IGBP-DIS SOIL DATA SET FOR PEDOTRANSFER
FUNCTION DEVELOPMENT

P. Tempel, N.H. Batjes and V.W.P. van Engelen
October 1996
Disclaimer

While every effort has been made to ensure that the data are accurate and reliable, ISRIC cannot assume liability for damages caused by in-accuracies in the data or as a result of the failure of the data to function on a particular system. ISRIC provides no warranty, expressed or implied, nor does an authorized distribution of the data set constitute such a warranty. ISRIC reserves the right to modify any information in this document and related data sets without notice.

Correct citation:


Inquiries:

Director
P.O. Box 353
6700 AJ Wageningen
The Netherlands
Telefax: +31-317-471700
E-mail: SOIL@ISRIC.NL
IGBP-DIS SOIL DATA SET FOR PEDOTRANSFER
FUNCTION DEVELOPMENT

P. Tempel, N.H. Batjes and V.W.P. van Engelen
(October 1996)

INTERNATIONAL SOIL REFERENCE AND INFORMATION CENTRE
Table of Contents

Abstract ........................................................................................................................................ ii

1 Introduction ................................................................................................................................. 1

2 Methodology ............................................................................................................................... 1
   2.1 Source of data ......................................................................................................................... 1
   2.2 List of attributes .................................................................................................................... 2
   2.3 Functional grouping of soil samples ....................................................................................... 2

3 Content of PTF data set .............................................................................................................. 3

4 Conclusions .................................................................................................................................. 3

Acknowledgements ....................................................................................................................... 4

References ....................................................................................................................................... 4

Appendices
Appendix 1 Structure of PTF data set (6)
Appendix 2 Main assumptions and defects (6)
Appendix 3 Characterization of data set (6)

Enclosure
Diskettes with data set

Abstract

At the request of the Global Soil Data Task (GSDT) of the Data and Information System of the International Geosphere Biosphere Programme (IGBP-DIS), ISRIC prepared a uniform soil data set for the development of pedotransfer functions. The necessary chemical and physical soil data have been derived from ISRIC’s Soil Information System (ISIS) and the CD-ROM of the Natural Resources Conservation Service (USDA-NRCS). All soil samples were clustered into functional groups based on soil textural class and (calculated) activity of the clay minerals (9 classes), while samples from organic soils and allophanic soils were flagged. The digital set contains data for 131,472 samples, originating from 20,920 profiles. It is presented as a comma delimited ASCII-file and in dBASE IV format.

GSDT will use the data set to develop a number of pedotransfer functions for often required, yet seldom measured soil properties. Interaction with the global soil and modelling community led to the identification of four soil properties as being especially important and urgently needed: soil organic carbon, soil total nitrogen, water holding capacity, and soil thermal properties.

Keywords: digital data sets; soil properties; pedotransfer functions; global change
1 Introduction

The Global Soil Data Task of IGBP-DIS is an international collaborative project with the objective of making accurate and appropriate data relating to soil properties accessible to the global change research community. The main collaborators are developers of major international soil data sets, such as the United States Department of Agriculture (USDA-NRCS), the Food and Agriculture Organization (FAO) of the United Nations, and the International Soil Reference and Information Centre (ISRIC), as well as individual soil scientists from a wide range of institutes. The tasks foreseen by the Global Soil Data Task (GSDT) include: building a global pedon database; developing methods to convert fundamental soil analytical properties into derived soil properties; and generating global gridded databases of derived soil properties (see Scholes et al., 1995).

Interaction of GSDT with the global soils and modelling community led to the identification of four soil properties as being especially important and urgently needed: soil organic carbon, soil total nitrogen, water holding capacity, and soil thermal properties. For this work, a uniform data set of measured soil properties is needed that can be used to derive a number of inferred soil properties where measured data are lacking. A classical method to cope with this lack of measured data has been to develop mathematical relationships, termed pedotransfer functions or PTFs (Bouma et al., 1986). For example, the prediction of $\theta_i$ from easily measured and widely available soil properties such as particle-size distribution and organic matter content (Saxton et al., 1986; Hudson, 1994; Batjes, 1996).

The current report presents a data set of soil properties that can be used for the development of pedotransfer functions, including soil moisture retention and soil thermal properties. The set was developed as a sequel to the workshop of the GSDT at Montpellier during which the minimum list of attributes was elaborated (Scholes, 1994).

2 Methodology

2.1 Source of data

The PTF data set includes profiles from ISRIC’s Soil Information System (ISIS) and from the CD-ROM of the Natural Resources Conservation Service (USDA-NRCS). Since both organisations use similar analytical procedures and produce comparable results (L.P. van Reeuwijk and J. Kimble, pers. comm.), records from the two data sets can be merged into one single file, thereby reducing the critical, yet often overlooked, issue of data comparability between disparate databases (Vogel, 1994; Pleijisjer, 1986). The various analytical methods used are coded in the data set; for a full explanation reference is made to the relevant laboratory manuals (Van Reeuwijk, 1993; Soil Survey Staff, 1996).

2.2 List of attributes
The attributes considered in the PTF data set include identifiers for profile, horizon type, upper and lower depth of the sample, soil classification (FAO-Unesco, USDA Soil Taxonomy, and version when known). Physical attributes, included when available, are: soil Munsell colour, percentage of fragments >2 mm, USDA particle size distribution, bulk density and water retention (at defined suctions). With respect to the chemical attributes the following have been included when available: organic carbon content, total nitrogen, soil reaction (pH-H₂O and pH-KCl), cation exchange capacity (at pH 7 and pH 8.2), effective CEC (in 1 M KCl), exchangeable acidity, exchangeable bases (Ca²⁺, Mg²⁺, K⁺ and Na⁺) and electrical conductivity (ECe). Contrary to the European Data Set on Soil Hydraulic Properties (SC-DLO, 1994), the PTF data set developed for IGBP-DIS does not include data on soil hydraulic conductivity, as these are not routinely measured in the ISRIC and USDA-NRCS laboratories.

All attributes considered in the PTF data set (and their units of measurement) are listed in Appendix 1.

The "base codes" of a number of attributes in the source databases are listed in Appendix 2. In transferring the source data to the uniform target data set, the source materials had to be taken at face value. The main assumptions underlying the compilation of the PTF data set as well as its main defects are also summarized in Appendix 2.

### 2.3 Functional grouping of soil samples

As recommended by the Montpellier workshop, all (horizon) data sets have been functionally grouped according to their textural class and the inferred activity of their clay minerals. In addition, samples originating from organic soils (i.e. Histosols) and allophanic soils (i.e. Andosols) were grouped as separate categories in view of their specific soil physical and chemical properties. The clustering procedure is schematically depicted in Figure 1.
Coarse  Medium  Fine

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>LAC</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CEC&lt;20 cmol·kg⁻¹</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MIX</td>
<td></td>
<td></td>
</tr>
<tr>
<td>20&lt; CEC&lt;62 cmol·kg⁻¹</td>
<td></td>
<td></td>
</tr>
<tr>
<td>HAC</td>
<td></td>
<td></td>
</tr>
<tr>
<td>62&lt; CEC cmol·kg⁻¹</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Organic Soils</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(OC&gt;16%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Allophanic Soils</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* CEC is cation exchange capacity, measured at pH7 in NH₄OAc; LAC stands for Low Activity Clay; MIX for Mixed Activity Clay; HAC for High Activity Clay. Activity, or CEC, of clay minerals is calculated by assuming a mean CEC of 350 cmol·kg⁻¹ for Organic Carbon (OC).

Textural classes are according to FAO (1974): coarse textured: sands, loamy sands and sandy loam with less than 18% clay, and more than 65 percent sand; medium textured: sandy loams, sandy clay loams, silt loams, silt, silty clay loams and clay loams with less than 35 percent clay and less than 65 percent sand; the sand fraction may be as high as 82 percent if a minimum of 18 percent of clay is present; fine textured: clay, silty clays, sandy clays, clay loams and silty clay loams with more than 35 percent clay.

Figure 1. Schematic representation of the procedure for clustering horizon samples in the PTF data set.

3 Content of PTF data set

The data set holds 715 profile descriptions from the ISIS and 20,205 from the NRCS-USDA source databases. This corresponds with 131,472 samples in total. The distribution over the various functional groupings is shown in figure 2.

Since all the attributes selected for the PTF data set were not necessarily measured for all samples in both source databases, there will be a number of gaps in the target database. Appendix 3 shows to which extent the records for the various attributes have been filled in the target database.
<table>
<thead>
<tr>
<th>Group</th>
<th>Nil 20,066 (15.3%)</th>
<th>Coarse 22,333 (17.0%)</th>
<th>Medium 65,131 (50.0%)</th>
<th>Fine 23,942 (18.2%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>LAC</td>
<td>60,077 (45.7%)</td>
<td>I 16,430 (12.5%)</td>
<td>II 35,063 (26.7%)</td>
<td>III 6,140 (4.7%)</td>
</tr>
<tr>
<td>MIX</td>
<td>28,743 (21.9%)</td>
<td>IV 847 (0.6%)</td>
<td>V 13,676 (10.4%)</td>
<td>VI 13,149 (10.0%)</td>
</tr>
<tr>
<td>HAC</td>
<td>660 (0.5%)</td>
<td>VII 49 (&lt;0.1%)</td>
<td>VIII 173 (0.1%)</td>
<td>IX 300 (0.2%)</td>
</tr>
<tr>
<td>Organic Soils</td>
<td>1,259 (1.0%)</td>
<td>X 1,259 (1.0%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Allophanic soils</td>
<td>1,510 (1.2%)</td>
<td>XI 1,510 (1.2%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NIL</td>
<td>39,223 (29.8%)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* 15.3% of the samples could not be grouped in one of the texture classes and for 29.8% of the records no grouping according to clay mineralogy could be made.

Figure 2. Distribution of the records over the functional groupings.

4 Conclusions

Being based on available (measured) data, the current data set inevitably includes some gaps. Another type of gaps is associated with the inadequate documentation of some coding protocols in the source databases, for example the uncertainty about which version of USDA Soil Taxonomy has been used. Despite these gaps, the data set will provide a useful basis for development of a wide range of pedotransfer functions for global land use planning and global change modelling purposes.

The current data set for PTF development complements other soil databases that were developed for similar applications, albeit using different data sources and data format (Madsen and Jones, 1995; Wösten et al., 1995; Batjes, 1995).

Acknowledgements

The Pedotransfer Function Development Data Set was developed at the request of, and with sponsorship by, the Data and Information System of the International Geosphere Biosphere Programme (IGBP-DIS).

ISRIC Working Paper 96/05
References


Appendix 1  Structure of PTF data set

The PTF data set is a Delimited Format ASCII file. Data is written character by character starting on the left. Each record ends with a carriage return and line feed. A comma separates each field and, in addition, double quotation marks surround character data.

The first two records in the data set are header lines, data starts in record three. Record one lists the names of all data fields as they appear in the data set. Record two consists of blocks of dashes ('-') that are of the same length as the field names.

The code in between brackets following the field name denotes the format of the field: type (Numeric or Character), width, and number of decimal places (for a numeric field). For example, C6 denotes a character field that is 6 positions wide, N5.2 denotes a numeric field that is 5 positions wide (including the decimal point) with two decimal places.

1. **RECNUM** (N6): The records in the dataset are sequentially numbered from 1 for the first record to 131,472 for the last record. RECNUM uniquely identifies each record in the dataset.

2. **DIS_PID** (C8): The DIS Pedon Identification code uniquely identifies each pedon in the dataset. A DIS_PID code is 8 characters wide and always starts with the three-character string "DIS", followed by a five-digit sequence number (e.g. "DIS00123").

3. **SOURCE_DB** (C5): The identification code for the originator of the pedon data. Valid codes are
   - "ISRIC" for the International Soil Reference and Information Centre, Wageningen, Netherlands, and
   - "SCS" for the national Soil Conservation Service, Lincoln (NE), United States.

4. **SOURCE_PID** (C7): The original Pedon Identification code.
   An SCS Pedon Identification number consists of the year in which the pedon was sampled multiplied by 10,000, to which the National Soil Survey Laboratory number is added.
   An ISRIC reference soil identification code is five characters wide and always starts with a two-character ISO 3166 country code, followed by a three-digit sequence number.

5. **DATE** (C4): The month (first two digits) and year (last two digits) in which the pedon was sampled (ISRIC) or completed (SCS).

6. **ISO_CODE** (C2): A two-character ISO 3166 country code for the country where the pedon is located.


8. **FAO74_DESC** (C20): FAO-Unesco 1974 soil unit name.


15. **TX_NAME** (C50): Soil Taxonomy (USDA/SCS) full classification name down to subgroup level, if not specified otherwise.

16. **TX_YEAR** (N4): The year of publication of the version of Soil Taxonomy...
<table>
<thead>
<tr>
<th>Code</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>TXTEX_CODE (C3)</td>
<td>Soil Taxonomy (USDA/SCS) texture code</td>
</tr>
<tr>
<td>TXTEX_DESC (C50)</td>
<td>Soil Taxonomy (USDA/SCS) texture description.</td>
</tr>
<tr>
<td>TXMIN_CODE (C2)</td>
<td>Soil Taxonomy (USDA/SCS) mineralogy code.</td>
</tr>
<tr>
<td>TXMIN_DESC (C25)</td>
<td>Soil Taxonomy (USDA/SCS) mineralogy description.</td>
</tr>
<tr>
<td>TOP_DEPTH (N3)</td>
<td>Depth of the top of the sample (cm).</td>
</tr>
<tr>
<td>BOT_DEPTH (N3)</td>
<td>Depth of the bottom of the sample (cm).</td>
</tr>
<tr>
<td>DESIGNTION (C8)</td>
<td>Horizon designation.</td>
</tr>
<tr>
<td>CLAY (N4.1)</td>
<td>Weight% of particles &lt; 0.002mm (clay) in fine earth fraction.</td>
</tr>
<tr>
<td>SILT (N4.1)</td>
<td>Weight% of particles 0.05-0.002 mm (silt) in fine earth fraction. SILT is equal to the sum of FSILT and CSILT when available.</td>
</tr>
<tr>
<td>FSILT (N4.1)</td>
<td>Weight% of particles 0.02-0.002 mm (fine silt) in fine earth fraction.</td>
</tr>
<tr>
<td>CSILT (N4.1)</td>
<td>Weight% of particles 0.05-0.02 mm (coarse silt) in fine earth fraction.</td>
</tr>
<tr>
<td>SAND (N4.1)</td>
<td>Weight% of particles 2.0-0.05 mm (sand) in fine earth fraction. SAND is equal to the sum of VFSAND, FSAND, MSAND, CSAND and VCSAND when available.</td>
</tr>
<tr>
<td>VFSAND (N4.1)</td>
<td>Weight% of particles 0.1-0.05 mm (very fine sand) in fine earth fraction.</td>
</tr>
<tr>
<td>FSAND (N4.1)</td>
<td>Weight% of particles 0.25-0.1 mm (fine sand) in fine earth fraction.</td>
</tr>
<tr>
<td>MSAND (N4.1)</td>
<td>Weight% of particles 0.5-0.25 mm (medium sand) in fine earth fraction.</td>
</tr>
<tr>
<td>CSAND (N4.1)</td>
<td>Weight% of particles 1.0-0.5 mm (coarse sand) in fine earth fraction.</td>
</tr>
<tr>
<td>VCSAND (N4.1)</td>
<td>Weight% of particles 2.0-1.0 mm (very coarse sand) in fine earth fraction.</td>
</tr>
<tr>
<td>MINFRAG (N5.1)</td>
<td>Weight% (whole soil) of mineral fragments &gt; 2 mm.</td>
</tr>
<tr>
<td>OC (N5.2)</td>
<td>Weight% of total organic carbon (Walkley-Black method).</td>
</tr>
<tr>
<td>N (N6.2)</td>
<td>Weight% of total nitrogen (Kjeldahl method).</td>
</tr>
<tr>
<td>CACO3 (N5.1)</td>
<td>Weight% of free CaCO3 in fine earth fraction.</td>
</tr>
<tr>
<td>CAX (N5.1)</td>
<td>Exchangeable Ca in cmol(+) kg(^{-1}).</td>
</tr>
<tr>
<td>MGX (N5.1)</td>
<td>Exchangeable Mg in cmol(+) kg(^{-1}).</td>
</tr>
<tr>
<td>NAX (N5.1)</td>
<td>Exchangeable Na in cmol(+) kg(^{-1}).</td>
</tr>
<tr>
<td>KX (N5.1)</td>
<td>Exchangeable K in cmol(+) kg(^{-1}).</td>
</tr>
<tr>
<td>SUMCAT (N5.1)</td>
<td>Sum of cations in cmol(+)) kg(^{-1}).</td>
</tr>
<tr>
<td>ALX (N5.1)</td>
<td>Exchangeable Al (1M KCl) in cmol(+) kg(^{-1}).</td>
</tr>
<tr>
<td>CEC7 (N5.1)</td>
<td>Cation Exchange Capacity of the soil at pH 7 in cmol(+) kg(^{-1}).</td>
</tr>
<tr>
<td>CECBAACL2 (N5.1)</td>
<td>Cation Exchange Capacity of the soil - BaCl2, pH 8.2 - in cmol(+) kg(^{-1}) (SCS method code 5A5).</td>
</tr>
<tr>
<td>ECEC (N6.2)</td>
<td>Effective CEC in cmol(+) kg(^{-1}).</td>
</tr>
<tr>
<td>ACIDX (N5.1)</td>
<td>NaOAc. extractable acidity at pH 8.2, in cmol(+) kg(^{-1}).</td>
</tr>
<tr>
<td>KCL_ACID (N5.1)</td>
<td>Extractable acidity (1M KCl) in cmol(+) kg(^{-1}).</td>
</tr>
<tr>
<td>ECSX (N5.2)</td>
<td>Electrical conductivity - saturation extract - in mmhos cm(^{-1}) at 25EC.</td>
</tr>
<tr>
<td>EC (N6.2)</td>
<td>Electrical conductivity - soil:water 1:2.5 - in mmhos cm(^{-1}) at 25EC.</td>
</tr>
<tr>
<td>EC5 (N6.2)</td>
<td>Electrical conductivity - soil:water 1:5 - in mmhos cm(^{-1}) at 25EC.</td>
</tr>
<tr>
<td>PHH2O (N4.1)</td>
<td>pH in the supernatant suspension of a soil-water mixture.</td>
</tr>
</tbody>
</table>
53. **PHCACL2** (N4.1): pH in a 1:2 soil-CaCl₂ suspension.
54. **PHKCL** (N4.1): pH in the supernatant suspension of a soil-1M KCl mixture.
55. **BULKOD** (N5.2): Bulk density - oven dry at 105°C - in kg dm⁻³.
56. **BULKPF3** (N5.2): Bulk density - at 1/3 bar suction - in kg dm⁻³.
57. **BULKFWC** (N5.2): Bulk density - at field water content - in kg dm⁻³.
58. **MCPF0** (N5.1): Moisture content at 0 water bar = pF0 (weight%).
59. **MCPF1** (N5.1): Moisture content at 0.01 bar = pF1.0 (weight%).
60. **MCPF15** (N5.1): Moisture content at 0.03 bar = pF1.5 (weight%).
61. **MCPF18** (N5.1): Moisture content at 0.06 bar = pF1.8 (weight%).
62. **MCPF2** (N5.1): Moisture content at 0.1 bar = pF2.0 (weight%).
63. **MCPF23** (N5.1): Moisture content at 0.2 bar = pF2.3 (weight%).
64. **MCPF25** (N5.1): Moisture content at 0.3 bar = pF2.5 (weight%).
65. **MCPF27** (N5.1): Moisture content at 0.5 bar = pF2.7 (weight%).
66. **MCPF3** (N5.1): Moisture content at 1.0 bar = pF3.0 (weight%).
67. **MCPF34** (N5.1): Moisture content at 2.5 bar = pF3.4 (weight%).
68. **MCPF42** (N5.1): Moisture content at 15 bar = pF4.2, on air dry soil (weight%).
69. **MCPF42W** (N5.1): Moisture content at 15 bar = pF4.2, on field moist soil (weight%).
70. **ABGLL** (N4): Atterberg liquid limit - percent water on a < 0.4 mm base.
71. **ABGPL** (N4): Atterberg plastic limit - percent water on a < 0.4 mm base.
72. **COLE** (N6.3): Coefficient of linear extensibility (air-dry or oven-dry to 1/3-bar tension).
73. **DCOLR1** (C10): First colour of dry soil (hue/value/chroma).
74. **DCOLR2** (C10): Second colour of dry soil (hue/value/chroma).
75. **DCOLR3** (C10): Third colour of dry soil (hue/value/chroma).
76. **MCOLR1** (C10): First colour of moist soil (hue/value/chroma).
77. **MCOLR2** (C10): Second colour of moist soil (hue/value/chroma).
78. **MCOLR3** (C10): Third colour of moist soil (hue/value/chroma).
79. **GROUPING1** (N2): Functional grouping of samples according to inferred activity of their clay minerals, organic carbon content, or allophanic properties. Valid codes are
   - "LAC" for Low Activity Clay (CEC # 20 cmol, kg⁻¹)
   - "MIX" for Mixed Activity Clay (20 < CEC # 64 cmol, kg⁻¹)
   - "HAC" for High Activity Clay (CEC > 64 cmol, kg⁻¹)
   - "ORG" for Organic soils (more than 16% organic carbon)
   - "ALL" for Allophanic soils
   - "NIL" for samples that do not qualify as either "LAC", "MIX", "HAC", "ORG" or "ALL".
80. **GROUPING2** (C4): Functional grouping of samples according to their textural class (FAO '74). Valid codes are
   - 0 : Nil. Textural class could not be established.
   - 1 : Coarse textured.
   - 2 : Medium textured.
   - 3 : Fine textured.
81. **GROUPING3** (C4): Stratification of samples based on GROUPING1 and GROUPING2. Valid codes range from T to XI, or blank (see figure 2).
Appendix 2 Main assumptions and defects

The PTF data set contains soil characterization and profile description data from the SCS Soil Survey Laboratory (Lincoln, NE - USA) database and the International Soil Reference and Information Centre’s ISIS database. Both databases are working databases, implying that their contents are volatile. The attribute values in the PTF data set reflect the current state of their sources. This may have lead to minor inconsistencies in the PTF data set.

In SCS data three special values are used as codes in the data fields:
- A -1 means the analysis was run, but the amount found was less than the minimum reporting amount (detection limit). Wherever possible, this code has been replaced by 0.
- A -2 means the analysis was run, but nothing was found. In a number of instances, this code has been replaced with 0.
- A -3 means the analysis was not run. ISRIC data uses the same code for analyses that were not run.

Soil Taxonomy class names are based on the Great Group and Subgroup codes as supplied by SCS on the May 1994 National Soil Characterization Data CD-ROM. Not all codes are currently active in the SCS database. Codes for which no Soil Taxonomy class name is available have not been changed in the SCS database to meet recent changes (J.Kimble of SCS, personal communication, 1996).

Soil Taxonomy class names are based on the Great Group and Subgroup codes as supplied by SCS on the May 1994 National Soil Characterization Data CD-ROM. Not all codes are currently active in the SCS database. Codes for which no Soil Taxonomy class name is available have not been changed in the SCS database to meet recent changes (J.Kimble of SCS, personal communication, 1996).

The ISRIC database makes use of the FAO SDBm 1995 Great Group and Subgroup codes.

Each record in the PTF data set represents a single sample. For attributes of horizon boundary-crossing samples (e.g. designation) a simple “greater part” rule has been applied.

Analytical methods for the various attributes are described in

The analytical methods used by ISRIC and SCS laboratories are nearly identical. Thus, attribute values from both sources are comparable. There are minor differences in methods for:
- pH-H₂O: pH-H₂O is determined in the supernatant suspension of a 1:2.5 (ISRIC) or 1:1 (SCS) soil-water mixture.
- pH-KCl: pH-KCl is determined in the supernatant suspension of a 1:2.5 (ISRIC) or 1:1 (SCS) soil-1M KCl mixture.
- ECEC: For SCS samples, the ECEC is calculated as the sum of SUMCAT and ALX.
Appendix 3 Characterization of data set

The following table shows the absolute and relative availability of all sample attributes in the PTF data set. The data set consists of 131,472 samples.

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Count</th>
<th>Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>DESIGNTION</td>
<td>42,883</td>
<td>33</td>
</tr>
<tr>
<td>CLAY</td>
<td>111,464</td>
<td>85</td>
</tr>
<tr>
<td>SILT</td>
<td>111,424</td>
<td>85</td>
</tr>
<tr>
<td>FSILT</td>
<td>110,176</td>
<td>84</td>
</tr>
<tr>
<td>CSILT</td>
<td>1,111,101</td>
<td>85</td>
</tr>
<tr>
<td>SAND</td>
<td>111,423</td>
<td>85</td>
</tr>
<tr>
<td>VFSAND</td>
<td>1,110,495</td>
<td>84</td>
</tr>
<tr>
<td>FSAND</td>
<td>110,490</td>
<td>84</td>
</tr>
<tr>
<td>MSAND</td>
<td>110,490</td>
<td>84</td>
</tr>
<tr>
<td>CSAND</td>
<td>110,467</td>
<td>84</td>
</tr>
<tr>
<td>VCSAND</td>
<td>110,490</td>
<td>84</td>
</tr>
<tr>
<td>MINFRAG</td>
<td>50,525</td>
<td>38</td>
</tr>
<tr>
<td>OC</td>
<td>111,140</td>
<td>85</td>
</tr>
<tr>
<td>N</td>
<td>40,820</td>
<td>31</td>
</tr>
<tr>
<td>CACO3</td>
<td>46,247</td>
<td>35</td>
</tr>
<tr>
<td>CAX</td>
<td>89,478</td>
<td>68</td>
</tr>
<tr>
<td>MGX</td>
<td>92,682</td>
<td>71</td>
</tr>
<tr>
<td>NAX</td>
<td>95,651</td>
<td>73</td>
</tr>
<tr>
<td>KX</td>
<td>95,757</td>
<td>73</td>
</tr>
<tr>
<td>SUMCAT</td>
<td>76,969</td>
<td>59</td>
</tr>
<tr>
<td>ALX</td>
<td>25,627</td>
<td>19</td>
</tr>
<tr>
<td>CEC7</td>
<td>91,885</td>
<td>70</td>
</tr>
<tr>
<td>CECBACL2</td>
<td>9</td>
<td>1</td>
</tr>
<tr>
<td>ECEC</td>
<td>20,094</td>
<td>15</td>
</tr>
<tr>
<td>ACIDX</td>
<td>66,868</td>
<td>51</td>
</tr>
<tr>
<td>Attribute</td>
<td>Count</td>
<td>Percent</td>
</tr>
<tr>
<td>-------------</td>
<td>--------</td>
<td>---------</td>
</tr>
<tr>
<td>KCL_ACID</td>
<td>1,546</td>
<td>1</td>
</tr>
<tr>
<td>ECSX</td>
<td>26,759</td>
<td>20</td>
</tr>
<tr>
<td>EC</td>
<td>3,546</td>
<td>3</td>
</tr>
<tr>
<td>EC5</td>
<td>38</td>
<td>1</td>
</tr>
<tr>
<td>PHH2O</td>
<td>108,703</td>
<td>83</td>
</tr>
<tr>
<td>PHCACL2</td>
<td>84,823</td>
<td>65</td>
</tr>
<tr>
<td>PHKCL</td>
<td>24,670</td>
<td>24</td>
</tr>
<tr>
<td>BULKOD</td>
<td>43,305</td>
<td>33</td>
</tr>
<tr>
<td>BULKPF3</td>
<td>40,000</td>
<td>30</td>
</tr>
<tr>
<td>BULKFWC</td>
<td>4,102</td>
<td>3</td>
</tr>
<tr>
<td>MCPF0</td>
<td>954</td>
<td>1</td>
</tr>
<tr>
<td>MCPF1</td>
<td>926</td>
<td>1</td>
</tr>
<tr>
<td>MCPF15</td>
<td>923</td>
<td>1</td>
</tr>
<tr>
<td>MCPF18</td>
<td>484</td>
<td>1</td>
</tr>
<tr>
<td>MCPF2</td>
<td>9,076</td>
<td>7</td>
</tr>
<tr>
<td>MCPF23</td>
<td>948</td>
<td>1</td>
</tr>
<tr>
<td>MCPF25</td>
<td>42,239</td>
<td>32</td>
</tr>
<tr>
<td>MCPF27</td>
<td>926</td>
<td>1</td>
</tr>
<tr>
<td>MCPF3</td>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td>MCPF34</td>
<td>909</td>
<td>1</td>
</tr>
<tr>
<td>MCPF42</td>
<td>108,571</td>
<td>83</td>
</tr>
<tr>
<td>MCPF42W</td>
<td>5,273</td>
<td>4</td>
</tr>
<tr>
<td>ABGLL</td>
<td>7,533</td>
<td>6</td>
</tr>
<tr>
<td>ABGPL</td>
<td>7,473</td>
<td>6</td>
</tr>
<tr>
<td>COLE</td>
<td>165</td>
<td>1</td>
</tr>
<tr>
<td>DCOLR1</td>
<td>26,272</td>
<td>20</td>
</tr>
<tr>
<td>DCOLR2</td>
<td>1,520</td>
<td>1</td>
</tr>
<tr>
<td>DCOLR3</td>
<td>153</td>
<td>1</td>
</tr>
<tr>
<td>MCOLR1</td>
<td>52,035</td>
<td>40</td>
</tr>
</tbody>
</table>
The next table shows the absolute and relative availability of all *pedon* attributes in the PTF data set. The data set consists of samples from 20,920 different pedons.

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Count</th>
<th>Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>DATE</td>
<td>8,783</td>
<td>42</td>
</tr>
<tr>
<td>FAO74_CODE</td>
<td>566</td>
<td>3</td>
</tr>
<tr>
<td>FAO74_DESC</td>
<td>566</td>
<td>3</td>
</tr>
<tr>
<td>FAO92_CODE</td>
<td>296</td>
<td>1</td>
</tr>
<tr>
<td>FAO92_DESC</td>
<td>296</td>
<td>1</td>
</tr>
<tr>
<td>SUB92_CODE</td>
<td>110</td>
<td>1</td>
</tr>
<tr>
<td>SUB92_DESC</td>
<td>110</td>
<td>1</td>
</tr>
<tr>
<td>TXGG_CODE</td>
<td>9,407</td>
<td>45</td>
</tr>
<tr>
<td>TXSG_CODE</td>
<td>9,317</td>
<td>45</td>
</tr>
<tr>
<td>TX_NAME</td>
<td>8,860</td>
<td>42</td>
</tr>
<tr>
<td>TX_YEAR</td>
<td>7,977</td>
<td>38</td>
</tr>
<tr>
<td>TXTEX_CODE</td>
<td>9,186</td>
<td>44</td>
</tr>
<tr>
<td>TXTEX_DESC</td>
<td>9,177</td>
<td>44</td>
</tr>
<tr>
<td>TXMIN_CODE</td>
<td>8,942</td>
<td>43</td>
</tr>
<tr>
<td>TXMIN_DESC</td>
<td>8,939</td>
<td>43</td>
</tr>
</tbody>
</table>

The STATS table - distributed with the PTF data set - contains a break down of the absolute and relative availability of all attributes in the PTF data set for the categories I to XI. The STATS table is a Delimited Format ASCII file (see appendix 1). The first two records in the file are header lines, data starts in record three. Record one lists the names of all data fields as they appear in the table. Record two consists of blocks of dashes ('-') that are of the same length as the field names.
Structure of the STATS table

1. **ATTRIBUTE** (C10): Attribute name.
2. **COUNT_ALL** (N6): Absolute availability of the attribute for the entire data set.
3. **PRCNT_ALL** (N3): Relative availability of the attribute for the entire data set (%).
4. **COUNT_I** (N6): Absolute availability of the attribute for category I.
5. **PRCNT_I** (N3): Relative availability of the attribute for category I (%).
6. **COUNT_II** (N6): Absolute availability of the attribute for category II.
7. **PRCNT_II** (N3): Relative availability of the attribute for category II (%).
8. **COUNT_III** (N6): Absolute availability of the attribute for category III.
9. **PRCNT_III** (N3): Relative availability of the attribute for category III (%).
10. **COUNT_IV** (N6): Absolute availability of the attribute for category IV.
11. **PRCNT_IV** (N3): Relative availability of the attribute for category IV (%).
12. **COUNT_V** (N6): Absolute availability of the attribute for category V.
13. **PRCNT_V** (N3): Relative availability of the attribute for category V (%).
14. **COUNT_VI** (N6): Absolute availability of the attribute for category VI.
15. **PRCNT_VI** (N3): Relative availability of the attribute for category VI (%).
16. **COUNT_VII** (N6): Absolute availability of the attribute for category VII.
17. **PRCNT_VII** (N3): Relative availability of the attribute for category VII (%).
18. **COUNT_VIII** (N6): Absolute availability of the attribute for category VIII.
19. **PRCNT_VIII** (N3): Relative availability of the attribute for category VIII (%).
20. **COUNT_IX** (N6): Absolute availability of the attribute for category IX.
21. **PRCNT_IX** (N3): Relative availability of the attribute for category IX (%).
22. **COUNT_X** (N6): Absolute availability of the attribute for category X.
23. **PRCNT_X** (N3): Relative availability of the attribute for category X (%).
24. **COUNT_XI** (N6): Absolute availability of the attribute for category XI.
25. **PRCNT_XI** (N3): Relative availability of the attribute for category XI (%).

For 6867 of the 20920 pedons (approx. 33%) in the PTF data set the exact geographic location (in degrees, minutes, and seconds) is available at ISRIC.