Project acronym	e-SOTER
Project full title	Regional pilot platform as EU contribution to a Global Soil Observing System
Project No	211578

Deliverable D3

1:1 million scale SOTER geometric databases of the terrain units for the windows in Europe, Morocco and S-China

April 2010

SEVENTH FRAMEWORK PROGRAMME
Environment

ENV.2007.4.1.3.3 Development of a Global Soil Observing System

COOPERATION

Document Information

Deliverable number	D3
Deliverable title	1:1 M scale SOTER geometric databases of the terrain units for the windows of Europe, Morocco and S-China
Period covered	n.a.
Due date of deliverable	28/02/2010
Actual date of deliverable	27/04/2010
Author(s)	Endre Dobos (Unimis), Anna Seres (Unimis), Péter Vadnai(Unimis), Joel Daroussin (INRA), Michael Bock (SciLand), Vincent van Engelen (ISRIC), Koos Dijkshoorn (ISRIC), Rainer Baritz (BGR), Jan Willer (BGR)
Participants	All WP1 partners
Work package	1
Work package title	Development of a landform and parent material platform methodology
Dissemination level	PU
Version	1.0

History of Versions

Version	Date	Status	Author	Approval level
1.0	26/04/2010	Version 1	Dobos	

Table of Contents

Table of Contents 3

1. Summary of Deliverable D3 5

1.1. General statements 5

2. Task 1.1. "The development of an artifact-free DEM" 7

2.1. SRTM DEM Processing/Elimination of Artefacts 7

2.1.1. What is SRTM DEM? 7

2.1.2. Why to process SRTM DEM before using in earth sciences? 7

2.1.3. How SRTM DEM was processed? 8

3. Task 1.2: Morphometric characterization of landforms, based on the existing SOTER landform criteria using elevation, slope, relief intensity and dissection parameters, and further developing the digital, SRTM-based procedure of Dobos et al (2005); and the creation of landform units. 13

3.1. The change in the PDD procedure 13

3.1.1 Automatically derived channel lines (water flow lines, stream lines, thalweg) 14

3.2 Removing the PDD as differentiating criteria from the terrain parameters list on the high relief areas 16

4. Task 1.3: Parent material classification 18

4.1 Task 1.3.2 Revised classification of parent materials relevant for soil development 18

4.1.1 General Scheme of the revised lithology structure 18

4.1.2 Rationale and framework conditions for the lithology classification 19

4.2. Task 1.3.1 Determination of soil parent material within the landform units, based on low-resolution satellite imagery (AVHRR, MODIS, SPOT Vegetation) and DEM data in combination with legacy soil parent material data 25

4.2.1. Covariates used to derive the thematic PM layers 26

4.2.2 Classification of the consolidated/unconsolidated areas 28

4.2.3. Classification of the consolidated material areas 30

4.2.4 Classification of the unconsolidated materials. 32

4.2.5 Development of the genetic layers 36

4.2.6 Finalizing the PM coverage 38
4.3. Comparison between modelled PM classes and legacy geological map data 40
4.3.1. Summary of the results from the PM modelling 40
4.3.2. Introduction 41
4.3.3. Database 43
4.3.4. Results 44
4.3.5. Discussion and conclusions 47
5. Task 1.4: Creation of terrain units by combining landform and soil parent material units 49
5.1 The hybrid method 49
5.1.1 The aggregation rule set 49
5.2 The e-SOTER approach 53
5.2.1 The aggregation rules 53
6. Dissemination and Exploitation 61
7. References 63
MODIS satellite images data sheet 64
Terrain parameters derived from SRTM DEM 68
Final layer stack, used for the classification 74
Landsat Data Processing 76
The finalization procedure for the PM polygon coverage 83

1. Summary of Deliverable D3

This deliverable of the e-SOTER project aims to describe a methodology to develop a SOTER-conform landform and parent material defined polygons at 1:1M scale. The procedure uses the combination of legacy data and other environmental covariates like SRTM and remotely sensed multitemporal images and Digital soil mapping tools. The developed procedure was tested on four windows, two ones in Europe, one in Morocco and one in Southern-China.

The deliverable has the following parts:

1. Four windows with the produced SOTER geometry based on the terrain and parent material data
2. One report describing the procedure
3. One ArcInfo “aml” package to run the procedure using SRTM and Legacy soil and parent material data and their description
4. One ArcGIS tool for developing the necessary parent material layers

1.1. General statements

The main aim of the e-SOTER project is to develop a methodology to overcome the data limitations and create a better harmonized, more comprehensive and consistent product using state of the art data processing, remote sensing and terrain modeling tools. The work has several antecedent steps, which the work proposed by the project description is based on. One of the major reference material was written by the Dobos et al (2005) and aimed to delineate physiographic units following the SOTER mapping scheme. This material has been tested previously and some slight modification has been applied on the methodology within the e-SOTER framework. However, this physiographic unit delineation approach is not complete without having parent material information to disaggregate the physiographic units into the so called SOTER-units, which are defined by both terrain and parent material information.

Work Package 1. of e-SOTER has aimed to develop a methodology (1) to import and spatially integrate existing soil parent material information with the SRTM-based physiographic units and (2) to develop a methodology to derive a minimum set of parent material information assisting the SOTER unit delineation. The work has been divided into sub tasks.

The objective of the first task **(T1.1)** is the development of an artifact-free DEM. DEMs derived from SRTM synthetic aperture radar measurements processed by interferometry suffer from *thermal noise.* This will be reduced using enhanced filter techniques developed by Selige *et al* (2006). Reflection of the radar signal by the vegetation canopy also produces a false surface; corrections will be applied using Europe-wide land cover data. The objectives of the second task **(T1.2)** is the morphometric characterization of landforms, based on the existing SOTER landform criteria using elevation, slope, relief intensity and dissection parameters, and further developing the digital, SRTM-based procedure of Dobos *et al* (Dobos *et al.* 2005); and the creation of landform units. The objectives of the third task are **(T1.3.1)** the determination of soil parent material within these landform units, based on low-resolution satellite imagery (AVHRR, MODIS, SPOT Vegetation) and DEM data in combination with legacy soil parent material data using a **(T1.3.2)** classification of parent materials relevant for soil development and based on the system developed by BGR (2005). The objective of the fourth task **(T1.4)** is the creation of terrain units by combining landform and soil parent material units. Research will focus on the generalization and aggregation processes and on methods of cleaning and structuring the geometric data sets for the four 1:1 million scale windows.

As a results of the subtasks Deliverable 1.1. had to be completed: a 1:1 million-scale SOTER geometric databases of the terrain units for the four windows: combining landform and parent material derived from 1) a quantitative, DEM-based algorithm for landform classification applied on a complete, artifact-free digital elevation model for Europe South of 60 degrees, for Morocco and for South China, and 2) methods to derive parent material classes and delineations using RS and legacy data and a simplified soil parent material classification relevant to soil development and performance.

2. Task 1.1. "The development of an artifact-free DEM"

2.1. SRTM DEM Processing/Elimination of Artefacts

2.1.1. What is SRTM DEM?

A Digital Elevation Model (DEM) is a generic term for digital topographic and/or bathymetric data, in all its various forms. It contains elevation data of a certain surface which can be above or below sea level and maybe include forest canopies and/or man-made terrain features. A special kind of a DEM is one which only provide elevation data of the solid surface of the earth. This bare-earth DEM is called a Digital Terrain Model (DTM).

In the year 2000 the Shuttle Radar Topography Mission (SRTM) took place. An international research consortium lead by NASA created the first homogeneous DEM for the planet earth. From space shuttle Endeavour a radar instrument sent signals to the ground. It received the reflection twice: on the space ship itself and at the end of a 60m long mast. From these two measurements the elevation of the surface could be derived with an horizontal resolution of approximately 30m. In this manner 80% of terrestrial surface of our planet could be covered. The product that mission achieved is of very high value for all earth sciences in general. With an reduced resolution of 90m all the data is provided for free in the WWW. See more details at the SRTM Homepage of NASA

2.1.2. Why to process SRTM DEM before using in earth sciences?

SRTM DEM derived from radar interferometry represent the earth's surface only in areas with low or sparse vegetation and low building densities. Particularly forest canopies cause problems for terrain analysis. If you are working in a continental scale this is not a problem. For e-SOTER however the plan is to work on much larger scales so the vegetation features represented in SRTM DEM come up. Especially in lowlands the differences in elevation are mainly caused by the change of forest - non forest areas. This can produce misinterpretations. To model the distribution of soil conditions e.g. for farming, hydrological planning or environmental protection purposes a DTM (not a DEM) is needed. For example you can simulate surface runoff to get an idea about the size and shape of watersheds which affects the soil properties inside. If forest canopies remain in the elevation model used for modelling of surface runoff a misled stream network will be derived and forests in lowlands will act as watersheds.

2.1.3. How SRTM DEM was processed?

The objective of SRTM DEN processing is the conversion of a DEM of the vegetation surface into a DEM of the earth's surface.

The presented figures demonstrate the effects of both using unprocessed SRTM DEM and the enhanced SRTM DTM for the modelling in context of e-SOTER.

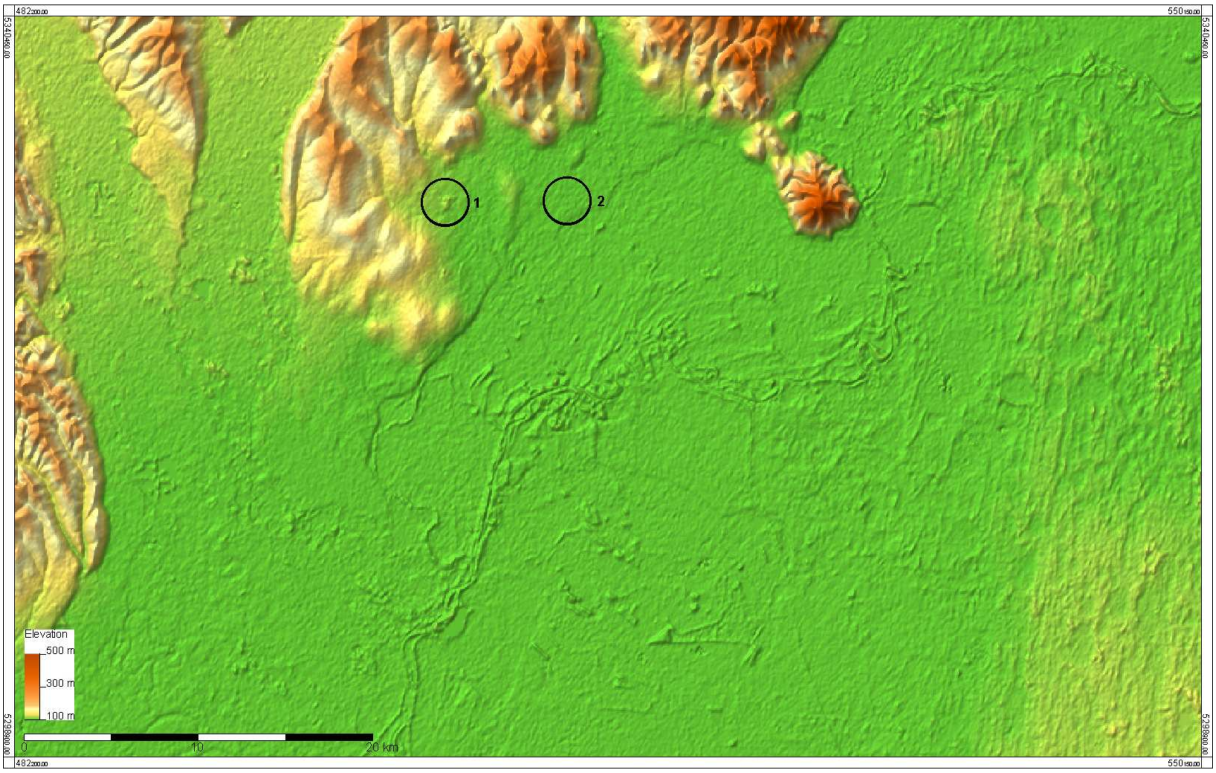


Figure 2.1. SRTM DEM Original

Figure 2.1. represents a shaded relief map derived from original SRTM elevation data (DEM). Besides forest canopies DEM produced with radar technology normally contain a typical “noise”. This can be recognized by an edgy appearance of the surface what does not exist in reality.

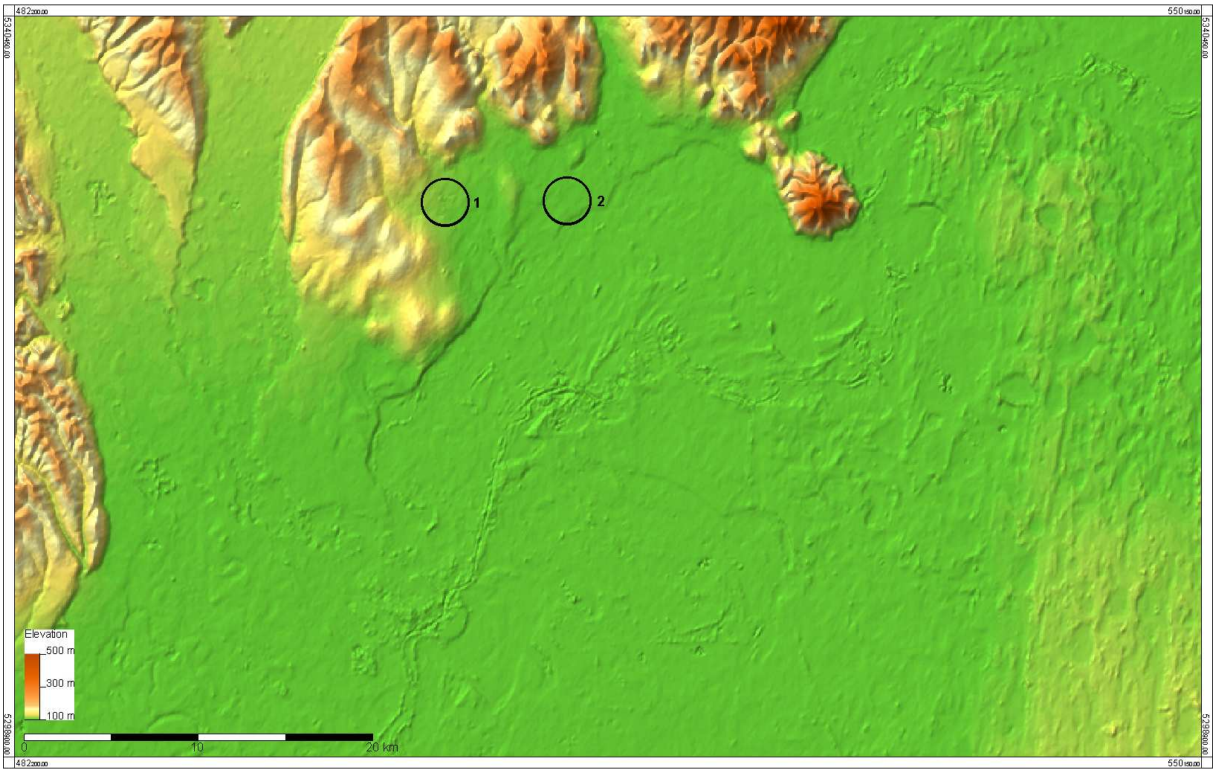


Figure 2.2 SRTM DEM processed

Figure 2.2. is based on the enhanced SRTM DTM data after the processing procedure. One difference of the result is the "noise" elimination using advanced filter technique. By using a special algorithm the surface is smoothed but the morphological features of the surface are preserved.

However the main advantage of the processed SRTM DTM is the removal of the features caused by forest canopy. For this objective the forest elevation has to be estimated. Calculating the elevation of forest canopies from SRTM DEM without having any information about tree/forest heights certainly is a "mission impossible". But the estimation of forest elevation - to a certain degree - is possible. For the estimation of forest elevation information about the location (not hight) of forest bodies is needed (fig. 2.3). The estimation of forest elevation is only possible at the forest borders. The flatter the terrain the better is the result of the estimation. The estimated elevation at the forest borders is then interpolated for the complete forest body – sure with inaccuracy within the centre of large forest bodies. At last the estimated values of the forest elevation are subtracted from original SRTM elevation and so the misguiding shapes caused by vegetation are removed from the data.

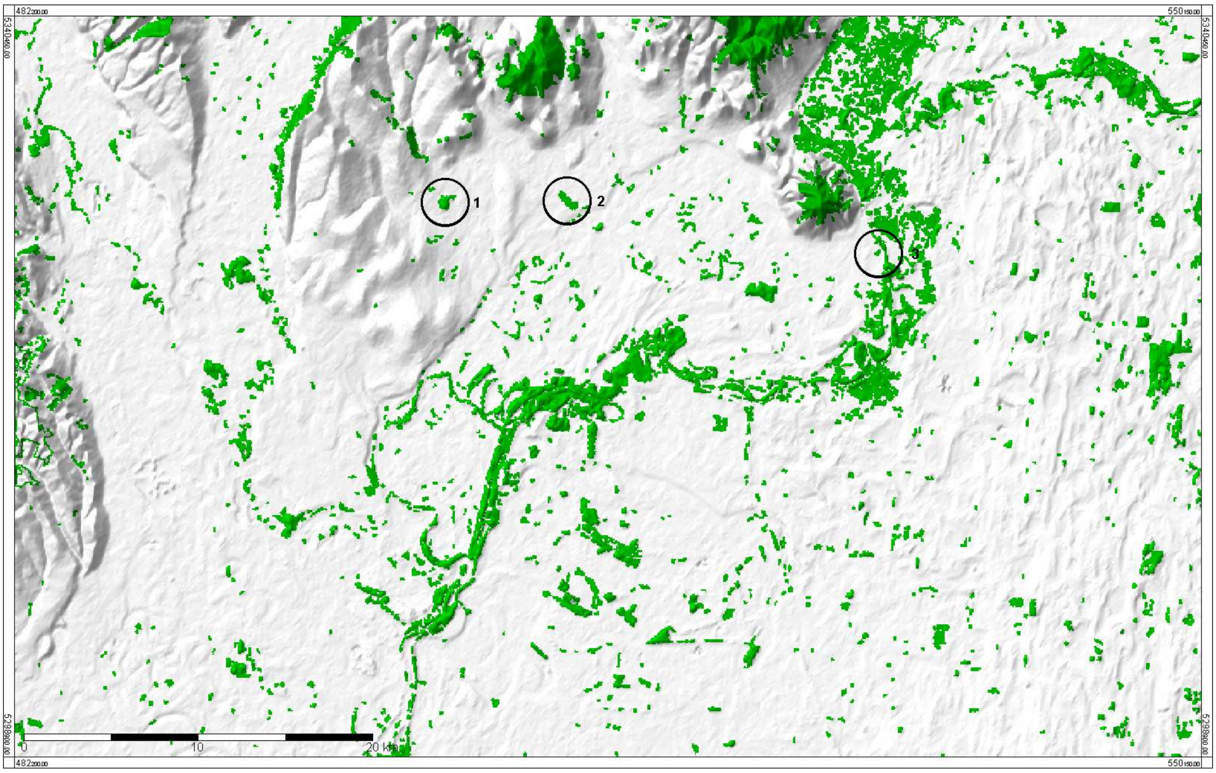


Figure 2.3. Forest Map

The information of the distribution of forest bodies is provided by the FMAP2000, a product of the Joint Research Centre (JRC) of the European Commission. The properties of that digital data make it very applicable for the usage in e-SOTER SRTM processing.

The developed algorithms now make use of the information in the following way: only for forest bodies provided by FMAP2000 the forest elevation is estimated. All areas outside stay untouched. However, one difficulty remains: not all forest bodies result in heightening the elevation of the DEM. Circle 1 shows a forest body which results in heightening, circle 2 shows a forest body without heightening the SRTM surface. This can be observed by comparing both circles (1 and 2) in figures 2.1. and 2.2. This meant that the information of FMAP2000 about the location of forest bodies could not be equated as the location of vegetation features in SRTM DEM. The clue is the combination: to search and find the "jump" in the elevation values only where forest bodies are located.

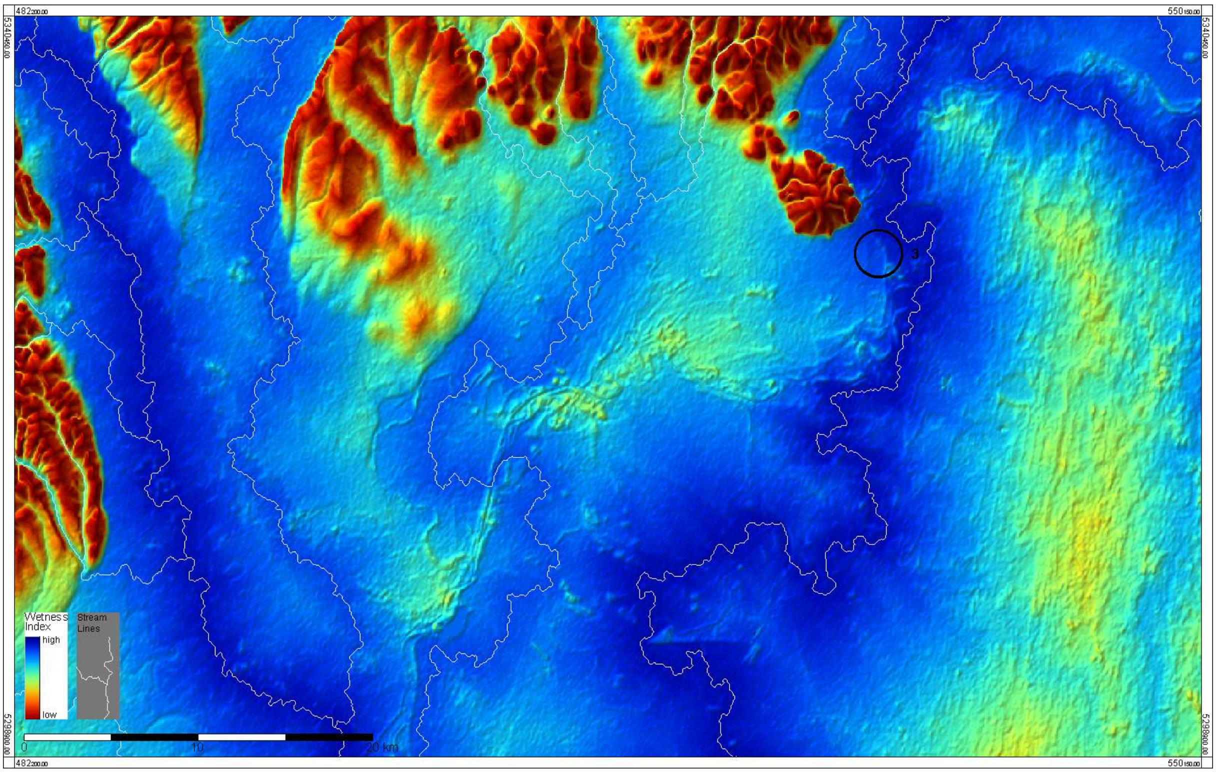


Figure 2.4. Potential Soil Wetness and Stream Lines – derived form SRTM DEM Original (fig. 2.1.)

Figure 2.4. and 2.5. show the terrain parameter 'potential soil wetness index' derived from the DEM/DTM. This is a useful parameter for questions that deal with distribution of water on and close below the surface. Although this parameter is theoretical and provides important information (it does not claim that certain areas are in reality 'wet' or 'dry', but it gives hints about the relative distribution). Sites with a potential better water supply can be delineated from others that have potential dryer conditions.

Also stream lines derived from DEM/DTM are presented in that figures. They show the course of theoretical linear water runoff on the DEM/DTM surface. These lines should meet the natural water streams like rivers but in fact they do not. This is because the modelled surface isn't the earth's surface. Minor deviations are unavoidable but should not go beyond the scope.

The two described data bases, the DEM and the DTM, result in crass different derivations in case of stream lines and wetness index. In the case of stream lines “dams” (in fact forest canopies in the DEM) block the runoff of the lines (see e.g. circle 3 in comparison with fig. 2.5), resulting in redirection of the complete stream network in the Great Hungarian Plane.

The forest canopies act as drainage divides. The Forest bodies also cause low values of wetness index in Tisza valley where high values are to be expected (see centre of fig. 2.4). Both derivations in fig. 2.4. show unacceptable results.

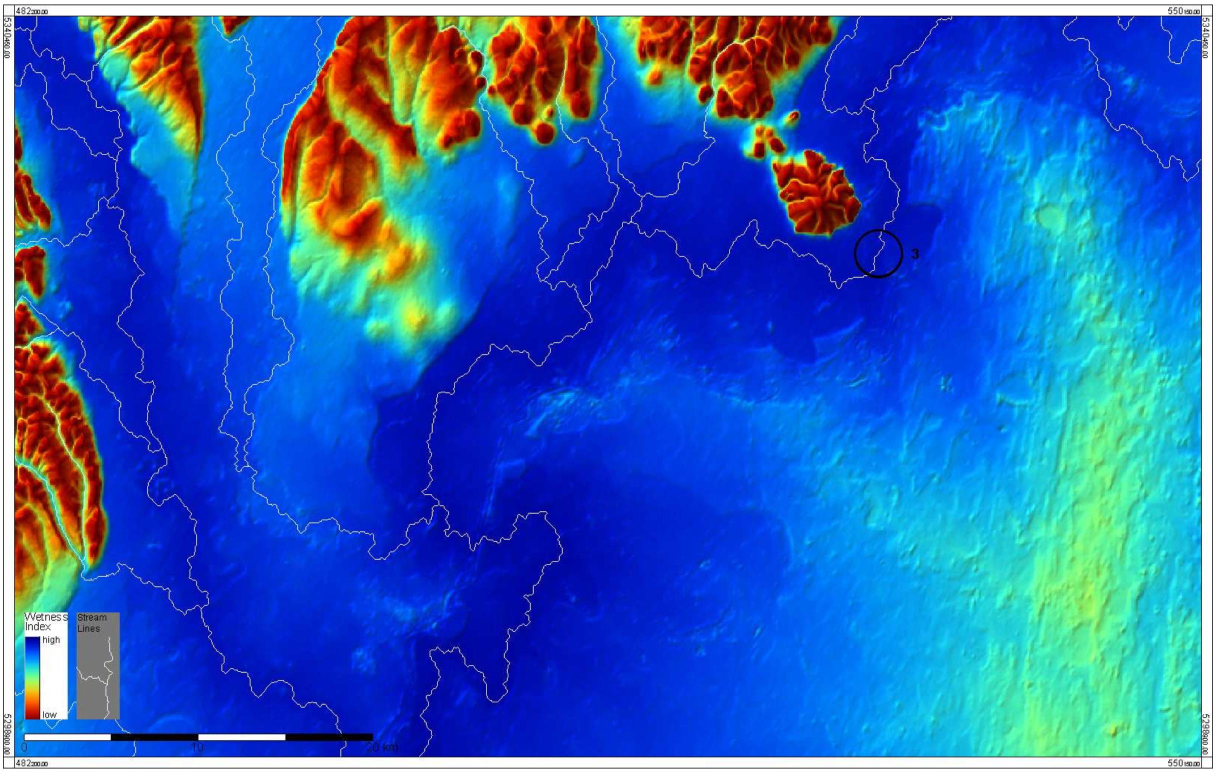


Figure 2.5. Potential Soil Wetness and Stream Lines – derived form SRTM DEM processed (fig. 2.2.)

Much more plausible are the results when the processed SRTM DTM is used. The low values of wetness index caused by forest bodies (centre of fig. 2.4.) mostly turned into high values in the Tisza valley. Only a few redirections occur for some stream lines. Plausible results of terrain analysis in earth sciences are now available and the designation of homogeneous e-SOTER landform units is improved.

3. Task 1.2: Morphometric characterization of landforms, based on the existing SOTER landform criteria using elevation, slope, relief intensity and dissection parameters, and further developing the digital, SRTM-based procedure of Dobos et al (2005); and the creation of landform units.

The developed procedure was following the reference material of Dobos et al. (2005). The only slight modifications that had to apply on the procedure was the (1) change in the PDD procedure and the (2) removal of PDD as a differentiating criteria for areas having higher than a 100 meter relief intensity.

3.1. The change in the PDD procedure

The Potential Drainage Density (PDD) is the terrain modeling based analogy of the dissection function used in the SOTER methodology. The general idea is based on the drainage line derivation methodology commonly used in the terrain modeling society. It has for major steps, the flow direction, flow accumulation, the thresholding of the flow accumulation to create the drainage line network, and the measurement of the drainage line density within defined sized area. The PDD, as it was originally developed, used the ArcInfo approach for the flow direction and flow accumulation definition on a filled (sink removal) SRTM. This procedure produces parallel straight lines for a perfectly flat areas, like the ones filled up by the ArcInfo fill algorithm. These straight lines produce very high PDD values for the corresponding area. The phenomenon is actually right, because this high PDD occur only in the filled depressions, where water flow in a convergent way and thus results in a naturally high PDD value anyway. However, this approach produces a slight geographic shift of the high–PDD patterns, because the higher elevated borders of the flat area, from where the drainage lines arrive, has lower PDD values up to the point where the drainage lines starts to form, approximately to the distance of the threshold cells. Therefore the higher elevated side will have a close to the threshold wide low-PDD corridor before the drainage lines starts to be formed.

Therefore, the channel line developing procedure has been changed and the old version of PDD was replaced with this newer version.

3.1.1 Automatically derived channel lines (water flow lines, stream lines, thalweg)

Remark: Up to date the open source GIS software SAGA is distributed as version 2.0.4. The channel lines module is at the moment only implemented in SAGA version 1 what means it is not executable in SAGA version 2.0.4. The migration of SAGA modules of scilands from SAGA version 1 to 2.0.4 is still under progress.

A water flow line (synonyms: channel line, stream line, thalweg) is a line, which is connecting the deepest points in an open hollow (valley) on the earth's surface. Considering a water impermeable surface, a continual sprinkling would lead to a linear discharge along this lines. This is true in the case of streams and rivers.

3.1.1.1 Flow lines as substitute and supplementation for a digital stream network

The automatically derived flow lines simulate the course of streams and rivers. Therefore, they can be used in Digital Terrain Models (DTM) to substitute not existing digital stream networks. In addition automatically derived flow lines constitute an important supplementation for existing stream networks. In dependence to the climate environment and the weather conditions, the discharge, in particular in small streams, can vary in a large scale. This means that the stream may perennial, periodical or episodically contains water. In particular episodic stream channels in the mountains, where discharge only occurs during seldom heavy rainfall, are often not represented in the digital stream networks. However, these episodic stream channels are relevant for the flood control and the protection of erosion. Automatically derived flow lines can show systematically, where liner superficial discharge may occur in side-valley and in slope depressions.

3.1.1.2 Water flow lines as relief framework line and local base level of erosion

Water flow lines represent beside the drainage divides the most important framework lines for the geomorphology and terrain analysis. Flow lines are the local base level of erosion, where all processes of transport (soil, water and solutes) on the slopes and in the drainage basin are related to.

3.1.1.3 Processes for the delineation of flow lines from Digital Terrain Models

Since many years different algorithms have been developed to determine flow lines on the basis of Digital Terrain Models. However, the results often kept insufficient because not every necessary aspect was considered to receive plausible flow lines. To find a remedy for this case, partner Scilands developed sophisticated techniques. Beside others, the following aspects were taken into account:

- Pre-processing of the Digital Terrain Model (for example to create 'exits' for closed hollows), to ensure a continual decline for the flow lines;
- The intensity of the convergence of the superficial discharge and the consideration of potential existing divergences (like they occur at alluvial fans);
- The simulation of discharge and the calculation of drainage basins using 'multiple flow'-methods;
- Using different criteria for the start and end points, resp. the course of flow lines, like for example minimum-convergence-index, minimum size of the drainage basin, minimum length of the flow line, maximum length of divergent segments;
- Hierarchic structure of flow lines by the size of drainage basins (which flow lines meets with which superordinate flow line).

3.1.1.4 Hierarchy of flow lines

All flow lines as vector data are provided with the attribute data as shown below:

ID	= No. (individual id) of the flow line
Tree_ID	= ID of flow line "tree" (drainage basin) to which the flow line belongs

Parent_ID = ID of the superordinate flow line

Order_No = hierarchy (order) No. of the flow line

Drainage Area = size of the drainage basin [km^2] of the flow line

notes:

1) There are *no objective criteria for the starting points of flow lines*. The start parameter (for example minimum-convergence-index, minimum size of the drainage basin) has to be adjusted optimal to the specific problem.

2) The created algorithms for the delineation of flow lines provide in contrast to many other methods plausible results even in flat sloped areas.

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

Previous tests of the Dobos 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, the relief intensity, slope, elevation and the 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 to form a final polygon system. This latter approach produces several analogue, but not perfectly fitting 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 different geomorphological units. Adding the PDD is just overcomplicating the procedure, while no add-on information is produced. Contrary to the high relief areas, PDD is one of the most significant parameter for the terrain differentiation on a low relief area, where the slope, relief and the elevation have only slight variations. Therefore the decision to pre-stratify the mapped area into high and low relief has been made. Threshold of 100 meter/square km 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 the PDD and all four were used together for the low relief areas. This approach significantly decreased the number of slave polygons top handle.

4. Task 1.3: Parent material classification

4.1 Task 1.3.2 Revised classification of parent materials relevant for soil development

4.1.1 General Scheme of the revised lithology structure

One of the major aims of the parent material classification revision was to simplify the existing one and adopt the system to take alternative PM input data where no complete PM information is available. One of the major limitations of the existing PM classification is the lack of knowledge to be used by a soil scientist on the field. The majority of the soil scientist has not enough field knowledge to differentiate and classify certain rock types, even if they occur in a none-mixed way. However, the situation is often further complicated by the processes of mixing, moving, erosion and deposition of the weathered material from the original formation. Classification of these different origin rocks and parent materials is the expertise and responsibility of the geologist. For a soil scientist, the most important is the weathered, unconsolidated material, from which the soil is actually forming of. Unfortunately, there is a great diversity of interpreting the PM in the national and international system, mixing the term with underlying rock and many national systems provides information on the latter one, rather than on the actual parent material. The US system and definition clearly defines, that parent material is only the unconsolidated material, from which the soil is forming on, and that is the material which can be described in the appropriate details by soil scientists. That is why the new classification is concentrating on the unconsolidated material and its major properties, like texture, carbonate status and genetics (Fig. 4.1.).

4.1.2 Rationale and framework conditions for the lithology classification

The new system to be developed within e-SOTER has to maintain its original mapping scheme, where the major components for the delineation of the geometric units are the physiography and the parent material. The major limitations of the traditional SOTER products, as it has been identified so far, is the inconsistency of the international SOTER coverages. SOTER is defined to incorporate existing data into a harmonized database. It is more like a correlation and 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 the case. The delineations of the SOTER units vary from country to country. Therefore, the effort to develop a more consistent way to delineate the units was started utilizing our state of the art terrain modeling knowledge and the newly emerged SRTM data. Using SRTM as a common input data makes sure that the procedure does not vary among the countries. These physiographic units have to be further subdivided by the Parent Material (PM) information. PM classification, like the soil one varies from country to country as well. Therefore, an international system has to have very general classes to be able to incorporate the national units (factor of globalness). And even if the classes are well defined, the polygons are coming from the national system, and inheriting their own, and often different way of delineations and interpretations of the classes. Importing these units immediately introducing significant spatial inconsistency into the database. The only way to avoid this problem is to develop PM coverages in controlled, quantitative way, or at least increase the quantitative portion within the whole procedure. It is rather important, because legacy data is often limited in existence or accessibility, hence SOTER cannot be completed for those areas. A quantitative procedure for PM delineations is urgently needed to complete SOTER where no PM information available and to increase the level of harmonization where legacy data can be incorporated into the database development.

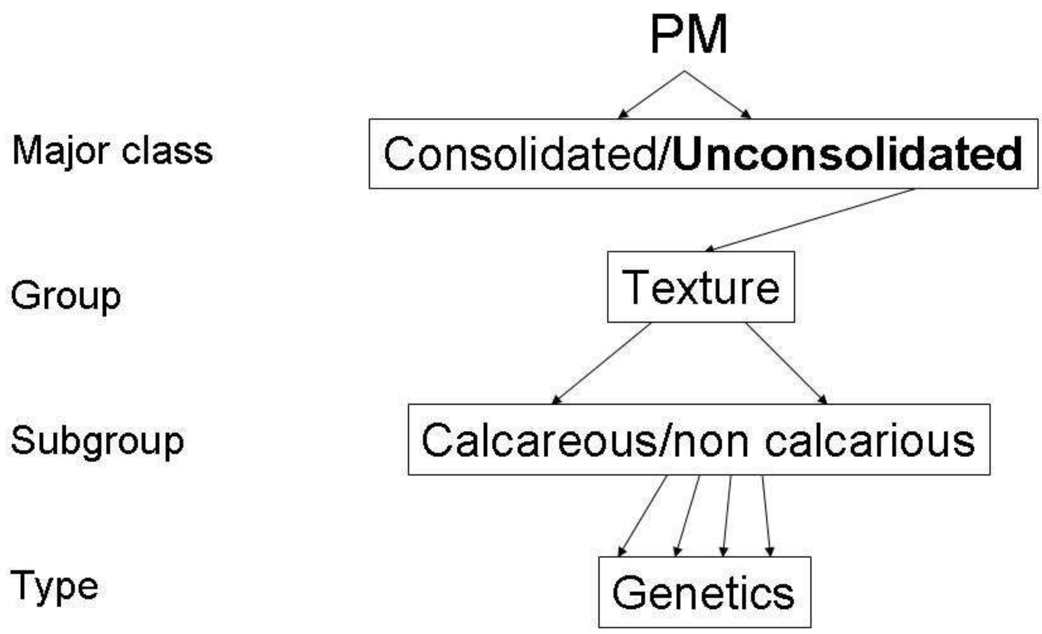


Fig 4.1. Classification scheme for the Unconsolidated PM classes

These are the properties that are described on the field with high level of reliability, and these are the ones that make the real differences in the soil formation process. However, this approach does not neglect the geological information, because the differences in chemical and crystal structure of the different consolidated rocks and their impact on the soil formation is still understood and appreciated. However, the detailed information can be imported only from existing, interpreted geology maps and databases, which are often none accessible on the field or for a significant portion of the Earth surface.

Therefore the first level of classification starts with the differentiation between the consolidated and unconsolidated parent materials.

Unconsolidated material: loose inorganic/organic material, that is by nature accumulated/deposited in a deeper stratum by water or ice (fluvial, estuarine, lacustrine, marine, glacial) or by mass movements (like the colluvial materials).

Consolidated material: solid rocks and its shallow weathering residuum, having mainly the typical mountain soil associations like bare rock/Leptosol/Cambisol, and by genetics it can be eluvial, colluvial or bare rock.

Remarks for the definitions:

The widening of the content with the weathering residuum is basically an unavoidable compromise, because the existing soil maps with parent material information for this kind of areas describe only the underlying rocks and gives no information on the properties of weathered material.

The lithological units of the existing maps often not even described in clear classes, only as associations or age groups, in which, for example, shale, sandstone and limestone can occur in the same group (like in a Triassic sea sediment). These three are extremely different as parent material, so there is no any reliable and reasonable way to describe them as one, because the combined class is so wide that it has no any add-on information any more.

The other reason for this definition is coming from the quantitative procedure. There is no reliable, available quantitative procedure to derive geology/lithology information in the detail required by any PM classification, like the previous SOTER one. They often occur as mountains with relatively dense vegetation (no way for RS applications) with higher relief and elevation, and very high and large scale diversity, which cannot be represented in a minimum of 25 km^2 unit (defined by the SOTER scale of 1:1M). The unconsolidated areas can be delineated relatively easily with quantitative procedures using RS and DEM data, but further separation of the lithology units is feasible only for areas having legacy data.

The new classification starts with the separation of consolidated/unconsolidated material using a quantitative approach. The consolidated areas are than further subdivided into bare and none bare rock. The non-bare rock area can have two subunits, eluvial and colluvial. However, the spatial mixing of these two is often to complex to differentiate between them (the only potential is to the give proportions within the geometric units). This is the detail, where the quantitative procedure has to stop (Level of Genetics, Fig. 4.2.)

In case of having legacy lithology data, the classification can go further and the level of information defined by the revised lithology classification can be filled in (Major class, Group, Type).

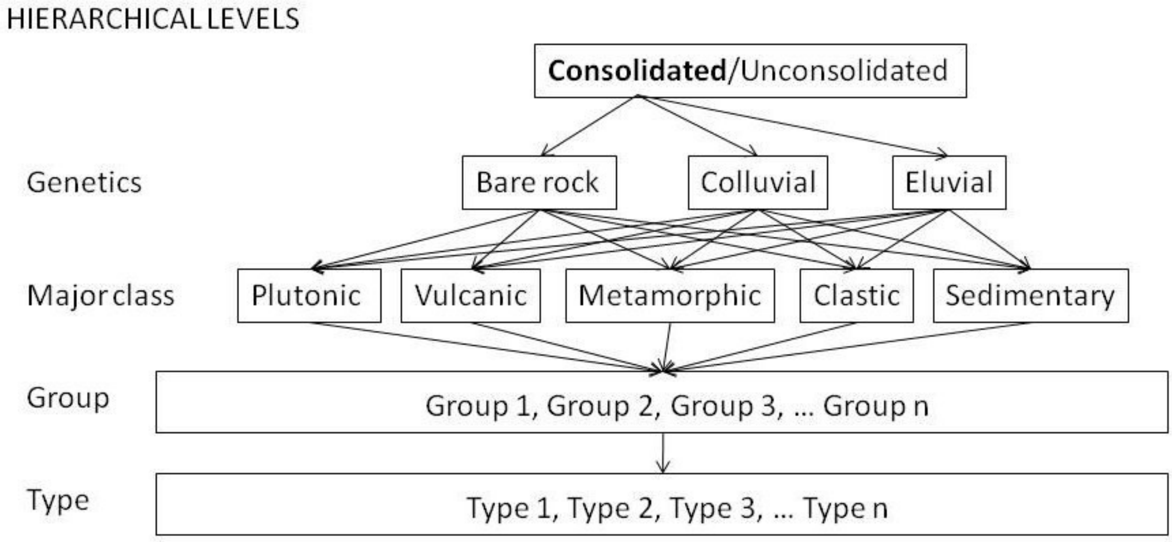
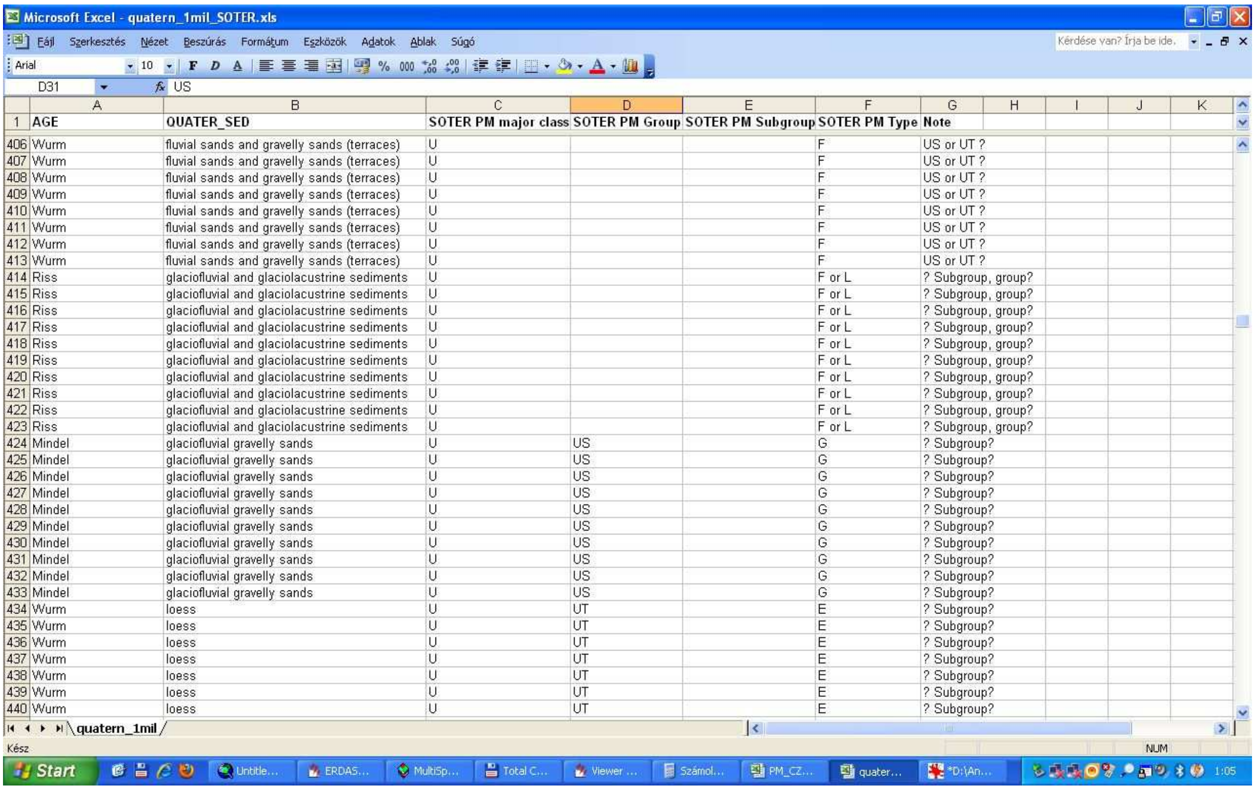


Fig. 4.2. Classification scheme for the Consolidated PM classes

Based on the rationales described above, a new, revised classification system was developed based on the existing SOTER structure and the new demands for adopting the system to a quantitative procedure. This system was tested using national databases and severe limitations were identified due to the hierarchy of the system. An example of the Czech quaternary database is shown in Fig. 4.3.



	AGE	QUATER_SED	SOTER PM major class	SOTER PM Group	SOTER PM Subgroup	SOTER PM Type	Note
406	Wurm	fluvial sands and gravelly sands (terraces)	U			F	US or UT ?
407	Wurm	fluvial sands and gravelly sands (terraces)	U			F	US or UT ?
408	Wurm	fluvial sands and gravelly sands (terraces)	U			F	US or UT ?
409	Wurm	fluvial sands and gravelly sands (terraces)	U			F	US or UT ?
410	Wurm	fluvial sands and gravelly sands (terraces)	U			F	US or UT ?
411	Wurm	fluvial sands and gravelly sands (terraces)	U			F	US or UT ?
412	Wurm	fluvial sands and gravelly sands (terraces)	U			F	US or UT ?
413	Wurm	fluvial sands and gravelly sands (terraces)	U			F	US or UT ?
414	Riss	glaciofluvial and glaciolacustrine sediments	U			F or L	? Subgroup, group?
415	Riss	glaciofluvial and glaciolacustrine sediments	U			F or L	? Subgroup, group?
416	Riss	glaciofluvial and glaciolacustrine sediments	U			F or L	? Subgroup, group?
417	Riss	glaciofluvial and glaciolacustrine sediments	U			F or L	? Subgroup, group?
418	Riss	glaciofluvial and glaciolacustrine sediments	U			F or L	? Subgroup, group?
419	Riss	glaciofluvial and glaciolacustrine sediments	U			F or L	? Subgroup, group?
420	Riss	glaciofluvial and glaciolacustrine sediments	U			F or L	? Subgroup, group?
421	Riss	glaciofluvial and glaciolacustrine sediments	U			F or L	? Subgroup, group?
422	Riss	glaciofluvial and glaciolacustrine sediments	U			F or L	? Subgroup, group?
423	Riss	glaciofluvial and glaciolacustrine sediments	U			F or L	? Subgroup, group?
424	Mindel	glaciofluvial gravelly sands	U	US		G	? Subgroup?
425	Mindel	glaciofluvial gravelly sands	U	US		G	? Subgroup?
426	Mindel	glaciofluvial gravelly sands	U	US		G	? Subgroup?
427	Mindel	glaciofluvial gravelly sands	U	US		G	? Subgroup?
428	Mindel	glaciofluvial gravelly sands	U	US		G	? Subgroup?
429	Mindel	glaciofluvial gravelly sands	U	US		G	? Subgroup?
430	Mindel	glaciofluvial gravelly sands	U	US		G	? Subgroup?
431	Mindel	glaciofluvial gravelly sands	U	US		G	? Subgroup?
432	Mindel	glaciofluvial gravelly sands	U	US		G	? Subgroup?
433	Mindel	glaciofluvial gravelly sands	U	US		G	? Subgroup?
434	Wurm	loess	U	UT		E	? Subgroup?
435	Wurm	loess	U	UT		E	? Subgroup?
436	Wurm	loess	U	UT		E	? Subgroup?
437	Wurm	loess	U	UT		E	? Subgroup?
438	Wurm	loess	U	UT		E	? Subgroup?
439	Wurm	loess	U	UT		E	? Subgroup?
440	Wurm	loess	U	UT		E	? Subgroup?

Fig. 4.3. The Czech SOTER “correlation”

Column B contains the original descriptive classes of the Czech data, while columns C, D, E and F represent the major class, group, subgroup and type levels of the revised classification. Keeping the hierarchy, as it is clearly visible on the datasheet, would erode a significant portion of the data, because there are missing steps in the description flow. On the other hand, these levels are not necessarily linked in order, each column represent a stand-alone level of information, hence disaggregating the hierarchy and using the different levels as different, equivalent-valued properties helps to maintain as much information as possible (Fig. 4.4.).

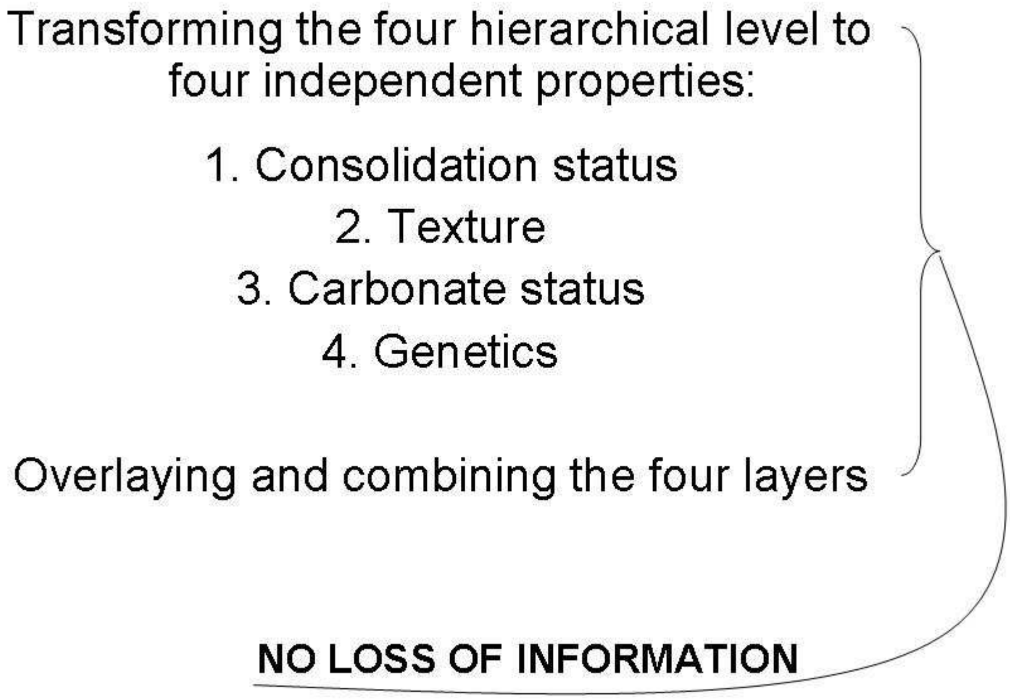


Fig. 4.4. Disaggregation of the hierarchy of the unconsolidated PM classification.

4.2. Task 1.3.1 Determination of soil parent material within the landform units, based on low-resolution satellite imagery (AVHRR, MODIS, SPOT Vegetation) and DEM data in combination with legacy soil parent material data

Remarks for this chapter:

There are three potential situations that the database developers are faced with when SOTER database has to be compiled.

(1) The first case assumes that there is existing and accessible legacy data for the whole area to map in an appropriate scale. This situation 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. This example was tested and studied by INRA, and a case study for incorporating the 1:1M European Soil Database information into the SOTER database is demonstrated (Thanks to Joel Daroussin). In this case, a close a full set of descriptive attributes can be loaded into the database.

(2) The second typical case is when there is limited data for only a certain portion of the mapping area. This situation requires a harmonization effort, input data development effort (digital mapping of PM info) and at the end the procedure to incorporate the info into the SOTER database. This is the most typical case for the windows. The limited coverage PM info is used as training and calibration info for creating a full coverage for the whole area. This 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.

(3) The third is the no data situation, when only general relationships and rules can be used to derive some delineations.

The first case have results in a close to complete set of PM information.

The second case has only a limited set of variables that can be derived with a relatively high level of reliability/accuracy. These are the consolidated status, the texture classes, and the genetic classes of fluvial, marine/estuarine, eolian, colluvial, eluvial, and bare rock. Large scale studies in a favorable environment can be done for further refinement and more attributes, but these are site specific and no generally usable procedure exist so far.

The third is the most limited case, when only the genetic classes listed above can be developed.

This subtask aimed to delineate PM units that follow the revised lithology structure of SOTER (described in the Task 1.3.2.). As it was agreed, PM is considered to be unconsolidated material, which is characterized by its texture, carbonate status and genetic classes (Fig 4.1.)

The three (four with the consolidation/unconsolidation) parameters require four separate layers to develop. The first layer is the consolidated/unconsolidated one. That has to be the first step in the procedure to stratify the area to be mapped into two groups. The two main areas than require different approaches and different classification steps.

4.2.1. Covariates used to derive the thematic PM layers

- RS images
    - MODIS-multi-temporal 8 days composites, 11 bands, visible to the thermal spectra, 5 dates covering the snow period, 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 meter resolution
        - SEE ANNEX I. for the derived parameters and band processing steps

  - Landsat images
    - SEE ANNEX IV. for the derived parameters and band processing steps
- Digital terrain model, SRTM
    - SEE ANNEX II. for the derived parameters and band processing steps

In order to strengthen the performance of the classification, multi-temporal images of none-altered MODIS bands were compiled into a 55 layers image representing the visible, NIR, MIR and thermal bands, and also to capture the temporal environmental conditions and changes that reveal to surface conditions and therefore to the soil/PM properties, like speed of wetting and drying out, cooling down or warming up, which are parameters strongly correlating with the texture, color, water content and water holding capacity. However, the 55 layers has a significant portion of information overlap, redundant info in the images, hence a PCA was used to decrease the number of input images and decorrelate the bands information. The best 15 PCA component was maintained and incorporated into the final image.

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. These band ratios were adopted to MODIS and were derived for each of the 5 dates, resulting an other 15 images, that have been added to the final image.

Previous studied also suggested to use surface temperature information, like the thermal bands of the MODIS (Bands 31, 32) and the LST (Land Surface Temperature) products (night and day) that have been derived from them. The daily temperate fluctuation is a function of the thermal capacity of the surface material, which is the function of the kind of material, texture, color and water content, basically the factors we are interested in. Therefore, a new normalized band combination was developed. The daily temperature difference 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 difference in potential energy intake from the sun. These were calculated for each dates as well.

SRTM data was used in combination with the MODIS derived layers as well. Annex II. gives the details for the derived parameters. The basic parameters are the followings:

Elevation (sinks are filled up to certain level)

  - Slope percent

- Relief Intensity
- Potential Drainage Density
- Groundwater level
- Topographic Wetness Index
- Upland/Lowland
- Convexity (not added to the basic image, used only for the colluvial image derivation)

The listed derivates are either used in the SOTER methodology already, or believed to add significant information for differentiating between the classified parameters. The SRTM images were degraded to the level of MODIS resolution and a 42 layers image containing the 15 PCA layers, 6 SRTM derivatives, 5 normalized LST difference images and 15 band ratios.

REMARK:

The developed procedure is demonstrated on the Central European Window.

4.2.2 Classification of the consolidated/unconsolidated areas

The first step in the process is the classification of the window into two classes, the consolidated and unconsolidated ones. The approach is based on traditional RS image processing/classification technique, Maximum likelihood supervised classification algorithm using the combined, 42 layers base image (ANNEX III.) Several direct approaches were evaluated, however no one had a good enough 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, when the vegetation masks out the PM signal of the images. These stochastic relationships can be well utilized in a supervised classification framework.

Training data was limited for the window (Fig. 4.5.). We used three small training areas for the Checz republic and the Hungarian part of the window. The data sources had to be interpreted and the training areas for the classes defined. The traditional PM classification varies a lot by countries, but almost none of them matches the definition of being unconsolidated. Parent material for the areas having consolidated rocks with shallow unconsolidated material /soil on the top is considered consolidated and named after its underlying material like, granite, basalt, etc., regardless of what the material is really coming from, whether it is mixed with other, in situ, or colluvial, etc. Anything, that has a bare rock, leptosol, cambisol soil association on a mountainous area considered solid rock. This fact had to be taken into consideration ad the training classes were merged this way, meaning that the colluvial, eluvial materials, in situ weathered relatively shallow material is not considered to be unconsolidated. Fig 4.6. shows the classification results for the Central European window.

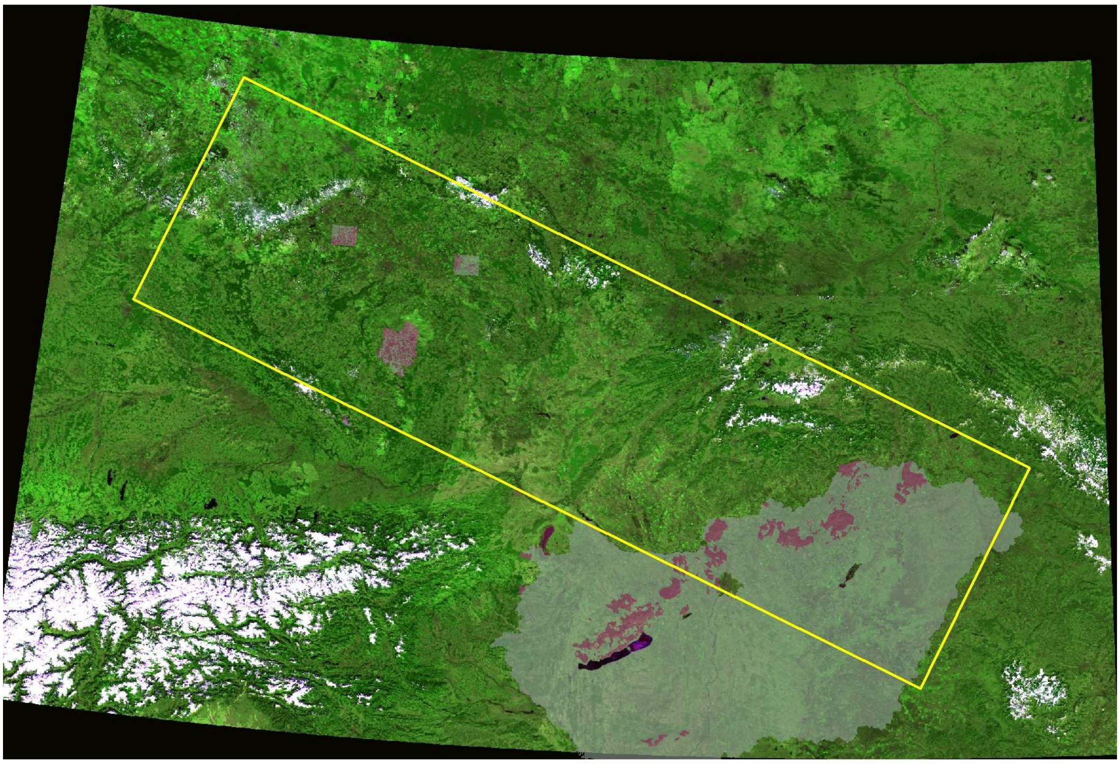


Fig. 4.5. The Central European window (yellow box) and the training areas.

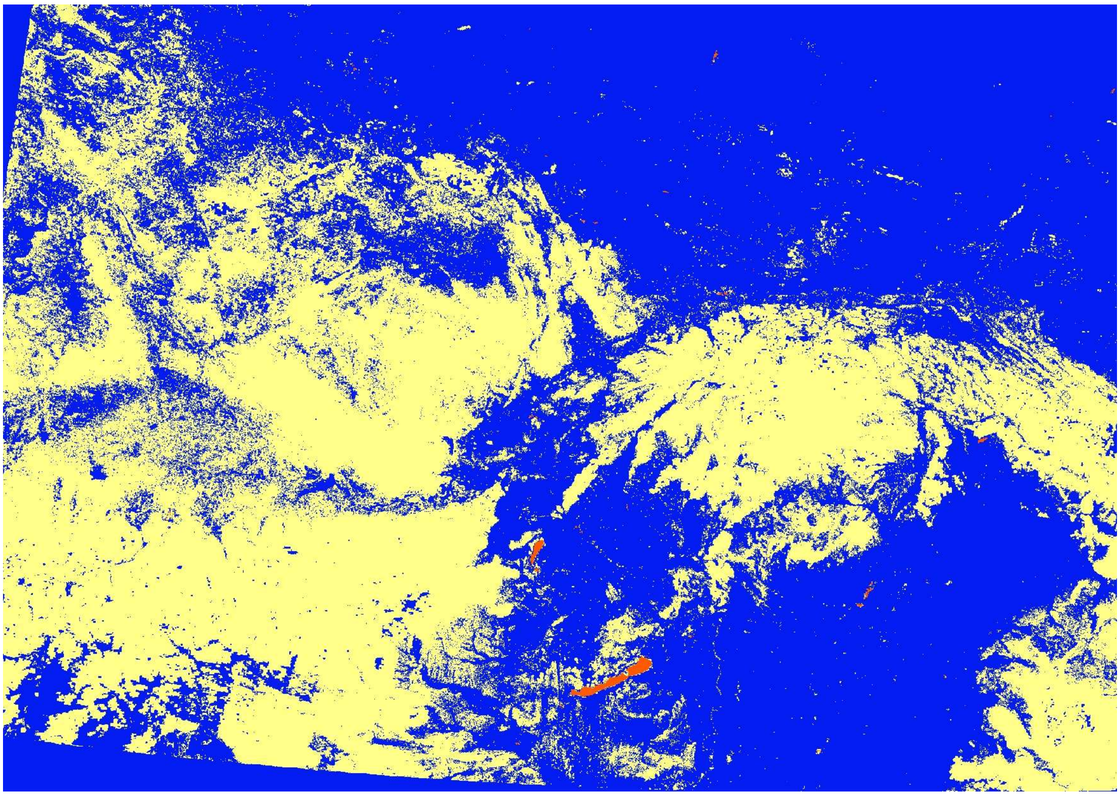


Fig 4.6. The result of the CE window classification. Blue means unconsolidated, yellow is consolidated PM, orange is water.

The classified image was used to stratify the area into the two major classes. The consolidated and unconsolidated parts are handled and classified differently from this point.

4.2.3. Classification of the consolidated material areas

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

4.2.3.1 Bare rock

The bare rock classification was done using the NDVI (Normalized Difference Vegetation Index) image generated from the peak of the vegetative period, like summer in the CE window, 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 and these were assigned the bare rock class. The value was set by comparing the images with Landsat and other high resolution images. This value and the procedure in general works very well in the temperate and tropical zones (CE, FR/UK and Chine windows). The only problematic window is the Moroccan one, where there is a significant portion of none-vegetated area due to the climate, even on a non-bare rock area. The procedure has to be refined after strong discussion and reinterpretation of the terms on a climatic zone basis. Bare rock areas often occupy small areas, spot, much less than the 25km2 threshold of the polygon limit. However, due to the significance of this class for soil issues and to the changing database format (being digital, where visualization is not a strong issue any more). We decided to keep the bare rock patterns, which are bigger than 5 km^2 . A two step focal majority function was used with the radiuses of 3 and 14 respectively to clean the image and remove the less than threshold sized patterns.

4.2.3.2 Colluvial areas:

The colluvial areas were delineated using the assumption that the colluvial material accumulates in lower sections of the slopes and often starts changing the shape of the slope to be concave. Based on this a two criteria system was set up with two thresholds:

- Curvature (filled z limit 20 SRTM) < 0
- Slope percent > 2 %

Running the selection algorithm results in the colluvial material image (Fig. 4.7.) The SRTM image based classification has a relatively high resolution for the area, with many slope scale patterns, much less than the minimum polygon size. Therefore this information was added to the PM polygon delineation, but will be added as descriptive percent cover within the polygons.

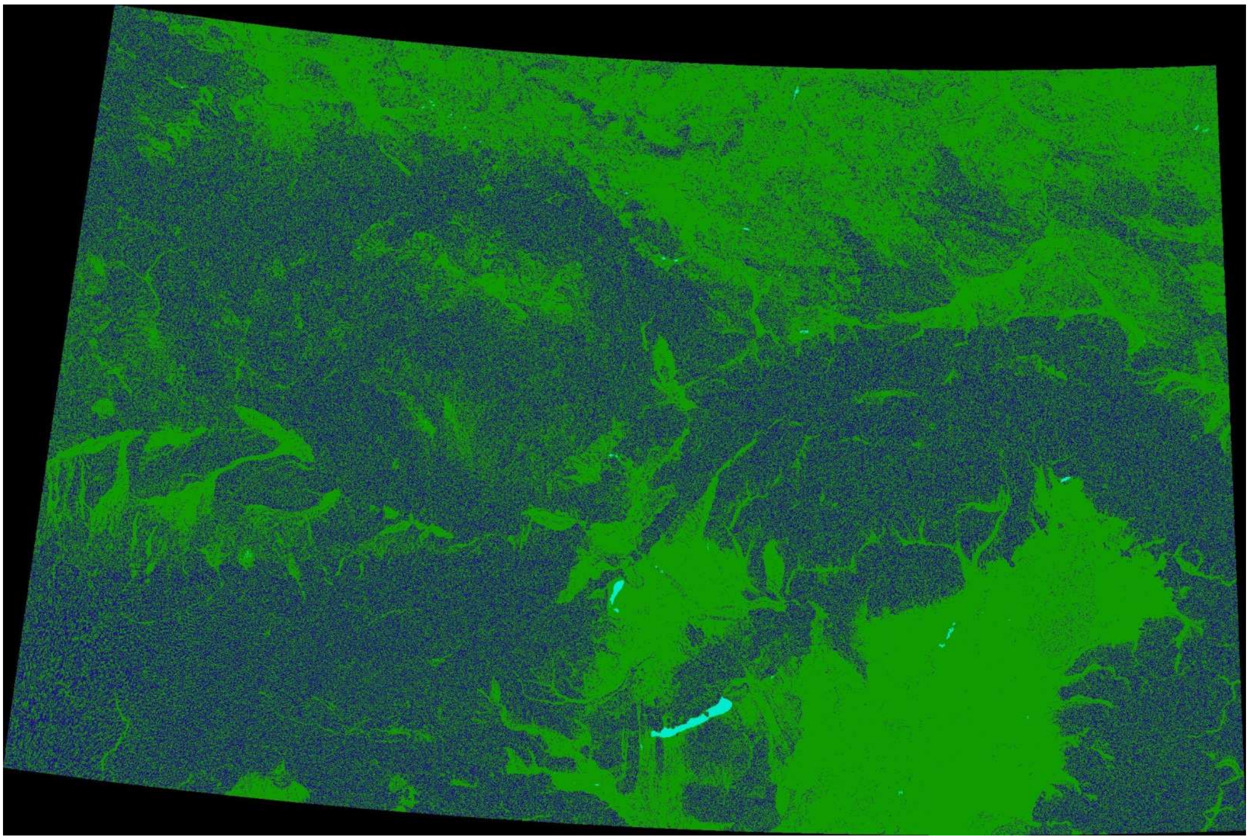


Figure 4.7. Colluvial areas (dark blue) of the Central European window (curvature<0, slope%>2)

4.2.4 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, eolian, lacustrine, marine, estuarine, glacial till, glaciofluvial

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

4.2.4.1 Developing the texture class layer

The texture classification was done the same way as the consolidated/unconsolidated layer, using the 42 layer combined image and training data for the supervised classification.

The coverage for the training area was the same as the one for the consolidated classification (Fig. 4.5.). The texture layer is shown on Fig 4.8. This step of the procedure requires much knowledge of the area (for validation purposes) and also of the use of the classification tools to achieve the best optimal results. No automatic approach can be developed, expert user is needed. Any kind of preprocessing, feature selection or extraction, and training data enhancement procedure are welcome to use.

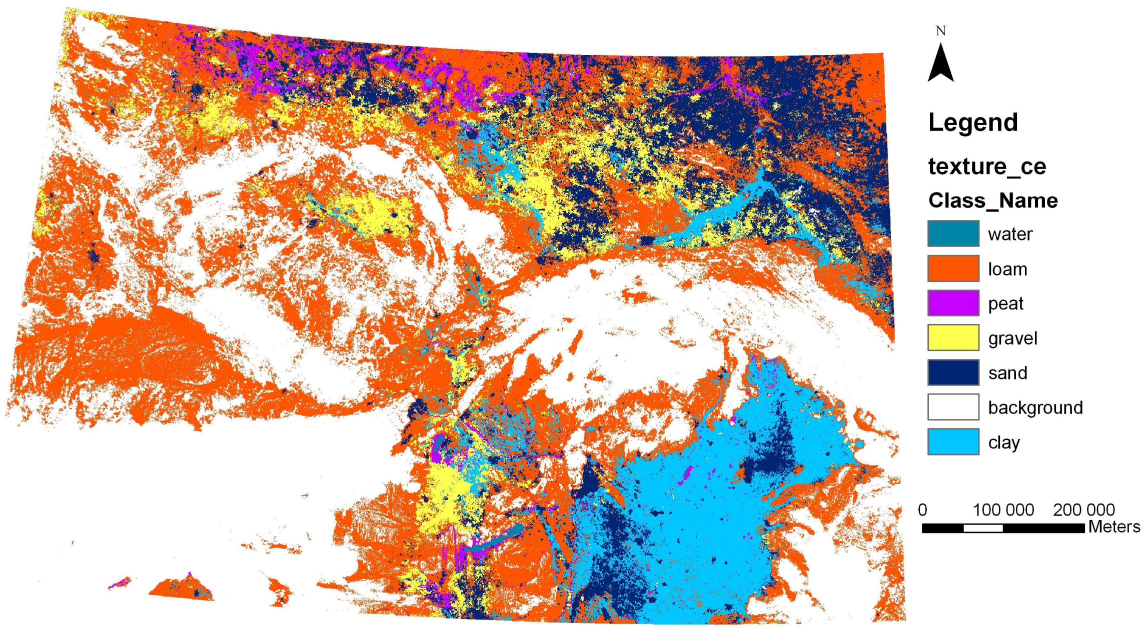


Fig. 4.8. The classified texture classes for the Central European Window.

Training data

The most critical part of the procedure is the training data that is often the most critical factor. The optimum case is when relatively high resolution training data is available with clear classes, equivalent with and correlated with the output classes. For the developed part of the World 1:100K to 1:250K data sources are commonly available ones, which contains aggregated but still concrete classes (not associations). These data sources can be used as direct inputs for the supervised classification.

Smaller scale training data, like 1:500K to1:1M is often less useful, especially because the level of thematic aggregation of the class units is very high, and no clear units can be identified within the units. This kind of information is structured into associations of clear units, with some proportional information added as well. This latter case is very common, especially because these small-scale maps are the only publicly available ones, larger scale maps cost much to develop or purchase for any purposes. Using these small-scale data sources requires some thematic preprocessing and quality check. Two approaches were developed for this project. A common approach to interpret and visualize the associations is to select the dominant member of the association. However, these dominant members are often not really dominant within the polygon, but simply the one occupying the highest portion of the polygon area. This problem can be handled with selecting only the pure polygons having only one member in the association. However, this is not really common, and often due to a less precise mapping approach. The percent coverage we can achieve this way is very low and the training dataset is often non-representative. Rocks/Soils naturally occurring in associations in smaller mixed patterns of the different types will be underrepresented, while purely occurring features, like a Rendzina soil in the limestone Eastern-Alps has higher representation in the database. Therefore a threshold of 80 % was set as a minimum percentage of the dominant soil type proportion within the polygon and only the polygons having real spatially dominant element were selected for training purposes. This approach was used to train the consolidated/unconsolidated and texture classifications for the French part of the window, where no data was available at all, but the internationally developed datasets, like the 1:1M soil database.

An alternative approach was the enhancement of the data using Landsat images. The processing steps of the Landsat images are described in Annex IV. Landsat images were preprocessed to remove the clouds, snow and vegetation covered areas to maintain only the bare ground areas. Geology/lithology driven band ratios were collected from the corresponding literature and were applied to the Landsat images. Landsat tiles were selected from the early vegetation periods and two –three of them were processed to cover the representative portions of the window in order to cover the thematic range of the PM classes for the whole window. These Landsat images were displayed as RGB images to highlight the major patterns (Fig. 4.9.). This procedure helps in the selection of the homogeneous polygons manually, only the polygons having clear, homogeneous Landsat occurrence patterns are used as training polygons.

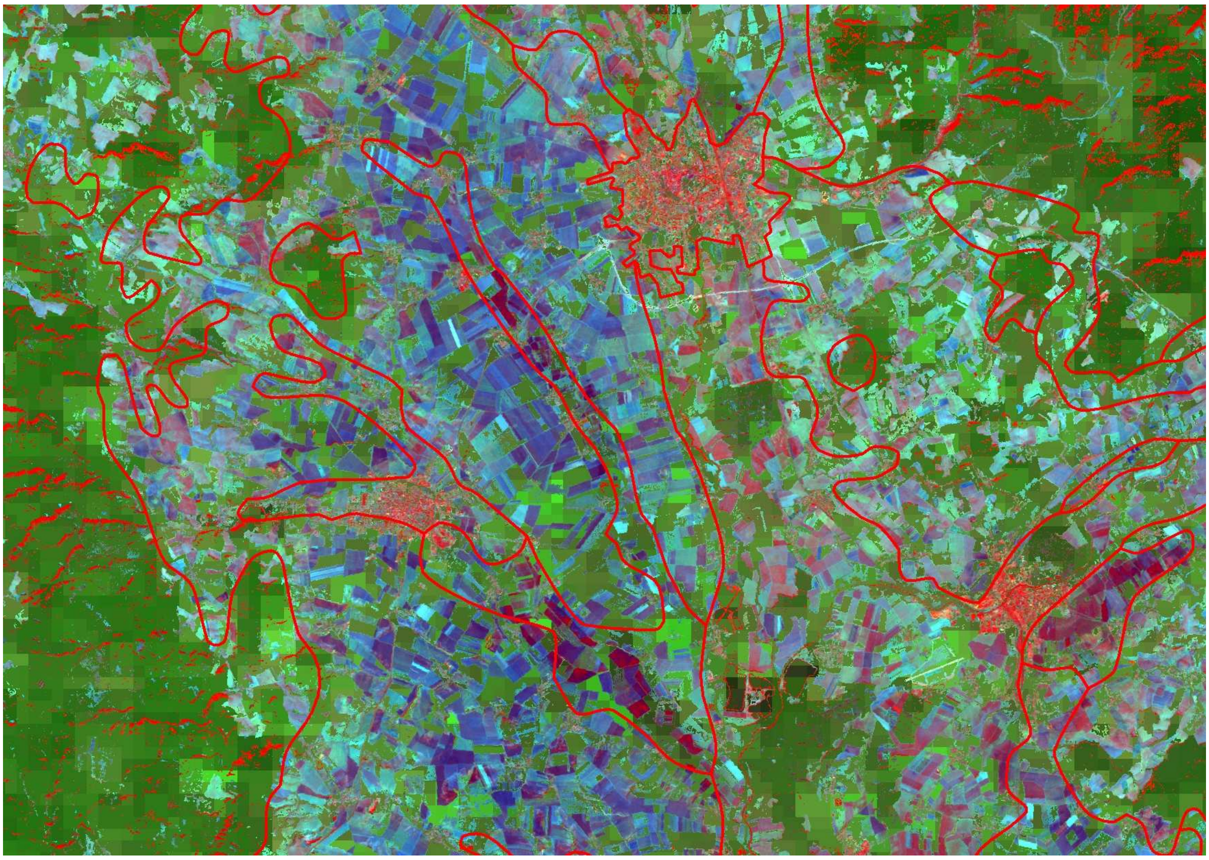


Fig. 4.9. The processed Landsat RGB image overlain by the Czech SOTER polygons. The green colors are the masked areas.

An alternative approach was a two-stepped classification. First the screened, filtered small-scale data was used as training data for the Landsat classification. (An assumption that the doinat element of the association is really spatially dominant as well was made. Therefore only a small portion of the polygon was selected as non-real member of the dominant class, and these pixels will be represent the tails of the probability distributions of the pixel values within the class histogram. The tails of the neighboring, overlapping class histograms are cut by the maximum likelyhood classification algorithm and the remaining class histogram will have a better fit to reality. In the last step, this classified Landsat image is used as training for the MODIS-SRTM classification.

4.2.5 Development of the genetic layers

The following genetic layers were developed:

- Fluvial/alluvial/lacustrine/glaciofluvial
    - Plain, low slope and low relief intensity areas, close to the groundwater level
- Marine and Estuarine
    - Follows the seashore line and characterized with 0 - 5 meter elevation along the seashore
- Eolian/older terraces/glacial till plain
    - Higher relief, higher above the groundwater level, not influenced by the fluvial activities.

4.2.5.. Fluvial/alluvial/lacustrine/glaciofluvial and the Eolian/Older terraces/Glacial till moraines areas

These are the areas, which have a plain, smooth surface, low relief intensity and an elevation close to the ground water/surface water level. These assumptions were translated to terrain modeling language. The final solution for the delineation has only one criterion, namely the closeness in vertical distance to the surface/groundwater level system. Ground/surface water level grid was simply extracted from the filled SRTM DEM and the difference is the depth to the ground water. For the derivation of the groundwater level see Annex II. In theory, the image selects the areas which are less than 3 meters elevation above the surface waters in the neighborhood. The procedure delineates the potentially flooded areas of both the rivers and the lakes, basically al kind of surface waters. That is why the fluvial and lacustrine sediments are combined in this classification. Fig. 4.10. shows the Fluvial areas of the Central European window. The glaciofluvial areas have similar appearance to the fluvial ones, only the source of the material to be deposited is different.

Eolian areas are the ones, which are free from flooding impact and also from significant ground water impact. Therefore, the non-fluvial areas are the potentially Eolian ones as well, with slightly higher relief intensity.

The most problematic genetic class is the Glacial till. Till can be till plain that is relatively plain (low relief), but lying higher than the fluvial areas. This can be identified with selecting areas with higher than 3 meter elevation above the water level, and have a plain/low relief surface. The glacial till moraines have much higher relief, sometimes similar to the Eolian areas and lies above the fluvial areas as well. That is why the Glacial till class was merged with the Eolian one if there is no more information to further specify the genetics of the polygons.

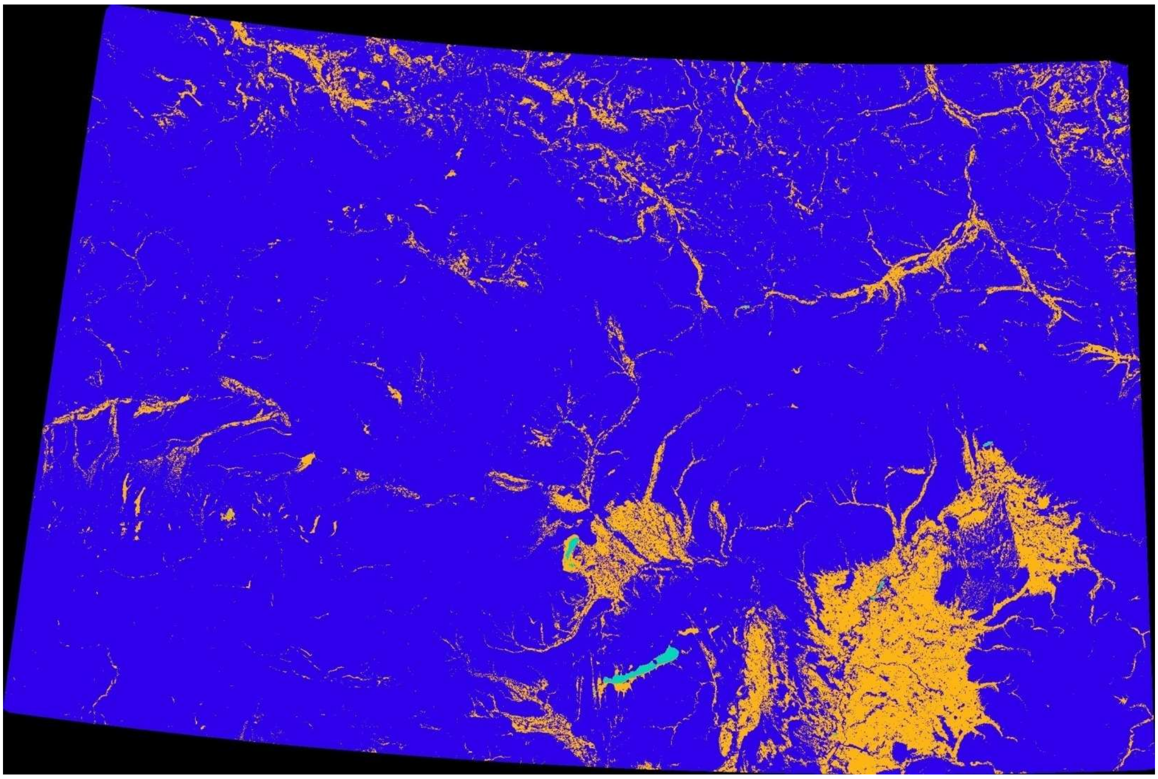


Fig. 4.10. Fluvial/Alluvial sediments (yellow) for the Central European window,

4.2.5.2 Marine and Estuarine class

The marine and estuarine sediments occur in the 0-5 meter zones along the seashore. The approach was to select the areas from the SRTM DEM which have a value between 0 and 5, and the contiguous areas lies along the sea shore. The procedure for delineating these areas is explained in the Annex II.

4.2.6 Finalizing the PM coverage

Input data:

- from satellite image classification
    - consolidated-unconsolidated in case of CE window
    - texture (consolidated, gravel, sand, loam, clay, peat, sapropel, diamicton, water)
- from SRTM DEM derivation
    - alluvial/eolian
    - bare rock
    - marine/estuarine

At the final stage of the PM development procedure the previously developed input layers had to be combined using certain priority rules. The procedure is described in Annex V. The resulting coverage is shown in Fig. 4.11.

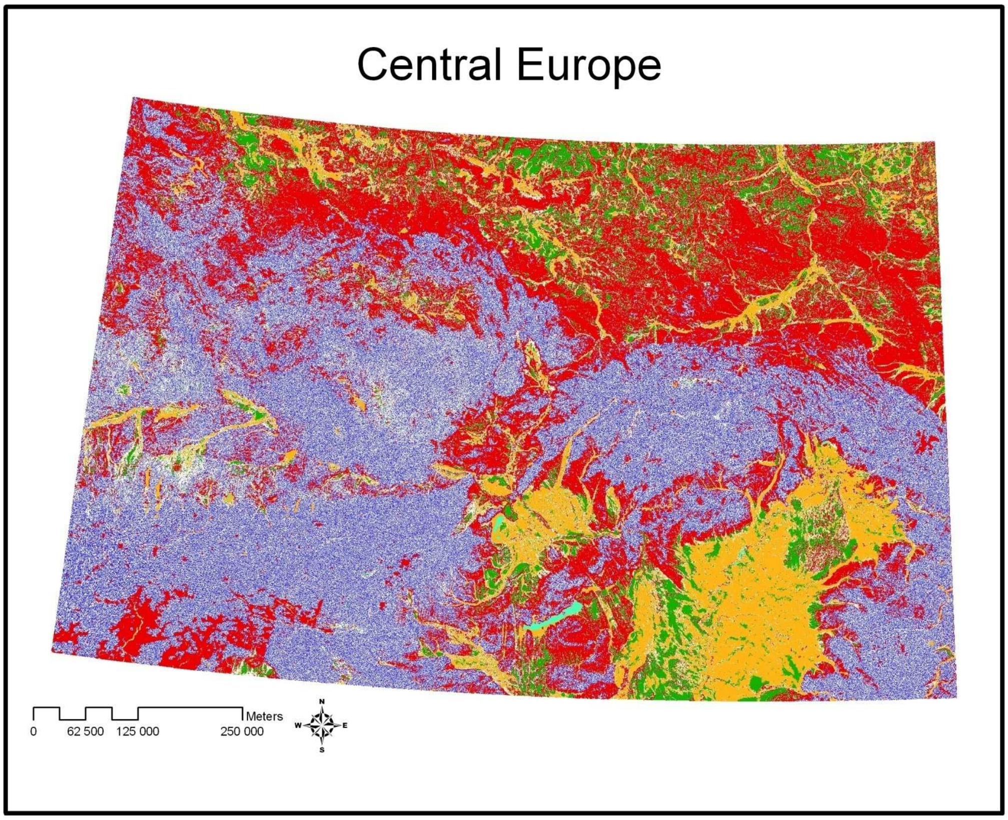


Fig. 4.11. The combined PM raster image for the Central European Window.

4.3. Comparison between modelled PM classes and legacy geological map data

4.3.1. Summary of the results from the PM modelling

The key data source for the modelled PM layer is based on a LANDSAT 42 bands layer images, combined and oriented based on training data. Before, it was necessary to enhance the Landsat images:

Removal of clouds, snow and vegetation covered areas to clearly identify the bare ground areas

Consideration of geology/lithology driven band ratios from the literature

Table 4.1. gives a summary of the terminology and modelled PM classes.

Table 4.1. Overview about the modelled PM classes

PM model		**PM legacy**	**comments**
parameter	**explanation**		
unconsolidated material	loose inorganic/organic material, that is by nature accumulated/deposited in a deeper stratum by water or ice (fluvial, esturine, lacustrine, marine, glacial) or by mass movements (like the colluvial materials).	loose sediments	definitions in legacy geological data agree with the modelled PM class.
consolidated material	solid rocks and its shallow weathering residuum, having mainly the typical mountain soil associations like bare rock/Leptosol/Cambisol, and by genetics it can be eluvial, colluvial or bare rock.	hard rock	- any soil developed from (shallow or deep) weathering residuum of a hard rock still has a consolidated parent material - in order to avoid data lacks from geology maps (parent rock is presented, but the weathered top is not important for geologists), eluvial/colluvial is counted “consolidated” in that it represents the weathered tops of solid rock
bare rock	Procedure: – LANDSAT band ratios to pronounce different geological / parent material features http://www.narss.sci.eg/uploads/Journal/06-12-2009.pdf – Masking vegetation/bare ground: low NDVI (< 0.8) (areas having no soil and	consolidated rock, of unknown area extent in geological maps	- geology does not differentiate between bare rock (at the ground surface) and weathered layers/shallow sediments on top of rock - delineation of PM-bare rock needs a ground check - management-induced lack of vegetation (e.g. clear cut) is

	thus no vegetative cover		patchy, threshold: area size > 5 km^2
none bare rock	either eluvial or colluvial; cannot be separated clearly with RS because of high spatial mixing	could be either consolidated or unconsolidated	
colluvial	curvature<0, slope%>2 (a plain to concave surface, with significant slope)		rational: colluvial material accumulates in lower sections of the slopes and often starts changing the shape of the slope to be concave
fluvial/alluvial sediments[1)]	terrain analysis: vertical distance to the surface/groundwater level system (channel line system)	unconsolidated sediment coupled with relief	identification of plain, low slope and low relief intensity, close to the groundwater level
eolian	inverse of the alluvial: higher relief, higher above the groundwater level, not influenced by the fluvial activities		
marine/estuary	follows the seashoreline and characterized with 0- 5 meter elevation along the seashore		
texture classes	– identifiy bare ground from LANDSAT – lithology driven band ratios – refined delineation using textures of dominating soils of soil map untis	as above, coupled with expert knowledge	No automatic approach is possible; the quality of the training data set and expert-knowledge is crucial

[1)] also includes: fluvial/alluvial/lacustrine/glaciofluvial and the Eolian/older terraces/glacial till moraines areas

It can be concluded that the PM modelling has built primarily on LANDSAT image processing combined with the modelling of relief data (SRTM). In the following, a comparison between the modelled PM classes and geological maps was conducted. This is the initial step of a series of investigations about the utilization of legacy geological map data for eSOTER and digital soil mapping.

4.3.2. Introduction

The primary focus of the eSOTER methodology development for PM mapping in the resolution of roughly 1:1,000,000 (or +/- 1 km) is to exploit options for using remote sensed data. Geological data from maps may support the refinement of the delineation procedures (which are based on vegetation indexes derived from satellite data), but may also contribute information about the kind of soil developed from weathering of stones (e.g. a loamy soil developed in the weathering zone of a basalt (basic igneous rock) is usually rich in nutrients, with Eutric Cambisols being typical under a temperate climate). The methodical options to explore remote sensing in vegetated areas such as central Europe is especially difficult, so that existing reconnaissance data may be an important indicator for the interpretation, and maybe also for the quality of the delineations, at least for some of the new PM classes.

Before comparisons of the new, modelled PM delineations are conducted, some important frame conditions need to be considered. These relate to the content and structure of parent material classifications for soil mapping. With regard to parent material, there are substantial differences between soil mapping and geological mapping. These frame conditions determine that quality of PM identified and presented in maps in both disciplines. These frame conditions are intensively investigated in eSOTER work package 3. There, the definitions and structure of the classification of rocks for use in soil mapping is being studied, and a revised list of parent rock is introduced.

For the comparison conducted here, some frame conditions need to be known:

1. The target resolution is roughly 1:1,000,000. This is also the scale of the global OneGeology mapping project, and the scale identified in the GEO workplan for the world-wide SOTER mapping. For that reason, the national geological maps were selected (Germany: 1;1,000,000, Slovakia and CZ: 1:500,000, Hungary: 1:500,000). At this point of the investigation, the feasibility of this research was applied to the Central European window first. Upon applicability of the method, the procedure will also be applied to the other windows.
2. It can be concluded from the definitions, that the PM of the soil often does not correlate with the parent rock presented in geological maps. This is especially true for land influenced by periglacial processes (deposition of aeolian and glaciofluvial materials, weathering and in-situ mixing, as well as lateral transport via solifluction and erosion). Very often the effects of these processes are not mapped in geology, but represent the material in which soils have developed, and in which plants root.

4.3.3. Database

Table 4.2. presents the geological map data used for this study.

Table 4.2.: List of geological maps used for the Central European Window

country	scale	# mapping units (MU)	Quaternary	characteristics for the purpose of soil PM	area [km^2]	mean [km^2 / MU]
Germany	1:1,000,000	120	no		58,985*)	492
Slovakia	1:500,000	283	no		85,170*)	301
	1:1,000,000	31	yes	quaternary geological map	43,554	1,405
CZ	1:500 000	20	no		76,452	3,823
	1:500 000	48	partly	parent material map	76,463	1,593
Hungary	1:500 000	86	yes		37,810	440

*) Map covers parts of adjacent countries

Figure 4.12. presents an overview about the geological data used.

Annex xy presents the legend of the CZ map. Before these maps could be used as comparison for the new PM classes, the mapping units needed to be re-classified. The re-classification has been conducted separately for each geological map.

The re-classification has aggregated the mapping units according to the following PM classes:

1. consolidated/unconsolidated
2. genesis (peat, rock, alluvial, eluvial, aeolian)
3. texture (gravel, sand, loam, clay, subclasses)

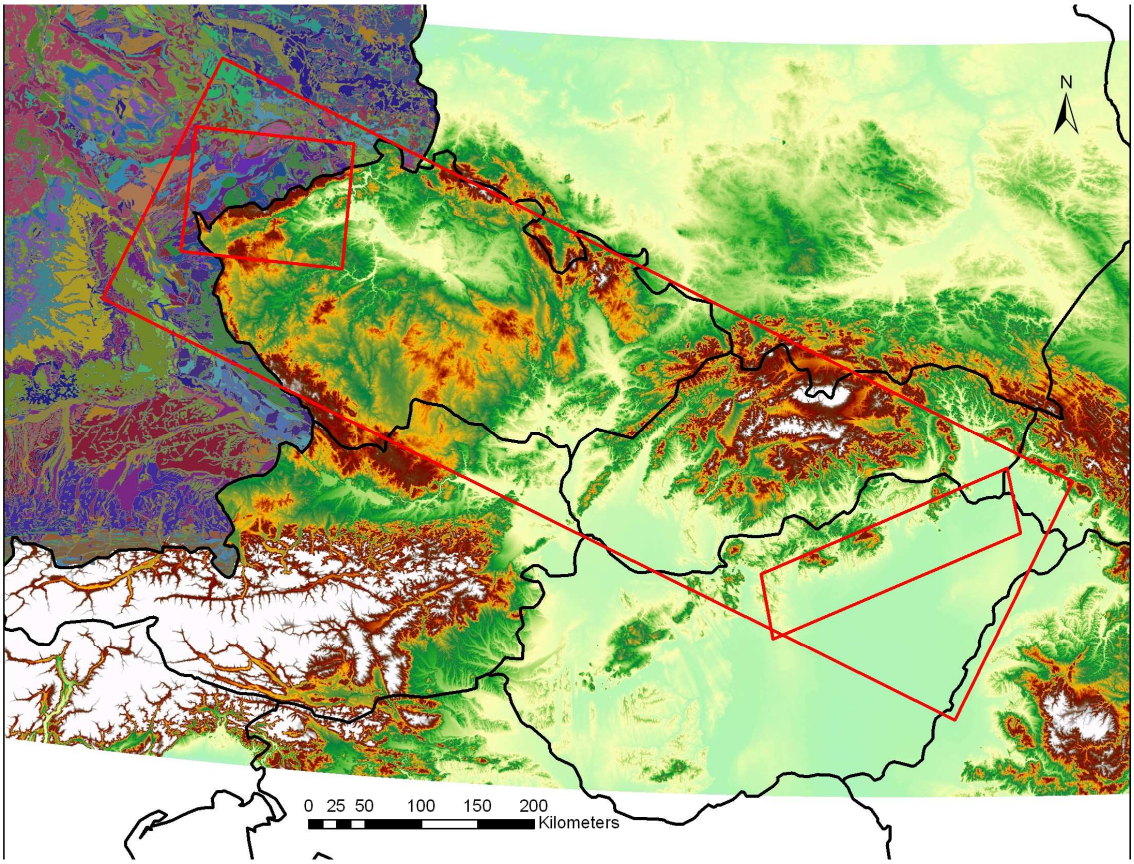


Figure 4.12. Geological maps used for the Central European Window

4.3.4. Results

Figures 4.13. to 4.16. present the results of the three types of new PM compared to the geological maps. The results present a map of the new PM classes from remote sensing (called model result), re-interpreted classes from Quaternary layers in geological maps where available, and the main geological mapping units.

4.3.4.1. Consolidated/Unconsolidated

Figure 4.13. presents the results from the comparison between modelled PM and geological maps. The modelled results are presented in the lower graph. It can be seen that the modelled results correspond quite well to those legacy maps which contain a Quaternary layer (CZ and Slovakia). Considering the geological approach to consolidated material, which does not

consider any weathered residuum on top of the solid rock, the area covered by consolidated parent material (rock) is much larger compared to the modelled PM consolidated. The modelled unconsolidated material includes shallow soils on top of consolidated or bedrock.

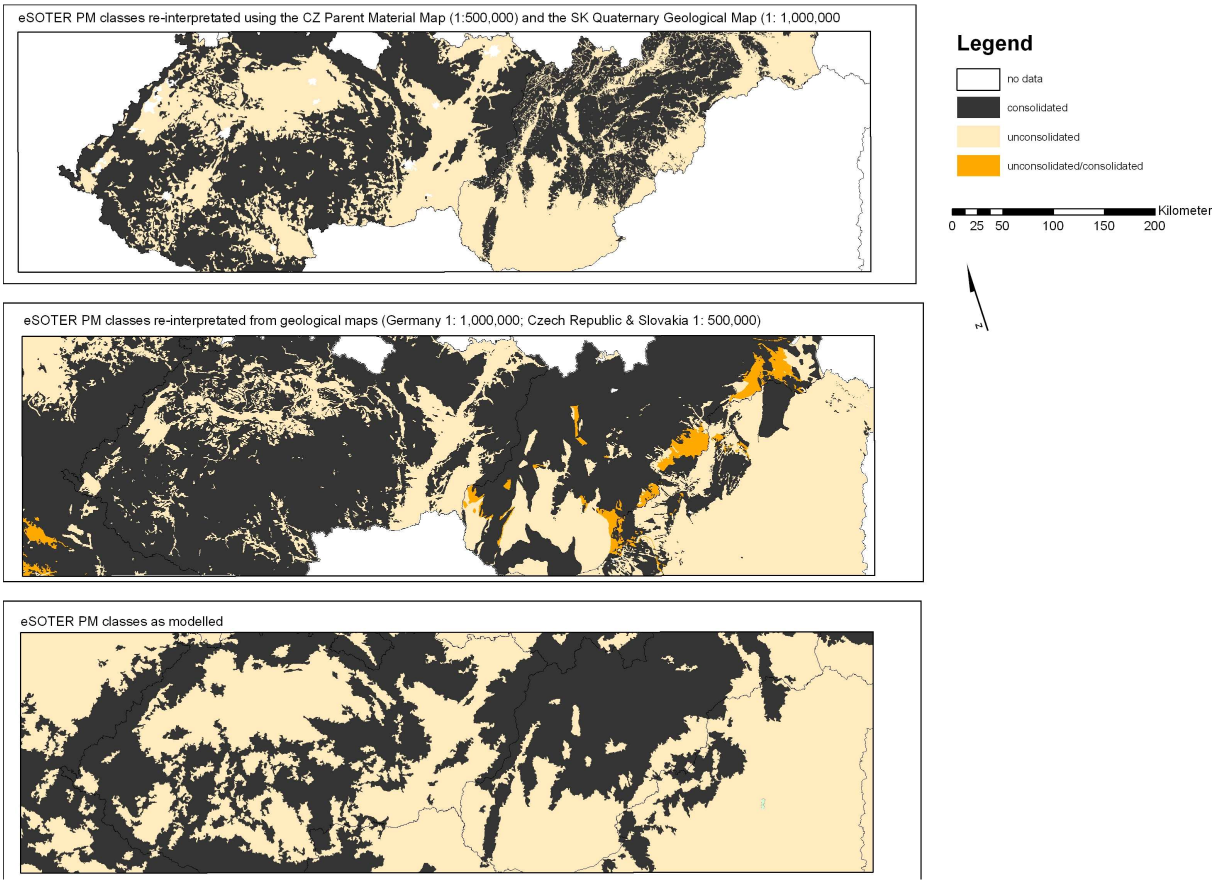


Figure 4.13. Comparison between modelled PM and geological maps: consolidated-unconsolidated

4.3.4.2. Genesis

Figure 4.14. shows substantial differences between the modelled classes and the geological maps. This is partly to be expected because according to the definition above, rock in the geological maps including the Quaternary maps (CZ and Slovakia) seems to correspond to the eluvial/colluvial modelled PM class. Thus eluvial/colluvial fully matches consolidated material in the geological maps, and also corresponds to the modelled consolidated PM class. Eolian has been modelled for the area covered by unconsolidated material, and also some of

the alluvial area, in geological maps. This means that the modelled PM class unconsolidated is either eolian or alluvial.

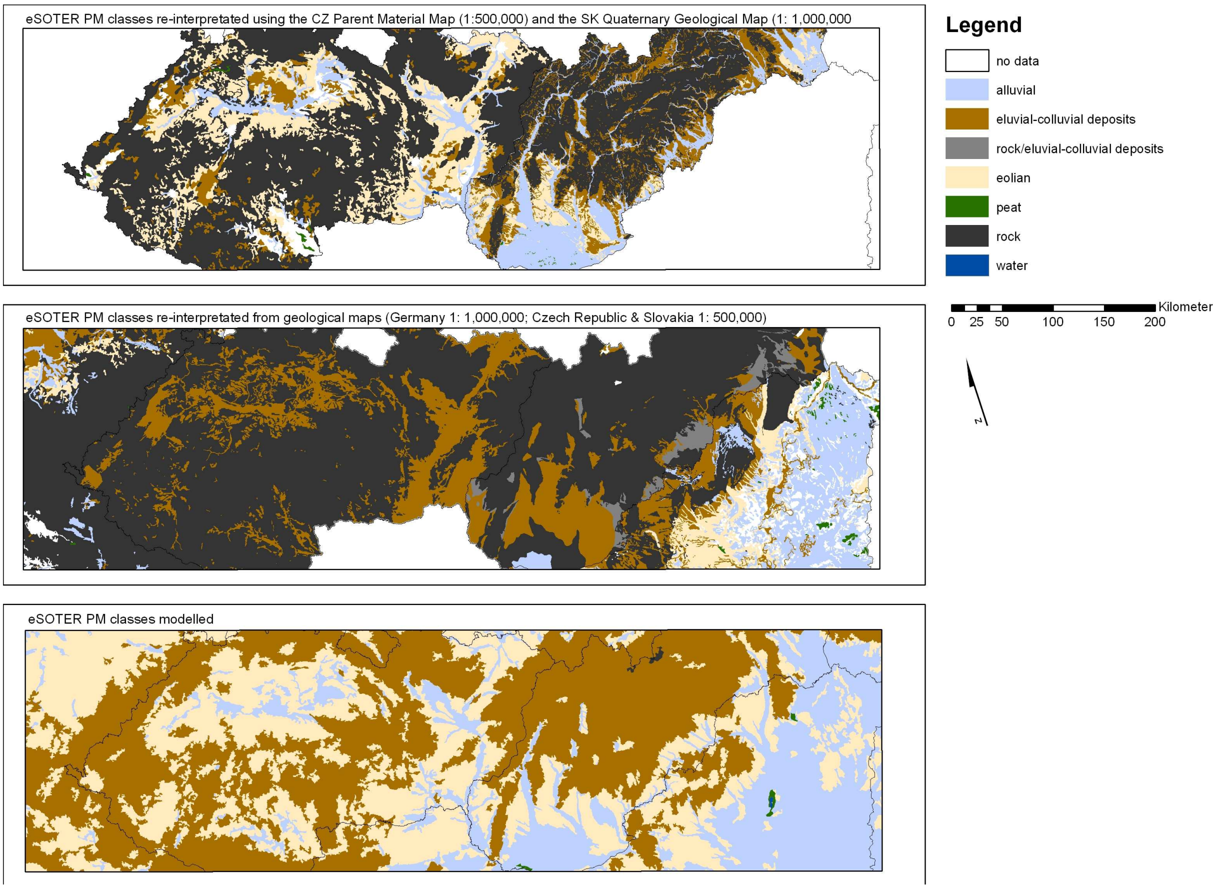


Figure 4.14. Comparison between modelled PM and geological maps: genesis/source of the PM

4.3.4.3. Texture

Texture has been reported in geological maps only for the unconsolidated material. It appears from Figure 4.15., that the area for which texture is available, corresponds to the modelled unconsolidated material, and exceeds the area covered in geological maps. This is to be expected and can be explained by differences in modelling eluvuial/colluvial material in relation to consolidated parent material (rock). The results on texture are futher discussed below.

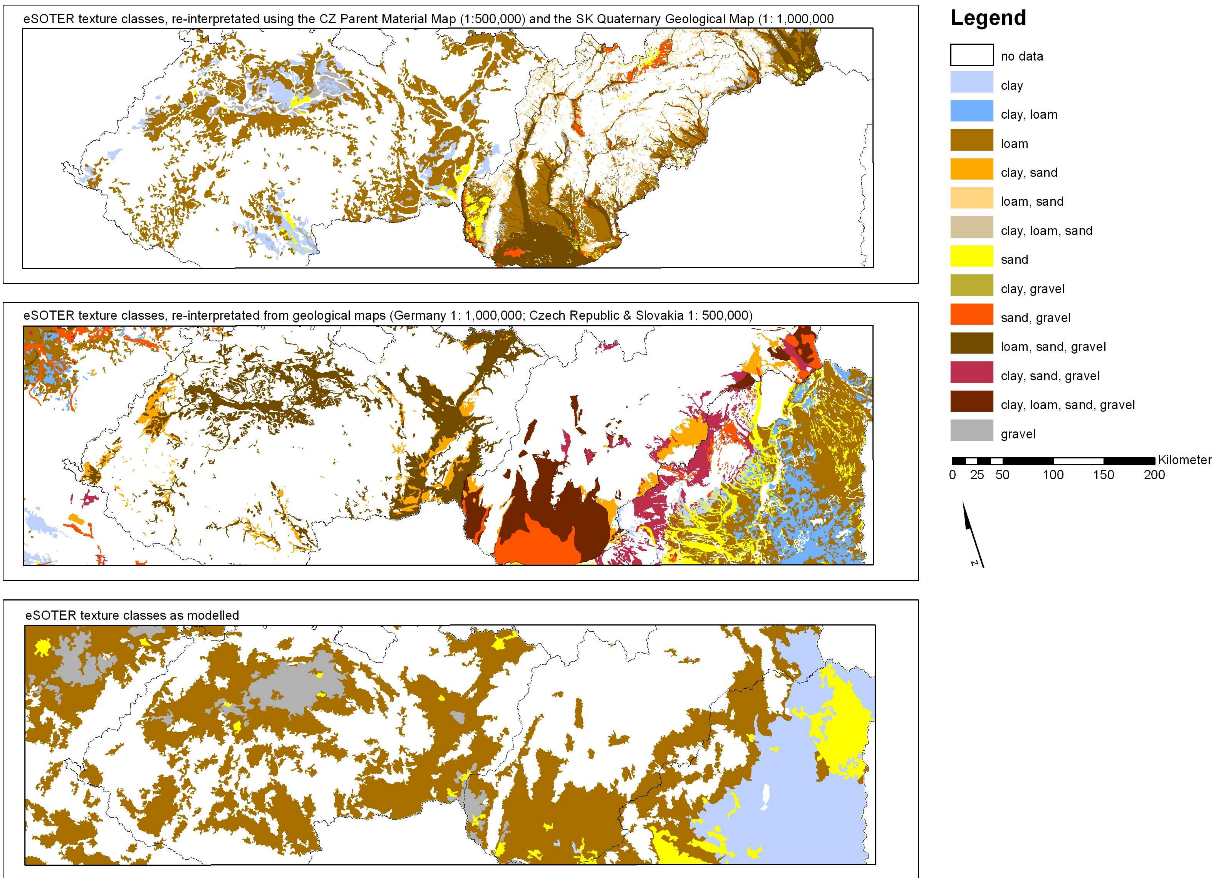


Figure 4.15. Comparison between modelled PM and geological maps: textures

4.3.5. Discussion and conclusions

Figure 4.16. discusses individual mapping units referred to the texture layers. It presents the mapping units of the geological maps, and compares the results of the modelled PM textures.

a) Where the CZ map mostly contains marls and calcareous clays, the modelled PM map contains gravel. The residuals of marls (as well as shales) are usually clay-rich, but could also yield shallow stony soils at exposed positions in the relief.
b) Flucial gravals and sands become loam in the modelled PM map.
c) Unconsolidated loamy material in the geological map becomes sandy in the PM map, and the proportion of clay is largely overestimated in the modelled PM map. Smaller

but important materials such as peat (which is usually intentously slightly overemphasized in order to keep the information in smaller scale maps) are lost in the modelled PM layer.

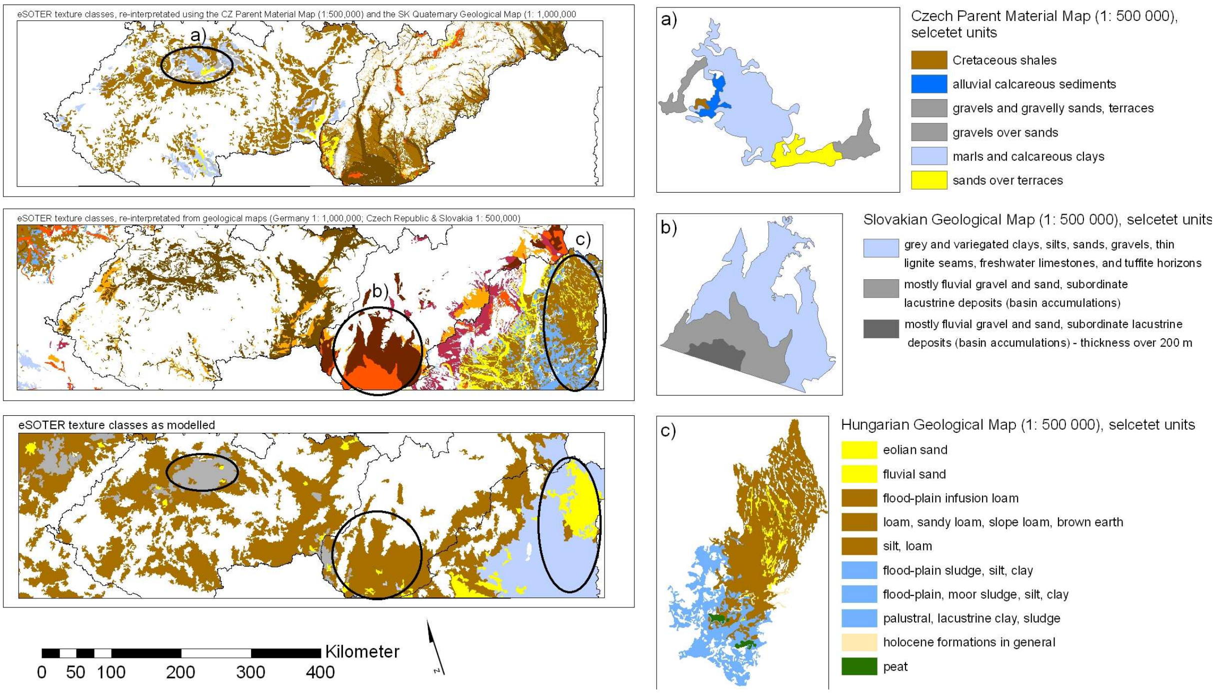


Figure 4.16. Detailed comparison between modelled PM and geological maps: textures

One of the critical aspects in using legacy PM data from re-interpreted geological maps is indeed the complexity of geological mapping units (e.g. in Schist areas). However, when looking at the comparisons on texture, fine textures are rarely mixed with coarse material. This is especially true for the Quaternary layers.

5. Task 1.4: Creation of terrain units by combining landform and soil parent material units

Two parallel procedures have been developed and employed within the Work Package. The first method, a hybrid one, delineates the terrain polygons with a pure digital soil mapping (DSM) based approach, while the parent material information for further dividing the SOTER-units are imported form traditionally made legacy PM data. The second approach simulates the limited/no info case when the whole polygon delineation is DSM-based.

5.1 The hybrid method

The first approach is based on the combination of SRTM derived physiographic units and legacy parent material data. This approach requires a pre-harmonization/correlation procedure for the thematic information. The base data was the 1:1M scale Eurasian soil database. The two polygon systems were combined to delineate the final SOTER-units. The new database had to be cleaned and small polygons to be removed/aggregated to reach the minimum size of the polygons. The aggregation procedure required a new rule system using a semantic distance based similarity test to drive the aggregation procedure and maintain the optimal homogeneity of the resulting polygons.

5.1.1 The aggregation rule set

This program is used to compute semantic distances between zones defined in a polygon coverage. These distances are needed by the programs GENERALIZE_POLYS and ELIMINATE_SIMILAR. When a polygon is below the area threshold argument of these programs, it has to be eliminated, i.e. aggregated to one of its neighbouring polygons. Choosing the neighbouring polygon into which to aggregation is done by selecting the most similar one. The similarity between two polygons is measured by their semantic distance. The distances are user defined: they can be defined manually, or inferred by rules, or computed by formulas.

In the case of the e-SOTER procedure used to generate SOTER Physiographic Units, the semantic distances are computed. The distance between each pair of neighbouring polygons is the Euclidean distance computed from 4 landform parameters which characterise each of the two polygons. The landform parameters are the slope (SLOPE), relief intensity (RI), potential drainage density (PDD), and hypsometry (HYPS). Each polygon is characterised by the normalized, mean value of each of these 4 parameters. The Euclidean distance (Ed) between each pair of polygons is calculated as:

Ed = SQRT(SQR(D1) + SQR(D2) + SQR(D3) + SQR(D4))

where

SQRT = square root

SQR = square

D1 = SLOPEPolyLeft – SLOPEPolyRight

D2 = RIPolyLeft – RIPolyRight

D3 = PDDPolyLeft – PDDPolyRight

D4 = HYPSOPolyLeft – HYPSOPolyRight

and where

SLOPEPolyLeft is the mean normalised slope value of the polygon standing to the left of the arc

SLOPEPolyRight is the mean normalised slope value of the polygon standing to the right of the arc and so forth for each of the four parameters.

The result is stored in the output <similarity table>. If the <similarity table> already exists, the <distance item> is updated for each pair of <key item> values that are listed in the table. If

the <similarity table> does not exist, it is first generated using the program GENERATE_SIMILARITY,
i.e. created and populated with all existing pairs of neighbouring <key item> values. Then the <distance item> is created and calculated.

In the case of the SOTER procedure used to generate SOTER Terrain Components, the semantic distances are computed. The distance between each pair of neighbouring polygon is computed from a set of parameters which characterize each of the two polygons. The parameters are the following: 4 physiographic parameters: a slope class, a relief intensity class, a potential drainage density class, and a hypsometry class; . a 2 level hierarchical SOTER lithology classification system; and a 3 level hierarchical soil classification system Each polygon is characterised by the value of each of these 6 parameters.
The rule applied in a former version of the procedure used to be:
small polygon is merged with the neighbouring polygon which has

- priority 1: the same SOIL at detailed level or
- priority 2: the same SOIL at intermediate level or
- priority 3: the same SOIL at coarse level (as recoded following F. Nachtergaele)

In the present version the rule becomes as suggested hereafter where priority is given first to soil name, then to lithology, and finally to physiography.
The semantic distance between 2 neighbouring polygons given by physiography is inversely proportional to the number of physiographic parameters that are equal:

- if none of the parameters are equal then the physiographic distance = 5
- if 1 of the parameters is equal then the physiographic distance = 4
- if 2 of the parameters are equal then the physiographic distance = 3
- if 3 of the parameters are equal then the physiographic distance = 2
- if all 4 parameters are equal then the physiographic distance = 1

The semantic distance between two neighbouring polygons given by lithology is inversely proportional to the level at which lithology is equal:

- if the lithology is equal at the finest level (LITHO) then the lithologic distance = 1

- else if the lithology is equal at the coarsest level (LITHO1) then the lithologic distance = 2
- else (when lithology is different between the 2 neighbouring polygons) the lithologic distance = 3

The semantic distance between 2 neighbouring polygons given by soil is inversely proportional to the level at which soil is equal:
if the soil is equal at the finest level (SOIL2) then

- the soil distance = 1
- else if the soil is equal at the intermediate level (SOIL1) then the soil distance = 2
- else if the soil is equal at the coarsest level (SOIL0) then the soil distance = 3
- else (when soil is different between the 2 neighbouring polygons) the soil distance = 4

The overall semantic distance between 2 neighbouring polygons is given by combining the physiographic, lithologic and soil distances. In principal, 2 polygons with different soils should not be merged. But this rule is relaxed to giving soil a higher priority (weight) over lithology and physiography (a soil boundary is stronger than a lithologic which in turn is stronger than a physiographic one). This translate to the following formula for combining distances:

combined distance = (soil distance * 100)+ (lithologic distance * 10) + physiographic distance

The result is stored in the output <similarity table>. If the <similarity table> already exists, the <distance item> is updated for each pair of <key item> values that are listed in the table. If the <similarity table> does not exist, it is first generated using the program GENERATE_SIMILARITY, i.e. created and populated with all existing pairs of neighbouring <key item> values. Then the <distance item> is created and calculated. For more information on programs GENERALIZE_POLYS, ELIMINATE_SIMILAR and GENERATE_SIMILARITY, see their respective documentation.

Description of the procedure can be found in the in the Dobos et al. (2005) and in the “e-SOTER–procedure-delivery-2010-04-07/readme” file of the attached aml package.

This approach was a typical case of the traditional SOTER approach, when existing data is translated and loaded into the SOTER framework. However, due to the DSM procedure for the terrain delineation, the final polygon system represents a coherent/consistent way of the terrain unit definition.

5.2 The e-SOTER approach

The second approach simulates the most common case, where data is limited, and covers only a small portion of the area to map. The generation procedure is described in Task1.3. (See fig. 1.3.1.8. for an example for the DSM derived PM layer. The Physiographic and PM units are combined to create homogeneous units. Small size units has to be eliminated via the aggregation procedure, which requires a new rule system using a semantic distance based similarity test to drive the aggregation procedure and maintain the optimal homogeneity of the resulting polygons.
The final polygons are shown in Figures 5.1 – 5.8.

5.2.1 The aggregation rules

In the case of the SOTER procedure used to generate SOTER Terrain Units, the semantic distances are computed. The distance between each pair of neighbouring polygon is computed from a set of parameters which characterize each of the two polygons. The parameters are the following:

- 4 physiographic parameters: a slope class, a relief intensity class, a potential drainage density class, and a hypsometry class;
- and a 2 level hierarchical SOTER lithology classification system.

Each polygon is characterised by the value of each of these 5 parameters.

The semantic distance between 2 neighbouring polygons given by physiography is inversely proportional to the number of physiographic parameters that are equal:

- if none of the parameters are equal then the physiographic distance = 5
- if 1 of the parameters is equal then the physiographic distance = 4

- if 2 of the parameters are equal then the physiographic distance = 3
- if 3 of the parameters are equal then the physiographic distance = 2
- if all 4 parameters are equal then the physiographic distance = 1

The semantic distance between 2 neighbouring polygons given by lithology is inversely proportional to the level at which lithology is equal:

- if the lithology is equal at the finest level (LITHO) then the lithologic distance = 1
- else if the lithology is equal at the coarsest level (LITHO1) then the lithologic distance = 2
- else (when lithology is different between the 2 neighbouring polygons) the lithologic distance = 3

The overall semantic distance between 2 neighbouring polygons is given by combining the physiographic and lithologic distances. In principal, 2 polygons with different lithogies should not be merged. But this rule is relaxed to giving lithology a higher priority (weight) over physiography (a lithogic boundary is stronger than a physiographic one). This translates to the following formula for combining distances:

combined distance = lithologic distance * 10 + physiographic distance

The result is stored in the output <similarity table>. If the <similarity table> already exists, the <distance item> is updated for each pair of <key item> values that are listed in the table. If the <similarity table> does not exist, it is first generated using the program GENERATE_SIMILARITY, i.e. created and populated with all existing pairs of neighbouring <key item> values. Then the <distance item> is created and calculated.

For more information on programs GENERALIZE_POLYS, ELIMINATE_SIMILAR and GENERATE_SIMILARITY, see their respective documentation.

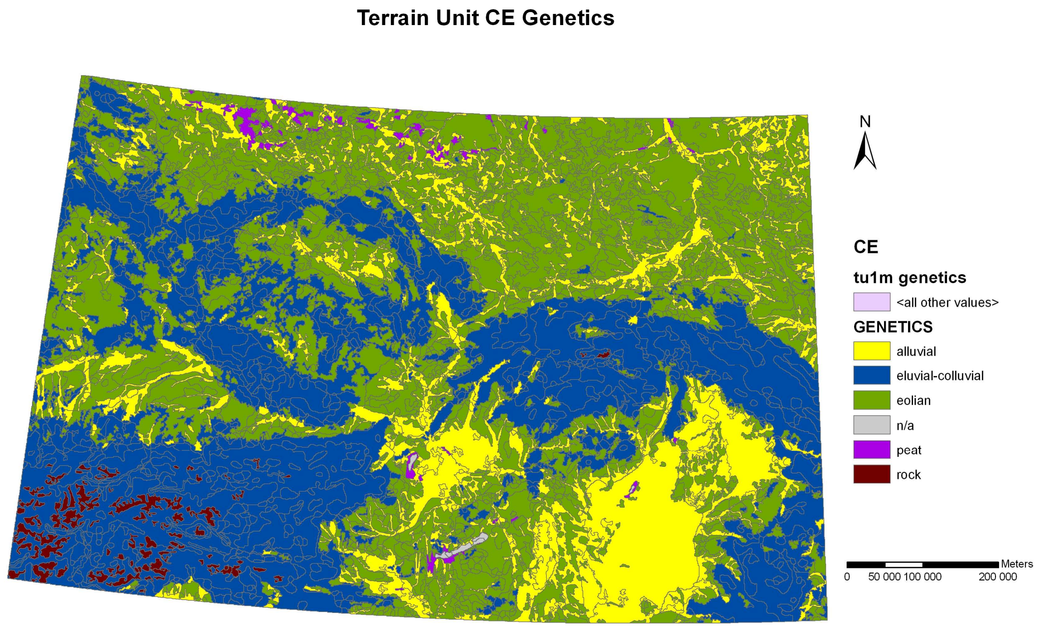


Fig 5.1. The genetic parent material classes of the Central European window.

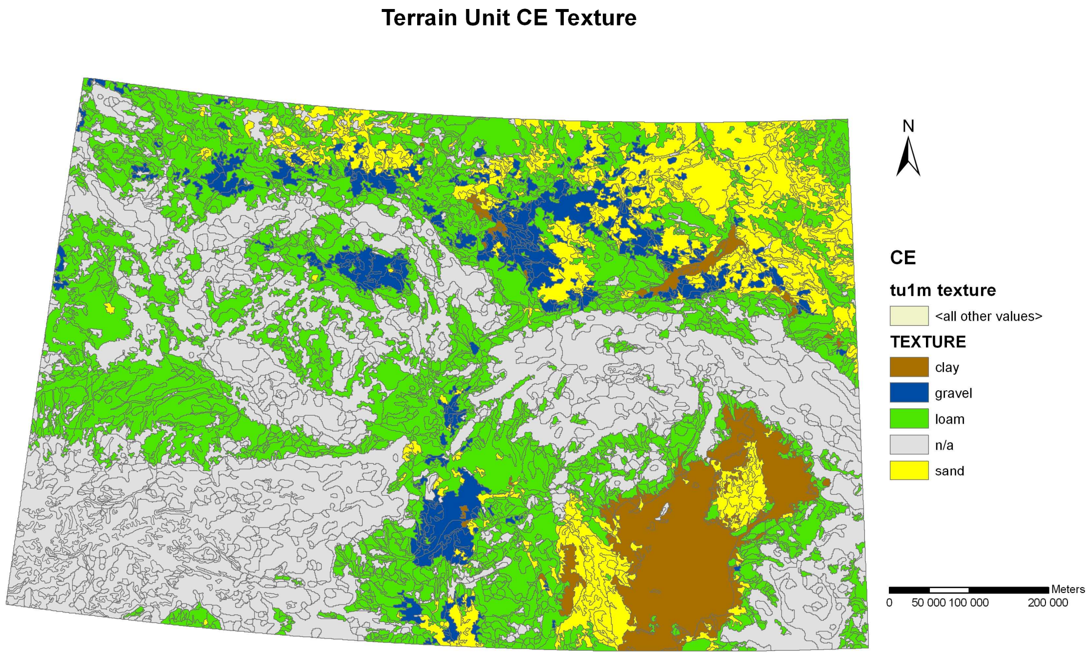


Fig 5.2. The texture classes of the Central European window.

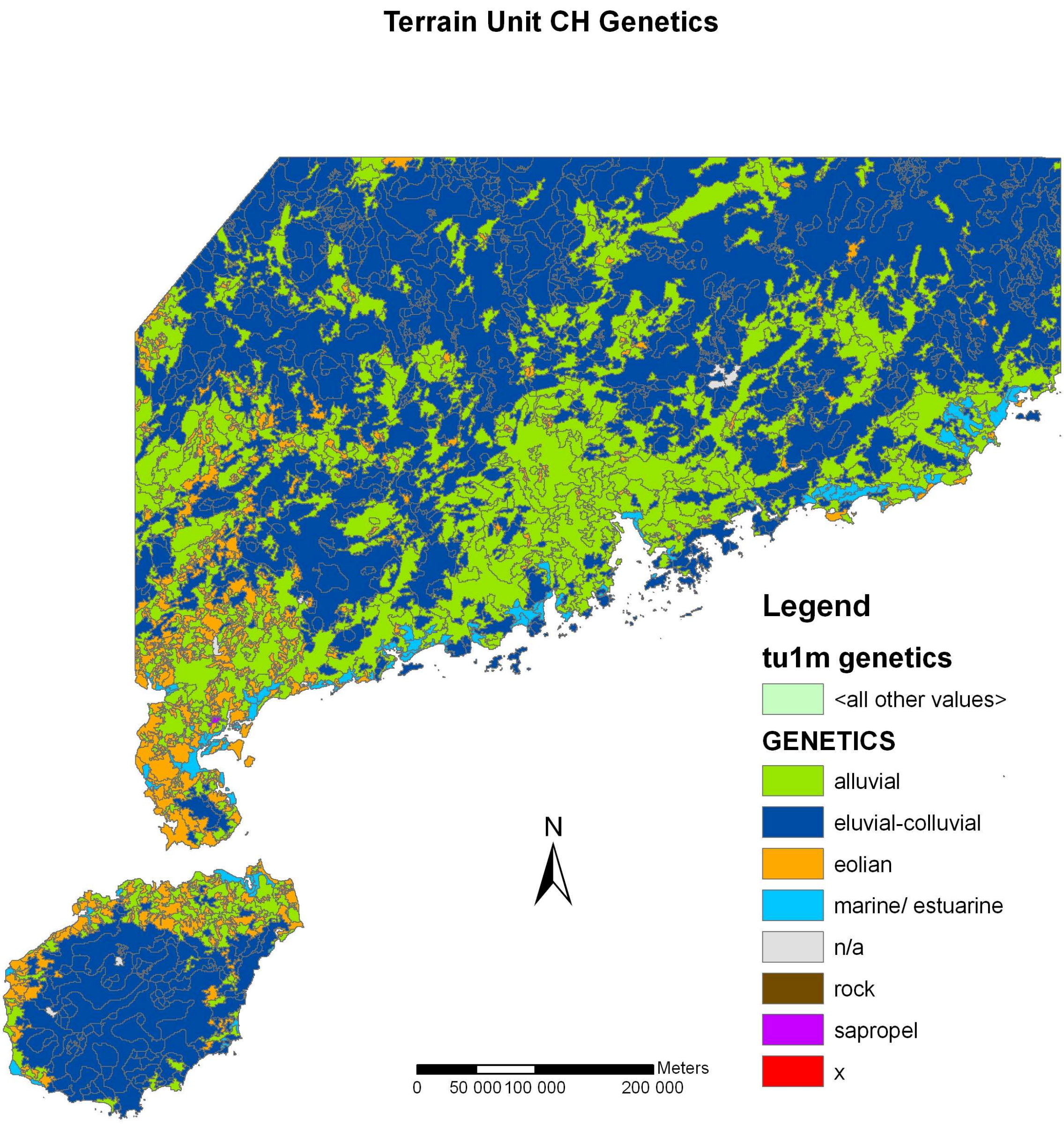


Fig 5.3. The genetic parent material classes of the China window.

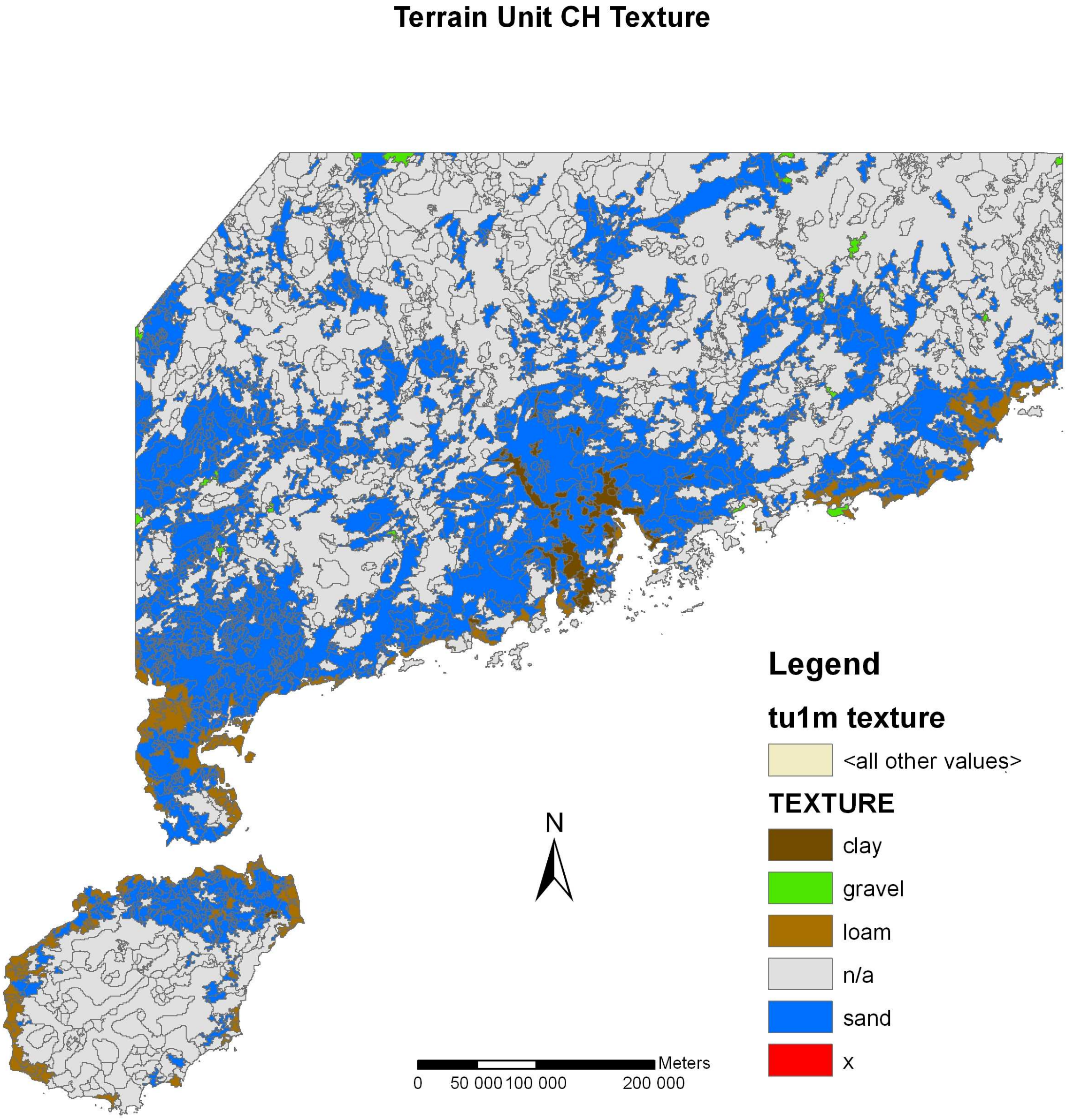


Fig 5.4. The texture classes of the China window.

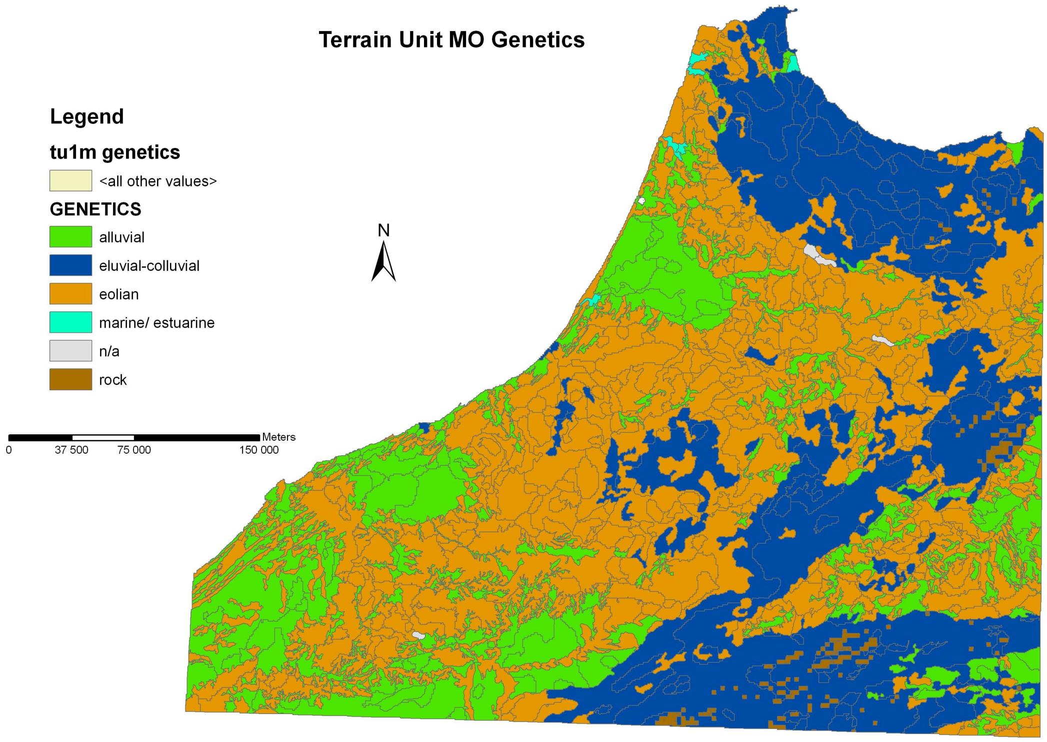


Fig 5.5. The genetic parent material classes of the Morocco window

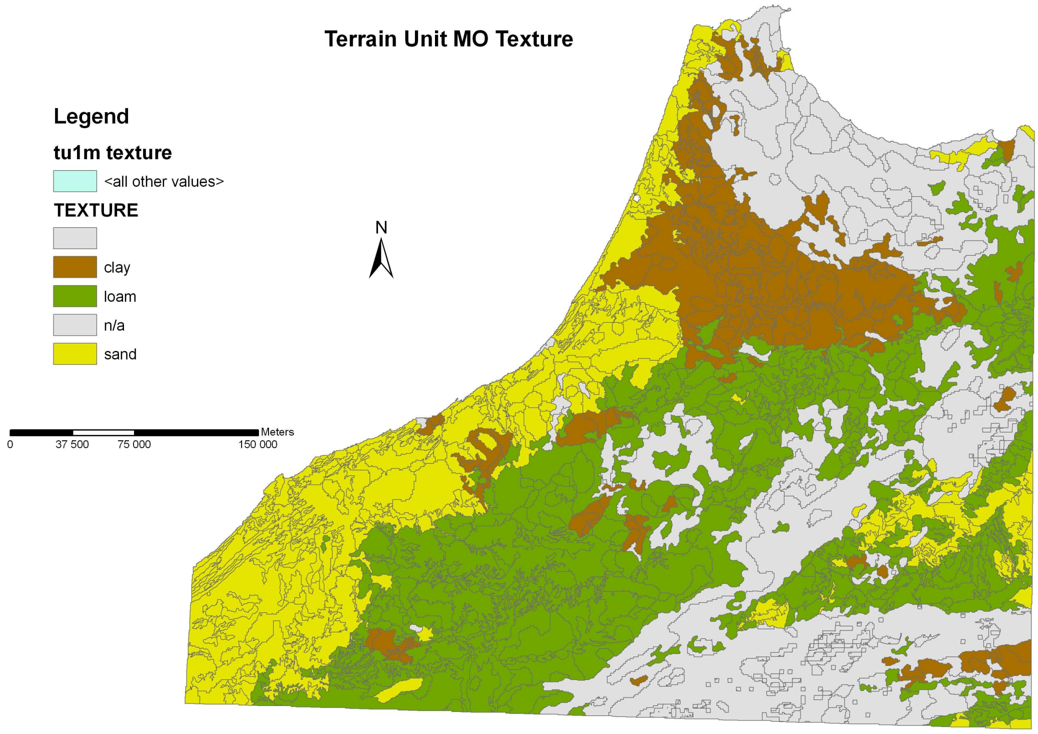


Fig 5.6. The texture classes of the Morocco window

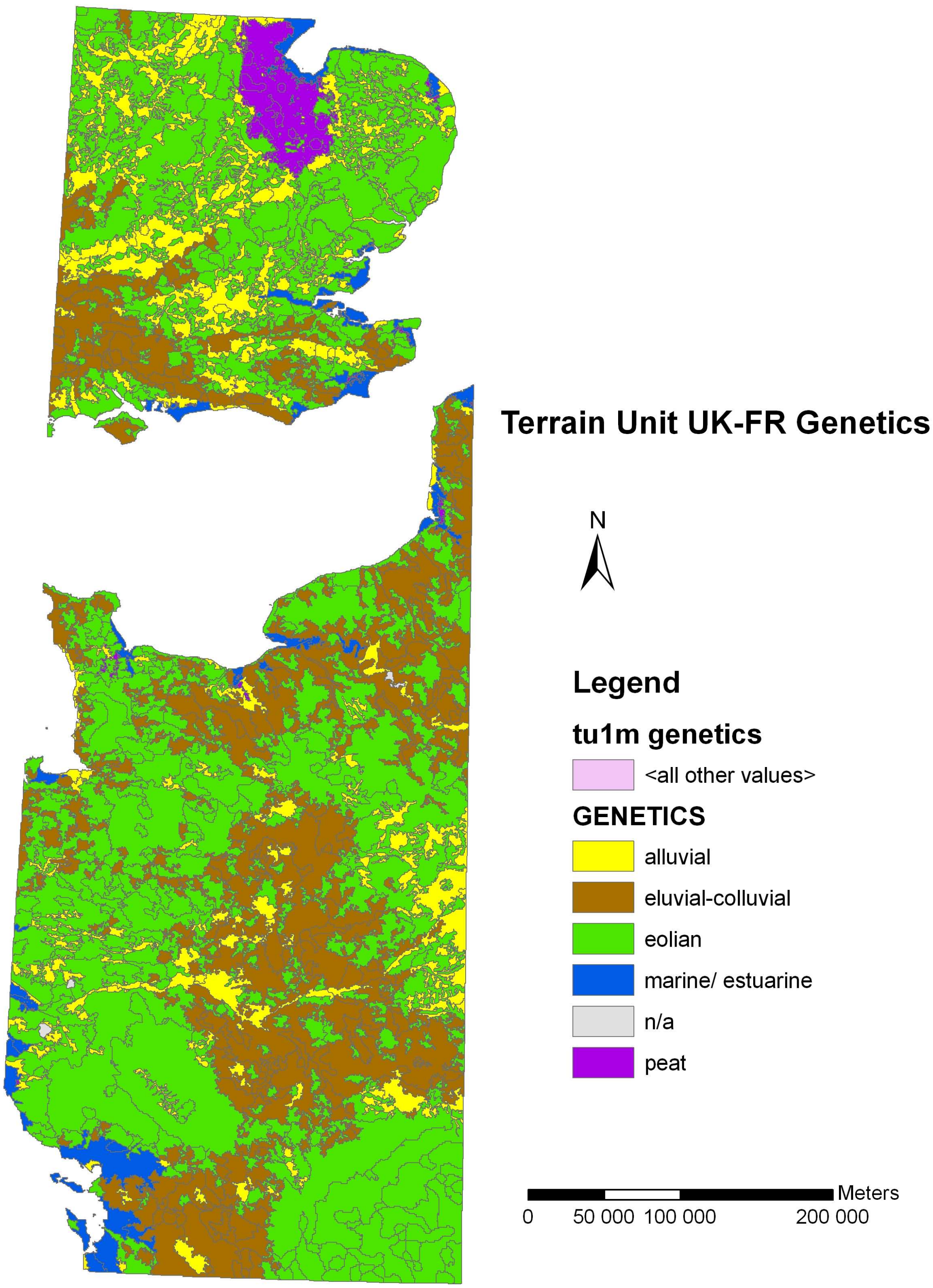


Fig 5.7. The genetic parent material classes of the Morocco window

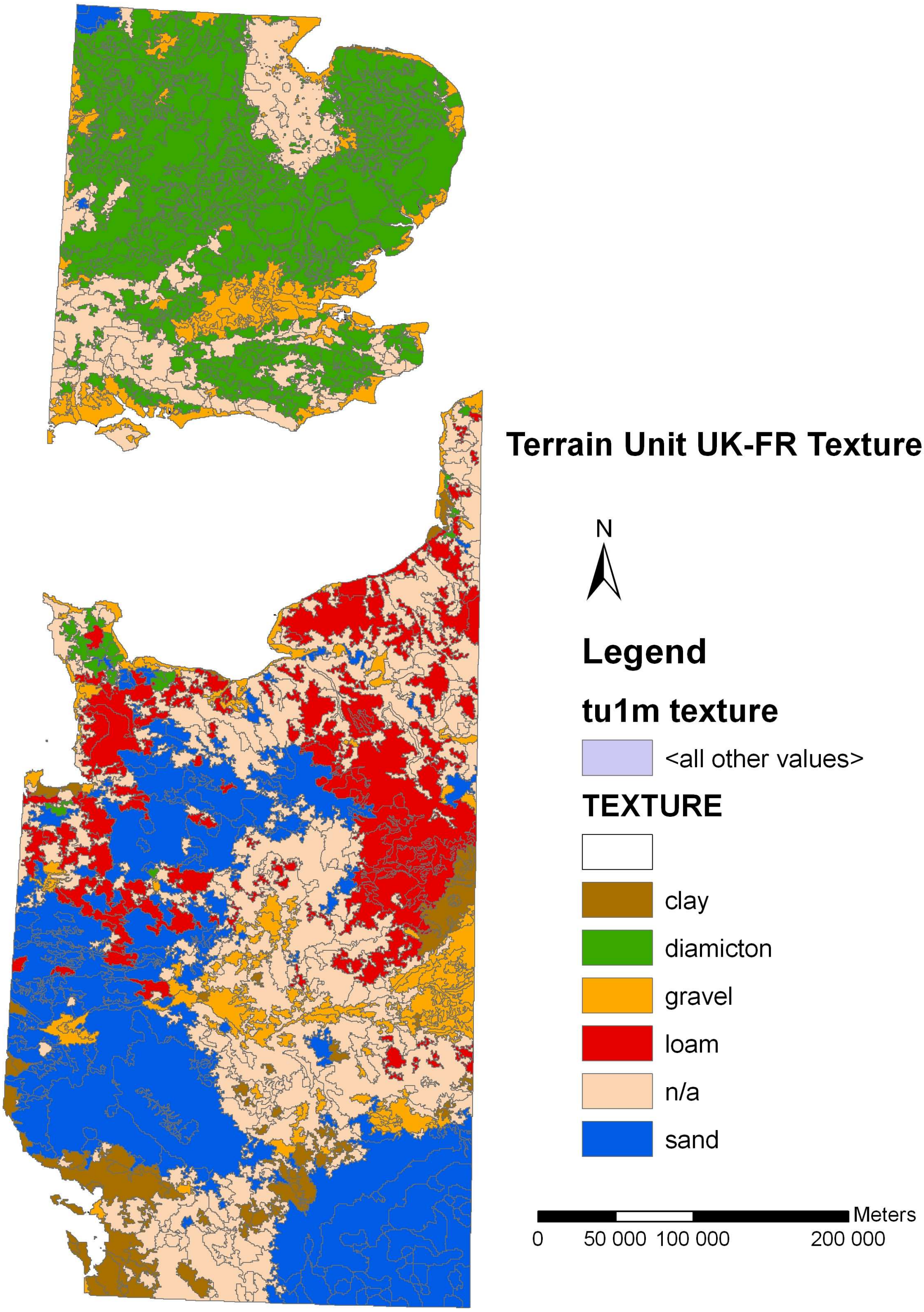


Fig 5.8. The texture classes of the UK/Fr window

6. Dissemination and Exploitation

The research of this WP has lead to the following publications:

1. Dobos,E., Micheli E., Kobza J. 2008. Transnational Soil Data Harmonization Database Development. 2008 Joint Annual meeting of GSA, SSSA, ASA, CSSA, GCAGS, HCS. October 5-9. Houston, USA

2. Dobos. E. Bialkó, T. Micheli. E., Kobza. J..2008. Legacy soil data harmonization and database development. 3rd Global Workshop on Digital Soil Mapping (DSM 2008 USA), September 30 – October 3. Utah State University,Logan, USA

3. Dobos E., Micheli, E., és Bialkó T. 2008. Határmenti talajadat harmonizació és adatbázis fejlesztés digitális talajtérképezési módszertan segítségével. IV. Magyar Földrajzi Konferencia. Debrecen. November 14-15.

4. Dobos E. 1, Micheli E. 2, Bialkó T. 3, Kobza J., 2009. Magyar-Szlovák határmenti talajadatharmonizációs módszertani fejlesztés és adatbázis építés. V. Kárpát-medencei Környezettudományi Konferencia. 2009. Március 26–28, Kolozsvár, Románia

5. Dobos E. 2009. Kis méretarányú talajtani adatbázisok digitalis fejlesztési lehetőségei. Erdélyi Magyar Tudományos Társaság. Április 2-5. Máramarossziget, Romania

6. Dobos E, Micheli E Kobza J. Legacy data integration into modern databases. 2009 National Cooperative Soil Survey National Conference Las Cruces.New Mexico, Majus 9-16.

7. Dobos E., Micheli E. 2009. Terrain Characterization in Support of the SOTER database development. International Annual Meeting of the Soil Science Society of America. November 1-5. Pittshburh. USA.

8. Dobos E., 2009. The WRB qualifiers as melting pots for digital soil mapping. Bridging the Centuries, 1909-2009, 16-17 September, Budapest,

9. Dobos, E., Bock, M., Köthe, R., Daroussin, J., Seres, A. and van Engelen, V.W.P. 2009. Landform mapping for SOTER at scale 1:1 million using SRTM-DEM. SOTER unit delineation using an SRTM-based procedure. Annual meeting of the Soil Science Society of East Africa. December 6-10. Moshi, Tanzania.

10. Seres A., Dobos E. 2009. Területhasználati térkép készítése műholdfelvételek alapján az SRTM magasságmodell pontosítására, HunDEM2009 Konferencia és kerekasztal Kiadványa, 2009. április 24; elfogadva, megjelenés alatt, 15p. (in press)

11. Seres A., Dobos E. 2009. Erdőborítottsági térkép készítése LANDSAT felvételek alapján az SRTM DEM pontosítása céljából, V. Kárpát-medencei környezettudományi konferencia, 2009. március 26–29., Kolozsvár, ISSN 1842-9815

12. Dobos E., Seres A. 2010. Digitális térképezési és távérzékeléses alapokon nyugvó térinformatikai eljárások kidolgozása egységes, a Föld egészére kiterjedő, talajtani adatbázisok fejlesztéséhez, in Jubileumi szakmai tudományos konferencia a Műszaki Földtudományi Kar megalakulásának 50. évfordulójára (in press)

7. References

Dobos E, 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, Institute for Environment and Sustainability, Joint Research Centre, Ispra

ANNEX I

MODIS satellite images data sheet

- DOWNLOADING satellite data from MODIS server (e4ftl01u.ecs.nasa.gov).
    - Downloaded composites:
        - MOD09A1.005
        - MOD11A2.005
    - Downloaded tiles:
        - CE window:
            - h18v03
            - h18v04
            - h19v03
            - h19v04
        - UK/FR window:
            - h17v03
            - h17v04
            - h18v03
            - h18v04
        - MO window:
            - h17v05
        - CH window:
            - h28v06
            - h28v07

- Downloaded dates. The downloaded dates should represent the vegetative period, changing environmental conditions (like soil wetness, temperature) during the year, cloud and snow-free images from every second month are selected
    - CE window:
        - 2008.02.02 – 2002.02.18
        - 2009.04.15
        - 2008.06.25
        - 2008.08.28

        - 2006.10.16
    - UK/FR window:
        - 2008.02.10
        - 2004.04.22
        - 2006.06.02
        - 2003.08.05
        - 2007.10.16
    - MO window:
        - 2002.02.02
        - 2008.04.30
        - 2008.06.09
        - 2007.08.21
        - 2007.10.16
        - 2000.12.10
    - CH window:
        - 2008.02.26
        - 2002.10.08
        - 2008.12.02

- Images in hdf, hdf.xml format
- IMPORTING images from hdf to img format with ERDAS Imagine
- LAYER SELECTION:
    - From MOD09A1: Band 1-7 (Layers 1-7)
    - From MOD11A2: Band 31-32 (Layers 9-10), LST (Land Surface Temperature) Day (Layer 1) and LST Night (Layer 5)
- LAYER STACK: The above mentioned layers were stacked with ERDAS Imagine with the resolution and the output type of the finer resolution image (MOD09A1). The result is a 11 layer, 500 meter image.
    - Layer 1: Band 1
    - Layer 2: Band 2
    - Layer 3: Band 3
    - Layer 4: Band 4

  - Layer 5: Band 5
  - Layer 6: Band 6
  - Layer 7: Band 7
  - Layer 8: LST Day
  - Layer 9: LST Night
  - Layer 10: Band 31
  - Layer 11: Band 32
- MOSAICING the four tiles for each date, and layer stacking all the channels of all dates, which results a usually 33-55-66 layer image depending on the number of cloud-free and snow-free dates.
- PCA: Principal component analysis (PCA) was run on the images to reduce the number of layers and the first 15 channels were kept.
- LST: A normalized temperature fluctuation layer was created for each dates using the following function: globmax(lstday)/lstday*(LST Day-LST Night)
- BAND RATIOS FROM THE LITERATURE (ORIGINALLY FOR LANDSAT):
  - 6/1
  - 1/3
  - 7/6

ANNEX II

Terrain parameters derived from SRTM DEM

Terrain parameters derived from STRM DEM

Elevation (filled srtm z limit: 20) :

1. Sinks were identified and filled up to 20 meters
2. "0" values were set to 0,0000001 to keep the 0 for representing NoData (it is important for the ERDAS Imagine and MultiSpec processing)

[con ([fele_uk_utm31] == 0, 0.000001, [fele_uk_utm31])]

3. export img, cellsize : 482,2727949 (to match the resolution of the MODIS image)

Slope percent:

1. ArcInfo slope function was used. "0" values were increased to 0.000001 to avoid any error due to dividing by 0 later in the procedure, or to keep the value "0" for representing the background
2. [slp_uk_utm31] + 0.000001
3. export img, cellsize : 482,2727949

NoData as 0

Relief (circle: radius 5)

1. Focalmax – Focalmin for a circle of 5 cells radius
2. "0" values were increased to 0.000001 to avoid any error due to dividing by 0 later in the procedure, or to keep the value "0" for representing the background

[relc5_uk_utm] + 0.000001

3. export img, cellsize : 482,2727949

NoData as 0

PDD:

1. Elevation grid to be resampled to 482,2727949 (MODIS resolution)

2. Creation of a mask for clipping the original coastline later in the procedure:

   Reclassification of the Elevation grid: Land area gets NoData while the sea pixels with NoData values get a value of 1. Raster to polygon transformation to create the sea polygon. (buffer zone is set 450m)

3. Clip resampled elevation by buffered water mask polygon, Clipping geometry = coastline

4. Add new field (“code”) for the Channel line (100 m threshold) layer and assign a value of “1” for all lines

5. Feature to raster : base field is “code”, cellsize: 482,2727949

6. Mosaic channel line raster & coastline raster

7. Focal sum: (circle: radius 15)

Groundwater level

1. Create a channel line system from the SRTM using a threshold of 50 pixels.

2. “Extraction by mask” the pixels of the Elevation grid by the channel line pixels.

3. Definition of the coastline

   Reclassification of the Elevation grid: Land area gets Nodata while the sea pixels with 0 values get a value of 1. Raster to polygon transformation to create the sea polygon (buffer zone is set to 90 m).

4. Clip(management) filled srtm, output extent a bufferelt tenger polygon, clipping geometry

5. Mosaic channel line masked SRTM raster (point 2.) & coastline raster (point 4.)
6. Raster to point transformation
7. Creation of the groundwater level using the interpolation to raster, natural neighbors function.
8. Extract the groundwater level grid from the Elevation to create the depth to groundwater level grid
9. “0” values were increased to 0.000001 to avoid any error due to dividing by 0 later in the procedure, or to keep the value “0” for representing the background

   con ([gwl_uk_utm31] < 0.000001, 0.000001, [gwl_uk_utm31])
10. Export img, cellsize : 482,2727949
11. NoData as 0

Topographic Wetness index (TWI)

1. ln(flow accumulation +0.000001 / slope percent +0.000001)
2. “0” values were increased to 0.000001 to avoid any error due to dividing by 0 later in the procedure, or to keep the value “0” for representing the background

con ([twi_uk_utm31] == 0, 0.000001, [twi_uk_utm31])

3. export img, cellsize : 482,2727949

NoData as 0

Upland/Lowland

1. [Focalmean (elevation)] (resolution 482,2727949m, radius 10 cells for the focalmean circle)

2. (Elevation) – (Focalmean)

Positive values represent lowland, while the negative ones are upland.

3. “0” values were increased to 0.000001 to avoid any error due to dividing by 0 later in the procedure, or to keep the value “0” for representing the background

con ([uplow_fmean-elev_482_uk_utm31] == 0, 0.000001, [uplow_fmean-elev_482_uk_utm31])

NoData as 0

Marine and Estuarine sediments

1. Reclassification of (Elevation) into three classes:

   below 0, 0 - 5, and above 5 meters

2. Focal majority (Circle range: 5.)
3. Con (isnull(focal majority), (reclassed srtm), (focal majority))
4. Reclassification of the new grid into the value “1” for the cells having 0 to 5 meter elevation, and value “NoData” for the rest of the cells.
5. Extract by mask with the Elevation to reset the coastline after the majority function caused shift.
6. Raster to polygon to create polygons with the desired elevation range (0-5)
7. Creation of sea polygon

   Reclassification of the Elevation grid: Land area gets Nodata while the sea pixels with 0 values get a value of 1. Raster to polygon transformation to create the sea polygon.

8. Selection of the 0-5 meter elevated areas along the coastline:

   a. select by location: with the “are within a distance of…sea polygon” function, buffer zone set to 10000 m).

9. "Create layer from selected features" to develop the final Marine /Estuarine layers

ANNEX III

Final layer stack, used for the classification

Layers:

- 1-15 MODIS PCA
- 16 elevation
- 17 slope
- 18 relief
- 19 TWI
- 20 groundwater level
- 21 PDD
- 22 upland-lowland
- 23 February LST
- 24 April LST
- 25 June LST
- 26 August LST
- 27 October LST
- 28 February band 6/1
- 29 February band 1/3
- 30 February band 7/6
- 31 April band 6/1
- 32 April band 1/3
- 33 April band 7/6
- 34 June band 6/1
- 35 June band 1/3
- 36 June band 7/6
- 37 August band 6/1
- 38 August band 1/3
- 39 August band 7/6
- 40 October band 6/1
- 41 October band 1/3
- 42 October band 7/6

ANNEX IV

Landsat Data Processing

A “two-stepped” classification was done in areas, where the input parent material information is very small scale or not thematic. Masking the better resolution LANDSAT images, so that only the bare ground areas are shown and classifying these based on the small-scale parent material data, gives the small scale maps more details and with this detailed dataset, the coarser resolution MODIS image can be classified later.

- DOWNLOADING the LANDSAT images from http://glovis.usgs.gov/ . Sample tiles representing all important major geologic features were selected for the windows. The images need to show the most bare ground with the least vegetation, so they should originate from dates when the vegetation is still very weak, small in the spring or has already dried out or had been harvested and there is no snow.
- Downloaded tiles and dates:

  - CE window:
  - tile: 187-026, date: 2000.10.23.
  - tile: 187-027, date: 2000.10.23.
  - tile: 189-026, date: 2000.10.21.
  - tile: 189-027, date: 2000.10.21.

  - UK window:
  - tile: 201-023, date: 2002.04.06.
  - tile: 202-024, date: 2002.03.28.

  - MO window:
  - tile: 200-036, date: 2008.08.15.
  - tile: 201-036, date: 2002.08.12.
  - tile: 201-037, date: 2002.08.12.

- CH window:
- tile: 120-043, date: 2001.12.24.
- tile: 120-044, date: 2001.12.24.
- tile: 122-044, date: 2001.11.20., 2002.01.07.
- tile: 123-043, date: 2001.12.29.
- tile: 123-044, date: 2001.12.29.
- tile: 124-045, date: 2002.11.05.

- IMPORTING the TIF files to IMG with ERDAS Imagine.
- REMOVING band footprint offset stripes (colored stripes on the eastern and western edges of LANDSAT images, which result from the footprints (location and spatial extent) of each band are not being exactly the same) with a model downloaded from:

  http://arsc.arid.arizona.edu/resources/image_processing/landsat/landsat.html

- BAND RATIOS: some RGB band ratio composites show the different geological / parent material features better, then the any of the original band composites. The band ratios used were:
  - http://www.narss.sci.eg/uploads/Journal/06-12-2009.pdf
    - band 5 / band 3
    - band 3 / band 1
    - band 7 / band 5
  - Vit penizek, JRC
    - band 3 / band 2
    - band 3 / band 7
    - band 5 / band 7
- LAYER STACKING the bands: first the original bands, then the band ratios:
  - For all windows:
    - 1: band 1
    - 2: band 2
    - 3: band 3

  - 4: band 4
  - 5: band 5
  - 6: band 61
  - 7: band 62
  - 8: band 7
  - 9: band ratio 3/1
  - 10: band ratio 3/2
  - 11: band ratio 3/7
  - 12: band ratio 5/3
  - 13: band ratio 5/7
  - 14: band ratio 7/5

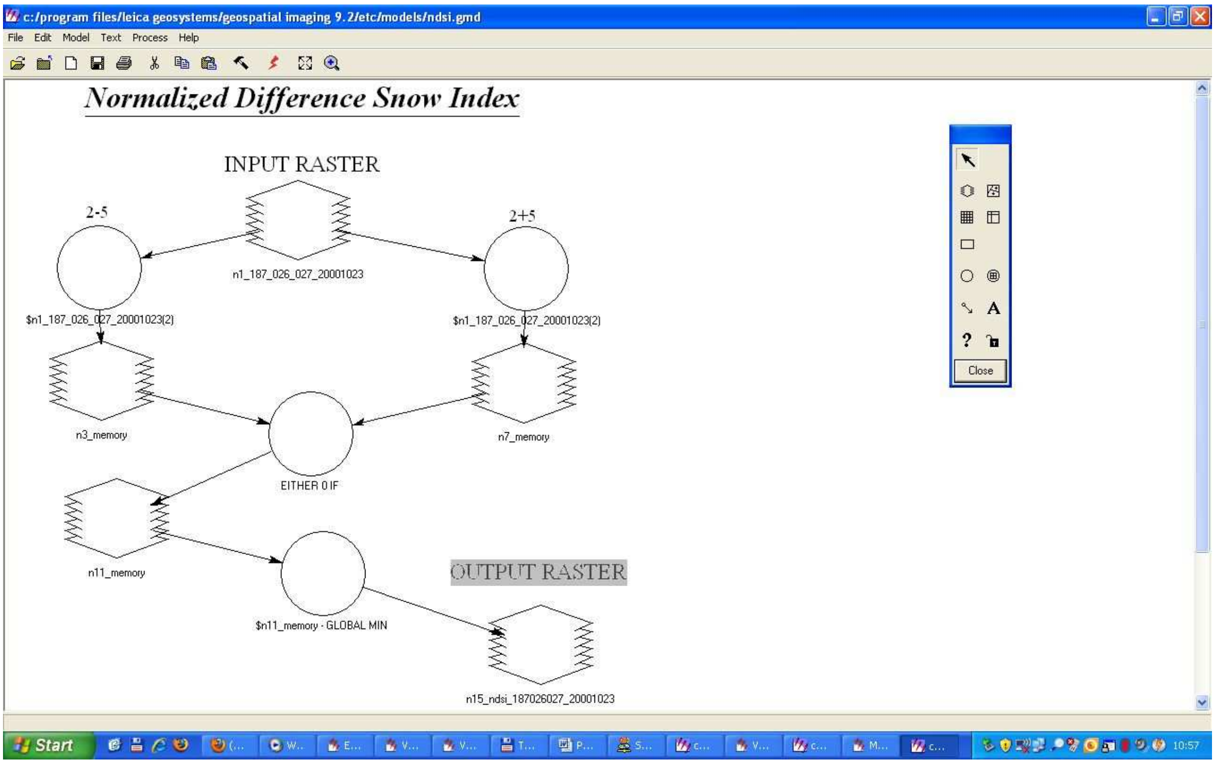
- MOSAICING the images using ERDAS Imagine mosaic wizard.
- MASKING: Areas which were covered with vegetation, snow or water were masked out and only the bare ground places were left. This LANDSAT classification procedure was not applied to China.
- MASKING OF THE VEGETATION was done by the following way:
  NDVI (Normalized Difference Vegetation Index) was calculated for the images (ERDAS - Interpreter – Spectral Enhancement – Indices and choosing NDVI on LANDSAT TM here and stretching it to unsigned 8 integer). By studying, comparing the NDVI, the original satellite image and a land cover map, a pixel value could be found, that could be used as a threshold, above which the pixels indicate the vegetated areas (sprouting or on the autumn image fallen leaved deciduous forests, coniferous forests, young wheat, corn and other crop seedlings). So pixel values above this threshold needed to be masked out of the image.

  NDVI threshold values were:

  - CE window: 150
  - UK window: 150
  - MO window: 130
- MASKING THE SNOW AND WATER:

NDSI (Normalized Difference Snow Index) was calculated for the images by the Model Maker function of ERDAS similar to the NDVI. http://www.gis.unbc.ca/projects/illpage/illpageone.html

[(2-5)/(2+5) – GlobMIN ((2-5)/(2+5)) / GlobMAX ((2-5)/(2+5)) – GlobMIN ((2-5)/(2+5))]*255



By studying the original LANDSAT image and the NDSI image, a threshold could again be found above which pixel values indicate snow or water, so these areas were again masked out of the already NDVI masked image.

NDSI threshold values:

- CE window: 185
- UK window: 150
- MO window: 185

- MASKING THE CLOUD COVER:

  Masking the cloud cover was done based on the previous method. The bands where the clouds have a high reflectance are band 1 and band 5 and the bands where they have the least reflectance are the thermal bands, so band 61 and band 62. From this the cloud mask was calculated the following way:

  (band1+band:5-band6-band7)/(band1+band5+band6+band7). this was then normalized, stretched to a 0-255 scale by the same function as in the previous calculations: (raster-GlobanMIN)/(GlobanMAX-GlobanMIN)*255.

  Pixels above the threshold value were masked out of the already NDVI, NDSI masked images.

  Cloud threshold values:

  - CE window: 110
  - UK window: no clouds on the image
  - MO window: 120

- TRANSLATING the geological map to the e-SOTER unconsolidated texture types in case of existing polygonal parent material data. After the translation, this parent material database and the masked LANDSAT image were masked with each other, so that only the not NoData cells are classified.

- CLASSIFICATION of the LANDSAT map using the polygonal parent material database as training data with the MultiSpec software in case of CE and UK windows. For windows with no existing thematic parent material database, such as Morocco, the classification was done manually, by drawing training areas based on the scanned parent material map, the 5/3, 3/1, 7/5 RGB composite of the masked LANDSAT image and the SRTM DEM.

- CLASSIFICATION of the MODIS image:

The classified LANDSAT image then was used as training data for the MODIS image in the final classification with the MultiSpec software.

At the end it was decided that classification based on LANDSAT images are only used in the windows where the available parent material information is only very small scale or not thematic. Like in case of Morocco, where the available data was only a 1:1 million scanned geological map.

- SOFTWARE USED:

ArcMap 9.3

ERDAS Imagine 9.2

MultiSpec 3.1

ANNEX V

The finalization procedure for the PM polygon coverage

Process:

1. RESAMPLE: Resampling the satellite image classified layers to 90m, which is the resolution of all the other SRTM derived data
2. FOCAL MAJORITY: The original rasters are usually full of scattered pixels, which have to be eliminated for the vectorizing process. Hence., a Focal majority function with a 3 cell radius circle moving window is applied to the images. (ArcMap focal statistics:circle neighborhood, 3 cell radius, statistics type majority, ignore NoData checked) In case of rock, water, peat and sapropel, which cannot be dissolved into other polygons later, the radius was set to 5 cells. – name fmaj)
3. CONDITIONAL: ArcMap gives NoData to cells, where there were equal number of cells from two dominant class in the neighborhood used by the focal majority calculation. These NoData holes should be filled. This was done with a conditional function in the raster calculator of ArcMap, taking the values from the original raster where the majority raster had NoData and using the majority raster's values at all other places. [Con (IsNull([majority raster]), original raster, majority raster) – name fmaj_nodata]
4. EXTRACT BY MASK: Cutting the focalmajority-nodata raster, so that all the input layers are the same size. This was done with ArcMap Extract by mask tool and the mask data was the original SRTM DEM.
5. JOIN: Attribute tables loose all their columns except ID and Value, so the rest of the attribute table fields had to be joined from the original raster based on the Value at the end. After the join the data always had to be exported in another name to save the changes.
6. COMBINE: After all layers were put to the same size, resolution and projection, they could be combined using ArcMap's spatial analyst combine.
7. FOCAL MAJORITY: The combined raster was also majority filtered with focal statistics, 5 cell radius circle, majority, ignore NoData.
8. CONDITIONAL: The NoData holes were again filled by using the values from the original raster where the majority filtered raster had no data values with a conditional in the raster calculator

9. EXTRACT BY MASK: As the focal majority increases the extent of the layer, this was also cut by ArcMap Extract by mask tool with the mask data being the SRTM DEM.
10. JOIN: The rest of the attribute table was joined by the Value to the majority filtered, NoData filled, cut raster.
11. ATTRIBUTE TABLE – ADD FIELD - FIELD CALCULATOR: Now the combined raster had an attribute table containing the codes from all the input layers. However the new e-SOTER procedure works with 4 classes to which the codes had to be converted. This was done by the Field Calculator in the ArcMap's Attribute table options. (The details of these 4 new classes can also be read at point 9.5 Parent Material in e-SOTER procedure by Joel Daroussin.) These 4 new classes are:
    a. SURFCOND: 30 char, string (Major class in the Revised hierarchy of lithology for e-SOTER)
        i. Attribute SURFCOND: surface condition of the parent material.

           - x Background polygon

           - n/a Not applicable

           Missing data

           - consolidated bare rock or eluvial/colluvial (dominantly autogene) material

           - unconsolidated unconsolidated material, loose sediment

           - water water
    b. GENETICS: 30 char, string (Type in the Revised hierarchy of lithology for e-SOTER)
        i. Attribute GENETICS: genetics classes.

           - x Background polygon

           - n/a Not applicable

           - Missing data

           - alluvial recent alluvial deposits

           - eluvial-colluvial eluvial or colluvial deposits

           - eolian eolian deposits or non-recent alluvial deposits

           - marine/estuarine marine or estuarine deposits

           - peat peat, lacustrine sediments or vegetation covered shallow water

- sapropel sapropel
- rock bare rock

c. TEXTURE: 30 char, string (Group in the Revised hierarchy of lithology for e-SOTER)

   i. Attribute TEXTURE: texture classes.

      - x Background polygon
      - n/a Not applicable
      - Missing data
      - clay clay
      - loam loam
      - sand sand
      - gravel gravelly sand
      -diamicton diamicton

d. CARBONATE: 20 char, string (Subgroup in the Revised hierarchy of lithology for e-SOTER)

   i. Attribute CARBONATE: carbonate status classes.

      - x Background polygon
      - n/a Not applicable
      - Missing data
      - calcareous calcareous material
      - non-calcareous non-calcareous material

12. CONVERSION RASTER TO POLYGON: Only the new variables are kept, all the previous fields are deleted. The combined raster was then converted to polygon by ArcMap's Raster to Polygon conversion. The input field for the conversion was the Value. Polygon simplification was not applied.
13. JOIN: The rest of the attribute table was again joined based on the Gridcode of the polygon data and the Value of the raster data.
14. CONVERSION FEATURE CLASS TO COVERAGE: The polygon parent material data was converted to coverage to meet the e-SOTER aggregation procedure requirements. The type was set to polygon, but everything was left default.

15. POLYGON LABELING: The coverage is not yet ready to use by the e-SOTER procedure as the polygons are not yet labeled. Labeling can be done in ArcInfo
    a. From the Arc: prompt in the folder where the coverage resides type :
    b. copy coverage_name save
    c. createlabels coverage_name
    d. build coverage_name
    e. Check the result in ArcMap
    f. kill save all