Working Paper and Preprint 96/05

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

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:

Tempel P., N.H. Batjes and V.W.P. van Engelen, 1966. IGBP-DIS soil data set for pedotransfer function development. Working paper and Preprint 96/05, International Soil Reference and information Centre (ISRIC), Wageningen.

Inquiries:

Director
P.O. Box 353
6700 AJ Wageningen
The Netherlands
Telefax: +31-317-471700
E-mail: SOIL@ISRIC.NL

Working Paper and Preprint 96/05

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

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 data base; 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 q_h 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; Pleijsier, 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-H_2O and pH-KCl), cation exchange capacity (at pH 7 and pH 8.2), effective CEC (in 1 *M* KCl), exchangeable acidity, exchangeable bases (Ca^{2+}, Mg^{2+}, 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
LAC CEC#20 $cmol_+ kg^{-1}$			
MIX 20< CEC# 62 $cmol_+ kg^{-1}$			
HAC 62 < CEC $cmol_+ kg^{-1}$			
Organic Soils (OC>16%)			
Allophanic soils			

* CEC is cation exchange capacity, measured at pH7 in NH_4OAc: 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 $cmol_c\ kg^{-1}$ 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.

	Nil 20,066 15.3%	Coarse 22,333 17.0%	Medium 65,131 50.0%	Fine 23,942 18.2%
LAC 60,077 45.7%		**I** 16,430 12.5%	**II** 35,063 26.7%	**III** 6,140 4.7%
MIX 28,743 21.9%		**IV** 847 0.6%	**V** 13,676 10.4%	**VI** 13,149 10.0%
HAC 660 0.5%		**VII** 49 <0.1%	**VIII** 173 0.1%	**IX** 300 0.2%
Organic Soils 1,259 1.0%			**X** 1,259 1.0%	
Allophanic soils 1,510 1.2%			**XI** 1,510 1.2%	
NIL 39,223 29.8%				

* 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).

References

Arnold, R.W., 1995. Role of soil survey in obtaining a global carbon budget. *In*: Soils and Global Change (eds. R. Lal, J. Kimble, E. Levine, and B.A. Stewart), pp. 257-263. Lewis Publishers, Boca Raton.

Arrouays, D., Jamagne, M. and Gaillard, H., 1993. Sur la possibilité d'estimer les propriétés de rétention en eau des sols limoneux lessivés hydromorphes du sud-ouest de la France à partir de leurs caractéristiques de constitution. *C.R. Acad. Agric. France*, 79: 111-121.

Batjes, N.H., 1995. A homogenized soil data file for global environmental research: A subset of FAO, ISRIC and NRCS profiles (Version 1.0). Working Paper and Preprint 95/10, ISRIC, Wageningen.

Batjes, N.H. 1996. Development of a global data set of soil moisture retention properties using pedotransfer rules. *Geoderma*, 71: 31-52.

Batjes, N.H. and Bridges, E.M., 1994. Potential emissions of radiatively active gases from soil to atmosphere with special reference to methane: development of a global database (WISE). *Journal of Geophysical Research* 99 (D8): 16,479-16,489.

FAO, 1995. Digital Soil Map of the World and Derived Soil Properties (version 3.5). FAO, Rome.

FAO-UNESCO, 1974. Soil Map of the World. Volume I: Legend. UNESCO, Paris.

Foussereau, Z., Hornsby, A.G. and Brown, R.B., 1993. Accounting for variability within map units when linking a pesticide fate model to soil survey. *Geoderma*, 60: 257-276.

Hudson, B.D., 1994. Soil organic matter and available water capacity. *J. Soil and Water Conservation*, 49: 189-194.

Ingram, S. and Gregory, P., (eds), 1996. Effects of global change on soils: implementation plan. Global Change and Terrestrial Ecosystems Report No. 12, GCTE, Wallingford.

Kern, J.S., 1995a. Geographic patterns of soil water holding capacity in the contiguous United States. *Soil Sci. Soc. Am. J.*, 59: 1126-1133.

Kern, J.S., 1995b. Evaluation of soil water retention models based on basic soil physical properties. *Soil Sci. Soc. Am. J.*, 59: 1134-1141.

Leenhardt, D., 1995. Errors in the estimation of soil water properties and their propagation through a hydrological model. *Land Use and Management*, 11: 15-21.

Madsen, H.B. and Jones, R.J.A., 1995. The establishment of a soil profile analytical database for the European Union. *In*: European Land Information Systems for Agro-Environmental Monitoring (eds. D. King, R.J.A. Jones and A.J. Thomasson), pp. 55-63. Office for Official Publications of the European Communities, Luxembourg.

Ngongo, L., King, D., Nicoullaud, B., Brisson, N. and Ruget, F., 1993. Estimation des erreurs de prédiction spatiale du déficit hydrique dues à la rastérisation des cartes des sols. *Bull. Rech. Agron. Gembloux*, 28: 223-239.

Pleijsier, L.K. (Ed.), 1986. Proceedings of an international workshop on the Laboratory Methods and Data Exchange programme. Technical Paper 13, ISRIC, Wageningen, pp. 6-9.

Van Reeuwijk, L.P., 1993. Procedures for soil analysis. Technical Paper 9, IRIC, Wageningen.

Saxton, K.E., Rawls, W.J., Romberger, J.S. and Papendick, R.I., 1986. Estimating generalized soil-water characteristics from texture. *Soil Sci. Soc. Am. J.*, 50: 1031-1036.

SC-DLO, 1994. Using existing soil data to derive hydraulic parameters for simulation models in environmental studies and in land use planning. Workshop Report, Winand Staring Centre, Wageningen.

Scholes, R.J., Skole, D. and Ingram, J.S., 1995. A global database of soil properties: proposal for implementation. IGBP-DIS Working Paper 10, International Geosphere Biosphere Program, Data and Information System, Paris.

Scholes, R.J., 1994. Workshop to define the specifications of the first data products of the Global Soil Data Task (GSDT) (Montpellier, 28 September - 1 October 1994). Draft Report, IGBP-DIS, Paris.

Shouse, P.J., Russell, W.B., Burden, D.S., Selim, H.M., Sisson, J.B. and van Genuchten, M.Th., 1995. Spatial variability of soil water retention functions in a silt loam. *Soil Science*, 159: 1-12.

Soil Survey Staff, 1996. Soil Survey Laboratory Methods Manual. Soil Survey Investigations Report 42 (Version 3.0). United States Department of Agriculture, Lincoln, NE.

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 (**N**umeric or **C**haracter), 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.
7. **FAO74_CODE** (C2): FAO-Unesco 1974 soil unit code.
8. **FAO74_DESC** (C20): FAO-Unesco 1974 soil unit name.
9. **FAO92_CODE** (C2): FAO-Unesco 1992 soil unit code.
10. **FAO92_DESC** (C20): FAO-Unesco 1992 soil unit name.
11. **SUB92_CODE** (C2): FAO-Unesco 1992 soil subunit prefix.
12. **SUB92_DESC** (C12): FAO-Unesco 1992 soil subunit name.
13. **TXGG_CODE** (C5): Soil Taxonomy (USDA/SCS) order/suborder/great group code.
14. **TXSG_CODE** (C4): Soil Taxonomy (USDA/SCS) subgroup prefix.
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

(USDA/SCS).

17. **TXTEX_CODE** (C3): Soil Taxonomy (USDA/SCS) texture code
18. **TXTEX_DESC** (C50): Soil Taxonomy (USDA/SCS) texture description.
19. **TXMIN_CODE** (C2): Soil Taxonomy (USDA/SCS) mineralogy code.
20. **TXMIN_DESC** (C25): Soil Taxonomy (USDA/SCS) mineralogy description.
21. **TOP_DEPTH** (N3): Depth of the top of the **sample** (cm).
22. **BOT_DEPTH** (N3): Depth of the bottom of the **sample** (cm).
23. **DESIGNTION** (C8): Horizon designation.
24. **CLAY** (N4.1): Weight% of particles < 0.002mm (clay) in fine earth fraction.
25. **SILT** (N4.1): 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.
26. **FSILT** (N4.1): Weight% of particles 0.02-0.002 mm (fine silt) in fine earth fraction.
27. **CSILT** (N4.1): Weight% of particles 0.05-0.02 mm (coarse silt) in fine earth fraction.
28. **SAND** (N4.1): 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.
29. **VFSAND** (N4.1): Weight% of particles 0.1-0.05 mm (very fine sand) in fine earth fraction.
30. **FSAND** (N4.1): Weight% of particles 0.25-0.1 mm (fine sand) in fine earth fraction.
31. **MSAND** (N4.1): Weight% of particles 0.5-0.25 mm (medium sand) in fine earth fraction.
32. **CSAND** (N4.1): Weight% of particles 1.0-0.5 mm (coarse sand) in fine earth fraction.
33. **VCSAND** (N4.1): Weight% of particles 2.0-1.0 mm (very coarse sand) in fine earth fraction.
34. **MINFRAG** (N5.1): Weight% (whole soil) of mineral fragments > 2 mm.
35. **OC** (N5.2): Weight% of total organic carbon (Walkley-Black method).
36. **N** (N6.2): Weight% of total nitrogen (Kjeldahl method).
37. **CACO3** (N5.1): Weight% of free $CaCO_3$ in fine earth fraction
38. **CAX** (N5.1): Exchangeable Ca in cmol(+) kg^{-1}.
39. **MGX** (N5.1): Exchangeable Mg in cmol(+) kg^{-1}.
40. **NAX** (N5.1): Exchangeable Na in cmol(+) kg^{-1}.
41. **KX** (N5.1): Exchangeable K in cmol(+) kg^{-1}.
42. **SUMCAT** (N5.1): Sum of cations in cmol(+) kg^{-1}.
43. **ALX** (N5.1): Exchangeable Al (1M KCl) in cmol(+) kg^{-1}.
44. **CEC7** (N5.1): Cation Exchange Capacity of the soil at pH 7 in cmol(+) kg^{-1}.
45. **CECBACL2** (N5.1): Cation Exchange Capacity of the soil - $BaCl_2$, pH 8.2 - in cmol(+) kg^{-1} (SCS method code 5A5).
46. **ECEC** (N6.2): Effective CEC in cmol(+) kg^{-1}.
47. **ACIDX** (N5.1): Na_4OAc. extractable acidity at pH 8.2, in cmol(+) kg^{-1}.
48. **KCL_ACID** (N5.1): Extractable acidity (1M KCl) in cmol(+) kg^{-1}.
49. **ECSX** (N5.2): Electrical conductivity - saturation extract - in mmhos cm^{-1} at 25EC.
50. **EC** (N6.2): Electrical conductivity - soil:water 1:2.5 - in mmhos cm^{-1} at 25EC.
51. **EC5** (N6.2): Electrical conductivity - soil:water 1:5 - in mmhos cm^{-1} at 25EC.
52. **PHH2O** (N4.1): pH in the supernatant suspension of a soil-water mixture.

53. **PHCACL2** (N4.1): pH in a 1:2 soil-$CaCl_2$ 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 EC - 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 1/3 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 '**I**' 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, bur 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).
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

- Soil Survey Staff, 1996. Soil Survey Laboratory Methods Manual. Soil Survey Investigations Report 42 (Version 3.0). United States Department of Agriculture, Lincoln, NE.
- Van Reeuwijk, L.P., 1993. Procedures for soil analysis. Technical Paper 9, IRIC, Wageningen.

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-H_2O:	pH-H_2O 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.

Attribute	Count	Percent
DESIGNTION	42,883	33
CLAY	111,464	85
SILT	111,424	85
FSILT	110,176	84
CSILT	1,111,101	85
SAND	111,423	85
VFSAND	1,110,495	84
FSAND	110,490	84
MSAND	110,490	84
CSAND	110,467	84
VCSAND	110,490	84
MINFRAG	50,525	38
OC	111,140	85
N	40,820	31
CACO3	46,247	35
CAX	89,478	68
MGX	92,682	71
NAX	95,651	73
KX	95,757	73
SUMCAT	76,969	59
ALX	25,627	19
CEC7	91,885	70
CECBACL2	9	1
ECEC	20,094	15
ACIDX	66,868	51

Attribute	Count	Percent
KCL_ACID	1,546	1
ECSX	26,759	20
EC	3,546	3
EC5	38	1
PHH2O	108,703	83
PHCACL2	84,823	65
PHKCL	24,670	24
BULKOD	43,305	33
BULKPF3	40,000	30
BULKFWC	4,102	3
MCPF0	954	1
MCPF1	926	1
MCPF15	923	1
MCPF18	484	1
MCPF2	9,076	7
MCPF23	948	1
MCPF25	42,239	32
MCPF27	926	1
MCPF3	4	1
MCPF34	909	1
MCPF42	108,571	83
MCPF42W	5,273	4
ABGLL	7,533	6
ABGPL	7,473	6
COLE	165	1
DCOLR1	26,272	20
DCOLR2	1,520	1
DCOLR3	153	1
MCOLR1	52,035	40

Attribute	Count	Percent
MCOLR2	5,554	4
MCOLR3	1,218	1

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.

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

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 **I**V.
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.