LABORATORIUM VOOR REGIONALE
BODEMKUNDE EN LANDEVALUATIE

SOIL SCIENCE DEPARTMENT
ORLEANS RESEARCH CENTER

Soil Survey and
Land Research Centre
CRANFIELD

LAND RESOURCE
PARTNERSHIP

A GEOGRAPHICAL KNOWLEDGE DATABASE ON SOIL PROPERTIES FOR ENVIRONMENTAL STUDIES

based on 1:1,000,000-scale EC soil map data

Final report of EC contract n° 3392004
Commission of the European Communities (DG XI)
November 1994

NATIONAL AGRONOMIC RESEARCH INSTITUTE
Soil Survey of France (SESCPF)
Research Centre of Orleans
Ardon 45160 OLIVET
FRANCE

UNIVERSITY OF GHENT
Geological Institute
Department of Regional Pedology
Krijgslaan 281
B 9000 GHENT
BELGIUM

SILSOE CAMPUS
Soil Survey and Land Research Centre (SSLRC)
Silsoe College, BEDFORD MK45 ADT
UNITED KINGDOM

LAND RESOURCE PARTNERSHIP
Soil Hydrology Consultancy, Land Systems Research
3 Axton Hill Pembroke Dyfed SA71 5HD
UNITED KINGDOM

CONTENTS

Preamble

This work is the result of a contract signed between the Commission of the European Communities (European Environmental Agency Task Force, DG XI) and a group of institutions: the National Agronomic Research Institute (INRA) in France, the Laboratory for Soil Science of the Ghent University in Belgium, and the Soil Survey and Land Research Centre (SSLRC) at Silsoe Campus and the Land Resource Partnership (LRP), both in England. The research carried out by this group relies heavily on previous work by the Soil & GIS Support Group within the framework of the EC MARS programme.

J. DAROUSSIN	INRA (F)
J.M. HOLLIS	SSLRC (UK)
M. JAMAGNE	INRA (F)
R.J.A. JONES	SSLRC (UK)
D. KING	INRA (F)
C. LE BAS	CLAIRE (F)
L. NGONGO	INRA (F)
A. THOMASSON	LRP (UK)
L. VANMECHELEN	GHENT UNIVERSITY (B)
E. VAN RANST	GHENT UNIVERSITY (B)

COORDINATION: D. KING (INRA, F)

1 - INTRODUCTION

Soil is a non-renewable natural resource, whose proper management is essential for both continued agricultural productivity and protection of the environment. Soil has a number of relevant ecological functions (Batjes, 1990): production of biomass, filtering, storage, gene reserve for biota and protection against exogenous processes. Soil vulnerability can be defined by the capacity to be harmed in one or more of its ecological functions. Major processes that influence soil vulnerability are acidification, eutrophization, pollution, salinization and erosion.

These problems on land use and soil conservation require increasingly accurate information on soil properties and their geographical location. An important point is to obtain harmonized data over the high diversity of regions. For the territory of the European Community, the Commission has suggested different approaches for twenty years. One of these is the publication of the EC soil map at scale 1:1,000,000 (CEC, 1985). Such a soil map provides answers to the above problems, thus helping in general decision making.

Until recently, spatial soil data were mainly available as paper-printed soil maps, which are graphically constrained and thus cannot hold an infinite quantity of information. They also have to present a simplified view of reality in order to be readable. Finally, most users agree that maps are often correct, but still may be difficult to read (Msanya *et al.*, 1987). Partly to resolve this problem, the EC soil map was computerized in 1986 (Platou *et al.*, 1986). This version was called "version 1.0", and enabled rapid processing and presentation of thematic documents that are more directly accessible to non-specialists. However, only the information on the map was included in the geographical database, and this resulted in a loss of data and risk of error in interpretation of soil properties (King *et al.*, 1993).

In 1990, the "Soil & GIS" support group of the EC MARS programme, decided to check and to improve the description of soil-map units from unpublished documents, from the national archives preserved at the University of Ghent. A computer data structure to receive such information *a posteriori*, and the database's improvement in term of quantity as well as quality was demonstrated. This version was called 2.0.

But this action was not sufficient to answer specific demands, particularly those concerned by environmental problems mentioned above. Many data were missing although they were implicit into present data. The Environment General Directorate (DGXI) thus asked to draw up procedures to facilitate the use of the EC soil database. The present report presents a method to translate data stored in the database to data needed for environmental purposes. This method is based on the concept of pedotransfer function (Bouma and Van Lanen, 1986). Due to the qualitative nature of the data, these functions are simple tables and we call them pedotransfer rule.

The combined pedotransfer rule form a sort of expert system that enables automatic interpretation of the soil map and its associated database. The advantage of this method is that the interpretations are explicit and can themselves be managed in a database, which we call a **knowledge database**. The rule presented hereafter are the result of consensus reached by the different participants. They can updated when necessary, either by adding new data in the future, or on the basis of particular features of certain regions.

The report consists of three parts: the first presents the methodology developed; the second describes each pedotransfer rule; and the third illustrates some validation and map-representation tests. The detailed lists of all rules are provided in appendix 6 in a standard format.

2 - METHODS

The method is based on creating a set of rules for estimating attributes needed for environmental work, based on attributes contained in the geographical database. First, we will examine the attributes used for input and exit of the system. Then, we will describe the typical structure of a rule with detailed explanations of the methodological choices adopted. Finally, we will propose a computer-based structure for automatic implementation of the rule.

2.1 - The attributes

2.1.1 - Input attributes

2.1.1.1 - Number of database version

The input attributes of pedotransfer rule correspond to attributes of the Version 2.0 database. They are descriptors of map units and are also called in the attribution document. Work is presently ongoing to correct and improve this database. Each country should soon provide new harmonized data (INRA-JRC, 1993). The new version will be called 3.0.

In view of the state of progress on Version 3.0 it was proposed to work with Version 2.0, for which we have completed and harmonized data for all of the EC and for some Central and East European countries. This will provide uniform short-term estimates at the scale of Europe. However, the work was oriented so as to render possible a refinement or complement of the pedotransfer rule as soon as Version 3.0 will be available. Certain of such rules were made for estimating input attributes that will become available in Version 3.0, such as organic surface carbon. Such estimates will then serve as validation tests once Version 3.0 will be ready.

2.1.1.2 - "Objects" for application of the rules

The geographic database consists of three geographic "objects" (King *et al.*, 1993): 1) Mapping polygons that show the geometry of soil units; 2) Soil Mapping Units (SMU) that combine polygons of the same soil type; 3) Soil Typological Units (STU) formed by the main soil types contained in the SMUs. SMUs and STUs are respectively described in two tables, showing horizontally the list of units and in columns the attributes describing such units.

Pedotransfer rules are applied at the level of STUs. In Version 2.0, certain attributes are attached to SMUs, such as parent material. A procedure of automatic allocation to an STU is planned for. For two materials that concern an SMU, parent material 1 is assigned to the dominant STU, i.e. the one with the largest surface area in the SMU. Parent material 2 is then assigned to the other STUs combined. In Version 3.0 we will dispose over all attributes at the level of STUs.

After completion of Version 3.0, it may be necessary to revise and enlarge the pedotransfer rules. This should not be a major problem in view of the ascendant compatibility planned for between versions 2.0 and 3.0 of the database. However, both rules and Version 3.0 will use a fourth data level that is the soil horizon. For this work, we have limited to two the maximum number of such horizons for each STU. The attributes describing such horizons will be directly attached to the STU and shall be called "Topsoil attribute" and "Subsoil attribute". In the present development of the knowledge database, the horizon thus is not considered as an "object" in its own right, but this point can be revised once Version 3.0 will be operational.

2.1.1.3 - Double notation of input attributes

For many input attributes a double notation is available, which enables indication of strong spatial variability within an STU. This is the case, for instance, for the texture and slope class. We have opted for selecting the attributes mentioned first and considered to be dominant in STUs. Tests that take the second values into account should be scheduled, in order to evaluate the sensitivity of pedotransfer rules to intra-unit spatial variability.

2.1.1.4 - List of input attributes

Definitions of the above-mentioned attributes are indicated in the working documents of the Soil & GIS working group and in the documents concerning the creation of the geographic database of European soils.

Table 1: List of input attributes of Version 2.0 (INRA, 1990)

Attributes in the version 2.0	**Definition**
FAO Soil name (SN)	cf. (FAO, 1975) and (CEC, 1985)
Topsoil texture class (TS)	1 Coarse 2 Medium 3 Fine/Medium 4 Fine 5 Very Fine
Slope (SL)	a Level (0-8%) b Sloping (8-15%) c Moderately steep (15-25%) d Steep (> 25%)
Parent Material (PM)	cf. (CEC, 1985), (INRA-JRC, 1993 Users guide)
Phase (PHA)	cf. (CEC,1985)
Land Use (U1)	cf. (INRA-JRC, 1993)
Surface percentage of SMU (%AREA)	% STU/SMU

Table 2: List of new input attributes of Version 3.0 (INRA-JRC, 1993)

Attributes	Classes
Depth to textural change (DT)	1 Textural change between 20 and 40 cm depth 2 Textural change between 40 and 60 cm depth 3 Textural change between 60 and 80 cm depth 4 Textural change between 80 and 120 cm depth 5 No textural change between 20 and 120 cm depth
Subsurface textural class (TD)	1 Coarse 2 Medium 3 Fine/Medium 4 Fine 5 Very Fine
Obstacle to roots (ROO)	1 No obstacle between 0 and 80 cm depth 2 Obstacle to roots between 60 and 80 cm 3 Obstacle to roots between 40 and 60 cm 4 Obstacle to roots between 20 and 40 cm
Presence of an impermeable layer (IL)	1 No impermeable layer within 150 cm 2 Impermeable layer between 80 to 150 cm 3 Impermeable layer between 40 to 80 cm 4 Impermeable layer within 40 cm
Water regime (WR)	1 Dry 2 Medium 3 Wet 4 Very Wet (cf. def. INRA-JRC, 1993)
Water management (WM)	cf. (INRA-JRC, 1993)

2.1.1.5 - External input attributes

Pedotransfer rules in certain cases require input attributes not present in the geographic database, e.g. the sum of temperatures. In other cases, notwithstanding the presence of a attribute in a soil database, it may be decided to use other, more precise, data sources for this attribute, e.g. land use, elevation. Such attribute are called "External". Their use implies the geographic combination of soil data with the geographic base of such external attributes. In order to harmonize the procedures, standard recoding of external attributes is planned, which themselves are structured according to the pedotransfer rules. The list of external input attributes is given in Table 3.

Table 3 : Name and class of external input attributes

Attributes	Classes
Regrouped land use class (USE)	C: Cultivated HG: Halophile Grassland MG: Managed Grassland SN: Natural and semi-natural land use
Accumulated mean annual temperature (AT)	H: High (≥ 3000°C) M: Medium (1800-3000°C) L: Low (< 1800°C)
Elevation (ALT)	U: Uplands and mountains L: Lowlands and intermediate altitudes

2.1.2 - Output attributes

Pedological output attributes were selected on the basis of the environmental parameters needed for the problems faced, e.g., hydrology of soil types for predicting catchment response to rainfall and standard percentage of run-off; location and sensitivity of wetlands; soil buffering capacity for predicting soil susceptibility; ecosystem and surface water deposition; vulnerability of ground- and surface water to pollution by agrochemicals and farm waste; soil erosion potential, etc.

Table 4 gives an overview of attributes required to develop expert systems to derive thematic maps of important environmental parameters. Some of the attributes required can be derived directly or recoded from the EC soil database; others are secondary or tertiary. Secondary attributes must be derived from primary attributes via pedotransfer rules, while tertiary attributes must be derived from a combination of primary and secondary attributes via pedotransfer rules. In theory it is possible to have only secondary attributes, but we prefer to retain the knowledge path of deriving attributes, in order to improve it for future versions of the rules. Moreover, it will allow easy updating of the rules with Version 3.0 of the geographical database.

Table 4: Attributes required to establish a database of major environmental parameters

ENVIRONMENTAL PARAMETERS	ATTRIBUTES REQUIRED
Predicting catchment overflow hazard	- Depth to an impermeable layer - Depth to a gleyed horizon - Hydrogeological class - Presence of a raw peaty topsoil - Maximum packing density
Location and sensitivity of wetlands	- Hydrogeological class - Soil wetness class - Depth to an impermeable layer
Soil buffering capacity (Loveland, 1990)	- Depth to rock - Topsoil textural class - Maximum CEC class - Maximum base saturation
Potential for the immobilization of radionuclides (Livens and Loveland, 1988)	- Soil class - Topsoil organic carbon content - Topsoil textural class - Subsoil clay mineralogical class
Potential for contamination by radon	- Soil parental material - Soil wetness class - Soil porosity class
Vulnerability of groundwater to pollution by agrochemicals and farm wastes (Hollis, 1990)	- Soil class - Hydrological class - Topsoil textural class - Depth to rock - Depth to a gleyed horizon - Depth to a impermeable layer - Soil adsorption capacity - Presence of a raw peaty topsoil

Table 4 (following)

ENVIRONMENTAL PARAMETERS	ATTRIBUTES REQUIRED
Vulnerability of surface water to pollution by agrochemicals and farm wastes (Hollis, 1991b)	- Slope class - Hydrogeological class - Depth to a gleyed horizon - Depth to a impermeable layer - Soil adsorption capacity - Soil porosity class - Presence of a raw peaty topsoil
Soil erosion potential (Palmer, 1993; King *et al.*,, 1993)	- Slope class - Topsoil textural class - Topsoil organic carbon content

The attributes selected for this work are listed in Table 5. They are grouped into four classes that respectively correspond to attributes of biological, chemical, mechanical and hydrological nature.

Table 5: Name and classes of final attributes selected with their required inputs

OUTPUT ATTRIBUTES	INPUT ATTRIBUTES		OUTPUT (CLASSES)
	BIOLOGICAL ATTRIBUTES		
Topsoil organic carbon content (OC_TOP) (0 - 25 cm)	SN TS USE AT	- FAO soil name - Topsoil textural class - Regrouped land use class - Accumulated mean temp.	H(igh): > 6.0% M(edium): 2.1-6.0% L(ow): 1.1-2.0% V(ery) L(ow): ≤ 1.0%
Presence of a raw peaty topsoil (PEAT)	SN	- FAO soil name	Y(es) N(o)
	CHEMICAL ATTRIBUTES		
Soil profile differentiation (DIFF)	SN	- FAO soil name	H(igh) differentiation L(ow) differentiation O: No differentiation
Profile Mineralogy (MIN)	SN	- FAO soil name	(C)hemical or Geochemical (M)echanical or Physical MC: Chemical and Mechanical ND: No Differentiation
Topsoil Mineralogy (MIN_TOP)	PM MIN	- Parental material - Profile Mineralogy	KQ: 1/1 minerals + quartz KX: 1/1 minerals + oxides & Hy. MK: 2/1 and 1/1 minerals M: 2/1 and 2/1/1 non swelling m. MS: Swelling and non s. 2/1 m. S: Swelling 2/1 minerals TV: Vitric materials TO: Andic materials
Subsoil Mineralogy (MIN_SUB)	PM MIN	- Parental material - Profile Mineralogy	
Topsoil Cation Exchange Capacity (CEC_TOP)	TS OC_TOP DIFF MIN	- Topsoil textural class - Topsoil organic carbon content - Soil profile differentiation - Profile Mineralogy	L(ow): < 15 cmol(+)kg-1 soil M(edium): 15-40 H(igh): > 40
Subsoil Cation Exchange Capacity (CEC_SUB)	TD MIN-SUB	- Subsurface textural class - Subsoil mineralogical class	
Topsoil Base saturation (BS_TOP)	SN USE	- FAO soil name - Regrouped land use class	L(ow): < 50 % M(edium): 50-75 % H(igh): > 75 %
Subsoil Base saturation (BS_SUB)	SN MIN_SUB	- FAO soil name - Subsoil mineralogical class	L(ow): < 50 % H(igh): > 50 %

Table 5 (following)

OUTPUT ATTRIBUTES	INPUT ATTRIBUTES	OUTPUT (CLASSES)
	MECHANICAL ATTRIBUTES	
Depth to rock (DR)	SN - FAO soil name PHA - Phase PM - Parental material	S(hallow): 0-40 cm M(oderate): 40-80 cm D(eep): 80-120 cm V(ery) D(eep): > 120 cm
Depth to rock corrected by phase (DRPH)	DR - Depth to rock PHA - Phase	
Subsurface textural class (TD)	SN - FAO soil name TS - Topsoil textural class	1 Coarse 2 Medium 3 Fine/Medium 4 Fine 5 Very Fine
Topsoil structure (STR_TOP)	USE - Regrouped land use class SN - FAO soil name	G(ood) N(ormal) P(oor) H(umic) or Peaty soil O : Peaty subsoil
Subsoil structure (STR_SUB)	SN - FAO soil name	
Topsoil Packing Density (PD_TOP)	STR_TOP - Topsoil structure class TS - Topsoil textural class USE - Regrouped land use class	L(ow): < 1.4 g/cm^3 M(edium): 1.4 - 1.75 g/cm^3 H(igh): > 1.75 g/cm^3
Subsoil Packing Density (PD_SUB)	SN - FAO soil name STR_SUB - Subsoil structure class TD - Subsoil textural class	
	HYDROLOGICAL ATTRIBUTES.	
Parent material hydrogeological type (PMH)	PM - Parental material	R, C, S, L, H, M (cf. definition § 3.
Depth to a gleyed horizon (DGH)	SN - FAO soil name	S(hallow): 0-40 cm M(oderate): 40-80 cm D(eep): 80-120 cm V(ery deep): > 120 cm
Depth to impermeable layer (DIMP)	SN - FAO soil name PD_SUB - Subsoil packing density TS - Topsoil textural class	S(hallow): < 80 cm D(eep): > 80 cm
Hydrological class (HG)	SN - FAO soil name ALT - Elevation PMH - Parent material hydrological class	HG1: soil with permeable substratu remote from groundwater: seldom w HG2: lowland soil affected by groundwater, seasonally or perman wet, or artificially drained HG3: soil with impermeable layers within 80 cm depth, seasonally or permanently wet HG4: soils of the uplands and mountains
Available Water Capacity of the topsoil (AWC_TOP)	TS - Topsoil textural class PD_TOP - Topsoil packing density	V(ery) H(igh): > 190 mm H(igh) : 140-189 mm M(edium) : 100-139 mm L(ow): < 99 mm
Easily Available Water Capacity of the topsoil (EAWC_TOP)	TS - Topsoil textural class PD_TOP - Topsoil packing density	
Available Water Capacity of the subsoil (AWC_SUB)	TD - Subsoil textural class PD_SUB - Subsoil packing density	V(ery) H(igh): > 190 mm H(igh) : 140-189 mm M(edium) : 100-139 mm L(ow): < 99 mm
Easily Available Water Capacity of the subsoil (EAWC_SUB)	TD - Subsoil textural class PD_SUB - Subsoil packing density	

For each output attribute of the pedotransfer rules, we have indicated the necessary input attributes for making the estimates. We also indicated the values of the classes adopted. The latter were fixed in a rather broad manner, in view of the low level of precision in the input attributes. The thresholds selected for class intervals are the result of a compromise between currently established values in Soil Science, and the possible level of precision at this scale. The adopted values may not correspond to the thresholds necessary for environmental problems. However, multiplication of the number of classes certainly would have diminished the reliability of the pedotransfer rule and would have rendered the system unusable.

In our work, we limited ourselves to estimating the soil parameters necessary for environmental problems. We did not draw risk (or vulnerability) maps; such work would require the combination of soil attributes with physical (climate, relief), agronomic (agricultural exploitation structure) and industrial (type and place of polluting emissions) variables. Each case would also require a fine analysis of the problem, modelling of the processes, selection of the tolerance threshold, and validation through experimental field work. The development of pedotransfer rules is a preliminary work for such investigations; it should facilitate a general application to local studies for all of Europe, providing a first estimate of the soil parameters needed for environmental models.

2.2 - Structure and options for application of pedotransfer rules

This section describes the structure that was adopted for implementation of the system, and defines the retained options. An example of a typical session that makes use of one of the rules is given in the appendix 1.

2.2.1 - Choice of the computer system

Implementation of the system takes place within the Arc/Info Geographical Information System (GIS) software package, using its macro-programming language AML (Arc Macro Language). The reasons for this choice are: 1) the database of available information (soil descriptions) is stored and managed within Arc/Info ; 2) the resulting data (environmental parameters) have to be stored and managed within Arc/Info for mapping display purposes ; and 3) this implementation had to be made within time and means limits that did not allow for the acquisition of - and staff training in - a specialized software.

The implementation is tailored for use within the general context of deriving new information from existing one via expert knowledge and could be used in any field of interest. But in our case, it is primarily meant to provide the European Environmental Agency with spatialized environmental indicators that can possibly be derived from the Soils Database.

2.2.2 - Dataset, objects, attributes, values, NODATA:

All the information available in the field of interest is stored in a so-called "dataset", e.g. the Soil Typological Units (STU) dataset. The dataset is physically stored as a dataset Info file, and holds information on a number of "objects", e.g. a number of soil types such as Luvisols, Cambisols, etc. Each object is physically stored as a line or record in the dataset Info file.

The objects in the dataset have a number of characteristics called "attributes", e.g. soil types have a soil name, a texture, etc. Each attribute is physically stored as a column in the dataset Info file. Each object in the dataset has a particular "value" for each of its attributes, e.g. such soil has a soil name Luvisol, a coarse texture, etc. Each value is physically stored at the intersection of the object's record and the corresponding column in the dataset Info file.

Values generally follow a coding scheme before being physically stored in the dataset, e.g. the soil name Luvisol is encoded and stored as "Lo", coarse texture is stored as "1", etc. Some objects might not be fully described when some of their attributes are unknown, e.g. unknown texture of a soil. An unknown value for an attribute is called a "NODATA" value. As there is no pre-defined way of coding and physically storing NODATA values in Info files, each attribute coding scheme will have to make provision for a NODATA value code, e.g. # will mean unknown texture.

2.2.3 - Rules, occurrences, input attributes, output attributes, facts

Soil Science experts of the working group provide the system with pedotransfer rules. These rules, using expert knowledge, permit to derive new needed information from the existing factual information, "fact", describing an object of the dataset; e.g. the soil depth of a particular soil type can be inferred from both its known soil name and its parent material. A rule is physically stored as a rule Info file. The whole of rules composes a set of rules and is physically stored as a rules Info database.

A rule can be seen as a statement of the form:

IF <available information is ...> THEN <new information is ...>
ELSE IF <available information is ...> THEN <new information is ...>
...
ELSE IF <available information is ...> THEN <new information is ...>

Each line in this statement is called an "occurrence" of the rule. An occurrence is physically stored as a line or record in the rule Info file.

An occurrence can be seen as a statement of the form:

IF (or ELSE IF)
<factual value for attribute i is w
and factual value for attribute j is x
...
and factual value for attribute n is y>
THEN
<inform the object with value z for a new attribute m>

where attributes i to n provide the factual information (values w to y of an object), and attribute m provides the new - inferred - information (with value z). Attributes providing the factual information are the "input attributes" to the rule. The attribute providing the new - inferred - information is called the "output attribute" from the rule. Input and output attributes are physically stored as columns in the rule Info file.

Example:

IF <soil name is "eutric Cambisol" and parent material is "450">
THEN <soil depth is "Medium">
ELSE IF <soil name is "eutric Cambisol" and parent material is "700">
THEN <soil depth is "Medium">
ELSE IF <soil name is "dystric Cambisol" and parent material is "500">
THEN <soil depth is "Deep">

As with the dataset, "values" are physically stored at the intersection of each record and the input and output attributes in the rule Info file.

Therefore pedotransfer rules tables are describing the link, established through expert knowledge, between input attributes from the Soils Database and output attributes. The structure of a typical table is given in Table 6. The columns on the left correspond to values taken by the input attributes; the central columns provide estimated values and their confidence level (see section 2.2.6) ; the right-hand columns contain management attributes and the references of rule occurrences (see section 2.2.9). The lines indicate the possible occurrences of the rule, based on the values (or combinations thereof) for the input variables in the Soils Database.

Table 6: Standard table for describing a pedotransfer rule.

Input Attributes					**Output Attributes**		**Reference Attributes**		
Regional Codes	i	j	....	n	Class	Confidence level	Authors	Date	Notes

Input attributes in a rule must have the same definition (name, type, size, etc.) and coding scheme as their corresponding attribute in the dataset.

2.2.4 - Inferences

An "Inference" is the action of producing a new derived information to an object according: a) to the available information it provides, and b) to the rule that is activated. It proceeds in 5 steps:

1. The input attributes are identified in the rule.
2. The values for these attributes are retrieved from the object in the dataset and constitute a fact.
3. Occurrence of the rule that matches the fact is searched for by sequentially skimming the rule's occurrences.
4. The output attribute definition and value are retrieved from the matching occurrence
5. and are added to the object in the dataset.

When a rule is activated on a dataset, inference will occur for each object of the dataset, one after the other. The result will be a new attribute in the dataset, one for the whole dataset, to hold the new inferred values, one for each object. An attribute of the dataset that has been previously inferred using a rule is further considered as storing available information. It can thus be used as an input attribute to other rules.

2.2.5 - Wild cards

It is difficult, if not impossible, for an expert to foresee all cases that can possibly occur in a set of available data. Furthermore, in some cases many different values of a fact will lead to the same conclusion, e.g. [IF <texture is sandy or loamy or ...> THEN ...]. A "wild card" mechanism allows the expert to define occurrences of rules that will match different facts. For example:

IF <soil name is "eutric Cambisol" and parent material is "450">

THEN <soil depth is "Medium">
ELSE IF <soil name is "eutric Cambisol" and parent material is "any other parent material">
THEN <soil depth is "Deep">

The "any other" wild card will, by convention, be denoted as a star character (*).

A fact for which an exact matching occurrence can be found will receive this occurrence's output attribute value. A fact for which an exact matching occurrence cannot be found, will receive the output attribute value of the last occurrence of the rule that matches, if it can be found with the wild card convention. This assumes that an expert will construct a rule by refining its occurrences, considering the most general cases before the most particular cases.

When no matching occurrence at all can be found for a fact, no value is provided to the output attribute, thus leaving it "blank" (or "0" (zero) depending on the output attribute's type). This can lead to confusion if blank (or 0) are possible normal output values. Therefore, having a fully "wild carded" occurrence as header of a rule, will "pick up" all facts for which no valid occurrence can be found and force the output value to, say, the NODATA value.

Using these specifications, the above example will become:
IF <soil name is "any soil name" and parent material is "any parent material">
THEN <soil depth is "unknown">
ELSE IF <soil name is "eutric Cambisol" and parent material is "any parent material">
THEN <soil depth is "Deep">
ELSE IF <soil name is eutric Cambisol and parent material is "450">
THEN <soil depth is "Medium">

It has been agreed that the last occurrence examined in the rule, will be the one to retain. As the occurrences are sequentially skimmed in the order of the lines of the table, i.e. from top to bottom, the construction of rules is designed to list the occurrences from the most general to the most detailed expert evaluations. For instance, if the input variable is "FAO Soil Name", the STU noted "Bge" will accept all following occurrences: "B**", "Bg*", "*g*", etc. The order of occurrences would be "B**", "Bg*", "Bge". If the STU soil name only contains code "B", the first occurrence will be applied; if it contains detailed information of the type "Bge", the third occurrence will be applied.

2.2.6 - Confidence level

Expert knowledge is subject to evolution. Furthermore, the available data, and the inferences that can be made using that information and the expert knowledge, have a certain level of reliability. It is thus necessary to have a mechanism that will allow all available information (or factual values) held in the dataset, and each inferred information (or output value) held in the rule database, to be complemented with an evaluation of its reliability.

The reliability of information is called its "confidence level". Confidence levels are held by confidence level attributes, one for each attribute of the dataset, and one for the output attribute of each rule. Each object in the dataset thus has a confidence level value for each of its attributes, and each occurrence of each rule has a confidence level value for its output attribute.

Four classes are proposed, ranging from "high", via "medium" and "low" to "very low". When the definition of input attributes enables the direct evaluation of an output attribute, the level is "high". On the other hand, if it is known that a very strong variation exists in the values of an output attribute, the "low" level is retained. "Very low" is used in the case of missing input attribute values.

So as to warn the users against a too abusive use of pedotransfer rules, it was decided that the confidence level of an output value should be the minimum of the confidence levels of all the input attributes and its corresponding occurrence.

When an inference takes place, the following 4 steps complement those listed above in section 2.2.4:

6. The output confidence level attribute definition is retrieved from the matching occurrence,
7. and is added to the object in the dataset.
8. The minimum (worst) confidence level value is retrieved from the confidence levels of all attributes implicated in the inference process (input confidence levels of the object, and output confidence level of the occurrence).
9. The resulting confidence level value is added to the output confidence level attribute in the object.

We have seen that an attribute of the dataset that previously was inferred using a rule, can be used as an input attribute to other rules. Its confidence level will be used in the same way as for any other input attribute.

2.2.7 - Missing data

In many cases, data are missing from the dataset because there are unknown input values to some objects. Two options then are open: the first consists in not evaluating the output attribute, which then itself becomes a missing attribute. The second proposes to output the best value found using the wild card convention, but with an imposed "Very Low" confidence level.

Use of wild cards in the case of missing input data carries the risk that information is generated that has never existed. The two options proposed above make it possible to retrace for each mapping unit the origin of its estimates. Checks are especially possible through the making of maps of the output "Confidence Level".

2.2.8 - Regionalizing of rules

In general the rules are drawn up for all of the European territory. For making estimates, no attributes were used that might cause a strong regional bias. To avoid any drift that, locally, might become dominant, a systematic input attribute called "Region" is planned. The selected geographical level is that of the European administrative regions, called NUTS II, but the stacked coding for administrative regions (NUTS 0 = country, NUTS I and NUTS II) enables the easy writing of a rule at the scale of a country. For instance, a rule that is specific for Italy will be noted "32*" in the "Region" column.

The rules can thus be completed by specific occurrences for countries, without modification of the initial general structure. As the occurrences are skimmed in a sequential fashion, displacement is always from the most general to the most specific case.

Although not used at present this option will enable revision or refinement of any rule with the help of regional expert knowledge. Its use will require the geographic combination of soil and administrative boundaries.

2.2.9 - Management and updating of rules

Three identifiers were added to the structure of the table describing a rule. The first gives a pointer to the author(s) of each occurrence. An authors' references table is kept up to date. The second attribute defines the date of establishing the occurrence. The third attribute gives a pointer to explanatory notes, defining the reasons for selecting a certain estimate.

Such management attributes give insight into the origin of the proposed estimate. Moreover, in case an occurrence is updated, it is avoided that an old occurrence has to be eliminated in order to be replaced with a new one. The new one will rather be placed sequentially behind the old one. During application of a rule, the last occurrence accepted is the one retained, which will enable to keep trace of the subsequent updates effected.

2.2.10 - Expert and class type rules

The rules described above are called "expert type rules" as opposed to "class type rules". The latter are simple reclassification or recoding rules. They are used in any of the following cases:

1. convert the Info data type of an input attribute in the dataset from an unauthorized to an authorized type (e.g. binary to clear numerical);
2. reduce the number of different values for an input attribute (e.g. reclass detailed land use classes into less detailed land use classes);
3. recode the values of an input attribute (e.g. change codes to a more "speaking" coding scheme);
4. any combination of the above cases.

Class type rules accept only one input attribute and produce one output attribute. The input attribute has no limitation as to its Info data type. The output attribute follows the same limitations as those applicable to expert type rules.

Class type rules do not follow the wild card convention. Wild cards may not be used there.

Class type rules do not make use of the confidence level of the input attribute if it exists in the dataset, whereas expert type rules use all available confidence levels to compute an output confidence level.

Class type rules may or not produce a confidence level attribute together with the output attribute, but expert type rules always produce a confidence level attribute.

2.2.11 - Tools

A toolbox was developed on the basis of these specifications for the creation, deletion, editing, management, description, report and inference of rules. The tools also maintain a dictionary for the rules database (see appendices 1 and 2), legends for input and output attributes, and a last rule edit historical file.

A tracing mechanism allows the detection of forward and backward dependencies. This means that when a rule is inferred, the tree of rules that are depending on its results can be traced forward in order to be fired in the correct sequence (see appendix 4). Conversely backward tracing chases all the rules on the results of which one rule is depending (see appendix 5).

Other utilities run compatibility controls between rules and the dataset, i.e. check input attributes in the rules against their corresponding in the dataset. This includes historical compatibility, i.e. date of last inference must be checked against date of last edit of a rule.

Plotting tools make use of the dictionary of the database, its legends, its controls for historical compatibility, and of the rules' output confidence levels. It also provides a mechanism for the proper generalization of the attributes describing STUs - which is the level of the Soils Database at which the rules are run - to the SMUs - which is the level that can be plotted on a map (see 2.1.1). Therefore, each map of the results of a rule inference represents the dominant value of the output attribute over the polygons and can be provided together with both its corresponding confidence level and purity maps.

Finally a "WHY" tool is provided to allow the user to interactively point to a location on the map and ask why a rule has provided such a result. It will then give a full explanation of the inference that lead to the result.

These tools are provided as a command line language. They should be considered as a prototype that could be fully implemented at a future stage using an appropriate expert system development software and an ad-hoc interface to Arc/Info.

3 - PRESENTING PEDOTRANSFER RULES

In the following chapter, we present a list of the main pedotransfer rules, justifying the estimates made by experts. A first group of rules corresponds to the coding of external attributes. Four other groups correspond to pedotransfer rules that provide access to the main parameters needed for environmental work, i.e. the biological, chemical, mechanical and hydrological properties of soils.

3.1 - Coding of external attributes

3.1.1 - Land use (USE)

Data describing land use for each Soil Typological Unit are clustered in four classes according to the level of agricultural activity: 1) cultivated (*C*), 2) halophile grassland (*HG*), 3) managed grassland (*MG*), and 4) semi-natural and natural land use (*SN*). Table 7 lists the codes of the Version 2.0 database and their final classes.

Table 7: Coding of land Use

USE class	corresponding land use	Soil database classes
C	Cultivated	2,3,6,7,12,13,14,15,16,17,20,21
HG	Halophile grassland	11
MG	Managed grassland	1
SN	Semi natural land use	4,5,8,9,10,18,19

3.1.2 - Elevation

Elevation is an important parameter for estimating the level of biological and geochemical activity. Each soil mapping unit contains the maximum and minimum elevations of its area. A simple method to estimate mean elevation of an STU is to use these two values. To have an accurate method, we can use the DTM available at the EC scale. In this case, it will be necessary to calculate the mean elevation for each polygon of the soil map, or to transfer the soil data on each cell of the elevation grid.

Pedotransfer rules, being qualitative functions, do not need elevation data with a high precision. It thus suggested to code elevation with only two classes: 1) Uplands and mountains (*U*); 2) Lowlands and intermediate elevations (*L*). Areas are uplands if their mean elevation is over 300 m in the northern part of Europe and over 1000 m in the southern part.

3.1.3 - Temperature

Temperature is not present in the soil database Version 2.0. This variable needs access to an external meteorological database. The agro-meteorological model *GOA* (Brisson *et al.*, 1992) was used to calculate the accumulated mean temperature from the January 1st to the December 31st *(AT)*. A 20-year average of *AT* was determined for each NUTS2 class. This allows regionalization of the rule based on climatological spatial variability. Three classes are distinguished that group Ireland, the United Kingdom, Denmark, the Benelux and N- and E-Germany in class *L*, the Mediterranean areas in class *H* and the rest in class *M*. This regionalization will be used to combine temperature data with other variables through pedotransfer rules.

Table 8: Classes of accumulated mean temperature (AT)

AT	**range (°C)**
H(igh)	> 3000
M(edium)	1800-3000
L(ow)	< 1800

3.2 - Biological variables

3.2.1 - Organic-carbon content of topsoil

Soil organic matter plays an important role in many environmental functions of the soil. It provides exchange sites for plant nutrients, it has a favourable effect on soil erosivity and water availability, it influences soil vulnerability to acid deposition, etc.

Topsoil is defined as the surface horizon, after mixing of the surface 25 cm, corresponding to the average thickness of the plowed layer. Four classes are determined related to the behaviour of topsoil.

Table 9: Organic carbon class of the top soil - (0-25 cm)

OC_TOP	**range (%)**
H(igh)	> 6.0
M(edium)	2.1-6.0
L(ow)	1.1-2.0
V(ery low)	≤ 1.0

Of the input attributes, soil type *(SOIL)* and texture *(TS1)* are directly obtained from the EC soil database. For reasons of convenience, the land-use classes describing STU are regrouped in four classes *(USE)* that are important to consider in the estimation of the organic carbon content (Table 7, section 3.1.1).

Climatological variation within the EC will certainly affect the amount of organic matter in the topsoil, and its organic-carbon content is thus estimated from external variables. It is necessary to combine the geographical data layers before applying the pedotransfer rule. This is the case especially for climate data, but also for land-use data if soil data are combined with, for example, CORINE land-cover data.

The first 13 lines of the pedotransfer rule define *OC_TOP* for all soil types based on texture, land use and AT. Generally, clay content is positively related to organic matter content and semi-natural vegetation is expected to supply more organic matter to the topsoil than agricultural land. Temperature affects the decomposition rate of organic matter. In practice, only Luvisols, Podzoluvisols, Acrisols and Cambisols are defined by these first 13 lines, because all other soil types, as well as some gleyic and chromic subtypes, are further redefined in the rule. Special attention is given to soil types with gleyic properties (Histosols, Gleysols and gleyic subunits). Anaerobic conditions in the soil reduce turnover rate of organic matter after resulting in high amounts of organic carbon in the topsoil.

3.2.2 - Peat

The presence of topsoil peat can easily be inferred from the FAO soil name, when the definition of a *Histic* horizon is used as the criterion to identify raw peaty topsoils. Histic horizons are present in Histosols, and in intergrades between Histosols, Podzols (Pgh) and Gleysols (Ghh). However, Fluvisols, Solonchaks and other Podzols and Gleysols can also have a thin histic horizon. It was opted to classify only Histosols and Histo-intergrades as soils with a raw peaty topsoil (class *Y*), while all soils in which the presence of a peaty topsoil is not excluded received class *N*, with a decreased level of confidence (*l*). All soils in which the presence of a histic horizon is excluded, are classified *N* with a high level of confidence (*h*). In view of its shortness, all occurrences of the pedotransfer rule for PEAT are shown on Table 10.

Table 10 : Pedotransfer rule for *PEAT*

REG	SN	PEAT	CL
***	***	N	h
***	O**	Y	h
***	Jt*	N	l
***	J*g	N	l
***	Zg*	N	l
***	Gh*	N	l
***	Ghh	Y	h
***	Pp*	N	l
***	Pg*	N	l
***	Pgh	Y	h
***	Ph*	N	m
***	Wh*	N	l

3.3 - Chemical attributes

3.3.1 Mineralogical differentiation

3.3.1.1 - Soil profile differentiation (DIFF & MIN)

For several attributes, it is important to know the degree of differentiation of soils, in particular for estimating the influence of the underlying material on the mineralogy of the soil profile. Such differentiation can be geochemical and mineralogical (*C*), and is relatively moderate in a temperate humid climate, except for Podzoluvisols, Podzols and certain Planosols that, in a Mediterranean climate, show stronger differentiation. Differentiation can also be mechanical (*M*) (illuviation), causing accumulation of constituent minerals. The influence of parent material is especially strong for the subsurface. The main groups concerned are Acrisols, Podzoluvisols, Ferralsols, Luvisols, Podzols, Planosols and all luvic, dystric, stagnic and spodic sub-groups.

Non or little differentiated soils (*ND*), present in certain cases an "alteration complex" in the upper horizons, leading to mineralogical changes such as microdivision and neoformation. For such soils that are weakly differentiated, the influence of the parent material commonly is felt throughout the whole soil profile. The main groups concerned are Cambisols, Rendzinas, Gleysols, Lithosols, Fluvisols, Arenosols, Regosols and Vertisols.

Two rules can be distinguished. The first estimates the intensity of differentiation (*DIFF*). Three classes are defined : High (*H*), Low (*L*) and Not differentiated (*O*). The second estimates its type (*MIN*). For STU having a high or a low differentiation we use the classes defined above (*C, M, ND*) or a mixed classe (*MC*). These two attributes are used as input to estimate the mineralogical classes for the topsoil and the subsoil

3.3.1.2 - Mineralogical clay classes

Mineralogical classes can be deduced from the FAO Soil Name (*SN*) and from the Parent Material (MP). Presently available knowledge provides a first simple classification:

1) Dominantly kaolinitic material (*K*) essentially consists of old weathering products from more aggressive climates, i.e. Tertiary, Pliocene and Early Quaternary times. This implies old soils and a long pedogenesis, corresponding to the cumulative effects of several climate fluctuations.

2) Dominantly smectitic material (*S*) appears to consist mostly of weathering products of sedimentary rock (marl, claystone and other petrified mud), of basic volcanic rock, or of old alluvial deposits, generally marine and accumulated in a confined environment. Most also contain a high proportion of interbedded 2/1 minerals.

3) Unconsolidated volcanic parent material, containing allophane or weakly crystallized minerals, must be considered as a special case.

4) The other minerals can only be classified as mixtures, noted as "Mixed Class" (*M*).

This is a first approach that needs completion. Class M in particular covers too large a field of distinct mineralogical groups. Moreover, the presence of quartz and iron oxides in the clay fraction should be considered. The final classes as proposed are shown in Table 11; a specific class for limy soil might be added as well.

Table 11: Mineralogical classes

Dominant minerals in the clay fraction	Class
1/1 minerals + quartz	KQ
1/1 minerals + oxides and hydroxides	KX
2/1 and 1/1 minerals	MK
2/1 and 2/1/ non-swelling minerals	M
Swelling and non-swelling 2/1 minerals	MS
Swelling 2/1 minerals	S
Vitric minerals	TV
Andic minerals	TO

For the differentiated soil unit, the mineralogical characterization of the soil surface will be (*MIN_TOP*) and that for the subsurface (*MIN_SUB*). For soil that is little or not differentiated, a single characterization is needed. So as to apply a single logic, it is lanned to apply both characterizations, even if this will lead to two identical rules in the case of undifferentiated profiles. Two pedotransfer rules are presented in Appendix 6, the first corresponding to an estimate of the attribute *MIN_TOP* and the second to *MIN_SUB*.

3.3.2 - Cation Exchange Capacity (CEC)

The cation exchange capacity (CEC) is a chemically significant property that reflects the buffer system in soils (Johnson and Todd ; Clayton et al., 1991). Exchange reactions control the ionic equilibrium between solid and liquid phases in soil. The CEC of soil materials is determined by the ability of clay-organic complexes to exchange cations. To assess a CEC-class, we have to rely on available information in the EC soil database, on clay and organic matter content of the soil material.

Many studies have evaluated the contribution of organic matter and clay contents to CEC (Addiscott, 1970 ; Wright and Jones, 1972 ; Genon and Dufey, 1991). This is not an easy task because clay-organic formations change the charge characteristics of soil components. It is complicated by the fact that CEC can be distinguished into a permanent (CEC_P), and a variable or pH-dependent (CEC_V) type. The variable charges responsible for CEC_V are mostly located in organic matter, although particular types of clay also possess pH-dependent charges (Tan and Dowling, 1984). Both pH and the valence of the neutralizing cation of the analytical procedure used to measure CEC, affect the variable charges on clay-organic complexes. Although the CEC of organic matter is a highly variable property, a mean value of 265 cmol(+) per kg organic carbon was used to estimate the contribution of organic matter to total soil CEC.

The CEC of the clay fraction or apparent CEC (aCEC) can vary considerably. It depends on clay mineralogy, particle size, pH, adsorption of organic molecules, etc. With the available information it is impossible to consider all properties for assessment of a CEC class. However, mineralogical class *(MIN)*, inferred from soil name and parent material, can be introduced in the pedotransfer rule for the CEC of the subsoil *(CEC_SUB)*. For estimation of the CEC of topsoil *(CEC_TOP)*, more weight is given to soil organic matter and texture than to clay mineralogy. Only two mineralogical classes are distinguished: 1) intensively differentiated soils due to geochemical weathering *(MIN* classes *C* and *MC*; *DIFF* class *H)*, and 2) all other soils.

With these considerations in mind, we tried to estimate the aCEC of various soil materials in the EC. A database was created with 623 soil horizons of more than 300 profiles found in the literature, for which CEC, clay content and organic matter content were given. Apparent CEC was calculated with the following equation:

$$aCEC = \left(CEC_{tot} - \frac{OC\%xCEC_{oc}}{100} \right) x \frac{100}{clay\%} \qquad \textbf{(1)}$$

with aCEC = apparent CEC [cmol(+)/kg clay]

CEC_{tot} = total CEC [cmol(+)/kg soil]

CEC_{OC} = CEC of the organic matter [= 265 cmol(+)/kg OC]

We realize that this is only a rough estimate of the CEC of clay.

Another database containing 96 horizons, was compiled of soil materials for which the apparent CEC was given. Together, 719 values of aCEC are available, of 390 A horizons, 257 B horizons and 64 E horizons.

Many of the estimated aCEC values are unreliable and were filtered out of the database. This unreliability can result from the mathematical operation in case of low clay content, or can be due to uncertain basic information (e.g. CEC_{OC} or textural data of volcanic soils). The following soil materials are disregarded in the estimation of aCEC:

- soil materials with less than 8% clay;
- soil materials with more than 6% OC;
- soil materials originating from volcanic parent material;
- Bs horizons of podzols.

Three classes are chosen to estimate CEC from the variables available in the soil database (Table 12).

Table 12: Classes for the attribute CEC.

CEC_TOP class	**range** (cmol(+)/kg soil)
L(ow)	< 15
M(edium)	15-40
H(igh)	> 40

3.3.2.1 - Cation Exchange Capacity of the topsoil (CEC_TOP)

Equation (1) can also be written as:

$$CEC_{tot} = \frac{(OC\%xCEC_{oc}) + (clay\%xaCEC)}{100} \qquad \textbf{(2)}$$

Equation (2) was used to derive a CEC class for topsoil of each possible combination of *OC_TOP* class and *T1* class, using the database averages of OC% and clay% per class. A value of 265 cmol(+) per kg organic carbon was used as CEC of the organic matter. The aCEC values of the database were critically analysed, to deduce a fairly reliable mean value.

From the remaining 271 A horizons in the filtered database, the calculated aCEC values are separated from the analyzed values and divided into two groups according to their mineralogy (Table 13). Certain soil types have undergone a more intensive chemical weathering and are expected to have a clay mineralogy containing higher amounts of kaolinite and oxides. Soil types with *MIN* class C or MC and *DIFF* class H, i.e. soil types A*, F*, D*, Po and Ph, W* except We, Lap and Lgp, are regarded as having a more advanced clay mineralogy.

Table 13: Average aCEC values for A horizons.

GROUP	number of samples	mean aCEC	stand. dev.
analysed aCEC values	96	53	11
estimated aCEC values			
- soil material that underwent intensive geochemical differentiation	35	36	21
- other soil materials	140	59	26

The mean estimated aCEC of all A horizons of the more reliable database is 55, which is similar to the mean value of the analysed samples. The mean aCEC of weathered soil is considerably lower than the mean value of the other materials, which enables differentiation between the two soil groups when estimating the CEC class. Finally, *CEC_TOP* classes can be obtained, applying equation (2) and considering a separate aCEC value for the more intensively weathered soil types (Table 14).

Table 14 - Assessment of *CEC_TOP*

OC_TOP class	T1 class	CEC_TOP class	confidence level
VL	1	L(L)	m(m)
	2/3	L(L)	m(m)
	4	M(M)	m(l)
	5	M(M)	l(l)
L	1	L(L)	m(m)
	2/3	L(L)	l(m)
	4	M(M)	m(l)
	5	H(M)	l(l)
M	1	L(L)	l(m)
	2/3	M(M)	l(l)
	4	M(M)	l(l)
	5	H(M)	l(l)
H	1	M(M)	l(m)
	2/3	M(M)	m(l)
	4	H(M)	l(l)
	5	H(H)	m(l)

Between brackets: soil types with *MIN* class C or MC and *DIFF* class H

The confidence level is deduced from the number of agreements during the validation. Table 14 can be converted into a pedotransfer rule for CEC_TOP, as described in section 2.2.1.

3.3.2.2 - Cation Exchange Capacity of the subsoil (CEC_SUB)

The contribution of organic matter to the total CEC of the subsoil can be neglected, simplifying equation (2) into:

$$CEC_{tot} = \frac{clay\% * aCEC}{100} \qquad \textbf{(3)}$$

However, the influence of parent material on the CEC is more pronounced and therefore more mineralogy classes are differentiated. The corresponding aCEC values are again obtained from the aCEC database from which only soil horizons directly overlying the C or R horizons were selected. Then, the unreliable aCEC records were filtered out and the remaining soil material was regrouped according to their *CEC_SUB* class. Average aCEC values per group are presented in Table 15.

Table 15: Average aCEC values for subsoil horizons

MIN_SUB CLASS	**number of samples**	**mean aCEC**	**stand. dev.**
M	196	51	22
MS	21	61	16
MK	10	47	16
KX	0	-	-
TV*	3	66	36
TO*	16	87	56

*: unreliable data

The aCEC values used for assessment of *CEC_SUB* were deduced from the mean database values, modified by expert judgement. The *CEC_SUB* classes obtained by applying equation 3 are presented in Table 16 and converted to a pedotransfer rule in Table 17.

Table 16: CEC values of mineralogical classes

MIN_SUB class	**corresponding aCEC value** cmol(+)/kg clay
KX	35
MK	45
M	55
MS	65
TV	35
TO	85

Table 17: Assessment of *CEC_SUB*

MIN_SUB class	**aCEC** cmol(+)/kg clay	**T2 class**	**CEC_SUB class**	**confidence level**
KX/TV	35	1	L	h
		2/3	L	m
		4	M	l
		5	M	m
MK	45	1	L	h
		2/3	L	l
		4	M	m
		5	M	l
M	55	1	L	m
		2/3	L	l
		4	M	m
		5	H	l
MS	65	1	L	m
		2/3	M	l
		4	M	m
		5	H	h
TO	85	1	L	l
		2/3	M	l
		4	H	l
		5	H	m

3.3.3 - Base saturation (*BS*)

The soil characteristic "base saturation" (BS), combined with CEC, provides information on the nutrient status of the soil and on its susceptibility to acidification. Soil with low BS and CEC is the most vulnerable to acid deposition, and must be indicated on the EC soil map.

In the FAO classification base saturation has a hierarchical position. The base saturation level of 50% is used as differentiating limit between Mollic and Umbric A horizons, and thus determines the soil classification on the highest level (soil unit) of several soil types. Subsoil base saturation determines the distinction between Luvisols and Acrisols, and between Eutric and Dystric (or Humic) subunits of many other soil units. Much information can be gained simply from the database attribute "FAO soil name" (*SN*).

However, land use has an important impact on the base status. This is especially true for topsoil that has a low natural base status, but in which base saturation was artificially increased for cultivation purposes through the application of fertilizers and/or lime. Therefore, the dominant land use class *(USE)* is used as input attribute in the rule for topsoil base saturation *(BS_TOP)*. Soil in cultivated areas often is classified without considering man-made alterations in the base status. This is taken into account by decreasing the confidence level of the *BS_TOP* output in those situations.

Parent material strongly influences the base status of the soil, particularly in the subsoil. It was decided to consider the subsoil mineralogy class *(MIN_SUB)* in the assessment of subsoil base saturation *(BS_SUB)*.

The variable BS_TOP is divided in three classes (Table 18).

Table 18: Base saturation classes.

BS_TOP class	**range (%)**
L(ow)	< 50
M(edium)	50-75
H(igh)	> 75

3.3.3.1 - Base Saturation of topsoil (BS_TOP)

Many topsoils, in particular in cultivated areas, have a base status close to 100%. A single 50% limit, adopted from the FAO classification, insufficiently differentiates managed topsoils. Therefore an additional 75% limit is proposed to split the large group of topsoils with a base saturation higher than 50%. The three classes *H*, *M* and *L* roughly correspond to the land use types of cultivated soil, managed grassland and semi natural vegetation, respectively.

3.3.3.2 - Base Saturation of subsoil (BS_SUB)

The available database information is insufficient to classify subsoil into three base saturation classes. The combined information on soil type and parent material provides a reliable distinction between subsoil with a high and a low base status, using the FAO limit of 50%.

In the first part of the pedotransfer rule, *BS_SUB* is determined by the mineralogy class. For many soil types, however, the FAO classification defines the base status of the subsoil and overrule the *BS_SUB* class based on mineralogy.

3.4 - Physical variables

3.4.1 - Depth to rock (DR)

For estimating soil depth we dispose over three input attributes: FAO soil name (*SN*), material (*MP*) and phase (*PHA*). We have defined the attribute soil depth as the boundary betwen unconsolidated soil and hard, continuous, coherent and mittle weathered material. It is assumed that the structure of the material is a primordial element as it is strongly linked to rooting possibilities. In certain cases this will be the R horizon and in others it is the C horizon. Common difficulties occur with fragipans or, at the other end of the scale, with sand or other material without coherence.

Very few indications on soil depth are given in the FAO legend. In general, details are provided for the presence of a diagnostic horizon above a certain depth limit. Soil depth is only clearly given for lithosols (<10cm) and the lithic phase (<50cm). The evaluations mentioned in the rule in appendix 6, result from an average estimate using values observed on type sections. Depth is mainly related to soil name, but can be qualified by the presence of a specific material.

3.4.2 - Volume of stones (VS)

When stones are numerous, it may be noted using phases. For two classes, Gravelly (*02*) and Stony (*03*), stone volume represents about 15% of the total volume. Phases are noted according to the quantity of stones at the surface of the soil. We can extent this estimation for the solum. If the nature of stones is calcareous, the percentage decreases to 10% due to the capacity of calcareous to retain water. The nature of stones is deduced from parental material. Two other phase classes are used decreasing the total soil volume. The concretionary phase (05) and the petrocalcic phase (06) are coded respectively with 10 and 20 %.

3.4.3- Subsurface textural class

Textural class of soil in Version 2.0 is texture of the topsoil. Textural class of the subsoil is estimated from the FAO Soil Name. One class is added to the textural class of the topsoil for mainly Luvisols and Planosols. Version 3.0 will be more precise than 2.0, giving directly the subsoil textural class for each STU.

Table 19: Rule for estimating subsoil textural class

Soil name	Topsoil textural class	Subsoil textural class
***	1	1
W**	1	4
***	2	2
L**	2	4
W**	2	4
***	3	3
L**	3	4
W**	3	4
***	4	4
***	5	5

3.4.4 - Soil Structure (STR_TOP and STR_SUB)

In order to estimate the soil Packing Density (*PD*), it is necessary to set up a separate procedure (Pedotransfer Rule) to estimate Structure Class beforehand.

Soil Structure of the subsoil is assessed entirely from the Soil Name. Most soils are assumed to have normal (*N*) structure. Good (*G*) structure is confined to soils dominated by strongly developed fine or very fine, porous, subangular peds, and all granular peds. Poor (*P*) structure implies mainly massive, coarse or medium, angular blocky or prismatic peds. Soils not evidently Good or Poor are classed as Normal.

Table 20: Soil Structure Class of the Sub_Soil (*STR_SUB*) estimated from Soil Name

Regional Code	Input - Soil Name (Letter Code)		STR_SUB	
			Class	Confidence
* * *	a)	Bea	Good	h
		Bh	"	
		Eo	"	
		Hc	"	
		HL	"	
		Hl	"	
		To	"	
		Tv	"	
		pL	"	
	b)	Bgg	Poor	m
		Bvg	"	
		Gds	"	
		Ges	"	
		Lgp	"	
		Lgs	"	
		Pgs	"	
		Wd	"	
		We	"	
	All other STU's		Normal	l

Soil Structure Class of the topsoil (*STR_TOP*) has the same rule, but also depends on Land Use as an input (cf. section 3.11). The main elements are :

1) All permanent grass and semi-natural uses (*USE=HG, MG, SN*) assumed to be GOOD
2) All Soil Names in Table 20a) assumed to be GOOD
3) All other uses and soils assumed to be Normal

3.4.5 - Packing Density (PD_TOP and PD_SUB)

Packing density (*PD*) is a fine earth value, excluding stone, calculated from measured soil data:

$$PD = \text{Bulk Density} + \frac{\text{Clay Content}}{100} \qquad (g/cm^3)$$

Low (L)	< 1.4 g/cm^3
Medium (M)	1.4 to 1.75 g/cm^3
High (H)	> 1.75 g/cm^3

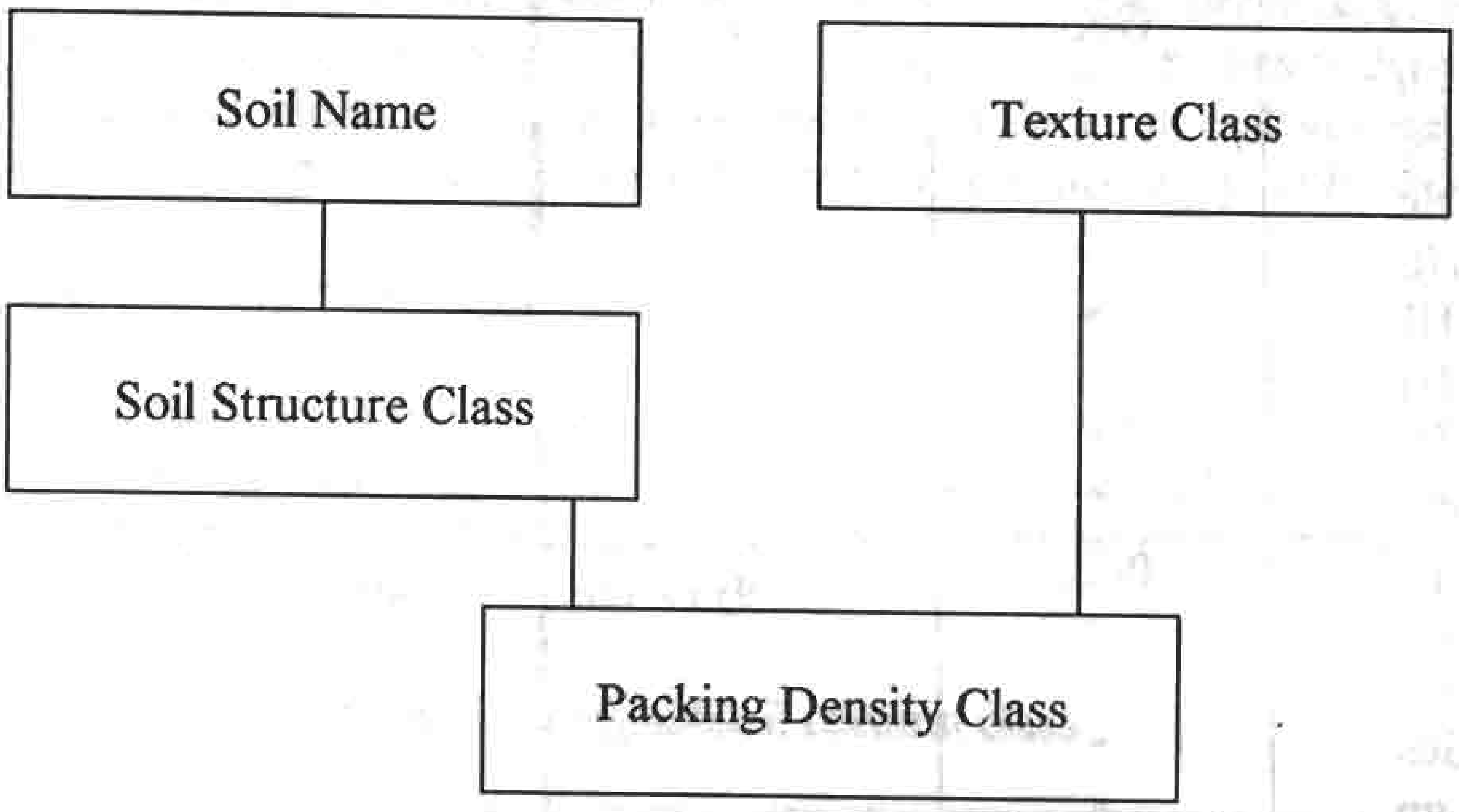
With caution, these classes can be broadly related to soil structural conditions as (Hall et al., 1977):

PD Class	Structure Class
Low (L)	Good
Medium (M)	Normal
High (H)	Poor

- provided a dominant 'normal' class is assumed.

The relationships are fairly robust for topsoils and medium textures, but less secure at the extremes of the texture range and in subsoils, where overburden pressures normally preclude low PD values.

This attribute is assessed from texture class, soil structure (cf above) and general horizonation (topsoil and subsoil) in the form of:



The general relationships of *PD* to texture and horizonation, under average land use and management conditions, are fairly clear. It is possible to set up a simple protocol to enable estimation of *PD* from the soil map legend. Obviously the extremes of land management conditions, with variation in soil structure and *PD*, will not be represented, but neither will they be shown on a 1:1,000,000 map, or indeed any map at a scale smaller than about 1:10,000. Jones and Thomasson (1993) describe the use of packing density as a parameter for environmental interpretation.

Table 21: Protocol to estimate *PD* Class from EC Texture Classes and Horizonation

EC Texture Class	**1 Coarse**	**2 Medium**	**3 Medium Fine**	**4 Very Fine**	**5 Fine**
Mineral Topsoils PD-TOP					
Normal structure. Use as default value in the absence of conflicting information	M	M	M	M	M
Good structure †	L	L	L	M	M
Humose or Peaty Topsoils #	L	L	L	L	L
Mineral Subsoils PD-SUB					
Normal structure and default value as above	M	M	M	H	H
Good structure †	L	M	M	M	H
Poor structure †	M	H	H	H	H

† See Table 20
# See Rule for attributes c and g
– Under arable land use these textures are usually Medium PD
NB For Fluvisols (J**), allocate low confidence for PD-SUB

Table 22: Estimating packing density

Regional Code	Input 1 Soil Structure	Input 2 Texture	Input 3 Horizon	Output PD Class	Confidence %
Availability of measured PD data	See Table 20	See Table 21	See Table 21	Low Medium High	M M M

3.5 - Hydrological variable

3.5.1 - Parent material: hydrogeological type (PMH)

To estimate the hydrological class, it is necessary to recode the attribute Parent Material. The definition of the new codes is indicated below:

R: hard, non- or weakly porous limestone (karstic), sandstone and crystalline rock with moderate storage capacity and high permeability because of well-developed fissure/joint systems.
C: chalk and soft limestone with bimodal porosity; microporous with moderate storage capacity, but well-developed fissure systems giving relatively high permeability.
S: weakly consolidated sandstone and unconsolidated sand and gravel with unimodal porosity; macroporous with large storage capacity and relatively high permeability.
L: weakly or unconsolidated microporous substratum with a low permeability and storage capacity.
H: hard massive rock with negligible permeability and storage capacity.
M soft massive substratum with negligible permeability and storage capacity.

3.5.2 - Depth to a gleyed horizon (DGH)

The attribute "Depth to a gleyed horizon" corresponds to the depth of a typical horizon with hydromorphic properties. The presence of such properties is either directly indicated by soil name, i.e. by the first letter of the FAO soil name, e.g. Gleysols, or by the presence of a suffix "g" (i.e. using the secondary and tertiary letters of the FAO soil name. Gleysols are defined by the presence of hydromorphic properties before 50 cm depth and thus are allocated to Class S (<40 cm) that is more restrictive than Class M (40-80 cm). It would have been possible to modify the depth classes to bring them into line with the FAO definitions, but we preferred to remain coherent with the definition of the attribute Depth to rock (*DR*). The suffix "g" does not indicate a particular depth and we have opted to attribute it to Class *M*. Histosols are treated as a separate case and are excluded from the rule.

3.5.3 - Depth to impermeable layer (DIMP)

Impermeable subsoil layers are identifiable from packing density and texture, using Table 23. Essentially, all mineral subsoil layers, with high packing density in texture classes 2, 3, 4 and 5, are considered to be impermeable. The criteria will require validation from national databases using EC texture classes, and air capacity values at 5 kPa equilibrium suction as a surrogate for K_{sat}.

Full application of these procedures should require more information, for example on parent material or depth to change of texture. At Version 2.0 level it is possible to identify only two classes of depth to impermeable layer:-

Class	**DIMP cm**
Shallow	< 80
Deep	> 80

The 'Shallow' class is defined as likely to occur in STU's having surface textures of classes 2, 3, 4 and 5 and Soil Names as listed in Table 23. All other STU's are considered as 'Deep'. The terminology used to define these two classes is different of "depth to rock" classes. In order to avoid confusion, we call the first DIMP class "*SM*" and the second class "*DVD*".

Table 23: Pedotransfer rule for depth to impermeable layer (DIMP)

Regional Code	Input 1	Output	Confidence Level
	Soil Name	DIMP	
***	Bv	SM	m
	Bgg	SM	m
	Bvc	SM	m
	Bvg	SM	m
	Dgs	SM	m
	Gds	SM	m
	Ges	SM	m
	Lgp(1)	SM	m
	Lgs(1)	SM	m
	Lv(1)	SM	m
	Lvk(1)	SM	m
	Vc*	SM	m
	Wd	SM	m
	We	SM	m

Notes:
(1) Other Luvisols (L**) with High PD-SUB are considered to have SM
(2) If surface texture class is 1 (Coarse), the confidence level is reduced to low (l)

3.5.4 - Hydrogeological Class (HG)

The data of the Version 2.0 database are insufficient to estimate the hydrological class for soils with a high accuracy. It needs a Version 3.0, or better, a soil profile database. Nevertheless, a simpler system, is offered in Table 24. A similar system has been tested and operated on 1:1,000,000 scale data in the UK (Thomasson, 1975). Table 24 allocates *HG* Classes 2 and 4, from STU Soil Name and elevation. *HG* Class 1 comprises all other STU's not listed in Tables 23 or 25.

Table 24: Hydrogeological classification

HG 1: Soil with permeable substratum, remote from groundwater; seldom wet. Subclasses, differentiated according to geological substratum where possible, are:

1R	Fractured rock (hard, fissured limestone, sandstone and crystaline rock)
1C	Chalk and soft limestone (bimodal pore systems)
1S	Sand, terrace deposits, soft sandstone (unimodal pore systems)

HG 2: Soil affected by groundwater; seasonally or permanently wet, or artificially drained (includes soil affected by climatic wetness as well as by general hydrology, needing Regional Codes for differentiation).

HG 3: Soil with impermeable layers within 80 cm depth; seasonally or permanently wet.

HG 4: Soil of uplands and mountains; generally above 300 m in northern Europe and above 1000 m in southern Europe. Subclasses differentiated according to soil water regime where possible are:

4W	Wet, with organic surface horizon
4D	Dry, stony and/or shallow soil depth

Table 25: Pedotransfer rule to allocate Hydrogeological Classes 2 and 4.

Regional Code	**Input 1 STU Name**	**Input 2 Elevation**	**Output HG Class**	**Confidence level**
***	Bef, Bg, Bgc,	***	2	m
	Bh, Bkf,		2	m
	Gc, Gd, Ge,		2	m
	Gef, Gh, Gmf		2	m
	Jcf, Jcg, Jd, Je,		2	m
	Jeg		2	m
	Od, Oe		2	m
	Pg		2	m
	Zg		2	m
	Gh	>300m	4W	h
	Od	N. Europe	4W	h
	Od		4W	h
	Pp		4W	h
	Pgs		4W	h
	1c	>300m	4D	m
	1d	N.Europe	4D	m
	1e	or	4D	m
	rO	>1000m	4D	m
	U	S.Europe	4D	m

Soils (STU's not listed in Tables 23 or 25, are assumed to be allocated to HG Class 1. In Table 26 subclasses of HG1 are allocated to these STU's according to the parent material class as listed in Database 2.0 (INRA 1990)

Table 26: Allocation of HG1 Subclasses

HG	1R	Fractured rock 211, 215, 250, 450-456, 700-750
HG	1C	Chalk and soft limestone 213, 214, 216, 220, 530
HG	1C	Sand, terrace deposits and soft sandstone 400-442, 500-523, 600-640

3.5.5 - Available water capacity (AWC)

The values of water-retention capacity are attributed to the suctions corresponding to the available reserves between the field capacity and the wilting point (water held between 5 and 1500 KPa), and to the easily available reserve (water held between 5 and 200 KPa). These values are estimated from texture and density. Specific values are attributed to the surface horizon, to take account of the effect of agronomical activity (organic-matter content and tilling). Three classes are proposed, but we can use the numerical values for making a more accurate evaluation for the solum as a whole.

Table 27: Available water capacity (AWC) and Easily available water capacity (EAWC)

Texture	Packing density	AWC (5 to 1500 KPa)	AWC class	EAWC (5 to 200 KPa)	EAWC class
1	A horizon	130	M	130	M
2	A horizon	180	H	180	H
3	A horizon	210	VH	210	VH
4	A horizon	170	H	170	H
5	A horizon	170	H	170	H
1	L	120	M	80	L
2	L	210	VH	140	M
3	L	220	VH	140	M
4	L	200	VH	150	H
5	L	210	VH	150	H
1	M	80	L	60	L
2	M	160	H	100	M
3	M	190	VH	120	M
4	M	150	H	80	L
5	M	150	H	80	L
1	H	80	L	60	L
2	H	120	M	70	L
3	H	130	M	70	L
4	H	130	M	70	L
5	H	130	M	70	L

The table 27 is described with four pedotansfer rules. The first table estimates the AWC for topsoil and the second table estimates the EAWC for topsoil too. The third and fourth tables estimate the same attributes (AWC and EAWC) for subsoil.

4 - MAPPING AND VALIDATION TEST

Mapping and validation tests only concern some attributes. The objective is to show, using examples, the results obtained with some pedotransfer rules. Validation tests are carried out with data from soil-profile databases. The latter being limited in Europe, the same data are often used for calibration and for verification of the pedotransfer rules. As the rules were established through expert knowledge, it is is necessary to check that they do conform to reality. Such validation only concerns point data; in order to add a spatial dimension, it is planned to provide map printouts. Maps are efficient means for visualizing data and geographic anomalies. By climbing the successive deductive steps of pedotransfer rules, it is possible to identify the origins of such estimates. Chapter 4 will need more work, based on the section databases or the maps containing independent regional data that were used for drawing up pedotransfer rules.

4.1 - Validation of some parameters

4.1.1 - Organic carbon content of topsoil

A quick validation of the *OC_TOP* rule was done on a small soil database consisting of ploughed layers (Ap horizons at least 25 cm thick) of 101 soils located throughout the EC. The results of the validation are presented in Table 28. The performance of the rule is quite satisfactory (61 out of 101 samples were estimated correctly). It should be mentioned that only soil used as arable land was considered; the rule was not checked on soil under semi-natural vegetation or permanent grassland.

Table 28 - Partial validation of the *OC_TOP* pedotransfer rule on soils with a plow layer. Matching estimated data (OC_TOP) with actual data (CLASS).

REGION	FAONAME	HORIZON	CLAY	OC	CLASS	OC TOP
S- and E-Greece	Bcc	Ap	50.00	0.81	VL	L
France	Bcc	Ap	20.90	1.24	L	L
Moselle (F)	Bd	Ap	8.90	2.18	M	L
Moselle (F)	Bd	Ap	14.10	2.71	M	L
Limagne (F)	Be	Ap	19.00	0.70	VL	L
Limagne (F)	Be	Ap	19.00	0.93	VL	L
W-Gutland (L)	Be	Ap	16.55	1.52	L	L
Vendée (F)	Be	Ap	26.90	1.72	L	L
Ardennes (L)	Be	Ap	23.00	2.70	M	L
Frankfurt (G)	Bea	Ap	5.50	4.00	M	L
Greece	Bec	Ap	28.90	1.40	L	VL
France	Bec	Ap	53.90	1.87	L	L
Greece	Bec	Ap	42.20	1.92	L	L
England	Bef	Ap	20.00	2.00	L	L
England	Bgc	Ap	40.00	3.50	M	M
Ardennes (L)	Bgg	Ap	27.40	3.55	M	L
S- and E-Greece	Bk	Ap	51.80	1.13	L	L
S- and E-Greece	Bk	Ap	38.00	1.28	L	L
S- and E-Greece	Bvc	Ap	46.00	0.80	VL	L
S- Central Italy	Bvc	Ap	38.80	2.20	M	L
W-Gutland (L)	Dd	Ap	11.00	1.28	L	L
NW-Italy	De	Ap	11.20	1.65	L	L
Vendée (F)	Dg	Ap	13.20	1.31	L	L
Moselle (F)	Dg	Ap	18.20	2.65	M	L
E-Jutland (DK)	E	Ap	15.40	2.49	M	M
England	Gm	Ap	11.00	1.40	L	M
S- and E-Greece	Gm	Ap	43.00	2.23	M	M
S- and E-Greece	Gm	Ap	52.00	2.41	M	M
S- and E-Greece	Gm	Ap	36.00	3.81	M	M
England	Gmf	Ap	46.00	15.00	H	H

Table 28 (following)-

REGION	FAONAME	HORIZON	CLAY	OC	CLASS	OC_TOP
Massif Central (F)	Hc	Ap	70.60	3.48	M	M
Frankfurt (G)	Hh	Ap	13.20	1.70	L	M
Moselle (F)	Hh	Ap	15.20	2.60	M	M
E-Jutland (DK)	Hl	Ap	12.90	1.98	L	M
E-Jutland (DK)	Hl	Ap	14.80	2.05	M	M
S- and E-Greece	Jc	Ap	43.40	1.69	L	M
S- and E-Greece	Jc	Ap	64.00	3.14	M	M
Greece	Jcg	Ap	39.20	1.16	L	M
Greece	Jcg	Ap	27.60	1.34	L	M
Marais Poitevin (F)	Jcg	Ap	64.20	1.68	L	M
Greece	Jcg	Ap	71.20	2.79	M	M
Marais Poitevin (F)	Jcg	Ap	63.50	3.12	M	M
Moselle (F)	Jd	Ap	5.20	1.22	L	L
S- and E-Greece	Jd	Ap	29.00	2.19	M	L
S- and E-Greece	Je	Ap	19.00	1.13	L	L
Vendée (F)	Lc	Ap	21.60	1.59	L	L
Canary Islands (SP)	Lf	Ap	48.10	1.60	L	L
England	Lg	Ap	27.00	1.70	L	L
W-Gutland (L)	Lgs	Ap	25.30	0.60	VL	L
W-Gutland (L)	Lgs	Ap	10.10	0.80	VL	L
W-Gutland (L)	Lgs	Ap	21.15	1.44	L	L
S- and E-Greece	Lk	Ap	32.00	0.81	VL	VL
S- and E-Greece	Lkc	Ap	13.20	1.30	L	VL
S- and E-Greece	Lkv	Ap	34.00	0.63	VL	VL
Limagne (F)	Lo	Ap	15.00	0.47	VL	L
Greece	Lo	Ap	27.60	0.70	VL	VL
Limagne (F)	Lo	Ap	18.00	0.87	VL	L
Flanders (B)	Lo	Ap	7.50	0.90	VL	L
Greece	Lo	Ap	34.40	0.93	VL	VL
Flanders (B)	Lo	Ap	7.50	0.94	VL	L
Flanders (B)	Lo	Ap	7.20	0.95	VL	L
W-Gutland (L)	Lo	Ap	14.00	0.99	VL	L
Flanders (B)	Lo	Ap	7.00	0.99	VL	L
N-France	Lo	Ap	18.40	1.00	VL	L
Flanders (B)	Lo	Ap	8.80	1.03	L	L
Flanders (B)	Lo	Ap	11.00	1.06	L	L
Marais Poitevin (F)	Lo	Ap	24.40	1.08	L	L
Flanders (B)	Lo	Ap	8.40	1.08	L	L
W-Gutland (L)	Lo	Ap	11.05	1.12	L	L
Denmark	Lo	Ap	10.80	1.12	L	L
W-Gutland (L)	Lo	Ap	33.25	1.13	L	L
Flanders (B)	Lo	Ap	8.60	1.20	L	L
Denmark	Lo	Ap	11.50	1.40	L	L
Vendée (F)	Lo	Ap	15.90	1.47	L	L
W-Gutland (L)	Lo	Ap	31.05	1.67	L	L
W-Gutland (L)	Lo	Ap	22.00	1.77	L	L
Flanders (B)	Lo	Ap	10.20	1.89	L	L
W-Gutland (L)	Lo	Ap	27.25	2.82	M	L
Fontainebleau (F)	Ph	Ap	3.00	1.57	L	M
Fontainebleau (F)	Ph	Ap	2.00	2.44	M	M
Denmark	Ph	Ap	2.50	3.62	M	M
S- and E-Greece	Rc	Ap	35.00	0.90	VL	L
Vendée (F)	Rc	Ap	34.80	1.70	L	L
Vendée (F)	Rc	Ap	27.60	2.90	M	L
S- and E-Greece	Re	Ap	21.00	1.03	L	L
Greece	V	Ap	74.00	1.28	L	L
Greece	V	Ap	41.00	1.86	L	L
S- and E-Greece	Vc	Ap	40.00	0.60	VL	L
Central Greece	Vc	Ap	61.30	0.82	VL	L
Central Greece	Vc	Ap	67.40	1.06	L	L
S- and E-Greece	Vc	Ap	44.00	1.80	L	L
Canary Islands (SP)	Vp	Ap	61.20	0.28	VL	L
Central Greece	Vp	Ap	60.10	0.93	VL	L
Greece	Vpc	Ap	63.00	1.30	L	L
Champagne (F)	Wd	Ap	5.60	0.54	VL	VL
Limagne (F)	We	Ap	12.00	0.52	VL	L
Limagne (F)	We	Ap	21.00	1.74	L	L
Champagne (F)	We	Ap	13.70	1.96	L	L
Canary Islands (SP)	Xk	Ap	63.00	0.38	VL	VL
Canary Islands (SP)	Xy	Ap	59.00	0.38	VL	VL
Netherlands	pL	Ap	6.00	2.70	M	M

Number of samples	101	agreement	60 %
		occurrence CLASS>OC_TOP	12 %
		occurrence CLASS<OC_TOP	28 %

4.1.2 - Cation exchange capacity of topsoil

The rule for *CEC_TOP* was checked on a database of 95 plowed layers (Ap horizons of at least 25 cm thick). The results of this validation are presented in Table 29. The performance of the rule is surprisingly good with 61 agreements on 95 samples, but these results are biased because the same soil materials were used for the estimation of the aCEC values.

Table 29: Partial validation of the *CEC_TOP* pedotransfer rule on soils with a plow layer. Matching estimated data (CEC_TOP) with actual data (CLASS).

FAONAME	HORIZON	OC_CLASS	T1	CEC	CEC_CLASS	CEC_TOP
Bcc	Ap	L	2	17	M	L
Bcc	Ap	L	4	35	M	M
Bd	Ap	L	2	11	L	L
Be	Ap	L	2	12	L	L
Be	Ap	L	2	12	L	L
Be	Ap	L	2	14	L	L
Be	Ap	L	2	15	M	L
Be	Ap	L	2	20	M	L
Bea	Ap	L	2	14	L	L
Bec	Ap	L	4	26	M	M
Bec	Ap	L	4	32	M	M
Bec	Ap	VL	2	34	M	L
Bgg	Ap	L	2	7	L	L
Bk	Ap	L	4	13	L	M
Bk	Ap	L	4	31	M	M
Bvc	Ap	L	4	22	M	M
Bvc	Ap	L	4	31	M	M
Dd	Ap	L	3	13	L	L
De	Ap	L	2	12	L	L
Dg	Ap	L	2	7	L	L
Dg	Ap	L	3	10	L	L
E	Ap	M	2	13	L	M
Gm	Ap	M	4	25	M	M
Gm	Ap	M	4	40	M	M
Gm	Ap	M	4	41	M	M
Hc	Ap	M	5	63	H	H
Hh	Ap	M	2	11	L	M
Hh	Ap	M	3	14	L	M
Hl	Ap	M	2	12	L	M
Hl	Ap	M	2	12	L	M
Jc	Ap	M	4	24	M	M
Jc	Ap	M	5	32	M	H
Jcg	Ap	M	2	12	L	M
Jcg	Ap	M	4	27	M	M
Jcg	Ap	M	5	31	M	H
Jcg	Ap	M	5	42	M	H
Jcg	Ap	M	5	51	H	H
Jd	Ap	L	1	6	L	L
Jd	Ap	L	2	17	M	L
Je	Ap	L	2	30	M	L
Lc	Ap	L	3	13	L	L
Lf	Ap	L	4	28	M	M
Lg	Ap	L	3	18	M	L
Lgs	Ap	L	1	10	L	L
Lgs	Ap	L	3	14	L	L
Lgs	Ap	L	3	23	M	L
Lk	Ap	VL	3	29	M	L
Lkc	Ap	VL	3	7	L	L
Lkv	Ap	VL	3	14	L	L
Lo	Ap	L	3	6	L	L
Lo	Ap	L	3	6	L	L
Lo	Ap	L	3	6	L	L

Table 29 (following)

FAONAME	HORIZON	OC CLASS	T1	CEC	CEC CLASS	CEC TOP
Lo	Ap	L	3	7	L	L
Lo	Ap	L	3	8	L	L
Lo	Ap	L	3	8	L	L
Lo	Ap	L	3	8	L	L
Lo	Ap	L	3	8	L	L
Lo	Ap	L	3	9	L	L
Lo	Ap	L	3	9	L	L
Lo	Ap	L	3	10	L	L
Lo	Ap	L	2	11	L	L
Lo	Ap	L	3	11	L	L
Lo	Ap	L	3	12	L	L
Lo	Ap	L	3	14	L	L
Lo	Ap	L	2	14	L	L
Lo	Ap	VL	3	17	M	L
Lo	Ap	L	3	17	M	L
Lo	Ap	L	2	23	M	L
Lo	Ap	VL	3	24	M	L
Lo	Ap	L	3	25	M	L
Lo	Ap	L	2	25	M	L
Lo	Ap	L	2	26	M	L
Lo	Ap	L	2	26	M	L
Ph	Ap	M	1	9	L	L
Ph	Ap	M	1	9	L	L
Ph	Ap	M	1	16	M	L
Rc	Ap	L	2	20	M	L
Rc	Ap	L	2	22	M	L
Rc	Ap	L	2	23	M	L
Re	Ap	L	2	11	L	L
V	Ap	L	4	27	M	M
V	Ap	L	5	37	M	M
Vc	Ap	L	4	29	M	M
Vc	Ap	L	4	31	M	M
Vc	Ap	L	5	32	M	M
Vc	Ap	L	5	57	H	M
Vp	Ap	L	5	37	M	M
Vp	Ap	L	5	67	H	M
Vpc	Ap	L	5	53	H	M
Wd	Ap	VL	1	4	L	L
We	Ap	L	2	6	L	L
We	Ap	L	2	7	L	L
We	Ap	L	2	12	L	L
Xk	Ap	VL	5	37	M	M
Xy	Ap	VL	4	37	M	M

Number of samples: 95
Agreement: 64%
Occurrence CEC_CLASS<CEC_TOP 11%
Occurrence CEC_CLASS>CEC_TOP 25%

4.1.3 - Cation exchange capacity of subsoil

The rule for *CEC_SUB* was validated on a database of 189 soil materials overlying a C or R horizon of soils located throughout the EC. The results of the validation are presented in Table 30.

Table 30: Validation of the *CEC_SUB* pedotransfer rule on soils overlying a C or R horizon. Matching estimated data (CEC_SUB) with actual data (CLASS).

	occurrence (%)	
	CEC_SUB	CLASS
CEC class L	62	60
CEC class M	31	36
CEC class H	7	4
Agreement	80	
CEC_SUB < CLASS	10	
CEC_SUB > CLASS	10	

4.1.4 - Packing density

Packing density was validated with 64 soil units of soil series of England. Predicted and measured PD are compared for topsoil and subsoil. The performance is poor (Table 31) due to the high variability of this attribute and the effect of agriculture. For the topsoil, only 45% of estimated values agree with measured values. For subsoil the agreement reaches 55%. Nevertheless, disagreements mainly are less than one class. Only less than 8% of the samples is over two classes.

Table 31: Validation of pedotransfer rule for Packing Density

Soil series	EC map unit	STU	Humic/ Peaty top	EC texture class	Land Use	Predicted structure		Predicted Ld		Measured Ld	
						topsoil	subsoil	topsoil	subsoil	topsoil	subsoil
Fladbury	1002	Jeg	N	(1/)4	PG	Good	Normal	M	H	L	M
Wallasea	1003	Jeg	N	4	PG	Good	Normal	M	H	L	H
Wisbech	1011	Jcg	N	1	Arable	Normal	Normal	M	M	M	M
Newchurch	1010	Jcg	N	(2/)4	P.G	Good	Normal	M	H	M	H
Windsor	1016	Ges	N	4	PG	Good	Poor	M	H	M	H
Foggathorpe	1018	Ges	N	4	PG	Good	Poor	M	H	L	H
Denshworth	1019	Ges	N	4	PG	Good	Poor	M	H	M	H
Brickfield	1029	Gds	N	2(/4)	PG	Good	Poor	L	H	L	M
Blackwood	1030	Gm	Y	1	Arable	Good	Normal	L	M	L(-M)	M
Downholland	1031	Gmf	Y	(2/)4	Arable	Good	Normal	M	H	M	M
Wilcocks	1033	Gh	Y	2(/4)	Semi-natural	Hum/peaty	Normal	L	M	L	M
Sandwich	1038	Rc	N	1	Semi-natural	Good	Normal	L	M	L	M
Beckfoot	1044	Rd	N	1	Semi-natural	Good	Normal	L	M	L	M
Bangor	1053	Id	N	(1/)2	Semi-natural	Good	Normal	L	M	L	Rock
Cuckney	1058	Ql	N	1	Arable	Normal	Normal	M	M	M	M
Worlington	1059	Ql	N	1	Semi-natural	Good	Normal	L	M	L	M
Newport	1060	Ql	N	1	Arable	Normal	Normal	M	M	M	M
Andover	1066	Eo	N	(2/)3	Arable	Good	Good	M	L	L	Chalk
Skiddaw	1073	U	N	1? 2	Semi-natural	Good	Normal	L	M	L	Rock
Waltham	2022	Be		(1/)2	Arable	Normal	Normal	M	M	M	M
South Petherton	2025	Be	N	(1/)2	Arable	Normal	Normal	M	M	L	M
Wick	2027	Be	N	(1/)2	Arable	Normal	Normal	M	M	M	M
Swaffham Prior	2033	Bec	N	2	Arable	Normal	Normal	M	M	M(-L)	M
Aberford	2035	Bec	N	2	Arable	Normal	Normal	M	M	M	M

Table 31 (following)

Soil series	EC map unit	STU	Humic/ Peaty top	EC texture class	Land Use	Predicted structure		Predicted Ld		Measured Ld	
						topsoil	subsoil	topsoil	subsoil	topsoil	subsoil
Badsey	2036	Bec	N	2	Arable	Normal	Normal	M	M	M	M
Elmton	2034	Bec	N	2(/4)	Arable	Normal	Normal	M	M	M	Lst
Blacktoft	2040	Bef	N	2? 3	Arable	Normal	Normal	M	M	M	M
Denbigh	2058	Bd	N	(1/)2	PG	Good	Normal	L	M	L	L
Rivington	2061	Bd	N	1(/2)	PG	Good	Normal	L	M	L	L
Milford	2062	Bd	N	(1/)2	PG	Good	Normal	L	M	L	M(-L)
East Keswick	2065	Bd	N	(1/)2	PG	Good	Normal	L	M	L	M(-L)
Malham	2063	Bd	N	(2/)3	PG	Good	Normal	L	M	L	M(-L)
Winskill	2071	Bds	N	1? 2	Semi-natural	Good	Good	L	L	L	L
Moretonhampstead	2070	Bds	N	(1/)2	Semi-natural	Good	Good	L	L	L	L
Manod	2069	Bds	N	2	Semi-natural	Good	Good	L	L	L	L
Hanslope	2082	Bgc	N	4	Arable	Normal	Normal	M	H	M	H
Evesham	2083	Bgc	N	4	Arable	Normal	Normal	M	H	M	H
Nercwys	2084	Bqg	N	2	PG	Good	Poor	L	H	L	M(-H)
Curtisden	2085	Bgg	N	3	Arable	Normal	Poor	M	H	M	M
Banbury	2106	Bc	N	(1/)2	Arable	Normal	Normal	M	M	M	M
Malling	3027	Lo	N	(1/)2	Arable	Normal	Normal	M	M	M	M
Burlingham	3026	Lo	N	2	Arable	Normal	Normal	M	M	M(-H)	H
Ardington	3025	Lo	N	2	Arable	Normal	Normal	M	M	M	M
Sutton	3009	Lo	N	2(/3)	Arable	Normal	Normal	M	M	M	M
Whimple	3024	Lo	N	2(/4)	Arable	Normal	Normal	M	M	M	H
Bromyard	3023	Lo	N	3	P.G.*	Good	Normal	L	M	L	M(-H)
Hamble	3004	Lo	N	3	Arable	Normal	Normal	M	M	M	M(-H)
Fyfield	3039	Lc	N	1	Arable	Normal	Normal	M	M	M	M
Barrow	3037	Lc	N	(1/)2	Arable	Normal	Normal	M	M	H	H
Carstens	3036	Lc	N	(2/)3	Arable	Normal	Normal	M	M	L	H(-M)
Blackwood var	3055	Lg	N	1	Arable	Normal	Normal	M	M	L	M
Park Gate	3051	Lg	N	3	Arable	Normal	Normal	M	M	M	M
Salop	3058	Lgs	N	2	PG	Good	Poor	L	H	L	H
Ragdale	3059	Lgs	N	(2/)4	Arable	Normal	Poor	M	H	M	H
Beccles	3060	Lgs	N	2	Arable	Normal	Poor	M	H	M	H
Croft Pascoe	3061	Lgs	N	3	Semi-natural	Good	Poor	L	H	L	M
Shirrell Heath	3081	Po	Y	1	Semi-natural	Hum/peaty	Normal	L	M	L	L(-M)
Belmont	3092	Pp	Y	(1/)2	Semi-natural	Hum/peaty	Normal	L	M	L	L
Hafren	3094	Pp	Y	(1/)2	Semi-natural	Hum/peaty	Normal	L	M	L	L
Sollom	3096	Pg	Y	1	Arable	Hum/peaty	Normal	L	M	L	M
Holiday's Hill	3097	Pgs	Y	1(/2)	Semi-natural	Hum/peaty	Poor	L	H	L	M(-H)
Adventurer's	4001	Oe	Y	-	Arable	Hum/peaty	Peaty	L	L	L	L
Turbary Moor	4005	Od	Y	-	Semi-natural	Hum/peaty	Peaty	L	L	L	L
Winnter Hill	4007	Od	Y	-	Semi-natural	Hum/peaty	Peaty	L	L	L	L

4.2 - Mapping of specific parameters

Mapping parameters are used for calculating the water reserve in soil. They are: 1) attributes deriving directly from the database (e.g. topsoil textural class), 2) attributes estimated from pedotransfer rules (e.g. depth to rock), 3) attributes calculated from other attributes deriving from pedotransfer rules.

4.2.1 - Mapping method: creating a legend for SMUs

Each estimate of an attribute in space is saddled with the problem of its map representation, which should provide a documents for helping in decision making that is easier to use than a simple table with numbers. The data that directly derive from the model are complex and cannot be shown directly on a map. A choice thus has to be made for simplifying the information and providing a document that is as close as possible to reality.

As most Soil Mapping Units (SMU) contain several Soil Typological Units (STU), it may be impossible to show multiple data on the map. A simple solution to this problem is to select for each SMU the so-called dominant STU, representing the greatest surface percentage. Certain SMUs are very complex and contain up to 11 typological units. In this case, the dominant unit may represent only a small percentage of the association, thus leading to errors when reading the map (Daroussin *et al.*, 1990; Ngongo *et al.*, 1992). Furthermore, certain mapping units can have highly diverging results for their STUs. Faced with such spatial variability, it is preferable not to give the results of dominant units at the publication scale of 1:1,000,000.

Instead, it is proposed to assign a colour to each of the classes of an attribute. The classification operation can be done on the level of STUs, and it will be easy to calculate the surface percentage taken up by each class. The selection of the dominant class will enable a much more precise representation of reality than the choice of a dominant STU.

Only a small percentage of the European Community surface corresponds to pure SMUs. For that reason, we propose to produce a set of maps that show confidence level as well as maps showing the "purity" concerning a specific subject. For the subject of water reserves, we adopted the graphic technique developed by Mori (1982), which affects a colour to each class. In this way, each mapping unit is assigned a colour that corresponds to its dominant class. If an SMU contains two classes of water reserve and the associated class has a lower value than the dominating class, the colour will be attenuated; in the opposite case, with an associated class that has a higher value, the colour will be more intense. In case three classes are present in a single SMU, a specific colour will be attributed, annulling the colour earlier attributed to the dominant class. In this case, associated classes below a surface threshold fixed at 30% are ignored. The proposed method makes it possible to show on a single document the estimated water reserves that are most represented in an area, as well as the type of reserves in the secondary units. In the case of strong spatial variability, the data are completely masked so as to avoid a wrong interpretation.

4.2.2 - Map of soil textural classes

The map of the dominant texture in each SMU is presented, as well as a map of the surface percentage covered by this texture within an SMU.

4.2.3 - Map of depth to rock classes

The depth-to-rock map is based on a pedotransfer rule that uses the attributes "FAO Soil name", "Parent Material" and "Phase" (cf. 3.4.1). A confidence-level map of the estimated new attribute is provided for all users of the map, as this is essential for the analysis and use of the results from pedotransfer rules.

4.2.4 - Map of the available soil water for plants

This attribute was chosen for its role in certain environmental problems, in particular those concerning the protection of water resources. In addition, earlier work carried out as part of the DG-VI MARS project, gave very rapid access to the data and rules for this attribute.

EUROPEAN UNION SOILS DATABASE

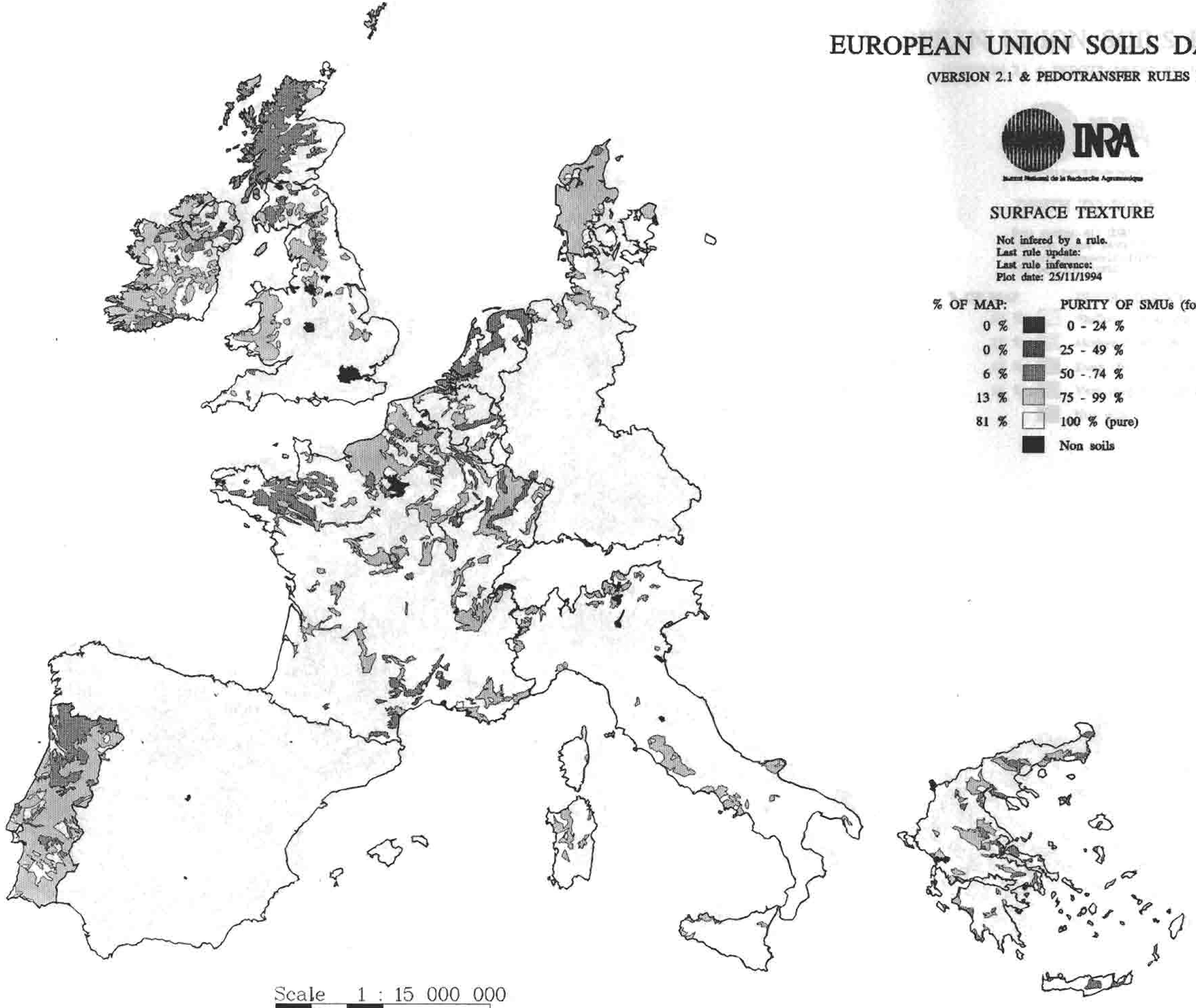
(VERSION 2.1 & PEDOTRANSFER RULES 1.0)

INRA

Institut National de la Recherche Agronomique

SURFACE TEXTURE

Not infered by a rule.
Last rule update:
Last rule inference:
Plot date: 25/11/1994



% OF MAP:		PURITY OF SMUs (for attribute TEXT):
0 %	■	0 - 24 %
0 %	▩	25 - 49 %
6 %	▨	50 - 74 %
13 %	▒	75 - 99 %
81 %	☐	100 % (pure)
	■	Non soils

Scale 1 : 15 000 000

EUROPEAN UNION SOILS DATABASE

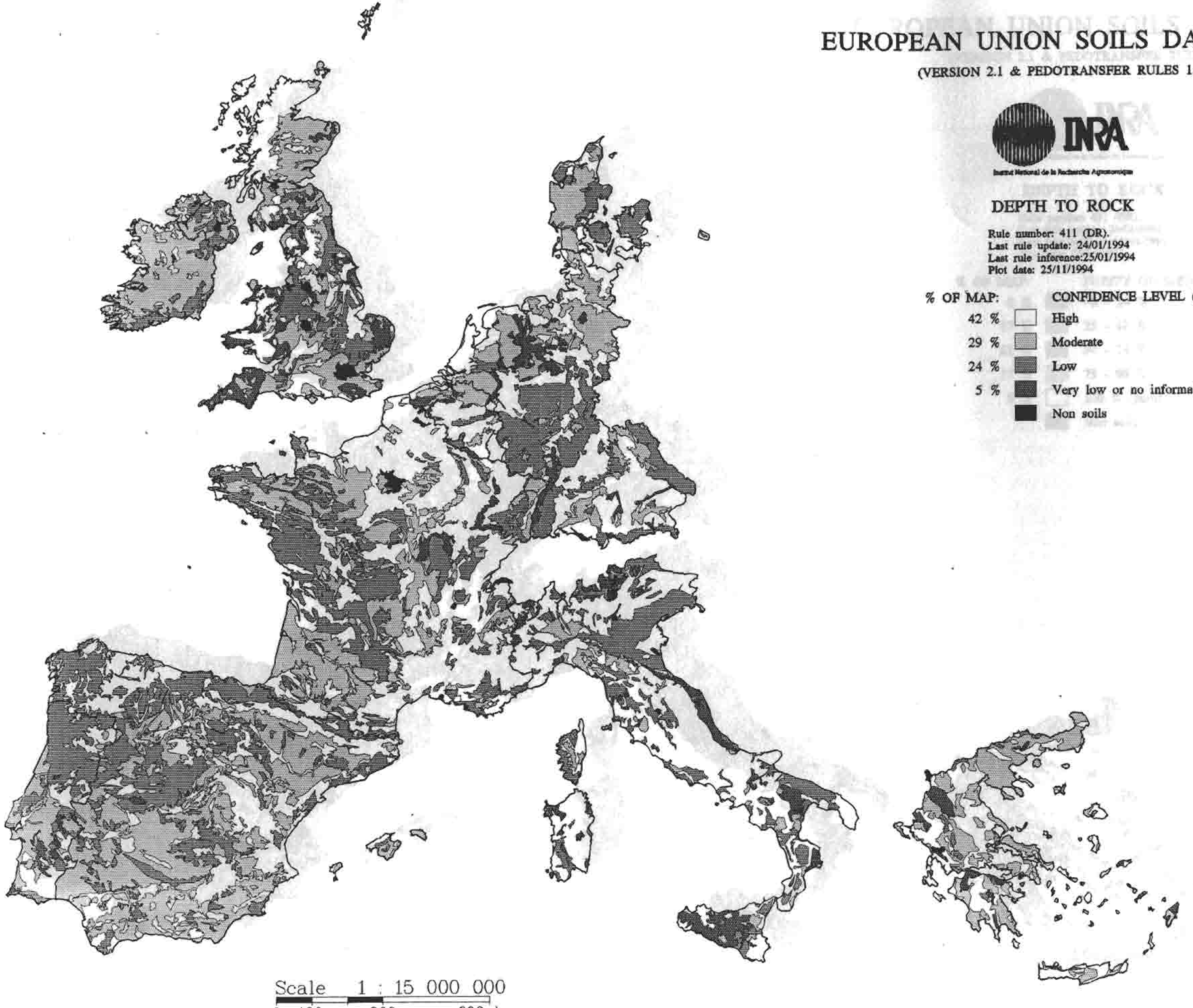
(VERSION 2.1 & PEDOTRANSFER RULES 1.0)

INRA
Institut National de la Recherche Agronomique

DEPTH TO ROCK

Rule number: 411 (DR).
Last rule update: 24/01/1994
Last rule inference:25/01/1994
Plot date: 25/11/1994

% OF MAP:		CONFIDENCE LEVEL (for attribute DR):
42 %	☐	High
29 %	▨	Moderate
24 %	▨	Low
5 %	▨	Very low or no information
	■	Non soils



EUROPEAN UNION SOILS DATABASE

(VERSION 2.1 & PEDOTRANSFER RULES 1.0)

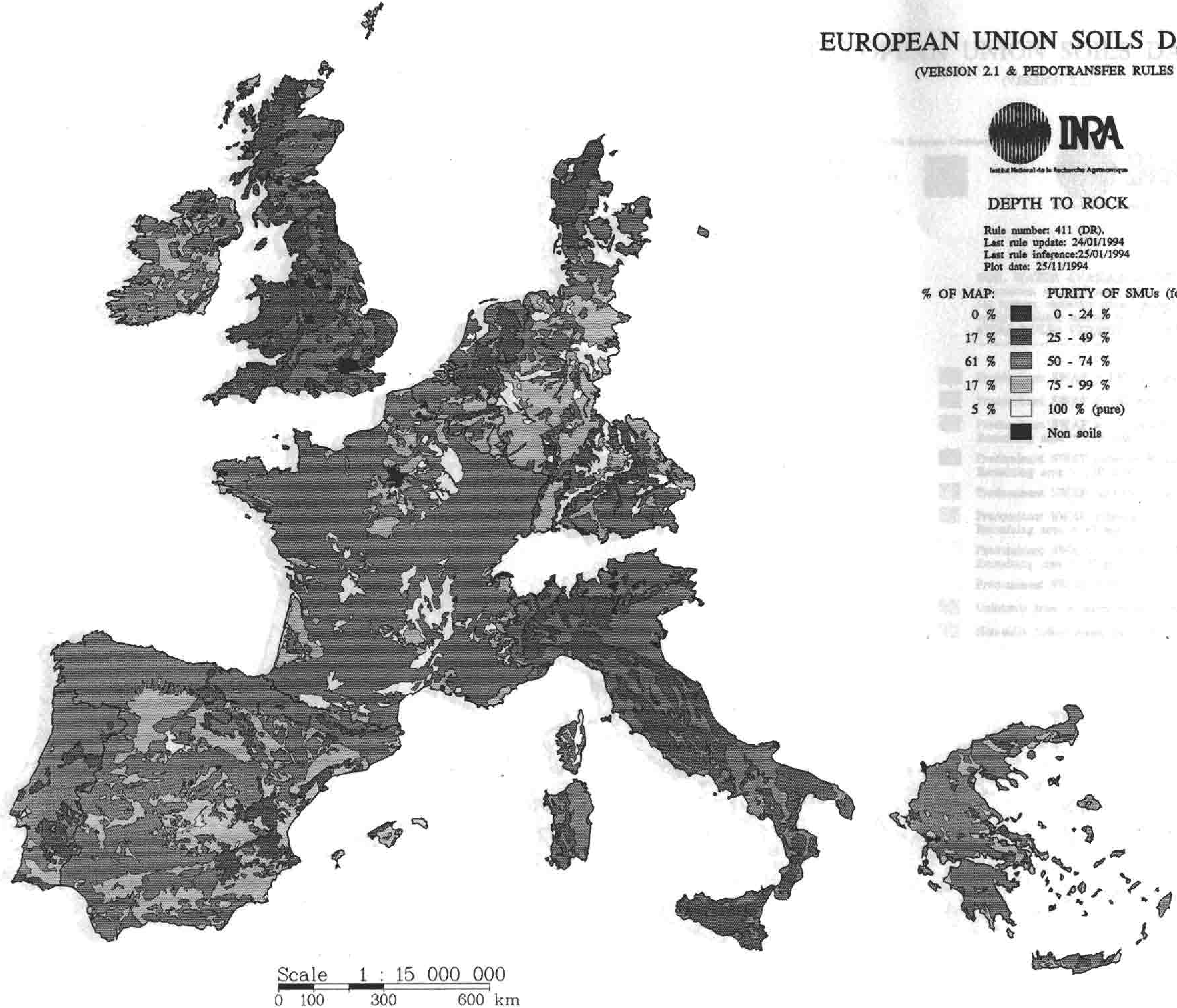
INRA

Institut National de la Recherche Agronomique

DEPTH TO ROCK

Rule number: 411 (DR).
Last rule update: 24/01/1994
Last rule inference:25/01/1994
Plot date: 25/11/1994

% OF MAP:		PURITY OF SMUs (for attribute DR):
0 %	■	0 - 24 %
17 %	■	25 - 49 %
61 %	▩	50 - 74 %
17 %	▨	75 - 99 %
5 %	☐	100 % (pure)
	■	Non soils



kaléidos

4.2.4.1 - Estimating soil-water reserves

Soil is a reservoir that absorbs water during rainy periods and can restitute it during dry periods. The water reserve available to plants classically is defined as the difference between the quantity of water in soil at field capacity (48 hours of drainage after a rainfall) and the quantity of water unavailable to plants (wilting point).

Both values can be measured in the field, but they can vary considerably as a result of environmental conditions (past rainfall, agricultural work, neighbourhood effects, etc.). For that reason, it is more reliable to use to use laboratory measurements of humidity values, which correspond to the application of two pressure-threshold values to soil samples that, if possible, should be undisturbed. Such threshold values can vary strongly depending on the authors, but we have taken -1500 kPa for the wilting point, and -5 kPa for the pressure simulating field capacity. We call this last value, measured in the laboratory, the retention capacity.

The method described above enables estimation of the water volume available for a soil sample. To evaluate the total water reserve of a soil, we sum the values obtained for each level or layer that was sampled, from the surface to a depth that corresponds to the lowest rooting level. The hydrous profiles measured in the field show that water uptake by plants decreases with increasing depth, in parallel with the reduction in the number of roots. To take account of this phenomenon, it is possible to limit the quantity of water really available at this depth to a threshold that lies below the wilting point. We selected a level of -200 kPa and called this the "easily available reserve".

Rapid inspection of the database shows the absence of data on the hydric properties of soil horizons as well as of information on general humidity values for the solum. Direct calculation thus is impossible because of a lack of data. To solve this problem, we propose an approach in two steps. First, a simplified model for estimating the water reserve in soil is drawn up, and then, using pedotransfer rules and the geographic database, the values of the parameters implicated in this model are estimated.

4.2.4.2 - Proposal for a crop-adjusted available water capacity model

Earlier work has attempted to limit the number of laboratory measurements necessary for determining humidity values, and tried to establish links between certain soil parameters and humidity values. We decided to start our work on a soil horizon considered as homogeneous, and then to work on an evaluation model for the entire solum.

* **To estimate a homogeneous horizon**, the standard way to write the equation for evaluating the available water in this horizon is:

$$Wh = E * (Wcr - Wpf) \qquad (1)$$

where:
Wh = available water reserve in the horizon
E = thickness of the horizon
Wcr = water volume at the retention capacity, e.g. -5 kPa
Wpf = water volume at the wilting point, e.g. -1500 kPa

Several authors proposed to link the water quantity retained in a soil sample at a given pressure, with the physico-chemical properties of this soil. The most discriminant characteristics are: grainsize distribution within the soil, its structure, and its organic-matter content. Mineralogy of the constituent clays also plays a role in the type of organization of the material and thus on its porosity, but we have ignored the last phenomenon in this study.

Grainsize distribution, or texture-class limit if a texture triangle is used, varies considerably between countries. A compilation was made on the basis of the five main texture classes of the FAO triangle, adapted to the EEC. We used the data from the English handbook, which contained the largest number of references (3600 measurements for 1000 profiles) (Hall *et al.*, 1977). Such references are easily changed in the list of rules, if another calculation with other references is made (cf. 3.5.5).

Structure and organic-matter content are two properties that influence soil porosity, thus determining water quantity and field capacity. Both characteristics commonly show great lateral variability, even within a single agricultural plot. For that reason, they are difficult to consider at the scale of 1:1,000,000, except for soil representing extreme cases, such as histosols and andosols. However, it is possible to take account of the vertical variation that generally is observed in most cultivated-soil profiles. We have retained several structural-quality classes, not only by distinguishing worked layers from the undisturbed soil below, but also by distinguishing three classes of structural development for the deeper layers.

The relationship between water present in a horizon at a given pressure and the characteristics of this horizon, can then be written as:

$$Wp = fp[Tx,St] \qquad (2)$$

where:

Wp = water quantity in a horizon at pressure p
Tx = texture class
St = quality of the structure
fp = pedotransfer function for pressure p

Equation [1] then is written as:

$$Wh = E * (fcr[Tx,St]-fpf[Tx,St]) \qquad (3)$$

A direct connection can be made between available water quantity and physical characteristics of the soil. The values obtained from the work by Hall *et al.* (1977) are shown in Table 27. The equation now is simplified to:

$$Wh = E * (fw[Tx,St]) \qquad (4)$$

where fw is the pedotransfer function that makes it possible to estimate the quantity of water available in a horizon, based on its texture (*TS*) and structure (*STR*).

*** To estimate a complete solum** is relatively easy in principle, once the data for horizons are available, as it suffices to add up the values found for each horizon in the root zone, limiting the sampling possibilities with depth. The geographic database does not contain precise data on the horizons present for each soil type. The solum is thus reconstituted into three layers that correspond respectively to: 1) a worked superficial layer; 2) a subsurface layer from which water is easily extracted by roots; 3) a deep layer from which it is difficult to extract water. For each layer, data from the database help to evaluate the textural properties, the organic-matter content and the structural conditions, thus enabling the application of equation [4] to each layer. Depending on the depth of the profile or that of the root system in question, one of the three layers may be truncated or absent. Without considering this depth-related limitation, the quantity of available water can be evaluated with the following equation:

$$Q = \Sigma i=1,3((Ei*Wi)) \quad (5)$$

where:

Q = quantity of water available for the solum as a whole
i = number of the layer
Ei = thickness of the layer
Wi = quantity of water available for layer i (or easily AW if $i = 3$).

The last parameter to be evaluated remains the depth exploited by roots. Soil depth can be contained either in a fragmentary manner in the soil name, e.g. lithosol has by definition a depth of less than 10 cm, or in the phase, e.g. the stony phase indicates a decrease in the total pore volume of the soil. Using data associated with parent material, this evaluation of total depth is further refined. Rooting depth will be calculated as being the minimum between this total-depth value (estimated with the rules) and the potential crop depth (data from field observations and corresponding to the genetic properties of the plant). This equation can be written as:

$$Pa = Min(Ps,Pr) \qquad (6)$$

where:

Pa = depth at which water is accessible to the plant
Ps = depth of soil, or of an obstacle to rooting
Pr = potential maximum depth of the root system.

Equations [4], [5], and [6] can be combined into a simplified model for estimating water reserve in a soil:

$$R = \Sigma i=1,min(Ps,Pr)((Ei*fi[Txi,Sti]) \quad (7))$$

where the same symbols were used as in equations [1] to [6].

The parameters of this equation (thickness, texture, etc.) are either directly obtained from the database, or are estimated with the help of pedotransfer rules.

4.2.4.3 - Pedotransfer rules for estimating parameters for the model

The equations above were used according to the rules laid down in Chapter 3. The final calculation for the solum is done independently of such rules. The parameter of maximum rooting depth was fixed at 150 cm for the map produced, but obviously can be changed according to the type of crop. The same is true for the depth at which the plant can no longer extract all the water from a soil, here chosen as 50 cm. After the last calculation, we obtained a value for each STU. Colours were attributed according to the method described in 4.2.1.

5 - CONCLUSIONS

The soil map of the European Communities and its associated geographic database, represent a knowledge potential that is based on many years of map-data collection and compilation in Europe. Such data have already been used in applications related to agriculture and the environment, thus showing the interest and importance of such knowledge as well as its limits. The main limitation is the difficulty of obtaining accurate data on soil parameters needed for environmental studies, when based only on synthetic attributes such as the soil name according to the FAO classification used.

The **objective** of our work was to propose an **automatic interpretation** of the data in the geographic database, leading to estimates for environmental use that are as reliable as possible. This meant that it was necessary to formalize as rules, the interpretations made instinctively by a well-versed reader when faced with a soil map.

Such rules are called **pedotransfer rules**, by analogy with pedotransfer functions that link the standard soil-analysis characteristics to more complex properties, such as hydrodynamic properties. The rules were formalized in a standard format, facilitating their handling and use. They were created by expert judgement based on a general knowledge of Soil Sciences. They can be associated to a region, in specific cases that are not mentioned as part of the initial EC soil map.

The results provided from such rules are only qualitative estimates. At the 1:1,000,000 scale it is difficult to provide precise information from the few data contained in a soil map, and care was taken to indicate the methodological limitations of our approach. This was done by attributing a confidence level to each rule, which can plot those areas for which the data are not very reliable. In addition, an indicator was created for internal spatial variability within Soil Mapping Units (SMU), indicating the surface percentage of Soil Typological Units (STU) that were used for the interpretation within an SMU. Tests of mapping the estimated attributes and certain other validations then enable a quantitative judgement of estimation quality.

The rules were drawn up from the attributes available in Version 2.0 of the EC geographic database. Version 3.1, planned for 1994, will provide access to increasingly plentiful and accurate data. At that stage, it will be necessary to modify some of the rules, to make them compatible with the new version. The statistical and cartographic validation of values estimated from the rules will continue under the new version. This will be helped by integration of the profile database that is currently being created.

The presently ongoing R&D work on the drawing up of more precise pedotransfer rules should lead to modification and even replacement of some of the rules proposed in this report. In fact, such rules cannot be considered as a final product, but precisely as a means for highlighting attributes and/or regions for which data are lacking and where a special effort has to be made to collect the missing information in the future.

REFERENCES

ADDISCOTT T.M., 1970 - A note on resolving soil cation exchange capacity into "mineral" and "organic" fractions. *J. Agric. Sci.*, 75: 365-367.

BATJES, N.H., 1990 - Macro-scale land evaluation using the 1:1 M world soils and terrain digital database. Working paper and preprint 90/9. *ISRIC*. 35p.

BOORMAN D., HOLLIS J.M., 1990 - Hydrology of soil types : A hydrologically-based classification of the soil in England and Wales. *Proceedings of the MAFF* Conference of River and Coastal Engineers. Loughborough University, 10-12 July, 1990.

BOUMA J., VAN LANEN H.A.J., 1986 - Transfer functions and threshold values : from soil characterics to land qualities. p. 106-110. In *Proceedings of the International Workshop on Quantified Land Evaluation Procedures*. 27/04-2/05/1986, Washington D.C..

BRISSON N., KING D., NICOULLAUD N., RUGET F., 1992 - A crop model for land suitability evaluation : a case study of the maize crop in France. *Eur. J. Agron.*, 1(3), 163-175.

CEC, 1985 - Soil map of the European Communities at 1:1 000 000. *CEC-DGVI*. Brussels-Luxembourg, 124p.

CLAYTON J.L., KENNEDY D.A., NAGEL T., 1991 - Soil response to acid deposition, Wind River Mountains, Wyoming: II. Column leaching studies. *Soil Sci. Soc. Am. J.* 55: 1433-1439.

DAROUSSIN, J., HARDY R., KING D., BORNAND M., 1990 - Organisation d'une banque cartographique en Science du Sol. In : *2ième Journées Nationales de l'Etude du Sol,* Orléans, 19-21/11/90.

FAO, 1975 - Soil map of the world at 1:5 000 000. Volume I. Europe. *UNESCO*, Paris, 62p.

GENON J.G., DUFEY J.E., 1991 - Contribution des argiles et de la matière organique à la capacité d'échange cationique des sols du sud-est de la Belgique. *Revue de l'Agriculture,* 44: 277-284.

HALL, D.G.M., REEVE M.J., THOMASSON A.J., WRIGHT V.F., 1977 -,Water retention porosity and density of field soils. *Soil Surv. Tech. Monogr.* 9, Harpenden. UK.

HOLLIS J.M., 1990 - Assessments of the vulnerability of aquifers and surface waters to contamination by pesticides. *SSLRC Research Report for MAFF project* a (v), 1989-1990. pp. 24.

HOLLIS J.M., 1991a - Mapping the vulnerability of aquifers and surface waters to pesticide contamination at the national/regional scale. In: Pesticides in Soils and water: *Current perspectives* (ed. A. Walker). BCPC monograph N° 47: 165-171.

HOLLIS J.M., 1991b - Proposed classification of the type of soil cover for a groundwater protection policy. *SSLRC Research Report for the National Rivers Authority*. Contract N° 82/2991. pp. 8.

INRA, 1990 - Elaboration d'une base de données géographiques des pédopaysages des Communautés Européennes. Final report of the CEC contract 3934-90-03 ED ISP F, JNES 1991, Orleans.

INRA-JRC, 1993 - Users guide for the elaboration of the EC soil database 3.1 version. INR Orleans, 10p.

JONES R.J.A., THOMASSON A.J., 1993 - Effects of soil-climate system interactions on th sustainability of land use : A European perspective. In : J.M. Kimble (Ed.), *Proceeedings of th International Soil Management Workshop*. Utilization of soil survey for sustainable land use. Ma 1993. USDA Soil Conservation Service, National Soil Survey Center, 39-52.

JONHSON D.W., TODD D.E., 1983 - Relationships among iron, aluminium, carbon and sulfate in variety of forest soils. *Soil Sci. Soc. Am.* J. 47: 792-800.

KING, D., LE BISSONNAIS Y., HARDY R., 1993 - Regional assessment of runoff and erosio risks. Example of the North/Pas de Calais region in France. In : *Farm lands erosion in temperat plans environments and hills*. Wichereck (Ed), Elsevier, 191-205.

KING D., DAROUSSIN J., TAVERNIER R., 1994 - Development of a soil geographical databas from the soil map of the European Communities, *Catena*, 21, 37-56.

LIVENS F.R., LOVELAND P.J., 1988 - The influence of soil properties on the environmenta mobility of Caesium. *Soil Use and Management*, Vol 4, N°3: 69-75.

LOVELAND P.J., 1990 - A soil buffering capacity map for England and Wales. SSLRC *Research report for department of the Environment*. Contract N° 82/2485. pp. 12.

MORI, A., 1982 - Notices explicatives de la carte des sols de France affectés par l'hydromorphie et de la carte des réserves en eau utile des sols de France. *INRA-SESCPF*, Paris, 54p.

MSANYA B.M., LANGOME R., LOPULISA C., 1987 - Testing and improvement of a questionnaire to users of soil maps. *Soil Survey and Land Evaluation* 7, 33-42.

NGONGO, L., KING D., NICOULLAUD B., BRISSON N., RUGET F., 1993 - The accuracy of methods for the selection of soil data in estimating regional moisture deficit from soil maps. *Pedologie*, XLIII-2, 285-299.

PALMER R.C., 1993 - 1:625,000 Map of the risk of soil erosion in England and Wales by water on land under winter cereal cropping. Soil Survey and Land Research Centre, Silsoe UK.

PLATOU S.W., NOOR A.M., MADSEN H.B., 1989 - Digitizing of the EC soil map. In : R.J.A. Jones and B. Biagi (eds). Computerization of land use data. CEC, Brussels, 12-24.

TAN K.H., DOWLING P.S., 1984 - Effect of organic matter on CEC due to permanent and variable charges in selected temperate region soils. *Geoderma*, 32: 89-101.

THOMASSON A.J., 1975 - Soils and field drainage. *Soil Survey Tecnical Monograph*, N° 7. Harperden.

WRIGHT W.R., JONES J.E., 1972 - Contributions of clay and organic matter to the cation exchange capacity of Maryland soils. *Soil Sci Soc. Am. Proc.*, 36: 115-118.