#### GLOBAL DIGITAL DATA SETS OF SOIL TYPE, SOIL TEXTURE, SURFACE SLOPE, AND OTHER PROPERTIES: DOCUMENTATION OF ARCHIVED DATA TAPE

Brad Staub and Cynthia Rosenzweig

October, 1987

NASA Goddard Space Flight Center
Institute for Space Studies
2880 Broadway
New York, NY 10025

#### ABSTRACT

This documentation describes the file structure and coding of four soils data sets derived from the Zobler (1986) world soil file. These data were digitized on a one-degree square grid, a design especially useful for large-area studies such as climate research with general circulation models. They are also suitable for studies in forestry, agriculture, soils, hydrology, and other research areas that can utilize soils information on this scale. All four data sets are based on the Food and Agriculture Organization (FAO, 1974) Soil Map of the World and on the Matthews (1983; 1984) vegetation data set. Data characteristics, coding scheme, and collection methodology are fully described by Zobler (1986).

The advantages of the Zobler world soil file arc: 1) Missing data have been filled in; 2) Subsidiary soils of the designated FAO soil association for each map unit are included; and 3) the data set may be easily converted to compatibility in nominal classification of land, land-ice and water cells with either the Matthews vegetation file or the FAO Soil Map of the World. Dot grid overlay was used to determine the largest map unit of each one-degree cell.

The Zobler world soil file differs from the Wilson (described in Wilson and Henderson-Sellers, 1985) soil data set in that more complete information on texture (9 classes), slope (9 classes), and phase (18) is included for each map unit (106 FAO soil types) comprising the largest area of each one-degree square land cell. This encoding provides more detailed information at each cell than the Wilson soil data set which has a classification scheme (21 types) based on combinations of three colors, three textures, and three drainage classes.

The Zobler world soil file also provides more detailed information about each map unit than the Gildea and Moore (described in Henderson-Sellers et al., 1986) FAOSOL global soil archive, because it includes the soil types designated by FAO as 'associated' and 'included'. \*Associated' soils occupy at least 20% of the soil association and 'included' soils occupy less than 20% of the soil association of each map unit. The Gildea and Moore global soil archive is at 0.5° x 0.5° resolution, while the Zobler world soil file is at 1°- x 1° resolution.

The fourth file on the tape is the SOILWRLD data set. It contains information on soil properties of land cells of both Matthews\* and FAO sources, reconciles land-classification differences between the two, and has missing data filled in. The first through third files on this tape are subsets of the SOILWRLD data set that have been ordered in a standard 360 x 180 data matrix geometry, have been partially recoded for ease of use, have missing data filled in, and agree with Matthews' (1984) nominal classification of land type (land, land-ice, water). The first file is a data set of codes for soil unit, land-ice, or water, for all the one-degree square cells of the earth. The second file is a data set of codes for texture, land-ice, or water for these same soil units. The third file is a data set of codes for slope, land-ice or water for these same soil units. When used together, these three data sets conveniently define a soil type, texture, and slope, for the one-degree square cells of the earth.

SECTION 1: Description of SOIL UNIT file, Data Set Name: SOIL-USER

This file consists of a (360, 180) matrix of character\*2 SOIL UNIT data based on the FAO Soil Map of the World, and compiled into digital form by Zobler. Each matrix element represents the dominant soil unit in a one-degree square cell of the earth's surface. This file was prepared from the SOILWRLD data set (the fourth file on this tape), with the addition of the land-ice cells of Antarctica. This file conforms exactly in number, location, and nominal classification (land, land-ice, water) to Matthews\* vegetation data set.

SECTION I.A: File Format and Coding, Irecl: 720 blksize: 3600 recfm: fb

This data set is read with the following FORTRAN statements, and with 'unit' set as a tape:

CHARACTER\*2 **SU(360**, 1801

DO 20 J=1,180 20 READ(unit, 101)(SU(I,J), I=1,360) 101 FORMAT(250A2,110A2)

The SOIL UNIT codes in this file are identical with the soil unit codes of the SOILWRLD file and the FAO map (see Appendix 1 for a list of the 106 soil unit names and their two-character codes). WATER cells are included in this file and are coded as '. (double blank). LAND-ICE cells are coded as '99' (ninety-nine)

SECTION I.B: Notes

(1) The latitude and longitude of the northwest corner of any one-degree square cell are determined by:

LAT = J - 90 LON = 1 - 181

Cell latitudes are 0 at the equator, increase positively northward to +90, and increase negatively southward to -89. Cell longitudes are 0 at the prime meridian, increase positively eastward to +179, and increase negatively westward to -180. This convention holds for all four data sets on this tape. See Appendix 2 for a discussion of map coordinates and matrix indices.

(2) See Appendix 3 for a discussion of simple printer-mapping routines for this data set.

#### SECTION 2: Description of TEXTURE file. Data Set Name: TEX.USER

This file consists of a (360.180) matrix of character\*1 TEXTURE data based on the FAO Soil Map of the World, and compiled into digital form by Zobler. Each matrix element represents the near-surface texture (upper 30 cm) of the dominant soil unit in a one-degree square cell of the earth's surface. This file was prepared from the SOILWRLD data set (the fourth file on this tape), with the addition of the land-ice cells of Antarctica. This file conforms exactly in number, location, and nominal classification (land, land-ice, water) to Matthews' vegetation data set.

SECTION 2.A: File Format and Coding. lrecl: 360 blksize:3600 recfm: f b

These data are read with the following FORTRAN statements, and with 'unit' set as a tape:

CHARACTER\* 1 TEX (360,180)

DO 20 J=1,180 20 READ(unit,105)(TEX(I,J); I=1,360) 105 FORMAT(250A1,110A1)

The TEXTURE codes in this data set are shown in Table 2.1.

Table 2.1. Texture codes and class in TEX data set.

texture code	texture class
'1' '2' '3' '4' '5' '6' '7' '8'	coarse medium fine coarse-medium coarse-fine medium-fine coarse-medium-fine dium-fine
	water

#### SECTION 2.B: Notes

(1) The 'organic' texture class is applied to those soil units, primarily Histosols, that have no

texture symbol on the FAO map, and that are organic soils (i.e., possessing an H or 0 master horizon). This texture class is not an FAO specification.

- (2) The longitude and latitude convention is identical to the Soil Unit data set, as described in Section 1.B.(1), and Appendix 2.
- (3) These data can be mapped, as shown in Appendix 3.

#### SECTION 3: Description of the SLOPE file. Data Set Name: SLP.USER

This file consists of a (360,180) matrix of character\*1 SLOPE data based on the FAO Soil Map of the World, and compiled into digital form by Zobler. Each matrix element represents the surface slope of the dominant soil unit in a one-degree square cell of the earth's surface. This file was prepared from the SOILWRLD data set (the fourth file on this tape), with the addition of the landice cells of Antarctica. This file conforms exactly in number, location, and nominal classification (land, land-ice, water) to the Matthews vegetation data set.

SECTION 3.A: File Format and Coding. lreel: 360 blksize: 3600 recfm: fb

These data are read with the following FORTRAN statements, and with 'unit' set as a tape:

CHARACTER\* 1 SLP (360.1801

DO 20 J=1,180 20 READ(unit,105)(SLP(I,J), I=1, 360) 105 FORMAT(250A1,110A1)

The slope codes in this data set are shown in Table 3.1.

Table 3.1 Slope codes and class in the SLP data set,

slope code	nercent slope class
'1' '2' '3' '4' '5' '6' '7'	O-8 8-30 greater than 30 O-30 O-8; greater than 30 8-30; greater than 30 O-8; 8-30; greater than 30 land-ice water

#### SECTION 3.B: Notes

- (1) The longitude and latitude convention is identical to the Soil Unit data set, as described in Section 1.B.(1), and Appendix 2.
- (2) These data may be easily plotted as shown in Appendix 3.

#### SECTION 4: Description of a WORLD SOIL FILE. Data Set Name: SOILWRLD.USER

This file is the SOILWRLD data set compiled by Zobler. It consists of 154 13 card image records of soils information obtained from the FAO Soil Map of the World. Each record contains information on soil type, near-surface texture, slope, and phase for the map unit comprising the largest area of a one-degree square land cell. The largest map unit was determined by dot grid overlay. Missing data have been filled in Conflicting nominal classification of some land cells between FAO and Matthews' data sets are identified, and coded to permit reconciliation according to either source. Water cells and Antarctica are excluded from this file; other non-soil land cells, such as debris fields, dunes, and land-ice, are included.

SECTION 4.A: File Format and Coding. lrecl: 80 blksize: 6160 recfm: fb

Each record consists of 9 fields in the order shown in Table 4.1.

Table 4.1 Data fields of a SOILWRLD data set record.

mnemonic d	lata type	field description
LAT	IS	LATITUDE of northwest corner of cell.
LON	15	LONGITUDE of northwest corner of cell.
s u	A2	Two-letter SOIL UNIT code (see Appendix 1) and also a ' ' (double blank) for land-ice; (see note (1), below).
SUBSID	A6	Associated and included SUBSIDIARY soil units (see note (2), below).
TEX	A3	TEXTURE codes (see note (3), below).
SLP	A3	SLOPE codes (see note (4), below).
PHS	A15	A string of one- and two-digit PHASE symbols, if any (see note (5), below).
SP(3)	A2	A three-element array of two-digit SPECIAL codes (see note (6), below) used to mask record attributes.
AREA	A3	A three-letter code that nominally identifies a broad regional location of the cell (see note (7), below).

Including the imbedded blanks between fields, the following FORTRAN statements (with 'unit' set as a tape) may be used to read one record:

INTEGER \*4 LAT. LON CHARACTER \*2 SU, SP(3) CHARACTER \*3 TEX, SLP, AREA CHARACTER \*6 SUBSID CHARACTER \* 15 PHS

10 READ (unit.1011 LAT, LON. SU, SUBSID, TEX, SLP, PHS. SP, AREA 101 FORMAT (215, 3X, A2, 1X, A6, 2X, A3, 2X, A3, IX, A15, 3(1X,A2), \*IX. A3)

To read all the records, place this READ statement within a DO loop of 1 to 15413, or under some other program control.

#### Section 4.B: Notes

- (1) The SU (soil unit) codes are the two-letter symbols for the FAO soil unit names of the dominant soil type of the map unit comprising the largest area within a one-degree land cell. These codes are identical with the codes in the first file on this tape (see Appendix 1). In addition, land-ice is coded as '' (double blank); recall that water cells are not included.
- (2) The SUBSID (subsidiary) codes are the FAO symbols for 'associated' and 'included' soil types of a soil association. The dominant soil type of the association is indicated in the SU field (note (1), above). 'Associated' soils occupy at least 20% of the area of the soil association, and 'included' soils occupy less than 20%. These codes are explained on individual FAO map sheets.
- (3) The TEX (texture) codes are for the upper 30 cm of the dominant soil unit of a cell (note (1), above). The codes are identical to the FAO symbols, with the addition of a ''(triple blank) for land-ice, and a '9 · (nine; double blank) for organic type soils having no texture symbol on the FAO map (see 2.B.1). Recall that water cells are not included. The texture classes and the codes are shown in Table 4.2.

Table 4.2. Texture codes and class in SOILWRLD data set.

texture code	texture class
'1	coarse medium fine coarse-medium coarse-fine medium-fine coarse-medium-fine organic land-ice

(4) The SLP (slope) codes are for the dominant soil unit of a cell (see note (1), above). The codes are identical to the FAO symbols with the addition of a ' ' (triple blank) for land-ice. Recall that water cells are not included. The slope classes and codes are shown in Table 4.3

Table 4.3. Slope codes and classes in SOILWRLD data set.

'A	Ope code	percent slope class
'AB . O-30 'AC . O-8; grate; than 30 'BC ' 8-30; greater than 30	'B' 'C' 'AB' 'AC' 'BC'	8-30 greater than 30 O-30 O-8; grate: than 30 8-30; greater than 30 O-8; 8-30; greater than 30

(5) Soil phases are properties of a soil that influence its use and management, but are not used to classify the soil. Soils may have none, one, or more phases. The PHS (phase) codes in this data set are one- and two-digit symbols, if any, separated by one or more blanks, not in ordinal sequence. The phase codes are shown in Table 4.4.

Table 4.4. Phase codes and meaning in SOILWRLD data set.

phase code	meaning	phase code meaning	
'1'	stony	'11'	phreact ic
'2'	lithic	'12'	cerrado

phase code	meaning	phase code	meaning
'3' '4' '5' '6' '7' '8' '9'	petric petroferric petrocalcic petrogypsic fragipan duripan saline	'13' '21' '22' '23' '24' '25' '26'	permafrost intermittent permafrost glacier (i.c land-ice) ponded dunes, sands rock debris

(6) The SP (special) codes serve two purposes: to identify land cells that have missing data in some of the fields, and to identify cells that have a conflicting nominal classification (land. land-ice, water) between Zobler and Matthews.

(6.1) Missing data. When soil unit, texture, or slope information for the map unit chosen to represent a land cell was missing on the FAO map, this information was added by Zobler and was used to fill in the SU. TEX. or SLP fields for such a record in the file. In addition, a code was placed in the SP field to indicate which fields were filled in (Table 4.5).

Table 4.5. SP codes for missing data in the SOILWRLD data set.

- '99' No soil information present on the map; SU. TEX. and SLP fields filled in.
- '88' Texture information missing; TEX field filled in.
- '77' Slope informat ion missing; SLP field filled in.

These codes allow a user of this data set to reproduce the actual FAO map unit chosen to characterize a cell by checking for these codes in the SP field and masking the appropriate filled in field(s). Note that a '99' indicates that a non-soil land cell (such as debris field or dunes) was present on the FAO map. In this instance, in addition to the filled in missing data, the phase field contains a code for the type of non-soil land unit that is indicated on the map.

(6.2) Nominal classification conflicts. The nominal classification conflicts between Zobler and

Matthews are of two kinds: land/water classification. and land-ice/land classification.

A land/water conflict occurred whenever Zobler found a cell to be absent on the FAO map and present in Matthews. In this instance the cell was added to the SOILWRLD data set as either landice or an interpreted soil, in conformity with the Matthews nominal classification of the cell. In addition, a code was placed in the SP field to indicate this condition (Table 4.6).

Table 4.6. SP codes for land/water conflicts.

- '44' Land cell not present on FAO map; classified as land-ice.
- '33' Land cell not present on FAO map; classified as soil; SU, TEX. and SLP fields filled-in.

A land-ice/land conflict occurred whenever Matthews and FAO disagreed on the land-ice classification of a cell. Records of this kind are not uniquely classified in this file, and so have attributes of both a soil cell and a land-ice (i.e. glacier) cell: the SU. TEX, and SLP fields have appropriate codes in them. and the PHS field also contains a glacier code. In addition, a code was placed in the SP field to indicate this condition (Table 4.7).

Table 4.7. SP codes for land-ice/land conflicts.

- '66' Cell classified as land-ice by Zobler, and as vegetation (soil implied) by Matthews.
- '55' Cell classified as soil by Zobler, and as land-ice by Matthews.

The user of this data set may choose how these records should be classified, either as Zobler-compatible, or as some other union of the two. According to the table below (Table 4.8), this decision amounts to either a row collapse, a column collapse, or other masking, respectively. (Note that as an aid to users already using the Matthews data sets, the first three data sets on this tape have been made Matthews-compatible).

Table 4.8. Zobler and Matthews compatibility.

#### ZOBLER

Soil Glacier

Soil true '66'
soil true
Glacier '55' true
ice

When these cells are resolved according to Matthews. a '66'-type all becomes a soil cell (delete the glacier code from the PHS field; retain all other fields), and a '55'-type all becomes a land-ice all (retain the glacier code in the PHS field; delete the SU. SUBSID, TEX. and SLP fields, and other non-glacier PHS codes). To reconcile these type cells according to Zobler, perform the same masking operations of the opposing cell types: a '66'-type all becoming a land-ice all, and a '55'-type all a soil all.

(7) The AREA codes indicate territorial regions and are used as broad guides to a all location on FAO map sheets. These codes are shown in Table 4.9.

Table 4.9. AREA codes and meaning in SOILWRLD data set.

AREA co	de meaning	AREA code	meaning
'AUS'	Australia. South Pacific	'SAM'	South Amerin
'AFR'	Africa	'NCA'	North Central Asia
'EUR'	Europe	'SAS'	Southern Asia
'NAM'	North America	'SBA'	South East Asia

<sup>(8)</sup> The latitude and longitude conventions are the same as those discussed in section 1.B.(1), and Appendix 2.

#### References

- FAO-UNESCO (1974) Soil Map of the World 1:5,000,000. UNESCO, Paris.
- Henderson-Sellers, A., M.F. Wilson, G. Thomas, R.E. Dickinson (1986) Current Global Land-Surface Data Sets for Use in Climate-Related Studies. NCAR Technical Note NCAR/TN-272+STR.
- Matthews, E. (1983) Global vegetation and land use: New high resolution data bases for climate studies. J. Clim. Appl. Meteor., 22, 474-487.
- Matthews, E. (1984) Vegetation. Land-Use and Seasonal Albedo Data Sets: Documentation of Archived Data Tape. NASA Technical Memorandum #86107.
- Wilson, M.F. and A. Henderson-Sellers (1985) A global archive of land cover and soils data for use in general circulation climate models. J. Clim., 5, 119-143.
- Zobler, L. (1986) A World Soil File for Global Climate Modeling. NASA Technical Memorandum #87802.

#### **Acknowledgments**

We thank Elaine Matthews for her help in preparing this report, Tina Cary for coding advice, and Leonard Zobler for his encouragement on soils parameterization in hydrology. Christopher Shashkin typed the report. This work was supported by the NASA Climate Program and EPA Contract R812962-010.

## APPENDIX 1: SOIL CLASSIFICATION UNITS AND CHARACTER CODES FAO-UNESCO (1974). SOIL MAP OF THE WORLD

#### Soil Units

J FLUVISOLS	Q ARENOSOLS	Z SOLONCHAKS
JE Eutric Fluvisols JC Calcaric Fluvisols JD Dystric Fluvisols	QC Cambic Arenosols QL Luvic Arenosols	ZO Orthic Solonchaks ZM Mollic Solonchaks
JT Thionic Fluvisols	QF Ferral ic Arcnosols QA Albic Arenosols	ZT Takyric Solonchaks ZG Gleyic Solonchaks
G GLEYSOLS	E RENDZINAS	S SOLONETZ
GE Eutric Gleysols GC Calcaric Gleysols GD Dystric Gleysols GM Mollic Gleysols GH Humic Gleysols GP Plinthic Gleysols	U RANKERS	s o Orthic Solonetz SM Mollic Solonetz SG Gleyic Solonetz
GX Gelic Gleysols	T ANDOSOLS	Y YERMOSOLS
R REGOSOLS	TO Ochric Andosols TM Mollic Andosols TH Humic Andosols TV Vitric Andosols	YH Haplic Yermosols YK Calcic Yermosols YY Gypsic Yermosols YL Luvic Yermosols
RE Eutric Regosols RC Calcaric Regosols RD Dystic Regosols		YT Takyric Yermosols
RX Gelic Regosols	V VERTISOLS	X XEROSOLS
I LITHOSOLS	VP Pellic Vertisols VC Chromic Vertisols	XH Hapl ic Xerosols XK Calcis Xerosols XY Gypsic Xerosols XL Luvic Xerosols
K KASTANOZEMS	H PHAEOZEMS	AL LUVIC ACTOSOIS
KH Haplic Kastanozems KK Calcic Kastanozems	HH Haplic Phacozems HC Calcic Phaeozems HL Luvic Phaeozems	B CAMBISOLS
KL Luvic Kastanozems	HG Gleyic Phaeozems	BE Eutric Cambisols BD Dystric Cambisols BH Humic Cambisols
C CHERNOZEMS	M GREYZEMS	BG Gleyic Cambisols BX Gelic Cambisols
CH Haplic Chernozems CK Calcic Chernozems CL Lywio Chernozems	MO Orthic Greyzems	BK Calcic Cambisols BC Chromic Cambisols
CL Luvic Chernozems CG Glossic Chernozems	MG Gleyic Greyzems	BV Vertic Cambisols BF Ferral ic Cambisols

#### APPENDIX 1 (cont'd.)

#### L LUVISOLS

# LO Orthic Luvisols LC Chromic Luvisols LK Calcic Luvisols LV Verlic Luvisols LF Ferric Luvisols LA Albic Luvisols LP Plinthic Luvisols LG Gleyic Luvisols

#### W PLANOSOLS

WE Eutric Planosols	
WD Dystric Planosols	;
WM Mollic Planosols	
WH Humic Planosols	
WS Solodic Planosols	
WX Gelic Planosols	

#### 0 HISTOSOLS

OE	Eutric Histosols
OD	Dystric Histosols
OX	Gelic Histosols

#### D PODZOLUVISOLS

DE	Eutric Podzoluvisols
DD	Dystric Podzoluvisols
DG	Gleyic Podzoluvisols

#### P- PODZOLS

PO Orthic Podrols
PL Leptiv Podtols
PF Ferric Podzols
PH Humic Podzols
PP Placic Podzols
PG Gleyic Podzols

#### F FERRALSOLS

FO Orthic Ferralsols
FX Xanthic Ferralsols
FR Rhodic Ferralsols
FH Humic Ferralsols
FA Acric Ferralsols
FP Plinthic Ferralsole

#### A ACRISOLS

AO Orthic Acrisols AF Ferric Acrisols AH Humic Acrisols AP Plinthic Acrisols AG Gleyic Acrisols

#### N NITOSOLS

NE Eutric Nitosols ND Dystric Nitosols NH Humic Nitosols

### APPENDIX 2: LATITUDE AND LONGITUDE CONVENTION

The latitude and longitude coordinates are for the northwest corner of a one-degree square  $\infty 11$ . Cell latitudes are 0 at the equator, increase positively northward to +90, and increase negatively southward to -89. Cell longitudes are 0 at the prime meridian, increase positively eastward to +179, and increase negatively westward to -180.

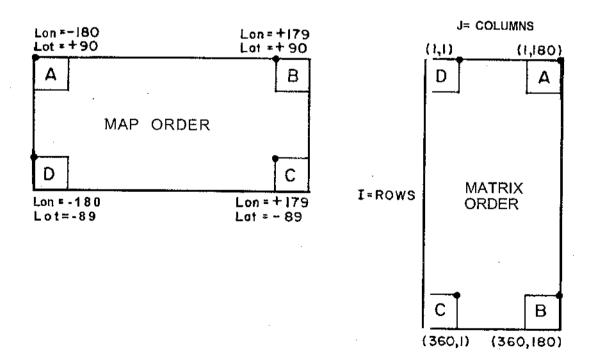
Cell latitude and longitude coordinates can be readily converted into matrix indices of I-longitude. J-latitude, for a (360,180) matrix of one-degree cells of the earth's surface, by the following FORTRAN statements:

$$I = LON + 181$$
  
 $J = LAT + 90$ 

This has been done in creating the first three files on this tape from the SOILWRLD data set.

Note that this places a longitude band of 180 cells from map order (south pole to north pole) into a row position (left to right) in the matrix. When pictured graphically (Fig. 1), this amounts to a 90 degree clockwise rotation of a map-order column of cells, to place them into a (360.180) matrix.

Fig. 1. Map and matrix order for cell latitude and longitude coordinates.



#### APPENDIX 3: PRINTER-MAP ROUTINES

Note that by using the string handling capabilities of FORTRAN77, line printer maps and other plots of the 28 major FAO soil units, water and land-ice, are easily made by printing the first character of the two-character matrix elements. For example, to printer-plot North America (see Fig. 2), delimit the region to be plotted by first recalling that;

and then by calculating the limits of the I,J indices to plot:

As seen in the routine below, plot from north at the top of the page, to south, by reverse order on the J outer loop; and from west at the left of the page, to right (don't exceed printer width), on inner loop I:

To plot the world, in four sections of 90 degrees longitude each, pole to pole, try something like this (with unit 6 set as your printer):

```
DO 40 K= 1.271.90

WRITE(6,103)

103 FORMAT('1')

DO 40 J=180,1,-1

40 PRINT *, (SU(I,I)(I:1), I*K,K+89)
```

To plot the TEX and SLP files, the string delimiter (1:1) is not needed (see Figs. 3 and 4).