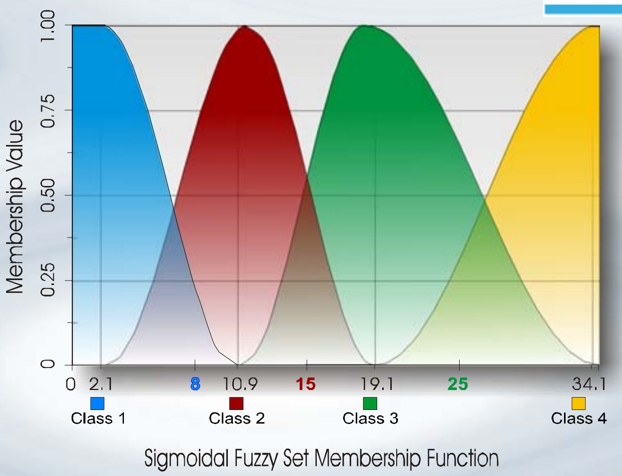
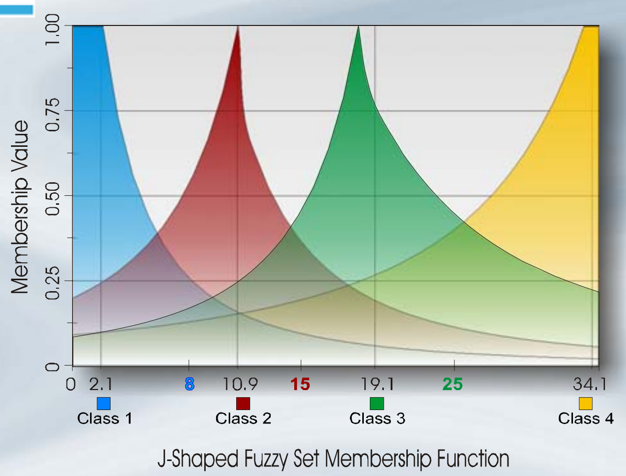




European Commission

JRC TECHNICAL REPORTS

Mapping Soil Typologies - Spatial Decision Support Applied to the European Soil Database

Roland Hiederer

2013

Report EUR 25932 EN

Joint Research Centre

European Commission
Joint Research Centre
Institute for Environment and Sustainability

Contact information
Roland Hiederer
Address: Joint Research Centre, Via Enrico Fermi 2749, TP 261, 21027 Ispra (VA), Italy
E-mail: roland.hiederer@jrc.ec.europa.eu
Tel.: +39 0332 78 57 98

http://ies.jrc.ec.europa.eu/
http://www.jrc.ec.europa.eu/

This publication is a Reference Report by the Joint Research Centre of the European Commission.

JRC80664

EUR 25932 EN

ISBN 978-92-79-29468-6 (pdf)
ISSN 1831-9424

doi:10.2788/87286

Luxembourg: Publications Office of the European Union, 2013

Table of Content

Page

List of Figure

Page

List of Tables

List of Acronyms

Acronym	Description
AOI	Area of Interest
ASCII	American Standard Code for Information Interchange
AVHRR	Advanced Very High Resolution Radiometer
CLC2000	Corine Land cover data 2000
CORINE	Coordination of information on the environment programme
CV	Coefficient of variation
DEM	Digital Elevation Model
EEA	European Environment Agency
EFTAS	European Free Trade Associations
ESDB	European Soil Database
ETRS89-LAEA	European Terrestrial Reference System 89, Lambert Azimuthal Equal Area projection
EU27	European Union of 27 Member States
FAO	Food and Agriculture Organization of the United Nations
FID	Feature identifier
FMF	Fuzzy set membership function
GIS	Geographic Information System
GISCO	Geographic Information System of the Commission
GLCC	Global Land cover Classification
ID	Identifier value
IPCC	Intergovernmental Panel on Climate Change
JRC	European Commission Joint Research Centre
MCE	Multi-criteria Evaluation
MF	Membership function
MOLA	Multi-objective Land Allocation
MULINO	Multi-sectoral Integrated and Operational decision support system for sustainable use of water resources at the catchment scale
NOAA	National Oceanic and Atmospheric Administration
NUTS	Nomenclature des Units Territoriales Statistiques
OAT	One-factor-at-a-time
PAT	Polygon Attribute Table
PTF	Pedo-transfer Function

Acronym	Description
PTR	Pedo-transfer Rule
PTRDB	Pedo-transfer Rule Database
RDBMS	Relational Database Management System
sDSS	Spatial Decision Support System
SGDBE	Soil Geographic Database of Eurasia
SMU	Soil Mapping Unit
SOTER	World SOil and TERrain Digital Database
SRTM	Shuttle Radar Topography Mission
STU	Soil Typological Unit
TMI	Topographic Moisture Index
USGS	United States Geological Survey
WRB	World Reference Base

1 INTRODUCTION

For many applications of modelling environmental conditions or developing scenarios for environmental change analysis, information on soil characteristics forms a vital component. Evaluations of the status of environmental conditions or modelling dynamics very often have a spatial dimension. This spatial dimension is required in the study of processes which include movements across a land surface, like soil erosion, or for which change statistics are required, such as transition matrices. The task is greatly facilitated, or indeed made possible, by having available the main input data in the form of spatial representations. Raster data formats are widely used for the modelling of movements through space and the storage of parameters, which change constantly over space and without pattern. Apart from the spatial aspects the data type used to record the soil characteristics is of significance to using the information for modelling purposes. To be widely applicable the data type of the soil characteristics should be recorded on ratio or interval scale rather than nominal classes. Depending on the soil characteristic this may not always be feasible, but the data type should be used to allow combining parameters by functions rather than sets of rules.

A source of uniform data on characteristics of European soils is available from the *European Soil Database* (ESDB) of the European Soil Bureau[1]. The soil information of the ESDB was collected by participating national institutions and underwent an extensive process of harmonizing the thematic content of recording the soil characteristics and ensuring spatial continuity along boundaries. For the spatial representation of the soil units a vector format is used with characteristics stored in tables in a database format. The conversion of the spatial representation from the vector to a raster format can be readily performed.

Considerably more effort is required to represent specific soil characteristics in a single spatial layer due to the relationship of one record in the spatial layer to multiple records in the table holding the soil typological data. The typological data represents a characteristic combination of several soil attributes, such as elevation range, texture and depth, and is recorded in tabular form. While the proportion of characteristic soil typologies is recorded for each typological unit, the database lacks explicit detailed information on the spatial distribution of the soil typologies within the boundaries of the spatial unit. While an average value for a soil characteristic of the spatial unit could be derived from aggregating the linked typological this procedure is restricted to those characteristics where the values are recorded as data on a ratio scale. The presentation of soil characteristics in form of a spatial layer is, therefore, intrinsically linked to the data type used to record the soil characteristics.

1 http://eusoils.jrc.ec.europa.eu/esdb_archive/ESDB/Index.htm

In this evaluation of the ESDB an attempt was made to position the soil typological units within the spatial units. Previous approaches to mapping soil attributes to the spatial layer either restricted the values to the dominant typological unit[2] (Panagos, 2006) or used general settings applied to all spatial and typological units of the database (Hiederer & Jones, 2009).

The approach followed in this study for the spatial allocation of typological units was to link the data from the soil typology table with thematically corresponding ancillary spatial data. In this respect the work performed contains some similarity to the method of predictive soil mapping (Scull, *et al*., 2003). For the spatial allocation a multi-criteria analysis within the framework of a *spatial decision support system* (sDSS) was used. Based on this method a single raster layer is generated for any of the soil characteristics, independently of the data type. This eliminates the need to aggregate data from multiple linked typological units to a typical value for a raster element. A spatially explicit allocation of typological units also greatly simplifies the mapping of soil characteristics and allows the use of the complete pool of information available in the ESDB.

The attributes defining a soil typological unit may be considered imprecise with respect to the membership of the typological unit to the attribute value or range of values; a typological unit may be mainly or completely found within an elevation range; most of the area may be clustered around a mean elevation with smaller areas scattered over the range given for elevation. That specifying an attribute by a single value or range of values without indication of any clusters may not be sufficient is recognized in the ESDB by giving a value for a dominant and a sub-dominant expression of the attribute, for example for slope.

To process imprecise and uncertain conditions fuzzy set theory is used. Applying fuzzy set theory to mapping soil properties is not new. An overview of uses of the approach is given by McBratney and Odeh (1997). Fuzzy sets and logic were also applied to land evaluation to estimate the suitability of an area for a specific land use (Chang & Burrough, 1987; Borrough, *et al*., 1992; Davidson, *et al*., 1994). In this study fuzzy set logic was used to define the membership of objects in a multi-criteria decision process. The multi-criteria decision process is more widely used to identify land suitability for a specific use. In this application of the method suitability is translated into the affinity of a set of parameters defining a mapping unit of the ESDB to suitable ancillary data in a Geographic Information System (Jiang and Eastman, 2000). The level of affinity between the typical soil characteristics and the ancillary data is then used to allocate to each grid in a raster layer a single soil typological unit, respecting the distribution of the typological units within the spatial unit specified in the ESDB.

The approach is fundamentally different from methods of digital soil mapping (McBratney, *et al*., 2003; Zhu, *et al*., 2010). It does not try to model a landscape or catena, such as used by the *SOil and TERrain Database*

---

[2] http://eusoils.jrc.ec.europa.eu/ESDB_Archive/ESDB_data_1k_raster_intro/ESDB_1k_raster_data_intro.html

(SOTER)[3] or the e-SOTER project[4]. Moreover, is does not rely on specifying factors of soil formation (Jenny, 1941) or genesis (Simonson, 1959). All defining data are taken form the ESDB and the links to ancillary spatial data only serve to support geographically positioning associations of typological characteristics with the spatial data. Where problems were encountered practical rather than theoretical solutions were applied.

3 http://www.isric.org/projects/soil-and-terrain-database-soter-programme
4 http://www.isric.org/projects/e-soter

2 EUROPEAN SOIL DATABASE

For information on soil characteristics the most detailed and harmonized spatial data set is available in the form of the *European Soil Database* (ESDB) of the European Soil Bureau (European Commission Joint Research Centre, 2003). The ESDB consists of a compilation of several integrated databases, each addressing very different aspects of soil properties. The main attribute databases used for mapping soil properties were the *Soil Geographical Database of Europe* (SGDBE) and the *Pedo-Transfer Rules Database* (PTRDB).

2.1 Components of ESDB

The SGDBE consists of several components: a spatial component given by a digitized soil map (SGDBE_4), non-spatial tables of related attributes (STU_SGDBE and STU_PTRDB) and information on linking soil attributes to the spatial units (STU_ORG). A graphical presentation of the database parts is given in Figure 1.

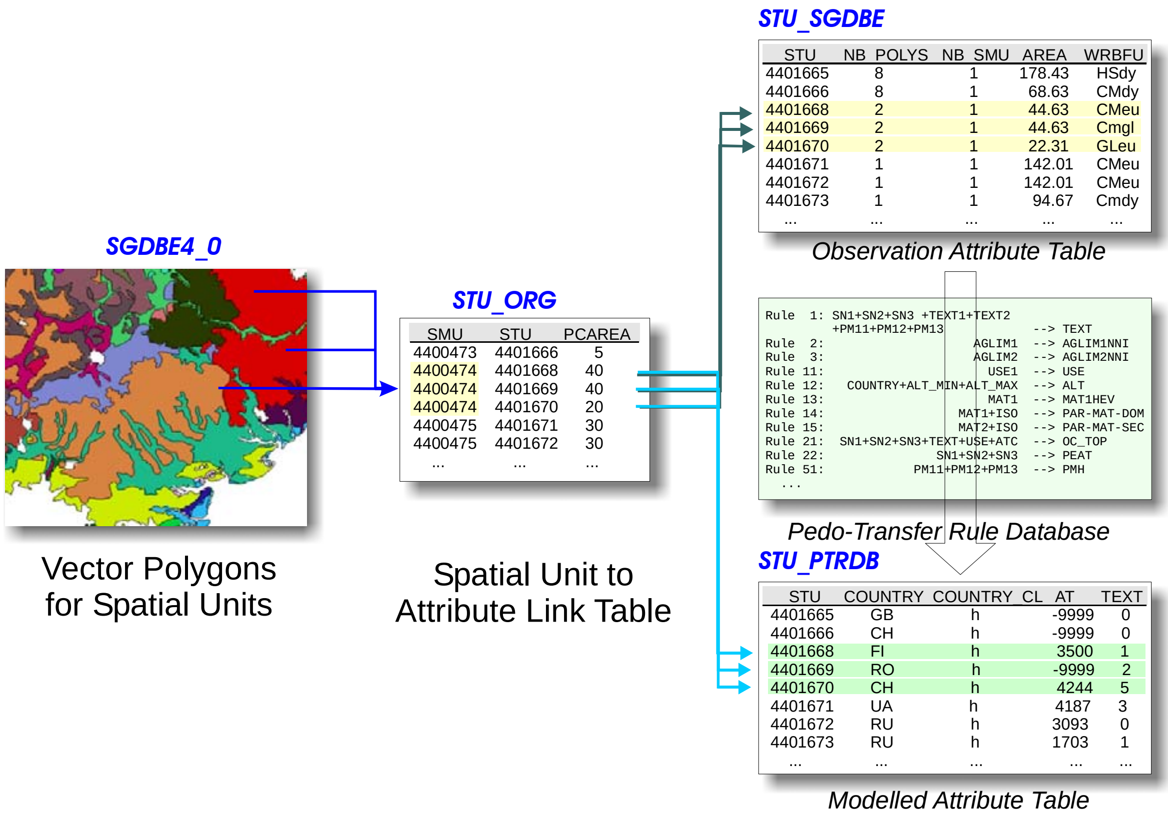


Figure 1: Data Structure of Soil Geographic Database of Europe

The digitized soil map contains a single layer of spatial units in form of 51,729 vector polygons. A *Spatial Mapping Unit* (SMU) of the database comprises one or several polygons, since the area belonging to an SMU is not necessarily continuous in space. The table associated with the vector layer defines unique identifier values (ID) for 3,856 SMUs. The 490 features linked to the SMU with the ID "-2" are defined as containing "No Information" (ESDB document SGDBE_ATTRICOD.DOC). The areas concerned are small islands, lakes, alpine regions and other areas not further identified, but without a consistent typology.

The SMU ID is used to establish a link of the spatial elements to the records of the attribute databases, which are referred to as *Soil Typological Units* (STUs). The relationship is of type one-to-many (1:n), i.e. one or more STUs may be linked to an SMU. Since the link between the SMU and associated STU(s) is inclusive and non-spatial, the geographic location of the STU within a SMU is not defined. Only the relative portion (areal percentage) of the appropriate STU(s) within the spatial unit is provided in the table [STU_ORG].

A special case is the table [SMU_SGDBE]. The table contains the typological data from the dominant STU within an SMU. The dominant STU is identified in the field [STU_DOM.SGDB4]. The table contains 3,856 records, but does not fully link to the [SGDBE4] features. SMUs with ID "-2" are given the code "0" for the dominant STU, which does not exist in the table [STU_SGDBE]. There can therefore only be a link between of at most 3,855 records and hence there is at least one record in the table [SMU_SGDBE] without correspondence in the spatial layer. The orphan link is caused by the record with the SMU ID "6", for which no correspondence is found in the parent table. Because the table [SMU_SGDBE] is only a limited view of the SGDBE it is not used in this evaluation.

The 5,262 records of the attribute tables [STU_SGDBE] and [STU_PTRDB] contain characteristics of the SMUs as defined by the STUs. An STU is composed of a typical association of specific soil attributes. The attribute table [STU_SGDBE] mainly stores the observed characteristics of the STUs of an SMU, such as the soil type according to various classification schemes. The table STU_PTRDB stores the results of applying the *Pedo-Transfer Rules* (PTR) of the PTRDB to the STUs of the observation data table.

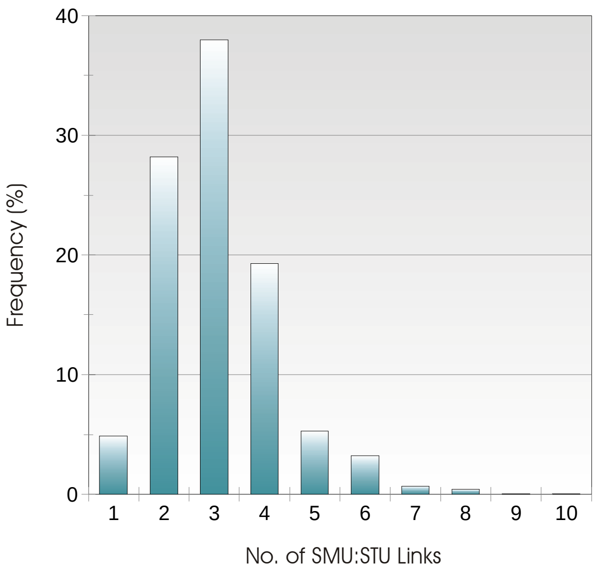
The PTRDB consists of a collection of rules stored in ASCII text files and a data table containing the results of applying the rules to the SGDBE. A PTR is designed to extend the range of soil parameters to properties not observed or measured in a soil sample. In principle, a PTR condenses the results obtained from field surveys to typical conditions, which were found to be associated with a specific soil property. The key parameters defining a property and the representative value for that property are identified through expert knowledge. Version 2.0 of the PTRDB contains rules for 39 soil attributes. The results of applying the rules are stored in the PTRDB table.

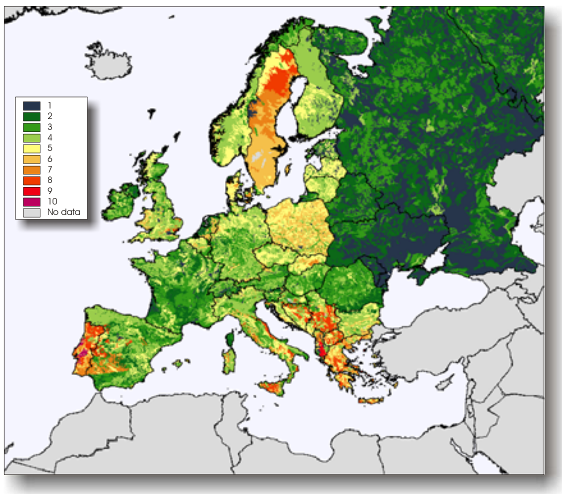
2.2 Relationship between Soil Mapping Units and Soil Typological Units

For the 3,856 SMUs of the spatial layer 11,811 links are defined to the 5,262 records of the STU database in the table [STUORG]. SMUs with ID "-2" in the spatial layer are not referenced in the attribute tables, while SMUs with the ID "6" of the attribute tables are not present in the spatial layer. As a consequence, any STUs defined as "Rock outcrops" (SOIL code: 6) are not linked to the spatial layer and the rule for referential integrity between the relationships of the tables is not met. The absence of a link is indicated by the entry 0 in the field [AREA/STU_SGBDE].

With the exception of SMUs with the ID "6" all SMUs of the spatial layer are linked to at least 1 STU. However, the STU with the code 3860126 of table [STU_SGDBE] has no correspondence in the tables [STUORG] or [SDGBE4]. Soil properties are defined for the STU and the SMU, to which the STU belongs, is located in the Czech Republic. As a consequence, there are 11,810 links with data from the spatial layer to the attributes given in the table [STU_SGDBE].

The number of STUs linked to an SMU ranges from 1 to a maximum of 10. The occurrence of the number of links between SMU and STU is presented in Figure 2 together with the spatial distribution of the number of SMU-STU links.





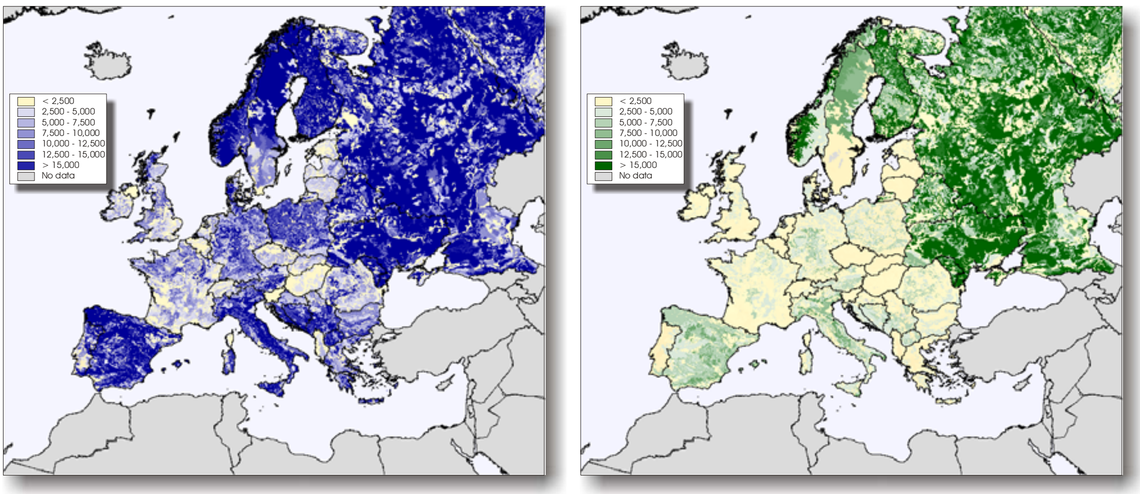
a) Frequency of SMU:STU Links *b) Spatial distribution of SMU:STU Links*

Figure 2: Frequency and Spatial Distribution of Links between SMU and STU

The average number of STUs linked to an SMU is three, which is also the most frequent number (1,465 occurrences, 38% of all links). On average the least number of links (2.65) are present for SMUs in Albania, the most (5.88) in Portugal. The smallest average SMU areas are those in Slovenia (191 km^2) while the largest SMUs are defined for Finland (average SMU area: 22,800 km^2).

Less than 1% of SMUs are defined by 7 or more STUs. More than 5 STUs are frequently defined for SMUs in Greece, Portugal, Sweden and parts of Spain, former Yugoslavia and southern Italy. For SMUs of the former Soviet Union the number of STUs linked to an SMU is lower than for SMUs of other European areas. The data for the soils of the former Soviet Union was added to the original Version 1.0 of the ESDB, which subsequently changed from the *Soil Geographical Database of the European Communities* of 1986 to the *Soil Geographical Database of Eurasia*.

A high number of linked STUs could indicate that the area of the SMU is defined with comparative detail. This assumption only holds, if the area covered by the SMUs is also similar and the variation of soil properties within an SMU is similar. For a more detailed presentation of the level of spatial detail, with which an SMU is described by soil typological information the spatial distribution of the size of the SMUs and the average size of STUs is presented in Figure 3.



a) Spatial distribution of SMU area *b) Spatial distribution of mean STU area*

Figure 3: Spatial Distribution of SMU and Mean STU Area

The graph shows the uneven distribution of SMU sizes within a country, but also the marked differences between countries with respect to the average size of the STUs. The SMU and STU areas vary most for Italy and Spain and to a lesser degree for Norway and Finland. Western and Central Europe are covered by SMUs with an area of by and large less than 15,000km^2. The SMUs of Northern Sweden, Finland and Eastern Europe are generally 30,000km^2 or more.

The mean area of the STUs linked to an SMU follows the area of the SMUs. They are relatively small for the region covered by the previous version of the soil database, for the most part below 3000 km^2, and large for the extended area, mainly above 10,000 km^2. SMUs are generally larger than the average size in Finland, Norway and Sweden, where the area covered by an STU regularly exceeds 3,000 km^2. The occurrence of SMU areas and then mean area of STUs linked to an SMU is presented in Figure 4.

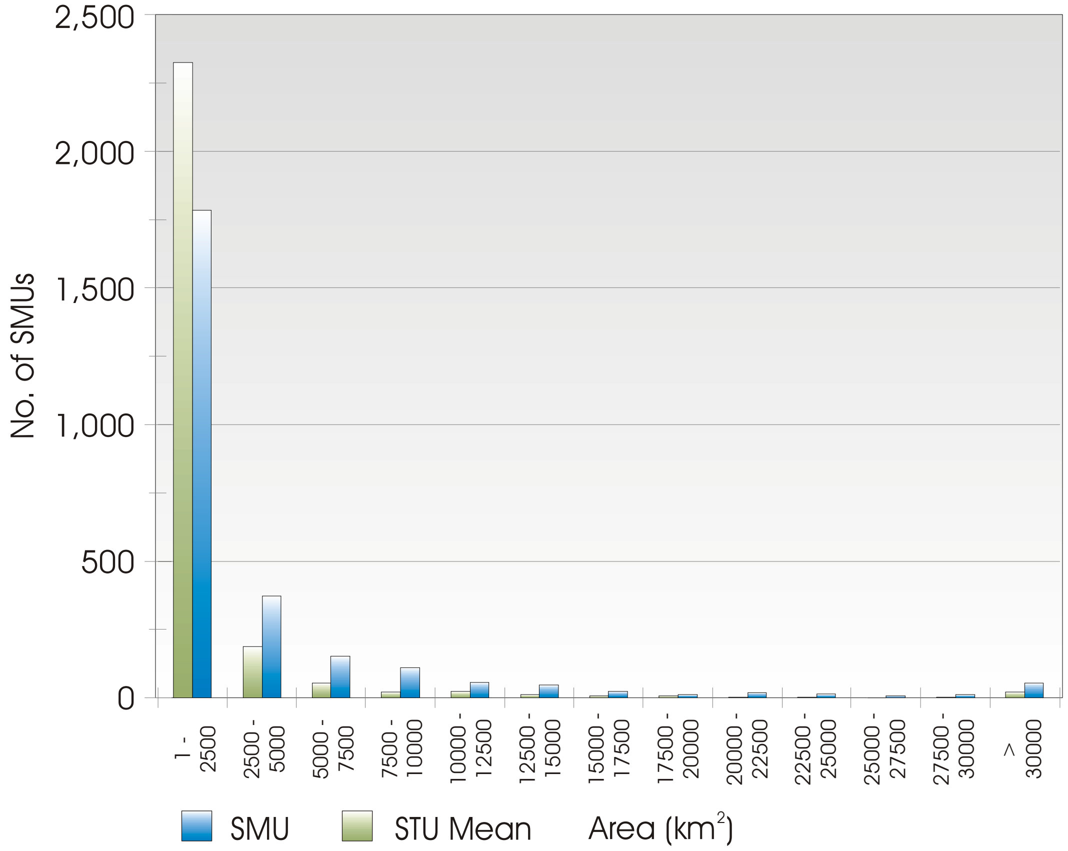


Figure 4: Frequency of SMU Area and Mean Area of STUs in SMU

The graph shows that most SMUs, over 66% of all SMUs, cover areas of 2,500 km^2 or less. The mean area of STUs is less than this size for over 83% of all SMUs. The occurrence of SMUs and the mean area of linked STUs decrease rapidly with size. For areas covering 30,000 km^2 or more 2.0% of all SMUs are found, while the frequency of the mean area of linked STUs of this size is 0.9%. However, the area covered by SMUs with an area of 30,000km^2 or more amounts to 33% of the total area covered by the SGDBE.

2.3 Mapping Soil Attributes

To be useful for an integrated analysis of the soil typological data with other spatial data the methodology for mapping attributes of the typological databases should address some particularities of the ESDB:

- Use of all information linked to a spatial unit in the typological database, not only the properties of the dominant STU.

- Express STU data units as physical parameters in the form of values on a ratio scale instead of categorical class values.

Several approaches of processing the soil typological data to obtain a spatial representation of a soil property can be use with varying degrees of complexity. Appropriate methods depend on the purpose of mapping a soil property, but also on the characteristics of the property and the storage format in the database.

2.3.1 Attribute Properties

In the SGDBE the geographic location of soils is defined by a single spatial layer of polygons, which delineate the SMUs, while corresponding soil typological properties are stored in the accompanying attribute tables. This structure of defining a single spatial layer and storing soil properties in a non-spatial tabular form allows for efficient data storage and simplifies the management overhead when up-dating STU information. However, any characteristic of a geographic location can only be identified through the link of the SMU with the corresponding STU(s) and thus be mapped as an attribute of the complete polygon comprising the SMU.

The one-to-many relationship between the spatial SMU and the tabulated attributes of the STU(s) restricts direct mapping to a property stored to a single STU for each SMU. Only those SMUs are fully defined by the soil typological attribute, to which just one STU is linked. However, there are commonly at least three STUs associated with an SMU and with a maximum of 10 STUs linked to an SMU. This arrangement and the lack of explicit information about the position of an STU within the spatial unit introduces a non-trivial element of complexity into the process of spatially representing soil typological properties.

As a consequence, soil typological properties of a location cannot be mapped exhaustively to a spatial layer through a direct link between the spatial unit layer and the soil property table. A graphical presentation of the ambiguity of the link between spatial and attribute data is given in Figure 5.

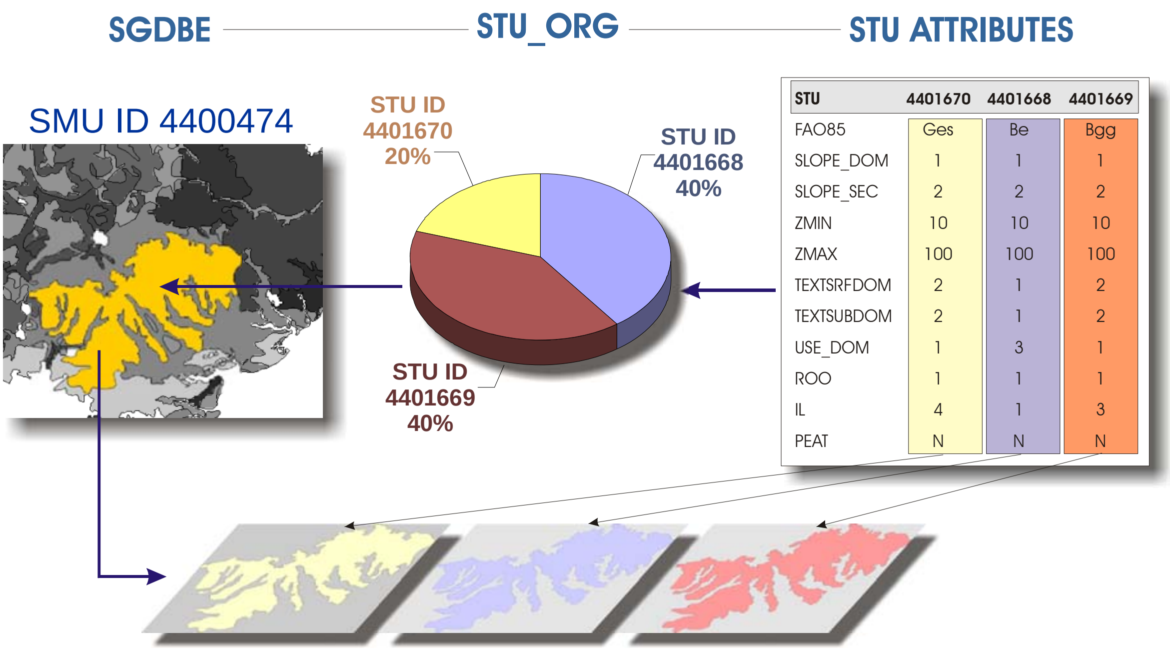


Figure 5: Relative STU Area Proportion in SMU

These limitations to directly mapping all typological properties of a spatial unit have led to some simplification when mapping soil characteristics from the attribute databases. Traditionally, when seeking to display soil properties, only the parameter value of the most widely present (dominant) STU within the SMU is displayed. This method of mapping attributes of only the dominant STU has several advantages. It can be used to map all attributes, including a representation of ordinal parameters, such as soil class, it allows retaining the data format of the parameter as stored in the attribute tables and it avoids the issue of aggregating data from 1:n relationships. Disadvantages of the method are that no dominant STU may exist, i.e. when the largest area in the spatial unit is not unique. In addition, in cases where three or more STUs are associated with an SMU the dominant STU may define less than 50% of the area of an SMU. Because of the selection nature of the method of mapping only the properties of the dominant STU the representation of a soil property is biased against properties expressed by sub-dominant soil typologies. This bias can limit the usefulness of the spatial attribute layer in statistical analyses of summary statistics and distribution of soil properties. The bias in soil properties can further propagate through models, in which the soil information forms a significant part of analysis.

A more comprehensive way of spatially representing soil properties stored in the attribute tables is to use the information of all STUs linked to an SMU. The contribution of each the STUs to the characteristics of the SMU can be accounted for by weighing the attribute value according to the proportional representation of the STU at a given geographic location. The relative portion of an STU within an SMU is given in the table [STU_ORG]. For parameters recorded using continuous units the calculation can be expressed as:

$$P_{SMU} = \sum_{i=1}^{n} P_{STU(SMU)_i} * A_{STU(SMU)_i}$$

where

- *P*: property value of STU
- *A*: relative portion of area covered by STU within SMU
- *i*: number of STUs linked to an SMU

Integrating parameter values form all STUs linked to an SMU does not improve the geographic position of the attributes within a spatial unit. For a better spatial representation of attributes additional information has to be integrated into the mapping procedure. One of the methods for improving attribute positioning exploits the morphological information found in the [STU_SGDBE] table, i.e. slope and elevation. By associating the STU records with an ancillary *digital elevation model* (DEM), the morphological information can be used to improve the location of one or more STUs within the SMU. This method has been applied to generate the European Soil Raster Data Set.

2.3.2 Attribute Data Types

The attributes transferred to the spatial data set cover the major soil characteristics from both the SGDBE and the PTRDB. For the purpose of processing data from the soil typological property fields in the attribute tables, the representation of a property can be grouped into three types of value scales:

- Type 1: ratio or interval scale in measured unit (numerical data items)
- Type 2: ordinal scale (ordered categorical data items)
- Type 3: nominal scales (descriptive categorical data items)

In addition to those three scales there is also an indicator for missing data, for which a single Type 0 is defined. The grouping of the soil properties into reporting scales used in this evaluation may not adhere in all aspects to the classical division, e.g. some interval data may be treated as ratio data, but is suitable for the purpose of this exercise.

The scale at which a soil typological property is reported determines the format type of the data field in the attribute table and subsequently the representation of a property as a spatial data layer. As long as the attribute databases are not exclusively linked to a geographic location and properties have to be weighted to comprehensively characterize a location only properties reported in ratio or interval scales can be expressed in single spatial layers. Soil typological data reported on a nominal scale can be mapped as numeric codes, but comprehensively only in as many layers as there are links of the STUs to the SMUs, i.e. using 10 spatial layers for each property.

An overview of the soil typological data by scale used to report a property as stored in the SDGBE is given in Table 1. By definition all soil properties of the PTRDB are reported on ordinal or cardinal scales.

Table 1: Soil Typological Property by Type of Reporting Scale

Type 1: Ratio or Interval	Type 2: Ordinal	Type 3: Nominal
AREA	SLOPE_DOM	WRBFU
ZMIN	SLOPE_SEC	FAO90FU
ZMAX	TEXTSRFDOM	FAO85FU
	TEXTSRFSEC	PARMADOM
	TEXTSUBDOM	MAT1
	TEXTSUBSEC	PARMASE
	TEXTSRFDOM	MAT2
	TEXTDEPCHG	USE_DOM
	ROO	USE_SEC
	CFL	AGLIM1
		AGLIM2
		IL
		WR
		WM1
		WM2

For field name definition: see file SGDBE_ATTRICOD.DOC of SGDBE.

The table shows that very few (3) properties of the [SMU_SGDBE] table are recorded on ratio or interval scale and these are not directly soil properties. Texture data are classified to an ordinal scale, although the property would qualify to be recoded as ratio data (interval with natural zero). Most data describing other soil properties, such as the parent material, are recoded on a nominal scale by their nature.

For data of Type 1 a proportional representation of the property can be directly calculated and the property can be represented in a single spatial data layer. Properties stored as Type 2 data first need to be transferred to intervals with an equal ranges before they can be processed to be represented in a single spatial data layer. Any of the properties stored as Type 3 data can be mapped as a set of multiple spatial layers, one for each STU, provided the numeric codes (integer values) are defined for the nominal data. Transferring properties stored on nominal scales to spatial data layers may be considered wasteful in terms of data storage requirements, but it allows processing PTRs in the spatial domain.

The soil property value formats of the attribute table fields are not always as unequivocal as shown in Table 1. Ratio data should be stored in a field with an appropriate type (float, real). Yet, some of the soil properties in the database contain a mixture of data format types, e.g. soil texture is coded by integer values, but also contains the alpha-numeric entry "*no texture*" which subsequently leads to the filed being defined as character format. Whenever such complex coding is encountered, the nominal codes are first separated from the ordinal entries and then represented in a separate binary spatial data layer.

2.3.3 Completeness of Attribute Information

The STU attribute table [STU_SGDBE] contains a total of 5,262 records. A summary of the completeness of the information given in the data fields is provided in Table 2.

In the evaluation of data completeness a distinction is made between blank entries and entries containing a defined code for missing data. For some fields, such as the soil classification classes, no specific code for missing data is defined. For other fields the code is either "0" or "-999" ([ZMIN], [ZMAX]). For the field [IL] one entry contained the code "9", which is not defined. This was taken to indicate missing data.

Included in the inventory of data completeness are the 6 STU, which are linked to non-soil areas. It is arguable if these STUs should contain any attribute data. Even when excluding the STUs of non-soil areas none of the fields contains a complete information cover.

Where a code for missing data is defined in the attribute specifications there is also a code indicating the absence of the attribute. For example, for the field containing information on obstacles to root ([ROO]) the code "0" is used to identify missing information and the code "1" to identify that there are no obstacles to roots. This logic is not universally applied. For the water management system indicators ([WM1], WM2]) codes are defined for cases where the information is not applicable (no agriculture; code "1") and where there is no water management system (code "2"). Yet, for [WM1] code "1" and for the code for the type of an existing water management system [WM2] codes "1" and "2" are not used. It appears that the code for missing data ("0") is used for [WM1] when the land use is not agriculture and for [WM2] when [WM1 is set to indicate no water management system.

Table 2: Completeness of STU Attribute Information in table [STU_SGDBE]

Field	Blank		No Data Code*		Completeness
	Records	*of total (%)*	*Records*	*of total (%)*	*%*
WRBFU	15	0.3	-	-	99.7
FAO90FU	3,025	57.5	-	-	42.5
FAO85FU	17	0.3	-	-	99.7
SLOPE_DOM	0	0.0	184	3.5	96.5
SLOPE_SEC	0	0.0	2,404	45.7	54.3
ZMIN	0	0.0	137	2.6	97.4
ZMAX	0	0.0	138	2.6	97.4
PARMADO	0	0.0	46	0.9	99.1
MAT1	143	2.7	-	-	97.3
PARMASE	0	0.0	2,463	46.8	53.2
MAT2	2,488	47.3	-	-	52.7
USE_DOM	0	0.0	139	2.6	97.4
USE_SEC	0	0.0	1,009	19.2	80.8
AGLIM1	0	0.0	183	3.5	96.5
AGLIM2	0	0.0	183	3.5	96.5
TEXTSRFDOM	0	0.0	145	2.8	97.2
TEXTSRFSEC	0	0.0	2,170	41.2	58.8
TEXTSUBDOM	0	0.0	1,975	37.5	62.5
TEXTSUBSEC	0	0.0	2,978	56.6	43.4
TEXTDEPCHG	0	0.0	717	13.6	86.4
ROO	0	0.0	305	5.8	94.2
IL	0	0.0	332	6.3	93.7
WR	0	0.0	269	5.1	94.9
WM1	0	0.0	927	17.6	82.4
WM2	0	0.0	4,157	79.0	21.0

* as specified in file SGDBE_attricod.txt.

More intricate is the interpretation of the available information for secondary attributes. It is not always evident whether the absence of any information indicates the absence of the attribute or missing information, as in the case of dominant and secondary parent material ([MAT1], [MAT2]), because there may well be no secondary attribute. Therefore, the fields for secondary attributes contain less information than the field with the dominant attribute. Conversely, for limitations to agricultural use ([AGLIM1], [AGLIM2]) the records with missing data are the same for both fields.

2.4 Soil Database in Raster Format

The purpose of creating a raster version of the ESDB was to provide a standardized set of basic soil properties in the form of consistent and readily available spatial data layers. The layers should be directly usable as input information to spatial models across a wide range of applications. Many models using soil properties as a major source of input, e.g. for modelling run-off, sedimentation, soil loss or land use change, rely on neighbourhood statistics and require data to be available in raster format.

In order to achieve the purpose of creating a raster version of the ESDB several aspects requiring specific attention were identified:

1. spatial unit rasterization for integration with complementary thematic spatial data;
2. improve the spatial positioning of typological attributes within a spatial unit;
3. use information of all typological units linked to a spatial unit for attribute mapping;
4. treatment of missing data.

On account of the intended use of the spatial layers and the nominal scale of the SMUs in the SGDBE (scale 1:1mio.) the attributes are stored as raster data layers in preference to a vector format with a grid size of 1 km. The methodology applied to reach the objectives of improved spatial positioning and representation of attributes is described hereafter.

2.4.1 Conversion of Vector to Raster Format

The soil database used was the "European Soil Database, Version 2.0". The Soil Geographical Database of Eurasia (SGDBE) was on Version 4 beta, 25/09/2001, the Pedo-Transfer Rules Database was Version 2.0 (files from 16.12.2003). The nominal scale of the data is 1:1mio. The scale is not uniform across all countries included in the database. The SMUs for Belarus, Russia, Ukraine, and Moldova were added to Version 2.0 from FAO project archives. The areas are covered at scale 1:2.5mio and the spatial units of the new areas are extent as far as the state boundaries. Assuming a constraint of ± 0.5 mm when digitizing the vector data the geometry could be resolved to a grid of 1,000 m. Indeed, the estimated positional accuracy of the data layer is estimated to range between 500 and 5,000 m (document: SGDBE_metadata.doc of European Soil Database, V2, SGDBE V.4 beta, 25/09/2001). This contrasts with the specifications for the polygon size of the vector layer given in the file SGDBE_dictiona.doc. In the file a resolution of 100m, i.e. an accuracy of 0.1 mm at the map scale, is stated. The minimum polygon area is given as 9 ha. In the *polygon attribute table* (PAT) to the SGDBE layer [SGDBE4_0.dbf] areas of the polygons are as computed by the GIS software used to manage the data. Of the 51,729 polygons 1,375 have an area of <100 ha (1 km^2). Summing up the size of the polygons to STUs, the minimum size is 209 ha. The grid size of 1,000 m for the raster layer seemed a reasonable compromise between the smallest area mapped and the positional accuracy of the data.

The vector layer is projected according the *Lambert Azimuth Equal Area* (LAEA) specifications used by the Eurostat *Geographical Information System at the COmmission* (GISCO) dataset[5]. For the raster dataset the vector layer was re-projected according to the specifications of the *European Terrestrial Reference System 89 LAEA* (ETRS-LAEA) projection. The latter corresponds to the recommendation of the Inspire initiative for representing data at the resolution of the data for area-based analysis (Annoni *et al*., 2003). The spatial extend of the final layer was determined by the need to cover the mainland area of all Member States of the European Union plus ascension countries. To be compatible with other thematic data layers the boundary parameters of the Corine Land Cover 2000 raster dataset of the *European Environment Agency* (EEA) were used to specify the geographic extend of the SMU layer. The projection and grid specifications are given in Table 3.

Table 3: Projection and Grid Specifications for Raster Layers

ITEM		SPECIFICATION
Reference System	Datum	European Terrestrial Reference System 1989 (ETRS89)
	Projection	Lambert Azimuth Equal Area (ETRS-LAEA)
Origin	Longitude	10.0
	Latitude	52.0
	X	4321000.0
	Y	3210000.0
Grid spacing		1000.0 m
Coverage	min. X	800000.0000
	max. X	7500000.0000
	min. Y	700000.0000
	max. Y	6500000.0000
Background value		-9000

Once the spatial parameters for the pixel size, projection and spatial extend are determined the process of rasterizing vector data is performed more or less by a single command in most GIS software packages. However, the output of the GIS transformation does not necessarily satisfy the objective of the task of rasterizing the SMU polygons and additional steps of data processing have to be applied.

[5] http://epp.eurostat.ec.europa.eu/portal/page/portal/gisco_Geographical_information_maps/introduction

2.4.2 Adjusting SMU Raster Layer

The SMU raster layer undergoes several stages of data adjustment for simplified processing, harmonization of geographic extent with reference data and corresponding thematic coverage with ancillary data. The objective of the adjustments is to allow using the spatial layers in overlay functions of a GIS without the need for further adjustments to avoid generating artefacts due to inconsistent geometric properties of features.

2.4.3 Recoding SMU Identifiers

The spatial units of the SGDBE are stored in vector format and consist of 51,729 polygons with as many feature identifiers (FID). In the vector data the separate polygons are joined by a common identifier to form a set of 3,856 SMUs. The identifiers of the SMUs are not stored as a continuous series of numbers starting at 1 with an increment of 1, but rather represent coded values with a maximum identifier of 4420577. By contrast, in the raster layer the SMU identifiers are stored directly as the pixel values. Thus, spatially non-continuous SMUs can be identified without the need for a link table. To link soil properties from the attribute tables to the raster layer the identifiers use an integer format. For some raster GIS packages these integer numbers are limited to 16bit, i.e. the maximum number is 32,767. This value is exceeded by both, the number of polygons and the SMU identifier used in the vector data. To allow using 16bit integer values as identifiers for the SMU raster layer a new identifier for all areas comprising a SMUs, i.e. not separating discontinuous areas of an SMU, was generated. The table STUORG links STUs to 3,856 SMUs. Defining a unique identifier for the spatial layer units as type short integer is therefore workable.

2.4.4 Adjusting SMU Geometry to GISCO Reference Data

The SGDBE covers most of Europe and Asian parts of the former Soviet Union, but not all. Notable exceptions are Malta and Cyprus, but also overseas areas, which signifies that the area of the European Union of 27 Member States (EU27) is not complete. Data are included for other countries of the *European Free Trade Associations* (EFTA), except for Iceland.

To allow the analysis of the soil data by administrative regions for statistical purposes the coverage of the SMU raster layer is adjusted to the area covered by the GISCO reference layer for administrative and/or statistical units[6]. Since the SGDBE includes also data for countries outside the areas for which a *Nomenclature of territorial units for statistics* (NUTS) is defined, such

[6] http://epp.eurostat.ec.europa.eu/portal/page/portal/gisco_Geographical_information_maps/geodata/reference

as the countries of the former Yugoslav Republic, the layer for countries is used at scale 1:1,000,000 (GISCO_CNTR_1M_2010).

The adjustment of the SGDBE spatial layer to the reference layer affects border regions of the overlap in the area covered, which are coastal zones and country boundaries of the former Soviet Union. In most cases the geographic extent differs by not more than one grid cell. Largely removed from the adjusted SMU layer are the tidal areas of Germany along the North Sea coastline (SMU/STU: 490002 / 490003). Not affected by the adjustment of the SMU cover are areas of inland water, such as estuaries or lakes or any internal country boundaries.

2.4.5 Removal of Non-Soil Land Cover Types

Once the SMU raster layer has been prepared to allow mapping soil typological information directly to the layer the next step in pre-processing is to make the layer content compatible with other thematic datasets for seamless integration. When comparable information is stored in more than one layer features may be shifted geographically between the layers and lead to areas with undetermined results in the output of models. The layers are considered compatible when features common to both datasets share the same geographic extent. This can be achieved by selecting one data set as a reference for the specific feature and transferring the feature to all other layers used in the integrated analysis.

The soil database contains common features with land use / cover types, for which no soil data are provided. The geographic extent of these areas is delineated by SMUs and can be identified in the fields containing the codes of the various soil classification schemes (table [STU_SGDBE]). The code for non-soil areas is a series of identical number, although the actual arrangement of the numbers varies between the fields. The codes used to denote areas without soil data in the three soil classification schemes are given in Table 4.

Table 4: Numeric Codes for Land Cover Types without Soil Data in Soil Classification Fields

Numeric Code			**Land Cover Type**
WRBFU	*FAO90FU*	*FAO85FU*	*Label*
1^1^1	1^11	111	Town
2^2^22^2	2^22	222	Soil disturbed by man
3^3^33^3	3^33	333	Water body
4^4^44^4	4^44	444	Marsh
5^5^55^5	5^55	555	Glacier
6^6^66^6	6^66	666	Rock outcrops

^ : space (ASCII code 32) between numbers

The reason for the organization of the numbers in the codes has been lost with time. According to the documentation and consistency of coding the code for "Town" in the field recording the soil type according to the *World Reference Base* (WRB) [WRBFU.STU_SGDBE] should be "1^1^11^1", but is in the field as given in table. The identification of features without soil data or non-soil areas is fully consistent between the fields. There are no mixed identifications between land cover types and no missing entries. No specific identifier is used for areas without data, i.e. the field entry is left empty. Therefore, the absence of a code in any of the fields for soil classes does not indicate a non-soil area or the absence of soil data. Although all fields for soil classes contain empty entries there are no records in the table [STU_SGDBE] where not at least one entry is given the fields. This condition allows identifying the areas without soil data by setting the numeric codes in a query in any of the soil classification fields.

The absence of soil data does not necessarily indicate land surface without soil. Surfaces, where soils have formed but without soil information due to a lack of a survey or accessibility are those of urban areas (soil code 1) and soils disturbed by man (soil code 2). The other 4 categories of land cover types coded in the soil classification field can be considered to designate areas without soil (water body, marsh, glacier or permanent snow and rock outcrops).

In the SGDBE areas of non-soil are largely assigned to a single SMU, which is characterized by one STU. The IDs for the SMUs and STUs range from 1 to 6 and are identical. The IDs follow the coding of the non-soil areas used in the fields to report the FAO soil type. An exception of the direct link between SMU and STU is made for "Rock outcrops". This land cover type is generally not the only STU linked to an SMU, but part of a composite characterization and often sub-dominant.

Contrary to the database for the SGDBE layer the delineation and identification of non-soil areas is not entirely consistent. For once, the spatial layer contains SMUs with the identifier "-2". This identifier is not recorded in the data tables and may therefore be interpreted as areas of missing data.

For non-soil areas in Switzerland the land cover linked to the SMU identifiers does not correspond to the land cover associated with the SMUs in other regions. This is noticeable in international water bodies, such as Lake Geneva, Lake Konstanz or Lago di Maggiore, which are divided into different SMUs following state boundaries. The problem concerns three non-soil land cover types, as given in Table 5.

Table 5: General Land Cover Representation in SGDBE and Deviation in Switzerland

SMU / STU	General Land Cover	Swiss Land Cover
ID	*Label*	*Local Correspondence*
1	Town	-
2	Soil disturbed by man	Water body
3	Water body	-
4	Marsh	Town
5	Glacier	Glacier & rock outcrops
6	Rock outcrops	-

For SMUs in Switzerland SMU / STU ID 2 is used to delineate water bodies, 3 for towns and 5 for a combination of glacier and rock outcrops. The local coding of the Swiss SMUs was standardized to the general scheme by re-coding the areas IDs in the spatial layer. This adjustment can be of consequence when linking other land cover data to the soil database.

In Sweden a code of "-2" is given to lakes instead of code "3". Since the SMU code "-2" is not linked to STU data it is interpreted as missing data. The STUs for missing areas are interpolated, while for lakes they are not. Therefore, the SMUs concerned were recoded in the SGDBE. Smaller islands are also assigned a value of "-2". These areas were kept to allow estimating SMUs for these areas.

The land cover types without soil data are also found in other land use / cover data, which could be used in integrated spatial analyses, or when overlying the soil data with a river network or layer of urban areas. For land cover / use types it can be expected that the more accurate geographic extent of the features is provided by the land cover dataset. For Europe the most comprehensive data is the CORINE Land Use / Cover data distributed by the European Environment Agency. It would thus be a possibility to use the CORINE data as a reference for the areas without soil data and adjust the SMUs accordingly. However, the reference data has to cover all areas of the soil database and remain invariable between data of the same thematic field. This is not the case for the various land use / cover products available. Rather than tying the geometry of the SMUs to one land use / cover product for the areas without soil data it seems more flexible to avoid using such areas in the soil database.

From the re-classified SMUs a layer of non-soil areas was generated. The layer is composed of SMUs with only non-soil land cover types. The SMUs concerned are given in Table 6.

Table 6: Areas where Absence of Soil is Specifically Defined

SGDBE		Land Cover	Non-Soil Layer	
SMU	*STU*	*[FAO85FU]*	*ID*	*Label*
-2	-	-	1	Missing
1	1	111	-*	Town
2	2	222	-*	Soil disturbed by man
3	3	333	3	Water body
70225	70022	333	3	Water body
4	4	444	4	Marsh
5	5	555	5	Glacier
460063	4600631	666	6	Rock outcrops
430023	430015	666	6	Rock outcrops
430023	430030	555	6	Rock outcrops

** Not used in non-soil layer.*

The non-soil layer excludes artificial surfaces ("Town") and other areas where soils are present, but not further characterized in by an STU. The layer includes all SMUs defined as water bodies and marsh land. In further includes areas of glaciers and rock outcrops which are covered completely by an SMU. Contrary to water bodies and marshes, not all such areas are therefore included in the non-soil layer, because glaciers are also found in SMU 337583 (50%) and rock outcrops in 62 SMUs with a relative cover from 2 to 90%. As a consequence, the non-soil layer does not cover all non-soil areas in the SGDBE.

The distribution of non-soil SMUs in the SGDBE is presented in Figure 1.

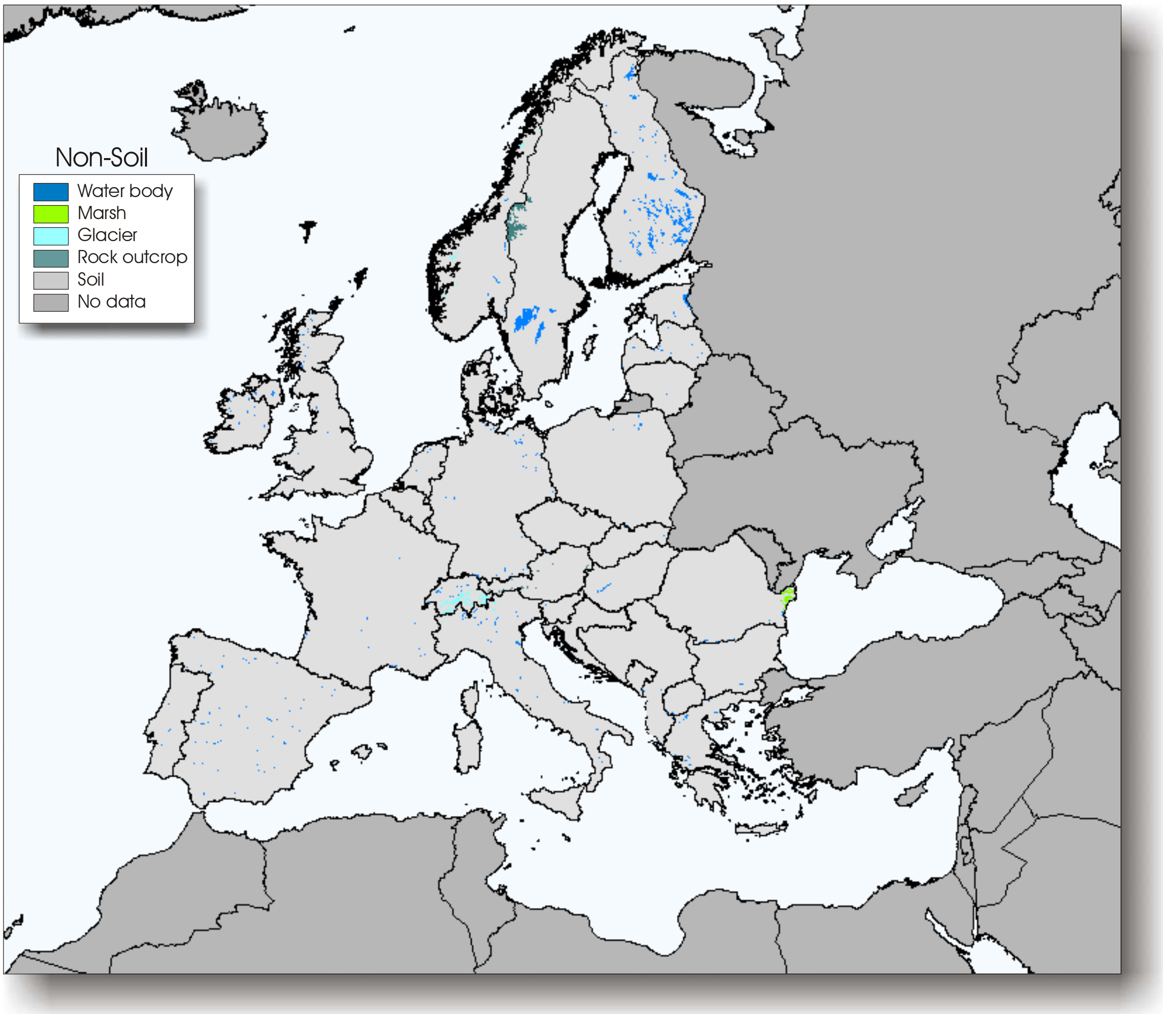


Figure 6: Distribution of SMUs Exclusively Covered by Non-soil STUs

The classification of the non-soil areas is not very reliable. As shown, the regions classified as glaciers contain also considerable areas of rock. The main use of the layer is to represent the minimum of areas without soil.

2.4.6 Estimation of Soil Code for Land Cover Areas without Soil Data

The total land surface area included in the adjusted SMU raster layer covers 4,935,649 grid cells. The area covered by soil data according to the codes of Table 4 are 4,822,293 grid cells or 97.7% of the land surface.

In order to use an external land use / cover data set for areas without soil information in the SGBDE the corresponding areas are first removed and the undefined areas are then assigned to areas with soil data. This procedure modifies the spatial layer but also the records in the SMU and STU database tables.

A SMU is removed from the raster layer when it comprises only land cover types without soil data. Where an SMU is partially covered by soil data the area of the associated STUs is proportionally adjusted to cover the total SMU area. Next, the remaining SMUs are geographically extended to cover the areas left undefined by the SMUs without soil data. The primary objective of the spatial extension of the soil information is thus not to estimate soil properties for areas, where no data are available, but to remove or reduce the influence of the imprecise information on the extent of land cover types on the results of spatial modelling.

Of the 5,258 STUs in the table [STU_SGDBE] 60 are defined as types without soil data. The number of SMUs with a link to these STUs is 404. Two SMUs (ID 337583 and 430023) are linked to two STUs containing codes for no soil data. Completely covered by STUs without soil data ([PCAREA] = 100) are 10 SMUs (SMU ID 430023 covered by two different codes for no soil data). The remaining 396 SMUs are partially covered by areas without soil data. The relative proportions of the STUs without soil data linked to SMUs ranges from 2 to 90%. The distribution of the proportion of SMUs linked to STUs without soil data is presented in Figure 7.

For most SMUs linked to STUs without soil data the proportion is < 20%. The SMUS are mainly located in Albania, Greece, Portugal, Spain and the west coast of the United Kingdom, with smaller areas occurring in Austria, Italy, Romania and Slovenia.

In total the area without soil data in the region covered in this study is 123,027 km^2, distributed across 12,991 clusters, which range from 1 km^2 to 6,402 km^2 in size. Many clusters concern areas of inland water or bare areas in alpine regions. The largest clusters where soil could be expected are the urban areas of Paris (1,845 km^2) and London (1,660 km^2).

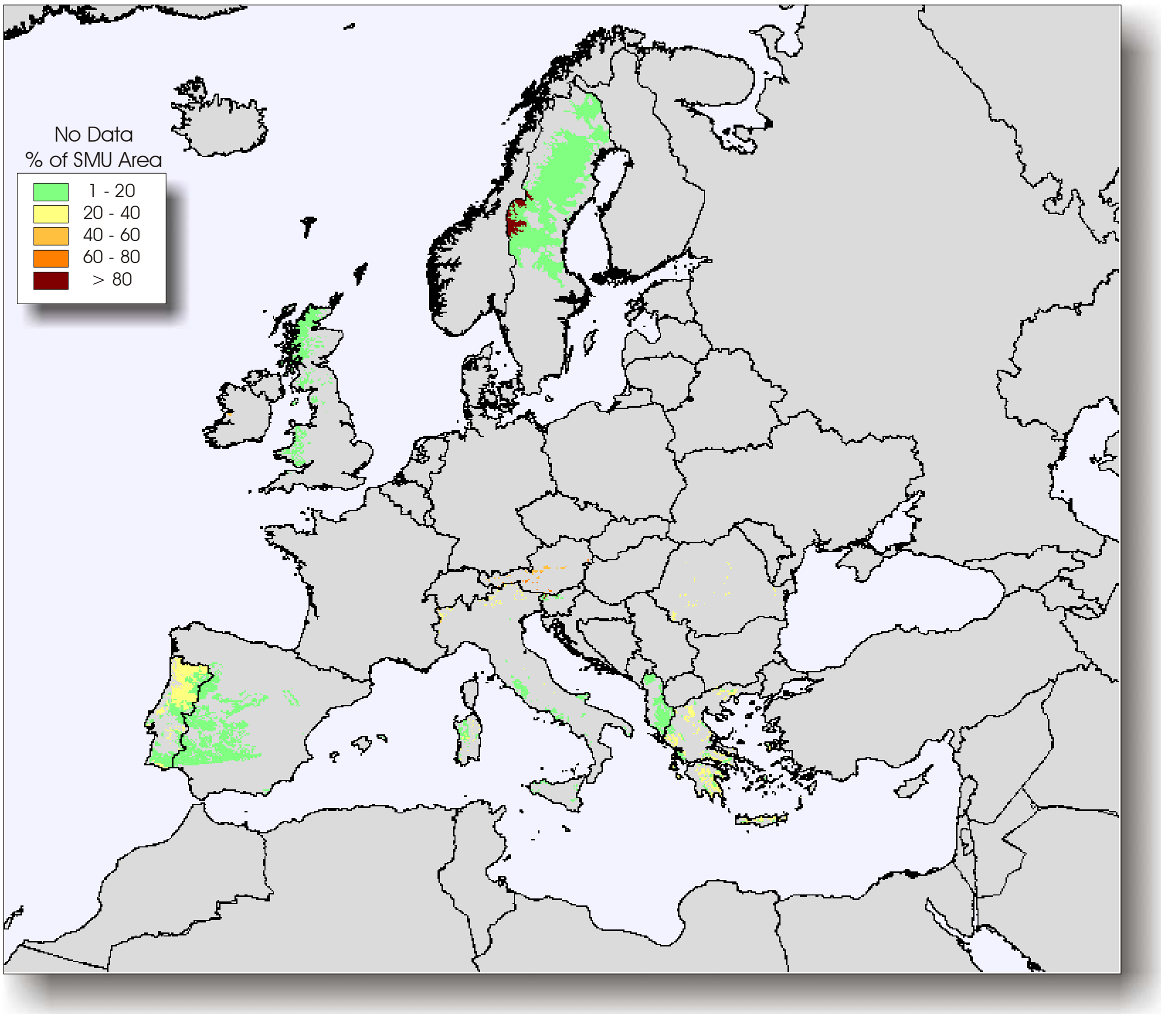


Figure 7: Proportion of the Area of SMU Linked to STUs with Codes Indicating Absence of Soil Data

The method applied to extend soil typologies to areas without such information uses the morphological similarities of neighbouring areas and ancillary data with complete cover of the surface area. For the land surface areas without soil data only the classes of urban areas (Class 1) and soils disturbed by man (Class 2) concern soil types. All other classes are likely to be left out from an analysis of soil properties. In the SGDBE missing soil data for urban areas and those disturbed by man are mainly located in relatively flat terrain and along rivers. This indicates that the soils could be estimated from extending SMUs along a flow network and corresponding topographic characteristics were used.

The data were processed in three consecutive steps:

- generate characteristics of topology
- establish relative proximity by friction over a cost surface
- assign undefined area to nearest SMU by relative proximity

Characteristics of the topology can be derived from a *digital elevation model* (DEM). The DEM used in this evaluation is based on the data obtained from the *Shuttle Radar Topography Mission* (SRTM). From the 90m SRTM DEM a product with 30 arc second grid spacing was derived as SRTM30[7]. For SRTM30 areas not covered by the instrument (below 56°South and 60°North) are filled in by data from the GTOPO30 DEM of Unites States Geological Survey[8]. Version 2 of SRTM30 was released in February, 2005 and was used in the evaluation[9].

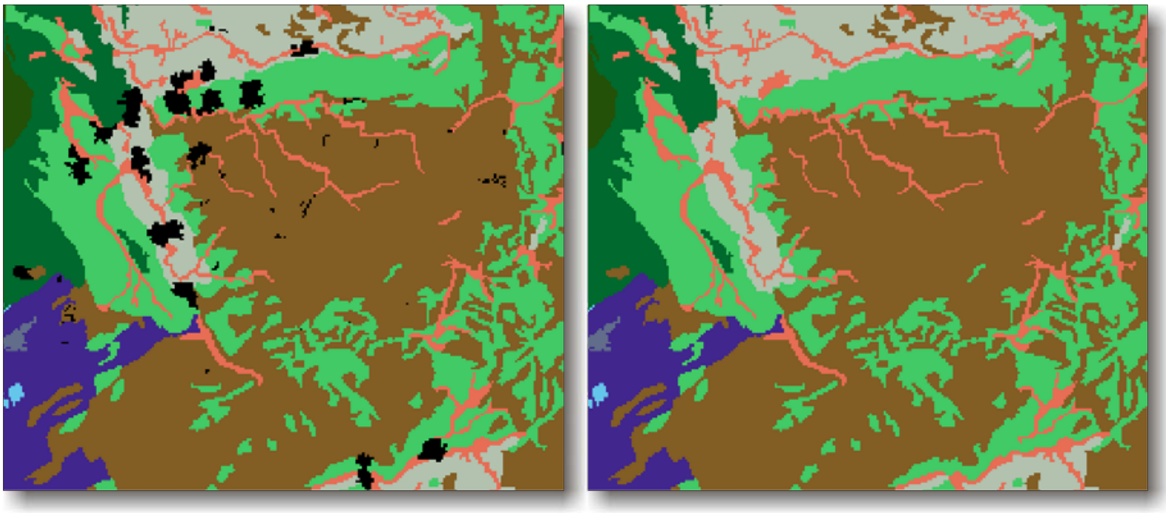
The spatial variability of the original elevation data very much reduces the use of the DEM in any analysis which tries to classify a location based on an assessment of neighbourhood statistics. The amount of neighbourhood variability was therefore reduced by using a Fourier analysis and removing the high frequency variability with a Gaussian filter. From the filtered DEM layers for slope and aspect, a flow network and the profile curvature are extracted. The flow network was derived from the DEM by the accumulated flow with a minimum set to an area of 25km^2. All parts of the flow network were set to a value of 1 and insert into the slope layer. The resulting layer was used as the friction magnitude for the cost analysis. The friction direction was derived from the aspect of the friction magnitude layer, where the direction was inverted under the assumption that soils are more likely to be similar down slope than upslope. A layer of maximum change in curvature served to define boundaries between areas with different slope and was used as an isotropic friction surface in the cost distance calculation.

Proximity was determined by the cost of moving over the three friction surfaces using an anisotropic analysis, i.e. where frictions can have different effects in different directions. The procedure fills in non-defined land areas only from neighbouring features. It also uses only the topographic information immediately adjacent to the data to be estimated. Therefore, it may be applicable to estimate features over small areas with spatial continuity, but for larger areas using a classification procedure is likely the better option.

An example of the result of the operation of filling in areas without soil data from adjacent features is shown in Figure 8. The extract shows the Ruhr area in Germany and goes as far south as the Main.

---

[7] Homepage: http://www2.jpl.nasa.gov/srtm/
[8] Homepage: http://eros.usgs.gov/#/Find_Data/Products_and_Data_Available/gtopo30_info
[9] Download page: http://dds.cr.usgs.gov/srtm/version2_1/SRTM30/



a) SGDBE SMU Coverage *b) Coverage after Filling SMU Areas*

Figure 8: SMU Cover before and after Filling Procedure

The areas of missing soil data, mainly urban areas, are shown in black on the map. On the whole, soil units are extended along river valleys where the topology is well expressed. In areas with a less well-defined topology the soil unit along the river is extended along the valley for some pixels, but then gives way to the dominant soil unit. On closer examination one reason for a feature not being extended along a river seems to be divergence geometry between the elevation data used to guide the extension and the soil unit data. On more undulating ground a shift of 1 or two pixels between the layers can already be responsible for a discontinuity in the extension of the soil unit.

The results from using the anisotropic procedure were compared to using an isotropic procedure (cost growth) for friction surfaces. In total a different feature was assigned to 5,239 to 7,871 grid cells or 4.3% to 6.4% of all grid cells without soil data, depending on the friction surface used. There seemed to be better agreement between the results when using the same isotropic friction surface in both procedures. The procedures were only compared by a subjective visual interpretation of the features connected along the flow network. No clear distinction could be found between the procedures. A more systematic evaluation would have required either reference data for the areas estimated or testing the procedures for other areas with similar conditions and with data. This work was not performed, because the aim of the processing step was not to generate a feature map for soil properties, but to allow using land use / cover data from other sources without creating artefacts due to differences in the delineation of non-soil areas. Therefore, specific note should be taken that the SMU layer with continuous feature cover has to be combined with layers on non-soil surface cover types from a land use / cover data set, at least covering the SMU feature layer of water surfaces, bare rock and glaciers or permanent snow fields.

2.4.7 Completion of Soil Classification Data

Features without data are not only present in the spatial layer when removing SMUs with land cover types, which are also found in land use / cover data, but also by missing information on soil properties in the attribute tables. Such missing information for STUs leads to an incomplete description of the associated SMUs. As a minimum requirement one would expect an STU to contain at least an entry for the soil type in one of the fields for the soil classification schemes for STUs related to soil rather than land use / cover. This lack of data occurs for a number of STUs where a soil classification code would be expected. Soil codes are given by three types of classifications schemes, yet none of the data for the soil classification schemes cover all STUs. Of the 5,262 STUs in the table [STU_SGDBE] soils are defined for the three soil classification schemes as given in Table 7.

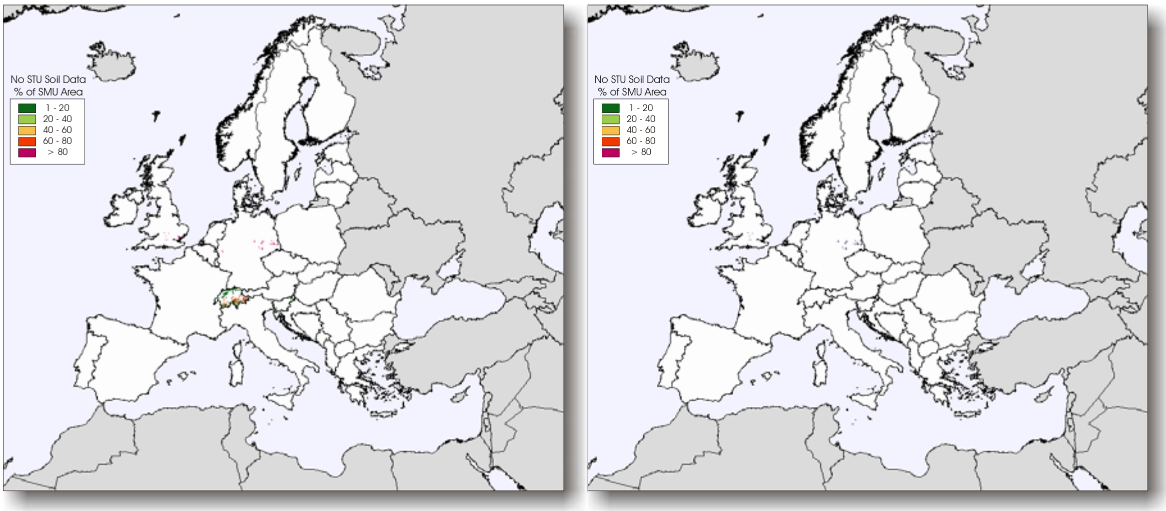
Table 7: Spatial and Typological Units with Missing Soil Classifications Codes

Soil Data	Soil Classification Field					
	WRBFU		FAO90FU		FAO85FU	
	AOI	*Total*	*AOI*	*Total*	*AOI*	*Total*
STUs without data	15	15	3,025	3,025	7	17
STUs with land use / cover Type 1, 2, 3, 4, 5, or 6	48	60	48	60	48	60
SMUs linked to STU(s) without data	20	20	832	3,025	12	409
SMUs linked to STU(s) with land use / cover Type 1, 2, 3, 4, 5, or 6	51	405*	51	832*	51	405*

* Area of 1 SMU is 0.
AOI: Area of Interest

The most complete cover with soil classification data is available for the FAO85 classification scheme. Of the 5,151 STUs linked to SMUs in the Area of Interest (AOI) 7 have no entry in the corresponding field. Affected by the blank entry in the STUs are 51 SMUs. For the entries in the WRBFU field 15 STUs have blank entries, affecting 20 SMUs. Most data are missing for the FAO90 classification (3,025). Although the difference between the STUs with blank entries in the WRB and the FAO90 classification is 8 the number of STUs where a blank entry is present in both fields is 12. For 9 of these STUs a classification according to FAO90 is available, but for 3 STUs, all located in the AOI, a blank entry is found in all fields for soil classification codes.

The number of STUs without soil data does not provide a reliable indicator of the area affected. For example, for WRB and FAO85 the number of STUs without soil information is quite comparable, although the number of SMUs affected is completely divergent. The areas affected by missing soil type information also largely different between classification schemes. The areas concerned by missing WRB and FAO85 codes are graphically shown in Figure 9.



a) SMU Area not described by WRB   *b) SMU Area not described by FAO85*

Figure 9: SMU Areas not Described by Soil Classification Code in STU Table

The surface area for which no soil code is given according to the WRB classification scheme is 6,426 km^2 in AOI. The area without data for the FAO85 classification scheme is 2,663 km^2 in AOI. The area for which all three field for soil classification schemes are blank covers 899 km^2.

Because the most complete information is available for the FAO85 soil classes and because the parameters used to define the rules of the PTR database frequently use FAO85 codes to specify soil properties it would appear reasonable to complete the missing soil information for this classification scheme. Where possible the FAO85 soil classis is defined based on information provided by the other classification schemes by corresponding codes.

FAO85 soil classification codes are only for the 12 STUs without such data. The class is derived from the correspondence where data for the FAO85 and at least one other classification scheme are available. Since there is no 1:1 relationship between soil classification schemes identifying a corresponding soil class can lead to ambiguity. For this study a measure to identify the correspondence was based on the frequency distribution obtained from a pair-wise comparison of soil classes. The selection of a FAO85 code was based on the entries in the fields [PARMADOM], [PARMASEC], [MAT1], [AGLIM1], [TEXTSRFDOM] and [IL]. In case some correspondence between existing data pairs and the properties of the missing STU could be established the most similar FAO85 soil class was assigned to the STU. A clear correspondence to the parent material could only be established for 2 cases. For other situations and where no particular similarity could be established, mainly due to a lack of data in the ancillary fields, the most frequently occurring class from the pair-wise combinations was assigned. The option of using the majority class would seem to reduce the risk of an unsuitable class being assigned. However, the reason for not assigning a soil code may have been the lack of information, which in turn may be caused by specific conditions of the STUs concerned and hence the assumption of reducing the risk of a wrong assignment would not hold. Other than numerically the combinations of soil classes for the three classification schemes in the STU database were not evaluated.

The list of completed FAO85 soil classes is given in Table 8.

Table 8: Completion of FAO85 Codes in STUs

Combinations			Assigned
STU	**WRB**	**FAO90**	**FAO85**
Code	*Code*	*Code*	*Code*
3860002	-	ATa	Dc
3860003	-	ATa	Dc
3860004	-	ATa	Dc
490218	-	ATc	Io
70016	ABeun	PDd	De
70004	CMca	CMc	Ba
70005	CMhu	CMd	Bd
70009	CRca	GLk	Gc
70011	HSge	HSi	Ox
70014	KScc	KSk	Kk
70015	LPum	LPd	Uo
70019	RGha	RG	R
70020	SNha	SNh	So
70021	SNha	SNh	So

A soil code according to the FAO85 classification scheme could be assigned to 14 of the STUs with missing data in the field. For the STUs with a blank entry in all classifications a different procedure was applied.

2.4.8 Estimation of Soil Typology in Absence of Soil Classification Codes

For the three STUs without information on the soil class in all fields (STUs 4210009, 4401565 and 4402004) also the ancillary fields used to find some correspondence only contain "0" entries. There is thus not much information available in the STU table to identify a soil class for these STUs and other sources have to be exploited.

The three STUs are linked to just three SMUs. Two of these STUs are the only linked typological units for two SMUs (SMU: 4404522 and 4405072), both are located in England. The remaining STU without soil class information is linked to an SMU in the Czech Republic (SMU: 4210003), which is defined to 70% by the STU. The location of the SMUs is given in Figure 10.

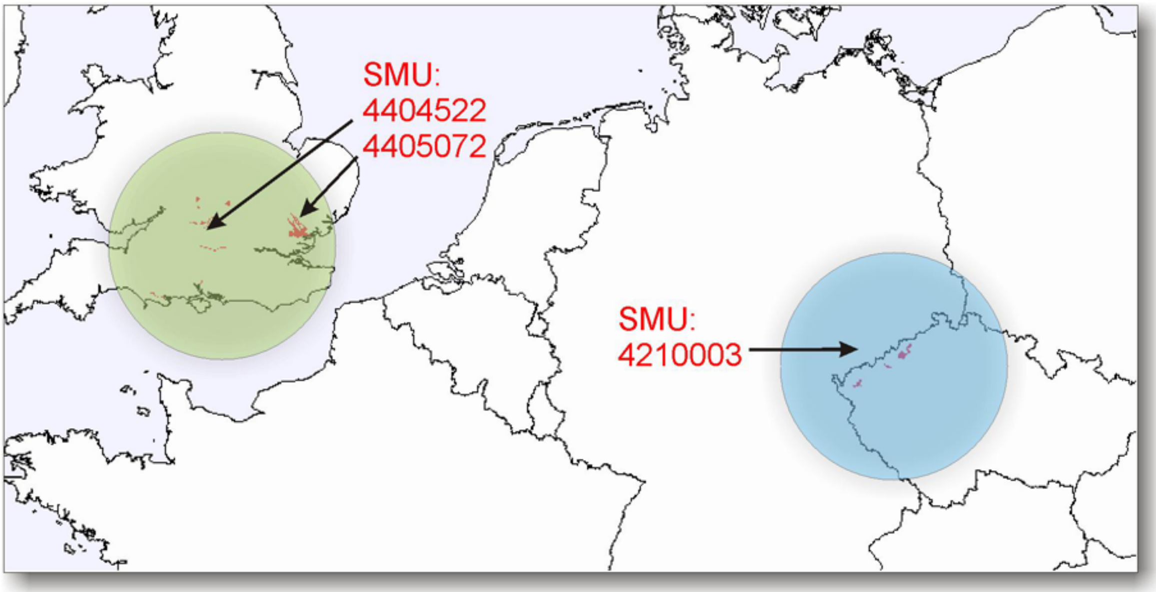


Figure 10: SMUs Linked to STUS without Soil Classification Code

The figure shows that the SMUs are not continuous, but spread across several spatially separate areas. A summary of the STUs without soil classification data and the SMUs affected is given in Table 9.

Table 9: Spatial and Soil Typological Units without Soil Classification Code

SMU	STU	Undefined Area	
		of SMU (%)	*km²*
4210003	4210009	70	173.4
4404522	4401565	100	197.8
4405072	4402004	100	526.5

The SMU in the Czech Republic is defined by 30% with the FAO85 soil code "R" (*Regosol*). The area of the neighbouring SMU 4210017 is defined to 30% by the FAO85 code "Be" (*Eutric Cambisol*). It was assumed possible that SMU 4210003 had the same composition as SMU except for the on component. An alternative assumption is that the missing soil typological data could be provided by the single STU not linked to any SMU, which is also located in the region. However, the elevation range given for the STU (800 to 2,400 m) does not correspond to the elevation of the SMU (approx. 200 to 500 m).

Consequently, SMU 4210003 was defined as the neighbouring SMU with the exception of exchanging the 30% covered by the *Eutric Cambisol* (STU 4210151) with the defined *Regosol* (STU 4210010).

For the two SMUs without soil classification data in England the information from similar areas was used as an approximation of the soil typological properties of the areas. In both cases a similar area is defined by SMU 4405071. The substitute SMU is linked to 8 STUs and may not be fully representative for the undefined SMUs, but was considered a likely composition.

For these changes the links in the table [STUORG] were modified to point to the substitute STUs.

3 IMPROVE SPATIAL POSITIONING OF ATTRIBUTES

The principle behind the process of adding a spatial component to the STUs is based on the assumption that the attribute information is sufficiently distinct to separate STUs within an SMU and that the attributes can be associated with ancillary spatial data to allow the STUs to be allocated to geographic location within an SMU. The method uses a multi-objective/multi-criteria evaluation decision support analysis (Triantaphyllou, *et al*., 1998). In an adaptation of the *multi-criteria evaluation* (MCE) method the customarily evaluated suitability of an objective for a location is replaced by the association of STU characteristics with ancillary spatial layers. The MCE objectives are given by the specified areas of STUs within linked SMUs. The decision for an objective is based on the evaluation of criteria. Where multiple criteria are defined the process combines these criteria into a single indicator to describe the closeness of the STU characteristics for a given location. For the ESDB there are generally several STUs linked to an SMU. For the actual geographic allocation of STUs a *multi-objective land allocation* (MOLA) procedure is applied. For objective competing for the same geographic location the MOLA tries to maximize the closeness of the STUs characteristics to geographic positions for the ensemble of STUs.

3.1 Multi-Criteria Evaluation Objectives

Decision support systems have been used for some time to provide a basis for making a choice between alternatives. One implementation of supporting decision-making using spatial data layers was developed under the project *Multi-sectoral Integrated and Operational decision support system for sustainable use of water resources at the catchment scale* (MULINO[10]; Giupponi, *et al*., 2004; Fassio, *et al*., 2005).

A decision is made following a set of rules, which are structured by objectives. In the case of improving the spatial positioning of attributes from the STU database the objectives are to allocate all STUs linked to an SMU to the most plausible geographic location. There are therefore as many objectives as there are STUs linked to an SMU. The objectives are in competition for a location, because the SMU area is limited. Conflicts in the choice of objectives are resolved in the MCE by ranking the level of affinity[11] of an objective with a location. The affinity between objectives and locations

[10] MULINO homepage: http://siti.feem.it/mulino/index.htm

[11] In this study pre-determined soil typologies with fixed proportions are arranged within a set area. This mode is quite different from assigning potential soil types according to landscape parameters or allocating a limited amount of land in a larger area. To avoid confusion with the latter modes of using MCE in this study the term "affinity" is preferred to "suitability".

is determined by a set of defined criteria. The choice is made through the evaluation of these criteria.

The score for objectives is based on a set of criteria, which define how closely an objective is related to a geographic location Criteria used in the decision process are divided into those defining affinity between the objectives and the geographic location (factors) and those that exclude areas from objectives being allocated (constraints). While criterion factors describe the relative affinity with or match of an STU to certain areas in the SMU constraints describe areas of incompatibility. In the GIS used for this study factors may be expressed on a continuous scale, but constraints are of data type Boolean. Criteria constrains, therefore, act as geographic masks on which the criterion factors take effect or which completely exclude areas from the allocating process.

3.2 Feasible Parameters for Use as Criteria

An STU is defined through a combination of a series of typical properties which separates one STU from another. The table [STU_SGDBE] defines 5,262 STUs by a unique combination of parameters defined in 34 attribute fields (plus 1 field for STU ID). The attributes describe very different categories of properties, as summarized in Table 10.

Table 10: Property Categories of Soil Typological Unit in Table [STU_SGDBE]

Property		**[STU_SGDBE] Table**
Category	*Sub-Category*	*Fields*
Spatial Unit		NB_POLYS, NB_SMU, AREA
Position	Topography	SLOPE_DOM, SLOPE_SEC, ZMIN, ZMAX
	Land Use	USE_DOM, USE_SEC, AGLIM1, AGLIM2
Soil	Soil Class	WRBFU, FAO90FU, FAO85FU
	Parent Material	PARMADO, PARMASE, MAT1, MAT2
	Texture	TEXTSRFDOM, TEXTSRFSEC, TEXTSUBDOM, TEXTSUBSEC
	Depth	ROO
	Profile	TEXTDEPCHG, IL
	Water Regime	WR, WM1, WM2

Fields on confidence levels not included.

In the table three main categories are distinguished. Attributes related to the spatial unit are not concerned with typological properties. The attributes related to the position of the typological unit in the landscape are only indirectly linked to soil typology. Included in the category were the fields containing the limitation to agricultural use of the STU, because it is frequently linked to topography. The remaining field cover various aspects of the soil typology.

Of the total of 5,262 STUs 5,130 differ in one or more attributes other than the fields related to the spatial unit or confidence level. Excluding units with non-soil entries in the soil classification fields a unique combination of attributes is found for 5,077 STUs. In this assessment STUs which differ by the combination of missing data are included.

For an improved spatial positioning of STU within the SMU the information on the position of the STU within the landscape can be used. The term landscape is used here to express aspects of topography and land use / cover. Information on slope and elevation originate form a single source of data, with slope being the change in elevation over space. Both can be related to soil type and can either facilitate or restrict the distribution of some soil types to certain topographic conditions, such as fluvial plains. Land use / cover can be related to elevation and slope, for example restricting agricultural land use on steeper ground. The soil type can influence the distribution of land use / cover through the water regime or available depth for roots.

The interpretation of the morphological and land use attributes to a landscape is graphically presented in Figure 11.

Attribute db

SMU ID	STU ID	PCAREA	FAO85	ZMIN	ZMAX	SLOPE_DOM	SLOPE_SEC	USE_DOM	USE_SEC	PEAT
3700004	3700015	5	Lk	30	130	2	1	3	1	N
3700004	3700016	5	Oe	20	100	1	1	1	19	Y
3700004	3700014	15	Gc	20	100	1	1	3	1	N
3700004	3700013	75	Bg	30	130	1	1	3	1	N

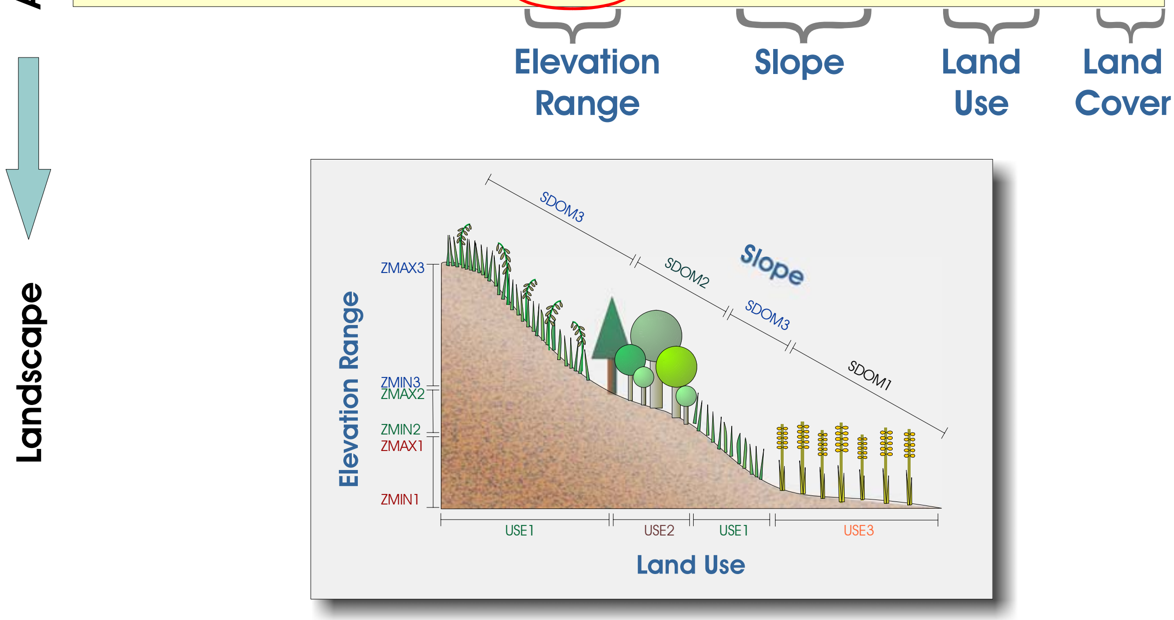


Figure 11: Positioning STUs within Landscape of Mapping Unit

Ancillary data on topography and land use / cover in form of spatial layers are available from several sources. The relationship between soil properties and surface features of the STUs, which define a specific SMU, and generalized relationships derived from all SMUs, may be utilized to geographically position an STU inside an SMU according to the distribution of the characteristic in the ancillary spatial data. The relationships between soil attributes can be described by functions, but the association of soil classes with landscape properties is often defined through rule systems based on expert knowledge.

Requirements for the use of ancillary information are a thematic equivalence to the criteria used in the analysis and spatial representation at a scale at least comparable to the resolution of the spatial unit of the SGDBE. Thematic equivalence also includes temporal correspondence, where appropriate. For example, the SGDBE contains information on land use for the STUs. In particular land use varies over time and the parameter is only useful in the MCA if both, the soil data and the land use map were surveyed at the same time or when applied to areas without land use changes.

The potential of spatially allocating STUs within an SMU is estimated by the number of unique links of the attributes with ancillary spatial coverage (slope, height and land use) to the SMUs. A summary of the links is present in Table 11.

Table 11: Shortfall of Combinations of STU Attributes Height, Slope and Land Use to Spatially Allocate STUs in SMUs

No. of STU – SMU Links	SMUs with Link No.	Parameter							
		SLOPE DOM	SLOPE SEC	ZMIN	ZMAX	USE DOM	USESEC	USEGEN DOM	USEGEN SEC
Links	*Count*	*Count*	*Count*	*Count*	*Count*	*Count*	*Count*	*Count*	*Count*
1	101	825	714	1283	1177	917	822	919	833
2	282	567	562	164	173	414	479	413	482
3	449	130	233	76	124	188	181	186	178
4	357	19	32	14	59	21	52	17	45
5	191			4	5	1	7	1	3
6	123				2				
7	24				1				
8	13								
9	0								
10	1								

No. of STU – SMU Links	Slope	Elevation	Land Use / Cover	Categorized Land Use / Cover
Links	*Count*	*Count*	*Count*	*Count*
1	635	1099	819	635
2	542	157	333	542
3	276	123	173	276
4	72	83	74	72
5	11	27	17	11
6	1	6	1	1
7		1		
8				
9				
10				

Shortfall of Conditions	All data		Excluding missing entries in all parameters	
Count	*Count*	*% of SMU*	*Count*	*% of SMUs*
0	620* (526**)	40.2* (36.4**)	585* (500**)	38.0* (34.6**)
1	495	32.1	475	30.8
2	273	17.7	273	17.7
3	99	6.4	97	6.3
4	34	2.2	31	2.0
5	13	0.8	14	0.9
6	6	0.4	6	0.4
7	1	01	1	0.1

* *Total number of SMUs: 1,541 in Area of Interest*
** *Number of SMUs with more than one linked STU: 1,447 in Area of Interest*

For the 1,541 SMUs in the AOI the majority (449 SMUs) are linked to three STUs. A 1:1 relationship is defined for 94 SMUs and more than the majority of links (4 to 10) are defined for 46% of the SMUs. The variability of a single STU attribute in characterizing unique conditions for the links with an SMU is largely unspecific for all attributes. Except for the secondary slope for more than half of the SMUs the linked STUs have the same attribute values. For these SMUs the information provided by a single attribute is insufficient to support the spatial allocation of the linked STUs. When combining the attribute dominant and secondary slope, minimum and maximum height and dominant and secondary land use the variability in the attributes increases. For the slope parameter the number of SUMs with only one value decreases from 825 (53.5%) for the dominant slope and 714 (46.3%) for the secondary slope to 635 SMUs (41.2%) for the combination of the two attributes.

The spatial distribution of the shortfall in the number of SMU:STU links with divergent STUs characteristics for all three parameters is presented in Figure 12.

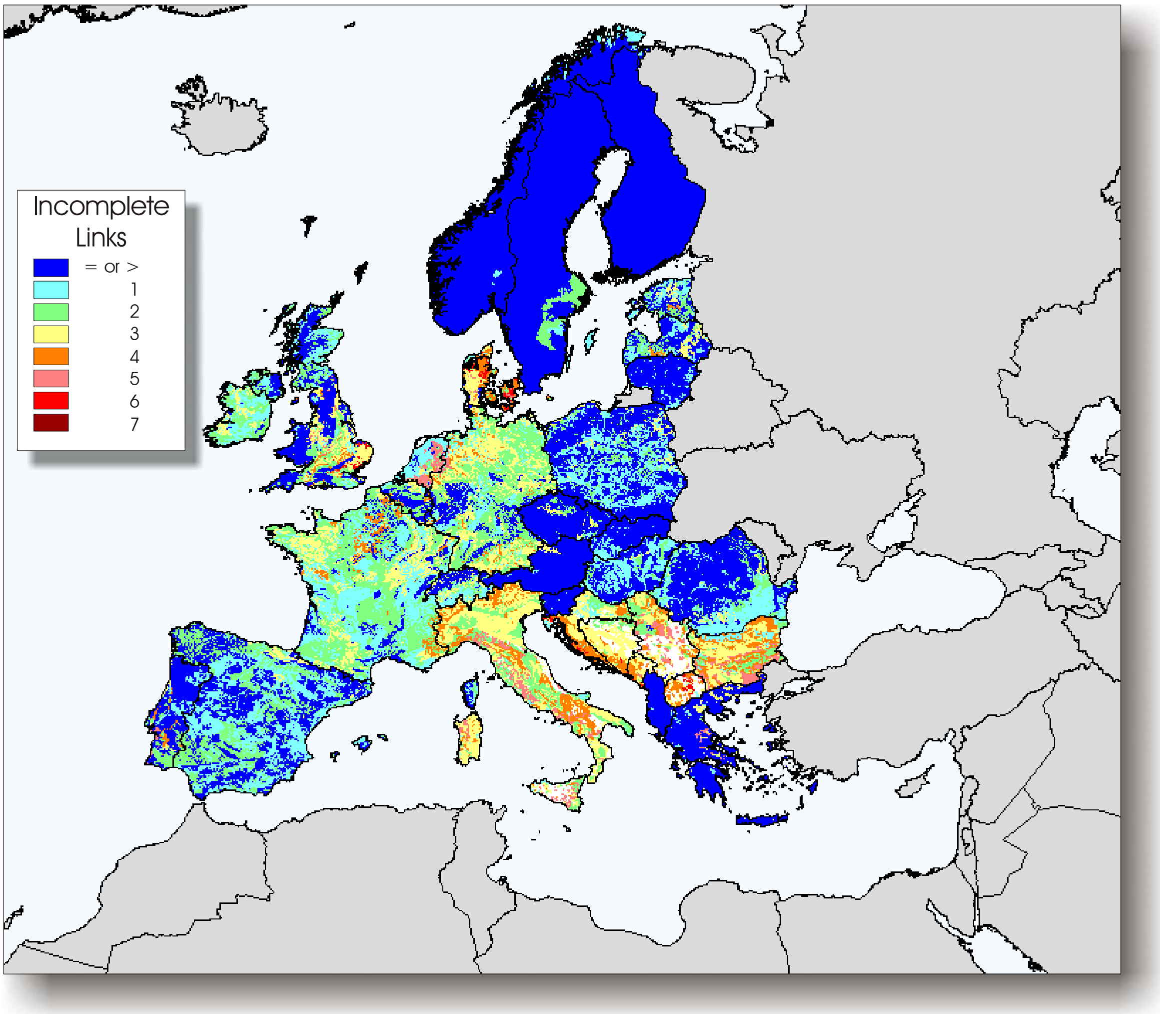


Figure 12: Shortfall of Disparate SMU:STU Links from Total No. of Links

When using all attributes 40.2% of the SMUs are characterized by as many conditions as there are links to STUs. For 32.1% of the SMUs the number of unique conditions of the attributes is one less than the number of links. In this calculation missing data are interpreted as a value different from an existing value. This treatment of missing data as containing some information (parameter not present) is not necessarily the correct treatment for all cases.

In the database there is no distinction between a situation when a parameter is not present and when it is not known whether it is present or not. Therefore, not in all cases can the denotation of missing data be unambiguously identified. This is the case for the secondary characteristics, in particular land use. The land use legend does not separate between instances of "no data" and "no secondary land use". Because the field has no empty entries it appears that where no information on the secondary land use was available the code "0" was set. Also, the secondary land use is set to equal the dominant land use for 43 of the 5,465 STUs linked to SMUs in the AOI.

When excluding missing data in all attributes the number of characterized SMUs is 585 or 38.0% of the SMUs in the AOI. To be consistent the portion of SMUs where missing data were excluded from the STUs is calculated over the total number of SMUs, not the number of SMUs without missing data for the attributes. When evaluating the potential of using the attributes for the spatial allocations of STUs it should be considered that in the AOI 94 SMUs are linked to a single STUs and therefore do not qualify for a spatial allocation.

From the evaluation of characterizing SMUs by slope, height and land use the shortfall of the shortfall of unique conditions very much restricts the spatial allocation of STUs. For a comprehensive spatial allocation of STUs more conditions characterizing the SMUs are needed. Increasing the number of conditions can be achieved by using more attributes and additional qualifiers for the attributes with ancillary spatial datasets.

Using additional attributes to improve the characterization of the STUs within an SMU is limited by the availability of suitable ancillary spatial data. For none of the other attributes such dataset were identified. As potential qualifying information for the slope parameter the attributes for soil depth (obstacle to roots [ROO], depth to impermeable layer [IL], water regime [WR] and depth to rock [DR]) were considered. Within a given area deeper soils and longer periods of wet conditions are generally associated with level slopes, while shallower soils and drier conditions tend to be positioned on steeper slopes. The relationship between obstacles to roots and the presence of an impermeable layer with slope may be more complex.

3.3 Decision Criterion Factors

The number of criteria, which could be derived from the elevation, slope and land cover/use parameters, was found to be limiting when solving conflicts of competing objectives for a geographic position, in particular for mapping units in Eastern Europe. Therefore, more generalized criteria had to be introduced to provide some measure of spatial differentiation for the typological units. Where such information is not available the typological units can still be assigned to geographic positions, but with a lower level of credence.

The primary criterion factors with scope restricted to a specific SMU are:

- STU Height ([ZMIN], [ZMAX])

- STU Slope ([SLOPE_DOM], [SLOPE_SEC])
- STU Land use ([USE_DOM], [USE_SEC])
- STU Depth ([ROO], [IL], [WR], [DR])

The available STU parameters characterizing an SMU may not be sufficiently distinct to define the position of the STUs within the SMU. Lack of distinctness can be caused by values covering largely overlapping ranges or by missing data for a parameter. To provide additional guidance for the allocation of STUs and avoid ambiguity caused by missing data factors of the mean or most frequent values were extracted from the STU database.

Secondary criterion factors are aggregations derived from all STUs:

- Mean height by FAO 85 soil class
- Mean dominant slope by FAO 85 soil class
- Mean sub-dominant slope by FAO 85 soil class

As a third level to resolve confusion in the position of the STUs individual STUs were mapped to the SMU and mean value for a topographic parameter were extracted from the DEM.

Tertiary criterion factors extracted from ancillary data for soil class are:

- Mean height in DEM by FAO 85 soil class
- Mean slope in DEM by FAO 85 soil class
- Distance to drainage network

Tertiary criteria are fairly general and unspecific as regards the geographic distribution or the relationship with other factors.

3.3.1 Criterion Factor Height

The height criterion refers to the ratio type values of the fields [ZMIN] and [ZMAX] of the typological database. The likely fuzziness in the parameter limits are accounted for by allowing for a deviation of 10m either side of the defined range.

To be useful in the spatial allocation the limits of the height parameters should cover a range of values, i.e. [ZMIN] < [ZMAX]. For the parameters the indicator for missing values is "-999". For the AOI missing height data for both parameters is found for 130 STUs. In one STU (STU ID: 400274) only the minimum height is given with a value for missing data for the maximum value. The missing maximum height was replaced by a value of 125m, which is estimated from the average difference between the maximum and minimum height (99.25m) for the soil type (*Hh*) for minimum heights between 10 and 100m.

More difficult to deal with are the 97 STUs where the minimum height equals the maximum height. Affected are 34 SMUs. Of these, one is linked to STUs which cover a range of height limits, while the other 33 SMUs are only linked to STUs with equal values for the two height parameters. An additional element of ambiguity is introduced by the STUs linked to an SMU having the same equal values, i.e. there is no variation in the height parameter for the whole SMU. In the AOI 88 SMUs are linked to a single STU with soil data. SMUs with links to more than one STU and variations in the height parameter are 398. Therefore, for 1,005 of the 1,491 SMUs with height data the values for minimum and maximum height do not differ between the linked STUs. As a consequence, for these SMUs the height parameter does not support the spatial allocation of STUs.

The presence of identical STU height values is contradicted by the values given for the dominant slope class, which for 31 STUs is either > 1 or 1 for the dominant slope and > 1 for the sub-dominant slope. As a consequence, for these SMUs the values for the height parameters are of no use to solving conflicts on the spatial allocation of STUs and the entries could be treated as missing data.

In the amended data the one typological parameter always specified is the code for the FAO85 soil classification scheme. Therefore, the means of the minimum and maximum height were defined for each soil class where the STU contained valid data.

3.3.2 Criterion Factor Slope

For the slope criterion the attribute database only contains ordinal values. Missing data are indicated by the value “0”. Within the STUs linked to an SMU of the AOI 66 STUs contain the indicator for missing data in the field [SLOPE_DOM] and 2,274 in the field [SLOPE_SEC]. For 7 STUs the dominant slope is 0 while the secondary slope has a non-zero entry, which is “4” for 6 STUs and “2” for one. Thus there are 59 STUs without slope data in either field.

It is not obvious from the data how entries for missing data for the secondary slope should be interpreted. Several interpretations are possible: for example a missing value may indicates that no secondary slope exists which differs from the dominant slope, or that a secondary slope was not be identified. There are also 181 STUs where the categories for the dominant and secondary slope category are equal. To provide a more uniform set of data for the secondary slope in the STU table all values for the secondary slope were set to the category of the dominant slope where a missing value was indicated.

- **Representation of Slope in Spatial Data**

  While the DEM elevation data generally corresponds with the STU height data a significant difference was found between the DEM and the STU data for terrain slope. The proportions of the distribution of areas for the slope classes was found to differ notably between the data sets. The general situation is summarized in Table 12.

Table 12: Distribution of Dominant Slope Classes in STU and DEM Data

SLOPE	STU		DEM*		
Dominant	**Proportion**	**Class Mean**	**Proportion**	**Class Mean**	**for STU Area Share**
Class	%	%	%	%	%
0 - 8	43.9	4.0	88.6 82.8	1.8 2.1	0.47 0.63
8 - 15	30.1	11.5	7.7 9.7	10.8 10.9	2.10 2.92
15 - 25	19.5	20.0	2.9 4.9	18.8 19.1	6.45 9.10
>25	6.5	35.0	0.7 2.6	30.6 34.1	17.60 25.47

* First row: filtered DEM
Second row: non-filtered DEM

The table shows that 43.9% of all STUs (area-adjusted) have slopes of Class 1, i.e. < 8%. In the DEM the proportion of the area belonging to Slope Class 1 is 88.6% (82.8 in non-filtered DEM). The central value of Slope Class 1 is 4.0% while the mean slope of areas < 8% is 1.8% in the filtered DEM (2.1 in the non-filtered DEM). For the allocation of STUs in the SMUs the share of the STUs in the database should be comparable to the share of the criterion on the spatial layer. The last column on the table gives the mean slope in the DEM for the share of the STUs in a slope class using a ranking procedure for slopes. The mean slope for the first 43.9% of ranked slopes in ascending order in the DEM is 0.47% (0.63% for non-filtered DEM). The mean slope for the next 30.1% is 2.10% (2.92%) and 6.45% (9.10%) for the next class. As a consequence, when allocating STUs by the slope class given in the STU table the procedure may very rapidly run out of suitable areas in the DEM for that slope class regardless of the filter option used for the DEM.

Some examples of situations frequently found are presented in Table 13.

Table 13: Examples of Distribution of Area by Slope Class in DEM Layer and STU Table

SMU	SLOPE		SHARE	PCAREA	
	Class	*Mean**	*DEM Area**	*[SLOPE_DOM]*	*[SLOPE_SEC]*
ID	*Code*	*%*	*%*	*%*	*%*
2317	1	2.48	86.16	45	70
2317	2	10.87	9.26	55	
2317	3	18.22	4.12		30
2317	4	28.80	0.47		
2320	1	3.51	81.02	5	15
2320	2	10.56	15.24	95	
2320	3	18.59	3.29		85
2320	4	29.41	0.44		
2380	1	1.72	99.79		
2380	2	8.61	0.21	95	
2380	3			5	95
2380	4				5

* for non-filtered DEM

In the first example the slope classes of the STUs (Classes 1 and 2 for dominant slope, 1 and 3 for secondary slope) are also present n the DEM layer. Areas with slopes according to Class 4 are found in the DEM, but with low cover (0.5% of SMU area). In the distribution of the areas belonging to the slope classes the data differ. The area with slope class 2 is almost twice as large in the DEM as the proportion given for the dominant slope of the STU with that slope class (86% vs. 45%). As a consequence, the area proportions also differ greatly for other slope classes. The proportion of the secondary slope is closer to the share of the class in the DEM (70%) and areas with higher slopes than class 2 are present.

The second example shows 81% of areas in the DEM as belonging to slope class 1, 95% of the dominant slope assigned to class 2 and 85% of the secondary slope assigned to class 3. Each slope parameter shows strong dominance for one class, but none of the STU values agree with the DEM data.

A situation similar to the one of the second case, but more extreme, is found for the third example. In the DEM layer almost all areas belong to slope class 1, but none of the STUs are assigned to this class. In the STU data the dominant slope is assigned with 95% to Class 2 and the secondary slope with 95% to class 3. Yet, areas with slope classes 2 are hardly found in the DEM and areas with Class 3 or 4 are not present.

The differences between the data in the typological database and the spatial layer can be evaluated by looking at the data from different angels. One approach is to calculate the mean SMU slope in the spatial layer for the interval defined by a slope class. This approach has been followed when comparing filtered with unfiltered DEM data. Another approach is to calculate the mean slope within an SMU for the area of the slope class given by the STU attribute [PCAREA.STUORG]. To find the area the slopes are ranked in the spatial data and the slope limit is increased until the area requirement is fulfilled. The mean slope is than extracted for this area. A comparison between the mean SMU slope when restricting the slope to the class range and to the class area is presented in Figure 13.

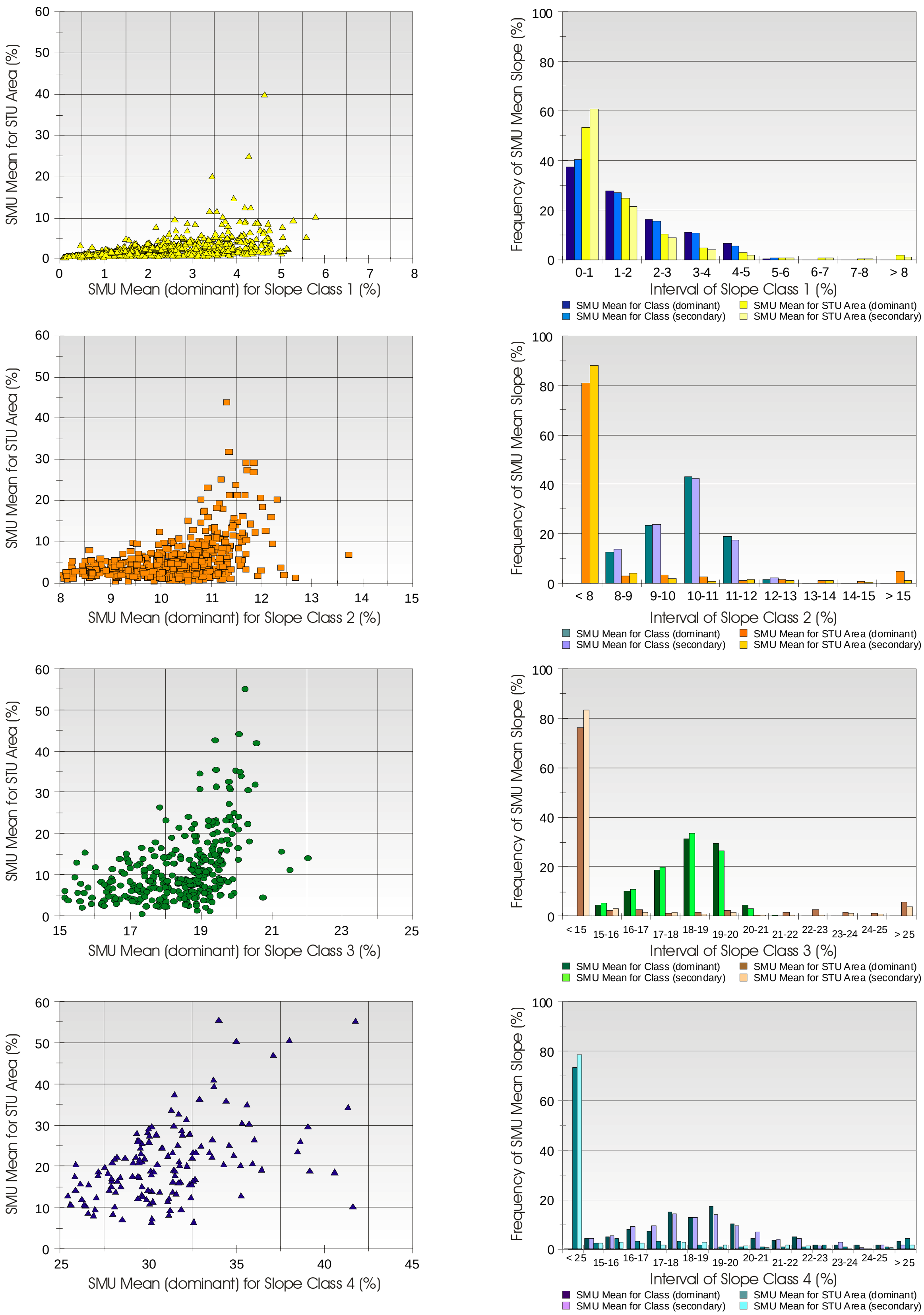


Figure 13: Comparison of Mean SMU Slope for Slope Class Range and STU Slope Class Area

The graphs on the left-hand side compare the mean SMU slope for range of slopes covered by a class (x-axis) with the mean SMU slope for the area of the class within an SMU (y-axis), only for the dominant slope. Except for Class 1 the range of mean slopes does not extend much beyond the mean slope for a class. The graphs on the right-hand show the relative frequency of the mean SMU slope for1%-intervals of a slope class. For Class 1 the frequency of the mean slope class decreases steadily with slope for both, the mean slope for the slope class and the STU area corresponding to the class. For all other slope classes the frequency of the mean SMU slope class indicates some variability around the general mean, while the mean SMU slope for the area is largely concentrated at the lowest limit of the slope range.

From the comparison it would appear that for a slope class the mean SMU slope tends to be lower than the central value of the class. There is, however, a strong difference between the area of a slope class and the mean SMU slope. As a consequence, the area covered by the range of slopes of a class is decidedly lower than the area indicated for the classes in the STU data.

- **Slope and Spatial Layer Processing**

To evaluate the effect of spatial resolution on the association between the STU slope attribute and the spatial resolution of the DEM the 90 m SRTM data was processed for the area covered. The version of SRTM used was the whole-filled seamless layer Version 4 available from the CGIAR-CSI SRTM 90m Database[12] (Reuter, *et al*., 2007). The mean slope was then extracted for each SMU. A comparison between the mean slope of the 1,000 m SRTM-derived DEM, the 90 m SRTM data and the STU attributes is presented in Figure 14.

[12] Jarvis, A., H.I. Reuter, A. Nelson, E. Guevara, 2008, Hole-filled SRTM for the globe Version 4, available from the CGIAR-CSI SRTM 90m Database (http://srtm.csi.cgiar.org)

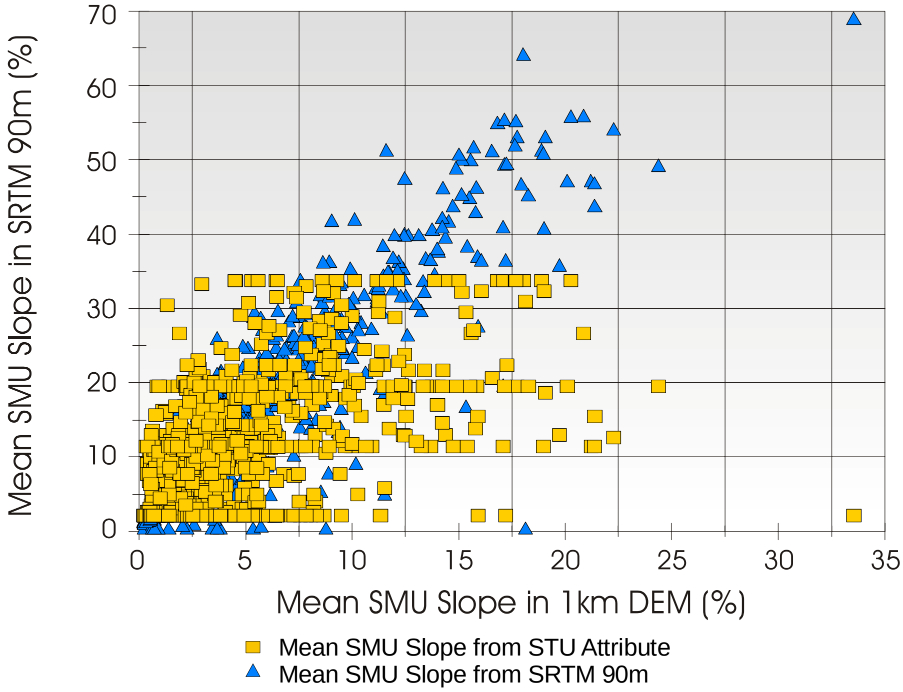


Figure 14: Comparison of Mean SMU Slope for 1km SRTM DEM with 90m SRTM DEM and aggregated STU Data

The graph shows a better correlation of the slope aggregated from STUs with the slope extracted from the 3 arc second (90 m) SRTM DEM than the 30 arc second (1,000 m) DEM. The x-coefficient of a linear relationship between the mean slopes (m = 2.9) is comparable to the x-coefficient of the relationship of the 1,000 m SRTM DEM with the aggregated STU values. This suggests that the association between STU and spatial layer for slope could potentially benefit from using the higher-resolution DEM or a 1,000 m layer with the characteristics of the 90 m DEM. Using the former option would introduce an inconsistency between the height and the slope layers, while using the later results in prohibitive processing times (data density > x 100).

However, the x-coefficient of the linear relationship between the mean slope for the SMUs between the processed 30 arc second and the 3 arc second SRTM data is not only a result of the resolution, but also of the filter applied. The Fourier filter is applied to improve the continuity of the flow network, but the resulting DEM has a reduced variability of the relief. The x-coefficient of the linear relationship between the mean SMU slope of the non-filtered 30 arc second and the 3 arc second DEMs re-sampled to 1,000 m is 1.06 (y-offset forced to 0) with a coefficient of determination of 0.99. When comparing the grid values instead of the mean slope for the SMU the linear relationship of the non-filtered 30 arc second and the 3 arc second DEM re-sampled to 1,000 m has an x-coefficient of 1.01 (r^2: 0.92). While the linear correlation of individual grid cells shows agreement in the gradient there is still some spread between the values, which is not evident in the mean SMU values, which are almost identical. The mean SMU slope of the 3 arc second DEM re-sampled to 1,000 m is related to the 3 arc second DEM at 100 m with a coefficient of 2.01 (r^2: 0.91). For the SMUs covered by SRTM in the AOI the relationships between the 30 arc second and the 3 arc second DEM for slope is presented in Figure 15.

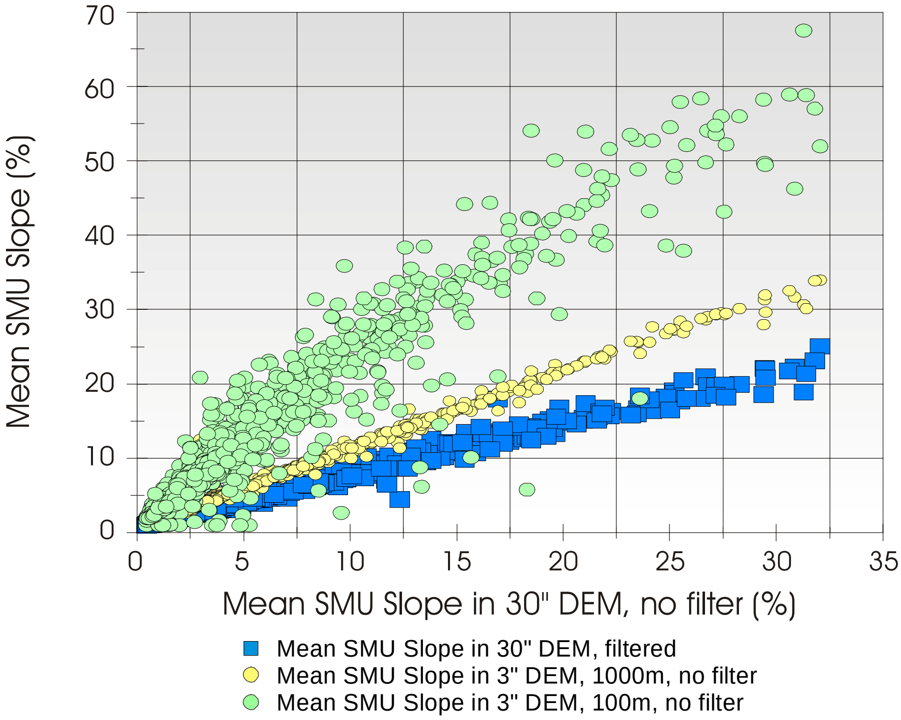


Figure 15: Comparison of Mean SMU Slope for 30" SRTM DEM with 3" SRTM DEM at 100 and 1,000 m Grid Spacing

Re-sampling the slope layer derived from the 3 arc second DEM does not significantly affect the mean SMU slope. Therefore, slope characteristics of an SMU depend on the resolution of the DEM used, the processing applied and the stage at which the slope is calculated with respect to re-sampling the layer.

Regardless of the processing applied it would appear that SMU height is better represented in the STU table than slope. This is not least a result of slope being recorded with only four categories. The nature of the slope data type (ordinal) is moderately masked when computing mean values.

- **Mean vs. Median SMU Slope**

  The difference in the frequency of the mean and the median slope for SMUs is given in Figure 16.

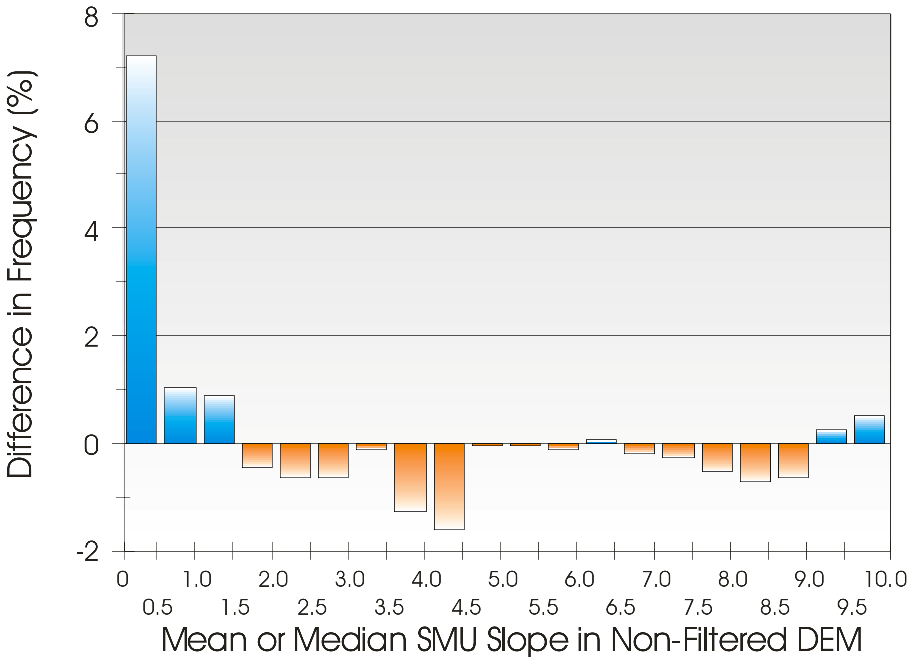


Figure 16: Difference in Frequency of Mean and Median Slope for SMUs in Non-Filtered DEM

The graph shows that the proportion of SMUs with a median slope of <0.5% (18.7% of all SMUs) is 7.3% higher than the proportion of SMUs with a mean slope (11.4% of all SMUs) in this range. Consequently, where the slope in the DEM is < 0.5% defining the central inflection points of the membership function based on the mean or the median may lead to different membership values, but less so for higher slopes.

- **Reducing Variability for Mean SMU Slope**

The relationship between the mean and the maximum slope for an SMU for the maximum value in the DEM, for a 10- and 5-percentile cut-off and a distance of 1 or 2 σ from the mean is presented in Figure 17.

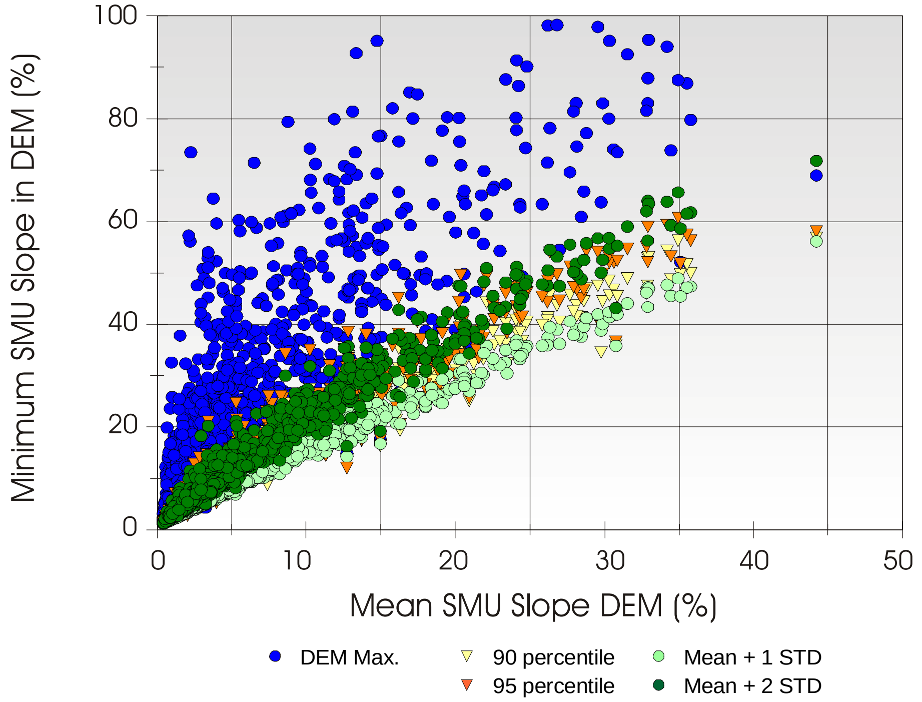


Figure 17: SMU Mean Slope compared to Maximum Slope and Maximum Limited by 90-, 95-percentile and 1 and 2 σ from Mean

The graph illustrates quite well the dependency of the maximum on the mean value when restricting the range. The slope data of the spatial layer also behaves according to expectations with comparable distribution of the maximum slope for a 5% cut-off and a distance of 2 σ from the mean and a maximum value which equals approx. 2 x the mean value. This is not the case for the minimum slope values, since the distribution of slope values is not normal. Instead of defining a restricted range of values from the standard deviation the percentiles would seem the more appropriate tool.

Reducing the range of slope values in the DEM by trimming affects class characteristics to varying degrees. A summary of the mean and maximum values for slope classes in the DEM are presented in Table 14.

Table 14: Distribution of Mean and Maximum Slope in DEM with Trimmed Range (non-filtered DEM)

SLOPE		All Data	5 % Trim	10 % Trim
Class	Range	Mean	Mean	Mean
	%	%	%	%
1	0 - 8	2.08	2.14	2.17
2	8 - 15	10.91	10.89	10.91
3	15 - 25	19.07	18.92	18.81
4	> 25	34.06	31.73	31.30
	Maximum	97.13	61.01	55.61

The mean and maximum values given in the table are derived from applying the 5% or 10% reduction in slope values to each SMU and then extracting the parameter from the collection of trimmed SMU data.

For slope classes with defined limits the effect of reducing the range of values is minimal. For Class 4, where the highest value is not pre-defined, the mean and maximum values of slope in the DEM decrease notably with the reduction in the range of values. The maximum slope is 97.13 in the DEM without restriction is 97.13%, which decreases to 61.01% when removing the upper 5% of data in each SMU and to 55.61% when removing the upper 10%. Similarly, the mean slope decreases from 34.06% (no restriction) to 31.73% (5% trim) and to 31.30 (10% trim).

The differences in the proportion of the area of a slope class in the typological database and the distribution of slope in the spatial layer result in the parameter being represented in the ancillary data only by approximation. In general, the extent of slopes shallower than defined by the class limits is significantly larger in the spatial data than the STUs. As a consequence, association the STU slope class with the corresponding area in an SMU leads to a low level of agreement.

3.3.3 Criterion Factor Land Use

While the topological criteria are considered constant over time land use is subject to change. As a consequence, the spatial allocation of an STU depends on the ancillary land use data employed and is therefore firmly linked to it. This can be problematic when estimating temporal changes in soil properties over a base period and when using different land use or cover data sets for two periods. For agricultural land the land use / cover also depends on the management status, in particular whether plots are irrigated or drained. In many parts drained peat lands are used for agriculture, but peat lands without drainage are unsuitable for this type of activity. The difference in management may thus lead to differing trends in `the spatial allocation associated with a single land use / cover criterion. Some of the agricultural land management information is provided in the fields for water management systems ([WM1], [WM2]) (see Table 2 for completeness).

The fields for land use ([USE_DOM] and [USE_SEC]) do not contain any blank entries. Instances without information are coded using a zero (0) entry. There are two STUs with a non-zero secondary land use, but a zero entry for the dominant land use (430023, 430027). For all other STUs a secondary land use only exists when there is also information on the dominant land use. For the dominant land use all types specified in the attribute documentation are used. In addition, the code "30" is used to identify bare areas. For the secondary land use all codes are used, except those for bare areas and code "22" (*Cultivos enarenados* (artificial soils for orchards in SE Spain)).

The land use/cover classes of the SGBDE were assigned to a catalogue of more generalized classes of the main types:

- Arable land, except rice
- Rice fields
- Permanent crops
- Pastures
- Grassland
- Forest
- Wetland
- Bush, shrub
- Bare or sparsely vegetated
- Other

These broad land use / cover classes were defined with the aim of close comparability with the *Intergovernmental Panel on Climate Change* (IPCC, 2006) land use types used to define the management system for estimating changes in soil organic carbon stock in croplands and grazing lands and for estimating organic carbon content in the topsoil by the pedo-transfer rule (PTR21) of the ESDB.

The re-classified information of the land use / cover attributes in the STU database was compared to land use / cover data with continuous spatial coverage for the AOI. The main such dataset considered to aid the spatial allocation procedure comes from the *Co-ordination of Information on the Environment* (CORINE) programme. The data are available as the *CORINE Land Cover* series (CLC) from the *European Environment Agency* (EEA)[13]. The series cover three main observation periods of 1990 (CLC90), 2000 (CLC2000) and 2006 (CLC2006). The nomenclature of the CORINE land cover legend allows aggregating the categories of Level 3 to those defined as broad land use / cover classes. To overlay with the rasterized SMUs of the SGDBE layer the 250m raster layers of Version 16 (04/2012)[14] were resample to 1km by pixel thinning (taking every 4th pixel). While images for CLC2000 and CLC2006 were acquired around the year indicated in the name the images

[13] European Environment Agency, Kongens Nytorv 6, 1050, Copenhagen K, Denmark. eea.enquiries@eea.europa.eu

[14] http://www.eea.europa.eu/data-and-maps/data/corine-land-cover-1990-raster-2

leading to the CLC90 series were acquired between 1986 and 1998, depending on the country covered (Kleeschulte, 2006).

Due to the origins of the ESDB (data originates from surveys well prior to 1990) the data from CLC90 would appear to be more suited to support the spatial allocation of STUs than data from the later periods. However, CLC90 data only partially covers the AOI. For areas not covered by CLC90 other sources of data are needed. One such data set is the *Global Land Cover Characterization* (GLCC) database from the *United States Geological Survey* (USGS)[15]. Evaluated for use was Version 2.0 for Eurasia in the Interrupted Goode Homolosine Projection with the Seasonal Land Cover Region legend[16] and the Global Ecosystem Legend[17].

Close to the image acquisition period of 1990 is also the global dataset compiled by the University of Maryland (Hansen, *et al*., 1998). The layer was compiled based on *National Oceanic and Space Administration* (NOAA) *Advanced Very High Resolution Radiomete*r (AVHRR) images acquired between 1981 and 1994. The land cover map is available in various combinations of spatial resolution (1km, 8km, 1 degree), projections (Goodes or geographic) and file formats (BSQ, GeoTIFF). Used was the 1km data in geographic coordinates as GeoTIFF format, as available from the Global Land Cover Facility[18].

The land use/cover information was compared based on the proportion of the classes by SMU. For the SGDBE the relative areas of the [PCAREA] field were merged according to the reclassified fields [USEDOM] and [USESEC]. For the spatial land use / cover data the relative proportions of the general classes was extracted from the SMU layer. For the restricted coverage of CLC90 only those SMUs were used in the comparison which cover at least 90% of an SMU with CLC90 data in the AOI. This resulted in 1,201 SMUs.

The re-classification of the CLC data to the generalized categories of land use / cover is given in Table 15.

---

[15] http://edc2.usgs.gov/glcc/globdoc2_0.php
[16] http://edcftp.cr.usgs.gov/pub/data/glcc/ea/goode/easlcr1_2g.img.gz
[17] http://edcftp.cr.usgs.gov/pub/data/glcc/ea/goode/eaoge1_2g.img.gz
[18] ftp://ftp.glcf.umd.edu/glcf/Global_Land_Cover/Global/1km/AVHRR_1km_LANDCOVER_1981_1994.GLOBAL.tif.gz

Table 15: Assignment of Corine Land Cover Classes to Generalized Land Use / Cover Categories

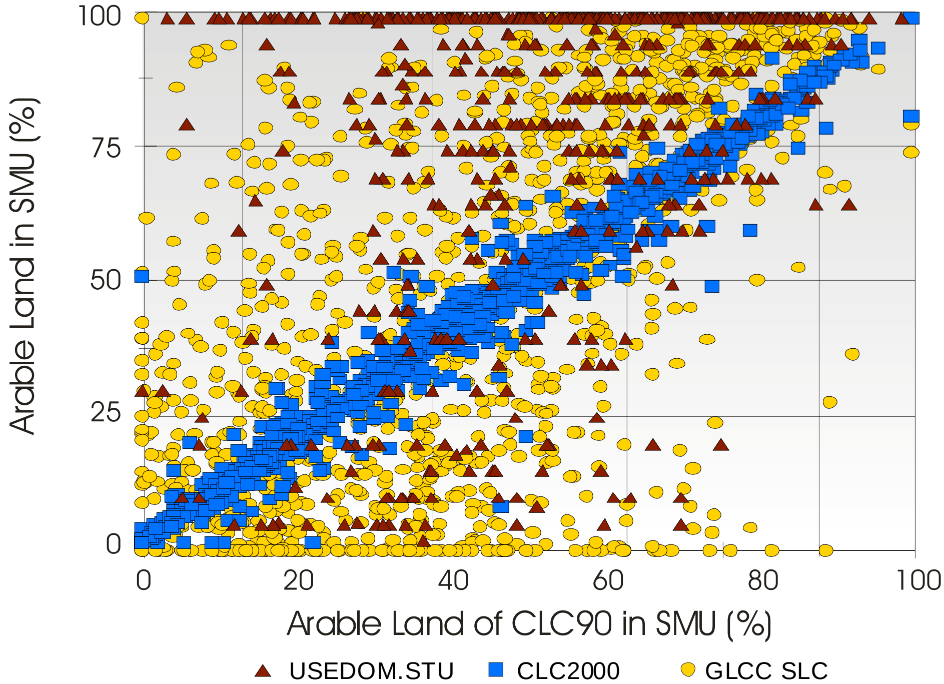
CORINE Class	Generalized Category	CORINE Class	Generalized Category
Continuous urban fabric	Artificial surfaces	Agro-forestry areas	Forests
Discontinuous urban fabric	Artificial surfaces	Broad-leaved forest	Forests
Industrial or commercial units	Artificial surfaces	Coniferous forest	Forests
Road and rail networks and associated land	Artificial surfaces	Mixed forest	Forests
Port areas	Artificial surfaces	Natural grasslands	Grassland
Airports	Artificial surfaces	Moors and heathland	Wetland
Mineral extraction sites	Artificial surfaces	Sclerophyllous vegetation	Bush, shrub
Dump sites	Artificial surfaces	Transitional woodland-shrub	Bush, shrub
Construction sites	Artificial surfaces	Beaches, dunes, sands	Bare or sparsely vegetated
Green urban areas	Artificial surfaces	Bare rocks	Bare or sparsely vegetated
Sport and leisure facilities	Artificial surfaces	Sparsely vegetated areas	Bare or sparsely vegetated
Non-irrigated arable land	Arable land, except rice	Burnt areas	Other
Permanently irrigated land	Arable land, except rice	Glaciers and perpetual snow	Other
Rice fields	Rice fields	Inland marshes	Wetland
Vineyards	Permanent crops	Peat bogs	Wetland
Fruit trees and berry plantations	Permanent crops	Salt marshes	Other
Olive groves	Permanent crops	Salines	Other
Pastures	Pastures	Intertidal flats	Background
Annual crops associated with permanent crops	Arable land, except rice	Water courses	Background
Complex cultivation patterns	Arable land, except rice	Water bodies	Background
Land principally occupied by agriculture, with significant areas of natural vegetation	Arable land, except rice	Coastal lagoons	Background
		Estuaries	Background
		Sea and ocean	Background

The GLCC layers were first assigned to the corresponding CLC classes and then to the generalized categories. This approach was found useful to help identifying the most appropriate assignment of the numerous mixed classes of the Seasonal Land Cover (253) and Global Ecosystems (96) legends.

Although the Seasonal Land Cover and the Global Ecosystems legends contain considerably more classes than the CLC legend there are some important omissions. There is no specific class for artificial areas and permanent crops are generally part of other classes.

The legend of the University of Maryland land cover data uses 14 land use / cover classes. Of these 6 relate to forested areas. When assigned to the generalized categories only artificial areas, cropland, grassland, forest, shrub and bare areas can be used. This lack of separating grassland and cropland into more detailed categories was found restricting the use of the data to supplement CLC90.

It would make very little sense to compare the classifications on a pixel-by-pixel basis. The geometric transformations alone would introduce small, but noticeable changes in the position of grid cells. Therefore, data aggregated to larger spatial units were used for the comparison. The proportion of the reclassified land use / cover datasets in the SMUs were compared to the information from the [USEDOM] and [USESEC] fields in the STU attribute table. A comparison of the categories "*Arable land*" and "*Forest*" is given in Figure 18.



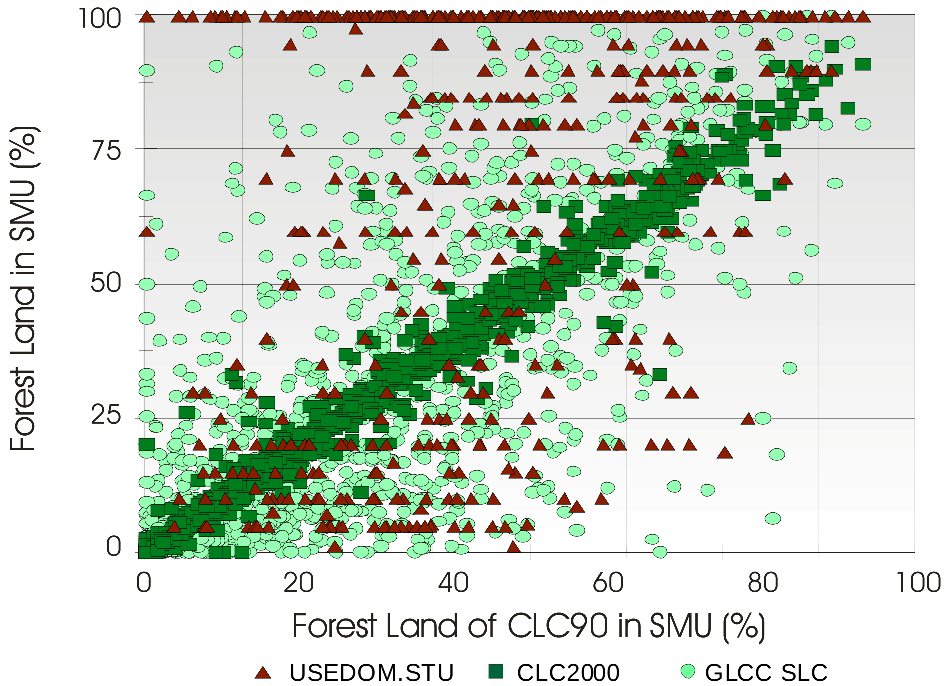
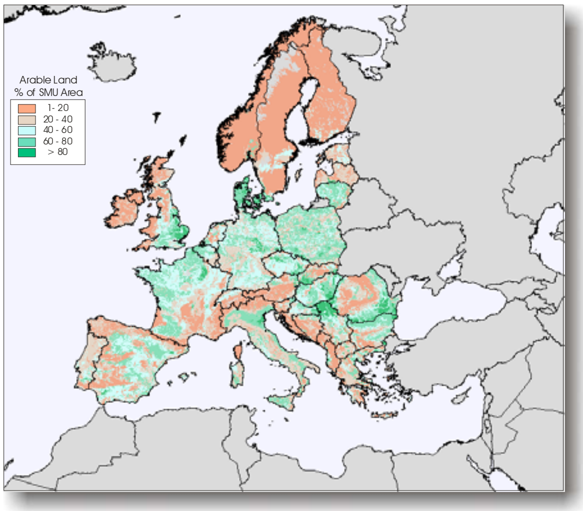


Figure 18: Comparison of Relative Areas of Arable and Forest Categories for STU, CLC2000 and GLCC with CLC90 Aggregated by SMU

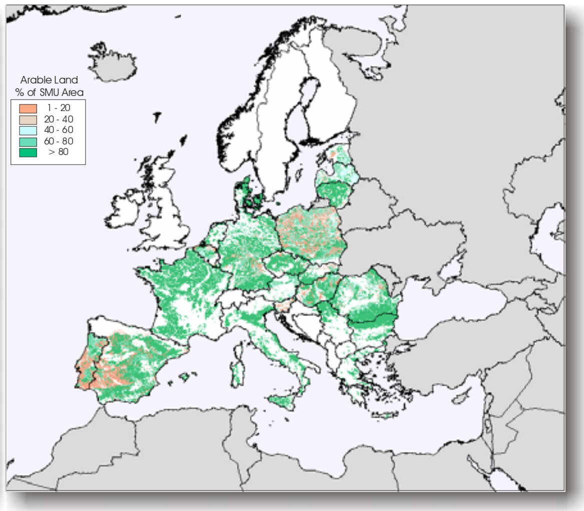
The graphs show a close correlation between the proportions of land use / cover categories between the CORINE datasets, but a large spread of values of the proportions derived from the GLCC and STU data to CLC90. The association of the 253 classes of the GLCC data to CORINE Level 3 classes was guided by a cross-classification with CLC90. Improvements in relating GLCC data more closely with CLC90 may well be achieved from investigating regional associations. However, using CLC90 as a reference to assign the GLCC classes and given the close relationship between CLC90 and CLC2000 areas not covered by CLC90 in the AOI may be better covered by CLC2000 then the GLCC data.

Although the comparison is interested in the relationship of the categories between the STU attribute data and the spatial land cover datasets the CLC90 shares in the SMUs were assigned to the x-axis. This approach was taken since the aggregated STU data frequently amounts to 0 or 100% and therefore very much limits the presentation of the variation in the land cover data in the figure. This absence or saturation of SMUs with land use / cover information is an unavoidable consequence of the nature of the attribute representation ion the database. The fields [USEDOM] and [USESEC] contain the land use categories, but no the share of the categories within the STU. In the process of aggregating relative areas of a land use / cover category from the STUs to the SMUs this absence of moderating share results in giving full weight to the category, i.e. treating it as completely covering the STU. Where all STUs of an SMU belong to the same category the portion of the category is then 100%. In case a land use / cover is not dominant in any of the STUs of an SMU the portion of the category is 0%. To some degree the information on the secondary land use could alleviate the tendency to extreme values. However, merging the two attributes requires defining the proportion of the dominant to the secondary land use /cover. It is further not known whether additional land use / cover categories are present in an STU. As a consequence, even setting a proportion such as 60% for the dominant and 40% for the secondary category may well be equally far from actual situation.

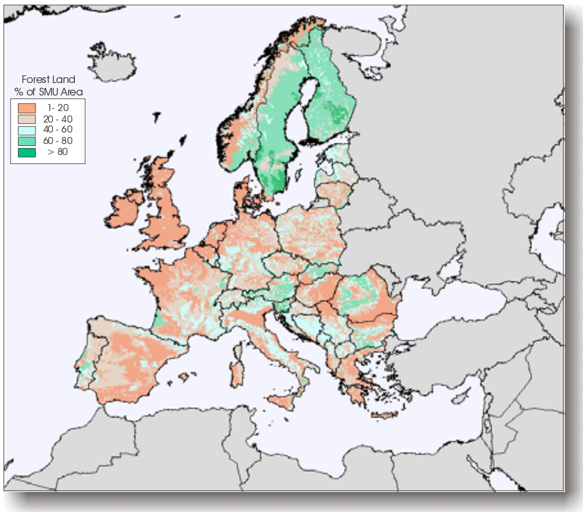
A further complication is that there seem to be variations in the delineation of SMUs according to land use / cover between data sources. This is illustrated in Figure 19.



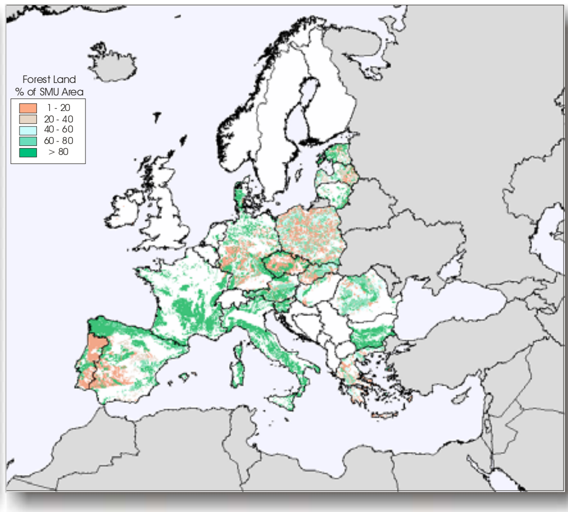
Portion of Arable Land in SMU (CLC2000)



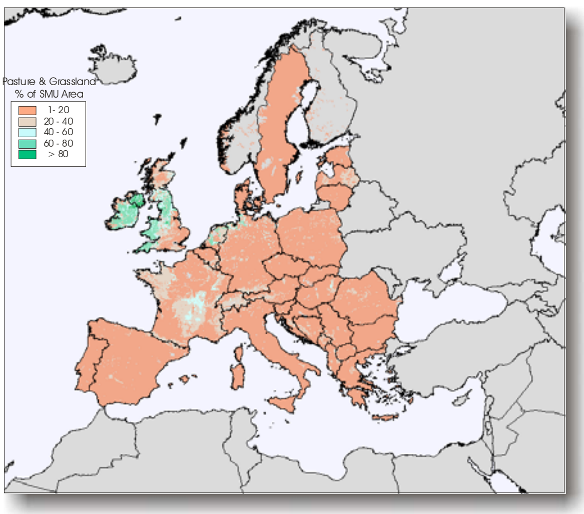
Portion of Arable Land in SMU from [USEDOM]



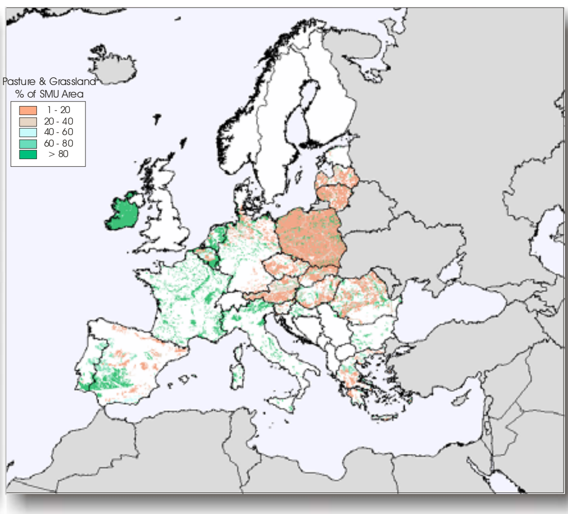
Portion of Forest Land in SMU (CLC2000)



Portion of Forest Land in SMU from [USEDOM]



Portion of Pasture & Grassland Land in SMU (CLC2000)



Portion of Pasture & Grassland in SMU from [USEDOM]

Figure 19: Comparison of Portion of Arable, Forest and Combined Pasture and Grassland for CLC2000 and STU Attribute Data Aggregated by SMU

The maps depict the portions of the land use / cover categories "Arable Land", "Forest" and the combined area of "Pasture" and "Grassland" for CLC2000 and the dominant land use of the STU for the SMUs. The categories "Pasture" and "Grassland" were grouped together for the purpose since there appears to be a fair amount of overlap between these land use / cover types. From a visual interpretation the maps indicate a close relationship of the general distribution of the categories between across the AOI, with the expected stronger contrast in the STU data. The maps also indicate that in several countries the SMUs contain STUs of just one category. In these SMUs setting proportions between the dominant and the secondary lad use / cover would not resolve the lack of information on the share of the land use / cover category in the STU database.

3.3.4 Criterion Factor Depth

It was further considered to link the various STU depths attributes to slope in the ancillary data layer. The depths of soils is expected to decrease with slope, while the period of water logged conditions increase with shallower slopes. Indicators of depth in the STU table are:

- obstacle to roots [ROO.STU_SGDBE]
- presence of an impermeable layer [IL.STU_SGDBE]
- dominant annual average soil water regime [WR.STU_SGDBE]
- depth to rock [DR.STU_PTR]

The relative frequency of a depth value by slope class for each of the STU and the PTR depth attributes is given in Figure 20.

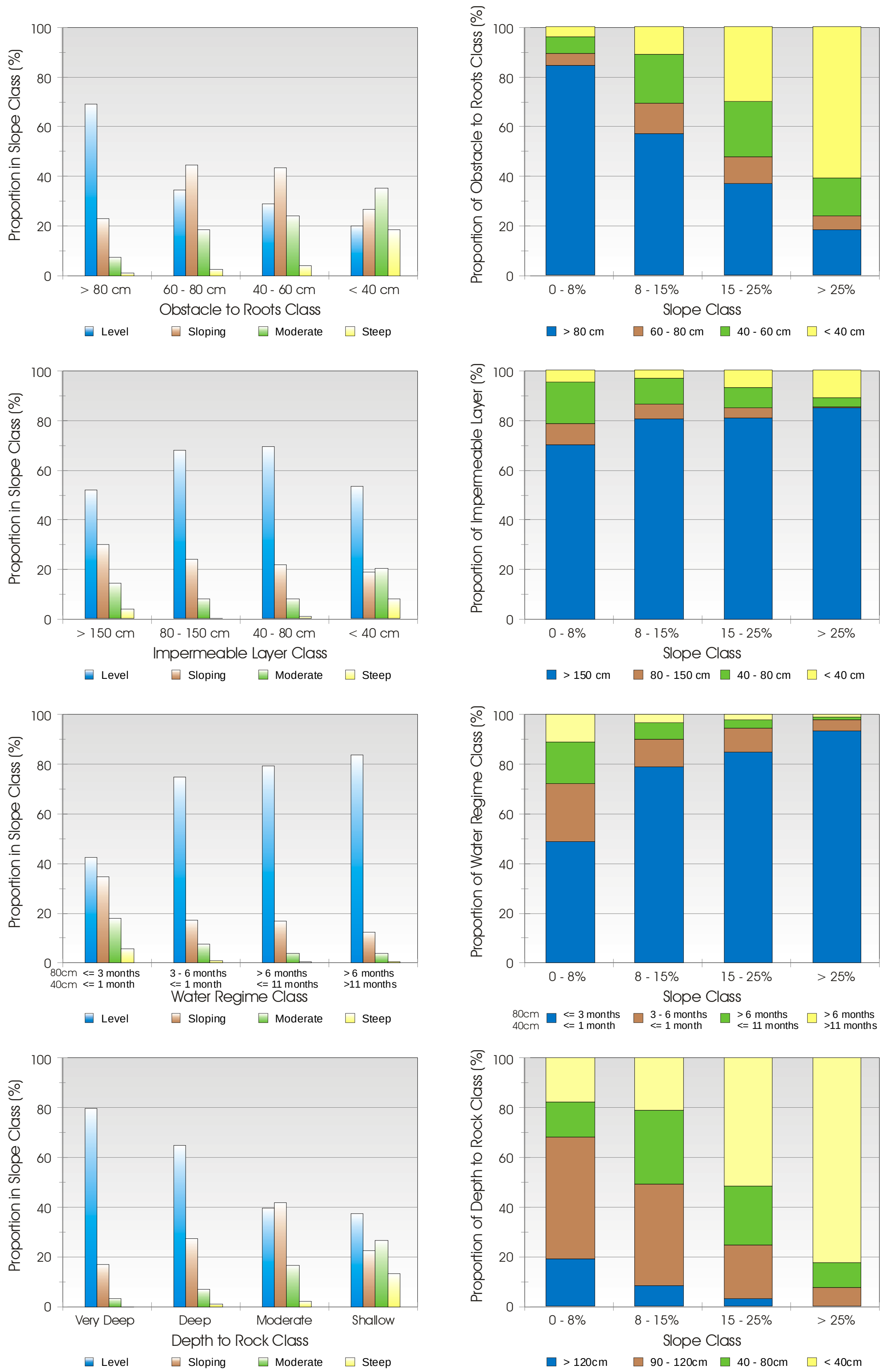


Figure 20: Relative Frequency of STU and PTR Depth Indicators by Slope Class

The bar charts show the relative occurrence of the slope classes by attribute class. Of the associations of the class "*No obstacle to roots between 0 and 80 cm*" for the attribute "*Obstacles to roots*" 68.9% are with slope class "*Level*" (0 – 8%), 23.0% with slope class "*Sloping*" (8 – 15%), 7.2% with slope class "*Moderate*" (15 – 25%) and 0.9% with slope class "*Steep*" (>25%). The stacked bar charts show the shares of the attribute classes within a slope class. Thus 84.6% of slope class "*Level*" are linked to the class "*No obstacle to roots between 0 and 80 cm*", 4.8% to class "*Obstacle to roots between 60 and 80 cm depth*", 6.5% to class "*Obstacle to roots between 40 and 60 cm depth*" and 4.1% to class "*Obstacle to roots between 20 and 40 cm depth*".

- **Obstacle to Roots**

  For the "*Obstacle to roots*" attribute the graph only includes values for distinct depths (codes "1" to "4") and not the general indicator for an obstacle given by code "5". Only one STU has an entry of code "6" ("*Obstacle to roots between 0 and 20 cm depth*"), for which no data on the slope class was recorded. Overall level slopes are dominated by soils with no obstacle to roots within at least the upper 80 cm. Obstacles to roots at shallower depth become increasingly frequent with increasing slope. The inverse trend is found for STUs with obstacles to roots within < 40 cm of the surface.

  The relationship between the obstacle-to-roots and the slope classes can be arranged as the distribution of the attribute across the slope classes. A graphical presentation of the distribution is given in Figure 21.

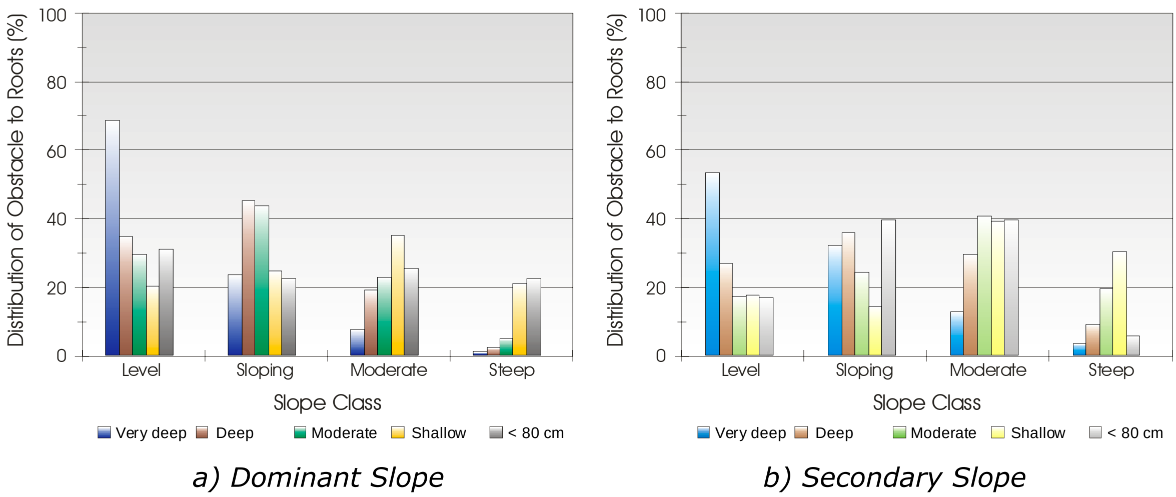


Figure 21: Distribution of Obstacle-to-Roots Classes over Slope Classes

The graph shows that 70% of very deep soils are aligned with level dominant slopes. Deep and moderately deep soils are mainly aligned with sloping dominant slopes and shallow soils tend to be aligned with moderately sloping or steep slopes. A comparable distribution of aligning obstacle-to-roots classes with slope is found for the secondary slope parameter. For completeness presented in the graph is also the distribution for the [ROO] class "*Obstacle to roots between 0 and 80 cm depth*" (code 5). There are only 36 STUs with this code) with a distribution close to the one for moderate slopes.

The parameter shows general trends for the portion of very deep soils to decrease with slope and the inverse trend for soils with obstacles to root closer to the surface. This general trend may be used as a criterion factor to support positioning STUs.

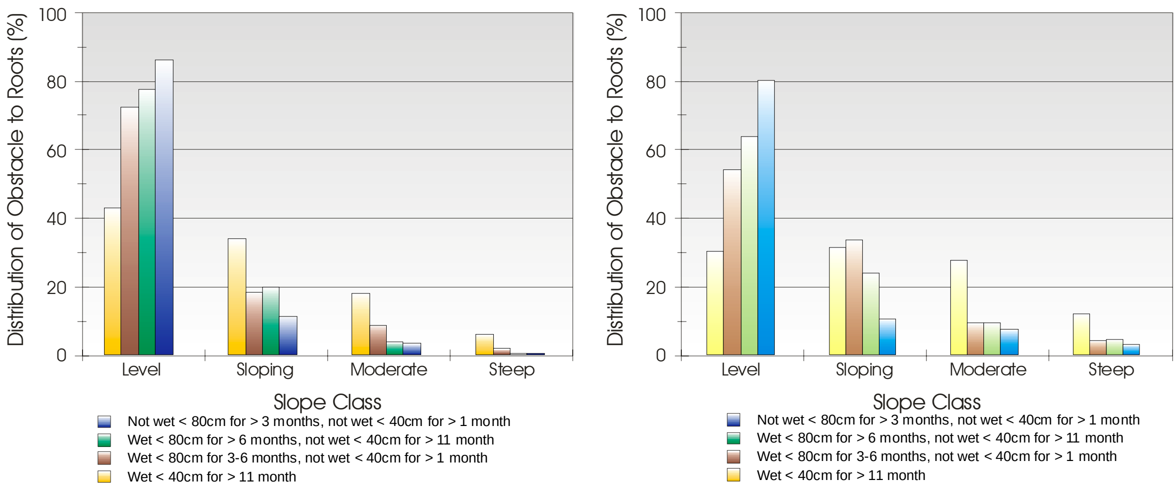
For the 314 STUs without information on the parameter a value was estimated from the most widely used value set for an FAO85 class. Assigning an FAO85 soil type to a class of obstacle to roots was based on the frequency of a class and the area covered. In most cases one class clearly dominated the soil type. Where data were completely missing for a soil type the obstacle-to-roots class was estimated from the higher soil category.

- **Impermeable Layer**

Less clear than for the depth of obstacles to roots is the relationship between the "*Impermeable layer*" attribute and slope. An impermeable layer may well be found at shallow depth in areas with level slopes. It would appear that this attribute is less suited to be associated with slope to serve as a parameter in the spatial allocation of STUs.

- **Water Regime**

The proportions of the classes of the attribute "*Water regime*" show a distinctly different change with slope between class 1 (*Not wet within 80 cm for over 3 months, nor wet within 40 cm for over 1 month*) and the other three classes of the attribute. The result of rearranging the data of Figure 20 to the distribution of the attribute across the slope classes is presented in Figure 22.



a) Dominant Slope *b) Secondary Slope*

Figure 22: Distribution of Water Regime Classes over Slope Classes

The graphs illustrate the rapidly decreasing presence of water-logged soils with increasing slope. For 333 STUs of 392 STUs with a [WR] value of class 4 ("*Wet within 40 cm depth for over 11 months*") 333 are aligned with a level dominant slope. Aligned with steep dominant slopes are 2 STUs (0.5%). On steep slopes 90% of the STUs are specified as [WR] class 1 (see Figure 20). These diverse and relatively specific trends in the distribution of the classes of the water regime attribute with slope should aid the geographic positioning of the classes specifying some period of a water-logged state when present in an SMU.

For the 275 STUs without information on the parameter a value was estimated from the most widely used value set for an FAO85 class. As for the obstacle-to-roots parameter assigning an FAO85 soil type to a class for the water regime was based on the frequency of a class and the area covered. Although in most cases one class dominated the soil type, there appeared to be more spread for some soils than for obstacles to roots. While for the obstacle to roots parameter 68% of the STUs were set to class 1 the figure was 5% lower for the class 1 of the water-regime parameter. For class 2 the portion of STUs is 11% higher for the water regime than the obstacles to roots[19]. Where data were completely missing for a soil type the water-regime class was estimated from the higher soil category.

- **Depth to Rock**

The attribute "*Depth to Rock*" is derived from a PTR. The conditions of the rule do not include slope (SLOPE_DOM.STU_SGDBE or SLOPE_DOM.STU_SGDBE) as a parameter and the attribute is thus an

[19] For "Obstacles to roots" parameter values are used for 5 classes, while 4 classes are used for the "Water regime" parameter. The relative frequency of classes occurring by STU is still comparable between the parameters since only 0.7% of STUs are assigned to class 5 of the "Obstacles to roots" parameter.

independent criterion. The advantage of using this attribute is the general availability of a value for an STU, with 96% of all STUs in the AOI with a complete set of data to define the conditions. However, the conditions are more specific only for *Cambisols* and *Luvisols*.

In the proportion and the distribution of the classes with slope the changes are comparable to those of the obstacle-to-roots attribute. When present the class of "*Very deep*" soils is aligned to 80% with level slopes and none with steep slopes. On steep slopes shallow soils, with a depth to rock of < 40 cm, are specified for 83% of the STUs. While the trends in the alignment of the attribute with slope are pronounced, they are close to those found for the obstacle-to-roots attribute.

The alignment of depth attributes in the fields [ROO.STU_SGDBE] and [WR.STU_SGDBE] with the dominant and secondary slope in the STU database suggests that the information may be useful to serve as criterion factors in the MCE. Less obvious is the use of the depth attributes [IL.STU_SGDBE] and [DR.STU_PTR]. For the impermeable layer data the lack of a generalized trend with slope could be expected. The PTR attribute may be better applied to the output of the sDSS than used as a criterion factor to allocate STUs.

3.3.5 FAO85-Aggregated Criterion Factor Height

For cases of STUs with missing height or slope data or where the values are not sufficiently distinct to characterize the affinity of the STU with geographic locations some measure of the relative position in the SMU can be estimated from the parameter values from other STUs with comparable characteristics. Therefore, height and slope parameters were computed for each of the FAO85 soil classes. For the range limits the minimum and maximum values of height and slope were extracted from the STUs by soil class. The topographic parameters mean height and slope were determined by using an area-weighted average as follows:

$$\overline{X}_{FAO85} = \frac{\sum_{i=1}^{n} X_{STU_i} * AREA_{STU.SMU_i}}{\sum_{i=1}^{n} AREA_{STU5.SMU_i}}$$

where

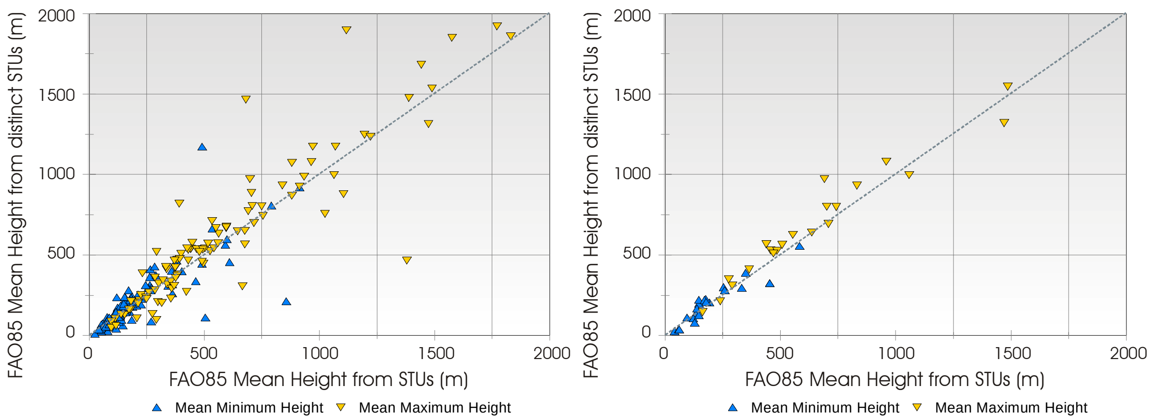
$\overline{X}_{FAO85}$	mean topographic parameter for FAO85 soil type
X_{STU_i}	topographic parameter in SMU
$AREA_{STU.SMU_i}$	area of STU in SMU (surface area or proportion)
n:	No. of STUs with FAO85 soil class

For 214 FAO85 soil classes of STUs assigned to SMUs in the AOI a range of height values could be determined, i.e. the Mean([*ZMIN*]) < Mean([*ZMAX*]), and no instances of equal values occurred. For 154 soil types a variation of values could be computed (n>1). For these the standard deviation around the maximum height value included the mean of the minimum height in 85 (55.2%) cases.

It could be argued that the height information is better defined in STUs which are used to adequately define SMUs. There should therefore be less variation in the height values. When evaluating the mean height by FAO85 soil class only for STUs which show diversity within SMUs the number of soil types with height data is reduced to 148. In this set of data more than 1 STU with a minimum and maximum height is found for 101 soil types. Cases where the minimum height is within 1 standard deviation from the maximum height are 43 (41.6%). Compared to the number of such cases in the non-restricted data (55.2%) this appears to be a notable reduction.

The standard deviation is related to the magnitude of the values. Therefore, a reduction in the variability of the height by FAO soil class was found to be closely linked to a reduction in the mean height. The differences in the mean height by soil class between using all STUs and only those STUs which provide distinct values to separate their position within SMUs they are linked to are presented in Figure 23.

The graphs show that for most soil classes the mean minimum and maximum heights are slightly higher when calculated from STUs with distinct values than when calculated from all STUs. The same tendency, but with less variation, is found when restricting the comparison to those soil types, which are characterized by at least 30 STUs.

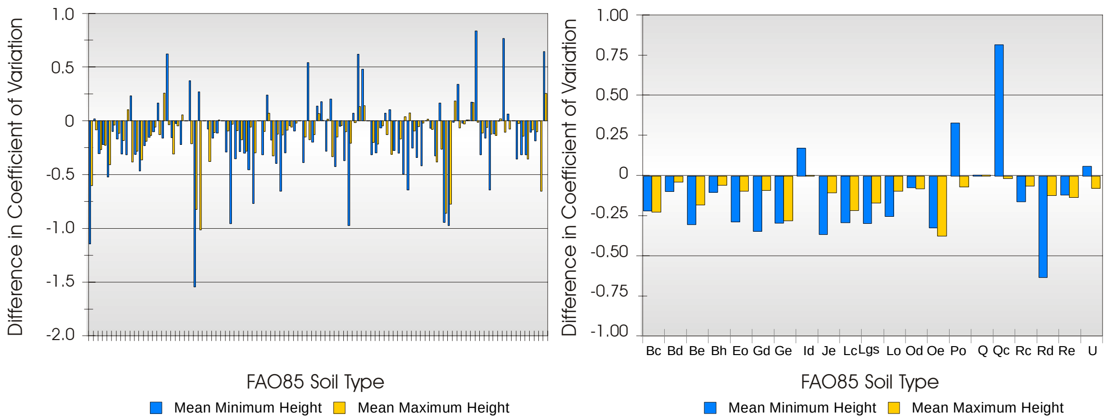


a) all STUs *b) mean from ≥30 STUs*

Figure 23: Mean of Minimum and Maximum Height by FAO85 Soil Class

A method to normalize the measure of variation for the magnitude of the values is the *coefficient of variation* (CV). Although in the definition of data types no specific distinction was made between data measured on a ratio scale or interval scale this separation is of relevance for computing the CV, which only provides a meaningful value for data on ratio scale. This is not the case for height (altitude). The CV was used as a relative indicator for changes in the variation between two treatments of including values for a parameter in the computations and to reduce the effect of any changes in the mean values.

The differences in the CV between the mean minimum and maximum height for all STUS and when using only distinct STUs for the 101 soil classes with distinct STU data and for soil classes with at least 30 distinct STUs are presented in



a) all STUs *b) mean from ≥30 STUs*

Figure 24: Coefficient of Variation for Mean of Minimum and Maximum Height by FAO85 Soil Class

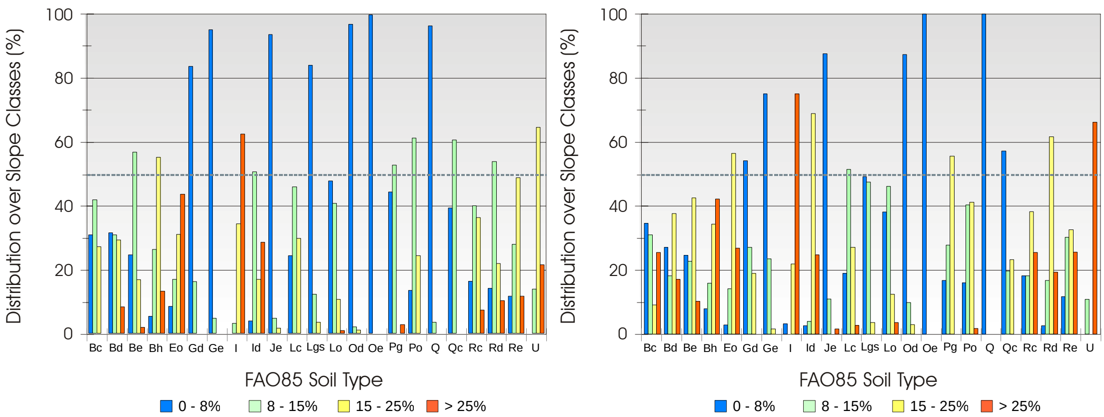
The graphs indicate a general decrease in the dispersion of height data in the distinct STUs as compared to height data taken from all STUs for both, the minimum and the maximum height. A notable exception is the data for the mean minimum height for the soil *Cambic Arenosol* (*Qc*). The mean minimum height changes little between the treatments (51.4 m vs. 44.6 m for distinct STUs) but the already high CV increases from 2.1 (all STUs) to 2.9 (only distinct STUs). There is no well-defined height range for this soil class in the data even when using all 95 STUs with data on minimum height, since the standard deviation is more than twice as high as the mean. However, for most soil classes the dispersion of values for the minimum and maximum height decreases when limiting the computation of the mean values to distinct STUs. In the definition of the factor these mean values could be given priority over means computed from using all STUs or where an insufficient number of distinct STUs are available to specify the mean.

3.3.6 FAO85-Aggregated Criterion Factor Slope

The method of aggregating the values for the slope parameter differs from the one used for height due to the data type (ordinal). As a result, rather than by a mean value the values are aggregated by the relative frequency of the slope types by soil class. Of the 214 soil classes with data for the dominant slope in 5,465 STUs 79 are only found for Class 1 (0 – 8%), 12 only for Class 2 (8 – 15%), 7 only for Class 3 (15 – 25%) and 1 only in Class 4 (> 25%). A preference (>50% of soil types in a single slope class) is found for 68 soil types. This leaves 47 soil types with <50% presence in any slope class. When using only STUs with distinct data within SMUs the number of soil types with data is reduced to 148. For 117 soil types one slope class accounts for >50% of distinct STUs with the slope data. However, this comparatively high number of soil types with a preference for one slope class is also the result of a reduced number of STUs (2,473). However, the characteristic slope is defined by only one STU for 48 soil types (39 for distinct STUs).

For the secondary slope parameter data from STUs also cover the 214 soil types. For 83 soil types a single slope class is found in the STU data and for 74 soil types a preference (>50% of STUs) for one slope class is found. When using only distinct STUs in the aggregation 107 soil types of the 148 soil types can be to one soil class or show a preference to occur in one soil class.

The distribution of the soil types across the dominant and secondary slope classes with more than 30 distinct STUs is presented in Figure 25.



a) dominant slope *b) secondary slope*

Figure 25: Distribution of FAO85 Soil Types by Slope Class for Soil Types with >30 Distinct STUs

The association of slope class and FAO85 soil type differs between the dominant and the secondary slope. For the secondary slope a shift towards steeper slopes was found. This tendency could be expected, because the prevalence of a given slope decreases with increasing steepness of the terrain.

For both slope parameters the use of data from only those STUs which distinctly define an SMU, instead of using all data, has sharpened the definition of FAO soil types. For the 23 soil types for which the class of the dominant slope is characterized by >30 distinct STUs the number of soil types with a frequency >50% is 11 (11 for secondary slope) when using all STUs and 16 (14 for secondary slope) when using only distinct STUs.

3.3.7 Other FAO85-Aggregated Criterion Factors

For the attributes on **land use** and **depth** no typical or mean values for a soil type were estimated. The changing and locally variable relationship between land use and soil type renders defining an association as a criterion ineffectual. Instead of defining a secondary depth criterion factor in order to cover all STUs with a value, the attribute "*Depth to rock*" of the PTR can be used, which is available for all STUs.

3.3.8 FAO85-Aggregated DEM-Based Criterion Factor Height

Tertiary criterion factors for height and slope are derived not from the STU table but from summary statistics from the ancillary DEM layer. They are intended to help resolve cases where the STU table contains no or largely indistinct data for the parameters used as criteria.

The SMU minimum and maximum height values extracted from the STU data are the minimum and maximum height of any STU linked to an SMU. The SMU minimum and maximum height values extracted from the DEM are the minimum and maximum elevation values of any grid cell within an SMU. The corresponding values extracted from the STUs were compared to those from the DEM for all SMUs and also for only those SMUs where the values for height in the STUs are sufficiently distinct to identify segments in the SMU.

For the height parameter the summary statistics are graphically compared in Figure 26.

The data for minimum height within an SMU show no relationship between the values extracted from the DEM and the values of the STUs linked to an SMU. The height limits are the extreme values found in the area covered by the SMU in a DEM, and could be defined by a single grid cell. This contrasts with the more generalized limits given in the STU table, which may contribute to the lack of a relationship for the minimum height. Since the extent of features in the SGDBE may vary slightly from the DEM layer a border effect, in particular along the land/sea boundary, cannot be excluded. To reduce such an effect a buffer of one grid cell was applied to the SMU layer and the height limits were from the DEM. This did not improve the relationship between the STU and the DEM data.

Another option followed was to remove from the assessment any SMU which contain areas on non-soil. Where an SMU contains STUs with non-soil areas, such as in alpine regions, the height limits extracted from the DEM would be in excess of those extracted from the STUs, because the non-soil STUs do not contain information on height limits. Also this approach did not result in a notable improvement in the relationship between STU and DEM information, neither for the minimum nor the maximum height.

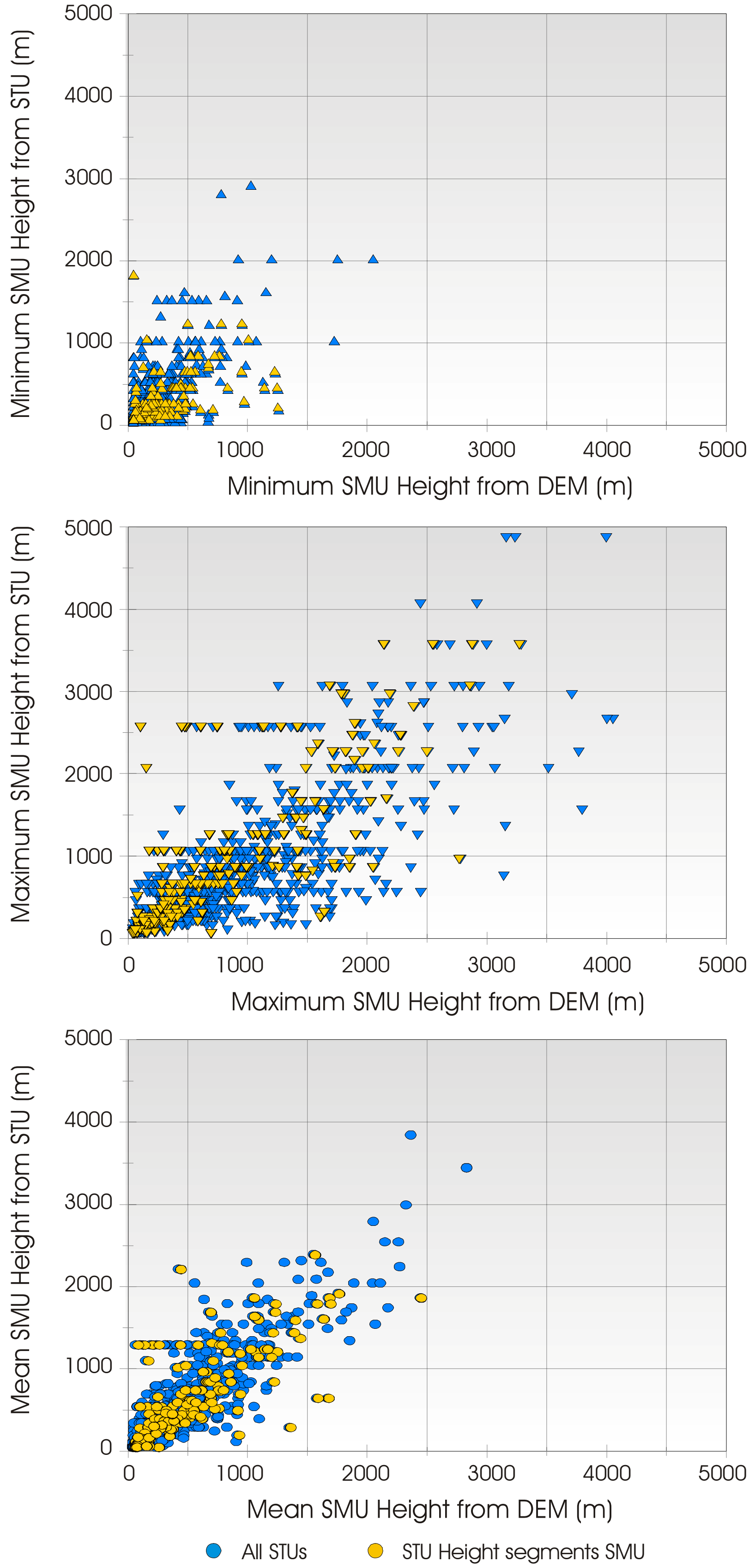


Figure 26: Comparison of Height Parameters for SMU from DEM and STUs

For the maximum height parameter some relationship between the STU and the DEM was found. As with the minimum height parameter the various options of processing the data did not result in an improvement of the relationship.

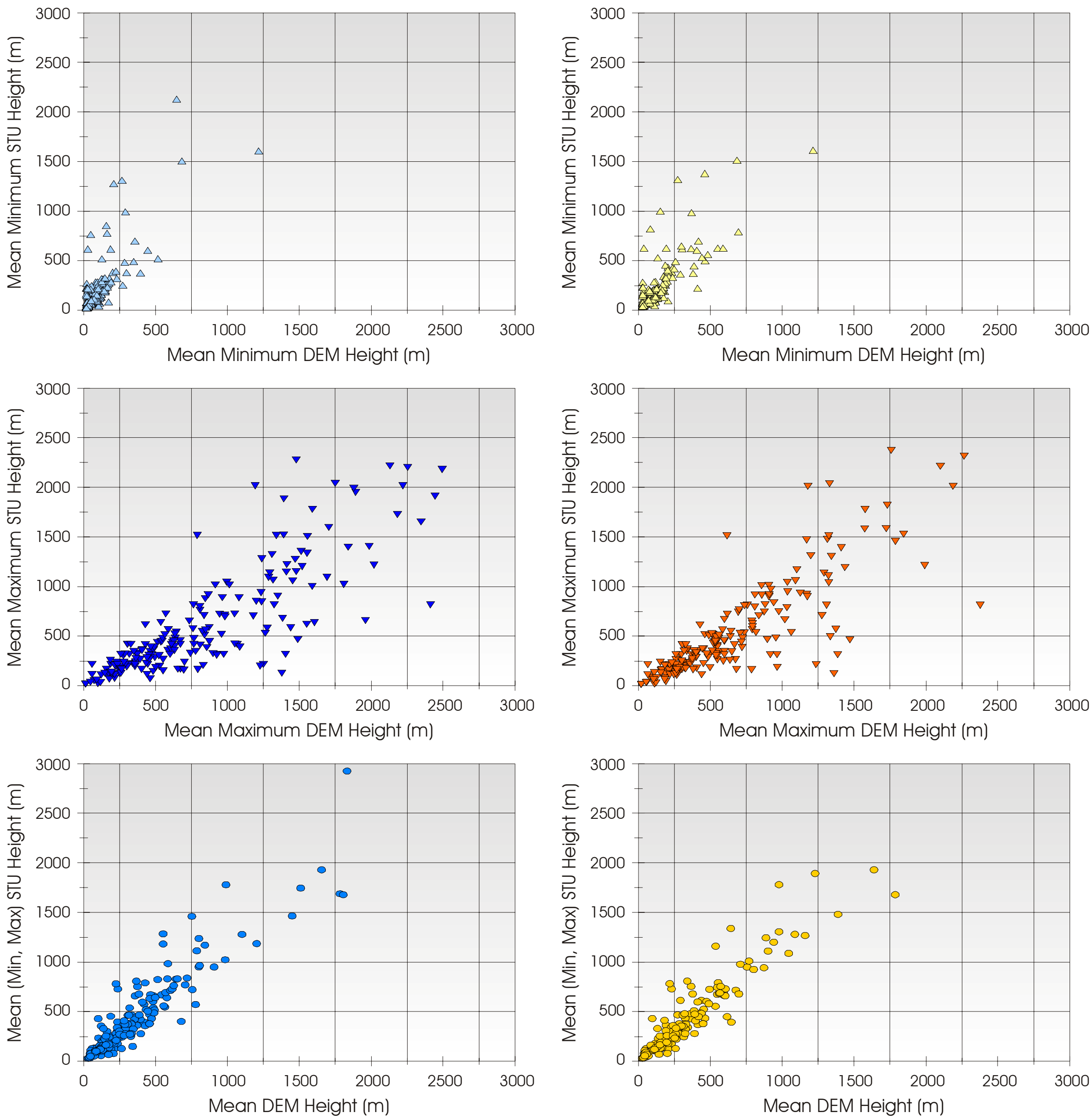
Values for the mean SMU height show a more distinct relationship. This would indicate that the height limits are not very well dealt with, but that there may be no particular trend to over- or underestimate the limits.

As for the STU data the minimum, maximum and mean height of the FAO85 soil classes were also extracted from the DEM. For each soil type the following statistics were determined for the STU and the DEM data:

- minimum and maximum height of soil type;
- surface area (in spatial layer) weighted mean of minimum and mean of maximum heights for soil type;
- relative area ([PCAREA]) weighted mean of minimum and mean of maximum heights for soil type;
- simple arithmetic mean of area-weighted minimum and maximum height;
- surface area (in spatial layer) weighted mean for mean height for DEM data;
- relative area ([PCAREA]) weighted mean for mean height for DEM data.

The weights given to computing the means of the minimum and maximum height as well as the average height can be derived from the surface are of the SMUs in the spatial layer or the relative proportions of the STUs in the SMUs, as given in the field [PCAREA.STUORG]. It may be argued which area to use as weighting factor when calculating mean height characteristics. Using the surface area emphasizes the spatial distribution of a soil type across height ranges while using the relative PCAREA parameter puts the accent on the number of occurrences independently of the actual area of the SMUs.

A visual comparison of using the two weighting factors on the mean minimum, maximum and the mean height (STU: AVG[AVG(*ZMIN*), AVG(*ZMAX*)]; DEM: AVG(*MEAN_HEIGHT*)) is presented in Figure 27.



a) Surface area weighted *b) PCAREA weighted*

Figure 27: Comparison of Minimum, Maximum and Mean Height Parameter for FAO85 Soil Class for STU and DEM Data from Surface and PCAREA Weighting Factors

The graphs do not indicate dramatic differences in the general relationship between weighting factors derived from the surface are and the proportional area in SMUs for any of the characteristics. However, for individual soil types changes in the relationship ensue. For the mean height of FAO85 soil types the method of calculating the characteristic changes the relationship between the DEM and the STU data more significantly. While the coefficient of determination (r^2) between the weighted mean height of SMUs in the DEM and the simple arithmetic mean of the weighted mean minimum and maximum heights of the STU data is 0.86 (surface area; 0.87 for PCAREA), the r^2 of relationship of the simple arithmetic mean of the weighted mean minimum and maximum heights of the DEM to the STU drops to 0.68 (surface area; 0.69 for PCAREA). Using the simple arithmetic mean of the mean minimum and maximum heights for the DEM data would appear to be not the first choice.

For SMUs with a sufficient number of diverse STUs to segment the area a graphical presentation of the data pairs of the surface area weighted mean height by FAO85 soil type between the STU and the DEM data is given in Figure 28.

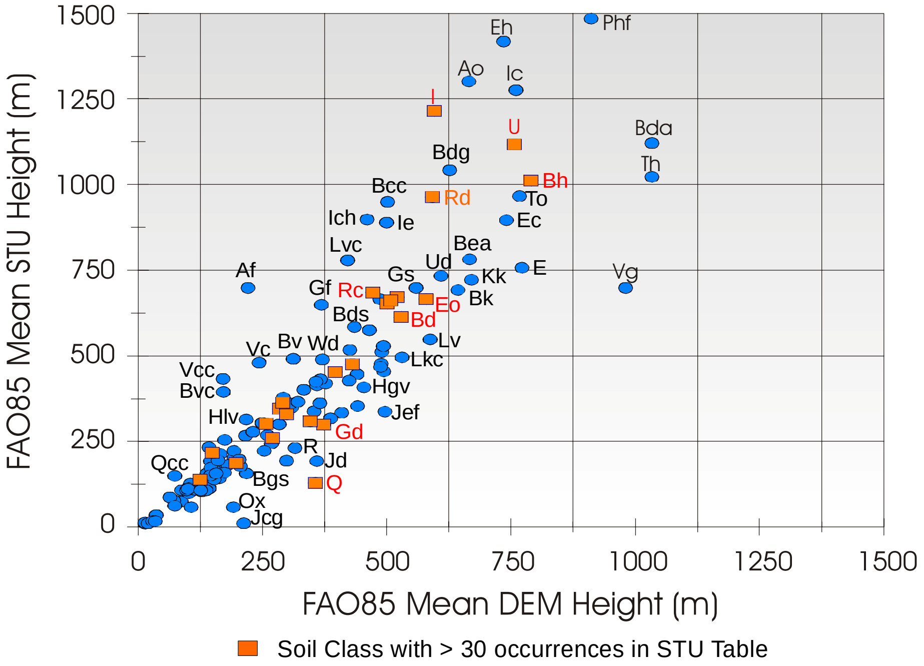
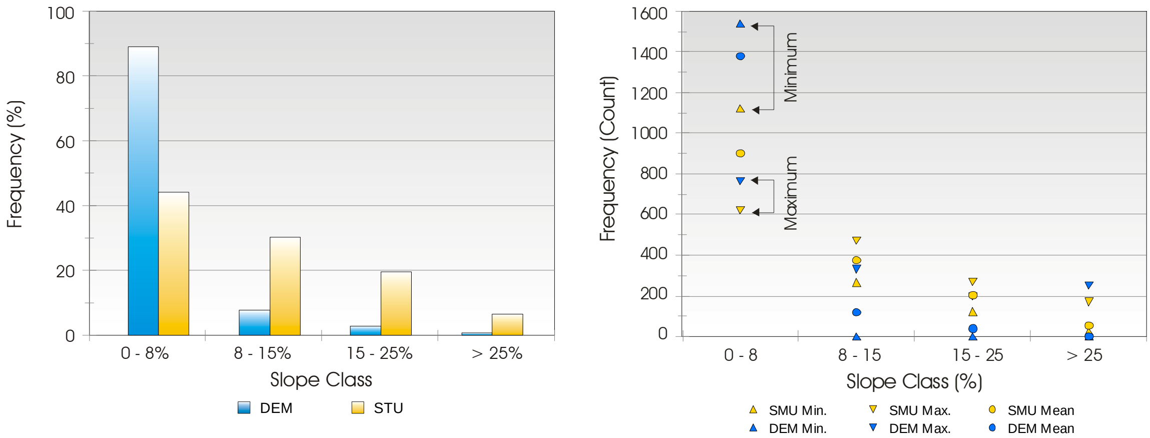


Figure 28: Comparison of Mean Height Parameter for FAO85 Soil Class for STU and DEM Data for SMUs with diverse STU data

The graph shows a relationship between the mean of the topographic parameters for the STU and the DEM by FAO85 soil class. Of interest in the data pairs is less that a relationship seems to exist, which can be expected since the topographic data in the STU table is related to the DEM of the SMU. More important is that the soil classes show some distinct positions in the relationship. The relationship is improved when using data from only those STUs for which the typological data is sufficiently diverse to provide information on the segmentation of areas within an SMU and from frequently occurring FAO85 soil types.

3.3.9 FAO85-Aggregated DEM-Based Criterion Factor Slope

In the attribute table slope is of an ordinal data type. This limits the comparison between the STU and DEM data to the classes defined for the parameter. The values extracted from the DEM were therefore aggregated to these classes. Results were compared based on the frequency of a slope parameter (minimum, maximum and mean) in one of the classes (intervals). A graphical presentation of the comparison is given in Figure 29.



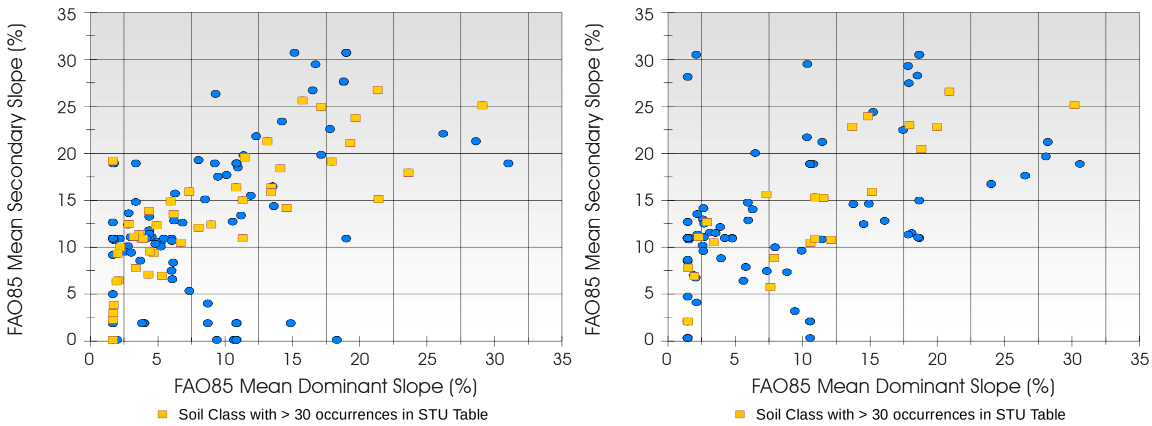
a) Slope from DEM *b) Slope Class Frequency*

Figure 29: Comparison of Slope Parameter for SMU from DEM and STU by Class Frequency

The frequency of slopes by class within the SMUs show a markedly different distribution between the slope extracted from the DEM and the dominant slope parameter for the STUs. For the DEM 89% of the area is assigned to slope class 1 (0-8%) compared to 44% for the STU data. The STU data shows a higher occurrence in all other slope classes. When looking at the frequency of the minimum, maximum and mean slope analogous relationships are found for most factors. The minimum slope in the DEM for the SMUs is in all but four SMUs in slope class 1 (0-8%). For the STU data the minimum slope class is more widely distributed and 25 SMUs have a minimum of > 25% in the STU database, but only one when using the DEM. For other slope classes there is less variety in the distribution of slopes between the data sources.

For FAO85 soil types the relationship between the DEM and the STU data was calculated for an area-weighted mean slope. The area used was taken from the surface area of the SMUs in the spatial layer to which the relative area of an STU within an SMU was applied.

For STU data the mean slope was estimated from assigning a typical value to each class. This value was based on the distribution of slopes in the DEM across the AOI. In the comparison of STU and DEM slope a distinction was made between using all STUs linked to SMUs of the AOI and only those STUs with diverse characteristics in the topological parameters used on the analysis of the data. A comparison of the mean dominant slope to the mean secondary slope for all STUs and an aggregation using only diverse STUs is presented in Figure 30.

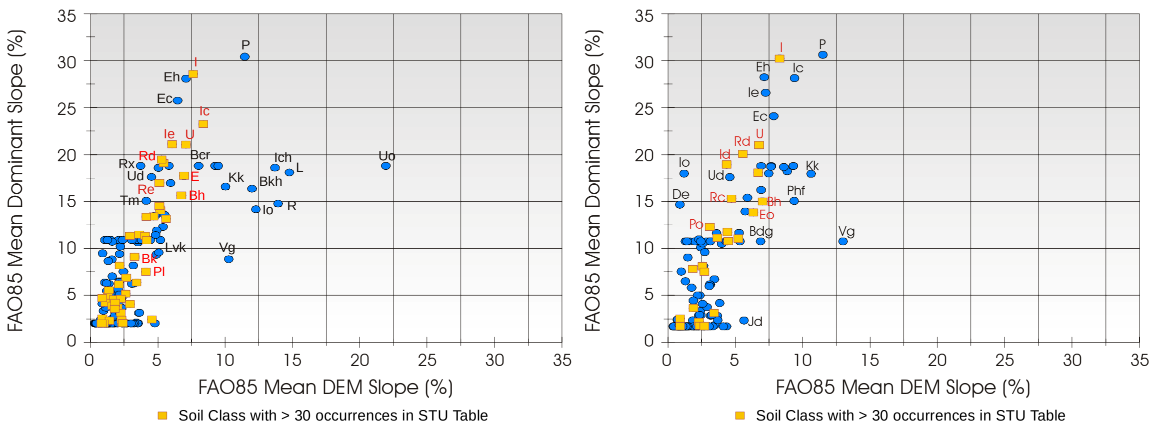


a) all STUs *b) STUs with diverse characteristics*

Figure 30: Comparison of Mean Slope for FAO85 Soil Type for STU Data

The graphs shows only a weak correlation between the mean dominant and the mean secondary slope for FAO85 soil types. The relationship is better defined for soil types with a higher frequency of occurrence in the STU data. Because dominant and secondary slope are used to provide an indicator of the diversity of characteristics of STUs limiting the comparison to those diverse STUs actually decreases the relationship between the mean dominant and secondary slope.

The relationship between the mean slope of FAO85 soil types extracted from the DEM and the mean dominant slope of the STU data is presented in Figure 31.



a) all STUs *b) STUs with diverse characteristics*

Figure 31: Comparison of Mean Slope for FAO85 Soil Type between DEM and STU Data

For both datasets the mean slope extracted form the DEM is considerably less than the slope given for STUs (x-coefficient: 3.1 for linear relationship). This is attributed to the resolution of the DEM (1,000 m). Common to both datasets is that with an increase in the number of relevant data for a soil type the relationship shows less variation for data pairs.

The relationship of mean height and slope between the STU and DEM data is influenced by the various steps applied for selecting relevant data, aggregating data to SMUs and calculating the mean values for FAO soil types. Furthermore, correlating the mean values does not carry any information about the variation of the parameter values around the mean. Thus, regression statistics of the relationship for the mean values between DEM and STU by FAO soil type should not be interpreted as presenting the relationship between the parameters for any soil type.

3.3.10 Other FAO85-Aggregated DEM-Based Criterion Factors

Besides height and slope the STU database does not contain any other data on topographic parameter. Investigating such parameters is therefore limited to comparing and disaggregating data at the level of the SMUs without recourse to STU data. Whether the SMUs are well defined by STUs is therefore irrelevant. The general characteristics of the topographic parameters can be estimated for soil types by using a weighted mean and variance. From these the confidence interval for the standard error of the mean can be derived to support separating soil types from each other.

The mean is calculated as:

$$\bar{x}_{FAO85} = \frac{\sum_{i=1}^{n} \bar{x}_{SMU_i} \bullet AREA_{FAO85.SMU_i}}{\sum_{i=1}^{n} AREA_{FAO85.SMU_i}}$$

where

$\bar{x}_{FAO85}$ parameter mean for FAO85 soil type

$\bar{x}_{SMU_i}$ parameter mean in SMU

$AREA_{FAO85.SMU_i}$ area of FAO85 in SMU (surface area or proportion)

n: No. of STUs with FAO85 soil class

The standard deviation of a parameter by FAO85 soil type is estimated from:

$$s_{FAO85} = \sqrt{\frac{\sum_{i=1}^{n} \left(s_{SMU_i}^2 + \bar{x}_{SMU_i}^2\right) \bullet AREA_{FAO85.SMU_i}}{\sum_{i=1}^{n} AREA_{FAO85.SMU_i}} - \bar{x}_{FAO85}^2}$$

where

s_{SMU} parameter standard deviation in SMU
$\bar{x}_{SMU_i}$ parameter mean in SMU
$\bar{x}_{FAO85}$ parameter mean for FAO85 soil type
$AREA_{FAO85.SMU_i}$ area of FAO85 in SMU (surface area or proportion)
n: No. of STUs with FAO85 soil class

- **Slope Aspect**

A topographic variable readily available from a DEM is slope aspect. For the AOI south-facing slopes should in general be warmer and drier than north-facing slopes, although there may be deviations from the rule by local effects. Commonly, aspect is defined as the direction of the maximum slope and is given using the azimuth designation, i.e. starting with 0 and increasing clockwise from north. The azimuth designation means that similarly facing slopes can differ by up to 360 degrees (slopes facing north). Therefore, an index for the degree of northerliness of a slope was computed. The index sets a value of 1 for slopes facing due north and a value of 0 for south-facing slopes, with a gradual change given by the azimuth in between. Aspect is not an attribute of the STU data and can only be established from the DEM for an SMU as a whole.

The mean degree of northerliness as defined above for the SMUs of the AOI is presented in Figure 32.

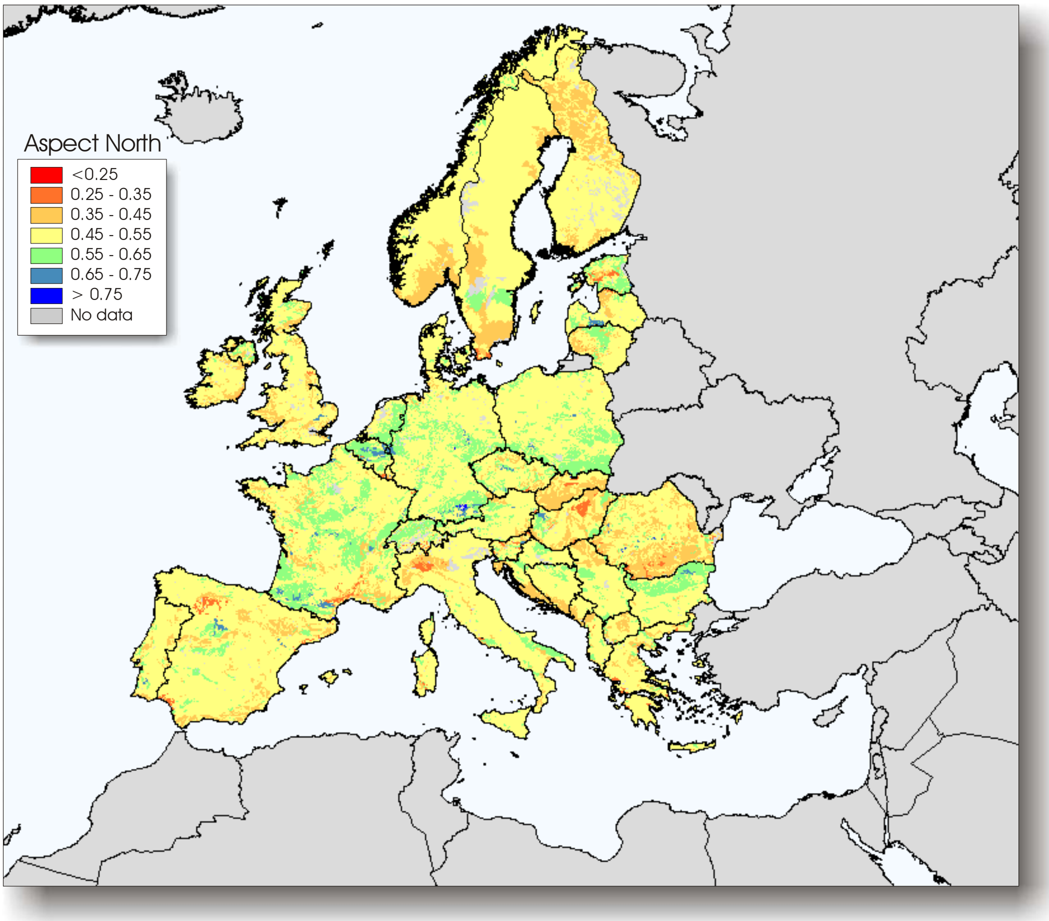


Figure 32: Map and Frequency of SMU Mean of Degree of Slopes Facing North

The map shows a distinct aspect for a number of SMUs, but no SMUs with only a northerly (value 1.0) or southerly (value 0.0) aspect. Notably southerly are SMUs on the slopes of the in the Cantabrian Mountains, the upper Po valley and northern Hungary. SMUs with a marked northerly aspect are found in the area of Zamora, French parts of the Pyrenees, Belgium and northern Alps. Distinct is also the change in aspect in the Danube valley in Bulgaria and Rumania.

The relative frequency of northern aspect in SMUs and the mean by FAO 85 soil type is presented in Figure 33.

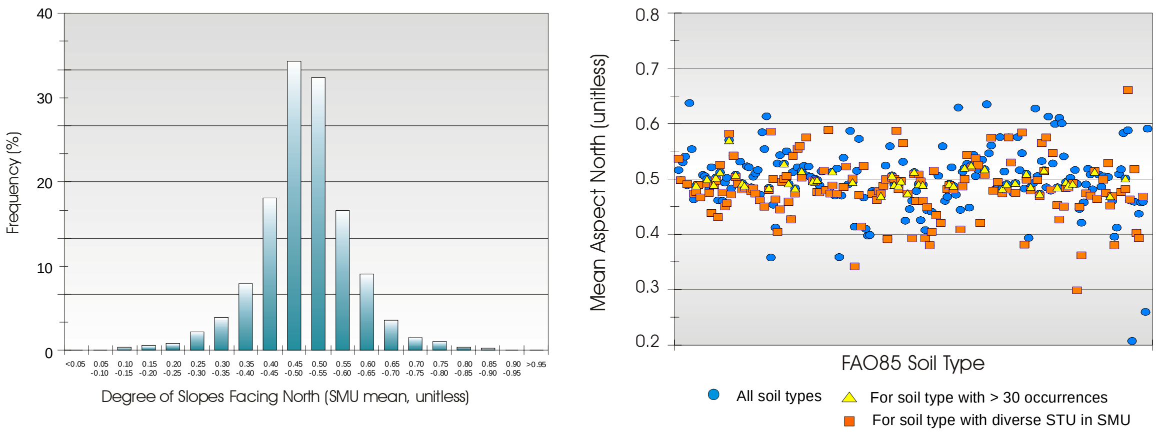


Figure 33: Relative Frequency of Northerly Aspect and Mean Northerly Aspect by FAO85 Soil Type

The frequency distribution shows that 50% of all SMUs are within ± 0.05 units from 0.5 and 62.5% are within ± 0.10 units. The standard deviation of the degree of north-facing slopes in the DEM for the AOI is 0.29, while the mean standard deviation within the SMUs is 0.27. Therefore, while there are some SMUs with a distinct aspect the delineation of the SMUs in general was not much affected by aspect. This was confirmed when estimating the mean slope by FAO 85 soil type. For soil types with a frequency of occurrence of 30 or more STUs no specific association with aspect could be determined. There is further no discernible difference between using all data and restricting the calculation of a mean value to those SMUs with diverse STUs. As a consequence, the simply using the degree to which slopes face north appears to be a poor criterion for the spatial allocation of STUs

- **Distance to Flow Network**

While the previously discussed topographic parameters are spatially discrete in that they only describe the grid cell (height) or the immediately neighbouring cells (slope, aspect), other topographic parameters are derived from treating the elevation data as a continuum. Since SMUs are often delineated following valleys characterizing SMUs by a surface flow parameter could be useful to characterize also the STUs. One such flow parameter is the accumulated upstream area of a grid cell. From setting a threshold for the area a flow network can be derived. Due to the very variable length of streams the accumulated upstream area itself is not a suitable parameter when comparing SMU characteristics. A degree of normalization between SMUs can be obtained by calculating the distance to the flow network, from which the average distance for an SMU can be determined.

A threshold of 100 km^2 was set for the accumulated upstream area to define the flow network. This size is exceeded by 90% of the SMUs in the AOI. The average distance to the flow network in the AOI is 4.5 km with a standard deviation of 3.2 km. Excluded were areas covered by SMUs without their own flow network. There areas largely consist of islands smaller than the threshold value set for the accumulated flow. Using the distance to these external flow networks would artificially increase the variability in the data.

The 95% confidence interval of the estimated mean by FAO85 soil type in the data with the mean distance to the runoff network with a minimum size of 100 km^2 is presented in Figure 34.

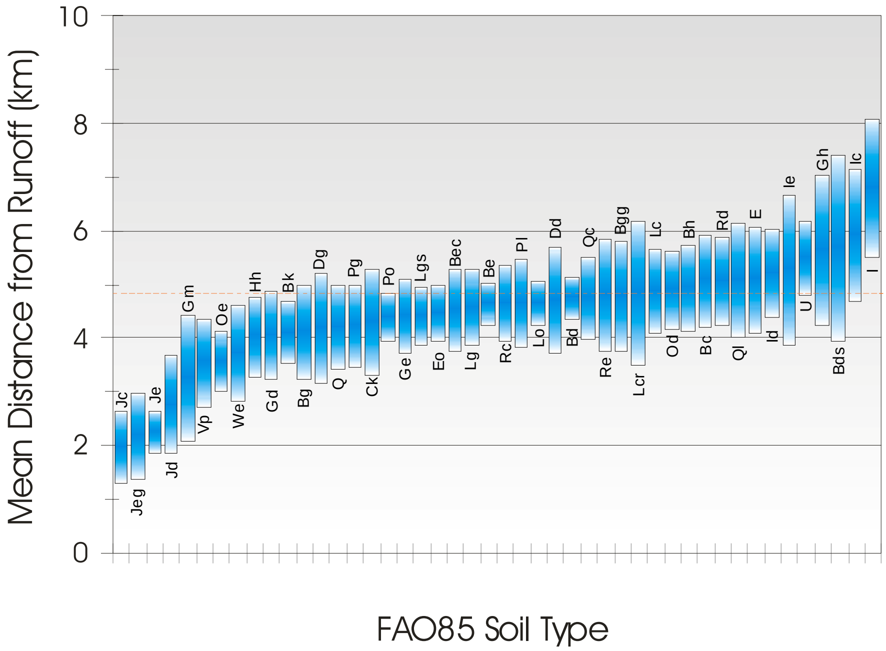


Figure 34: Confidence Interval (95%) of Estimated Mean Distance by FAO85 Soil Type to Runoff Network (>100km^2 catchment area)

Shown in the graph are the 46 FAO85 soil types with more than 30 occurrences in the SMUs of the AOI, ordered by the estimated mean. The graphs shows that *Fluvisols* are clearly closer to the runoff network than other soils, with *Lithosols* being found at the other end. Most soil types are, however, close to the mean distance of 4.7 km. A conceptual problem of using the distance to the runoff or drainage network is that no values can be estimated for grid cells which are located on the network.

- **Topographic Moisture Index**

Another index based on surface flow is the *Topographic Moisture Index* (TMI) (Beven & Kirby, 1979). The formulation is equivalent to the wetness index used by Burrough and McDonnell (1998):

$$TMI = \ln\left(\frac{As}{\tan(\beta)}\right)$$

where

As upstream area per unit contour length
β local slope angle

(Tenenbaum, *et al*., 2006)

For the TMI the mean for the AOI is 5.6 units with a standard deviation of 2.44 units. The mean TMI for the SMUs is 5.78 grid units with a standard deviation of 1.94. Also the TMI indicates a predisposition in the delineation of SMUs to follow flow accumulations. For the TMI no drainage network was defined and a value can be calculated for any SMU. The 95% confidence interval of the estimated mean TMI by FAO85 soil type is presented in Figure 35.

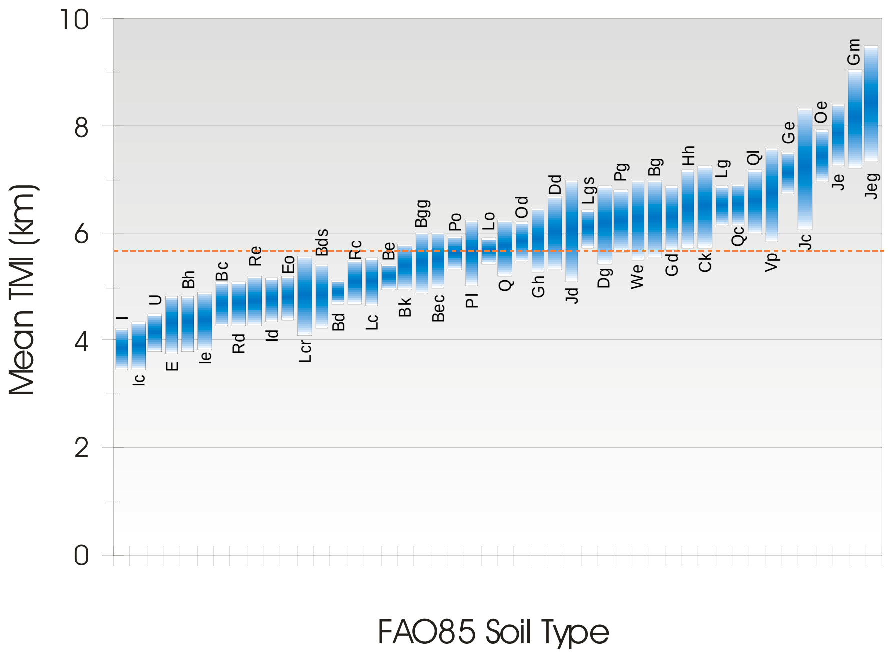


Figure 35: Confidence Interval (95%) of Estimated Mean TMI by FAO85 Soil Type

The 46 soil types show a generally similar arrangement of estimated mean TMI values to the arrangement of the mean distance to the runoff network. Quite different is the order of the soil types by estimated mean. *Lithosols* generally have lower TMI values than other soil types while *Fluvisols* are found at the higher end. For the TMI the confidence intervals show less overlap between soil types than for the distance to the runoff network. This should allow a better spatial separation of STUs within SMUs when using the TMI instead of the simple distance to a drainage network.

3.4 Decision Criteria Constraints

Due to the binary nature decision criterion constraints set unconditional limits to the areas to which STUs can be assigned. The spatial layers of constraint defined in the MCE are therefore integral elements of the final spatial layer. This characteristic of constraints raises conceptual and practical issues. When allocating only those STUs that describe soils the constraints define areas without soil, such as bare rock, permanent snow fields or water. These non-soil areas should no vary much over time since changes in the geographic distribution of such constraints would stand for geographic changes in the occurrence of soil[20]. An alternative processing approach is to include STUs describing non-soil areas in the MCE. Yet, this approach may introduce inconsistencies in the occurrence of non-soil areas between the MCE result and other land cover / use data, such as Corine LC data. Using this approach is hindered by the unreliable classification of some non-soil areas in the ESDB. A third option would be to ignore non-soil areas in the MCE and add such areas after the process. However, this approach runs the risk of allocating STUs to areas which are then concealed by non-soil land cover types.

None of the options satisfies all needs. For this study it was decided to exclude non-soil areas from the MCE and define these areas as constraints. For the delineation of the spatial extent of the layer of SMUs with exclusively non-soil cover is used in preference to the Corine LC layer. This approach is more in accordance with the study objective of giving priority to the soil data and using other data as ancillary information for the geographic allocation of the data rather than substituting them. Using the non-soil layer means that artificial surfaces (classes "*Town*" and "*Soil disturbed by man*") are not defined as constraints to allocating STUs.

3.5 Criterion Standardization

Criterion factors are generally defined for parameters with continuous range of values (ratio or interval scale). This allows using transitional ranges instead of using fixed threshold with Boolean membership for the degree to which a value belongs to a criterion. The nature of the transitional affiliation can be defined by fuzzy set membership functions.

Before the criteria can be compared in the MCE the factors are processed in several stages:

- **Standardization Function**

  When comparing factors measured on different scales the factor values need to be standardized to a common range. Rather than using fixed limits or linear scaling to define the relationship between the STU

[20] Areas considered to be covered by soil may actually change over time, e.g. following the construction of dams or at excavation sites.

(evidence) and the ancillary (decision) data the membership is defined by fuzzy set functions.

Depending on the characteristics of the data, such as given ranges or mean values, a type of fuzzy set membership function (sigmoidal, J-shaped, Boolean, etc.) is defined for a criterion. The four inflection points of the fuzzy set membership functions are defined in the units and ranges of the transformed data.

- **Transformation from STU Values to Ancillary Data**

  As shown there is no direct and complete representation of STU parameters in the ancillary data. The minimum and maximum height often only cover part of the range of the elevation in the area covered by an SMU in the DEM. For slope a decidedly different distribution between STU and DEM was found. To meaningfully relate the STU data to the ancillary data and define the inflection points for the fuzzy set membership functions the STU data are transformed. The transformation method applied depends on the representation of the criterion factor in the STU database.

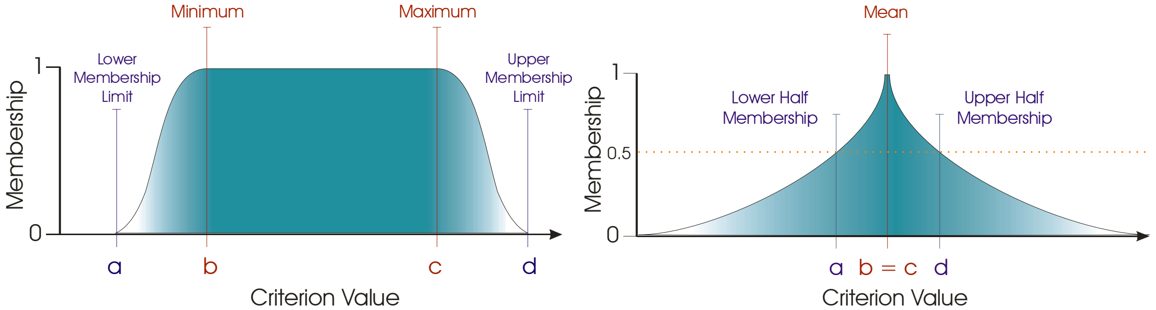
- **Factor Weight**

  The standardized factors are then weighted according to their importance in defining the suitability of a location for the objectives.

The correlation of the soil typological data to the ancillary data is closely linked to the definition of the inflection points of the membership function. The stages of preparing the MCE data are presented in detail hereafter.

3.5.1 Membership Functions for Criterion Factors

Criterion scores are calculated using fuzzy logic to define a membership function (MF). For the topographic parameters either sigmoidal or J-shaped MF is used. A graphical example of the MFs is shown in Figure 36.



a) sigmoidal *b) J-shaped*

Figure 36: Sigmoidal and J-Shaped Membership Functions

The fuzzy set membership functions (FMFs) shown in the graph are symmetrical, but applied in the MCE were also non-symmetrical functions and monotonically increasing / decreasing functions.

A sigmoidal MF with four control points is used where a criterion covers a range of values with the same magnitude for the membership. This type of MF is used for attribute data of ordinal type, such as for the slope parameter in the STU table and where minimum and maximum height are aggregated to mean values. For each category full membership is defined for the range of values between the inner inflection points. Values for the start and end control points need to be set to cover the possible range of values. For values beyond the outer control points the membership becomes 0.

The symmetrical J-shaped function defines a single point to full membership and all other factors to a membership value that varies with the criterion values in the spatial units. In contrast to the sigmoidal function under practical conditions the J-shaped function only approaches a membership value of 0, but does not reach it. Thus, the J-shaped MF can be applied where the outer control points are not well defined. This is the case for factors characterizing FAO85 soil types rather than STU attributes. The inner inflection points can be set to the mean factor value and the distance to the outer inflection points is defined by a measure of the variability of values around the mean, such as the standard deviation. The two sides of the function are not necessarily symmetric since the distance *a* -> *b, c* may differ from the distance *b, c* -> *d*.

The choice of a MF to be used for a criterion factor is not always apparent. For factors related to SMUs a sigmoidal membership function seems the appropriate function for the factor membership. For an SMU the range of values in the area is known from the ancillary data and the outer control points can be set accordingly. For factors aggregated to FAO 85 soil types the maximum range is given by the values in the AOI, not by an SMU. Setting the outer control points to these extreme values may render the function ineffective for a number of SMUs, because only a fraction of the range in present in the area of the SMU. This may also be the case when using a narrow range for the outer control points. To use a sigmoidal MF for factors aggregated to FAO 85 soil types the outer control points are therefore set to 2 standard deviations from the mean, instead of 1 standard deviation used for the J-shaped membership function.

3.5.2 Consolidate Relationship of Soil Typological to Ancillary Data for Factors

For a criterion factor the relationship between the STU attribute and the corresponding ancillary data can be improved (height) or even has to be established (slope) when the parameters of MF are specified. The nature of the STU attribute data types, mainly classified or on nominal scale, makes defining the relationships to ancillary data on a ratio scale more complex.

- **Membership Function for Height**

One option of defining the FMF for height is to set the inner control points to the minimum and maximum height of the STU data as found in the database. The outer control point *a* is set to the minimum of height (MIN(E_{SMU})) or elevation ($ZMIN_{SMU}$) and control point *d* to the maximum of height (MAX(E_{SMU})) or elevation ($ZMAX_{SMU}$) in an SMU. This option maintains the original data, but sets a wide range of values for full membership and the outer control points.

To account for differences between the typological and DEM data values for the height parameter of the STU were associated with elevation in the DEM by defining for each SMU a linear function, which relates the STU height to the DEM elevation. Slope and constant of the function are defined by the minimum and maximum height/elevation data pairs of an SMU. For each SMU the transformed height Z^t is then found by the linear equation:

$$Z^t_{STU.SMU} = m_{SMU} \cdot Z_{STU.SMU} + c_{SMU}$$

with

$$m_{SMU} = \frac{MAX(E_{SMU}) - MIN(E_{SMU})}{MAX(Z_{SMU}) - MIN(Z_{SMU})}$$

$$c_{SMU} = MAX(E_{SMU}) - m_{SMU} \cdot MAX(Z_{SMU})$$

where

Z^t	transformed STU height
$MAX(E_{SMU})$	maximum elevation value in DEM for an SMU
$MIN(E_{SMU})$	minimum elevation value in DEM for an SMU
$MAX(Z_{SMU})$	maximum STU height value for an SMU
$MIN(Z_{SMU})$	minimum STU height value for an SMU

When using a sigmoidal FMF the control points should be set to cover the range of elevation of the attribute data. In this case, the external control points should be set to cover at least the minimum and maximum elevation of the SMU in the DEM. To avoid setting the value of the external control points to the internal inflection points, which would result in an error when processing the data, the range for the external control points is extended by 1m on either end.

This method assumes that the ranges given for the STUs are present in the DEM in the area covered by the SMU, at least with a partial overlap. The requirement may be expressed as:

$ZMIN_{STU.SMU}$ < MAX(DEM_{SMU}) OR $ZMAX_{STU.SMU}$ > MIN(DEM_{SMU})

For the AOI the first condition is not met by 73 STUs and 55 STUs do not meet the second condition. At least partial overlap in height exists for all SMUs with valid data. Partial overlap is sufficient to define all inflection points when using this method. Using a J-shaped MF the external control points can be set to values smaller than the range of the elevation in the DEM, because the membership will not be set to zero for any value.

To decrease the influence of outliers on the function relating the height values of the STU data to elevation in the SMU, the range of values used in defining the function can be reduced by excluding a percentile of values at the extreme ends (trimming). In the evaluation the lower and upper range of elevation data was trimmed by 5% (Z^{t5}, 90% of values) and 10% (Z^{t10}, 80% of values).

The control points of the MF are then set by a combination of the transformed values. The control point values depend on the processing applied:

- When using a J-shaped MF the central control points are set to the minimum of [$ZMIN^{t5}_{SMU}$, $ZMIN^{t10}_{SMU}$] and control point *b* to the maximum of [$ZMIN^{t5}_{SMU}$, $ZMIN^{t10}_{SMU}$]. Point *c* is set to the minimum of [$ZMAX^{t5}_{SMU}$, $ZMAX^{t10}_{SMU}$], while point *d* is set to the maximum of [$ZMAX^{t5}_{SMU}$, $ZMAX^{t10}_{SMU}$].
- The control points of a sigmoidal MF need to cover the whole range of elevation data within an SMU in the DEM. Therefore, the values for the outer control points is given by the transfer function derived from elevation data without trimming the range (Z^{t0}). The value for control point *a* is set to the minimum of [$ZMIN^{t0}_{SMU}$, $ZMIN^{t10}_{SMU}$] and control point *b* to the maximum of [$ZMIN^{t0}_{SMU}$, $ZMIN^{t10}_{SMU}$]. Point *c* is set to the minimum of [$ZMAX^{t0}_{SMU}$, $ZMAX^{t10}_{SMU}$], while point *d* is set to the maximum of [$ZMAX^{0}_{SMU}$, $ZMAX^{t10}_{SMU}$].

Under the first option the values for the control points may contain values from the original STU. Instead, under the second option only transferred values are used, although the lowest and highest elevation values appear at least once as values of an external control point.

- **Membership Function for Slope (dominant and secondary)**

Given the differences in the proportion of the area of a slope class in the typological database and the distribution of slope in the spatial layer, adhering strictly to the upper and lower limits defining the slope classes to define the control points of the fuzzy set membership function will lead to only partial overlap of areas or even none at all.

With the generally low correlation between the typological slope data and the ancillary spatial data derived from the DEM the nominal STU values were compared to transformed values with the aim of improving the association between the data sets. Considered as processing option for the control points were:

a) retain nominal class limits for central control points, extend range of external control points (retain nominal range);

b) adjust the general distribution of mean slope and class limits to the range of slope in the ancillary data (proportional scaling).

Not considered was the option of adjusting slope class limits to maintain the STU area. Under this option the share of a slope class in the total area of an SMU [PCAREA.STU_ORG] is transferred to a corresponding area in the DEM of ranked slope values. The reason for disregarding the option is that the limits defining the slope classes differ frequently and considerably from the nominal values and vary between SMUs.

Retaining the nominal class limits as values for the central control points requires flexibility in adjusting the external control points of the MF. This is particularly relevant when using a sigmoidal MF, because the membership values are set to zero beyond the outer control points (see Figure 37). The advantage of the method is that the original data are largely retained, but that the decision may be based on very low membership values.

When using **proportionally scaling** the association between the STU and the DEM slope values is approximated by matching the general relationship between the STU data and the DEM across the AOI to the specific conditions of an SMU. This can be realized by either scaling the class limits or the class mean values using a linear function.

Proportional scaling of class limits leads to a MF with a plateau ($b < c$) for full membership. It was achieved by adjusting the parameters of the general distribution of slope classes to the nominally present minimum and maximum slopes in an SMU, as given by the linked STUs. The same procedure was used to scale the class mean values.

The approach to scaling class limits is very similar to the one used to scale the STU height data. For a given SMU the limits of the slope classes given by the STU attribute are scaled to the range of slope values present in the DEM. However, height ranges are set for each STU, while the ranges for slope are set globally as classes. Thus, the ranges of height may overlap within an SMU, but the association with slope defines exclusive ranges.

For scaling mean values of a slope class the central control points are set to the same value for full membership ($b = c$). The class mean provides the values for the central control points, while the outer control points (a, d) are set to the mean of the neighbouring classes. A graphical presentation of the functions is given in Figure 37.

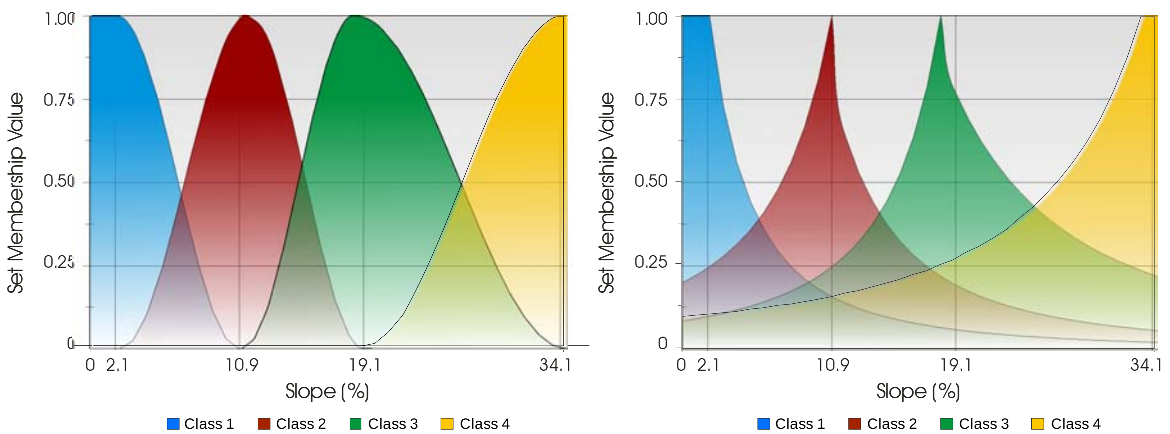


Figure 37: Sigmoidal and J-Shaped Fuzzy Set Membership Function for Slope Classes Applied to DEM (Non-filtered)

For the sigmoidal and J-shaped MF with common values for the central control points (b = c) the graph shows the mean slope in the DEM for the four slope classes and the zones of overlap between the classes when setting the external control points to the mean of the neighbouring classes. The cross-over values between membership classes are close to 0.5 and the nominal class limits. When using a sigmoidal MFs the external control points have to be adjusted to the range of values in the spatial data (not shown), regardless of whether the central points are adjusted or not. This is not necessarily the case when using a J-shaped MF since the membership values are > 0 across the whole range of slope values in the spatial layer.

The generally low frequency of high values for slope in the DEM may lead to calculating ranges for the central control points with poor representation in the DEM layer. The function can be made more robust by removing a percentile of data at the low and high end. This can be attempted by excluding values outside a fixed range of standard deviations from the mean or by excluding a number of values based on a percentile. However, this in not generally the case. Instead, the distribution of slope values is frequently not normal and asymmetric, mostly with positive skew. For the mean value the asymmetry in the distribution of the values can be addressed, e.g. by using the median for the central value instead of the arithmetic mean. For the generalized mean slope of a class not much difference was found between the mean and the median and the arithmetic mean was used. To limit the range of values to define the parameters of the transfer function the use of the standard deviation calculated directly from the slope of the DEM is not a suitable approach and the range of data was trimmed using percentiles.

There does not appear to be one method which satisfies all demands on the type of FMF to be used. Which method to use depends on the definition of the evaluation task. This evaluation of the MCE/MOLA method for spatial allocation concentrated on three options:

a) membership function with inner control points set to nominal lower and upper limit of slope class *n* (Sn_{LOW}, Sn_{HIGH}), forming a plateau for membership, and the outer control points to the mean slope of the adjacent slope class ($S(n-1)_{MEAN}$, $S(n+1)_{MEAN}$) without further adjustment;

b) membership function with inner ($S^t n_{LOW}$, $S^t n_{HIGH}$) and outer ($S^t(n-1)_{MEAN}$, $S^t(n+1)_{MEAN}$) control points adjusted by linear transfer function derived from trimming DEM data by 5% at low and high end;

c) membership function with central control points set to mean slope for class ($S^t n_{MEAN}$) and outer control points to the mean of the adjacent slope class ($S^t(n-1)_{MEAN}$), ($S^t(n+1)_{MEAN}$), with all control points adjusted by linear transfer function derived from trimming DEM data by 5% at low and high end.

In the first option the values of the STU are used as found in the database. The membership value is set to 1 across the whole range of values between the nominal class limits. To account for slope values present in the DEM but not covered by the STU classes an element of fuzziness is added to the function in setting the outer control points to the mean of the neighbouring slope class.

The second option allows for some adjustment of the class limits by the range of values found in the spatial layer. The linear function used to shift the control points is based on spatial data with a limited range, excluding 5% of values at either end of the value scale.

The third option uses a FMF where both central control points are set to the mean value of the slope class and the external control points to the mean value of the neighbouring class. A linear function is used to shift all control points, based on spatial data with a limited range, excluding 5% of values at either end of the scale in the DEM data for an SMU.

- **Membership Function for FAO85-Aggregated Height**

For the area-weighted average height for an FAO85 soil type a J-shaped MF is used. The outer control points are set to the mean minimum (*a*) and mean maximum (*d*) height. Full membership (*b*, *c*) is given for the mean height.

The data considered for aggregating height information by FAO 85 soil type were:

a) area-weighted aggregation of **STU height** information ([ZMIN.STU_SGDBE], [ZMAX.STU_SGDBE]);

b) area-weighted aggregation of linearly **transformed STU height** information ($ZMIN^t$, $ZMAX^t$);

c) area-weighted aggregation of SMU **DEM elevation** information (MIN[DEM_{SMU}]; MAX[DEM_{SMU}]).

Aggregation options a) and b) use information on the basis of STUs associated with an SMU. Information from the DEM covers the SMU as a whole. In the aggregation of DEM data the minimum and maximum elevation in the area covered by the SMU substitute the minimum [ZMIN] and maximum [ZMAX] STU values.

The relationship between the mean elevation data for FAO 85 soil types and the height data derived from the recorded STU and the transformed STU data is presented in Figure 38.

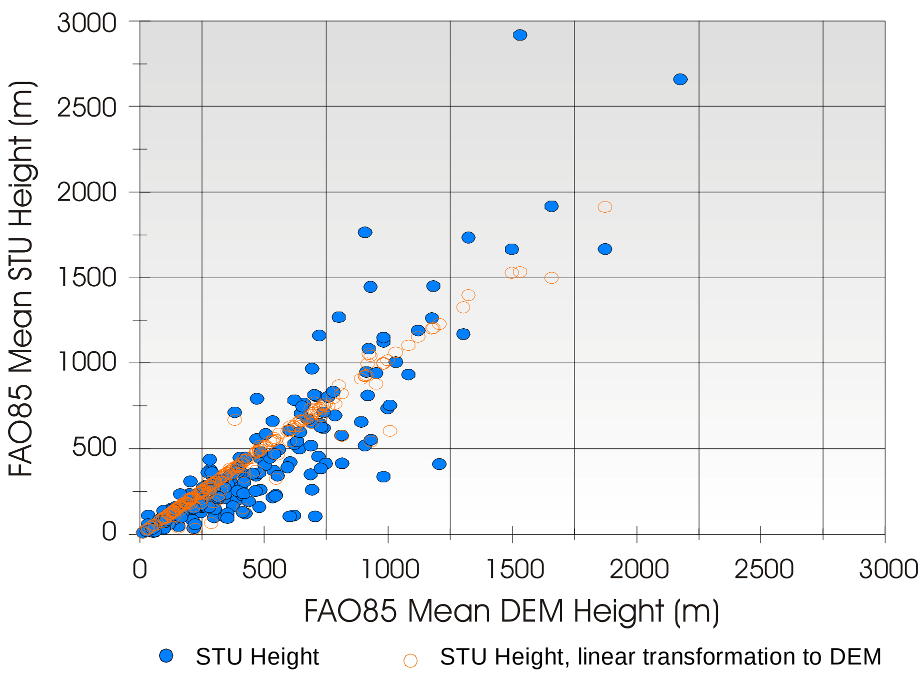


Figure 38: Comparison of Mean Elevation of DEM with Mean Recorded and Transformed STU Height for FAO 85 Soil Types

There is a close relationship between the mean elevation of FAO 85 soil types when the minimum and maximum values are based on the DEM. The relationship is less defined between the DEM-based mean values and the aggregated height values recorded in the STU database.

There would appear to be a fair amount of redundancy when using all three data aggregation options. The closet relationship with the STU height and the DEM elevation data is probably available from aggregating the transformed height data to FAO 85 soil types. Included in the height criterion for FAO 85 classes are also "*Rock outcrops*" to allow positioning the land cover type in complex SMUs.

For the FMF the inner control points (*b*, *c*) are always set to the area-weighted mean height estimated for an FAO85 soil type. The external control points *a* and *d* for the for the J-shaped MF are set to ± 1 σ from the estimated mean. For the sigmoidal MF the external control points are set to ± 2 σ from the estimated mean[21].

The area-weighted mean height may be derived from all DEM data of the area covered by an SMU or from data trimmed at the lower and higher end of the elevation range. The mean FAO 85 heights calculated from the different SMU mean elevation data resulted in almost identical values. The standard deviation was computed from the full range of elevation data in the DEM rather than the reduced data since the parameter is used to define the external control points of the FMF.

- **Membership Function for FAO85-Aggregated Criterion Factor Slope**

The difference in slope between the STU and the DEM is such that applying an average slope derived from the STU data to the DEM as a criterion factor is of little practical use. Too often the average slope of the soil types would be outside the range of values for the slope derived from the DEM.

The central control point (*b*, *c*) for the mean slope of FAO85 soil types is set to the area-weighted estimated mean slope. The external control points (*a*, *d*) are set at one standard deviation from the mean. The standard deviation was chosen over the mean minimum and maximum slope for the external inflection points because the latter values were generally close to extreme conditions (zero or > 25%). Such values are of very limited use as external inflection points for a J-shaped membership function, which set the membership to 0.5 foe the corresponding values.

Analogous to estimating the mean FAO 85 height the mean slope can be estimated from all DEM data within the area covered by an SMU or from data trimmed at the lower and higher end of the range. When computing the mean FAO 85 slope as the area-weighted mean SMU slope the resulting values differed by insignificant amounts. However, when using the mean slope of a slope class where the values in the DEM were trimmed by 10% (lower and higher end) and scaled to the range of slope values in the DEM ($S^{t10}n_{MEAN}$) the estimates for the mean FAO 85 slope differed. The differences in the estimates are presented in Figure 39.

---

[21] The standard deviation is used here instead of percentiles, because the values are aggregated from mean values and are assumed to approximate a normal distribution.

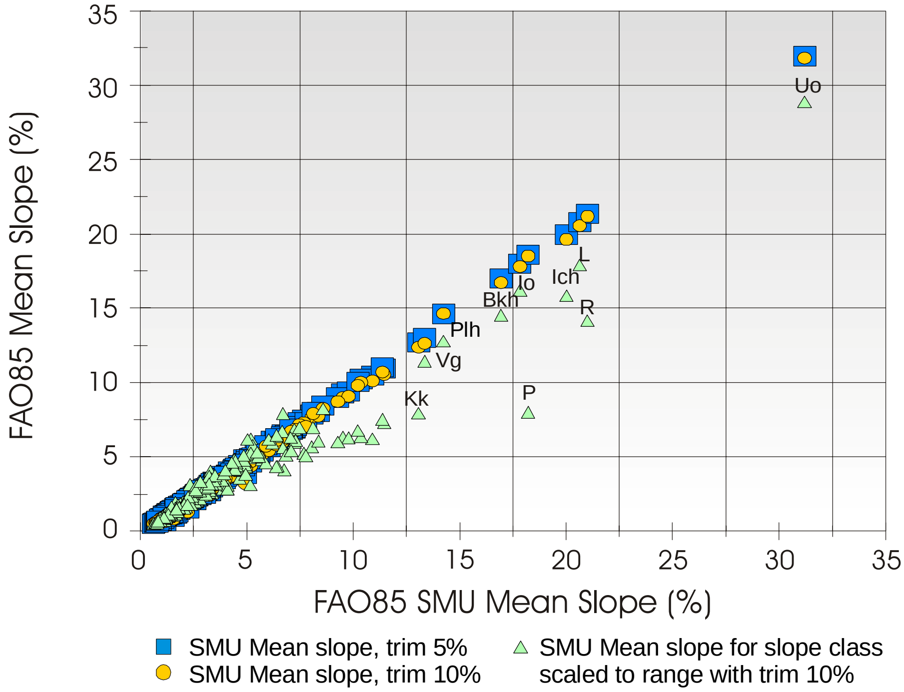


Figure 39: FAO 85 Mean Slope Estimated from various Processing Options for Calculating SMU Mean Slope

The graph shows lower mean FAO 85 slope estimates when using the trimmed and scaled mean values calculated for each slope class within an SMU. The difference is attributed to the more effective reduction in higher slope values when performed for each slope class within an SMU instead of generally for the SMU as a whole. The values for the external control points were set by the estimate of the standard deviation calculated from the full range of values in the DEM was used.

- **Membership Function for FAO85-Aggregated Criterion Factor TMI**

Although the TMI factor is aggregated for FAO 85 soil types it generally uses a monotonically changing membership function. The reason is the spatial distribution of the TMI values. Areas on or close to the river have higher values of TMI than those farther away. Deviating from height or slope the soils related to the TMI are not expected to be dominant within a range of values. Rather, soils are either generally closer to the flow network or found at a distance. This behaviour is therefore modelled by monotonically increasing membership functions for soils closer to the flow network (higher TMI values; generally for *Fluvisols*) and monotonically decreasing for soils found at a distance from the river network. Whether an increasing or decreasing function should be used is largely determined by the standard deviation of the TMI. When zero is included at 2 σ from the mean a monotonically decreasing function is used, otherwise the inflection points are set for a monotonically increasing function.

To avoid generating an STU layer from the TMI line network and to allow for a broader coverage if TMI values in flat areas the TMI line network was modified by a distance function as:

$$TMI_d = MAX\left(TMI \bullet e^{s \bullet d}, TMI\right)$$

where

TMI_d	distance-weighted TMI
s	slope (%)
d	distance (1000m)

The distance-weighted TMI_d is the maximum value of the TMI and the function output. The application of the function is further limited to slopes < 3%.

- **Membership Function for Criterion Factor Land Use and Cover**

Information on land use is nominal, and discrete, by nature. The criterion factors are therefore not defined by FMFs, but as Boolean conditions. Most land use data can be defined as criterion factor, some may serve to define constraints. Defining constraints directly from the spatial land use layer is not recommended, because the area demand from the STUs may differ from the distribution of the factor in the ancillary data.

Technically, the information on land use in the STU table can be linked to the general land use / cover classes of the Corine LC data. Yet, without a history of land use and changes allocating STUs by only the current land use could lead to STUs being geographically shifted depending on changes in land use.

An association was establish between largely permanent types of land cover. For the fields [USE_DOM.STU_SGDBE] and [USE_SEC.STU_SGDBE] these are the land uses "*Moor*" and "*Bare land*". Contrary to the categories "111" to "666" the STUs with these land use categories also have additional data on soil attributes, such as a soil type. For the 143 STUs of "*Moor*" the soil type is "*Histosol*" for 114 STUs. Less obvious is the combination of 6 STUs with "*Moor*" as the dominant land use type and "*Ranker*" as soil type. "*Bare land*" is the dominant land use given for 39 STUs. Of these 20 are "*Regosols*" and 17 are "*Lithosols*", two are given as "*Orthic Podzol*". "*Moors*" of the STU database can be associated with the Corine LC classes "322" (*Moors and heathland*) and "412" (*Peat bogs*). "*Bare land*" can be associated with CLC class "333" (*Sparsely vegetated areas*), but not class "332" (*Bare rock*), because the class has soil data attached. "*Moor*" is given as secondary land use for 49 STUs, of which 10 also have this category as the dominant land use. For 35 STUs the soil type is "*Histosol*". The category "*Bare land*" is not used as a secondary land use.

Apart from the information of the land use fields [USE_DOM.STU_SGDBE] and [USE_SEC.STU_SGDBE] the information on bare land given in the field [FAO85FU.STU_SGDBE] could be associated with the Corine LC data. The non-soil class "*Rock outcrop*" (Code "666") can be allied with the CLC class 332 (*Bare rock*). Areas of rock outcrops are generally not specifically delineated in the SGDBE as individual SMUs but occur in combination with STUs with soil data. They are therefore treated as a soil type.

With the differences found for land use between the ESDB and the ancillary spatial data only the associations for *peat*, in it various forms, and for *bare land* were set as criterion factors. For both factors a monotonically decreasing sigmoidal membership function is defined based on the distance to such an area in the spatial data. Full membership is set to the area of the land use in the ancillary and no membership to areas at a distance > 5 km.

- **Membership Function for Criterion Factor Obstacle to Roots**

The option applied to define the FMF for the obstacles-to-roots criterion factor follows the procedure used for the slope criterion factor with a sigmoidal MF and an adjustment of the control points to the range of values in the spatial data using a linear transfer function.

- **Membership Function for Criterion Factor Water Regime**

For the set membership function of the water regime criterion factor a sigmoidal function is used with adjusted control points. In a deviation from definition of the obstacle-to-roots criterion factor the membership function of the first class ("*Not wet within 80 cm for over 3 months, nor wet within 40 cm for over 1 month*") is monotonically increasing, since this class becomes more dominant with slope.

- **Summary of Transformations and Control Point Settings**

The FMF and parameters for control points for the criterion factors of the multi-criteria decision process are given in Table 16.

Table 16: Membership Function Parameters for Criterion Factors

<table>
<tr><th rowspan="2">Criterion</th><th>Member-ship Function</th><th colspan="4">Transformation and Control Point Values</th></tr>
<tr><th>Type</th><th>a</th><th>b</th><th>c</th><th>d</th></tr>
<tr><td rowspan="6">Height</td><td colspan="5">Retain STU class limits for ZMIN and ZMAX</td></tr>
<tr><td>J-shaped</td><td>MIN[$ZMIN_{SMU}$, MIN[E_{SMU}]] -1</td><td>MAX[$ZMIN_{SMU}$, MIN[E_{SMU}]]</td><td>MIN[$ZMAX_{SMU}$, MAX[E_{SMU}]]</td><td>MAX[$ZMAX_{SMU}$, MAX[E_{SMU}]]+1</td></tr>
<tr><td>sigmoidal</td><td>MIN[$ZMIN^{t0}_{SMU}$, $ZMIN^{t5}_{SMU}$]-1</td><td>MAX[$ZMIN^{t0}_{SMU}$, $ZMIN^{t5}_{SMU}$]</td><td>MIN[$ZMAX^{t100}_{SMU}$, $ZMAX^{t95}_{SMU}$]</td><td>MAX[$ZMAX^{t100}_{SMU}$, $ZMAX^{t95}_{SMU}$]+1</td></tr>
<tr><td colspan="5">Linear transformation with 5% and 10% trim of elevation range:
$$Z^{t}_{STU.SMU} = m_{SMU(t)} \cdot Z_{STU.SMU} + c_{SMU(t)}$$</td></tr>
<tr><td>J-shaped</td><td>MIN[$ZMIN^{t5}_{SMU}$, $ZMIN^{t10}_{SMU}$] -1</td><td>MAX[$ZMIN^{t5}_{SMU}$, $ZMIN^{t10}_{SMU}$]</td><td>MIN[$ZMAX^{t5}_{SMU}$, $ZMAX^{t10}_{SMU}$]</td><td>MAX[$ZMAX^{t5}_{SMU}$, $ZMAX^{t10}_{SMU}$]+1</td></tr>
<tr><td>sigmoidal</td><td>MIN[$ZMIN^{t0}_{SMU}$, $ZMIN^{t10}_{SMU}$]-1</td><td>MAX[$ZMIN^{t0}_{SMU}$, $ZMIN^{t10}_{SMU}$]</td><td>MIN[$ZMAX^{t0}_{SMU}$, $ZMAX^{t10}_{SMU}$]</td><td>MAX[$ZMAX^{t0}_{SMU}$, $ZMAX^{t10}_{SMU}$]+1</td></tr>
<tr><td rowspan="5">Dominant Slope</td><td colspan="5">Retain nominal class limits, apply mean for $S4_{HIGH}$:
$SDOM1_{LOW}$ = 0, $SDOM2_{LOW}$ = 8, $SDOM3_{LOW}$ = 15, $SDOM4_{LOW}$ = 25
$SDOM1_{HIGH}$ = 8, $SDOM2_{HIGH}$ = 15, $SDOM3_{HIGH}$ = 25, $SDOM4_{HIGH}$ = 34.06</td></tr>
<tr><td>J-shaped</td><td>$SDOM(n\text{-}1)_{LOW}$
for n=1: $SDOM1_{HIGH}$</td><td>$SDOMn_{LOW}$
for n=1: $SDOM1_{HIGH}$</td><td>$SDOMn_{HIGH}$
for n=4: $SDOM4_{LOW}$</td><td>$SDOM(n+1)_{HIGH}$
for n=4: $SDOM4_{LOW}$</td></tr>
<tr><td colspan="5">Linear transformation with 10% trim:
$$SDOM^{t10}_{STU.SMU} = m_{SMU(P10)} \cdot SDOM_{STU.SMU} + c_{SMU(P10)}$$</td></tr>
<tr><td>J-shaped</td><td>$SDOM^{t10}(n\text{-}1)_{MEAN}$
for n=1: $SDOM^{t10}1_{HIGH}$</td><td>$SDOM^{t10}n_{LOW}$
for n=1: $SDOM^{t10}1_{HIGH}$</td><td>$SDOM^{t10}n_{HIGH}$
for n=4: $SDOM^{t10}4_{LOW}$</td><td>$SDOM^{t10}(n+1)_{MEAN}$
for n=4: $SDOM^{t10}4_{LOW}$</td></tr>
<tr><td>sigmoidal</td><td>$SDOM^{t10}(n\text{-}1)_{MEAN}$
for n=1: $SDOM^{t10}1_{MEAN}$</td><td colspan="2">$SDOM^{t10}(n)_{MEAN}$</td><td>$SDOM^{t10}(n+1)_{MEAN}$
for n=4: $SDOM^{t10}4_{MEAN}$</td></tr>
<tr><td rowspan="5">Secondary Slope</td><td colspan="5">Retain nominal class limits, apply mean for $S4_{HIGH}$:
$SSEC1_{LOW}$ = 0, $SSEC2_{LOW}$ = 8, $SSEC3_{LOW}$ = 15, $SSEC4_{LOW}$ = 25
$SSEC1_{HIGH}$ = 8, $SSEC2_{HIGH}$ = 15, $SSEC3_{HIGH}$ = 25, $SSEC4_{HIGH}$ = 34.06</td></tr>
<tr><td>J-shaped</td><td>$SSEC(n\text{-}1)_{LOW}$
for n=1: $SSEC1_{HIGH}$</td><td>$SSECn_{LOW}$
for n=1: $SSEC1_{HIGH}$</td><td>$SSECn_{HIGH}$
for n=4: $SSEC4_{LOW}$</td><td>$SSEC(n+1)_{HIGH}$
for n=4: $SSEC4_{LOW}$</td></tr>
<tr><td colspan="5">Linear transformation with 10% trim:
$$SSEC^{t10}_{STU.SMU} = m_{SMU(P10)} \cdot SSEC_{STU.SMU} + c_{SMU(P10)}$$</td></tr>
<tr><td>J-shaped</td><td>$SSEC^{t10}(n\text{-}1)_{MEAN}$
for n=1: $SSEC^{t10}1_{HIGH}$</td><td>$SSEC^{t10}n_{LOW}$
for n=1: $SSEC^{t10}1_{HIGH}$</td><td>$SSEC^{t10}n_{HIGH}$
for n=4: $SSEC^{t10}4_{LOW}$</td><td>$SSEC^{t10}(n+1)_{MEAN}$
for n=4: $SSEC^{t10}4_{LOW}$</td></tr>
<tr><td>sigmoidal</td><td>$SSEC^{t10}(n\text{-}1)_{MEAN}$
for n=1: $SSEC^{t10}1_{MEAN}$</td><td colspan="2">$SSEC^{t10}(n)_{MEAN}$</td><td>$SSEC^{t10}(n+1)_{MEAN}$
for n=4: $SSEC^{t10}4_{MEAN}$</td></tr>
<tr><td rowspan="2">Average Height by FAO85</td><td colspan="5">Area-weighted elevation mean with 10% trim:
$$\bar{E}_{FAO85.P10} = \frac{\sum \bar{E}_{P10}.SMU \cdot AREA_{FAO85}SMU}{\sum AREA_{FAO85}}$$</td></tr>
<tr><td>J-shaped</td><td>$\bar{E}_{FAO85.P10} - 1\sigma$</td><td colspan="2">$\bar{E}_{FAO85.P10}$</td><td>$\bar{E}_{FAO85.P10} + 1\sigma$</td></tr>
<tr><td>Average Slope by FAO85</td><td colspan="5">Area-weighted slope mean with 10% trim for general FAO85 slope:
$$\overline{SDOM}_{FAO85(s)} = \frac{\sum SDOM^{t10}n_{FAO85(s)}.SMU \cdot AREAn_{FAO85(s)}.SMU}{\sum AREAn_{FAO85(s)}}$$</td></tr>
</table>

Criterion	Membership Function *Type*	Transformation and Control Point Values *a*	*b*	*c*	*d*
	J-shaped	$\overline{SDOM}_{FAO85(s)} - 1\sigma$	$\overline{SDOM}_{FAO85(s)}$		$\overline{SDOM}_{FAO85(s)} + 1\sigma$
Average TMI by FAO85			$\overline{TMI}_{FAO85} = \frac{\sum TMI \cdot AREA.STU}{\sum AREA.STU}$		
	J-shaped	$\overline{TMI}_{FAO85} - 1\sigma$	$\overline{TMI}_{FAO85}$		$\overline{TMI}_{FAO85} + 1\sigma$
	sigmoidal	$\overline{TMI}_{FAO85} - 2\sigma$	$\overline{TMI}_{FAO85}$		$\overline{TMI}_{FAO85} + 2\sigma$
Moor, Heath, Peat		Distance to 322.CLC (Moors and heathland) + 412.CLC (Peat bogs)			
	sigmoidal	0	0	0	5000
Bare		Distance to 332.CLC (Bare rock)			
	sigmoidal	0	0	0	5000
Obstacles to Roots		ROO mapped to slope Linear transformation of slope with 10% trim: $S_{SMU}^{t10} = m_{SMU(P10)} \cdot S_{SMU} + c_{SMU(P10)}$			
	J-shaped	$S^{t10}(n-1)_{MEAN}$ for $n=1$: $S^{t10}1_{MEAN}$	$S^{t10}(n)_{MEAN}$		$S^{t10}(n+1)_{MEAN}$ for $n=4$: $S^{t10}4_{MEAN}$
	sigmoidal	$S^{t10}(n-1)_{MEAN}$ for $n=1$: $S^{t10}1_{MEAN}$	$S^{t10}(n)_{MEAN}$		$S^{t10}(n+1)_{MEAN}$ for $n=4$: $S^{t10}4_{MEAN}$
Water Regime		WR mapped to slope Linear transformation of slope with 10% trim: $S_{SMU}^{t10} = m_{SMU(P10)} \cdot S_{SMU} + c_{SMU(P10)}$			
	J-shaped	$S^{t10}(n-1)_{MEAN}$ for $n=1$: $S^{t10}1_{MEAN}$	$S^{t10}(n)_{MEAN}$		$S^{t10}(n+1)_{MEAN}$ for $n=4$: $S^{t10}4_{MEAN}$
	sigmoidal	$S^{t10}(n-1)_{MEAN}$ for $n=1$: $S^{t10}1_{MEAN}$	$S^{t10}(n)_{MEAN}$		$S^{t10}(n+1)_{MEAN}$ for $n=4$: $S^{t10}4_{MEAN}$

To improve the correlation between the STU attributes used as criterion factors and the values in the ancillary data the data in the STU database were transformed. The main transformations applied are:

- linear function, used to relate STU parameter spatial layer;
- area-weighted mean, used to relate FAO 85 soil types to spatial layer;
- distance to land cover feature in spatial layer.

The type of transformation applied to a factor is given in Table 16. For the STU height parameter two processing options were considered:

- retain the minimum and maximum height values as given in the STU database;
- linearly transform the minimum and maximum height.

The options may be used to compare the results when using the original data with the modified ranges. For the transformed data the values of the outer control points are set to the full range of data or the 5% or 10% trimmed range, depending on the membership function used.

For the STU slope parameters (dominant and secondary) similar options to height of retaining the original data and transformed values for the control points were used. The upper limit of slope class 3 is set to the mean of slope class 4 (34.06%). Transformed values for control points for slope are generally based on trimming the range of values by 10% at the lower and higher end. When using a J-shaped function different values are used for the central control points to define a plateau. For the sigmoidal function a common value is used for eh central control points.

For all FMFs for STU parameters the functions for the outermost range of values is monotonically decreasing (low range) or increasing (high range). For example, slope class 1 is always monotonically decreasing and class 4 always monotonically increasing. In case either class is not present in the data classes 2 and / or 3 may change the function form. If only a single class is present the function is symmetrical.

For criterion factors defined for FAO85 classes an area-weighted transformation was applied. For height the FAO85 mean was calculated from the SMU mean elevation, which was derived from a range of values trimmed by 10%. The mean values for the FAO85 slope factor were computed from the mean values of the dominant slope as determined for the STU slope factor.

The values for the control points for the criterion factors "*Obstacles to roots*" and "*Water Regime*" were based on the transformation approach applied for the STU slope factors. The values were calculated independently of the transformed STU slope values. In the definition of the parameters for the linear transfer function the values for the independent variable is set to the full nominal range of values for a 10% trim (0% – 55.61%). This decision was taken to decrease the dependence of the distribution of membership values on the number of different classes present in an SMU and better align the distribution of the membership values in an SMU to the general distribution of the factor.

For the two criterion factors linked to the land use layer (bare areas, moor land) the values of the control points are defined by the proximity to these features. A monotonically decreasing function is used with the external control point *d* set to 5,000m. This distance should be sufficient to allow for geometric differences between the SGDBE and the land use layer.

Another issue leading to diversity of the processing setup are the conditions under which a fuzzy set membership function should be defined for a criterion factor. One may distinguish the following situations:

a) No values are recorded for a criterion factor of an SMU:

   No membership function can be defined.

b) A value is set for all cases of a criterion factor in an SMU:

If all values of a criterion factor are identical also the membership values are identical for all STUs of an SMU. Consequently, the membership values do not contribute to MCE for the spatial positioning of STUs.

If the values differ for at least one of the STUs a membership function should be defined for all STUs of an SMU.

c) A value is set for some but not all cases of a criterion factor in an SMU:

In a situation where data are missing for some STUs, but not all, a membership functions should be defined for the STUs with data even if the values of STUs with data are identical.

For the processing carried out in this evaluation FMFs were defined for two situations: one defines a function whenever data are available for a criterion factor and one only when the values of the control points differ between the criterion factors of an SMU.

3.5.3 Evaluation of Standardized Criteria (MCE)

The standardized criterion factors produced by applying fuzzy sets provide a measure of affinity between individual STU attributes and the related decision layers. To express the affinity of an STU with a grid cell the affinities of the criterion factors pertaining to an STU are evaluated and aggregated into a single value. Since the factor affinity is expressed on a ratio scale an evaluation can be more refined than a simple AND or OR assessment of Boolean layers. In this study the standardized criterion factors were assigned weights and the merged using the method of a *weighted linear combination* (WLC).

- **Criterion Factor Weights**

For each criterion the weight is set according to the perceived relative significance of the factor for the objective. The weights are determined by creating a matrix of all combinations of criteria, separately for factors and constraints. For each factor pair the relative significance for allocating areas is then rated. The matrix of relative significance is analyzed using the Analytical Hierarchical Process (Saaty, 1977). The performance of this method may not necessarily minimize errors (Triantaphyllou, *et al*., 1990), but is implemented by the GIS used[22].

The weights used in the process depending for the emphasis given to a group of data are given in Table 17,

[22] Idrisi performs the analysis under the module "Weight".

Table 17: Criterion Weights of Weighted Linear Combination

Decision Criterion	**Factor Weight %**
Factor	
STU height (w_1)	28.9
STU dominant slope (w_2)	20.6
STU secondary slope (w_3)	5.1
FAO 85 STU height (w_4)	17.3
FAO 85 dominant slope (w_5)	9.7
FAO 85 TMI (w_6)	8.3
STU obstacle to roots (w_7)	5.8
STU water regime (w_8)	4.3
TOTAL	**100.0**
STU land use "Moor" (w_9)	30.0
STU land use "Bare land" (w_{10})	30.0
Constraint	
All areas defined without soil(w_{11})	-

Consistency Ratio: 0.05

An indicator of the consistency of the ratings is given in form of the consistency ratio, which represents the probability that the ratings were randomly assigned (Eastman, 2012; p.134).

The weights for STU height and the dominant slope account for almost 50% of all weights. They have been given prominence because, together with the land use factors, they are the most direct association between the values of an STU parameter and the spatial layer data. The factors "*Moor*" and "*Bare land*" are only present in relatively few STUs. They were not included in the matrix of factor weights to avoid affecting the rating. The weights for these factors was set higher than for any other factor because of their prominence in the evaluation of the STU affinity with the spatial decision layer of land use. When an STU contains the factors "*Moor*" or "*Bare land*" the sum of the factor weights is re-scaled to 1.0. This method is also applied when data are missing for one or more criterion factors of an STU.

Criteria constraints are of type Boolean and no weights are set for them. The various constraints for the allocation of STUs were combined to a single layer, which includes the land/sea mask.

- **Weighted Linear Combination**

The criterion factors scores are multiplied with their relative weights to derive the overall affinity *A* of an STU (object) which is calculated as:

$$A = \sum_{i=1}^{n} SCS_i \times w_i$$

with

SCS_i:	standardized criterion score of factor *i*
w_i:	criterion weight of factor *i*

In the WLC the factor weights determine the level of trade-off between factors. The method represents an average level of risk since the final affinity value is neither one where affinity is found in all factors (low risk), nor where it is highest for a single factor (high risk). For factors "*Moor*" and "Bare land" the level of trade-off with other factors could be set lower than for conditions when these factors are not present in an SMU. However, in order to modify the level of trade-off between factor scores a different approach to the MCE than WLC has to be used. The method of *ordered weighted average* (OWA) allows control over the level of trade-off, but was found to be outside the scope of this study.

3.6 Spatial Positioning of Soil Typological Units (MOLA)

The evaluation of the factor scores provides a single value for the affinity of an STU (object) with a geographic position in an SMU. More than one STU of an SMU is likely to claim the same areas in the SMU. Since only one STU should be assigned to any one grid position the STUs are in conflict for claims of grid cells. The method offered by the GIS package for solving conflicting claims is available in form of the multi-objective land allocation (MOLA) module as available in the Idrisi GIS (Eastman, *et al.*, 1993). The method is based on ranking the objectives and assigning an objective by iteratively balancing the weighted claims for areas.

- **Ranking**

The aggregated factor score for each objective is transferred to a byte binary layer of ranked scores. To improve solving ties of ranking a second layer of factor scores is used. The first layer order scores in descending order, while the second orders the scores in ascending order. The rule defined to automate the procedure was to use the scores of the factor following the factor assessed from Table 17. In the GIS package used the ranked scores are standardized by a histogram equalization. This methods is used in preference to standard scores, which rely on the values being normally distributed. Standardizing the STU scores is a pre-requisite for the allocation of STUs.

- **Allocation**

  The aim of the allocation procedure is the to maximize the overall level of affinity for an STU within the SMU within the limits of the claims for area. The iterative procedure is applies the following steps:

  - assign non-competing regions to objective;
  - divide competing areas between objectives according to distance to ideal point;
  - re-assess objective areas according the areal requirements.

  The division of areas with competing objectives is performed according to the distance of the weighted score to the ideal point for that objective.

  For each STU the claim for grid cells is derived from the relative area of the STU in the SMU. The area claims are set 1% below the nominal values to avoid impasses in the allocation of areas, but at least 1 grid cell. The procedure itself uses an areal tolerance to decide when the allocation of objects to geographic positions has been achieved. In the runs this value was set to 5 grid cells for a mapping unit. This was found to be a practical compromise between processing time and the strictness of allocating the STU area claim to geographic positions.

  With the competition of STUs for geographic positions the procedure can result in assigning an STU to a geographic position, for which it does not have the highest value of affinity. The compromises in allocating STUs allow optimizing the overall level of affinity of all STUs within an SMU.

- **Computational Limitations**

  The procedure could not be applied to SMU where the area of an STU was below the grid size of the raster layer used. In the vector layer the minimum size of an SMU was set to 9 ha. By comparison, a grid cell in the raster layer is 100ha. Affected from the limitation are 17 SMUs. The total are of these SMUs is 237 km^2 and ranges for individual SMUs from 6 to 91 km^2. To these SMUs the dominant STU was assigned.

Processing the attribute data and setting the criterion factors for the MCE/MOLA was largely performed using a RDBMS. The implementation of the MCE/MOLA procedure was performed using the GIS. For automating data processing the GIS provides a script language. However, the language lacks elements of structure and allows only a single level for subroutines. The conditions for processing the spatial data were therefore analyzed in a separate programming environment, which generated also the scripts to run the MCE/MOLA procedures under the GIS.

4 RESULTS

The product generated by the evaluation of the decision criteria and the allocation process is a spatial layer with STUs assigned to individual grid positions instead of the compound area of the SMUs. In the resulting layer the proportions of STUs within an SMU are retained within the limits of the tolerances set for the allocation process.

The STU layer can be generated from a large number of options set for processing the data. Depending on the processing options and parameters set for the MCE the positioning of the STUs within an SMU was found to be variable.

4.1 Single-Layer STU MAP

The new STU spatial layer directly links the attributes of the SGDBE_STU and SGDBE_PTR tables to geographic positions. The corresponding data model is presented in Figure 40.

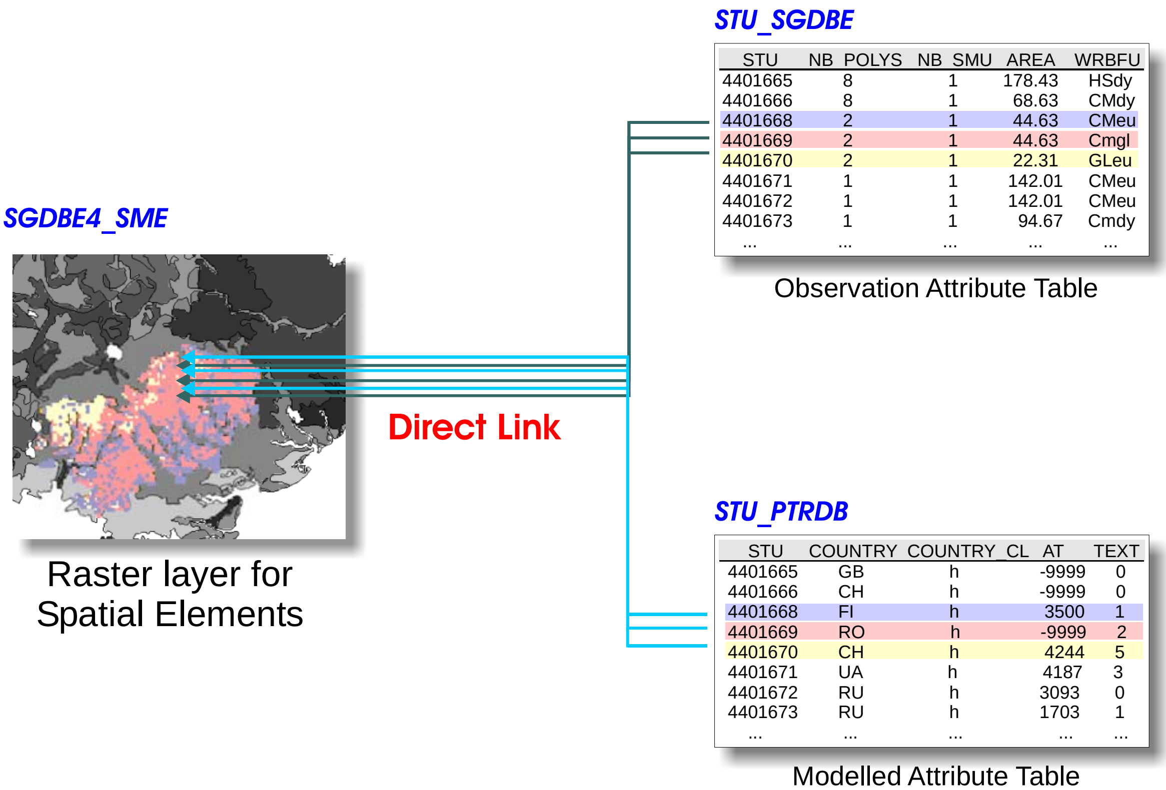
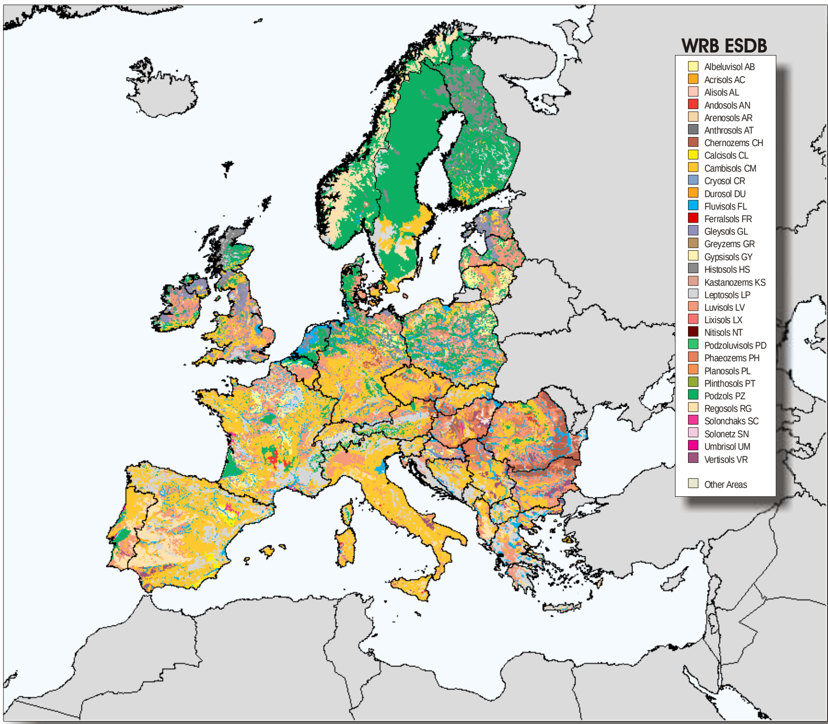


Figure 40: Data Links between Spatial Mapping Elements and Attribute Tables

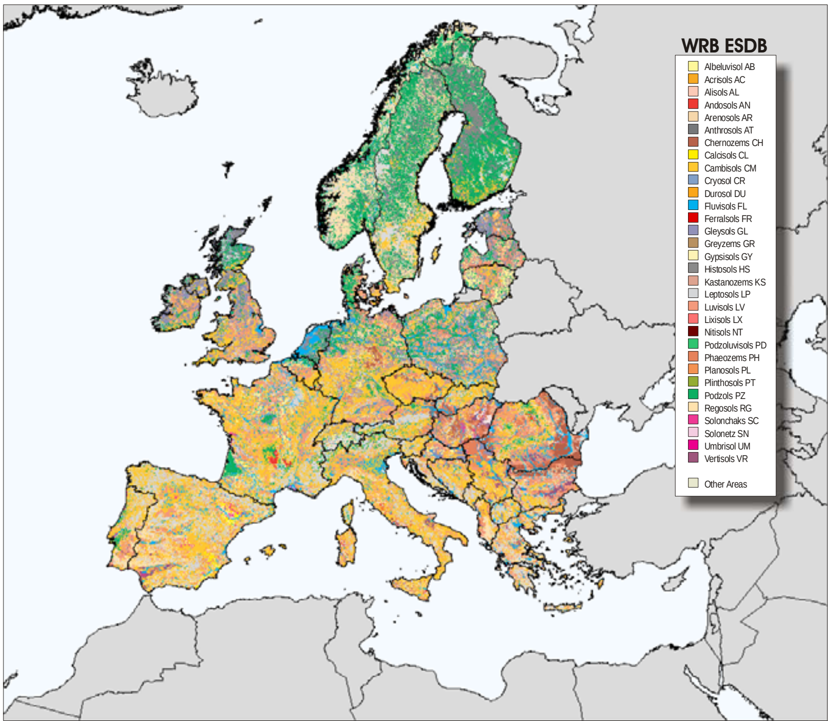
The data model eliminates the intermediate STU_ORG table by shifting the distribution of STUs to the spatial layer. Each grid cell of the spatial layer is thus linked to only one STU. This allows mapping the STU attributes directly to a single layer.

Mapping soil attributes to a single layer instead of 10 composite layers avoids the need to aggregate the values within an SMU to represent the full range of the soil typology for an area. Direct mapping to a single STU layer is of particularly relevance to mapping data stored on ordinal or nominal scales. While data available on ratio scales may be aggregated using a weighted mean data on nominal scales are not suitable to be aggregated using an arithmetic method.

An example of mapping a soil attribute recorded on the nominal scale is the soil type of the WRB. The WRB Level 1 soil types were mapped using the dominant STU of an SMU and the STU spatial layer. The resulting layers are shown in Figure 41.



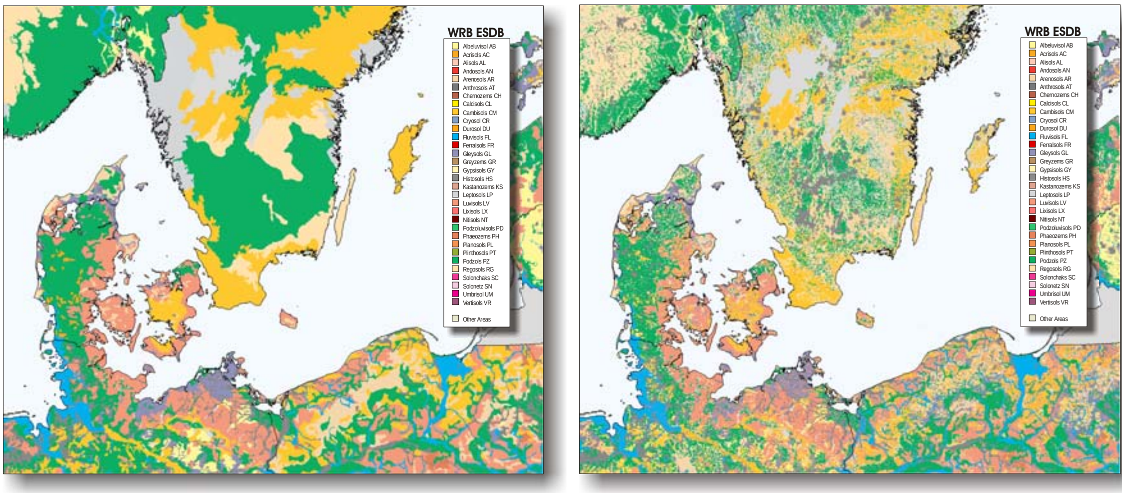
a) dominant STU to SMU



b) all STUs to STU map

Figure 41: Mapping WRB Level 1 Soil Type for Dominant STU in an SMU and for STU Spatial Layer

The map extract on the top of Figure 41 shows the WRB Level 1 soil type for the dominant STU of an SMU. SMUs in Sweden and Norway are comparatively large in size and linked to an above-average No. of STUs. This contrasts with the spatially more detailed definition of SMUs in Denmark and Germany, to which a lower No. of typological units is linked. As a consequence, the change in detail when mapping the soil classes to the STU map is most noticeable in the larger SMUs. This change in detail depending on the SMU size is depicted in the comparison of dominant STU and spatial STU mapping of WRB codes for Denmark and Sweden, as shown in Figure 42.



a) dominant STU to SMU *b) all STUs to STU map*

Figure 42: Mapping WRB Level 1 Soil Type for Dominant STU in an SMU and for STU Spatial Layer for Denmark and Southern Sweden

The changes in the distribution of WRB codes from mapping the dominant STU to the spatial STU layer are noticeable in Denmark, northern Germany and Poland, but not to the degree of the changes in Southern Sweden and Norway. The more extensive changes in the allocation of soil types are due the number of STUs in an SMU and the size of the SMUs. In Denmark the average size of an STU (SMU) is 437 ha (2,076 ha), in Poland 1,075 ha (5,468 ha), in Sweden 1,218 ha (6,978 ha) and in Germany 1,233 ha (4,649 ha). Thus, for Sweden and Poland more STUs are allocated within a larger area than in Denmark or Germany, which results in the more extensive changes between the layer of the dominant and the allocated STUs (see also Figure 2 and Figure 3).

One particular item of interest may be the location of organic soils (*Histosols*) in the spatial STU map, which are largely absent when mapping only the dominant STU. The area of organic soils is generally sub-dominant in the SMUs in Sweden. Therefore, deriving statistics from the dominant STU are biased against organic soils. Yet, the information is available in the STU attribute table.

The dominant STUs for the SMUs cover 77 WRB soil types, while there are 98 WRB soil types in the mapped STUs. Increasing the number of mapped WRB soil types by 27% seems notable, but it may have limited effect in practical terms, because of the area concerned. A list of the WRB soil types concerned and their area is given in Table 18.

Table 18: WRB Soil Types included in STU Map in Addition to Mapping Dominant STU

WRB *Code*	**Area** *km²*	**WRB** *Code*	**Area** *km²*	**WRB** *Code*	**Area** *km²*
ACfr	178	CMha	140	KSlv	143
ACpl	38	FLha	75	PLha	210
ALpl	78	FLti	289	RGge	2,914
ANmo	1,231	GLhi	4,063	RGha	318
ANvi	941	HS	146	SCty	5
ATpa	3,473	HSge	184	VRgl	35
CHgl	962	KSha	44	VRha	1,559

The areas covered by the additional STUs ranges from 5 km^2 (*PLha*) to 4,063 km^2 (*GLhi*). In total, 17,023 km^2 are concerned, which corresponds to 0.3% of the total AOI.

4.2 Assessment of Results

Assessing the results of the allocation of STUs very much depends on the availability of references for comparison. A distinction can be made between the ESDB data, the model approach and the implementation. Errors in one area propagate through to the final result. A standard method of assessing modelled outputs is to compare the results to external references. Such references could be national soil maps. However, simply contrasting the STU map with national soil maps to validate the results of the model used cannot separate the performance of the allocation process from deviations in the soil data between the ESDB and national datasets.

A comparison of the ESDB data with external sources is outside the scope of this study. The delineation of SMUs is used as found in the SGDBE as area the STUs assigned to an SMU. Neither is the quality of the ancillary data (DEM, CLC) evaluated. As a consequence, comparing the final STU map to an independent reference map of soil properties does not provide an insight into the procedures applied and the parameters set for processing the data.

The properties of the attributes of the STUs were evaluated in the preparatory phase of the study. While the actual values of the data are not assessed the effectiveness of providing information to the process of disaggregating STUs can be appraised. The mere presence of attribute data does not imply that the values given are of practical use. The degree to which the position of an STU within an SMU is determined by the criterion factors depends on the number of distinct memberships between the STUs. To fully separate *n* STUs the same number of independent functions is needed. This degree of separation could be achieved if more STU attributes could be linked to ancillary spatial data. However, there is very little spatial data available on attributes such as texture, soil depths or water regime. Ground data from point observations are potentially valuable in an assessment of the ESDB data, but a point-based comparison of the STU map with ground data is an ineffectual approach to evaluating the disaggregation method.

While it is not obvious how the uncertainty in the STU map can be assessed, the assessment of the STU map has to rely on an evaluation of the robustness of the procedures and parameters used. This can be achieved by using a sensitivity analysis. Dedicated software to support performing a sensitivity analysis can be used, such as the JRC SimLab software[23].

A list of the variables set during processing and the range of alternative values is presented in Table 19.

[23] SimLab and information on uncertainty and sensitivity analysis is presented at: http://ipsc.jrc.ec.europa.eu/?id=752

Table 19: Processing Variables and Settings

Variables	Option
DEM data processing	• not filtered • filtered
Method for relating STU to DEM height	• direct • transformed
Method for relating STU to DEM slope	• by class range • by ranked value
Method for aggregating minimum and maximum height	• absolute value • average value
Method for aggregating mean height	• weighted mean • min-max mean
Method for aggregating minimum and maximum slope (common to dominant and secondary)	• absolute value • weighted mean
Method for aggregating mean slope (common to dominant and secondary)	• weighted mean • min-max mean
Method for deciding on fluvial/non-fluvial STU	• 1 σ • 2 σ
Fuzzy set membership function type	• J-shaped • sigmoidal
Factor diversity within SMU	• all data • only diverse data
Distance for external control points	• 1 σ • 2 σ
Criterion factor weights	• by correlation • STU emphasis • FAO emphasis
Ranking process	• secondary sort list
MOLA objective weight	• 0 < value < 1
Tolerance for objective area requirement	• fixed limit • function by area
Tolerance for allocation area	• fixed limit • function by area

Not considered in the list were the options for the methodological aspects, such as the MCE method, the method for finding criterion factor weights or the weights for the MOLA procedure.

The assessment of the sensitivity of the method is kept simple by changing *one factor at a time* (OAT) and by looking at the output in its entirety. This unsophisticated approach was selected because MCE and MOLA are run on individual SMUs with settings specific to each SMU. Some of the evaluation and allocation parameters are set globally, such as the criterion factor weights, while others change with the SMU-STU combination, such as the external control points for the fuzzy set membership function. It would require considerable effort to vary more parameters, given the time it takes for a single run to finish.

Because of the processing overhead the number of variables (dimensionality) was kept to the main factors which are considered of influence on the result. Varied were the criterion factor weights, the type of fuzzy membership function and the distance of the external control points from the mean for aggregated factors. A summary of the parameters varied is given in Table 20.

Given the nature of the weights (sum = 1.0) a single weight cannot be changed unless one or more other weights are also altered. The weights were therefore varied according to settings by group. For all sets of weights the AHP method was employed to arrive at the factor weights.

The type of FMF was altered between J-shaped and sigmoidal. Except for the TMI factor (monotonically increasing or decreasing) all functions were to some degree symmetric. The external control points of the sigmoidal function set to a distance of either 1 or 2 σ from the mean.

Other variables set in the procedure, such as the weights or minimum area of the MOLA procedure, were not modified. These settings were considered to be of less influence on the mapping outcome than the three variables varied in the assessment. This reduces the dimensionality of the runs. For the OAT method to be practicable such a reduction in dimensionality is necessary since each run of the procedure takes one day, not including the time needed to prepare the data. The number of runs increases exponentially with the parameters varied and, without an uncertainty analysis, becomes impractical.

Table 20: Variation in Fragmentation Index as a Consequence of Settings for Key Processing Parameters

Criterion Factor Weights		Fuzz Membership Function: J-shaped	Fuzz Membership Function: sigmoidal	External Control Points
	correlation of factor with ancillary spatial data		53,989 0.221	*1 σ*
		52,625 0.248	53,553 0.222	*2 σ*
	giving weight to STU data		53,721 0.221	*1 σ*
		52,770 0.242	53,352 0.220	*2 σ*
	giving weight to aggregated data		54,226 0.219	*1 σ*
		52,594 0.248	53,821 0.219	*2 σ*

n No. of grid cells not allocated
m Fragmentation Index

In the absence of suitable data to assess the allocation output a comparative evaluation was performed, based on the number of grid cells not assigned and the fragmentation index[24] of the SMU-STUs combinations in the layer. The J-shaped fuzzy set membership functions consistently resulted in a lower number of non-allocated grid cells and a higher value for the fragmentation index. Changes to the criterion weights did not result in a discernible difference in the number of non-allocated grid cells or the fragmentation index. For the settings of the sigmoidal functions using 2 σ instead of 1 σ for

[24] Fragmentation Index = (n-1)/(c-1)
where n = number of different classes present in the kernel
c = number of cells in kernel (after: Eastman, 2012)

the external control points resulted in a lower number of non-allocated grid cells, but not to the extent of changing to a J-shaped MF for the criterion factors.

A cross-tabulation of the maps against the using the WRB classes was found to yield very little diversity between the treatments with respect to the Kappa coefficient of agreement. Thus, improving the understanding of the method of allocating STUs by assessing how changes in key parameters affect the number of non-allocated grid cells and the fragmentation index did not shed any light on the factors driving the allocation process. This asks for a more detailed investigation.

- **Criterion Factor Weight**

  The criterion factor weights under the three options of emphasising a group of factors are summarized in Table 21.

Table 21: Criterion Factor Weights by Factor Group

Criterion Factor	**Emphasis**		
	Correlation with Ancillary Data	**STU Attribute**	**FAO Aggregation**
	Weight	*Weight*	*Weight*
Height	0.3060	0.3198	0.1735
Dominant Slope	0.0985	0.2447	0.1413
Secondary Slope	0.0448	0.0816	0.0517
FAO 85 Height	0.2516	0.1721	0.3247
FAO 85 Slope	0.1236	0.0619	0.1008
FAO 85 TMI	0.1756	0.1198	0.2080
Moorland	0.3000	0.3000	0.3000
Bare Areas	0.3000	0.3000	0.3000
Consistency ratio	*0.07*	*0.04*	*0.04*

The relative importance of the pair-wise comparison matrix when emphasizing the STU attribute data sets a relative weight of 1/3 to the aggregated criterion factors in relation to the STU attributes. When emphasis is put on the aggregated factors the relationship is inversed.

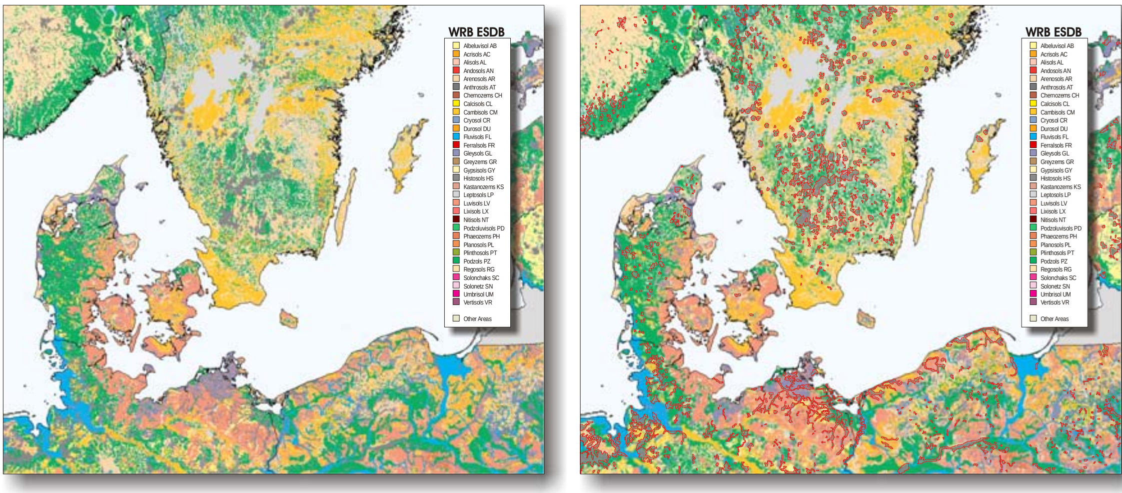
The assessment of the method sensitivity to parameter variations was performed without including the obstacle-to-roots and the water-regime factors. In any of the settings the combined weight was below 10% and not considered to lead to noteworthy changes.

- **Fuzzy Membership Function**

  The FMF type was only varied for the aggregated criterion factors. For the STU factors all inflection points can be determined and the external control points can be set to values which cover the complete range in the ancillary data for the SMU. For the aggregated factors the inflections points are set globally and therefore only approximate the range of values in the ancillary data in the SMU area.

  All FMF leave some areas without STU allocation. This peculiarity could be removed by setting tighter limits to the area requirements in the MOLA and forcing all areas to be assigned to an STU. However, the processing time increases drastically when setting lower thresholds for the area requirements and, in rare cases when the SMU area is small, the process can get locked without achieving the target areas.

  A visual comparison of the influence of the settings of FMFs to run the MCE, either J-shaped or sigmoidal, is presented in Figure 43.



a) J-shaped function *b) sigmoidal function (with peat)*

Figure 43: Mapping WRB Level 1 Soil Type from J-shaped and Sigmoidal Fuzzy Membership Functions for FAO-aggregated factors (Denmark, Southern Sweden), with Moorland/Peat Highlighted

In general, setting the membership of the FAO-aggregated factors to a sigmoidal MF gives more compact areas for the STUs than using a J-shaped function. The average fragmentation index for the WRB codes of the dominant STU soil type is 0.047, it is 0.238 (0.248) for the J-shaped function and 0.213 (0.222) for the sigmoidal function of FAO-aggregated factors for a 3x3 kernel (STU codes).

One reason for the higher level of fragmentation when using a J-shaped function may be the steep decline in membership around the central inflection points of the J-shaped function. Small differences in factor values may thus lead to relatively larger differences in membership. The sigmoidal function defines a more gradual change in membership around the central inflection points. This characteristics may better represent the distribution of factor values around the value of the central control points. However, the membership value of any value outside the external control points of the sigmoidal function is set to zero. The effect of the factor therefore depends on the definition of the values for the external control points. When changing the external control points to 1 σ of the mean the overall fragmentation index for WRB soil types decreases to 0.148 (0.221 for STUs).

On the map processed using the sigmoidal membership function areas identified as moorland or peat are highlighted b the red outline. The distribution of these areas contrasts strongly with the map of the dominant WRB soil type presented is Figure 41 a). The geographic allocation of moorland is guided by the peat land cover type, while respecting the areas of peat of the STU database.

- **External Control Point Settings**

Alternative setting of the external control points were only considered for factors from aggregated data and where sigmoidal MFs were used. The external control points were defined with a distance to the mean of either 1 or 2 standard deviations (σ). Setting a narrow limit for the external control points ($\bar{x} \pm 1\sigma$) give sharper boundaries for the affinity of a factor in the ancillary data, but in case of a sigmoidal function also caries a greater risk of only partially covering the range of values in the ancillary data. A wider limit for the external control points ($\bar{x} \pm 2\sigma$) reduces the risk of not covering the ancillary data, but the membership of the factors may become indistinct.

For the settings used to in the analysis no clear trend could be identified. There would appear to be a tendency for lower non-allocated areas when using the 2 σ settings, although not as pronounced as the difference between using J-shaped and sigmoidal functions.

- **Distinctness in Local Membership Values**

The distinctness of membership values for the criterion factors was evaluated by relating the maximum membership value provided by the MCE to the sum of all membership values, using the following equation:

$$c = 1 - \frac{MS_{\max} - \frac{\sum_{i-1}^{n} MS_i}{n}}{1 - \frac{1}{n}}$$

where

MS	membership value after MCE
MS_{max}	maximum membership value
n	No. of objects

(adapted from: Eastman, 2012)

In the equation the number of objects corresponds to the number of STUs which are linked to the SMU at a location and which compete for allocation to a grid cell.

In Eastman (2012) the value c is termed "*classification uncertainty*". However, the value c is a relative measure of the lack of commitment of the object with the highest membership value compared to all membership values of all objects. In this sense the relative lack of commitment is related to indicators of dominance rather than uncertainty.

For a configuration of the MCE using a J-shaped MF (non-filtered DEM, only diverse SMU data) the average lack of commitment is 81%. A similar level of pre-eminence, or lack thereof, for a criterion is obtained when configuring the MCE for sigmoidal MF. However, the spatial distribution of the lack of commitment varies considerably between the two configurations.

A graphical presentation of the variation in the lack of commitment between the sigmoidal and the J-shaped configuration is given in Figure 44.

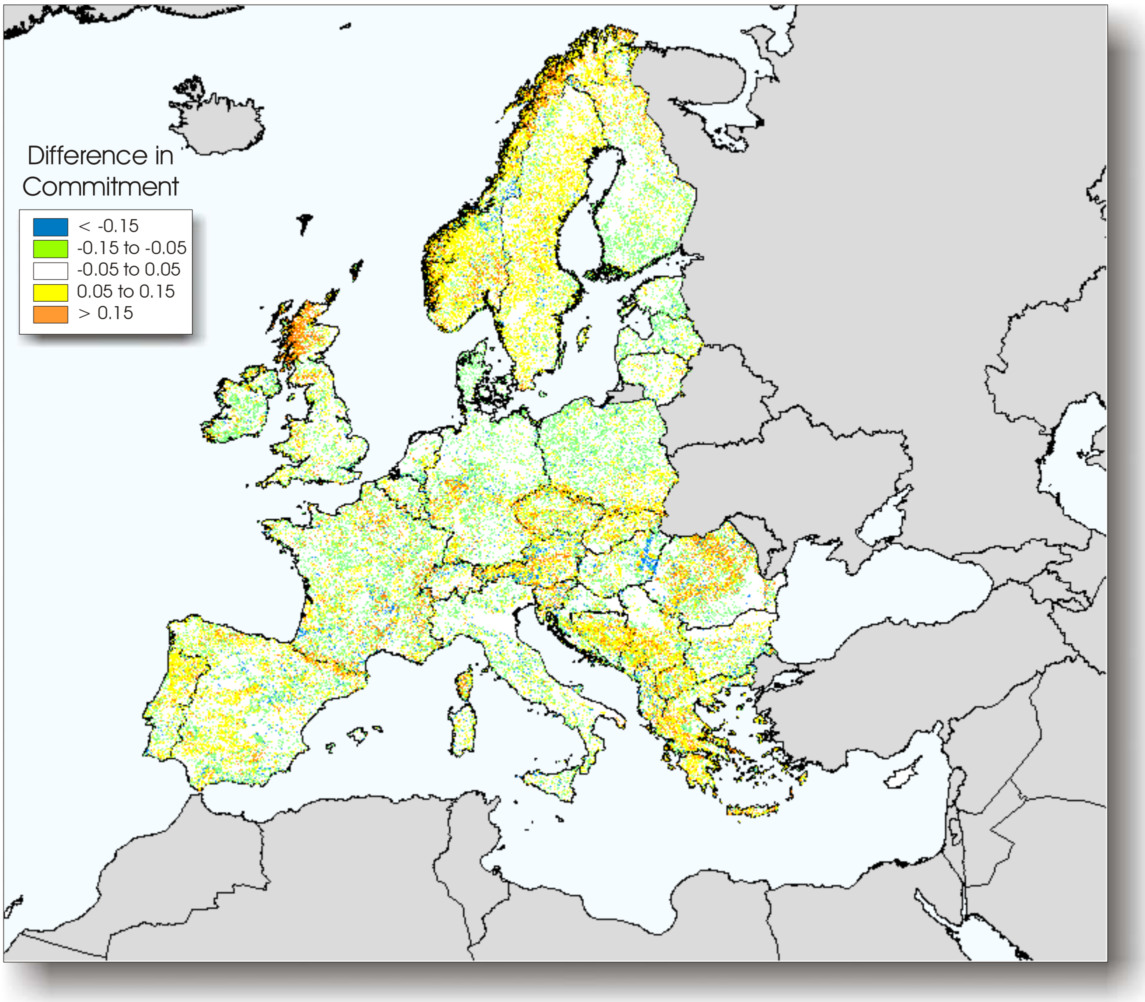


Figure 44: Difference in Lack of Commitment of Object to Allocation between a Sigmoidal and a J-shaped Configuration for MCE/MOLA Processing

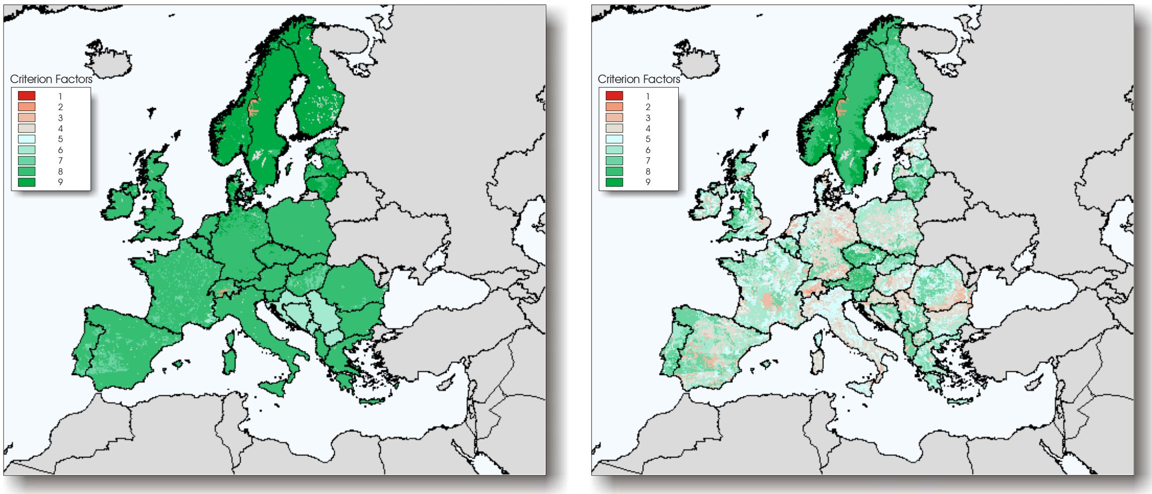
A higher level of commitment to an objective for the settings of J-shaped MF is found in particular in Scotland and to a lesser degree in Norway, Sweden, Austria and most Eastern European Countries. The sigmoidal MF settings give higher values mainly in undulating areas, such as Ireland, Finland Germany and Poland. In most flat areas the differences in commitment between the settings are minor. Reducing the number of criterion factors in the MCE/MOLA reduces the lack of commitment mainly in sloping areas and increases it in flat areas. Thus, more criterion factors do not necessarily lead to a better distinction of objects.

The lack of distinguishing between objectives in the flat areas is at least in part a consequence of using morphological differences as the basis for defining the criterion factors and such differences are by nature slight in flat areas. Another condition leading to the low level of commitment is the diversity of the data available for the MCE/MOLA procedure.

- **Treatment of Data Diversity**

  For the criterion factors used the attribute database contains data for most STUs (see Table 2). Yet, STUs of an SMU are distinct from other STUs of the same SMU only by varying the value of some attributes. In consequence, the same membership value applies to the SMU as a whole. Accordingly, there is no variation in membership values to support the spatial allocation process. Including these cases of spatially non-specific membership values in the processing results in higher values from the MCE, but lower scores in the commitment of an STU to be allocated.

  The consequence of the two options for treating diversity in the definition of criterion factors on the number of criterion factors per SMU available for the MCE/MOLA procedure is graphically presented in Figure 45.



a) all values *b) diverse values*

Figure 45: Maximum Number of Criterion Factors for STUs within SMU

The graph shows a distinct decrease in the number of criterion factors for the STUs of an SMU in most parts of the AOI. Only for SMUs in Norway and Sweden, and to a lesser degree for Austria, are the conditions for a factor within an SMU defined with diverse values. As a consequence, the number of decisive FMFs for a criterion factor to allocate STUs is considerably less than the number of data entries in the attribute table.

The level of identical entries in a field varies between attributes. An overview of the data diversity for the criterion factors is presented in Table 22.

Table 22: SMU Attribute Data and Diversity of Values

Factor	**Available Data**		**Diverse Data**
	Attribute – STU – SMU Combinations	**Attribute – SMU Combinations**	**Unique Attribute – SMU Combinations**
STU Attribute	*No.*	*No.*	*No.*
FAO 85	5,522	5,472	5,369
ZMIN + ZMAX*	5,370	2,304	1,210
SDOM	5,442	2,408	1,574
SSEC**	5,449	2,593	1,878
ROO	5,208	2,376	1,554
WR	5,244	2,518	1,808

* ZMIN > -400 and ZMAX > -400
** Modified according to SDOM

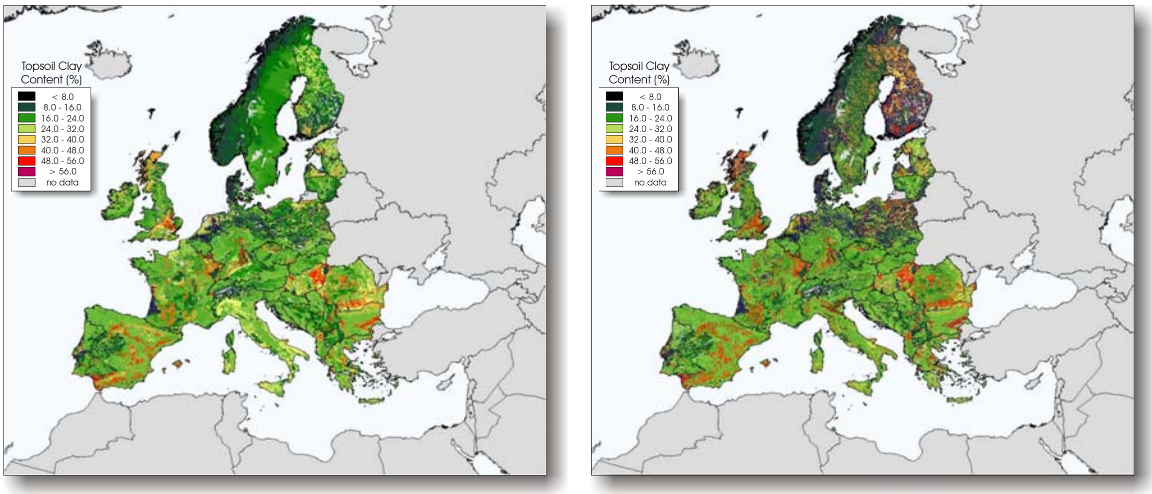
The loss of diverse data available to allocate STUs within an SMU is most prominent for the height attribute. Of the 5,370 unique combinations of attribute value + STU + SMU remain 2,304 combinations of attribute value + SMU, out of a total of 1,543 SMUs in the AOI. The number of cases with diverse combinations of attribute value and SMU affected is 1,210. The number of SMUs with more than one diverse value of the height attribute is 399 or 25.9% of the total number of SMUs. This lack of diversity in the attribute data very much restricts the allocation process. The effect is apparent in the relative lack of commitment after the MCE/MOLA of an STU to a spatial position.

4.3 MCE/MOLA STU Layer

The variations of the numerous factors influencing the spatial allocation of STUs tended to show a higher level of commitment for the J-shaped FMF as compared to the sigmoidal FMF in many sloping regions and a reversal for flat areas. This complementary behaviour was exploited to generate the final STU layer from a combination of the results from a J-shaped and a sigmoidal FMF. The results were combined by first computing the average level of commitment per SMU. The FMF with the highest average commitment was then used provide the STU allocation.

The average level of lack of commitment *c* in the merged layer is 0.804. (0.814 when using J-shaped and 0.822 when using sigmoidal FMF). The overall improvement in the level of commitment for an object by merging the results is rather modest, but could not be achieved by a single setting for the process. Reducing the number of criterion factors to 5 (height, dominant slope, secondary slope, FAO height and FAO slope) increases the commitment (0.776 for J-shaped FMF), but decreases the range of membership values for objects.

Mapping parameters for individual STU values instead of aggregated values by SMU changes the distribution of the parameter within the spatial layer. As an example the clay content in the topsoil is mapped by SMU and by STU. The clay content is derived from the Harmonized World Soil Database (HWSD; FAO/IIASA/ISRIC/ISSCAS/JRC, 2012). For Europe the HWSD is based on the ESDB and the spatial layer is largely identical. However, texture in the ESDB is provided as classes, while the HWSD translated the classes into interval data of texture content (in %). Using the interval data allows computing the area-weighted texture class content for an SMU and for individual STUs. The resulting layers are presented in Figure 46.



a) SMU Aggregation *b) STU Mapping*

Figure 46: Comparison of Topsoil Clay Content Aggregated to SMU and by Spatially Allocated STU

Areas with low clay content (< 8%) are 8.3% in the layer of aggregated values compared to 16.9% in the STU layer. Areas of clay content between 24 – 32% are 4.3% in the SMU layer and 1.6% in the STU layer. Areas with higher clay content are generally more widely found in the STU layer, as a consequence of the spatial dispersion of the STUs. These differences in the distribution of texture class content, which are more pronounced at local scale, may affect the results obtained from models using such data, e.g. when soil water capacity or movements are modelled.

5 SUMMARY

The ESDB contains a wide range of data defining typical combinations of soil and land characteristics. This table of soil typologies is complemented by a geographic component in which areas of distinct collections of soil typologies are delimited. The soil typologies are therefore only attributable to an area as part of the collection, but not to specific geographic positions. In addition to pedological parameters the soil typologies are also defined by parameters related to topography and land cover. For these parameters corresponding spatial data exists. The evaluation of the ESDB data evaluated to what degree such ancillary spatial data could be used to improve the geographic position of individual soil typologies.

5.1 Soil Typological Parameters for Criterion Factors

The initial task was to assess the attributes of the soil typologies for their suitability of defining explicit and definitive links to the spatial database. The most complete data is available for the soil type. However, with the extension to Eastern European countries and the former Russian Federation, the soil types are available as either WRB or FAO 85 classifications, in some cases for both. For this evaluation it was decided to use the FAO 85 data as main parameter, which was found to offer the most complete set of data. The area of interest is thus limited to Europe. For the few cases where an FAO 85 entry was not available a suitable value was introduced into the database.

The soil typological attributes, which appear to be readily available to be linked to spatial layers, are height and slope.

- **Height**

  The soil typology database provides the height range for a typological units as ratio values. These values could be related to elevation data of a DEM. To improve the coverage of typological height with the DEM a linear transfer function was defined for each mapping unit.

- **Slope**

  In contrast to height, the slope attribute is recorded as an ordinal type with data ranges according to a classification scheme. A comparison of the distribution of the classes in the typological database and the spatial layer showed a strong tendency for lower values in the slope derived from the DEM compared to the typological data. Due to the data type of the slope attribute using a function, analogous to height, was not appropriate. To relate the typological slope attribute to the slope derived from the DEM the order of classes was therefore transferred according to the proportions of typological units in the mapping unit. The lack of a direct association of the slope classes in the typological data and the slope derived from the DEM may be attributed to some degree to the processing applied to reduce local variability in elevation rather than spatial resolution. The relationship between the typological data and the DEM for the slope parameter was found to depend strongly on the processing applied to the DEM data to derive the slope characteristics.

- **Depth**

  Investigated was also the relationship of attribute data on soil depth with topography. That no better links between depth and slope were found is attributed to the data type used for slope in the typological database (classes) and the weak association with the DEM-derived slope.

- **Land Use and Cover**

  Land use or cover could have a strong link with soil properties. Yet, because the procedure establishes a link of cause-effect between the typological attributes and the spatial data and due to the relatively rapid changes of land use and cover only those land cover types were included as criterion factors for which a corresponding typological entry as a soil type exists. These were moor lands / peat and bare areas.

- **Topographic Moisture Index**

  Comparatively strong were the associations between soil types and distance to the river network. A tendency to have less variation in the association than the simple distance to the river network was the topographic moisture index. The TMI was used as the only criterion factor in the evaluation for which no equivalent attribute exists in the typological database.

- **Other Topographic Features**

  It was further investigated whether soil types could be associated with other topographic features which were not recoded as typological attributes. However, the associations found between soil type and topographic features, including aspect, were indistinct. This should not be interpreted that such links would not exist, only that none could be substantiated in the evaluation.

There remain some open issues related to the completeness of the information in the soil database. One issue is the information content of the typological attributes. In the data for height and slope it would appear that in cases the range of values founding the spatial layers is not reflected in the typological units. Furthermore, also where data are entered there is a lack of distinctness between typological units. It is not evident whether this lack is the consequence of the generally mixed occurrence of the typological units or generalized reporting.

5.2 Multi-Criterion Evaluation and Object Allocation

As criterion factors two types were defined:

- factors relating soil typological attributes directly to a spatial layer (minimum and maximum height, dominant and secondary slope, moor land, bare areas) and
- factors aggregated by FAO 85 soil type (mean height, mean dominant slope, TMI).

For the attribute-related factors the minimum and maximum values within a mapping area can be accurately identified. This is not the case for the aggregated factors, which are defined at the level of the soil types and are derived from more than one mapping unit. Hence, a sigmoidal FMF can be defined for those attribute factors that cover the full range of values of the mapping unit. However, for factors from aggregated attributes full coverage of the range of values in a mapping unit is not warranted. This situation may be covered using a J-shaped FMF. This function type was found to define sharper boundaries of membership than the sigmoidal function and leads to a higher level of commitment of the typological units to a geographic position.

The study could demonstrate that the principals of multi-criteria evaluation and object allocation can be applied to the ESDB to support the geographic positioning of typological units. It also found that substantial and elaborate preparation were needed to prepare the data to serve as criterion factors. These preparations took much longer than actually performing the MCE and the MOLA. The processing itself was made more convoluted because each mapping unit has its own set of parameters was performed individually.

An unresolved issue remains the assessment of the uncertainty of the typological unit map. This is due to the lack of information of the uncertainty of the ESDB typological data and mapping units. Data from ground surveys or national maps are of little use when assessing the map, since the uncertainty in the input data cannot be separated from the uncertainty of the method. A sensitivity analysis is restricted by the time needed to process the data. Varied were only those parameters which were considered of main influence on the result (criterion weight, fuzzy membership function, setting of external function control points). As expected, a sigmoidal membership function produces a spatially less fractured map of typological unit than the J-shaped function, which also applies to setting narrower distances for the external control points. Visually, the criterion weights influence the distribution of the typological units, but not in any specific direction. The fragmentation index varied less with the criterion weights than the fuzzy set membership function.

In its present state the spatial layer of STUs provides a simplified solution to using all soil characteristics rather than only the dominant soil typological unit, although aggregating the results to mapping units is recommended. One viable option to follow is the mapping of peat areas, where a combination of soil and land use, and possibly climate data as additional factors, may allow a improved delineation of peat areas.

5.3 Recommendations

The results from the evaluation of using a spatial decision support approach to mapping soil typological properties of the ESDB has identified several specific areas, which could be investigated in view of an improvement in the spatial allocation of typological units.

- **Extend Links between Soil Typological and Ancillary Spatial Data**

  To improve the results of the mapping procedure for typological units it would be useful to develop more links between soil typological attributes and thematic data available in spatial form. Possible characteristics would be soil depth and parent material. The source of such data could be spatially interpolated ground surveys.

- **Condense Range for Set Membership Control Points**

  Reducing the span of values covered between the control points of the fuzzy set membership function increases the dominance of the membership value of one criterion factor over the sum of the membership values of all criterion factors. The range covered by the control points could be tightened by increasing the level of correlation between the typological data with the ancillary spatial layers.

- **Alternative Information Processing**

  With the present definition of the criterion factors very different approaches to mapping typological units could be applied. The definition of the control points of the fuzzy set membership can be used to develop signatures for procedures generally applied to classifying images, such as probability-based classifiers or neural networks.

- **Uncertainty Assessment**

  Alternative processing procedures may be used to demonstrate the technical feasibility, but they do not provide answers to the main obstacle of applying any of these methods, which is an assessment of uncertainty of the resulting map of soil typological units.

The ESDB contains a wealth of information, which is somewhat buried under the sheer amount of data. Thus, apart from these specific areas of augmenting the spatial allocation of the soil typological units, the evaluation of the soil data suggests that investigating methods for separating spurious and indistinct data from more particular and well-defined information can be recommended.

Acknowledgements

Many thanks for the invaluable background on the development of the ESDB and continuous support given to the project to Robert Jones, Cranfield University. The author is further very grateful for the editorial support provided by Tore Tollefsen, European Commission Joint Research Centre, for trying to make the writing more accessible where the phrasing was particularly convoluted.

References

Annoni, A., C. Luzet, E. Gubler, and J. Ihde (2003) Map Projections for Europe, EUR 20120 EN, Luxembourg, 131pp. http://www.ec-gis.org/sdi/publist/pdfs/annoni-etal2003eur.pdf

Beven, K.J. and M.J. Kirkby (1979) A physically-based variable contributing area model of basin hydrology. Hydrological Sciences Bulletin 24:43–69. http://iahs.info/hsj/240/hysj_24_01_0043.pdf

BURROUGH, P.A., R. A. MACMILLAN and W. van DEURSEN (1992) Fuzzy classification methods for determining land suitability from soil profile observations and topography. Journal of Soil Science, Vol. 43: 2. pp. 193–210.

Burrough, P.A. and R.A. McDonnell (1998) Principles of Geographical Information Systems, 2nd Edition. Oxford: Oxford University Press. 352pp. ISBN: 978-0-19-823365-7.

Chang, L. and P.A. Burrough (1987) Fuzzy reasoning: a new quantitative aid for land evaluation. Soil Surv. Land Eval. (7). pp. 69-80.

CEC (1985) Soil map of the European Communities at 1:1,000,000. CEC-DGVI. Brussels, Belgium. 124 pp.

Daroussin J. and D. King (1996) A Pedotransfer Rules Database to interpret the Soil Geographical Data Base of Eurasia for environmental purposes. In: The use of pedotransfer in soil hydrology research in Eurasia, workshop proceedings, Orléans, France, 10-12. October 1996. pp. 25-40. http://eusoils.jrc.ec.europa.eu/esdb_archive/ESDBv2/popup/pt_meta.htm

Davidson, D.A., S.P. Theocharopoulos and R.J. Bloksma (1994) A land evaluation project in Greece using GIS and based on Boolean and fuzzy set methodologies. International Journal of Geographical Information Systems, Vol.8: 4. pp. 369-384. doi: 10.1080/02693799408902007

Eastman, J.R. (2012) Idrisi Selva Manual – Manual Verion 17. Clarks Labs, Clark University, 950 Main Street, Worcester, MA, 01610-1477 USA. 322pp.

Eastman, J.R. (2006) Idrisi Andes: Guide to GIS and Image Processing. Clarks Labs, Clark University, 950 Main Street, Worcester, MA, 01610-1477 USA. 327pp.

Eastman, J.R., P.A.K. Keym, J. Toledano and W. Jin (1993) GIS and Decision Making – Volume IV. United Nations Institute for Training and Research. UNITAR European Office, Paliais des Nations, 1211 Geneva, Switzerland. 112pp.

European Commission Joint Research Centre (2003) European Soil Database – Distribution Version 2.0. European Commission Joint Research Centre, Ispra, Italy. http://eusoils.jrc.ec.europa.eu/ESDB_Archive/ESDBv2/index.htm

Fassio, A., C. Giupponi, R. Hiederer and C. Simota (2005) A Decision Support Tool for Simulating the Effects of Alternative Policies Affecting Water Resources: an Application at the European Scale. Journal of Hydrology 304, pp. 462-476.

Finke, P., R. Hartwich, R. Dudal, J. Ibàñez, M. Jamagne, D. King, L. Montanarella and N. Yassoglou (2001) Georeferenced Soil Database for Europe – Manual of Procedures, Version 1.1. EUR 18092 European Commission Joint Research Centre, Institute for Environment and Sustainability, Italy. 184pp.

Giupponi C., J. Mysiak, A. Fassio and V. Cogan (2004) MULINO-DSS: a computer tool for sustainable use of water resources at the catchment scale. Mathematics and Computers in Simulation, 64 (2004). pp.13-24. http://www.sciencedirect.com/science/article/pii/S0378475403001174

Global Land Cover 2000 database. European Commission, Joint Research Centre (2003) http://bioval.jrc.ec.europa.eu/products/glc2000/glc2000.php

Intergovernmental Panel on Climate Change (2006) 2006 Guidelines for National Greenhouse Gas Inventories. Egglestone, S., L. Buendia, K. Miwa, T. Ngara and K. Tanabe (Eds.). IPCC/OECD/IEA/IGES, Hayama, Japan.
http://www.ipcc-nggip.iges.or.jp/public/2006gl

Hansen, M., R. DeFries, J.R.G. Townshend, and R. Sohlberg (2000), Global land cover classification at 1km resolution using a decision tree classifier, International Journal of Remote Sensing. Vol.21. pp. 1331-1365.
http://www.glcf.umd.edu/library/guide/ijrs21_p1331.pdf

Hansen, M., R. DeFries, J.R.G. Townshend and R. Sohlberg (1998) UMD Global Land Cover Classification, 1 Kilometer, 1.0, Department of Geography, University of Maryland, College Park, Maryland, 1981-1994.

Hiederer, R. and R.J.A. Jones. (2009) Development of a Spatial European Soil Property Data Set. EUR 23839 EN. Luxembourg: Office for Official Publications of the European Communities. 30pp.
http://eusoils.jrc.ec.europa.eu/esdb_archive/eusoils_docs/other/EUR23839.pdf

FAO/IIASA/ISRIC/ISSCAS/JRC (2012) Harmonized World Soil Database (version 1.2). FAO, Rome, Italy and IIASA, Laxenburg, Austria.
http://webarchive.iiasa.ac.at/Research/LUC/External-World-soil-database/HTML/

Jenny, H. (1941) Factors of soil formation. McGraw-Hill Book Company, New York, London. 191pp. http://soilslab.cfr.washington.edu/S-7/Publications/Jenny-Factors%20of%20Soil%20Formation.pdf

Jiang, H. and J.R. Eastman (2000) Application of fuzzy measures in multi-criteria evaluation in GIS. International Journal of Geographical Information Science, Vol.14: 2. pp. 173-184.
DOI: 10.1080/136588100240903

Jones, R.J.A., R. Hiederer, E. Rusco, P.J. Loveland, and L. Montanarella (2003) Topsoil organic carbon in Europe. Proceedings of the 4th European Congress on Regional Geoscientific Cartography and Information Systems, 17-20 June 2003, Bologna, Emilia Romagna, Direzione Generale Ambiente e Difesa del Suolo e della Costa, Servizio Geologico, Sismico e dei Suoli. pp. 249-251.

Jones, R.J.A, R. Hiederer, E. Rusco, P.J. Loveland and L. Montanarella (2005) Estimating organic carbon in the soils of Europe for policy support. European Journal of Soil Science, October 2005. pp. 655-671. http://onlinelibrary.wiley.com/doi/10.1111/j.1365-2389.2005.00728.x/abstract;jsessionid=5B4CAAC4E16223480C1089B2F0DAA5F8.d01t04

King D., J. Daroussin and R. Tavernier (1994) Development of a soil geographical database from the soil map of the European Communities. Catena (21). pp. 37-56.

Kleeschulte, S. (2006) European land cover mapping. Presentation at the North American Land Cover Summit, Washington DC, September 2006.

Lagacherie, P., A. McBratney and M. Voltz (Eds.) (2006) Digital Soil Mapping - An Introductory Perspective. Elsevier, Amsterdam, The Netherlands. 658pp. ISBN: 978-0-444-52958-9.

Panagos, P. (2006) The European soil database. Geo: International, July/Aug 2006 Volume 5, Issue 7. pp. 32-33.

McBratney, A.B. and I.O.A. Odeh (1997) Application of fuzzy sets in soil science: fuzzy logic, fuzzy measurements and fuzzy decisions. Geoderma (77). pp. 85-113.

McBratney, A.B., M.L. Mendonça Santos and B. Minasny (2003). On digital soil mapping. Geoderma 117:1–2. pp. 3–52. doi:10.1016/S0016-7061(03)00223-4 http://www.sciencedirect.com/science/article/pii/S0016706103002234

Reuter H.I, A. Nelson and A. Jarvis (2007) An evaluation of void filling interpolation methods for SRTM data, International Journal of Geographic Information Science. 21:9. pp. 983-1008.

Saaty, T.L. (1977) A Scaling Method for Priorities in Hierarchical Structures. J. Math. Psychology, 15. pp. 234-281. http://www.sciencedirect.com/science/article/pii/0022249677900335

Scull, P. J. Franklin, O.A. Chadwick and D. McArthur (2003) Predictive soil Mapping: a review. Progress in Physical Geography 27:2. pp. 171–197.

Simonson, R.W. (1959) Outline of a generalized theory of soil genesis. Soil Science Society of America Journal 23.2 (1959). pp. 152-156.

Skov, F. and J-C. Svenning (2003) Predicting plant species richness in a managed forest. Forest Ecology and Management 180. pp. 583–593. http://www.sciencedirect.com/science/article/pii/S0378112702006461

Tenenbaum, D.E., L.E. Band, S.T. Kenworthy and C.L. Tague (2006) Analysis of soil moisture patterns in forested and suburban catchments in

Baltimore, Maryland, using high-resolution photogrammetric and LIDAR digital elevation datasets. Hydrol. Process. 20, pp. 219–240. http://onlinelibrary.wiley.com/doi/10.1002/hyp.5895/abstract

Triantaphyllou, E., B. Shu, S. Nieto Sanchez and T. Ray (1998) Multi-Criteria Decision Making: An Operations Research Approach. Encyclopedia of Electrical and Electronics Engineering. Vol. 15. pp. 175-186. http://www.csc.lsu.edu/trianta/EditedBook_CHAPTERS/EEEE1.pdf

Triantaphyllou, E., P.M. Pardalos, and S.H. Mann (1990) The problem of determining membership values in fuzzy sets in real world situations. In: Operations Research and Artificial Intelligence: The Interaction of Problem Solving Strategies. D.E. Brown and C.C. White III (Eds.). Kluwer Academic Publishers. pp. 197-214. http://www.csc.lsu.edu/trianta/EditedBook_CHAPTERS/EstimMembershipFuzzySetsByTRIANTA1.pdf

Van Liedekerke, M., A. Jones and P. Panagos (2006) ESDBv2 Raster Library - a set of rasters derived from the European Soil Database distribution v2.0 (published by the European Commission and the European Soil Bureau Network, CD-ROM, EUR 19945 EN). http://eusoils.jrc.ec.europa.eu/ESDB_Archive/ESDB_data_1k_raster_intro/ESDB_1k_raster_data_intro.html

Zhu, A-Xing, F. Qi, A. Moore and J.E. Burt (2010) Prediction of soil properties using fuzzy membership values. Geoderma, 158:3–4. pp. 199-206. ISSN 0016-7061, 10.1016/j.geoderma.2010.05.001. http://www.sciencedirect.com/science/article/pii/S0016706110001497

European Commission
EUR 25932 EN – Joint Research Centre – Institute for Environment and Sustainability

Title: **Mapping Soil Typologies - Spatial Decision Support Applied to European Soil Database**

Author: Roland Hiederer

Luxembourg: Publications Office of the European Union

2013 – 127 pp. – 21.0 x 29.7 cm

EUR – Scientific and Technical Research series – ISSN 1831-9424 (online)

ISBN 978-92-79-29468-6 (pdf)

doi:10.2788/87286

Abstract

For many applications of modelling environmental conditions or processes soil characteristics are needed in form of spatial data ready to be integrated in a GIS. A source of uniform data on characteristics of European soils is available from the European Soil Database (ESDB) of the European Soil Bureau. The soil information was collected by participating national institutions and underwent an extensive process of harmonizing the thematic content of recording the soil characteristics and ensuring spatial continuity along boundaries. In the database a many-to-1 link is used to relate soil characteristics to the geographic .layer. Thus, considerably effort is required to represent specific soil characteristics in a single spatial layer.

In this evaluation of the ESDB an attempt was made to position the soil typological units within the spatial units. The approach followed in this study for the spatial disaggregation of typological units was to link the data from the soil typology table with thematically corresponding ancillary spatial data. For the spatial allocation a multi-criteria analysis within the framework of a spatial decision support system (sDSS) was used. This approach allows generating single raster layers for any of the soil characteristics, independently of the data type. This greatly simplifies the mapping of soil characteristics and allows the use of the complete pool of information available in the ESDB. However, in the absence of an estimation of uncertainties in the source data issues of model uncertainties remain to be addressed.

LB-NA-25932-EN-N

As the Commission's in-house science service, the Joint Research Centre's mission is to provide EU policies with independent, evidence-based scientific and technical support throughout the whole policy cycle.

Working in close cooperation with policy Directorates-General, the JRC addresses key societal challenges while stimulating innovation through developing new standards, methods and tools, and sharing and transferring its know-how to the Member States and international community.

Key policy areas include: environment and climate change; energy and transport; agriculture and food security; health and consumer protection; information society and digital agenda; safety and security including nuclear; all supported through a cross-cutting and multi-disciplinary approach.

Publications Office

ISBN 978-92-79-29468-6