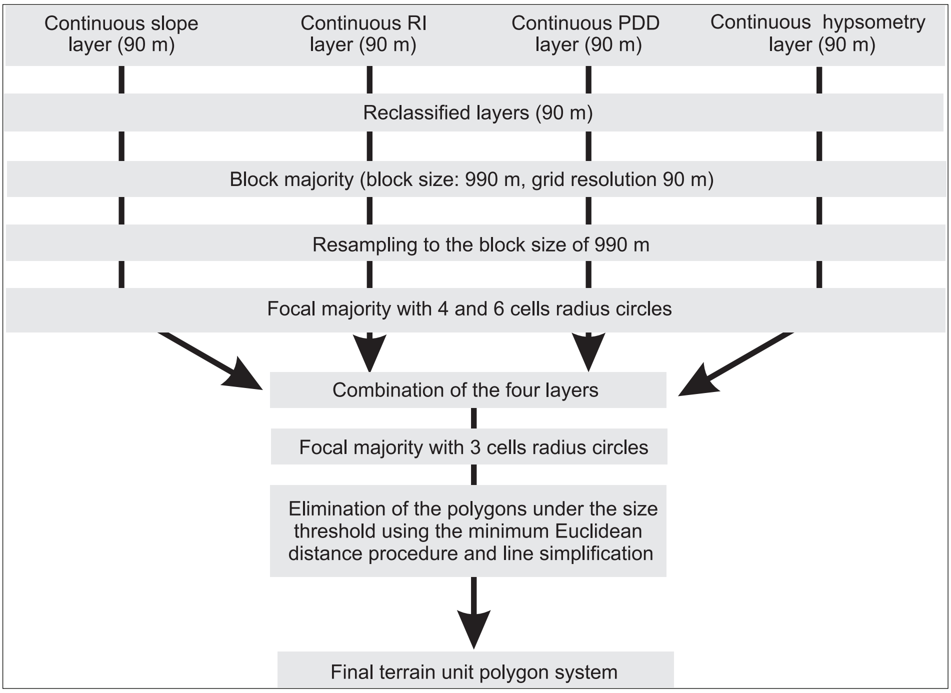


Fig. 3. Flowchart of the Terrain Unit delineation. RI stands for relief intensity while PDD means potential drainage density

SOTER Terrain Unit delineation

SOTER Unit delineation is based on two primary soil formation phenomena: terrain and lithology. Each SOTER Unit represents a unique combination of terrain and soil characteristics. The two major differentiating criteria are applied in a step-by-step manner, leading to a more detailed identification of the land area under consideration. Physiography is the first differentiating criterion to be used to characterize a SOTER Unit. The term physiography is used in this context as the description of the landforms on the Earth's surface. It can be best described as identifying and quantifying the major landforms, on the basis of the dominant gradient of their slopes and their relief intensity. The use of these variables, in combination with a hypsometric (absolute altitude above sea level) classification and a factor characterizing the degree of dissection, can make for a broad subdivision of an area and delineate it on the map. (Dobos, E. and Daroussin, J. (2007) gives more information on the role of dissection on the terrain parameterization.) Further subdivision of the SOTER Unit according to the lithology (parent material) needs to be done to complete the delineation procedure.

Until recently, manual methods were used to delineate SOTER Units, the geometric objects of the SOTER database. The availability of DEM makes it feasible to use a quantitative approach. Worstell, B. (2000) has proposed an approach using DEM for SOTER Terrain Unit delineation based on Edwin Hammond's (1954) landform classification methods. Worstell, B. has adapted and modified his methods to create a quantitative procedure to classify landforms on a regional scale.

A new, quantitative method for creating a SOTER database has been suggested by Dobos, E. and Montanarella, L. (2004) using the 1 km resolution SRTM30 data as the base DEM. Although the procedures were promising, they suggested more research and quality check on the results.

The aim of the present study was to develop a quantitative method to derive terrain classes that match the criteria of the SOTER Manual of Procedure (ISRIC, 1993).

According to the manual, four terrain attributes are used to define the SOTER Terrain Unit: hypsometry (elevation), slope percentage, relief intensity and dissection. The GIS layers of these attributes were derived from the digital elevation model by translating and reformatting the terrain class characteristics given by the SOTER Manual.

These four layers are combined to produce the complex landform classification. This combined layer was then vectorized, and finally generalized to achieve the polygon size limit appropriate for the 1:1M and 1:5M scales of the database to be produced.

The class limits of these attributes are defined more or less quantitatively in the Manual, ("Attribute coding"), except for the dissection for which only qualitative definitions are given. Changes in the class borders were implemented as well: they have been proposed and agreed upon in the SOTER Procedure Modification workshop. The dissection class limits were derived from the Potential Drainage Density (PDD) layers (Dobos, E. *et al*. 2000) via empirical approach.

Coarse resolution DEM tends to generalize the land surface and eliminate the micro- and meso-scale features of the surface, drastically decreasing the slope and relief values of the area. Therefore, a relatively fine resolution DEM (SRTM) was used to maintain the higher scale landscape elements and to derive the terrain descriptor values for the area. The resulting variables show much more detail than what the targeted scales are capable of handling. Thus, a generalization and aggregation procedure was used to obtain the appropriate resolution.

Slope

The slope layer was derived from the SRTM data using the slope function available from the ArcInfo® GRID module. This function uses the average maximum technique (Burrough, P.A. 1986).

The SOTER modified classification scheme for the slopes is shown in *Table 1.* The use of SLOPE function resulted in a continuous slope layer, which was reclassified according to the SOTER classes in (*Table 1*). This kind of classification rarely results in distinct borders between the classes, necessary for defining polygons of practical size. It gives too much spatial detail, which cannot be represented at the target scale ("salt and pepper" effect). Therefore, adequate filters have to be used to derive the slope layer with homogeneous pattern of slope classes and the resolution has to be degraded to reach the spatial detail needed for the targeted 1:1M and 1:5M scales.

Table 1. The SOTER modified slope classification scheme

Slope class	Range of slope percentages
1	0–2
2	2–5
3	5–10
4	10–15
5	15–30
6	30–45
7	above 45

Raster-based generalization procedure

The main steps of the generalization procedure are shown in *figures 3* and *4*. The appropriate spatial resolution of the grid for a 1:1M scale target database is around 1x1 km. Two options were considered for degrading the resolution, *(i)* averaging the cell values within 1 km^2 area or, *(ii)* taking the majority class and assigning it to the spatially degraded cell (blocks). The majority class of the area characterizes the landscape better than the average value. Therefore, the blocking approach was applied to define the majority class within a square shape area with the size of 11 by 11 cells (990x990 m). The resulting grid remains with the original 90 m resolution. Therefore it was then resampled to the target resolution of 990 meters, which was decided by the authors to be appropriate for the target 1:1M scale.

The resulting grid still had some salt and pepper effect, having a mixture of stand-alone cells or small contiguous areas, especially on the transition zones between the classes *(Figure 4C).* This phenomenon represents a significant problem when representative polygons are to be drawn with a minimum polygon-size requirement. To overcome this problem, a filter was applied to that grid layer by using a majority value function with a 4 cells radius circle *(Figure 4D).* The function takes the most frequently occurring class within the specified neighbour-

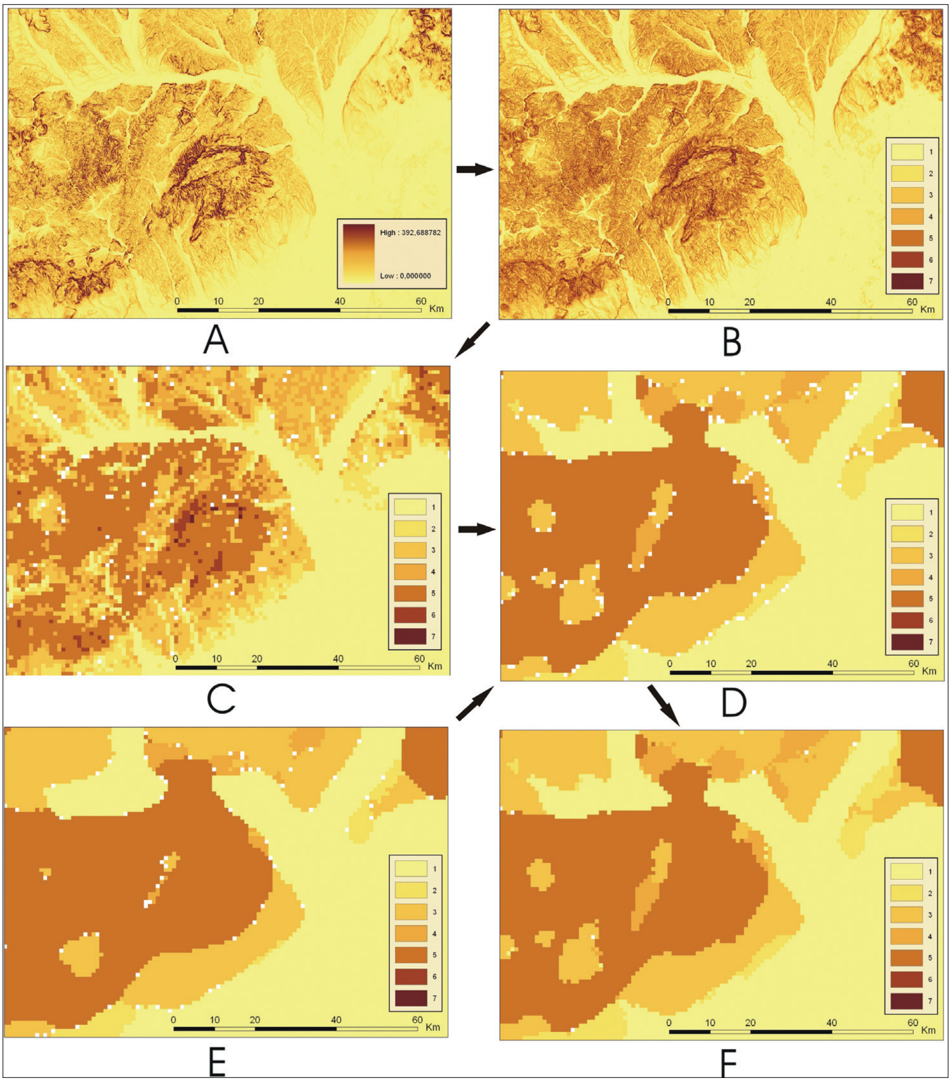


Fig. 4. Flowchart explaining the creation of the slope grid illustrated with an example from the Bükk Mountains, pilot area No. 2. *Table 1.* provides the corresponding slope classes. – A = Continuous slope layer; B = Classified slope layer; C = Resampled; classified slope layer; D = Filtered with majority filter; 4 cells radius; E = Filtered with majority filters; 6 cells radius; F = The final slope layer

hood and assigns it to the centre cell of the moving window. This relatively small neighbourhood of 4 cells was chosen not to over-generalize the landscape, while having enough area to smooth the grid to the required level. In case of equal representation of two or more classes within the specified neighbourhood, a "*no* *data*" value is given to the centre cell, which has to be filled to achieve complete/continuous coverage. To achieve this, the same filter was used again but with a bigger, 6 cells radius *(Figure 4E)*, circle, and the "*no data*" cells in the previous step were replaced with values from this grid layer. Applying these two steps creates patches having sizes appropriate to that target scale *(Figure 4F).*

Relief Intensity (RI)

Relief Intensity (RI) is one of the most significant discriminating terrain factors in the SOTER Procedure. RI is defined as the difference in altitude between the highest and lowest points within a specified distance. It is used in three different places in the procedure: (1) in the major landform description, (2) in the hypsometry characterization and finally, (3) in the dissection characterization. A simplified classification has been suggested by the SPM workshop *(Table 2)* and introduced a new unit for measuring RI, which is now expressed in *m/[area of a 1 km diameter circle]*. For practical reasons in the implementation of the method, the 1 km diameter is approximated with a 990 m diameter because, when using 90 m resolution projected SRTM data, it can thus be expressed simply with a 5 cells radius circle (5 cells radius = 11 cells diameter = 990 metres diameter circle).

The algorithm defines a circle shape neighbourhood with a radius of 5 cells and identifies the highest and the lowest points within that area. The difference between these two points is assigned to the centre cell of the moving processing window as an RI value.

The rest of the procedure was basically the same as in the one used for deriving the slope layer. High resolution, 90 m SRTM data was used to create the original RI layer, which was later classified according to the classes listed in *(Table 2)*. Then this classified image was generalized following the generalization procedure described above.

Table 2. The Relief Intensity classification as proposed by the SOTER Procedure Modification workshop

Relief Intensity class	Elevation range in meters within a 990 m diameter circle
1	0–50
2	50–100
3	100–300
4	above 300

Dissection (PDD)

The degree of dissection is difficult to quantify with traditional methods (ISRIC, 1993). The use of DEM makes it feasible to derive an artificial drainage/valley network, which characterizes the landscape dissection. Dobos, E. *et al*. (2000) developed an index called the Potential Drainage Density (PDD) and a function to compute the PDD. The function derives a drainage network from the DEM and measures the network's density within a predefined sized neighbourhood. The nature of (and the procedure for creating) the PDD layer are described by Dobos, E. and Daroussin, J. (2007).

The data was processed in two steps in this study. In the first step, a DEM-based drainage network was derived by thresholding flow-accumulation values. Cells having a flow-accumulation value higher than this threshold were considered as drainage ways. These drainage cells were assigned a value of 1, while all other cells were set to "*no data*". In the second step, a size for a moving window was selected, and a count of the drainage-way cells within the window was assigned to the centre cell. The result is the PDD value. The higher is the PDD value, the more dissected is the terrain.

The procedure suggested here requires three parameters to be set: (1) the flow accumulation threshold for the drainage network derivation, (2) the radius of the circle for the counting window and (3) the class limits for the reclassification of the continuous PDD image.

Choosing a flow accumulation threshold to build the drainage network

The threshold value was set to 100. In the case of the SRTM data, it translates to approximately a 1 km^2 catchment's area for a drainage line to start. Lower values create a very dense network with too many details, while higher thresholds decrease the pattern density, thus disguising some necessary details.

Choosing a radius for the circle to count drainage cells

The radius was set to 20 pixels/cells. Previously several radii were tested. The rule of thumb is that a too small radius is not able to deliver meaningful information about the general landscape. The resulting images show buffer-like zones along the drainage lines with relatively big portions of the image having "*no data*". Choosing a radius that is too large tends to over-generalize the image, there again masking general landscape characteristics. The optimal choice is the smallest circle, which is still big enough to pick up at least one drainage cell. After several trials and errors, the radius value of 20 was found to be appropriate for the drainage network derived from the SRTM data using the flow-accumulation threshold value of 100.

Choosing class limits for the reclassification of the continuous PDD image

The dissection measurement unit as defined in the "Manual of Procedure" (ISRIC, 1993) is the length of permanent and seasonal streams and rivers within a

1 km² neighbourhood. Three dissection classes are distinguished: 0–10, 10–25, and above 25 km/km². In this study, using these class limits would have been meaningless because the approach is different in a sense that we use "potential" drainage lines *versus* actual ones. In the digital procedure, only two classes were defined with value ranges as given in *Table 3.*

Table 3. The PDD class ranges

Class	PDD value range
1: less dissected areas, convex surfaces	0–90
2: more dissected areas or depressions, concave surfaces	above 90

The procedure then followed the same line as for the other data/model/map layers by applying the generalization and filtering procedure described above for slope in order to derive the final image.

Hypsometry (elevation)

The original SOTER Procedure Manual suggests a two steps procedure for elevation classification. The first step divides the area into three general relief types, namely *(i)* the level lands, *(ii)* the sloping lands and *(iii)* the steep sloping lands. The second step further divides each of these three types into elevation subclasses but using a different classification scheme for each type. This two steps classification system was simplified by the SPM workshop and a new one was introduced which was used here as well.

After the SRTM image was reclassified using the classes in *Table 4*, it followed the same generalization and filtering procedure as the other layers to derive the final hypsometry image *(Figure 6D).*

Table 4. The Hypsometry classes suggested by the SPM workshop.

Table 4. The Hypsometry classes suggested by the SPM workshop

Hypsometry class	Altitude ranges (meters above sea level)
1	Up to 10
2	10–50
3	50–100
4	100–200
5	200–300
6	300–600
7	600–1500
8	1500–3000
9	3000–5000
10	above 5000

Removing PDD as differentiating criterion from the terrain parameters list on the high relief areas

Previous tests of the Dobos, E. *et al.* (2005) procedure have indicated some need for modification. By nature, the quantitative procedure interprets the landscape based on four different stand-alone terrain parameters: relief intensity, slope, elevation and dissection. These four were found to be the most significant factors to identify natural landscape units. However, when the geomorphologic unit delineation is manually done, the interpreter has a complex view on the landscape and units are formed in his mind, not necessarily taking the quantitative thresholds into consideration. The interpreter aims to find the best-corresponding complex units as one, while the quantitative procedure creates four sets of delineations and combines them to form a final polygon system. This latter approach produces several analogue, but not perfectly matching lines, almost similar, but often not the identical delineations of the same units, resulting parallel, redundant approaches in the procedure, and a lot of extra work for aggregating the slave polygons. That was the case for the PDD, where it was used on a high relief area. On the mountainous and hilly regions, slope, relief and elevation do a perfect job for differentiating between the geomorphological units. Involving PDD just means to overcomplicate the procedure, while no additional information is produced. Contrary to the high relief areas, PDD is one of the most significant parameters for the terrain differentiation on a low relief area, where the slope, relief and the elevation have only slight variations. Therefore a decision has been made to pre-stratify the mapped area into high and low relief. Threshold of 100 m/km^2 was chosen to classify the area into the two groups. Elevation, slope and relief were used for the high relief areas, while these three were completed with PDD and all four were used together for the low relief areas. This approach significantly decreased the number of slave polygons top handle.

Database compilation

The next step in the database development was to combine the information from the above four SRTM-derived terrain thematic layers (slope, RI, PDD and hypsometry) into a single grid constituted of Terrain Units. Expectedly this grid had many small and meaningless patterns. It was therefore filtered using a majority function with a 3 cells radius circle shape neighbourhood to create a more homogeneous appearance of the Terrain Units. This grid was then vectorized to create polygon coverage.

Vector-based generalization of the raw SOTER Terrain Units polygon system

Figure 5 shows a close view on part of the raw SOTER Terrain Units polygon system (black thick lines and red thin lines). It is evident that even after the many steps of filtering, there are still a lot of small polygons, which are below the size limit specified in the SOTER Procedure Manual (ISRIC, 1993) (hatched polygons on *Figure 5*). According to the Manual, the minimum area of a polygon

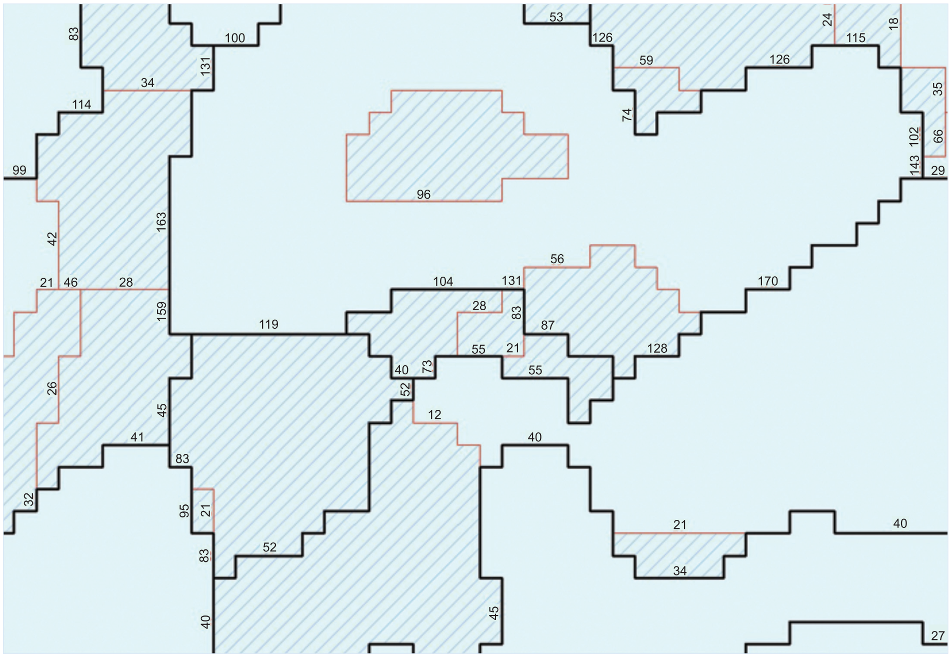


Fig. 5. The minimum Euclidean distance-based aggregation procedure. The red numbers are the Euclidean distances calculated between one polygon and its neighbour. The hatched polygons are those that have a surface area below the threshold (to be aggregated to one of their neighbours). The red thin lines are the polygon borders removed by the algorithm for aggregating

must be 25 km^2 if represented on a 1:1M map and 625 km^2 at a 1:5M scale. Any polygon smaller than these thresholds, depending on the target scale, has to be aggregated with one of its neighbouring polygons, preferably with the one that has the most similar terrain characteristics. To evaluate this similarity in a quantitative manner, the Euclidean distance between each polygon pair is calculated using the same four thematic terrain layers as produced above. These distances (red numbers on *Figure 5*) are then used to select the neighbouring polygon that has the minimum Euclidean distance from the polygon to aggregate.

In the first step, the four continuous landform parameter grids (slope, RI, PDD and hypsometry) are normalized to a range of 0 to 1,000 using equation (1). This step is important to give equal weights to all the four landform parameters.

$$X_{norm} = \frac{X - X_{\min}}{X_{range}} * 1000 \qquad (1)$$

where *Xnorm* is the normalized value for parameter X in the processing cell,*X* is the original value for parameter X in the processing cell, *Xmin* is the minimum value of parameter X calculated from all cells in the grid layer,and *Xrange* is the value range of parameter X (*Xmax – Xmin*) calculated from all cells in the grid layer.

At this stage, each cell is characterized by the four normalized values $SLOPE_{norm}$, RI_{norm}, PDD_{norm} and $HYPSO_{norm}$. In the second step, the mean values of these four normalized landform parameters are computed for and assigned to each polygon. Thus, each polygon is characterized by the four normalized mean values $SLOPE_{Poly}$, RI_{Poly}, PDD_{Poly} and $HYPSO_{Poly}$. In the third step, the Euclidean distance (*Ed*) is calculated as a measure of similarity for each neighbouring polygon pair using equation (2) and assigned to the arc that divides that pair.

$$Ed = \sqrt{D_1^2 + D_2^2 + D_3^2 + D_4^2} \qquad (2)$$

where $D_1 = SLOPE_{PolyLeft} - SLOPE_{PolyRight}$

$D_2 = RI_{PolyLeft} - RI_{PolyRight}$

$D_3 = PDD_{PolyLeft} - PDD_{PolyRight}$

$D_4 = HYPSO_{PolyLeft} - HYPSO_{PolyRight}$

and where $SLOPE_{PolyLeft}$ is the mean normalized slope value of the polygon standing to the left of the arc

$SLOPE_{PolyRight}$ is the mean normalized slope value of the polygon standing to the right of the arc and so forth for each of the four parameters.

At this stage, each arc holds a quantitative estimation of the similarity between the two polygons that it separates in terms of landform characteristics (red numbers on *Figure 5*).

In the fourth and final step, the algorithm can now select – for each of the polygons that are candidate for aggregation – the arc to delete that has the lowest Euclidean distance value of all arcs making up the polygon. Deleting that arc results in aggregating the small polygon to its most similar neighbour.

Aggregating several small polygons together may result in a new polygon that still has an area below the specified threshold. Therefore, the procedure is iterative. It starts again from the second step, i.e. calculating the mean values of the 4 normalized landform parameters for each newly aggregated polygon, calculating the Euclidean distances between the pairs and eliminating the arc with the smallest Euclidean distance value. The procedure is repeated until the resulting polygon system remains stable.

The procedure was implemented as a standalone AML tool, which can either be downloaded from the JRC's Soil and Waste Unit homepage or provided by the authors on request.

Results and Discussion

Figure 6 shows the four terrain variables derived from the SRTM DEM grid for the eastern half of the Carpathian basin. The resulting data sets were found to be realistically characterizing the terrain. Although, the centre and south-west-

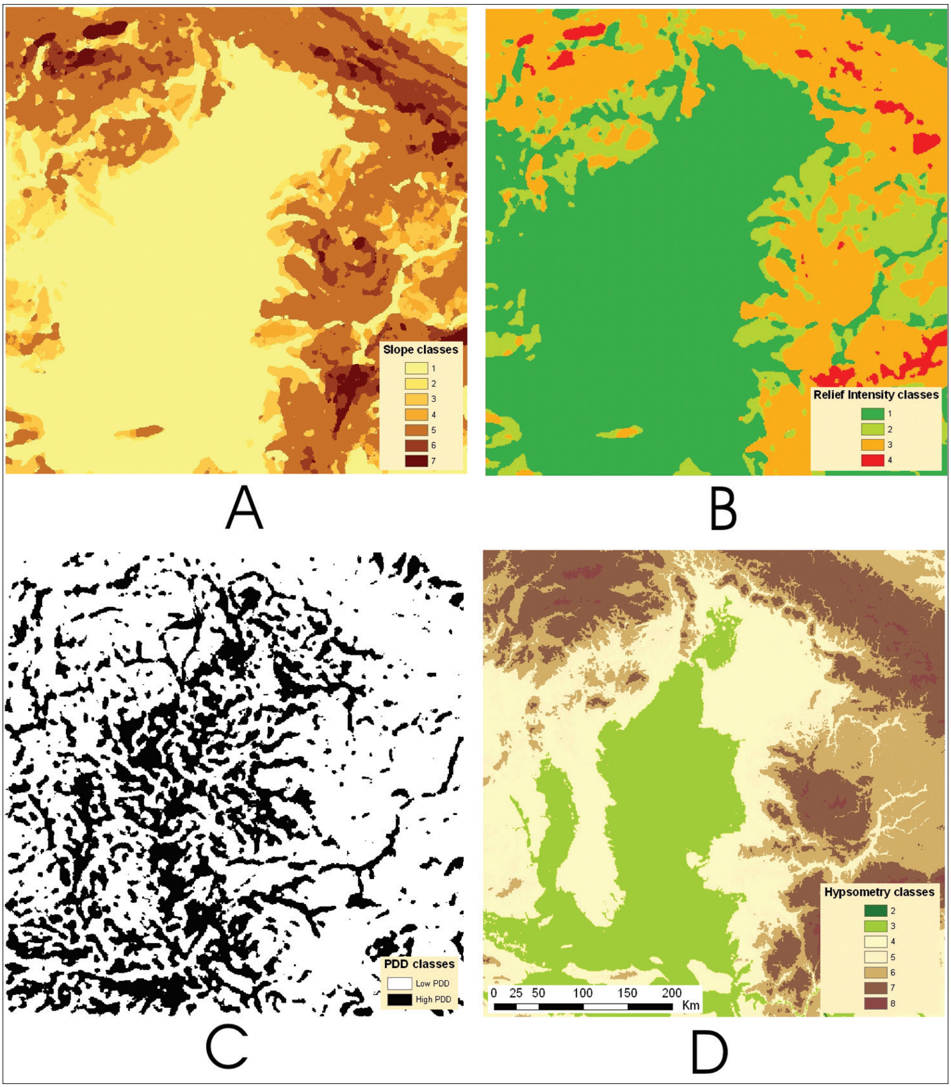


Fig. 6. Maps of the four classified landform parameters in the pilot area. Please refer to the method section for the class meanings. –A = Slope classes; B = Relief Intensity classes; C = PDD classes; D = Hypsometry

ern part is a low relief area (Alföld) with little terrain variability, expressed as only a few meters range in elevation, yet this variability is the most important factor controlling the soil formation processes. In contrast, it is evident from all four derived terrain variable images that the ring of the Carpathians surrounding the plain shows strong variations in the terrain. The challenge of this study was to capture all the significant natural physiographic variation on both plain and hilly areas by applying a quantitative, consistent procedure using the 3 arc second resolution SRTM DEM.

Dobos, E. *et al.* (2004) used a coarser resolution 30 arc seconds SRTM DEM (~1 km) for delineating SOTER Terrain Units. They concluded that the spatial detail needed for deriving a 1:1M scale database is well achieved with such a 1 by 1 km resolution dataset. However, a coarse spatial resolution DEM tends to "overgeneralize" the terrain features by decreasing their value ranges.

The approach developed here is an attempt to take advantage of the availability of a fine spatial resolution DEM. The thematic terrain variable layers were derived from the 90 m resolution DEM to keep as much details of the landscape information as possible. These resulting images were classified based on the SOTER criteria. The classification was followed by the generalization procedure. In the first stage of the generalization procedure, the parameter grids were resampled to coarser resolution. This could have been done by averaging the fine resolution (non-classified) cell values to the coarser resolution one and then classify this spatially degraded grid. But doing so would have taken us back to producing unrealistic classes, which may not even exist in the area. Instead, a "blocking" approach was chosen to achieve this step of the generalization procedure. Blocking consists in assigning the most frequently occurring class value within a block of cells to the entire block. Doing so has several advantages over averaging: 1) it maintains the "most important" information within a block, 2) it maintains the variability inherent to the entire dataset, and 3) it is a valid operation on class values (non-numerical). The first step of the generalization procedure for the 90 by 90 m cell images thus consisted in majority blocking to 11 by 11 cells blocks (990 by 990 m) followed by simply resampling the blocked images to 990 by 990 m cell size images. The method is illustrated on *Figure 4* was applied to the four landform parameter images. As it can be seen from *the Figure 4C*, the method keeps the original classes within the area, showing no trend to shift them downward. Further filtering the image (*Figure 4D*) however shows a tendency to enlarge solid, homogeneous patterns and extend their area over the neighbouring, more heterogeneous areas. This is why filtering must use a relatively small moving window size to minimize this effect.

Both the BLOCKMAJORITY and the FOCALMAJORITY functions have a negative side effect. When more than one majority class occurs within a block or focal neighbourhood (two or more classes are found in the same number of cells within the processing window), the function cannot decide among them and assigns "*no data*" to either all cells within that block (BLOCKMAJORITY case) or to the centre cell of the focal processing window (FOCALMAJORITY case) (see the white spots in *Figure 4D*). Introducing a second filtering step with a larger window is a turnaround "trick" to get rid of almost all these "*no data*" cells by simply picking up for them a value that is found within a larger neighbourhood. The insignificant portion of "*no data*" cells that still remain even after applying this method is treated in the later stages of the generalization procedure, namely with using the polygon aggregation algorithm. It is a more "intelligent" and target oriented procedure and it uses the cell properties directly, instead of estimating the cell class based on its neighbourhood characteristics. However, running the algorithm on a database having a lot of tiny neighbouring polygons with often missing bigger, dominant adjacent polygons to join, the product of the aggregation would be influenced much more by order of polygon processing than by the semantic characteristics of the polygon and would result in a random like aggregation of these polygons. That is why filtering was necessary first to decrease the number of polygons to an acceptable level. The polygon aggregation function was then used to deal with the rest of the "*no data*" polygons. Thus, a raster based filtering and a vector based aggregation was found to be the optimal combination for generalization.

New classification schemes were introduced by the SPM workshop for the four landform parameters derived from DEM data. This paper does not aim to discuss their benefits and impacts. The only conclusion made from the visual interpretation of the polygon system is that the new scheme follows well the geomorphologic units of the landscape and creates meaningful delineation. The changes in the class limits for slope and hypsometry were found to be a great improvement in this study. The new slope classes make a smoother and better discrimination of the land. The classes are grouped around the most frequently occurring values, therefore making a more balanced distribution among them. This phenomenon is even more evident with the hypsometry classes. New classes were introduced to improve differentiation among the landscape units on the low-lying plain areas, where the elevation above the sea level is almost the only significant terrain factor besides the PDD.

Three of the four terrain variables, namely the slope, RI and hypsometry, have not any contribution to the plain area characterization *(Figure 6).* The hypsometry classes contribute a little, but just "accidentally" here because the Great Hungarian Plain lies along the 100 m altitude class limit. The only parameter, which contributes to the landscape unit delineation on a plain area is the PDD. The slope and RI parameters, complemented with the hypsometry, are very efficient in characterizing the landscape of the hilly and mountainous lands. The combination of these three elements creates a very detailed physiographic characterization even without considering the PDD parameter. A large portion of the information provided by the PDD for these higher relief regions is already delivered by the other three parameters. However it is evident that the PDD carries much additional information over the other three parameters, but this appears mainly on a higher scale, not appropriate to the target scales of this project. As opposed to the high relief areas, only the PDD can provide meaningful information for unit delineation within the plain areas by highlighting the depressions and low lying areas where wetness potentially occurs. Using the three dissection classes suggested in the Manual (ISRIC, 1993) would result in a lot of details and small patterns not adequate at the target scales. Instead, only two classes were created and adjusted to delineate both the low lying areas and valley bottoms in hilly and mountainous regions, and the depressions in plain areas. The class boundaries were defined empirically by testing different setups and matching them to real physiographic features.

The pilot area selected for the study was quite challenging due to its very complex natural and anthropogenic geomorphologic patterns. Starting from the second half of the nineteenth century an extensive dike system was built along the major rivers of the Great Hungarian Plain to prevent a huge area from annual floods and to expand agricultural land. At that time, it was often difficult to identify the major watercourse because huge areas were completely covered by inundation or flooding water appearing as temporal lakes. Nowadays the geomorphologic setup of the Plain still resembles to the one from before the dike system. Huge low lying areas, depressions, narrow sand barriers, sand dunes, loess plateaus and old, abandoned river beds create a mosaic of geomorphologic patterns *(Figure 7).* This picture was further diversified by the man made dike and channel structure. Dikes along the Tisza River are captured in the SRTM DEM data and are visible on the DEM image as well (see the brown linear pattern along the Tisza River, pointed with two arrows in *Figure 7).* This of course has a significant impact on the drainage network determination necessary for deriving the PDD. The PDD is generally successful at detecting depressions and local heights. However, the area along the Tisza River is quite problematic. Dikes, especially where the surrounding area is relatively high as well are often taken as heights, preventing the water from flowing through to the river and are thus classified as local elevated areas. This phenomenon extends the area of the natural heights over the low lying ones (see the polygon peninsula extending along the Tisza and pointed by the two arrows). In other cases, where the low lying character of the surrounding area is strongly expressed, thus collecting a lot of drainage lines, these lines are trapped along the dikes on both sides of the river. This creates a high drainage density for those areas, despite the existence of the dikes (see the lower left

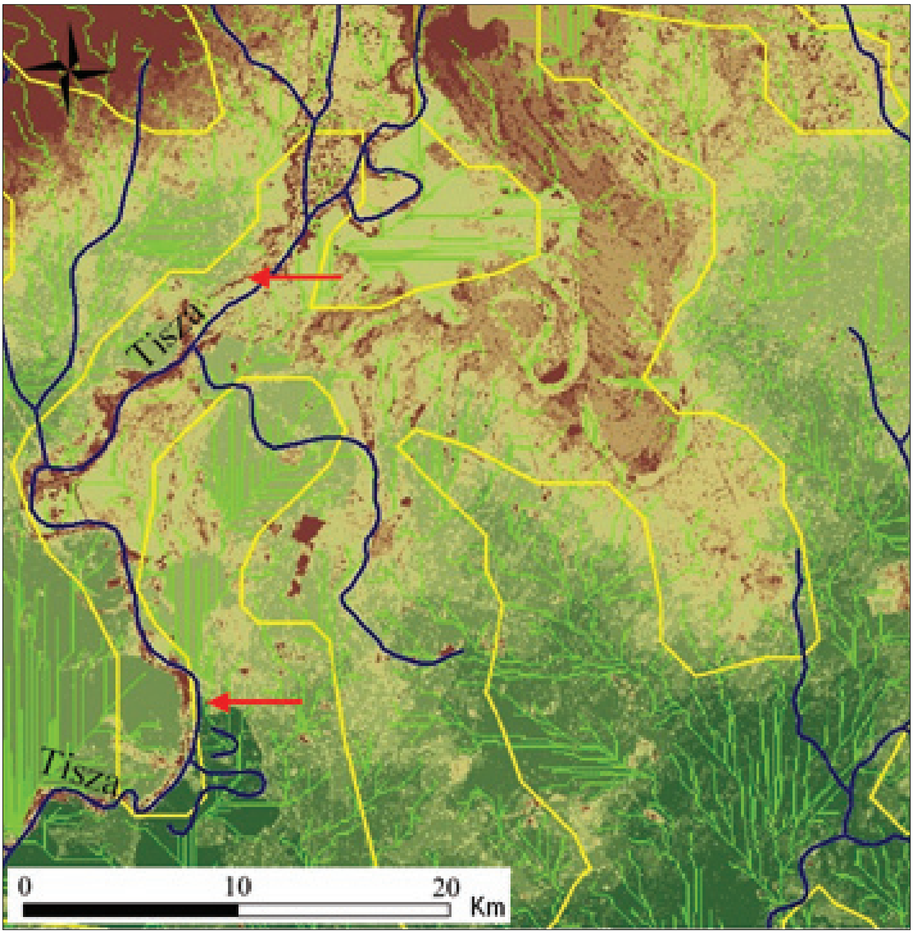


Fig. 7. A 50 by 50 km zoom into the Tisza Valley on the Great Hungarian Plain. The 15 m elevation range (81 to 96 m above sea level) is represented with green to brown colours. Yellow lines show the resulting 1:5 million SOTER polygon system borders. Blue lines show the actual drainage system (rivers, streams and channels) whereas the green ones show the drainage system derived from the SRTM DEM as part of the PDD procedure. The red arrows highlight some of the dikes along the Tisza River

corner of *Figure 7*). In contrast, these are classified as low lying areas. Although this effect hampers the continuous delineation of the Tisza River channel and shifts the balance towards the low PDD area over the high one, it is not as erroneous decision as it may appear. Indeed, dikes have a great impact on the surface water flow, which impact is reflected on the image as well.

The RI classification has changed a lot after the SPM workshop, but mainly in a formal way. The various RI units that were used in different contexts within the previous methodology were replaced with one common unit: *m/[area of a 1 km diameter circle]*. This new unit still maintains a link to the original SOTER RI units while it is easier to handle within a GIS and meets the definition used by geographers for that measure.

The final step in the procedure aims at generalizing the polygon system so that it meets the requirements of the SOTER Manual regarding the target scales: polygons that are below the minimum size limits of 25 and 625 km^2 at the 1:1 and 1:5 million scales, respectively, must be eliminated and

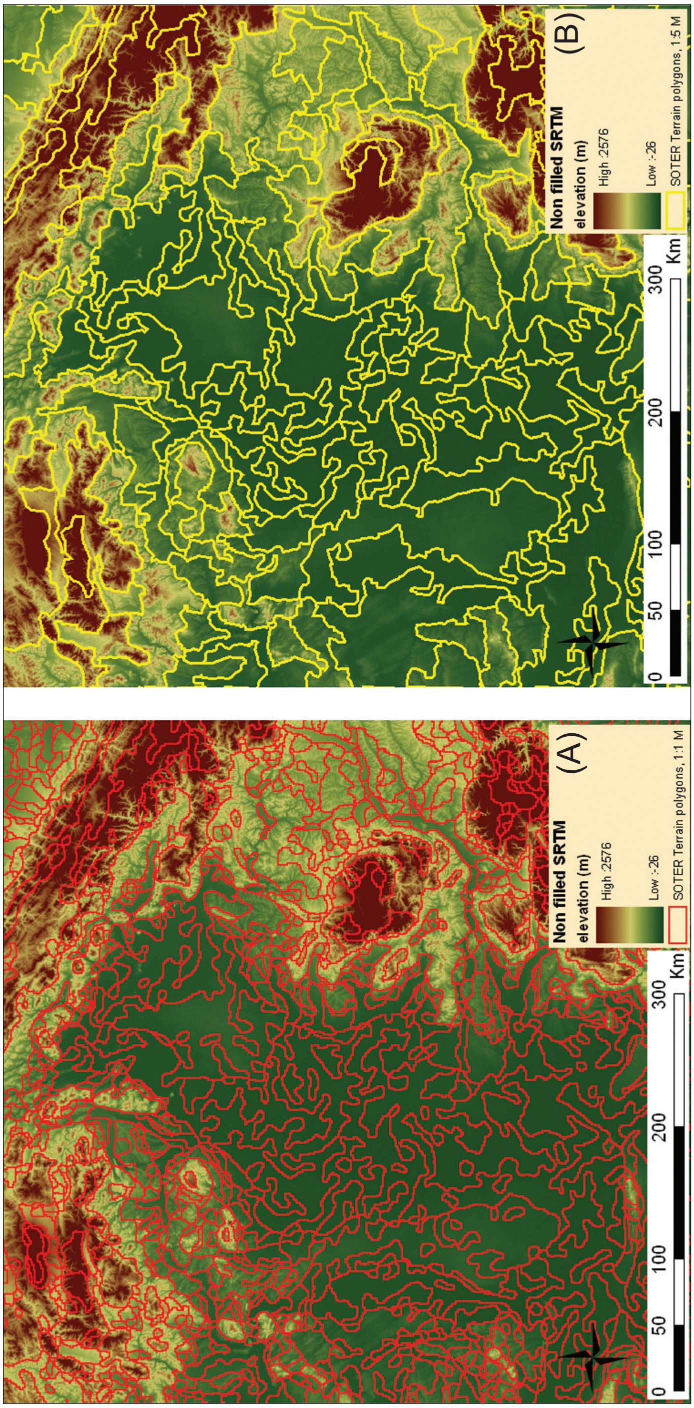


Fig. 8. The generalized polygon system derived from SRTM DEM data to produce the 1:1M (A) and 1:5M (B) scales SOTER Terrain Units for the Alföld and for part of the Carpathians

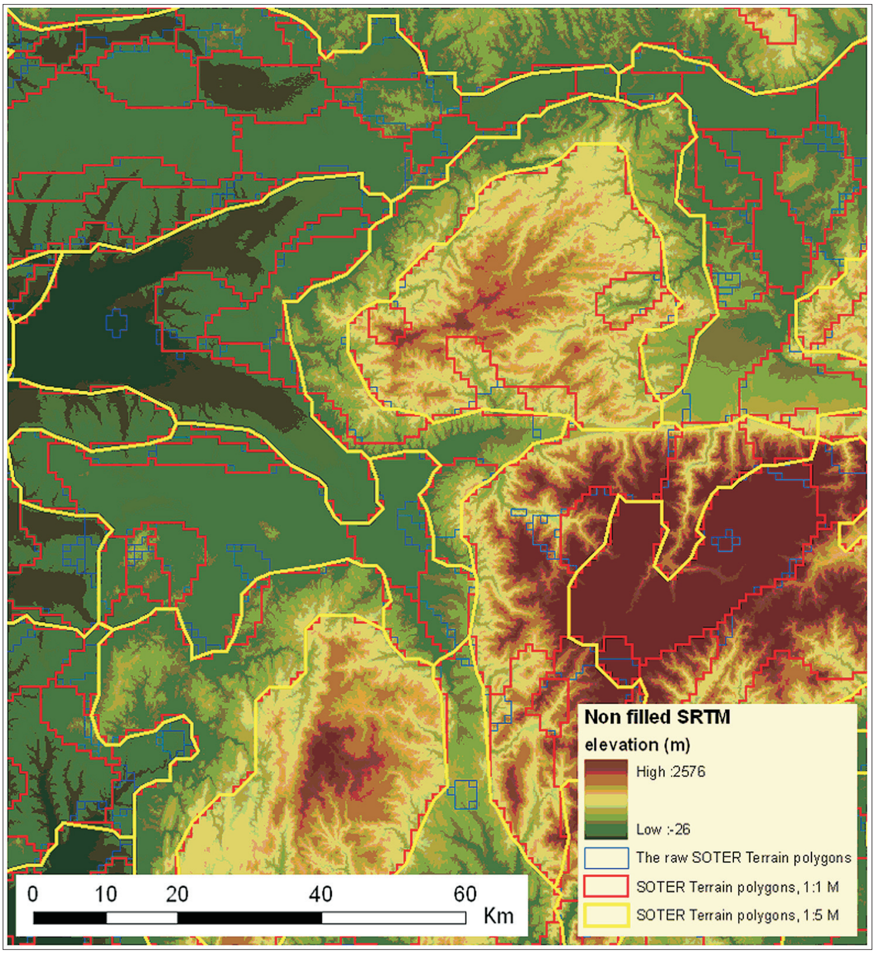


Fig. 9. The SOTER Terrain Unit structure aggregated to 1:1M and 1:5M scales

polygon borders must be simplified. Within the pilot area *(Figure 8)* the lower right corner was zoomed into *(Figure 9)*. It shows all three polygon systems together to give an idea of the effect of the procedure: the raw polygon system as derived directly from the grids (blue lines), the polygon systems produced for the 1:1 million (red lines) and 1:5 million (yellow lines) target scales are displayed.

The procedure eliminates small polygons by aggregating each of them with its most similar neighbour. Similarity between a small polygon and its neighbours is measured by considering the Euclidean distance between their respective landform characteristics. The most similar neighbour, namely the one that is at the smallest Euclidean distance, is selected for aggregation.

On our dataset, despite reducing the number of polygons by 82% at a scale of 1:1M aggregation still moderately altered the original polygon system. Only small, few cell-sized polygons were eliminated, while the overall look of the polygon system was well-kept. At that scale, most of the landscape units that are visible on the SRTM image have been successfully delineated. With hardly any exception, all the blue polygons in *Figure 9* that were aggregated to form the red ones were correctly assigned, followed the landform units of the area.

As expected, aggregation to the 1:5M scale resulted in a more drastic change of the polygon system, reducing the number of polygons by 98%. Characteristic andsometimes very different landscape units had to be combined in a meaningful way. This required a lot of compromise, especially when a small intrusion had to be eliminated but none of its neighbours was really similar. Nevertheless, the algorithm was generally performed successfully and no major problem was identified on the resulting polygon system. Many of the linear features such as bottoms of the valleys were kept in the output and representative (to the geomorphologic units) Terrain Units were formed, appropriate for the target scale.

Generalizing polygons means also generalizing the borders of the polygons that remain after aggregation so that their resolution meets the target scale. The corresponding tolerances (200 and 1000 m, respectively) were chosen according to cartographic common sense (0.2 mm on a map at both scales). Generalizing the arcs is achieved through line simplification. Among the two algorithms provided within the ArcInfo® GENERALIZE command, in our case and only by visual inspection of the results, the Douglas, D.H. and Peucker, T.K. algorithm (1973) appeared to perform better than the pattern recognition one. A "cosmetic" side effect of the line simplification is that the stair-like appearance of the arcs due to the vectorization of the original 990 by 990 m grid cells is smoothed out. Applying both polygon aggregation and line simplification reduced the number of vertices (coordinate pairs) in our dataset by 53 and 89% at the 1:1M and 1:5M scales, respectively, thus dramatically reducing the database volume. Regardless of the scales, the aggregation method produces more or less homogeneous Terrain Units giving satisfying results.

Conclusions

The development of a quantitative procedure for compiling a harmonized, consistent database is seen as a promising way to speed up the completion process of the global SOil and TERrain project. Many segments of the digital soil mapping technology have been made available since the late 1990s. Numerous studies were carried out to test the usefulness of digital elevation data for soil survey and characterization and much knowledge has been accumulated on this topic. The Shuttle Radar Topography Mission (SRTM) project developed global digital elevation model coverage, which is now freely available and easily accessible for use. This emerging set of data and technology can help creating a common platform for the global SOTER database development.

The authors used this opportunity to develop a procedure and test it in the context of the SOTER project for the European Union. This pilot study aimed at making a step forward on this road by creating a methodology for incorporating DEM into the SOTER procedure.

This paper describes this new quantitative method for creating a SOTER Terrain Unit polygon system. The method is designed for mapping large areas of the world quickly and cost effectively. The resulting SOTER database will have the advantages of quantitatively derived databases, namely consistency, homogeneity, limited data generalization problems, and it will avoid edge-matching and harmonization problems. The procedure is based on the SOTER Manual specifications and is meant to be compatible with the datasets formally developed using the traditional way. But it also incorporates the procedural changes, which have occurred since 1995, when the last revision of the Manual was published.

The procedure has been tested on a representative pilot area covering the eastern half of the Carpathian Basin. The delineation of the terrain features is appropriate to the targeted scales. Meaningful and homogeneous geomorphologic units were identified at the 1:1M scale in the test area. More complex but still uniform units were identified at the 1:5M scale as well.

The method is used to develop the SOTER database for the Member States of the European Union. Further refinement and characterization of the Terrain Units will be done using the Soil Geographical Database of Eurasia at 1:1M scale (King, D. *et al*., 2002).

Acknowledgement: This study was supported by the European Commission 7th Framework program, the Hungarian National Science Foundation (OTKA, 34210), the Bolyai Foundation and the French National Agronomic Research Institute (INRA). The authors wish to express their thanks to Vincent van Engelen, Freddy Nachtergaele, Norm Bliss, Koos Dijkshoorn and Jan Nemecek for their scientific and technical support.

REFERENCES

Batjes, N.H. 1990. *Macro-scale land evaluation using the 1:1 M World Soils and Terrain Digital Database.* Working paper and preprint 90/9, Int. Soil Reference and Information Centre, Wageningen, The Netherlands.

Batjes, N.H. and Bridges, E.M. 1997. *Implementation of a soil degradation and vulnerability database for Central and Eastern Europe.* Proceedings of an International Workshop. Wageningen, 1–3 October 1997. ISRIC Wageningen. The Netherlands.

Burrough, P.A. 1986. *Principles of Geographical Information Systems for Land Resources Assessment.* New York, Oxford University Press 50 p.

Dobos, E. and Daroussin, J. 2007. Calculation of potential drainage density index (PDD). In *Digital Terrain Modelling. Development and Applications in a Policy Support Environment. European Commission*. Eds. Peckham, R. and Jordán, Gy. Springer, 283–295.

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. 2005. *An SRTM based procedure to delineate SOTER Terrain units on the 1:1 and 1:5 million scales.* EUR 21571 EN, Office for Official publications of the European Communities, Luxembourg 55 p. [2].

Dobos, E., Micheli, E. Baumgardner, M.F., Biehl, L. and Helt, T. 2000. Use of combined digital elevation model and satellite radiometric data for regional soil mapping. *Geoderma* 97. 367–391.

Douglas, D.H. and Peucker, T.K. 1973. Algorithms for the Reduction of the Number of Points Required to Represent a Digitized Line or Its Caricature. *Canadian Cartographer* 10. (2).

ESRI Inc. 1996. *Automation of Map Generalization: The Cutting-Edge Technology*. Technical paper. White Papers section of ArcOnline. [3]

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

Hammond, E.H. 1954. Small scale continental landform maps. *Annals of Association of American Geographers* 44. 33–42.

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

King, D., Saby, N., Le Bas, C., Nachtergaele, F., Van Engelen, V., Eimberck, M., Jamagne, M., Lambert, J.J., Bridges, M., Reinhard, R. and Montanarella, L. 2002. *A method for generalization of a soil geographical database: the example of a transfer of the European database EUSIS at 1:1M to the world SOTER program at 1:5M.* 17th World Congress of Soil Science, Bangkok, Thailand.

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.

REFERRED WEBSITES

[1] ftp://edcsgs9.cr.usgs.gov/pub/data/srtm/).
[2] http://srtm.csi.cgiar.org/SELECTION/inputCoord.asp
[3] http://arconline.esri.com.

NOW AVAILABLE!

Ukraine in Maps

Edited by

Kocsis, K., Rudenko, L. and Schweitzer, F.

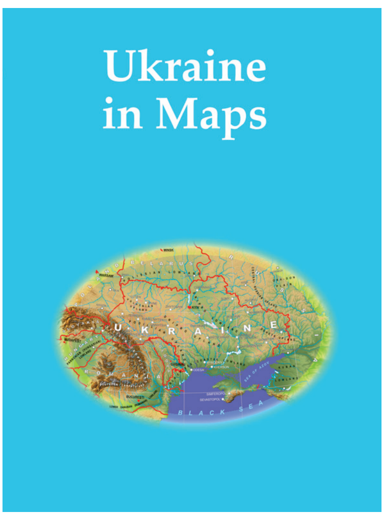
Institute of Geography National Academy of Sciences of Ukraine

Geographical Research Institute Hungarian Academy of Sciences. Budapest, 148 p.

Kyiv–Budapest, 2008

Since the disintegration of the USSR, the Western world has shown an ever-growing interest in Ukraine, its people and its economy. As the second-largest country in Europe, Ukraine has a strategic geographical position at the crossroads between Europe and Asia. It is a key country for the transit of energy resources from Russia and Central Asia to the European Union, which is one reason why Ukraine has become a priority partner in the neighbourhood policy of the EU. Ukraine has pursued a path towards the democratic consolidation of statehood, which encompasses vigorous economic changes, the development of institutions and integration into European and global political and economic structures. In a complex and controversial world, Ukraine is building collaboration with other countries upon the principles of mutual understanding and trust, and is establishing initiatives aimed at the creation of a system that bestows international security.

This recognition has prompted the Institute of Geography of the National Academy of Sciences of Ukraine (Kyiv) and the Geographical Research Institute of the Hungarian Academy of Sciences (Budapest) to initiate cooperation, and the volume entitled "Ukraine in Maps" is the outcome of their joint effort. The intention of this publication is to make available the results of research conducted by Ukrainian and Hungarian geographers, to the English-speaking public. This atlas follows in the footsteps of previous publications from the Geographical Research Institute of the Hungarian Academy of Sciences. Similar to the work entitled South Eastern Europe in Maps (2005, 2007), it includes 64 maps, dozens of figures and tables accompanied by an explanatory text, written in a popular, scientific manner. The book is an attempt to outline the geographical setting and geopolitical context of Ukraine, as well as its history, natural environment, population, settlements and economy. The authors greatly hope that this joint venture will bring Ukraine closer to the reader and make this neighbouring country to the European Union more familiar, and consequently, more appealing.



Price: EUR 15.00

Order: Geographical Research Institute HAS Library H-1388 Budapest, POB. 64.

E-mail: magyar@sparc.core.hu