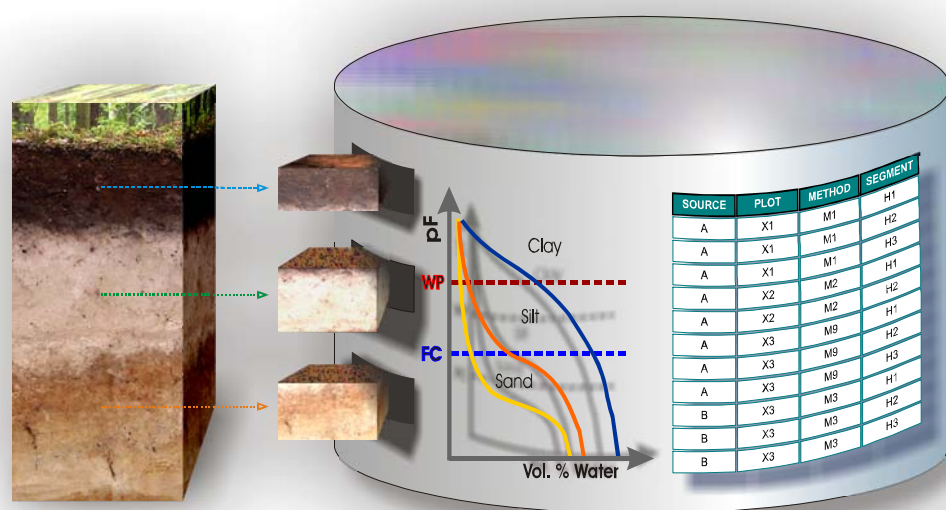


Extending Geographic and Thematic Range of SPADE/M with HYPRES Soil Profile Data

Roland Hiederer



EUR 24971 EN - 2011

The mission of the JRC-IES is to provide scientific-technical support to the European Union's policies for the protection and sustainable development of the European and global environment.

European Commission
Joint Research Centre
Institute for Environment and Sustainability

Contact information

R. Hiederer
European Commission Joint Research Centre
Institute for Environment and Sustainability
Via Enrico Fermi, 2749 - 21027 - Ispra (VA) – Italy
E-mail: roland.hiederer@jrc.ec.europa.eu

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

Legal Notice

Neither the European Commission nor any person acting on behalf of the Commission is responsible for the use which might be made of this publication.

***Europe Direct is a service to help you find answers
to your questions about the European Union***

Freephone number (*):

00 800 6 7 8 9 10 11

(*) Certain mobile telephone operators do not allow access to 00 800 numbers or these calls may be billed.

A great deal of additional information on the European Union is available on the Internet. It can be accessed through the Europa server <http://europa.eu/>

JRC 67188

EUR 24971 EN

ISBN 978-92-79-21592-6 (print)

ISBN 978-92-79-21593-3 (PDF)

ISSN 1018-5593 (print)

ISSN 1831-9424 (online)

doi: 10.2788/8209

Publications Office of the European Union

© European Union, 2011

Reproduction is authorised provided the source is acknowledged.

Printed in Italy

This document may be cited as follows:

Hiederer, R. (2011) Extending Geographic and Thematic Range of SPADE/M with HYPRES Soil Profile Data. EUR 24971 EN. Publications Office of the European Union. 43pp.

European Commission Joint Research Centre
Institute for Environment and Sustainability
TP 261
21027 Ispra (VA)
Italy

Table of Contents

	Page
1 INTRODUCTION	7
2 DATABASE OVERVIEW	9
2.1 DATABASE ORGANIZATION	9
2.2 TABLES AND RELATIONSHIPS	10
2.2.1 Table: BASICDATA	11
2.2.2 Table: SOIL_PROPS	11
2.2.3 Table: HYDRAULIC_PROPS	11
2.2.4 Table: RAWPSD	11
2.2.5 Tables: RAWRET and RAWK	11
2.2.6 Tables: LOCALNGR_TYPE and EUROSTANDARD	12
2.3 ATTRIBUTES	13
3 ADAPTATIONS OF HYPRES DATA	17
3.1 CONVERSION OF LIST DATA IN BASICDATA TABLE	17
3.1.1 Sample Date (SAMPLE_DATE)	17
3.1.2 Data Source (SOURCE)	17
3.1.3 Plot Country (COUNTRY)	18
3.1.4 Local Name for Sample Site Area (LOC_NAME)	19
3.1.5 Site Coordinate Projection (PROJ)	19
3.1.6 Soil Name as Given by Source (SOIL) and According to FAO (FAO)	19
3.1.7 Depth of Groundwater Level (GWL_NM, GWL_HI and GWL_LO)	20
3.1.8 Dominant Land Use (LU)	21
3.1.9 Dominant Parent Material (PM)	21
3.1.10 Depth to Obstruction for Rooting (D_ROO_X, D_ROC_X and D_OTH_X)	22
3.1.11 Origin of Plot Observations (ORIGIN)	22
3.1.12 Comment on Plot Observations (COM_PL)	22
3.2 CONVERSION OF LIST DATA IN SOIL_PROPS TABLE	23
3.2.1 Horizon Name as given by Source (HOR_NAME)	23
3.2.2 Percentage of Stones and Gravel (GRAVEL)	23
3.2.3 Structure Class (STRUCT)	23
3.2.4 Sodium Adsorption Ratio < 4 (AR_NA_X)	24
3.3 CONTINUOUS NUMERIC DATA	24
3.3.1 Conversion to Geographic Coordinates (COOR_X, COOR_Y)	24
3.3.2 Variations in Plot Locations	28

3.3.3	<i>Plot Altitude (ALT)</i>	30
3.3.4	<i>Normal Groundwater Level (GWL_N_M)</i>	31
3.3.5	<i>Rooting Depth and Depth to Obstructions to Rooting (D_ROO, D_ROC and D_OTHOBS)</i>	31
3.3.6	<i>Upper and Lower Horizon Limits (HOR_START, HOR_END)</i>	31
3.3.7	<i>Particle Content (CLAY, SILT_1/2 and SAND_1/2/3)</i>	32
3.3.8	<i>Organic Matter and Carbon (ORG_MAT and ORG_C)</i>	35
3.3.9	<i>Soil Chemical Properties</i>	36
3.3.10	<i>Soil Water Retention (WC_1, WC_2, WC_3, WC_4, WC_FC)</i>	36
3.3.11	<i>Total Porosity (POR_TOT)</i>	37
3.3.12	<i>Bulk Density (BD)</i>	37
4	SOIL PROFILE SUITABILITY FOR SPADE/M DATABASE	39
4.1	CONCEPTUAL LIMITATIONS TO EXTENDING PROFILE DATA SET	39
4.2	ENHANCEMENTS OF SPADE/M BY HYPRES	40
4.3	GEOGRAPHIC COVERAGE OF PROFILES	40
4.4	RANGE OF FAO SOIL TYPES	43
5	CONCLUSIONS	47

List of Figures

	Page
Figure 1: Organization of HYPRES Database Used in Study	10
Figure 2: Distance between Plots from EUROSTANDARD Coordinates to Re-calculated Locations	28
Figure 3: Relative Frequency Distribution of Organic Matter Content in Horizon Records	35
Figure 4: Relative Frequency Distribution of Bulk Density in Horizon Records.....	38
Figure 5: Geographic Distribution of SPADE/M and Supplementary HYPRES Profiles.....	41
Figure 6: Plot Density for Combined SPADE/M and HYPRES Profile Sites on 1 km Grid.....	43
Figure 7: Relative Distribution of FAO85 Level 1 Soil Types for SGDBE (Europe), SPADE/M and HYPRES Profile Data	44
Figure 8: Relative Distribution of FAO85 Level 3 Soil Types for SGDBE (Europe), SPADE/M and HYPRES Profile Data	45
Figure 9: Diffusion of SPADE/M and HYPRES Profiles for Dystric Cambisol and Eurtric Fluvisol on Distribution of the Soil Types in SGDBE	46

List of Tables

	Page
Table 1: Site Data	13
Table 2: Profile Horizon Data.....	14
Table 3: Countries Referenced in BASICDATA	18
Table 4: Entries in FAO_SOIL Field without Correspondence in FAO Classification	20
Table 5: Summary of Plots with Geographic Location	30
Table 6: Occurrence of Particle Sizes in HYPRES RAWPSD Table	33
Table 7: Distance to Nearest Profile for SPADE/M, HYPRES and Combined Data Sets	42

List of Acronyms

ACRONYM	TEXT
CLC	Corine Land Cover
DBMS	Database management system
ESDB	European Soil Database
DHDN	Deutsches Hauptdreiecksnetz
ETRS89	European Terrestrial Reference System 1989
FAO	Food and Agriculture Organization of the United Nations
GISCO	Geographical Information System at the COMmission
HYPRES	Hydraulic Properties of European Soils
ID	Identifier
LAEA	Lambert Azimuth Equal Area projection
LAT/LONG	Latitude and longitude in geographic coordinates
NTF	Nouvelle Triangulation de la France
SOC	Soil organic carbon
SMU	Soil mapping unit
SOM	Soil organic matter
PTF	Pedo-transfer function
PTR	Pedo-transfer rule
RDBMS	Relational database management system
SGDBE	Soil Geographic Database of Eurasia
SPADE/M	Soil Profile Analytical Database of Europe of Measured parameters
UNSODA	Database of unsaturated soil hydraulic properties
WGS84	World Geodetic System Datum 84
WRB	World Reference Base for Soil Resources

Explanatory Notes:

The naming convention for the locations of the data on the ground is, in order of detail: site, plot, profile and horizon. A site corresponds to an area which is uniform with respect to the general characteristics. A plot is a sub-area within a site with a geometric form, such as a square or circle, and limited extent. Soil profiles are sampled within the extent of a plot at specific points. Horizons are the pedological horizon within a sampled profile and arranged in vertical sequence of depth.

Depending on the sample procedure and the method used to store the data there may be no distinction between some of the location attributes. Where only one profile has been sampled the geographic position is the same for the site, plot and profile. Site and plot may, therefore, be indistinguishable and, although representing an area, are located by a point. Where several profiles are sampled within a plot the data can be merged to represent the plot characteristics. A particular situation can arise when profiles are sampled along a transect. In case all samples remain within the area characterized for the site / plot the data can be merged. Where the site / plot characteristics differ along the transect the profile samples belong to different they may have been sampled in close proximity.

In practice the physical site / plot characteristics may be repeated for the profile sampled in close proximity and the only difference is the plot name. The approach of assigning a plot to each profile changes not only the character of the plot geographic unit from an area to a point, but also defines the profile as a distinct instance, which makes it difficult to identify the profile as belonging to a series of samples which characterize the plot area. In SPADE/M a site should only contain a single plot and a plot should only be linked to a single soil profile. Where several measurements are taken for a profile or horizon only the aggregated values are used and reported in the method on data aggregation. The profile data are sufficiently widely spaced to avoid auto-correlation and allow treating the data as independent observations.

1 INTRODUCTION

The database of *Hydraulic Properties of European Soils* (HYPRES)¹ contains detailed measured properties of soil profiles for a large range of soil types located in 12 European countries (Wösten, *et al.*, 1999; Wösten *et al.*, 1998). The data from the *Soil Profile Analytical Database of Europe of measured profiles* (SPADE/M) of the *European Soil Database* (ESDB) is intended to extend the qualitative data of the *Soil Geographic Database of Eurasia* (SGDBE) with quantitative values to improve the modelling of soil properties. Adding soil profile data from HYPRES to the SPADE/M database has been undertaken with the aim of broadening the basis of measured soil properties available from a single database with harmonized data. The profiles are intended to support the definition of pedo-transfer rules (PTRs) and function (PTFs) to extend the range of parameters characterizing soils and the validation of model runs, in particular the results from spatial interpolations.

The HYPRES and SPADE/M data sets were assembled and used to define generalized rules and functions to estimate specific physical soil properties from more widely available soil characteristics. With respect to these types of properties estimated from the PTRs and PTFs the data sets differ notably. The SPADE/M profiles were intended to delineate the conditions for PTRs on common soil properties, which supplement the range of soil properties stored in the SGDBE of the ESDB (European Commission Joint Research Centre, 2003). The data derived from the 39 PTRs of the pedo-transfer rule database are stored in the attribute table of the SGDBE. The HYPRES database is listed as one of the databases of the European Soil Database, but was subsequently not included in the version distributed. The profiles of the database were used to specify the parameters of PTFs to characterize soil hydrological properties from more easily measured soil data, such as soil texture (sand, silt, and clay) and organic matter content. The database contains raw and standardized data where the standardization of hydraulic data was achieved by fitting the Mualem-van Genuchten model parameters to the individual $\theta(h)$ and $K(h)$ hydraulic properties.

Given the differences in the intended use of the soil profile data assembled in the two data sets the characterization of the soil profiles also differ with respect to the properties recorded, the measurement units and the database model. To increase the number of measured soil profile data of the SPADE/M data set with HYPRES profiles the data from the latter has to be standardized to be adjusted to adhere to the specifications of the SPADE/M data set. The standardization process involves conversions of units or reference systems, such as the plot co-ordinate transformation or the extraction of properties from comment fields, but also conceptual adaptations of the method used to characterize a soil horizon in the database. The outcome of the standardization process is a series of soil profiles which can be seamlessly added to the SPADE/M data set.

¹ DG XII, Human Capital and Mobility ~ CHRX-CT94-0639

2 DATABASE OVERVIEW

In the naming convention of aspects of a relational database and the management system, the document follows the standard used by the Paradox^{®2} *Relational Database Management System* (RDBMS). Data are arranged in form of tables with records as rows and fields as columns. For each field a format type is specified. Access to tables is governed by an index. The primary index is also referred to as the table key and can be a composite index. Relationships between tables are defined by links. Joins are created when querying data, not when defining relationships. The naming convention adopted and the format types of Paradox[®] vary to a small degree from those used by other RDBMSs, but should be readily accessible. For describing the relational data model a simple schematic representation is used and any reference to entities, domains or definitions specific to a particular RDBMS are avoided.

2.1 Database Organization

The SPADE/M database was presented in detail in Hiederer (2010) and Hiederer *et al.* (2006). The HYPRES database³ is described in several articles and in detail in the final project report (Wösten, *et al.*, 1998). The HYPRES data used in the project were provided by Allan Lilly, James Hutton Institute, Aberdeen in April, 2011. The original HYPRES data are stored in an Oracle[®] database. For reasons of convenience of transferring the database the Microsoft[®] Access format was used. The Microsoft[®] Access version of the tables was created in February, 2009.

The database is organized into several tables. There are two principal tables, one containing the information related to the site where a soil profile has been sampled (BASICDATA) and one table containing the measurements of the pedological horizons of the profile (SOILPROPS). Additional information is given in the tables containing the raw data on the distribution of the particle size (RAWPSD), on soil water retention (RAWRET) and on hydraulic conductivity by pressure head $K(h)$ (RAWK). Two additional tables (LOCALNGR_TYPE and EUROSTANDARD) were included in the data received to provide supplementary information on geo-referencing the plot positions. These ancillary tables are not part of the original HYPRES data set.

The organization of the data tables with the link fields is presented in Figure 1.

² PARADOX[®] V7.0 ©Borland International Inc.

³ See also: <http://www.macaulay.ac.uk/hypres/index.html>

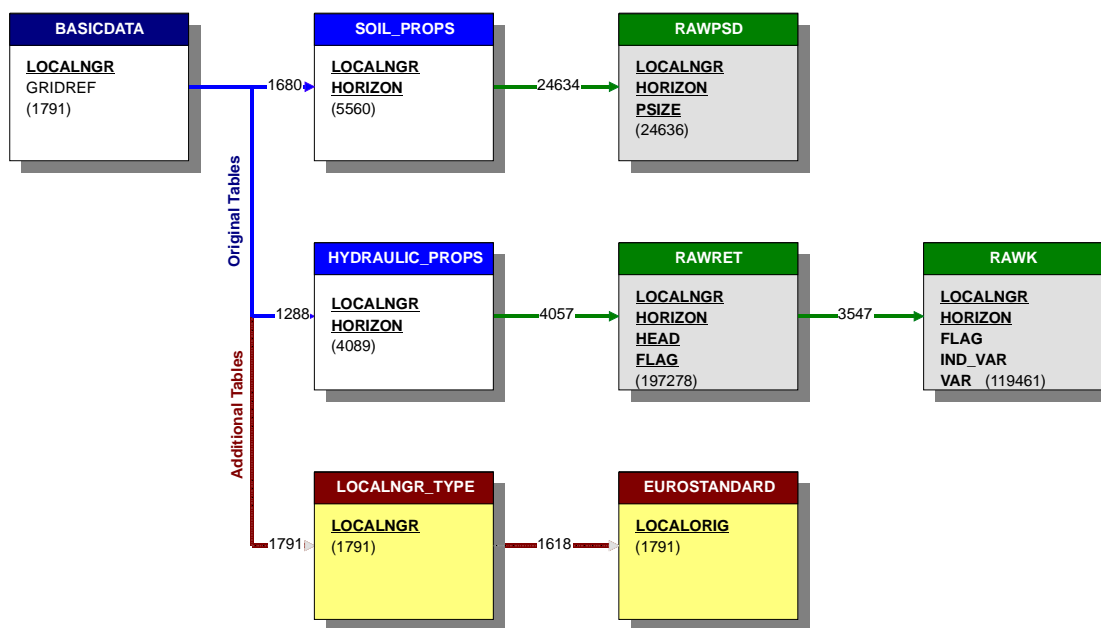


Figure 1: Organization of HYPRES Database Used in Study

The links of the two additional tables shown in the graph are set to the key fields of the tables. Indicated in the links are also the records in a table and the links of the child to the parent table on the key field of the parent table as used in the study.

2.2 Tables and Relationships

Subsequent to the final project report published in 1998 the database has undergone significant changes with new fields added to tables. The LOCALNGR_TYPE and EUROSTANDARD tables are ancillary to the database and shown here due to their relevance to the information they contain and which was made use of in the study.

Some of the additional fields are of only ancillary character, but the field LOCALNGR replaces the field GRIDREF for forming the table primary index. The LOCALNGR field contains the merge of the plot latitude and longitude for un-projected plot coordinates or the x-y values for geo-referenced data, where such data are given or, in the absence of the coordinates, a unique integer value. The function of the field as primary index was confirmed through personal communication. Using the LOCALNGR field to define table indices and links still requires particular care to be applied to address the problems of duplicate records and issues of referential integrity in the table relations.

2.2.1 Table: BASICDATA

For the BASICDATA table the key field or primary index is thus set to the LOCALNGR field. The table contains a total of 1,791 records, which are uniquely identified by the field. For the version of the table provided the GRIDREF field is not suitable to form the primary index, because it contains empty and duplicate entries. To allow a single field to act as primary key the LOCALNGR field was introduced at a later stage.

2.2.2 Table: SOIL_PROPS

The SOIL_PROPS table is indexed on a combined key made from the LOCALNGR and HORIZON fields. It contains 5,560 records on horizons for 1,680 sites (unique LOCALNGR). As a consequence, 111 sites of the BASICDATA do not have any data on the properties of the soil horizons.

2.2.3 Table: HYDRAULIC_PROPS

The HYDRAULIC_PROPS table contains 4,089 records. As the SOILPROPS table it is indexed on the LOCALNGR and HORIZON fields. The table contains the hydrological data for the profile horizons for 1,288 sites of the BASICDATA table. The hydraulic properties can be directly linked to the soil horizon properties where the raw soil hydrological data could be parameterized using the Mualem-van Genuchten parameterization procedure. Thus, 1,471 horizons have no standardized parameters.

2.2.4 Table: RAWPSD

The RAWPSD table contains 24,636 records with data on the distribution of the particle size within a horizon. When not using the SEQNO field in the table key 2 duplicate records were found in the table. The table links to 3,908 horizons from 1,180 sites in the SOIL_PROPS table.

2.2.5 Tables: RAWRET and RAWK

These two tables store the raw data as collated during the HYPRES project. They contain data on soil moisture characteristics and the K/θ relationship. These data were subsequently parameterized and the parameter set was stored in the HYDRAULICS_PROPS table.

The 197,278 records of the RAWRET table cannot be uniquely identified as they represent output data from the experimental procedures to determine the K/θ relationship, some of which would have been repeated. Hence, the table contains 186,149 unique records, while 9,359 records are duplicates with up to 10 copies from the experimental data. The table records describe 1,579 sites and 4,969 horizons. Of these 1,274 sites with 4,057 horizons can be linked to the HYDRAULIC_PROPS table.

The RAWK table, which links to the RAWRET table, contains 119,461 records of which 118,633 are unique. Also for this table the duplicate entries arise from experimental data. No primary index can be defined for the table due to duplicate records. The table contains 4,969 profile horizons (LOCALNGR + HORIZON), of which 3,547 can be linked to the RAWRET table. As a consequence, no link can be established for the hydraulic conductivity for 1,422 profile horizons. For 25 profile horizons of the RAWRET table no data on the hydraulic conductivity are included in the RAWK table.

The RAWRET and RAWK tables were only intended to hold the unprocessed, hydrological data in the event that new parameterisation methods become available or that there was a question over the existing parameterisation data held in the HYDRAULIC_PROPS table. Data from the RAWK table were not extracted to the SPADE/M data set and the table was not included in any queries.

2.2.6 Tables: LOCALNGR_TYPE and EUROSTANDARD

The additional tables LOCALNGR_TYPE and EUROSTANDARD differ in several aspects from the other tables. The tables were not part of the standard HYPRES data set and were used to investigate the possibility of generating a standardized geo-reference for the plots. The work was done at the end of the project period, then discarded and subsequently not resumed. The tables are presented here because they were used as ancillary data for the standardization of the HYPRES geo-references to the SPADE/M specifications.

The EUROSTANDARD table contains 1,726 records. In the table a field ID seems to be set to the form the key. Alternatively, the field NO_PROFILE can also form the primary key. The field COUNTRY contains a text string for the country rather than the numeric code used in the table LOCALNGR_TYPE and BASICDATA with the same name. A similar variation is found for the field COORD_TYPE, where a numeric code is used in the EUROSTANDARD table while the tables LOCALNGR_TYPE and BASICDATA use alpha-numeric entries. The EUROSTANDARD table is linked to the LOCALNGR_TYPE table using the LOCALNGR and LOCAL_ORIG fields. This link references 1,618 sites in the LOCALNGR_TYPE table. Therefore, 49 sites of the EUROSTANDARD are not referenced in the LOCALNGR_TYPE table using this link, and subsequently also not in the BASICDATA table. The LOCALNGR_TYPE and the EUROSTANDARD tables differ from the other tables of the database since they are of

intermediate character and neither table is mentioned in the project report (Wösten, *et al.*, 1998).

2.3 Attributes

The database does not contain any tables describing the fields or units used. For any information on parameter codes and measurements related to analysis methods or data processing the documentation has to be used. An overview of the table describing the sample site with their fields and formats is given Table 1.

Table 1: Site Data

BASICDATA			LOCALNGR_TYPE			EUROSTANDARD		
FIELD_NAME	FORMAT	DIM	FIELD_NAME	FORMAT	DIM	FIELD_NAME	FORMAT	DIM
GRIDREF	Text	30	LOCALNGR	Text	25	ID	Number	10
NAME	Text	40	FAO_SOIL	Text	10	NO_PROFIL	Double	
FAO_SOIL	Text	10	COUNTRY	Text	3	X	Double	
COUNTRY	Text	3	COORD_TYPE	Text	30	Y	Double	
LOCALNGR	Text	25	FAO	Text	40	SOIL	Text	5
LOCALMSSG	Text	80	LOCALMSSG	Text	80	COUNTRY	Text	5
LOCALSERIES	Text	30				REGION	Text	5
TOP_DEPTH_GW	Text	6				AUTEUR	Double	
BOT_DEPTH_GW	Text	6				LOCAL_ORIG	Text	25
SITEDESCRIP	Text	255				COORD_TYPE	Double	
SAMPLEDATE	Date/Time					X1	Double	
ANNRAIN	Text	5				X2	Double	
AVE_JAN_TEMP	Text	5				X3	Double	
AVE_JUL_TEMP	Text	5				Y1	Double	
CONTACT_NAME	Text	30				Y2	Double	
CONTACT_ADDR	Text	255				Y3	Double	
ESS								
EMAIL	Text	40				SYSTEME	Double	
PUBLICN	Text	255				COORD1	Text	30
COMMENTS1	Text	255				COORD2	Text	30
COMMENTS2	Text	255				GISCO	Text	126
KEYWORDS	Text	255						
NUMBER_HOR	Number	5						
RATING	Number	5						
RATED_BY	Text	30						
SEQNO	Double							
FAO	Text	40						
COORD_TYPE	Text	30						

Italics: fields modified or added after project report (Wösten, *et al.*, 1998)

Tables LOCALNGR_TYPE and EUROSTANDARD are not part of the original HYPRES dataset

The tables, fields and formats related to information of the soil horizons are summarized in Table 2.

Table 2: Profile Horizon Data

SOILPROPS			HYDRAULIC_PROPS			RAWK		
<i>FIELD_NAME</i>	<i>FORMAT</i>	<i>DIM</i>	<i>FIELD_NAME</i>	<i>FORMAT</i>	<i>DIM</i>	<i>FIELD_NAME</i>	<i>FORMAT</i>	<i>DIM</i>
GRIDREF	Text	30	GRIDREF	Text	30	GRIDREF	Text	30
HORIZON	Text	7	LOCALNGR	Text	25	HORIZON	Text	7
0	Double		HORIZON	Text	7	FLAG	Text	1
BOT_DEPTH	Double		DMVG_SAT	Double		IND_VAR	Text	5
STRUCTURE1	Text	45	DMVG_RES	Double		VAR	Double	
STRUCTURE2	Text	45	DMVG_ALPHA	Double		COND	Double	
USCLAY	Double		DMVG_N	Double		SEQNO	Double	
USSILT	Double		DMVG_M	Double		LOCALNGR	Text	25
USSAND	Double		DMVG_L	Double		ATTILA_ID	Double	
KSAT	Double		DMVG_KS	Double				
SATWAT	Double		THETA0	Double				
BULK_DEN	Double		THETA10	Double				
PARTICLE_DEN	Double		THETA20	Double				
POROSITY	Double		THETA50	Double				
ORG_MAT	Double		THETA100	Double				
TEXTURE	Text	6	THETA200	Double				
MVG_SAT	Double		THETA250	Double				
MVG_RESID	Double		THETA500	Double				
MVG_ALPHA	Double		THETA1000	Double				
MVG_N	Double		THETA2000	Double				
MVG_M	Double		THETA5000	Double				
MVG_L	Double		THETA10000	Double				
MVG_KS	Double		THETA15000	Double				
COMMENTS	Text	255	THETA16000	Double				
KEYWORDS	Text	255	COND0	Double				
SEQNO	Double		COND10	Double				
LOCALNGR	Text	25	COND20	Double				
FLAG_50	Text	6	COND50	Double				
HYDROCLASS	Text	5	COND100	Double				
			COND200	Double				
			COND250	Double				
			COND500	Double				
			COND1000	Double				
			COND2000	Double				
			COND5000	Double				
			COND10000	Double				
			COND15000	Double				
			COND16000	Double				
			LOCALNGRHOR	Text	55			

Italics: subsequently added to provide unique primary key or for intermediate processing purposes

All key and link fields are defined as alpha-numeric format. This is a workable option, but care needs to be taken when the entries contain leading spaces or when the RDBMS used distinguish between small and capital letters, i.e. when the key becomes case sensitive.

3 ADAPTATIONS OF HYPRES DATA

The plot and profile data from the HYPRES data set differs from the SPADE/M data set with respect to the depth ranges assessed within a soil profile, the parameters reported and units used. The processing and changes applied to adapt the HYPRES data to the specifications of the SPADE/M database concern categorical as well as continuous numeric data and are presented by the fields of the receiving SPADE/M database.

3.1 *Conversion of List Data in BASICDATA Table*

The conversion of data belonging to ordered lists (tuples) mainly concerns observations at the site level. In HYPRES these data are stored in the BASICDATA table. In HYPRES data defined as categorical or ordinal in SPADE/M are often stored as free text and are stored in different fields. Hence, the method used to extract and arrange the information was to search in each field for the entries of the list and transfer the result to the particular field.

3.1.1 **Sample Date (SAMPLE_DATE)**

Where a sample date is given the validity of the dates were checked and all were found to correspond to valid entries. The date of sampling soil profiles was not known or not recorded for 725 records.

3.1.2 **Data Source (SOURCE)**

For all data from the HYPRES database the entry for the source parameter is set to “HYPRES”. The origin of the organizations or persons providing the profile data is not extracted and transferred to the SPADE/M field. If needed such information will have to be retrieved from the original database (BASICDATA.CONTACT_NAME or BASICDATA.CONTACT_ADDRESS).

3.1.3 Plot Country (COUNTRY)

All sites are assigned a value for the COUNTRY field. The entries in the HYPRES database follow the international phone code, with an additional code to separate England & Wales, Northern Ireland and Scotland in the UK. In the conversion a single code was assigned to the UK regions (44).

The database contains references to sites in 15 countries. The countries and the number of sites referenced are given in Table 3.

Table 3: Countries Referenced in BASICDATA

Country	Plots		Comment
	No.	Relative (%)	
Belgium	158	8.8	
Denmark	93	5.2	
France	54	3.0	
Germany	834	46.6	
Greece	60	3.4	
Italy	157	8.8	
Netherlands	153	8.5	
Poland	1	0.1	not mentioned in final report
Portugal	20	1.1	
Russia	11	0.6	not mentioned in final report
Slovakia	14	0.8	
Spain	21	1.2	
Sweden	6	0.3	
Switzerland	15	0.8	not mentioned in final report
United Kingdom	194	10.8	England, Wales, Scotland and Northern Ireland listed individually
Total	1791	100.0	

The documentation to the database lists 12 countries for the location of plots. Data from sites in Poland, Russia and Switzerland were taken from the *Database of unsaturated soil hydraulic properties* (UNSODA) (Leij, *et al.*, 1994) and added after the compilation of the final project report. Also given in the table is the relative distribution of referenced plots by country. Almost half of all sites (46.6%) are located in Germany and 83.5% of the sites are located in 5 countries.

3.1.4 Local Name for Sample Site Area (LOC_NAME)

The description of the area where a sample site is located was mainly retrieved from the field BASICDATA.NAME. Some relevant information was also found in the field LOCALMSSG and in a few cases also the field LOCALSERIES. The entries were examined manually and relevant information was transferred to the LOC_NAME field. An entry in the field could be provided for 1,197 records. Any additional information on the site was preserved as far as it was relevant to the site location.

3.1.5 Site Coordinate Projection (PROJ)

All site coordinates were un-projected to latitudes and longitudes using the functionality of a GIS. The coordinates are reported conform to the geodetic datum of the European Terrestrial Reference System 1989 (ETRS 1989) and thus follows the recommendations of INSPIRE⁴. Further details on the process are given in the section on the conversion of site locations to geographic coordinates.

3.1.6 Soil Name as Given by Source (SOIL) and According to FAO (FAO)

The soil names extracted from the data largely originate from the field BASICDATA.FAO. In most cases they are general descriptions of the soil. The soil codes are stored in the field BASICDATA.FAO_SOIL. These codes should correspond to the 1974 or 1990 FAO soil classification schemes. This was generally the case, although some exceptions of non-conform codes were also found. These cases are listed in Table 4.

⁴ Directive 2007/2/EC of the European Parliament and of the Council of 14 March 2007 establishing an Infrastructure for Spatial Information in the European Community (INSPIRE) 14.03.2007

Table 4: Entries in FAO_SOIL Field without Correspondence in FAO Classification

FAO_SOIL Code	FAO Name	New Code	Cases
	Gleyic Arenosol	ARg	1
Bl	Luvic Cambisol	B	194
Chc	Calcario-Haplic Chernozem	Ch	1
GM	Mollic Gleysol	Gm	1
Hhc	Calcario-Haplic Phaeozem	Hh	2
Jg		J	18
Jgg		J	7
Jh	Humic Fluvisol	FLu	5
Js	Sodic* Fluvisol	FL	2
ND	Haplic Calcisol	CLh	3
ND	Haplic Luvisol	LVh	1
Qg	Gleyic Arenosol	ARg	1
Vk	Calcic Vertisol	VRk	1

* According to LOCALNGR_TYPE.FAO entry.

When assigning new FAO codes precedence was given to the soil name information. In case no matching code was found for the 1974 FAO nomenclature the 1990 system was used. When neither system provided a suitable code the higher classification level was assigned. This was the case for example for code *Js*, which is given as *sodic Fluvisol* in the LOCALNGR_TYPE.FAO field, but which has no correspondence in the FAO90 codes of the SGDBE and was therefore assigned a more generic Level 1 code (*FL*). No entries in the FAO_SOIL field were found for 133 records. Where no information on the soil name could be found no codes were assigned. A code according to FAO74 or FAO90 could be assigned to 1,659 records for plots.

3.1.7 Depth of Groundwater Level (GWL_NM, GWL_HI and GWL_LO)

The BASICDATA table provides information on the depth to the groundwater as values on the high and low range in the fields TOP_DEPTH_GW and BOT_DEPTH_GW. The fields were assigned to the mean highest (GWL_HI) and mean lowest (GWL_LO) fields of the SPADE/M database. The field containing data on the normal groundwater table (GWL_NM) has no equivalent in HYPRES and no data could be transferred to the field.

Both fields in the BASICDATA table are defined in alpha-numeric format to accommodate a text indicator for missing data. Two indicators are used, "NA" (TOP_DEPTH_GW: 573; BOT_DEPTH_GW: 989) and "ND" (TOP_DEPTH_GW: 573; BOT_DEPTH_GW: 1,022). A numeric entry in one or both fields was given in

241 cases. In 19 cases the upper value for the ground water depth is identical to the lower value. For 30 records the lower depth value is higher than the value for the upper depth. Where the lower level value was greater than the higher value level the entries were exchanged. Where only a single value was recorded the values were left as found.

The continuous numeric values were then classified according to the 5 classes with 50 cm intervals used by SPADE/M and the resulting values transferred to the `GWL_HI` (229 values) and `GWL_LO` (196) fields. A value for both ground water levels could be stored for 184 plots.

3.1.8 Dominant Land Use (LU)

The dominant land use at the site of the profile is not recorded in any particular field of the `BASICDATA` table. Information on the land use is distributed across several fields, mainly in the fields `LOCALMSSG` and `SITEDESCRIP`. Other fields contain land use information in isolated cases. The descriptions of the land uses were first standardized to a common spelling. The entries were then assigned to a class of the land use according to the most appropriate item of the list. The list of land use items is based on the `CORINE` Land Cover (CLC) classification. The 64 items cover all 3 levels of the CLC nomenclature. A land use according to the list could be given for 1,108 plots.

A particular uncertainty in assigning the free-form description of the land use of the `BASICDATA` table to a CLC class is posed by descriptions indicating a type of grassland. The terms meadow, grazing, pasture and grassland, the latter also as temporary or permanent, were used to characterize the land use at the sample site in the table. It is not evident from the data that these terms have been applied consistently between plots. This uncertainty in the definition of grassland is not uncommon and the separation between the various types of grassland is fluid. This makes stratifying data according to land use, for example to support the definition of a PTR for soil organic carbon, a more demanding task. Where the land use information does not separate arable land from grassland the profile data cannot be used for this specific purpose.

3.1.9 Dominant Parent Material (PM)

Information on the dominant parent material is dispersed over several alpha-numeric fields of the `BASICDATA` table. The detail on the parent material ranges from minimal reference to the soil texture to very comprehensive portrayals. Any references to the parent material were assigned to one of the 127 items of the parent material list specified for the SPADE/M database. The list reproduces the classification of parent material (MAT) as specified for the `SGDBE`. It contains all 3 levels of the classification scheme in a single list.

The procedure employed to transfer the free-form information to an item of the parent material list follows the approach used for the land use information. First, relevant

information was extracted from the various fields and collected. Second, the spelling of the descriptions was harmonized and thirdly a list item was assigned to the temporary entries. Using the transfer table the sites with suitable information were then assigned an item of the parent material list. For 791 sites a reference that could be used to specify the parent material was transferred to the field.

3.1.10 Depth to Obstruction for Rooting (D_ROO_X, D_ROC_X and D_OTH_X)

No values or other data were transferred to these fields, which have no equivalent in the HYPRES database.

3.1.11 Origin of Plot Observations (ORIGIN)

In SPADE/M the field ORIGIN refers to the aggregation method used to report the horizon properties (e.g. from single representative profile or from more than one profile), not the institutional source of the data. Such information is not available from the HYPRES database. Instead, for a number of plots the aggregation method used to determine water retention or hydraulic conductivity are mentioned. Because the method of aggregation available in the HYPRES database does not relate to the same soil profile properties as the SPADE/M data and because it was found difficult to consistently structure the methods used to determine the aggregation method at site level any extractable information was retained as a comment, but not transferred to the ORIGIN field.

3.1.12 Comment on Plot Observations (COM_PL)

Additional remarks on site locations and attributes were included in the LOC_NAME field of the BASICDATA table and included in the COM_PL field. For other observations related to site characteristics, such as land use and parent material, the relevant information was transferred to the corresponding fields.

3.2 Conversion of List Data in SOIL_PROPS Table

The SPADE/M table VAL_LIST contains 4 parameters which are linked to the characterization of the soil profile specified for a site. Corresponding data in the HYPRES database are stored in the SOIL_PROPS table.

3.2.1 Horizon Name as given by Source (HOR_NAME)

An entry for the field is given for all records. However, some entries are not indicative of the horizon properties. Such entries are “NA” (1), “ND” (134), “ND n” (60) and numeric entries from 1 to 7 (155). Entries without numeric prefixes were preceded by a leading space. This space was preserved when transferring the entries to the table of listed values.

The overlap of horizon names with the SPADE/M data is small. For 9 entries the HYPRES horizon names are identical to those of the SPADE profiles. This difference is very much a consequence of the divergent methods used to code the profile horizons between the databases. The HYPRES database uses a set of 7 fixed positions to record horizon data while the horizons of the SPADE/M database use a free-form format. A strict comparison, including the leading spaces stipulated by the FAO coding rules, leads to the low count of common entries. In the course of the study it was found that, while these leading spaces are part of the coding rules and should be retained, they can be lost when moving the data between software. These unintentional concatenated entries produce different joins between the tables and can confuse the data analysis.

3.2.2 Percentage of Stones and Gravel (GRAVEL)

There is no specific field defined in HYPRES to store information on the percentage of stones or gravel in the soil. For some profiles (48) the comments included in fields with descriptive information contain an indication of the parameter in form of “stoniness negligible”. This comment is understood to correspond to the SPADE category “*Very few, <5% by volume*”.

3.2.3 Structure Class (STRUCT)

The soil structure is described as the primary (field: STRUCTURE1) and secondary (field: STRUCTURE2) structure following FAO guidelines. SPADE/M uses only one field to characterize the horizon structure and the corresponding information was taken from the primary structure field. The information is given in form of a descriptive text

without a specific naming convention. The description of the soil structure does not directly match the 10 categories of the SPADE/M data. At times the HYPRES descriptions are more detailed than the categories available by providing an additional grading of the structure, such as weak, moderate or strong. This additional information could not be transferred to the classification of the SGDBE, which is used by SPADE/M.

Meaningful data for the primary structure (not entry “ND” or “not available”) was given for 2,126 records. Where the horizon structure was given as peat the horizon attribute was set to an absence of structure, as peats do not have a structure in many cases.

3.2.4 Sodium Adsorption Ratio < 4 (AR_NA_X)

Data on the sodium adsorption rate are not stored in the HYPRES database.

3.3 Continuous Numeric Data

Site and horizon characteristics recorded in form of continuous numeric values (integer or float type). The methods used to adjust the data are presented in the order of the RPARA_ID field of the DIC_RANGE table.

3.3.1 Conversion to Geographic Coordinates (COOR_X, COOR_Y)

For integration with spatial data the soil profiles need to be geographically positioned. The geographic positioning allows an appreciation of the geographic spread of the measured profile data and is essential when the data is used to validate the estimates generated by spatial modelling soil parameters.

In HYPRES V. 1.0, which is described in the final project report, the geographic position is given in the GRIDREF field. The field contains data converted from national projections to a projection using the Lambert Azimuth as projection system. No further information on the type of Lambert Azimuth projection (equal area, conformal conic) or the parameters (datum, origin of latitude and longitude, false northing and easting) are given. It was assumed, and later confirmed, that the projection used was Lambert Azimuth Equal Area, as defined by the *Geographical Information System at the Commission* (GISCO) at the time (origin longitude: 9.0°; origin latitude: 48°). This hypothesis was verified by plotting the coordinates based on the projection.

Subsequent to the initial coordinate transformation additional processing was carried out on the parameter which resulted in introducing the LOCALNGR as a key field and new

coordinates as recorded in the ancillary EUROSTANDARD table. The table contains steps of the conversion from the initial values to the entries in the fields COORD1 and COORD2, but no information on the projection parameters used to convert the data or the common reference system used. Some information on the projection system of the initial coordinates is given in the COORD_TYPE field of the ancillary LOCALNGR_TYPE table. The information provided gives an indication of the projection system used, but is inadequate to reproduce the coordinate transformation.

Rather than transforming the initial coordinates to another projection the site positions were transformed to a non-projected geographic reference system. The European Terrestrial Reference System 1989 (ETRS89) was adopted which is compliant with the INSPIRE specifications (Annoni, *et al.*, 2001). One option would have been to un-project the coordinates given in the table EUROSTANDARD. However, the available information was found insufficient with respect to the projection parameters used to convert the coordinates reliably. The task of converting the coordinates to a common system was therefore completely redone, using the initial data wherever possible.

The merged coordinates in the field LOCALNGR were separated into X and Y coordinates or longitude and latitude. The information in the fields BASICDATA.COORD_TYPE and LOCALNGR_TYPE.COORD_TYPE was then used to find the projection system of the original data. Finding appropriate projection systems for the coordinate transformation was a tedious and non-trivial task with a substantial element of trial-and-error. The results are not without uncertainties, in particular because there are differences to the data in the ancillary EUROSTANDARD table. The situations encountered and solutions applied are described hereafter by country.

- *Greece*

For Greece 60 plots are referenced in the table BASICDATA. Positions in geographic coordinates are given for 58 plots. Of those, 1 plot is located outside Greece (Xyloupoli Thessaloniki) and 1 outside Europe. The longitude of the latter was modified from 27.4° to 37.4°. This step seemed justified by the location given for the plot (Nafplio, Tiryntha). Still, the coordinates set the plot into the sea. For the plots near Spata, Attica the original coordinates set the plots into the sea. This is mainly due to the value set for minutes in the latitude field, which for the location is 58 instead of 38. The coordinates are also unusual in that the seconds for all 6 plots of the site increase from 0 to 5.

- *The Netherlands*

A total of 153 plots are given for the Netherlands. Coordinates in the Dutch National Grid (Rijksdriehoekstelsel Grid) are recorded for 133 plots and in LAT/LONG values for 1 plot. For the projected data the values for the X and Y coordinates were reversed. The false origin used was in X: 155000 and in Y: 463000.

- ***Belgium***

The number of plots in Belgium in the database is 158. Coordinates are given for 141 plots, all projected according to the Belgium National Grid. The projection parameters used for the transformation were based on the Datum Géodésique 1950 definition (LCC1950).

- ***France***

Profile data in France were indicated for 54 plots. For 51 plots coordinates for a projection are given, while for 2 plots geographic coordinates are recorded. No information on the projection is given in the table. The plots could be positioned according to the Lambert II étendu specifications. One of the plots with geographic coordinates had only data for degrees and minutes. For both plots the minute value was 75, which indicates that the coordinates were reported in decimals or that a typing error occurred.

- ***Spain***

For Spain 21 plots are referenced in the BASICDATA table. All plots have geographic coordinates recorded, although the X-coordinates are positive values notwithstanding that all are located West of Greenwich. For 10 plots in the region of Alora the coordinates are not recorded as DEG/MIN/SEC. Rather, the data seem to be decimals, because they exceed a value of 59. Yet, even when assuming that decimal notation was used the geographic position of the plots is nowhere near Alora or even in Spain. Those plots were therefore excluded from being assigned coordinates. For the plot near Pago Macharnudo the coordinates in the database (36° 44' 50" North; 06° 30' 20" West) are only within the larger environs of the site (36° 54' 0" North, 6° 3' 0").

- ***Italy***

Of the 155 plots given for Italy 153 are recorded in UTM-33N with 2 plots without data on the geographic position. No particular conditions in transforming the projected coordinates were encountered.

- ***Slovakia***

The 14 plots for Slovakia have data on the plot locations for 13 plots in form of geographic coordinates. The data are not very precise, frequently lacking a value for seconds. The ancillary EUROSTANDARDS table does not use the same entries in the field LOCALNGR. As a consequence, the records cannot be linked to the table BASICDATA using this configuration.

- ***Denmark***

Coordinates in UTM are given for 88 of the 93 plots in Denmark. These coordinates could be converted to the geographic system using UTM zone 32.

- ***Sweden***

The 6 plots for Sweden had geographic coordinates recorded. While the format indicates DEG/MIN/SEC the data for the seconds are 0 in all cases but one.

- ***Germany***

With 834 sites Germany is the largest single contributor of soil profiles to the database. For 27 sites no data on the position of the plot is provided. For 10 sites the position is given in geographic coordinates. The data values for these plots lack information on seconds or the values recorded are 0 or 1 for both, latitude and longitude. For the remaining 797 plots the COORD_TYPE entries are either “Gauss-Kruger” or “Krassowski-Ellipsoid”⁵.

Identifying the appropriate projection for the German plots was an arduous task. Where in the information field the reference to the “Gauss-Kruger” projection was made this was interpreted to signify the use of the Bessel 1841 ellipsoid with the datum Potsdam. The information “Krassowski-Ellipsoid” was interpreted as specifying the Krassowsky ellipsoid with the datum Pulkov, using the Gauss-Kruger projection. All sites using the Krassowsky ellipsoid are located in the former German Democratic Republic. The data should be based on the 3° projection system. However, as deduced from obvious mapping errors when using the projection parameters, it was found that the coordinates relate to the 6° projection system. One plot in the Oderbruch area was likely to be recorded with incorrect X-coordinates (3664220, probably 3464220). For another plot without further information on the site the Y-coordinate (2825650) places the plot outside Germany.

- ***Portugal***

The location of all 20 plots for the mainland of Portugal were recorded in geographic coordinates. No particular situations on the plot locations were found.

- ***United Kingdom***

The locations on the 190 plots in the UK were recorded according to the specifications of the UK National Grid based on the OSGB36 datum for plots in Scotland, England and Wales. For Scotland the notation used the grid lettering of the 100km squares. For the 8 plots located in Northern Ireland the Irish National Grid specifications were used. For 4 plots no data on the location were recorded in the tables.

While the plot coordinates could be established from the data for plots in Scotland and Northern Ireland the link between the plot recorded in the

⁵ The spelling of Gauss-Kruger and Krassovsky varies between GIS applications. In this document the spelling of the IDRISI® GIS is used (IDRISI is a registered Trademark of Clark University).

BASICDATA table and the coordinates assigned in the EUROSTANDARD table seems to have been rearranged at some stage, leading to arbitrary locations of plots in those areas.

Plots located in Poland (1), Russia (11) and Switzerland (15) were added to the HYPRES database from the UNSODA database. Those plots lack adequate information on their geographic location and on several pedological parameters.

3.3.2 Variations in Plot Locations

The difference in geographic location to the coordinates given in the COORD1 and COORD2 fields of the EUROSTANDARD table were approximated using the old GISCO LAEA projection system. A graphical presentation of the distance in plot locations is given in Figure 2.

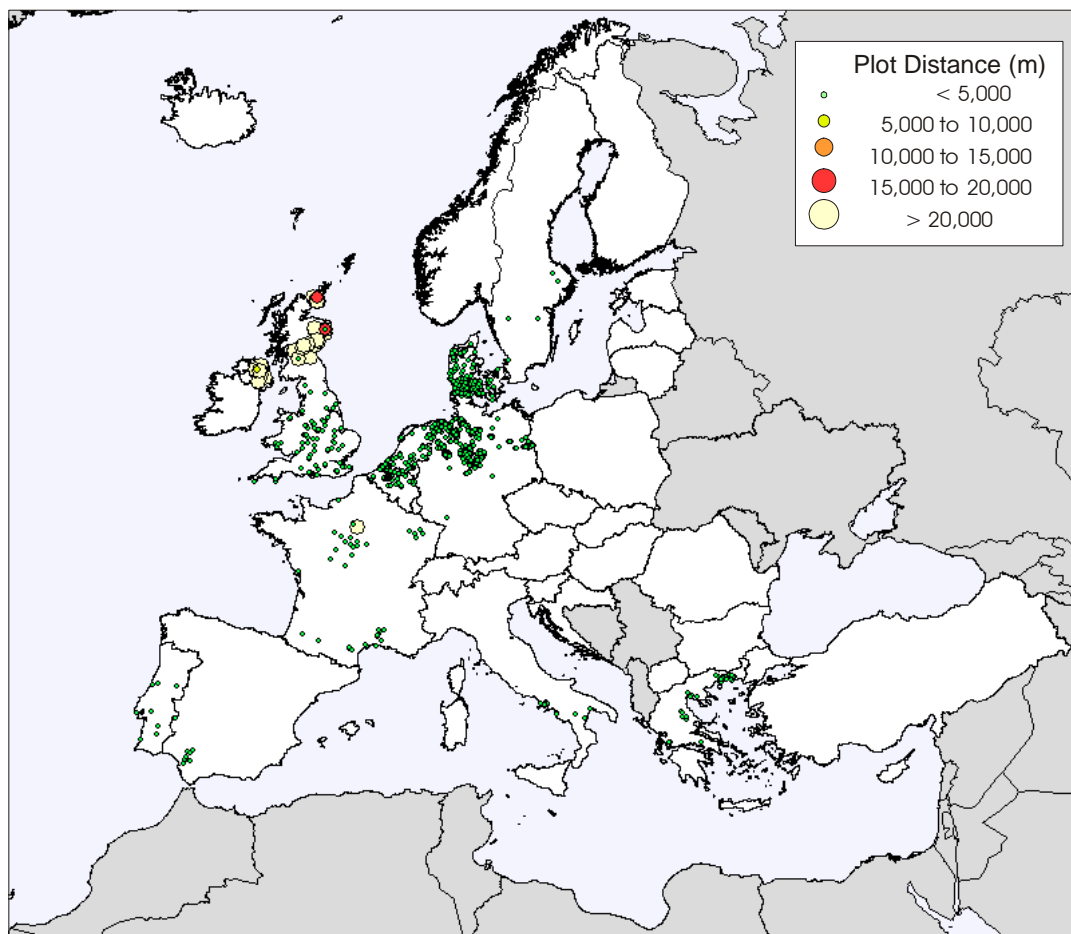


Figure 2: Distance between Plots from EUROSTANDARD Coordinates to Re-calculated Locations

Using the LOCALNGR field coordinates from 1,382 plots could be compared. For most of the plots the variation between the plot locations provided in the EUROSTANDARD table and the re-calculated values is $< 1,000$ m. For one plot in France the difference of 57 km results from the interpretation of the geographic coordinates as decimals. All other differences over 5 km are restricted to plots in Northern Ireland and Scotland. In these regions the coordinates between the two calculations agree, but are assigned to different plots. It appears that at one stage during processing the data for the EUROSTANDARD table the site IDs got scrambled.

Using the Bessel Ellipsoid for plots in Germany the distance between the two sets of site locations is 1,500m. This constant difference indicates a systematic source in the parameters used for the conversion of the coordinates from the *Deutsches Hauptdreiecksnetz* (DHDN) with Potsdam Datum and Bessel 1841 ellipsoid using Gauss-Kruger projection. The source of the variation could not be identified by the study and in the absence of any additional information no further steps to reduce the variation were taken. Using the parameters of the 6° projection for the Krassowsky ellipsoid and Pulkov datum resulted in close proximity of the plot locations.

Where the location of a plot was recoded as geographic latitudes and longitudes the lack of information on seconds makes specifying precise locations an impossible task. Moreover, where all seconds are given as 0 values the data suggests an accuracy in locating plots, which is not realistic on the ground, and may have been the consequence of protecting the exactly location of a plot. The values of the GRIDREF field are given to a precision of 1m, but in the absence of seconds for the site LAT/LONG coordinates the uncertainty in the position of the site is in the region of 7 km at a latitude of 50°N.

When taking into account the uncertainty over the projection parameters used to specify the original coordinates and the imprecision of the coordinates in the database the location of the plots can only be considered approximate. This uncertainty in plot locations of several kilometers are carried on when using a spatial link to compare the data with other profile databases but also when locating the profiles on spatial layers. This severely limits the use of the data for monitoring and validation purposes, which rely on the comparison of point data.

A summary of plots with geographic information by country is presented in Table 5.

Table 5: Summary of Plots with Geographic Location

Country	Total Plots	Location in LOCALNGR	Plots Retained	
	<i>No.</i>	<i>No.</i>	<i>No.</i>	<i>%</i>
Belgium	158	141	141	89.2
Denmark	93	88	88	94.6
France	54	53	53	98.1
Germany	834	792	791	94.8
Greece	60	58	50	83.3
Italy	157	153	153	97.5
Netherlands	153	134	134	87.6
Poland*	1	0	0	0.0
Portugal	20	20	20	100.0
Russia*	11	0	0	0.0
Slovakia	14	13	13	92.9
Spain	21	21	10	47.6
Sweden	6	6	6	100.0
Switzerland*	15	0	0	0.0
United Kingdom	194	190	190	97.9
TOTAL	1791	1669	1649	92.1

* Countries not included in final project report

Of the total number of 1,791 plots in the database plot coordinates could be extracted from the field LOCALNGR for 1,669 plots. For 1,649 plots the location could be translated into geographic coordinates and placed within the country boundaries. The overall rate of plots with reasonable coordinates is 92.1%. There are large variations of the figure by country, ranging from 47.6 (Spain) to 100.0 % (Portugal, Sweden).

It should be noted that some plots with large deviations were assigned different coordinates in the EUROSTANDARD table. In cases where a site name is given the coordinates in the table seem to be more appropriate to represent the site position, e.g. in Greece and Spain. Yet, there is no reference in the documentation as to the source of the new coordinates and the data recorded in the tables are inconsistent. In these cases the strict adherence to transparent processing may not have given the best result.

3.3.3 Plot Altitude (ALT)

There is no field in the HYPRES database dedicated to record the altitude of the site at which the profiles were taken. The altitude is at times provided in the comment fields and was extracted when given as a numeric value. This could be done for 55 plots.

3.3.4 Normal Groundwater Level (GWL_N_M)

Instead of the normal groundwater level HYPRES contains fields for the upper and lower depth of a variable groundwater table. These figures were classified to the mean highest (GWL_HI) and mean lowest (GWL_LO) levels of the groundwater of the categorical data. Taking the mean of the two values to represent the normal groundwater level was not considered a viable option.

3.3.5 Rooting Depth and Depth to Obstructions to Rooting (D_ROO, D_ROC and D_OTHOB)

These fields of the SPADE/M database have no correspondence in the HYPRES data. There are at times qualitative statements in the comments, but no numeric data that could be transferred.

3.3.6 Upper and Lower Horizon Limits (HOR_START, HOR_END)

The profile horizons of HYPRES are intended at deriving comprehensive hydrological parameters rather than record a profile by the pedological horizons as found in the field. Therefore, for a number of profiles the sample depth rather than the profile thickness was recorded. As a consequence, the profile description contains gaps in the horizon depth and duplications for a pedological horizon or depth segment with different depth limits. This structure of the horizon data is not compatible with the SPADE/M data model. In SPADE/M a profile description is continuous, i.e. without gaps or overlaps in the depth parameters. The sequence of the horizons is given by the horizon identifier, which is unique for each plot. Therefore, to transfer the HYPRES data to the SPADE/M model some conditions on the data were introduced:

- **Step 1**
The upper limit of the horizon is not negative. This also excludes horizons where no upper limit is given (blank entry) or where the code “-9” is used to signify an unknown upper limit (remaining records: 5,365).
- **Step 2**
The lower horizon limit is specified, i.e. not “-9” or blank (remaining records: 5,225).
- **Step 3**
The upper limit of the horizon must be greater than the lower limit (remaining records: 5,211).
- **Step 4**
The upper limit of the profile horizon is less than 10 cm (remaining records:

3,281). The limit of 10 cm for describing the profile was used because setting a restriction on the upper limit of the profile to start at 0 cm is met by only 154 profiles.

Next, all profiles without an overlap in depth between horizons were separated from those with a depth overlap between one or more horizons (records without horizon overlap: 1,963; records with horizon overlap: 1,318). Affected by multiple data for the same or overlapping depth ranges are 195 plots.

Contrary to other soil profile descriptions, such as for BioSoil (Hiederer, *et al.*, 2011), the HYPRES profiles start at the surface with a value of “0” and counting depth with positive values. An organic layer is specifically identified. For profiles recording soil parameters by sample depth instead of horizons the starting depth of the uppermost sample depth of the profile given in the SOIL_PROPS.TOP_DEPTH field differs from “0”. It is not clear if in these cases an organic layer has been omitted from the profile description, which limits the comparability of data with soil profiles sampled under forest.

3.3.7 Particle Content (CLAY, SILT_1/2 and SAND_1/2/3)

The SOIL_PROPS table contains the relative distribution for the principal particle classes. The particle sizes given in the table were interpolated from raw data of the RAWPSD table to follow the FAO specifications for particle sizes:

- **USCLAY (< 2 μm)**

For the clay content no data as indicated by the code “-9” is recorded for 660 records. One record contains a value of “-1”, which is interpreted to indicate missing data. A value of zero is given for 157 records.

- **USSILT (2 – 50 μm)**

For the silt fraction the indicator for no data is set for 667 records. A value of zero is given for 22 records.

- **USSAND (50 - 2,000 μm)**

The sand fraction contains 667 entries for no data (“-9”) and in 1 case a value of zero is given.

A value of zero for a particle class was interpreted as indicating the result of an analysis, i.e. the absence of particles falling into the size class, and not the absence of any data. Under this assumption a value for all tree particle classes is given for 4,427 records. In 5 cases the sum of the particle contents does not add to 100.0. The deviations are within \pm

0.1% in 33 cases and could be attributed to rounding inaccuracies where values are recorded to 2 decimals. The extremes in the sum of the particle content are given by 94.80% and 125.70%. The latter (LOCALNGR: 161025 428730) appears to be caused by an entry for clay in one horizon (“*IA h b*”) which belongs to another horizon (“*IC wg a*” and “*IC wg b*”) of the profile.

The measured distribution of particle sizes is given in the RAWPSD table. The table contains data for 3,908 horizons where the particle content is reported with a total of 30 different particle sizes. An overview over the number of occurrences for the particle sizes in the table is given in Table 6.

Table 6: Occurrence of Particle Sizes in HYPRES RAWPSD Table

Particle Size	Occurrence	Particle Size	Occurrence
<i>μm</i>	<i>No.</i>	<i>μm</i>	<i>No.</i>
0.2	188	75.0	17
1.0	2	100.0	488
2.0	3,860	105.0	144
4.0	18	125.0	1,623
5.0	36	150.0	148
6.0	61	180.0	4
6.3	529	200.0	2,839
8.0	17	210.0	144
10.0	568	250.0	117
16.0	163	300.0	127
20.0	2,903	500.0	263
32.0	18	600.0	456
50.0	767	630.0	2,091
60.0	660	1,000.0	213
63.0	2,316	2,000.0	3,856

The SPADE/M data model allows storing more than one value for the silt (2 values) and sand (3 values) particle classes. It would therefore be possible to convert a more detailed distribution of measured particle size data to the SPADE/M database using the RAWPSD instead of the data from the SOIL_PROPS table. This task is made more difficult by the fact that particle sizes were reported in different ways. In most cases the field PCENT.RAWPSD contains the proportion of a particle size for a given diameter specified in the PSIZE.RAWPSD field. The sum of all proportions should then come to 100%. However, there are also 618 cases where the proportion represents an accumulated value to the particle size and/or where the proportion is not given as a percentage but a proportional value from 0 – 1. As a consequence, the data from the RAWPSD table need to be harmonized before they can be used or transferred to another

structure. This was done for the development of the PTFs derived from the HYPRES data and the results are stored in the SOIL_PROS table.

To maintain some of the information on measured particle sizes in the SPADE/M data the number of measured particle sizes was reduced to the maximum number available for clay (1), silt (2) and sand (3).

- ***Clay***

For clay the limit of the particle size was set to 2 μm . A value for this particle size was given for 3,848 records. When clay contents were given also for smaller particle sizes the data were aggregated into a single value. In 2 cases a value for clay content was given only for a 1 μm size. These cases were not included when transferring the data to the SPADE/M structure.

- ***Silt***

For silt the data were aggregated to 20 μm or to 50, 60 and 63 μm . For the larger particle sizes data were reported for only one of the sizes in the table. Therefore, any silt content $> 20 \mu\text{m}$ could be aggregated to the largest particle size for which data exist and lead to no more than 2 grades of silt.

- ***Sand***

For sand the main size limits were determined by the relative frequency of data for a size. The values used were 100 or 125, 200 or 250, 1,000 or 2,000 μm . In one case the particle content is given for a maximum size of 630 μm . This value was not transferred to the SPADE/M database because no corresponding ID for the size is defined.

A particular situation is given by zero (“0”) entries in a field. In the procedure used such entries were interpreted as a measurement at a specified particle size for which no particle content was found, i.e. where a parameter was measured but not present or below the detection limit of the method used. The condition was mainly found for clay and sand fractions, but not for silt. As a consequence, when a value of “0” was given for a sand fraction the values of smaller particle sizes were aggregated to this size. For example, in case a particle content > 0 is given for 1,000 μm and a value of 0 is given for 2,000 μm the final value is the aggregation of particle sizes to 2,000 μm and the method indicates 2,000 μm as the particle size. In case a value was given for a particle size of 1,000 μm but the entry for the 2,000 μm was blank it was assumed that the fine earth fraction was limited to the sum of the particles with a diameter up to 1,000 μm .

With 1,116 occurrences the number of horizons with data on the particle content was restricted compared to the size of the table (3,908 records). A reason leading to the reduction in transferable data was the number of multiple entries for a soil profile. While the SPADE/M data model allows storing multiple profiles for the same plot and could also store more than one measurement of the same parameter for a horizon covering an identical depth interval the latter condition runs against the principle of the

database. The same applies to cases where horizons partially overlay in depth for the same profile. To increase the data on particle size distribution for the HYPRES data the estimated values of the SOIL_PROPS table can be used.

3.3.8 Organic Matter and Carbon (ORG_MAT and ORG_C)

The SOIL_PROPS table contains data on soil organic matter (SOM) instead of soil organic carbon (SOC). A code indicating no data (“-9”) is set for 1,115 entries, while a value of zero (45 entries) or greater (4,400 entries) is given for value entries.

The relative frequency distribution of the OM content in the 4,445 records with data is graphically presented in Figure 3.

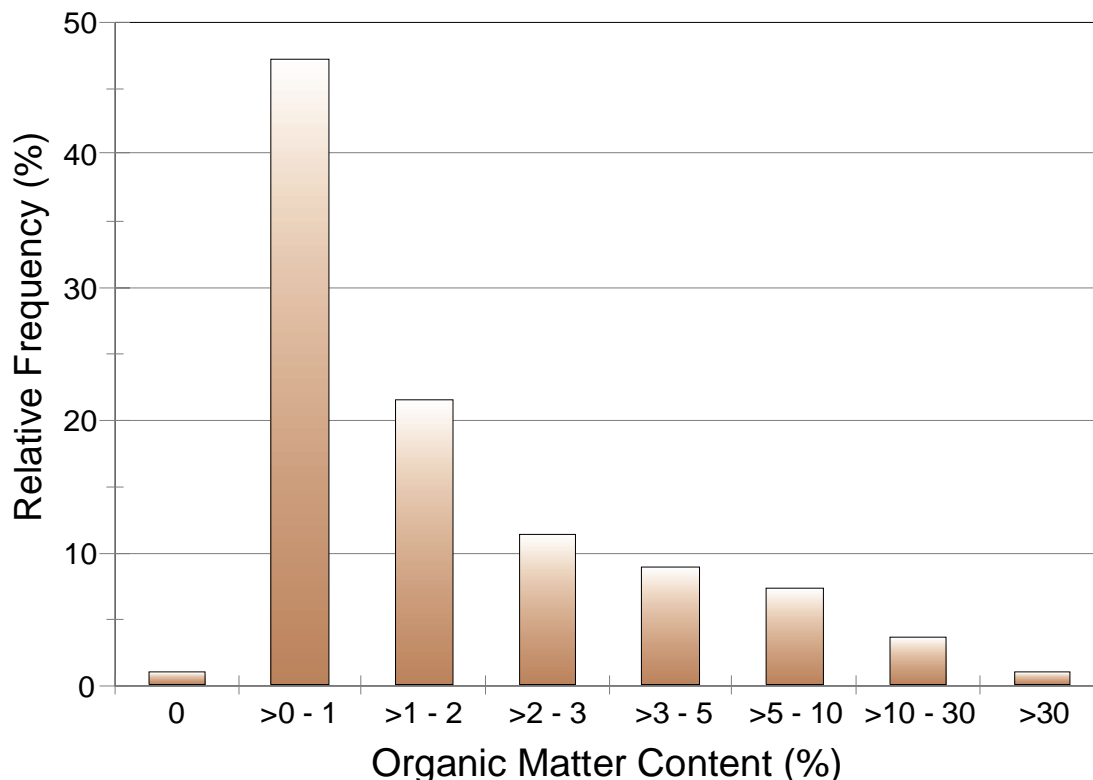


Figure 3: Relative Frequency Distribution of Organic Matter Content in Horizon Records

A value $\leq 1\%$ SOM is given for 2,081 records, between 1% and 2% SOM for 914 records and between 2% and 5% SOM for 835 records. A value $> 30\%$ SOM is given for 37 records. The maximum organic matter content is 90.0%, which is given in 2 cases. In 50% of the records of the table the SOM content is $\leq 1.1\%$. This distribution

shows relatively low SOM contents (corresponding to 0.64% SOC) and reflects the prevalence of profile sites on arable land and from the subsoil.

For 1,986 horizons the information in one of the comment fields provides the value of the coefficient used to convert the measured organic carbon into organic matter (1.72 or 1.724). This information was used to provide data on SOC to be stored in the ORG_C field. From the comments on the method used to measure the soil organic matter or carbon content it emerges that in more cases organic carbon was measured than those for which a coefficient has been given. However, only data where the coefficient was specifically stated were converted to SOC content.

3.3.9 Soil Chemical Properties

The database does not contain information on the chemical soil properties, such as pH, cation exchange capacity or base saturation.

3.3.10 Soil Water Retention (WC_1, WC_2, WC_3, WC_4, WC_FC)

Data on soil water retention are stored in the RAWRET table. The table contains 11,260 different values for the pressure head at which the moisture content was measured. This compares to the 16 distinct values specified in the SPADE/M methods table, of which a combination of 5 (4 free pressure heads and water retention at field capacity) can be used for a given soil profile.

From the frequency distribution of the pressure heads in the RAWRET table pressure heads of 0, 10, 100 and 15,000 cm were used to fill the fields WC_1 to WC_4. The pressure head of the HYPRES data is given in *cm* as unit, while SPADE/M uses *kPa* to describe the unit of the method. For the water retention fields the heads in *cm* may be approximated by using a conversion factor of 0.1⁶. For the water retention at field capacity a suction pressure of -33 *kPa* is widely applied. However, data for this pressure value is given for only 22 cases. In contrast, the measured water retention at a suction pressure of 300 *cm* is given for 1,747 cases. The latter value was therefore used to provide the water retention at field capacity (WC_FC), although the method indicates a value of 33 *kPa*.

All data were extracted from the RAWRET table. For WC_1 (0 *cm* pressure head) the data of the saturated water retention in the field SOIL_PROPS.SATWAT could also have been used. The RAWRET table includes a flag field to indicate whether measurements were made in the laboratory (“*l*”) or the field (“*f*”). The table also contains flag values “*s*”, which is only set for a limited number of standard pressure

⁶ 1 *cm* = 0.0980665 *kPa*

heads (0, 4, 30, 60, 300, 3,000, 6,000 and 15,000 cm). The flag implies that the value came from the SATWAT field in the SOIL_PROPS table.

3.3.11 Total Porosity (POR_TOT)

The measured total porosity is recorded in the SOIL_PROPS.PORISITY field either as a proportion of the total volume or a proportion of the soil fraction with particles sizes $< 2,000 \mu\text{m}$. The information on the volume the values relate to are stored in the comment field and can, therefore, only be extracted by a largely manual procedure.

No data on porosity, indicated by “-9”, is available for 1,927 records. For 9 records a value of -0.09 was given. These records were excluded from being transferred, leaving 3,624 records.

For 464 records a comment field specified that the total porosity value equals the value for the saturated water content (“*porosity=satwat*”). This condition is a generalization and the separate values are usually found in the table. The remark was included in the comments on the horizon (COM_HOR field), but not enforced when processing the data.

3.3.12 Bulk Density (BD)

For bulk density a value not coded by “-9” was given for 4,840 records. Of these, 23 records had entries of zero for the parameter. This is not a viable value for the parameter and the entries were subsequently excluded from being further processed. The maximum value for bulk density was 2.65 g cm^{-3} , which was reported for one horizon. The value indicates solid rock and was not included in the transfer.

The relative distribution of the remaining horizon bulk density values is given in Figure 4.

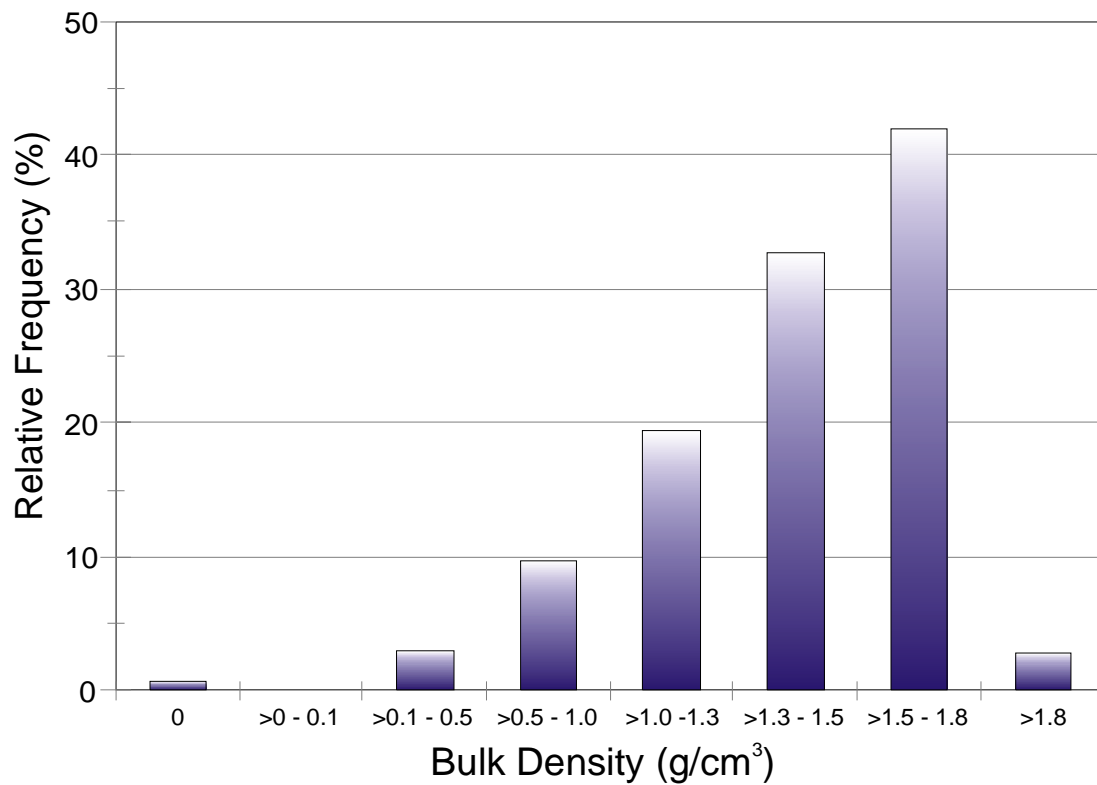


Figure 4: Relative Frequency Distribution of Bulk Density in Horizon Records

The graph shows a strong prevalence of including soil horizons with a bulk density of $> 1.3 \text{ g cm}^{-3}$ (75%). For 42% a bulk density is given of $> 1.5 \text{ g cm}^{-3}$. The distribution with an incline towards higher values of BD is consistent with the low values for SOM/SOC and the prevalence of profile samples from the subsoil.

4 SOIL PROFILE SUITABILITY FOR SPADE/M DATABASE

Suitable HYPRES soil profile data were transferred to the SPADE/M V2 database. The data model is a normalization of the standard model of the relational model using plot-horizon tables (Hiederer, 2010). It is designed for flexibility and on the basis of separating categorical from continuous data.

Qualitative or categorical data are stored in the LIST tables. They can be related to nominal or ordinal scales, although no distinction between the type of scales is made in the database. A finite list of possible values (tuples) for categorical data is defined in the VAL_LIST table for both, plot and horizon data.

Quantitative or continuous numeric data are stored in the RANGE tables. They represent measurements over a range of magnitudes. No distinction is made between measurements on interval or ratio scales. All continuous data are linked to a measurement unit and method, which are defined in the corresponding dictionary tables.

4.1 Conceptual Limitations to Extending Profile Data Set

The SOIL_PROPS table contains 5,560 records and the information on the vertical position and extent of the pedological horizons within the profile. A conceptual complication to transferring the HYPRES horizon data to the SPADE/M data base is caused by the presence of replicate measurement for a given horizon and profile. These horizons have unique names within the profile they belong to, and can thus be referenced unambiguously, but a one-to-many relationship between profiles and horizons is a divergence from the concept of presenting soil profiles in the SPADE/M database.

In SPADE/M a soil profile is characterized by a continuous and non-overlapping vertical sequence of horizons. The profile data may be the result of the analysis of several profiles or horizons, but in this case only the aggregated data are used to characterize the profile and stored in the database. In contrast, in the HYPRES database a soil profile can be characterized by duplicate, overlapping or a non-continuous sequence of horizons. For some of the parameters an aggregated value could be approximated by the mean, e.g. for the distribution of the particle size, but the calculation of a mean is not applicable for parameters expressed as discrete values, e.g. texture. The range of parameters measured or the methods used to establish a parameter may further vary between horizons of the same profile which can render even the aggregation of parameters given as continuous numeric values an inadequate task.

Where horizons are discontinuous or have partial overlap a method of data interpolation needs to be applied. The potential uncertainties and inconsistencies introduced by aggregating data from multiple horizon measurements were found to off-set any advantage of enlarging the basis of measured soil profiles. Therefore, profiles with multiple measurements for the same vertical position within the profile were excluded from being extracted. However, profiles with gaps in the vertical position of horizons were included.

After applying the rules of selecting profiles data from 673 profiles with 1,963 horizons were used. Of the profiles with multiple horizons for identical depth layers 175 profiles with 664 horizons were retained. For multiple horizons the HORIZON field entry was adapted by removing the counter from the specifier, usually a letter in alphabetic order. For completeness the inspection of the data by field was performed on the whole database.

4.2 Enhancements of SPADE/M by HYPRES

Adding soil profiles to the SPADE/M data set should extend the coverage of measured data to areas without profiles. The new profiles should also broaden the range of soil types with profile data and increase the number of measurements for any given soil type.

4.3 Geographic Coverage of Profiles

The geographic coverage of the profiles in SPADE/M and the profiles added to the data set from HYPRES is presented in Figure 5.

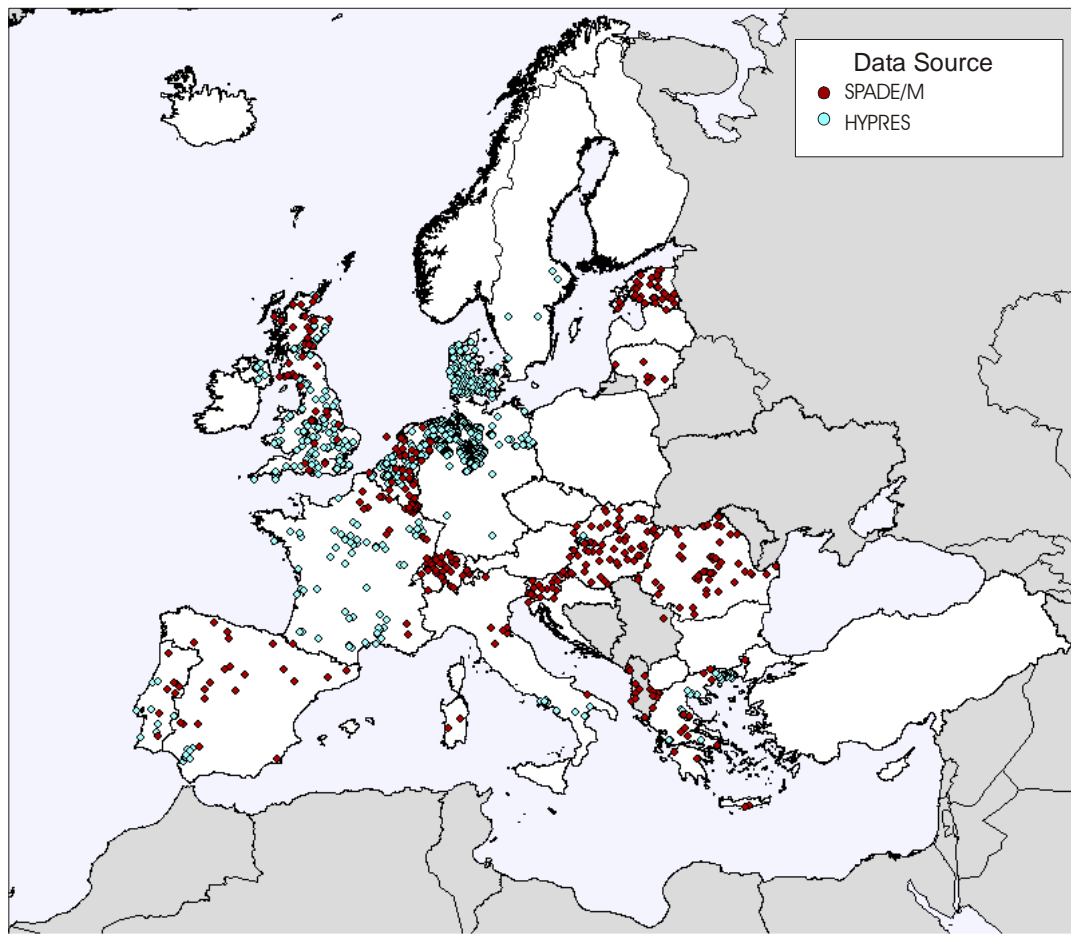


Figure 5: Geographic Distribution of SPADE/M and Supplementary HYPRES Profiles

The graph shows that data from HYPRES add profiles to regions previously not covered, notably Denmark and Germany, but also an increase in the density of profile data in other countries, such as Portugal, Spain and the United Kingdom.

The larger coverage of European regions with measured profile data could improve the definition of the conditions in a PTR although the PTRs of the ESDB do not explicitly include a geographic stratification. The increase in coverage should allow an improved appreciation of typical characteristics of a soil type under a wider range of conditions. The appreciation of typical characteristics is simplified when the profiles are taken at a distance at which the properties are no longer auto-correlated. This minimum distance varies depending on site conditions. In the field the distance can be anisotropic, i.e. depend on the direction.

For the SPADE/M, the HYPRES and the combined data sets the distance to the nearest plot was determined. The result aggregated to classes is presented in Table 7.

Table 7: Distance to Nearest Profile for SPADE/M, HYPRES and Combined Data Sets

Distance <i>m</i>	Data Set		
	<i>SPADE/ M</i>	<i>HYPRES</i>	<i>Combined Profiles</i>
≤ 1	24	28	52
1 - 10	0	82	92
10 - 100	0	396	423
100 – 1,000	2	469	473
1,000 – 10,000	98	475	585
> 10,000	346	198	493

The coordinates of the SPADE/M profiles indicate a distance between profiles of 1 m or less for 24 sites. The number is caused by 12 pairs of plots with identical coordinates in Estonia (1 pair), Romania (1 pair), Slovenia (1 pair) and the UK (9 pairs). In 8 cases the duplicates are for the same soil type and in the other 4 cases for similar soil types. It would appear that these duplicates are replicate measurements of profiles taken at a site. The large majority of SPADE/M profiles (346) are located at distances of more than 10 km.

The distances between the nearest HYPRES profiles are very differently distributed. The coordinates show 28 sites with a distance of just 1 m. About 60% of the HYPRES sites for profiles are at a distance of 1,000 m or less to the nearest other site. This figure is 50% for the combined data set. When compared to the mapping detail of the SGDBE, which can be reasonably mapped to 1 km grid resolution, a distance of less than 1,000 m indicates multiple sites per grid cell and potentially a significant degree of auto-correlation of the data.

The plot density in terms of the number of plots on a 1 km raster grid is presented in Figure 6.

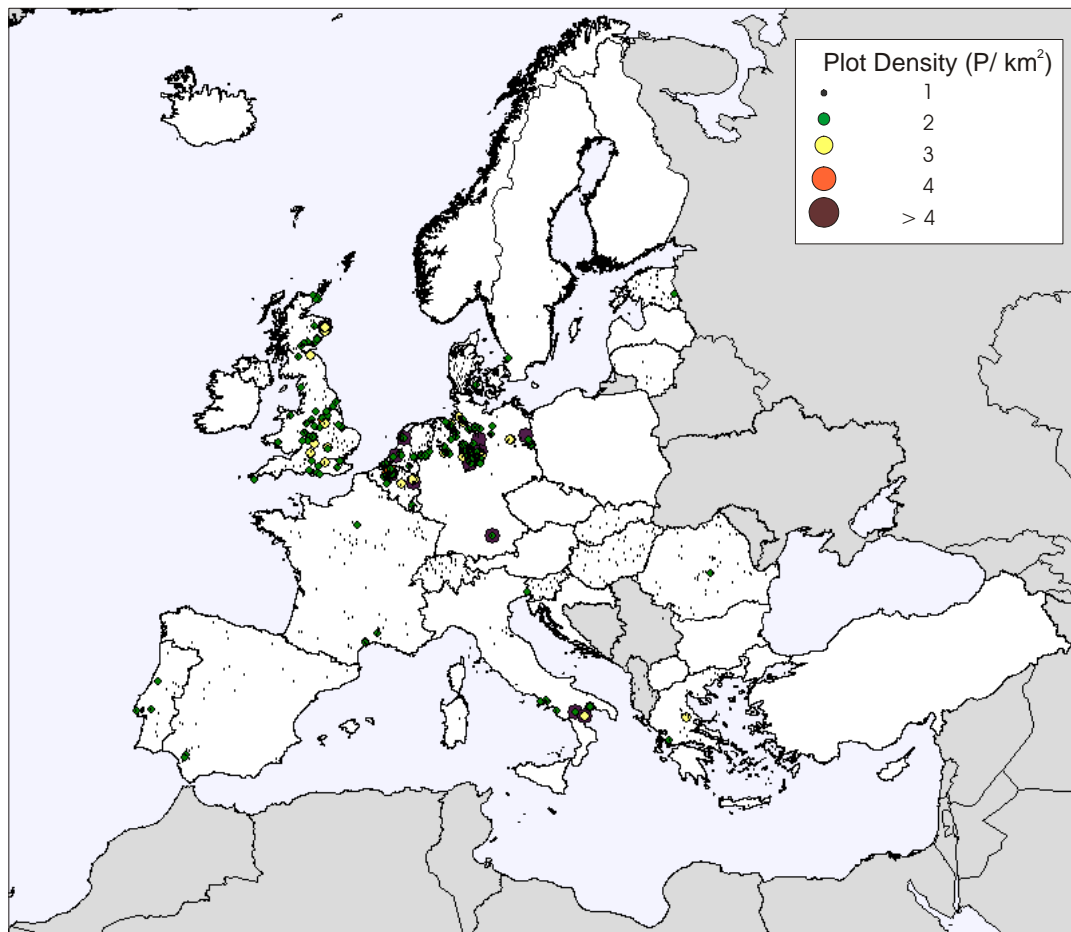


Figure 6: Plot Density for Combined SPADE/M and HYPRES Profile Sites on 1 km Grid

The graph shows that the close proximity of plots is largely restricted to specific regions. Most grid cells with 2 or more sites are found in Belgium, Germany, The Netherlands and the UK. The possibility of the profile characteristics being related should be considered when analyzing the data. That a high number of plots per area does not necessarily result in close proximity is demonstrated in countries such as Denmark or other countries covered by SPADE/M data.

4.4 Range of FAO Soil Types

Adding profiles to the SPADE/M data set should also increase the representation of soil types by providing representative profile data for a soil type or a sufficient number of profiles for a given soil type to cover the range of values typical for a soil type. This would support to refine the conditional rules of a PTR.

Whether the available profile data are adequate to define the conditions for a PTR or parameters of a PTF depends on the soil characteristic to be modelled. As an approximation one may assume that soils with a wider geographic distribution may also be found under a wider range of climatic conditions and land use/cover types. Therefore, more profiles would be needed to characterize the soil than for a soil with very limited spread.

Most PTRs of the SGDBE include the soil type code according to the FAO85 classification scheme as a rule parameter. The relative occurrence of a FAO85 soil type in a map derived from the SGDBE was compared to the proportion of the soil type in the profile data sets. For the analysis the *soil mapping units* (SMUs) of the spatial layer of the SGDBE was mapped to a 1 km grid raster using the ERTS89 LAEA projection. The area evaluated was restricted to the countries covered by the map on topsoil organic carbon (Europe without Cyprus, Iceland, Belarus, Ukraine, Moldova, Russian Federation and European part of Turkey).

A comparison of the relative occurrence of FAO85 Level 1 soil type codes in the data sets is presented in Figure 7.

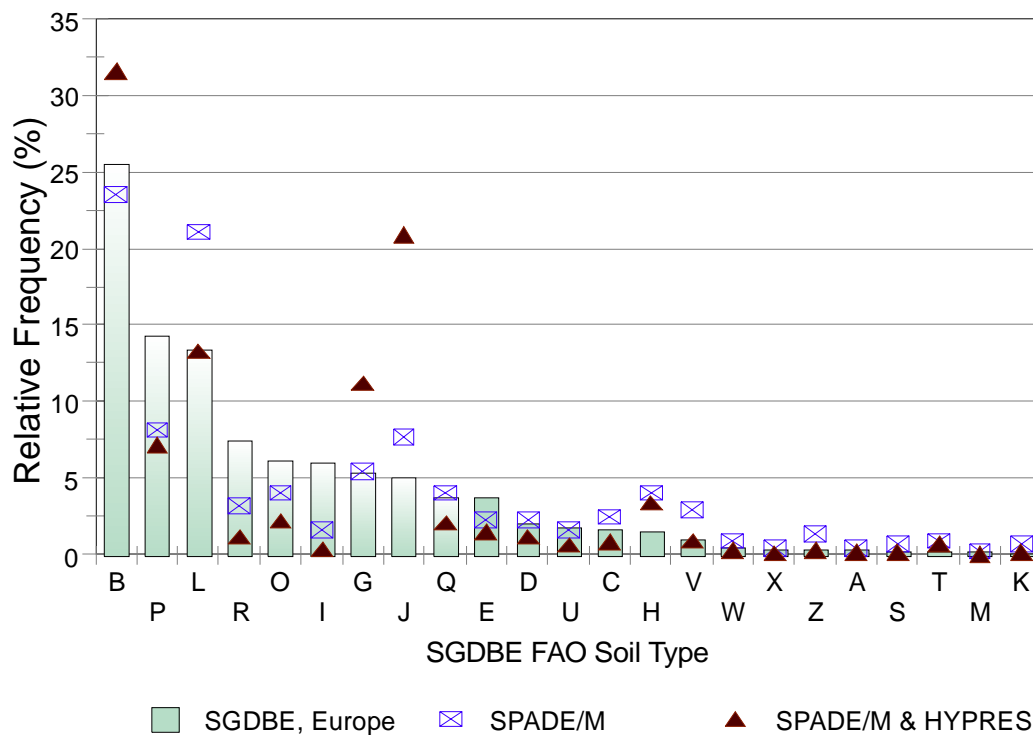


Figure 7: Relative Distribution of FAO85 Level 1 Soil Types for SGDBE (Europe), SPADE/M and HYPRES Profile Data

With over 25% the most widespread soils are *Cambisols* (B), followed by *Podzols* (P) with 14% coverage. Together with *Luvissols* (L) these types account for more than 50% of all soils in Europe.

The profile data from the SPADE/M and the combined SPADE/M & HYPRES data sets are depicted as the relative frequency of a profile belonging to the soil type. The relative frequency of the profile soil types of the SPDAE/M data set roughly follows the distribution of the soils in the spatial data. Including the HYPRES profiles the representation of Level 1 soil types in the data set is affected by the strong presence of profiles of *Gelysols* (*G*) and *Fluvisols* (*J*). While the SPADE/M profiles cover all Level 1 soil with profiles, no profiles in the HYPRES data are available for *Lithosols* (*I*), *Kastanozem* (*K*), *Greyzem* (*M*), *Solonetz* (*S*), *Xerosols* (*X*) and *Solonchak* (*Z*). These soils account for 6.9% of the soils in the analysis area of the SGDBE, although only *Lithosols* (5.9%) occur with a relative distribution > 1%.

PTRs may make use of more detailed levels of soil type codes, such as PTR21 for estimating SOC. Therefore, the distribution of soil types in the SGDBE with the relative frequency in the profile data sets was therefore compared for FAO85 Level 3 codes. The results are presented in Figure 8.

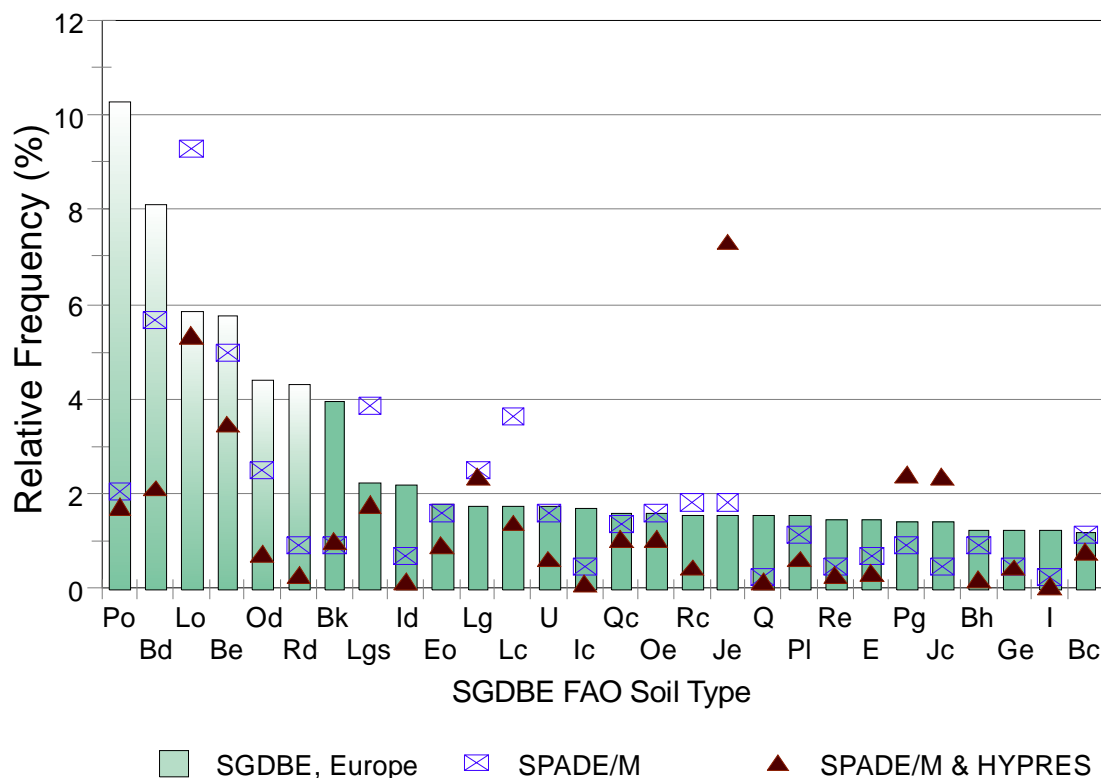


Figure 8: Relative Distribution of FAO85 Level 3 Soil Types for SGDBE (Europe), SPADE/M and HYPRES Profile Data

The graph shows FAO85 Level 3 soil types with a relative coverage of 1% or more in the area analyzed. The most frequently occurring Level 3 soil type in the area analyzed is the *orthic Podzol* (*Po*, 10%), followed by *dystric Cambisols* (*Bd*, 8%) and *orthic Luvisols* (*Lo*, 6%).

The general relationship between the distribution of Level 3 soil types in the map data and the relative frequency of the soil types in the profile data sets is not significantly different from Level 1 soil types. The relative frequency of profile soil types is very much influenced by number of profiles with a *eutric Fluvisol* (*Je*, 7.3%), which is found for 144 profiles in the HYPRES data set and which covers 1.6% of the mapped area.

The positions of the measured profile data on the map of *dystric Cambisols* (*Bd*) and *eutric Fluvisols* (*Je*) is depicted in Figure 9.

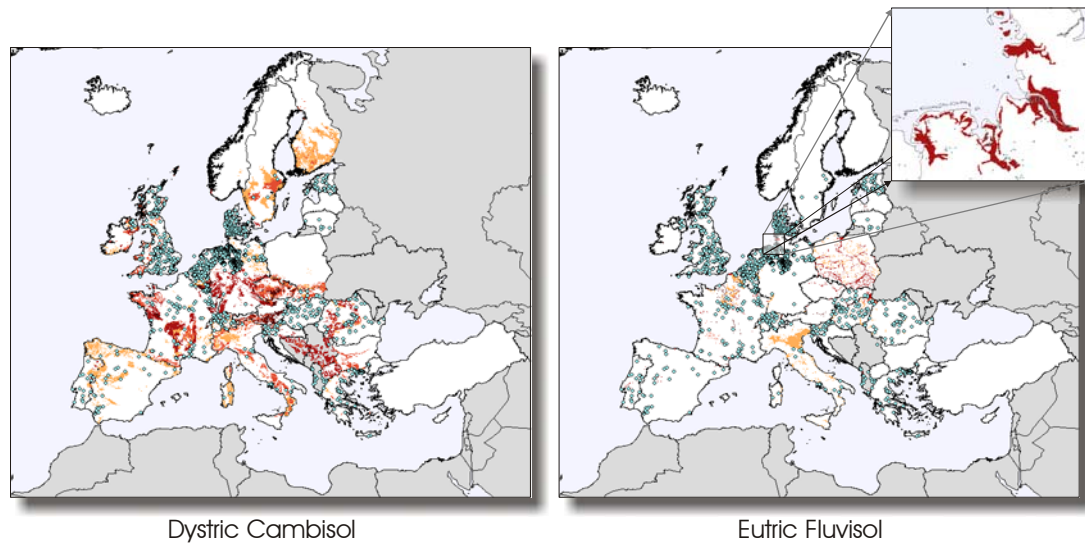


Figure 9: Diffusion of SPADE/M and HYPRES Profiles for Dystric Cambisol and Eurtric Fluvisol on Distribution of the Soil Types in SGDBE

The spread of the *dystric Cambisols* is largely covered by profile data, although not by data from northern Italy, Austria, Czech Republic and Finland. In contrast, the much more limited extent of areas with *eutric Fluvisols* are dominated by a dense number of profiles from sites in northern Germany (see inset). For these profiles in close proximity the aggregation of the characteristics into fewer profiles with a wider distance between the sites could be more suitable to the task of supporting the definition of PTRs from the measured data.

5 CONCLUSIONS

The investigation into the possibility of using HYPRES soil profile data to extend the geographic and thematic range of the SPADE/M data set has been more complex than anticipated. Both data sets containing measurements of properties describing soil profiles and a comparable arrangement of the data in form of two tables forming a plot-profile relationship, but the principles underlying the data stored in the tables are very different between the databases. The SPADE/M data were assembled with the intention of supporting the definition of PTRs to attributes not included in the list of observed or measured parameters of the SGDBE. The profiles included in the HYPRES data set were assembled with the aim of defining the parameters of hydrological PTFs. The range of data stored in the data sets shows partial overlap. The HYPRES data are more detailed in representing soil horizons by particle size and conductivity, but lack the results from the chemical analysis. SPADE/M profiles use a wider range of attributes, including the results of the soil chemical analysis, but less detail is stored on some physical properties and none on conductivity.

The differences in the concepts between SPADE/M and HYPRES of the profile data included in the data sets limits the number of profiles that could be profitably used to extend the geographic and thematic range of the profiles of the SPADE/M data set. Profiles with overlapping or duplicate measurements were excluded from further evaluation. Data from profiles in close geographic proximity could be aggregated to represent the plot area as a single profile. Selecting the profiles to be included to represent the plot and the aggregation method to be used to define the typical horizons, their depths and characteristics is a task which cannot be fully automated and beyond the scope of this evaluation.

To improve the level of integration between the data sets the information reported in free text form on parameters used to define PTRs, such as parent material, soil type and land use/cover, were transferred to comply with the classification schemes of the PTR database of the ESDB. This standardization of the data to a common classification simplifies testing the PTRs of the ESDB with measured profile data.

Acknowledgements

The author would like to thank Allan Lilly (James Hutton Institute, Aberdeen) for kindly providing the HYPRES data used in this evaluation and his detailed comments on the many data issues. Attila Nemes (University of Maryland, Beltsville) provided critical explanations on the organization of the tables and the changes to the fields and additional tables. On the content of the data and the interpretation of their use the author is indebted to Henk Wösten (Alterra, Wageningen) for his patience, availability and always practical advice. Many thanks also to Marc van Liedekerke (European Commission JRC, Ispra) for his great support in adapting the text to make it accessible to a wider audience.

References

- Annoni, A., C. Luzet, E. Gubler and J. Ihnde (2001) *Map Projections for Europe*. European Commission Joint Research Centre, Ispra, Italy. EUR 20120 EN. 131pp.
- European Commission Joint Research Centre (2003) *The European Soil Database V. 2.0*. On-line distribution through:
http://eusoils.jrc.ec.europa.eu/ESDB_Archive/ESDB/index.htm
- Hiederer, R., E. Michéli and T. Durrant (2011) *Evaluation of BioSoil Demonstration Project - Soil Data Analysis*. EUR 24729 EN. Publications Office of the European Union. 155pp.
- Hiederer, R. (2010) *Data Update and Model Revision for Soil Profile Analytical Database of Europe of Measured Parameters (SPADE/M2)*. EUR 24333 EN. Luxembourg: Office for Official Publications of the European Communities. 55pp.
- Hiederer, R., R.J.A. Jones and J. Daroussin (2006) Soil Profile Analytical Database for Europe (SPADE): Reconstruction and Validation of the Measured Data (SPADE/M). *Geografisk Tidsskrift, Danish Journal of Geography* 106(1). p. 71-85.
- Ishtiaq A., F.D. Whisler, J. Iqbal, J.N. Jenkins and J.M. Mckinion (2004) Soil Physical Properties Web Database for GOSSYM and GLYCIM Crop Simulation Models. *Agron. J.* 96:pp1706–1710.
- Leij, F.J., W.J. Alves, M.Th. van Genuchten and J.R. Williams. 1994. *Unsaturated Soil hydraulic Database, UNSODA 1.0 user's manual*. Office of Research and Development. United States Environmental Protection Agency. Ada. Oklahoma
- Wösten, J.H.M., A. Lilly, A. Nemes and C. Le Bas (1999) Development and use of database of hydraulic properties of European soils. *Geoderma* 90. pp169–185.
- Wösten, J.H.M., A. Lilly, A. Nemes and C. Le Bas (1998) *Using existing soil data to derive hydraulic parameters for simulation models in environmental studies and in land use planning*. Final report on the European Union funded project. Report 156, DLO Winand Staring Centre, Wageningen, The Netherlands. 106pp.

European Commission

EUR 24971 EN – Joint Research Centre – Institute for Environment and Sustainability

Title: Extending Geographic and Thematic Range of SPADE/M with HYPRES Soil Profile Data

Author(s): R. Hiederer

Publications Office of the European Union

2011 – 43 pp. – 21.0 x 29.7 cm

EUR – Scientific and Technical Research series – ISSN 1018-5593 (print) ISSN 1831-9424 (online)

ISBN 978-92-79-21592-6 (print)

ISBN 978-92-79-21593-3 (online)

doi: 10.2788/8209

Abstract

The measured soil profile data of the *Hydraulic Properties of European Soils* (HYPRES) were evaluated for their potential use in extending the geographic coverage and thematic range of the profiles of the *Soil Profile Analytical Database of Europe of measured profiles* (SPADE/M) database. The aim of increasing the number of measured profiles is to improve the definition of pedo-transfer rules (PTRs) to extend the range of parameters characterizing soils and the validation of model runs.

The HYPRES and SPADE/M databases follow different concepts in the compilation of soil profiles. These differences were reflected in the organization of storing profile data in the databases. A specific conceptual problem to extending the SPADE/M data set is posed by recording repeated measurements in the HYPRES database for horizons of a profile and multiple profiles for a plot. The two data sets also differ with respect to the properties recorded, the measurement units and the database model.

To increase the number of measured soil profile data of the SPADE/M data set with HYPRES profiles the data from the latter has to be standardized to be adjusted to comply with the specifications of the SPADE/M data set. The standardization process involves conversions of units or reference systems, such as the plot co-ordinate transformation or the extraction of properties from comment fields, but also conceptual adaptations of the method used to characterize a soil horizon in the database. The outcome of the standardization process is a series of soil profiles which can be seamlessly added to the SPADE/M data set.

How to obtain EU publications

Our priced publications are available from EU Bookshop (<http://bookshop.europa.eu>), where you can place an order with the sales agent of your choice.

The Publications Office has a worldwide network of sales agents. You can obtain their contact details by sending a fax to (352) 29 29-42758.

The mission of the JRC is to provide customer-driven scientific and technical support for the conception, development, implementation and monitoring of EU policies. As a service of the European Commission, the JRC functions as a reference centre of science and technology for the Union. Close to the policy-making process, it serves the common interest of the Member States, while being independent of special interests, whether private or national.



LB-NA-24971-EN-N



ISBN 978-92-79-21593-3



9 789279 215933